Fundamental Limits of Covert Communication over MIMO AWGN Channel

Amr Abdelaziz and C. Emre Koksal
Department of Electrical and Computer Engineering
The Ohio State University
Columbus, Ohio 43201
{abdelaziz.7, koksal.2}@osu.edu

Abstract—Fundamental limits of covert communication have been studied for different models of scalar channels. It was shown that, over \( n \) independent channel uses, \( \mathcal{O}(\sqrt{n}) \) bits can transmitted reliably over a public channel while achieving an arbitrarily low probability of detection (LPD) by other stations. This result is well known as the square-root law and even to achieve this diminishing rate of covert communication, all existing studies utilized some form of secret shared between the transmitter and the receiver. In this paper, we establish the limits of LPD communication over the MIMO AWGN channel. In particular, using relative entropy as our LPD metric, we study the maximum codebook size for which the transmitter can guarantee reliability and LPD conditions are met. We first show that, the optimal codebook generating input distribution under \( \delta \)-PD constraint is the zero-mean Gaussian distribution. Then, assuming channel state information (CSI) on only the main channel at the transmitter, we derive the optimal input covariance matrix, hence, establishing scaling laws of the codebook size. We evaluate the codebook scaling rates in the limiting regimes for the number of channel uses (asymptotic block length) and the number of antennas (massive MIMO). We show that, in the asymptotic block-length regime, square-root law still holds for the MIMO AWGN. Meanwhile, in massive MIMO limit, the codebook size, while it scales linearly with \( \sqrt{n} \), it scales exponentially with the number of transmitting antennas. The practical implication of our result is that MIMO has the potential to provide a substantial increase in the file sizes that can be covertly communicated subject to a reasonably low delay.

Index Terms—LPD communication, Covert MIMO Communication, MIMO physical layer security, LPD Capacity.

I. INTRODUCTION

Conditions for secure communication under a passive eavesdropping attack fall in two broad categories: 1) low probability of intercept (LPI), 2) low probability of detection (LPD). Communication with LPI requires the message exchanged by two legitimate parties to be kept secret from an illegitimate adversary. Meanwhile, LPD constrained communication is more restrictive as it requires the adversary to be unable to decide whether communication between legitimate parties has taken place. Fundamental limits of LPD constrained communication over scalar AWGN has been established in [1] where the square-root law for LPD communication was established. Assuming a shared secret of sufficient length between transmitter and receiver, square-root law states that, over \( n \) independent channel uses of an AWGN channel, transmitter can send \( \mathcal{O}(\sqrt{n}) \) bits reliably to the receiver while keeping arbitrary low probability of detection at the adversary. In this paper, we study the fundamental limits of communication with LPD over MIMO AWGN channels.

Consider the scenario in which a transmitter (Alice) wishes to communicate to a receiver (Bob) while being undetected by a passive adversary (Willie) when all nodes are equipped with multiple antennas. To that end, Alice wish to generate a codebook that satisfies both reliability, in terms of low error probability \( \epsilon \), over her channel to Bob and, in the same time, ensures, a certain maximum PD, namely \( \delta \), at Willie. Denote the maximum possible size of such codebook by \( K_n(\delta, \epsilon) \). In this paper, we are interested in establishing the fundamental limits of \( K_n(\delta, \epsilon) \) in the asymptotic length regime and in the limit of large number of transmitting antenna. First we show that, the maximum codebook size is attained when the codebook is generated according to zero mean circular symmetric complex Gaussian distribution. We establish this result building upon the the Principle Minimum Relative Entropy [2] and Information Projection [3].

Some of our findings can be summarized as follows. For an isotropic Willie channel, we show that Alice can transmit \( \mathcal{O}(N \sqrt{n}/M) \) bits reliably in \( n \) independent channel uses, where \( N \) and \( M \) are the number of active eigenmodes of Bob and Willie channels, respectively. Further, we evaluate \( \delta \)-PD rates in the limiting regimes for the number of channel uses (asymptotic block length) and the number of antennas (massive MIMO). We show that, while the square-root law still holds for the MIMO AWGN, the number of bits that can be transmitted covertly scales exponentially with the number of transmitting antennas. The practical implication of our result is that, MIMO has the potential to provide a substantial increase in the file sizes that can be covertly communicated subject to a reasonably low delay.

The contributions of this work can be summarized as follows:

- Using the Principle Minimum Relative Entropy [2] and
Information Projection [3], we show that the $K_n(\delta, \epsilon)$ is achievable when the codebook is generated according to zero mean complex Gaussian distribution in MIMO AWGN channels.

- With the availability of only the main CSI to Alice, we evaluate the optimal input covariance matrix under the assumption that Willie channel satisfies a bounded spectral norm constraint [4], [5]. Singular value decomposition (SVD) precoding is shown to be the optimal signaling strategy and the optimal water-filling strategy is also provided.

- We evaluate the block-length and massive MIMO asymptotics for $K_n(\delta, \epsilon)$. We show that, while the square-root law cannot be avoided, $K_n(\delta, \epsilon)$ scales exponentially with the number of antennas. Thus, MIMO has the potential to provide a substantial increase in the file sizes that can be covertly communicated subject to a reasonably low delay.

**Related Work.** Fundamental limits of covert communication have been studied in literature for different models of scalar channels. In [6], LPD communication over the binary symmetric channel was considered. It was shown that, square-root law holds for the binary symmetric channel, yet, without requiring a shared secret between Alice and Bob when Willie channel is significantly noisier. Further, it was shown that Alice can transmit $O(n)$ bits taking the advantage from Willie’s uncertainty about his own channel transition probabilities. Recently in [7], LPD communication was studied from a resolvability perspective for the discrete memoryless channel (DMC). Wherein, a trade-off between the secret length and asymmetries between Bob and Willie channels has been considered. Later in [8], the exact capacity (using relative entropy instead of total variation distance as LPD measure) of DMC and AWGN have been characterized. For a detailed comparison between recent results on LPD communication for different channel models, secret key length, LPD security metric and achievable LPD rate readers may refer to Table II in [6]. LPD communication over MIMO fading channel was first studied in [9]. Under different assumption of CSI availability, the author derived the average power that satisfies the LPD requirement. However, the authors did not obtain the square-root law, since the focus was not on the asymptotics of the bound. Recently in [10], LPD communication with multiple antennas at both Alice and Bob is considered when Willie has only a single antenna over Rayleigh fading channel. An approximation to the LPD constrained rate when Willie employs a radiometer detector and has uncertainty about his noise variance was presented. However, a full characterization of the capacity of MIMO channel with LPD constraint was not established. Despite not explicitly stated, the assumption of keeping the codebook generated by Alice secret from Willie (or at least a secret of sufficient length [1], [7]) is common in all aforementioned studies of covert communication. Without this assumption, LPD condition cannot be met along with arbitrarily low probability of error at Bob. That is because, when Willie is informed about the codebook, he can decode the message using the same decoding strategy as that of Bob [1]. Only in [6], square-root law was obtained over binary symmetric channel without this assumption when Willie channel is significantly noisier than that of Bob, i.e., when there is a positive secrecy rate over the underlying wiretap channel. Despite the availability of the codebook at Willie, [6] uses the total variation distance as the LPD metric.

