Information Theoretic Contributions to Covert Communications

David Kibloff

N° d'ordre NNT : 2019LYSEI070

THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LYON
opérée au sein de
l'INSA de Lyon

École Doctorale N° 160
Électronique, Électrotechnique et Automatique
(ED EEA)

Spécialité / discipline de doctorat :
Génie Électrique

Soutenue publiquement le 17/09/2019, par :
David Kibloff

Contributions Théoriques sur les Communications Furtives

Devant le jury composé de :

Rapporteurs :		
GUILLEN I FABREGAS Albert	Universitat Pompeu Fabra	Espagne
ROUMY Aline	INRIA	France
Examinateurs :		
CLAVIER Laurent	IMT Lille Douai	France
FIJALKOW Inbar	Université de Cergy-Pontoise	France
GORCE Jean-Marie	INSA Lyon	France
WANG Ligong	CNRS	France
Directeur de thèse :		
VILLEMAUD Guillaume	INSA Lyon	France
Encadrant :		
PERLAZA Samir	INRIA	France
Membre invité :		
COSQUER Ronan	Direction générale de l'armement	France

UNIVERSITY OF LYON
Doctoral School of Electronics, Electrotechnics and Automation
(ED EEA)

THESIS

presented in partial fulfilment of the requirements
for the degree of Doctor of Philosophy from the University of Lyon,
the 17/09/2019.

Specialization: Electrical Engineering

David Kibloff

Information Theoretic Contributions to Covert Communications

Jury:

Reviewers:		
GUILLEN I FABREGAS Albert	Universitat Pompeu Fabra	Spain
ROUMY Aline	INRIA	France
Examiners:		
CLAVIER Laurent	IMT Lille Douai	France
FIJALKOW Inbar	Université de Cergy-Pontoise	France
GORCE Jean-Marie	INSA Lyon	France
WANG Ligong	CNRS	France
Supervisor:		
VILLEMAUD Guillaume	INSA Lyon	France
Advisor:		
PERLAZA Samir	INRIA	France
Guest:		
COSQUER Ronan	Direction générale de l'armement	France

UNIVERSITÉ DE LYON
Électronique, Électrotechnique et Automatique
(ED EEA)

THÈSE

présentée publiquement pour l'obtention
du diplôme de Docteur de l'Université de Lyon,
le 17/09/2019.

Spécialité : Génie Électrique

David Kibloff

Contributions Théoriques sur les Communications Furtives

Jury :

Rapporteurs :		
GUILLEN I FABREGAS Albert	Universitat Pompeu Fabra	Espagne
ROUMY Aline	INRIA	France
Examinateurs :		
CLAVIER Laurent	IMT Lille Douai	France
FIJALKOW Inbar	Université de Cergy-Pontoise	France
GORCE Jean-Marie	INSA Lyon	France
WANG Ligong	CNRS	France
Directeur de thèse :		
VILLEMAUD Guillaume	INSA Lyon	France
Encadrant :		
PERLAZA Samir	INRIA	France
Membre invité :		
COSQUER Ronan	Direction générale de l'armement	France

Abstract

The problem of covert communications, also known as communications with low-probability of detection has gained interest in the information theory community in the last years. Since Bash *et al.* showed in 2012 that the square-root law applied in the point-to-point case for such communications systems, the number of contributions on the topic did not cease to grow. In this thesis, two new problems of covert communications are introduced. First, the problem of covert communications over a point-to-point link where a warden observes only a fraction of channel outputs in order to try to detect the communications is studied. An achievability bound in the finite block-length regime is derived for this problem. Second, the problem of embedding covert information into a given broadcast code is introduced. Given a broadcast code to transmit a common message to two receivers, the goal is to determine the maximum number of information bits that can be reliably sent to one receiver while remaining covert with respect to the other receiver. For this problem, both an achievability and converse bound in the asymptotic block-length regime are derived for a particular class of channels, *i.e.*, symmetric channels. Together these bounds characterize the maximum number of information bits that can be covertly embedded in a given broadcast code for symmetric channels.

Résumé

L'étude des communications furtives, aussi connues sous le nom de communications avec faible probabilité de détection, a connu un regain d'intérêt dans la communauté Théorie de l'Information dans les années passées. Depuis que Bash *et al.* ont montré en 2012 que les communications point-à-point sous contrainte de furtivité obéissent à une loi en racine carrée, le nombre de contributions dans ce domaine n'a cessé de croître. Dans cette thèse, deux nouveaux problèmes de communications furtives sont présentés. Premièrement, les communications furtives sur les liens point-à-point sont étudiées quand l'adversaire observe uniquement une fraction des sorties de canal pour essayer de détecter la communication. Une borne de faisabilité pour une longueur finie de blocs est obtenue pour ce problème. Deuxièmement, le problème d'introduction d'information furtive dans un code de broadcast existant est présenté. Etant donné un code de broadcast pour transmettre de l'information à deux récepteurs, le but de cette étude est de déterminer le nombre maximum de bits d'information qui peuvent être envoyés de manière fiable à l'un des récepteurs tout en étant furtifs pour l'autre récepteur. Pour ce problème, une borne de faisabilité et une borne d'impossibilité sont obtenues dans le régime asymptotique pour une classe particulière de canaux, *i.e.*, les canaux symétriques. Ces deux bornes caractérisent le nombre maximal de bits d'information qui peuvent être introduits de manière furtive dans le code de broadcast donné pour des canaux symétriques.

Education

2010 : B.Sc. at INSA Lyon, France

2012 : M.Eng. at INSA Lyon, France

Publications

[1] F. Hutu, D. Kibloff, G. Villemaud, and Jean-Marie Gorce. Experimental validation of a wake-up radio architecture. In *IEEE Radio and Wireless Symposium (RWS)*, pages 155–158, Austin, TX, USA, Jan. 2016.

[2] D. Kibloff, S. M. Perlaza, G. Villemaud, and L. S. Cardoso. On the duality between state-dependent channels and wiretap channels. In *Proc. of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Washington, DC, USA, Dec. 2016.

[3] D. Kibloff, S. M. Perlaza, and L. Wang. Broadcast codes can be enhanced to perform covert communications. Technical Report 9249, INRIA Grenoble - Rhône-Alpes, Jan. 2019.

[4] D. Kibloff, S. M. Perlaza, and L. Wang. Embedding covert information on a given broadcast code. In *IEEE International Symposium on Information Theory (ISIT)*, Paris, France, Jul. 2019.

Contents

Acronyms xv

Notation xvi

Synthèse des contributions majeures xvii
1. Introduction . . . xvii
2. Etat de l'art . . . xviii
   2.1. Résultats théoriques principaux sur les communications furtives . . . xx
3. Résultat principal . . . xxiii
   3.1. Modèle . . . xxiv
   3.2. Faisabilité des communications furtives . . . xxviii
   3.3. Impossibilité des communications furtives . . . xxxii
   3.4. Résultat principal . . . xxxvii
   3.5. Exemples . . . xxxvii
4. Conclusion . . . xlii

1. Introduction 1

2. State of the Art 3
2.1. Main Results on Covert Communications . . . 4
   2.1.1. Point-to-point Channels . . . 5
   2.1.2. Point-to-point Channels with Jammers . . . 6
   2.1.3. Multiple-User Channels . . . 6
   2.1.4. Coding for Covert Communications . . . 7
2.2. Detailed Results for Canonical Information Theoretic Channels . . . 7
   2.2.1. Point-to-point Channels . . . 7
   2.2.2. Point-to-point Channels with States . . . 16
   2.2.3. Broadcast Channels . . . 22
   2.2.4. Multiple-Access Channels . . . 28
2.3. Conclusion . . . 31

3. Covert Communications Type II 33
3.1. System Model . . . 33
   3.1.1. Channel Model . . . 33
   3.1.2. Covert Codes . . . 35
   3.1.3. Fundamental Limits . . . 36
3.2. An achievable bound for Binary Memoryless Channels . . . 36
3.3. Discussion . . . 39
3.4. Conclusion . . . 39

4. Embedding Covert Information into a Given Broadcast Code **41**
4.1. System Model . . . 42
4.1.1. Broadcast Codes . . . 43
4.1.2. Induced Codes . . . 44
4.1.3. Covert Codes . . . 45
4.2. Examples of Impossible Covert Communications . . . 46
4.3. Achievability of Covert Communications . . . 49
4.4. Impossibility of Covert Communications . . . 53
4.4.1. Auxiliary Results . . . 53
4.4.2. Finite Block-length Results . . . 56
4.4.3. Asymptotic Result . . . 57
4.5. Main Result . . . 58
4.6. Examples . . . 58
4.7. Conclusion . . . 63

5. Conclusion **65**

A. Auxiliary Results **67**

B. Proof of Proposition 1 **69**

C. Proof of Proposition 2 **73**

D. Proof of Lemma 9 **77**

E. Proof of Lemma 3 **81**

F. Proof of Lemma 4 **83**

G. Proof of Proposition 3 **87**

H. Proof of Lemma 10 **101**

I. Proof of Lemma 11 **109**

J. Proof of Proposition 4 **115**

K. Proof of Lemma 5 **123**

L. Proof of Lemma 6 **125**

M. Proof of Proposition 5 **127**

N. Proof of Proposition 6 **131**

O. Proof of Lemma 7 **143**

Bibliography **145**

List of Figures

1. Système de communication point-à-point sécurisé. xix
2. Canal broadcast dégradé avec messages furtifs à l'utilisation de canal $t \in \{1, 2, \ldots, n\}$, où $d_1 : \mathcal{Y}_1^n \to \mathcal{W} \times \hat{\mathcal{W}}$ dénote la fonction de décodage au Récepteur 1 et $d_2 : \mathcal{Y}_2^n \to \mathcal{W}$ dénote la fonction de décodage au Récepteur 2. xxiv
3. Canal broadcast dégradé verifiant (69) et (70), à l'utilisation de canal $t \in \{1, 2, \ldots, n\}$. xxxviii
4. Limite fondamentale $\lim_{n\to\infty} \frac{\log_2\left(\hat{M}^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)}{\sqrt{n}}$ en fonction de la probabilité p_1, pour $\delta = 0.005$. xxxix
5. Limite fondamentale $\lim_{n\to\infty} \frac{\log_2\left(\hat{M}^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)}{\sqrt{n}}$ en fonction de la probabilité p_2, pour $\delta = 0.005$. xxxix
6. Canal broadcast dégradé vérifiant (69) et (70), à l'utilisation de canal $t \in \{1, 2, \ldots, n\}$. xl
7. Limite fondamentale $\lim_{n\to\infty} \frac{\log_2\left(\hat{M}^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)}{\sqrt{n}}$ en fonction de la probabilité p_1, pour $\delta = 0.005$. xli
8. Limite fondamentale $\lim_{n\to\infty} \frac{\log_2\left(\hat{M}^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)}{\sqrt{n}}$ en fonction de la probabilité p_2, pour $\delta = 0.005$. xlii

2.1. Point-to-point secrecy system. 3
2.2. Point-to-point channel with a warden. 8
2.3. Point-to-point channel with states and warden. 16
2.4. Broadcast channel with a warden. 22
2.5. Broadcast channel with covert messages. 25
2.6. Point-to-point channel with a warden. 28

3.1. Point-to-point channel with a warden. 34

4.1. 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. 42
4.2. Degraded erasure broadcast channel at channel use $t \in \{1, 2, \ldots, n\}$. 47
4.3. Degraded typewriter broadcast channel at channel use $t \in \{1, 2, \ldots, n\}$. . . . 49
4.4. Degraded broadcast channel satisfying (4.77) and (4.78) at channel use $t \in \{1, 2, \ldots, n\}$. 59
4.5. 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 $\delta = 0.005$. 60
4.6. 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_2, for $\delta = 0.005$. 60

4.7. Degraded broadcast channel satisfying (4.77) and (4.78) at channel use $t \in \{1, 2, \ldots, n\}$. 61
4.8. 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 $\delta = 0.005$. 62
4.9. 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 $\delta = 0.005$. 63

Acronyms

AWGN	Additive White Gaussian Noise
BC	Broadcast Channel
BMC	Binary Memoryless Channel
CSI	Channel State Information
DMC	Discrete Memoryless Channel
MAC	Multiple Access Channel
WTC	Wiretap Channel

Notation

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

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

The Kullback-Leibler divergence between P_X and Q_X is denoted by

$$D(P_X||Q_X) = \sum_{x \in \mathcal{X}} P_X(x) \log\left(\frac{P_X(x)}{Q_X(x)}\right). \tag{2}$$

Finally, the total variation distance between P_X and Q_X is given by

$$\|P_X - Q_X\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x \in \mathcal{X}} |P_X(x) - Q_X(x)|. \tag{3}$$

Whenever a second random variable Y is considered, P_{XY} 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 $\mathbb{E}_{Y|X=x}[\cdot]$ and $\mathbb{V}_{Y|X=x}[\cdot]$.

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 $N(x, x'|\boldsymbol{x}, \boldsymbol{x}') \triangleq \sum_{t=1}^{n} \mathbb{1}_{\{x=x_t\}} \mathbb{1}_{\{x'=x'_t\}}$. Given a set of m indices $\mathcal{A} = \{a_1, a_2, \ldots, a_m\} \subseteq \{1, 2, \ldots, n\}$ and an n-length vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, $\boldsymbol{X}_{\mathcal{A}}$ denotes the m-length vector $\boldsymbol{X}_{\mathcal{A}} = (X_{a_1}, X_{a_2}, \ldots, X_{a_m})$. Given $\ell \in \mathbb{N}$, $\ell \neq n$, an ℓ-dimensional vector of random variables is denoted by $\boldsymbol{X}_{(1:\ell)} = (X_1, X_2, ..., X_\ell) \in \mathcal{X}^\ell$ and its realization is denoted by $\boldsymbol{x}_{(1:\ell)} = (x_1, x_2, ..., x_\ell)$.

Synthèse des contributions majeures

1. Introduction

La sécurité des systèmes de communications est un sujet de recherche actif depuis de nombreuses années, autant du point de vue théorique que du point de vue pratique, et ce particulièrement dans les systèmes sans fils. En effet, espionner une ligne filaire n'est possible que si l'espion a un accès physique à la ligne de communication afin de mettre celle-ci sur écoute. A l'inverse, la nature radiodiffusée des réseaux sans fils permet à des utilisateurs malveillants d'espionner ou de perturber la communication. Dans de tels réseaux, le simple fait d'avoir un récepteur permet l'espionnage du système. Ainsi, en raison de l'augmentation inédite du nombre d'appareils sans fils connectés autour du globe, la sécurité des systèmes de communication sans fils connaît une demande croissante. Dans cette thèse, l'étude portera sur les problèmes rencontrés quand un utilisateur malveillant espionne la communication.

D'un point de vue théorique, la sécurité des systèmes de communication a été d'abord étudiée dans l'article fondateur de Shannon [1], qui introduit le concept de sécurité parfaite. D'après Shannon, afin d'avoir une communication sécurisée en présence d'un espion, le message et l'observation de l'espion doivent être statistiquement indépendants. Cette notion de sécurité parfaite est assez contraignante, et donc, elle a été relaxée dans la littérature. Ainsi, les concepts dérivés de sécurité faible, sécurité forte, sécurité effective et sécurité sémantique ont été introduits et étudiés. Toutes ces mesures de sécurité garantissent différents niveaux d'indépendance statistique entre le message et l'observation de l'espion.

Une autre façon d'assurer la non-décodabilité du message par l'espion est de garantir que celui-ci ne puisse pas détecter la communication elle-même. Ce problème est connu sous le nom de communications furtives ou communications à faible probabilité de détection. Dans ce problème, le codage des message doit garantir que le meilleur détecteur que l'espion pourrait détenir échoue presque systématiquement à détecter la communication. C'est-à-dire, dans le cas le plus simple d'une communication point-à-point, la sortie de canal de l'espion doit ressembler au bruit qu'il observe lorsqu'il n'y a pas de communication. Ce genre de contrainte de sécurité est bien plus stricte que les concepts de sécurité discutés précédemment.

Néanmoins, les communications furtives trouvent leurs applications, parmi d'autres, dans le domaine militaire. Par exemple, considérons un scénario de bataille dans lequel un général envoie des ordres à ses troupes. Dans certaines circonstances, il pourrait être crucial pour l'adversaire de savoir que des ordres ont été envoyés, même si ceux-ci ne sont pas explicitement connus, justifiant le recours à des communications furtives. Une autre application concerne la conception des véhicules furtifs, tels les sous-marins par exemple, qui peuvent communiquer sans révéler leur présence, ou pire, leur position. Les communications furtives trouvent également des applications dans le journalisme d'investigation, où la transmission secrète des données est

cruciale pour la coopération des journalistes. Enfin, dans la vie quotidienne, les communications furtives pourraient être utilisées pour échanger des données critiques telles que les données médicales dans le contexte de l'Internet des Objets, ou bien pour le paiement sans fil des cartes de crédit.

L'étude théorique des modèles de systèmes de communication furtifs est d'un grand intérêt. Elle permet de fournir au concepteur d'un système de communication des bornes sur le débit maximal auquel les données pourront être envoyées de façon fiable et furtive. Ainsi, les ingénieurs dédiés à la conception du code peuvent évaluer à quel point leur produit présente des performances proches des performances optimales.

Cette thèse présente deux contributions théoriques principales. Premièrement, le problème des communications furtives de type II est introduit. Il s'agit d'un problème de communication furtive point-à-point traditionnel, dans lequel l'espion est autorisé à choisir une fraction des sorties de canal pour détecter la communication. Pour ce problème, une borne de faisabilité – une borne inférieure sur le débit maximal auquel la transmission peut être fiable et furtive – est présentée pour un code de longueur finie. Deuxièmement, le problème d'intégration d'information furtive dans un code broadcast est présenté. Etant donné un code broadcast pour transmettre un message commun à deux récepteurs, l'objectif dans ce problème est de déterminer le nombre maximal de bits d'information qui peuvent être envoyés de façon fiable à l'un des récepteurs sans que l'autre ne le détecte, en plus du message commun. Dans cette thèse, une borne de faisabilité et une borne d'impossibilité – une borne supérieure – sur le nombre maximal de bits d'informations qui peuvent être intégrés de manière furtive dans un code broadcast donné sont établies pour des codes de longueur infinie, et pour une classe de canaux particuliers, *i.e.*, les canaux symétriques. Dans cette synthèse des résultats en français, seul le second problème est abordé.

Aucun de ces deux problèmes n'a été étudié précédemment. Le problème des communications furtives de type II généralise le problème des communications furtives sur un canal point-à-point. La borne de faisabilité présentée dans cette thèse est un premier pas vers la résolution de ce modèle généralisé. Le problème d'intégration d'information furtive dans un code broadcast donné est nouveau, du fait que le code broadcast soit donné. A l'inverse, [2, 3] traitent le problème du design conjoint d'un code qui permet l'envoi d'un message commun et d'un message furtif à l'un des récepteurs. Une fois encore, la caractérisation du nombre maximal de bits d'information qui peuvent intégrés à un code broadcast de manière furtive constitue un premier pas vers la résolution du problème pour des canaux discrets sans mémoire arbitraires.

Le reste de cette partie suit le plan suivant. La section 2 présente brièvement l'état de l'art. La section 3 est dédiée au résultat principal de cette thèse, c'est-à-dire, la caractérisation du nombre maximal de bits d'information qui peuvent intégrés à un code broadcast de manière furtive. Enfin, la section 4 conclue cette synthèse des résultats.

2. Etat de l'art

La sécurité des systèmes de communication est un problème de longue date en théorie de l'information. Shannon lui-même a introduit le le premier modèle d'étude pour la sécurité des systèmes de communication [1]. Dans ce problème, le transmetteur a pour but de communiquer avec un récepteur sur un canal sans bruit en présence d'un adversaire, l'espion, qui observe également une version non-buitée du signal transmis (voir Fig. 2.1). Shannon a introduit dans son artivle la notion de *sécurité parfaite* pour traiter ce problème et a défini la sécurité parfaite

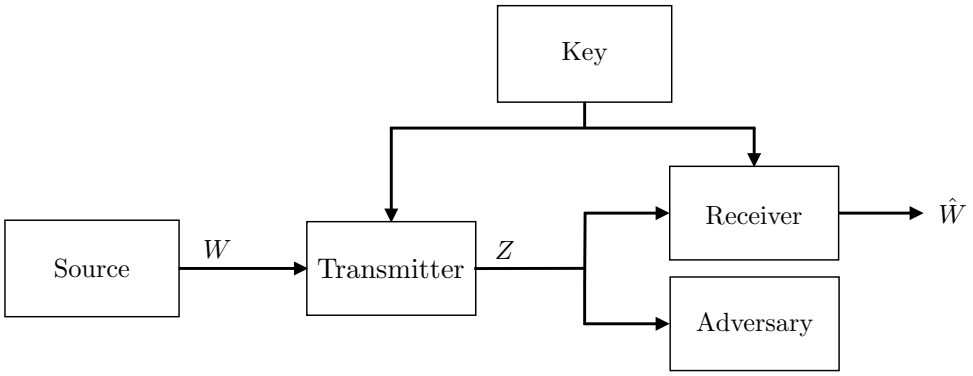

FIGURE 1. : Système de communication point-à-point sécurisé.

comme suit :

$$I(W; \boldsymbol{Z}) = 0, \tag{4}$$

où $W \in \mathcal{W}$ est le message secret, $\mathcal{W}$ est l'ensemble des messages, et $\boldsymbol{Z} \in \mathcal{Z}^n$ est la sortie de canal de l'espion, avec $\mathcal{Z}$ l'alphabet de sortie de canal de l'espion, et $n \in \mathbb{N}$ la longueur du code. En d'autres termes, étant donné son observation de sortie de canal $\boldsymbol{Z}$, l'espion ne doit pas être capable d'inférer quoi que ce soit à propos du message W, a fortiori, il ne doit pas être capable de décoder W. Shannon a conclu que pour satisfaire une telle contrainte, une clé secrète partagée entre l'émetteur et le récepteur était requise, et l'entropie de la clé doit être égale à l'entropie de la source de messages. En d'autres termes, en supposant les messages et clés distribués uniformément, le nombre de messages et de clés doit être le même.

Plus tard, Wyner introduit le canal wiretap [4], modèle dans lequel le canal n'est plus sans bruit, et dans lequel l'espion observe une version dégradée du signal reçu par le récepteur. Il a également relaxé la contrainte de sécurité de Shannon et introduit la notion de *sécurité faible* définie comme suit :

$$\lim_{n\to\infty} \frac{1}{n} I(W; \boldsymbol{Z}) = 0. \tag{5}$$

Avec ces hypothèses de travail, Wyner a pu montrer qu'une communication sécurisée sans clé était faisable.

Csiszár a introduit le concept de *sécurité forte* [5], dans le but de renforcer la mesure de Wyner, et l'a défini comme suit :

$$\lim_{n\to\infty} I(W; \boldsymbol{Z}) = 0. \tag{6}$$

En effet, l'abandon de la normalisation renforce la mesure, dans le sens où les performances guaranties en termes de sécurité sont meilleures.

