# Broadcast Codes Can Be Enhanced to Perform Covert Communications

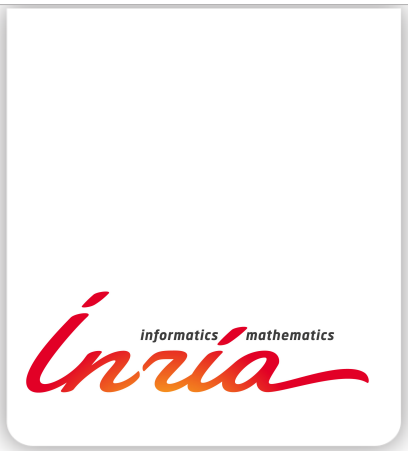
David Kibloff, Samir Perlaza, Ligong Wang


# Broadcast Codes Can Be Enhanced to Perform Covert Communications

**David Kibloff, Samir M. Perlaza and Ligong Wang**

**RESEARCH REPORT**
**N° 9249**
January 2019
Project-Team MARACAS

ISSN 0249-6399 ISRN INRIA/RR--9249--FR+ENG




# Broadcast Codes Can Be Enhanced to Perform Covert Communications

David Kibloff, Samir M. Perlaza and Ligong Wang

Project-Team MARACAS

Research Report n° 9249 — January 2019 — 70 pages

**Abstract:** Given a code used to send a message to two receivers through a degraded discrete memoryless broadcast channel (DM-BC), the sender wishes to alter the codewords to achieve the following goals: (*i*) the original broadcast communication continues to take place, possibly at the expense of a tolerable increase of the decoding error probability; and (*ii*) an additional covert message can be transmitted to the stronger receiver such that the weaker receiver cannot detect the existence of this message. The main results are: (*a*) feasibility of covert communications is proven by using a random coding argument for general DM-BCs; and (*b*) necessary conditions for establishing covert communications are described and an impossibility (converse) result is presented for a particular class of DM-BCs. Together, these results characterize the asymptotic fundamental limits of covert communications for this particular class of DM-BCs within an arbitrarily small gap.

**Key-words:** Covert Communication, Low-Probability of Detection, Information-Theoretic Security, Broadcast Channel.

---

D. Kibloff and S. M. Perlaza are with the Laboratoire CITI, a joint laboratory between the Institut National de Recherche en Informatique et en Automatique (INRIA), the Université de Lyon and the Institut National de Sciences Apliquées (INSA) de Lyon. 6 Av. des Arts 69621 Villeurbanne, France. (david.kibloff, samir.perlaza@inria.fr)

L. Wang is with the Laboratoire ETIS, a joint laboratory between the Université Paris Seine, Université de Cergy-Pontoise, ENSEA and CNRS. 6, avenue du Ponceau 95014 Cergy-Pontoise, France (e-mail: ligong.wang@ensea.fr).

This research was supported in part by the European Commission under Marie Sklodowska-Curie Individual Fellowship No. 659316 (CYBERNETS) and the French Ministry of Défense through the Direction Générale de l'Armement (DGA).

**RESEARCH CENTRE**
**GRENOBLE – RHÔNE-ALPES**

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

**Résumé :** Etant donné un code utilisé pour transmettre un message à deux récepteurs à travers un canal broadcast discret et sans mémoire, le transmetteur souhaite altérer les mots-code dans les buts suivants: $(i)$ continuer la transmission du message original, éventuellement au prix d'une augmentation de la probabilité d'erreur de décodage; et $(ii)$ transmettre un message additionnel furtif au récepteur le plus fort de sorte que le récepteur le plus faible ne puisse pas détecter l'existence de ce message. Les résultats principaux sont les suivants: $(a)$ la faisabilité de la communication furtive est prouvée en utilisant un argument de codages aléatoires; et $(b)$ les conditions nécessaires pour établir des communications furtives sont décrites et un résultat d'impossibilité est présenté pour une classe de canaux broadcast discrets sans mémoire. Ensemble, ces résultats caractérisent la limite asymptotique fondamentale des communications furtives pour cette classe particulière de canaux.

**Mots-clés :** Communications furtives, Sécurité de la couche physique, Canal Broadcast

# Contents

# 1 Introduction

Covert communications refer to scenarios in which legitimate parties aim at communicating while keeping an adversary unaware of the existence of the communication. In point-to-point channels, reliable covert communications are subject to a fundamental limit that states that only $O(\sqrt{n})$ bits can be transmitted in $n$ channel uses [1, 2, 3, 4].

Two different covert communication problems have been studied within the context of broadcast channels [5, 6, 7]. In [5], the sender tries to send two covert messages to two receivers. In [6] and [7], the sender sends a common non-covert message to both receivers, and tries to simultaneously send a covert message to one of the receivers. That is, the other receiver cannot know whether or not a covert message is being sent.

The current work is related to [6] and [7]. The focus is on the problem of embedding a covert message in a non-covert broadcast code. Some of the main differences between this problem and the one in [6] and [7] are:

- In [6] and [7], the non-covert broadcast code and the covert code are designed together by the transmitter. This potentially allows the transmitter to choose a non-covert code on which it is easy to embed a covert code. Alternatively, the current work assumes that the non-covert code is given and cannot be changed, making the achievability proof more difficult.[1]

- In [6] and [7] there is a separate covertness criterion conditional on every non-covert message. In this work, only one covertness criterion on the overall distribution is adopted. This difference considerably complicates the proof of the converse. In fact, a general proof of the converse using the Kullback-Leibler divergence as the covertness criterion is still an open problem. Alternatively, in this work, the total variation distance is used by adapting some techniques from [8]. Interestingly, the proof of the converse is shown to be tight for a class of channels satisfying certain symmetry properties.

In a nutshell, it is shown that in this scenario, it is possible to covertly transmit $O(\sqrt{n})$ bits in $n$ channel uses by modifying an existing broadcast code. Moreover, the proposed transmission rate is shown to be asymptotically optimal for a class of discrete memoryless broadcast channels (DM-BCs).

The remaining of this report is organized as follows. Section 2 and Section 3 present respectively the notation and the system model. Section 5 establishes preliminary results on the probability of detecting covert communications. Section 6 presents an achievability scheme. Section 4 exposes examples in which covert communications can not be achieved. Finally, Section 7 establishes a converse result and Section 9 concludes this work.

# 2 Notation

Throughout this report, random variables are denoted by uppercase letters, *e.g.* , $X$, and their realizations are denoted by lowercase letters, *e.g.* , $x$. Sets are denoted by calligraphic letters, *e.g.* , $\mathcal{X}$. The probability distribution of the random variable $X$ is denoted by $P_X$ unless specified otherwise. The expected value and the variance evaluated with respect to the probability distribution $P_X$ are respectively denoted by $\mathbb{E}_X[\cdot]$ and $\mathbb{V}_X[\cdot]$. The complementary cumulative distribution function of a standard Gaussian random variable evaluated at $x \in \mathbb{R}$ is denoted by $Q(x)$. Given two distributions $P_X$ and $Q_X$, $P_X \ll Q_X$ denotes the fact that $P_X$ is absolutely

[1] A technical condition is that the given non-covert code must have a positive error exponent; see (82).

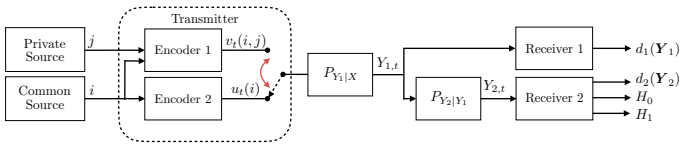



Figure 1: Degraded broadcast channel with covert messages at channel use $t \in \{1, 2, \ldots, n\}$, where $d_1 : \mathcal{Y}_1^n \to \mathcal{W} \times \hat{\mathcal{W}}$ denotes the decoding function at Receiver 1 and $d_2 : \mathcal{Y}_2^n \to \mathcal{W}$ denotes the decoding function at Receiver 2.

continuous with respect to $Q_X$. Assuming that the probability mass functions $P_X$ and $Q_X$ have countable support $\mathcal{X}$, the function $\chi_k\left(P_X, Q_X\right)$, with $k \in \mathbb{N}$, is

$$\chi_k\left(P_X, Q_X\right) = \sum_{x \in \mathcal{X}} \frac{\left(P_X(x) - Q_X(x)\right)^k}{Q_X(x)^{k-1}}. \tag{1}$$

Whenever a second random variable $Y$ is considered, $P_{X\,Y}$ and $P_{Y|X}$ denote respectively the joint probability distribution of the pair $(X, Y)$ and the conditional probability distribution of $Y$ given $X$. Given a realization $x \in \mathcal{X}$, the expected value and the variance evaluated with respect to the conditional probability distribution $P_{Y|X}(\cdot|x)$ (also denoted as $P_{Y|X=x}$) are respectively denoted by $\mathbb{E}_{Y|X=x}\left[\cdot\right]$ and $\mathbb{V}_{Y|X=x}\left[\cdot\right]$.

Given an integer $n$, an $n$-dimensional vector of random variables is denoted by a bold uppercase letter, *e.g.* , $\boldsymbol{X} = (X_1, X_2, ..., X_n) \in \mathcal{X}^n$ and its realization is denoted by a bold lowercase letter, *e.g.* , $\boldsymbol{x} = (x_1, x_2, ..., x_n)$. The number of occurrences of the symbol $x \in \mathcal{X}$ in the vector $\boldsymbol{x} \in \mathcal{X}^n$ is denoted by $N(x|\boldsymbol{x}) \triangleq \sum_{t=1}^n \mathbb{1}_{\{x=x_t\}}$. Similarly, the number of joint occurrences of the pair $(x, x') \in \mathcal{X}^2$ in the pair of vectors $(\boldsymbol{x}, \boldsymbol{x}') \in \mathcal{X}^{2n}$ is denoted by $N(x, x'|\boldsymbol{x}, \boldsymbol{x}') \triangleq \sum_{t=1}^n \mathbb{1}_{\{x=x_t\}} \mathbb{1}_{\left\{x'=x'_t\right\}}$.

# 3 System Model

Consider a three-party communication system in which a transmitter simultaneously sends information to two receivers through a noisy communication medium. In this work, the noisy communication medium is described by a product random transformation

$$(\mathcal{X}^n, \mathcal{Y}_1^n \times \mathcal{Y}_2^n, P_{\boldsymbol{Y}_1\boldsymbol{Y}_2|\boldsymbol{X}}), \tag{2a}$$

where $n \in \mathbb{N}$ is the block-length; the alphabets $\mathcal{X}$, $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are finite; and $\boldsymbol{Y}_1 = (Y_{1,1}, Y_{1,2}, \ldots, Y_{1,n}) \in \mathcal{Y}_1^n$, $\boldsymbol{Y}_2 = (Y_{2,1}, Y_{2,2}, \ldots, Y_{2,n}) \in \mathcal{Y}_2^n$ and $\boldsymbol{X} = (X_1, X_2, \ldots, X_n) \in \mathcal{X}^n$ are $n$-dimensional vectors of random variables. In particular, given an input $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, the output $(\boldsymbol{y}_1, \boldsymbol{y}_2)$ with $\boldsymbol{y}_k = (y_{k,1}, y_{k,2}, \ldots, y_{k,n})$ for all $k \in \{1, 2\}$ is observed with probability:

$$P_{\boldsymbol{Y}_1\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}_1, \boldsymbol{y}_2|\boldsymbol{x}) = \prod_{t=1}^n P_{Y_1|X}(y_{1,t}|x_t) P_{Y_2|Y_1}(y_{2,t}|y_{1,t}). \tag{2b}$$

That is, the channel is degraded and memoryless.

Given the random transformation in (2), the Transmitter is given a *broadcast code* (Encoder 2 in Figure 1) to transmit a message intended to both receivers at a fixed rate. Often, this message index is referred to as the *common message*. Section 3.1 formally defines these codes.

Each codeword of a broadcast code can be altered to generate a set of new codewords. Hence, by redefining the decoding sets at one receiver (Receiver 1 in Figure 1), it is possible to build a new code (Encoder 1 in Figure 1) to transmit two messages: *(a)* the common message at the same rate as the original broadcast code, possibly at the expense of a higher probability of error; and *(b)* a message exclusively intended to Receiver 1. Often, this message index is referred to as the *private message* and the new code is referred to as an *induced code*. These codes are formally defined in Section 3.2.
An induced code might satisfy some additional constraints on the transmission of the private message, e.g., a covertness constraint. A covertness constraint consists in rendering the non-intended receiver of the private message (Receiver 2) unable to determine whether or not a private message is being transmitted. That is, Receiver 2 is unable to determine whether the codeword being transmitted belongs to either the broadcast code or the induced code. An induced code that satisfies a covertness constraint is referred to as a *covert code* and is formally defined in Section 3.3.

The objective of this work is to determine the maximum information rate at which private messages can be transmitted by using a covert code induced from a given broadcast code.

## 3.1 Broadcast Codes

The common message index to be sent from the Transmitter to both receivers is a realization of a random variable $W$ that is uniformly distributed in the set

$$\mathcal{W} = \{1, 2, \ldots, M\}, \tag{3}$$

with $M \in \mathbb{N}$. To send a common message index within $n$ channel uses, the Transmitter uses an $(n, M)$-broadcast code.

**Definition 1** ($(n, M)$-broadcast code)**.** *Given $M \in \mathbb{N}$ and a block-length $n \in \mathbb{N}$, an $(n, M)$-broadcast code for the random transformation in* (2) *is a system*

$$\left\{\Big(\boldsymbol{u}(1), \mathcal{D}_1(1), \mathcal{D}_2(1)\Big), \Big(\boldsymbol{u}(2), \mathcal{D}_1(2), \mathcal{D}_2(2)\Big), \ldots, \Big(\boldsymbol{u}(M), \mathcal{D}_1(M), \mathcal{D}_2(M)\Big)\right\}, \tag{4}$$

*where for all $(i, j, k) \in \mathcal{W}^2 \times \{1, 2\}$, with $i \neq j$:*

$$\boldsymbol{u}(i) = (u_1(i), u_2(i), \ldots, u_n(i)) \in \mathcal{X}^n, \tag{5a}$$
$$\mathcal{D}_k(i) \cap \mathcal{D}_k(j) = \emptyset, \quad \textit{and} \tag{5b}$$
$$\bigcup_{l=1}^{M} \mathcal{D}_k(l) \subseteq \mathcal{Y}_k^n. \tag{5c}$$

*The vectors $\boldsymbol{u}(1)$, $\boldsymbol{u}(2)$, ..., $\boldsymbol{u}(M)$ and the sets $\mathcal{D}_k(1)$, $\mathcal{D}_k(2)$, ..., $\mathcal{D}_k(M)$ in* (4) *are respectively the codewords and the decoding sets at receiver $k$.*

Given a broadcast code represented by the system in (4), the Transmitter uses the codeword $\boldsymbol{u}(i)$ to transmit the message index $i \in \mathcal{W}$. At channel use $t$, with $t \in \{1, 2, \ldots, n\}$, the Transmitter inputs the symbol $u_t(i)$ to the channel. For all $k \in \{1, 2\}$, Receiver $k$ observes the output $\boldsymbol{y}_k = (y_{k,1}, y_{k,2}, \ldots, y_{k,n})$ after $n$ channel uses and determines that the symbol $i$ was transmitted if it satisfies the decoding rule:

$$\boldsymbol{y}_k \in \mathcal{D}_k(i). \tag{6}$$

The average decoding error probability associated to the given broadcast code at receiver $k$, denoted by $\lambda_k \in [0,1]$, is

$$\lambda_k = \frac{1}{M}\sum_{i=1}^{M} \Pr\left[\boldsymbol{Y}_k \in \mathcal{D}_k^{\mathsf{c}}(i) \middle| \boldsymbol{X} = \boldsymbol{u}(i)\right], \tag{7}$$

where the probability operator applies with respect to the marginal $P_{\boldsymbol{Y}_k|\boldsymbol{X}}$ of the joint distribution in (2b); and $\mathcal{D}_k^{\mathsf{c}}(i)$ in (7) represents the complement of $\mathcal{D}_k(i)$ with respect to $\mathcal{Y}_k^n$.

**Definition 2** ($(n, M, \epsilon)$-broadcast code)**.** *Let $\epsilon \in [0,1]$ be fixed and consider an $(n, M)$-broadcast code $\mathcal{C}$ described by (4). The broadcast code $\mathcal{C}$ is said to be an $(n, M, \epsilon)$-broadcast code if*

$$\max(\lambda_1, \lambda_2) < \epsilon. \tag{8}$$

## 3.2 Induced Codes

Let the private message index be represented by a random variable $\hat{W}$, independent of $W$ and uniformly distributed over

$$\hat{\mathcal{W}} = \{1, 2, \ldots, \hat{M}\}, \tag{9}$$

with $\hat{M} \in \mathbb{N}$. Assume that a broadcast code denoted by $\mathcal{C}$ is given and is represented by the system in (4). The transmitter uses an $(n, \mathcal{C}, \hat{M})$-induced code to transmit both the common and private message indices.

**Definition 3** ($(n, \mathcal{C}, \hat{M})$-induced code)**.** *Given $\hat{M} \in \mathbb{N}$ and an $(n, M)$-broadcast code $\mathcal{C}$ described by (4), an $(n, \mathcal{C}, \hat{M})$-induced code is a system*

$$\left\{ \left(\boldsymbol{v}(1,1), \mathcal{D}_1(1,1), \mathcal{D}_2(1)\right), \left(\boldsymbol{v}(1,2), \mathcal{D}_1(1,2), \mathcal{D}_2(1)\right), \ldots, \left(\boldsymbol{v}(M,\hat{M}), \mathcal{D}_1(M,\hat{M}), \mathcal{D}_2(M)\right) \right\}, \tag{10}$$

*where for all $(i,k,j,l) \in \mathcal{W}^2 \times \hat{\mathcal{W}}^2$, with $(i,j) \neq (k,l)$, the following holds:*

$$\boldsymbol{v}(i,j) = (v_1(i,j), v_2(i,j), \ldots, v_n(i,j)) \in \mathcal{X}^n, \tag{11a}$$

$$\mathcal{D}_1(i,j) \cap \mathcal{D}_1(k,l) = \emptyset, \tag{11b}$$

$$\bigcup_{p=1}^{M}\bigcup_{q=1}^{\hat{M}} \mathcal{D}_1(p,q) \subseteq \mathcal{Y}_1^n. \tag{11c}$$

*The vectors $\boldsymbol{v}(1,1)$, $\boldsymbol{v}(1,2)$, ..., $\boldsymbol{v}(M,\hat{M})$ are the codewords; the sets $\mathcal{D}_1(1,1)$, $\mathcal{D}_1(1,2)$, ..., $\mathcal{D}_1(M,\hat{M})$ are the decoding sets at Receiver 1; and the sets $\mathcal{D}_2(1)$, $\mathcal{D}_2(2)$, ..., $\mathcal{D}_2(M)$ are the same decoding sets at Receiver 2 as in the $(n, M)$-broadcast code $\mathcal{C}$.*

Given an $(n, \mathcal{C}, \hat{M})$-induced code denoted by $\hat{\mathcal{C}}$ and described by (10), the Transmitter uses the codeword $\boldsymbol{v}(i,j)$ to transmit the common message index $i \in \mathcal{W}$ and the private message index $j \in \hat{\mathcal{W}}$. At channel use $t$, with $t \in \{1, 2, \ldots, n\}$, the Transmitter inputs the symbol $v_t(i,j)$ to the channel. At the end of $n$ channel uses, Receiver $k$ observes the output $\boldsymbol{y}_k = (y_{k,1}, y_{k,2}, \ldots, y_{k,n})$, with $k \in \{1,2\}$. Receiver 1 declares that the pair $(i,j) \in \mathcal{W} \times \hat{\mathcal{W}}$ was transmitted if $(i,j)$ satisfies the decoding rule:

$$\boldsymbol{y}_1 \in \mathcal{D}_1(i,j). \tag{12}$$

Alternatively, the decoding rule of Receiver 2 remains being that in (6), with $k = 2$, *i.e.*, the same as in the broadcast code $\mathcal{C}$.

The average decoding error probability associated to the induced code $\hat{\mathcal{C}}$ at receiver $k$ is denoted by $\hat{\lambda}_k \in [0,1]$, with $k \in \{1,2\}$. That is,

$$\hat{\lambda}_1 = \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i,j) | \boldsymbol{X} = \boldsymbol{v}(i,j)\right], \text{ and} \tag{13}$$

$$\hat{\lambda}_2 = \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \Pr\left[\boldsymbol{Y}_2 \in \mathcal{D}_2^{\mathsf{c}}(i) | \boldsymbol{X} = \boldsymbol{v}(i,j)\right], \tag{14}$$

where the probability operators apply with respect to the conditional marginals $P_{\boldsymbol{Y}_1|\boldsymbol{X}}$ and $P_{\boldsymbol{Y}_2|\boldsymbol{X}}$ of the joint distribution in (2b), respectively. The sets $\mathcal{D}_1^{\mathsf{c}}(i,j)$ and $\mathcal{D}_2^{\mathsf{c}}(i)$ represent the complement of $\mathcal{D}_1(i,j)$ and $\mathcal{D}_2(i)$ with respect to $\mathcal{Y}_1^n$ and $\mathcal{Y}_2^n$, respectively. Using this notation, the definition of an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code is presented hereunder.

**Definition 4** ($(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code)**.** *Let $\hat{\epsilon} \in [0,1]$ be fixed. Consider an $(n, M)$-broadcast code $\mathcal{C}$ described by* (4) *and an $(n, \mathcal{C}, \hat{M})$-induced code $\hat{\mathcal{C}}$ described by* (10)*. The induced code $\hat{\mathcal{C}}$ is said to be an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code if* $\max(\hat{\lambda}_1, \hat{\lambda}_2) < \hat{\epsilon}$.

In order to guarantee that for all $\boldsymbol{y}_k \in \mathcal{Y}_k^n$, with $k \in \{1,2\}$, there always exists a message index $i \in \mathcal{W}$ that satisfies the decoding rule (6), the inclusion in (5c) is assumed with equality. Note that in the case in which the set $\mathcal{Y}_k^n \setminus (\mathcal{D}_k(1) \cup \mathcal{D}_k(2) \cup \ldots \cup \mathcal{D}_k(M))$ is not empty, the channel output vectors therein always induce a decoding error at receiver $k$. Therefore, given any $j \in \mathcal{W}$, replacing the set $\mathcal{D}_k(j)$ by $\mathcal{D}'_k(j) = \mathcal{D}_k(j) \cup (\mathcal{Y}_k^n \setminus (\mathcal{D}_k(1) \cup \mathcal{D}_k(2) \cup \ldots \cup \mathcal{D}_k(M))$ does not increase the average decoding error probability. Thus, there is no loss of generality in studying a system in which (5c) holds with equality. Without any loss of generality, the inclusion in (11c) is assumed with equality for an analogous reason.
One of the central parameters to characterize an $(n, \mathcal{C}, \hat{M})$-induced code $\hat{\mathcal{C}}$ described by (10) is the number of times a component of a codeword $\boldsymbol{u}(i)$ from $\mathcal{C}$ differs from that of the induced codeword $\boldsymbol{v}(i,j)$ from $\hat{\mathcal{C}}$, with $(i,j) \in \mathcal{W} \times \hat{\mathcal{W}}$. This quantity is referred to as the *weight of the codeword* $\boldsymbol{v}(i,j)$.

**Definition 5** (Weight of the codeword $\boldsymbol{v}(i,j)$)**.** *Given an $(n, M)$-broadcast code $\mathcal{C}$ represented by the system in* (4)*, consider an $(n, \mathcal{C}, \hat{M})$-induced code $\hat{\mathcal{C}}$ represented by the system in* (10)*. For all $(i,j) \in \mathcal{W} \times \hat{\mathcal{W}}$, the weight of the codeword $\boldsymbol{v}(i,j)$, denoted by $\omega(i,j)$, is:*

$$\omega(i,j) = \sum_{t=1}^{n} \mathbb{1}_{\{u_t(i) \neq v_t(i,j)\}}. \tag{15}$$

Another parameter is the number of times that a given symbol $x$ is observed at a given component of a codeword from $\mathcal{C}$ and at the same component of a codeword from $\hat{\mathcal{C}}$ another symbol is observed. This quantity is referred to as the *weight of the symbol $x$*.

**Definition 6** (Weight of the Symbol $x$)**.** *Given an $(n, M)$-broadcast code $\mathcal{C}$ represented by the system in* (4)*, consider an $(n, \mathcal{C}, \hat{M})$-induced code $\hat{\mathcal{C}}$ represented by the system in* (10)*. For all $x \in \mathcal{X}$, the weight of the symbol $x$, denoted by $\omega(x)$, is*

$$\omega(x) = \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{t=1}^{n} \mathbb{1}_{\{u_t(i)=x\}} \mathbb{1}_{\{u_t(i) \neq v_t(i,j)\}}. \tag{16}$$

The codes $\mathcal{C}$ and $\hat{\mathcal{C}}$ induce several empirical probability mass functions that are relevant for the analysis of covert codes. These functions are defined hereunder.

**Definition 7** (Empirical Probability Distributions)**.** *Given an $(n, M)$-broadcast code $\mathcal{C}$ represented by the system in* (4)*, consider an $(n, \mathcal{C}, \hat{M})$-induced code $\hat{\mathcal{C}}$ represented by the system in* (10)*. For all $(x, \hat{x}) \in \mathcal{X}^2$,*

- *the empirical channel input probability distribution induced by the broadcast code $\mathcal{C}$, denoted by $\bar{P}_X$, is*

$$\bar{P}_X(x) \triangleq \frac{1}{nM} \sum_{i=1}^{M} N(x|\boldsymbol{u}(i)); \tag{17}$$

- *the empirical joint probability distribution induced by the two codes $\mathcal{C}$ and $\hat{\mathcal{C}}$ on $\mathcal{X}^2$, denoted by $\bar{P}_{X\hat{X}}$, is*

$$\bar{P}_{X\hat{X}}(x, \hat{x}) \triangleq \frac{1}{nM\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} N(x, \hat{x}|\boldsymbol{u}(i), \boldsymbol{v}(i, j)); \tag{18}$$

- *the empirical probability with which a symbol $x$ in a codeword from $\mathcal{C}$ is changed into a symbol $\hat{x} \neq x$ in a codeword from $\hat{\mathcal{C}}$, denoted by $\tilde{P}_{\hat{X}|X}$, is:*

$$\hat{P}_{\hat{X}|X}(\hat{x}|x) \triangleq \frac{\sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{t=1}^{n} \mathbb{1}_{\{x=u_t(i)\}} \mathbb{1}_{\{\hat{x}=v_t(i,j)\}} \mathbb{1}_{\{x\neq\hat{x}\}}}{\sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{t=1}^{n} \mathbb{1}_{\{x=u_t(i)\}} \mathbb{1}_{\{u_t(i)\neq v_t(i,j)\}}}, \tag{19}$$

  *and*

$$\operatorname{supp} \hat{P}_{\hat{X}|X=x} = \mathcal{X} \setminus \{x\}; \tag{20}$$

- *the empirical probability with which a symbol $x$ in a codeword from $\mathcal{C}$ is changed to any other symbol to generate a codeword in $\hat{\mathcal{C}}$, denoted by $\theta(x)$, is*

$$\theta(x) \triangleq 1 - \bar{P}_{\hat{X}|X}(x|x), \tag{21}$$

  *where $\bar{P}_{\hat{X}|X}(x|x)$ is such that*

$$\bar{P}_{\hat{X}X}(x, x) = \bar{P}_X(x) \bar{P}_{\hat{X}|X}(x|x). \tag{22}$$

The next lemma establishes an expression of the empirical probability $\bar{P}_{X\hat{X}}$ in (18) in terms of $\bar{P}_X$ in (17), $\hat{P}_{\hat{X}|X}$ in (19) and $\theta$ in (21).

