Covert Communication Over a Compound Discrete Memoryless Channel

Mehrasa Ahmadipour*, Sadaf Salehkalaibar*, Mohammad Hossein Yassaee† and Vincent Y. F. Tan‡

* ECE Department, College of Engineering, University of Tehran, Tehran, Iran, Emails: {m.ahmadipour,s.saleh}@ut.ac.ir
† Institute for Research in Fundamental Sciences (IPM), Tehran, Iran, Email: yassaee@ipm.ir
‡ Department of ECE, National University of Singapore, Email: vtan@nus.edu.sg

Abstract—In this paper, we study covert communication over a compound discrete memoryless channel (DMC). There are two channel states in which one of them is arbitrarily chosen and remains fixed during the transmission. The objective is to reliably send a message from the transmitter to the receiver. An adversary who is observing the channel output should not be able to infer the channel state. Two covertness metrics are considered. In the first metric, covertness is measured using the KL-divergence between the channel output marginal of each state with a fixed distribution. Different cases where such a distribution can be specified, are studied. The optimal transmission rate of each case is established. In the second metric, the covertness is measured by using the total variation distance of the channel output marginals of the two states. Upper and lower bounds on the optimal transmission covert rate are derived. The bounds match for a special case and characterize the optimal throughput.

I. INTRODUCTION

Covert communication aims to conceal the status of whether the transmitter is sending a message. An adversary observing the channel output, should not be able to find out whether the transmitter is communicating a message or the all-zeroes sequence, which represents the no communication status. It can be shown, under some conditions, that the maximum amount of information that can be transmitted scales with the “square root” of blocklength. The square root law has been investigated for Gaussian channels [1], discrete memoryless channels (DMCs) [2], [3], classical-quantum [4], multiple-access [5] and broadcast channels [6]. Fundamental limits of covert communication from the second-order perspective [7] and channels with state [8] have been established. In [9], covert communication over compound binary symmetric channels was studied but the setting is different from that in Sec. III here. Specifically, the best throughput to deniability (i.e., the adversary cannot learn whether communication is occurring) and reliably communicate was established, whereas our main contribution is to establish bounds on the throughput under the constraint that the adversary cannot learn the channel state.

In this paper, we consider covert communication over a compound DMC (see Fig. 1). There are two channel states \( s \in \{1, 2\} \) where at each state, a specific conditional channel pmf \( W_s(\cdot|\cdot) \) is induced. The channel state is fixed through the transmission and is known at the transmitter. The goal of the communication is to reliably send a message from the transmitter to the receiver while an adversary observing the channel output cannot infer the channel state.

Two covertness metrics are considered in this work. In the first, we assume that there exist two input distributions \( P_1 \) and \( P_2 \) which induce the same channel output marginal when they are fed into channels \( W_1 \) and \( W_2 \), respectively. The common output distribution is denoted by \( Q_0 \) and covertness is measured by the KL-divergence between the channel output marginal at each state \( Q_n^1 \) and \( Q_n^2 \) (the \( n \)-fold product of \( Q_0 \)). The input distributions \( P_1 \) and \( P_2 \) can be either deterministic or non-deterministic. Notice the difference between this formulation for a compound channel and that for a DMC in previous works [2], [3]. In our setup, an “off” symbol which represents the state of transmitter when there is no input to the channel, does not necessarily exist. This is because \( Q_0 \) denotes the common channel output marginal between the two states and it does not necessarily correspond to the output marginal of a specific input symbol. Thus, an adversary observing the channel output may learn that the transmitter is sending a message. However, the main goal of covert communication over a compound channel is primarily to mask the channel state. For this setup, we consider all possible cases that a distribution \( Q_0 \) can be defined for covert communication. We also characterize the optimal covert rate of each case.

The second covertness metric is more pertinent to the compound DMC. In this setup, covertness is measured by the total variation distance between the channel output marginals of two states \( Q_n^1 \) and \( Q_n^2 \). Providing a converse is challenging here since we are required to work with non-product distributions \( Q_n^1 \) and \( Q_n^2 \). We obtain upper and lower bounds on the optimal transmission throughput and show that the square root law holds. We also present a special case where the two bounds match and characterize the optimal transmission rate.

Fig. 1. Covert communication over a compound DMC

\[ M \xrightarrow{Tx} X^n \xrightarrow{W^n} Y^n \xrightarrow{Rx} M \]

Hypothesis 1: \( Q_n^1 \)
Hypothesis 2: \( Q_n^2 \)
Notation: The KL-divergence, chi-squared, and total variation distance between two distributions $P$ and $Q$ are denoted by $D(P || Q)$, $\chi^2(P || Q)$ and $d_{TV}(P, Q)$ respectively. The notation $w(x^n)$ denotes the Hamming weight of sequence $x^n$. The binary entropy function is shown by $h_b(\cdot)$. The $n$-fold product of $Q_0$ is denoted by $Q_0^\times n$.

II. COVERT COMMUNICATION OVER A COMPOUND CHANNEL

A. System Model

Consider covert communication over a compound DMC with input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, see Fig. 1. The compound channel consists of two channel pmfs $W_s(\cdot | \cdot)$, $s \in \{1, 2\}$ where the channel state $s$ is available at the encoder. We denote the input distribution on $\mathcal{X}$ by $P_s$ and the output distribution on $\mathcal{Y}$ for each possible channel pmf $W_s(\cdot | \cdot)$ by $Q_s$. The message set is shown by $\mathcal{M}$. The transmitter and the receiver choose a random code of blocklength $n$ for each message set $\mathcal{M}$, using a sufficiently long secret key $K$ that is shared between them. The transmitter sends an $n$-length input $X^n$ over the channel using a state-dependent encoding function $f_s(n) : \mathcal{M} \rightarrow X^n$ where

$$X^n = f_s(n)(M).$$

The receiver uses a decoding function $g_s(n) : \mathcal{Y}^n \rightarrow \mathcal{M}$ which maps the channel output to an estimated message as follows:

$$\hat{M} = g_s(n)(\mathcal{Y}^n).$$

The message $M$ uniformly drawn from $\mathcal{M}$, together with the channel pmf $W_s(\cdot | \cdot)$ induces a distribution $Q_s^n(\cdot)$ on $\mathcal{Y}^n$.

In the conventional covert communication over a single-state point-to-point channel, a fixed distribution which is denoted by $Q_0$, is chosen to represent the single-letter channel output marginal when the transmitter is not sending a message. The covertness is measured by the “distance” between the output marginal of the channel when a message is sent and the distribution $Q_0^\times n$. For covert communication over a compound channel, the covertness is defined by the distance between the channel output marginal at each state and a fixed distribution $Q_0^\times n$. However, for the case of compound channel, the distribution $Q_0$ can be specified through several definitions which are discussed in the following. Here, the fixed distribution $Q_0$ does not necessarily represent the state of the transmitter when it has no input to the channel.

Case 1: There exist two non-deterministic distributions $P_1$ and $P_2$ such that

$$\sum_x P_1(x)W_1(y|x) = \sum_x P_2(x)W_2(y|x), \forall y \in \mathcal{Y}. \tag{3}$$

1The adversary knows the distribution which is used for generating the codebook but does not know the chosen codebook by the transmitter and receiver.

2A deterministic distribution $P$ is one in which there exists a symbol $x^*$ such that $P(x^n) = 1$. A non-deterministic distribution is one that is not deterministic.

In this case, we define $Q_0$ to represent the above common channel output marginal between the two states. Here, an adversary who is looking at the channel output, cannot find out the channel state. Thus, the distribution $Q_0$ masks the state of the communication. It will be discussed in the following Examples 1 and 2, that the communication rate is positive.

Here, there does not necessarily exist an input symbol whose channel output marginal is as (3). Thus, the adversary looking at the channel may distinguish the state of transmitter, i.e., whether it is sending a message. However, it does not know the state of the communication.

Case 2: Condition (3) is satisfied only if $P_1$ and $P_2$ are deterministic distributions. Thus, there exists an off-symbol denoted by ‘0’ such that

$$W_1(y|0) = W_2(y|0), \forall y \in \mathcal{Y}. \tag{4}$$

Here, $Q_0$ denotes the above common channel output marginal between the two channel states. The distribution $Q_0$ masks the channel state and also represents that the transmitter is not sending a message for both states. It will be discussed in the following Example 3, that the communication rate is zero in this case.