Ensuite, Hou, *et al.* ont introduit la notion de *sécurité effective* [6], définie comme suit :

$$\lim_{n\to\infty} D(P_{W\boldsymbol{Z}} || P_W Q_{\boldsymbol{Z}}) = 0, \tag{7}$$

où $Q_{\boldsymbol{Z}}$ est la distribution attendue par l'espion lorsque l'émetteur n'envoie pas de messages utiles. Cette mesure regroupe deux critères, nommés confusion de l'espion et discrétion du

message (qui est d'une certaine façon liée au problème des communications furtives).

Enfin, Bellare *et al.* ont introduit la notion de *sécurité sémantique*, définie comme suit :

$$\lim_{n\to\infty} \max_{P_W} I(W; \boldsymbol{Z}) = 0, \tag{8}$$

où la distribution du message est arbitraire. Ceci contraste avec les travaux précédents dans lesquels le message est généralement supposé comme étant uniformément distribué.

Devant ce nombre croissant de mesures garantissant la non-decodabilité du message, un nouveau problème a émergé en théorie de l'information : le problème des communications furtives. Dans ce problème, l'émetteur et le récepteur doivent communiquer de façon à ce que l'adversaire ne puisse pas détecter la communication. L'adversaire a pour but non pas de décoder le message, comme dans le cas du canal wiretap, mais de détecter s'il y a une communication. Ainsi, pour distinguer les deux modèles, l'adversaire est nommé l'espion lorsqu'il a pour but de décoder le message, et le surveillant lorsqu'il a pour but de détecter la communication. La contrainte de furtivité est bien plus stricte que les contraintes de sécurité évoquées précédemment. La motivation pour l'étude de ce genre de problème vient entre autre des applications militaires pour lesquelles l'existence d'une transmission peut-être une information d'importance capitale.

2.1. Résultats théoriques principaux sur les communications furtives

Dans cette section, une description des résultats principaux sur les communications furtives est présentée. D'abord, les canaux point-à-point sont étudiés. Ensuite, les canaux point-à-point avec brouilleurs sont présentés. Enfin, les canaux à utilisateurs multiples et les résultats de codage pour les communications furtives sont exposés.

Canaux Point-à-Point

Les communications furtive ont été introduites en théorie de l'information par Bash *et al.* [7, 8]. Dans ces articles, les auteurs montrent que, comme en stéganographie, les communications furtives sur des canaux point-à-point Gaussiens obéissent à une loi en racine carrée. En d'autres termes, le nombre de bits d'information qui peuvent être envoyés en $n \in \mathbb{N}$ utilisations de canal est de l'ordre de la racine carrée de n. Pour montrer ce résultat, les auteurs utilisent un code qui utilise une clé de longueur $O(\sqrt{n}\log(n))$. Ces résultats montrent que la notion traditionnelle de débit ($R = \frac{\log(M)}{n}$, où $M \in \mathbb{N}$ est le nombre de messages) ne peut pas être utilisée puisque cette quantité tend vers zéro. Ainsi, dans les problèmes de communications furtives, la quantité qui est souvent étudiée est la fraction $\frac{\log(M)}{\sqrt{n}}$, qui tend vers une constante pour la plupart des canaux.

Plus tard, ces résultats ont été affinés par Che *et al.* [9, 10], Bloch [11] et Wang *et al.* [12]. Dans [9, 10], la constante exacte qui caractérise la limite du ratio $\frac{\log(M)}{\sqrt{n}}$ pour les canaux binaires symétriques est obtenue. Dans [11, 12], la limite du ratio $\frac{\log(M)}{\sqrt{n}}$ est caractérisée de manière exacte pour des canaux discrets sans mémoire, en utilisant des preuves différentes. De plus, [11] caractérise les conditions dans lesquelles aucune clé secrète n'est requise et présente, pour les canaux discrets sans mémoire, une amélioration de la longueur de clé lorsque son utilisation est nécessaire. Le procédé de communication qui y est introduit requiert une clé de longueur $O(\sqrt{n})$. Enfin, [12] caractérise également la limite exacte du ratio $\frac{\log(M)}{\sqrt{n}}$ pour

les canaux à bruit blanc Gaussien additif. De plus, Tahmasbi *et al.* ont obtenu dans [13] des bornes supérieures et inférieures sur le débit qui peut être atteint sous diverses contraintes de furtivité pour une longueur de code finie.

Les canaux Gaussiens à temps continu sont étudiés par Wang [14]. Dans cet article, il est montré que lorsque le bruit est blanc, un débit positif d'information furtive peut être atteint quand il n'y a pas de contraintes de bande passante sur l'entrée du canal. Au contraire, lorsqu'on s'intéresse au cas où la bande passante est limitée, les procédés de communication furtive requièrent que le nombre de bits transmis grandisse au plus comme la racine carrée du temps de transmission total.

Les canaux point-à-point à états ont été étudiés par Lee *et al.* [15, 16]. Dans ces articles, des expressions analytiques du débit maximum atteignable sous contrainte de furtivité sont présentées dans le cas où l'information sur l'état du canal est accessible de manière causale ou non causale pour les canaux discrets sans mémoire, et dans le cas où l'information sur l'état du canal est accessible de manière non causale pour les canaux Gaussiens. Curieusement, il existe des canaux à états pour lesquels le débit sous contrainte de furtivité est strictement positif dans le régime asymptotique. Ceci contraste avec les résultats obtenus pour les canaux point-à-point discrets sans mémoire ni état et les canaux point-à-point Gaussiens sans état.

Dans l'article [17], les canaux point-à-point non-cohérents sont étudiés par Tahmasbi *et al.* En particulier, des canaux à attenuation de Rayleigh rapide sont considérés. Il est montré dans cet article que la loi en racine carrée est encore vérifiée.

Dans les articles [18, 19, 20], Soltani *et al.* étudient les limites fondamentales de l'insertion de paquets furtifs. Les auteurs considèrent un premier émetteur envoyant des paquets sur un canal à destination d'un récepteur légitime dans un intervalle de temps donné. Une autre paire émetteur-récepteur est présente et son objectif est l'envoi de paquets – l'insertion de paquets – sur le même canal de manière non détectable pour un adversaire externe. Il est montré dans cet article que la loi en racine carrée est vérifiée. C'est à dire, si le nombre total de paquets envoyés par le premier émetteur est n, le nombre de paquets qui peuvent être insérés de manière furtive croît en racine carrée de n.

Dans l'article [21], Soltani *et al.* considèrent le problem d'insertion de bits dans des paquets. Les auteurs considèrent qu'un premier émetteur envoie n paquets à un premier récepteur. Les paquets sont relayés à ce premier récepteur par trois entités. La première a pour objectif d'insérer des bits furtifs dans les paquets, dont la charge utile est supposée non saturée. La seconde est un adversaire qui vise à détecter les bits insérés dans les paquets. La troisième a pour objectif de décoder les bits furtifs insérés dans les paquets. Il est montré dans cet article que si le nombre de paquets transmis est n, alors, il est possible d'insérer un nombre de bits dans ces paquets qui est de l'ordre de la racine carrée de n.

Les canaux point-à-point Poissoniens ont été étudiés par Wang [22]. Il montre que pour des canaux point-à-point Poissoniens à temps continu sans contrainte sur le pic de puissance en entrée, la capacité du canal sous contrainte de furtivité est infinie.

Tahmasbi *et al.* ont introduit les exposants d'erreur pour les communications furtives sur les canaux point-à-point [23]. Ils prouvent dans cet article des bornes supérieures et inférieures sur l'exposant d'erreur. Pour certains regimes de communication, ces bornes se rejoignent.

Dans l'article [24], Tahmasbi *et al.* étudient la faisabilité de la génération furtive de clé secrète. Il y est montré que la génération furtive de clé secrète est possible pour certains modèles. De plus, dans les modèles pour lesquels la capacité de clé secrète furtive égale la capacité du canal sous contrainte de furtivité, il est montré que la confidentialité de la clé est une conséquence naturelle. C'est à dire que, pour de tels modèles, si la furtivité du protocole

de génération de clé est assurée, alors la clé et la sortie de canal observée par l'adversaire sont statistiquement indépendantes dans le régime asymptotique.

Canaux Point-à-Point avec Brouilleurs

Les canaux point-à-point sous contrainte de furtivité avec un brouilleur coopératif ont été introduits par Soltani *et al.* [25, 26]. Dans l'article [25], l'étude considère plusieurs brouilleurs coopératifs voués à aider l'émetteur à réaliser une communications furtive tandis que plusieurs adversaires scrutent le canal. Les brouilleurs sont supposés aléatoirement localisés suivant une distribution de Poisson en deux dimensions et les adversaires sont supposés être distribués aléatoirement de manière uniforme et indépendante. Dans cette configuration, une amélioration de la loi en racine carrée peut être observée pour la plupart des canaux. En effet, il est montré dans cet article que cette amélioration dépend de la densité du processus ponctuel, du coefficient de propagation, et du nombre d'adversaires.

Dans les articles [27, 28], Zheng *et al.* considèrent les communications furtives en présence d'un brouilleur adverse. En présence d'un tel brouilleur, il est montré qu'indépendamment des caractéristiques du canal, une clé secrète de longueur $\Omega(\log(n))$ est requise. De plus, bien que le canal soit brouillé, la loi en racine carrée est toujours vérifiée.

Dans l'article [29], Sobers *et al.* étudient la présence d'un brouilleur coopérant non informé. Le brouilleur est non informé dans le sens où il n'est pas coordonné avec l'émetteur. Cet article révèle des scenarios dans lesquels la loi en racine carrée n'est plus vérifiée, et pour lesquels $O(n)$ bits peuvent être envoyés en n utilisations de canal.

Canaux à Utilisateurs Mutiples

Les communications furtives sur les canaux broadcast discrets sans mémoire ont été introduites dans [2, 3] par Arumugam *et al.* Le modèle consiste en un émetteur envoyant une information privée à un récepteur uniquement et une information commune aux deux récepteurs. La situation étudiée est celle dans laquelle le récepteur à qui le message privé n'est pas adressé doit être incapable de détecter la communication de ce message privé. Les codes pour la transmission du message privé et commun sont supposés conjointement conçus. Dans ce cas, les auteurs caractérisent de manière exacte la quantité d'information maximale qui peut être introduite dans le code qui transmet l'information commune. Ce résultat vérifie une nouvelle fois la loi en raciné carrée des communications furtives.

Tan *et al.* ont considéré un autre type de canal broadcast dans [30]. Leur modèle consiste en un émetteur qui envoie deux messages indépendants aux deux récepteurs en présence d'un adversaire qui veut détecter la communication. Dans ce cas, la loi en raciné carrée des communications furtives est vérifiée une nouvelle fois.

Dans le contexte des communications furtives, le canal à accès multiple a été étudié par Arumugam *et al.* [31, 32]. Le modèle est constitué de deux transmetteurs envoyant de l'information à un récepteur légitime en présence d'un adversaire qui veut détecter la communication. Dans l'article [31, 32], les auteurs caractérisent la région de capacité du canal à accès multiple à 2 et K utilisateurs sous contrainte de furtivité. Les auteurs montrent que pour ce genre de canal à accès multiple, la loi en raciné carrée des communications furtives est vérifiée.

Arumugam *et al.* ont également étudié le canal à relais sous contrainte de furtivité [33]. Les auteurs de cet article considèrent un canal à relais dégradé avec deux adversaires indépendants.

Un des adversaires est sur le lien entre l'émetteur et le relais tandis que le second est sur le lien entre le relais et le récepteur. Les auteurs caractérisent le ratio optimal $\frac{\log(M)}{\sqrt{n}}$ dans le régime asymptotique où n croît infiniment, et où $M \in \mathbb{N}$ et $n \in \mathbb{N}$ sont respectivement le nombre de messages et le nombre d'utilisations de canal. Dans un tel scenario, la loi en raciné carrée des communications furtives est vérifiée.

Coder pour les Communications Furtives

Bloch *et al.* [34] et Kadampot *et al.* [35, 36] ont commencé l'étude de la conception de codes pour réaliser des communications furtives sur des canaux point-à-point. Ils montrent que la modulation en position d'impulsion et des variations autour de cette modulation permettent d'atteindre des débits optimaux. Dans les articles [35, 36], la construction du code a l'avantage d'avoir une faible complexité algorithmique.

3. Résultat principal

Dans le contexte des canaux broadcast, deux types de problèmes de communications furtives ont été étudiés [30, 2, 3]. Dans l'article [30], l'émetteur veut envoyer deux messages furtifs indépendants à deux récepteurs. Dans les articles [2] et [3], l'émetteur envoie un message commun non furtif aux deux récepteurs, et essaie simultanément d'envoyer un message furtif à l'un des récepteurs. C'est-à-dire, l'autre récepteur ne doit pas pouvoir détecter si un message furtif à été envoyé ou non.

La présente étude est liée au problème présenté dans [2] et [3]. L'intérêt est porté sur le problème d'introduction d'information furtive dans un code broadcast non-furtif. Les différences principales entre le problème présenté ici et celui dans [2] et [3] sont :

- Dans les articles [2] et [3], le code broadcast non furtif et le code furtif sont construits conjointement par l'émetteur. Ceci autorise éventuellement le choix d'un code non furtif dans lequel il est facile d'introduire un code furtif. L'étude présente suppose que le code broadcast non furtif est donné et ne peut pas être changé, rendant la preuve de faisabilité plus difficile.[1]

- Dans les articles [2] et [3] il y a une contrainte de furtivité différente conditionnée sur les messages communs non furtifs. Dans la présente étude, une unique contrainte de furtivité sur l'ensemble du code est considérée. Cette différence complique considérablement la preuve du converse. La preuve d'un converse général utilisant une contrainte de furtivité mettant en jeu la divergence de Kullback-Leibler est encore un problème ouvert. Dans cette étude, la variation totale est utilisée comme critère de furtivité en adaptant des techniques de [13]. La borne du converse se montre proche de la borne de faisabilité pour une classe de canaux satisfaisant certaines propriétés de symétrie.

Il est montré que dans ce scenario, il est possible de transmettre $O(\sqrt{n})$ bits en n utilisations de canal en modifiant un code broadcast existant. De plus, il est montré que le débit atteignable est asymptotiquement optimal pour une classe de canaux broadcast discrets sans mémoire.

[1] Une condition technique est que le code non furtif donné ait un exposant d'erreur positif ; voir (4.52).

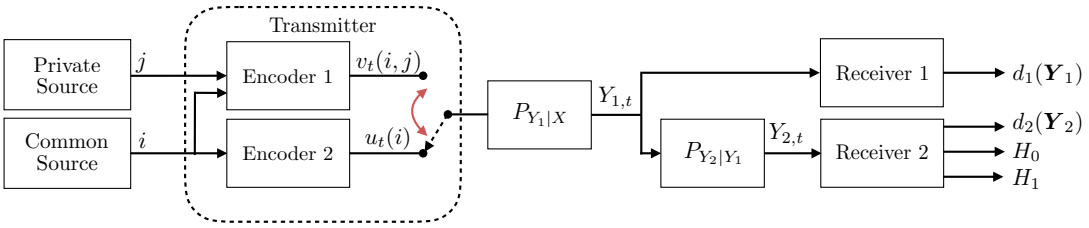

FIGURE 2. : Canal broadcast dégradé avec messages furtifs à l'utilisation de canal $t \in \{1, 2, \ldots, n\}$, où $d_1 : \mathcal{Y}_1^n \to \mathcal{W} \times \hat{\mathcal{W}}$ dénote la fonction de décodage au Récepteur 1 et $d_2 : \mathcal{Y}_2^n \to \mathcal{W}$ dénote la fonction de décodage au Récepteur 2.

3.1. Modèle

Soit un système de communication à trois utilisateurs dans lequel un émetteur envoie simultanément de l'information à deux récepteurs à travers un canal de communication. Dans cette étude, le canal est décrit comme suit :

$$(\mathcal{X}^n, \mathcal{Y}_1^n \times \mathcal{Y}_2^n, P_{\boldsymbol{Y}_1\boldsymbol{Y}_2|\boldsymbol{X}}), \tag{9a}$$

où $n \in \mathbb{N}$ est la durée de la communication en utilisations de canal (longueur de bloc) et les alphabets $\mathcal{X}$, $\mathcal{Y}_1$ et $\mathcal{Y}_2$ sont finis. Etant donné une entrée de canal $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, la sortie de canal $(\boldsymbol{y}_1, \boldsymbol{y}_2)$, avec $\boldsymbol{y}_k = (y_{k,1}, y_{k,2}, \ldots, y_{k,n})$ pour tout $k \in \{1, 2\}$, est observée au récepteur k avec probabilité :

$$P_{\boldsymbol{Y}_1\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}_1, \boldsymbol{y}_2|\boldsymbol{x}) \triangleq \prod_{t=1}^{n} P_{Y_1|X}(y_{1,t}|x_t) P_{Y_2|Y_1}(y_{2,t}|y_{1,t}). \tag{9b}$$

C'est-à-dire, le canal est dégradé et sans mémoire.

Etant donné le canal en (9a), l'émetteur utilise un *code broadcast* (Encoder 2 sur la Figure 2) pour transmettre un message destiné aux deux récepteurs à un débit fixé. Ce message est souvent appelé le *message commun*.

Chaque mot-code d'un code broadcast peut être modifié pour générer un nouvel ensemble de mot-codes. Ainsi, en redéfinissant les ensembles de décodage à un récepteur (Receiver 1 sur la Figure 2), il est possible de construire un nouveau code (Encoder 1 sur la Figure 2) qui transmet deux messages : *(a)* le message commun au même débit que le code broadcast original, éventuellement au prix d'une probabilité d'erreur de décodage accrue ; et *(b)* un message exclusivement destiné au récepteur 1. Ce message est souvent appelé le *message privé* et le nouveau code est appelé un *code induit*.

Un code induit peut satisfaire des contraintes additionnelles sur la transmission du message privé, e.g., une contrainte de furtivité, une contrainte de sécurité, une contrainte de transmission simultanée d'information et d'énergie, etc. Cette étude s'intéresse à une contrainte de furtivité consistant à rendre le second récepteur incapable de déterminer si un message privé est transmis ou non au premier récepteur. C'est-à-dire, le second récepteur est incapable de déterminer si le mot-code utilisé provient du code broadcast original ou du code induit. Un code induit satisfaisant un telle contrainte de furtivité est appelé un *code furtif*.

L'objectif de cette étude est de déterminer la quantité maximale d'information qui peut être transmise de façon furtive étant donné un code broadcast initial arbitraire.

Codes Broadcast

L'index du message commun, qui doit être transmis de l'émetteur aux deux récepteurs, est une réalisation d'une variable aléatoire W uniformément distribuée dans l'ensemble

$$\mathcal{W} \triangleq \{1, 2, \ldots, M\}, \tag{10}$$

où $M \in \mathbb{N}$. Pour envoyer un message commun en n utilisations de canal, l'émetteur utilise un (n, M, ϵ)-code broadcast.

Définition 1 ((n, M, ϵ)-code broadcast)**.** *Etant donné $M \in \mathbb{N}$, $\epsilon \in [0, 1]$ et une longueur de bloc $n \in \mathbb{N}$, un (n, M, ϵ)-code broadcast pour le canal en* (9) *est un système*

$$\left\{ \left(\boldsymbol{u}(1), \mathcal{D}_1(1), \mathcal{D}_2(1) \right), \left(\boldsymbol{u}(2), \mathcal{D}_1(2), \mathcal{D}_2(2) \right), \ldots, \left(\boldsymbol{u}(M), \mathcal{D}_1(M), \mathcal{D}_2(M) \right) \right\}, \tag{11}$$

qui vérifie pour tout $(i, j, k) \in \mathcal{W}^2 \times \{1, 2\}$, avec $i \neq j$:

$$\boldsymbol{u}(i) \triangleq (u_1(i), u_2(i), \ldots, u_n(i)) \in \mathcal{X}^n, \tag{12a}$$

$$\mathcal{D}_k(i) \cap \mathcal{D}_k(j) = \emptyset, \tag{12b}$$

$$\bigcup_{l=1}^{M} \mathcal{D}_k(l) \subseteq \mathcal{Y}_k^n, \qquad \textit{et} \tag{12c}$$

$$\frac{1}{M} \sum_{i=1}^{M} \Pr\left[\boldsymbol{Y}_k \in \mathcal{D}_k^{\mathsf{c}}(i) \middle| \boldsymbol{X} = \boldsymbol{u}(i) \right] \leqslant \epsilon. \tag{12d}$$

L'opérateur en (12d) s'applique avec la la distribution marginale $P_{\boldsymbol{Y}_k|\boldsymbol{X}}$ de la fonction de masse conjointe en (9b) ; et $\mathcal{D}_k^{\mathsf{c}}(i)$ en (12d) représente le complément de $\mathcal{D}_k(i)$ par rapport à l'ensemble $\mathcal{Y}_k^n$.

Etant donné un code broadcast représenté par le système en (11), l'émetteur utilise le mot-code $\boldsymbol{u}(i)$ pour transmettre l'indice de message $i \in \mathcal{W}$. A l'utilisation de canal t, où $t \in \{1, 2, \ldots, n\}$, l'émetteur envoie le symbole $u_t(i)$ à travers le canal. Après n utilisations de canal, le récepteur k, où $k \in \{1, 2\}$, observe la sortie de canal $\boldsymbol{y}_k = (y_{k,1}, y_{k,2}, \ldots, y_{k,n})$ et détermine que l'indice de message i a été transmis s'il satisfait la règle de décodage suivante :

$$\boldsymbol{y}_k \in \mathcal{D}_k(i). \tag{13}$$

La probabilité moyenne d'erreur de décodage au récepteur k associée au code broadcast donné, dénotée par λ_k, est donné dans le membre gauche de l'inéquation (12d).

Codes Induits

Supposons que le message privé est représenté par une variable aléatoire $\hat{W}$, indépendante de W et uniformément distribuée sur l'ensemble

$$\hat{\mathcal{W}} \triangleq \{1, 2, \ldots, \hat{M}\}, \tag{14}$$

où $\hat{M} \in \mathbb{N}$. Supposons également qu'un code broadcast dénoté par $\mathcal{C}$ est donné et est représenté par le système en (11). L'émetteur utilise un $(n, \mathcal{C}, \hat{M})$-code induit pour transmettre à la fois

le message commun et le message privé.

