Covert MIMO Communications Under Variational Distance Constraint

Shi-Yuan Wang, *Student Member, IEEE*, and Matthieu R. Bloch, *Senior Member, IEEE*

***Abstract*— The problem of covert communication over Multiple-Input Multiple-Output (MIMO) Additive White Gaussian Noise (AWGN) channels is investigated, in which a transmitter attempts to reliably communicate with a legitimate receiver while avoiding detection by a passive adversary. The covert capacity of the MIMO AWGN channel is characterized under a variational distance covertness constraint when the MIMO channel matrices are static and known. The characterization of the covert capacity is also extended to a class of channels in which the legitimate channel matrix is known but the adversary's channel matrix is only known up to a rank and a spectral norm constraint.**

***Index Terms*— Physical-layer security, covert communication, Multiple-Input Multiple-Output (MIMO)-Additive White Gaussian Noise (AWGN) channels, variational distance, compound channels.**

I. Introduction

COVERT communications, also known as communications with low probability of detection, have long been used to transmit sensitive information without raising suspicion. While technologies such as spread-spectrum communications have been widely deployed, the information-theoretic limits of covert communication had not been investigated until recently. Much of the interest has been spurred by the discovery of a *square-root law* [1], which limits the scaling with the coding blocklength n of the number of reliable and covert communication bits over memoryless channels to $\mathcal{O}(\sqrt{n})$. In other words, the standard capacity of covert communications is zero but the number of bits still grows with the blocklength. The optimal constant behind the $\mathcal{O}(\sqrt{n})$ scaling then plays the role of the *covert capacity* and has been characterized for many channels, including Discrete Memoryless Channels (DMCs) and AWGN channels, using both relative entropy [2], [3] and variational distance [4], [5] as a covertness metric. Covert communications often require secret keys as an enabling resource, the amount of which can be characterized [3]; in particular, no secret keys are required when the legitimate receiver obtains better observations than the adversary [6]. Refined characterizations of the message and key sizes for finite length [7], [8] and second-order asymptotics [4] are also known, although they are often not complete. Recent advances include the characterization of the covert capacity in network information theory problems [9]–[12], quantum channels [13], [14], low-complexity code constructions [15]–[19], and system-level considerations highlighting how to allocate resources in the presence of covertness constraints [20], [21]. Particularly relevant to the present work, there have been attempts at studying MIMO-AWGN channels when measuring covertness using relative entropy as a covertness metric [22]–[24].

Covertness must be measured in terms of a metric that captures how different the statistics of the observations are in presence and in absence of communication. Relative entropy has been a popular choice [2], [3] because of its convenient analytical properties; however, variational distance is the metric that is operationally relevant to the performance of the adversary's detector [4]. In this work, we therefore use variational distance to measure covertness, which requires specific techniques, especially in the converse proof.

The contributions of the present work are twofold. 1) We revisit the MIMO-AWGN channel model of [22]–[24] and, under the assumption that the null space of the main and adversary's channel matrices are trivial, we obtain a closed-form of the covert capacity with variational distance as the covertness metric. Our approach extends the techniques developed in [4], [5] and the crux of contribution is the converse proof. 2) We investigate the problem of covert communication over compound MIMO-AWGN channels, in particular, the situation in which the adversary's channel matrix is only known up to a rank constraint and a spectral norm constraint [22]–[25]. Our approach differs from the analysis in [22]–[25] and borrows ideas from [26] to avoid implicit constraints on the adversary's operation when dealing with uncountable compound channels.

A preliminary version of these results was presented in [27] but without complete proofs. The present work offers self-contained and detailed proofs.

II. Channel Model

A. Notation

Both log and exp should be understood in base e; hence, all information-theoretic quantities are nats. Calligraphic letters are used for sets and $|\cdot|$ denotes their cardinality. $(\cdot)^\dagger$ denotes the Moore-Penrose inverse of a matrix. $\mathbf{M} \succeq \mathbf{0}$ denotes a positive semi-definite matrix $\mathbf{M}$. $\mathbb{H}(\cdot)$, $h(\cdot)$, $\mathbb{I}(\cdot;\cdot)$, and $h_b(\cdot)$ denote the usual entropy, differential entropy, mutual information, and binary entropy function, respectively.

Manuscript received January 16, 2021; revised May 6, 2021 and July 26, 2021; accepted August 11, 2021. Date of publication September 6, 2021; date of current version September 16, 2021. This work was supported by the National Science Foundation under Award 1527387 and Award 1910859. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ashish Khisti. *(Corresponding author: Shi-Yuan Wang.)*

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: shi-yuan.wang@gatech.edu; matthieu.bloch@ece.gatech.edu).

Digital Object Identifier 10.1109/TIFS.2021.3110048

For a continuous alphabet Ω and any two distributions $P,\ Q$ with densities $f_P,\ f_Q$, respectively, the variational distance between P and Q is defined as $\mathbb{V}(P,Q) \triangleq \frac{1}{2}\int_\Omega |f_P(x) - f_Q(x)|dx$ or equivalently $\mathbb{V}(P,Q) = \sup_{\mathcal{S}\subseteq\Omega}|P(\mathcal{S}) - Q(\mathcal{S})|$. The relative entropy between P and Q is defined as $\mathbb{D}(P \parallel Q) \triangleq \int_\Omega f_P(x)\log\frac{f_P(x)}{f_Q(x)}dx$. Pinsker's inequality ensures that $\mathbb{V}(P,Q)^2 \leqslant \frac{1}{2}\min(\mathbb{D}(P\parallel Q), \mathbb{D}(Q\parallel P))$. Let $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ be jointly distributed random variables according to $P\cdot W$, where P has density f_P, and $W : (x, y) \mapsto W(y|x)$ is a transition probability from $\mathcal{X} \to \mathcal{Y}$ with density f_W. We define the marginal distribution of Y as $P \circ W$ with density $\int_\mathcal{X} f_W(y|x)f_P(x)dx$.

Moreover, for two integers $\lfloor a\rfloor$ and $\lceil b\rceil$ such that $\lfloor a\rfloor \leqslant \lceil b\rceil$, we define $[\![a,b]\!] \triangleq \{\lfloor a\rfloor, \lfloor a\rfloor+1,\dots,\lceil b\rceil-1,\lceil b\rceil\}$; otherwise $[\![a,b]\!] \triangleq \emptyset$. For any $x \in \mathbb{R}$, we also define the Q-function $Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}dx$ and its inverse function $Q^{-1}(\cdot)$.

B. System Model

We consider a MIMO-AWGN channel in which a transmitter (Alice) with N_a antennas attempts to reliably communicate with a legitimate receiver (Bob) with N_b antennas in the presence of a passive adversary (the warden Willie) equipped with N_w antennas. We assume that Bob and Willie possess more antennas than Alice, i.e., $N_a \leqslant N_b$ and $N_a \leqslant N_w$. Bob and Willie's received signals at every channel use are then

$$\mathbf{y} = \mathbf{H}_b\mathbf{x} + \mathbf{n}_b \quad \text{and } \mathbf{z} = \mathbf{H}_w\mathbf{x} + \mathbf{n}_w, \tag{1}$$

respectively, where $\mathbf{x} \in \mathbb{R}^{N_a}$ is Alice's transmitted signal and $\mathbf{H}_b$ and $\mathbf{H}_w$ are Bob's and Willie's channel matrices, assumed known to everyone. We further assume that both matrices have full rank, i.e., $m = \operatorname{rank}(\mathbf{H}_b) = \operatorname{rank}(\mathbf{H}_w) = N_a$. Hence, both channel matrices can be decomposed with a Generalized Singular Value Decomposition (GSVD) [28], [29] as

$$\begin{aligned}\mathbf{H}_b &= \mathbf{U}'_b\Sigma_b\mathbf{\Omega}^{-1}\Psi^\intercal = \mathbf{U}_b\Lambda_b\mathbf{V}^\intercal,\\ \mathbf{H}_w &= \mathbf{U}'_w\Sigma_w\mathbf{\Omega}^{-1}\Psi^\intercal = \mathbf{U}_w\Lambda_w\mathbf{V}^\intercal,\end{aligned} \tag{2}$$

where $\Psi \in \mathbb{R}^{N_a\times N_a}$, $\mathbf{U}'_b \in \mathbb{R}^{N_b\times N_b}$, and $\mathbf{U}'_w \in \mathbb{R}^{N_w\times N_w}$ are orthogonal, $\mathbf{\Omega} \in \mathbb{R}^{m\times m}$ is lower triangular and nonsingular, and $\mathbf{V}^\intercal \triangleq \mathbf{\Omega}^{-1}\Psi^\intercal$. Both $\Sigma_b \in \mathbb{R}^{N_b\times m}$ and $\Sigma_w \in \mathbb{R}^{N_w\times m}$ are diagonal with positive elements, $\{\lambda_{b,j}\}_{j=1}^m$ and $\{\lambda_{w,j}\}_{j=1}^m$, respectively. We truncate $\mathbf{U}'_b$ and $\mathbf{U}'_w$ into $\mathbf{U}_b \in \mathbb{R}^{N_b\times m}$ and $\mathbf{U}_w \in \mathbb{R}^{N_w\times m}$, and define $\Lambda_b = \operatorname{diag}\left(\{\lambda_{b,j}\}_{j=1}^m\right)$ and $\Lambda_w = \operatorname{diag}\left(\{\lambda_{w,j}\}_{j=1}^m\right)$. The noise vectors $\mathbf{n}_b \in \mathbb{R}$ and $\mathbf{n}_w \in \mathbb{R}$ are realizations of AWGN distributed according to $\mathcal{N}\left(\mathbf{0}, \sigma_b^2\mathbf{I}_{N_b}\right)$ and $\mathcal{N}\left(\mathbf{0}, \sigma_w^2\mathbf{I}_{N_w}\right)$, respectively, assumed known to everyone.

Furthermore, for $n \in \mathbb{N}^*$, we define the innocent symbol corresponding to the absence of communication as $\mathbf{x}_0 = \mathbf{0}$; the output distributions induced by the innocent symbol at Bob and Willie are denoted $P_0 \triangleq \mathcal{N}\left(\mathbf{0}, \sigma_b^2\mathbf{I}_{N_b}\right)$ and $Q_0 \triangleq \mathcal{N}\left(\mathbf{0}, \sigma_w^2\mathbf{I}_{N_w}\right)$, respectively. The associated product distributions are denoted by $P_0^{\otimes n} = \prod_{i=1}^n P_0$ and $Q_0^{\otimes n} = \prod_{i=1}^n Q_0$.

Remark 1: We assume that both $\mathbf{H}_b$ and $\mathbf{H}_w$ have a trivial null space equal to $\{\mathbf{0}\}$. If this were not the case, the presence of a null space would result in the following scenarios. If $\mathbf{H}_w$ has a non-trivial null space, Alice can overcome the square-root law by steering her beam in the corresponding directions [22]–[24]. If $\mathbf{H}_b$ has a non-trivial null space, Alice has no incentive to use the corresponding directions and would simply ignore them. We offer further discussion in Appendix A.

C. Problem Formulation

Alice transmits a uniformly-distributed message $W \in [\![1, M_n]\!]$ by encoding it into a codeword $\mathbf{X}^n = [\mathbf{X}_1\ \dots\ \mathbf{X}_n] \in \mathbb{R}^{N_a\times n}$ of blocklength n with the aid of a uniformly-distributed secret key $S \in [\![1, K_n]\!]$ shared with Bob. The resulting code is called an (n, M_n, K_n)-code $\mathcal{C}$, assumed known to everyone. Whether Alice communicates or not is controlled by $\phi \in \{0,1\}$, with $\phi = 1$ indicating the transmission. Upon observing $\mathbf{Y}^n = [\mathbf{Y}_1\ \dots\ \mathbf{Y}_n] \in \mathbb{R}^{N_b\times n}$, Bob uses his knowledge of the secret key to form a reliable estimate $\widehat{W}$ of W. Reliability is measured by the maximal average probability of error

$$P_e^{(n)} \triangleq \max_s \bar{P}_e^{(n)}(s) + \mathbb{P}\big(\widehat{\phi} = 1|\phi = 0\big), \tag{3}$$

where $\bar{P}_e^{(n)}(s) \triangleq \mathbb{P}\big(W \neq \widehat{W}|S = s, \phi = 1\big)$, and we define $\bar{P}_e^{(n)} \triangleq \mathbb{E}_S\big\{\mathbb{P}\big(W \neq \widehat{W}|S, \phi = 1\big)\big\} + \mathbb{P}\big(\widehat{\phi} = 1|\phi = 0\big)$. In contrast, Willie's objective is to detect whether Alice is transmitting based on the observations $\mathbf{Z}^n = [\mathbf{Z}_1\ \dots\ \mathbf{Z}_n] \in \mathbb{R}^{N_w\times n}$ via a hypothesis test $T(\mathbf{Z}^n)$. In particular, Willie expects $Q_0^{\otimes n}$ when there is no transmission between Alice and Bob (i.e., the null hypothesis) and $\widehat{Q}^n$ when the transmission occurs (i.e., the alternative hypothesis), where $\widehat{Q}^n$ is the output distribution induced by the code $\mathcal{C}$ used by Alice and Bob, $\forall\ \mathbf{z}^n \in \mathbb{R}^{N_w\times n}$,

$$\widehat{Q}^n\left(\mathbf{z}^n\right) = \frac{1}{M_nK_n}\sum_{\ell=1}^{M_n}\sum_{k=1}^{K_n} W_{\mathbf{Z}|\mathbf{X}}^{\otimes n}\left(\mathbf{z}^n|\mathbf{x}^{(\ell k)n}\right). \tag{4}$$

In the sequel, we use $\mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n})$ as our covertness metric. When testing the null hypothesis $Q_0^{\otimes n}$ against the alternative hypothesis $\widehat{Q}^n$, any test $T(\mathbf{Z}^n)$ conducted by Willie on the observations $\mathbf{Z}^n$ satisfies $1 \geqslant \alpha + \beta \geqslant 1 - \mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n})$, where α and β are the probabilities of false alarm and missed detection, respectively, and the lower bound can be achieved by an optimal test [30, Theorem 13.1.1]. In addition, the trade-off $\alpha + \beta = 1$ is achieved with blind tests that do not use the observations. Consequently, making $\mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n})$ vanish amounts to rendering the adversary's hypothesis test effectively blind and hence achieves covertness.

Definition 1: A reliable and covert throughput $r \in \mathbb{R}_+$ is achievable with corresponding key throughput $k \in \mathbb{R}_+$, if there exists a sequence of (n, M_n, K_n, δ)-codes with increasing blocklength n such that

$$\liminf_{n\to\infty}\frac{\log M_n}{\sqrt{nd}} \geqslant r, \quad \limsup_{n\to\infty}\frac{\log M_nK_n}{\sqrt{nd}} \leqslant r + k, \tag{5}$$

and

$$\lim_{n\to\infty}P_e^{(n)} = 0, \quad \mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n}) \leqslant \delta, \tag{6}$$

where $d = Q^{-1}\left(\frac{1-\delta}{2}\right)$. The covert capacity C_{covert} is the supremum of achievable throughputs r.

Note that, in our definition, we normalize the message and key size by $\sqrt{nd}$ instead of the usual choice, n; this is essential to unveil the square-root law behind the covertness and is justified a posteriori by the results in Section III. Intuitively, the square-root law exists, for we are hiding messages in "statistical noise", whose standard deviation behaves as $\mathcal{O}(1/\sqrt{n})$ [3, Section III.A].

Remark 2: Our use of $\mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n})$ is motivated by the following considerations. As a strong converse states that the optimal code rate has no dependency to the target constraint in the first asymptotics, it is clear that there is no strong converse for the value of δ with respect to (w.r.t.) the covert throughput, i.e., the covert throughput depends on the value δ through $d = Q^{-1}\left(\frac{1-\delta}{2}\right)$, which is directly related to $\mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n})$. This can be seen from Definition 1 and Theorem 1 where the notion of throughput depends on the covertness metric; hence, the choice of covertness metric matters. Many earlier works [2], [3], [22]–[24] measure covertness using the relative entropy $\mathbb{D}(\widehat{Q}^n \parallel Q_0^{\otimes n})$. Unfortunately, relative entropy is only a loose proxy for variational distance since Pinsker's inequality is not tight [4] and is then less directly related to the operational test of the adversary. Furthermore, both $\mathbb{D}(\widehat{Q}^n \parallel Q_0^{\otimes n})$ and $\mathbb{D}(Q_0^{\otimes n} \parallel \widehat{Q}^n)$ could in principle be used but, depending on which metric is chosen, different conclusions regarding the optimal signaling over AWGN channels can be reached [31].

Remark 3: Our model does not include a power constraint on the channel input. This is justified since we only consider channel matrices with trivial null space and since any power constraint on the input is weaker than the covertness constraint [2, Section V.]. Previous works [22]–[24] impose the power constraint precisely because they allow non-trivial null spaces.

III. Main Results

Theorem 1: The covert capacity of a MIMO-AWGN channel with full knowledge of the channel matrices is

$$C_{covert} = \frac{\sigma_w^2}{\sigma_b^2}\sqrt{2\text{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}. \quad (7)$$

The covert capacity is achievable with key throughput

$$R_{key} = \sqrt{\frac{2}{\text{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}} \times\left(\text{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right) - \frac{\sigma_w^2}{\sigma_b^2}\text{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)\right)^+, \quad (8)$$

where $(x)^+ \triangleq \max(x, 0)$.