**Lemma 1.** *Given an $(n, M)$-broadcast code $\mathcal{C}$ represented by the system in* (4)*, consider an $(n, \mathcal{C}, \hat{M})$-induced code $\hat{\mathcal{C}}$ represented by the system in* (10)*. For all $(x, \hat{x}) \in \mathcal{X}^2$, it holds that*

$$\bar{P}_{X\hat{X}}(x, \hat{x}) = \bar{P}_X(x) \Big( (1 - \theta(x)) \, \mathbb{1}_{\{x=\hat{x}\}} + \theta(x) \hat{P}_{\hat{X}|X}(\hat{x}|x) \Big). \tag{23}$$

*Proof:* The proof of Lemma 1 is presented in Appendix B. ∎

The next lemma relates for all $x \in \mathcal{X}$, the parameter $\theta(x)$ in (21) to the average weight $\omega(x)$ in (16).

**Lemma 2.** *Given an $(n, M)$-broadcast code $\mathcal{C}$ represented by the system in (4), consider an $(n, \mathcal{C}, \hat{M})$-induced code $\hat{\mathcal{C}}$ represented by the system in (10). For all $x \in \mathcal{X}$, it holds that*

$$\omega(x) = n\bar{P}_X(x)\theta(x). \tag{24}$$

*Proof:* The proof of Lemma 2 is presented in Appendix C. ∎

The following lemma is an immediate consequence of Lemma 2 and both Definition 5 and Definition 6.

**Lemma 3.** *Given an $(n, M)$-broadcast code $\mathcal{C}$ represented by the system in (4), consider an $(n, \mathcal{C}, \hat{M})$-induced code $\hat{\mathcal{C}}$ represented by the system in (10). Then, it holds that*

$$n \sum_{x \in \mathcal{X}} \bar{P}_X(x)\theta(x) = \sum_{x \in \mathcal{X}} \omega(x) = \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \omega(i, j).$$

## 3.3 Covert Codes

Consider an $(n, M, \epsilon)$-code described by (4) and denoted by $\mathcal{C}$. Consider also an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code denoted by $\hat{\mathcal{C}}$ and described by (10). For all $k \in \{1, 2\}$, let $Q_{\boldsymbol{Y}_k}$ and $R_{\boldsymbol{Y}_k}$ be respectively the probability distribution functions of the channel output vector $\boldsymbol{Y}_k$ when the broadcast code $\mathcal{C}$ is used and when the induced code $\hat{\mathcal{C}}$ is used. That is, for all $\boldsymbol{y} \in \mathcal{Y}_k^n$,

$$Q_{\boldsymbol{Y}_k}(\boldsymbol{y}) \triangleq \frac{1}{M} \sum_{i=1}^{M} P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i)), \text{ and} \tag{25}$$

$$R_{\boldsymbol{Y}_k}(\boldsymbol{y}) \triangleq \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}(i, j)), \tag{26}$$

where $P_{\boldsymbol{Y}_k|\boldsymbol{X}}$ is the marginal obtained from (2b). Using this notation a covert code is defined hereunder.

**Definition 8** ($(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code)**.** *Given $\delta \in [0, 1]$ and an $(n, M, \epsilon)$-broadcast code $\mathcal{C}$ described by (4), an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code described by (10) is said to be an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code if*

$$\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} \leqslant \delta, \tag{27}$$

*where $Q_{\boldsymbol{Y}_2}$ and $R_{\boldsymbol{Y}_2}$ are respectively defined in (25) and (26).*

Let $Q_{W\boldsymbol{Y}_2}$ and $R_{W\boldsymbol{Y}_2}$ be two distributions such that, for all $(i, \boldsymbol{y}) \in \mathcal{W} \times \mathcal{Y}_2^n$,

$$Q_{W\boldsymbol{Y}_2}(i, \boldsymbol{y}) \triangleq \frac{1}{M} Q_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i), \text{ and} \tag{28}$$

$$R_{W\boldsymbol{Y}_2}(i, \boldsymbol{y}) \triangleq \frac{1}{M} R_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i), \tag{29}$$

with

$$Q_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i) \triangleq \prod_{t=1}^{n} P_{Y_2|X}(y_t|u_t(i)), \text{ and} \tag{30}$$

$$R_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i) \triangleq \frac{1}{\hat{M}} \sum_{j=1}^{\hat{M}} \prod_{t=1}^{n} P_{Y_2|X}(y_t|v_t(i, j)). \tag{31}$$

Note that the marginal distributions $Q_{\boldsymbol{Y}_2}$ and $R_{\boldsymbol{Y}_2}$ are respectively in (25) and (26).

Using this notation, the following lemma highlights that replacing the constraint $\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} < \delta$ in (27) by the constraint $\|Q_{W\boldsymbol{Y}_2} - R_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} < \delta$ is equivalent up to an additive constant.

**Lemma 4.** *Given an $(n, M, \epsilon)$-broadcast code $\mathcal{C}$ described by (4), any $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code described by (10) satisfies*

$$\|Q_{W\boldsymbol{Y}_2} - R_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} \leqslant \|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} + \epsilon + \hat{\epsilon}, \tag{32}$$

*where the distributions $Q_{\boldsymbol{Y}_2}$, $R_{\boldsymbol{Y}_2}$, $Q_{W\boldsymbol{Y}_2}$ and $R_{W\boldsymbol{Y}_2}$ are defined in (25), (26), (28) and (29), respectively.*

*Proof:* The proof of Lemma 4 is presented in Appendix E. ■

In the remainder of this work, it is assumed that the induced codes satisfy $R_{\boldsymbol{Y}_2} \ll Q_{\boldsymbol{Y}_2}$. Otherwise, a covert transmission of private messages is impossible for some values of $\delta \in [0, 1]$. This is illustrated by the following lemma.

**Lemma 5.** *Consider the random transformation in (2), an $(n, M)$-broadcast code $\mathcal{C}$ and an $(n, \mathcal{C}, \hat{M})$-induced code respectively described in (4) and (10) such that $Q_{\boldsymbol{Y}_2} \not\ll R_{\boldsymbol{Y}_2}$. Then,*

$$\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} \geqslant \frac{1}{2}\left(1 - \Pr\left[\boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2}\right]\right), \tag{33}$$

*where the probability is calculated under the assumption that $\boldsymbol{Y}_2 \sim R_{\boldsymbol{Y}_2}$, and where $Q_{\boldsymbol{Y}_2}$ and $R_{\boldsymbol{Y}_2}$ are respectively defined in (25) and (26).*

The proof of Lemma 5 is presented in Appendix D.

Given that $\Pr\left[\boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2}\right] > 0$ in (33), it follows that covert communications can not be achieved for values of $\delta < \frac{1}{2}\left(1 - \Pr\left[\boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2}\right]\right)$.
Finally, the analysis is restricted to induced-codes that satisfy $R_{\boldsymbol{Y}_2} \neq Q_{\boldsymbol{Y}_2}$. This guarantees that there exists no induced-code that can perfectly mimic the channel output distribution $Q_{\boldsymbol{Y}_2}$ induced by the broadcast code at Receiver 2. Otherwise, the problem is trivial and covert communications are always achievable.

The information rate at which information can be covertly transmitted to Receiver 1 using an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code is $\frac{\log_2(\hat{M})}{n}$ bits per channel use. Thus, given the broadcast code $\mathcal{C}$, a fundamental limit on the rate at which information can be covertly transmitted is given by the largest possible $\hat{M}$ for which an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code exists. This notion is formalized by the following definition.

**Definition 9** (Largest covert code's size)**.** *Fix a pair $(\hat{\epsilon}, \delta) \in [0, 1]^2$ and consider an $(n, M, \epsilon)$-broadcast code $\mathcal{C}$. The largest covert code's size induced by $\mathcal{C}$, denoted by $\hat{M}^*(n, \mathcal{C}, \hat{\epsilon}, \delta)$, is:*

$$\hat{M}^*(n, \mathcal{C}, \hat{\epsilon}, \delta) = \max\left\{\hat{M} \in \mathbb{N} : \exists\, (n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)\text{-covert code}\right\}.$$

# 4 Examples of Impossible Covert Communications

This section provides two examples in which covert communications can not be achieved for arbitrary values of $\delta \in [0, 1]$. Later, a more general impossibility result is presented.

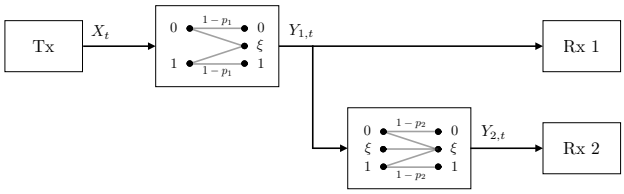



Figure 2: Degraded erasure broadcast channel at channel use $t \in \{1, 2, \ldots, n\}$.

**Example 1.** *Assume that the random transformation in* (2) *is such that* $\mathcal{X} = \{0, 1\}$*;* $\mathcal{Y}_1 = \mathcal{Y}_2 = \{0, \xi, 1\}$*; and for all* $x \in \mathcal{X}$*, the conditional probability distributions* $P_{Y_1|X}$ *and* $P_{Y_2|Y_1}$ *respectively satisfy:*

$$P_{Y_1|X}(x|x) = 1 - P_{Y_1|X}(\xi|x) = 1 - p_1, \tag{34a}$$
$$P_{Y_1|X}(x|1-x) = 0, \tag{34b}$$

*and*

$$P_{Y_2|Y_1}(x|x) = 1 - P_{Y_2|Y_1}(\xi|x) = 1 - p_2, \tag{35a}$$
$$P_{Y_2|Y_1}(x|1-x) = 0, \tag{35b}$$
$$P_{Y_2|Y_1}(\xi|\xi) = 1, \tag{35c}$$

*with* $(p_1, p_2) \in [0, \frac{1}{2}]^2$.

Figure 2 depicts the channel in Example 1. Note that the probability distribution $P_{Y_2|X}$ verifies that for all $x \in \mathcal{X}$:

$$\begin{aligned} P_{Y_2|X}(x|x) = 1 - P_{Y_1|X}(\xi|x) &= 1 - p_1 - p_2 + p_1 p_2 \\ &= 1 - p, \end{aligned} \tag{36}$$

with

$$p = p_1 + p_2 - p_1 p_2. \tag{37}$$

Given an $(n, M, \epsilon)$-broadcast code $\mathcal{C}$ described by (4) and an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code described by (10), let $\mathcal{T}_{ij}$ and $\bar{\mathcal{T}}_{ij}$ be respectively defined for all $(i, j) \in \mathcal{W} \times \hat{\mathcal{W}}$ by

$$\mathcal{T}_{ij} = \{t \in \{1, 2, \ldots, n\} : u_t(i) \neq v_t(i, j)\}, \text{ and} \tag{38}$$
$$\bar{\mathcal{T}}_{ij} = \{t \in \{1, 2, \ldots, n\} : u_t(i) = v_t(i, j)\}. \tag{39}$$

Note that the cardinalities of the sets $\mathcal{T}_{ij}$ and $\bar{\mathcal{T}}_{ij}$ respectively satisfy

$$|\mathcal{T}_{ij}| = \omega(i, j), \qquad \text{and} \tag{40}$$
$$|\bar{\mathcal{T}}_{ij}| = n - \omega(i, j). \tag{41}$$

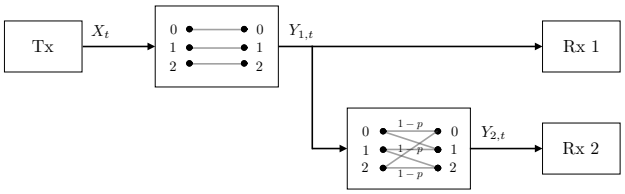



Figure 3: Degraded typewriter broadcast channel at channel use $t \in \{1, 2, \ldots, n\}$.

Within this context, the term $\Pr[\boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2}]$ in (33) can be upper bounded as follows:

$$\begin{aligned}
\Pr[\boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2}] &= \Pr[\boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2} \cap \operatorname{supp} R_{\boldsymbol{Y}_2}] \\
&= \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{\substack{\boldsymbol{y} \in \operatorname{supp} Q_{\boldsymbol{Y}_2} \\ \cap \operatorname{supp} R_{\boldsymbol{Y}_2}}} \prod_{t=1}^{n} P_{Y_2|X}(y_t|v_t(i,j)) \\
&= \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{\substack{\boldsymbol{y} \in \operatorname{supp} Q_{\boldsymbol{Y}_2} \\ \cap \operatorname{supp} R_{\boldsymbol{Y}_2}}} \prod_{s \in \mathcal{T}_{ij}} P_{Y_2|X}(y_s|v_s(i,j)) \prod_{r \in \bar{\mathcal{T}}_{ij}} P_{Y_2|X}(y_r|u_r(i)) \\
&\leqslant \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{\substack{\boldsymbol{y} \in \operatorname{supp} Q_{\boldsymbol{Y}_2} \\ \cap \operatorname{supp} R_{\boldsymbol{Y}_2}}} \prod_{s \in \mathcal{T}_{ij}} P_{Y_2|X}(y_s|v_s(i,j)). \qquad (42)
\end{aligned}$$

Note that for all $\boldsymbol{y} \in \operatorname{supp} Q_{\boldsymbol{Y}_2} \cap \operatorname{supp} R_{\boldsymbol{Y}_2}$ and for all $t \in \{1, 2, \ldots, n\}$ for which $u_t(i) \neq v_t(i,j)$ for some $(i,j) \in \mathcal{W} \times \hat{\mathcal{W}}$, it holds that $y_t = \xi$, which implies that $P_{Y_2|X}(y_t|v_t(i,j)) = p$. Hence, it holds from (42) that

$$\begin{aligned}
\Pr[\boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2}] &\leqslant \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} p^{|\mathcal{T}_{ij}|} \\
&\overset{(a)}{=} \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} p^{\omega(i,j)} \\
&\leqslant p^{\omega_{\min}}, \qquad (43)
\end{aligned}$$

where *(a)* follows from (40); and

$$\omega_{\min} \triangleq \min_{(i,j) \in \mathcal{W} \times \hat{\mathcal{W}}} \omega(i,j). \qquad (44)$$

Finally, it follows from Lemma 5 that the total variation $\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}}$ verifies

$$\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} \geqslant \frac{1}{2}\left(1 - p^{\omega_{\min}}\right). \qquad (45)$$

The above lower bound shows that the constraint in (27) can not be satisfied for values of $\delta \leqslant \frac{1}{2}(1 - p^{\omega_{\min}})$.

**Example 2.** *Consider the random transformation in* (2) *such that* $\mathcal{X} = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1, 2\}$*, and such that for all* $x \in \mathcal{X}$*, the conditional probability distributions* $P_{Y_1|X}$ *and* $P_{Y_2|Y_1}$ *respectively satisfy:*

$$P_{Y_1|X}(x|x) = 1, \qquad (46)$$

*and*

$$P_{Y_2|Y_1}(x|x'') = 0, \tag{47a}$$
$$P_{Y_2|Y_1}(x|x) = 1 - P_{Y_2|Y_1}(x'|x) = 1 - p, \tag{47b}$$

*with* $x' = x + 1 \mod |\mathcal{X}|$, $x'' = x + 2 \mod |\mathcal{X}|$, *and* $p \in [0, \frac{1}{2}]$.

Figure 3 depicts the channel in Example 2. Given that $P_{Y_1|X}(x|x) = 1$ for any $x \in \mathcal{X}$, it follows that $P_{Y_2|X=x} = P_{Y_2|Y_1=x}$.

Note that given an arbitrary $(n, M, \epsilon)$-broadcast code $\mathcal{C}$ of the form in (4) and any $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code, the inequality in (43) holds, which implies

$$\Pr\left[\boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2}\right] \leqslant p^{\omega_{\min}}. \tag{48}$$

The above lower bound shows that the constraint in (27) can not be satisfied for values of $\delta \leqslant \frac{1}{2}\left(1 - p^{\omega_{\min}}\right)$.

# 5 Probability of Detecting Covert Communications

Throughout this section, consider a given $(n, M, \epsilon)$-broadcast code $\mathcal{C}$ described by the system in (4) and an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code $\hat{\mathcal{C}}$ described by the system in (10). Section 5.1 and Section 5.2 provide respectively a lower bound and an upper bound on the probability of incorrectly determining that the covert code $\hat{\mathcal{C}}$ is used instead of the broadcast code $\mathcal{C}$.

## 5.1 A Lower Bound

Consider a hypothesis test in which Receiver 2 aims to determine whether the broadcast code $\mathcal{C}$ (hypothesis $H_0$) or the covert code $\hat{\mathcal{C}}$ (hypothesis $H_1$) is used upon the observation of the channel output $\boldsymbol{Y}_2$:

$$\begin{cases} H_0 : \boldsymbol{Y}_2 \sim Q_{\boldsymbol{Y}_2} \\ H_1 : \boldsymbol{Y}_2 \sim R_{\boldsymbol{Y}_2}, \end{cases} \tag{49}$$

where $Q_{\boldsymbol{Y}_2}$ and $R_{\boldsymbol{Y}_2}$ are respectively given in (25) and (26).

Denote by $\alpha \in [0, 1]$ and $\beta \in [0, 1]$ the type-I and type-II error probabilities associated with a decision rule $T : \mathcal{Y}_2^n \to \{0, 1\}$ of the form

$$T(\boldsymbol{y}) \triangleq \begin{cases} 0 & \text{if } H_0 \text{ is accepted,} \\ 1 & \text{if } H_1 \text{ is accepted.} \end{cases} \tag{50}$$

That is,

$$\alpha \triangleq \Pr\left[T(\boldsymbol{Y}_2) = 1\right], \text{ and} \tag{51}$$
$$\beta \triangleq \Pr\left[T(\boldsymbol{Y}_2) = 0\right], \tag{52}$$

where the probability operator in (51) applies assuming that $\boldsymbol{Y}_2 \sim Q_{\boldsymbol{Y}_2}$ and the probability operator in (52) applies assuming that $\boldsymbol{Y}_2 \sim R_{\boldsymbol{Y}_2}$.

Using this notation, the following corollary, which is an immediate consequence of Lemma 8 in Appendix A, introduces a property of covert codes.

**Corollary 1.** *Given an $(n, M, \epsilon)$-broadcast code $\mathcal{C}$, any $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code satisfies*

$$\alpha + \beta \geqslant 1 - \|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}}\,, \tag{53}$$

*with $\alpha$ and $\beta$ respectively defined in* (51) *and* (52)*, for all decision rules $T : \mathcal{Y}_2^n \to \{0, 1\}$ of the form* (50)*.*

Note that Corollary 1 highlights the relevance of the parameter $\delta$ in Definition 8. Essentially, the smaller the parameter $\delta$, the higher the probability of failing to determine whether the broadcast code or the covert code is used.

## 5.2 An Upper Bound

In this section, given an $(n, M, \epsilon)$-broadcast code $\mathcal{C}$ and an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code $\hat{\mathcal{C}}$, an upper bound on the type-I and type-II error probabilities at Receiver 2 are presented. The underlying assumption is that Receiver 2 performs perfect decoding of the common message index $i \in \mathcal{W}$. This assumption is essentially improving the capability of Receiver 2 for detecting a covert communication. Thus, the upper bound obtained under this assumption is rather loose.
Under these assumptions, the hypothesis test run by Receiver 2 to determine whether or not a covert communication occurs is the following:

$$\begin{cases} H_0 : \boldsymbol{Y}_2 \sim Q_{\boldsymbol{Y}_2|W=i}, \\ H_1 : \boldsymbol{Y}_2 \sim R_{\boldsymbol{Y}_2|W=i}, \end{cases} \tag{54}$$

where the distributions $Q_{\boldsymbol{Y}_2|W=i}$ and $R_{\boldsymbol{Y}_2|W=i}$ are respectively defined in (30) and (31).
Denote by $\alpha_i \in [0, 1]$ and $\beta_i \in [0, 1]$ the type-I and type-II error probabilities associated with a decision rule $T_i \ : \mathcal{Y}_2^n \to \{0, 1\}$ of the form

$$T_i(\boldsymbol{y}) \triangleq \begin{cases} 0 & \text{if } H_0 \text{ is accepted,} \\ 1 & \text{if } H_1 \text{ is accepted.} \end{cases} \tag{55}$$

That is,

$$\alpha_i \triangleq \Pr\left[T_i(\boldsymbol{Y}_2) = 1\right], \text{ and} \tag{56}$$

$$\beta_i \triangleq \Pr\left[T_i(\boldsymbol{Y}_2) = 0\right], \tag{57}$$

where the probability operator in (56) applies assuming that $\boldsymbol{Y}_2 \sim Q_{\boldsymbol{Y}_2|W=i}$ and the probability operator in (57) applies assuming that $\boldsymbol{Y}_2 \sim R_{\boldsymbol{Y}_2|W=i}$. The next proposition establishes upper-bounds on $\alpha_i$ in (56) and $\beta_i$ in (57), under certain conditions.

**Proposition 1.** *Assume that for all pairs $(x, x') \in \mathcal{X}^2$ such that $x \neq x'$, the random transformation in* (2) *satisfies the following conditions:*

$$\chi_2\left(P_{Y_2|X=x}, P_{Y_2|X=x'}\right) = d, \tag{58}$$

$$D\left(P_{Y_1|X=x} || P_{Y_1|X=x'}\right) = \ell, \tag{59}$$

*where $(d, \ell) \in \mathbb{R}_+^2$. Let $\nu \in \mathbb{N}$ be such that for all $(i, j) \in \mathcal{W} \times \hat{\mathcal{W}}$,*

$$\omega(i, j) \geqslant \nu. \tag{60}$$

*Then, for all $i \in \mathcal{W}$, it follows that:*

$$\alpha_i \leqslant Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right) + \frac{c_3}{\sqrt{n}}, \tag{61}$$

*and*

$$\beta_i \leqslant Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right) + \frac{c_4}{\sqrt{n}}, \tag{62}$$

*where $c_3$ and $c_4$ are constants.*

*Proof:* The proof of Proposition 1 is presented in Appendix G. ■

Note that the binary symmetric channel is an example of channel satisfying (58) and (59). Another example is presented in Section 8.

# 6 Achievability of Covert Communications

In this section, a lower bound on the largest code's size (Definition 9) given an $(n, M, \epsilon)$-broadcast code, denoted by $\mathcal{C}$, is established. The construction of this result is presented in three parts using a random coding argument. In the first part, a probability distribution to randomly generate $(n, \mathcal{C}, \hat{M})$-induced codes is chosen. Often, this distribution is referred to as the *generating distribution*. This distribution is expressed in terms of some parameters, which are referred to as the *generating parameters*. In the second part, the average of the decoding error probability (denoted by $\hat{\Lambda}$) over all possible $(n, \mathcal{C}, \hat{M})$-induced codes that can be generated by the generating distribution is upper-bounded. This upper-bound is expressed in terms of the generating parameters, which proves that there must exist at least one $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code for which $\hat{\epsilon} < \hat{\Lambda}$. In the third part, the generating parameters are chosen in order to satisfy the covertness constraint in (27) for a fixed $\delta$, which allows the construction of an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code.

**Part I: Generation of Induced Codes**

Consider an $(n, M, \epsilon)$-broadcast code $\mathcal{C}$ for the random transformation in (2) described by the system in (4). Consider also the parameters $\hat{M} \in \mathbb{N}$; $K \in [0, \sqrt{n}]$ and a conditional probability distribution $\tilde{P}_{\hat{X}|X}$ such that, for all $x \in \mathcal{X}$,

$$\operatorname{supp} \tilde{P}_{\hat{X}|X=x} = \mathcal{X} \setminus \{x\}. \tag{63}$$

Using the parameters $K$ and $\tilde{P}_{\hat{X}|X}$, let $P_{\hat{X}|X}$ be a conditional probability distribution such that for all $(x, \hat{x}) \in \mathcal{X}^2$,

$$P_{\hat{X}|X}(\hat{x}|x) \triangleq (1-\theta)\, \mathbb{1}_{\{x=\hat{x}\}} + \theta \tilde{P}_{\hat{X}|X}(\hat{x}|x), \tag{64}$$

with

$$\theta \triangleq \frac{K}{\sqrt{n}}. \tag{65}$$

Often, the parameters $\hat{M} \in \mathbb{N}$; $K \in [0, \sqrt{n}]$; and $\tilde{P}_{\hat{X}|X}$ are referred to as the *generating parameters*. For all $i \in \{1, 2, \ldots, M\}$, generate $\hat{M}$ codewords

$$\boldsymbol{v}(i,1), \boldsymbol{v}(i,2), \ldots, \boldsymbol{v}(i,\hat{M}) \tag{66}$$

to form the codebook of an $(n, \mathcal{C}, \hat{M})$-induced code. For all $j \in \{1, 2, \ldots, \hat{M}\}$, the codeword $\boldsymbol{v}(i,j)$ is the realization of a random variable following the probability distribution $P_{\hat{\boldsymbol{X}}|\boldsymbol{X}=\boldsymbol{u}(i)}$ such that for all $\hat{\boldsymbol{x}} \in \mathcal{X}^n$,

$$P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}\left(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)\right) \triangleq \prod_{t=1}^{n} P_{\hat{X}|X}\left(\hat{x}_t|u_t(i)\right), \tag{67}$$

where $\boldsymbol{u}(1)$, $\boldsymbol{u}(2)$, …, $\boldsymbol{u}(M)$ are the codewords of the given broadcast code $\mathcal{C}$. In the following, the distribution $P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}$ is referred to as the *generating distribution*.

To complete the generation of the $(n, \mathcal{C}, \hat{M})$-induced code, the decoding sets must be specified. Receiver 2 uses the decoding sets

$$\mathcal{D}_2(1), \mathcal{D}_2(2), \ldots, \mathcal{D}_2(M) \tag{68}$$

of the given broadcast code $\mathcal{C}$, according to the decoding rule in (6), with $k = 2$.

For all $(\boldsymbol{x}, \hat{\boldsymbol{x}}, \boldsymbol{y}) \in \mathcal{X}^{2n} \times \mathcal{Y}_k^n$, let $\imath_k(\hat{\boldsymbol{x}}; \boldsymbol{y}|\boldsymbol{x})$, with $k \in \{1, 2\}$, be defined by

$$\imath_k(\hat{\boldsymbol{x}}; \boldsymbol{y}|\boldsymbol{x}) \triangleq \log_2 \left( \frac{P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{\sum\limits_{\boldsymbol{x}' \in \mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{x}'|\boldsymbol{x}) P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}')} \right). \tag{69}$$

At Receiver 1, upon the reception of the channel output $\boldsymbol{y} \in \mathcal{Y}_1^n$, Receiver 1 declares that the index pair $(i, j) \in \mathcal{W} \times \hat{\mathcal{W}}$ was transmitted according to the decoding rule in (12), with

$$\mathcal{D}_1(i,j) = \Big\{ \boldsymbol{y} \in \mathcal{Y}_1^n : \imath_1\Big(\boldsymbol{v}(i,j), \boldsymbol{y}|\boldsymbol{u}(i)\Big) \geqslant n\eta \Big\} \setminus \bigcup_{(k,\ell) \in \Gamma(i,j)} \mathcal{D}_1(k,l), \tag{70}$$

where $\eta \in \mathbb{R}$ is a parameter whose exact value is determined later; and

$$\Gamma(i,j) \triangleq \{1, 2, \ldots, i\} \times \{1, 2, \ldots, j-1\}. \tag{71}$$

Note that the codewords in (66), the decoding sets in (68) and the decoding sets in (70) form an $(n, \mathcal{C}, \hat{M})$-induced code.