Définition 2 ($(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-code induit)**.** *Etant donné $\hat{M} \in \mathbb{N}$, $\hat{\epsilon} \in [0,1]$, et un (n, M, ϵ)-code broadcast $\mathcal{C}$ décrit par* (11)*, un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-code induit est un système*

$$\Big\{ \left(\boldsymbol{v}(1,1), \mathcal{D}_1(1,1), \mathcal{D}_2(1)\right), \left(\boldsymbol{v}(1,2), \mathcal{D}_1(1,2), \mathcal{D}_2(1)\right), \ldots, \left(\boldsymbol{v}(M,\hat{M}), \mathcal{D}_1(M,\hat{M}), \mathcal{D}_2(M)\right) \Big\} \tag{15}$$

qui vérifie pour tout $(i,k,j,l) \in \mathcal{W}^2 \times \hat{\mathcal{W}}^2$, avec $(i,j) \neq (k,l)$:

$$\boldsymbol{v}(i,j) \triangleq (v_1(i,j), v_2(i,j), \ldots, v_n(i,j)) \in \mathcal{X}^n, \tag{16a}$$

$$\mathcal{D}_1(i,j) \cap \mathcal{D}_1(k,l) = \emptyset, \tag{16b}$$

$$\bigcup_{p=1}^{M} \bigcup_{q=1}^{\hat{M}} \mathcal{D}_1(p,q) \subseteq \mathcal{Y}_1^n, \tag{16c}$$

$$\frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i,j) | \boldsymbol{X} = \boldsymbol{v}(i,j)\right] \leqslant \hat{\epsilon}, \tag{16d}$$

$$\frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \Pr\left[\boldsymbol{Y}_2 \in \mathcal{D}_2^{\mathsf{c}}(i) | \boldsymbol{X} = \boldsymbol{v}(i,j)\right] \leqslant \hat{\epsilon}. \tag{16e}$$

Les opérateurs en (16d) et (16e) s'appliquent respectivement avec les distributions marginales conditionnelles $P_{\boldsymbol{Y}_1|\boldsymbol{X}}$ et $P_{\boldsymbol{Y}_2|\boldsymbol{X}}$ de la fonction de masse conjointe en (9b). Les ensembles $\mathcal{D}_1^{\mathsf{c}}(i,j)$ et $\mathcal{D}_2^{\mathsf{c}}(i)$ représentent respectivement les compléments des ensembles $\mathcal{D}_1(i,j)$ et $\mathcal{D}_2(i)$ par rapport à $\mathcal{Y}_1^n$ et $\mathcal{Y}_2^n$.

Etant donné un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-code induit, dénoté $\hat{\mathcal{C}}$ et décrit par (15), l'émetteur utilise le mot-code $\boldsymbol{v}(i,j)$ pour transmettre le message commun $i \in \mathcal{W}$ et le message privé $j \in \hat{\mathcal{W}}$. A l'utilisation de canal t, où $t \in \{1, 2, \ldots, n\}$, l'émetteur envoie les symbole $v_t(i,j)$ à travers le canal. Après n utilisations de canal, le récepteur k observe la sortie de canal $\boldsymbol{y}_k = (y_{k,1}, y_{k,2}, \ldots, y_{k,n})$, où $k \in \{1,2\}$. Le récepteur 1 déclare que la paire $(i,j) \in \mathcal{W} \times \hat{\mathcal{W}}$ a été transmise si (i,j) vérifie la règle de décodage suivante :

$$\boldsymbol{y}_1 \in \mathcal{D}_1(i,j). \tag{17}$$

Le récepteur 2 détermine que le message commun i a été transmis s'il vérifie (13), avec $k = 2$, *i.e.*, le récepteur 2 utilise la même règle de décodage que le code broadcast initial $\mathcal{C}$.
La probabilité moyenne d'erreur de décodage au récepteur k associée au code induit $\hat{\mathcal{C}}$ est notée $\hat{\lambda}_k$. Les membres gauches de (16d) et (16e) définissent respectivement $\hat{\lambda}_1$ et $\hat{\lambda}_2$.

Remarque 1. *Pour garantir qu'il existe un message $i \in \mathcal{W}$ satisfaisant la règle de décodage en* (13) *pour tout $\boldsymbol{y}_k \in \mathcal{Y}_k^n$, avec $k \in \{1,2\}$, l'inclusion en* (12c) *est supposée vérifiée avec égalité. Dans le cas où l'ensemble $\mathcal{Y}_k^n \setminus (\mathcal{D}_k(1) \cup \mathcal{D}_k(2) \cup \ldots \cup \mathcal{D}_k(M))$ n'est pas vide, les vecteurs de sorties de canal qui appartiennent à cet ensemble induisent toujours une erreur de décodage au récepteur k. Ainsi, pour tout $j \in \mathcal{W}$, remplacer l'ensemble $\mathcal{D}_k(j)$ par $\mathcal{D}'_k(j) = \mathcal{D}_k(j) \cup (\mathcal{Y}_k^n \setminus (\mathcal{D}_k(1) \cup \mathcal{D}_k(2) \cup \ldots \cup \mathcal{D}_k(M)))$ n'augmente pas la probabilité d'erreur moyenne. Donc, il n'y a pas de perte de généralité en étudiant un système dans lequel l'équation* (12c) *est vérifiée avec égalité. Sans perte de généralité, l'inclusion en* (16c) *est supposée vérifiée avec égalité pour des raisons analogues.*

Codes Furtifs

Soient un (n, M, ϵ)-code broadcast décrit par (11) et dénoté $\mathcal{C}$ et un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-code induit décrit par (15) et dénoté $\hat{\mathcal{C}}$. Pour tout $k \in \{1, 2\}$, les fonctions de masses $Q_{\boldsymbol{Y}_k}$ et $R_{\boldsymbol{Y}_k}$ sont respectivement les distributions du vecteur de sorties de canal $\boldsymbol{Y}_k$ quand le code broadcast $\mathcal{C}$ est utilisé et quand le code induit $\hat{\mathcal{C}}$ est utilisé. C'est-à-dire, pout tout $\boldsymbol{y} \in \mathcal{Y}_k^n$,

$$Q_{\boldsymbol{Y}_k}(\boldsymbol{y}) \triangleq \frac{1}{M} \sum_{i=1}^{M} P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i)), \text{ et} \tag{18}$$

$$R_{\boldsymbol{Y}_k}(\boldsymbol{y}) \triangleq \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}(i,j)), \tag{19}$$

où $P_{\boldsymbol{Y}_k|\boldsymbol{X}}$ est la distribution marginale de la fonction de masse conjointe en (9b). Soit un test d'hypothèses dans lequel le récepteur 2 vise à déterminer si le code broadcast $\mathcal{C}$ est utilisé (hypothèse H_0) ou si le code induit $\hat{\mathcal{C}}$ est utilisé (hypothèse H_1) d'après son observation de la sortie de canal $\boldsymbol{Y}_2$:

$$\begin{cases} H_0 : \boldsymbol{Y}_2 \sim Q_{\boldsymbol{Y}_2} \\ H_1 : \boldsymbol{Y}_2 \sim R_{\boldsymbol{Y}_2}, \end{cases} \tag{20}$$

où $Q_{\boldsymbol{Y}_2}$ et $R_{\boldsymbol{Y}_2}$ sont respectivement données en (18) et (19).

Notons $\alpha \in [0, 1]$ et $\beta \in [0, 1]$ les probabilité d'erreur de type I et de type II associées au test $T : \mathcal{Y}_2^n \to \{0, 1\}$ de la forme

$$T(\boldsymbol{y}) \triangleq \begin{cases} 0 & \text{si } H_0 \text{ est acceptée,} \\ 1 & \text{si } H_1 \text{ est acceptée.} \end{cases} \tag{21}$$

C'est-à-dire,

$$\alpha \triangleq \Pr\left[T(\boldsymbol{Y}_2) = 1\right], \text{ and} \tag{22}$$

$$\beta \triangleq \Pr\left[T(\boldsymbol{Y}_2) = 0\right], \tag{23}$$

où l'opérateur en (22) s'applique avec $\boldsymbol{Y}_2 \sim Q_{\boldsymbol{Y}_2}$ et l'opérateur en (23) s'applique avec $\boldsymbol{Y}_2 \sim R_{\boldsymbol{Y}_2}$.

D'après [37, Theorem 13.1.1], il est vérifié que

$$\alpha + \beta \geqslant 1 - \|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}}, \tag{24}$$

avec égalité pour le test optimal, et où α et β sont respectivement définis en (22) et (23), pour tout test $T : \mathcal{Y}_2^n \to \{0, 1\}$ de la forme (21).
D'après l'inégalité (24), il suit que plus la variation totale $\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}}$ est petite, plus la probabilité d'échouer à déterminer si le code broadcast ou si le code induit est utilisé est élevée. Ainsi, un code furtif est défini comme suit.

Définition 3 ($(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-code furtif)**.** *Etant donné $\delta \in [0, 1]$ et un (n, M, ϵ)-code broadcast $\mathcal{C}$ décrit par* (11)*, un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-code induit décrit par* (15) *est un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-code furtif si*

$$\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} \leqslant \delta, \tag{25}$$

où $Q_{\boldsymbol{Y}_2}$ *et* $R_{\boldsymbol{Y}_2}$ *sont respectivement définies en* (18) *et* (19).

Dans la suite de cette étude, il est supposé que les codes induits vérifient $R_{\boldsymbol{Y}_2} \ll Q_{\boldsymbol{Y}_2}$. Sinon, la transmission d'information privée de manière furtive est impossible pour certaines valeurs de $\delta \in [0, 1]$.
Enfin, l'analyse est restreinte aux code-induits qui vérifient $R_{\boldsymbol{Y}_2} \neq Q_{\boldsymbol{Y}_2}$. Ceci garantit qu'il n'existe pas de code induit qui puisse imiter parfaitement la fonction de masse $Q_{\boldsymbol{Y}_2}$ de la sortie de canal induite par le code broadcast au récepteur 2. Sinon, le problème est trivial et la transmission de manière furtive est toujours faisable.

Le taux de transmission auquel l'information peut être envoyée au récepteur 1 de manière furtive en utilisant un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-code furtif est $\frac{\log_2(\hat{M})}{n}$ bits par utilisation de canal. Ainsi, étant donné le code broadcast initial $\mathcal{C}$, une limite fondamentale sur le débit auquel l'information peut être transmise de manière furtive est donnée par le plus grand $\hat{M}$ pour lequel il existe un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-code furtif. Cette notion est formalisée par la définition suivante.

Définition 4 (Taille de code furtif maximale)**.** *Etant donné une paire* $(\hat{\epsilon}, \delta) \in [0, 1]^2$ *et un* (n, M, ϵ)*-code broadcast* $\mathcal{C}$*, la taille de code furtif maximale, dénotée* $\hat{M}^*(n, \mathcal{C}, \hat{\epsilon}, \delta)$*, est :*

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

3.2. Faisabilité des communications furtives

Dans cette section, étant donné un (n, M, ϵ)-code broadcast noté $\mathcal{C}$, une borne inférieure sur la taille maximale sur code furtif (Définition 4) est établie en adaptant des techniques présentées dans [11] et [12]. La construction de ce résultat est présentée en trois parties. Dans la première partie, une fonction de masse est choisie pour générer aléatoirement un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-code induit. Souvent cette fonction de masse est appelée la *distribution generatrice*. Cette distribution est exprimée en fonction de paramètres appelés *paramètres générateurs*. Dans la deuxième partie, les paramètres générateurs sont choisis de manière à satisfaire la contrainte de furtivité en (25) pour un δ fixé, ce qui permet l'obtention d'un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-code furtif. Dans la troisième partie, il est montré que les probabilités moyennes d'erreur de décodage (notées $\hat{\Lambda}_k$, avec $k \in \{1, 2\}$) associées au code furtif admettent chacune une borne supérieure. Ces bornes sont exprimées en fonction des paramètres générateurs, prouvant que le $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-code furtif vérifie $\hat{\Lambda}_k < \hat{\epsilon}$ pour tout $k \in \{1, 2\}$.

Partie I : Génération aléatoire du code induit

Soient un (n, M, ϵ)-code broadcast $\mathcal{C}$ pour le canal en (9), décrit par le système en (11), $\hat{M} \in \mathbb{N}$; $K \in [0, \sqrt{n}]$ des paramètres ; et $\tilde{P}_{\hat{X}|X}$ une distribution conditionnelle telle que pour tout $x \in \mathcal{X}$,

$$\operatorname{supp} \tilde{P}_{\hat{X}|X=x} \subseteq \mathcal{X} \setminus \{x\}. \tag{26}$$

Etant donnés le paramètre K et la distribution $\tilde{P}_{\hat{X}|X}$, soit $P_{\hat{X}|X}$ une fonction de masse conditionnelle telle que pour tout $(x, \hat{x}) \in \mathcal{X}^2$,

$$P_{\hat{X}|X}(\hat{x}|x) \triangleq (1 - \theta)\, \mathbb{1}_{\{x=\hat{x}\}} + \theta \tilde{P}_{\hat{X}|X}(\hat{x}|x), \tag{27}$$

with

$$\theta \triangleq \frac{K}{\sqrt{n}}. \tag{28}$$

Souvent, les paramètres $\hat{M}, K$ et $\tilde{P}_{\hat{X}|X}$ sont appelés les paramètres générateurs.

Le dictionnaire d'un code induit est obtenu en générant pour tout $i \in \{1, 2, \ldots, M\}$, les $\hat{M}$ mot-codes

$$\boldsymbol{v}(i,1), \boldsymbol{v}(i,2), \ldots, \boldsymbol{v}(i,\hat{M}). \tag{29}$$

Pour tout $j \in \{1, 2, \ldots, \hat{M}\}$, le mot-code $\boldsymbol{v}(i,j)$ est la réalisation d'une variable aléatoire distribuée selon la fonction de masse $P_{\hat{\boldsymbol{X}}|\boldsymbol{X}=\boldsymbol{u}(i)}$ qui vérifie pour tout $\hat{\boldsymbol{x}} \in \mathcal{X}^n$,

$$P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}\left(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)\right) \triangleq \prod_{t=1}^{n} P_{\hat{X}|X}\left(\hat{x}_t|u_t(i)\right), \tag{30}$$

où $\boldsymbol{u}(1)$, $\boldsymbol{u}(2)$, …, $\boldsymbol{u}(M)$ sont les mot-codes du code broadcast initial $\mathcal{C}$. Dans la suite de cette preuve, la fonction de masse $P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}$ est appelée la distribution génératrice.

Pour terminer la génération du $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-code induit, les ensembles de décodage doivent être spécifiés. Le récepteur 2 utilise les ensembles de décodage

$$\mathcal{D}_2(1), \mathcal{D}_2(2), \ldots, \mathcal{D}_2(M) \tag{31}$$

du code broadcast initial $\mathcal{C}$, et la règle de décodage en (13), avec $k = 2$.

Pour tout $(\boldsymbol{x}, \hat{\boldsymbol{x}}, \boldsymbol{y}) \in \mathcal{X}^{2n} \times \mathcal{Y}_k^n$ et tout $k \in \{1, 2\}$, définissons $\imath_k(\hat{\boldsymbol{x}}; \boldsymbol{y}|\boldsymbol{x})$ comme suit :

$$\imath_k(\hat{\boldsymbol{x}}; \boldsymbol{y}|\boldsymbol{x}) \triangleq \log_2 \left(\frac{P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{\sum\limits_{\boldsymbol{x}' \in \mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{x}'|\boldsymbol{x}) P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}')} \right). \tag{32}$$

A la réception de la sortie de canal $\boldsymbol{y} \in \mathcal{Y}_1^n$, le récepteur 1 déclare que la paire $(i,j) \in \mathcal{W} \times \hat{\mathcal{W}}$ a été transmise d'après la règle de décodage en (17), où

$$\mathcal{D}_1(i,j) = \left\{ \boldsymbol{y} \in \mathcal{D}_1(i) : \imath_1\Big(\boldsymbol{v}(i,j), \boldsymbol{y}|\boldsymbol{u}(i)\Big) \geqslant n\eta \right\} \setminus \bigcup_{k<j} \mathcal{D}_1(i,j), \tag{33}$$

avec $\eta \in \mathbb{R}$ un paramètre dont la valeur exacte sera précisée plus tard. Les mot-codes en (29) et les ensembles de décodage en (31) et (33) forment un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-code induit.

Partie II : Analyse de furtivité

Cette partie s'intéresse à l'obtention de conditions sur les paramètres générateurs qui garantissent que le $(n, \mathcal{C}, \hat{M}, L, \hat{\epsilon})$-code induit généré est un $(n, \mathcal{C}, \hat{M}, L, \hat{\epsilon}, \delta)$-code furtif.

Soient $Q_{W\boldsymbol{Y}_2}$ et $S_{W\boldsymbol{Y}_2}$ deux fonctions de masse telles que pour tout $(i, \boldsymbol{y}) \in \mathcal{W} \times \mathcal{Y}_2^n$,

$$Q_{W\boldsymbol{Y}_2}(i, \boldsymbol{y}) \triangleq \frac{1}{M} Q_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i), \text{ et} \tag{34}$$

$$S_{W\boldsymbol{Y}_2}(i, \boldsymbol{y}) \triangleq \frac{1}{M} S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i), \tag{35}$$

où

$$Q_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i) \triangleq \prod_{t=1}^{n} P_{Y_2|X}(y_t|u_t(i)), \text{ et} \tag{36}$$

$$S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i) \triangleq \prod_{t=1}^{n} \sum_{\hat{x}\in\mathcal{X}} P_{\hat{X}|X}(\hat{x}|u_t(i)) P_{Y_2|X}(y_t|\hat{x}). \tag{37}$$

Notons $\hat{\Lambda}_k$, où $k \in \{1,2\}$, la probabilité moyenne d'erreur de décodage au récepteur k sur tous les dictionnaires possibles. Le lemme suivant établit une borne supérieure sur la variation totale $\|Q_{W\boldsymbol{Y}_2} - S_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}}$.

Lemme 1. *Etant donné un (n, M, ϵ)-code broadcast $\mathcal{C}$ décrit par* (11)*, la variation totale $\|Q_{W\boldsymbol{Y}_2} - S_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}}$ vérifie*

$$\|Q_{W\boldsymbol{Y}_2} - S_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} \leqslant \|Q_{\boldsymbol{Y}_2} - S_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} + \epsilon + \hat{\Lambda}_2, \tag{38}$$

où les fonctions de masse $Q_{\boldsymbol{Y}_2}$, $Q_{W\boldsymbol{Y}_2}$ et $S_{W\boldsymbol{Y}_2}$ sont respectivement définies en (18)*,* (34) *et* (35)*, et $S_{\boldsymbol{Y}_2}(\boldsymbol{y}) = \sum_{i=1}^{M} S_{W\boldsymbol{Y}_2}(i, \boldsymbol{y})$, pour tout $\boldsymbol{y} \in \mathcal{Y}_2^n$.*

Preuve: La preuve du Lemme 1 est présentée dans l'Annexe F. ■

La proposition suivante décrit les conditions sur les paramètres générateurs qui garantissent que le $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-code induit généré est un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-code furtif.

Proposition 1. *Un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-code induit est un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-code furtif si*

$$\theta \leqslant \frac{2Q^{-1}\left(\frac{1-\delta-\epsilon-\hat{\epsilon}+\sqrt{c_n}-\frac{c}{\sqrt{n}}}{2}\right)}{\sqrt{n \sum_{x\in\mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x})}}, \tag{39}$$

où c est une constante positive et $c_n = 2^{-b\sqrt{n}} + 2n \log_2\left(\frac{2}{\mu_0}\right) \exp\left(-a\sqrt{n}\right)$ avec a et b deux constantes positives et $\mu_0 = \min_{(x,y)\in\mathcal{X}\times\mathcal{Y}_2} \sum_{\hat{x}\in\mathcal{X}} P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x})$.

Preuve: La preuve de la Proposition 1 est présentée dans l'Annexe G et utilise des éléments de preuve provenant de [13, Lemma 8]. ■

Partie III : Analyse de la probabilité d'erreur de décodage

Soit $\tilde{R}_{Y_k|X}(y|x)$ la fonction de masse telle que pour tout $k \in \{1,2\}$ et pour tout $(x, y) \in \mathcal{X} \times \mathcal{Y}_k$,

$$\tilde{R}_{Y_k|X}(y|x) \triangleq \sum_{\hat{x}\in\mathcal{X}} \tilde{P}_{\hat{X}|X}(\hat{x}|x) P_{Y_k|X}(y|\hat{x}). \tag{40}$$

Définissons également $\bar{D}(\tilde{P}_{\hat{X}|X})$ et $\bar{\chi}_{2,k}(\tilde{P}_{\hat{X}|X})$, avec $k \in \{1,2\}$, respectivement comme suit :

$$\bar{D}_1(\tilde{P}_{\hat{X}|X}) \triangleq \sum_{x\in\mathcal{X}} \sum_{\hat{x}\in\mathcal{X}} \bar{P}_X(x) \tilde{P}_{\hat{X}|X}(\hat{x}|x) D\big(P_{Y_1|X=\hat{x}} || P_{Y_1|X=x}\big), \tag{41}$$

et

$$\bar{\chi}_{2,k}(\tilde{P}_{\hat{X}|X}) \triangleq \sum_{x\in\mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_k|X=x}, P_{Y_k|X=x}). \tag{42}$$

Proposition 2. *Soit un (n, M, ϵ)-code broadcast $\mathcal{C}$ pour le canal en (9). Il existe toujours un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-code furtif qui vérifie pour tout $\xi > 0$ arbitrairement petit*

$$\begin{aligned}\frac{\log_2\left(\hat{M}\right)}{n} &\geqslant \max_{\theta, \tilde{P}_{\hat{X}|X}} (1-\xi)\theta \bar{D}_1(\tilde{P}_{\hat{X}|X}) \\ &= \max_{\tilde{P}_{\hat{X}|X}} (1-\xi) \frac{2Q^{-1}\left(\frac{1-\delta-\epsilon-\hat{\epsilon}+\sqrt{c_n}-\frac{c}{\sqrt{n}}}{2}\right)}{\sqrt{n\bar{\chi}_{2,2}(\tilde{P}_{\hat{X}|X})}} \bar{D}_1(\tilde{P}_{\hat{X}|X}) \end{aligned} \tag{43}$$

Preuve: La preuve de la Proposition 2 est présentée dans l'Annexe J. ∎

Dans le régime asymptotique où la longueur de bloc n tend vers l'infini, la Proposition 2 conduit au théorème suivant.