A. Converse Proof for Variational Distance

Proposition 1: Consider a sequence of covert MIMO-AWGN communication schemes for the model in (1) with increasing blocklength $n \in \mathbb{N}^*$, characterized by $\epsilon_n \triangleq P_e^{(n)}$ and $\delta \geqslant \mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n})$. If $\lim_{n\to\infty}\epsilon_n = 0$ and $\lim\limits_{n\to\infty} M_n = \infty$, then we have

$$\liminf_{n\to\infty}\frac{\log M_n}{\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)} \leqslant \frac{\sigma_w^2}{\sigma_b^2}\sqrt{2\text{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}. \quad (9)$$

Proof: The proof extends the techniques developed in [1], [4], [5] by constructing a test for Willie that is simple enough to be analyzed yet powerful enough to obtain a tight bound. Since we do not have any knowledge of the specific code exploited in the converse proof, the crux of our converse is to show that there cannot be too many high-power codewords, for otherwise the covertness would be compromised. We then analyze the maximal size of the low-power subcode, which is "good" in the sense that both reliability and covertness can be ensured. We first recall the Berry-Esseen Theorem.

Theorem 2 (Berry-Esseen Theorem): Let $X_1, \ldots, X_n$ be independent random variables such that for $k \in [\![1, n]\!]$, we have $\mathbb{E}\{X_k\} = \mu_k$, $\sigma_k^2 = \text{Var}(X_k)$, and $t_k = \mathbb{E}\left\{|X_k - \mu_k|^3\right\}$. If we define $\sigma^2 = \sum_{k=1}^n \sigma_k^2$ and $T = \sum_{k=1}^n t_k$, then we have

$$\left|\mathbb{P}\left(\sum_{k=1}^n (X_k - \mu_k) \geqslant \lambda\sigma\right) - Q(\lambda)\right| \leqslant \frac{6T}{\sigma^3}. \quad (10)$$

1) Lower Bound on Covertness Metric: We start by establishing a lower bound relating the covertness metric to the minimum received power of codewords at Bob within a given code $\mathcal{M}$. Consider a simple hypothesis testing problem with two hypotheses H_0 and H_1 corresponding to distributions $Q_0^{\otimes n}$ and $\widehat{Q}^n$, respectively. We define a sub-optimal power detector

$$T\left(\mathbf{z}^n\right) \triangleq \mathbb{1}\left\{\sum_{i=1}^n S_i > \tau\right\}, \quad (11)$$

where $S_i \triangleq S(\mathbf{z}_i) \triangleq \|\mathbf{H}_b(\mathbf{H}_w)^\dagger \mathbf{z}_i\|_2^2$, and the threshold τ will be specified later. The intuition behind the test is to realign Willie's observations with those of Bob. Note that, $\mathbf{H}_w^\mathsf{T}\mathbf{H}_w$ is invertible because of the full-rank assumption. Hence, we rewrite the test S_i using the GSVD as

$$S_i = \left(\mathbf{H}_b(\mathbf{H}_w)^\dagger \mathbf{z}_i\right)^\mathsf{T}\left(\mathbf{H}_b(\mathbf{H}_w)^\dagger \mathbf{z}_i\right) = \hat{\mathbf{z}}_i^\mathsf{T}\hat{\mathbf{z}}_i, \quad (12)$$

where $\hat{\mathbf{z}}_i = \Lambda_b\Lambda_w^{-1}\mathbf{U}_w^\mathsf{T}\mathbf{z}_i$. The following lemma, which is proved in Appendix B, characterizes upper bounds for both the false-alarm and the missed-detection probabilities.

Lemma 1: Consider a specific code $\mathcal{M}$ with codewords indexed by k, $\mathbf{x}^{(k)n} = \left[\mathbf{x}_1^{(k)} \ \ldots \ \mathbf{x}_n^{(k)}\right] \in \mathcal{C}$. By defining $P_* \triangleq \min_k \left\|\mathbf{H}_b\mathbf{x}^{(k)n}\right\|_F^2 = \min_k \text{tr}\left(\Lambda_b^2\mathbf{P}^{(k)}\right)$ the minimum power of Bob's received codewords, and setting the detection threshold to $\tau = \frac{P_*}{2} + n\sigma_w^2\text{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)$,

$$\alpha \leqslant Q\left(\frac{P_*}{2\sqrt{2n\text{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}\sigma_w^2}\right) + \frac{B_0}{\sqrt{n}}, \quad (13)$$

$$\beta \leqslant Q\left(\frac{P_*}{2\sqrt{2n\text{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}\sigma_w^2}\right)$$

$$+\frac{P_*^2\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)}{4\sqrt{\pi}n^{3/2}\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)^{3/2}\sigma_w^4}+\frac{B_1}{\sqrt{n}}, \tag{14}$$

where B_0 and B_1 are some constants independent of n.

Hence, the covertness metric can be lower-bounded as

$$\begin{aligned}\mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n}) &\geqslant 1-\alpha-\beta\\ &\geqslant 1-2Q\left(\frac{P_*}{2\sqrt{2n\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}\sigma_w^2}\right)\\ &\quad-\frac{P_*^2\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)}{4\sqrt{\pi}n^{3/2}\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)^{3/2}\sigma_w^4}\\ &\quad-\frac{B_0+B_1}{\sqrt{n}},\end{aligned} \tag{15}$$

which only depends on the code through the minimum power of Bob's received codewords.

2) Existence of a Good Sub-Code: For a covert code $\mathcal{C}$, we develop a bound for the maximum power of a non-empty low-power sub-code in the following lemma, which is proved in Appendix C. The key idea is to use (15) to analyze the covertness for the high-power sub-code and argue the existence of a low-power sub-code.

Lemma 2: For any covert channel code $\mathcal{C}$, given a decreasing sequence $\{\gamma_n\}_{n=1}^{\infty}$ with $\gamma_n \in (0,1)$, $\lim_{n\to\infty}\gamma_n = 0$, there exists a subset of codewords $\mathcal{C}^{(\ell)}$ such that $\left|\mathcal{C}^{(\ell)}\right| \geqslant \gamma_n|\mathcal{C}|$ and $\|\mathbf{H}_b\mathbf{x}^n\|_F^2 \leqslant A\sqrt{n}$, where

$$\begin{aligned}A &\triangleq 2\sqrt{2\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}\sigma_w^2\\ &\quad\times Q^{-1}\left(\frac{1-\delta}{2}-\frac{\nu^2\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)}{4\sqrt{\pi}n\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)^{3/2}\sigma_w^4}-\gamma_n\right),\end{aligned} \tag{16}$$

and ν depends on the channel.

3) Upper Bound on Covert Message Size Within a Good Sub-Code: The code $\mathcal{C}$ can be partitioned into K_n sub-codes $\mathcal{C}_s$ indexed by the key value s for all $s \in [\![1, K_n]\!]$ such that $\mathcal{C} = \cup_{s\in[\![1,K_n]\!]}\mathcal{C}_s$, and the size of each sub-code is M_n. Let $\mathcal{C}_s^{(\ell)} \triangleq \mathcal{C}_s \cap \mathcal{C}^{(\ell)}$. By the pigeonhole principle, there exists a sub-code $\mathcal{C}_s$ satisfying $\left|\mathcal{C}_s^{(\ell)}\right| \geqslant \gamma_n M_n$. Furthermore, since the average probability of error of $\mathcal{C}_s$ is at most ϵ_n, we have $\bar{P}_e^{(n)}\left(\mathcal{C}_s^{(\ell)}\right) \leqslant \frac{\epsilon_n}{\gamma_n}$, which vanishes in the limit of large n upon choosing $\{\gamma_n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty}\frac{\epsilon_n}{\gamma_n} = 0$.

Let $\widetilde{W}$ denote the uniformly distributed variable over the messages in $\mathcal{C}_s^{(\ell)}$. By standard techniques, we therefore have

$$\log\left|\mathcal{C}_s^{(\ell)}\right| = \mathbb{H}\big(\widetilde{W}|S=s\big) \tag{17}$$

$$= \mathbb{I}\big(\widetilde{W};\mathbf{Y}^n|S=s\big)+\mathbb{H}\big(\widetilde{W}|\mathbf{Y}^nS=s\big) \tag{18}$$

$$\leqslant \mathbb{I}\big(\widetilde{W};\mathbf{Y}^n|S=s\big)+\left[\frac{\epsilon_n}{\gamma_n}\log\left|\mathcal{C}_s^{(\ell)}\right|+h_b\left(\frac{\epsilon_n}{\gamma_n}\right)\right] \tag{19}$$

$$\leqslant \mathbb{I}\big(\mathbf{X}^n;\mathbf{Y}^n|S=s\big)+\left[\frac{\epsilon_n}{\gamma_n}\log\left|\mathcal{C}_s^{(\ell)}\right|+h_b\left(\frac{\epsilon_n}{\gamma_n}\right)\right] \tag{20}$$

$$\leqslant n\mathbb{I}\big(\bar{\mathbf{X}};\bar{\mathbf{Y}}\big)+\frac{\epsilon_n}{\gamma_n}\log\left|\mathcal{C}_s^{(\ell)}\right|+1, \tag{21}$$

where the random variables $\bar{\mathbf{X}}$ and $\bar{\mathbf{Y}}$ have distributions

$$\Pi_{\bar{\mathbf{X}}}(\mathbf{x}) \triangleq \frac{1}{n}\sum_{i=1}^{n}\Pi_{\mathbf{X}_i}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\left|\mathcal{C}_s^{(\ell)}\right|}\sum_{\mathbf{x}^n\in\mathcal{C}_s^{(\ell)}}\mathbb{1}\{\mathbf{x}=\mathbf{x}_i\}$$

and $P_{\bar{\mathbf{X}}\bar{\mathbf{Y}}} \triangleq \Pi_{\bar{\mathbf{X}}}W_{\mathbf{Y}|\mathbf{X}}$.

Let $\mathbb{E}\big\{\bar{\mathbf{X}}\bar{\mathbf{X}}^{\intercal}\big\} = \mathbf{Q}_n$. Note that $\mathbb{E}\big\{\bar{\mathbf{Y}}\bar{\mathbf{Y}}^{\intercal}\big\} = \mathbf{H}_b\mathbf{Q}_n\mathbf{H}_b^{\intercal}+\sigma_b^2\mathbf{I}_{N_b}$. Then,

$$\mathbb{I}\big(\bar{\mathbf{X}};\bar{\mathbf{Y}}\big) = h\left(\bar{\mathbf{Y}}\right)-h\left(\bar{\mathbf{Y}}|\bar{\mathbf{X}}\right) \tag{22}$$

$$\leqslant \frac{1}{2}\log\left|\mathbf{I}_{N_b}+\frac{1}{\sigma_b^2}\mathbf{H}_b\mathbf{Q}_n\mathbf{H}_b^{\intercal}\right| \tag{23}$$

$$= \frac{1}{2}\mathrm{tr}\left(\log\left(\mathbf{I}_{N_b}+\frac{1}{\sigma_b^2}\mathbf{H}_b\mathbf{Q}_n\mathbf{H}_b^{\intercal}\right)\right) \tag{24}$$

$$\overset{(a)}{\leqslant} \frac{1}{2\sigma_b^2}\mathrm{tr}\left(\mathbf{H}_b\mathbf{Q}_n\mathbf{H}_b^{\intercal}\right) \overset{(b)}{\leqslant} \frac{A}{2\sigma_b^2\sqrt{n}}, \tag{25}$$

where (a) follows since for any $\mathbf{A} \succcurlyeq \mathbf{0}$ and $\|\mathbf{A}\|_2 \leqslant 1$, $\mathrm{tr}\left(\log\left(\mathbf{I}+\mathbf{A}\right)\right) = \sum_i\log\left(1+\lambda_i(\mathbf{A})\right) \leqslant \sum_i\lambda_i(\mathbf{A}) = \mathrm{tr}\left(\mathbf{A}\right)$, where $\{\lambda_i(\mathbf{A})\}_i$ is the set of eigenvalues of $\mathbf{A}$ and we have used $\log(1+x) \leqslant x$ for all $x > 0$, and (b) follows from the definition of $\mathcal{C}^{(\ell)}$ and $\mathrm{tr}\left(\mathbf{H}_b\mathbf{Q}_n\mathbf{H}_b^{\intercal}\right) = \frac{1}{n\left|\mathcal{C}_s^{(\ell)}\right|}\sum_{\mathbf{x}^n\in\mathcal{C}_s^{(\ell)}}\|\mathbf{H}_b\mathbf{x}^n\|_F^2$. Combining (16), (21), (25), and the fact that $\lim_{n\to\infty}\gamma_n = 0$, we have

$$\log\left|\mathcal{C}_s^{(\ell)}\right| \leqslant \frac{\sqrt{2n\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}\frac{\sigma_w^2}{\sigma_b^2}Q^{-1}\left(\frac{1-\delta}{2}\right)+\mathcal{O}(1)}{1-\frac{\epsilon_n}{\gamma_n}}. \tag{26}$$

We further choose the sequence $\{\gamma_n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty}-\frac{\log\gamma_n}{\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)} = 0$. Finally, we obtain

$$\liminf_{n\to\infty}\frac{\log M_n}{\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)} \leqslant \liminf_{n\to\infty}\frac{\log\left|\mathcal{C}_s^{(\ell)}\right|-\log\gamma_n}{\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)} \tag{27}$$

$$= \frac{\sigma_w^2}{\sigma_b^2}\sqrt{2\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}. \tag{28}$$

■

Unfortunately, we have not found a matching converse argument for the key throughput.

B. Achievability Proof for Variational Distance

Proposition 2: Consider a MIMO-AWGN covert communication channel in (1). There exist covert communication *schemes such that*

$$\lim_{n\to\infty} \frac{\log M_n}{\sqrt{n} Q^{-1}\left(\frac{1-\delta}{2}\right)} \geqslant \frac{\sigma_w^2}{\sigma_b^2}\sqrt{2\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}, \tag{29}$$

$$\lim_{n\to\infty} \frac{\log M_n K_n}{\sqrt{n} Q^{-1}\left(\frac{1-\delta}{2}\right)} \leqslant \sqrt{\frac{2}{\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}}\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right), \tag{30}$$

$$\lim_{n\to\infty} P_e^{(n)} = 0, \quad \mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n}) \leqslant \delta. \tag{31}$$

Proof: Our proof follows [2], [3], [5] to construct a Binary Phase-Shift Keying (BPSK) code achieving the desired throughput pair. Note that we could also use a Gaussian codebook, but this would require extra care to deal with the power of codewords.

1) Covert Stochastic Process [3]: We introduce another input process $\Pi_{\mathbf{Q}_n}$ with covariance matrix $\mathbf{Q}_n$ and its associated distribution at the output of channel $W_{\mathbf{Z}|\mathbf{X}}$, $Q_n \triangleq \Pi_{\mathbf{Q}_n} \circ W_{\mathbf{Z}|\mathbf{X}}$. Additionally, the associated product distributions are $\Pi_{\mathbf{Q}_n}^{\otimes n} = \prod_{i=1}^n \Pi_{\mathbf{Q}_n}$ and $Q_n^{\otimes n} = \prod_{i=1}^n Q_n$. The achievability proof decomposes the covertness metric $\mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n})$ into two pieces, $\mathbb{V}(\widehat{Q}^n, Q_n^{\otimes n})$ and $\mathbb{V}(Q_n^{\otimes n}, Q_0^{\otimes n})$ by the triangle inequality. The former term is related to the channel output approximation problem, and we rely on the channel resolvability to analyze its behavior [3], [32]. We upper-bound the latter term by a covertness constraint $\delta - \frac{1}{\sqrt{n}}$. Essentially, this constraint makes $Q_n^{\otimes n}$ asymptotically indistinguishable from the output distribution of the innocent symbol $Q_0^{\otimes n}$; accordingly, $Q_n^{\otimes n}$ is called a *covert stochastic process*. The rationale for introducing such a process is to find a proxy to control the discrepancy captured by the covertness metrics by a carefully designed covariance matrix $\mathbf{Q}_n$, which is the counterpart of low-weight codewords designed in the covert communication scheme over DMCs [3], [4].