**Part II: Decoding Error Probability Analysis**

Denote respectively by $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ the average of $\hat{\lambda}_1$ in (13) and $\hat{\lambda}_2$ in (14) over all possible induced codes that can be obtained from the generating distribution in (67). That is,

$$\begin{aligned} \hat{\Lambda}_1 &= \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{\hat{\boldsymbol{x}} \in \mathcal{X}^n} \frac{P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i))}{M\hat{M}} \Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i,j) | \boldsymbol{X} = \hat{\boldsymbol{x}}\right], \\ \text{and} \qquad \hat{\Lambda}_2 &= \sum_{i=1}^{M} \sum_{\hat{\boldsymbol{x}} \in \mathcal{X}^n} \frac{P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i))}{M} \Pr\left[\boldsymbol{Y}_2 \in \mathcal{D}_2^{\mathsf{c}}(i) | \boldsymbol{X} = \hat{\boldsymbol{x}}\right]. \end{aligned} \tag{72a}$$

Let for all $k \in \{1, 2\}$ and for all pairs $(x, y) \in \mathcal{X} \times \mathcal{Y}_k$, $\tilde{R}_{Y_k|X}(y|x)$ be the distribution

$$\tilde{R}_{Y_k|X}(y|x) \triangleq \sum_{\hat{x} \in \mathcal{X}} \tilde{P}_{\hat{X}|X}(\hat{x}|x) P_{Y_k|X}(y|\hat{x}). \tag{73}$$

Using this notation, the following proposition establishes an upper-bound on $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$.

**Proposition 2.** *For all $k \in \{1, 2\}$ the average error probability $\Lambda_k$ in (72) satisfies*

$$\hat{\Lambda}_k \leqslant \max\left( \Pr\left[\imath_1\left(\hat{\boldsymbol{X}}; \boldsymbol{Y}_1 | \boldsymbol{u}(W)\right) \leqslant n\eta\right], \epsilon\Big(1 + \max_{(x,y) \in \mathcal{X} \times \mathcal{Y}_2} \frac{\tilde{R}_{Y_2|X}(y|x) - P_{Y_2|X}(y|x)}{P_{Y_2|X}(y|x)} \theta\Big)^n \right), \tag{74}$$

*where $\epsilon$ is the average decoding error probability of the broadcast code $\mathcal{C}$; $\eta$ is in (70); and the probability operator in (74) is with respect to the joint distribution $P_{W\hat{\boldsymbol{X}}\boldsymbol{Y}_1}$, for which*

$$P_{W\hat{\boldsymbol{X}}\boldsymbol{Y}_1}(i, \hat{\boldsymbol{x}}, \boldsymbol{y}) = \frac{1}{M} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)) P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}}), \tag{75}$$

*for all $(i, \hat{\boldsymbol{x}}, \boldsymbol{y}) \in \mathcal{W} \times \mathcal{X}^n \times \mathcal{Y}_1^n$.*

*Proof:* The proof of Proposition 2 is presented in Appendix J. ■

**Remark 1.** *Proposition 2 suggests that the average probabilities of error over all possible* $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$*-induced codes at Receiver* 1 *and Receiver* 2*, i.e.,* $\hat{\Lambda}_1$ *and* $\hat{\Lambda}_2$ *respectively, are bigger than the average decoding error probability* $\epsilon$ *of the given broadcast code* $\mathcal{C}$*.*

Using the result in Proposition 2, it is possible to determine the conditions on $\hat{\epsilon}$, $\eta$, $\theta$, $\hat{M}$ and $\tilde{P}_{\hat{X}|X}$, such that at least one $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code exists. The following proposition describes these conditions.

**Proposition 3.** *There always exists an* $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$*-induced code for the random transformation in* (2) *that satisfies*

$$\frac{\log_2\left(\hat{M}\right)}{n} \geqslant \sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}} \bar{P}_X(x)\Big(\theta\tilde{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big) - \theta^2\chi_2\left(\tilde{R}_{Y_1|X=x}, P_{Y_1|X=x}\right)\Big) \tag{76}$$

*if the parameters* $\theta$ *and* $\tilde{P}_{\hat{X}|X}$ *are chosen such that*

$$\hat{\epsilon} < \epsilon\left(1 + \max_{(x,y)\in\mathcal{X}\times\mathcal{Y}_2} \frac{\tilde{R}_{Y_2|X}(y|x) - P_{Y_2|X}(y|x)}{P_{Y_2|X}(y|x)}\theta\right)^n. \tag{77}$$

*Proof:* The proof of Proposition 3 is presented in Appendix K. ■

**Part III: Generation of Covert Codes**

This final part focuses on determining the conditions on the generating parameters to satisfy the covertness constraint in (27). The following proposition describes such conditions.

**Proposition 4.** *There always exists an* $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$*-covert code, with* $\hat{M}$ *satisfying* (76)*, if* $\theta$ *and* $\tilde{P}_{\hat{X}|X}$ *satisfy* (77) *and*

$$\theta \leqslant \frac{2Q^{-1}\left(\frac{1-\delta-\epsilon-\max\{\hat{\Lambda}_1,\hat{\Lambda}_2\}-\frac{c}{\sqrt{n}}}{2}\right)}{\sqrt{n\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x})}}, \tag{78}$$

*where* $c$ *is a positive constant.*

*Proof:* The proof of Proposition 4 is presented in Appendix M. ■

From Proposition 3 and Proposition 4, it follows that the rate at which the covert message is transmitted, *i.e.*, $\frac{\log_2\left(\hat{M}\right)}{n}$, can be optimized by properly choosing the values of the parameters $\theta$ and $\tilde{P}_{\hat{X}|X}$. Proposition 5 follows immediately from this observation and can be presented more compactly by using the following notation:

$$\bar{D}(P_{Y_1|X}) \triangleq \sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}} \bar{P}_X(x)\tilde{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big), \tag{79}$$

and for all $k \in \{1, 2\}$,

$$\bar{\chi}_2(\tilde{R}_{Y_k|X}, P_{Y_k|X}) \triangleq \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_k|X=x}, P_{Y_k|X=x}). \tag{80}$$

**Proposition 5.** *Consider an $(n, M, \epsilon)$-broadcast code $\mathcal{C}$ for the random transformation in (2). Then, there always exists an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code that satisfies*

$$\frac{\log_2\left(\hat{M}\right)}{n} \geqslant \max_{\theta, \tilde{P}_{\hat{X}|X}} \theta \bar{D}(P_{Y_1|X}) - \theta^2 \bar{\chi}_2(\tilde{R}_{Y_1|X}, P_{Y_1|X}), \tag{81}$$

*where $c$ a positive constant and the optimization domain is the set of all $\theta$ and all $\tilde{P}_{\hat{X}|X}$ that jointly satisfy (77) and (78).*

In the asymptotic block-length regime, Proposition 5 leads to the following theorem.

**Theorem 1.** *Consider a sequence $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$, ..., of $(n, M_n, \epsilon_n)$-broadcast codes for the random transformation in (2), with $n \in \{1, 2, \ldots\}$ and*

$$\epsilon_n \leqslant \exp(-\zeta n), \tag{82}$$

*for some fixed positive real $\zeta$. Then, there always exists a sequence of $(n, \mathcal{C}_n, \hat{M}_n, \hat{\epsilon}_n, \delta)$-covert codes with $\lim_{n\to\infty} \hat{\epsilon}_n = 0$ such that*

$$\lim_{n\to\infty} \frac{\log_2\left(\hat{M}_n^*(n, \mathcal{C}_n, \hat{\epsilon}_n, \delta)\right)}{\sqrt{n}} \geqslant \max_{\tilde{P}_{\hat{X}|X}} \frac{2Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{\bar{\chi}_2(\tilde{R}_{Y_2|X}, P_{Y_2|X})}} \bar{D}(P_{Y_1|X}). \tag{83}$$

*Proof:* Consider an infinite sequence of positive reals $K_1 < K_2 < K_3$, ... and an infinite sequence of reals $\hat{\epsilon}_1 > \hat{\epsilon}_2 > \ldots > 0$, such that, for all $n \in \mathbb{N}$,

$$K_n \triangleq \frac{2Q^{-1}\left(\frac{1-\delta-\epsilon_n-\hat{\epsilon}_n-\frac{c}{\sqrt{n}}}{2}\right)}{\sqrt{\bar{\chi}_2(\tilde{R}_{Y_k|X}, P_{Y_k|X})}}. \tag{84}$$

In particular, for all $n \in \mathbb{N}$,

$$K_n < \frac{2Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{\bar{\chi}_2(\tilde{R}_{Y_k|X}, P_{Y_k|X})}}. \tag{85}$$

Note that if $\zeta$ in (82) satisfies the following condition

$$\zeta > \max_{\tilde{P}_{\hat{X}|X}} \ln\left(1 + \frac{2Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{n\bar{\chi}_2(\tilde{R}_{Y_k|X}, P_{Y_k|X})}} \max_{(x,y)\in\mathcal{X}\times\mathcal{Y}_2} \frac{\tilde{R}_{Y_2|X}(y|x) - P_{Y_2|X}(y|x)}{P_{Y_2|X}(y|x)}\right), \tag{86}$$

where the maximization is performed over all possible conditional distributions $\tilde{P}_{\hat{X}|X}$, then it follows that (77) is always satisfied. Hence, from Proposition 5, it holds that for a fixed $n$ and $\zeta$ satisfying (86), there always exists an $(n, \mathcal{C}_n, \hat{M}_n, \hat{\epsilon}_n, \delta)$-covert code such that

$$\frac{\log_2\left(\hat{M}_n\right)}{\sqrt{n}} \geqslant K_n \bar{D}(P_{Y_1|X}) - \frac{K_n^2}{\sqrt{n}} \bar{\chi}_2(\tilde{R}_{Y_1|X}, P_{Y_1|X}). \tag{87}$$

In the asymptotic block-length regime, the condition in (86) holds for all $\zeta > 0$, which immediately implies that

$$\lim_{n\to\infty} \frac{\log_2\left(\hat{M}_n\right)}{\sqrt{n}} \geqslant \frac{2Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{\bar{\chi}_2(\tilde{R}_{Y_2|X}, P_{Y_2|X})}} \bar{D}(P_{Y_1|X}). \tag{88}$$

The proof is completed by optimizing the right-hand side of (88) over all possible conditional distributions $\tilde{P}_{\hat{X}|X}$. ∎

# 7 Impossibility of Covert Communications

Given an $(n, M, \epsilon)$-broadcast code $\mathcal{C}$, this section introduces an upper bound on the ratio between the largest covert code's size $\hat{M}^*(n, \mathcal{C}, \hat{\epsilon}, \delta)$ and the square-root of the block-length, i.e., $\frac{\log_2\left(\hat{M}^*(n,\mathcal{C},\hat{\epsilon},\delta)\right)}{\sqrt{n}}$, in the asymptotic block-length regime. The following section introduces some preliminary results in the finite block-length regime that are crucial for proving the main result of this section.

## 7.1 Preliminary Results

Using Fano's inequility [9], the following proposition presents for all $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code $\hat{\mathcal{C}}$ an upper-bound on $\log_2\left(\hat{M}\right)$ in terms of the empirical probability mass functions induced by both the original code $\mathcal{C}$ and the covert code $\hat{\mathcal{C}}$.

**Proposition 6.** *Consider an $(n, M, \epsilon)$-broadcast code $\mathcal{C}$, described by the system in (4), for the random transformation in (2). Then, every $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code satisfies that*

$$\log_2\left(\hat{M}\right) \leqslant \frac{1}{1-\hat{\epsilon}}\Bigg(1 + n\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\bar{P}_X(x)\theta(x)\hat{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big) + \bar{P}_X(x)\frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x}, P_{Y_1|X=x})\Bigg). \tag{89}$$

*Proof:* The proof of Proposition 6 is presented in Appendix P. ∎

A central observation for proving the main result of this section is that given a covert code, a covert sub-code can be obtained by choosing the codewords whose weight (Definition 5) is bounded. More importantly, the cardinality of the set of upper-bounded-weight codewords can be lower-bounded. This result is presented by the following proposition.

**Proposition 7.** *Let $\eta > 0$ be arbitrarily small. Consider an $(n, M, \epsilon)$-broadcast code $\mathcal{C}$, described by the system in (4), for the random transformation in (2). Assume that the random transformation in (2) satisfies (58) and (59). Then, every $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code described by the system in (10) can be formed by two sub-codes. One sub-code whose codewords are in the set*

$$\tilde{\mathcal{W}} = \left\{\boldsymbol{v}(i,j) : \omega(i,j) < 2\sqrt{\frac{n}{d}}Q^{-1}\left(\frac{1-\delta-\eta}{2}\right), 1 \leqslant i \leqslant M, \text{ and } 1 \leqslant j \leqslant \hat{M}\right\}, \tag{90}$$

*and another sub-code whose codewords are in the set*

$$\tilde{\mathcal{W}}^{\mathsf{c}} = \left\{\boldsymbol{v}(i,j) : \omega(i,j) \geqslant 2\sqrt{\frac{n}{d}}Q^{-1}\left(\frac{1-\delta-\eta}{2}\right), 1 \leqslant i \leqslant M, \text{ and } 1 \leqslant j \leqslant \hat{M}\right\}. \tag{91}$$

*Moreover,*

$$\left|\tilde{\mathcal{W}}\right| > M\hat{M}\left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon - \hat{\epsilon}\right), \tag{92}$$

*where $c$ is a constant.*

*Proof:* The proof of Proposition 7 is presented in Appendix Q. ∎

## 7.2 Main Result

The following theorem introduces the main result of this section.

**Theorem 2.** *Consider a sequence $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$, ..., of $(n, M_n, \epsilon_n)$-broadcast codes for the random transformation in (2), with $\lim_{n\to\infty} \epsilon_n = 0$. Assume that the random transformation in (2) satisfies (58) and (59). Then, for any sequence $\hat{\mathcal{C}}_1$, $\hat{\mathcal{C}}_2$, $\hat{\mathcal{C}}_3$, ... of $(n, \mathcal{C}_n, \hat{M}_n, \hat{\epsilon}_n, \delta)$-covert codes with $\lim_{n\to\infty} \hat{\epsilon}_n = 0$, it holds that*

$$\lim_{n\to\infty} \frac{\log_2\left(\hat{M}_n^*(n, \mathcal{C}_n, \hat{\epsilon}_n, \delta)\right)}{\sqrt{n}} < \frac{2\ell}{\sqrt{d}} Q^{-1}\left(\frac{1-\delta-\eta}{2}\right), \tag{93}$$

*with $\eta > 0$ arbitrarily small.*

*Proof:* For all $n \in \mathbb{N}$, it follows from Proposition 7 that the covert sub-code of the covert code $\hat{\mathcal{C}}_n$ with codewords in the set

$$\tilde{\mathcal{W}}_n = \left\{\boldsymbol{v}(i,j) : \omega(i,j) < 2\sqrt{\frac{n}{d}} Q^{-1}\left(\frac{1-\delta-\eta}{2}\right), 1 \leqslant i \leqslant M_n, \text{ and } 1 \leqslant j \leqslant \hat{M}_n\right\}, \tag{94}$$

satisfies

$$\left|\tilde{\mathcal{W}}_n\right| > M_n \hat{M}_n \left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon_n - \hat{\epsilon}_n\right). \tag{95}$$

From Proposition 6, it holds that $\left|\tilde{\mathcal{W}}_n\right|$ is upper bounded as follows

$$\log_2\left(\frac{\left|\tilde{\mathcal{W}}_n\right|}{M_n}\right) \leqslant \frac{1}{1-\hat{\epsilon}_n}\left(1 + n\sum_{x\in\mathcal{X}} \ell \bar{P}_X(x)\theta(x) + \bar{P}_X(x)\frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x}, P_{Y_1|X=x})\right), \tag{96}$$

that is,

$$\log_2\left(\hat{M}_n\left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon_n - \hat{\epsilon}_n\right)\right) \leqslant \frac{1}{1-\hat{\epsilon}_n}\left(1 + n\sum_{x\in\mathcal{X}} \ell \bar{P}_X(x)\theta(x) + \bar{P}_X(x)\frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x}, P_{Y_1|X=x})\right), \tag{97}$$

which implies that

$$\begin{aligned}
\log_2\left(\hat{M}_n\right) &\leqslant \frac{1}{1-\hat{\epsilon}_n}\left(1+n\sum_{x\in\mathcal{X}}\ell\bar{P}_X(x)\theta(x)+\bar{P}_X(x)\frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\right)\\
&\quad-\log_2\left(\frac{\eta}{2}-\frac{c}{\sqrt{n}}-\epsilon_n-\hat{\epsilon}_n\right)\\
&\overset{(a)}{\leqslant} \frac{1}{1-\hat{\epsilon}_n}\left(1+\sum_{x\in\mathcal{X}}\ell\omega(x)+\omega(x)^3\max_{x'\in\mathcal{X}}\frac{\chi_3(\hat{R}_{Y_1|X=x'},P_{Y_1|X=x'})}{6n^2\bar{P}_X(x')^2}\right)\\
&\quad-\log_2\left(\frac{\eta}{2}-\frac{c}{\sqrt{n}}-\epsilon_n-\hat{\epsilon}_n\right)\\
&\overset{(b)}{\leqslant} \frac{1}{1-\hat{\epsilon}_n}\left(1+2\frac{\ell\sqrt{n}}{\sqrt{d}}Q^{-1}\left(\frac{1-\delta-\eta}{2}\right)+\left(\frac{2\sqrt{n}}{\sqrt{d}}Q^{-1}\left(\frac{1-\delta-\eta}{2}\right)\right)^3\right.\\
&\quad\left.\cdot\sum_{x\in\mathcal{X}}\max_{x'\in\mathcal{X}}\frac{\chi_3(\hat{R}_{Y_1|X=x'},P_{Y_1|X=x'})}{6n^2\bar{P}_X(x')^2}\right)-\log_2\left(\frac{\eta}{2}-\frac{c}{\sqrt{n}}-\epsilon_n-\hat{\epsilon}_n\right)\\
&= \frac{1}{1-\hat{\epsilon}_n}\left(1+2\frac{\ell\sqrt{n}}{\sqrt{d}}Q^{-1}\left(\frac{1-\delta-\eta}{2}\right)+\frac{4\left|\mathcal{X}\right|}{3\sqrt{n}\sqrt{d}^3}Q^{-1}\left(\frac{1-\delta-\eta}{2}\right)^3\right.\\
&\quad\left.\cdot\max_{x'\in\mathcal{X}}\frac{\chi_3(\hat{R}_{Y_1|X=x'},P_{Y_1|X=x'})}{\bar{P}_X(x')^2}\right)-\log_2\left(\frac{\eta}{2}-\frac{c}{\sqrt{n}}-\epsilon_n-\hat{\epsilon}_n\right),
\end{aligned}\tag{98}$$

where $c$ is constant that depends only on the parameters of the random transformation in (2), *(a)* follows from Lemma 3, and *(b)* follows from the fact that

$$\omega(x)\leqslant\sum_{x\in\mathcal{X}}\omega(x)=\sum_{i=1}^{M_n}\sum_{j=1}^{\hat{M}_n}\frac{\omega(i,j)}{M_n\hat{M}_n}\leqslant 2\frac{\sqrt{n}}{\sqrt{d}}Q^{-1}\left(\frac{1-\delta-\eta}{2}\right).\tag{99}$$

The proof is completed by dividing both hand-sides of (98) by $\sqrt{n}$ and taking the limit. ■

Note that for channels satisfying (58) and (59), the right-hand side of (88) reduces to

$$\frac{2\ell}{\sqrt{d}}Q^{-1}\left(\frac{1-\delta}{2}\right).\tag{100}$$

Recalling that $\eta$ in (93) can be chosen arbitrarily small, it follows that, for such channels, the asymptotic bounds in Theorem 5 and Theorem 2 are tight, *i.e.*, (100) gives the optimal scaling constant for $\log_2\left(\hat{M}_n^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)$ with respect to $\sqrt{n}$.

# 8 Example

This section presents an example to illustrate the results in Theorem 5.

**Example 3.** *Consider the random transformation in* (2) *such that* $\mathcal{X}=\mathcal{Y}_1=\mathcal{Y}_2=\{0,1,2\}$, *and such that for all* $(x,x')\in\mathcal{X}^2$ *with* $x\neq x'$, *the conditional probability distributions* $P_{Y_1|X}$ *and* $P_{Y_2|Y_1}$ *respectively satisfy:*

$$P_{Y_1|X}(x|x)=1-2P_{Y_1|X}(x'|x)=1-2p_1,\ \textit{and}\tag{101}$$
$$P_{Y_2|Y_1}(x|x)=1-2P_{Y_2|Y_1}(x'|x)=1-2p_2,\tag{102}$$

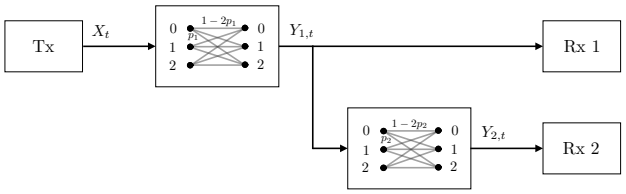



Figure 4: Degraded broadcast channel satisfying (58) and (59) at channel use $t \in \{1, 2, \ldots, n\}$.

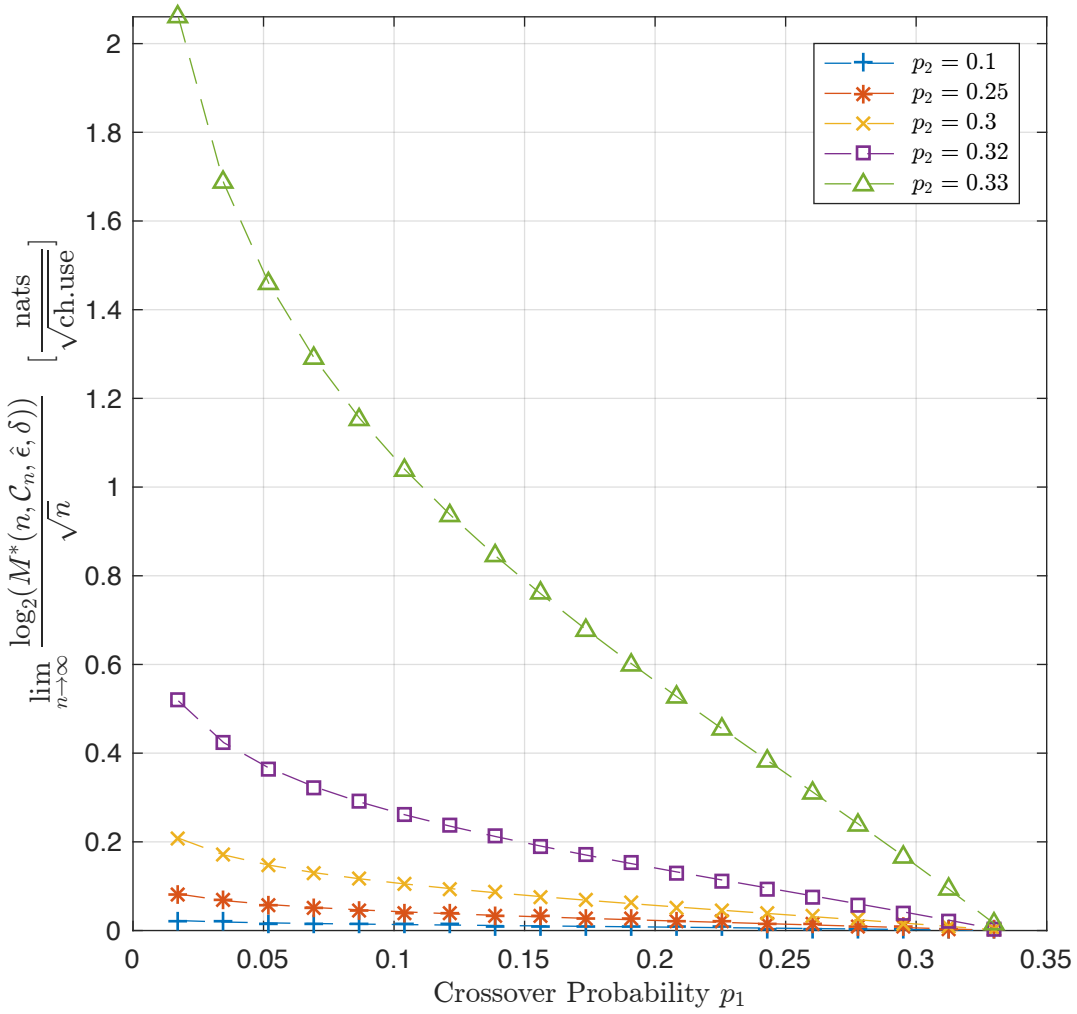



Figure 5: Fundamental limit $\lim_{n\to\infty} \frac{\log_2\left(\hat{M}^*(n, \mathcal{C}_n, \hat{\epsilon}_n, \delta)\right)}{\sqrt{n}}$ as a function of the crossover probability $p_1$, for $\delta = 0.005$.

*with* $(p_1, p_2) \in ]0, \frac{1}{3}[^2$.

Figure 4 depicts the channel in Example 3. The probability distribution $P_{Y_2|X}$ verifies that

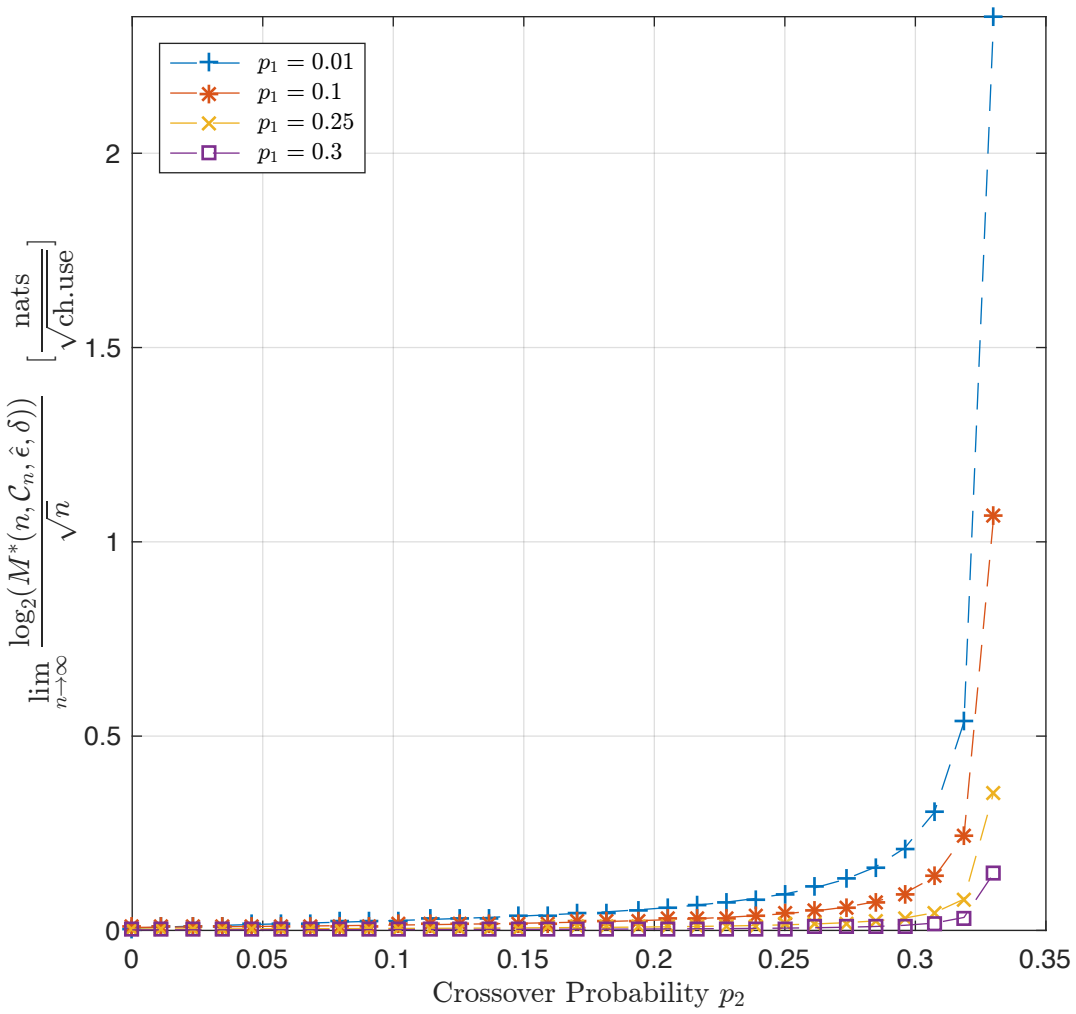



Figure 6: Fundamental limit $\lim_{n\to\infty} \frac{\log_2\left(\hat{M}^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)}{\sqrt{n}}$ as a function of the crossover probability $p_1$, for $\delta = 0.005$.

for all $(x, x') \in \mathcal{X}^2$ such that $x \neq x'$:

$$\begin{aligned} P_{Y_2|X}(x|x) = 1 - 2P_{Y_2|X}(x'|x) &= 1 - 2\left(p_1 + p_2 - 3p_1p_2\right) \\ &= 1 - 2p, \end{aligned} \tag{103}$$

with

$$p = p_1 + p_2 - 3p_1p_2. \tag{104}$$

The following lemma quantifies the expressions $\chi_2(P_{Y_2|X=x'}, P_{Y_2|X=x})$ and $D\left(P_{Y_2|X=x'}||P_{Y_2|X=x}\right)$ for any pair $(x, x') \in \mathcal{X}^2$ with $x \neq x'$.