Théorème 1. *Soit une séquence $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$, ... de (n, M_n, ϵ_n)-codes broadcast pour le canal en (9), avec $n \in \{1, 2, \ldots\}$ et*

$$\epsilon_n \leqslant \exp(-\zeta n), \tag{44}$$

pour un réel positif fixe ζ. Alors, il existe toujours une séquence de $(n, \mathcal{C}_n, \hat{M}_n, \hat{\epsilon}_n, \delta)$-codes furtifs vérifiant $\lim_{n\to\infty} \hat{\epsilon}_n = 0$ telle que pour tout $\xi > 0$ arbitrairement petit,

$$\lim_{n\to\infty} \frac{\log_2\left(\hat{M}_n^*(n, \mathcal{C}_n, \hat{\epsilon}_n, \delta)\right)}{\sqrt{n}} \geqslant \max_{\tilde{P}_{\hat{X}|X}} (1-\xi) \frac{2Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{\bar{\chi}_{2,2}(\tilde{P}_{\hat{X}|X})}} \bar{D}(\tilde{P}_{\hat{X}|X}). \tag{45}$$

Preuve: Soient une séquence infinie de réels positifs $K_1 < K_2 < K_3, \ldots$ et une séquence infinie de réels $\hat{\epsilon}_1 > \hat{\epsilon}_2 > \ldots > 0$, telles que pour tout $n \in \mathbb{N}$,

$$K_n \triangleq \frac{2Q^{-1}\left(\frac{1-\delta-\epsilon_n-\hat{\epsilon}_n+\sqrt{c_n}-\frac{c}{\sqrt{n}}}{2}\right)}{\sqrt{\bar{\chi}_{2,2}(\tilde{P}_{\hat{X}|X})}}. \tag{46}$$

En particulier, pour tout $n \in \mathbb{N}$, il est vrai que

$$K_n < \frac{2Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{\bar{\chi}_{2,2}(\tilde{P}_{\hat{X}|X})}}. \tag{47}$$

Notons que si ζ en (44) satisfait la condition

$$\begin{aligned}\zeta > \max\Bigg\{ &\max_{\tilde{P}_{\hat{X}|X}} \ln\left(1 + \frac{2Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{n\bar{\chi}_{2,2}(\tilde{P}_{\hat{X}|X})}} \max_{(x,y)\in\mathcal{X}\times\mathcal{Y}_2} \frac{\tilde{R}_{Y_2|X}(y|x) - P_{Y_2|X}(y|x)}{P_{Y_2|X}(y|x)}\right), \\ &\max_{\tilde{P}_{\hat{X}|X}} \ln\left(1 + \frac{2Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{n\bar{\chi}_{2,2}(\tilde{P}_{\hat{X}|X})}} \max_{(x,y)\in\mathcal{X}\times\mathcal{Y}_2} \frac{\tilde{R}_{Y_1|X}(y|x) - P_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)}\right)\Bigg\} \end{aligned} \tag{48}$$

où l'optimisation prend en compte toutes les distributions conditionnelles $\tilde{P}_{\hat{X}|X}$ possibles, alors il découle de la Proposition 2 que pour un n fixé, il existe toujours un $(n, \mathcal{C}_n, \hat{M}_n, \hat{\epsilon}_n, \delta)$-code

furtif tel que

$$\frac{\log_2\left(\hat{M}_n\right)}{\sqrt{n}} \geqslant (1-\xi)K_n\bar{D}_1(\tilde{P}_{\hat{X}|X}). \tag{49}$$

Dans le régime asymptotique, la condition en (48) est vérifiée pour tout $\zeta > 0$, ce qui implique directement

$$\lim_{n\to\infty} \frac{\log_2\left(\hat{M}_n\right)}{\sqrt{n}} \geqslant (1-\xi)\frac{2Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{\bar{\chi}_{2,2}(\tilde{P}_{\hat{X}|X})}}\bar{D}_1(\tilde{P}_{\hat{X}|X}). \tag{50}$$

La preuve est complétée en optimisant le membre de droite de l'inégalité (50) sur toutes les fonctions de masse $\tilde{P}_{\hat{X}|X}$ possibles. ■

3.3. Impossibilité des communications furtives

Etant donné un (n, M, ϵ)-code broadcast $\mathcal{C}$, cette section présente une borne supérieure sur la ratio entre la taille de code furtif maximale $\hat{M}^*(n, \mathcal{C}, \hat{\epsilon}, \delta)$ et la racine carrée de la longueur de bloc, *i.e.*, $\frac{\log_2\left(\hat{M}^*(n,\mathcal{C},\hat{\epsilon},\delta)\right)}{\sqrt{n}}$, dans le régime asymptotique où la longueur de bloc tend vers l'infini. La section suivante présente des résultats préliminaires en longueur de bloc finie qui sont cruciaux pour prouver le résultat principal de cette section.

Résultats auxiliaires

Un des paramètres centraux pour décrire un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-code induit $\hat{\mathcal{C}}$ décrit par (15) est le nombre de fois qu'un composant d'un mot-code $\boldsymbol{u}(i)$ du code $\mathcal{C}$ diffère du composant correspondant dans le mot-code induit $\boldsymbol{v}(i, j)$ du code $\hat{\mathcal{C}}$, où $(i, j) \in \mathcal{W} \times \hat{\mathcal{W}}$. Cette quantité est appelée le *poids du mot-code* $\boldsymbol{v}(i, j)$. Un autre paramètre d'intérêt est le nombre de fois où le symbole $x \in \mathcal{X}$ apparaît dans les mot-codes de $\mathcal{C}$ and n'apparaît pas dans les composants correspondant des mot-codes de $\hat{\mathcal{C}}$. Cette quantité est appelée le *poids du symbole* x.

Définition 5 (Poids)**.** *Etant donné un (n, M, ϵ)-code broadcast $\mathcal{C}$ représenté par le système en* (11)*, soit un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-code induit $\hat{\mathcal{C}}$ représenté par le système en* (15)*. Pour tout $(i, j) \in \mathcal{W} \times \hat{\mathcal{W}}$, le poids du mot-code $\boldsymbol{v}(i, j)$, noté $\omega(i, j)$, est :*

$$\omega(i, j) \triangleq \sum_{t=1}^{n} \mathbb{1}_{\{u_t(i)\neq v_t(i,j)\}}. \tag{51}$$

Pour tout $x \in \mathcal{X}$, le poids du symbole x, noté $\omega(x)$, est :

$$\omega(x) \triangleq \frac{1}{M\hat{M}} \sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\sum_{t=1}^{n} \mathbb{1}_{\{u_t(i)=x\}}\mathbb{1}_{\{u_t(i)\neq v_t(i,j)\}}. \tag{52}$$

Les codes $\mathcal{C}$ et $\hat{\mathcal{C}}$ induisent différentes fonctions de masse pertinentes pour l'analyse des codes furtifs. Ces fonctions sont définies ci-dessous.

Définition 6 (Distributions empiriques)**.** *Etant donné un (n, M, ϵ)-code broadcast $\mathcal{C}$ représenté par le système en* (11)*, soit un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-code induit $\hat{\mathcal{C}}$ représenté par le système en* (15)*. Pour tout $(x, \hat{x}) \in \mathcal{X}^2$,*

- *la fonction de masse empirique induite à l'entrée du canal par le code broadcast $\mathcal{C}$, notée $\bar{P}_X$, est*

$$\bar{P}_X(x) \triangleq \frac{1}{nM} \sum_{i=1}^{M} N(x|\boldsymbol{u}(i)); \tag{53}$$

- *la fonction de masse empirique conjointe induite à l'entrée du canal par les deux codes $\mathcal{C}$ et $\hat{\mathcal{C}}$ sur $\mathcal{C}$ and $\hat{\mathcal{C}}$, notée $\bar{P}_{X\hat{X}}$, est*

$$\bar{P}_{X\hat{X}}(x,\hat{x}) \triangleq \frac{1}{nM\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} N(x,\hat{x}|\boldsymbol{u}(i),\boldsymbol{v}(i,j)); \tag{54}$$

- *la probabilité empirique avec laquelle un symbole x dans un mot-code de $\mathcal{C}$ est changé un en autre symbole $\hat{x} \neq x$ dans un mot-code de $\hat{\mathcal{C}}$, notée $\hat{P}_{\hat{X}|X}$, est*

$$\hat{P}_{\hat{X}|X}(\hat{x}|x) \triangleq \frac{\sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{t=1}^{n} \mathbb{1}_{\{x=u_t(i)\}} \mathbb{1}_{\{\hat{x}=v_t(i,j)\}} \mathbb{1}_{\{x\neq\hat{x}\}}}{\sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{t=1}^{n} \mathbb{1}_{\{x=u_t(i)\}} \mathbb{1}_{\{u_t(i)\neq v_t(i,j)\}}}, \tag{55}$$

et

$$\operatorname{supp} \hat{P}_{\hat{X}|X=x} \subseteq \mathcal{X} \setminus \{x\}; \tag{56}$$

- *la probabilité empirique avec laquelle un symbole x dans un mot-code de $\mathcal{C}$ est changé pour n'importe quel autre symbole dans un mot-code de $\hat{\mathcal{C}}$, notée $\theta(x)$, est*

$$\theta(x) \triangleq 1 - \bar{P}_{\hat{X}|X}(x|x), \tag{57}$$

où $\bar{P}_{\hat{X}|X}(x|x)$ est telle que

$$\bar{P}_{\hat{X}X}(x,x) = \bar{P}_X(x)\bar{P}_{\hat{X}|X}(x|x). \tag{58}$$

Le lemme suivant établit des relations entre les fonctions de masse empiriques et les poids.

Lemme 2. *Etant donné un (n, M, ϵ)-code broadcast $\mathcal{C}$ représenté par le système en* (11)*, soit un $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-code induit $\hat{\mathcal{C}}$ représenté par le système en* (15)*. Toute paire $(x, \hat{x}) \in \mathcal{X}^2$ vérifie*

$$\bar{P}_{X\hat{X}}(x,\hat{x}) = \bar{P}_X(x) \left((1-\theta(x))\, \mathbb{1}_{\{x=\hat{x}\}} + \theta(x) \hat{P}_{\hat{X}|X}(\hat{x}|x) \right), \tag{59}$$

$$\omega(x) = n\bar{P}_X(x)\theta(x), \quad \textit{et} \tag{60}$$

$$n \sum_{x\in\mathcal{X}} \bar{P}_X(x)\theta(x) = \sum_{x\in\mathcal{X}} \omega(x) = \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \omega(i,j). \tag{61}$$

Preuve: La preuve du Lemme 2 est présentée dans l'Annexe K. ∎

Soient $Q_{W\boldsymbol{Y}_2}$ et $R_{W\boldsymbol{Y}_2}$, deux fonctions de masse telles que pour tout $(i, \boldsymbol{y}) \in \mathcal{W} \times \mathcal{Y}_2^n$,

$$Q_{W\boldsymbol{Y}_2}(i, \boldsymbol{y}) \triangleq \frac{1}{M} Q_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i), \text{ et} \tag{62}$$

$$R_{W\boldsymbol{Y}_2}(i, \boldsymbol{y}) \triangleq \frac{1}{M} R_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i), \tag{63}$$

où

$$Q_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i) \triangleq \prod_{t=1}^{n} P_{Y_2|X}(y_t|u_t(i)), \text{ et} \tag{64}$$

$$R_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i) \triangleq \frac{1}{\hat{M}} \sum_{j=1}^{\hat{M}} \prod_{t=1}^{n} P_{Y_2|X}(y_t|v_t(i,j)). \tag{65}$$

Les distributions marginales $Q_{\boldsymbol{Y}_2}$ et $R_{\boldsymbol{Y}_2}$ sont respectivement en (18) et (19).

Le lemme suivant révèle que remplacer la contrainte $\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} < \delta$ en (25) par la contrainte $\|Q_{W\boldsymbol{Y}_2} - R_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} < \delta$ est équivalent à une constante additive près.

Lemme 3. *Etant donné un (n, M, ϵ)-code broadcast $\mathcal{C}$ décrit par* (11)*, tout $(n, \mathcal{C}, \hat{M}, L, \hat{\epsilon})$-code induit décrit par* (15) *vérifie*

$$\|Q_{W\boldsymbol{Y}_2} - R_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} \leqslant \|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} + \epsilon + \hat{\epsilon}, \tag{66}$$

où les fonctions de masse $Q_{\boldsymbol{Y}_2}$, $R_{\boldsymbol{Y}_2}$, $Q_{W\boldsymbol{Y}_2}$ et $R_{W\boldsymbol{Y}_2}$ sont définies respectivement en (18)*,* (19)*,* (62) *et* (63)*.*

Preuve: La preuve du Lemme 3 est présentée dans l'Annexe L. ■

Résultat en longueur de bloc finie

Définissons pour tout $k \in \{1, 2\}$ et tout $(x, y) \in \mathcal{X} \times \mathcal{Y}_k$,

$$\hat{R}_{Y_k|X}(y|x) = \sum_{\hat{x} \in \mathcal{X}} \hat{P}_{\hat{X}|X}(\hat{x}|x) P_{Y_k|X}(y|\hat{x}). \tag{67}$$

En utilisant l'inégalité de Fano[38], la proposition suivante établit pour tout $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-code furtif $\hat{\mathcal{C}}$ que $\log_2\left(\hat{M}\right)$ admet une borne supérieure qui s'exprime en fonction des distributions empiriques induites à la fois par le code broadcast initial $\mathcal{C}$ et le code furtif $\hat{\mathcal{C}}$.

Proposition 3. *Soit un (n, M, ϵ)-code broadcast $\mathcal{C}$, pour le canal en* (9) *décrit par le système en* (11)*. Alors, tout $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-code furtif vérifie*

$$\log_2\left(\hat{M}\right) \leqslant \frac{1}{1-\hat{\epsilon}} \Bigg(1 + \sum_{x \in \mathcal{X}} \bar{P}_X(x) \Bigg(n \sum_{\hat{x} \in \mathcal{X}} \theta(x) \hat{P}_{\hat{X}|X}(\hat{x}|x) \cdot D\Big(P_{Y_1|X=\hat{x}} || P_{Y_1|X=x}\Big) + \frac{\theta(x)^3}{6} \chi_3(\hat{R}_{Y_1|X=x}, P_{Y_1|X=x}) \Bigg) \Bigg). \tag{68}$$

Preuve: La preuve de la Proposition 3 est présentée dans l'Annexe M and découle de l'inégalité de Fano [38]. ■

Une observation centrale pour prouver le résultat principal de cette section est qu'étant donné un code furtif, un sous-code peut être obtenu en choisissant les mot-codes dont le poids (Définition 5) est borné. Plus important, pour une classe particulière de canaux, la cardinalité de l'ensemble de mot-codes dont le poids admet une borne supérieure fixée admet une borne inférieure. Ce résultat est présenté par la proposition suivante, qui est inspirée de [13, Lemma 12].

Proposition 4. *Soit $\eta > 0$ arbitrairement petit et supposons que pour toute paire $(x, x') \in \mathcal{X}^2$ telle que $x \neq x'$, le canal en (9) satisfait les conditions suivantes :*

$$\chi_2\left(P_{Y_2|X=x}, P_{Y_2|X=x'}\right) = d, \quad \text{et} \tag{69}$$

$$D\left(P_{Y_1|X=x} || P_{Y_1|X=x'}\right) = \ell, \tag{70}$$

où $(d, \ell) \in \mathbb{R}_+^2$.

Soit un (n, M, ϵ)-code broadcast $\mathcal{C}$, décrit par le système en (11), pour ce canal. Alors, tout $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-code furtif décrit par le système en (15) peut être divisé en deux sous-codes. Un sous-code dont les mot-codes sont dans l'ensemble

$$\tilde{\mathcal{W}} = \left\{ \boldsymbol{v}(i,j) \colon \omega(i,j) < 2\sqrt{\frac{n}{d}} Q^{-1}\Big(\frac{1-\delta-\eta}{2}\Big), 1 \leqslant i \leqslant M, \text{ et } 1 \leqslant j \leqslant \hat{M} \right\}, \tag{71}$$

et un autre sous-code dont les mot-codes sont dans l'ensemble

$$\tilde{\mathcal{W}}^{\mathsf{c}} = \left\{ \boldsymbol{v}(i,j) \colon \omega(i,j) \geqslant 2\sqrt{\frac{n}{d}} Q^{-1}\Big(\frac{1-\delta-\eta}{2}\Big), 1 \leqslant i \leqslant M, \text{ et } 1 \leqslant j \leqslant \hat{M} \right\}. \tag{72}$$

De plus,

$$\left|\tilde{\mathcal{W}}\right| > M\hat{M}\left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon - \hat{\epsilon}\right), \tag{73}$$

où c est une constante.

Preuve: La preuve de la Proposition 6 est présentée dans l'Annexe N. ∎

Un example de canal vérifiant (69) et (70) est le canal binaire symétrique. Un autre exemple sera abordé plus loin.

Résultat asymptotique

Le théorème suivante présente le résultat principal de cette section.

Théorème 2. *Soit une séquence $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$, ... de (n, M_n, ϵ_n)-codes broadcast pour le canal en (9), avec $\lim_{n\to\infty} \epsilon_n = 0$. Supposons que le canal en (9) vérifie (69) and (70). Alors, toute séquence $\hat{\mathcal{C}}_1$, $\hat{\mathcal{C}}_2$, $\hat{\mathcal{C}}_3$, ... de $(n, \mathcal{C}_n, \hat{M}_n, \hat{\epsilon}_n, \delta)$-codes furtifs avec $\lim_{n\to\infty} \hat{\epsilon}_n = 0$ vérifie*

$$\lim_{n\to\infty} \frac{\log_2\left(\hat{M}_n^*(n, \mathcal{C}_n, 1, \hat{\epsilon}_n, \delta)\right)}{\sqrt{n}} < \frac{2\ell}{\sqrt{d}} Q^{-1}\left(\frac{1-\delta-\eta}{2}\right), \tag{74}$$

où $\eta > 0$ est arbitrairement petit.

Preuve: Pour tout $n \in \mathbb{N}$, il suit de la Proposition 4 que le sous-code du code furtif $\hat{\mathcal{C}}_n$

dont les mot-codes appartiennent à l'ensemble

$$\tilde{\mathcal{W}}_n = \left\{ \boldsymbol{v}(i,j) \colon \omega(i,j) < 2\sqrt{\frac{n}{d}} Q^{-1}\Big(\frac{1-\delta-\eta}{2}\Big), 1 \leqslant i \leqslant M_n, \text{ et } 1 \leqslant j \leqslant \hat{M}_n \right\}, \tag{75}$$

vérifie

$$\left|\tilde{\mathcal{W}}_n\right| > M_n \hat{M}_n \left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon_n - \hat{\epsilon}_n \right). \tag{76}$$

Ainsi, il découle de l'équation (76) que pour tout indice $i \in \mathcal{W}$, il y a en moyenne $\hat{M}_n \left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon_n - \hat{\epsilon}_n \right)$ mot-codes dans le sous-code. Donc, la Proposition 3 s'applique, et il s'ensuit que

$$\begin{aligned} &\log_2 \left(\hat{M}_n \left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon_n - \hat{\epsilon}_n \right) \right) \leqslant \\ &\qquad \frac{1}{1-\hat{\epsilon}_n} \left(1 + n \sum_{x \in \mathcal{X}} \ell \bar{P}_X(x) \theta(x) + \bar{P}_X(x) \frac{\theta(x)^3}{6} \chi_3(\hat{R}_{Y_1|X=x}, P_{Y_1|X=x}) \right), \end{aligned} \tag{77}$$

ce qui implique que

$$\begin{aligned} \log_2 \left(\hat{M}_n \right) &\leqslant \frac{1}{1-\hat{\epsilon}_n} \left(1 + n \sum_{x \in \mathcal{X}} \ell \bar{P}_X(x) \theta(x) + \bar{P}_X(x) \frac{\theta(x)^3}{6} \chi_3(\hat{R}_{Y_1|X=x}, P_{Y_1|X=x}) \right) \\ &\quad - \log_2 \left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon_n - \hat{\epsilon}_n \right) \\ &\overset{(a)}{\leqslant} \frac{1}{1-\hat{\epsilon}_n} \left(1 + \sum_{x \in \mathcal{X}} \ell \omega(x) + \omega(x)^3 \max_{x' \in \mathcal{X}} \frac{\chi_3(\hat{R}_{Y_1|X=x'}, P_{Y_1|X=x'})}{6n^2 \bar{P}_X(x')^2} \right) \\ &\quad - \log_2 \left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon_n - \hat{\epsilon}_n \right) \\ &\overset{(b)}{\leqslant} \frac{1}{1-\hat{\epsilon}_n} \left(1 + 2\frac{\ell\sqrt{n}}{\sqrt{d}} Q^{-1}\left(\frac{1-\delta-\eta}{2}\right) + \left(\frac{2\sqrt{n}}{\sqrt{d}} Q^{-1}\left(\frac{1-\delta-\eta}{2}\right) \right)^3 \right. \\ &\quad \left. \cdot \sum_{x \in \mathcal{X}} \max_{x' \in \mathcal{X}} \frac{\chi_3(\hat{R}_{Y_1|X=x'}, P_{Y_1|X=x'})}{6n^2 \bar{P}_X(x')^2} \right) - \log_2 \left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon_n - \hat{\epsilon}_n \right) \\ &= \frac{1}{1-\hat{\epsilon}_n} \left(1 + 2\frac{\ell\sqrt{n}}{\sqrt{d}} Q^{-1}\left(\frac{1-\delta-\eta}{2}\right) + \frac{4\,|\mathcal{X}|}{3\sqrt{n}\sqrt{d}^3} \right. \\ &\quad \left. \cdot Q^{-1}\left(\frac{1-\delta-\eta}{2}\right)^3 \max_{x' \in \mathcal{X}} \frac{\chi_3(\hat{R}_{Y_1|X=x'}, P_{Y_1|X=x'})}{\bar{P}_X(x')^2} \right) - \log_2 \left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon_n - \hat{\epsilon}_n \right), \end{aligned} \tag{78}$$

où c est une constante qui dépend uniquement des paramètres du canal en (9). Notons que

(a) est une conséquence du Lemme 5, et *(b)* suit du fait que pour tout $x \in \mathcal{X}$,

$$\begin{aligned}\omega(x) &\leqslant \sum_{v\in\mathcal{X}} \omega(v) \\ &= \sum_{i=1}^{M_n}\sum_{j=1}^{\hat{M}_n} \frac{\omega(i,j)}{M_n\hat{M}_n} \\ &\leqslant 2\frac{\sqrt{n}}{\sqrt{d}} Q^{-1}\left(\frac{1-\delta-\eta}{2}\right). \end{aligned} \tag{79}$$

La preuve est complétée en divisant les deux membres de (78) par $\sqrt{n}$ et en prenant la limite quand n tend vers l'infini. ■

3.4. Résultat principal

Pour les canaux satisfaisant (69) et (70), le membre de droite de (50) se simplifie comme suit :

$$(1-\xi)\frac{2\ell}{\sqrt{d}} Q^{-1}\left(\frac{1-\delta}{2}\right). \tag{80}$$

En se rappelant que ξ et η en (74) peuvent être choisis arbitrairement petits, il s'ensuit que, pour des canaux symétriques, les bornes asymptotiques présentées dans le Théorème 1 et le Théorème 2 sont arbitrairement proches, *i.e.*, l'équation (80) donne la constante optimale qui caractérise la limite du ratio entre $\log_2\left(\hat{M}_n^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)$ et $\sqrt{n}$ quand la longueur de bloc n tend vers l'infini.

Théorème 3. *Soit une séquence $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$, ... de (n, M_n, ϵ_n)-codes broadcast pour le canal en* (9) *tel que* (69) *et* (70) *sont vérifiées, avec $n \in \{1, 2, \ldots\}$ et*

$$\epsilon_n \leqslant \exp(-\zeta n), \tag{81}$$

pour un réel positif fixe ζ. Alors,

$$\lim_{n\to\infty} \frac{\log_2\left(\hat{M}_n^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)}{\sqrt{n}} = \frac{2\ell}{\sqrt{d}} Q^{-1}\left(\frac{1-\delta}{2}\right). \tag{82}$$

L'obtention d'un converse général pour des canaux qui ne satisfont pas les conditions de symétrie en (69) et (70) reste un problème ouvert à ce jour. Par ailleurs, une question intéressante est de savoir si la variation totale utilisée dans cette étude peut être remplacée par une divergence de Kullback-Leibler.