2) Random Code Generation: We decompose the channel into m parallel sub-channels defined by the GSVD precoding with the input alphabet $\widetilde{\mathcal{X}} \triangleq \{-a_{n,1}, 0, a_{n,1}\} \times \cdots \times \{-a_{n,m}, 0, a_{n,m}\}$, where $m = \mathrm{rank}\left(\mathbf{H}_b\right) = \mathrm{rank}\left(\mathbf{H}_w\right) = N_a$. Throughout the section, tildes refer to the operations over the parallel sub-channels. Let $M_n, K_n \in \mathbb{N}^*$. Alice independently generates $M_n K_n$ codewords $\tilde{\mathbf{x}}^n\left(\ell, k\right) \in \prod_{j=1}^m \{-a_{n,j}, a_{n,j}\}^n$ jointly over all the sub-channels with $\ell \in [\![1, M_n]\!]$ and $k \in [\![1, K_n]\!]$, according to the distribution $\prod_{j=1}^m \Pi_{\rho_{n,j}}$ such that $\Pi_{\rho_{n,j}}\left(a_{n,j}\right) = \Pi_{\rho_{n,j}}\left(-a_{n,j}\right) = \frac{1}{2}$, and $\Pi_{\rho_{n,j}}\left(0\right) = 0$, where $\{\rho_{n,j}\}$ is a set of non-negative real numbers defined as

$$\rho_{n,j} \triangleq \frac{\tau_j Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{n}} = a_{n,j}^2, \quad \forall j \in [\![1, m]\!], \tag{32}$$

and $\{\tau_j\}_{j=1}^m$ is determined later via an optimization program. We define two diagonal matrices $\mathbf{P}_n$ and $\mathbf{T}$ with $\{\rho_{n,j}\}_{j=1}^m$ and $\{\tau_j\}_{j=1}^m$ as the diagonal entries, respectively. For simplicity, we stack codewords into $\tilde{\mathbf{x}}^n\left(\ell, k\right) \in \mathbb{R}^{m\times n}$. Alice then employs the precoding matrix $(\mathbf{V}^{\intercal})^{-1}$ to form $\mathbf{x}^n = (\mathbf{V}^{\intercal})^{-1}\tilde{\mathbf{x}}^n$, and therefore the input covariance matrix after the precoding is $\mathbf{Q}_n = (\mathbf{V}^{\intercal})^{-1}\mathbf{P}_n\mathbf{V}^{-1}$, where we design $\mathbf{P}_n$ carefully as in (32). Bob and Willie postprocess their observations from channel outputs $\mathbf{y}^n \in \mathbb{R}^{N_b\times n}$ and $\mathbf{z}^n \in \mathbb{R}^{N_w\times n}$ by transforming them via $\mathbf{U}_b^{\intercal}$ and $\mathbf{U}_w^{\intercal}$ to get $\tilde{\mathbf{y}}^n \in \mathbb{R}^{m\times n}$ and $\tilde{\mathbf{z}}^n \in \mathbb{R}^{m\times n}$, respectively. There is no loss of generality in making this assumption for Willie, as the post-processing $\mathbf{U}_w^{\intercal}$ performs an orthogonal transform and then discards the components in the observations corresponding to Null($\mathbf{H}_w^{\intercal}$), which only contain noise. These operations result in, $\forall i \in [\![1, n]\!]$, $\tilde{\mathbf{y}}_i = \Lambda_b\tilde{\mathbf{x}}_i + \tilde{\mathbf{n}}_{b,i}$, and $\tilde{\mathbf{z}}_i = \Lambda_w\tilde{\mathbf{x}}_i + \tilde{\mathbf{n}}_{w,i}$, where $\tilde{\mathbf{n}}_b \sim \mathcal{N}\left(\mathbf{0}, \sigma_b^2\mathbf{I}_m\right)$ and $\tilde{\mathbf{n}}_w \sim \mathcal{N}\left(\mathbf{0}, \sigma_w^2\mathbf{I}_m\right)$. From the perspective of the m sub-channels $\left(\widetilde{\mathcal{X}}, W_{\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}}, \widetilde{\mathcal{Y}}, W_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}, \widetilde{\mathcal{Z}}\right)$ with $\widetilde{\mathcal{Y}} = \mathbb{R}^m$, and $\widetilde{\mathcal{Z}} = \mathbb{R}^m$, we therefore define the following statistics: $\Pi_{\mathbf{P}_n}\left(\tilde{\mathbf{x}}\right) = \prod_{j=1}^m \Pi_{\rho_{n,j}}\left(\tilde{x}_j\right)$, $\Pi_{\mathbf{P}_n}^{\otimes n} = \prod_{i=1}^n \Pi_{\mathbf{P}_n}$, $\widetilde{P}_n = \Pi_{\mathbf{P}_n} \circ W_{\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}}$, $\widetilde{P}_n^{\otimes n} = \prod_{i=1}^n \widetilde{P}_n$, $\widetilde{Q}_n = \Pi_{\mathbf{P}_n} \circ W_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}$, and $\widetilde{Q}_n^{\otimes n} = \prod_{i=1}^n \widetilde{Q}_n$. Note that because of the parallelness of sub-channels, we can derive simple forms as follows:

$$\widetilde{P}_n = \prod_{j=1}^m \left(\frac{1}{2}\mathcal{N}\left(-\lambda_{b,j}a_{n,j}, \sigma_b^2\right) + \frac{1}{2}\mathcal{N}\left(\lambda_{b,j}a_{n,j}, \sigma_b^2\right)\right),$$

$$\widetilde{Q}_n = \prod_{j=1}^m \left(\frac{1}{2}\mathcal{N}\left(-\lambda_{w,j}a_{n,j}, \sigma_w^2\right) + \frac{1}{2}\mathcal{N}\left(\lambda_{w,j}a_{n,j}, \sigma_w^2\right)\right).$$

In the sequel, we also use the notation $\widetilde{P}_{n,j} \triangleq \frac{1}{2}\mathcal{N}\left(-\lambda_{b,j}a_{n,j}, \sigma_b^2\right) + \frac{1}{2}\mathcal{N}\left(\lambda_{b,j}a_{n,j}, \sigma_b^2\right)$ and $\widetilde{Q}_{n,j} \triangleq \frac{1}{2}\mathcal{N}\left(-\lambda_{w,j}a_{n,j}, \sigma_w^2\right) + \frac{1}{2}\mathcal{N}\left(\lambda_{w,j}a_{n,j}, \sigma_w^2\right)$ to represent distributions at each sub-channel in the above decomposition. Similarly, we also use $\widetilde{P}_{0,j} \triangleq \mathcal{N}\left(0, \sigma_b^2\right)$ and $\widetilde{Q}_{0,j} \triangleq \mathcal{N}\left(0, \sigma_w^2\right)$. Hence, $\widetilde{P}_0 = \prod_{j=1}^m \widetilde{P}_{0,j}$ and $\widetilde{Q}_0 = \prod_{j=1}^m \widetilde{Q}_{0,j}$.

3) Channel Reliability Analysis:

Lemma 3: By choosing

$$\log M_n = (1-\xi)\frac{\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)}{2\sigma_b^2}\mathrm{tr}\left(\Lambda_b^2\mathbf{T}\right), \tag{33}$$

the average probability of error satisfies

$$\mathbb{E}\left\{\bar{P}_e^{(n)}\right\} \leqslant e^{-\theta_1\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)}, \tag{34}$$

where $\xi \in (0, 1)$, and $\theta_1 > 0$.

The proof is provided in Appendix D.

4) Covertness Analysis:

Lemma 4: By choosing $\mathbf{T}$ such that

$$\frac{1}{4\sigma_w^4}\mathrm{tr}\left(\Lambda_w^4\mathbf{T}^2\right) \leqslant 2 - \frac{C}{\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)}, \tag{35}$$

for some $C > 0$, and

$$\log M_n K_n = (1+\xi)\frac{\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)}{2\sigma_w^2}\mathrm{tr}\left(\Lambda_w^2\mathbf{T}\right), \tag{36}$$

the expected covertness metric is bounded as follows:

$$\mathbb{E}\{\mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n})\} \leqslant \delta + e^{-\theta_2\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)} - \frac{1}{\sqrt{n}}, \tag{37}$$

where $\xi \in (0, 1)$, and $\theta_2 > 0$ are some constants.

The proof is provided in Appendix E.

5) Identification of a Specific Code: Choosing ξ, $\log M_n$ and $\log K_n$ to satisfy Lemma 8 and Lemma 9, Markov's inequality allows us to conclude that there exists at least one specific code $\mathcal{C}$ with n large enough and appropriate constants $\xi_1, \xi_2 > 0$ such that $\bar{P}_e^{(n)} \leqslant e^{-\xi_1 \sqrt{n} Q^{-1}\left(\frac{1-\delta}{2}\right)}$ and $\mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n}) \leqslant \delta - \frac{1}{\sqrt{n}} + e^{-\xi_2 \sqrt{n} Q^{-1}\left(\frac{1-\delta}{2}\right)}$. Although a code $\mathcal{C}$ with vanishing $\bar{P}_e^{(n)}$ does not necessary satisfy the reliability constraint (3), which requires $P_e^{(n)}$ to vanish as n goes to infinity, the following lemma from [5] gives us such a guarantee by merely rearranging the codewords in $\mathcal{C}$.

Lemma 5: Suppose a code $\mathcal{C}$ contains K_n sub-codes of size M_n such that $\bar{P}_e^{(n)} \leqslant \epsilon_n$ and $\mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n}) \leqslant e^{-\xi_2 \sqrt{n} Q^{-1}\left(\frac{1-\delta}{2}\right)} + \delta - \frac{1}{\sqrt{n}}$. Then, there exists a code $\mathcal{C}'$ containing K_n' sub-codes of size M_n' such that $P_e^{(n)} \leqslant \epsilon_n'$ and $\mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n}) \leqslant e^{-\xi_2 \sqrt{n} Q^{-1}\left(\frac{1-\delta}{2}\right)} + \delta - \frac{1}{\sqrt{n}}$. In particular, $\lim_{n\to\infty} \epsilon_n = \lim_{n\to\infty} \epsilon_n' = 0$, $\lim_{n\to\infty} \frac{M_n'}{M_n} = 1$, and $\lim_{n\to\infty} \frac{K_n'}{K_n} = 1$.

6) Constellation Power Design: Next, we design the optimal constellation points that result in the largest achievable message set size satisfying the covertness constraint. We formalize our optimization program by combining (33) and (35) as follows:

$$\max_{\mathbf{T} \succcurlyeq \mathbf{0}} \frac{1}{2\sigma_b^2} \operatorname{tr}\left(\Lambda_b^2 \mathbf{T}\right), \tag{38a}$$

$$\text{s.t. } \frac{1}{4\sigma_w^4} \operatorname{tr}\left(\Lambda_w^4 \mathbf{T}^2\right) \leqslant 2 - \frac{C}{\sqrt{n} Q^{-1}\left(\frac{1-\delta}{2}\right)}. \tag{38b}$$

To solve this, we regard the term $\frac{C}{\sqrt{n} Q^{-1}\left(\frac{1-\delta}{2}\right)}$ as a perturbation. Consider the optimization

$$\max_{\mathbf{T} \succcurlyeq \mathbf{0}} \frac{1}{2\sigma_b^2} \operatorname{tr}\left(\Lambda_b^2 \mathbf{T}\right), \tag{39a}$$

$$\text{s.t. } \frac{1}{4\sigma_w^4} \operatorname{tr}\left(\Lambda_w^4 \mathbf{T}^2\right) \leqslant 2. \tag{39b}$$

The optimal Lagrange multiplier μ and solution $\mathbf{T}$ to (39) are $\mu = \frac{\sigma_w^2}{2\sqrt{2}\sigma_b^2} \sqrt{\operatorname{tr}\left(\Lambda_b^4 \left(\Lambda_w^{-1}\right)^4\right)}$, and $\mathbf{T} = 2\sqrt{2}\sigma_w^2 \frac{\Lambda_b^2 (\Lambda_w^{-1})^4}{\sqrt{\operatorname{tr}\left(\Lambda_b^4 \left(\Lambda_w^{-1}\right)^4\right)}}$, respectively. Let ρ and ρ' denote the optimal objective values of (39) and (38), respectively. By the sensitivity analysis [33, Ch. 8.5], we have

$$\rho \geqslant \rho' \geqslant \rho - \mathcal{O}\left(\frac{1}{\sqrt{n} Q^{-1}\left(\frac{1-\delta}{2}\right)}\right), \tag{40}$$

which shows the perturbation is negligible as n goes to infinity. Consequently,

$$\lim_{n\to\infty} \frac{\log M_n}{\sqrt{n} Q^{-1}\left(\frac{1-\delta}{2}\right)} = (1-\xi) \frac{\sigma_w^2}{\sigma_b^2} \sqrt{2 \operatorname{tr}\left(\Lambda_b^4 \left(\Lambda_w^{-1}\right)^4\right)}, \tag{41}$$

$$\lim_{n\to\infty} \frac{\log M_n K_n}{\sqrt{n} Q^{-1}\left(\frac{1-\delta}{2}\right)} = (1+\xi) \frac{\sqrt{2} \operatorname{tr}\left(\Lambda_b^2 \left(\Lambda_w^{-1}\right)^2\right)}{\sqrt{\operatorname{tr}\left(\Lambda_b^4 \left(\Lambda_w^{-1}\right)^4\right)}}. \tag{42}$$

Since for any $\xi \in (0, 1)$, there exists a scheme satisfying all the requirements, we can therefore make ξ arbitrarily small, i.e., $\xi \to 0^+$. The result then follows.

The reader might wonder why the optimal solution to (39) is not the usual water-filling solution. This is a unique phenomenon due to the covertness constraint. The usual water-filling solution would encourage the use of high power in the sub-channels in which Bob has better observations than Willie. In contrast, the square-root law discourages the use of high power, as allocating too much power to sub-channels that have better observations increases the risk of detection. Hence, one should not expect the water-filling solution to appear here. Specifically, our power allocation uses all the sub-channels and suggests that each sub-channel j contribute $(1+\xi) \frac{\sqrt{2}\lambda_{b,j}^2 \lambda_{w,j}^{-2}}{\sqrt{\operatorname{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}} \sqrt{n} Q^{-1}\left(\frac{1-\delta}{2}\right)$ to $\log M_n K_n$, which is aligned with the direction of the diagonal elements of $\left(\Lambda_b^2 \left(\Lambda_w^{-1}\right)^2\right)$. ■

IV. Covert Communication With Unknown Warden Channel State

We now assume that only partial channel state information of Willie's channel is available. Specifically, all parties know the exact $\mathbf{H}_b$, while Alice only knows that $\mathbf{H}_w$ belongs to the following uncertainty set:

$$\mathcal{S} \triangleq \{\mathbf{H}_w = \mathbf{U}_w \Lambda_w \mathbf{V}^\intercal : \|\Lambda_w\|_2 \leqslant \lambda_0, \\ \times m = \operatorname{rank}(\mathbf{H}_w) = \operatorname{rank}(\mathbf{H}_b) = N_a\}, \tag{43}$$

where $\mathbf{U}_w$ is known to Willie and hence can be canceled by post-processing [25]. Thus, the set $\mathcal{S}$ contains all the channels that are fully aligned with the main channel and for which the singular-value matrix is less than or equal to $\Lambda_0 \triangleq \lambda_0 \mathbf{I}_m$. The channel realization is fixed during the transmission period. This model corresponds to a quasi-static scenario where the adversary cannot be closer to the transmitter than a certain protection distance [25].

For an (n, M_n, K_n, δ)-code $\mathcal{C}$ designed for the compound channel induced by $\mathcal{S}$, the covertness metric at Willie is $\sup_{\mathbf{H}_w \in \mathcal{S}} \mathbb{V}(\widehat{Q}_{\mathbf{H}_w}^n, Q_0^{\otimes n})$, where $\widehat{Q}_{\mathbf{H}_w}^n$ is the distribution when communication occurs over the channel realization $\mathbf{H}_w$,

$$\widehat{Q}_{\mathbf{H}_w}^n\left(\mathbf{z}^n\right) = \frac{1}{M_n K_n} \sum_{\ell=1}^{M_n} \sum_{k=1}^{K_n} W_{\mathbf{Z}|\mathbf{X}}^{\otimes n}\left(\mathbf{z}^n | \mathbf{x}^{(\ell k) n}\right), \tag{44}$$

$\forall \mathbf{z}^n \in \mathbb{R}^{N_w \times n}$, and $W_{\mathbf{Z}|\mathbf{X}=\mathbf{x}} \sim \mathcal{N}\left(\mathbf{H}_w \mathbf{x}, \sigma_w^2 \mathbf{I}_{N_w}\right)$.

We show that the compound covert capacity is equal to the worst-case covert capacity at channel realization $\mathbf{U}_w \Lambda_0 \mathbf{V}^\intercal$. Here we only present the achievability proof for the compound covert capacity under the variational distance, in which we show that there exists a compound covert code achieving the worst-case covert capacity. The converse proof follows from the fact that the worst-case covert capacity within the uncertainty set $\mathcal{S}$ upper-bounds the compound covert capacity, for a compound covert code also works on the worst-case channel realization by definition as pointed out in [25, Corollary 1].

Proposition 3: Consider a compound MIMO-AWGN covert communication channel in (1) *and the uncertainty set $\mathcal{S}$ in* (43) *containing all possible channel realizations of the warden. There exist covert communication schemes such that*

$$\lim_{n\to\infty}\frac{\log M_n}{\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)} \geqslant \frac{\sigma_w^2}{\sigma_b^2}\sqrt{2\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_0^{-1}\right)^4\right)}, \tag{45}$$

$$\lim_{n\to\infty}\frac{\log M_n K_n}{\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)} \leqslant \sqrt{\frac{2}{\mathrm{tr}\left(\Lambda_b^4\right)}}\mathrm{tr}\left(\Lambda_b^2\right), \tag{46}$$

$$\lim_{n\to\infty}P_e = 0, \quad \sup_{\mathbf{H}_w\in\mathcal{S}}\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q_0^{\otimes n}) \leqslant \delta. \tag{47}$$

Note that (46) does not depend on Willie's channel. As mentioned previously, the power allocation makes each subchannel j contribute $(1+\xi)\frac{\sqrt{2}\lambda_{b,j}^2\lambda_0^{-2}}{\sqrt{\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_0^{-1}\right)^4\right)}}\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)$ to $\log M_n K_n$, which is aligned with the direction of the diagonal elements of $\left(\Lambda_b^2\left(\Lambda_0^{-1}\right)^2\right)$. Since we show that the worst-case channel capacity at Λ_0 is achievable, the fact that Λ_0 is isotropic makes the result independent of Willie's channel.

