One-shot achievability and converse bounds of Gaussian random coding in AWGN channels under covert constraint

Xinchun Yu*, Shuangqing Wei† and Yuan Luo*

*School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China
†School of Electrical Engineering and Computer Science, Louisiana State University, Baton Rouge, USA
Emails: moonyuyu@sjtu.edu.cn, swei@lsu.edu, yuanluo@sjtu.edu.cn

Abstract—The achievability and converse bounds on the throughput of covert communication over AWGN channels are investigated in this paper. By re-visiting [8], several new achievability bounds on maximal and average probability of error based on random coding scheme are presented, which leads to results on achievability bounds when the codewords are generated from Gaussian distribution and then selected from a subset under maximal power constraint. The bounds provide us the framework for analyzing the maximal throughput under covert constraint of total variation distance.

I. INTRODUCTION

Covert communication has earned much attention in recent years where the adversary should have a low probability of detection (LPD) of the transmitted message. The information theory for covert communication was first characterized on AWGN channels in [1] and DMCs in [2] [3], and later in [4] and [5] on BSC and MIMO AWGN channels, respectively. It has been shown that covert communication follows from the following square root law [1]. A consequence of SRL is that the average power per channel use should decrease as \( n \) increases and the asymptotic capacity is zero. To characterize the throughput more accurately, it is necessary to consider the second-order asymptotic. In [6] [7], the first and second order asymptotics are investigated over discrete memoryless channels when the discrimination metrics are relative entropy, total variation distance (variational distance) and missed detection probability at a fixed significance level, respectively.

In this work, we pose a problem which is a sub-problem of covert communication in non-asymptotic scenario: as title suggests with the model, if Gaussian codewords are utilized under maximal power constraint \( P(n) \) which is a decreasing function of the blocklength, how much throughput shall we expect? We are interested in Gaussian codewords because the total variation distance at the adversary is relatively easy to manipulate in this circumstance. There are no existing general framework in the literature to utilize. In [8], the achievability bound over AWGN channel is obtained from a codebook on the surface of \( n \)-dimensional sphere with radius \( \sqrt{n}P \) and \( P \) is fixed. The information density function is the same for all the codewords, while it is not the case of Gaussian codebook. Moreover, if a codebook is generated directly from Gaussian distribution, there will be no guarantee that the power constraint is satisfied in the finite block regime. The dilemma impels us to go through the work [8]. To get the bounds, the techniques of random generation, picking from a special subset, and composite hypothesis testing are integrated and utilized. The contributions of our work are listed as follows:

- New achievability bounds (with both maximal probability of error and average probability of error) are obtained by revising techniques introduced in [8] for cases with random coding and input constraints. These bounds are useful in the preceding analysis.
- New converse and achievability bounds are found on the channel coding rate when the codewords are generated from Gaussian distribution and selected from a set with maximal power constraint.

Due to space limitations, some of the proofs are omitted and can be found in an extended version of this paper [9].
II. THE MODEL OF COVERT COMMUNICATION OVER AWGN CHANNEL

An \((n, 2nR)\) code for the Gaussian covert communication channel consists of a message set \(W \in \mathcal{W} = \{1, \ldots, 2nR\}\), an encoder at the transmitter Alice \(f_n : \mathcal{W} \rightarrow \mathbb{R}^n, w \mapsto x^n\), and a decoder at the legitimate user Bob \(g_n : \mathbb{R}^n \rightarrow \mathcal{W}, y^n \mapsto \hat{w}\). Meanwhile, a decoder is at an adversary Willie \(h_n : \mathbb{R}^n \rightarrow \{0, 1\}, z^n \mapsto 0/1\). The error probability of the code is defined as \(P_e^n = Pr[g_n(f_n(W)) \neq W]\). The channel model is defined by

\[
\begin{align*}
y_i &= x_i + N_{B,i}, i = 1, \ldots, n \\
z_i &= x_i + N_{W,i}, i = 1, \ldots, n
\end{align*}
\]  

where \(x^n = \{x_i\}_{i=1}^n, y^n = \{y_i\}_{i=1}^n, z^n = \{z_i\}_{i=1}^n\) denote Alice’s input codeword, the legitimate user Bob’s observation and the adversary Willie’s observation, respectively. \(N_{B,i}, i = 1, \ldots, n,\) is independent identically distributed (i.i.d) according to \(\mathcal{N}(0, \sigma_B^2)\). The quantity \(N_{W,i}, i = 1, \ldots, n,\) is independent of \(N_{B,i}\) and is i.i.d according to \(\mathcal{N}(0, \sigma_W^2)\). For convenience, it is assumed that \(\sigma_B^2 = \sigma_W^2 = 1\). Each codeword is randomly selected from a subset of candidate codewords. Each coordinate of these candidates is i.i.d generated from \(\mathcal{N}(0, P(n))\). The detail of selection will be discussed later. The adversary is aware of the fact that the codebook is generated from Gaussian distribution \(\mathcal{N}(0, P(n))\) with blocklength \(n\) but he don’t know the specific codebook. The adversary Willie tries to determine whether Alice is communicating \((h_n = 1)\) or not \((h_n = 0)\) by statistical hypothesis test. Alice, who is active about her choice, is obligated to seek for a code such that \(\lim_{n \rightarrow \infty} P_e^n \rightarrow 0\) and \(\lim_{n \rightarrow \infty} P(h_n = 0) \rightarrow \frac{1}{2}\).