Case 3: Condition (3) is satisfied only if $P_1$ is deterministic while $P_2$ can be non-deterministic. In this case, we have:

$$W_1(y|0) = \sum_x P_2(x)W_2(y|x), \forall y \in \mathcal{Y}. \tag{5}$$

Here, $Q_0$ denotes the above common channel output marginal. Besides the fact that $Q_0$ masks the channel state, it also represents the fact that the transmitter has no input to the channel when $S = 1$. A similar case can be considered when $P_2$ is deterministic and $P_1$ is non-deterministic.

Case 4: There does not exist any pair of distributions $P_1$ and $P_2$ such that Condition (3) is satisfied. In this case, covert communication is not possible.

As discussed in the above Cases 1–3, a distribution $Q_0$ over $\mathcal{Y}$ can be fixed. As such we can define covert rates as follows:

Definition 1: An $([|M|, \epsilon, \delta_1, \delta_2, \beta]$-code consists of encoding and decoding functions $(f_1^{(n)} : \mathcal{M} \rightarrow \mathcal{X}^n, f_2^{(n)} : \mathcal{M} \rightarrow \mathcal{X}^n, g^{(n)} : \mathcal{Y}^n \rightarrow \mathcal{M})$, such that the probability of error satisfies

$$P_e \triangleq \max_{s=1,2} \max_{m \in \mathcal{M}} \Pr[\hat{M} \neq M | S = s, M = m] \leq \epsilon. \tag{6}$$

and the following covertness conditions hold

$$D(Q_s^n || Q_0^\times n) \leq \delta_s, s \in \{1, 2\}. \tag{7}$$

For $(\epsilon, \delta_1, \delta_2, \beta) \in R_+^4$, a covert rate $L \in R_+$ is said to be $(\epsilon, \delta_1, \delta_2, \beta)$-achievable if there exists a sequence of $([|M|, \epsilon, \delta_1, \delta_2, \beta]$-codes such that

$$\lim_{n \rightarrow \infty} \frac{\log |M|}{n^\beta} \geq L. \tag{8}$$

The supremum of all $(\epsilon, \delta_1, \delta_2, \beta)$-achievable $L$ is denoted by $L_\beta^*(\epsilon, \delta_1, \delta_2)$. Here, we are interested in the cases of $\beta = \frac{1}{2}$ and $\beta = 1$. When $L_1^*(\epsilon, \delta_1, \delta_2) > 0$, positive communication rate is feasible.

3There is also an implicit maximization over the set of key values in (6).
**B. Examples**

In this section, examples are provided for the cases discussed in the previous section. The following propositions provide the optimality results for Cases 1 and 2.

**Proposition 1:** Assume that Case 1 holds. We have

\[ L^*_1(0, \delta_1, \delta_2) = \min_{s \in \{1, 2\}} \max_{P_s} I(P_s, W_s), \]

(9)

where the maximum is over all distributions \( P_1 \) and \( P_2 \) such that Condition (3) is satisfied.

Proof: Similar to [2, Proposition 1].

**Proposition 2:** Assume that Case 2 holds. We have

\[ L^*_2(0, \delta_1, \delta_2) = \min_{s \in \{1, 2\}} \max_{P_s; P_s(0)=0} \sqrt{2\delta_s} \cdot D \left( W_s \| Q_0 \| P_s \right), \]

(10)

where \( Q_s \) denotes the output marginal of the channel \( W_s \) when the input distribution is \( P_s \).

Proof: Similar to [2, Theorem 2].

**Example 1:** Suppose that \( \mathcal{X} = \mathcal{Y} = \{0, 1\} \). At channel state \( S = 1 \), we assume that there is a binary symmetric channel, BSC \( \left( \frac{1}{2} \right) \) and at channel state \( S = 2 \), we have a BSC \( \left( \frac{3}{4} \right) \). For this example, the non-deterministic input distributions \( P_1(0) = P_1(1) = 1/2 \) and \( P_2(0) = P_2(1) = 1/2 \) satisfy Condition (3), where \( Q_0(0) = Q_0(1) = 1/2 \). Clearly, the distribution \( Q_0 \) is not induced by any of the input symbols of both channel states. Thus, an off-symbol which denotes the state of the transmitter when it is not communicating a message, cannot be defined here. An adversary may be able to infer whether the transmitter is sending a message. However, as discussed in the previous section, the goal of the communication is primarily to mask the channel state.

Evaluating the rate in (9) for the above example yields the following optimal rate:

\[ L^*_1(0, \delta_1, \delta_2) = 1 - h_b \left( \frac{1}{3} \right). \]

(11)

**Example 2 (Gaussian Setup):** Suppose that at each channel state, the channel is given by

\[ Y = X + Z_s, \quad Z_s \sim \mathcal{N}(0, \sigma^2_s), \]

(12)

where \( \sigma^2_1 > \sigma^2_2 \), \( X, Y, Z_s \in \mathbb{R} \) and \( X \) is independent of \( Z_s \). We impose an average power constraint on the input such that \( \mathbb{E} \left[ |X|^2 \right] \leq P \) for some \( P > 0 \). In the following, we discuss the scenario in which there exist two non-deterministic distributions \( P_1 \) and \( P_2 \) such that Condition (3) holds. Thus, a positive rate can be achieved for this example. Consider Proposition 1 and let the second moment of the distribution \( P_s \) be denoted by \( \rho_s \). We know that the zero-mean Gaussian random variable maximizes \( I(P_s, W_s) \) among all distributions of the same second moment. Thus, we get:

\[ I(P_s, W_s) \leq \frac{1}{2} \log \left( 1 + \frac{\rho_s}{\sigma^2_s} \right), \]

(13)

\[ \rho_1 + \sigma^2_1 = \rho_2 + \sigma^2_2, \]

(14)

which yields the following optimal positive rate:

\[ L^*_1(0, \delta_1, \delta_2) = \frac{1}{2} \log \left( 1 + \frac{P}{\sigma^2_1} \right). \]

(15)

The distributions \( P_1 \) and \( P_2 \) satisfy (3) if the following holds:

\[ \rho_1 + \sigma^2_1 = \rho_2 + \sigma^2_2, \]

(14)

Notice the difference of this example with the Gaussian setup of [2, Section V]. In the proposed example, a positive rate can be achieved at each channel state. However, in the Gaussian example of [2, Section V], covert communication is feasible at zero rate. As discussed in the previous section, the difference comes from the fact that the goal of covert communication over a compound channel is to mask the state, while in the conventional model, the aim is to keep the adversary oblivious about the state of transmitter, i.e., whether or not a message is sent.

**Example 3:** We provide an example of Case 2, see Fig. 2. Here, we have \( \mathcal{X} = \{0, 1\} \) and \( \mathcal{Y} = \{0, 1, 2\} \). Notice that the unique distributions satisfying (3) are given by \( P_1(0) = P_2(0) = 1 \) and \( P_1(1) = P_2(1) = 0 \). Thus, the symbol “0” is considered as the off-symbol. Evaluating (10) for the proposed example with \( \delta_1 = 1 \) yields \( L^*_1(0, 1, 1) = 1.26 \).

**III. ENFORCING COVERTNESS BY CONSTRAINING THE TV DISTANCE BETWEEN THE OUTPUT DISTRIBUTIONS**

In the previous section, we defined the covertness with respect to an i.i.d. distribution \( Q_0^{\infty} \). In general, such an i.i.d. distribution may not exist. As discussed previously, the primary goal of the covert communication over a compound DMC is to mask the channel state. In this section, we change the covertness metric so that it is more pertinent to the compound channel. We define the covertness to be the total variation distance between the channel output marginals of the two states.

**A. System Model**

We suppose that \( \mathcal{X} = \{0, 1\} \) where 0 is the off-symbol. In this case, the output distribution induced by each input symbol is denoted by

\[ \tilde{Q}_1(\cdot) \triangleq W_1(\cdot | 1), \quad \tilde{Q}_2(\cdot) \triangleq W_2(\cdot | 1), \]

(16)

and

\[ Q_0(\cdot) \triangleq W_1(\cdot | 0) = W_2(\cdot | 0), \]

(17)
where we have assumed that the same output distribution is induced by the off-symbol at each channel state. As discussed in the previous section (see Case 2), this assumption ensures us that covert communication is possible over the compound channel in the sense that \(L^*_1(\epsilon, \delta_1, \delta_2) > 0\).