3.5. Exemples

Cette section présente des exemples illustrant le résultat du Théorème 3.

Exemple 1 (Canal binaire symétrique)**.** *Soit le canal en* (9) *tel que $\mathcal{X} = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1\}$, et tel que pour toute paire $(x, x') \in \mathcal{X}^2$ où $x \neq x'$, les distributions conditionnelles de probabilité $P_{Y_1|X}$ et $P_{Y_2|Y_1}$ vérifient respectivement :*

$$P_{Y_1|X}(x|x) = 1 - P_{Y_1|X}(x'|x) = 1 - p_1, \text{ et} \tag{83}$$
$$P_{Y_2|Y_1}(x|x) = 1 - P_{Y_2|Y_1}(x'|x) = 1 - p_2, \tag{84}$$

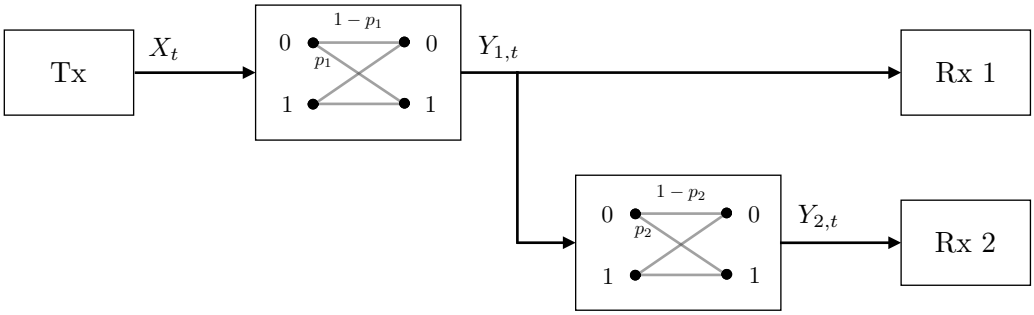

FIGURE 3. : Canal broadcast dégradé verifiant (69) et (70), à l'utilisation de canal $t \in \{1, 2, \ldots, n\}$.

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

La Figure 3 illustre le canal de l'Exemple 1. La distribution de probabilité $P_{Y_2|X}$ vérifie pour toute paire $(x, x') \in \mathcal{X}^2$ où $x \neq x'$:

$$P_{Y_2|X}(x|x) = 1 - P_{Y_2|X}(x'|x) = 1 - p, \tag{85}$$

avec

$$p = p_1 + p_2 - 2p_1p_2. \tag{86}$$

Ainsi, pour toute paire $(x, x') \in \mathcal{X}^2$ où $x \neq x'$,

$$\chi_2(P_{Y_2|X=x'}, P_{Y_2|X=x}) = \frac{(1-2p)^2}{p(1-p)}, \tag{87}$$

$$D\big(P_{Y_1|X=x'} || P_{Y_1|X=x}\big) = (1-2p_1)\log_2\left(\frac{1-p_1}{p_1}\right), \tag{88}$$

Il s'ensuit immédiatement du Théorème 3 que

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

où p est en (92).

L'expression précédente est tracée en fonction des probabilités p_1 et p_2 respectivement sur la Figure 4 et la Figure 5, avec $\delta = 0.005$.

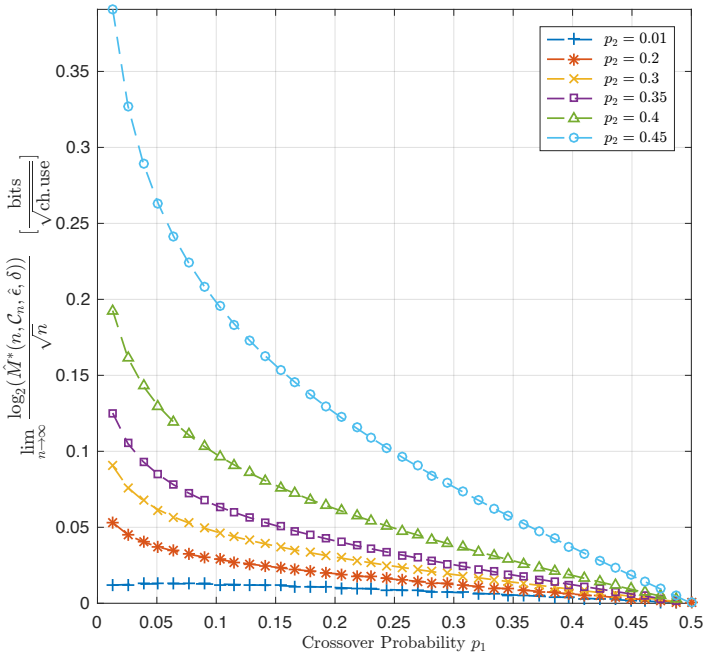

FIGURE 4. : Limite fondamentale $\lim_{n\to\infty} \frac{\log_2\left(\hat{M}^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)}{\sqrt{n}}$ en fonction de la probabilité p_1, pour $\delta = 0.005$.

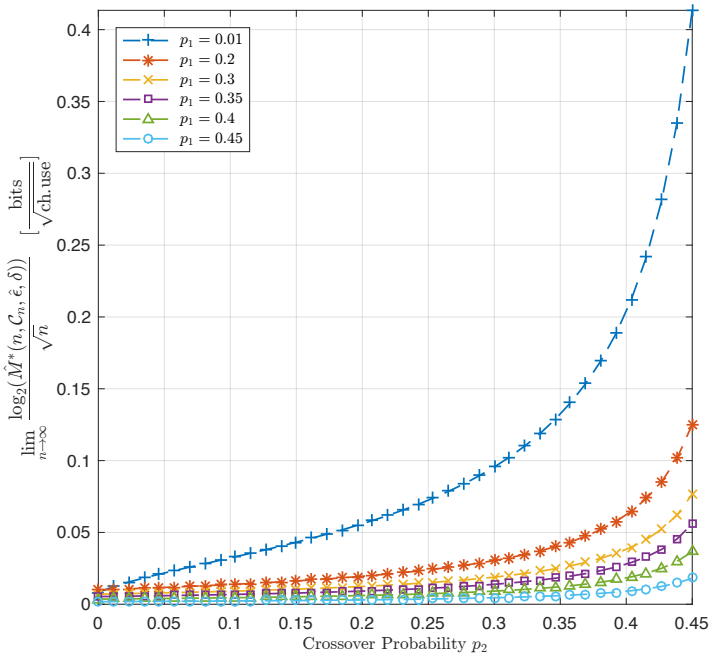

FIGURE 5. : Limite fondamentale $\lim_{n\to\infty} \frac{\log_2\left(\hat{M}^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)}{\sqrt{n}}$ en fonction de la probabilité p_2, pour $\delta = 0.005$.

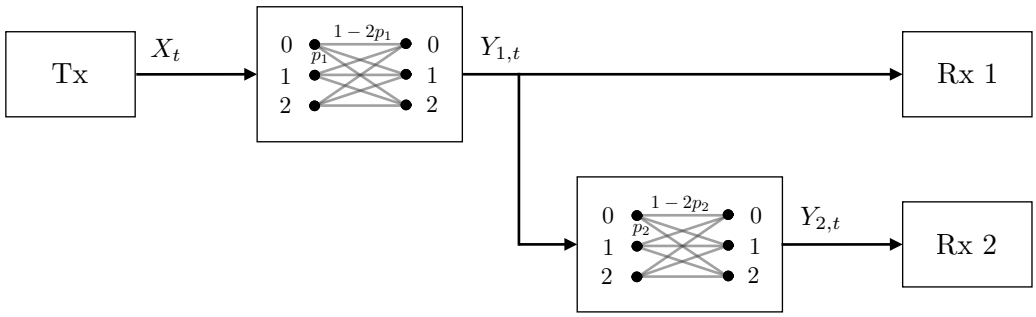

FIGURE 6. : Canal broadcast dégradé vérifiant (69) et (70), à l'utilisation de canal $t \in \{1, 2, \ldots, n\}$.

Exemple 2. *Soit le canal en* (9) *tel que* $\mathcal{X} = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1, 2\}$*, et tel que pour toute paire* $(x, x') \in \mathcal{X}^2$ *où* $x \neq x'$*, les distributions conditionnelles* $P_{Y_1|X}$ *et* $P_{Y_2|Y_1}$ *vérifient respectivement :*

$$P_{Y_1|X}(x|x) = 1 - 2P_{Y_1|X}(x'|x) = 1 - 2p_1, \text{ et} \tag{89}$$
$$P_{Y_2|Y_1}(x|x) = 1 - 2P_{Y_2|Y_1}(x'|x) = 1 - 2p_2, \tag{90}$$

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

La Figure 6 illustre le canal de l'Exemple 2. La fonction de masse $P_{Y_2|X}$ vérifie pour toute paire $(x, x') \in \mathcal{X}^2$ où $x \neq x'$:

$$\begin{aligned} P_{Y_2|X}(x|x) = 1 - 2P_{Y_2|X}(x'|x) &= 1 - 2\left(p_1 + p_2 - 3p_1p_2\right) \\ &= 1 - 2p, \end{aligned} \tag{91}$$

avec

$$p = p_1 + p_2 - 3p_1p_2. \tag{92}$$

Le lemme suivant quantifie les expressions $\chi_2(P_{Y_2|X=x'}, P_{Y_2|X=x})$ et $D\big(P_{Y_2|X=x'}||P_{Y_2|X=x}\big)$ pour toute paire $(x, x') \in \mathcal{X}^2$ telle que $x \neq x'$.

Lemme 4. *Considérons l'Exemple 2. Toute paire* $(x, x') \in \mathcal{X}^2$ *telle que* $x \neq x'$ *vérifie*

$$\chi_2(P_{Y_2|X=x'}, P_{Y_2|X=x}) = \frac{(3p-1)^2(1-p)}{p(1-2p)}, \tag{93}$$
$$D\big(P_{Y_1|X=x'}||P_{Y_1|X=x}\big) = (1-3p_1)\log_2\left(\frac{1-2p_1}{p_1}\right), \tag{94}$$

où p *est défini en* (4.94).

Preuve: La preuve du Lemme 4 est présentée dans l'Annexe O. ■

La proposition suivante suit directement du Lemme 4 et du Théorème 3.

Proposition 5. *Considérons l'Exemple 2 et considérons une séquence de* (n, M_n, ϵ_n)*-codes broadcast, avec* $n \in \{1, 2, \ldots\}$*, notés respectivement* $\mathcal{C}_1$*,* $\mathcal{C}_2$*, ..., telle que* $\lim_{n\to\infty} \epsilon_n = 0$*. Alors,*

$$\lim_{n\to\infty} \frac{\log_2\left(\hat{M}_n^*(n, \mathcal{C}_n, L_n, \hat{\epsilon}_n, \delta)\right)}{\sqrt{n}} = 2Q^{-1}\left(\frac{1-\delta}{2}\right)\sqrt{\frac{p(1-2p)}{1-p}}\frac{1-3p_1}{1-3p}\log_2\left(\frac{1-2p_1}{p_1}\right). \tag{95}$$

L'expression précédente est tracée en fonction des probabilités p_1 et p_2 respectivement sur la Figure 7 et la Figure 8, avec $\delta = 0.005$.

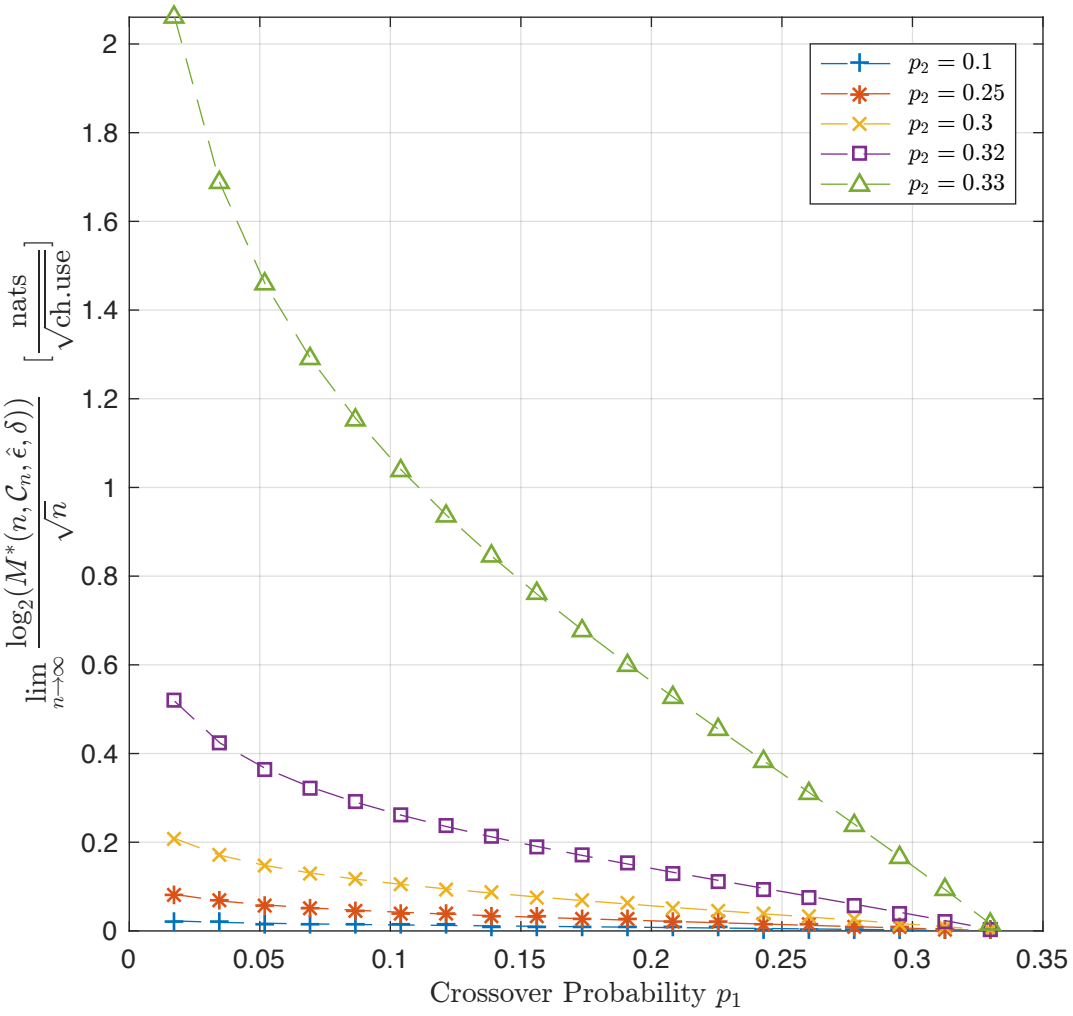

FIGURE 7. : Limite fondamentale $\lim_{n\to\infty} \frac{\log_2\left(\hat{M}^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)}{\sqrt{n}}$ en fonction de la probabilité p_1, pour $\delta = 0.005$.

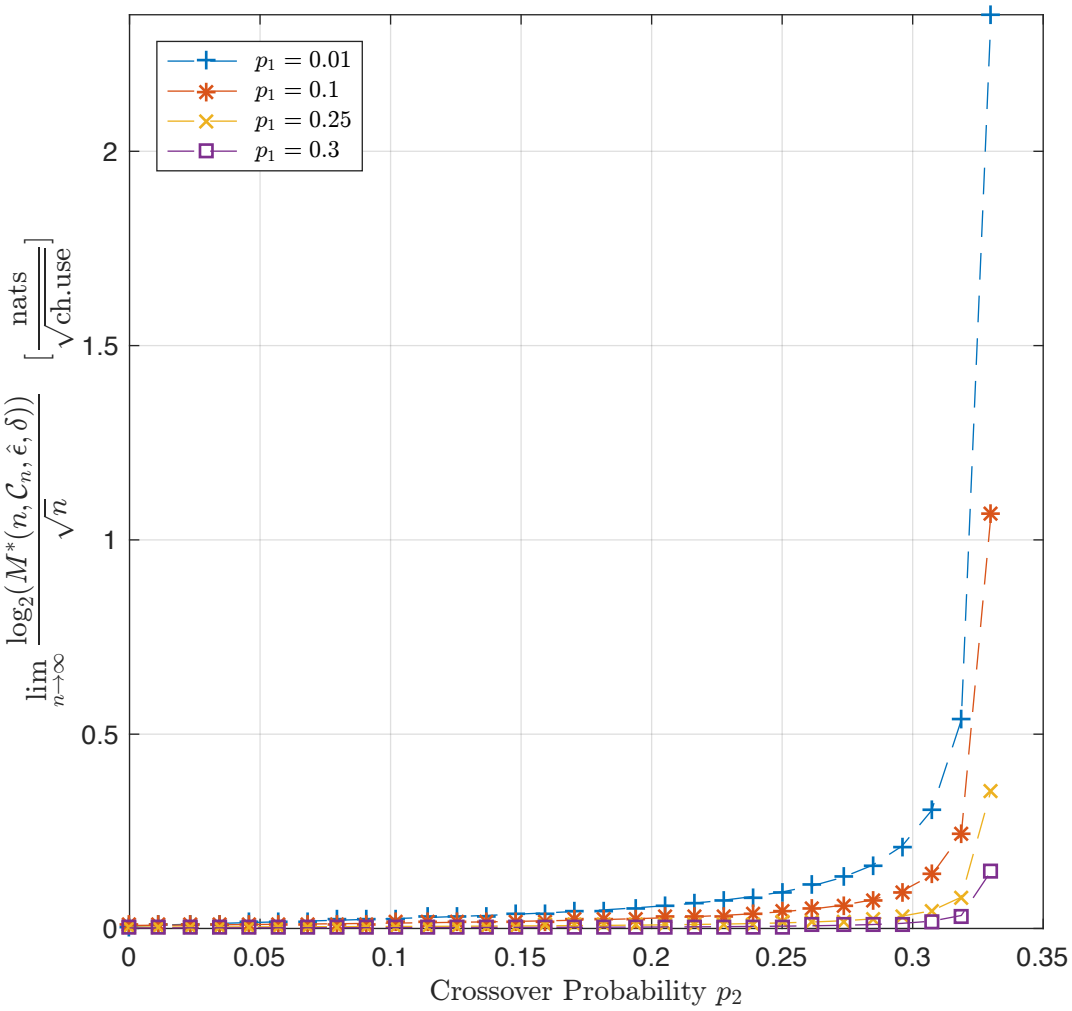

FIGURE 8. : Limite fondamentale $\lim_{n\to\infty} \frac{\log_2\left(\hat{M}^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)}{\sqrt{n}}$ en fonction de la probabilité p_2, pour $\delta = 0.005$.

4. Conclusion

Dans cette thèse deux nouveaux problèmes sont présentés, bien qu'un seul d'entre eux fasse l'objet de cette synthèse. Premièrement, cette thèse introduit le problème de transmission furtive d'information sur un canal point-à-point lorsqu'un adversaire observe uniquement une fraction des sorties de canal, nommé communications furtive de type II. Ce modèle généralise le problème de communication furtive sur un canal point-à-point. Une borne inférieure sur le débit atteignable pour un code de longueur finie est obtenue dans le Chapitre 3 pour ce problème. Cette borne révèle deux régimes de communication : un régime dans lequel la capacité du canal point-à-point est atteignable et un second régime dans lequel la loi en racine carrée des communications furtives s'applique au taux de transmission. Cette étude constitue une première étape vers la résolution du modèle général, *i.e.*, la détermination du débit maximum auquel l'information peut être transmise de manière fiable et furtive simultanément. Ce problème, qui n'est pas l'objet de cette synthèse, est détaillé au Chapitre 3.

Deuxièmement, le problème d'introduction d'information furtive dans un code broadcast donné est présenté. Contrairement aux travaux précédents, cette étude suppose un code broadcast arbitraire donné, ce qui rend la preuve de faisabilité plus difficile. Une borne de faisabilité et un converse sont obtenus dans le régime asymptotique dans la section 3 et dans le Chapitre 4 pour une classe de canaux particulière, *i.e.*, les canaux vérifiant certaines conditions de symétrie. Ces bornes permettent de caractériser le nombre maximal de bits d'information qui peuvent être introduits dans un code broadcast donné pour des canaux symmétriques. Ce travail constitue une première étape vers la résolution du problème pour des canaux discrets sans mémoire arbitraires.

Ces deux contributions ouvrent un certain nombre de perspectives tant du point de vue théorique que du point de vue des applications. D'un point de vue théorique, la contribution présentée dans ce document sur les communications furtives de type II laisse ouvert le problème de l'obtention d'un converse pour des codes de longueur finie. Le converse n'étant pas connu, il est possible que la borne de faisabilité présentée dans ce document puisse être améliorée, bien que le terme de premier ordre soit optimal. De plus, la caractérisation du taux de transmission optimal sous des contraintes de furtivité différentes de la divergence de Kullback-Leibler est également un problème ouvert. Par ailleurs, la contribution sur les canaux broadcast laisse deux problèmes ouverts : déterminer une borne de faisabilité et une borne de converse pour des canaux discrets sans mémoire généraux. La caractérisation du nombre maximal de bits d'information qui peuvent être introduits de manière furtive en considérant d'autres contraintes de furtivité que la variation totale est également un problème ouvert. Il est intéressant de noter que le problème d'introduction d'information furtive dans un code broadcast donné est une instance d'un problème plus général. Dans les canaux multi-utilisateurs, les codes broadcast peuvent être modifiés pour remplir d'autres fonctionnalités, e.g., transmission simultanée d'énergie et d'information, sécurité de la couche physique (au sens traditionel), etc. Dans les deux problèmes, seulement les canaux discrets et sans mémoire sont considérés. Ainsi, le cas du canal Gaussien est également un problème ouvert. De plus, ces bornes permettent aux designers de codes de développer des codes qui peuvent atteindre ces bornes. Comme il est discuté dans [34, 35, 36], la modulation en position d'impulsions et certaines variations autour de cette modulation semblent de bonnes candidates pour atteindre le débit optimal dans le cas d'une communication point-à-point. Ceci pourrait éventuellement se vérifier aussi dans le cas du canal broadcast.

En termes d'applications, le problème des communications furtive de type II ouvre une voie vers le design d'une multiplicité de systèmes de communication qui peuvent transmettre l'information de manière furtive. Par exemple, les communications furtives pourraient être utilisées pour les signaux de contrôle du réseau, qui sont des signaux de faible débit souvent détournés par des attaquants, ce qui induit des failles de sécurité. D'un autre côté, le problème d'introduction d'information furtive dans un code broadcast donné ouvre deux perspectives principales. Premièrement, ce problème ouvre la voie vers l'amélioration de systèmes de communications existants dans le but de leur ajouter un service, dans ce cas, transmettre un message privé additionnel de manière furtive. Deuxièmement, les bornes obtenues pour ce problème montrent qu'il existe des malwares pouvant affecter l'émetteur de sorte que celui-ci transmette des informations quelconques vers un tiers. Dans ce cas, la fuite d'information ne peut pas être détectée en inspectant le réseau mais seulement en vérifiant le code au niveau de l'émetteur, ce qui peut ne pas être aisé dans certains cas.