In all of the aforementioned works, square-root law was reported to be beaten only in situations where channel statistics are imperfectly known to Willie or when he is unaware with the time in which communication between Alice and Bob takes place. In this paper, the obtained result does not build on the ignorance of the adversary of his own channel or noise statistics, yet, it takes the sole advantage of the spatial dimension of MIMO channel.

**II. SYSTEM MODEL AND PROBLEM STATEMENT**

In the rest of this paper we use boldface uppercase letters for vectors/matrices. Meanwhile, $(\cdot)^*$ denotes conjugate of complex number, $(\cdot)^\dagger$ denotes conjugate transpose, $\mathbf{I}_N$ denotes identity matrix of size $N$, $\text{tr}(\cdot)$ denotes matrix trace operator, $|\mathbf{A}|$ denotes the determinant of matrix $A$ and $\mathbf{I}_{m \times n}$ denotes a $m \times n$ matrix of all 1’s. We say $\mathbf{A} \succeq \mathbf{B}$ when the difference $\mathbf{A} - \mathbf{B}$ is positive semi-definite. The mutual information between two random variables $x$ and $y$ denoted by $I(x;y)$ while $\lim_{n \to \infty}$ denotes the limit inferior. We use the standard order notation $f(n) = O(g(n))$ to denote an upper bound on $f(n)$ that is asymptotically tight, i.e., there exist a constant $m$ and $n_0 > 0$ such that $0 \leq f(n) \leq mg(n)$ for all $n > n_0$.

**A. Communication Model**

We consider the MIMO channel scenario in which a transmitter (Alice) with $N_a \geq 1$ antennas aims to communicate to a receiver (Bob) having $N_b \geq 1$ antennas without being detected by a passive adversary, (Willie) equipped with $N_w \geq 1$ antennas. The discrete baseband equivalent channel for the signal received by each of the legitimate destination, $y$, and the adversary, $z$, are given as follows:

$$
\begin{align*}
\mathbf{y} &= \mathbf{H}_b \mathbf{x} + \mathbf{e}_b,
\mathbf{z} &= \mathbf{H}_w \mathbf{x} + \mathbf{e}_w,
\end{align*}
$$

where $\mathbf{x} \in \mathbb{C}^{N_a \times 1}$ is the transmitted signal vector constrained by an average power constraint $\mathbb{E}[\text{tr}(\mathbf{x}\mathbf{x}^\dagger)] \leq P$. Also, $\mathbf{H}_b \in \mathbb{C}^{N_b \times N_a}$ and $\mathbf{H}_w \in \mathbb{C}^{N_w \times N_a}$ are the channel coefficients matrices between Alice, Bob and Willie respectively. Throughout this paper, unless otherwise noted, $\mathbf{H}_b$ and $\mathbf{H}_w$ are assumed deterministic, also, we assume that $\mathbf{H}_b$ is known to all parties, meanwhile, $\mathbf{H}_w$ is known only to Willie. We define $N \triangleq \min\{N_a, N_b\}$ and $M \triangleq \min\{N_a, N_w\}$. Finally, $\mathbf{e}_b \in \mathbb{C}^{N_b \times 1}$ and $\mathbf{e}_w \in \mathbb{C}^{N_w \times 1}$ are an independent zero mean circular symmetric complex Gaussian random vectors for both destination and adversary channels respectively, where, $\mathbf{e}_b \sim \mathcal{CN}(0, \sigma_b^2 \mathbf{I}_{N_b})$ and $\mathbf{e}_w \sim \mathcal{CN}(0, \sigma_w^2 \mathbf{I}_{N_w})$. 
We further assume $H_w$ lies in the set of matrices with bounded spectral norms:
\[
S_w = \{H_w : \|H_w\|_{op} \leq \sqrt{w}\}
\]
where $\|A\|_{op}$ is the operator (spectral) norm of $A$, i.e., the maximum eigenvalue of $A$. The set $S_w$ incorporates all possible $W_w$ that is less than or equal to $\gamma_w I$ (in positive semidefinite sense) with no restriction on its eigenvectors, where $I$ is a diagonal matrix with the first $M$ diagonal elements equal to 1 while the rest $N_a - M$ elements of the diagonal are zeros. Observe that, $\|W_w\|_{op}$ represents the largest possible power gain of Willie channel. Unless otherwise noted, throughout this paper we will assume that $H_w \in S_w$.

B. Problem Statement

Our objective is to establish the fundamental limits of reliable transmission over Alice to Bob MIMO channel, constrained by the low detection probability at Willie. Scalar AWGN channel channel have been studied in [8], we use a formulation that follow closely the one used therein while taking into consideration the vector nature of the MIMO channel. Alice employ a stochastic encoder with blocklength $n N_a$, where $n$ is the number of channel uses, for message set $M$ consists of:

1) An encoder $M \mapsto C^n N_a$, $m \mapsto x^n$ where $x \in C^{N_a}$.

2) A decoder $C^n N_a \mapsto M$, $y^n \mapsto \hat{m}$ where $y \in C^{N_a}$.