**Definition 2:** An \((|M|, n, \epsilon, \delta)\)-code consists of encoding and decoding functions \((f^{(n)}_1 : M \rightarrow X^n, f^{(n)}_2 : M \rightarrow X^n, g^{(n)} : Y^n \rightarrow M)\), such that the probability of error satisfies (6) and the total variation distance between \(Q^n_1\) and \(Q^n_2\) is upper bounded as

\[
d_{TV}(Q^n_1, Q^n_2) \leq \delta. \tag{18}
\]

For \((\epsilon, \delta) \in \mathbb{R}^2_+\), a covert rate \(L \in \mathbb{R}_+\) is said to be \((\epsilon, \delta)\)-achievable if there exists a sequence of \((|M|, n, \epsilon, \delta)\)-codes such that

\[
\liminf_{n \to \infty} \frac{\log |M|}{\sqrt{n}} \geq L. \tag{19}
\]

The supremum of all \((\epsilon, \delta)\)-achievable covert rates \(L\) is denoted as \(L^*(\epsilon, \delta)\).

**B. Results for \(L^*(\epsilon, \delta)\)**

In this section, we present upper and lower bounds on \(L^*(\epsilon, \delta)\) when the covertness metric is considered to be the total variation distance between the channel outputs.

**Theorem 1:** Define

\[
\Omega(\gamma_1, \gamma_2) \triangleq \gamma_1^2 \chi_2(Q_1||Q_0) + \gamma_2^2 \chi_2(Q_2||Q_0) - 2\gamma_1\gamma_2\rho(\tilde{Q}_1, \tilde{Q}_2, Q_0). \tag{20}
\]

where

\[
\rho(\tilde{Q}_1, \tilde{Q}_2, Q_0) \triangleq \mathbb{E}_{Q_0} \left[ \frac{(\tilde{Q}_1(Y) - Q_0(Y))(\tilde{Q}_2(Y) - Q_0(Y))}{Q_0(Y)^2} \right]. \tag{21}
\]

Then, \(L^*(\epsilon, \delta)\) is lower bounded as follows:

\[
L^*(\epsilon, \delta) \geq \max_{\gamma_1, \gamma_2} \min_{s \in \{1, 2\}} \gamma_s D(\tilde{Q}_s||Q_0), \tag{22}
\]

where the maximization is over all positive \(\gamma_1, \gamma_2\) such that

\[
\Omega(\gamma_1, \gamma_2) \leq 2\Phi^{-1}(\frac{1 + \delta}{2})^2. \tag{23}
\]

**Proof:** See Section III-C1. □

The solution of the optimization problem in (22) is provided in Appendix C.

**Theorem 2:** Assume that

\[
\min \{\chi_2(Q_1||Q_0), \chi_2(Q_2||Q_0)\} \geq \rho(\tilde{Q}_1, \tilde{Q}_2, Q_0). \tag{24}
\]

Then, \(L^*(0, \delta)\) is upper bounded as follows:

\[
L^*(0, \delta) \leq \max_{s \in \{1, 2\}} \frac{2\Phi^{-1}(\frac{1 + \delta}{2})}{\sqrt{\Delta}} \cdot D(\tilde{Q}_s||Q_0), \tag{25}
\]

where

\[
\Delta = \Delta(\tilde{Q}_1, \tilde{Q}_2, Q_0) \triangleq \mathbb{E}_{Q_0} \left[ \left( \frac{\tilde{Q}_1(Y) - \tilde{Q}_2(Y)}{Q_0(Y)} \right)^2 \right]. \tag{26}
\]

**Proof:** See Section III-C2. □

**Theorem 3 (Optimality Result):** Assume that

\[
\min \{\chi_2(Q_1||Q_0), \chi_2(Q_2||Q_0)\} \geq \rho(\tilde{Q}_1, \tilde{Q}_2, Q_0), \tag{27}
\]

and

\[
D(\tilde{Q}_1||Q_0) = D(\tilde{Q}_2||Q_0) \triangleq \Delta. \tag{28}
\]

Then,

\[
L^*(0, \delta) = \frac{2\Phi^{-1}(\frac{1 + \delta}{2})}{\sqrt{\Delta}} \cdot \Delta. \tag{29}
\]

**Proof:** The achievability follows from Theorem 1 and considering the fact that \(\gamma_1 = \gamma_2 = \frac{2}{\sqrt{\Delta}} \Phi^{-1}(\frac{1 + \delta}{2})\) satisfies inequality (23). The converse follows from Theorem 2 and the assumptions in (27) and (28). □

The setup of Example 3 satisfies the constraints (27) and (28). Choosing \(\delta = 0.2\) for this example, Theorem 3 yields \(L^*(0, 0.2) = 0.3399\). For other values of \(r \triangleq \tilde{Q}_2(0) = \tilde{Q}_2(1),\) the upper and lower bounds on \(L^*(0, 0.2)\) are plotted in Fig. 3. It can be verified that the two bounds match for \(r = 0.1\).

**C. Proofs**

Due to space constraints, only proof sketches are provided. Complete proofs can be found at [10].

1) **Proof of Theorem 1:** Fix a large blocklength \(n\) and choose positive \(\gamma_s\) such that (23) is satisfied. Define \(\mu_{s,n} \triangleq \frac{s}{n}\). We generate an i.i.d. codebook \(C_s = \{x^n_s(m) : m \in M\}\) according to the binary input distribution \(P_{s,n}\) where

\[
P_{s,n}(x) \triangleq \begin{cases} 
\mu_{s,n} & x = 1 \\
1 - \mu_{s,n} & x = 0 
\end{cases} \tag{30}
\]

For \(s \in \{1, 2\}\), let \(Q_{s,n}\) denote the output marginal of the channel \(W_s\) when the input distribution is \(P_{s,n}\). The distribution \(Q_{s,n}\) can be written as follows:

\[
Q_{s,n} = (1 - \mu_{s,n})Q_0 + \mu_{s,n}Q_s. \tag{31}
\]

Denote the \(n\)-fold product of \(Q_{s,n}\) by \(Q_{s,n}^n\). The transmitter upon observing the message \(m \in M\), sends the codeword
\( x^n(m) \) over the channel. Given \( y^n \), the decoder uses a threshold rule à la Feinstein’s rule [11, Lemma 3.4.1] which chooses the message \( m \) as follows:

\[
\frac{1}{\sqrt{n}} \log \frac{W_s(y^n|x^n(m))}{Q^n_s(y^n)} \geq \sqrt{n}I(P_{s,n}, W_s) + \xi,
\]

(32)

for \( \xi > 0 \) (arbitrarily small). The analysis of rate constraint is provided in Appendix A. We also note from [3] that

\[
I(P_{s,n}, W_s) = \frac{\gamma_n}{\sqrt{n}} D(\hat{Q}_s||Q_0) + \Theta \left( \frac{1}{n} \right),
\]

(33)

The following lemma provides a single-letter characterization of the covertness condition.

**Lemma 1:** We have

\[
d_{TV}(Q_1^n, Q_2^n) \leq 2\Phi \left( \frac{1}{2} \sqrt{\Omega(\gamma_1, \gamma_2)} \right) - 1 + O \left( \frac{1}{n} \right).
\]

(34)

**Proof:** See Appendix B.

Combining the rate constraint given in (32)–(33) and upper bounding (34) in Lemma 1 by \( \delta \) and rearranging yields the bounds in (22)–(23).