**Lemma 6.** *Consider Example 3. For all pairs $(x, x') \in \mathcal{X}^2$, with $x \neq x'$, it holds that*

$$\chi_2(P_{Y_2|X=x'}, P_{Y_2|X=x}) = \frac{(3p-1)^2(1-p)}{p(1-2p)}, \tag{105}$$

$$D\big(P_{Y_1|X=x'} || P_{Y_1|X=x}\big) = (1-3p_1)\log_2\left(\frac{1-2p_1}{p_1}\right), \tag{106}$$

*where $p$ is defined in* (104).

*Proof:* The proof of Lemma 6 is presented in Appendix F. ∎

The following proposition follows immediately from Lemma 6 and Theorem 5,.

**Proposition 8.** *Consider Example 3 and consider a sequence of $(n, M_n, \epsilon_n)$-broadcast codes, with $n \in \{1, 2, \ldots\}$, denoted respectively by $\mathcal{C}_1$, $\mathcal{C}_2$, ..., such that $\lim_{n\to\infty} \epsilon_n = 0$. Then, there always exists a sequence of $(n, \mathcal{C}_n, \hat{M}_n, \hat{\epsilon}_n, \delta)$-covert codes with $\lim_{n\to\infty} \hat{\epsilon}_n = 0$, such that*

$$\lim_{n\to\infty} \frac{\log_2\left(\hat{M}_n^*(n, \mathcal{C}_n, \hat{\epsilon}_n, \delta)\right)}{\sqrt{n}} \geqslant 2Q^{-1}\left(\frac{1-\delta}{2}\right)\sqrt{\frac{p(1-2p)}{1-p}}\frac{1-3p_1}{1-3p}\log_2\left(\frac{1-2p_1}{p_1}\right). \tag{107}$$

The left hand side of (107) is plotted as a function of the probability $p_1$ and $p_2$ in Figure 5 and in Figure 6, respectively, with $\delta = 0.005$.

# 9 Conclusion

So far, a tight converse for general DM-BCs, i.e., those that do not necessarily satisfy the conditions in (58) and (59) is still an open problem. An interesting question is whether the total variation distance used in the current work can be replaced by the Kullback-Leibler divergence.

Finally, it is interesting to highlight that the problem introduced in this report is an instance of a more general problem. In multi-user channels, broadcast codes can be altered to perform other functionalities, e.g., simultaneous energy and information transmission to an energy harvester, physical-layer secrecy, etc.

# A Auxiliary Results

This section introduces some auxiliary results that play a key role in the following appendices.

**Theorem 3** (Berry-Esseen Theorem)**.** *Let $X_1, X_2, \ldots, X_n$ be independent random variables such that for all $t \in \{1, 2, \ldots, n\}$,*

$$\mu_t = \mathbb{E}_{X_t}[X_t], \tag{108}$$

$$\sigma_t^2 = \mathbb{E}_{X_t}\left[X_t^2\right] - \mu_t^2, \tag{109}$$

$$\phi_t = \mathbb{E}_{X_t}\left[|X_t - \mu_t|^3\right]. \tag{110}$$

*Then, it holds for all $\lambda \in \mathbb{R}$ that*

$$\left|\Pr\left[\sum_{t=1}^{n} X_t - \mu_t \geqslant \sigma\lambda\right] - Q(\lambda)\right| \leqslant \frac{c_0\phi}{\sigma^3}, \tag{111}$$

*where*

$$\mu = \sum_{t=1}^{n}\mu_t, \quad \sigma^2 = \sum_{t=1}^{n}\sigma_t^2, \quad \text{and} \quad \phi = \sum_{t=1}^{n}\phi_t. \tag{112}$$

The best value of the constant $c_0$ is $c_0 = 0.4748$ [10].

**Lemma 7.** *Let $P_X$ and $P_Y$ be two probability distribution functions on a common finite support $\mathcal{Z}$. Let also $X$ and $Y$ be two random variables following the distributions $P_X$ and $P_Y$, respectively. Then,*

$$\|P_X - P_Y\|_{\mathrm{TV}} = \Pr\left[P_X(X) \geqslant P_Y(X)\right] - \Pr\left[P_X(Y) \geqslant P_Y(Y)\right], \tag{113}$$

*where the probability operators apply with respect to $P_X$ and $P_Y$, respectively.*

*Proof:* The proof consists in the following algebraic manipulations:

$$\begin{aligned}
\|P_X - P_Y\|_{\mathrm{TV}} &= \frac{1}{2}\sum_{z\in\mathcal{Z}} |P_X(z) - P_Y(z)| \\
&= \frac{1}{2}\sum_{\substack{z\in\mathcal{Z}:\\ P_X(z)\geqslant P_Y(z)}} P_X(z) - P_Y(z) + \frac{1}{2}\sum_{\substack{z\in\mathcal{Z}:\\ P_X(z)\leqslant P_Y(z)}} P_Y(z) - P_X(z) \\
&= \frac{1}{2}\sum_{z\in\mathcal{Z}} P_X(z)\mathbb{1}_{\{P_X(z)\geqslant P_Y(z)\}} - \frac{1}{2}\sum_{z\in\mathcal{Z}} P_Y(z)\mathbb{1}_{\{P_X(z)\geqslant P_Y(z)\}} \\
&\quad + \frac{1}{2}\sum_{z\in\mathcal{Z}} P_Y(z)\mathbb{1}_{\{P_X(z)\leqslant P_Y(z)\}} - \frac{1}{2}\sum_{z\in\mathcal{Z}} P_X(z)\mathbb{1}_{\{P_X(z)\leqslant P_Y(z)\}} \\
&= \frac{1}{2}\Big(\Pr\left[P_X(X) \geqslant P_Y(X)\right] - \Pr\left[P_X(Y) \geqslant P_Y(Y)\right]) \\
&\quad + \Pr\left[P_X(Y) \leqslant P_Y(Y)\right] - \Pr\left[P_X(X) \leqslant P_Y(X)\right]\Big) \\
&= \Pr\left[P_X(X) \geqslant P_Y(X)\right] - \Pr\left[P_X(Y) \geqslant P_Y(Y)\right],
\end{aligned} \tag{114}$$

and this completes the proof. ∎

Consider a hypothesis test designed to determine whether the distribution $P_X$ (hypothesis $H_0$) or the distribution $Q_X$ (hypothesis $H_1$) is used upon the observation of $X$:

$$\begin{cases} H_0 : X \sim P_X \\ H_1 : X \sim Q_X. \end{cases} \tag{115}$$

Denote by $\alpha \in [0,1]$ and $\beta \in [0,1]$ the type-I and type-II error probabilities associated with a decision rule $T : \mathcal{X} \to \{0,1\}$ of the form

$$T(x) \triangleq \begin{cases} 0 & \text{if } H_0 \text{ is accepted,} \\ 1 & \text{if } H_1 \text{ is accepted.} \end{cases} \tag{116}$$

That is,

$$\alpha \triangleq \Pr\left[T(X) = 1 | H_0\right], \text{ and} \tag{117}$$

$$\beta \triangleq \Pr\left[T(X) = 0 | H_1\right]. \tag{118}$$

Given the above hypothesis test, the following lemma establishes a lower bound on the total variation $\|P_X - Q_X\|_{\mathrm{TV}}$ in terms of the type-I and type-II error probabilities.

**Lemma 8** (Minimum Total Variation)**.** *Given the hypothesis test in* (115)*, it holds that*

$$\|P_X - Q_X\|_{\mathrm{TV}} \geqslant 1 - \alpha - \beta, \tag{119}$$

*with equality for the optimal decision rule $T$.*

*Proof:* The proof of Lemma 8 essentially relies on applying the inequality $|x| \geqslant x$, for all $x \in \mathbb{R}$, after judiciously developing the total variation $\|P_X - Q_X\|_{\mathrm{TV}}$. Then, identifying the type-I and type-II error probabilities in the resulting expression yields the lower bound in (53).

Given the hypothesis test in (49), let $\mathcal{A} \subseteq \mathcal{X}$ be an arbitrary acceptance region for the hypothesis $H_0$. That is, the decision rule $T$ in (50) is such that $T(x) = \mathbb{1}_{\{x \notin \mathcal{A}\}}$. Then, it follows that

$$\begin{aligned}
\|P_X - Q_X\|_{\mathrm{TV}} &= \frac{1}{2}\sum_{x\in\mathcal{X}} |P_X(x) - Q_X(x)| \\
&= \frac{1}{2}\sum_{x\in\mathcal{A}} |P_X(x) - Q_X(x)| + \frac{1}{2}\sum_{x\in\mathcal{A}^{\mathrm{c}}} |P_X(x) - Q_X(x)| \\
&\geqslant \frac{1}{2}\left(\sum_{x\in\mathcal{A}} P_X(x) - Q_X(x) + \sum_{x\in\mathcal{A}^{\mathrm{c}}} Q_X(x) - P_X(x)\right) \\
&= \frac{1}{2}\left(1 - \sum_{x\in\mathcal{A}^{\mathrm{c}}} P_X(x) - \sum_{x\in\mathcal{A}} Q_X(x) + 1 - \sum_{x\in\mathcal{A}} Q_X(x) - \sum_{x\in\mathcal{A}^{\mathrm{c}}} P_X(x)\right) \\
&= 1 - \sum_{x\in\mathcal{A}} Q_X(x) - \sum_{x\in\mathcal{A}^{\mathrm{c}}} P_X(x) \\
&= 1 - \Pr\left[T(X) = 0 | H_1\right] - \Pr\left[T(X) = 1 | H_0\right] \\
&= 1 - \alpha - \beta,
\end{aligned} \tag{120}$$

with $\alpha$ and $\beta$ respectively in (51) and (52). Note that (120) holds with equality if the acceptance region is chosen such that $\mathcal{A} = \{x \in \mathcal{X} | P_X(x) \geqslant Q_X(x)\}$. Combining the lower-bound in (120) and the covertness constraint in (27) yields (53) and completes the proof. ■

**Lemma 9.** *Let $(x, a, b) \in \mathbb{R}_+^3$ and $n \in \mathbb{N}$. Then, it holds that*

$$Q\left(\frac{x - \frac{a}{\sqrt{n}}}{2\sqrt{x + \frac{b}{\sqrt{n}}}}\right) \leqslant Q\left(\frac{\sqrt{x}}{2}\right) + \frac{2a + b}{4\sqrt{2\pi n(x + b)}}. \tag{121}$$

*Proof:* Note that

$$\begin{aligned}
Q\left(\frac{x-\frac{a}{\sqrt{n}}}{2\sqrt{x+\frac{b}{\sqrt{n}}}}\right) &= Q\left(\frac{\sqrt{x}}{2}\,\frac{1-\frac{a}{x\sqrt{n}}}{\sqrt{1+\frac{b}{x\sqrt{n}}}}\right)\\
&= Q\left(\frac{\sqrt{x}}{2}\,\frac{1-\frac{a}{x\sqrt{n}}+\sqrt{1+\frac{b}{x\sqrt{n}}}-\sqrt{1+\frac{b}{x\sqrt{n}}}}{\sqrt{1+\frac{b}{x\sqrt{n}}}}\right)\\
&= Q\left(\frac{\sqrt{x}}{2}\left(1+\frac{1-\frac{a}{x\sqrt{n}}-\sqrt{1+\frac{b}{x\sqrt{n}}}}{\sqrt{1+\frac{b}{x\sqrt{n}}}}\right)\right)\\
&\overset{(a)}{\leqslant} Q\left(\frac{\sqrt{x}}{2}\left(1+\frac{1-\frac{a}{x\sqrt{n}}-1-\frac{b}{2x\sqrt{n}}}{\sqrt{1+\frac{b}{x\sqrt{n}}}}\right)\right)\\
&= Q\left(\frac{\sqrt{x}}{2}\left(1-\frac{\frac{a}{x\sqrt{n}}+\frac{b}{2x\sqrt{n}}}{\sqrt{1+\frac{b}{x\sqrt{n}}}}\right)\right)\\
&= Q\left(\frac{\sqrt{x}}{2}\left(1-\frac{2a+b}{2x\sqrt{n\left(1+\frac{b}{x\sqrt{n}}\right)}}\right)\right)\\
&\overset{(b)}{\leqslant} Q\left(\frac{\sqrt{x}}{2}\right)+\frac{2a+b}{4\sqrt{2\pi xn\left(1+\frac{b}{x\sqrt{n}}\right)}}\\
&\leqslant Q\left(\frac{\sqrt{x}}{2}\right)+\frac{2a+b}{4\sqrt{2\pi n\left(x+\frac{b}{\sqrt{n}}\right)}}\\
&\leqslant Q\left(\frac{\sqrt{x}}{2}\right)+\frac{2a+b}{4\sqrt{2\pi n\,(x+b)}}, \qquad (122)
\end{aligned}$$

where *(a)* follows since $\sqrt{1+x}\leq 1+\frac{x}{2}$ for all $x\geqslant -1$; and *(b)* follows since $Q(x-y)\leqslant Q(x)+\frac{y}{\sqrt{2\pi}}$ for all $(x,y)\in\mathbb{R}_+^2$ such that $0\leqslant y\leqslant x$. ■

# B Proof of Lemma 1

Note that from (18), it follows that

$$\begin{aligned}
\bar{P}_{X\hat{X}}(x,\hat{x}) &= \frac{1}{nM\hat{M}}\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\sum_{t=1}^{n}\mathbb{1}_{\{x=u_t(i)\}}\mathbb{1}_{\{\hat{x}=v_t(i,j)\}} \\
&= \frac{1}{nM\hat{M}}\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\sum_{t=1}^{n}\mathbb{1}_{\{x=u_t(i)\}}\mathbb{1}_{\{\hat{x}=v_t(i,j)\}}\mathbb{1}_{\{x=\hat{x}\}} \\
&\quad + \frac{1}{nM\hat{M}}\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\sum_{t=1}^{n}\mathbb{1}_{\{x=u_t(i)\}}\mathbb{1}_{\{\hat{x}=v_t(i,j)\}}\mathbb{1}_{\{x\neq\hat{x}\}} \\
&= \sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\frac{N(x,\hat{x}|\boldsymbol{u}(i),\boldsymbol{v}(i,j))}{nM\hat{M}}\mathbb{1}_{\{x=\hat{x}\}} + \frac{\omega(x)}{n}\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\sum_{t=1}^{n}\frac{\mathbb{1}_{\{x=u_t(i)\}}\mathbb{1}_{\{\hat{x}=v_t(i,j)\}}\mathbb{1}_{\{x\neq\hat{x}\}}}{\omega(x)M\hat{M}} \\
&\overset{(a)}{=} \sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\frac{N(x,x|\boldsymbol{u}(i),\boldsymbol{v}(i,j))}{nM\hat{M}}\mathbb{1}_{\{x=\hat{x}\}} + \frac{\omega(x)}{n}\hat{P}_{\hat{X}|X}(\hat{x}|x) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\frac{\sum_{t=1}^{n}\mathbb{1}_{\{x=u_t(i)\}}\mathbb{1}_{\{x=v_t(i,j)\}}}{nM\hat{M}}\mathbb{1}_{\{x=\hat{x}\}} + \frac{\omega(x)}{n}\hat{P}_{\hat{X}|X}(\hat{x}|x) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\frac{\sum_{t=1}^{n}\mathbb{1}_{\{x=u_t(i)\}}\left(1-\mathbb{1}_{\{x\neq v_t(i,j)\}}\right)}{nM\hat{M}}\mathbb{1}_{\{x=\hat{x}\}} + \frac{\omega(x)}{n}\hat{P}_{\hat{X}|X}(\hat{x}|x) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\frac{N(x|\boldsymbol{u}(i)) - \sum_{t=1}^{n}\mathbb{1}_{\{x=u_t(i)\}}\mathbb{1}_{\{x\neq v_t(i,j)\}}}{nM\hat{M}}\mathbb{1}_{\{x=\hat{x}\}} + \frac{\omega(x)}{n}\hat{P}_{\hat{X}|X}(\hat{x}|x) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\frac{N(x|\boldsymbol{u}(i))}{nM\hat{M}}\mathbb{1}_{\{x=\hat{x}\}} - \sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\frac{\sum_{t=1}^{n}\mathbb{1}_{\{x=u_t(i)\}}\mathbb{1}_{\{x\neq v_t(i,j)\}}}{nM\hat{M}}\mathbb{1}_{\{x=\hat{x}\}} \\
&\quad + \frac{\omega(x)}{n}\hat{P}_{\hat{X}|X}(\hat{x}|x) \\
&\overset{(b)}{=} \bar{P}_X(x)\mathbb{1}_{\{x=\hat{x}\}} - \frac{\omega(x)}{n}\mathbb{1}_{\{x=\hat{x}\}} + \frac{\omega(x)}{n}\hat{P}_{\hat{X}|X}(\hat{x}|x) \\
&= \bar{P}_X(x)\left(1-\frac{\omega(x)}{n\bar{P}_X(x)}\right)\mathbb{1}_{\{x=\hat{x}\}} + \frac{\omega(x)}{n}\hat{P}_{\hat{X}|X}(\hat{x}|x) \\
&\overset{(c)}{=} \bar{P}_X(x)\left((1-\theta(x))\,\mathbb{1}_{\{x=\hat{x}\}} + \theta(x)\hat{P}_{\hat{X}|X}(\hat{x}|x)\right), \qquad (123)
\end{aligned}$$

where *(a)* follows from (19); *(b)* follows from (16); and *(c)* follows from Lemma 2. This completes the proof. ■

# C Proof of Lemma 2

Given an $(n, M)$-broadcast code $\mathcal{C}$ and an $(n, \mathcal{C}, \hat{M})$-induced code, note that any message index pair $(i, j) \in \mathcal{W} \times \hat{\mathcal{W}}$ satisfies for all $x \in \mathcal{X}$:

$$\sum_{t=1}^{n} \mathbb{1}_{\{x=u_t(i)\}} = \sum_{t=1}^{n} \mathbb{1}_{\{x=u_t(i)\}} \left( \mathbb{1}_{\{x=v_t(i,j)\}} + \mathbb{1}_{\{x\neq v_t(i,j)\}} \right). \tag{124}$$

Developping the right hand-side of (124) and dividing the two hand-sides by the block-length $n$ yields

$$\frac{N(x|\boldsymbol{u}(i))}{n} = \frac{N(x, x|\boldsymbol{u}(i), \boldsymbol{v}(i,j))}{n} + \sum_{t=1}^{n} \frac{\mathbb{1}_{\{x=u_t(i)\}} \mathbb{1}_{\{x\neq v_t(i,j)\}}}{n}. \tag{125}$$

Therefore, summing over all pairs of message indices $(i, j) \in \mathcal{W} \times \hat{\mathcal{W}}$, and normalizing by the total number of messages $M \cdot \hat{M}$ yields

$$\begin{aligned} \bar{P}_X(x) &= \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \frac{N(x, x|\boldsymbol{u}(i), \boldsymbol{v}(i,j))}{nM\hat{M}} + \sum_{t=1}^{n} \frac{\mathbb{1}_{\{x=u_t(i)\}} \mathbb{1}_{\{x\neq v_t(i,j)\}}}{nM\hat{M}}. \\ &= \bar{P}_{X\hat{X}}(x, x) + \frac{\omega(x)}{n}. \end{aligned} \tag{126}$$

Thus, it follows that

$$\begin{aligned} \omega(x) &= n \left( \bar{P}_X(x) - \bar{P}_{X\hat{X}}(x, x) \right) \\ &= n\bar{P}_X(x) \left( 1 - \bar{P}_{\hat{X}|X}(x|x) \right) \\ &= n\bar{P}_X(x)\theta(x), \end{aligned} \tag{127}$$

where the last equality follows from the definition of $\theta(x)$ in (21). This completes the proof. ■

# D Proof of Lemma 5

Note that from the triangle inequality, it follows that

$$\begin{aligned} \|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} &= \frac{1}{2} \sum_{\boldsymbol{y} \in \mathcal{Y}_2} |Q_{\boldsymbol{Y}_2}(\boldsymbol{y}) - R_{\boldsymbol{Y}_2}(\boldsymbol{y})| \\ &= \frac{1}{2} \sum_{\boldsymbol{y} \in \operatorname{supp} Q_{\boldsymbol{Y}_2}} |Q_{\boldsymbol{Y}_2}(\boldsymbol{y}) - R_{\boldsymbol{Y}_2}(\boldsymbol{y})| + \frac{1}{2} \sum_{\substack{\boldsymbol{y} \in \operatorname{supp} R_{\boldsymbol{Y}_2} \setminus \\ \operatorname{supp} Q_{\boldsymbol{Y}_2}}} |Q_{\boldsymbol{Y}_2}(\boldsymbol{y}) - R_{\boldsymbol{Y}_2}(\boldsymbol{y})| \\ &\geqslant \frac{1}{2} \left( 1 - \Pr\left[ \boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2} \right] + \Pr\left[ \boldsymbol{Y}_2 \notin \operatorname{supp} Q_{\boldsymbol{Y}_2} \right] \right) \\ &\geqslant \frac{1}{2} \left( 1 - \Pr\left[ \boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2} \right] \right), \end{aligned} \tag{128}$$

where the random variable $\boldsymbol{Y}_2$ is distributed according to $R_{\boldsymbol{Y}_2}$.

# E Proof of Lemma 4

Let $\bar{W} \in \mathcal{W}$ be a random variable that represents the decoded message index at Receiver 2. Consider the joint distributions $Q_{\bar{W}\boldsymbol{Y}_2}$ and $R_{\bar{W}\boldsymbol{Y}_2}$ such that, for all pairs $(i, \boldsymbol{y}) \in \mathcal{W} \times \mathcal{Y}_2^n$,

$$Q_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) = Q_{\boldsymbol{Y}_2}(\boldsymbol{y}) Q_{\bar{W}|\boldsymbol{Y}_2}(i|\boldsymbol{y}), \tag{129}$$

$$\text{and} \quad R_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) = R_{\boldsymbol{Y}_2}(\boldsymbol{y}) R_{\bar{W}|\boldsymbol{Y}_2}(i|\boldsymbol{y}), \tag{130}$$

where $Q_{\boldsymbol{Y}_2}$ and $R_{\boldsymbol{Y}_2}$ are the marginal channel output distributions respectively in (25) and (26), and

$$Q_{\bar{W}|\boldsymbol{Y}_2}(i|\boldsymbol{y}) = R_{\bar{W}|\boldsymbol{Y}_2}(i|\boldsymbol{y}) = \mathbb{1}_{\{\boldsymbol{y} \in \mathcal{D}_2(i)\}}. \tag{131}$$

Consider also the joint distributions $Q_{W\boldsymbol{Y}_2}$ and $R_{W\boldsymbol{Y}_2}$ respectively in (28) and (29). Note that

$$\begin{aligned} \|R_{W\boldsymbol{Y}_2} - Q_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} = \frac{1}{2} \sum_{i=1}^{M} \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} \big| & R_{W\boldsymbol{Y}_2}(i, \boldsymbol{y}) + R_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) - R_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) \\ & - Q_{W\boldsymbol{Y}_2}(i, \boldsymbol{y}) + Q_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) - Q_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) \big| \\ \leqslant \big\|R_{\bar{W}\boldsymbol{Y}_2} - Q_{\bar{W}\boldsymbol{Y}_2}\big\|_{\mathrm{TV}} & + \big\|Q_{W\boldsymbol{Y}_2} - Q_{\bar{W}\boldsymbol{Y}_2}\big\|_{\mathrm{TV}} + \big\|R_{W\boldsymbol{Y}_2} - R_{\bar{W}\boldsymbol{Y}_2}\big\|_{\mathrm{TV}} \end{aligned} \tag{132}$$

where the last inequality follows from the triangle inequality. The remainder of the proof consists in establishing an upper-bound on each of the three terms in the right hand-side of (132).

First, note that

$$\begin{aligned} \big\|R_{\bar{W}\boldsymbol{Y}_2} - Q_{\bar{W}\boldsymbol{Y}_2}\big\|_{\mathrm{TV}} &= \frac{1}{2M} \sum_{i=1}^{M} \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} \mathbb{1}_{\{\boldsymbol{y} \in \mathcal{D}_2(i)\}} \left| R_{\boldsymbol{Y}_2}(\boldsymbol{y}) - Q_{\boldsymbol{Y}_2}(\boldsymbol{y}) \right| \\ &\overset{(a)}{=} \frac{1}{2} \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} \left| R_{\boldsymbol{Y}_2}(\boldsymbol{y}) - Q_{\boldsymbol{Y}_2}(\boldsymbol{y}) \right| \\ &= \|R_{\boldsymbol{Y}_2} - Q_{\boldsymbol{Y}_2}\|_{\mathrm{TV}}, \end{aligned} \tag{133}$$

where *(a)* holds since (5c) is assumed with equality.

Note also that

$$\begin{aligned} \big\|Q_{W\boldsymbol{Y}_2} - Q_{\bar{W}\boldsymbol{Y}_2}\big\|_{\mathrm{TV}} &= \frac{1}{2} \sum_{i=1}^{M} \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} Q_{\boldsymbol{Y}_2}(\boldsymbol{y}) \left| Q_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y}) - Q_{\bar{W}|\boldsymbol{Y}_2}(i|\boldsymbol{y}) \right| \\ &= \frac{1}{2} \sum_{i=1}^{M} \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} Q_{\boldsymbol{Y}_2}(\boldsymbol{y}) \left| Q_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y}) - \mathbb{1}_{\{\boldsymbol{y} \in \mathcal{D}_2(i)\}} \right| \\ &= \frac{1}{2} \sum_{i=1}^{M} \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} Q_{\boldsymbol{Y}_2}(\boldsymbol{y}) \Big( \mathbb{1}_{\{\boldsymbol{y} \in \mathcal{D}_2(i)\}} \big(1 - Q_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y})\big) + \mathbb{1}_{\{\boldsymbol{y} \in \mathcal{D}_2^{\mathrm{c}}(i)\}} Q_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y}) \Big) \\ &= \frac{1}{2} \sum_{i=1}^{M} \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} \Big( Q_{\boldsymbol{Y}_2}(\boldsymbol{y}) \mathbb{1}_{\{\boldsymbol{y} \in \mathcal{D}_2(i)\}} - Q_{\boldsymbol{Y}_2}(\boldsymbol{y}) Q_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y}) \mathbb{1}_{\{\boldsymbol{y} \in \mathcal{D}_2(i)\}} \\ &\quad + Q_{\boldsymbol{Y}_2}(\boldsymbol{y}) Q_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y}) \mathbb{1}_{\{\boldsymbol{y} \in \mathcal{D}_2^{\mathrm{c}}(i)\}} \Big) \\ &\overset{(a)}{=} \frac{1}{2} \left( 1 - 1 + 2 \sum_{i=1}^{M} \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} Q_{\boldsymbol{Y}_2}(\boldsymbol{y}) Q_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y}) \mathbb{1}_{\{\boldsymbol{y} \in \mathcal{D}_2^{\mathrm{c}}(i)\}} \right) \\ &\leqslant \epsilon, \end{aligned} \tag{134}$$

where *(a)* holds since (5c) holds with equality. Note that since (11c) is assumed with equality, the equality in (131) ensures that the same steps can be followed with the total variation $\left\|R_{W\boldsymbol{Y}_2} - R_{\bar{W}\boldsymbol{Y}_2}\right\|_{\mathrm{TV}}$. This yields

$$\left\|R_{W\boldsymbol{Y}_2} - R_{\bar{W}\boldsymbol{Y}_2}\right\|_{\mathrm{TV}} \leqslant \hat{\epsilon}. \tag{135}$$

Plugging (133)–(135) into (132) completes the proof. ∎

# F Proof of Lemma 6

Note that the probability distribution $P_{Y_2|X}$ verifies for all $(x, x') \in \mathcal{X}^2$ such that $x \neq x'$:

$$\begin{aligned} P_{Y_2|X}(x|x) = 1 - 2P_{Y_2|X}(x'|x) &= 1 - 2\left(p_1 + p_2 - 3p_1p_2\right) \\ &= 1 - 2p, \end{aligned} \tag{136}$$

with $p = p_1 + p_2 - 3p_1p_2$.