Proof: We extends the proof in [25], using ideas from [26]. The idea of the proof in [25] is to extend the result of compound secrecy capacity for uncountably infinite compound DMCs to continuous alphabets through a sequence of successively finer quantizers, which quantize the input and output alphabets at all parties. The compound secrecy rate derived from quantized alphabets can be made arbitrarily close to the compound secrecy capacity with a sufficiently fine quantizer. Unfortunately, this process requires the adversary to obey the quantization rule and implicitly assumes that the adversary should cooperate with Alice and Bob. We propose a small correction that circumvents the issue by considering an adversary that directly operates on the channel output without quantization, and directly analyzes the difference in terms of covertness induced by a code between two close channel states.

A. Discretization

Since the uncertainty set $\mathcal{S}$ described in (43) is uncountable, we first discretize $\mathcal{S}$ to construct a countably finite uncertainty set $\mathcal{S}_n$ with a suitable choice of discretization level and discretization points.

Note that since the uncertainty set $\mathcal{S}$ is subject to the spectral norm constraint, which results in an m-dimensional hypercube with length λ_0 on each side, a natural way to discretize is to uniformly slice $\mathcal{S}$ into 2^{mn} hypercubic regions with length $\epsilon_n \triangleq \lambda_0 2^{-n}$ on each side. The discretization points constructing the set $\mathcal{S}_n$ are chosen as follows:

$$\begin{aligned}\mathcal{S}_n \triangleq \{&\mathbf{H} = \mathbf{U}\Lambda_J\mathbf{V}^\intercal : \Lambda_J = \mathrm{diag}\left(j_1\epsilon_n, \dots, j_m\epsilon_n\right), \\ &\times J = (j_1,\dots,j_m), j_\ell \in [\![1, 2^n]\!], \forall \ell \in [\![1,m]\!]\},\end{aligned}$$

where J is an index for the elements in $\mathcal{S}_n$, and since $\mathbf{U}$ is known to Willie, we henceforth omit its impact in the remaining.

Each discretization point $\mathbf{H}_J$ is associated with a neighborhood

$$\mathcal{S}_{J,n} \triangleq \{\widetilde{\mathbf{H}} = \mathbf{U}\widetilde{\Lambda}\mathbf{V}^\intercal : \Lambda_J \succcurlyeq \widetilde{\Lambda}, \|\Lambda_J - \widetilde{\Lambda}\|_2 < \epsilon_n\}, \tag{48}$$

which covers a portion of the original uncertainty set. By construction, $\cup_{\mathbf{H}_J\in\mathcal{S}_n}\mathcal{S}_{J,n} = \mathcal{S}$. As discussed in previous sections, without loss of generality, we directly investigate the parallel sub-channels described by Λ_b and Λ_w.

B. Approximation

Consider a BPSK constellation $\rho_{n,j} \triangleq \frac{\tau_j Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{n}}$ for all $j \in [\![1,m]\!]$ with $\{\tau_j\}_{j=1}^m$ defined in a way similar to (32). Let $\mathbf{P}_n = \frac{Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{n}}\mathbf{T}$. For any of the above neighborhoods, the covertness metric at any channel realization $\widetilde{\mathbf{H}}$ is close to that measured at the corresponding discretization point $\mathbf{H}$. Precisely, we show that the difference of covertness metric between them vanishes fast with respect to the blocklength n in the following lemma, which is proved in Appendix F.

Lemma 6: For any $\widetilde{\mathbf{H}} \in \mathcal{S}_{J,n}$ and its associated discretization point $\mathbf{H} \in \mathcal{S}_n$,

$$|\mathbb{V}(\widehat{Q}^n_{\widetilde{\mathbf{H}}}, Q_0^{\otimes n}) - \mathbb{V}(\widehat{Q}^n_{\mathbf{H}}, Q_0^{\otimes n})| \leqslant \mathcal{O}\left(n^{\frac{3}{4}}e^{-n\log 2}\right). \tag{49}$$

Thus, the covertness metric of any point in $\mathcal{S}$ can be closely approximated by some discretization point in $\mathcal{S}_n$ for a sufficiently large n.

C. Existence

To show the existence of a compound code applicable for the entire uncertainty set $\mathcal{S}$, note that by the property of supremum and our approximation argument (49), for n sufficiently large, we have

$$\begin{aligned}&\left|\sup_{\mathbf{H}_w\in\mathcal{S}}\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q_0^{\otimes n}) - \max_{\mathbf{H}_w\in\mathcal{S}_n}\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q_0^{\otimes n})\right| \\ &\qquad\leqslant \mathcal{O}\left(n^{\frac{3}{4}}e^{-n\log 2}\right).\end{aligned} \tag{50}$$

Accordingly, we choose $\max_{\mathbf{H}_w\in\mathcal{S}_n}\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q_0^{\otimes n})$ as our optimization constraint by using (50). This modification only causes a small perturbation in the throughput, which is negligible in the limit of large n. Furthermore, we have

$$\begin{aligned}&\max_{\mathbf{H}_w\in\mathcal{S}_n}\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q_0^{\otimes n}) \\ &\quad\leqslant \max_{\mathbf{H}_w\in\mathcal{S}_n}\mathbb{V}(Q^{\otimes n}_{\mathbf{H}_w}, Q_0^{\otimes n}) + \max_{\mathbf{H}_w\in\mathcal{S}_n}\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q^{\otimes n}_{\mathbf{H}_w}).\end{aligned} \tag{51}$$

Note that for any $\mathbf{H}$, $\widetilde{\mathbf{H}} \in \mathcal{S}$ such that $\mathbf{H} - \widetilde{\mathbf{H}} \succcurlyeq \mathbf{0}$ and for n large enough, $\mathbb{V}(Q^{\otimes n}_{\mathbf{H}}, Q_0^{\otimes n}) - \mathbb{V}(Q^{\otimes n}_{\widetilde{\mathbf{H}}}, Q_0^{\otimes n}) \geqslant 2Q\left(\sqrt{\frac{1}{2}\sum_{j=1}^m\frac{\widetilde{\lambda}_j^4\tau_j^2}{4\sigma_w^4}}Q^{-1}\left(\frac{1-\delta}{2}\right)\right) - 2Q\left(\sqrt{\frac{1}{2}\sum_{j=1}^m\frac{\lambda_j^4\tau_j^2}{4\sigma_w^4}}Q^{-1}\left(\frac{1-\delta}{2}\right)\right) \geqslant 0$, where $\{\lambda_j\}_{j=1}^m$ and

$\{\widetilde{\lambda}_j\}_{j=1}^m$ are the diagonal elements of Λ and $\widetilde{\Lambda}$ corresponding to $\mathbf{H}$ and $\widetilde{\mathbf{H}}$ defined in (43), respectively. To ensure

$$\max_{\mathbf{H}_w \in \mathcal{S}_n} \mathbb{V}(Q^{\otimes n}_{\mathbf{H}_w}, Q_0^{\otimes n}) \leqslant \delta - \frac{1}{\sqrt{n}}, \tag{52}$$

for large enough n, by using (124) and (125) in Appendix E, we have the optimization constraint

$$\frac{1}{4\sigma_w^4}\mathrm{tr}\left(\Lambda_0^4\mathbf{T}^2\right) \leqslant 2 - \frac{\bar{C}}{\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)} \tag{53}$$

for some $\bar{C} > 0$. Also, choosing

$$\log M_n K_n = (1+\xi)\,\frac{1}{2\sigma_w^2}\mathrm{tr}\left(\Lambda_0^2\mathbf{T}\right)\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right), \tag{54}$$

ensures that $\max_{\mathbf{H}_w\in\mathcal{S}_n}\mathbb{E}_{\mathcal{C}}\{\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q^{\otimes n}_{\mathbf{H}_w})\}$ vanishes in $\exp\left(-\mathcal{O}\left(\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)\right)\right)$, where it follows from (129) and (130) in Appendix E.

We next show that for any BPSK random code, $\max_{\mathbf{H}_w\in\mathcal{S}_n}\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q^{\otimes n}_{\mathbf{H}_w})$ can be upper-bounded by $\max_{\mathbf{H}_w\in\mathcal{S}_n}\mathbb{E}_{\mathcal{C}}\{\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q^{\otimes n}_{\mathbf{H}_w})\}$.

Lemma 7: For any generated code, set

$$\begin{aligned}\mathcal{E} \triangleq \Bigg\{&\max_{\mathbf{H}_w\in\mathcal{S}_n}\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q^{\otimes n}_{\mathbf{H}_w})\\ &\times \leqslant \max_{\mathbf{H}_w\in\mathcal{S}_n}\mathbb{E}_{\mathcal{C}}\{\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q^{\otimes n}_{\mathbf{H}_w})\} + \alpha_n\Bigg\},\end{aligned} \tag{55}$$

where $\alpha_n = o\left(\frac{1}{\sqrt{n}}\right)$*. Then,* $\mathbb{P}(\mathcal{E}) \geqslant 1 - |\mathcal{S}_n|\exp\left(-2M_nK_n\alpha_n^2\right)$*.*

Proof: We have

$$\begin{aligned}\mathbb{P}(\mathcal{E}) &\overset{\text{(a)}}{\geqslant} 1 - \sum_{\mathbf{H}\in\mathcal{S}_n}\mathbb{P}\Bigg(\mathbb{V}(\widehat{Q}^n_{\mathbf{H}}, Q^{\otimes n}_{\mathbf{H}})\\ &\qquad\geqslant \max_{\mathbf{H}_w\in\mathcal{S}_n}\mathbb{E}_{\mathcal{C}}\{\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q^{\otimes n}_{\mathbf{H}_w})\} + \alpha_n\Bigg) &(56)\\ &\overset{\text{(b)}}{\geqslant} 1 - \sum_{\mathbf{H}\in\mathcal{S}_n}\mathbb{P}(\mathbb{V}(\widehat{Q}^n_{\mathbf{H}}, Q^{\otimes n}_{\mathbf{H}}) \geqslant \mathbb{E}_{\mathcal{C}}\{\mathbb{V}(\widehat{Q}^n_{\mathbf{H}}, Q^{\otimes n}_{\mathbf{H}})\} + \alpha_n) &(57)\\ &\overset{\text{(c)}}{\geqslant} 1 - |\mathcal{S}_n|\exp\left(-2M_nK_n\alpha_n^2\right), &(58)\end{aligned}$$

where (a) follows from the union bound, (b) follows since $\max_{\mathbf{H}_w\in\mathcal{S}_n}\mathbb{E}_{\mathcal{C}}\{\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q^{\otimes n}_{\mathbf{H}_w})\} \geqslant \mathbb{E}_{\mathcal{C}}\{\mathbb{V}(\widehat{Q}^n_{\mathbf{H}}, Q^{\otimes n}_{\mathbf{H}})\}$ for any $\mathbf{H} \in \mathcal{S}_n$, and (c) follows from McDiarmid's Theorem [4, Lemma 2]. ■

Hence, we have $\mathbb{P}(\mathcal{E}) \to 1$ as $n \to \infty$ since, with our choice in (54), $\exp(-M_nK_n) = \exp\left(-\exp\left(\mathcal{O}\left(\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)\right)\right)\right)$ and $|\mathcal{S}_n| = \exp(\mathcal{O}(n))$. If n is large enough, with overwhelming probability, we can rewrite (51) as follows:

$$\begin{aligned}&\max_{\mathbf{H}_w\in\mathcal{S}_n}\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q_0^{\otimes n})\\ &\quad\leqslant \max_{\mathbf{H}_w\in\mathcal{S}_n}\mathbb{V}(Q^{\otimes n}_{\mathbf{H}_w}, Q_0^{\otimes n})\\ &\qquad + \exp\left(-\mathcal{O}\left(\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)\right)\right) + \alpha_n.\end{aligned} \tag{59}$$

As a result, we show that for n large enough, there exists a random code $\mathcal{C}_c$ generated according to the constraints (53) and (54), which is also a compound covert code for the whole discretized uncertainty set $\mathcal{S}_n$, i.e.,

$$\begin{aligned}&\max_{\mathbf{H}_w\in\mathcal{S}_n}\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q_0^{\otimes n})\\ &\quad\leqslant \delta - \frac{1}{\sqrt{n}} + \exp\left(-\mathcal{O}\left(\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)\right)\right) + \alpha_n &(60)\\ &\quad= \delta - \frac{1}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right) \leqslant \delta - \frac{D}{\sqrt{n}} &(61)\end{aligned}$$

is ensured for sufficiently large n and some $D > 0$.

Eventually, combining the above with the triangle inequality and (50), for n large enough, we can develop a bound similar to (37) as follows:

$$\sup_{\mathbf{H}_w\in\mathcal{S}}\mathbb{V}(\widehat{Q}^n_{\mathbf{H}_w}, Q_0^{\otimes n}) \leqslant \delta - \frac{D}{\sqrt{n}} + \mathcal{O}\left(n^{\frac{3}{4}}e^{-n\log 2}\right) \leqslant \delta. \tag{62}$$

Therefore, $\mathcal{C}_c$ is also a compound covert code for the entire uncertainty set $\mathcal{S}$.

D. Constellation Power Design

The power design follows the same steps as in (38)-(42). To find the optimal design point of $\mathbf{T}$, we first ignore the $\mathcal{O}\left(\frac{1}{\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)}\right)$ term in the (53) and include it later as a perturbation as in (40). By solving an optimization program similar to (39), we obtain

$$\lim_{n\to\infty}\frac{\log M_n}{\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)} = (1-\xi)\,\frac{\sigma_w^2}{\sigma_b^2}\sqrt{2\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_0^{-1}\right)^4\right)}. \tag{63}$$

Therefore, by using (62), we can further normalize (63) and obtain (45). (46) follows similarly. ■

V. Comparison and Discussion

We now compare our results with the ones obtained when measuring covertness with relative entropy in [23], [24]. There are several key distinctions between the present work and [23], [24]: 1) We analyze covertness in terms of variational distance, which is a more operationally relevant covertness metric, and leads to a higher number of covert bits. 2) We do not assume that Alice and Bob use a *large amount of key* to create independent and identically distributed (i.i.d.) codewords, and resort instead to a channel resolvability analysis and a conservative amount of secret key $S \in [\![1, K_n]\!]$ shared between Alice and Bob. 3) We develop a complete characterization of the covert capacity, which is only implicitly defined in [23], [24] through an optimization problem that depends on the blocklength n [23, Appendix B]. 4) We do not require the covert bits to be secret. In our opinion, requiring the covert bits to be secret makes the problem closer to a wiretap channel and we want to exclusively focus on covertness. 5) We do not investigate in depth what happens when channel matrices have non-trivial null spaces, except for a short discussion in Remark 1 and Appendix A. In our opinion, these situations are not particularly difficult to analyze because the optimal signaling schemes are rather straightforward.

Note that [23] does not completely fit into our framework since the codebook is assumed to be secret from Willie in [23, Theorems 1 and 2] and Willie directly observes an i.i.d. stochastic process. Nevertheless, one can still compare resulting rates.