Alice choose a message $M$ from $M$ uniformly at random to transmit to Bob. Let us denote by $H_0$ the null hypothesis under which Alice is not communicating and denote by $P_0$ the probability distribution of Willie’s observation under the null hypothesis. Conversely, let $H_1$ be the true hypothesis under which Alice is transmitting her chosen message $M$ and let $P_1$ be the probability distribution of Willie’s observation under the true hypothesis. Further, define type I error $\alpha$ to be the probability of mistakenly accepting $H_1$ and type II error $\beta$ to be the probability of mistakenly accepting $H_0$. For the optimal hypothesis test generated by Willie we have [11]
\[
\alpha + \beta = 1 - \mathcal{V}(P_0, P_1),
\]
where $\mathcal{V}(P_0, P_1)$ the total variation distance between $P_0$ and $P_1$ and is given by
\[
\mathcal{V}(P_0, P_1) = \frac{1}{2} \|p_0(x) - p_1(x)\|_1,
\]
where $p_0(x)$ and $p_1(x)$ are, respectively, the densities of $P_0$ and $P_1$ and $\|\cdot\|_1$ is the $L_1$ norm. The variation distance between $P_0$ and $P_1$ is related to the Kullback–Leibler Divergence (relative entropy) by the well known Pinsker’s inequality [12]:
\[
\mathcal{V}(P_0, P_1) \leq \sqrt{\frac{1}{2} D(P_0 \parallel P_1)}
\]
where
\[
D(P_0 \parallel P_1) = \mathbb{E}_{P_0} [\log P_0 - \log P_1].
\]

Note that, in $n$ independent channel uses, we have
\[
D(P_0^n \parallel P_1^n) = n D(P_0 \parallel P_1)
\]
by the chain rule of relative entropy. Accordingly, for Alice to insure certain performance of Willie’s optimal detector, she needs to bound $\mathcal{V}(P_0^n, P_1^n)$ above by an $\delta$ of her choice. Consequently, she insures that the sum of error probabilities at Willie is bounded as $\alpha + \beta \geq 1 - \delta$. Using (5), Alice can achieve her goal by designing her signaling strategy (based on the amount of information available) subject to
\[
D(P_0 \parallel P_1) \leq \frac{2\delta^2}{n}.
\]

Throughout this paper, we adopt (8) as our LPD metric. Thus, the input distribution used by Alice to generate the codebook has to satisfy (8). As in [8], our goal is to find the maximum value of $\log |M|$ for which a random codebook of length $n N_a$ exists that satisfies (8) and whose average probability of error is at most $\epsilon$. We will further require $\epsilon$ to be made arbitrary small as $n$ increase. We denote this maximum by $K_n(\delta, \epsilon)$ and we define
\[
L \triangleq \lim_{\epsilon \downarrow 0} \lim_{n \rightarrow \infty} \frac{K_n(\delta, \epsilon)}{\sqrt{2n \delta^2}}.
\]

Note that $L$ has units $\sqrt{n \text{bits}}$. We are interested in the characterization of $L$ under different conditions of Bob and Willie channels in order to derive scaling laws of number of covert bits over MIMO AWGN channel. We first give the following Proposition which provide a general expression for $L$ by extending Theorem 1 in [8] to the MIMO AWGN channel with infinite input and output alphabet.

**Proposition 1.** For the considered MIMO AWGN channel,
\[
L = \max_{\{f_n(x)\}} \lim_{n \rightarrow \infty} \sqrt{\frac{n}{2\delta^2}} \mathcal{I}(f_n(x), f_n(y|x))
\]
\[
\text{Subject to: } D(P_0^n \parallel P_1^n) - 2\delta^2 \leq 0
\]
where $\{f_n(x)\}$ is a sequence of input distributions over $C^{N_a}$ and $\mathbb{E}_{n}[\cdot]$ denotes the expectation with respect to $f_n(x)$.

**Proof.** First, using the encoder/decoder structure described above, we see that the converse part of Theorem 1 in [8] can be directly applied here. Meanwhile, the achievability part there was derived based on the finiteness of input and output alphabet. It was not generalized to the continuous alphabet input over scalar AWGN channel. Rather, the achievability over AWGN channel was shown for Gaussian distributed input in Theorem 5. Here, we argue that, showing achievability for Gaussian distributed input is sufficient and, hence, we give achievability proof in Appendix A that follow closely the proof of Theorem 5 in [8]. The reason is that, Gaussian distribution is indeed optimal for (10) for all $n$. Unlike the non LPD constrained capacity which attains its maximum when the underlying input distribution is zero mean complex Gaussian, it is not straightforward to infer what input distribution is optimal.
However, using the Principle Minimum Relative Entropy [2] and Information Projection [3], we can verify that, the distribution $P_1$ that minimizes $D(P_0 \parallel P_1)$ is the zero mean circular symmetric complex Gaussian distribution whenever $P_0$ follow the same distribution. This, in turn, requires $f_n(x)$ to follow the same distribution for all $n$. Fortunately, it is well known that the zero mean circular symmetric complex Gaussian distribution maximizes the objective function (11) whenever $e_b$ follow the same distribution [13], [14].

Further, we provide a more convenient expression for $L$ in the following Theorem which provides an extension of Corollary 1 in [8] to the MIMO AWGN channel.

**Theorem 1.** For the considered MIMO AWGN channel,

$$L = \lim_{n \to \infty} \sqrt{\frac{n}{2\delta^2}} \max_{f_n(x)} \mathcal{I}(f_n(x), f_n(y|x))$$

Subject to: $D(P_0^n \parallel P_1^n) - 2\delta^2 \leq 0$

where $f_n(x)$ is the input distribution over $\mathbb{C}^{N_x}$ and $\mathbb{E}[\cdot]$ denotes the expectation with respect to $f_n(x)$.

**Proof.** The proof is given in Appendix B.

Now, since we now know that zero mean circular symmetric complex Gaussian input distribution is optimal, the only remaining task is to characterize the covariance matrix, $Q = \mathbb{E}[xx^H]$ of the optimal input distribution. Accordingly, (11) can be rewritten as:

$$L = \lim_{n \to \infty} \sqrt{\frac{n}{2\delta^2}} \max_{Q \succeq 0} \log \left| I_{N_w} + \frac{W_bQ}{\sigma_b^2} \right|$$

Subject to: $D(P_0^n \parallel P_1^n) - 2\delta^2 \leq 0$,

where $W_b = H_b^H H_b$. Further, we can evaluate the relative entropy at Willie as follows:

$$D(P_0 \parallel P_1) = \log \left| \frac{1}{\sigma_w^2} H_w Q H_w^H + I_{N_w} \right| + \text{tr} \left\{ \frac{1}{\sigma_w^2} H_w Q H_w^H + I_{N_w} \right\}^{-1} - N_w.$$  

(13)

In this paper, we are mainly concerned with characterizing $L$ when Alice knows only $H_b$. To that end, let us define:

$$C_{pd}(\delta) = \max_{Q \succeq 0 \atop \text{tr}(Q) \leq P} \log \left| I_{N_w} + \frac{W_bQ}{\sigma_b^2} \right|$$

Subject to: $D(P_0^n \parallel P_1^n) - 2\delta^2 \leq 0$.