Note also that due to the nature of the channel, $\chi_2(P_{Y_2|X=x'}, P_{Y_2|X=x})$ verifies for all $(x, x') \in \mathcal{X}^2$ such that $x \neq x'$:

$$\begin{aligned} \chi_2(P_{Y_2|X=x'}, P_{Y_2|X=x}) &= \sum_{y\in\mathcal{Y}_2} \frac{\left(P_{Y_2|X}(y|x') - P_{Y_2|X}(y|x)\right)^2}{P_{Y_2|X}(y|x)} \\ &= \sum_{y\in\mathcal{Y}_2} \frac{\left(P_{Y_2|X}(y|x') - P_{Y_2|X}(y|x)\right)^2}{P_{Y_2|X}(y|x)} \\ &= \frac{\left(P_{Y_2|X}(x|x') - P_{Y_2|X}(x|x)\right)^2}{P_{Y_2|X}(x|x)} + \frac{\left(P_{Y_2|X}(x'|x') - P_{Y_2|X}(x'|x)\right)^2}{P_{Y_2|X}(x'|x)} \\ &\quad + \frac{\left(P_{Y_2|X}(x|x') - P_{Y_2|X}(x'|x)\right)^2}{P_{Y_2|X}(x'|x)} \\ &= \frac{(3p-1)^2}{1-2p} + \frac{(1-3p)^2}{p} \\ &= \frac{(3p-1)^2(1-p)}{p(1-2p)}, \end{aligned} \tag{137}$$

and $D\big(P_{Y_1|X=x'}||P_{Y_1|X=x}\big)$ verifies:

$$\begin{aligned} D\big(P_{Y_1|X=x'}||P_{Y_1|X=x}\big) &= \sum_{y\in\mathcal{Y}_2} P_{Y_1|X}(y|x') \log_2\left(\frac{P_{Y_1|X}(y|x')}{P_{Y_1|X}(y|x)}\right) \\ &= P_{Y_1|X}(x'|x') \log_2\left(\frac{P_{Y_1|X}(x'|x')}{P_{Y_1|X}(x'|x)}\right) + P_{Y_1|X}(x|x') \log_2\left(\frac{P_{Y_1|X}(x|x')}{P_{Y_1|X}(x|x)}\right) \\ &= (1-2p_1)\log_2\left(\frac{1-2p_1}{p_1}\right) + p_1 \log_2\left(\frac{p_1}{1-2p_1}\right) \\ &= (1-3p_1)\log_2\left(\frac{1-2p_1}{p_1}\right). \end{aligned} \tag{138}$$

# G Proof of Proposition 1

Given a fixed block-length $n \in \mathbb{N}$, an $(n, M, \epsilon)$-broadcast code $\mathcal{C}$ and an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code, consider for all message indices $i \in \mathcal{W}$ the set $\hat{\mathcal{W}}_i$ defined in (299).

Consider the distribution $\hat{R}_{Y_k|X}$ such that for all pairs $(x,y) \in \mathcal{X} \times \mathcal{Y}_2$ and $k \in \{1,2\}$,

$$\hat{R}_{Y_k|X}(y|x) = \sum_{\hat{x} \in \mathcal{X}} \hat{P}_{\hat{X}|X}(\hat{x}|x) P_{Y_k|X}(y|x), \tag{139}$$

with $\hat{P}_{\hat{X}|X}$ is in (19).

Define also for all $\boldsymbol{y} \in \mathcal{Y}_2^n$

$$B(\boldsymbol{y}) = \sum_{t=1}^{n} A(u_t(i), v_t(i,j^*), y_t), \tag{140}$$

with

$$A(x,\hat{x},y) = \frac{P_{Y_2|X}(y|\hat{x}) - P_{Y_2|X}(y|x)}{P_{Y_2|X}(y|x)}. \tag{141}$$

and $j^* \in \operatorname{argmax}_{j \in \hat{\mathcal{W}}_i} \Pr\left[B(\boldsymbol{Y}_2) < \tau | \hat{W} = j\right]$.

Consider the decision rule $T : \mathcal{Y}_2^n \to \{0,1\}$ of the form in (50) such that for all $\boldsymbol{y} \in \mathcal{Y}_2^n$,

$$T(\boldsymbol{y}) = \mathbb{1}_{\{B(\boldsymbol{y}) \geqslant \tau\}}, \tag{142}$$

with $\tau \in \mathbb{R}_+$ an arbitrary threshold.

Consider first the type-I error probability $\alpha_i$ in (56). It follows from the choice of $T$ in (142) that

$$\begin{aligned} \alpha_i &= \Pr\left[B(\boldsymbol{Y}_2) \geqslant \tau\right] \\ &= \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} \prod_{t=1}^{n} P_{Y_2|X}(y_t|u_t(i)) \mathbb{1}_{\{B(\boldsymbol{y}) \geqslant \tau\}}. \end{aligned} \tag{143}$$

For all $i \in \mathcal{W}$ and $t \in \{1,2,\ldots,n\}$, define the random variable

$$Z_{it} = A(u_t(i), v_t(i,j^*), Y), \tag{144}$$

where $Y$ is distributed according to $P_{Y_2|X=u_t(i)}$.

Then, it follows from (143) and the definition of the random variable $Z_{it}$ in (144) that

$$\alpha_i = \Pr\left[\sum_{t=1}^{n} Z_{it} \geqslant \tau\right]. \tag{145}$$

Denote by $\mu_{it}$, $\sigma_{it}$ and $\phi_{it}$ the first, second and third absolute moments of the random variable $Z_{it}$, respectively, *i.e.*,

$$\mu_{it} = \sum_{y \in \mathcal{Y}_2} P_{Y_2|X}(y|u_t(i)) A(u_t(i), v_t(i,j^*), y), \tag{146}$$

$$\sigma_{it}^2 = \sum_{y \in \mathcal{Y}_2} P_{Y_2|X}(y|u_t(i)) A(u_t(i), v_t(i,j^*), y)^2 - \mu_{it}^2, \tag{147}$$

and

$$\phi_{it} = \sum_{y \in \mathcal{Y}_2} P_{Y_2|X}(y|u_t(i)) \left|A(u_t(i), v_t(i,j^*), y) - \mu_{it}\right|^3. \tag{148}$$

**Lemma 10.** *Consider the random variables $Z_{it}$ in (144) with $i \in \mathcal{W}$ and $t \in \{1, 2, \ldots, n\}$. Then, it holds that*

$$\mu_i = \sum_{t=1}^{n} \mu_{it} = 0, \tag{149}$$

$$\sigma_i^2 = \sum_{t=1}^{n} \sigma_{it}^2 = nd, \tag{150}$$

$$\phi_i = \sum_{t=1}^{n} \phi_{it} \leqslant n\phi_i^*, \tag{151}$$

*with $d$ in (58) and*

$$\phi_i^* \triangleq \max_{t \in \{1,2,\ldots,n\}} \phi_{it}. \tag{152}$$

*Proof:* The proof of Lemma 10 is presented in Appendix H. ∎

It follows from (145) that

$$\begin{aligned}
\alpha_i &= \Pr\left[\sum_{t=1}^{n} Z_{it} - \mu_i \geqslant \sigma_i \frac{\tau - \mu_i}{\sigma_i}\right] \\
&\overset{(a)}{\leqslant} Q\left(\frac{\tau - \mu_i}{\sigma_i}\right) + c_0 \frac{\phi_i}{\sigma_i^3} \\
&\overset{(b)}{\leqslant} Q\left(\frac{\tau}{\sqrt{nd}}\right) + \frac{nc_0\phi_i^*}{\sqrt{nd}^3} \\
&\overset{(c)}{=} Q\left(\frac{\tau}{\sqrt{nd}}\right) + \frac{c_3}{\sqrt{n}},
\end{aligned} \tag{153}$$

where *(a)* follows from the Berry-Esseen Theorem (Theorem 3); *(b)* follows from Lemma 10; and *(c)* follows with

$$c_3 \triangleq c_0 \phi_i^* d^{-\frac{3}{2}}. \tag{154}$$

Consider now the type-II error probability $\beta_i$ in (57). It follows from the choice of $T$ in (142) that

$$\begin{aligned}
\beta_i &= \Pr\left[B(\boldsymbol{Y}_2) \leqslant \tau\right] \\
&= \frac{1}{|\hat{\mathcal{W}}_i|} \sum_{j \in \hat{\mathcal{W}}_i} \Pr\left[B(\boldsymbol{Y}_2) < \tau \middle| \hat{W} = j\right] \\
&\leqslant \max_{j \in \hat{\mathcal{W}}_i} \Pr\left[B(\boldsymbol{Y}_2) < \tau \middle| \hat{W} = j\right].
\end{aligned} \tag{155}$$

Let $j^*$ be

$$j^* \in \operatorname*{argmax}_{j \in \hat{\mathcal{W}}_i} \Pr\left[B(\boldsymbol{Y}_2) < \tau \middle| \hat{W} = j\right], \tag{156}$$

and define for all $t \in \{1, 2, \ldots, n\}$ the random variable

$$\hat{Z}_{it} = A(u_t(i), v_t(i, j^*), \hat{Y}), \tag{157}$$

where $\hat{Y}$ is distributed according to $P_{Y_2|X=v_t(i,j^*)}$. Then, it follows from (155) and the definition of the random variable $\hat{Z}_{it}$ in (157) that

$$\beta_i \leqslant \Pr\left[\sum_{t=1}^{n} \hat{Z}_{it} < \tau\right]. \tag{158}$$

Denote by $\hat{\mu}_{it}$, $\hat{\sigma}_{it}$ and $\hat{\phi}_{it}$ the first, second and third absolute moments of the random variable $\hat{Z}_{it}$, respectively, *i.e.*,

$$\hat{\mu}_{it} = \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))A(u_t(i),v_t(i,j^*),y), \tag{159}$$

$$\hat{\sigma}^2_{it} = \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))A(u_t(i),v_t(i,j^*),y)^2 - \hat{\mu}^2_{it}, \tag{160}$$

and

$$\hat{\phi}_{it} = \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))\left|A(u_t(i),v_t(i,j^*),y) - \hat{\mu}_{it}\right|^3. \tag{161}$$

**Lemma 11.** *Consider the random variables $\hat{Z}_{it}$ in (157) with $i \in \mathcal{W}$ and $t \in \{1,2,\ldots,n\}$. Then, it holds that*

$$\hat{\mu}_i = \sum_{t=1}^{n} \hat{\mu}_{it} = \omega(i,j^*)d, \tag{162}$$

$$\hat{\sigma}^2_i = \sum_{t=1}^{n} \hat{\sigma}^2_{it} \leqslant nd + \omega(i,j^*)\left|d'\right|, \tag{163}$$

$$\hat{\sigma}^2_i \geqslant n\left(d - \hat{\mu}^*_i\right) + \omega(i,j^*)d'', \tag{164}$$

$$\hat{\phi}_i = \sum_{t=1}^{n} \hat{\phi}_{it} \leqslant n\hat{\phi}^*_i, \tag{165}$$

*with $d$ in (58),*

$$\hat{\phi}^*_i \triangleq \max_{t\in\{1,2,\ldots,n\}} \hat{\phi}_{it}, \tag{166}$$

$$\hat{\mu}^*_i \triangleq \max_{t\in\{1,2,\ldots,n\}} \hat{\mu}^2_{it}, \tag{167}$$

$$d' \triangleq \max_{x\in\mathcal{X}} \chi_3(P_{Y_2|X=x}, P_{Y_2|X=u_t(i)}), \tag{168}$$

*and*

$$d'' \triangleq \min_{x\in\mathcal{X}} \chi_3(P_{Y_2|X=x}, P_{Y_2|X=u_t(i)}). \tag{169}$$

*Proof:* The proof of Lemma 11 is presented in Appendix I. ∎

It follows from (158) that

$$\begin{aligned}
\beta_i &\leqslant \Pr\left[\sum_{t=1}^{n} \hat{Z}_{it} - \hat{\mu}_i < \hat{\sigma}_i \frac{\tau - \hat{\mu}_i}{\hat{\sigma}_i}\right] \\
&= 1 - \Pr\left[\sum_{t=1}^{n} \hat{Z}_{it} - \hat{\mu}_i \geqslant \hat{\sigma}_i \frac{\tau - \hat{\mu}_i}{\hat{\sigma}_i}\right] \\
&\overset{(a)}{\leqslant} Q\left(-\frac{\tau - \hat{\mu}_i}{\hat{\sigma}_i}\right) + c_0 \frac{\hat{\phi}_i}{\hat{\sigma}_i^3} \\
&= Q\left(\frac{\hat{\mu}_i - \tau}{\hat{\sigma}_i}\right) + c_0 \frac{\hat{\phi}_i}{\hat{\sigma}_i^3} \\
&\overset{(b)}{\leqslant} Q\left(\frac{\omega(i,j^*)d - \tau}{\sqrt{nd + \omega(i,j^*)\,|d'|}}\right) + \frac{n c_0 \hat{\phi}_i^*}{\sqrt{n\,(d - \hat{\mu}_i^*) + \omega(i,j^*)\,|d'|}^3} \\
&\overset{(c)}{\leqslant} Q\left(\frac{\omega(i,j^*)d - \tau}{\sqrt{nd + \omega(i,j^*)\,|d'|}}\right) + \frac{c_4}{\sqrt{n}},
\end{aligned} \tag{170}$$

where $d'$ is in (168), *(a)* follows from the Berry-Esseen Theorem (Theorem 3); *(b)* follows from Lemma 11; and *(c)* follows with $c_4$ a positive constant

$$c_4 \triangleq c_0 \hat{\phi}^* \left(d - \hat{\mu}_i^*\right)^{-\frac{3}{2}}. \tag{171}$$

Finally, choosing $\tau$ such that

$$\tau = \frac{\nu d}{2}, \tag{172}$$

and plugging it into (153) and (170) yields respectively

$$\alpha_i \leqslant Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right) + \frac{c_3}{\sqrt{n}}, \tag{173}$$

and

$$\begin{aligned}
\beta_i &\leqslant Q\left(\frac{\omega(i,j^*)d - \frac{\nu d}{2}}{\sqrt{nd + \omega(i,j^*)\,|d'|}}\right) + \frac{c_4}{\sqrt{n}} \\
&\leqslant Q\left(\frac{\nu d}{2\sqrt{nd + \omega(i,j^*)\,|d'|}}\right) + \frac{c_4}{\sqrt{n}},
\end{aligned} \tag{174}$$

where the last inequality follows from the fact that by definition of $\hat{\mathcal{W}}_i$, $\omega(i,j^*) \geqslant \sqrt{n}\nu$.

Note that

$$\begin{aligned}
Q\left(\frac{\nu d}{2\sqrt{nd+\omega(i,j^*)\,|d'|}}\right) &= Q\left(\frac{\nu d}{2\sqrt{n(d+\frac{\omega(i,j^*)}{n}\,|d'|)}}\right)\\
&= Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\frac{1}{\sqrt{1+\frac{\omega(i,j^*)|d'|}{nd}}}\right)\\
&\overset{(a)}{\leqslant} Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\left(1-\frac{\omega(i,j^*)\,|d'|}{2nd}\right)\right)\\
&\overset{(b)}{\leqslant} Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right)+\frac{\nu\omega(i,j^*)\,|d'|}{4n\sqrt{2\pi nd}},\\
&\leqslant Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right)+\frac{\nu^2\,|d'|}{4n\sqrt{2\pi nd}},\\
&\leqslant Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right)+\frac{\nu^2\,|d'|}{4\sqrt{n}\sqrt{2\pi d}},
\end{aligned} \tag{175}$$

where *(a)* follows from the fact that $(1+x)^{-\frac{1}{2}} \geqslant 1-\frac{x}{2}$ for all $x \in \mathbb{R}_+$; and *(b)* follows from the fact that for all $0 \leqslant y \leqslant x$, it holds that $Q(x-y) \leqslant Q(x)+\frac{y}{\sqrt{2\pi}}$.

Thus, by letting

$$c_5 \triangleq c_4 + \frac{\nu^2\,|d'|}{4\sqrt{2\pi d}}, \tag{176}$$

it follows that

$$\beta \leqslant Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right)+\frac{c_5}{\sqrt{n}}. \tag{177}$$

This completes the proof. ∎

# H Proof of Lemma 10

For all $(i,j^*) \in \mathcal{W} \times \operatorname{argmax}_{j\in\hat{\mathcal{W}}_i} \Pr\left[B(\boldsymbol{y}) < \tau \middle| \hat{W}=j\right]$ and all $t \in \{1,2,\ldots,n\}$, note that

$$\begin{aligned}
\mu_{it} &= \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|u_t(i))A(u_t(i),v_t(i,j^*),y)\\
&= \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*)) - P_{Y_2|X}(y|u_t(i))\\
&= 0,
\end{aligned} \tag{178}$$

$$\begin{aligned}
\sigma_{it}^2 &= \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|u_t(i))A(u_t(i),v_t(i,j^*),y)^2 - \mu_{it}^2\\
&= \sum_{y\in\mathcal{Y}_2} \frac{\left(P_{Y_2|X}(y|v_t(i,j^*)) - P_{Y_2|X}(y|u_t(i))\right)^2}{P_{Y_2|X}(y|u_t(i))}\\
&\overset{(a)}{=} d,
\end{aligned} \tag{179}$$

where *(a)* follows from (58). Finally,

$$\begin{aligned}\phi_{it} &= \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|u_t(i))\left|A(u_t(i),v_t(i,j^*),y)-\mu_{it}\right|^3\\ &= \sum_{y\in\mathcal{Y}_2} \frac{\left|P_{Y_2|X}(y|v_t(i,j^*))-P_{Y_2|X}(y|u_t(i))\right|^3}{P_{Y_2|X}(y|u_t(i))^2}\\ &\leqslant \phi_i^*, \end{aligned} \tag{180}$$

with $\phi_i^*$ in (152).

Therefore, it follows that

$$\mu_i = \sum_{t=1}^{n}\mu_{it} = 0, \tag{181}$$

$$\sigma_i^2 = \sum_{t=1}^{n}\sigma_{it}^2 = nd, \tag{182}$$

$$\text{and}\quad \phi_i = \sum_{t=1}^{n}\phi_{it} \leqslant n\phi_i^*. \tag{183}$$

This completes the proof. ■

# I Proof of Lemma 11

For all $(i,j^*)\in\mathcal{W}\times\operatorname{argmax}_{j\in\hat{\mathcal{W}}_i}\Pr\left[B(\boldsymbol{y})<\tau\middle|\hat{W}=j\right]$, and all $t\in\{1,2,\ldots,n\}$, note that

$$\hat{\mu}_{it} = \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))A(u_t(i),v_t(i,j^*),y), \tag{184}$$

$$\hat{\sigma}_{it}^2 = \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))A(u_t(i),v_t(i,j^*),y)^2 - \hat{\mu}_{it}^2, \tag{185}$$

$$\begin{aligned}\text{and}\quad \hat{\phi}_{it} &= \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))\left|A(u_t(i),v_t(i,j^*),y)-\hat{\mu}_{it}\right|^3\\ &\leqslant \hat{\phi}_i^*, \end{aligned} \tag{186}$$

with $\hat{\phi}_i^*$ in (166).

Therefore, it follows that

$$\begin{aligned}
\hat{\mu}_i &= \sum_{t=1}^{n} \hat{\mu}_{it} \\
&= \sum_{t=1}^{n} \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))A(u_t(i),v_t(i,j^*),y) \\
&= \sum_{t=1}^{n} \sum_{y\in\mathcal{Y}_2} \mathbb{1}_{\{u_t(i)\neq v_t(i,j^*)\}} P_{Y_2|X}(y|v_t(i,j^*))A(u_t(i),v_t(i,j^*),y) \\
&= \sum_{t=1}^{n} \sum_{y\in\mathcal{Y}_2} \mathbb{1}_{\{u_t(i)\neq v_t(i,j^*)\}} A(u_t(i),v_t(i,j^*),y)\left(P_{Y_2|X}(y|v_t(i,j^*)) - P_{Y_2|X}(y|u_t(i))\right) \\
&= \sum_{t=1}^{n} \sum_{y\in\mathcal{Y}_2} \frac{\left(P_{Y_2|X}(y|v_t(i,j^*)) - P_{Y_2|X}(y|u_t(i))\right)^2}{P_{Y_2|X}(y|u_t(i))} \mathbb{1}_{\{u_t(i)\neq v_t(i,j^*)\}} \\
&\overset{(a)}{=} \omega(i,j^*)d,
\end{aligned} \tag{187}$$

where *(a)* follows from (58). It also follows that,

$$\begin{aligned}
\hat{\sigma}_i^2 &= \sum_{t=1}^{n} \hat{\sigma}_{it}^2 \\
&= \sum_{t=1}^{n} \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))\left(A(u_t(i),v_t(i,j^*),y)^2 - \hat{\mu}_{it}^2\right) \\
&\leqslant \sum_{t=1}^{n} \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))A(u_t(i),v_t(i,j^*),y)^2 \\
&= \sum_{t=1}^{n} \mathbb{1}_{\{u_t(i)=v_t(i,j)\}} \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|u_t(i))A(u_t(i),v_t(i,j^*),y)^2 \\
&+ \sum_{t=1}^{n} \mathbb{1}_{\{u_t(i)\neq v_t(i,j)\}} \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))A(u_t(i),v_t(i,j^*),y)^2 \\
&= \sum_{t=1}^{n} \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|u_t(i))A(u_t(i),v_t(i,j^*),y)^2 \\
&\quad + A(u_t(i),v_t(i,j^*),y)^2 \mathbb{1}_{\{u_t(i)\neq v_t(i,j)\}}(P_{Y_2|X}(y|v_t(i,j^*)) - P_{Y_2|X}(y|u_t(i))) \\
&\overset{(a)}{\leqslant} nd + \omega(i,j^*)d' \\
&\leqslant nd + \omega(i,j^*)\left|d'\right|,
\end{aligned} \tag{188}$$

where *(a)* follows from (58), with $d'$ in (168). In addition, $\hat{\sigma}_i^2$ also satisfies:

$$\begin{aligned}
\hat{\sigma}_i^2 &= \sum_{t=1}^{n} \hat{\sigma}_{it}^2 \\
&\geqslant nd + \omega(i,j^*)d'' - \sum_{t=1}^{n} \hat{\mu}_{it}^2 \\
&\geqslant n(d - \hat{\mu}_i^*) + \omega(i,j^*)d'',
\end{aligned} \tag{189}$$

with $\hat{\mu}_i^*$ in (167). Finally, it also holds that

$$\hat{\phi}_i = \sum_{t=1}^{n} \hat{\phi}_{it} \leqslant n\hat{\phi}_i^*. \tag{190}$$

This completes the proof. ∎

# J Proof of Proposition 2

From (72a), it follows that

$$\begin{aligned}
\hat{\Lambda}_1 &= \sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n} \frac{P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i))}{M\hat{M}} \Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i,j)|\hat{\boldsymbol{X}}=\hat{\boldsymbol{x}}\right] \\
&= \sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n}\sum_{\boldsymbol{y}\in\mathcal{Y}_1^n} \frac{P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i))}{M\hat{M}} P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}}) \mathbb{1}_{\{\boldsymbol{y}\in\hat{\mathcal{D}}_1^{\mathsf{c}}(i,j)\}} \\
&= \sum_{i=1}^{M}\sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n}\sum_{\boldsymbol{y}\in\mathcal{Y}_1^n} \frac{P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)) P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{M} \mathbb{1}_{\{\imath_1(\hat{\boldsymbol{x}};\boldsymbol{y}|\boldsymbol{u}(i))\leqslant n\eta\}} \\
&= \Pr\left[\imath_1(\hat{\boldsymbol{X}};\boldsymbol{Y}_1|\boldsymbol{u}(W)) \leqslant n\eta\right].
\end{aligned} \tag{191}$$

From (72a), it follows that

$$\begin{aligned}
\hat{\Lambda}_2 &= \sum_{i=1}^{M}\sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n} \frac{P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i))}{M} \Pr\left[\boldsymbol{Y}_2 \in \mathcal{D}_2^{\mathsf{c}}(i)|\hat{\boldsymbol{X}}=\hat{\boldsymbol{x}}\right] \\
&= \sum_{i=1}^{M}\sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n} \frac{P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i))}{M} P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}}) \mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2^{\mathsf{c}}(i)\}} \\
&= \sum_{i=1}^{M}\sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n} \frac{P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i))}{M} P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}}) \frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))} \mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2^{\mathsf{c}}(i)\}} \\
&= \sum_{i=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n} \frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}{M} \mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2^{\mathsf{c}}(i)\}} \sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)) \frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))} \\
&= \sum_{i=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n} \frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}{M} \mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2^{\mathsf{c}}(i)\}} \left(\sum_{\hat{x}_1\in\mathcal{X}} P_{\hat{X}|X}(\hat{x}_1|u_1(i)) \frac{P_{Y_2|X}(y_1|\hat{x}_1)}{P_{Y_2|X}(y_1|u_1(i))}\right) \\
&\quad \cdot \left(\sum_{\hat{x}_2\in\mathcal{X}} P_{\hat{X}|X}(\hat{x}_2|u_2(i)) \frac{P_{Y_2|X}(y_2|\hat{x}_2)}{P_{Y_2|X}(y_2|u_2(i))}\right) \cdot \ldots \cdot \left(\sum_{\hat{x}_n\in\mathcal{X}} P_{\hat{X}|X}(\hat{x}_n|u_n(i)) \frac{P_{Y_2|X}(y_n|\hat{x}_n)}{P_{Y_2|X}(y_n|u_n(i))}\right) \\
&= \sum_{i=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n} \frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}{M} \mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2^{\mathsf{c}}(i)\}} \prod_{t=1}^{n}\sum_{\hat{x}\in\mathcal{X}} P_{\hat{X}|X}(\hat{x}|u_t(i)) \frac{P_{Y_2|X}(y_t|\hat{x})}{P_{Y_2|X}(y_t|u_t(i))} \\
&= \sum_{i=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n} \frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}{M} \mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2^{\mathsf{c}}(i)\}} \prod_{t=1}^{n}\sum_{\hat{x}\in\mathcal{X}} \left((1-\theta)\mathbb{1}_{\{\hat{x}=u_t(i)\}} + \theta\tilde{P}_{\hat{X}|X}(\hat{x}|u_t(i))\right) \\
&\quad \cdot \frac{P_{Y_2|X}(y_t|\hat{x})}{P_{Y_2|X}(y_t|u_t(i))}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{M} \sum_{\boldsymbol{y}\in\mathcal{Y}_2^n} \frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}{M} \mathbb{1}_{\left\{\boldsymbol{y}\in\mathcal{D}_2^{\epsilon}(i)\right\}} \prod_{t=1}^{n} \left(1+\theta \frac{\tilde{R}_{Y_2|X}(y_t|u_t(i)) - P_{Y_2|X}(y_t|u_t(i))}{P_{Y_2|X}(y_t|u_t(i))}\right) \\
&\leqslant \sum_{i=1}^{M} \sum_{\boldsymbol{y}\in\mathcal{Y}_2^n} \frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}{M} \mathbb{1}_{\left\{\boldsymbol{y}\in\mathcal{D}_2^{\epsilon}(i)\right\}} \left(1+\max_{(x,y)\in\mathcal{X}\times\mathcal{Y}_2} \frac{\tilde{R}_{Y_2|X}(y|x) - P_{Y_2|X}(y|x)}{P_{Y_2|X}(y|x)}\theta\right)^n \\
&= \epsilon \left(1+\max_{(x,y)\in\mathcal{X}\times\mathcal{Y}_2} \frac{\tilde{R}_{Y_2|X}(y|x) - P_{Y_2|X}(y|x)}{P_{Y_2|X}(y|x)}\theta\right)^n . \qquad (192)
\end{aligned}
$$