Enfin, dans les deux problèmes, l'addition de brouilleurs coopérants pourrait être utilisée pour améliorer le débit de communication comme il est discuté dans [25]. Ainsi, les systèmes

de communication full-duplex sont des candidats intéressants pour l'implémentation de tels systèmes de communications furtives.

— 1 —

Introduction

Security of communications systems, from theory to application, has been a long standing problem especially in wireless systems. Indeed, eavesdropping on a wired line requires that the eavesdropper has access to the communication line and can wiretap it. In contrast, the broadcast nature of wireless communications opens the door to malicious users that may eavesdrop on the communication or disrupt it. In this case, the simple fact of having a receiver is enough to eavesdrop on a system. Hence, due to the unprecedented growth of wireless devices connected around the globe, there is a growing demand for security of wireless systems. In this thesis, the focus is on problems that arise when a malicious user eavesdrops the communication.

From a theoretical perspective, the security of communications systems was first studied in Shannon's landmark paper [1], which introduces the concept of perfect secrecy. According to Shannon, in order to have a secure communication system against an eavesdropper, the message and the eavesdropper's observation must be statistically independent. This notion of perfect secrecy is quite stringent, and thus, it was later weakened to secrecy constraints such as weak secrecy, strong secrecy, effective secrecy or semantic secrecy. All these secrecy metrics ensure different levels of statistical independence between the message and the eavesdropper's observation.

Another way to ensure the non-decodability of the message at the eavesdropper is to guarantee that the latter will not be able to detect the communication itself. This problem is known as covert communications or communications with low-probability of detection. In this problem, the coding of the message should be done in such a way that the most powerful detector at the eavesdropper will almost always fail to detect the communication. That is, in the simplest point-to-point case, the eavesdropper's channel output should look like the channel output generated when only noise is present. It is worth noting that this kind of constraint is much more stringent than the secrecy metrics mentioned above.

Nevertheless, covert communication systems find applications in military settings among others. For instance, consider a battlefield scenario where a general sends orders to its troops. In some circumstances, it might be crucial for the adversary to know that orders have been sent even though the orders themselves are not known. Hence, the need to covertly transmit the orders. Another application is the design of stealth vehicles, such as submarines for instance,

that have the ability to communicate without revealing their presence or, even worse, their position. Covert communications also find applications in investigation journalism where the secret transmission of data is crucial to the cooperation of journalists between them. Finally, in the day-to-day life, covert communications could also be used to communicate critical data such as medical data in the context of Internet of Things devices, or to wirelessly pay with debit cards.

The theoretical study of the different instances of covert communications problems is of great interest. It provides the designer of communications systems with bounds on the maximum rate at which information can be simultaneously reliably and covertly sent. This paves the way for coding engineers to benchmark their designs.

This thesis presents two main theoretical contributions. First, the problem of covert communications type II is introduced. This is a traditional instance of the covert communication problem over a point-to-point link, where the adversary chooses a fraction of the channel outputs to perform its detection of the communication. For this problem, an achievable bound – a lower bound on the maximum rate at which information can be reliably and covertly transmitted – in the finite block-length regime is presented. Second, the problem of embedding covert information into a given broadcast code is introduced. Given a broadcast code to transmit a common message to two receivers, the goal of this problem is to determine the maximum number of information bits that can be reliably sent to one receiver while being covert with respect to the other receiver. In this thesis, an achievable bound and a converse bound – an upper-bound – on the maximum number of information bits that can be covertly embedded into a given broadcast code are established in the asymptotic block-length regime for a particular class of channels. Together, these bounds characterize the maximum number of information bits that can be covertly embedded into a given broadcast code for a particular class of channels, *i.e.*, symmetric channels.

None of these two problems were studied in the literature before. The problem of covert communications type II generalizes the problem of covert communications over point-to-point links. The achievability bound presented in this thesis is a first step towards the resolution of this generalized model. The problem of embedding covert communications into a given broadcast code is new in the sense that the broadcast code is assumed to be given. In contrast, [2, 3] treat the problem of jointly designing a code to transmit common information to two receiver and additional covert information to one of the two receivers. Again, the characterization of the maximum number of information bits that can be covertly embedded into a given broadcast code for symmetric channels constitutes a first step towards solving the problem for arbitrary discrete memoryless channels.

The remainder of this thesis unfolds as follows. Chapter 2 presents a state of the art on covert communications. Chapter 3 introduces the problem of covert communications type II and establishes an achievable bound in the finite block-length regime. Then, Chapter 4 introduces the problem of embedding covert information into a given broadcast code and characterizes the maximum number of information bits that can be covertly embedded into a given broadcast code for a particular class of channels. Finally, Chapter 5 concludes this thesis.

— 2 —

State of the Art

Security has been a long standing problem in the information theory community. Shannon himself introduced the first problem of secrecy in [1]. In this problem, the transmitter aims at communicating with a receiver over a noiseless channel in the presence of an adversary, the eavesdropper, that also observes a noiseless channel output (see Figure 2.1). Shannon introduced the notion of *perfect secrecy* to treat this problem, and defined perfect secrecy as follows:

$$I(W; \boldsymbol{Z}) = 0, \tag{2.1}$$

where $W \in \mathcal{W}$ is the secret message, with $\mathcal{W}$ the set of messages, and $\boldsymbol{Z} \in \mathcal{Z}^n$ is the eavesdropper's channel output, with $\mathcal{Z}$ the eavesdropper's channel output alphabet and $n \in \mathbb{N}$ the block-length. That is, given its channel output observation $\boldsymbol{Z}$, the eavesdropper should not be able to infer anything about the message W, a fortiori, he should not be able to decode W. Shannon concluded that to satisfy such a constraint, a secret-key was required, and in particular, the entropy of the key should equal that of the message source. That is, assuming the message indices and key indices uniformly distributed, there should be the same number of key indices as the number of message indices.

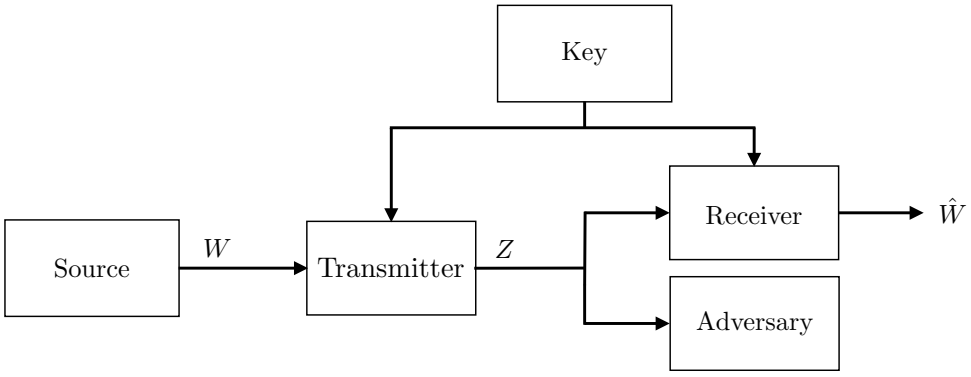

Figure 2.1.: Point-to-point secrecy system.

Later, Wyner introduced the wiretap channel in [4], in which the channel is not anymore noiseless, and in particular, the eavesdropper observes a degraded version of the legitimate receiver's channel output. He also introduced a looser secrecy metric which is of the form

$$\lim_{n\to\infty} \frac{1}{n} I(W; \boldsymbol{Z}) = 0, \tag{2.2}$$

known as *weak secrecy*. With these assumptions, Wyner was able to show that keyless secret communication are achievable.

Csiszár introduced *strong secrecy* in [5] in order to strengthen the metric of Wyner, and defined it as

$$\lim_{n\to\infty} I(W; \boldsymbol{Z}) = 0. \tag{2.3}$$

Abandoning the normalization of the mutual information indeed strengthen the metric.

Then, Hou *et al.* introduced in [6] the notion of *effective secrecy*, defined as

$$\lim_{n\to\infty} D(P_{W\boldsymbol{Z}} || P_W Q_{\boldsymbol{Z}}) = 0, \tag{2.4}$$

where $Q_{\boldsymbol{Z}}$ is the distribution expected by the eavesdropper when the source is not communicating useful messages. This measure gather two criterions, namely the confusion of the eavesdropper and the stealth of the message (which in some sense related to covert communications).

Finally, Bellare *et al.* introduced the notion of *semantic secrecy* [39], defined as

$$\lim_{n\to\infty} \max_{P_W} I(W; \boldsymbol{Z}) = 0, \tag{2.5}$$

in which the distribution of the message can be anything. This contrasts with precedent work in which the message is usually assumed to be uniformly distributed.

In front of this growing number of metrics ensuring non-decodability of the message emerged a new problem in information theory: the problem of covert communications. In this problem, the transmitter and the receiver should communicate in such a way that the adversary should not be able to detect the existence of communications. The adversary, in contrast with the wiretap setup does not aim at decoding the transmitted message but aims at detecting the transmission. Hence, to distinguish the two setup, the adversary is named the eavesdropper when it tries to decode the message and the warden when it tries to detect the communication. The covertness constraint is a much more stringent requirement than a security constraint. It finds motivation, among others, in military applications in which the very existence of a transmission can be an information of capital importance.

This chapter reviews existing information theoretic results on covert communications. In the first part of the chapter (Section 2.1), a brief description of the main contributions in covert communications is given for different channels. Then in a second part (Section 2.2), more details are provided with a unified notation for the canonical information theoretic channels.

2.1. Main Results on Covert Communications

In this section a description of the main results on covert communications is presented. First, point-to-point channels are reviewed. Then, point-to-point channels with jammers are

presented. Afterwards, channels with multiple users are considered. Finally, results on coding for covert communications are reviewed.

2.1.1. Point-to-point Channels

Covert communications channels were first introduced in information theory by Bash *et al.* in [7, 8]. In the aforementioned papers, the authors show that, as in steganography, covert communications over AWGN point-to-point links are subject to the so-called square-root law. That is, the number of information bits that can be sent over $n \in \mathbb{N}$ channel uses scales with the square-root of n. To show this result, the authors use a communication scheme that involves a key of length $O(\sqrt{n}\log(n))$. These results show that the traditional notion of rate ($R = \frac{\log(M)}{n}$, with $M \in \mathbb{N}$ the number of messages) cannot be used since this quantity tends to zero. Hence, in covert communications problems, the quantity that is often studied is the ratio $\frac{\log(M)}{\sqrt{n}}$, which tends to a constant for most channels.

Later, these results have been refined by Che *et al.* in [9, 10], Bloch in [11] and Wang *et al.* in [12]. In [9, 10], the exact constant characterizing the limit of the ratio $\frac{\log(M)}{\sqrt{n}}$ for binary symmetric channels is established. In [11, 12], the limit of the ratio $\frac{\log(M)}{\sqrt{n}}$ is exactly characterized for DMCs using different achievability proof techniques. Indeed, the proof in [11] relies on channel resolvability and channel reliability (see [40]) whereas the proof in [12] relies on a one-shot achievability analysis. In addition, [11] characterizes the conditions for which no secret-key is required and shows an improvement on the key-length for DMCs when the key is needed. The communication scheme therein requires a key of length $O(\sqrt{n})$. Finally, [12] also charaterizes the exact limit of the ratio $\frac{\log(M)}{\sqrt{n}}$ for AWGN channels. In addition, Tahmasbi *et al.* have derived in [13] upper and lower bounds on the finite block-length rate that can be achieved under various covertness constraints.

Continuous-time AWGNs are studied by Wang in [14]. Therein, it is shown that when the noise is white, a positive covert information rate is achievable when there is no bandwidth constraint on the input. In contrast, in the band-limited case, covert communication schemes require that the number of transmitted bits grow at most as the square root of the total communication time.

Point-to-point channels with states have been studied by Lee *et al.* in [15, 16]. Therein, closed-form expressions of the maximum achievable covert rate are given in the case of causal and non-causal channel state information (CSI) for DMCs and in the case of non-causal CSI for AWGN channels. Interestingly, there exist state-dependent channels for which the rate is strictly positive, which contrasts with point-to-point DMCs and point-to-point AWGN channels without states.

In [17], non coherent point-to-point channels are studied by Tahmasbi *et al.* In particular, fast Rayleigh fading channels are considered. Therein, it is shown that the square-root law still holds, and that the optimal rate is achieved with an amplitude-constrained input distribution over a finite number of mass points that include 0.

In [18, 19, 20], Soltani *et al.* study the fundamental limits of covert packet insertion. The authors consider a first transmitter sending packets on a channel to a legitimate receiver in a given time interval. Another transmitter-receiver pair is present and aims at communicating packets – inserting packets – on the same channel while remaining undetected by an external warden. Therein, it is shown that the square-root law still applies. That is, if the total number of packets sent by the first transmitter is n, the number of packets that can be covertly inserted

scales with the square-root of n.

In [21], Soltani *et al.* consider the problem of bit insertion in packets. The authors consider a first transmitter sending n packets to a first receiver. The packets are relayed to the first receiver by three entities. The first one aims at inserting covert bits into the packets that are assumed to have available payload space. The second one is a warden that wishes to detect the bit insertion in the packets. The third one aims at decoding the covert bits inserted into the packets. It is shown therein that if the number of packets is n, then, on the order of square-root of n bits can be inserted in the packets.

Poisson point-to-point channels have been studied by Wang in [22]. It is shown that for continuous-time Poisson point-to-point channels without peak-power constraint, the covert communication capacity is infinite.

Tahmasbi *et al.* have introduced error exponents for covert communications over point-to-point channels in [23]. Therein, they prove upper and lower bounds on the error exponent that are shown to match in a certain regime.

In [24], Tahmasbi *et al.* study the feasibility of covert secret key generation. It is shown in this paper that covert secret key generation is possible for some models. In addition, in models for which the covert secret key capacity equals the covert capacity of the channel, it is shown that secrecy comes "for free". That is, for these models, if covertness of the key generation protocol is ensured, then the key and the adversary's channel output appear statistically independent in the regime where the block-length grows large.

2.1.2. Point-to-point Channels with Jammers

Point-to-point covert communications channels with friendly jammers have been introduced by Soltani *et al.* in [25, 26]. In [25], the study considers several friendly jammers willing to help the transmitter to achieve covert communications whereas several wardens are scrutinizing the channel. The jammers are assumed to be randomly located according to a two-dimensional point-process whereas the wardens are assumed to be uniformly and independently distributed at random. In this case, an improvement can be observed with respect to the square-root law that holds for most point-to-point channels. Indeed, it is shown in this paper that the improvement depends on the density of the point-process, on the path-loss exponent, and on the number of wardens.

In [27, 28], Zheng *et al.* consider communications with an adversarial jammer (with respect to the legitimate transmitter-receiver pair). In the presence of such a jammer, it is shown in the paper that regardless of the channel characteristics, a secret key of length $\Omega(\log(n))$ is required. In addition, when the channel is adversarially jammed, the square-root law still applies.

In [29], Sobers *et al.* consider the presence of an uninformed friendly jammer and assume block fading channels. The jammer is uninformed in the sense that it is not coordinated with the transmitter. This paper reveals scenarios in which the square-root law does not hold and in which $O(n)$ bits can be sent in n channel uses.

2.1.3. Multiple-User Channels

Covert communications over discrete memoryless broadcast channels (BC) have been introduced in [2, 3] by Arumugam *et al.* The model consists of a transmitter sending private information to one receiver and common information to two receivers. The setup considered is that in

which the non-intended receiver of the private message should remain unaware of the existence of the private message. It is considered in this paper that the codes for common information transmission and covert private information transmission are jointly designed. In this case, the authors explicit the exact maximum amount of information that can be embedded in the code to transmit common information. This result verifies once again the square-root law for covert communications.

Tan *et al.* considered another kind of broadcast channel in [30]. Their model consists of a transmitter sending two messages to two receivers (one for each receiver) in the presence of an additional warden. In this case, the square-root law also applies. In addition, the authors show that time-sharing between the two receivers is optimal.

In the context of covert communications, the multiple-access channel (MAC) has been studied by Arumugam *et al.* in [31, 32]. The model consists of two transmitters willing to transmit covert information to a legitimate receiver in the presence of a warden. In [31, 32] the authors fully characterize the covert capacity region of the 2-user discrete memoryless MAC and the K-user discrete memoryless MAC, respectively. The authors show that for this instance of MAC, the square-root law applies.

Arumugam *et al.* also studied the relay-channel under covertness constraint in [33]. The authors of this paper consider a degraded relay channel with two non-colluding wardens. One is on the link between the transmitter and the relay and the second is on the link between the relay and the receiver. The authors characterize in the asymptotic block-length regime the optimal ratio $\frac{\log(M)}{\sqrt{n}}$, where $M \in \mathbb{N}$ and $n \in \mathbb{N}$ are the number of messages and the number of channel uses, respectively. In this scenario, the square-root law applies.

2.1.4. Coding for Covert Communications

Bloch *et al.* [34] and Kadampot *et al.* [35, 36] started to investigate how to code information over point-to-point channels to perform covert communications. They show that pulse-position modulation and variations around pulse-position modulation achieve the best achievable rate. In [35, 36], the code construction presents the advantage of having low computational complexity.

2.2. Detailed Results for Canonical Information Theoretic Channels

2.2.1. Point-to-point Channels

In this section, existing results for point-to-point channels are reviewed. First, a general channel model is described. Then, results for two particular cases are presented: the additive white Gaussian noise (AWGN) channels and the discrete memoryless channels (DMCs).

System Model

Consider a three-party communication system in which a transmitter sends information to a legitimate receiver while a second receiver (the warden) observes the channel aiming to determine whether or not communication occurs between the legitimate parties. The noisy communication medium is described by a product random transformation

$$\left(\mathcal{X}^n, \mathcal{Y}^n, \mathcal{Z}^n, P_{\boldsymbol{YZ}|\boldsymbol{X}}\right), \tag{2.6a}$$

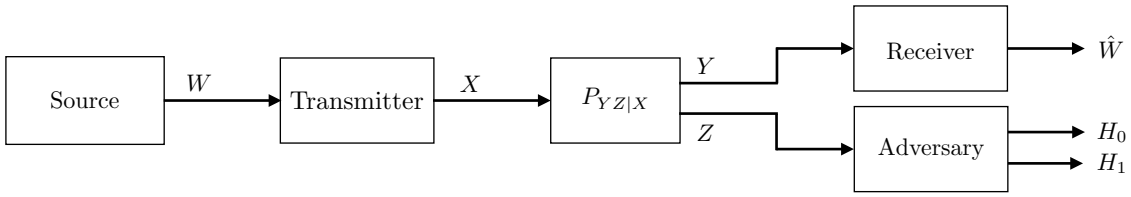

Figure 2.2.: Point-to-point channel with a warden.

where $n \in \mathbb{N}$ is the block-length; $\mathcal{X}$ is the input alphabet, $\mathcal{Y}$ and $\mathcal{Z}$ are the output alphabets. That is, given an input vector $\boldsymbol{x}$, the outputs $\boldsymbol{y}$ at the legitimate Receiver and $\boldsymbol{z}$ at the warden are observed with probability

$$P_{\boldsymbol{YZ}|\boldsymbol{X}}(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{x}) = \prod_{t=1}^{n} P_{YZ|X}(y_t, z_t|x_t), \tag{2.6b}$$

where $P_{YZ|X}$ is a random transformation from $\mathcal{X}$ to $\mathcal{Y} \times \mathcal{Z}$ given as a parameter of the problem. From (2.6b), it follows that the channel is memoryless. This channel is represented in Figure 2.2.

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

$$\mathcal{W} \triangleq \{1, 2, \ldots, M\}, \tag{2.7}$$

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

Definition 1 ((n, M, ϵ)-code)**.** *Given $(M, n) \in \mathbb{N}^2$ and $\epsilon \in [0, 1]$, an (n, M, ϵ)-code is a system*

$$\{(\boldsymbol{u}(1), \mathcal{D}(1)), (\boldsymbol{u}(2), \mathcal{D}(2)), \ldots, (\boldsymbol{u}(M), \mathcal{D}(M))\}, \tag{2.8}$$

where for all $(i, j) \in \mathcal{W}^2$ with $i \neq j$,

$$\boldsymbol{u}(i) = (u_1(i), u_2(i), \ldots, u_n(i)) \in \mathcal{X}^n, \tag{2.9}$$

$$\mathcal{D}(i) \cap \mathcal{D}(j) = \emptyset, \tag{2.10}$$

$$\bigcup_{k=1}^{M} \mathcal{D}(k) \subseteq \mathcal{Y}^n, \text{ and} \tag{2.11}$$

$$\frac{1}{M} \sum_{i=1}^{M} \Pr\left[\boldsymbol{Y} \in \mathcal{D}^{\mathsf{c}}(i) \middle| \boldsymbol{X} = \boldsymbol{u}(i)\right] \leqslant \epsilon. \tag{2.12}$$

The probability in (2.12) applies with respect to the marginal $P_{\boldsymbol{Y}|\boldsymbol{X}}$ of the joint distribution in (2.6); and $\mathcal{D}^{\mathsf{c}}(i)$ in (2.12) represents the complement of $\mathcal{D}(i)$ with respect to $\mathcal{Y}^n$.

Given a code represented by the system in (2.8), 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. The Receiver observes the output $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ after n channel uses and determines that the message index i was transmitted if it satisfies the decoding rule

$$\boldsymbol{y} \in \mathcal{D}(i). \tag{2.13}$$

The average decoding error probability associated to the code at the Receiver is given by the term in the left-hand side of (2.12). This system is depicted in Figure 2.2.

In the remainder of this section on point-to-point channels, the focus will be on (n, M, ϵ)-codes that satisfy a covertness constraint, *i.e.*, the adversary or the warden must remain unaware of the transmission. These covert codes are formally described in the next section.

Covert Codes

Assume that the input alphabet $\mathcal{X}$ contains an "off" symbol denoted by x_0. Let $Q_{\boldsymbol{Z}}$ and $R_{\boldsymbol{Z}}$ be respectively the probability distributions

$$Q_{\boldsymbol{Z}}(\boldsymbol{z}) = P_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x_0}), \text{ and} \tag{2.14}$$

$$R_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{M}\sum_{i=1}^{M} P_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{u}(i)), \tag{2.15}$$

where $\boldsymbol{x_0} = (x_0, x_0, \ldots, x_0)$ denotes an n-dimensional vector that consists exclusively of "off" symbols. Consider a hypothesis test in which the warden aims to determine whether the Transmitter is off (hypothesis H_0) or the (n, M, ϵ)-code (hypothesis H_1) is used upon an observation $\boldsymbol{z}$ of the channel output. That is,

$$\begin{cases} H_0 : \boldsymbol{Z} \sim Q_{\boldsymbol{Z}} \\ H_1 : \boldsymbol{Z} \sim R_{\boldsymbol{Z}}, \end{cases} \tag{2.16}$$

where $Q_{\boldsymbol{Z}}$ and $R_{\boldsymbol{Z}}$ are respectively given in (2.14) and (2.15).