The hypothesis test of Willie in covert communication is performed on his received signal \(z^n\) which is a sample of random vector \(Z^n\). The null hypothesis \(H_0\) corresponds to the situation that Alice doesn’t transmit and consequently \(Z^n\) has output probability distribution \(P_0\). Otherwise, the received vector \(Z^n\) has output probability distribution \(P_1\) which depends on the input distribution. The rejection of \(H_0\) when it is true will lead to a false alarm with probability \(\alpha\). The acceptance of \(H_0\) when it is false is considered to be a miss detection with probability \(\beta\). The aim of Alice is to decrease the success probability of Willie’s test by increasing \(\alpha + \beta\), and meanwhile obtain reliable communication with Bob. The effect of the optimal test is usually measured by the total variation distance (TVD) \(V_T(P_1, P_0)\) which is \(1 - (\alpha + \beta)\) [10]. The total variation distance between two probability measures \(P\) and \(Q\) on a sigma-algebra \(\mathcal{F}\) of subsets of the sample space \(\Omega\) is defined as

\[
V_T(P, Q) = \sup_{A \in \mathcal{F}} |P(A) - Q(A)|.
\]  

In literature, KL divergence, rather than TVD, has often by adopted as a metric [1] [3]. In addition, KL divergence between two Gaussian distributions, subject to presence or absence of a Gaussian codeword from Alice, have been employed to quantify covertness of Alice’s transmission schemes. Our interest has been to find both achievability and converse bounds in the finite block regime with any given \(n\) under a constraint of an imposed upper-bound on TVD between two Gaussian distributions. As a result of such bound on TVD, the transmission power \(P(n)\) of Alice is a function of \(n\). Next, we introduce the framework to attain such one-shot bounds.

III. GENERAL RESULTS ON ACHIEVABILITY BOUND UNDER MAXIMAL POWER CONSTRAINT

A. Preliminary

In this section, we introduce the notions. Let two sets \(A\) and \(B\) be input and output sets of a communication system with conditional probability measure \(P_{Y|X} : A \rightarrow B\). A codebook is a set \((c_1, \ldots, c_2) \in A^M\). An encoder is a function from \([M] = \{1, \ldots, M\} \rightarrow A^M; W \mapsto c_W\) and the decoder is defined as \(P_{W|Y} : B \rightarrow \{0, 1, \ldots, M\}\) (‘0’ indicates “error”) where \(\hat{W}\) is a random variable represents the index \(\hat{W}\). There are two kinds of metric for the error probability to judge the quality of a code, i.e., average error probability and maximal error probability, which are defined as follows:

\[
i. \quad P_e \triangleq \mathbb{P} \left[ W \neq \hat{W} \right].
\]

\[
ii. \quad P_{e,\max} \triangleq \max_{m \in [M]} \mathbb{P} \left[ \hat{W} \neq m | W = m \right].
\]

When a codebook and its decoder satisfy the condition \(i\) (\(ii\)), they are called an \((M, c)\) code with average error probability (maximal error probability).

B. Achievability Bounds in General Setting

As our codewords are generated from Gaussian distribution, we should rely on the achievability bound of random coding scheme under a given input distribution \(F_X\). However, the existing results in the literature, such as the bounds provided in (108) and (127) in [8] cannot be directly employed in our scenario because Gaussian codewords will bring us much complexity. In order to get some available achievable results for our scenario, it is necessary to revisit and go through Yury’s result (mainly Theorem 21 in [8]) to make it applicable for our case. Specifically, our achievable results are based on Part C of Section III in [8]. The code is randomly and sequentially constructed by showing there exists some codeword in a set whose error probability is small by dependence testing in every step from the fact that the error probability is small in average of the available choices. The decoding procedure of the code is determined as sequential dependence testing. Our results are based on combined application of the following elements:

- The codebooks are randomly chosen from a set \(\mathcal{F}\) which is a subset induced by some constraint on the whole space \(A\).
- For each codeword \(x\), there is an associated threshold \(\gamma(x)\) in the dependence testing.

Lemma 1. For any distribution \(P_X\) on \(A\), and any measurable function \(\gamma : A \rightarrow [0, \infty]\), there exists a code with
$M$ codewords in the set $F$ whose maximal error probability satisfies
\[
\epsilon_P[X] \leq E[X[P(i(x; Y) \leq \log \gamma(x)) \cdot 1_{\{x \in F\}}]
+(M-1)\sup_{x \in F}P_Y[i(x; Y) \leq \log \gamma(x)]
\]
(6)

**Proof.** The operation of the decoder is the following sequential decoding process. It computes $i(c_j; y)$ for the received channel output $y$ and selects the first codeword $c_j$ which satisfies $i(c_j; y) > \log \gamma(c_j)$. For the first codeword, the conditional probability of error under the decoding rule is
\[
\epsilon_1(x) = P[i(x; Y) \leq \log \gamma(x) | X = x]
\]
(7)

once the codeword $x$ is chosen. Since $x$ is chosen from $F$, we have
\[
\epsilon_1(c_1) \leq E[\epsilon_1(x) | F]
= \sum_{x \in F} P[i(x; Y) \leq \log \gamma(x) | X = x] P(X = x | F)
= \sum_{x \in F} P(X = x, x \in F) P[i(x; Y) \leq \log \gamma(x)]
\]
(8)