The proof is completed by verifying that for all $k \in \{1,2\}$,

$$
\hat{\Lambda}_k \leqslant \max\left(\Pr\left[\imath_1\left(\hat{\boldsymbol{X}};\boldsymbol{Y}_1|\boldsymbol{u}(W)\right) \leqslant n\eta\right], \epsilon\left(1+\max_{(x,y)\in\mathcal{X}\times\mathcal{Y}_2} \frac{\tilde{R}_{Y_2|X}(y|x) - P_{Y_2|X}(y|x)}{P_{Y_2|X}(y|x)}\theta\right)^n\right). \qquad (193)
$$

■

# K Proof of Proposition 3

Let $\upsilon \in (0, \frac{1}{2})$ and $\xi > 0$ be two parameters whose exact values will be defined later. Assume that for all $(i,j) \in \mathcal{W}\times\hat{\mathcal{W}}$, the parameter $\eta$ in (70) is chosen such that

$$
\eta = \sup\left\{b\in\mathbb{R} : \Pr\left[\imath_1\left(\hat{\boldsymbol{X}};\boldsymbol{Y}_1|\boldsymbol{u}(W)\right) < n\left(b-\xi\right)\right] < \upsilon\right\}, \qquad (194)
$$

where the probability operator in (194) applies with respect to the distribution $P_{W\hat{\boldsymbol{X}}\boldsymbol{Y}_1}$ in (75). From Proposition 2, it follows that if $\upsilon$ is chosen such that

$$
\upsilon = \epsilon\left(1+\max_{(x,y)\in\mathcal{X}\times\mathcal{Y}_2} \frac{\tilde{R}_{Y_2|X}(y|x) - P_{Y_2|X}(y|x)}{P_{Y_2|X}(y|x)}\theta\right)^n, \qquad (195)
$$

where $\epsilon$ is the probability of error of the broadcast code $\mathcal{C}$ and $\theta$ is the parameter in (65), then, for all $k \in \{1,2\}$,

$$
\hat{\Lambda}_k \leqslant \upsilon. \qquad (196)
$$

The remainder of the proof consists in calculating the supremum on the right hand-side of (194) subject to (195). Consider the following positive constants

$$
\gamma_1 \triangleq \min_{(\hat{x},x,y)\in\mathcal{X}^2\times\mathcal{Y}_1} \imath_1\left(\hat{x},y|x\right) \qquad (197)
$$

$$
\gamma_2 \triangleq \max_{(\hat{x},x,y)\in\mathcal{X}^2\times\mathcal{Y}_1} \imath_1\left(\hat{x},y|x\right) \qquad (198)
$$

$$
\gamma \triangleq \gamma_2 - \gamma_1, \qquad (199)
$$

which depend only on the parameters of the random transformation in (2). Note that the random variables $\imath_1\left(\hat{X}_1, Y_{1,1}|u_1(W)\right)$, $\imath_1\left(\hat{X}_2, Y_{1,2}|u_2(W)\right)$, …, $\imath_1\left(\hat{X}_n, Y_{1,n}|u_n(W)\right)$ are mutually independent and bounded, *i.e.*, for all $t \in \{1,2,\ldots,n\}$,

$$
\gamma_1 \leqslant \imath_1\left(\hat{X}_n, Y_{1,n}|u_n(W)\right) \leqslant \gamma_2, \qquad (200)
$$

where $\gamma_1$ and $\gamma_2$ are defined in (197) and (198), respectively. Therefore, these random variables are also sub-Gaussian with sub-Gaussian parameter $\gamma$ in (199). Note also that the random variable

$$\frac{1}{n}\imath_1\left(\hat{\boldsymbol{X}};\boldsymbol{Y}_1|\boldsymbol{u}(W)\right)=\frac{1}{n}\sum_{t=1}^{n}\imath_1\left(\hat{x}_t,y_t|u_t(i)\right) \tag{201}$$

exhibits an expectation that satisfies:

$$\begin{aligned}
&\mathbb{E}_{W\hat{\boldsymbol{X}}\boldsymbol{Y}_1}\left[\frac{1}{n}\sum_{t=1}^{n}\imath_1\left(\hat{X}_t,Y_{1,t}|u_t(W)\right)\right]\\
&=\frac{1}{nM}\sum_{i=1}^{M}\sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n}\sum_{\boldsymbol{y}\in\mathcal{Y}_1^n}P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i))P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})\sum_{t=1}^{n}\imath_1\left(\hat{x}_t,y_t|u_t(i)\right)\\
&=\frac{1}{nM}\sum_{t=1}^{n}\sum_{i=1}^{M}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}P_{\hat{X}|X}(\hat{x}|u_t(i))P_{Y_1|X}(y|\hat{x})\imath_1\left(\hat{x},y|u_t(i)\right)\\
&=\frac{1}{nM}\sum_{t=1}^{n}\sum_{i=1}^{M}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\sum_{x\in\mathcal{X}}\mathbb{1}_{\{x=u_t(i)\}}P_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})\imath_1\left(\hat{x},y|x\right)\\
&=\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\bar{P}_X(x)P_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})\imath_1\left(\hat{x},y|x\right)\\
&=I(\hat{X};Y_1|X),
\end{aligned} \tag{202}$$

where $\bar{P}_X$ is defined in (17). Therefore, for all $b$ for which

$$b-\xi-I(\hat{X};Y_1|X)<0, \tag{203}$$

it holds from Hoeffding's inequality [11] that

$$\begin{aligned}
\Pr\left[\imath_1\left(\hat{\boldsymbol{X}};\boldsymbol{Y}_1|\boldsymbol{u}(W)\right)<n\left(b-\xi\right)\right]&=\Pr\left[\frac{1}{n}\sum_{t=1}^{n}\imath_1\left(\hat{X}_t,Y_{1,t}|u_t(W)\right)<b-\xi\right],\\
&<\exp\left(-\frac{n\left(I(\hat{X};Y_1|X)-b+\xi\right)^2}{2\gamma^2}\right).
\end{aligned} \tag{204}$$

Alternatively, for all $b$ for which

$$b-\xi-I(\hat{X};Y_1|X)\geqslant 0, \tag{205}$$

it follows that:

$$\begin{aligned}
\Pr\left[\imath_1\left(\hat{\boldsymbol{X}};\boldsymbol{Y}_1|\boldsymbol{u}(W)\right)<n\left(b-\xi\right)\right]&=\Pr\left[\frac{1}{n}\sum_{t=1}^{n}\imath_1\left(\hat{X}_t,Y_{1,t}|u_t(W)\right)<b-\xi\right]\\
&>1-\exp\left(-\frac{n\left(I(\hat{X};Y_1|X)-b+\xi\right)^2}{2\gamma^2}\right).
\end{aligned} \tag{206}$$

From (204), it holds that for all $b$ such that

$$b<I(\hat{X};Y_1|X)+\xi-\sqrt{-\frac{2\gamma^2\log_2\left(v\right)}{n}}, \tag{207}$$

then

$$\exp\left(-\frac{n\left(I(\hat{X};Y_1|X)-b+\xi\right)^2}{2\gamma^2}\right)<\upsilon, \tag{208}$$

which implies that $\Pr\left[\imath_1\left(\hat{\boldsymbol{X}};\boldsymbol{Y}_1|\boldsymbol{u}(W)\right)<n\left(b-\xi\right)\right]<\upsilon$. Alternatively, from (206), it holds that for all $b$, such that

$$b>I(\hat{X};Y_1|X)+\xi+\sqrt{-\frac{2\gamma^2\log_2\left(\upsilon\right)}{n}}, \tag{209}$$

then given that $0<\upsilon\leqslant\frac{1}{2}$, it follows that

$$1-\exp\left(-\frac{n\left(I(\hat{X};Y_1|X)-b+\xi\right)^2}{2\gamma^2}\right)>\upsilon, \tag{210}$$

which implies that $\Pr\left[\imath_1\left(\hat{\boldsymbol{X}};\boldsymbol{Y}_1|\boldsymbol{u}(W)\right)<n\left(b-\xi\right)\right]>\upsilon$. Hence, the supremum $\eta$ in (194) can be lower bounded as follows:

$$\begin{aligned}\eta>&\sup\left\{b\in\mathbb{R}_+:b<I(\hat{X};Y_1|X)+\xi-\sqrt{-\frac{2\gamma^2\log_2\left(\upsilon\right)}{n}}\right\},\\ =&I(\hat{X};Y_1|X)+\xi-\sqrt{-\frac{2\gamma^2\log_2\left(\upsilon\right)}{n}},\end{aligned} \tag{211}$$

where the inequality in (211) follows from the fact that there might exists a $b$ such that $\left|b-\xi-I(\hat{X};Y_1|X)\right|<\sqrt{-\frac{2\gamma^2\log_2(\upsilon)}{n}}$ for which $\Pr\left[\imath_1\left(\hat{\boldsymbol{X}};\boldsymbol{Y}_1|\boldsymbol{u}(W)\right)<n\left(b-\xi\right)\right]<\upsilon$. Assume now that $\xi$ is chosen such that

$$\xi=\sqrt{-\frac{2\gamma^2\log_2\left(\upsilon\right)}{n}}. \tag{212}$$

Hence, under the assumptions on $\upsilon$ and $\xi$ in (212), it follows from (211) that

$$\eta>I(\hat{X};Y_1|X). \tag{213}$$

Assume now that the number of codewords $\hat{M}$ is chosen such that:

$$\eta=\frac{\log_2\left(\hat{M}\right)}{n}. \tag{214}$$

Then, from (213), it holds that

$$\frac{\log_2\left(\hat{M}\right)}{n}>I(\hat{X};Y_1|X). \tag{215}$$

Finally, note that from the choice of $\eta$ in (194), for all $k\in\{1,2\}$, the probability of error $\hat{\Lambda}_k$ in (74) satisfies

$$\hat{\Lambda}_k<\upsilon. \tag{216}$$

This implies that there always exists an $(n,\mathcal{C},\hat{M},\hat{\epsilon})$-induced code whose number of codewords $\hat{M}$ satisfies (215) and its decoding error probability satisfies $\hat{\epsilon}\leqslant\upsilon$. The following lemma completes the proof.

**Lemma 12.** *The mutual information $I(\hat{X};Y_1|X)$ in (215) satisfies*

$$I(\hat{X};Y_1|X) \geqslant \sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}} \bar{P}_X(x)\Big(\theta\tilde{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big) - \theta^2\chi_2\left(\tilde{R}_{Y_1|X=x}, P_{Y_1|X=x}\right)\Big), \tag{217}$$

*where the probability distribution $\tilde{R}_{Y_1|X}$ is defined in (73).*

*Proof:* The proof of Proposition 12 is presented in Appendix L. ■

■

# L Proof of Lemma 12

The proof of Lemma 12 consists in expressing a lower bound for the mutual information in (215) in terms of the generating parameters $\theta$ and $\tilde{P}_{\hat{X}|X}$. Note that the mutual information in (215) is with respect to the probability distribution $\bar{P}_X P_{\hat{X}|X} P_{Y_1|X}$, with $\bar{P}_X$ defined in (17); $P_{\hat{X}|X}$ in (64) and $P_{Y_1|X}$ in (2b). That is,

$$\begin{aligned}
&I(\hat{X};Y_1|X)\\
&= \sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1} \bar{P}_X(x)P_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})\log_2\left(\frac{P_{Y_1|X}(y|\hat{x})}{\sum_{\hat{x}\in\mathcal{X}}P_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})}\right)\\
&= \sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}} \bar{P}_X(x)P_{\hat{X}|X}(\hat{x}|x)\Bigg(D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big)\\
&\quad + \sum_{y\in\mathcal{Y}_1} P_{Y_1|X}(y|\hat{x})\log_2\left(\frac{P_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)+\theta\left(\tilde{R}_{Y_1|X}(y|x)-P_{Y_1|X}(y|x)\right)}\right)\Bigg)\\
&= \sum_{x\in\mathcal{X}} \bar{P}_X(x)\Bigg(\theta\sum_{\hat{x}\in\mathcal{X}}\tilde{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big)\\
&\quad - \sum_{y\in\mathcal{Y}_1} P_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})\log_2\left(1+\theta\frac{\tilde{R}_{Y_1|X}(y|x)-P_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)}\right)\Bigg)\\
&\overset{(a)}{\geqslant} \sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}} \bar{P}_X(x)\Bigg(\theta\tilde{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big)\\
&\quad - \sum_{y\in\mathcal{Y}_1} P_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})\theta\frac{\tilde{R}_{Y_1|X}(y|x)-P_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)}\Bigg)\\
&= \sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}} \bar{P}_X(x)\Bigg(\theta\tilde{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big)\\
&\quad - \sum_{y\in\mathcal{Y}_1}\Big((1-\theta)P_{Y_1|X}(y|x)+\theta\tilde{P}_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})\Big)\theta\frac{\tilde{R}_{Y_1|X}(y|x)-P_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)}\Bigg)\\
&= \sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}} \bar{P}_X(x)\Bigg(\theta\tilde{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big) - \sum_{y\in\mathcal{Y}_1}\theta^2\frac{\left(\tilde{R}_{Y_1|X}(y|x)-P_{Y_1|X}(y|x)\right)^2}{P_{Y_1|X}(y|x)}\Bigg)\\
&= \sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}} \bar{P}_X(x)\Big(\theta\tilde{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big) - \theta^2\chi_2\left(\tilde{R}_{Y_1|X=x}, P_{Y_1|X=x}\right)\Big)
\end{aligned} \tag{218}$$

where *(a)* follows from the inequality $\log_2(1+x) \leqslant x$ for all $x > -1$. This completes the proof. ■

# M Proof of Proposition 4

Let $S_{W\boldsymbol{Y}_2}$ be a distribution such that, for all $(i, \boldsymbol{y}) \in \mathcal{W} \times \mathcal{Y}_2^n$,

$$S_{W\boldsymbol{Y}_2}(i,\boldsymbol{y}) \triangleq \frac{1}{M} S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i), \tag{219}$$

with

$$\begin{aligned} S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i) &\triangleq \sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)) P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}}) \\ &= \sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n} \prod_{t=1}^{n} P_{\hat{X}|X}(\hat{x}_t|u_t(i)) P_{Y_2|X}(y_t|\hat{x}_t) \\ &= \prod_{t=1}^{n} \sum_{\hat{x}\in\mathcal{X}} P_{\hat{X}|X}(\hat{x}|u_t(i)) P_{Y_2|X}(y_t|\hat{x}). \end{aligned} \tag{220}$$

Consider the distribution $Q_{W\boldsymbol{Y}_2}$ in (28). From Lemma 7 (in Appendix A), it follows that the total variation $\|S_{W\boldsymbol{Y}_2} - Q_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}}$ verifies

$$\begin{aligned} &\|S_{W\boldsymbol{Y}_2} - Q_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} \\ &\quad = \Pr\left[S_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2S}) \geqslant Q_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2S})\right] - \Pr\left[S_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2Q}) \geqslant Q_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2Q})\right], \end{aligned} \tag{221}$$

where the first probability operator in the left hand-side of (221) applies assuming that $(W, \boldsymbol{Y}_{2S})$ follows the joint distribution $S_{W\boldsymbol{Y}_2}$; and the second applies assuming that $(W, \boldsymbol{Y}_{2Q})$ follows the joint distribution $Q_{W\boldsymbol{Y}_2}$.

For all $(x, y) \in \mathcal{X} \times \mathcal{Y}_2$ let $B : \mathcal{X} \times \mathcal{Y}_2 \to \mathbb{R}$ be

$$B(x,y) \triangleq \log_2\left(1 + \theta C(x,y)\right), \tag{222}$$

where

$$C(x,y) \triangleq \frac{\tilde{R}_{Y_2|X}(y|x) - P_{Y_2|X}(y|x)}{P_{Y_2|X}(y|x)}. \tag{223}$$

Then, note that

$$\begin{aligned}
&\Pr\left[S_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2S}) \geqslant Q_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2S})\right] \\
&= \Pr\left[\frac{S_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2S})}{Q_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2S})} \geqslant 1\right] \\
&= \Pr\left[\log_2\left(\frac{S_{\boldsymbol{Y}_2|W}(\boldsymbol{Y}_{2S}|W)}{Q_{\boldsymbol{Y}_2|W}(\boldsymbol{Y}_{2S}|W)}\right) \geqslant 0\right] \\
&= \Pr\left[\log_2\left(\frac{\sum\limits_{\hat{\boldsymbol{x}}\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(W))P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{Y}_{2S}|\hat{\boldsymbol{x}})}{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{Y}_{2S}|\boldsymbol{u}(W))}\right) \geqslant 0\right] \\
&= \Pr\left[\sum_{t=1}^{n}\log_2\left(\frac{\sum\limits_{\hat{x}\in\mathcal{X}} P_{\hat{X}|X}(\hat{x}|u_t(W))P_{Y_2|X}(Y_{2S,t}|\hat{x})}{P_{Y_2|X}(Y_{2S,t}|u_t(W))}\right) \geqslant 0\right] \\
&= \Pr\left[\sum_{t=1}^{n}\log\left(\frac{(1-\theta)P_{Y_2|X}(Y_{2S,t}|u_t(W))}{P_{Y_2|X}(Y_{2S,t}|u_t(W))} + \frac{\theta\sum_{\hat{x}\in\mathcal{X}}\tilde{P}_{\hat{X}|X}(\hat{x}|u_t(W))P_{Y_2|X}(Y_{2S,t}|\hat{x})}{P_{Y_2|X}(Y_{2S,t}|u_t(W))}\right) \geqslant 0\right] \\
&= \Pr\left[\sum_{t=1}^{n}\log\left(1 + \frac{\theta\left(\tilde{R}_{Y_2|X}(Y_{2S,t}|u_t(W)) - P_{Y_2|X}(Y_{2S,t}|u_t(W))\right)}{P_{Y_2|X}(Y_{2S,t}|u_t(W))}\right) \geqslant 0\right] \\
&= \Pr\left[\sum_{t=1}^{n}\log_2\left(1+\theta C(u_t(W),Y_{2S,t})\right) \geqslant 0\right] \\
&= \Pr\left[\sum_{t=1}^{n} B(u_t(W),Y_{2S,t}) \geqslant 0\right]. \qquad (224)
\end{aligned}$$

Following similar steps, it can be shown that

$$\Pr\left[S_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2Q}) \geqslant Q_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2Q})\right] = \Pr\left[\sum_{t=1}^{n} B(u_t(W),Y_{2Q,t}) \geqslant 0\right]. \qquad (225)$$

Plugging (224) and (225) into (221) yields

$$\|S_{W\boldsymbol{Y}_2} - Q_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} = \Pr\left[\sum_{t=1}^{n} B(u_t(W),Y_{2S,t}) \geqslant 0\right] - \Pr\left[\sum_{t=1}^{n} B(u_t(W),Y_{2Q,t}) \geqslant 0\right]. \qquad (226)$$

The remainder of the proof consists in obtaining a lower-bound and an upper-bound on the first and second terms in the right hand-side of (226), respectively.
Consider the first term in the right hand-side of (226). For all $t \in \{1,2,\ldots,n\}$, let $\hat{\mu}_t$, $\hat{\sigma}_t^2$ and $\hat{\phi}_t$ be the first moment, second moment and third absolute moment of the random variable

$$B(u_t(W),Y_{2S,t}) = \log_2\left(1+\theta C(u_t(W),Y_{2S,t})\right). \qquad (227)$$

That is,

$$\hat{\mu}_t = \mathbb{E}_{WY_{2S,t}}\left[B(u_t(W),Y_{2S,t})\right], \qquad (228)$$

$$\hat{\sigma}_t^2 = \mathbb{E}_{WY_{2S,t}}\left[B(u_t(W),Y_{2S,t})^2\right] - \hat{\mu}_t^2, \quad \text{and} \qquad (229)$$

$$\hat{\phi}_t = \mathbb{E}_{WY_{2S,t}}\left[\left|B(u_t(W),Y_{2S,t}) - \hat{\mu}_t\right|^3\right]. \qquad (230)$$

Using this notation, let $\hat{\mu}$, $\hat{\sigma}^2$ and $\hat{\phi}$ be

$$\hat{\mu} \triangleq \sum_{t=1}^{n} \hat{\mu}_t, \quad \hat{\sigma}^2 \triangleq \sum_{t=1}^{n} \hat{\sigma}_t^2, \text{ and } \hat{\phi} \triangleq \sum_{t=1}^{n} \hat{\phi}_t. \tag{231}$$

The following lemma characterizes $\hat{\mu}$, $\hat{\sigma}^2$ and $\hat{\phi}$.

**Lemma 13.** *The terms $\hat{\mu}$, $\hat{\sigma}^2$ and $\hat{\phi}$ in* (231) *satisfy*

$$\hat{\mu} \geqslant \frac{K^2}{2} \sum_{x \in \mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}} |c_1|, \tag{232}$$

$$\hat{\sigma}^2 \leqslant K^2 \sum_{x \in \mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}} |c_2|, \tag{233}$$

$$\hat{\sigma}^2 \geqslant K^2 \sum_{x \in \mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}} |c_3|, \tag{234}$$

$$\hat{\phi} \leqslant \frac{K^3}{\sqrt{n}} c_4. \tag{235}$$

*where $c_1, c_2, c_3$ and $c_4$ are constants that depend only on the random transformation in* (2).

*Proof:* The proof of Lemma 13 is presented in Appendix N. ∎

From Lemma 13, it follows that

$$\begin{aligned}
&\Pr\left[\sum_{t=1}^{n} B(u_t(W), Y_{2S,t}) \geqslant 0\right] \\
&= \Pr\left[\sum_{t=1}^{n} B(u_t(W), Y_{2S,t}) - \hat{\mu} \geqslant -\hat{\sigma}\frac{\hat{\mu}}{\hat{\sigma}}\right] \\
&\overset{(a)}{\geqslant} Q\left(-\frac{\hat{\mu}}{\hat{\sigma}}\right) - c_0 \frac{\hat{\phi}}{\hat{\sigma}^3} \\
&\overset{(b)}{\geqslant} Q\left(\frac{-\frac{K^2}{2}\sum_{x \in \mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_1|}{\sqrt{K^2 \sum_{x \in \mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_2|}}\right) \\
&\quad - \frac{\frac{K^3}{\sqrt{n}} c_0 c_4}{\left(K^2 \sum_{x \in \mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_3|\right)^{\frac{3}{2}}} \\
&\overset{(c)}{\geqslant} 1 - \frac{c_{14}}{\sqrt{n}} - Q\left(\frac{\frac{K^2}{2}\sum_{x \in \mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_1|}{\sqrt{K^2 \sum_{x \in \mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_2|}}\right)
\end{aligned} \tag{236}$$

where *(a)* follows from the Berry-Esseen Theorem (Theorem 3); *(b)* follows from Lemma 13; and *(c)* follows with

$$c_{14} \triangleq \max c_0 c_4 \left( \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2\Big(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}\Big) - \frac{K}{\sqrt{n}}\,|c_3| \right)^{-\frac{3}{2}}, \tag{237}$$

where the maximization is over all possible conditional distributions $\tilde{P}_{\hat{X}|X}$ and $n \in \mathbb{N}$ subject to

$$\sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2\Big(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}\Big) - \frac{K}{\sqrt{n}}\,|c_3| > 0. \tag{238}$$

Note that $c_{14}$ depends only on the random transformation in (2). Consider the second term in the right hand-side of (226). For all $t \in \{1, 2, \ldots, n\}$, let $\mu_t$, $\sigma_t^2$ and $\phi_t$ be the first, second and third absolute moments of the random variable

$$B(u_t(W), Y_{2Q,t}) = \log_2\left(1 + \theta C(u_t(W), Y_{2Q,t})\right). \tag{239}$$

That is,

$$\mu_t \triangleq \mathbb{E}_{WY_{2Q,t}}\left[B(u_t(W), Y_{2Q,t})\right], \tag{240}$$

$$\sigma_t^2 \triangleq \mathbb{E}_{WY_{2Q,t}}\left[B(u_t(W), Y_{2Q,t})^2\right] - \mu_t^2, \quad \text{and} \tag{241}$$

$$\phi_t \triangleq \mathbb{E}_{WY_{2Q,t}}\left[\left|B(u_t(W), Y_{2Q,t}) - \mu_t\right|^3\right]. \tag{242}$$

Using this notation, let $\mu$, $\sigma^2$ and $\phi$ be

$$\mu \triangleq \sum_{t=1}^{n} \mu_t, \quad \sigma^2 \triangleq \sum_{t=1}^{n} \sigma_t^2, \text{ and } \phi \triangleq \sum_{t=1}^{n} \phi_t. \tag{243}$$

The following lemma characterizes $\mu$, $\sigma^2$ and $\phi$.

**Lemma 14.** *The terms $\mu$, $\sigma^2$ and $\phi$ in* (243) *satisfy*

$$\mu \leqslant \frac{-K^2}{2} \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}\,|c_5|, \tag{244}$$

$$\sigma^2 \leqslant K^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}\,|c_6|, \tag{245}$$

$$\sigma^2 \geqslant K^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}\,|c_7|, \tag{246}$$

$$\phi \leqslant \frac{K^3}{\sqrt{n}} c_8, \tag{247}$$

*where $c_5, c_6, c_7$ and $c_8$ are constants that depend only on the random transformation in* (2).

*Proof:* The proof of Lemma 14 is presented in Appendix O. ∎

From Lemma 14, it follows that

$$\begin{aligned}
&\Pr\left[\sum_{t=1}^{n} B(u_t(W), Y_{2Q,t}) \geqslant 0\right] \\
&= \Pr\left[\sum_{t=1}^{n} B(u_t(W), Y_{2Q,t}) - \mu \geqslant -\sigma\frac{\mu}{\sigma}\right] \\
&\overset{(a)}{\leqslant} Q\left(\frac{-\mu}{\sigma}\right) + c_0\frac{\phi}{\sigma^3} \\
&\overset{(b)}{\leqslant} Q\left(\frac{\frac{K^2}{2}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_5|}{\sqrt{K^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_7|}}\right) \\
&+ \frac{\frac{K^3}{\sqrt{n}}c_0c_8}{\sqrt{K^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_7|}^3} \\
&\overset{(c)}{\leqslant} Q\left(\frac{\frac{K^2}{2}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_5|}{\sqrt{K^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_7|}}\right) + \frac{c_{10}}{\sqrt{n}},
\end{aligned} \tag{248}$$

where *(a)* follows from the Berry-Esseen Theorem (Theorem 3); and *(b)* follows from Lemma 14; and *(c)* follows with

$$c_{10} \triangleq \max c_0c_8\left(\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2\big(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}\big) - \frac{K}{\sqrt{n}}|c_7|\right)^{-\frac{3}{2}}, \tag{249}$$

where the maximization is over all possible conditional distributions $\tilde{P}_{\hat{X}|X}$ and $n \in \mathbb{N}$ subject to

$$\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2\big(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}\big) - \frac{K}{\sqrt{n}}|c_7| > 0. \tag{250}$$

Note that $c_{10}$ depends only on the random transformation in (2).