(14)

Clearly, $L = \lim_{n \to \infty} \sqrt{\frac{n}{2\delta^2}} C_{pd}(\delta)$. In what follows, we characterize $C_{pd}(\delta)$ and, hence, $L$ under different models of $H_b$ and $H_w$.

### III. Characterization of $C_{pd}(\delta)$

With uncertainty about Willie’s channel, $H_w \in S_w$, it is intuitive to think that Alice should design her signaling strategy against the worst (stronger) possible Willie channel. We first derive the worst case Willie channel, then, we establish the saddle point property of the considered class of channels in the form of $\min \max = \max \min$, where the maximum is taken over all admissible input covariance matrices and the minimum is over all $H_w \in S_w$. Thus, we show that $C_{pd}(\delta)$ equals to the $C_{pd}(\delta)$ evaluated at the worst possible $H_w$.

#### A. Worst Willie Channel and Saddle Point Property

To characterize $C_{pd}(\delta)$ when $H_w \in S_w$, we need first to establish the worst case $C_{pd}(\delta)$ denoted by $C_{pd}^w(\delta)$. Suppose we have obtained $C_{pd}(\delta)$ for every possible state of $H_w$, then, $C_{pd}^w(\delta)$ is the minimum $C_{pd}(\delta)$ over all possible state of $H_w$.

First, let us define

$$\mathcal{R}(W_w, Q, \delta) = \log \left| I_{N_w} + \frac{W_wQ}{\sigma_b^2} \right|.$$  

(15)

We give $C_{pd}^w(\delta)$ in the following proposition.

**Proposition 2.** Consider the class of channels in (2), for any $Q \succeq 0$ satisfies $\text{tr}(Q) \leq P$ and $W_w \in S_w$ we have:

$$C_{pd}^w(\delta) = \min_{W_w \in S_w} \max_{Q \succeq 0 \atop \text{tr}(Q) \leq P} \mathcal{R}(W_w, Q, \delta)$$

Subject to: $D(P_0^n \parallel P_1^n) - 2\delta^2 \leq 0$.

$$= \max_{Q \succeq 0 \atop \text{tr}(Q) \leq P} \mathcal{R}(\gamma, I, Q, \delta)$$

Subject to: $D(P_0^n \parallel P_1^n) - 2\delta^2 \leq 0$  

(16)

i.e., the worst Willie channel is isotropic.

**Proof.** See Appendix B in [15].

Proposition 2 establishes $C_{pd}^w(\delta)$. The following proposition proves that $C_{pd}(\delta) = C_{pd}^w(\delta)$ by establishing the saddle point property of the considered class of channels.

**Proposition 3.** (Saddle Point Property.) Consider the class of channels in (2), for any $Q \succeq 0$ satisfies $\text{tr}(Q) \leq P$ and $W_w \in S_w$ we have:

$$C_{pd}(\delta) = \min_{W_w \in S_w} \max_{Q \succeq 0 \atop \text{tr}(Q) \leq P} \mathcal{R}(W_w, Q, \delta)$$

Subject to: $D(P_0^n \parallel P_1^n) - 2\delta^2 \leq 0$.

$$= \max_{Q \succeq 0 \atop \text{tr}(Q) \leq P} \mathcal{R}(W_w, Q, \delta)$$

Subject to: $D(P_0^n \parallel P_1^n) - 2\delta^2 \leq 0$.

(17)

**Proof.** By realizing that, for any feasible $Q$, the function $D(P_0 \parallel P_1(W_w))$ is monotonically increasing in $W_w$, we
have that
\[
\min_{W_w \in S_w} \mathcal{R}(W_w, Q, \delta) = \mathcal{R}(\gamma_w I, Q, \delta)
\]
Subject to: \(D(P^n_0 \parallel P^n_1) - 2\delta^2.\) (18)
Hence, the required result follows by using proposition 2. ■

B. Evaluation of \(C_{pd}(\delta)\)

In light of the saddle point property established in the previous section, in this section we characterize \(C_{pd}(\delta)\) by solving (17) for the optimal signaling strategy, \(Q^\star\). We give the main result of this section in the following theorem.

**Theorem 2.** The eigenvalue decomposition of the capacity achieving input covariance matrix that solves (14) is given by \(Q^\star = U_b A L^\dagger_b\) where \(U_b \in \mathbb{C}^{N_a \times N_a}\) is the matrix whose columns are the right singular vectors of \(H_b\) and \(A\) is a diagonal matrix whose diagonal entries, \(\lambda_{ii}\), are given by the solution of
\[
\lambda = (\sigma^2_b \lambda^{-1}_i (W_b) + \Lambda_{ii})^{-1}
\]
\[+ \eta \left( \left( \frac{\sigma^2_w}{\gamma_w} + \Lambda_{ii} \right)^{-2} - \left( \frac{\sigma^2_w}{\gamma_w} + \Lambda_{ii} \right)^{-1} \right) \] (19)

where \(\lambda\) and \(\eta\) are constants determined from the constraints \(\text{tr} \{Q\} \leq P\) and (8), respectively. Moreover,
\[
C_{pd}(\delta) = \sum_{i=1}^{N} \log \left( 1 + \frac{\lambda_i(W_b)}{\sigma^2_b \gamma_w} \right) \] (20)

where \(\lambda_i\) is the \(i\)th non-zero eigenvalue of \(W_b\).