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

$$T(\boldsymbol{z}) \triangleq \begin{cases} 0 & \text{if } H_0 \text{ is accepted,} \\ 1 & \text{if } H_1 \text{ is accepted.} \end{cases} \tag{2.17}$$

That is,

$$\alpha \triangleq \Pr\left[T(\boldsymbol{Z}) = 1\right], \text{ and} \tag{2.18}$$

$$\beta \triangleq \Pr\left[T(\boldsymbol{Z}) = 0\right], \tag{2.19}$$

where the probability in (2.18) applies assuming that $\boldsymbol{Z} \sim Q_{\boldsymbol{Z}}$ and the probability in (2.19) applies assuming that $\boldsymbol{Z} \sim R_{\boldsymbol{Z}}$.

Lemma 1 ([37, Theorem 13.1.1])**.** *The test to distinguish between the two distributions in* (2.16) *satisfies:*

$$\|Q_{\boldsymbol{Z}} - R_{\boldsymbol{Z}}\|_{\mathrm{TV}} \geqslant 1 - \alpha - \beta, \tag{2.20}$$

with equality for the optimal test.

Lemma 2 ([41, Lemma 11.6.1])**.** *Given two distributions P_X and Q_X on $\mathcal{X}$, it holds that*

$$\|P_X - Q_X\|_{\mathrm{TV}} \leqslant \sqrt{\frac{1}{2} D(P_X||Q_X)}. \tag{2.21}$$

A consequence of Lemma 1 and Lemma 2 is

$$1-\alpha-\beta\leqslant\sqrt{\frac{1}{2}D(Q_{\boldsymbol{Z}}||R_{\boldsymbol{Z}})}. \tag{2.22}$$

The above results suggests that the total variation and the Kullback-Leibler divergence are suitable covertness metrics. Indeed by guaranteeing that the total variation or Kullback-Leibler divergence are small enough, one ensures that the hypothesis test will almost always fail. That is, the type-I and type-II error probabilities almost sum up to one. In addition, covertness could also be guaranteed by lower bounding the type-II error probability β for a fixed type-I error probability α. This also leads to failures of the test in almost every case. These observations lead to the following definitions of covert codes.

Definition 2 ($(n,M,\epsilon,\delta)_{\mathrm{KL}}$-covert code)**.** *Given $\delta\in[0,1]$, an (n,M,ϵ)-code described by* (2.8) *is said to be an $(n,M,\epsilon,\delta)_{\mathrm{KL}}$-covert code if*

$$D(Q_{\boldsymbol{Z}}||R_{\boldsymbol{Z}})\leqslant\delta, \tag{2.23}$$

where $Q_{\boldsymbol{Z}}$ and $R_{\boldsymbol{Z}}$ are respectively defined in (2.14) *and* (2.15)*.*

Definition 3 ($(n,M,\epsilon,\delta)_{\mathrm{TV}}$-covert code)**.** *Given $\delta\in[0,1]$, an (n,M,ϵ)-code described by* (2.8) *is said to be an $(n,M,\epsilon,\delta)_{\mathrm{TV}}$-covert code if*

$$\|Q_{\boldsymbol{Z}}-R_{\boldsymbol{Z}}\|_{\mathrm{TV}}\leqslant\delta, \tag{2.24}$$

where $Q_{\boldsymbol{Z}}$ and $R_{\boldsymbol{Z}}$ are respectively defined in (2.14) *and* (2.15)*.*

Note that from Lemma 2, it follows that an $(n,M,\epsilon,\delta)_{\mathrm{KL}}$-covert code is an $(n,M,\epsilon,\sqrt{\frac{1}{2}\delta})_{\mathrm{TV}}$-covert code.

Definition 4 ($(n,M,\epsilon,\delta,\alpha)_{\beta}$-covert code)**.** *Given $(\alpha,\delta)\in[0,1]^2$, an (n,M,ϵ)-code described by* (2.8) *is said to be an $(n,M,\epsilon,\delta,\alpha)_{\beta}$-covert code if*

$$1-\alpha-\delta\leqslant\beta_{\alpha}(Q_{\boldsymbol{Z}},R_{\boldsymbol{Z}}), \tag{2.25}$$

where $Q_{\boldsymbol{Z}}$ and $R_{\boldsymbol{Z}}$ are respectively defined in (2.14) *and* (2.15) *and $\beta_{\alpha}(Q_{\boldsymbol{Z}},R_{\boldsymbol{Z}})=\inf_{\mathcal{T}\subset\mathcal{Z}^n:Q_{\boldsymbol{Z}}(\mathcal{Z}^n\setminus\mathcal{T})\leqslant\alpha}R_{\boldsymbol{Z}}(\mathcal{T})$.*

Covert Codes with Secret Key

To covertly exchange a message, the Transmitter and the Receiver might share a secret key $A\in\mathcal{A}\triangleq\{1,2,\ldots,L\}$, with $L\in\mathbb{N}$, unknown to the warden. This follows conventional principles in security known as Kerckhoffs' principles [42]. Hence, the Transmitter and the Receiver use an (n,M,L,ϵ)-code defined as follows.

Definition 5 ((n,M,L,ϵ)-code)**.** *Given $(M,L,n)\in\mathbb{N}^3$ and $\epsilon\in[0,1]$, an (n,M,L,ϵ)-code is a system*

$$\Big\{(\boldsymbol{u}(1,1),\mathcal{D}(1,1)),(\boldsymbol{u}(1,2),\mathcal{D}(1,2)),\ldots,(\boldsymbol{u}(M,L),\mathcal{D}(M,L))\Big\}, \tag{2.26}$$

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

$$\boldsymbol{u}(i,j) = (u_1(i,j), u_2(i,j), \ldots, u_n(i,j)) \in \mathcal{X}^n, \tag{2.27}$$

$$\mathcal{D}(i,j) \cap \mathcal{D}(k,j) = \emptyset, \tag{2.28}$$

$$\bigcup_{p=1}^{M} \mathcal{D}(p,j) \subseteq \mathcal{Y}^n, \text{ and} \tag{2.29}$$

$$\frac{1}{MK} \sum_{i=1}^{M} \sum_{j=1}^{K} \Pr\left[\boldsymbol{Y} \in \mathcal{D}^{\mathsf{c}}(i,j) \middle| \boldsymbol{X} = \boldsymbol{u}(i,j)\right] \leqslant \epsilon, \tag{2.30}$$

The probability in (2.30) is with respect to the marginal $P_{\boldsymbol{Y}|\boldsymbol{X}}$ of the joint distribution in (2.6); and $\mathcal{D}^{\mathsf{c}}(i,j)$ in (2.30) represents the complement of $\mathcal{D}(i,j)$ with respect to $\mathcal{Y}^n$.

Given a code represented by the system in (2.26), the Transmitter uses the codeword $\boldsymbol{u}(i,j)$ to transmit the message index $i \in \mathcal{W}$ using the key index $j \in \mathcal{A}$. At channel use t, with $t \in \{1,2,\ldots,n\}$, the Transmitter inputs the symbol $u_t(i,j)$ to the channel. The Receiver observes the output $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ after n channel uses and determines that the message index i was transmitted using key index j if it satisfies the decoding rule:

$$\boldsymbol{y} \in \mathcal{D}(i,j). \tag{2.31}$$

The average decoding error probability associated to the code is given by the term in the left-hand side of (2.30).

As in the case without key, Lemma 1 and Lemma 2 lead to the following different definitions of covert codes.

Definition 6 ($(n,M,L,\epsilon,\delta)_{\mathrm{KL}}$-covert code)**.** *Given* $\delta \in [0,1]$*, an* (n,M,L,ϵ)*-code described by* (2.8) *is said to be an* $(n,M,L,\epsilon,\delta)_{\mathrm{KL}}$*-covert code if*

$$D(Q_{\boldsymbol{Z}} || R_{\boldsymbol{Z}}) \leqslant \delta, \tag{2.32}$$

where $Q_{\boldsymbol{Z}}$ *and* $R_{\boldsymbol{Z}}$ *are respectively defined in* (4.10) *and* (4.11)*.*

Definition 7 ($(n,M,L,\epsilon,\delta)_{\mathrm{TV}}$-covert code)**.** *Given* $\delta \in [0,1]$*, an* (n,M,L,ϵ)*-code described by* (2.8) *is said to be an* $(n,M,L,\epsilon,\delta)_{\mathrm{TV}}$*-covert code if*

$$\|Q_{\boldsymbol{Z}} - R_{\boldsymbol{Z}}\|_{\mathrm{TV}} \leqslant \delta, \tag{2.33}$$

where $Q_{\boldsymbol{Z}}$ *and* $R_{\boldsymbol{Z}}$ *are respectively defined in* (4.10) *and* (4.11)*.*

Definition 8 ($(n,M,L,\epsilon,\delta,\alpha)_{\beta}$-covert code)**.** *Given* $(\alpha,\delta) \in [0,1]^2$*, an* (n,M,L,ϵ)*-code described by* (2.8) *is said to be an* $(n,M,L,\epsilon,\delta,\alpha)_{\beta}$*-covert code if*

$$1 - \alpha - \delta \leqslant \beta_\alpha(Q_{\boldsymbol{Z}}, R_{\boldsymbol{Z}}), \tag{2.34}$$

where $Q_{\boldsymbol{Z}}$ *and* $R_{\boldsymbol{Z}}$ *are respectively defined in* (4.10) *and* (4.11) *and* $\beta_\alpha(Q_{\boldsymbol{Z}}, R_{\boldsymbol{Z}}) = \inf_{\mathcal{T} \subset \mathcal{Z}^n : Q_{\boldsymbol{Z}}(\mathcal{Z}^n \setminus \mathcal{T}) \leqslant \alpha} R_{\boldsymbol{Z}}(\mathcal{T})$*.*

Fundamental Limits

The information rate at which information can be covertly transmitted to the Receiver using a covert code is $\frac{\log(M)}{n}$ bits per channel use. Thus, a fundamental limit on the rate at which information can be covertly transmitted is given by the largest possible M for which a covert code exists. This notion is formalized by the following definition for different covertness constraints when no secret key is required.

Definition 9 (Largest covert code's size)**.** *Fix a pair* $(\epsilon, \delta) \in [0,1]^2$*. The largest covert code's sizes for the different covertness criteria are:*

$$M^*_{\mathrm{KL}}(n,\epsilon,\delta) \triangleq \max\left\{M \in \mathbb{N} : \exists\, (n, M, \epsilon, \delta)_{\mathrm{KL}}\text{-}\textit{covert code}\right\}, \tag{2.35}$$

$$M^*_{\mathrm{TV}}(n,\epsilon,\delta) \triangleq \max\{M \in \mathbb{N} : \exists\, (n, M, \epsilon, \delta)_{\mathrm{TV}}\text{-}\textit{covert code}\}, \ \textit{and} \tag{2.36}$$

$$M^*_{\beta}(n,\epsilon,\delta,\alpha) \triangleq \max\left\{M \in \mathbb{N} : \exists\, (n, M, \epsilon, \delta, \alpha)_{\beta}\text{-}\textit{covert code}\right\}. \tag{2.37}$$

In case a secret key is used, the fundamental limits are characterized by the following quantities.

Definition 10 (Largest covert code's size and smallest key size)**.** *Fix a pair* $(\epsilon, \delta) \in [0,1]^2$*. The largest covert code's sizes for the different covertness criteria are:*

$$M^*_{\mathrm{KL}}(n,L,\epsilon,\delta) \triangleq \max\{M \in \mathbb{N} : \exists\, (n, M, L, \epsilon, \delta)_{\mathrm{KL}}\text{-}\textit{covert code}\}, \tag{2.38}$$

$$M^*_{\mathrm{TV}}(n,L,\epsilon,\delta) \triangleq \max\{M \in \mathbb{N} : \exists\, (n, M, L, \epsilon, \delta)_{\mathrm{TV}}\text{-}\textit{covert code}\}, \ \textit{and} \tag{2.39}$$

$$M^*_{\beta}(n,L,\epsilon,\delta,\alpha) \triangleq \max\{M \in \mathbb{N} : \exists\, (n, M, L, \epsilon, \delta, \alpha)_{\beta}\text{-}\textit{covert code}\}. \tag{2.40}$$

The smallest key sizes for the different covertness criteria are:

$$L^*_{\mathrm{KL}}(n,M,\epsilon,\delta) \triangleq \min\{L \in \mathbb{N} : \exists\, (n, M, L, \epsilon, \delta)_{\mathrm{KL}}\text{-}\textit{covert code}\}, \tag{2.41}$$

$$L^*_{\mathrm{TV}}(n,M,\epsilon,\delta) \triangleq \min\{L \in \mathbb{N} : \exists\, (n, M, L, \epsilon, \delta)_{\mathrm{TV}}\text{-}\textit{covert code}\}, \ \textit{and} \tag{2.42}$$

$$L^*_{\beta}(n,M,\epsilon,\delta,\alpha) \triangleq \min\{L \in \mathbb{N} : \exists\, (n, M, L, \epsilon, \delta, \alpha)_{\beta}\text{-}\textit{covert code}\}. \tag{2.43}$$

Main Results in AWGN Channels

Assume that $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathbb{R}$. Assume also that the random transformation in (2.6) satisfies:

$$P_{YZ|X}(y,z|x) = P_{Y|X}(y|x)P_{Z|X}(z|x), \tag{2.44a}$$

with

$$P_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{(y-x)^2}{2\sigma_Y^2}\right), \text{ and} \tag{2.44b}$$

$$P_{Z|X}(z|x) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp\left(-\frac{(z-x)^2}{2\sigma_Z^2}\right), \tag{2.44c}$$

where $\sigma_Y \in \mathbb{R}$ and $\sigma_Z \in \mathbb{R}$. That is, the channels are additive Gaussian noise channels with noise variances σ_Y^2 and σ_Z^2 respectively.

The problem of covert communications was first introduced in [7] and [8]. Therein, it is shown that one can covertly transmit on the order of $\sqrt{n}$ bits per n uses of the channel.

Theorem 1 ([7, 8])**.** *Consider the random transformation in* (2.6) *subject to* (2.44)*. In the asymptotic block-length regime, given a key A of sufficient length, it holds that*

$$\frac{\log\left(M^*_{\mathrm{KL}}(n, L, \epsilon, \delta)\right)}{\sqrt{n}} = O(1). \tag{2.45}$$

In addition, [8] provides a code construction that requires a key of length $O(\sqrt{n}\log(n))$. Theorem 1 suggests that if more than $O(\sqrt{n})$ bits were sent over the channel, either the covertness constraint or the reliability constraint would not be satisfied. That is, either the warden often detecs the communications, or the Receiver often fails to decode the covert message.

In [12] the exact constant that characterizes the ratio $\frac{\log(M)}{\sqrt{n}}$ for AWGN channels is derived.

Theorem 2 ([12])**.** *Consider the random transformation in* (2.6) *subject to* (2.44)*, and suppose that $Y = Z$. Then, it holds that*

$$\lim_{n\to\infty} \frac{\log\left(M^*_{\mathrm{KL}}(n, L, \epsilon, \delta)\right)}{\sqrt{n}} = \sqrt{\delta}. \tag{2.46}$$

Note that in [12], it is assumed that an infinitely long key is used to perform covert communications.

Theorem 2 establishes the optimal scaling constant of $\log\left(M^*_{\mathrm{KL}}(n, L, \epsilon, \delta)\right)$ with respect to $\sqrt{n}$.

Main Results in Discrete Memoryless Channels

In this section, the alphabets $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ are assumed to be countable. In addition, the random transformation in (2.6) is assumed to satisfy the product decomposition in (2.44a).

First Order Asymptotics

The first order asymptotics of covert communications for DMCs have been studied in [9, 10, 11] and [12]. First, results for binary channels are given. Then, extensions of these results to general DMCs are presented.

In [11], the first order constant for binary channels is obtained.

Theorem 3 ([11])**.** *Consider the random transformation in* (2.6)*. If $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{0, 1\}$, it holds that*

$$\lim_{n\to\infty} \frac{\log\left(M^*_{\mathrm{KL}}(n, L, \epsilon, \delta)\right)}{\sqrt{n}} = \sqrt{\frac{2\delta}{\chi_2(P_{Z|X=1}, P_{Z|X=0})}} D\big(P_{Y|X=1}||P_{Y|X=0}\big). \tag{2.47}$$

The above theorem characterizes the optimal scaling constant of $\log\left(M^*_{\mathrm{KL}}(n, L, \epsilon, \delta)\right)$ with respect to $\sqrt{n}$. The achievability scheme proposed in the proof of Theorem 3 uses a key that satisfies:

$$\lim_{n\to\infty} \frac{\log\left(L^*_{\mathrm{KL}}(n, M, \epsilon, \delta)\right)}{\sqrt{n}} = \sqrt{\frac{2\delta}{\chi_2(P_{Z|X=1}, P_{Z|X=0})}} \left[D\big(P_{Z|X=1}||P_{Z|X=0}\big) - D\big(P_{Y|X=1}||P_{Y|X=0}\big)\right]^+, \tag{2.48}$$

where $[\cdot]^+$ is $\max\{\cdot, 0\}$. From (2.48), it follows that when the warden observes the same output as the legitimate Receiver or a degraded output with respect to that of the legitimate Receiver, covert communications can be achieved without sharing a secret key.

This result extends to arbitrary DMCs as shown in [12] and in [11]:

Theorem 4 ([12], [11])**.** *Consider the random transformation in* (2.6)*.*

$$\lim_{n\to\infty} \frac{\log\left(M^*_{\mathrm{KL}}(n,\epsilon,\delta)\right)}{\sqrt{n}} = \max_{\tilde{P}_X:\tilde{P}_X(x_0)=0} \sqrt{\frac{2\delta}{\chi_2(\tilde{P}_Z, P_{Z|X=x_0})}} \sum_{x\in\mathcal{X}} \tilde{P}_X(x) D\big(P_{Y|X=x}||P_{Y|X=x_0}\big), \tag{2.49}$$

where $\tilde{P}_Z(z) = \sum_{x\in\mathcal{X}} \tilde{P}_X(x) P_{Z|X}(z|x)$*.*

In [12], it assumed that $Y = Z$ and thus no secret key is required to perform covert communications according to (2.48). However, the authors assume that an arbitrarily long key is available since the key analysis is not in the scope of the paper. On the other hand, [11] presents an achievability scheme in which the key length satisfies:

$$\lim_{n\to\infty} \frac{\log\left(L^*_{\mathrm{KL}}(n,M,\epsilon,\delta)\right)}{\sqrt{n}} \leqslant \sqrt{\frac{2\delta}{\chi_2(\tilde{P}_Z, P_{Z|X=x_0})}} \left[\sum_{x\in\mathcal{X}} \tilde{P}_X(x) D\big(P_{Z|X=x}||P_{Z|X=x_0}\big) - D\big(P_{Y|X=x}||P_{Y|X=x_0}\big)\right]^+, \tag{2.50}$$

where $\tilde{P}_X$ is the distribution that optimizes the left-hand side of (2.49) and $\tilde{P}_Z(z) = \sum_{x\in\mathcal{X}} \tilde{P}_X(x) P_{Z|X}(z|x)$.

The same conclusions as in the binary case can be found in the general DMC case. That is, Theorem 4 gives the optimal scaling constant of $\log\left(M^*(n,\epsilon,\delta)\right)$ with respect to $\sqrt{n}$ and from (2.50), it follows that the secret-key is not required if the warden observes a degraded channel output with respect to that of the legitimate Receiver.

Second Order Asymptotics

The second order asymptotics have been studied in [13] for binary channels with a variety of covertness constraints. In this section, results are given considering the maximum decoding error probability.

The following theorem characterizes the covert communication rate under a Kullback-Leibler covertness constraint.

Theorem 5 ([13])**.** *Consider the random transformation in* (2.6)*. If* $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{0,1\}$*, it holds that*

$$\frac{\log\left(M^*_{\mathrm{KL}}(n,\epsilon,\delta)\right)}{\sqrt{n}} = \sqrt{\frac{2\delta}{\chi_2(P_{Z|X=1}, P_{Z|X=0})}} D\big(P_{Y|X=1}||P_{Y|X=0}\big) - \sqrt{\sqrt{\frac{2\delta}{\chi_2(P_{Z|X=1}, P_{Z|X=0})}} \mathbb{V}_{Y|X=1}\left[\log\left(\frac{P_{Y|X=1}(Y)}{P_{Y|X=0}(Y)}\right)\right]} \frac{Q^{-1}(\epsilon)}{n^{\frac{1}{4}}} + O\left(\frac{\log(n)}{\sqrt{n}}\right). \tag{2.51}$$

In Theorem 5, note that the first term in the right-hand side of (2.51) is the optimal scaling constant of $\log\left(M^*(n,\epsilon,\delta)\right)$ with respect to $\sqrt{n}$. The interest of an expression such as (2.51)

is that it is a finite block-length result. That is, the result is valid not only in the asymptotic block-length regime, but also for smaller values of n. In this case, given n finite, one can compute a strictly positive covert communication rate $\frac{\log(M)}{n}$.

The following theorem characterizes the covert communication rate under a total variation covertness constraint.

Theorem 6 ([13])**.** *Consider the random transformation in* (2.6)*. If* $\mathcal{X}=\mathcal{Y}=\mathcal{Z}=\{0,1\}$*, it holds that*

$$\frac{\log\left(M^*_{\mathrm{TV}}(n,\epsilon,\delta)\right)}{\sqrt{n}} \leqslant \frac{2\Gamma D\big(P_{Y|X=1}||P_{Y|X=0}\big)}{\sqrt{\chi_2(P_{Z|X=1},P_{Z|X=0})}} - \sqrt{\frac{2\Gamma \mathbb{V}_{Y|X=1}\left[\log\left(\frac{P_{Y|X=1}(Y)}{P_{Y|X=0}(Y)}\right)\right]}{\sqrt{\chi_2(P_{Z|X=1},P_{Z|X=0})}}}\frac{Q^{-1}(\epsilon)}{n^{\frac{1}{4}}} + O\left(\frac{\log(n)}{\sqrt{n}}\right), \tag{2.52}$$

and

$$\begin{aligned}\frac{\log\left(M^*_{\mathrm{TV}}(n,\epsilon,\delta)\right)}{\sqrt{n}} \geqslant & \frac{2\Gamma D\big(P_{Y|X=1}||P_{Y|X=0}\big)}{\sqrt{\chi_2(P_{Z|X=1},P_{Z|X=0})}} \\ & -n^{-\frac{1}{4}}\left(\sqrt{\frac{2\Gamma \mathbb{V}_{Y|X=1}\left[\log\left(\frac{P_{Y|X=1}(Y)}{P_{Y|X=0}(Y)}\right)\right]}{\sqrt{\chi_2(P_{Z|X=1},P_{Z|X=0})}}}Q^{-1}(\epsilon) + \frac{2\sqrt{\pi}\exp\left(\frac{\Gamma^2}{2}\right)D\big(P_{Y|X=1}||P_{Y|X=0}\big)}{\sqrt{\Gamma}\chi_2(P_{Z|X=1},P_{Z|X=0})^{\frac{1}{4}}}\right) \\ & +O\left(\frac{\log(n)}{\sqrt{n}}\right),\end{aligned} \tag{2.53}$$

where $\Gamma = Q^{-1}\left(\frac{1-\delta}{2}\right)$.