Combining (226), (248), and (236) yields

$$\begin{aligned}\|S_{W\boldsymbol{Y}_2} - Q_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} &\geqslant 1 - \frac{c_{10}+c_{14}}{\sqrt{n}} - Q\left(\frac{\frac{K^2}{2}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_1|}{\sqrt{K^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_2|}}\right)\\ &\quad -Q\left(\frac{\frac{K^2}{2}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_5|}{\sqrt{K^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_7|}}\right)\\ &\geqslant 1 - \frac{c_{10}+c_{14}}{\sqrt{n}} - 2Q\left(\frac{\frac{K^2}{2}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_1|}{\sqrt{K^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_2|}}\right).\end{aligned} \tag{251}$$

From Definition 8, it follows that the total variation $\|Q_{\boldsymbol{Y}_2} - S_{\boldsymbol{Y}_2}\|_{\mathrm{TV}}$ verifies

$$\begin{aligned}\delta &\geqslant \|Q_{\boldsymbol{Y}_2} - S_{\boldsymbol{Y}_2}\|_{\mathrm{TV}}\\ &\overset{(a)}{\geqslant} \|S_{W\boldsymbol{Y}_2} - Q_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} - \epsilon - \max\{\hat{\Lambda}_1, \hat{\Lambda}_2\}\\ &\overset{(b)}{\geqslant} 1 - \frac{c_{10}+c_{14}}{\sqrt{n}} - \epsilon - \max\{\hat{\Lambda}_1, \hat{\Lambda}_2\} - 2Q\left(\frac{\frac{K^2}{2}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_1|}{\sqrt{K^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_2|}}\right)\\ &\geqslant 1 - \frac{c_{10}+c_{14}}{\sqrt{n}} - \epsilon - \max\{\hat{\Lambda}_1, \hat{\Lambda}_2\} - \frac{\frac{K^3}{\sqrt{n}}(4|c_1| + |c_2|)}{4\sqrt{2\pi n\left(\sum_{x\in\mathcal{X}}K^2\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_2|\right)}}\\ &\quad -2Q\left(\frac{K}{2}\sqrt{\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2\left(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}\right)}\right),\\ &= 1 - \frac{c_{10}+c_{14}}{\sqrt{n}} - \epsilon - \max\{\hat{\Lambda}_1, \hat{\Lambda}_2\} - \frac{K^2(4|c_1| + |c_2|)}{4n\sqrt{2\pi\left(\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K}{\sqrt{n}}|c_2|\right)}}\\ &\quad -2Q\left(\frac{K}{2}\sqrt{\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2\left(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}\right)}\right),\\ &\overset{(c)}{\geqslant} 1 - \frac{c_{10}+c_{14}}{\sqrt{n}} - \frac{c_{15}}{n} - \epsilon - \max\{\hat{\Lambda}_1, \hat{\Lambda}_2\} - 2Q\left(\frac{K}{2}\sqrt{\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2\left(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}\right)}\right),\\ &\overset{(d)}{\geqslant} 1 - \frac{c}{\sqrt{n}} - \epsilon - \max\{\hat{\Lambda}_1, \hat{\Lambda}_2\} - 2Q\left(\frac{K}{2}\sqrt{\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2\left(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}\right)}\right),\end{aligned} \tag{252}$$

where *(a)* is due to Lemma 4; *(b)* is due to (251); *(c)* follows from Lemma 9, with

$$c_{15} \triangleq \frac{K^2\left(4\left|c_1\right|+\left|c_2\right|\right)}{4\sqrt{2\pi\left(\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})+\frac{K}{\sqrt{n}}\left|c_2\right|\right)}}; \tag{253}$$

and *(d)* follows with

$$c = c_{10}+c_{14}+c_{15}. \tag{254}$$

From (252), it follows that

$$\frac{K}{2}\sqrt{\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2\left(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x}\right)} \leqslant Q^{-1}\left(\frac{1-\delta-\epsilon-\max\{\hat{\Lambda}_1,\hat{\Lambda}_2\}-\frac{c}{\sqrt{n}}}{2}\right), \tag{255}$$

that is,

$$K \leqslant \frac{2Q^{-1}\left(\frac{1-\delta-\epsilon-\max\{\hat{\Lambda}_1,\hat{\Lambda}_2\}-\frac{c}{\sqrt{n}}}{2}\right)}{\sqrt{\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})}}. \tag{256}$$

This completes the proof. ■

# N Proof of Lemma 13

Note that for all $t\in\{1,2,\ldots,n\}$, it holds that

$$\hat{\mu}_t = \frac{1}{M}\sum_{i=1}^{M}\sum_{y\in\mathcal{Y}_2}\sum_{\hat{x}\in\mathcal{X}}P_{\hat{X}|X}(\hat{x}|u_t(i))P_{Y_2|X}(y|\hat{x})B(u_t(i),y), \tag{257}$$

$$\hat{\sigma}_t^2 = \frac{1}{M}\sum_{i=1}^{M}\sum_{y\in\mathcal{Y}_2}\sum_{\hat{x}\in\mathcal{X}}P_{\hat{X}|X}(\hat{x}|u_t(i))P_{Y_2|X}(y|\hat{x})B(u_t(i),y)^2-\hat{\mu}_t^2, \tag{258}$$

and

$$\hat{\phi}_t = \frac{1}{M}\sum_{i=1}^{M}\sum_{y\in\mathcal{Y}_2}\sum_{\hat{x}\in\mathcal{X}}P_{\hat{X}|X}(\hat{x}|u_t(i))P_{Y_2|X}(y|\hat{x})\left|B(u_t(i),y)-\hat{\mu}_t\right|^3. \tag{259}$$

Thus, it follows that

$$
\begin{aligned}
\hat{\mu} &= \sum_{t=1}^{n} \hat{\mu}_t \\
&= \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{y \in \mathcal{Y}_2} \sum_{\hat{x} \in \mathcal{X}} P_{\hat{X}|X}(\hat{x}|u_t(i)) P_{Y_2|X}(y|\hat{x}) B(u_t(i), y) \\
&= \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{y \in \mathcal{Y}_2} \sum_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \mathbb{1}_{\{x = u_t(i)\}} P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) B(x, y) \\
&= n \sum_{x \in \mathcal{X}} \sum_{\hat{x} \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \bar{P}_X(x) P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \log_2 \left(1 + \theta C(x, y)\right) \\
&= n \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \bar{P}_X(x) \Big( (1-\theta) P_{Y_2|X}(y|x) \log_2 \left(1 + \theta C(x, y)\right) \\
&\quad + \theta \sum_{\hat{x} \in \mathcal{X}} \tilde{P}_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \log_2 \left(1 + \theta C(x, y)\right) \Big) \\
&= n \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \log_2 \left(1 + \theta C(x, y)\right) \\
&\quad + \theta (\tilde{R}_{Y_2|X}(y|x) - P_{Y_2|X}(y|x)) \log_2 \left(1 + \theta C(x, y)\right) \Big) \\
&= n \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \sum_{k=1}^{\infty} \frac{(-1)^{k+1} \theta^k}{k} C(x, y)^k \\
&\quad + \theta (\tilde{R}_{Y_2|X}(y|x) - P_{Y_2|X}(y|x)) \sum_{k=1}^{\infty} \frac{(-1)^{k+1} \theta^k}{k} C(x, y)^k \\
&= n \sum_{x \in \mathcal{X}} \bar{P}_X(x) \sum_{k=2}^{\infty} \frac{(-1)^{k+1} \theta^k}{k} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \\
&\quad + n \sum_{x \in \mathcal{X}} \bar{P}_X(x) \sum_{k=1}^{\infty} \frac{(-1)^{k+1} \theta^{k+1}}{k} \chi_{k+1}(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \\
&= n \sum_{x \in \mathcal{X}} \bar{P}_X(x) \sum_{k=2}^{\infty} \frac{(-1)^{k+1} \theta^k}{k} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \\
&\quad + n \sum_{x \in \mathcal{X}} \bar{P}_X(x) \sum_{k=2}^{\infty} \frac{(-1)^{k} \theta^k}{k-1} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \\
&= n \sum_{x \in \mathcal{X}} \bar{P}_X(x) \sum_{k=2}^{\infty} \frac{(-1)^{k} \theta^k}{k(k-1)} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \\
&= n \frac{\theta^2}{2} \sum_{x \in \mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 \sum_{x \in \mathcal{X}} \bar{P}_X(x) \sum_{k=3}^{\infty} \frac{(-1)^{k} \theta^{k-3}}{k(k-1)} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}).
\end{aligned}
\tag{260}
$$

Note that since the random variable $B(u_t(W), Y_{2S,t})$ is bounded, its expectation is finite. Thus,

it follows that the second term in (260) is also finite. Let $c_1$ be defined as

$$c_1 \triangleq \min \sum_{x\in\mathcal{X}} \bar{P}_X(x) \sum_{k=3}^{\infty} \frac{(-1)^k \theta^{k-3}}{k(k-1)} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}), \tag{261}$$

where the minimization is over all possible values of $\theta \in (0,1)$ and all possible conditional distributions $\tilde{P}_{\hat{X}|X}$. Note that $c_1$ depends only on the parameters of the channel. It follows that

$$\begin{aligned}
\hat{\mu} &\geqslant n\frac{\theta^2}{2} \sum_{x\in\mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 c_1 \\
&\geqslant n\frac{\theta^2}{2} \sum_{x\in\mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - n\theta^3 \left|c_1\right| \\
&= \frac{K^2}{2} \sum_{x\in\mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}} \left|c_1\right|.
\end{aligned} \tag{262}$$

Similarly, it holds that

$$\begin{aligned}
\hat{\sigma}^2 &= \sum_{t=1}^{n} \hat{\sigma}_t^2 \\
&= \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x\in\mathcal{X}} \sum_{\hat{x}\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \mathbb{1}_{\{x=u_t(i)\}} P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) B(x,y)^2 - \hat{\mu}_t^2 \\
&\leqslant n \sum_{x\in\mathcal{X}} \sum_{\hat{x}\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) B(x,y)^2 \\
&= n \sum_{x\in\mathcal{X}} \sum_{\hat{x}\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \log_2\left(1+\theta C(x,y)\right)^2 \\
&= n \sum_{x\in\mathcal{X}} \sum_{\hat{x}\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \left(\sum_{k=1}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} C(x,y)^k\right)^2 \\
&= n \sum_{x\in\mathcal{X}} \sum_{\hat{x}\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \Bigg(\theta C(x,y) + 2\theta C(x,y) \sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} C(x,y)^k \\
&\quad + \left(\sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} C(x,y)^k\right)^2\Bigg) \\
&= n \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) \left((1-\theta) P_{Y_2|X}(y|x) + \theta \tilde{R}_{Y_2|X}(y|x)\right) \theta^2 C(x,y)^2 + n \sum_{x\in\mathcal{X}} \sum_{\hat{x}\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) \\
&\quad \cdot P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \left(2\sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^{k+1}}{k} C(x,y)^{k+1} + \left(\sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} C(x,y)^k\right)^2\right)
\end{aligned}$$

$$
\begin{aligned}
&= n\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)\Big(P_{Y_2|X}(y|x)+\theta\big(\tilde{R}_{Y_2|X}(y|x)-P_{Y_2|X}(y|x)\big)\Big)\theta^2C(x,y)^2\\
&\quad+n\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})\Bigg(2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k+1}}{k}C(x,y)^{k+1}\\
&\quad+\left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k}}{k}C(x,y)^{k}\right)^2\Bigg)\\
&= n\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)\Big(\theta^2\chi_2(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})+\theta^3\chi_3(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})\Big)+n\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)\\
&\quad\cdot P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})\Bigg(2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k+1}}{k}C(x,y)^{k+1}+\left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k}}{k}C(x,y)^{k}\right)^2\Bigg)\\
&= n\theta^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})+n\theta^3\sum_{x\in\mathcal{X}}\bar{P}_X(x)\Bigg(\chi_3(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})\\
&+\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})\Bigg(2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-2}}{k}C(x,y)^{k+1}+\left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k}C(x,y)^{k}\right)^2\Bigg)\Bigg).
\end{aligned}
\tag{263}
$$

Note that since $B(x,y)$ is bounded, the upper-bound in (263) is finite. Thus, the terms in (263) are also finite. Let $c_2$ be defined by

$$
\begin{aligned}
c_2\triangleq\max\sum_{x\in\mathcal{X}}\bar{P}_X(x)\Bigg(&\chi_3(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})+\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})\\
&\cdot\Bigg(2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-2}}{k}C(x,y)^{k+1}+\left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k}C(x,y)^{k}\right)^2\Bigg)\Bigg),
\end{aligned}
\tag{264}
$$

where the maximization is over all values of $\theta\in(0,1)$ and all possible conditional distributions $\tilde{P}_{\hat{X}|X}$. Using this notation, it follows that

$$
\begin{aligned}
\hat{\sigma}^2&\leqslant n\theta^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})+n\theta^3c_2\\
&\leqslant n\theta^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})+n\theta^3\left|c_2\right|\\
&=K^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})+\frac{K^3}{\sqrt{n}}\left|c_2\right|.
\end{aligned}
\tag{265}
$$

It also holds that

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{M}\sum_{t=1}^{n}\sum_{i=1}^{M}\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\mathbb{1}_{\{x=u_t(i)\}}P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})B(x,y)^2-\hat{\mu}_t^2\\
&= n\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})B(x,y)^2-\sum_{t=1}^{n}\hat{\mu}_t^2\\
&= n\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})B(x,y)^2\\
&\quad-\sum_{t=1}^{n}\Big(\frac{1}{M}\sum_{i=1}^{M}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}P_{\hat{X}|X}(\hat{x}|u_t(i))P_{Y_2|X}(y|\hat{x})B(u_t(i),y)\Big)^2\\
&\geqslant n\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})B(x,y)^2\\
&\quad-\sum_{t=1}^{n}\frac{1}{M}\sum_{i=1}^{M}\Big(\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}P_{\hat{X}|X}(\hat{x}|u_t(i))P_{Y_2|X}(y|\hat{x})B(u_t(i),y)\Big)^2\\
&= n\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})B(x,y)^2\\
&\quad-n\sum_{x\in\mathcal{X}}\bar{P}_X(x)\left(\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})B(x,y)\right)^2\\
&= n\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})B(x,y)^2-n\sum_{x\in\mathcal{X}}\bar{P}_X(x)\\
&\quad\cdot\left(\Big(P_{Y_X|X}(y|x)+\theta\Big(\tilde{R}_{Y_2|X}(y|x)-P_{Y_2|X}(y|x)\Big)\Big)\log_2\left(1+\theta C(x,y)\right)\right)^2\\
&= n\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})B(x,y)^2-n\sum_{x\in\mathcal{X}}\bar{P}_X(x)\\
&\quad\cdot\left(\Big(P_{Y_2|X}(y|x)+\theta\Big(\tilde{R}_{Y_2|X}(y|x)-P_{Y_2|X}(y|x)\Big)\Big)\sum_{k=1}^{\infty}\frac{(-1)^{k+1}\theta^k}{k}C(x,y)^k\right)^2\\
&= n\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})B(x,y)^2\\
&\quad-n\sum_{x\in\mathcal{X}}\bar{P}_X(x)\left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^k}{k}\chi_k(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})\right.\\
&\quad\left.+\sum_{k=1}^{\infty}\frac{(-1)^{k+1}\theta^{k+1}}{k}\chi_{k+1}(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})\right)^2\\
&= n\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})B(x,y)^2\\
&\quad-n\sum_{x\in\mathcal{X}}\bar{P}_X(x)\left(\sum_{k=2}^{\infty}\frac{(-1)^{k}\theta^{k}}{k(k-1)}\chi_k(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})\right)^2
\end{aligned}
$$

$$
\begin{aligned}
&= n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\Bigg(\chi_3(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \\
&\quad + \sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2} P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})\Bigg(2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-2}}{k}C(x,y)^{k+1} \\
&\quad + \left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k}C(x,y)^{k}\right)^2\Bigg)\Bigg) - n\sum_{x\in\mathcal{X}}\bar{P}_X(x)\left(\sum_{k=2}^{\infty}\frac{(-1)^{k}\theta^{k}}{k(k-1)}\chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x})\right)^2 \\
&= n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\Bigg(\chi_3(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \\
&\quad + \sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2} P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})\Bigg(2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-2}}{k}C(x,y)^{k+1} + \left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k}C(x,y)^{k}\right)^2 \\
&\quad - \Big(\sum_{k=2}^{\infty}\frac{(-1)^{k}\theta^{k-\frac{3}{2}}}{k(k-1)}\chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x})\Big)^2\Bigg)\Bigg). \qquad (266)
\end{aligned}
$$

Note that since $B(x,y)$ is bounded, $\hat{\sigma}^2$ is finite. Thus all the terms in (266) are also finite. Let $c_3$ be defined by

$$
\begin{aligned}
c_3 = \min \sum_{x\in\mathcal{X}} \bar{P}_X(x)\Bigg(&\chi_3(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2} P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x}) \\
&\cdot\Bigg(2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-2}}{k}C(x,y)^{k+1} + \left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k}C(x,y)^{k}\right)^2 \\
&- \Big(\sum_{k=2}^{\infty}\frac{(-1)^{k}\theta^{k-\frac{3}{2}}}{k(k-1)}\chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x})\Big)^2\Bigg)\Bigg), \qquad (267)
\end{aligned}
$$

where the minimization is over all possible values of $\theta \in (0,1)$ and all possible conditional distributions $\tilde{P}_{\hat{X}|X}$. Plugging (267) into (266) yields

$$
\begin{aligned}
\hat{\sigma}^2 &\geqslant n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 c_3 \\
&\geqslant n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - n\theta^3 \left|c_3\right| \\
&= K^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}\left|c_3\right|. \qquad (268)
\end{aligned}
$$

Finally,

$$\begin{aligned}
\hat{\phi} &= \sum_{t=1}^{n} \hat{\phi}_t \\
&= \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x \in \mathcal{X}} \sum_{\hat{x} \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \mathbb{1}_{\{x = u_t(i)\}} P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \left| \log_2 \left(1 + \theta C(x,y)\right) - \hat{\mu}_t \right|^3 \\
&= \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x \in \mathcal{X}} \sum_{\hat{x} \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \mathbb{1}_{\{x = u_t(i)\}} P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \Bigg| \log_2 \left(1 + \theta C(x,y)\right) \\
&\quad - \frac{1}{M} \sum_{i'=1}^{M} \sum_{\hat{x}' \in \mathcal{X}} \sum_{y' \in \mathcal{Y}_2} P_{\hat{X}|X}(\hat{x}'|u_t(i')) P_{Y_2|X}(y'|\hat{x}') \log_2 \left(1 + \theta C(u_t(i'), y')\right) \Bigg|^3 \\
&\leqslant \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x \in \mathcal{X}} \sum_{\hat{x} \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \mathbb{1}_{\{x = u_t(i)\}} P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \Bigg| \log_2 \left(1 + \theta C(x,y)\right) \\
&\quad - \log \left( \frac{1}{M} \sum_{i'=1}^{M} \sum_{\hat{x}' \in \mathcal{X}} \sum_{y' \in \mathcal{Y}_2} P_{\hat{X}|X}(\hat{x}'|u_t(i')) P_{Y_2|X}(y'|\hat{x}') \left(1 + \theta C(u_t(i'), y')\right) \right) \Bigg|^3 \\
&= \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x \in \mathcal{X}} \sum_{\hat{x} \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \mathbb{1}_{\{x = u_t(i)\}} P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \Bigg| \log_2 \left(1 + \theta C(x,y)\right) \\
&\quad - \log \left( 1 + \frac{\theta^2}{M} \sum_{i'=1}^{M} \chi_2(\tilde{R}_{Y_2|X=u_t(i')}, P_{Y_2|X=u_t(i')}) \right) \Bigg|^3 \\
&\leqslant \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x \in \mathcal{X}} \sum_{\hat{x} \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \mathbb{1}_{\{x = u_t(i)\}} P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \Bigg| \log_2 \left(1 + \theta C(x,y)\right) \Bigg|^3 \qquad (269) \\
&= n \sum_{x \in \mathcal{X}} \sum_{\hat{x} \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \bar{P}_X(x) P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \Bigg| \log_2 \left(1 + \theta C(x,y)\right) \Bigg|^3 \\
&= n \sum_{x \in \mathcal{X}} \sum_{\hat{x} \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \bar{P}_X(x) P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \Bigg| \sum_{k=1}^{\infty} \frac{(-1)^{k+1} \theta^k}{k} C(x,y)^k \Bigg|^3 \\
&= n\theta^3 \sum_{x \in \mathcal{X}} \sum_{\hat{x} \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \bar{P}_X(x) P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \Bigg| \sum_{k=1}^{\infty} \frac{(-1)^{k+1} \theta^{k-1}}{k} C(x,y)^k \Bigg|^3 . \qquad (270)
\end{aligned}$$

Note that the upper-bound in (269) is finite since $B(x,y)$ is bounded. Thus, the expression in (270) is also finite. Let $c_4$ be defined by

$$c_4 \triangleq \max \sum_{x \in \mathcal{X}} \sum_{\hat{x} \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \bar{P}_X(x) P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \Bigg| \sum_{k=1}^{\infty} \frac{(-1)^{k+1} \theta^{k-1}}{k} C(x,y)^k \Bigg|^3 , \qquad (271)$$

where the maximization is over all $\theta$ and all distributions $\tilde{P}_{\hat{X}|X}$. Using this notation, it follows that

$$\begin{aligned}
\hat{\phi} &\leqslant n\theta^3 c_4 \\
&= \frac{K^3}{\sqrt{n}} c_4 . \qquad (272)
\end{aligned}$$

This completes the proof. ∎

# O Proof of Lemma 14

Note that for all $t \in \{1, 2, \ldots, n\}$, it holds that

$$\mu_t = \frac{1}{M} \sum_{i=1}^{M} \sum_{y \in \mathcal{Y}_2} P_{Y_2|X}(y|u_t(i)) B(u_t(i), y), \tag{273}$$

$$\sigma_t^2 = \frac{1}{M} \sum_{i=1}^{M} \sum_{y \in \mathcal{Y}_2} P_{Y_2|X}(y|u_t(i)) B(u_t(i), y)^2 - \mu_t^2, \tag{274}$$

and

$$\phi_t = \frac{1}{M} \sum_{i=1}^{M} \sum_{y \in \mathcal{Y}_2} P_{Y_2|X}(y|u_t(i)) |B(u_t(i), y) - \mu_t|^3 . \tag{275}$$

Thus, it follows that

$$\begin{aligned}
\mu &= \sum_{t=1}^{n} \mu_t \\
&= \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{y \in \mathcal{Y}_2} P_{Y_2|X}(y|u_t(i)) B(u_t(i), y) \\
&= \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{y \in \mathcal{Y}_2} \sum_{x \in \mathcal{X}} \mathbb{1}_{\{x = u_t(i)\}} P_{Y_2|X}(y|x) B(x, y) \\
&= n \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \log_2 (1 + \theta C(x, y)) \\
&= n \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \sum_{k=1}^{\infty} \frac{(-1)^{k+1} \theta^k}{k} C(x, y)^k \\
&= n \sum_{x \in \mathcal{X}} \bar{P}_X(x) \sum_{k=2}^{\infty} \frac{(-1)^{k+1} \theta^k}{k} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \\
&= -n \frac{\theta^2}{2} \sum_{x \in \mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \\
&\quad + n\theta^3 \sum_{x \in \mathcal{X}} \bar{P}_X(x) \sum_{k=3}^{\infty} \frac{(-1)^{k+1} \theta^{k-3}}{k} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}).
\end{aligned} \tag{276}$$

Note that since the random variable $B(u_t(W), Y_{2Q,t})$ is bounded, its expectation is finite. Thus, it follows that the second term in (276) is also finite. Let $c_5$ be defined as

$$c_5 \triangleq \max \sum_{x \in \mathcal{X}} \bar{P}_X(x) \sum_{k=3}^{\infty} \frac{(-1)^k \theta^{k-3}}{k(k-1)} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}), \tag{277}$$

where the maximization is over all possible values of $\theta \in (0, 1)$ and all possible conditional distributions $\tilde{P}_{\hat{X}|X}$. Using this notation, it follows that

$$\begin{aligned}
\mu &\leqslant -n \frac{\theta^2}{2} \sum_{x \in \mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 |c_5| \\
&= -\frac{K^2}{2} \sum_{x \in \mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}} |c_5| .
\end{aligned} \tag{278}$$

Similarly, it holds that

$$\begin{aligned}
\sigma^2 &= \sum_{t=1}^{n} \sigma_t^2 \\
&= \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \mathbb{1}_{\{x=u_t(i)\}} P_{Y_2|X}(y|u_t(i)) B(u_t(i),y)^2 - \mu_t^2 \\
&\leqslant \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \mathbb{1}_{\{x=u_t(i)\}} P_{Y_2|X}(y|x) B(x,y)^2 \qquad (279) \\
&= n \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \log_2 \left(1+\theta C(x,y)\right)^2 \\
&= n \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \left( \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} C(x,y)^k \right)^2 \\
&= n \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \Bigg( \theta^2 C(x,y)^2 + 2\theta C(x,y) \sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} C(x,y)^k \\
&\quad + \left( \sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} C(x,y)^k \right)^2 \Bigg) \\
&= n \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \Bigg( \theta^2 C(x,y)^2 + 2 \sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^{k+1}}{k} C(x,y)^{k+1} \\
&\quad + \left( \sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} C(x,y)^k \right)^2 \Bigg) \\
&= n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \\
&\quad \cdot \left( 2 \sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^{k+1}}{k} C(x,y)^{k-2} + \left( \sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k} C(x,y)^k \right)^2 \right). \qquad (280)
\end{aligned}$$

Note that since $B(x,y)$ is bounded, the upper-bound in (279) is finite. Hence, the terms in (280) are also finite. Let $c_6$ be defined by

$$\begin{aligned}
c_6 \triangleq \max \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \Bigg( & 2 \sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^{k+1}}{k} C(x,y)^{k-2} \\
& + \left( \sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k} C(x,y)^k \right)^2 \Bigg), \qquad (281)
\end{aligned}$$

where the maximization is over all possible values of $\theta \in (0,1)$ and all possible conditional

distributions $\tilde{P}_{\hat{X}|X}$. Using this notation, it follows that

$$\begin{aligned}\sigma^2 &\leqslant n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 c_6 \\ &\leqslant n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 \left|c_6\right| \\ &\leqslant K^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}\left|c_6\right|. \end{aligned} \tag{282}$$

On the other hand, it also holds that

$$\begin{aligned}\sigma^2 &= \frac{1}{M}\sum_{t=1}^{n}\sum_{i=1}^{M}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2} \mathbb{1}_{\{x=u_t(i)\}} P_{Y_2|X}(y|x)B(x,y)^2 - \mu_t^2 \\ &\geqslant n\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{Y_2|X}(y|x)B(x,y)^2 - \sum_{t=1}^{n}\mu_t^2 \\ &= n\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{Y_2|X}(y|x)B(x,y)^2 \\ &\quad - \sum_{t=1}^{n}\left(\sum_{i=1}^{M}\sum_{y\in\mathcal{Y}_2}\frac{1}{M}P_{Y_2|X}(y|u_t(i))\log_2\left(1+\theta C(u_t(i),y)\right)\right)^2 \\ &\geqslant n\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{Y_2|X}(y|x)B(x,y)^2 \\ &\quad - \sum_{t=1}^{n}\sum_{i=1}^{M}\frac{1}{M}\left(\sum_{y\in\mathcal{Y}_2}P_{Y_2|X}(y|u_t(i))\cdot\log_2\left(1+\theta C(u_t(i),y)\right)\right)^2 \\ &= n\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{Y_2|X}(y|x)B(x,y)^2 - n\sum_{x\in\mathcal{X}}\bar{P}_X(x)\left(\sum_{y\in\mathcal{Y}_2}P_{Y_2|X}(y|x)\log_2\left(1+\theta C(x,y)\right)\right)^2 \\ &= n\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{Y_2|X}(y|x)B(x,y)^2 \\ &\quad - n\sum_{x\in\mathcal{X}}\bar{P}_X(x)\left(\sum_{y\in\mathcal{Y}_2}P_{Y_2|X}(y|x)\sum_{k=1}^{\infty}\frac{(-1)^{k+1}\theta^k}{k}C(x,y)^k\right)^2 \\ &= n\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{Y_2|X}(y|x)B(x,y)^2 \\ &\quad - n\sum_{x\in\mathcal{X}}\bar{P}_X(x)\left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^k}{k}\chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x})\right)^2 \\ &= n\theta^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{Y_2|X}(y|x) \\ &\quad \cdot\left(2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-2}}{k}C(x,y)^{k+1} + \left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k}C(x,y)^k\right)^2\right) \\ &\quad - n\sum_{x\in\mathcal{X}}\bar{P}_X(x)\left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^k}{k}\chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x})\right)^2 \end{aligned}$$

$$= n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2} \bar{P}_X(x)P_{Y_2|X}(y|x)$$
$$\cdot\left(2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-2}}{k}C(x,y)^{k+1} + \left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k}C(x,y)^{k}\right)^2\right.$$
$$\left.-\left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k}\chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x})\right)^2\right). \tag{283}$$