**Proof.** See Appendix C in [15]. ■

The result of Theorem 2 provides the full characterization of \(C_{pd}(\delta)\) of the considered class of channels. It can be seen that, the singular value decomposition (SVD) preceding [13] is the optimal signaling strategy except for the water filling strategy in (19) which is chosen to satisfy both power and LPD constraints. Unlike both MIMO channel without security constraint and MIMO wiretap channel, transmission with full power is, indeed, not optimal. Let
\[
P_{th} = \sum_{i} \lambda_{ii},\] (21)

be the maximum total power that is transmitted by Alice. An equivalent visualization of our problem is that Alice need to choose a certain power threshold, \(P_{th}\), to satisfy the LPD constraint. However, again, \(P_{th}\) is distributed along the eigenvectors using conventional water filling solution. Although it is not straightforward to obtain a closed form expression for \(C_{pd}(\delta)\) and, hence, \(L\), we could obtain both upper and lower bounds on \(C_{pd}(\delta)\) which leads to upper and lower bounds on \(L\). Based on the obtained bounds, we gives the square-root law for MIMO AWGN channel in the following Theorem.

**Theorem 3** (Square-root Law of MIMO AWGN channel). For the considered class of channels, the following bounds on \(C_{pd}(\delta)\) holds
\[
\sum_{i=1}^{N} \log \left( 1 + \frac{\sqrt{2} \sigma^2_w \delta \lambda_i(W_b)}{\sigma^2_b \gamma_w \sqrt{4M}} \right) \leq C_{pd}(\delta) \leq \sum_{i=1}^{N} \log \left( 1 + \frac{\sqrt{2} \sigma^2_w \xi \delta \lambda_i(W_b)}{\sigma^2_b \gamma_w \sqrt{4M}} \right) \] (22)

where \(\xi \geq 1\) is a function of \(\delta\) that approaches 1 as \(\delta\) goes to 0. Moreover,
\[
\sum_{i=1}^{N} \sigma^2_b \lambda_i(W_b) \geq L \leq \sum_{i=1}^{N} \sigma^2_b \lambda_i(W_b) \] (23)

Accordingly, Alice can transmit a maximum of \(O(N \sqrt{n/M})\) bits reliably to Bob in \(n\) independent channel uses while keeping Willie’s sum of error probabilities lower bounded by 1 – \(\delta\).

**Proof.** We give both achievability and converse results in Appendix D in [15]. ■

Theorem 3 extends the square-root law for scalar AWGN channel to the MIMO AWGN channel. The interesting result here is that, the gain in covert rate scales linearly with number of active eigenmodes of Bob channel. Meanwhile, it scales down with the square-root of the number of active eigenmodes over Willie channel. This fact is of great importance in case of massive MIMO limit. Further, the bounds on \(L\) in (23) can, with small effort, generate the result of Theorem 5 in [8] by setting \(N = M = 1\), \(\lambda(W_b) = \gamma_w\) and \(\sigma_b = \sigma_w\).

It worth mentioning that, in some practical situations, compound MIMO channel can be too conservative for resource allocation. In particular, the bounded spectral norm condition in (2) not only leads us to the worst case Willie channel, but it also does not restrict its eigenvectors leaving the beamforming strategy used by Alice (SVD precoding) to be of insignificant gain in protecting against Willie. Although we believe that the eigenvectors of Willie channel plays an important role in the determination of the achievable covert rate, the ignorance of Alice about Willie channel leaves the compound framework as our best option.

IV. UNIT RANK MIMO CHANNEL

As pointed out in the previous section, the distinction between the eigenvectors of Bob and Willie channels would have a considerable effect on the achievable covert rate. However, unavailability of Willie’s CSI left the compound framework as the best model for \(H_w\). In this section, we consider the case when either \(H_w\) or both \(H_w\) and \(H_b\) are of unit rank. This scenario not only models the case when both Bob and Willie have a single antenna, but it also models the case when they have a strong line of sight with Alice. Moreover, this scenario allows us to evaluate the effect of the eigenvectors of \(H_w\) and \(H_b\) on the achievable covert rate.

---

1 Using Mathematica, \(\Lambda_{ii}\) was found to be an expression of almost 30 lines which does not provide the required insights here.
A. Unit Rank Willie Channel

In this section we analyze the scenario in which only Willie channel is of unit rank. In this case, we can write \( H_w = \lambda_w^{1/2} v_w u_w^* \), where \( v_w \in \mathbb{C}^{N_w} \) and \( u_w \in \mathbb{C}^{N_w} \) are the left and right singular vectors of \( H_w \). Accordingly, \( W_w = \lambda_w u_w u_w^* \) and the product \( W_w Q \) has only one non-zero eigenvalue. The nonzero eigenvalue \( \lambda \) is of unit rank (which is the case when \( \lambda \) has only one non-zero eigenvalue. In this section we analyze the scenario in which only Willie channel. Accordingly, \( \lambda \) is the achievable covert rate increases linearly by \( \lambda \). This fact means that, the achievable covert rate increases linearly by \( \lambda \). Accordingly, \( \lambda \) is the non LPD constrained capacity of Alice to Bob channel. That is because, unlike the scenario of this case. That is due to the technical difficulty in setting tight upper bound on the power received by Willie when Bob channel has higher rank. Also, we see that unit rank channel offers better covert rate than that shown under the compound settings for Willie channel. That is because, unlike the scenario of this section, compound settings does not restrict the eigenvectors of Willie’s channel.

V. COVERT COMMUNICATION WITH MASSIVE MIMO

In Theorem 4, it was shown that Alice can communicate at full rate with Bob without being detected by Willie whenever \( \cos(\theta) = 0 \). In this section, we study the behavior of covert rate as the number of antennas scale which we call the massive MIMO limit. In particular, the high beamforming capability of massive MIMO system can provide substantial gain in the achievable LPD rate. However, a quantitative relation between the achievable LPD rate and the number of transmitting antennas seems to be unavailable. More precisely, how does the achievable LPD rate scale with the number of transmitting antennas? Before we answer this question, we state some necessary basic results on the inner product of unit vectors in higher dimensions [16].

A. Inner Product of Unit Vectors in Higher Dimensions

In this section, we reproduce some established results on the inner product of unit vectors in higher dimensions.