If we assume that $j-1$ codewords $\{c_1\}_{i=1}^{j-1}$ have been chosen. Denote
\[
D_{j-1} = \cup_{i=1}^{j-1} \{y : i(c_i; y) > \log \gamma(c_i)\} \subseteq B.
\]
(9)

The conditional probability of error that $x \in F$ is chosen to be the $j$th codeword is
\[
\epsilon_j(c_1, \cdots, c_{j-1}) = 1 - P[i(x; Y) > \log \gamma(x) | D_{j-1} | X = x].
\]
(10)

The expectation of error is actually the conditional expectation as follows,
\[
\mathbb{E}[\epsilon_j(c_1, \cdots, c_{j-1}, X) | F]
= \sum_{x \in F} P[i(x; Y) \leq \log \gamma(x) \cup D_{j-1} | X = x] P(X = x | F)
\leq \sum_{x \in F} P[i(x; Y) \leq \log \gamma(x) | X = x] P(X = x | F) + P_Y(D_{j-1})
\leq \sum_{x \in F} P(X = x) P[i(x; Y) \leq \log \gamma(x)]
\]
(11)

Thus, there exists a codeword $c_j \in F$ such that
\[
\epsilon_j(c_1, \cdots, c_{j-1}, c_j) \text{ satisfies }
\epsilon_j P[X] \leq \sum_{x \in F} P(X = x) P[i(x; Y) \leq \log \gamma(x)]
+(j-1)P_Y[i(x; Y) > \log \gamma(x)]
\]
(12)

In particular, the maximal error probability should satisfy
\[
\epsilon_P[X] \leq \sum_{x \in F} P(X = x) P[i(x; Y) \leq \log \gamma(x)]
+(M-1)\sup_{x \in F}P_Y[i(x; Y) > \gamma(x)]
\]
(13)

\[
= \sum x \in F \sup_{x \in F} P_Y[i(x; Y) > \log \gamma(x)]
\]
(14)

**Proof.** As we have shown in Lemma 1 that there exists a codebook $\{c_j\}_{j=1}^M$, the conditional error probability given the $j$th codeword satisfies
\[
\epsilon_j \leq \sum_{x \in F} P(X = x) P[i(x; Y) \leq \log \gamma(x)]
\]
(15)

Consider all possible permutations of the index $j$ and the codewords are equiprobable, the average error probability satisfies
\[
\epsilon \leq \frac{\sum_{x \in F} P(X = x) P[i(x; Y) \leq \log \gamma(x)]}{P_X[F]}
+ \frac{M-1}{2} \sup_{x \in F} P_Y[i(x; Y) > \gamma(x)]
\]
(16)

Consequently, we have
\[
\epsilon_P[X] \leq \sum_{x \in F} P(X = x) P[i(x; Y) \leq \log \gamma(x)]
+ \frac{M-1}{2} P_X[F] \cdot \sup_{x \in F} P_Y[i(x; Y) > \gamma(x)]
\]
(17)

**Remark.** The above lemma is in general weaker than the next one. However, it is more convenient to evaluate when the computation of the expectation of $P_Y[i(x; Y) > \gamma(x)]$ over $x \in F$ is difficult.

**Lemma 2.** For any distribution $P_X$ on $A$, and any measurable function $\gamma : A \mapsto [0, \infty]$, there exists a code with $M$ codewords in the set $F$ whose average error probability satisfies
\[
\epsilon_P[X] \leq \sum_{x \in F} P(X = x) P[i(x; Y) \leq \log \gamma(x)]
+(M-1)\sup_{x \in F}P_Y[i(x; Y) > \gamma(x)]
\]
(13)

\[
= \sum x \in F \sup_{x \in F} P_Y[i(x; Y) > \log \gamma(x)]
\]
(14)

**Proof.** As we have shown in Lemma 1 that there exists a codebook $\{c_j\}_{j=1}^M$, the conditional error probability given the $j$th codeword satisfies
\[
\epsilon_j \leq \sum_{x \in F} P(X = x) P[i(x; Y) \leq \log \gamma(x)]
\]
(15)

Consider all possible permutations of the index $j$ and the codewords are equiprobable, the average error probability satisfies
\[
\epsilon \leq \frac{\sum_{x \in F} P(X = x) P[i(x; Y) \leq \log \gamma(x)]}{P_X[F]}
+ \frac{M-1}{2} \sup_{x \in F} P_Y[i(x; Y) > \gamma(x)]
\]
(16)

Consequently, we have
\[
\epsilon_P[X] \leq \sum_{x \in F} P(X = x) P[i(x; Y) \leq \log \gamma(x)]
+ \frac{M-1}{2} P_X[F] \cdot \sup_{x \in F} P_Y[i(x; Y) > \gamma(x)]
\]
(17)

**Remark.** The above lemma is in general weaker than the next one. However, it is more convenient to evaluate when the computation of the expectation of $P_Y[i(x; Y) > \gamma(x)]$ over $x \in F$ is difficult.