Note that since $B(x,y)$ is bounded, $\sigma^2$ is finite. Thus all the terms in (283) are also finite. Let $c_7$ be defined by

$$c_7 \triangleq \min \sum_{x\in\mathcal{X}} \bar{P}_X(x)\left(\chi_3(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2} P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})\right.$$
$$\cdot\left(2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-2}}{k}C(x,y)^{k+1} + \left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k}C(x,y)^{k}\right)^2\right.$$
$$\left.\left.-\Big(\sum_{k=2}^{\infty}\frac{(-1)^{k}\theta^{k-\frac{3}{2}}}{k(k-1)}\chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x})\Big)^2\right), \tag{284}$$

where the minimization is over all possible values of $\theta \in (0,1)$ and all possible conditional distributions $\tilde{P}_{\hat{X}|X}$. Plugging (284) into (283) yields

$$\sigma^2 \geqslant n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 c_7$$
$$\geqslant K^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}\left|c_7\right|. \tag{285}$$

Finally,

$$\begin{aligned}
\phi &= \sum_{t=1}^{n} \phi_t \\
&= \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \mathbb{1}_{\{x=u_t(i)\}} P_{Y_2|X}(y|x) \left| B(x,y) - \mu_t \right|^3 \\
&= \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \mathbb{1}_{\{x=u_t(i)\}} P_{Y_2|X}(y|x) \Bigg| B(x,y) - \frac{1}{M} \\
&\quad \cdot \sum_{i'=1}^{M} \sum_{y'\in\mathcal{Y}_2} P_{Y_2|X}(y'|u_t(i')) \log_2\left(1+\theta C(u_t(i'),y')\right) \Bigg|^3 \\
&\leqslant \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \mathbb{1}_{\{x=u_t(i)\}} P_{Y_2|X}(y|x) \Bigg| B(x,y) - \\
&\quad \cdot \log_2\left(1+\theta \frac{1}{M} \sum_{i'=1}^{M} \sum_{y'\in\mathcal{Y}_2} P_{Y_2|X}(y'|u_t(i')) C(u_t(i'),y')\right) \Bigg|^3 \\
&\leqslant \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \mathbb{1}_{\{x=u_t(i)\}} P_{Y_2|X}(y|x) \left| B(x,y) \right|^3 \qquad (286) \\
&= n \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \left| \log_2\left(1+\theta C(x,y)\right) \right|^3 \\
&= n \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \left| \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} C(x,y)^k \right|^3 \\
&= n\theta^3 \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \left| \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\theta^{k-1}}{k} C(x,y)^k \right|^3 . \qquad (287)
\end{aligned}$$

Note that the upper-bound in (286) is finite since $B(x,y)$ is bounded. Thus, the expression in (287) is also finite. Let $c_8$ be defined by

$$c_8 \triangleq \max \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \left| \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\theta^{k-1}}{k} C(x,y)^k \right|^3 , \qquad (288)$$

where the maximization is over all possible values of $\theta \in (0,1)$ and all possible conditional distributions $\tilde{P}_{\hat{X}|X}$. Using this notation, it follows that

$$\begin{aligned}
\phi &\leqslant n\theta^3 c_8 \\
&= \frac{K^3}{\sqrt{n}} c_8 . \qquad (289)
\end{aligned}$$

This completes the proof. ■

# P Proof of Proposition 6

Note that

$$\begin{aligned}
&\log_2\left(\hat{M}\right)\\
&= H(\hat{W})\\
&\overset{(a)}{=} H(\hat{W}|W)\\
&\overset{(b)}{\leqslant} I(\hat{W};\boldsymbol{Y}_1|W) + 1 + \hat{\epsilon}\log_2\left(\hat{M}\right)\\
&\overset{(c)}{=} I(\hat{\boldsymbol{X}};\boldsymbol{Y}_1|\boldsymbol{X}) + 1 + \hat{\epsilon}\log_2\left(\hat{M}\right)\\
&= H(\boldsymbol{Y}_1|\boldsymbol{X}) - H(\boldsymbol{Y}_1|\boldsymbol{X},\hat{\boldsymbol{X}}) + 1 + \hat{\epsilon}\log_2\left(\hat{M}\right)\\
&= \sum_{t=1}^{n} H(Y_{1,t}|\boldsymbol{X}, Y_{1,1}, Y_{1,2}, \ldots, Y_{1,t-1}) - H\big(Y_{1,t}|\boldsymbol{X},\hat{\boldsymbol{X}}, Y_{1,1}, Y_{1,2}, \ldots, Y_{1,t-1}\big) + 1 + \hat{\epsilon}\log_2\left(\hat{M}\right)\\
&\overset{(d)}{=} \sum_{t=1}^{n} H(Y_{1,t}|X_t) - H(Y_{1,t}|\hat{X}_t, X_t) + 1 + \hat{\epsilon}\log_2\left(\hat{M}\right)\\
&= nI(\hat{X};Y_1|X) + 1 + \hat{\epsilon}\log_2\left(\hat{M}\right),
\end{aligned} \tag{290}$$

where *(a)* follows from the independence between $W$ and $\hat{W}$; *(b)* follows from Fano's inequality [9]; *(c)* follows from the fact that the mapping from the set of message indices to the codewords is deterministic and bijective in both the broadcast code $\mathcal{C}$ and the covert code $\hat{\mathcal{C}}$; and *(d)* follows from the fact that the channel is memoryless.
Note that the mutual information in (290) is computed with respect to a joint distribution $Q_{X\hat{X}Y_1}$, where for all triplets $(x,\hat{x},y) \in \mathcal{X}^2 \times \mathcal{Y}_1$,

$$Q_{X\hat{X}Y_1}(x,\hat{x},y) \triangleq \bar{P}_X(x)\bar{P}_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x}); \tag{291}$$

the empirical conditional distribution $\bar{P}_{\hat{X}|X}$ is obtained from both (17) and (18); and $P_{Y_1|X}$ is the marginal of the joint distribution in (2b).

In order to calculate the mutual information in (290), let $Q_{Y_1|X}$ be the conditional marginal of the joint distribution $Q_{X\hat{X}Y_1}$ in (291), that is, for all pairs $(x,y) \in \mathcal{X} \times \mathcal{Y}_1$:

$$\begin{aligned}
Q_{Y_1|X}(y|x) &= \sum_{\hat{x}\in\mathcal{X}} \bar{P}_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})\\
&\overset{(a)}{=} (1-\theta(x))\,P_{Y_1|X}(y|x) + \theta(x)\sum_{\hat{x}\in\mathcal{X}} \hat{P}_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})\\
&\overset{(b)}{=} (1-\theta(x))\,P_{Y_1|X}(y|x) + \theta(x)\hat{R}_{Y_1|X}(y|x),
\end{aligned} \tag{292}$$

where *(a)* follows from Lemma 1 and *(b)* follows with $\hat{R}_{Y_1|X}(y|x)$ in (139).

Using (291) and (292), the mutual information in (290) satisfies

$$\begin{aligned}
&I(\hat{X};Y_1|X)\\
&=\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\bar{P}_X(x)\bar{P}_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})\cdot\log_2\left(\frac{P_{Y_1|X}(y|\hat{x})}{Q_{Y_1|X}(y|x)}\right)\\
&=\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\bar{P}_X(x)\bar{P}_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})\log_2\left(\frac{P_{Y_1|X}(y|\hat{x})P_{Y_1|X}(y|x)}{Q_{Y_1|X}(y|x)P_{Y_1|X}(y|x)}\right)\\
&=\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\bar{P}_X(x)\bar{P}_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})\left(\log_2\left(\frac{P_{Y_1|X}(y|\hat{x})}{P_{Y_1|X}(y|x)}\right)-\log_2\left(\frac{Q_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)}\right)\right)\\
&=\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\bar{P}_X(x)\bar{P}_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})\Bigg(\log_2\left(\frac{P_{Y_1|X}(y|\hat{x})}{P_{Y_1|X}(y|x)}\right)\\
&\quad-\log_2\left(\frac{(1-\theta(x))\,P_{Y_1|X}(y|x)+\theta(x)\hat{R}_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)}\right)\Bigg)\\
&=\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\bar{P}_X(x)\bar{P}_{\hat{X}|X}(\hat{x}|x)\Bigg(D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big)\\
&\quad-\sum_{y\in\mathcal{Y}_1}P_{Y_1|X}(y|\hat{x})\log_2\left(1+\theta(x)\frac{\hat{R}_{Y_1|X}(y|x)-P_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)}\right)\Bigg).
\end{aligned}\tag{293}$$

The last term in (293) can be approximated using a Taylor expansion of $\log_2(1+x)$ at $x=0$. For all $k\in\{1,2\}$, let $A_k:\mathcal{X}\times\mathcal{Y}_k\to\mathbb{R}$ be defined by

$$A_k(x,y)=\frac{\hat{R}_{Y_k|X}(y|x)-P_{Y_k|X}(y|x)}{P_{Y_k|X}(y|x)}.\tag{294}$$

Then, the second term in the right hand-side of (293) can be written as follows:

$$\begin{aligned}
&\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\bar{P}_X(x)\bar{P}_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})\log_2\left(1+\theta(x)A_1(x,y)\right)\\
&=\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\bar{P}_X(x)\left((1-\theta(x))\mathbb{1}_{\{x=\hat{x}\}}+\theta(x)\hat{P}_{\hat{X}|X}(\hat{x}|x)\right)P_{Y_1|X}(y|\hat{x})\log_2\left(1+\theta(x)A_1(x,y)\right)\\
&=\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\bar{P}_X(x)\Bigg(P_{Y_1|X}(y|\hat{x})\log_2\left(1+\theta(x)A_1(x,y)\right)\\
&\quad+\theta(x)\left(\sum_{\hat{x}\in\mathcal{X}}\hat{P}_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})-P_{Y_1|X}(y|x)\right)\log_2\left(1+\theta(x)A_1(x,y)\right)\Bigg)\\
&=\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\bar{P}_X(x)\Bigg(P_{Y_1|X}(y|\hat{x})\log_2\left(1+\theta(x)A_1(x,y)\right)+\theta(x)\left(\hat{R}_{Y_1|X}(y|x)-P_{Y_1|X}(y|x)\right)\\
&\quad\cdot\log_2\left(1+\theta(x)A_1(x,y)\right)\Bigg)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\bar{P}_X(x)\Bigg(P_{Y_1|X}(y|\hat{x})\sum_{k=1}^{\infty}\frac{(-1)^{k+1}\theta(x)^k}{k}A_1(x,y)^k + \theta(x)\left(\hat{R}_{Y_1|X}(y|x) - P_{Y_1|X}(y|x)\right)\\
&\quad\cdot\sum_{k'=1}^{\infty}\frac{(-1)^{k'+1}\theta(x)^{k'}}{k'}A_1(x,y)^{k'}\Bigg)\\
&= \sum_{x\in\mathcal{X}}\bar{P}_X(x)\Bigg(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta(x)^k}{k}\chi_k(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\\
&\quad+\sum_{k'=1}^{\infty}\frac{(-1)^{k'+1}\theta(x)^{k'+1}}{k'}\chi_{k'+1}(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\Bigg)\\
&= \sum_{x\in\mathcal{X}}\bar{P}_X(x)\Bigg(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta(x)^k}{k}\chi_k(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\\
&\quad+\sum_{k'=2}^{\infty}\frac{(-1)^{k'}\theta(x)^{k'}}{k'-1}\chi_{k'}(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\Bigg)\\
&= \sum_{x\in\mathcal{X}}\bar{P}_X(x)\sum_{k=2}^{\infty}\frac{(-1)^{k}\theta(x)^k}{k(k-1)}\chi_k(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\\
&= \sum_{x\in\mathcal{X}}\bar{P}_X(x)\left(\frac{\theta(x)^2}{2}\chi_2(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})+\sum_{k=3}^{\infty}\frac{(-1)^{k}\theta(x)^k}{k(k-1)}\chi_k(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\right)\\
&\geqslant \sum_{x\in\mathcal{X}}\bar{P}_X(x)\left(\frac{\theta(x)^2}{2}\chi_2(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})-\frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\right)\\
&\geqslant -\sum_{x\in\mathcal{X}}\bar{P}_X(x)\frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x},P_{Y_1|X=x}).
\end{aligned}
\tag{295}
$$

Therefore, from (293) and (295) it follows that

$$
\begin{aligned}
&I(\hat{X};Y_1|X)\\
&\overset{(a)}{\leqslant}\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\bar{P}_X(x)\left(\bar{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big)+\frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\right)\\
&\overset{(b)}{=}\sum_{x\in\mathcal{X}}\bar{P}_X(x)(1-\theta(x))D\big(P_{Y_1|X=x}||P_{Y_1|X=x}\big)+\bar{P}_X(x)\theta(x)\sum_{\hat{x}\in\mathcal{X}}\hat{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big)\\
&\quad+\bar{P}_X(x)\frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\\
&=\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\bar{P}_X(x)\theta(x)\hat{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big)+\bar{P}_X(x)\frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x},P_{Y_1|X=x}).
\end{aligned}
\tag{296}
$$

Finally, from (290) and (296), it follows that

$$
\begin{aligned}
\log_2\left(\hat{M}\right)\leqslant\frac{1}{1-\hat{\epsilon}}\Bigg(1+n\sum_{x\in\mathcal{X}}\bar{P}_X(x)\Bigg(&\theta(x)\sum_{\hat{x}\in\mathcal{X}}\hat{P}_{\hat{X}|X}(\hat{x}|x)\cdot D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big)\\
&+\frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\Bigg)\Bigg).
\end{aligned}
\tag{297}
$$

This completes the first part of the proof.

# Q Proof of Proposition 7

From Lemma 4, it follows that given an $(n, M, \epsilon)$-broadcast code, any $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code satisfies:

$$\begin{aligned}
\delta &\geqslant \|R_{\boldsymbol{Y}_2} - Q_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} \\
&\geqslant \|R_{W\boldsymbol{Y}_2} - Q_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} - \epsilon - \hat{\epsilon} \\
&= \frac{1}{2M} \sum_{i=1}^{M} \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} \left| \frac{1}{\hat{M}} \sum_{j=1}^{\hat{M}} P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}(i,j)) - P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i) \right| - \epsilon - \hat{\epsilon} \\
&= \frac{1}{M} \sum_{i=1}^{M} \left\| R_{\boldsymbol{Y}_2|W=i} - Q_{\boldsymbol{Y}_2|W=i} \right\|_{\mathrm{TV}} - \epsilon - \hat{\epsilon}, \qquad (298)
\end{aligned}$$

where the distributions $Q_{\boldsymbol{Y}_2|W=i}$ and $R_{\boldsymbol{Y}_2|W=i}$ are respectively defined in (30) and (31). For all message indices $i \in \mathcal{W}$, consider the set

$$\hat{\mathcal{W}}_i = \left\{ j \in \hat{\mathcal{W}} : \omega(i,j) \geqslant \nu \right\}, \qquad (299)$$

where $\nu$ will be specified later. Note that $\hat{\mathcal{W}}_i$ and $\hat{\mathcal{W}}_i^c$ form a partition of the set $\hat{\mathcal{W}}$. Let $R_{\boldsymbol{Y}_2|W}^{(\hat{\mathcal{W}}_i)}$ and $R_{\boldsymbol{Y}_2|W}^{(\hat{\mathcal{W}}_i^c)}$ be respectively defined by

$$R_{\boldsymbol{Y}_2|W}^{(\hat{\mathcal{W}}_i)}(\boldsymbol{y}|i) \triangleq \frac{1}{\left|\hat{\mathcal{W}}_i\right|} \sum_{j \in \hat{\mathcal{W}}_i} \prod_{t=1}^{n} P_{Y_2|X}(y_t|v_t(i,j)), \text{ and} \qquad (300)$$

$$R_{\boldsymbol{Y}_2|W}^{(\hat{\mathcal{W}}_i^c)}(\boldsymbol{y}|i) \triangleq \frac{1}{\left|\hat{\mathcal{W}}_i^c\right|} \sum_{j \in \hat{\mathcal{W}}_i^c} \prod_{t=1}^{n} P_{Y_2|X}(y_t|v_t(i,j)). \qquad (301)$$

Consider that the transmission of covert communications occurs by using the sub-code whose codewords have lower-bounded weight, i.e., $v(i,j)$ with $i \in \mathcal{W}$ and $j \in \hat{\mathcal{W}}_i$. Under this consideration, the test run by Receiver 2 to determine whether or not private messages are being sent is

$$\begin{cases} H_0 : \boldsymbol{Y}_2 \sim Q_{\boldsymbol{Y}_2|W=i}, \\ H_1 : \boldsymbol{Y}_2 \sim R_{\boldsymbol{Y}_2|W=i}^{(\hat{\mathcal{W}}_i)}, \end{cases} \qquad (302)$$

where the distributions $Q_{\boldsymbol{Y}_2|W=i}$ and $R_{\boldsymbol{Y}_2|W=i}^{(\hat{\mathcal{W}}_i)}$ are respectively defined in (30) and (300).
Denote by $\hat{\alpha}_i \in [0,1]$ and $\hat{\beta}_i \in [0,1]$ the type-I and type-II error probabilities associated with a decision rule $T_i : \mathcal{Y}_2^n \to \{0,1\}$ of the form

$$T_i(\boldsymbol{y}) \triangleq \begin{cases} 0 & \text{if } H_0 \text{ is accepted,} \\ 1 & \text{if } H_1 \text{ is accepted.} \end{cases} \qquad (303)$$

That is,

$$\hat{\alpha}_i \triangleq \Pr\left[T_i(\boldsymbol{Y}_2) = 1\right], \text{ and} \qquad (304)$$

$$\hat{\beta}_i \triangleq \Pr\left[T_i(\boldsymbol{Y}_2) = 0\right], \qquad (305)$$

where the probability operator in (304) applies assuming that $\boldsymbol{Y}_2 \sim Q_{\boldsymbol{Y}_2|W=i}$ and the probability operator in (305) applies assuming that $\boldsymbol{Y}_2 \sim R_{\boldsymbol{Y}_2|W=i}^{(\hat{\mathcal{W}}_i)}$.

Note also that for all message indices $i \in \mathcal{W}$, it follows that

$$
\begin{aligned}
&\left\|R_{\boldsymbol{Y}_2|W=i} - Q_{\boldsymbol{Y}_2|W=i}\right\|_{\mathrm{TV}} \\
&= \frac{1}{2}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\left|P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i)) - \frac{1}{\hat{M}}\sum_{j=1}^{\hat{M}} P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}(i,j))\right| \\
&= \frac{1}{2}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\left|\frac{1}{\hat{M}}\sum_{j\in\hat{\mathcal{W}}_i}\Big(P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i)) - P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}(i,j))\Big)\right. \\
&\quad \left. -\frac{1}{\hat{M}}\sum_{j\in\hat{\mathcal{W}}_i^{\mathsf{c}}}\big(P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}(i,j)) - P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))\big)\right| \\
&\overset{(a)}{\geqslant} \frac{1}{2}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\left|\frac{1}{\hat{M}}\sum_{j\in\hat{\mathcal{W}}_i}\big(P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i)) - P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}(i,j))\big)\right| \\
&\quad -\left|\frac{1}{\hat{M}}\sum_{j\in\hat{\mathcal{W}}_i^{\mathsf{c}}}\big(P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}(i,j)) - P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))\big)\right| \\
&= \frac{1}{2}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\left(\frac{|\hat{\mathcal{W}}_i|}{\hat{M}}\left|\frac{1}{|\hat{\mathcal{W}}_i|}P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i)) - \frac{1}{|\hat{\mathcal{W}}_i|}\sum_{j\in\hat{\mathcal{W}}_i} P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}(i,j))\right|\right. \\
&\quad \left. -\frac{|\hat{\mathcal{W}}_i^{\mathsf{c}}|}{\hat{M}}\left|\frac{1}{|\hat{\mathcal{W}}_i^{\mathsf{c}}|}\sum_{j\in\hat{\mathcal{W}}_i^{\mathsf{c}}} P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}(i,j)) - P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))\right|\right) \\
&\overset{(b)}{=} \frac{|\hat{\mathcal{W}}_i|}{\hat{M}}\left\|R^{(\hat{\mathcal{W}}_i)}_{\boldsymbol{Y}_2|W=i} - Q_{\boldsymbol{Y}_2|W=i}\right\|_{\mathrm{TV}} - \frac{|\hat{\mathcal{W}}_i^{\mathsf{c}}|}{\hat{M}}\left\|R^{(\hat{\mathcal{W}}_i^{\mathsf{c}})}_{\boldsymbol{Y}_2|W=i} - Q_{\boldsymbol{Y}_2|W=i}\right\|_{\mathrm{TV}} \\
&\overset{(c)}{\geqslant} \frac{|\hat{\mathcal{W}}_i|}{\hat{M}}\left\|R^{(\hat{\mathcal{W}}_i)}_{\boldsymbol{Y}_2|W=i} - Q_{\boldsymbol{Y}_2|W=i}\right\|_{\mathrm{TV}} - \frac{|\hat{\mathcal{W}}_i^{\mathsf{c}}|}{\hat{M}} \\
&\overset{(d)}{\geqslant} \frac{|\hat{\mathcal{W}}_i|}{\hat{M}}\left(1-\hat{\alpha}_i-\hat{\beta}_i\right) - \frac{|\hat{\mathcal{W}}_i^{\mathsf{c}}|}{\hat{M}} \\
&= \frac{\hat{M}-|\hat{\mathcal{W}}_i^{c}|}{\hat{M}}\left(1-\hat{\alpha}_i-\hat{\beta}_i\right) - \frac{|\hat{\mathcal{W}}_i^{\mathsf{c}}|}{\hat{M}} \\
&\geqslant \left(1-\hat{\alpha}_i-\hat{\beta}_i\right) - 2\frac{|\hat{\mathcal{W}}_i^{\mathsf{c}}|}{\hat{M}} \\
&\overset{(e)}{\geqslant} \left(1 - 2Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right) - \frac{c_{15}}{\sqrt{n}}\right) - 2\frac{|\hat{\mathcal{W}}_i^{\mathsf{c}}|}{\hat{M}},
\end{aligned}
\tag{306}
$$

where $\nu$ will be specified later, $c_{15}$ is a constant, *(a)* is a consequence of the triangle inequality; *(b)* follows from the definition of $R^{(\hat{\mathcal{W}}_i)}_{\boldsymbol{Y}_2|W}$ and $R^{(\hat{\mathcal{W}}_i^{\mathsf{c}})}_{\boldsymbol{Y}_2|W}$ in (300) and (301) respectively; *(c)* follows since $\left\|R^{(\hat{\mathcal{W}}_i^{\mathsf{c}})}_{\boldsymbol{Y}_2|W=i} - Q_{\boldsymbol{Y}_2|W=i}\right\|_{\mathrm{TV}} \leqslant 1$; *(d)* follows from Lemma 8 in Appendix A; and *(e)* follows from Proposition 1.

Plugging (298) into (306) yields

$$
\delta \geqslant 1 - 2Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right) - \frac{c_{15}}{\sqrt{n}} - 2\sum_{i=1}^{M}\frac{|\hat{\mathcal{W}}_i^{\mathsf{c}}|}{M\hat{M}} - \epsilon - \hat{\epsilon}.
\tag{307}
$$

Choosing $\nu$ in (299) such that

$$\nu = \frac{2\sqrt{n}}{\sqrt{d}} Q^{-1}\left(\frac{1-\delta-\eta}{2}\right), \tag{308}$$

with $\eta \in (0, 1-\delta)$, it follows that

$$\begin{aligned} 2\sum_{i=1}^{M} \frac{|\hat{\mathcal{W}}_i^{\mathfrak{c}}|}{M\hat{M}} &\geqslant 1 - 2Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right) - \frac{c_{15}}{\sqrt{n}} - \delta - \epsilon - \hat{\epsilon} \\ &= 1 - 1 + \delta + \eta - \frac{c_{15}}{\sqrt{n}} - \delta - \epsilon - \hat{\epsilon} \\ &= \eta - \frac{c_{15}}{\sqrt{n}} - \epsilon - \hat{\epsilon}, \end{aligned} \tag{309}$$

which implies

$$\frac{\left|\tilde{\mathcal{W}}\right|}{M} = \sum_{i=1}^{M} \frac{|\hat{\mathcal{W}}_i^{\mathfrak{c}}|}{M} \geqslant \hat{M}\left(\frac{\eta}{2} - \frac{c_6}{\sqrt{n}} - \epsilon - \hat{\epsilon}\right), \tag{310}$$

with $c_6 = \frac{c_{15}}{2}$. This completes the proof. ■

# 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] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 2945–2949.

[3] M. Bloch, "Covert communications over noisy channels: A resolvability perspective," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2334–2354, May 2016.

[4] L. Wang, G. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3493–3503, Jun. 2016.

[5] V. Y. F. Tan and S. Lee, "Time-division transmission is optimal for covert communication over broadcast channels," *CoRR*, vol. abs/1710.09754, 2017. [Online]. Available: http://arxiv.org/abs/1710.09754

[6] K. S. K. Arumugam and M. R. Bloch, "Covert communication over broadcast channels," in *Proc. of IEEE Information Theory Workshop (ITW)*, Kaohsiung, Taiwan, Nov. 2017, pp. 299–303.

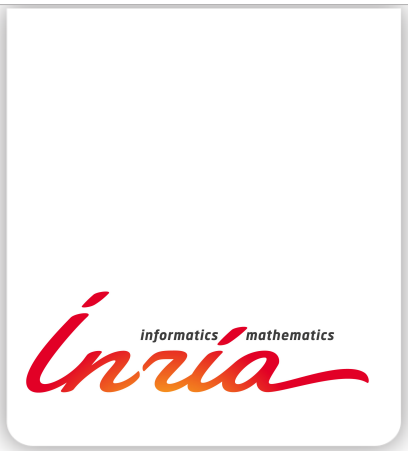
[7] ——, "Embedding covert information in broadcast communications," *CoRR*, vol. abs/1808.09556, 2018. [Online]. Available: http://arxiv.org/abs/1808.09556

[8] M. Tahmasbi and M. R. Bloch, "Second-order asymptotics in covert communication," 2017. [Online]. Available: http://arxiv.org/abs/1703.01362

[9] R. Fano, *Transmission of Information: A Statistical Theory of Communication*, 1st ed. MIT Press, 1961.

[10] I. Shevtsova, "On the absolute constants in the Berry-Esseen type inequalities for identically distributed summands," *ArXiv e-prints*, Nov. 2011.

[11] R. Vershynin, *High-Dimensional Probability*, 1st ed. Cambridge University Press, 2018.




**RESEARCH CENTRE**
**GRENOBLE – RHÔNE-ALPES**

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399