Lemma 1. [Proposition 1 in [16]] Let \( a \) and \( b \) any two vectors in the unit sphere in \( \mathbb{C}^p \) chosen uniformly at random. Let \( \theta = \cos^{-1}(\langle a, b \rangle) \) be the angle between them. Then

\[
\Pr \left( \left| \theta - \frac{\pi}{2} \right| \leq \zeta \right) \geq 1 - K \sqrt{p} (\cos \zeta)^{p-2}
\]
for all $p \geq 2$ and $\zeta \in \left(0, \frac{\pi}{2}\right)$ where $K$ is a universal constant.

The statement of Lemma 1 states that, the probability that any two vectors chosen uniformly at random being orthogonal increases exponentially fast with the dimension $p$. In particular, for small $\zeta$ we have $\cos \zeta$ close to 1. Hence, it can be stated that, in the higher dimension limit, $p \to \infty$, every two vectors chosen uniformly at random are nearly orthogonal.

Corollary 1. Let $a$ and $b$ any two vectors in the unit sphere in $\mathbb{C}^p$ chosen uniformly at random and let $\theta = \cos^{-1}(\langle a, b \rangle)$ be the angle between them. Let $A, B \in \mathbb{C}^{p \times p}$ be two matrices of unit rank generated as $A = \lambda_a a a^\dagger$ and $B = \lambda_b b b^\dagger$. Then, the eigenvalue of the product $\lambda(AB)$ approaches $0$ exponentially fast with the dimension $p$.

**Proof.** It can be easily verified that $\lambda(AB) = \lambda_a \lambda_b \cos \theta$. Using Lemma 1, we see that, the probability that $\theta$ approaches $\pi/2$ increases exponentially with $p$. Hence, the probability that $\cos \theta$ approaches $0$ increases in the same order. Then so is $\cos^2 \theta$. \hfill \blacksquare

B. Massive MIMO Limit

In the previous section it was demonstrated that, in higher dimensions every two independent vectors chosen uniformly at random are orthogonal with very high probability. More generally, using spherical invariance, given $u_b$, for any $u_w$ chosen uniformly at random in $\mathbb{C}^{N_a}$, the result of Lemma 1 still holds. This scenario typically models the scenario when Alice knows her channel to Bob, meanwhile, she models $u_w$ as a uniform random unit vector. Of course, the number of antennas at Alice is a physical resource which can not be compared to $n$ that can approach $\infty$ very fast. The more interesting question is, how fast $\cos^2 \theta$ approaches $0$ as $N_a$ increase. As illustrated in corollary 1, we know that $\cos \theta$ approaches $0$ exponentially fast with $N_a$. Consequently, we conclude that $\cos^2 \theta$, also, approaches $0$ exponentially fast with $N_a$. For proper handling of the scaling of $K_n(\delta, \epsilon)$ in massive MIMO limit, let us define

$$S = \lim_{n,N_a \to \infty} N_a^2 \sqrt{2 \pi} \Pr(C_{pd}(\delta) = C).$$

Observe that, following Proposition 1 and Theorem 1, we can show that

$$S = \lim_{n,N_a \to \infty} N_a \sqrt{\frac{n}{2 \pi} C_{pd}(\delta)}. \quad (31)$$

We give the result of this scenario in the following Theorem.

**Theorem 5.** Assume that $\text{rank} \{H_b\} = \text{rank} \{H_c\} = 1$. Given $u_b$, for any $u_w$ chosen uniformly at random, we have $C_{pd}(\delta)$ is as given in Theorem 4 and

$$S = \infty. \quad (32)$$

Moreover, $K_n$ grows like $2^g(N_a-2)/2 \sqrt{\frac{n}{K^2 N_a}}$ where $K$ is a universal constant and $g = \left(\sqrt{2 \pi} \delta \sqrt{\frac{n}{N_a}} P \right).$

**Proof.** Combining the result of Theorem 4 and Corollary 1, multiplying (27) by $N_a \sqrt{\frac{n}{2 \pi} C_{pd}(\delta)}$ and taking the limit as both of $n$ and $N_a$ tend to infinity we obtain

$$\lim_{n \to \infty} \lim_{N_a \to \infty} N_a \sqrt{\frac{n}{2 \pi} C_{pd}(\delta)} = \lim_{N_a \to \infty} N_a \frac{\delta^2 \lambda_b}{\sigma_b^2 \lambda_w \cos^2(\theta)} = \infty,$$

where the last equality follow since $\cos^2(\theta) \to 0$ as $N_a \to \infty$. On the other hand, we also can verify that

$$\lim_{n \to \infty} \lim_{N_a \to \infty} N_a \sqrt{\frac{n}{2 \pi} C_{pd}(\delta)} = \infty. \quad (33)$$

To show how does $K_n$ scale in this massive MIMO limit, we first note that, for fixed $N_a$, $K_n$ scales like $\sqrt{N_a}$ as well. Also note that, $S = \infty$ implies that $K_n$ scales faster than $N_a$. Even though, it does not provide how does it scale with $N_a$. Note that, Alice can fully utilize her channel while keeping the LPD constraint (8) satisfied when the quantity $C_{pd}(\delta) = C$. Thus, we show the scaling of $C_{pd}(\delta)$ up to $C$. Following the proof of Theorem 3, we can obtain the following bound on $P_{th}$:

$$P_{th} \leq \min \left\{ \frac{\sqrt{2 \pi} \delta}{\sqrt{N_a} \cos^2(\theta)} P \right\}, \quad (35)$$

then, we see that $P_{th} = P$, and hence $C_{pd}(\delta) = C$, when

$$P \leq \frac{\sqrt{2 \pi} \delta}{\sqrt{N_a} \cos^2(\theta)} \quad (36)$$

equivalently,

$$\cos^2(\theta) \leq \frac{\sqrt{2 \pi} \delta}{\sqrt{N_a} \cos^2(\theta)} \quad (37)$$

de, and hence,

$$\Pr(C_{pd}(\delta) = C) \geq 1 - K \sqrt{\frac{N_a}{(1-g)(N_a-2)/2}} \geq 1 - K \sqrt{\frac{N_a}{2^{-g(N_a-2)/2}}} \quad (38)$$

where $g = \left(\sqrt{2 \pi} \delta \sqrt{\frac{n}{N_a}} P \right)$. Note that, the first inequality in (38) follows by setting $\zeta$ in Lemma 1 equal to the RHS of (37) and using $\cos(\pi/2 - x) = \sin(x)$ and $\sin(\cos^{-1}(x)) = \sqrt{1-x^2}$. Meanwhile, the second inequality follows by realizing that [17]

$$(1-g)^{N_a-2}/2 \leq 2^{-g(N_a-2)/2} \quad (39)$$

It can be seen that, the probability that $C_{pd}(\delta) = C$ scales as $2^{g(N_a-2)/2}/K \sqrt{N_a}$ up to 1. \hfill \blacksquare

Theorem 5 states that Alice can communicate at full rate to Bob while satisfying the LPD constraint (8). Note that, the limit in both orders yields $S = \infty$. 