**Lemma 3.** For any distribution $P_X$ on $A$, and any measurable function $\gamma : A \mapsto [0, \infty]$, there exists a code with
$M$ codewords in the set $F$ whose average error probability satisfies

$$
\epsilon P_X[F] \leq EX[P(i(x; Y) \leq \log \gamma(x)) \cdot 1\{x \in F\}]
+ \frac{M-1}{2} EX[P_Y(i(x; Y) > \log \gamma(x)) \cdot 1\{x \in F\}].
$$

(18)

Proof. In the step (11) in the proof of Lemma 1, we rewrite it as

$$
\mathbb{E}[c_j(c_1, \ldots, c_{j-1}, X)|F]
= \sum_{x \in F} P[i(x; Y) \leq \log \gamma(x)]P(X = x|F)
\leq \sum_{x \in F} P[i(x; Y) \leq \log \gamma(x)]P(X = x|F) + P_Y(D_{j-1})
\leq \sum_{x \in F} P[i(x; Y) \leq \log \gamma(x)]P(X = x|F)
\sum_{j=1}^{J-1} P_Y[i(x; Y) > \gamma(x)]
+ \sum_{x \in F} P_Y[i(x; Y) > \gamma(x)]
\leq \sum_{x \in F} P[i(x; Y) \leq \log \gamma(x)]P(X = x|F)
\sum_{j=1}^{J-1} P_Y[i(x; Y) > \gamma(x)]
+ \frac{M-1}{2} EX[P_Y(i(x; Y) > \log \gamma(x)) \cdot 1\{x \in F\}].
$$

(19)

Consider all possible orders we choose these codewords from $F$, the indexes is irrelevant. The average error probability should satisfy

$$
\epsilon \leq \frac{\sum_{x \in F} P(X = x)P[i(x; Y) \leq \log \gamma(x)]}{P_X[F]}
+ \frac{1}{M} \sum_{j=1}^{J-1} \sum_{x \in F} P_Y[i(x; Y) > \log \gamma(x)]P(X = x|F)
+ \frac{1}{M} \sum_{x \in F} P(X = x)P[i(x; Y) \leq \log \gamma(x)]
+ \frac{M-1}{2} \frac{EX[P_Y(i(x; Y) > \log \gamma(x)) \cdot 1\{x \in F\}]}{P_X[F]}.
$$

(20)

Thus, we have proved (18).

This lemma extends Lemma 19 in [8] by choosing codewords from a subset $F$.

C. Further Results on Achievability Bounds

Now consider the binary hypothesis test between $P_{Y|X=x}$ and $P_Y$. Let us introduce the detection probability

$$
P_{Y|X=x}[i(x; Y) > \log \gamma(x)] \geq 1 - \epsilon + \tau(x)
$$

with $0 < \tau(x) < \epsilon$ of Neyman-Pearson hypothesis tests with decision threshold $\log \gamma(x)$ when the sending codeword is $x$. The details of Neyman-Pearson hypothesis testing can be found in Appendix A. Note that for a particular codeword $x$, we do have a separate threshold $\gamma(x)$, and the resulting false alarm probability for this particular $x$ is $P_Y[i(x; Y) > \gamma(x)] = \beta_{1-\epsilon+\tau(x)}$. It is obvious that a lower bound on the decision threshold $\gamma(x)$ is equivalent to a lower bound on the detection probability $1 - \epsilon + \tau(x)$.

The following theorem is a combination of random coding, and selection $\gamma(x)$ for each $x$ in the set of $F$ such that for each $x$ in this set, and we have the detection probability in favor of $P_{Y|X=x}$ over $P_Y$ lower bounded by $1 - \epsilon + \tau(x)$ with an additional constraint: $\tau(x) < \epsilon$ for all $x \in F$.

Theorem 1. For any input distribution $P_X$ on $A$ and measurable function $\tau: A \rightarrow [0, \infty]$, there exists a code with $M$ codewords in the set $F \subseteq A$ such that the maximal error probability $\epsilon$ satisfies

$$
M \geq \frac{E_x[\tau(x) \cdot 1\{x \in F\}]}{\sup_{x \in F} \beta_{1-\epsilon+\tau(x)}(x, Q_Y)}
$$

(22)

$\beta_{1-\epsilon+\tau(x)}(x, Q_Y)$ is the minimum probability of error under hypothesis $Q_Y$ if the probability of error when $x$ is sent is not larger than $1 - \epsilon + \tau(x)$.

Proof. The theorem is an application of Lemma 1 with Dependent testing is substituted by composition of Neyman-Pearson tests. We have

$$
M \geq \frac{\epsilon P_X[F] - EX[P(i(x; Y) \leq \log \gamma(x)) \cdot 1\{x \in F\}]}{\sup_{x \in F} P_Y[i(x; Y) > \log \gamma(x)]}
$$

(23)

from Lemma 1, and $P_Y[i(x; Y) > \log \gamma(x)] = \beta_{1-\epsilon+\tau(x)}$ from above analysis, then $P_Y[i(x; Y) \leq \log \gamma(x)] = \epsilon - \tau(x)$. The conclusion is obvious.

Corollary 1. If $F$ is the subset of $F$ that satisfies $\tau(x) \geq \tau_0$, then the bound is rewritten as

$$
M \geq \frac{\tau_0 P_X[F]}{\sup_{0 < \tau_0 < \epsilon, x \in F} \beta_{1-\epsilon+\tau(x)}(x, Q_Y)}
$$

(24)

Note that it is slightly different from (127) in [8], i.e., the parameter in the detection probability is codeword dependent.

Remark. When the constraint set of $F$ in Theorem 25 is the set of vectors sharing the same length, the weaker lower bound in (127) becomes trivial as $Q_X(F) = 0$ when $Q_Y$, which is a Gaussian, is induced by a Gaussian input.

The following two achievability bounds of average error probability are direct applications of Lemma 2 and Lemma 3.