Note that the result obtained in [23, Theorem 2] is not a closed-form expression, for it involves the characterization of "normalized KL divergence" defined in [23, (15)]. This characterization remains incomplete therein, for the authors do not exactly solve the power allocation problem in [23, Appendix B]. Hence, the optimal scaling L in [23, Theorem 2], which plays the role of the *covert capacity*, has an implicit dependency on the blocklength n. To make a fair comparison, we specialize the result in the same scenario as ours (i.e., the signal, the channel matrices, and the AWGN are all real number, we consider the same full-rank assumption and our assumption on numbers of antennas, and the covertness requirement is $\mathbb{D}(\widehat{Q}^n \parallel Q_0^{\otimes n}) \leqslant \delta$) to obtain

$$\max_{\mathbf{P}_n \succcurlyeq \mathbf{0}} \frac{1}{2} \log \left| \mathbf{I}_m + \frac{1}{\sigma_b^2} \Lambda_b \mathbf{P}_n \Lambda_b^{\intercal} \right|, \tag{64a}$$

$$\text{s.t. } \frac{1}{2} \mathrm{tr} \left(\frac{1}{\sigma_w^2} \Lambda_w \mathbf{P}_n \Lambda_w^{\intercal} - \log \left(\mathbf{I}_m + \frac{1}{\sigma_w^2} \Lambda_w \mathbf{P}_n \Lambda_w^{\intercal} \right) \right) \leqslant \frac{\delta}{n}, \tag{64b}$$

By following the same line of reasoning as in [2, Section V] and the sensitivity analysis in the proof of Proposition 2, we can actually solve the power allocation problem of [23, Appendix B] and (64) to obtain the first-order asymptotics. Because the power vanishes with n, we also introduce the notation $\rho_{n,j} \triangleq \tau_j \sqrt{\frac{\delta}{n}}$ for all $j \in [\![1, m]\!]$. The power allocation problem, (64a)-(64b), reduces to

$$\max_{\mathbf{T} \succcurlyeq \mathbf{0}} \frac{1}{2\sigma_b^2} \mathrm{tr} \left(\Lambda_b^2 \mathbf{T} \right), \tag{65a}$$

$$\text{s.t. } \frac{1}{4\sigma_w^4} \mathrm{tr} \left(\Lambda_w^4 \mathbf{T}^2 \right) \leqslant 1, \tag{65b}$$

where we have ignored higher-order terms vanishing with n because of the sensitivity analysis. The optimal solution to (65) is $\mathbf{T} = 2\sigma_w^2 \frac{\Lambda_b^2 (\Lambda_w^{-1})^4}{\sqrt{\mathrm{tr}\left(\Lambda_b^4 \left(\Lambda_w^{-1}\right)^4\right)}}$. We therefore express the covert capacity under a relative entropy metric as follows:

$$\lim_{n\to\infty} \frac{\log M_{n,D}}{\sqrt{n\delta}} = \frac{\sigma_w^2}{\sigma_b^2} \sqrt{\mathrm{tr}\left(\Lambda_b^4 \left(\Lambda_w^{-1} \right)^4 \right)}. \tag{66}$$

Hence, the first-order asymptotics of the optimal covert throughput under a relative entropy constraint $\mathbb{D}(\widehat{Q}^n \parallel Q_0^{\otimes n}) \leqslant \delta$ and a variational distance constraint $\mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n}) \leqslant \delta$ can be expressed as

$$\lim_{n\to\infty} \frac{\log M_{n,D}(\delta)}{\sqrt{n}} = \frac{\sigma_w^2}{\sigma_b^2} \sqrt{\mathrm{tr}\left(\Lambda_b^4 \left(\Lambda_w^{-1} \right)^4 \right) \delta} \triangleq f_D(\delta), \text{ and} \tag{67}$$

$$\lim_{n\to\infty} \frac{\log M_{n,V}(\delta)}{\sqrt{n}} = \frac{\sigma_w^2}{\sigma_b^2} \sqrt{2\mathrm{tr}\left(\Lambda_b^4 \left(\Lambda_w^{-1} \right)^4 \right)} Q^{-1} \left(\frac{1-\delta}{2} \right) \triangleq f_V(\delta), \tag{68}$$

respectively. Note that the above results are consistent with the ones for the AWGN channel under a relative entropy metric [2, Theorem 5] and a variational distance metric [5] if we consider Single-Input Single-Output (SISO) channels with unit gain.

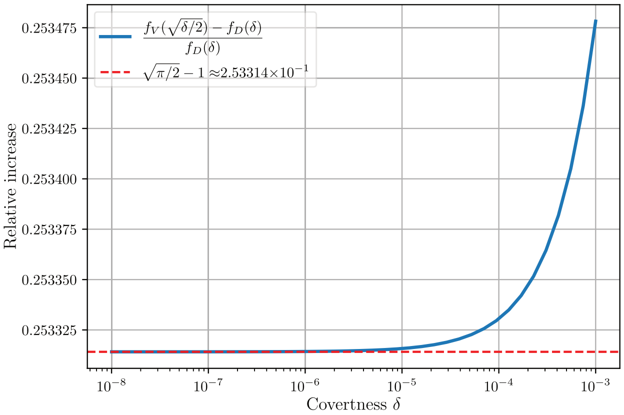

Fig. 1. Relative increase of the first-order asymptotics of optimal covert throughput as a function of covertness.

As remarked in [4, Remark 2], since $2\mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n}) \leqslant \mathbb{D}(\widehat{Q}^n \parallel Q_0^{\otimes n})$ by Pinsker's inequality, requiring $\mathbb{D}(\widehat{Q}^n \parallel Q_0^{\otimes n}) \leqslant \delta$ is more stringent than requiring $\mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n}) \leqslant \sqrt{\delta/2}$. Consequently,

$$f_D(\delta) \leqslant f_V(\sqrt{\delta/2}). \tag{69}$$

We illustrate the above relation with a simple numerical example. We consider a compound channel case in which both channel matrices are 4×4 real matrices and $\lambda_0 = 0.05$, the main channel has generalized singular value $\Lambda_b = \mathrm{diag}\,(0.385, 0.214, 0.172, 0.028)$, and $\sigma_w^2 = 2\sigma_b^2 = 0.001$. As shown in Fig. 1, using the variational distance metric results in at least a 25% relative increase in covert throughput. In fact, by the Maclaurin series for the inverse error function erf^{-1}, we have

$$\begin{aligned} \lim_{\delta\to 0} \frac{f_V(\sqrt{\delta/2})}{f_D(\delta)} &= \lim_{\delta\to 0} \frac{\sqrt{2} Q^{-1}\left(\frac{1-\sqrt{\delta/2}}{2}\right)}{\sqrt{\delta}} \\ &= \lim_{\delta\to 0} \frac{2\mathrm{erf}^{-1}(\sqrt{\delta/2})}{\sqrt{\delta}} \\ &= \lim_{x\to 0} \frac{2\mathrm{erf}^{-1}(x)}{\sqrt{2x^2}} \geqslant \sqrt{2} \lim_{x\to 0} \frac{\frac{\sqrt{\pi}}{2} x + \mathcal{O}(x^3)}{x} \\ &= \sqrt{\frac{\pi}{2}}. \end{aligned} \tag{70}$$

Appendix A
Further Discussion of the Assumption on Rank and Numbers of Antennas

In this section, we provide a more detailed discussion of our full-rank assumption for channel matrices $\mathbf{H}_w$ and $\mathbf{H}_b$ using a GSVD and the analysis of [28, Section II.A]. Without any assumption on the channel matrices, we first define the following subspaces

$$\begin{aligned}\mathcal{S}_b &\triangleq \mathrm{Null}(\mathbf{H}_b)^{\perp} \cap \mathrm{Null}(\mathbf{H}_w),\\ \mathcal{S}_{b,w} &\triangleq \mathrm{Null}(\mathbf{H}_b)^{\perp} \cap \mathrm{Null}(\mathbf{H}_w)^{\perp},\\ \mathcal{S}_w &\triangleq \mathrm{Null}(\mathbf{H}_b) \cap \mathrm{Null}(\mathbf{H}_w)^{\perp},\\ \mathcal{S}_n &\triangleq \mathrm{Null}(\mathbf{H}_b) \cap \mathrm{Null}(\mathbf{H}_w),\end{aligned}$$

which correspond to whether the signal in the subspaces can be observed by Bob or Willie. Let

$$m \triangleq \mathrm{rank}\left(\begin{bmatrix}\mathbf{H}_b\\ \mathbf{H}_w\end{bmatrix}\right),$$

$p \triangleq \dim(\mathcal{S}_b)$, and $q \triangleq \dim(\mathcal{S}_{b,w})$. Clearly, $N_a \geqslant m$, $\dim(\mathcal{S}_n) = N_a - m$, and $\dim(\mathcal{S}_w) = m - p - q$. Both channel matrices can be decomposed with a GSVD as

$$\begin{aligned}\mathbf{H}_b &= \mathbf{U}_b' \Sigma_b [\mathbf{\Omega}^{-1} \quad \mathbf{0}_{m\times(N_a-m)}]\Psi^{\mathsf{T}},\\ \mathbf{H}_w &= \mathbf{U}_w' \Sigma_w [\mathbf{\Omega}^{-1} \quad \mathbf{0}_{m\times(N_a-m)}]\Psi^{\mathsf{T}},\end{aligned}$$

where $\Psi \in \mathbb{R}^{N_a\times N_a}$, $\mathbf{U}_b' \in \mathbb{R}^{N_b\times N_b}$, and $\mathbf{U}_w' \in \mathbb{R}^{N_w\times N_w}$ are orthogonal, $\mathbf{\Omega} \in \mathbb{R}^{m\times m}$ is lower triangular and nonsingular, and

$$\Sigma_b = \begin{array}{c c} & \begin{array}{ccc} m-p-q & q & p \end{array}\\ \begin{array}{c} N_b-p-q\\ q\\ p \end{array} & \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \Lambda_b & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \end{array}$$

$$\Sigma_w = \begin{array}{c c} & \begin{array}{ccc} m-p-q & q & p \end{array}\\ \begin{array}{c} m-p-q\\ q\\ N_w+p-m \end{array} & \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \Lambda_w & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \end{array},$$

with $\Lambda_b = \mathrm{diag}\left(\{\lambda_{b,j}\}_{j=1}^{q}\right)$ and $\Lambda_w = \mathrm{diag}\left(\{\lambda_{w,j}\}_{j=1}^{q}\right)$. Note that the following cases do not contribute to the investigation of the square-root law. 1) If $\mathcal{S}_b$ is not trivial, Alice can steer her signal in these directions without being detected by Willie, and therefore overcome the square-root law (i.e., the covert capacity as defined in (5) is unbounded). The use of these directions contributes nothing to the covertness metric, and a power constraint must be active for the rates to be finite. The optimal power allocation is then the usual water-filling solution. We exclude this case by assuming $p = 0$. 2) If $\mathcal{S}_w$ is not trivial, Alice would avoid using these directions, since the signals in those directions cannot be observed by Bob. We exclude this case by assuming $m = p + q$. 3) If $\mathcal{S}_n$ is not trivial, Alice would similarly avoid using these directions, and we also exclude this case by assuming $N_a = m$.

In summary, only $\mathcal{S}_{b,w}$ is relevant to the square-root law. By excluding the above three cases, if $N_a \leqslant N_b$ and $N_a \leqslant N_w$, we have the full-rank assumption $m = \mathrm{rank}(\mathbf{H}_b) = \mathrm{rank}(\mathbf{H}_w) = N_a$. If $N_a > N_b$ or $N_a > N_w$, some of the mentioned null spaces may not be trivial, and this still falls into the scenarios we point out in Remark 1, which does not affect the investigation of the square-root law. For instance, if $N_a > N_b$, then either $N_a > m$ or $N_a = m$ is true, and the former case corresponds to a non-trivial $\mathcal{S}_n$. If we also have $N_a = m$, since $m = N_a > N_b \geqslant q$, at least one of $m > p + q$ or $p > 0$ is true, which corresponds to the non-trivial $\mathcal{S}_w$ or $\mathcal{S}_b$. Hence, we also impose the assumption that both Bob and Willie possess more antennas than Alice. The assumptions on numbers of antennas and rank are without loss of generality, and the reason is simply to exclude some perhaps less interesting cases to reveal the constant before the square-root instead of preventing any technical difficulty.

APPENDIX B
PROOF OF LEMMA 5

Proof: Under the null hypothesis H_0, for all $i \in [\![1, n]\!]$ $\mathbf{Z}_i \sim \mathcal{N}\left(\mathbf{0}, \sigma_w^2 \mathbf{I}_{N_w}\right)$, so that $\hat{\mathbf{Z}}_i \sim \mathcal{N}\left(\mathbf{0}, \sigma_w^2 \Lambda_b^2 \left(\Lambda_w^{-1}\right)^2\right) \in \mathbb{R}^m$. Hence, we have the following statistics:

$$\mu_0 = \sum_{i=1}^{n} \mathbb{E}_{Q_0}\{S_i\} = n\sigma_w^2 \mathrm{tr}\left(\Lambda_b^2 \left(\Lambda_w^{-1}\right)^2\right), \tag{71}$$

$$\sigma_0^2 = \sum_{i=1}^{n} \mathrm{Var}(S_i) = 2n\sigma_w^4 \mathrm{tr}\left(\Lambda_b^4 \left(\Lambda_w^{-1}\right)^4\right), \tag{72}$$

$$t_{0,i} = \mathbb{E}_{Q_0}\left\{|S_i - \mu_{0,i}|^3\right\} = \mathcal{O}\left(1\right), \quad t_0 = \sum_{i=1}^{n} t_{0,i} = \mathcal{O}\left(n\right), \tag{73}$$

where $\mu_{0,i} \triangleq \mathbb{E}_{Q_0}\{S_i\}$. We use the Berry-Esseen Theorem to obtain an upper bound for the probability of false alarm as follows:

$$\begin{aligned}\alpha &= \mathbb{P}_{H_0}\left(\sum_{i=1}^{n} S_i \geqslant \tau\right) &(74)\\ &\leqslant Q\left(\frac{\tau - n\sigma_w^2 \mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)}{\sqrt{2n\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}\sigma_w^2}\right) + \frac{6t_0}{\sigma_0^3} &(75)\\ &\leqslant Q\left(\frac{\tau - n\sigma_w^2 \mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)}{\sqrt{2n\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}\sigma_w^2}\right) + \frac{B_0}{\sqrt{n}}. &(76)\end{aligned}$$

The bound on the probability of false alarm (13) follows by applying the threshold $\tau = \frac{P_*}{2} + n\sigma_w^2 \mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)$.

Similarly, under the hypothesis H_1, we know that given a codeword $\mathbf{x}^{(k)n}$ transmitted over the channel $W_{\mathbf{Z}|\mathbf{X}}^{\otimes n}$, for every $i \in [\![1, n]\!]$ $\mathbf{Z}_i|\mathbf{X}_i = \mathbf{x}_i^{(k)} \sim \mathcal{N}\left(\mathbf{H}_w\mathbf{x}_i^{(k)}, \sigma_w^2 \mathbf{I}_{N_w}\right)$, so that $\hat{\mathbf{Z}}_i|\tilde{\mathbf{X}}_i = \tilde{\mathbf{x}}_i^{(k)} \sim \mathcal{N}\left(\Lambda_b\tilde{\mathbf{x}}_i^{(k)}, \sigma_w^2\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)$, where $\tilde{\mathbf{x}}_i^{(k)} = \mathbf{V}^{\mathsf{T}}\mathbf{x}_i^{(k)}$. Let $\tilde{\mathbf{x}}^{(k)n} \triangleq \mathbf{V}^{\mathsf{T}}\mathbf{x}^{(k)n}$ and $\mathbf{P}^{(k)} \triangleq \sum_{i=1}^{n} \tilde{\mathbf{x}}_i^{(k)}\tilde{\mathbf{x}}_i^{(k)\mathsf{T}}$. Hence, we have the following statistics:

$$\begin{aligned}\mu_1^{(k)} &= \sum_{i=1}^{n} \mathbb{E}_{\widehat{Q}}\left\{S_i|\mathbf{X}_i = \mathbf{x}_i^{(k)}\right\}\\ &= \mathrm{tr}\left(\Lambda_b^2\mathbf{P}^{(k)}\right) + n\sigma_w^2\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right), &(77)\end{aligned}$$

$$\begin{aligned}\sigma_1^{(k)2} &= \sum_{i=1}^{n} \mathrm{Var}\left(S_i|\mathbf{X}_i=\mathbf{x}_i^{(k)}\right)\\ &= 4\sigma_w^2\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^2\mathbf{P}^{(k)}\right)+2n\sigma_w^4\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right),\end{aligned} \tag{78}$$

$$\begin{aligned}t_{1,i}^{(k)} &= \mathbb{E}_{\widehat{Q}}\left\{|S_i-\mu_{1,i}^{(k)}|^3|\mathbf{X}_i=\mathbf{x}_i^{(k)}\right\}=\mathcal{O}\left(1\right),\\ t_1^{(k)} &= \sum_{i=1}^{n} t_{1,i}^{(k)}=\mathcal{O}\left(n\right),\end{aligned} \tag{79}$$

where $\mu_{1,i}^{(k)} \triangleq \mathbb{E}_{\widehat{Q}}\left\{S_i|\mathbf{X}_i=\mathbf{x}_i^{(k)}\right\}$. We use again the Berry-Esseen Theorem to obtain an upper bound for the probability of missed detection. For the k-th codeword, by defining $\beta^{(k)} = \mathbb{P}_{H_1}\left(\sum_{i=1}^{n} S_i<\tau|\mathbf{X}^n=\mathbf{x}^{(k)n}\right)$, we have (80), shown at the bottom of the next page, where (a) follows since Λ_b, Λ_w, $\mathbf{P}^k \succeq \mathbf{0}$, and since $\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^2\mathbf{P}^{(k)}\right) \leqslant \mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)\mathrm{tr}\left(\Lambda_b^2\mathbf{P}^{(k)}\right)$. By setting the detection threshold to $\tau = \frac{P_*}{2}+n\sigma_w^2\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)$, we further obtain (81), shown at the bottom of the next page, where (a) follows since $\mathrm{tr}\left(\Lambda_b^2\mathbf{P}^{(k)}\right)-\frac{P_*}{2} \geqslant \frac{\mathrm{tr}\left(\Lambda_b^2\mathbf{P}^{(k)}\right)}{2} \geqslant \frac{P_*}{2}$ and $\frac{\mathrm{tr}\left(\Lambda_b^2\mathbf{P}^{(k)}\right)}{\sqrt{a\mathrm{tr}\left(\Lambda_b^2\mathbf{P}^{(k)}\right)+b}} \geqslant \frac{P_*}{\sqrt{aP_*+b}}$, $\forall a,b>0$, (b) follows from $\frac{1}{\sqrt{1+x}} \geqslant 1-\frac{x}{2}$, $\forall x>0$, and (c) follows from $Q\left(x-y\right)=Q\left(x\right)+\int_{x-y}^{x}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)du \leqslant Q\left(x\right)+\frac{y}{\sqrt{2\pi}}$ $\forall\ 0<y<x$. Note that the upper bound (81) is independent of the codeword $\mathbf{x}^{(k)n}$, and hence we can use the bound on all the $\beta^{(k)}$'s. Therefore, we obtain

$$\begin{aligned}\beta &= \mathbb{P}_{H_1}\left(\sum_{i=1}^{n} S_i<\tau\right)=\frac{1}{|\mathcal{M}|}\sum_{\mathbf{x}^{(k)n}\in\mathcal{M}}\beta^{(k)}\\ &\leqslant Q\left(\frac{P_*}{2\sqrt{2n\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}\sigma_w^2}\right)\\ &\quad+\frac{P_*^2\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)}{4\sqrt{\pi}n^{3/2}\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)^{3/2}\sigma_w^4}+\frac{B_1}{\sqrt{n}}.\end{aligned} \tag{82}$$