---

**2017 IEEE Conference on Communications and Network Security (CNS)**
As $N_a \to \infty$, the radiation pattern of a wireless MIMO transmitter becomes extremely directive (pencil beam), we can call this the wired limit of wireless MIMO communication. In this wired limit, Willie can not detect Alice transmission unless he wiretapped this virtual wire. The result of Theorem 5 provides a rigorous quantitative characterization to the intuitive wired limit of wireless MIMO communication. In principle, it answer the fundamental question: How fast the LPD constrained rate increase with number of antennas at Alice? It can be seen that the probability that Alice fully utilizes the channel scales like $2^{g(N_a - 2)/2}/K\sqrt{N_a}$ up to 1 using the same justification given after Lemma 1. In fact, it can be directly deduced from Theorems 4 and 5 that, in the limit of large $N_a$ Alice can transmit $O(n)$ bits in $n$ independent channel uses while satisfying the LPD constraint. The following numerical example demonstrates the covert rates in massive MIMO limit with and without a shared secret between Alice and Bob.

**Example 1.** Assume that Alice intend to use the channel for $n = 10^3$ times over a channel of bandwidth of $10^3$ Hz, hence, $\sqrt{n} = 3.1623 \times 10^4$. Suppose that Alice is targeting $\epsilon = 10^{-2}$. Let $\sigma^2_w = \sigma^2_b = 10^{-2}$ and $\lambda_w = \lambda_b = 10^{-3}$. Assume that Alice is targeting $SNR = 15dB$ at Bob, hence, $P = 316.228$. Then, for $N_a = 100$ it can be verified that, Alice can transmit $O(n)$ covert bits instead of $O(\sqrt{n})$. Observe that, Alice needed only $N_a = 100$ to communicate covertly at near full rate to Bob. Also note that, at $6GHz$, two dimensional array of 100 elements can fit within an area of a single sheet of paper. See Fig. (1) for the relation between the $\epsilon$-PD capacity and number of transmitting antenna for different values of number of antennas at Willie with $\epsilon = 10^{-2}$.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{The relation between achievable covert rate (plotted in log scale), in bits per second, and $N_a$ for different values of $N_w$, with target $\epsilon = 10^{-2}$. It shows that Alice can communicate near full rate with $N_a$ around 100.}
\end{figure}

**VI. CONCLUSION**

We have established the limits of LPD communication over the MIMO AWGN channel. In particular, using relative entropy as our LPD metric, we studied the maximum codebook size, $K_n(\delta, \epsilon)$, for which Alice can guarantee reliability and LPD conditions are met. We first showed that, the optimal codebook generating input distribution under $\delta$-PD constraint is the zero-mean Gaussian distribution. We based our arguments on the the principle of minimum relative entropy. For an isotropic Willie channel, we showed that Alice can transmit $O(N\sqrt{n}/M)$ bits reliably in $n$ independent channel uses, where $N$ and $M$ are the number of active eigenmodes of Bob and Willie channels, respectively. Further, we evaluated the scaling rates of $K_n(\delta, \epsilon)$ in the limiting regimes for the number of channel uses (asymptotic block length) and the number of antennas (massive MIMO). We showed that, while the square-root law still holds for the MIMO AWGN, the number of bits that can be transmitted covertly scales exponentially with the number of transmitting antennas. More precisely, for a unit rank MIMO channel, we show that $K_n(\delta, \epsilon)$ scales as $2^{g(N_a - 2)/2}\sqrt{\frac{n}{K^2N_a}}$.

The practical implication of our result is that, MIMO has the potential to provide a substantial increase in the file sizes that can be covertly communicated subject to a reasonably low delay.

**REFERENCES**

[1] B. A. Bash, D. Goeckel, and D. Towsley, “Limits of reliable communication with low probability of detection on awgn channels,” IEEE Journal on Selected Areas in Communications, vol. 31, no. 9, pp. 1921–1930, 2013.

[2] A. D. Woodbury, “Minimum relative entropy, bayes and kapur,” Geophysical Journal International, vol. 185, no. 1, pp. 181–189, 2011.

[3] I. Csiszár and F. Matus, “Information projections revisited,” IEEE Transactions on Information Theory, vol. 49, no. 6, pp. 1474–1490, 2003.

[4] R. F. Schaefer and S. Loyka, “The secrecy capacity of compound gaussian mimo wiretap channels,” IEEE Transactions on Information Theory, vol. 61, no. 10, pp. 5535–5552, 2015.

[5] A. Abdelaziz, A. Elbeyouni, C. E. Koksal, and H. El Gamal, “On the compound MIMO wiretap channel with mean feedback,” in 2017 IEEE International Symposium on Information Theory (ISIT) (ISIT’2017), Aachen, Germany, Jun. 2017.

[6] P. H. Che, M. Bakshi, and S. Jaggi, “Reliable deniable communication: Hiding messages in noise,” in Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on. IEEE, 2013, pp. 2945–2949.

[7] M. R. Bloch, “Covert communication over noisy channels: A resolvability perspective,” IEEE Transactions on Information Theory, vol. 62, no. 5, pp. 2334–2354, 2016.

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

[9] A. O. HERO, “Secure space-time communication,” IEEE Transactions on Information Theory, vol. 49, no. 12, pp. 3235–3249, 2003.

[10] S. Lee, R. J. Baxley, J. B. McMahon, and R. S. Frazier, “Achieving positive rate with undetectable communication over mimo rayleigh channels,” in Sensor Array and Multichannel Signal Processing Workshop (SAM), 2014 IEEE 8th. IEEE, 2014, pp. 257–260.

[11] E. L. Lehmann and J. P. Romano, Testing statistical hypotheses. Springer Science & Business Media, 2006.

[12] T. M. Cover and J. A. Thomas, Elements of information theory. John Wiley & Sons, 2012.