Theorem 2. For any distribution $P_X$ on set $A$ there exists a code with $M$ codewords in $F$ and average probability of error satisfying

$$
M \geq \frac{2 \cdot E_x[\tau(x) \cdot 1\{x \in F\}]}{\sup_{x \in F} \beta_{1-\epsilon+\tau(x)}(x, Q_Y)}
$$

(25)

Theorem 3. For any distribution $P_X$ on set $A$ there exists a code with $M$ codewords in $F$ and average probability of error satisfying

$$
M - \frac{1}{2} \geq \frac{E_x[\tau(x) \cdot 1\{x \in F\}]}{E_x[\beta_{1-\epsilon+\tau(x)}(x, Q_Y) \cdot 1\{x \in F\}]}.
$$

(26)

Remarks. * In above results, $\tau(x)$ is a function of the candidates of the codewords, which depends on the specific $x$. It should be within $(0, \epsilon)$ and depends on the encoding scheme and $x$. Moreover, any function
which satisfies \( \tau(x) \in (0, \epsilon) \) should be feasible. This is different from (108) or (127) in [8]. Tighter bounds can be obtained by searching the optimal function \( \tau(x) \) to maximize the right-hand side of all these bounds. For example, the bound in (22) can be further optimized as

\[
M \geq \sup_{x \in \mathbb{F}} \frac{E_x [\tau(x) \cdot 1_{\{x \in \mathbb{F}\}}]}{\tau(x) \sup_{x \in \mathbb{F}} \beta_{1-\epsilon+\tau(x)}(x, Q_Y)}
\]  

\( (27) \)

If we choose \( \tau(x) = \tau \) for some constant \( \tau \), then the achievability bound will be the same as (127).

- In Theorem 3, if we can get an upper bound \( H(R) \) of \( \beta \) which is a concave function of \( R = R(x) \), then we can use Jensen’s Inequality to get a lower bound

\[
\frac{M - 1}{2} \geq \frac{E_x [\tau(x) \cdot 1_{\{x \in \mathbb{F}\}}]}{H(E_x[R])}.
\]

\( (28) \)

- In Theorem 3, if we use the subset \( \mathbb{F} \) with constraint that \( \tau(x) > \tau_0 \) instead and \( \sup_{x \in \mathbb{F}} \beta_{1-\epsilon+\tau(x)}(x, Q_Y) \cdot P_X[F] \)

\[
\geq \sup_{0 < n < \epsilon} \sup_{x \in \mathbb{F}} \tau_0
\]

\( (29) \)

This bound is \( \frac{1}{P_S[F]} \) times larger than the bound with maximal error probability in (24).

IV. EVALUATION OF THE BOUNDS IN AWGN CHANNELS

A. Converse Bounds

In this section, we focus on evaluating the converse bound for covert communication over AWGN channels. In general, a converse bound is independent of the input distribution and the construction of the code. In the scenario of covert communication, it is assumed that the length of the code as well as the power level is known by the adversary, but the adversary can not still effectively detect the communication provided the power is low enough. Hence, we prove the converse bound by providing a converse bound under a maximal power constraint \( P(n) \), i.e., each codeword \( c_i \in X^n \) should satisfy: \( \|c_i\|^2 \leq nP(n) \) where \( P(n) \) is a decreasing function of \( n \).

For the converse bound under maximal probability of error, a general conclusion (Theorem 31 in [8]) under binary hypothesis test is

\[
M \leq \inf_{Q_Y} \sup_{x \in \mathbb{F}} \frac{1}{\beta_{1-\epsilon}(x, Q_Y)}.
\]

\( (30) \)

The distribution \( Q_Y \) is determined as \( \mathcal{N}(0, (1 + P(n))I_n) \) and the first infimum is canceled. However, we need to find the \( x \) in \( \mathbb{F} \) which minimizes \( \beta_{1-\epsilon}(x, Q_Y) \). This is rather involved.

From a part of the conclusion in Lemma 39 in [8]:

\[
M_m^*(n, \epsilon, P(n)) \leq M_m^*(n + 1, \epsilon, P)
\]

\( (31) \)

regardless of whether \( \epsilon \) is an average or maximal probability of \( \epsilon \). If a converse bound under equal power constraint has been obtained, the above inequality would imply a converse bound under maximal power constraint. In [8], the converse bound under equal power constraint is proved in Theorem 65. However, we could not use it directly since the power \( P \) is decreasing with \( n \) in our case. Thus, in our proof, the auxiliary distribution \( P_{Y_m} \) is chosen as \( P_{Y_m} = \mathcal{N}(0, (1 + P(n)I_n)) \). The proof is similar with the proof of Theorem 65 in [8] except that the power \( P(n) \) is decreasing with \( n \) due to covert constraint.