■

Appendix C
Proof of Lemma 6

Proof: We partition the code $\mathcal{C}$ into two different sub-codes, a low-power sub-code $\mathcal{C}^{(\ell)}$ and a high-power sub-code $\mathcal{C}^{(h)}$, where $\mathcal{C}^{(\ell)} \triangleq \{\mathbf{x}^n\in\mathcal{C}:\|\mathbf{H}_b\mathbf{x}^n\|_F^2 \leqslant A\sqrt{n}\}$, $\mathcal{C}^{(h)} \triangleq \mathcal{C}\backslash\mathcal{C}^{(\ell)}$. The output distributions induced by these two sub-codes are

$$\widehat{Q}^{(\ell)}\left(\mathbf{z}^n\right)=\frac{1}{\left|\mathcal{C}^{(\ell)}\right|}\sum_{\mathbf{x}^n\in\mathcal{C}^{(\ell)}}W_{\mathbf{Z}|\mathbf{X}}^{\otimes n}\left(\mathbf{z}^n|\mathbf{x}^n\right),$$

$$\text{and } \widehat{Q}^{(h)}\left(\mathbf{z}^n\right)=\frac{1}{\left|\mathcal{C}^{(h)}\right|}\sum_{\mathbf{x}^n\in\mathcal{C}^{(h)}}W_{\mathbf{Z}|\mathbf{X}}^{\otimes n}\left(\mathbf{z}^n|\mathbf{x}^n\right), \tag{83}$$

respectively. Note that $\widehat{Q}^n=\frac{|\mathcal{C}^{(\ell)}|}{|\mathcal{C}|}\widehat{Q}^{(\ell)}+\frac{|\mathcal{C}^{(h)}|}{|\mathcal{C}|}\widehat{Q}^{(h)}$. For a code $\mathcal{C}$ such that $\mathbb{V}(\widehat{Q}^n,Q_0^{\otimes n}) \leqslant \delta$, we have

$$\delta \geqslant \mathbb{V}(\widehat{Q}^n,Q_0^{\otimes n}) \tag{84}$$

$$\overset{(a)}{\geqslant} \frac{\left|\mathcal{C}^{(h)}\right|}{|\mathcal{C}|}\mathbb{V}(\widehat{Q}^{(h)},Q_0^{\otimes n})-\frac{\left|\mathcal{C}^{(\ell)}\right|}{|\mathcal{C}|}\mathbb{V}(\widehat{Q}^{(\ell)},Q_0^{\otimes n}) \tag{85}$$

$$=\mathbb{V}(\widehat{Q}^{(h)},Q_0^{\otimes n})-\frac{\left|\mathcal{C}^{(\ell)}\right|}{|\mathcal{C}|}(\mathbb{V}(\widehat{Q}^{(h)},Q_0^{\otimes n})+\mathbb{V}(\widehat{Q}^{(\ell)},Q_0^{\otimes n})) \tag{86}$$

$$\overset{(b)}{\geqslant} \mathbb{V}(\widehat{Q}^{(h)},Q_0^{\otimes n})-2\frac{\left|\mathcal{C}^{(\ell)}\right|}{|\mathcal{C}|} \tag{87}$$

$$\begin{aligned}&\overset{(c)}{\geqslant} 1-2Q\left(\frac{A\sqrt{n}}{2\sqrt{2n\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}\sigma_w^2}\right)\\ &\quad-\frac{A^2n\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)}{4\sqrt{\pi}n^{3/2}\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)^{3/2}\sigma_w^4}-\frac{B_0+B_1}{\sqrt{n}}\\ &\quad-2\frac{\left|\mathcal{C}^{(\ell)}\right|}{|\mathcal{C}|}\end{aligned} \tag{88}$$

$$\begin{aligned}&=\delta+\frac{2\nu^2\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)}{4\sqrt{\pi n}\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)^{3/2}\sigma_w^4}-\frac{B_0+B_1}{\sqrt{n}}\\ &\quad-\frac{A^2\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)}{4\sqrt{\pi n}\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)^{3/2}\sigma_w^4}+2\gamma_n-2\frac{\left|\mathcal{C}^{(\ell)}\right|}{|\mathcal{C}|}\end{aligned} \tag{89}$$

$$\overset{(d)}{\geqslant} \delta+2\gamma_n-2\frac{\left|\mathcal{C}^{(\ell)}\right|}{|\mathcal{C}|} \tag{90}$$

where (a) follows from

$$\frac{\left|\mathcal{C}^{(h)}\right|}{|\mathcal{C}|}\mathbb{V}(\widehat{Q}^{(h)},Q_0^{\otimes n}) \tag{91}$$

$$=\frac{1}{2}\left\|\left(\widehat{Q}^n-Q_0^{\otimes n}\right)-\frac{\left|\mathcal{C}^{(\ell)}\right|}{|\mathcal{C}|}\left(\widehat{Q}^{(\ell)}-Q_0^{\otimes n}\right)\right\|_1 \tag{92}$$

$$\leqslant \mathbb{V}(\widehat{Q}^n,Q_0^{\otimes n})+\frac{\left|\mathcal{C}^{(\ell)}\right|}{|\mathcal{C}|}\mathbb{V}(\widehat{Q}^{(\ell)},Q_0^{\otimes n}), \tag{93}$$

(b) follows since the variational distance between any two distributions is upper bounded by 1, (c) follows from (15), and (d) follows by choosing ν to satisfy

$$\begin{aligned}&\frac{2\nu^2\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)}{4\sqrt{\pi n}\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)^{3/2}\sigma_w^4}-\frac{B_0+B_1}{\sqrt{n}}\\ &\quad-\frac{A^2\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)}{4\sqrt{\pi n}\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)^{3/2}\sigma_w^4}>0.\end{aligned} \tag{94}$$

Hence, we can bound the cardinality of the low-power sub-code $\mathcal{C}^{(\ell)}$ from below as $\left|\mathcal{C}^{(\ell)}\right| \geqslant \gamma_n|\mathcal{C}|$, which shows the existence of such a low-power sub-code. ■

Appendix D
Proof of Lemma 8

Proof: We specialize [3, Lemma 3] and [4, Lemma 1] into the following lemma.

Lemma 8: For any $\gamma > 0$,

$$\mathbb{E}\left\{\bar{P}_e^{(n)}\right\} \leqslant M_n e^{-\gamma}\left(1+\mathbb{E}_{P_n^{\otimes n}}\left\{\frac{P_n^{\otimes n}\left(\mathbf{Y}^n\right)}{P_0^{\otimes n}\left(\mathbf{Y}^n\right)}\right\}\right) + \mathbb{P}_{\Pi_{\mathbf{Q}_n}^{\otimes n} W_{\mathbf{Y}|\mathbf{X}}^{\otimes n}}\left(\log\frac{W_{\mathbf{Y}|\mathbf{X}}^{\otimes n}\left(\mathbf{Y}^n|\mathbf{X}^n\right)}{P_0^{\otimes n}\left(\mathbf{Y}^n\right)} \leqslant \gamma\right), \quad (95)$$

where $\bar{P}_e^{(n)}$ is the average probability of error.

We first analyze the first term on the right-hand side of (95) as follows:

$$\mathbb{E}_{P_n}\left\{\frac{P_n\left(\mathbf{Y}\right)}{P_0\left(\mathbf{Y}\right)}\right\} \overset{\text{(a)}}{=} \prod_{j=1}^{m}\mathbb{E}_{\widetilde{P}_{n,j}}\left\{\frac{\widetilde{P}_{n,j}\left(\tilde{Y}_j\right)}{\widetilde{P}_{0,j}\left(\tilde{Y}_j\right)}\right\} \quad (96)$$

$$= \prod_{j=1}^{m}\cosh\left(\frac{\lambda_{b,j}^2\rho_{n,j}}{\sigma_b^2}\right) \overset{\text{(b)}}{\leqslant} \exp\left(\frac{\sum_{j=1}^{m}\lambda_{b,j}^4\rho_{n,j}^2}{2\sigma_b^4}\right), \quad (97)$$

where (a) follows from the facts that we apply the orthogonal transform $\mathbf{U}_b^{'\mathsf{T}}$ to the observation $\mathbf{Y}$, each sub-channel is independent, and this mapping is one-to-one and onto. Note that after the orthogonal transform, we could truncate the last $N_b - m$ components, since they contain pure noise (as they correspond to the null space of $\mathbf{H}_b^{\mathsf{T}}$) and thus do not affect the decoding process. (b) follows from $\cosh(x) \leqslant e^{\frac{x^2}{2}}$ and the exponential property. Therefore, we know that

$$\mathbb{E}_{P_n^{\otimes n}}\left\{\frac{P_n^{\otimes n}\left(\mathbf{Y}^n\right)}{P_0^{\otimes n}\left(\mathbf{Y}^n\right)}\right\} \leqslant \exp\left(\frac{n\sum_{j=1}^{m}\lambda_{b,j}^4\rho_{n,j}^2}{2\sigma_b^4}\right) = \mathcal{O}\left(1\right). \quad (98)$$

$$\beta^{(k)} \leqslant Q\left(\frac{\mathrm{tr}\left(\Lambda_b^2\mathbf{P}^{(k)}\right) + n\sigma_w^2\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right) - \tau}{\sqrt{4\sigma_w^2\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^2\mathbf{P}^{(k)}\right) + 2n\sigma_w^4\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}}\right) + \frac{6t_1^{(k)}}{\sigma_1^{(k)3}}$$
$$\overset{\text{(a)}}{\leqslant} Q\left(\frac{\mathrm{tr}\left(\Lambda_b^2\mathbf{P}^{(k)}\right) + n\sigma_w^2\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right) - \tau}{\sqrt{4\sigma_w^2\mathrm{tr}\left(\Lambda_b^2\mathbf{P}^{(k)}\right)\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right) + 2n\sigma_w^4\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}}\right) + \frac{B_1}{\sqrt{n}}, \quad (80)$$

$$\beta^{(k)} \overset{\text{(a)}}{\leqslant} Q\left(\frac{\frac{P_*}{2}}{\sqrt{4P_*\sigma_w^2\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right) + 2n\sigma_w^4\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}}\right) + \frac{B_1}{\sqrt{n}}$$
$$= Q\left(\frac{P_*}{2\sqrt{2n\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}\sigma_w^2}\frac{1}{\sqrt{1+\frac{2P_*\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)}{n\sigma_w^2\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}}}\right) + \frac{B_1}{\sqrt{n}}$$
$$\overset{\text{(b)}}{\leqslant} Q\left(\frac{P_*}{2\sqrt{2n\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}\sigma_w^2}\left(1-\frac{P_*\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)}{n\sigma_w^2\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}\right)\right) + \frac{B_1}{\sqrt{n}}$$
$$\overset{\text{(c)}}{\leqslant} Q\left(\frac{P_*}{2\sqrt{2n\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)}\sigma_w^2}\right) + \frac{P_*^2\mathrm{tr}\left(\Lambda_b^2\left(\Lambda_w^{-1}\right)^2\right)}{4\sqrt{\pi}n^{3/2}\mathrm{tr}\left(\Lambda_b^4\left(\Lambda_w^{-1}\right)^4\right)^{3/2}\sigma_w^4} + \frac{B_1}{\sqrt{n}}, \quad (81)$$

Next, we turn to analyze the last term of (95). Similarly, with the above orthogonal transform, note that

$$\log \frac{W_{\mathbf{Y}|\mathbf{X}}^{\otimes n}\left(\mathbf{Y}^n|\mathbf{X}^n\right)}{P_0^{\otimes n}\left(\mathbf{Y}^n\right)} = \sum_{j=1}^{m} \log \frac{W_{\tilde{Y}^{(j)}|\tilde{X}^{(j)}}^{\otimes n}\left(\tilde{Y}_j^n|\tilde{X}_j^n\right)}{\widetilde{P}_{0,j}^{\otimes n}\left(\tilde{Y}_j^n\right)} \tag{99}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{\lambda_{b,j}\tilde{X}_{ij}\tilde{Y}_{ij}}{\sigma_b^2} - \frac{\lambda_{b,j}^2\tilde{X}_{ij}^2}{2\sigma_b^2}\right). \tag{100}$$

Since each sub-channel is independent and $\tilde{Y}_{ij}|\tilde{X}_{ij} = \tilde{x}_{ij} \sim \mathcal{N}\left(\lambda_{b,j}\tilde{x}_{ij}, \sigma_b^2\right)$ and $\tilde{x}_{ij} \in \{-a_{n,j}, a_{n,j}\}$ for every j, we have $\sum_{j=1}^{m} \log \frac{W_{\tilde{Y}(j)|\tilde{X}(j)}\left(\tilde{Y}_{ij}|\tilde{X}_{ij}=\tilde{x}_{ij}\right)}{\widetilde{P}_{0,j}\left(\tilde{Y}_{ij}\right)} \sim \sum_{j=1}^{m}\mathcal{N}\left(\frac{\lambda_{b,j}^2\rho_{n,j}}{2\sigma_b^2}, \frac{\lambda_{b,j}^2\rho_{n,j}}{\sigma_b^2}\right)$. Therefore, by setting $\gamma = (1-\epsilon)\, n \sum_{j=1}^{m} \frac{\lambda_{b,j}^2\rho_{n,j}}{2\sigma_b^2}$, where $\epsilon \in (0,1)$, and using Hoeffding's inequality, we have

$$\mathbb{P}_{W_{\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}=\tilde{\mathbf{x}}^n}^{\otimes n}}\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\log \frac{W_{\tilde{Y}^{(j)}|\tilde{X}^{(j)}}\left(\tilde{Y}_{ij}|\tilde{x}_{ij}\right)}{\widetilde{P}_0\left(\tilde{Y}_{ij}\right)} \leqslant n\left(1-\epsilon\right)\sum_{j=1}^{m}\frac{\lambda_{b,j}^2\rho_{n,j}}{2\sigma_b^2}\right)$$

$$\leqslant \exp\left(-n\sum_{j=1}^{m}\frac{\epsilon^2\lambda_{b,j}^2\rho_{n,j}}{8\sigma_b^2}\right). \tag{101}$$

Then, we have

$$\mathbb{P}_{\Pi_{\mathbf{Q}_n}^{\otimes n} W_{\mathbf{Y}|\mathbf{X}}^{\otimes n}}\left(\log \frac{W_{\mathbf{Y}|\mathbf{X}}^{\otimes n}\left(\mathbf{Y}^n|\mathbf{X}^n\right)}{P_0^{\otimes n}\left(\mathbf{Y}^n\right)} \leqslant \gamma\right) \tag{102}$$

$$= \sum_{\tilde{\mathbf{x}}^n \in \prod_{j=1}^{m}\{-a_{n,j}, a_{n,j}\}^n} \Pi_{\mathbf{P}_n}^{\otimes n}\left(\tilde{\mathbf{x}}^n\right)$$

$$\mathbb{P}_{W_{\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}=\tilde{\mathbf{x}}^n}^{\otimes n}}\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\log \frac{W_{\tilde{Y}^{(j)}|\tilde{X}^{(j)}}\left(\tilde{Y}_{ij}|\tilde{x}_{ij}\right)}{\widetilde{P}_0\left(\tilde{Y}_{ij}\right)} \leqslant n\left(1-\epsilon\right)\sum_{j=1}^{m}\frac{\lambda_{b,j}^2\rho_{n,j}}{2\sigma_b^2}\right)$$

$$\leqslant \exp\left(-n\sum_{j=1}^{m}\frac{\epsilon^2\lambda_{b,j}^2\rho_{n,j}}{8\sigma_b^2}\right). \tag{103}$$

Eventually, by combining (95), (98), and (103), we have

$$\mathbb{E}\left\{\bar{P}_e^{(n)}\right\} \leqslant \exp\left(-n\sum_{j=1}^{m}\frac{\epsilon^2\lambda_{b,j}^2\rho_{n,j}}{8\sigma_b^2}\right) + M_n e^{-\gamma}\left(1+\mathcal{O}(1)\right). \tag{104}$$