2) **Proof of Theorem 2:** For any two codebooks

\[
C_s = \{ x^n_s(m) : m \in M \}, \quad s \in \{1, 2\},
\]

(35)

we define the following parameters:

\[
\mu_{\ell,n} \triangleq \min_{s \in \{1, 2\}} \min_{m \in M} \frac{1}{n} w(x^n_s(m)),
\]

(36)

\[
\mu_{\ell,n} \triangleq \max_{s \in \{1, 2\}} \max_{m \in M} \frac{1}{n} w(x^n_s(m)),
\]

(37)

where \( w(x^n_s(m)) \) denotes the Hamming weight of the codeword \( x^n_s(m) \) and for \( s \in \{1, 2\} \), let

\[
\Gamma_s \triangleq \mathbb{E}_{Q_0} \left[ \frac{(\hat{Q}_s(Y) - Q_0(Y))(\hat{Q}_s(Y) - \hat{Q}_2(Y))^2}{Q_0(Y)^3} \right],
\]

(38a)

\[
D_s \triangleq \mathbb{E}_{Q_0} \left[ \frac{(\hat{Q}_s(Y) - Q_0(Y))(\hat{Q}_s(Y) - \hat{Q}_2(Y))}{Q_0(Y)^2} \right],
\]

(38b)

\[
V_s^n \triangleq \Delta + \mu_{\ell,n} \cdot |\Gamma_s|.
\]

(38c)

Notice that (24) implies that \( D_1 \geq 0 \) and \( D_2 \leq 0 \). Define

\[
\tau \triangleq \frac{n\mu_{\ell,n}}{2} (D_2 + D_1).
\]

(39)

We relate the total variation distance to a hypothesis testing problem as follows. Consider the following inequality:

\[
d_{TV}(Q_1^n, Q_2^n) \geq 1 - \alpha_n - \beta_n,
\]

(40)

where \( \alpha_n \) and \( \beta_n \) are false alarm and missed detection probabilities of a (possibly suboptimal) hypothesis test for the following setup. Under the null hypothesis \( H = 1 \), \( Y^n \) is distributed according to \( Q^n_1 \) and under the alternative hypothesis \( H = 2 \), it is distributed according to \( Q^n_2 \). Define the following suboptimal test:

\[
\mathcal{A}(y^n) \triangleq \left\{ \sum_{i=1}^{n} \mathcal{T}_{test}(y_i) > \tau \right\},
\]

(41)

where

\[
\mathcal{T}_{test}(y_i) \triangleq \frac{Q_1(y_i) - \hat{Q}_2(y_i)}{Q_0(y_i)}.
\]

(42)

The following lemma provides upper bounds on the false alarm and missed detection probabilities of the hypothesis test.

**Lemma 2:** We have:

\[
\alpha_n \leq 1 - \Phi \left( \frac{1}{2} \mu_{\ell,n} \sqrt{n}\Delta \right) + \frac{\sqrt{n}\mu_{\ell,n}\mu_{\ell,n} |\Gamma_1|}{4\sqrt{2}\pi\Delta} + O \left( \frac{1}{\sqrt{n}} \right),
\]

(43)

and

\[
\beta_n \leq 1 - \Phi \left( \frac{1}{2} \mu_{\ell,n} \sqrt{n}\Delta \right) + \frac{\sqrt{n}\mu_{\ell,n}\mu_{\ell,n} |\Gamma_2|}{4\sqrt{2}\pi\Delta} + O \left( \frac{1}{\sqrt{n}} \right).
\]

(44)

**Proof:** See Appendix D.

Combining (43) and (44) with (40), and defining \( \omega_{\ell,n} \triangleq n\mu_{\ell,n} \) and \( \omega_{H,n} \triangleq n\mu_{H,n} \) as the minimum and maximum Hamming weights of the codewords, we obtain

\[
d_{TV}(Q_1^n, Q_2^n) \geq 2\Phi \left( \frac{1}{2} \omega_{\ell,n} \sqrt{\frac{2\pi}{\Delta}} \right) - 1 - \omega_{\ell,n} \omega_{H,n} (|\Gamma_1| + |\Gamma_2|) + O \left( \frac{1}{\sqrt{n}} \right).
\]

(45)

The proof is followed by showing the existence of a subcodebook with size \( |M|_n \) such that the Hamming weight of codewords is smaller than \( 2\sqrt{2}\Phi^{-1} \left( \frac{1}{2} + \tau \right) + O \left( \frac{1}{\sqrt{n}} \right) \). It is concluded by applying Fano’s inequality using this subcodebook (see Appendix E).

**References**


APPENDIX A

ANALYSIS OF RATE CONSTRAINT FOR THEOREM 1

The analysis of the rate constraint uses Feinstein’s theorem [11, Lemma 3.4.1] which yields the following:

\[
\liminf_{n \to \infty} \frac{\log |\mathcal{M}|}{\sqrt{n}} \leq p \liminf_{n \to \infty} \frac{1}{\sqrt{n}} \log \frac{W_s(Y^n | X^n)}{Q_s^n(Y^n)}. \tag{46}
\]

The expectation of the above term is given in the following:

\[
\mathbb{E} \left[ \frac{1}{\sqrt{n}} \log \frac{W_s(Y^n | X^n)}{Q_s^n(Y^n)} \right] = \sqrt{n}I(P_{s,n}, W_s). \tag{47}
\]

One can show that as \( n \to \infty \):

\[
\frac{1}{\sqrt{n}} \log \frac{W_s(Y^n | X^n)}{Q_s^n(Y^n)} - \sqrt{n}I(P_{s,n}, W_s) \xrightarrow{p} 0, \tag{48}
\]

where \( \xrightarrow{p} \) means convergence in probability. To prove this, we calculate the following variance:

\[
\mathbb{V} \left[ \frac{1}{\sqrt{n}} \log \frac{W_s(Y^n | X^n)}{Q_s^n(Y^n)} \right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{V} \left[ \frac{W_s(Y_i | X_i)}{Q_s^n(Y_i)} \right] = \mathbb{V} \left[ \log \frac{W_s(Y_i | X_i)}{Q_s^n(Y_i)} \right] \tag{49}
\]

\[
\leq \mathbb{E}_{P_{s,n}W_s} \left[ \left( \frac{W_s(Y_i | X_i)}{Q_s^n(Y_i)} \right)^2 \right] = P_{s,n}(0) \mathbb{E}_{Q_{s,0}} \left[ \left( \frac{Q_0(Y)}{Q_s^n(Y)} \right)^2 \right] \tag{51}
\]

\[
+ \sum_{x \neq 0} P_{s,n}(x) \mathbb{E}_{W_s(.|x)} \left[ \left( \frac{W_s(Y|x)}{Q_s^n(Y)} \right)^2 \right]. \tag{52}
\]

Notice that

\[
P_{s,n}(0) \to 1, \quad \text{and} \quad P_{s,n}(x) \to 0, \quad \forall x \neq 0, \tag{53}
\]

yields the following:

\[
Q_{s,n} \to Q_0. \tag{54}
\]

Combining the above with (52) implies that the variance of estimation in (48) tends to zero as \( n \) goes to infinity. Thus, by Chebyshev’s inequality, we obtain

\[
\liminf_{n \to \infty} \frac{\log |\mathcal{M}|}{\sqrt{n}} \leq \liminf_{n \to \infty} \sqrt{n}I(P_{s,n}, W_s). \tag{55}
\]

APPENDIX B

PROOF OF LEMMA 1

Consider the following identity for the total variation distance:

\[
d_{TV}(Q_1^n, Q_2^n) = P_{Q_1^n} \left( \log \frac{Q_1^n}{Q_2^n} \geq 0 \right) - P_{Q_2^n} \left( \log \frac{Q_1^n}{Q_2^n} \geq 0 \right) \tag{56}
\]

\[
= P_{Q_1^n} \left( \sum_{i=1}^{n} \log \frac{Q_1(Y_i)}{Q_2(Y_i)} \geq 0 \right) - P_{Q_2^n} \left( \sum_{i=1}^{n} \log \frac{Q_1(Y_i)}{Q_2(Y_i)} \geq 0 \right). \tag{57}
\]

Now, we bound each of the probabilities in (57). First, we analyze the first probability. The mean and variance of \( \sum_{i=1}^{n} \log \frac{Q_1(Y_i)}{Q_2(Y_i)} \) are given by the following:

\[
\mathbb{E} \left[ \sum_{i=1}^{n} \log \frac{Q_1(Y_i)}{Q_2(Y_i)} \right] = nD(Q_1 || Q_2) \triangleq nd_1, \tag{58}
\]

and

\[
\mathbb{V} \left[ \sum_{i=1}^{n} \log \frac{Q_1(Y_i)}{Q_2(Y_i)} \right] = n \left( \sum_{y} Q_1(y) \left( \log \frac{Q_1(y)}{Q_2(y)} \right)^2 - D^2(Q_1 || Q_2) \right) \tag{59}
\]

\[
\triangleq n\nu_1. \tag{60}
\]