Theorem 4. For the AWGN channel with \( P = P(n) \) which is a decreasing function of \( n, \) and \( \epsilon \in (0, 1) \) and equal power constraint under a given \( n: \) each codeword \( c_i \in X^n \) satisfies \( \|c_i\|^2 \leq nP(n), \) we have (maximal probability of error)

\[
\log M_m^*(n, \epsilon, P(n)) \leq nC(n) - \sqrt{nV(n)}Q^{-1}(\epsilon) + \frac{1}{2} \log n + O(1).
\]

(32)

The proof is almost the same as Theorem 65 in [8] except that normal approximation is slightly different. The details can be found in [9]. The basis of the proof is the Central Limit Theorem and Berry Esseen Theorem still work provided the power of each coordinate remains constant under any given block length. Actually, when the power is irrelevant with \( n, \) the information rate density will approach a fixed Gaussian distribution whose expectation is the capacity and the variance is channel dispersion. When the power level is decreasing with \( n, \) the information rate density will approach Gaussian distributions of different expectations and variances. The key of the proof is that when the power is decreasing with \( n, \) the expectations and the variances are closer and closer so that we can approximate them by choosing sufficiently large \( n. \) From (31), (32) and Taylor’s theorem, we have a converse bound under maximal power constraint (maximal probability of error)

\[
\log M_m^*(n, \epsilon, P(n)) \leq nC(n) - \sqrt{nV(n)}Q^{-1}(\epsilon) + \frac{1}{2} \log n + O(1).
\]

(33)

B. Achievability Bound

In this section, the achievability bound under maximal probability of error is discussed. Generally speaking, the framework of the proof on achievability bound here is similar as the proof of Theorem 67 in [8]. However, there are several differences.

1. Firstly, in the generation of the codebook, the input distribution is determined as zero-mean Gaussian distribution with a specific variance \( \mu P(n) \). The dependence on both \( \mu \) and \( P(n) \) stems from controlling the TVD at the adversary.

2. Secondly, the codewords are drawn from an specific set \( \mathbb{F} \) which is a subset of a \( n \)-dimension sphere and varies with \( n \).
3) Thirdly, as in the proof of the converse bound, the power \(P\) is decreasing with \(n\), so that we should be very cautious when dealing with normal approximation. Because of the first reason, the general formula (108) in [8] which is used to prove the achievability bound in Theorem 67 is not applicable here, and neither is formula (127). In the proof of Theorem 67 in [8], the same dependent test threshold \(\tau(x)\) will lead to the same false alarm probability \(\beta(x)\) for all codewords \(x\) in the set \(\mathcal{F}\) under equal power constraint (the set \(\mathcal{F}\) was defined to be the surface of \(n\)-dimensional sphere in [8]), which is not satisfied under maximal power constraint. Hence, the conclusion of Theorem 1 is available. The difficulty in evaluating Theorem 1 is that we have to integrate the \(\tau(x)\) over the whole \(n\)-dimensional space. Instead, we use Corollary 1.

**Theorem 5.** For the AWGN channel with noise \(N(0, 1)\) and any \(0 < \epsilon < 1\), there exists an \((n, M, \epsilon)\) code (maximal probability of error) chosen from a set \(\mathcal{F}\) of codewords whose coordinates are i.i.d. \(N(0, \mu P(n))\), \(0 < \mu < 1\) and satisfy:

1. \(\|x\|_2^2 \leq nP(n)\)
2. \(\tau_0 \leq \tau_n(R) \leq \frac{n}{n+1} \epsilon\).

Let \(x = [\sqrt{R}, \ldots, \sqrt{R}]\), \(C_n = \frac{1}{2} \log(1 + \mu P(n))\),
\[\tau_n(R) = \frac{B_{\mu}([P, R])}{\frac{\mu}{n}}\]
\[T_n(P, R) = E\left[\left(\frac{\log e}{2(1 + MP)}[\mu P + 2\sqrt{R}Z_i - \mu PZ_i^2]\right)^2\right],\]
\[\hat{V}_n(P, R) = \left(\frac{\log e}{2(1 + P)}\right)^2(4R + 2P^2) = V\left(\frac{2R + P^2}{2P + P^2}\right),\]
where \(Z_i\)'s are i.i.d. standard normal, then we have (maximal probability of error)
\[
\log M_n^*(n, \epsilon, P(n)) \geq \sup_{0 < \tau_0 < \epsilon} \{nC_n + \frac{n(R - \mu P(n))log e}{2(1 + \mu P(n))} + \sqrt{n\hat{V}_n(P, R^*)Q^{-1}}(1 - \epsilon + \frac{2B_{\mu}([P, R^*])}{\sqrt{n}}) + \log \tau_0 + \log P_X[F_n] + \frac{1}{2} \log n - \log \left[\frac{2\log 2}{\sqrt{2\pi V_{\mu}(P, R^*)}} + 4B_{\mu}(P, R^*)\right]\}.
\]
(34)

The quantity \(R^*\) satisfies \(x^*_{0} = [\sqrt{R^*}, \ldots, \sqrt{R^*}] \in \mathcal{F}\) and maximizes (30).

The details of the proof could be found in Appendix A.

**Remarks.**

- If we could prove that \(R = P(n)\) satisfies \(\tau_0 \leq \tau_n(R) < \epsilon\) and maximizes \(\beta_1 - \epsilon - \tau_n(R)\) or (31), then the achievability bound could be
\[
\log M_n^*(n, \epsilon, P(n)) \geq nC_n - \sqrt{n\hat{V}_n(n)}Q^{-1}(\epsilon) + \frac{1}{2} \log n + \sup_{\tau_0} \{\log \tau_0 + \log P_X[F_n] + O(1)\}.
\]

From the expression of (31), as a function of \(R\) with fixed \(n\), the derivative of it is positive when \(n\) is sufficiently large and there exists some \(\tau_0\) for \(\epsilon > \tau_n(R) > \tau_0\) holds. Hence, we have
\[
\log M_n^*(n, \epsilon, P(n)) \geq nC_n - \sqrt{n\hat{V}_n(n)}Q^{-1}(\epsilon) + \frac{1}{2} \log n + \log \tau_0 + \log P_X[F_n] + O(1).
\]
holds for some \(\tau_0\). The above claim holds when \(n\) is sufficiently large.