[13] E. Telatar, “Capacity of multi-antenna gaussian channels,” European transactions on telecommunications, vol. 10, no. 6, pp. 585–595, 1999.

[14] D. Tse and P. Viswanath, Fundamentals of wireless communication. Cambridge university press, 2005.

[15] A. Abdelaziz and C. E. Koksal, “Fundamental limits of covert communication over mimo awgn channel,” arXiv preprint arXiv:1705.02303, 2017.
Accordingly, of a random vector denotes the $n$th extension of the probability density function of a random vector $x$. Thus, we need to show

$$\frac{1}{\sqrt{n}} \log \frac{f^{\infty}(z^n|\Sigma)}{f^{\infty}(z^n)} = -\sqrt{n} I(f_n(x), f_n(z|x)) \to 0$$

in probability as $n$ tends to infinity. First observe that,

$$f^{\infty}(z^n|\Sigma) = \prod_{i=1}^{n} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (z_i - x_i)^\top \Sigma^{-1} (z_i - x_i) \right\},$$

$$f^{\infty}(z^n) = \prod_{i=1}^{n} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} z_i^\top \Sigma^{-1} z_i \right\},$$

where $\Sigma_0 = \sigma_0^2 I_{N_w}$, $\Sigma_1 = H_q Q H_q^\top + \sigma_0^2 I_{N_w}$ and let $Q = \mathbb{E} [xx]\}$.\} be chosen such that $\log |\Sigma| \geq 0$ a decreasing function of $n$. Then,

$$\frac{f^{\infty}(z^n|\Sigma)}{f^{\infty}(z^n)} = |\Sigma_1^{-\frac{1}{2}}| \times \exp \left\{ \sum_{i=1}^{n} \text{tr}(\Sigma_1^{-1} z_i z_i^\top) - \sum_{i=1}^{n} \text{tr}(\Sigma_0^{-1} e_i e_i^\top) \right\}.$$ (43)

Accordingly,

$$\frac{1}{\sqrt{n}} \log \frac{f^{\infty}(z^n|\Sigma)}{f^{\infty}(z^n)} = -\sqrt{n} \log |\Sigma_1^{-\frac{1}{2}}| + \frac{1}{\sqrt{n}} \left( \sum_{i=1}^{n} \text{tr}(\Sigma_1^{-1} z_i z_i^\top) - \sum_{i=1}^{n} \text{tr}(\Sigma_0^{-1} e_i e_i^\top) \right),$$ (44)

whose expectation can be found as

$$\mathbb{E} \left[ \frac{1}{\sqrt{n}} \log \frac{f^{\infty}(z^n|\Sigma)}{f^{\infty}(z^n)} \right] = -\sqrt{n} \log |\Sigma_1^{-\frac{1}{2}}| + \frac{1}{\sqrt{n}} \left( \sum_{i=1}^{n} \text{tr}(\Sigma_1^{-1} \Sigma_1) - \sum_{i=1}^{n} \text{tr}(\Sigma_0^{-1} \Sigma_0) \right).$$ (45)

It then follows by Chebyshev’s inequality that, for any constant $a > 0$,

$$\Pr \left( \frac{1}{\sqrt{n}} \log \frac{f^{\infty}(z^n|\Sigma)}{f^{\infty}(z^n)} - \sqrt{n} I(f_n(x), f_n(z|x)) \geq a \right) \leq \frac{1}{a^2} \text{var} \left( \frac{1}{\sqrt{n}} \log \frac{f^{\infty}(z^n|\Sigma)}{f^{\infty}(z^n)} \right)$$ (46)

and, it remains to show that

$$\lim_{n \to \infty} \text{var} \left( \frac{1}{\sqrt{n}} \log \frac{f^{\infty}(z^n|\Sigma)}{f^{\infty}(z^n)} \right) = 0.$$ (47)

Note that,

$$\text{var} \left( \frac{1}{\sqrt{n}} \log \frac{f^{\infty}(z^n|\Sigma)}{f^{\infty}(z^n)} \right) = \frac{1}{n} \sum_{i=1}^{n} \text{var} \left( z_i^\top \Sigma_1^{-1} z_i - e_i^\top \Sigma_0^{-1} e_i \right)$$

$$= \text{var} \left( z_i^\top \Sigma_1^{-1} z_i - e_i^\top \Sigma_0^{-1} e_i \right)$$

$$= \mathbb{E} \left[ (z_i^\top \Sigma_1^{-1} z_i - e_i^\top \Sigma_0^{-1} e_i) (z_i^\top \Sigma_1^{-1} z_i - e_i^\top \Sigma_0^{-1} e_i)^\top \right]$$

$$= \mathbb{E} \left[ z_i^\top \Sigma_1^{-1} z_i - e_i^\top \Sigma_0^{-1} e_i \right]$$

$$- \text{var} \left( z_i^\top \Sigma_1^{-1} z_i - e_i^\top \Sigma_0^{-1} e_i \right).$$ (48)

Now observe that, since $Q \to 0$ as $n$ tends to infinity, we can verify that each term in (48) tends to zero as $n$ tends to infinity.

\section*{Appendix B

Proof of Theorem 1

We show that the limit in (11) always exists. Note that, for every $n$, $f_n(x)$ is zero mean Gaussian. Let $Q_n = \mathbb{E}_n [xx]\}$ be $Q_{Q_n} = \arg \max_{Q \succeq 0} I(f_n(x), f_n(y|x))$\}$ subject to $\text{tr}(Q) \leq P$ where the maximum is subject to (8). Hence, we have

$$\max_{\text{tr}(Q) \leq P} I(f_n(x), f_n(y|x)) = \log \left| I + \frac{H_q Q H_q^\top}{\sigma_b^2} \right|.$$ (50)

Now, we have two cases to consider:

1. $D(P_0 \parallel P_1) = 0$. In this case, $\delta$ can be made 0 causing the limit to be infinity.
2. $D(P_0 \parallel P_1) > 0$. In this case, $Q_{Q_n}$ has to be a decreasing function of $n$, otherwise, the constraint (8) can not be met. In this case, the limit is $\geq 0$ and $< \infty$.

In either cases, the limit exist and, hence, limit can be used in place of limit inferior and, also, the order of limit and maximum can be interchanged.