Employing the Berry-Esseen theorem [12, Theorem 1.6], we obtain

\[
P_{Q_1^n} \left( \sum_{i=1}^{n} \log \frac{Q_1(Y_i)}{Q_2(Y_i)} \geq 0 \right) \leq \Phi \left( \frac{\sqrt{nd_1}}{\sqrt{\nu_1}} \right) + \frac{6t_1}{\sqrt{\nu_1}}, \tag{61}
\]

where

\[
t_1 \triangleq \mathbb{E}_{Q_1} \left[ \left| \log \frac{Q_1(Y)}{Q_2(Y)} - D(Q_1 || Q_2) \right|^3 \right]. \tag{62}
\]

Similarly, we get:

\[
P_{Q_2^n} \left( \sum_{i=1}^{n} \log \frac{Q_1(Y_i)}{Q_2(Y_i)} \geq 0 \right) \geq 1 - \Phi \left( \frac{\sqrt{nd_2}}{\sqrt{\nu_2}} \right) - \frac{6t_2}{\sqrt{\nu_2}}, \tag{63}
\]

where we define:

\[
d_2 \triangleq D(Q_2 || Q_1), \tag{64}
\]

\[
\nu_2 \triangleq \sum_{y} Q_2(y) \left( \log \frac{Q_1(y)}{Q_2(y)} \right)^2 - D^2(Q_2 || Q_1), \tag{65}
\]

\[
t_2 \triangleq \mathbb{E}_{Q_2} \left[ \left| \log \frac{Q_1(Y)}{Q_2(Y)} + D(Q_2 || Q_1) \right|^3 \right]. \tag{66}
\]
Now, consider the following set of approximations:
\begin{align}
d_1 &= D(Q_1 || Q_2) \\
 &= \sum_y Q_1(y) \log \frac{Q_1(y)}{Q_2(y)} \\
 &= \frac{1}{2} \mathbb{E}_{Q_2} \left[ \left( \frac{\mu_{1,n}(\hat{Q}_1 - Q_0) - \mu_{2,n}(\hat{Q}_2 - Q_0)}{(1 - \mu_{2,n})Q_0 + \mu_{2,n}Q_2} \right)^2 \right] \\
&\quad + o(\mu_{1,n}^2 + \mu_{2,n}^2),
\end{align}
(67)

where (75) follows because \( (75) \) follows because we have:
\begin{align}
\sum_y Q_1(y) \log \frac{Q_1(y)}{Q_2(y)} &= \sum_y Q_1(y) \log \frac{Q_1(y)}{\frac{Q_1(y) + \hat{Q}_1(y)}{2}} \\
&= \sum_y \left( (1 - \mu_{1,n})Q_0(y) + \mu_{1,n}\hat{Q}_1(y) \right) \log \frac{\left( (1 - \mu_{1,n})Q_0(y) + \mu_{1,n}\hat{Q}_1(y) \right)}{\left( (1 - \mu_{2,n})Q_0(y) + \mu_{2,n}\hat{Q}_2(y) \right)} \\
&\quad \times \left( (1 - \mu_{1,n})Q_0(y) + \mu_{1,n}\hat{Q}_1(y) \right) \\
&= \mathbb{E}_{Q_0} \left[ \left( \frac{\mu_{1,n}(\hat{Q}_1 - Q_0) - \mu_{2,n}(\hat{Q}_2 - Q_0)}{Q_0} \right)^2 \right] \\
&\quad + o(\mu_{1,n}^2 + \mu_{2,n}^2),
\end{align}
(69)

as \( \mu_{s,n} \to 0 \). Combining (61), (71), (77) and definition of \( v_1 \) in (60), we have:
\begin{align}
\Phi \left( \frac{\sqrt{n}d_1}{\sqrt{\eta_1}} \right) &= \Phi \left( \frac{1}{2} \sqrt{\frac{n\tilde{\Omega}(\mu_{1,n}, \mu_{2,n})}{\eta_1}} \right) \\
&\quad + o(\mu_{1,n}^2 + \mu_{2,n}^2).
\end{align}
(78)

Similarly,
\begin{align}
\Phi \left( \frac{\sqrt{n}d_2}{\sqrt{\eta_2}} \right) &= \Phi \left( \frac{1}{2} \sqrt{\frac{n\tilde{\Omega}(\mu_{1,n}, \mu_{2,n})}{\eta_2}} \right) \\
&\quad + o(\mu_{1,n}^2 + \mu_{2,n}^2).
\end{align}
(79)

Combining (57), (61), (63), (78) and (79), we obtain the following:
\begin{align}
d_{TV}(Q_1^{\times n}, Q_2^{\times n}) &\leq \Phi \left( \frac{\sqrt{n}d_1}{\sqrt{\eta_1}} \right) + \Phi \left( \frac{\sqrt{n}d_2}{\sqrt{\eta_2}} \right) - 1 \\
&\quad + \frac{6\delta_1}{\sqrt{n\eta_1^2}} + \frac{6\delta_2}{\sqrt{n\eta_2^2}} \\
&\leq 2\Phi \left( \frac{1}{2} \sqrt{\frac{n\tilde{\Omega}(\mu_{1,n}, \mu_{2,n})}{\eta_1}} \right) - 1 \\
&\quad + O \left( \frac{1}{\sqrt{n}} \right).
\end{align}
(81)

Considering the fact that \( \gamma_s = \sqrt{n}\mu_{s,n} \), completes the proof of the lemma.

\section*{APPENDIX C}
\section*{SOLUTION OF THE OPTIMIZATION PROBLEM (22)}

Define:
\begin{align}
a_1 &\triangleq \chi_2(\hat{Q}_1 || Q_0), \\
a_2 &\triangleq \chi_2(\hat{Q}_2 || Q_0), \\
b &\triangleq -2\mathbb{E}_{Q_0} \left[ (\hat{Q}_1 - Q_0) \cdot (\hat{Q}_2 - Q_0) - \hat{Q}_1 \cdot \hat{Q}_2 - Q_0 \cdot Q_0 \right], \\
c &\triangleq 2\Phi^{-1} \left( \frac{1}{2} \right)^2,
\end{align}
(84)

Thus, the inequality in (23) can be written as follows:
\begin{align}
a_1\gamma_1^2 + a_2\gamma_2^2 + b\gamma_1\gamma_2 &\leq c.
\end{align}
(88)

The optimization problem in (22) is given by the following:
\begin{align}
\max_{\gamma_1, \gamma_2} \min_{s} \gamma_s d_s, \\
\text{s.t. : } a_1\gamma_1^2 + a_2\gamma_2^2 + b\gamma_1\gamma_2 &\leq c.
\end{align}
(89)

The above optimization problem is equivalently written as:
\begin{align}
\max_{\gamma_1, \gamma_2} \min_{0 \leq \xi \leq 1} \xi_1 d_1 + (1 - \xi)\gamma_2 d_2, \\
\text{s.t. : } a_1\gamma_1^2 + a_2\gamma_2^2 + b\gamma_1\gamma_2 &\leq c.
\end{align}
(91)