- The right-hand side of the constraint \(\tau_0 \leq \tau_n^R(R) \leq \frac{1}{n+1} \epsilon\) is to ensure that \(R_0\) is compact. \(\tau_0 \leq \gamma_n^R(R) < \epsilon\) is sufficient for the existence of \(\mathcal{F}\) and the upper bound (31) of \(\beta_1 - \epsilon - \tau_n(R)\).

The parameter \(\mu\) satisfies \(\mu \in (0, 1)\). If \(\mu\) is away from 1, the set \(\mathcal{F}\) is almost the whole space and the codewords are almost i.i.d. Gaussian distributed. Nevertheless, \(\mu\) can be slightly less than 1 and the codewords are still behaving like Gaussian codewords for moderate large \(n\) due to sphere hardening effect [11]. To some extent, the codebook can be regarded as Gaussian codebook. The utilization of sphere hardening effect is important for controlling the TVD at the adversary. The details about choosing \(\mu\) will be discussed in [9].

**ACKNOWLEDGMENT**

This work was supported by National Natural Science Foundation of China under Grant 61871264.

**APPENDIX A**

**PROOF OF THEOREM 5**

**Proof.** Some definitions and notions used in the proof are

1. The codeword is now \(x^n \in A \triangleq \mathbb{R}^n\).
2. \(Y^n|X^n=x^n\) stands for the condition probability of \(Y^n\) when the codeword \(x^n\) is sent.
3. \(\mathcal{F} = \{x^n; \|x_i\|^2 \leq nP(n)\}\). It is clear that \(\mathcal{F}\) is a subset of \(\mathcal{F}\).

**Generation of the codebook:** The process of generation is the same as in Lemma 1 and the dependent test is substituted by Neyman-Pearson test as in Theorem 1. For each \(n\), the distribution \(P_X = N(0, \mu P(n)I_n)\) where \(0 < \mu < 1\) will be determined later, i.e., each coordinate of these candidates is i.i.d. drawn from \(N(0, \mu P(n))\). Each codeword is randomly chosen from the set \(\mathcal{F}\) following the steps in Lemma 1. The conclusion is an application of Corollary 1. The details are as follows.

Since the candidates of these codewords are generated from \(N(0, \mu P(n)I_n)\), the auxiliary distribution \(P_{Y^n} = N(0, (1 + \mu P(n))I_n)\) as our bounds are based on binary hypothesis test between \(P_{Y^n|X^n-x^n}\) and \(P_{Y^n}\). It is necessary to evaluate \(\beta_0^n(x^n, N(0, (1 + \mu P(n))I_n))\) with a given detection probability \(\alpha\). Assume \(x^n = [\sqrt{R(n)}, \ldots, \sqrt{R(n)}]\), because spherical symmetry will lead to the same \(\beta\) with given \(\alpha\) on the surface with radius \(\sqrt{nR}\). Under \(P_{Y^n}\) and \(P_{Y^n|X^n=x^n}\), the expressions of \(\beta_0^n\) and \(\alpha = 1 + \tau(n) - \epsilon\) are
\[
\beta_1 - \epsilon - \tau(n, R) = \mathbb{P}[G_n(R) \geq \gamma(n, R)]
\]
(32)
\[nC_\mu(n) + \frac{n(R^* - \mu P(n)) \log e}{2(1 + \mu P(n))} + \sqrt{n\hat{V}_\mu(P(n), R^*)}Q^{-1}\left(1 - \epsilon + \frac{2B_\mu(P(n), R^*)}{\sqrt{n}}\right)
+ \frac{1}{2} \log n + \log \tau_0 + \log P_X[\hat{F}_n] - \log \left[\frac{2\log 2}{\sqrt{2\pi \hat{V}_\mu(P, R)}} + 4B_\mu(P, R)\right].\] (30)

And \(B_\mu(P) = \frac{6T_n(P)}{V_n(P)^{3/2}}.\)

Let \(\hat{S}_i = S_i - E(S_i),\) then

\[H_n = nC_\mu(n) + \frac{nR\log e}{2(1 + \mu P)} - \frac{n\mu P \log e}{2(1 + \mu P)} - \sum_{i=1}^{n} \hat{S}_i.\] (39)

Let \(\alpha_n^\mu = 1 - \epsilon + 2\tau_n^\mu(R),\) \(\zeta_n^\mu = \sqrt{n\hat{V}_\mu(n)Q^{-1}(\alpha_n^\mu)}\) and

\[\log \gamma_n = nC_\mu(P(n)) + \frac{n(R - \mu P) \log e}{2(1 + \mu P)} + \sqrt{n\hat{V}_\mu(P(n))Q^{-1}(\alpha_n^\mu)} + \frac{2\log 2}{\sqrt{2\pi \hat{V}_\mu(P, R)}} + 4B_\mu(P, R).\] (40)

We have

\[T_\mu(P) \leq \left[\frac{P \log e}{2(1 + \mu P)^{3/2}}[C_1 + C_2/P + \frac{16\sqrt{2}C_3}{\pi P^2}]\right].\] (41)

where \(C_1, C_2, C_3\) are positive constants. Hence, as \(P(n) \to 0,\)

\[0 < B_\mu(P(n)) < \left(\frac{\log e}{2(1 + \mu P)^{3/2}}\right)^3 (4R + 2\mu^2 P^2)^{3/2}\] (42)

Then, \(\frac{n \mu P}{\sqrt{n}}\) tends to 0 as \(n \to \infty,\) \(\alpha_n^\mu\) is certainly less than 1 when \(n\) is sufficiently large and the definition of \(\zeta_n^\mu\) is meaningful.