Hence, by using (32), if we choose

$$\log M_n = (1-\omega)(1-\epsilon)\, n \sum_{j=1}^{m}\frac{\lambda_{b,j}^2\rho_{n,j}}{2\sigma_b^2}$$

$$= (1-\xi)\sum_{j=1}^{m}\frac{\lambda_{b,j}^2\tau_j\sqrt{n}\,Q^{-1}\left(\frac{1-\delta}{2}\right)}{2\sigma_b^2}, \tag{105}$$

where $\omega \in (0,1)$ and $\xi = \frac{1}{2}[(1+\omega)(1+\epsilon)-(1-\omega)(1-\epsilon)] > 0$, the result follows. ■

Appendix E
Proof of Lemma 9

Proof: In the following, we apply the triangle inequality to upper-bound the covertness metric, put a direct constraint on $\mathbb{V}(Q_n^{\otimes n}, Q_0^{\otimes n})$, and show the remaining term vanishes exponentially fast. First note that, by the triangle inequality, we have

$$\mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n}) \leqslant \mathbb{V}(\widehat{Q}^n, Q_n^{\otimes n}) + \mathbb{V}(Q_n^{\otimes n}, Q_0^{\otimes n}). \tag{106}$$

We first analyze the second term on the left-hand side of (106). By the basic property of the variational distance, we have

$$\mathbb{V}(Q_n^{\otimes n}, Q_0^{\otimes n})$$

$$= \mathbb{P}_{Q_n^{\otimes n}}\left(Q_n^{\otimes n}\left(\mathbf{Z}^n\right) \geqslant Q_0^{\otimes n}\left(\mathbf{Z}^n\right)\right) - \mathbb{P}_{Q_0^{\otimes n}}\left(Q_n^{\otimes n}\left(\mathbf{Z}^n\right) \geqslant Q_0^{\otimes n}\left(\mathbf{Z}^n\right)\right) \tag{107}$$

$$= \mathbb{P}_{Q_n^{\otimes n}}\left(\sum_{i=1}^{n}\log\frac{Q_n\left(\mathbf{Z}_i\right)}{Q_0\left(\mathbf{Z}_i\right)} \geqslant 0\right) - \mathbb{P}_{Q_0^{\otimes n}}\left(\sum_{i=1}^{n}\log\frac{Q_n\left(\mathbf{Z}_i\right)}{Q_0\left(\mathbf{Z}_i\right)} \geqslant 0\right) \tag{108}$$

$$\overset{(a)}{=} \mathbb{P}_{\widetilde{Q}_n^{\otimes n}}\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\log\frac{\widetilde{Q}_{n,j}\left(\tilde{Z}_{ij}\right)}{\widetilde{Q}_{0,j}\left(\tilde{Z}_{ij}\right)} \geqslant 0\right) - \mathbb{P}_{\widetilde{Q}_0^{\otimes n}}\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\log\frac{\widetilde{Q}_{n,j}\left(\tilde{Z}_{ij}\right)}{\widetilde{Q}_{0,j}\left(\tilde{Z}_{ij}\right)} \geqslant 0\right), \tag{109}$$

where (a) follows since we apply the orthogonal transform $\mathbf{U}_w'^{\mathsf{T}}$ to the per-channel-use observation $\mathbf{Z}_i$, and this mapping does not reduce the variational distance (i.e., the equality of data-processing inequality holds). Note that after the orthogonal transform, we could truncate the last $N_w - m$ components, since they contain pure noise (as they correspond to the null space of $\mathbf{H}_w^{\mathsf{T}}$). Then, for every $i \in [\![1, n]\!]$,

$$\mu_{1j} \triangleq \mathbb{E}_{\widetilde{Q}_{n,j}}\left\{\log\frac{\widetilde{Q}_{n,j}\left(Z_i\right)}{\widetilde{Q}_{0,j}\left(Z_i\right)}\right\} \tag{110}$$

$$= \mathbb{E}_{\widetilde{Q}_{n,j}}\left\{-\frac{\lambda_{w,j}^2\rho_{n,j}}{2\sigma_w^2} + \log\left(\cosh\left(\frac{\lambda_{w,j}a_{n,j}Z_i}{\sigma_w^2}\right)\right)\right\} \tag{111}$$

$$= \frac{\lambda_{w,j}^4\rho_{n,j}^2}{4\sigma_w^4} + \mathcal{O}\left(\rho_{n,j}^3\right), \tag{112}$$

$$\sigma_{1j}^2 \triangleq \mathrm{Var}\left(\log\frac{\widetilde{Q}_{n,j}\left(Z_i\right)}{\widetilde{Q}_{0,j}\left(Z_i\right)}\right) \tag{113}$$

$$= \mathbb{E}_{\widetilde{Q}_{n,j}}\left\{\log^2\frac{\widetilde{Q}_{n,j}\left(Z_i\right)}{\widetilde{Q}_{0,j}\left(Z_i\right)}\right\} - \mathcal{O}\left(\rho_{n,j}^4\right) \tag{114}$$

$$= \frac{\lambda_{w,j}^4\rho_{n,j}^2}{2\sigma_w^4} + \mathcal{O}\left(\rho_{n,j}^3\right), \tag{115}$$

$$t_{1j} \triangleq \mathbb{E}_{\widetilde{Q}_{n,j}}\left\{\left|\log\frac{\widetilde{Q}_{n,j}\left(Z_i\right)}{\widetilde{Q}_{0,j}\left(Z_i\right)} - \mu_{1j}\right|^3\right\} = \mathcal{O}\left(\rho_{n,j}^3\right). \quad (116)$$

Therefore, by the Berry-Esseen Theorem, we have

$$\begin{aligned}\mathbb{P}_{\widetilde{Q}_n^{\otimes n}}&\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\log\frac{\widetilde{Q}_{n,j}\left(\tilde{Z}_{ij}\right)}{\widetilde{Q}_{0,j}\left(\tilde{Z}_{ij}\right)} \geqslant 0\right)\\ &\leqslant Q\left(-\sqrt{\frac{n}{2}\sum_{j=1}^{m}\frac{\lambda_{w,j}^4\rho_{n,j}^2}{4\sigma_w^4}}\right) + \frac{6n\sum_{j=1}^{m}t_{1j}}{\left(n\sum_{j=1}^{m}\sigma_{1j}^2\right)^{3/2}} \quad (117)\\ &= 1 - Q\left(\sqrt{\frac{n}{2}\sum_{j=1}^{m}\frac{\lambda_{w,j}^4\rho_{n,j}^2}{4\sigma_w^4}}\right) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right). \quad (118)\end{aligned}$$

Similarly, for every $i \in [\![1, n]\!]$,

$$\mu_{0j} \triangleq \mathbb{E}_{\widetilde{Q}_{0,j}}\left\{\log\frac{\widetilde{Q}_{n,j}\left(Z_i\right)}{\widetilde{Q}_{0,j}\left(Z_i\right)}\right\} = -\frac{\lambda_{w,j}^4\rho_{n,j}^2}{4\sigma_w^4} + \mathcal{O}\left(\rho_{n,j}^3\right), \quad (119)$$

$$\sigma_{0j}^2 \triangleq \mathrm{Var}\left(\log\frac{\widetilde{Q}_{n,j}\left(Z_i\right)}{\widetilde{Q}_{0,j}\left(Z_i\right)}\right) = \frac{\lambda_{w,j}^4\rho_{n,j}^2}{2\sigma_w^4} + \mathcal{O}\left(\rho_{n,j}^3\right), \quad (120)$$

$$t_{0j} \triangleq \mathbb{E}_{\widetilde{Q}_{0,j}}\left\{\left|\log\frac{\widetilde{Q}_{n,j}\left(Z_i\right)}{\widetilde{Q}_{0,j}\left(Z_i\right)} - \mu_{1j}\right|^3\right\} = \mathcal{O}\left(\rho_{n,j}^3\right). \quad (121)$$

We therefore have

$$\begin{aligned}\mathbb{P}_{\widetilde{Q}_0^{\otimes n}}&\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\log\frac{\widetilde{Q}_{n,j}\left(\tilde{Z}_{ij}\right)}{\widetilde{Q}_{0,j}\left(\tilde{Z}_{ij}\right)} \geqslant 0\right)\\ &\geqslant Q\left(\sqrt{\frac{n}{2}\sum_{j=1}^{m}\frac{\lambda_{w,j}^4\rho_{n,j}^2}{4\sigma_w^4}}\right) - \frac{6n\sum_{j=1}^{m}t_{0j}}{\left(n\sum_{j=1}^{m}\sigma_{0j}^2\right)^{3/2}} \quad (122)\\ &= Q\left(\sqrt{\frac{n}{2}\sum_{j=1}^{m}\frac{\lambda_{w,j}^4\rho_{n,j}^2}{4\sigma_w^4}}\right) - \mathcal{O}\left(\frac{1}{\sqrt{n}}\right). \quad (123)\end{aligned}$$

Eventually, we find an upper bound for (109) as follows:

$$\begin{aligned}\mathbb{V}(Q_n^{\otimes n}, Q_0^{\otimes n}) \leqslant 1 - 2Q&\left(\sqrt{\frac{n}{2}\sum_{j=1}^{m}\frac{\lambda_{w,j}^4\rho_{n,j}^2}{4\sigma_w^4}}\right)\\ &+ \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) \leqslant \delta - \frac{1}{\sqrt{n}}, \quad (124)\end{aligned}$$

where the $-\frac{1}{\sqrt{n}}$ term is added to ensure that the $\mathbb{V}(Q_n^{\otimes n}, Q_0^{\otimes n})$ term is less than δ for n large enough. Equivalently, we impose a covertness constraint

$$\frac{1}{4\sigma_w^4}\mathrm{tr}\left(\Lambda_w\mathbf{T}\Lambda_w^{\intercal}\Lambda_w\mathbf{T}\Lambda_w^{\intercal}\right) \leqslant 2 - \frac{C}{\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)}, \quad (125)$$

for some $C > 0$ and the constraint (125) would be the main concern in the power design optimization. Combining (106) with (124), we therefore have, for n large enough,

$$\begin{aligned}\mathbb{V}(\widehat{Q}^n, Q_0^{\otimes n}) &\leqslant \mathbb{V}(\widehat{Q}^n, Q_n^{\otimes n}) + \mathbb{V}(Q_n^{\otimes n}, Q_0^{\otimes n})\\ &\leqslant \mathbb{V}(\widehat{Q}^n, Q_n^{\otimes n}) + \delta - \frac{1}{\sqrt{n}}. \quad (126)\end{aligned}$$

For the term $\mathbb{V}(\widehat{Q}^n, Q_n^{\otimes n})$ in (126), our analysis follows from [3, Lemma 5], and we recall the following lemma.

Lemma 9: For any $\theta > 0$,

$$\begin{aligned}\mathbb{E}\{\mathbb{V}(\widehat{Q}^n, Q_n^{\otimes n})\} \leqslant \mathbb{P}_{\Pi_{\mathbf{Q}_n}^{\otimes n}W_{\mathbf{Z|X}}^{\otimes n}}&\left(\log\frac{W_{\mathbf{Z|X}}^{\otimes n}\left(\mathbf{Z}^n|\mathbf{X}^n\right)}{Q_0^{\otimes n}\left(\mathbf{Z}^n\right)} \geqslant \theta\right)\\ &+ \frac{1}{2}\sqrt{\frac{e^{\theta}}{M_nK_n}}. \quad (127)\end{aligned}$$

Similarly, we apply the orthogonal transform $\mathbf{U}_w'^{\intercal}$ to the observations and decompose them into observations on each sub-channel; we obtain

$$\log\frac{W_{\mathbf{Z|X}}^{\otimes n}\left(\mathbf{Z}^n|\mathbf{X}^n\right)}{Q_0^{\otimes n}\left(\mathbf{Z}^n\right)} = \sum_{i=1}^{n}\sum_{j=1}^{m}\log\frac{W_{\tilde{Z}^{(j)}|\tilde{X}^{(j)}}\left(\tilde{Z}_{ij}|\tilde{X}_{ij}\right)}{\widetilde{Q}_{0,j}\left(\tilde{Z}_{ij}\right)}. \quad (128)$$

Since $\tilde{x}_{ij} \in \{-a_{n,j}, a_{n,j}\}$, $\sum_{j=1}^{m}\log\frac{W_{\tilde{Z}^{(j)}|\tilde{X}^{(j)}}\left(\tilde{Z}_{ij}|\tilde{X}_{ij}=\tilde{x}_{ij}\right)}{\widetilde{Q}_{0,j}\left(\tilde{Z}_{ij}\right)}$ $\sim \sum_{j=1}^{m}\mathcal{N}\left(\frac{\lambda_{w,j}^2\rho_{n,j}}{2\sigma_w^2}, \frac{\lambda_{w,j}^2\rho_{n,j}}{\sigma_w^2}\right)$. Therefore, by setting $\theta = (1+\epsilon)n\sum_{j=1}^{m}\frac{\lambda_{w,j}^2\rho_{n,j}}{2\sigma_w^2}$, and using Hoeffding's inequality, we have $\mathbb{P}_{\Pi_{\mathbf{P}_n}^{\otimes n}W_{\mathbf{Z|X}}^{\otimes n}}\left(\log\frac{W_{\mathbf{Z|X}}^{\otimes n}(\mathbf{Z}^n|\mathbf{X}^n)}{Q_0^{\otimes n}(\mathbf{Z}^n)} \geqslant \theta\right) \leqslant \exp\left(-\frac{\epsilon^2n\sum_{j=1}^{m}\lambda_{w,j}^2\rho_{n,j}}{8\sigma_w^2}\right)$. Therefore, $\mathbb{E}\{\mathbb{V}(\widehat{Q}^n, Q_n^{\otimes n})\} \leqslant \exp\left(-\frac{\epsilon^2n\sum_{j=1}^{m}\lambda_{w,j}^2\rho_{n,j}}{8\sigma_w^2}\right) + \frac{1}{2}\sqrt{\frac{e^{\theta}}{M_nK_n}}$. Eventually, recalling (32), if we choose

$$\begin{aligned}\log M_nK_n &= (1+\omega)(1+\epsilon)n\sum_{j=1}^{m}\frac{\lambda_{w,j}^2\rho_{n,j}}{2\sigma_w^2}\\ &= (1+\xi)\sum_{j=1}^{m}\frac{\lambda_{w,j}^2\tau_j\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)}{2\sigma_w^2}, \quad (129)\end{aligned}$$

then

$$\mathbb{E}\{\mathbb{V}(\widehat{Q}^n, Q_n^{\otimes n})\} \leqslant e^{-\theta_2\sqrt{n}Q^{-1}\left(\frac{1-\delta}{2}\right)}, \quad (130)$$

for some appropriate choice of $\theta_2 > 0$. The result follows by combining (126) and (130). ■

Appendix F
Proof of Lemma 12

Proof: We start by using $\mathbb{V}(\widehat{Q}_{\mathbf{H}}^n, Q_0^{\otimes n})$ to approximate $\mathbb{V}(\widehat{Q}_{\widetilde{\mathbf{H}}}^n, Q_0^{\otimes n})$. For a fixed $\mathbf{P}_n$, by the triangle inequality,

$$|\mathbb{V}(\widehat{Q}_{\widetilde{\mathbf{H}}}^n, Q_0^{\otimes n}) - \mathbb{V}(\widehat{Q}_{\mathbf{H}}^n, Q_0^{\otimes n})| \leqslant \mathbb{V}(\widehat{Q}_{\widetilde{\mathbf{H}}}^n, \widehat{Q}_{\mathbf{H}}^n). \quad (131)$$