Since the above optimization problem is convex, we have:
\begin{align}
\min_{0 \leq \xi \leq 1} \max_{\gamma_1, \gamma_2} \xi_1 d_1 + (1 - \xi)\gamma_2 d_2, \\
\text{s.t. : } a_1\gamma_1^2 + a_2\gamma_2^2 + b\gamma_1\gamma_2 &\leq c.
\end{align}
(93)
Define
\[
G = \xi \gamma_1 d_1 + (1 - \xi) \gamma_2 d_2 - \lambda (a_1 \gamma_1^2 + a_2 \gamma_2^2 + b \gamma_1 \gamma_2 - c).
\]
(95)
Taking the derivative of \(G\) with respect to \(\gamma_1\) and \(\gamma_2\) and letting it be zero to get the optimal values \(\gamma_1^*\) and \(\gamma_2^*\), we obtain
\[
2a_1 \gamma_1^* + b \gamma_2^* = \frac{\xi d_1}{\lambda},
\]
(96)
\[
2a_2 \gamma_2^* + b \gamma_1^* = \frac{(1 - \xi) d_2}{\lambda}.
\]
(97)
This yields:
\[
\gamma_1^* = \frac{2a_2 \xi d_1 - b(1 - \xi) d_2}{\lambda(4a_1 a_2 - b^2)} \triangleq \tilde{\gamma}_1,
\]
(98)
\[
\gamma_2^* = \frac{2a_1 (1 - \xi) d_2 - \xi bd_1}{\lambda(4a_1 a_2 - b^2)} \triangleq \tilde{\gamma}_2.
\]
(99)
From (92), we know that
\[
\lambda \geq \frac{a_1 \tilde{\gamma}_1^2 + a_2 \tilde{\gamma}_2^2 + b \gamma_1 \gamma_2}{c}.
\]
(100)
Thus, we have:
\[
\min_{0 \leq \xi \leq 1} \max_{\gamma_1, \gamma_2} \xi \gamma_1 d_1 + (1 - \xi) \gamma_2 d_2 \\
= \min_{0 \leq \xi \leq 1} \sqrt{\frac{c}{a_1 \tilde{\gamma}_1^2 + a_2 \tilde{\gamma}_2^2 + b \gamma_1 \gamma_2}} (\xi \gamma_1 d_1 + (1 - \xi) \gamma_2 d_2).
\]
\[
= \min_{0 \leq \xi \leq 1} \frac{4c ((1 - \xi) a_1 d_2^2 - \xi (1 - \xi) bd_1 d_2 + \xi^2 a_2 d_1^2)}{(4a_1 a_2 - b^2)} \]
\[
= \left\{ \begin{array}{ll}
d_1 d_2 \sqrt{c} & b > -2 \min \{ \frac{a_1 d_2}{d_1}, \frac{a_2 d_1}{d_2} \} \\
\frac{4c \min \{ a_1 d_2, a_2 d_1 \}}{4a_1 a_2 - b^2} & b < -2 \min \{ \frac{a_1 d_2}{d_1}, \frac{a_2 d_1}{d_2} \}
\end{array} \right.
\]
(101)
\[
\alpha_n \leq \max_{\pi_X} \mathbb{P}_{W_1^n(\cdot|x^n)} \left[ \sum_{i=1}^{n} \mathbb{T}(Y_i) \leq \tau \right],
\]
(109)
where \(x^n\) above refers to any vector with type \(\pi_X\). Denoting the maximizing type in (109) by \(\pi^*\), (109) can be written as follows:
\[
\alpha_n \leq \mathbb{P}_{W_1^n(\cdot|x^n)} \left[ \sum_{i=1}^{n} \mathbb{T}(Y_i) \leq \tau \right],
\]
(110)
Since \(\mathbb{P}_{W_1^n(\cdot|x^n)} \left[ \sum_{i=1}^{n} \mathbb{T}(Y_i) \leq \tau \right]\) remains the same for each \(x^n\) with the same type, and \(\{P_{X^n}^{\tau} \{C_{s,\pi_X}\} \}_{\pi_X} \in \mathbb{P}^n\) is a probability distribution, we have
\[
\alpha_n \leq \max_{\pi_X} \mathbb{P}_{W_1^n(\cdot|x^n)} \left[ \sum_{i=1}^{n} \mathbb{T}(Y_i) \leq \tau \right],
\]
(111)
where \(x^n\) is any vector with type \(\pi^*\). Notice that \(\pi^*\) relates to the Hamming weight of the codeword \(x^n\) as follows:
\[
\pi^*(x) = \left\{ \begin{array}{ll}
\frac{w(x^n)}{n} & x = 1 \\
1 - \frac{w(x^n)}{n} & x = 0
\end{array} \right.
\]
(112)
\[
\mu_{1,n} = \frac{w(x^n)}{n}.
\]
(113)
Since the channel is memoryless, \(\{\mathbb{T}(Y_i)\}_{i=1}^{n}\) are mutually independent so the Berry-Esseen theorem [12, Theorem 1.6] can be applied to upper bound the probability in (110). For every \(x^n\) with type \(\pi^*\), we calculate \(\sum_{i=1}^{n} \mathbb{E}_{W_1^n(\cdot|x^n)}(\mathbb{T}(Y_i))\) and \(\sum_{i=1}^{n} \mathbb{V}(\mathbb{T}(Y_i))\) in the following. First, we calculate the expectation as follows,
\[
\sum_{i=1}^{n} \mathbb{E}_{W_1^n(\cdot|x^n)}(\mathbb{T}(Y_i))
\]
(114)
\[
\sum_{i=1}^{n} \mathbb{E}_{W_1^n(\cdot|x^n)}(\hat{Q}_1(Y_i) - \hat{Q}_2(Y_i)) \cdot \mathbb{Q}_0(Y_i)
\]
(115)
\[
n^2 \mu_{1,n} \sum_{y} \hat{Q}_1(y) \hat{Q}_2(y) - \hat{Q}_2(y) \hat{Q}_1(y) \mathbb{Q}_0(y)
\]
(116)
\[ = n\mu_{1,n} D_1, \]  
where (115) follows because when \( x_i^* = 0 \), we have \( W_i(Y_i | x_i^*) = Q_0(Y_i) \) and the expectation in the summand is equal to zero. Next, we calculate the variance as follows

\[
\sum_{i=1}^{n} V[\mathcal{T}_{\text{test}}(Y_i)] = \sum_{i=1}^{n} E \left[ (\mathcal{T}_{\text{test}}(Y_i))^2 \right] - \sum_{i=1}^{n} (E[\mathcal{T}_{\text{test}}(Y_i)])^2,
\]

(119)

Now, we calculate each expectation term in (119) as follows,

\[
\sum_{i=1}^{n} E \left[ (\mathcal{T}_{\text{test}}(Y_i))^2 \right] = \sum_{i=1}^{n} \mathbb{E}_{W_i(\cdot | x_i^*)} \left[ (\mathcal{T}_{\text{test}}(Y_i))^2 \right]
\]

(120)

\[
= \sum_{i:x_i^* \neq 1} \mathbb{E}_{\tilde{Q}_i} \left[ \left( \frac{\tilde{Q}_1(Y_i) - \tilde{Q}_2(Y_i)}{Q_0(Y_i)} \right)^2 \right] + \sum_{i:x_i^* = 0} \mathbb{E}_{\tilde{Q}_0} \left[ \left( \frac{\tilde{Q}_1(Y_i) - \tilde{Q}_2(Y_i)}{Q_0(Y_i)} \right)^2 \right]
\]

(121)

\[
= n\mu_{1,n} \sum_{y} \frac{\tilde{Q}_1(y)(\tilde{Q}_1(y) - \tilde{Q}_2(y))^2}{Q_0(y)} + n(1 - \mu_{1,n}) \Delta,
\]

(122)

and

\[
\sum_{i=1}^{n} (E[\mathcal{T}_{\text{test}}(Y_i)])^2 = \sum_{i=1}^{n} \left( \mathbb{E}_{W_i(\cdot | x_i^*)} [\mathcal{T}_{\text{test}}(Y_i)] \right)^2
\]

(123)

\[
= \sum_{i=1}^{n} \left( \mathbb{E}_{W_i(\cdot | x_i^*)} \left[ \frac{\tilde{Q}_1(Y_i) - \tilde{Q}_2(Y_i)}{Q_0(Y_i)} \right] \right)^2
\]

(124)

\[
= \sum_{i:x_i^* \neq 1} \mathbb{E}_{\tilde{Q}_i} \left[ \left( \frac{\tilde{Q}_1(Y_i) - \tilde{Q}_2(Y_i)}{Q_0(Y_i)} \right)^2 \right] + \sum_{i:x_i^* = 0} \mathbb{E}_{\tilde{Q}_0} \left[ \left( \frac{\tilde{Q}_1(Y_i) - \tilde{Q}_2(Y_i)}{Q_0(Y_i)} \right)^2 \right]
\]