As \(\hat{S}_i, i = 1, \ldots, n\) are i.i.d zero- mean variables with variance \(\hat{V}(P(n)),\) Berry-Esseen Theorem implies that

\[\Pr\left[\sum_{i=1}^{n} \hat{S}_i \leq \zeta_n^\mu\right] \geq \alpha_n^\mu - \frac{B_\mu(P(n))}{\sqrt{n}}.\] (43)

for all \(x_n^\mu\) with the same radius \(\sqrt{nR(n)}\) in the space.

From Lemma 47 in [8], an upper bound (44) of \(\beta_{\alpha_{\mu} + \tau_\mu(R)}(x_n^\mu, Q_Y)\) for \(x_n^\mu \in F_n\) is found for each \(n.\)

From now on, the blocklength \(n\) is sufficiently large and fixed so that \(\alpha_n^\mu = 1 - \epsilon + 2\tau_n^\mu(R)\) is less than 1. Note that \(R\) varies in \([0, P(n)],\) to utilize Corollary 1, the following statements are important.

(a) From (42), \(B_\mu(P, R)\) is positive and bounded, \(\tau_n^\mu(R) = \frac{B_\mu(P, R)}{\sqrt{n}}\) will be sufficiently small when \(n\) is large. We
can always find some \( \tau_0 \) so that the set \( \mathcal{F}_n \) is nonempty, and there are sufficiently many points in \( \mathcal{F}_n \).

(b) Since \( P \) is a function of \( n \), \( \tau_n^R(R) \) is a continuous function of \( R \) with fixed \( n \). Considering the mapping from \([0, P(n)]\) of \( \tau_n^R(R) \) to \( \mathbb{R} \), it is proper since \([0, P(n)]\) is compact and \( \tau_n^R(R) \) is continuous.\(^1\) When the range of \( \tau_n^R(R) \) is constrained to be \([0, \frac{1089}{n}\]) \( \epsilon \), the preimage \( R \) of \( \tau_n^R(R) \) is a compact set. We denote it as \( R_n \).

(c) From the expression (31), it is a continuous function of \( R \). Consequently, we can always find some \( R^* \in R_n \) which maximizes (31).

(d) As the set \( \mathcal{F}_n \) is determined by \( \tau_0 \) and \( \sqrt{nR^*} \) is radius of the point \( x_{R^*} = [\sqrt{R^*}, \cdots, \sqrt{R^*}] \) in \( \mathcal{F}_n \). The value \( R^* \) depends on \( \tau_0 \).

(e) There are many choices of \( \tau_0 \), and we choose one which maximizes (30). The choice will lead to the tightest achievable bound for the throughput.

Consequently, we have proved that a codebook exists which satisfies maximal power constraint and the lower bound of the size satisfies (34).

**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, Sep. 2013.

[2] M. R. Bloch, “Covert Communication Over Noisy Channels: A Resolvability Perspective,” IEEE Trans. Inf. Theory, Vol. 62, No. 5, pp. 2334-2354, May 2016.

[3] 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, May 2016.

[4] P. H. Che, M. Bakshi, S. Jaggi, “Reliable Deniable Communication: Hiding Messages in Noise,” IEEE Int. Symp. Inf. Theory (ISIT2013), Istanbul, Turkey, pp. 2945-2949, Jul. 2013.

\(^1\) A continuous map: \( f : X \to Y \) between topological spaces is called proper if for every compact subspace \( K \subseteq Y \), the pre-image \( f^{-1}(K) \) is compact. When \( X \) is compact and \( Y \) is Hausdorff, then every continuous map \( f : X \to Y \) is proper.

[5] A. Abdelaziz and C. E. Koksal, “Fundamental limits of covert communication over MIMO AWGN channel,” 2017 IEEE Conference on Communications and Network Security (CNS), pp. 1-9, Las Vegas, NV, 2017.

[6] M. Tahmasbi and M. R. Bloch, “Second-Order Asymptotics of Covert Communications over Noisy Channels,” in IEEE Int. Symp. Inf. Theory (ISIT2016), Barcelona, Spain, pp. 2224-2228, Jul. 2016.

[7] M. Tahmasbi and M. R. Bloch, “First and Second Order Asymptotics in Covert Communications,” IEEE Trans. Inf. Theory, DOI 10.1109/TIT.2018.2878526, Oct. 2018.

[8] Y. Polyanskiy, H. V. Poor and S. Verdu, “Channel Coding Rate in the Finite Blocklength Regime,” IEEE Trans. Inf. Theory, Vol. 56, No. 5, pp. 2307-2358, May 2015.

[9] X. Yu, S. Wei and Y. Luo, “One-shot achievability and converse bounds of Gaussian random coding in AWGN channels under covert constraints,” arXiv preprint arXiv:1909.11324.

[10] E. Lehmann and J. Romano, Testing Statistical Hypotheses, 3rd ed. New York: Springer, 2005.

[11] J. Hamkins and K. Zeger “Gaussian Source Coding With Spherical Codes,” IEEE Trans. Inf. Theory, Vol. 48, No. 11, pp. 2980-2989, Nov 2002.