Note that for a given code $\mathcal{C}$ and $\mathbf{x}^{(\ell k)n} \in \mathcal{C}$,

$$\begin{aligned}&\mathbb{V}(\widehat{Q}_{\widetilde{\mathbf{H}}}^n, \widehat{Q}_{\mathbf{H}}^n)\\ &\quad\leqslant \frac{\sum_{\ell=1}^{M_n}\sum_{k=1}^{K_n}}{M_nK_n}\mathbb{V}(\widetilde{W}_{\mathbf{Z|X}}^{\otimes n}\left(\mathbf{Z}^n|\mathbf{x}^{(\ell k)n}\right), W_{\mathbf{Z|X}}^{\otimes n}\left(\mathbf{Z}^n|\mathbf{x}^{(\ell k)n}\right))\\ &\quad\overset{(a)}{=} \frac{\sum_{\ell=1}^{M_n}\sum_{k=1}^{K_n}}{M_nK_n}\mathbb{V}(\widetilde{W}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}^{\otimes n}\left(\tilde{\mathbf{Z}}^n|\tilde{\mathbf{x}}^{(\ell k)n}\right), W_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}^{\otimes n}\left(\tilde{\mathbf{Z}}^n|\tilde{\mathbf{x}}^{(\ell k)n}\right)),\end{aligned}$$

where (a) follows since the orthogonal transformation preserves the variational distance. To characterize the behavior of $\mathbb{V}(\widetilde{W}^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{Z}}^n|\tilde{\mathbf{x}}^{(\ell k)n}), W^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{Z}}^n|\tilde{\mathbf{x}}^{(\ell k)n}))$, we follow the proof of [26, Lemma 5] and consider a specific codeword $\tilde{\mathbf{x}}^n$,

$$\begin{aligned}
&2\mathbb{V}(\widetilde{W}^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{Z}}^n|\tilde{\mathbf{x}}^n), W^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{Z}}^n|\tilde{\mathbf{x}}^n)) \\
&= \int_{\tilde{\mathbf{z}}^n} \left|\widetilde{W}^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}^n|\tilde{\mathbf{x}}^n) - W^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}^n|\tilde{\mathbf{x}}^n)\right| d\tilde{\mathbf{z}}^n \qquad (132)\\
&= \int_{\sum_{i=1}^n \|\tilde{\mathbf{z}}_i - \widetilde{\Lambda}\tilde{\mathbf{x}}_i\|_2^2 \geqslant r_n^2} \left|\widetilde{W}^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}^n|\tilde{\mathbf{x}}^n) - W^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}^n|\tilde{\mathbf{x}}^n)\right| d\tilde{\mathbf{z}}^n \\
&\quad + \int_{\sum_{i=1}^n \|\tilde{\mathbf{z}}_i - \widetilde{\Lambda}\tilde{\mathbf{x}}_i\|_2^2 < r_n^2} \left|\widetilde{W}^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}^n|\tilde{\mathbf{x}}^n) - W^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}^n|\tilde{\mathbf{x}}^n)\right| d\tilde{\mathbf{z}}^n \qquad (133)\\
&\leqslant \mathbb{P}_{\widetilde{W}^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}}\left(\sum_{i=1}^n \|\tilde{\mathbf{z}}_i - \widetilde{\Lambda}\tilde{\mathbf{x}}_i\|_2^2 \geqslant r_n^2\right) \\
&\quad + \mathbb{P}_{W^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}}\left(\sum_{i=1}^n \|\tilde{\mathbf{z}}_i - \widetilde{\Lambda}\tilde{\mathbf{x}}_i\|_2^2 \geqslant r_n^2\right) \\
&\quad + \int_{\sum_{i=1}^n \|\tilde{\mathbf{z}}_i - \widetilde{\Lambda}\tilde{\mathbf{x}}_i\|_2^2 < r_n^2} \left|\widetilde{W}^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}^n|\tilde{\mathbf{x}}^n) - W^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}^n|\tilde{\mathbf{x}}^n)\right| d\tilde{\mathbf{z}}^n, \qquad (134)
\end{aligned}$$

where $r_n \triangleq \sqrt{nm(1+\epsilon)\sigma_w^2} + \epsilon_n\|\tilde{\mathbf{x}}^n\|_F$, which is close to $\sqrt{nm(1+\epsilon)\sigma_w^2}$ as n grows to infinity, and $\epsilon \in (0,1)$. Note that the first term of (134) can be bounded by the concentration inequality for the sub-exponential random variables as follows:

$$\begin{aligned}
&\mathbb{P}_{\widetilde{W}^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}}\left(\sum_{i=1}^n \|\tilde{\mathbf{z}}_i - \widetilde{\Lambda}\tilde{\mathbf{x}}_i\|_2^2 \geqslant r_n^2\right) \\
&\qquad \leqslant \exp\left(-\frac{(r_n^2 - nm\sigma_w^2)^2}{8nm}\right) = \exp\left(-\mathcal{O}(n)\right). \qquad (135)
\end{aligned}$$

Also, by the triangle inequality for the Frobenius norm, we have

$$\begin{aligned}
\left\|\tilde{\mathbf{z}}^n - \widetilde{\Lambda}\tilde{\mathbf{x}}^n\right\|_F &\leqslant \left\|\tilde{\mathbf{z}}^n - \Lambda\tilde{\mathbf{x}}^n\right\|_F + \left\|\left(\Lambda - \widetilde{\Lambda}\right)\tilde{\mathbf{x}}^n\right\|_F \qquad (136)\\
&\leqslant \left\|\tilde{\mathbf{z}}^n - \Lambda\tilde{\mathbf{x}}^n\right\|_F + \sqrt{\epsilon_n^2 \sum_{i=1}^n \|\tilde{\mathbf{x}}_i\|_2^2} \qquad (137)\\
&= \left\|\tilde{\mathbf{z}}^n - \Lambda\tilde{\mathbf{x}}^n\right\|_F + \epsilon_n\left\|\tilde{\mathbf{x}}^n\right\|_F. \qquad (138)
\end{aligned}$$

Therefore, $\sum_{i=1}^n \left\|\tilde{\mathbf{z}}_i - \widetilde{\Lambda}\tilde{\mathbf{x}}_i\right\|_2^2 \geqslant r_n^2$ implies that $\sum_{i=1}^n \left\|\tilde{\mathbf{z}}_i - \Lambda\tilde{\mathbf{x}}_i\right\|_2^2 \geqslant (r_n - \epsilon_n\|\tilde{\mathbf{x}}^n\|_F)^2 = nm(1+\epsilon)\sigma_w^2$. Accordingly, we can use the concentration inequality for the sub-exponential random variables and obtain $\mathbb{P}_{W^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}}\left(\sum_{i=1}^n \left\|\tilde{\mathbf{z}}_i - \widetilde{\Lambda}\tilde{\mathbf{x}}_i\right\|_2^2 \geqslant r_n^2\right) \leqslant \exp\left(-\frac{nm\epsilon^2}{8}\right)$. We next investigate the variation between densities $W^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}$ and $\widetilde{W}^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}$ caused by the difference between $\mathbf{H}$ and $\widetilde{\mathbf{H}}$ as follows:

$$\begin{aligned}
&\left|\log \frac{W^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}^n|\tilde{\mathbf{x}}^n)}{\widetilde{W}^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}^n|\tilde{\mathbf{x}}^n)}\right| \\
&\quad = \frac{1}{2\sigma_w^2}\left|\sum_{i=1}^n \left(\|\tilde{\mathbf{z}}_i - \widetilde{\Lambda}\tilde{\mathbf{x}}_i\|_2^2 - \|\tilde{\mathbf{z}}_i - \Lambda\tilde{\mathbf{x}}_i\|_2^2\right)\right| \qquad (139)\\
&\quad \overset{(a)}{\leqslant} \frac{1}{2\sigma_w^2}\left(\left|\sum_{i=1}^n\sum_{j=1}^m 2(\tilde{z}_{ij} - \widetilde{\lambda}_j\tilde{x}_{ij})(\epsilon_n\tilde{x}_{ij})\right| \right.\\
&\qquad \left. + \left|\sum_{i=1}^n\sum_{j=1}^m (\epsilon_n\tilde{x}_{ij})^2\right|\right) \qquad (140)\\
&\quad \overset{(b)}{\leqslant} \frac{1}{\sigma_w^2}\sqrt{\sum_{i=1}^n \|\tilde{\mathbf{z}}_i - \widetilde{\Lambda}\tilde{\mathbf{x}}_i\|_2^2\epsilon_n^2\|\tilde{\mathbf{x}}^n\|_F^2} + \frac{1}{2\sigma_w^2}\epsilon_n^2\|\tilde{\mathbf{x}}^n\|_F^2, \qquad (141)
\end{aligned}$$

where (a) follows from the definition of $\mathcal{S}_{J,n}$ and the triangle inequality, and (a) follows from the Cauchy-Schwartz inequality. Then, for the last term of (134), we proceed as follows:

$$\begin{aligned}
&\int_{\sum_{i=1}^n \|\tilde{\mathbf{z}}_i - \widetilde{\Lambda}\tilde{\mathbf{x}}_i\|_2^2 < r_n^2} \left|\widetilde{W}^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}^n|\tilde{\mathbf{x}}^n) - W^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}^n|\tilde{\mathbf{x}}^n)\right| d\tilde{\mathbf{z}}^n \\
&\overset{(a)}{\leqslant} \int_{\sum_{i=1}^n \|\tilde{\mathbf{z}}_i - \widetilde{\Lambda}\tilde{\mathbf{x}}_i\|_2^2 < r_n^2} \widetilde{W}^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}^n|\tilde{\mathbf{x}}^n)\left(f_n + \mathcal{O}\left(f_n^2\right)\right) d\tilde{\mathbf{z}}^n \qquad (142)\\
&\overset{(b)}{\leqslant} f_n + \mathcal{O}\left(f_n^2\right) = \mathcal{O}\left(n^{\frac{1}{2}}\left\|\tilde{\mathbf{x}}^n\right\|_F e^{-n\log 2}\right), \qquad (143)
\end{aligned}$$

where we let $f_n = \frac{1}{\sigma_w^2}\sqrt{r_n^2\epsilon_n^2\|\tilde{\mathbf{x}}^n\|_F^2} + \frac{1}{2\sigma_w^2}\epsilon_n^2\|\tilde{\mathbf{x}}^n\|_F^2$, since $\sum_{i=1}^n \left\|\tilde{\mathbf{z}}_i - \widetilde{\Lambda}\tilde{\mathbf{x}}_i\right\|_2^2 < r_n^2$, (a) and (b) follow since $\left|1 - \frac{W^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}^n|\tilde{\mathbf{x}}^n)}{\widetilde{W}^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{z}}^n|\tilde{\mathbf{x}}^n)}\right| \leqslant \max\{e^{f_n} - 1, 1 - e^{-f_n}\} \leqslant f_n + \mathcal{O}\left(f_n^2\right)$. Therefore, we have, for n large enough, $\mathbb{V}(\widetilde{W}^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{Z}}^n|\tilde{\mathbf{x}}^n), W^{\otimes n}_{\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}}(\tilde{\mathbf{Z}}^n|\tilde{\mathbf{x}}^n)) \leqslant \mathcal{O}\left(n^{\frac{1}{2}}\|\tilde{\mathbf{x}}^n\|_F e^{-n\log 2}\right)$, and from (132), we obtain

$$\mathbb{V}(\widehat{Q}^n_{\widetilde{\mathbf{H}}}, \widehat{Q}^n_{\mathbf{H}}) \leqslant \sum_{\ell=1}^{M_n}\sum_{k=1}^{K_n}\frac{1}{M_nK_n}\mathcal{O}\left(n^{\frac{1}{2}}\left\|\tilde{\mathbf{x}}^{(\ell k)n}\right\|_F e^{-n\log 2}\right). \qquad (144)$$

For any BPSK code generated independently according to $\Pi^{\otimes n}_{\mathbf{P}_n}$, the power of generated codewords are fixed, and therefore we have

$$\mathbb{V}(\widehat{Q}^n_{\widetilde{\mathbf{H}}}, \widehat{Q}^n_{\mathbf{H}}) \leqslant \mathcal{O}\left(n^{\frac{3}{4}}e^{-n\log 2}\right). \qquad (145)$$

Eventually, combining (131) with (145), the result follows. ■

REFERENCES

[1] 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.

[2] 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.

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

[4] M. Tahmasbi and M. R. Bloch, "First- and second-order asymptotics in covert communication," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2190–2212, Apr. 2019.

[5] Q. E. Zhang, M. R. Bloch, M. Bakshi, and S. Jaggi, "Undetectable radios: Covert communication under spectral mask constraints," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 992–996.

[6] P. H. Che, M. Bakshi, and S. Jaggi, “Reliable deniable communication: Hiding messages in noise,” in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 2945–2949.

[7] S. Yan, B. He, X. Zhou, Y. Cong, and A. L. Swindlehurst, “Delay-intolerant covert communications with either fixed or random transmit power,” *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 1, pp. 129–140, Jan. 2019.

[8] X. Yu, S. Wei, and Y. Luo, “Finite blocklength analysis of Gaussian random coding in AWGN channels under covert constraint,” *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1261–1274, 2021.

[9] K. S. K. Arumugam and M. R. Bloch, “Covert communication over a K-user multiple-access channel,” *IEEE Trans. Inf. Theory*, vol. 65, no. 11, pp. 7020–7044, Nov. 2019.

[10] K. S. K. Arumugam and M. R. Bloch, “Embedding covert information in broadcast communications,” *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 10, pp. 2787–2801, Oct. 2019.

[11] V. Y. F. Tan and S-H Lee, “Time-division is optimal for covert communication over some broadcast channels,” *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 5, pp. 1377–1389, May. 2019.

[12] K.-H. Cho and S.-H. Lee, “Treating interference as noise is optimal for covert communication over interference channels,” *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 322–332, 2021.

[13] B. A. Bash *et al.*, “Quantum-secure covert communication on bosonic channels,” *Nature Commun.*, vol. 6, no. 1, p. 8626, Oct. 2015.

[14] L. Wang, “Optimal throughput for covert communication over a classical-quantum channel,” in *Proc. IEEE Inf. Theory Workshop*, Sep. 2016, pp. 364–368.

[15] G. Frèche, M. Bloch, and M. Barret, “Polar codes for covert communications over asynchronous discrete memoryless channels,” *Entropy*, vol. 20, no. 1, p. 3, Dec. 2017.

[16] I. A. Kadampot, M. Tahmasbi, and M. R. Bloch, “Multilevel-coded pulse-position modulation for covert communications,” in *Proc. IEEE Int. Symp. Inf. Theory*, Vail, CO, USA, Jan. 2018, pp. 1864–1868.

[17] I. A. Kadampot, M. Tahmasbi, and M. R. Bloch, “Codes for covert communication over additive white Gaussian noise channels,” in *Proc. IEEE Int. Symp. Inf. Theory*, Paris, France, Jul. 2019, pp. 977–981.

[18] M. Lamarca and D. Matas, “A non-linear channel code for covert communications,” in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Marrakesh, Morocco, Apr. 2019, pp. 1–7.

[19] Q. Zhang, M. Bakshi, and S. Jaggi, “Covert communication with polynomial computational complexity,” *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1354–1384, Mar. 2020.

[20] J. Hu, S. Yan, X. Zhou, F. Shu, J. Li, and J. Wang, “Covert communication achieved by a greedy relay in wireless networks,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4766–4779, Jul. 2018.

[21] T.-X. Zheng, H.-M. Wang, D. W. K. Ng, and J. Yuan, “Multi-antenna covert communications in random wireless networks,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1974–1987, Mar. 2019.

[22] A. Abdelaziz and C. E. Koksal, “Fundamental limits of covert communication over MIMO AWGN channel,” in *Proc. IEEE Conf. Commun. Netw. Secur.*, Las Vegas, NV, USA, Oct. 2017, pp. 1–9.

[23] A. Bendary, A. Abdelaziz, and C. E. Koksal, “Achieving positive covert capacity over MIMO AWGN channels,” *IEEE J. Sel. Areas Inf. Theory*, vol. 2, no. 1, pp. 149–162, Mar. 2021.

[24] A. Bendary and C. E. Koksal, “Order-optimal scaling of covert communication over MIMO AWGN channels,” in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Avignon, France, Jun. 2020, pp. 1–9.

[25] R. F. Schaefer and S. Loyka, “The secrecy capacity of compound Gaussian MIMO wiretap channels,” *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5535–5552, Oct. 2015.

[26] X. He and A. Yener, “MIMO wiretap channels with unknown and varying eavesdropper channel states,” *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6844–6869, Nov. 2014.

[27] S.-Y. Wang and M. R. Bloch, “Covert MIMO communications under variational distance constraint,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Los Angeles, CA, USA, Jun. 2020, pp. 828–833.

[28] A. Khisti and G. Womell, “Secure transmission with multiple antennas—Part II: The MIMOME wiretap channel,” *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5515–5532, Nov. 2010.

[29] C. C. Paige and M. A. Saunders, “Towards a generalized singular value decomposition,” *SIAM J. Numer. Anal.*, vol. 18, no. 3, pp. 398–405, 1981.

[30] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses* (Springer Texts in Statistics). New York, NY, USA: Springer, 2006.

[31] 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.

[32] T. Han and S. Verdú, “Approximation theory of output statistics,” *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[33] D. G. Luenberger, *Optimization by Vector Space Methods*, 1st ed. New York, NY, USA: Wiley, 1997.

Shi-Yuan Wang (Student Member, IEEE) received the B.S.E. degree in electrical engineering from the National Taiwan University, Taiwan, in 2018. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Georgia Institute of Technology.

Matthieu R. Bloch (Senior Member, IEEE) received the degree in engineering from Supélec, Gif-sur-Yvette, France, the M.S. degree in electrical engineering from Georgia Institute of Technology, Atlanta, in 2003, the Ph.D. degree in engineering science from the Université de Franche-Comté, Besançon, France, in 2006, and the Ph.D. degree in electrical engineering from Georgia Institute of Technology in 2008. From 2008 to 2009, he was a Post-Doctoral Research Associate with the University of Notre Dame, South Bend, IN, USA. Since July 2009, he has been with the Faculty of the School of Electrical and Computer Engineering, Georgia Institute of Technology. He is currently a Professor with the School of Electrical and Computer Engineering, Georgia Institute of Technology. He has coauthored the textbook *Physical-Layer Security: From Information Theory to Security Engineering* (Cambridge University Press). His research interests are in the areas of information theory, error-control coding, wireless communications, and cryptography. He has served on the organizing committee for several international conferences. He was a co-recipient of the IEEE Communications Society and the IEEE Information Theory Society 2011 Joint Paper Award. He was the Chair of the Online Committee of the IEEE Information Theory Society from 2011 to 2014 and has been on the Board of Governors of the IEEE Information Theory Society since 2016 and serves as the 2nd Vice President. He was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY from 2016 to 2019. He has been an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY since 2019.