(125)

\[
= n\mu_{1,n} \sum_{y} \frac{\tilde{Q}_1(y)(\tilde{Q}_1(y) - \tilde{Q}_2(y))^2}{Q_0(y)} + n(1 - \mu_{1,n}) \Delta
\]

(126)

Combining (119), (123) and (129), we get

\[
V[Z_n(x^{*n})] = n \left( \mu_{1,n} \sum_{y} \frac{\tilde{Q}_1(y)(\tilde{Q}_1(y) - \tilde{Q}_2(y))^2}{Q_0(y)} + (1 - \mu_{1,n}) \Delta - \mu_{1,n} D_1^2 \right)
\]

(130)

\[
\triangleq nV_{1,n}.
\]

(131)

We now bound the sum of the third absolute moments

\[
\sum_{i=1}^{n} \mathbb{E}_{W_i(\cdot | x_i^*)} \left[ (\mathcal{T}_{\text{test}}(Y_i) - \mathbb{E}[\mathcal{T}_{\text{test}}(Y_i)])^3 \right].
\]

(132)

We know that \( \mathcal{T}_{\text{test}}(y_i) = \frac{\tilde{Q}_1(y_i) - \tilde{Q}_2(y_i)}{q_{0i}(y_i)} \). Hence,

\[
|\mathcal{T}_{\text{test}}(y_i)| \leq \frac{2}{\min_{y_i} Q_0(y_i)} \triangleq 2/\eta < \infty.
\]

(133)

By the triangle inequality,

\[
|\mathcal{T}_{\text{test}}(Y_i) - \mathbb{E}[\mathcal{T}_{\text{test}}(Y_i)]| \leq 4/\eta, \text{ a.s.}
\]

(134)

Combining (132) and (134), we obtain

\[
\sum_{i=1}^{n} \mathbb{E}_{W_i(\cdot | x_i^*)} \left[ (\mathcal{T}_{\text{test}}(Y_i) - \mathbb{E}[\mathcal{T}_{\text{test}}(Y_i)])^3 \right]
\]

(135)

\[
\leq \frac{64n}{\eta^3} \triangleq nT.
\]

Notice that \( V_{1,n} \) in (131) is further upper bounded as follows:

\[
V_{1,n} \leq \left( \mu_{1,n} \sum_{y} \frac{\tilde{Q}_1(y)(\tilde{Q}_1(y) - \tilde{Q}_2(y))^2}{Q_0(y)} + (1 - \mu_{1,n}) \Delta \right)
\]

(136)

\[
= \Delta + \mu_{1,n} \Gamma_1
\]

(137)

\[
\leq \Delta + \mu_{1,n} |\Gamma_1|
\]

(138)

\[
\triangleq V_{1,n}^*.
\]

(139)

Thus, from the Berry-Esseen theorem [12, Theorem 1.6], we get the following:

\[
\mathbb{P} \left[ \sum_{i=1}^{n} \mathcal{T}_{\text{test}}(Y_i) \leq \tau \right] \leq 1 - \Phi \left( -\tau + n\mu_{1,n} D_1 \sqrt{nV_{1,n}} \right) + \frac{6T}{V_{1,n}^* \sqrt{n}}.
\]

(140)

Combining (109) and (140), we get the following:

\[
\alpha_n \leq \max_{\mu_{1,n}} 1 - \Phi \left( \frac{-\tau + n\mu_{1,n} D_1 \sqrt{nV_{1,n}}}{\sqrt{nV_{1,n}}} \right) + O \left( \frac{1}{\sqrt{n}} \right),
\]

(141)

where the maximization is over all \( \mu_{1,n} \) such that \( \mu_{1,n} \leq \mu_{n} \leq \mu_{H,n} \). Now, consider the missed detection probability as follows:

\[
\beta_n = \mathbb{P}_{Q^2} \left[ \sum_{i=1}^{n} \mathcal{T}_{\text{test}}(Y_i) > \tau \right].
\]

(142)
Following similar steps leading to (109), we can write the missed detection probability as:

$$\beta_n \leq \max_{\mu_{2,n}} 1 - \Phi \left( \frac{\tau - n\mu_{2,n}D_2}{\sqrt{nV_{2,n}}} \right) + \frac{6T}{V_{2,n}^2 \sqrt{n}}.$$  

(143)

where the maximization is over all \(\mu_{2,n}\) such that \(\mu_{l,n} \leq \mu_{2,n} \leq \mu_{l,n}\) and we define:

$$V_{2,n} \triangleq \left( \mu_{2,n} \sum_y \tilde{Q}_2(y)(\tilde{Q}_1(y) - \tilde{Q}_2(y))^2 + (1 - \mu_{2,n})\Delta - \mu_{2,n}\left( \sum_y \tilde{Q}_2(y)(\tilde{Q}_1(y) - \tilde{Q}_2(y))^2 \right) \right).$$  

(144)

We can further upper bound \(V_{2,n}\) as follows:

$$V_{2,n} \leq \mu_{2,n} \sum_y \tilde{Q}_2(y)(\tilde{Q}_1(y) - \tilde{Q}_2(y))^2 + (1 - \mu_{2,n})\Delta$$  

(145)

$$= \Delta + \mu_{2,n}\Gamma_2$$  

(146)

$$\leq \Delta + \mu_{l,n}|\Gamma_2|$$  

(147)

$$\triangleq V_{2,n}.$$  

(148)

The proof is followed by upper bounding the false alarm and missed detection probabilities using the choice of \(\tau\) in (39):

$$\alpha_n \leq \max_{\mu_{l,n}} 1 - \Phi \left( \frac{-\tau + n\mu_{l,n}D_1}{\sqrt{nV_{1,n}}} \right) + O \left( \frac{1}{\sqrt{n}} \right)$$  

(149)

$$= \max_{\mu_{l,n}} 1 - \Phi \left( \frac{-n\mu_{l,n}D_1 + n\mu_{l,n}D_1}{\sqrt{nV_{1,n}}} \right) + O \left( \frac{1}{\sqrt{n}} \right)$$  

(150)

$$= 1 - \Phi \left( \frac{n\mu_{l,n}D_1}{\sqrt{nV_{1,n}}} \right) + O \left( \frac{1}{\sqrt{n}} \right)$$  

(151)

$$\leq 1 - \Phi \left( \frac{\sqrt{n}\mu_{l,n}(D_1 - D_2)}{2\sqrt{V_{1,n}}} \right) + O \left( \frac{1}{\sqrt{n}} \right)$$  

(152)

$$= 1 - \Phi \left( \frac{\sqrt{n}\mu_{l,n}\Delta}{2\sqrt{V_{1,n}}} \right) + O \left( \frac{1}{\sqrt{n}} \right)$$  

(153)

$$\leq 1 - \Phi \left( \frac{\sqrt{n}\mu_{l,n}\Delta}{2\sqrt{V_{1,n}}} \right) + O \left( \frac{1}{\sqrt{n}} \right)$$  

(154)

$$= 1 - \Phi \left( \frac{\sqrt{n}\mu_{l,n}\Delta}{2\sqrt{(\Delta + \mu_{l,n}|\Gamma_1|)}} \right) + O \left( \frac{1}{\sqrt{n}} \right)$$  

(155)

$$\leq 1 - \Phi \left( \frac{1}{\sqrt{n}} \Delta \left( \frac{1 - \frac{\mu_{l,n}|\Gamma_1|}{2\Delta}}{2\Delta} \right) + O \left( \frac{1}{\sqrt{n}} \right) \right)$$  

(156)

$$\leq 1 - \Phi \left( \frac{1}{\sqrt{n}} \Delta + \frac{\sqrt{n}\mu_{l,n}\mu_{H,n}|\Gamma_1|}{4\sqrt{2\pi}} \right) + O \left( \frac{1}{\sqrt{n}} \right),$$  

(157)

where

- (152) follows because \(D_1 \geq 0\);
- (153) follows because \(D_1 - D_2 = \Delta\);
- (154) follows because \(V_{1,n} \leq V_{1,n}^*\);
- (156) follows from \(\frac{1}{\sqrt{n\Delta}} \geq 1 - \frac{x}{\sqrt{2\pi}}\);
- (157) follows because \(1 - \Phi(x - y) \leq 1 - \Phi(x) + \frac{y}{\sqrt{2\pi}}\) for all \(0 < y < x\).

With the choice of \(\tau\) in (39), \(\beta_n\) in (141) can be upper bounded as follows:

$$\beta_n \leq \max_{\mu_{2,n}} 1 - \Phi \left( \frac{\tau - n\mu_{2,n}D_2}{\sqrt{nV_{2,n}}} \right) + O \left( \frac{1}{\sqrt{n}} \right)$$  

(158)

$$= \max_{\mu_{2,n}} 1 - \Phi \left( \frac{n\mu_{l,n}(-D_2 + D_1) + n(n\mu_{l,n} - \mu_{2,n})D_2}{\sqrt{nV_{2,n}}} \right)$$  

(159)

$$\leq 1 - \Phi \left( \frac{n\mu_{l,n}(-D_2 + D_1)}{\sqrt{nV_{2,n}}} \right) + O \left( \frac{1}{\sqrt{n}} \right)$$  

(160)

$$= 1 - \Phi \left( \frac{n\mu_{l,n}\Delta}{2\sqrt{V_{2,n}}} \right) + O \left( \frac{1}{\sqrt{n}} \right)$$  

(161)

$$\leq 1 - \Phi \left( \frac{n\mu_{l,n}\Delta}{2\sqrt{V_{2,n}}} \right) + O \left( \frac{1}{\sqrt{n}} \right)$$  

(162)

$$\leq 1 - \Phi \left( \frac{1}{2\mu_{l,n}\sqrt{n}\Delta} + \frac{\sqrt{n}\mu_{l,n}\mu_{H,n}|\Gamma_2|}{4\sqrt{2\pi}} \right) + O \left( \frac{1}{\sqrt{n}} \right),$$  

(163)

where

- (160) follows because \(D_2 \leq 0\) and hence, \((\mu_{l,n} - \mu_{2,n})D_2 \geq 0\);
- (161) follows because \(D_1 - D_2 = \Delta\);
- (162) follows because \(V_{2,n} \leq V_{2,n}^*\);
- (163) follows because \(V_{2,n}^* = \Delta + \mu_{H,n}|\Gamma_2|\) and also from the fact that \(1 - \Phi(x - y) \leq 1 - \Phi(x) + \frac{y}{\sqrt{2\pi}}\) for all \(0 < y < x\).

This completes the proof of lemma.

**APPENDIX E**

**EXISTENCE OF A SUB-CODEBOOK WITH SIZE \(|M|/\sqrt{n}\)**

In the following, we first show that there exists a sub-codebook which satisfies (45) and its size is at least \(|M|/\sqrt{n}\).

Define the following set of codewords for some \(\gamma > 0\):

$$D_s \triangleq \left\{ x^n_s \in C_s : w(x^n_s) \leq 2\sqrt{\frac{n}{\Delta}} \Phi^{-1} \left( \frac{1 + \delta}{2} + \gamma + \frac{E}{\sqrt{n}} \right) \right\},$$

(164)
where $E$ is chosen such that
\[ E > \left( \frac{\Gamma_1 + |\Gamma_2|}{8\sqrt{2\pi\Delta}} \right) \left( \frac{1 + \delta}{2} + \gamma \right)^2. \] (165)

Let $\hat{Q}_s^n$ and $\overline{Q}_s^n$ be the induced output distributions for codes $D_s$ and $C_s \setminus D_s$, respectively. The distribution $Q_s^n$ can be written as follows:
\[ Q_s^n = \xi_s \hat{Q}_s^n + (1 - \xi_s)\overline{Q}_s^n, \] (166)
where $\xi_s \triangleq \frac{|D_s|}{|M|}$. Without loss of generality, assume that $\xi_1 \geq \xi_2$. Then, we get the following:
\[ \delta \geq d_{TV}(Q_s^n, Q_2^n) \geq \frac{1}{2} \sum_{y^n} |Q_s^n - Q_2^n| \] (167)
\[ = \frac{1}{2} \sum_{y^n} \left| \xi_1 \hat{Q}_1^n + (1 - \xi_1)\overline{Q}_1^n - \xi_2 \hat{Q}_2^n - (1 - \xi_2)\overline{Q}_2^n \right| \] (168)
\[ \geq d_{TV}(\overline{Q}_1^n, \overline{Q}_2^n) - \xi_1 d_{TV}(\hat{Q}_1^n, \overline{Q}_1^n) - \xi_2 d_{TV}(\hat{Q}_2^n, \overline{Q}_2^n), \] (169)
where the last inequality follows from the triangle inequality.

We know that for any $x_s^n \in C_s \setminus D_s$,
\[ w(x_s^n) \geq 2\sqrt{\frac{n}{\Delta}} \Phi^{-1} \left( \frac{1 + \delta}{2} + \gamma + \frac{E}{\sqrt{n}} \right). \] (171)

Combining (45) with (171) and considering (165), we can further lower bound the total variation distance as follows:
\[ d_{TV}(\overline{Q}_1^n, \overline{Q}_2^n) \geq \delta + \frac{2E}{\sqrt{n}} + 2\gamma - \omega_{\Delta, nW/n}(|\Gamma_1| + |\Gamma_2|) \] (172)
\[ + O\left( \frac{1}{\sqrt{n}} \right) \]
\[ \geq \delta + 2\gamma, \] (173)
where the last inequality follows from (165). Uniting (170) and (173) yields
\[ \delta \geq \delta + 2\gamma - \xi_1 - \xi_2. \] (174)

If we choose $\gamma = \frac{1}{\sqrt{n}}$, we obtain
\[ \xi_1 + \xi_2 \geq \frac{2}{\sqrt{n}}. \] (175)

In summary, from the assumption $\xi_1 \geq \xi_2$, we obtain a set of codewords with size $|D_1| \geq \frac{|M|}{\sqrt{n}}$ and the Hamming weight of these codewords is given by
\[ w(x_1^n) \leq 2\sqrt{\frac{n}{\Delta}} \Phi^{-1} \left( \frac{1 + \delta}{2} + \frac{E}{\sqrt{n}} + \gamma \right) \] (176)
\[ = 2\sqrt{\frac{n}{\Delta}} \Phi^{-1} \left( \frac{1 + \delta}{2} \right) + O\left( \frac{1}{\sqrt{n}} \right) \] (177)
\[ \triangleq \psi(n, \delta) + O\left( \frac{1}{\sqrt{n}} \right). \] (178)

Using the sub-codebook $D_s$, we proceed to upper bound the rate in the following. Define the following set of codewords:
\[ D_{s,i} = \{ x_s^n \in D_s : w(x_s^n) = i \}. \] (179)

The codewords in the sub-codebook $|D_{s,i}|$ have the same type which we denote by $\pi^i_s$. This sub-codebook is with maximum probability of error not larger than $\epsilon$. Using Fano’s inequality, we can write the following set of inequalities for $s \in \{1, 2\}$:
\[ (1 - \epsilon) \log |D_{s,i}| - 1 \leq I(X^n; Y_s^s) \] (180)
\[ \leq \sum_{i=1}^n I(X_{s,i}; Y_i) \] (181)
\[ \leq nI(\pi^i_s, W_s) \] (182)
\[ = iD(\hat{Q}_s||Q_0). \] (183)

Next, we continue to upper bound the size of the message set as follows:
\[ \log \frac{|M|}{\sqrt{n}} \leq \max_s \log |D_s| \] (184)
\[ = \max_s \log \left( \sum_{i=0}^1 |D_{s,i}| \right) \] (185)
\[ \leq \max_s \log \left( \sum_{i=0}^1 2^{iD(\hat{Q}_s||Q_0)} \right) \] (186)
\[ \leq \max_s \log \left( \psi(n, \delta)2^{\psi(n, \delta)D(\hat{Q}_s||Q_0)} \right) \] (187)
\[ \leq \psi(n, \delta) \cdot \max_s D(\hat{Q}_s||Q_0) + O(1) \] (188)
\[ \leq 2\sqrt{\frac{n}{\Delta}} \Phi^{-1} \left( \frac{1 + \delta}{2} \right) \cdot \max_s D(\hat{Q}_s||Q_0) + O(1). \] (189)

Dividing both sides by $\sqrt{n}$ and taking limits in $n$ completes the proof.