The following theorem characterizes the covert communication rate under a covertness constraint on the optimal probability of missed detection.

Theorem 7 ([13])**.** *Consider the random transformation in* (2.6)*. If* $\mathcal{X}=\mathcal{Y}=\mathcal{Z}=\{0,1\}$*, it holds that*

$$\begin{aligned}\frac{\log\left(M^*_{\beta}(n,\epsilon,\delta,\alpha)\right)}{\sqrt{n}} \leqslant & \frac{(\Lambda+\Psi)D\big(P_{Y|X=1}||P_{Y|X=0}\big)}{\sqrt{\chi_2(P_{Z|X=1},P_{Z|X=0})}} \\ & -\sqrt{\frac{(\Lambda+\Psi)\mathbb{V}_{Y|X=1}\left[\log\left(\frac{P_{Y|X=1}(Y)}{P_{Y|X=0}(Y)}\right)\right]}{\sqrt{\chi_2(P_{Z|X=1},P_{Z|X=0})}}}\frac{Q^{-1}(\epsilon)}{n^{\frac{1}{4}}} + O\left(\frac{\log(n)}{\sqrt{n}}\right),\end{aligned} \tag{2.54}$$

and

$$\begin{aligned}\frac{\log\left(M^*_{\beta}(n,\epsilon,\delta,\alpha)\right)}{\sqrt{n}} \geqslant & \frac{(\Lambda+\Psi)D\big(P_{Y|X=1}||P_{Y|X=0}\big)}{\sqrt{\chi_2(P_{Z|X=1},P_{Z|X=0})}} \\ & -n^{-\frac{1}{4}}\left(\sqrt{\frac{(\Lambda+\Psi)\mathbb{V}_{Y|X=1}\left[\log\left(\frac{P_{Y|X=1}(Y)}{P_{Y|X=0}(Y)}\right)\right]}{\sqrt{\chi_2(P_{Z|X=1},P_{Z|X=0})}}}Q^{-1}(\epsilon)\right. \\ & \left.+\frac{\sqrt{2\pi}\left(\exp\left(\frac{\Lambda^2}{2}\right)+\exp\left(\frac{\Psi^2}{2}\right)\right)D\big(P_{Y|X=1}||P_{Y|X=0}\big)}{\sqrt{\Lambda+\Psi}\chi_2(P_{Z|X=1},P_{Z|X=0})^{\frac{1}{4}}}\right)+O\left(\frac{\log(n)}{\sqrt{n}}\right),\end{aligned} \tag{2.55}$$

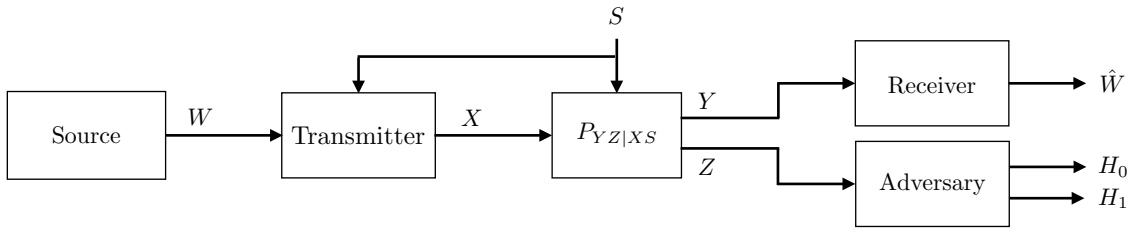

Figure 2.3.: Point-to-point channel with states and warden.

where $\Lambda = Q^{-1}(1-\delta-\alpha)$ *and* $\Psi = Q^{-1}(\alpha)$.

Theorem 6 and Theorem 7 give upper and lower bounds on the finite block-length rate achievable under different covertness constraints. Note that the first term (first oder term) in the upper and lower bounds given in the two theorems are the same. That is, in the asymptotic block-length regime, the first order term is the optimal scaling constant of $\log\left(M^*(n,\epsilon,\delta)\right)$ with respect to $\sqrt{n}$.

2.2.2. Point-to-point Channels with States

In this section, existing results on point-to-point channels are reviewed. First, the channel model is described. Then, results for the AWGN channels and discrete memoryless channels (DMC) are presented.

System Model

Consider a three-party communication system in which a transmitter sends information to a legitimate receiver while a second receiver (the warden) observes the channel and aims to determine whether or not communication occurs between the legitimate parties. The noisy communication medium is described by a product random transformation

$$\left(\mathcal{X}^n, \mathcal{S}^n, \mathcal{Y}^n, \mathcal{Z}^n, P_{\boldsymbol{YZ}|\boldsymbol{XS}}\right), \tag{2.56a}$$

where $n \in \mathbb{N}$ is the block-length; $\mathcal{X}$ is the input alphabet, $\mathcal{S}$ is the state alphabet, and $\mathcal{Y}$ and $\mathcal{Z}$ are the output alphabets. Given an input vector $\boldsymbol{x}$ and a state vector $\boldsymbol{s}$, the outputs $\boldsymbol{y}$ and $\boldsymbol{z}$ are observed at the legitimate receiver and at the warden, respectively, with probability

$$P_{\boldsymbol{YZ}|\boldsymbol{XS}}(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{x},\boldsymbol{s}) = \prod_{t=1}^{n} P_{YZ|XS}(y_t, z_t|x_t, s_t), \tag{2.56b}$$

with $P_{YZ|XS}$ a given parameter of the problem. That is, the channel is memoryless. Assume that the state $\boldsymbol{S}$ is distributed over $\mathcal{S}^n$ according to $P_{\boldsymbol{S}}$ with

$$P_{\boldsymbol{S}}(\boldsymbol{s}) = \prod_{t=1}^{n} P_S(s_t), \tag{2.57}$$

with P_S a given parameter of the problem. This channel is represented in Figure 2.3.

To covertly exchange a message, the Transmitter and the Receiver might share a secret key $A \in \mathcal{A} = \{1, 2, \ldots, L\}$, with $L \in \mathbb{N}$, unknown to the warden. The message index to be sent

from the Transmitter to the Receiver is a realization of a random variable W that is uniformly distributed in the set

$$\mathcal{W} \triangleq \{1, 2, \ldots, M\}, \tag{2.58}$$

with $M \in \mathbb{N}$.

Two scenarios might arise. Channel state information may indeed be causally or non-causally known at the Transmitter.

Causal State Information at the Transmitter

To send a message index within n channel uses, the Transmitter uses an (n, M, L, ϵ)-code defined as follows.

Definition 11 ((n, M, L, ϵ)-code)**.** *Given $(M, L, n) \in \mathbb{N}^3$ and $\epsilon \in [0, 1]$, an (n, M, L, ϵ)-code is a system*

$$\Big\{(\boldsymbol{u}(1, 1, \boldsymbol{s}_1), \mathcal{D}(1, 1)), (\boldsymbol{u}(1, 2, \boldsymbol{s}_1), \mathcal{D}(1, 2)), \ldots, (\boldsymbol{u}(M, L, \boldsymbol{s}_{|\mathcal{S}|^n}), \mathcal{D}(M, L))\Big\}, \tag{2.59}$$

where for all $(i, k, j, \boldsymbol{s}) \in \mathcal{W}^2 \times \mathcal{A} \times \mathcal{S}^n$ with $i \neq k$,

$$\boldsymbol{u}(i, j, \boldsymbol{s}) = (u_1(i, j, s_1), u_2(i, j, s_2), \ldots, u_n(i, j, s_n)) \in \mathcal{X}^n, \tag{2.60}$$

$$\mathcal{D}(i, j) \cap \mathcal{D}(k, j) = \emptyset, \tag{2.61}$$

$$\bigcup_{p=1}^{M} \mathcal{D}(p, j) \subseteq \mathcal{Y}^n, \text{ and} \tag{2.62}$$

$$\frac{1}{ML} \sum_{i=1}^{M} \sum_{j=1}^{L} \sum_{\boldsymbol{s} \in \mathcal{S}^n} P_{\boldsymbol{S}}(\boldsymbol{s}) \Pr\Big[\boldsymbol{Y} \in \mathcal{D}^{\mathsf{c}}(i, j) \Big| \boldsymbol{X} = \boldsymbol{u}(i, j, \boldsymbol{s}), \boldsymbol{S} = \boldsymbol{s}\Big] \leqslant \epsilon, \tag{2.63}$$

The probability in (2.63) is with respect to the marginal $P_{\boldsymbol{Y}|\boldsymbol{X}\boldsymbol{S}}$ of the joint distribution in (2.56); and $\mathcal{D}^{\mathsf{c}}(i, j)$ in (2.63) represents the complement of $\mathcal{D}(i, j)$ with respect to $\mathcal{Y}^n$.

Given a code, the Transmitter uses the codeword $\boldsymbol{u}(i, j, \boldsymbol{s})$ to transmit the message index $i \in \mathcal{W}$ under state $\boldsymbol{s} \in \mathcal{S}^n$ with the key index $j \in \mathcal{A}$. At channel use t, with $t \in \{1, 2, \ldots, n\}$, the Transmitter inputs the symbol $u_t(i, j, s_t)$ to the channel. The Receiver observes the output $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ after n channel uses and determines that the message index i was transmitted if it satisfies the decoding rule:

$$\boldsymbol{y} \in \mathcal{D}(i, j), \tag{2.64}$$

where $j \in \mathcal{A}$ is the key index. Note that the state is not known at the Receiver.

The average decoding error probability associated to the code is given in the left-hand side of (2.63).

Non-Causal State Information at the Transmitter

In this case, the Transmitter gets non-causally the knowledge of the vector $\boldsymbol{s}$. That is, at the beginning of the transmission the Transmitter knows the full vector $\boldsymbol{s}$. To send a message index within n channel uses, the Transmitter uses an (n, M, L, ϵ)-code defined as follows.

Definition 12 ((n, M, L, ϵ)-code)**.** *Given $(M, L, n) \in \mathbb{N}^3$, an (n, M, L, ϵ)-code is a system*

$$\Big\{(\boldsymbol{u}(1,1,\boldsymbol{s}_1), \mathcal{D}(1,1)), (\boldsymbol{u}(1,2,\boldsymbol{s}_1), \mathcal{D}(1,2)), \ldots, (\boldsymbol{u}(M,L,\boldsymbol{s}_{|\mathcal{S}|^n}), \mathcal{D}(M,L))\Big\}, \tag{2.65}$$

where for all $(i, k, j, \boldsymbol{s}) \in \mathcal{W}^2 \times \mathcal{A} \times \mathcal{S}^n$ with $i \neq k$,

$$\boldsymbol{u}(i,j,\boldsymbol{s}) = (u_1(i,j,\boldsymbol{s}), u_2(i,j,\boldsymbol{s}), \ldots, u_n(i,j,\boldsymbol{s})) \in \mathcal{X}^n, \tag{2.66}$$

$$\mathcal{D}(i,j) \cap \mathcal{D}(k,j) = \emptyset, \text{ and} \tag{2.67}$$

$$\bigcup_{p=1}^{M} \mathcal{D}(p,j) \subseteq \mathcal{Y}^n, \text{ and} \tag{2.68}$$

$$\lambda \triangleq \frac{1}{ML} \sum_{i=1}^{M} \sum_{j=1}^{L} \sum_{\boldsymbol{s} \in \mathcal{S}^n} P_{\boldsymbol{S}}(\boldsymbol{s}) \Pr\Big[\boldsymbol{Y} \in \mathcal{D}^{\mathsf{c}}(i,j) \Big| \boldsymbol{X} = \boldsymbol{u}(i,j,\boldsymbol{s}), \boldsymbol{S} = \boldsymbol{s}\Big] \leqslant \epsilon, \tag{2.69}$$

The probability in (2.69) is with respect to the marginal $P_{\boldsymbol{Y}|\boldsymbol{X}\boldsymbol{S}}$ of the joint distribution in (2.56); and $\mathcal{D}^{\mathsf{c}}(i,j)$ in (2.69) represents the complement of $\mathcal{D}(i,j)$ with respect to $\mathcal{Y}^n$.

Given a code, the Transmitter uses the codeword $\boldsymbol{u}(i,j,\boldsymbol{s})$ to transmit the message index $i \in \mathcal{W}$ under state $\boldsymbol{s} \in \mathcal{S}^n$ with the key index $j \in \mathcal{A}$. At channel use t, with $t \in \{1, 2, \ldots, n\}$, the Transmitter inputs the symbol $u_t(i,j,\boldsymbol{s})$ to the channel. The Receiver observes the output $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ after n channel uses and determines that the message index i was transmitted if it satisfies the decoding rule in (2.64) where $j \in \mathcal{A}$ is the key index.
The average decoding error probability associated to the code is given in (2.63). Figure 2.3 depicts this system.

In the remainder of this section, the focus will be on (n, M, L, ϵ)-codes that satisfy a covertness constraint, *i.e.*, the adversary or the warden must remain unaware of the transmission. These covert codes are formally described in the next section.

Covert Codes

Let $Q_{\boldsymbol{Z}}$ and $R_{\boldsymbol{Z}}$ be respectively the probability distributions

$$Q_{\boldsymbol{Z}}(\boldsymbol{z}) = \sum_{\boldsymbol{s} \in \mathcal{S}^n} P_{\boldsymbol{S}}(\boldsymbol{s}) P_{\boldsymbol{Z}|\boldsymbol{X}\boldsymbol{S}}(\boldsymbol{z}|\boldsymbol{x_0}, \boldsymbol{s}), \text{ and} \tag{2.70}$$

$$R_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{ML} \sum_{i=1}^{M} \sum_{j=1}^{L} \sum_{\boldsymbol{s} \in \mathcal{S}^n} P_{\boldsymbol{S}}(\boldsymbol{s}) P_{\boldsymbol{Z}|\boldsymbol{X}\boldsymbol{S}}(\boldsymbol{z}|\boldsymbol{u}(i,j,\boldsymbol{s}), \boldsymbol{s}), \tag{2.71}$$

where $\boldsymbol{x_0} = (x_0, x_0, \ldots, x_0)$ denotes a vector that consists exclusively of "off" symbols. Consider a hypothesis test in which the warden aims to determine whether the Transmitter is off (hypothesis H_0) or the (n, M, ϵ)-code (hypothesis H_1) is used upon the observation of the channel output $\boldsymbol{Z}$:

$$\begin{cases} H_0 : \boldsymbol{Z} \sim Q_{\boldsymbol{Z}} \\ H_1 : \boldsymbol{Z} \sim R_{\boldsymbol{Z}}, \end{cases} \tag{2.72}$$

where $Q_{\boldsymbol{Z}}$ and $R_{\boldsymbol{Z}}$ are respectively given in (2.70) and (2.71).

Denote by $\alpha \in [0,1]$ and $\beta \in [0,1]$ the type-I and type-II error probabilities associated with

a decision rule $T : \mathcal{Z}^n \to \{0,1\}$ of the form

$$T(\boldsymbol{z}) \triangleq \begin{cases} 0 & \text{if } H_0 \text{ is accepted,} \\ 1 & \text{if } H_1 \text{ is accepted.} \end{cases} \tag{2.73}$$

That is,

$$\alpha \triangleq \Pr\left[T(\boldsymbol{Z}) = 1\right], \text{ and} \tag{2.74}$$

$$\beta \triangleq \Pr\left[T(\boldsymbol{Z}) = 0\right], \tag{2.75}$$

where the probability in (2.74) applies assuming that $\boldsymbol{Z} \sim Q_{\boldsymbol{Z}}$ and the probability in (2.75) applies assuming that $\boldsymbol{Z} \sim R_{\boldsymbol{Z}}$.

Definition 13 ($(n, M, L, \epsilon, \delta)_{\mathrm{KL}}$-covert code)**.** *Given $\delta \in [0,1]$, an (n, M, ϵ)-code described by* (2.59) *or* (2.65) *is said to be an $(n, M, L, \epsilon, \delta)_{\mathrm{KL}}$-covert code if*

$$D(Q_{\boldsymbol{Z}} || R_{\boldsymbol{Z}}) \leqslant \delta, \tag{2.76}$$

where $Q_{\boldsymbol{Z}}$ and $R_{\boldsymbol{Z}}$ are respectively defined in (2.70) *and* (2.71)*.*

Fundamental Limits

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

Definition 14 (Largest covert code's size and smallest key size)**.** *Fix a pair $(\epsilon, \delta) \in [0,1]^2$. The largest covert code's size, denoted by $M^*_{\mathrm{KL}}(n, L, \epsilon, \delta)$, is:*

$$M^*_{\mathrm{KL}}(n, L, \epsilon, \delta) \triangleq \max\{M \in \mathbb{N} : \exists\, (n, M, L, \epsilon, \delta)_{\mathrm{KL}}\text{-covert code}\}.$$

The smallest key size, denoted by $L^*_{\mathrm{KL}}(n, M, \epsilon, \delta)$, is:

$$L^*_{\mathrm{KL}}(n, M, \epsilon, \delta) \triangleq \min\{L \in \mathbb{N} : \exists\, (n, M, L, \epsilon, \delta)_{\mathrm{KL}}\text{-covert code}\}.$$

Main Results in Discrete Memoryless Channels with Causal State Information at the Transmitter

Channels with states have been studied in [16] and [15].

When the channel state is causally available at the Transmitter, that is, when the Transmitter has the knowledge of the state S_t at time t, the following upper-bound holds.

Theorem 8 ([16])**.** *Given $\frac{\log(L)}{n} \geqslant 0$, it holds that*

$$\lim_{n\to\infty} \frac{\log\left(M^*_{\mathrm{KL}}(n, L, \epsilon, \delta)\right)}{n} \leqslant \max_{P_V,\ x(v,s)} I(V;Y), \tag{2.77}$$

with V an auxiliary random variable that satisfies $|\mathcal{V}| \leqslant \min\{|\mathcal{X}| + |\mathcal{Y}| + |\mathcal{Z}| - 2, (|\mathcal{X}| - 1)|\mathcal{S}| + 1\}$, and $P_Z = P_{Z|X=x_0}$.

When the channel state is causally available at the Transmitter, the following lower-bound holds.

Theorem 9 ([16]). *It holds that*

$$\lim_{n\to\infty} \frac{\log\left(M^*_{\mathrm{KL}}(n, L, \epsilon, \delta)\right)}{n} \geqslant \max_{P_V,\ x(v,s)} I(V;Y), \tag{2.78}$$

with V an auxiliary random variable that satisfies $|\mathcal{V}| \leqslant \min\{|\mathcal{X}| + |\mathcal{Y}| + |\mathcal{Z}| - 1, (|\mathcal{X}| - 1)|\mathcal{S}| + 2\}$, $P_Z = P_{Z|X=x_0}$, and

$$\frac{\log(L)}{n} > I(V;Z) - I(V;Y). \tag{2.79}$$

Main Results in Discrete Memoryless Channels with Non-causal Channel State Information at the Transmitter

When the channel state is non-causally available at the Transmitter, that is, when the Transmitter has the knowledge of the state sequence $\boldsymbol{S}$ at the beginning of the transmission, the following upper-bound holds.

Theorem 10 ([16]). *Given $\frac{\log(L)}{n} \geqslant 0$, it holds that*

$$\lim_{n\to\infty} \frac{\log\left(M^*_{\mathrm{KL}}(n, L, \epsilon, \delta)\right)}{n} \leqslant \max_{P_U, x(u,s)} I(U;Y) - I(U;S), \tag{2.80}$$

with U an auxiliary random variable that satisfies $|\mathcal{U}| \leqslant \min\{|\mathcal{X}| + |\mathcal{Y}| + |\mathcal{Z}| + |\mathcal{S}| - 3, |\mathcal{X}||\mathcal{S}|\}$, and $P_Z = P_{Z|X=x_0}$.

When the channel state is non-causally available at the Transmitter, the following lower-bound holds.

Theorem 11 ([16]). *It holds that*

$$\lim_{n\to\infty} \frac{\log\left(M^*_{\mathrm{KL}}(n, L, \epsilon, \delta)\right)}{n} \geqslant \max_{P_U, x(u,s)} I(U;Y) - I(U;S), \tag{2.81}$$

with U an auxiliary random variable that satisfies $|\mathcal{U}| \leqslant \min\{|\mathcal{X}| + |\mathcal{Y}| + |\mathcal{Z}| + |\mathcal{S}| - 2, |\mathcal{X}||\mathcal{S}| + 1\}$, and $P_Z = P_{Z|X=x_0}$, and

$$\frac{\log(L)}{n} > I(U;Z) - I(U;Y). \tag{2.82}$$

Interestingly, in the DMC with channel state information at the Transmitter, covert communications might be possible at strictly positive rates. This contrasts with the results obtained for the DMC where the rate vanishes with the block-length n.

Main Results in AWGN Channels

Assume that $S \sim \mathcal{N}(0, T)$, with $T \in \mathbb{R}_+$, and assume that the random transformation in (2.56) satisfies:

$$P_{YZ|XS}(y, z|x, s) = P_{Y|XS}(y|x, s) P_{Z|XS}(z|x, s), \tag{2.83}$$

with

$$P_{Y|XS}(y|x,s) = \frac{1}{\sqrt{2\pi(\sigma^2+T)}} \exp\left(-\frac{(y-x-s)^2}{2(\sigma^2+T)}\right), \tag{2.84}$$

and

$$P_{Z|XS}(z|x,s) = \frac{1}{\sqrt{2\pi(\sigma^2+T)}} \exp\left(-\frac{(z-x-s)^2}{2(\sigma^2+T)}\right). \tag{2.85}$$

Assume also that the channel input is subject to a power constraint of the form

$$\mathbb{E}_X\left[X^2\right] \leqslant P, \tag{2.86}$$

with $P \in \mathbb{R}_+$.

Define also

$$\gamma^* \triangleq \min\left\{1, \frac{P}{2T}\right\}, \tag{2.87}$$

$$T* \triangleq (1-\gamma^*)^2 T, \quad \text{and} \tag{2.88}$$

$$P^* \triangleq T - T^*. \tag{2.89}$$

When the channel state is causally available at the Transmitter, that is, when the Transmitter has the knowledge of the state S_t at time t, the following bound hold.

Theorem 12 ([16])**.** *If the secret key rate satisfies*

$$\frac{\log(L)}{n} > \frac{1}{2}\log\left(1+\frac{P^*}{T^*+\sigma^2}\right) - \frac{1}{2}\log\left(1+\frac{P^*}{T^*+1}\right), \tag{2.90}$$

then, the covert capacity with causal channel state information at the Transmitter is lower-bounded as

$$\lim_{n\to\infty} \frac{\log\left(M^*_{\mathrm{KL}}(n,L,\epsilon,\delta)\right)}{n} \geqslant \frac{1}{2}\log\left(1+\frac{P^*}{T^*+1}\right), \tag{2.91}$$

When the channel state is non causally available at the Transmitter, that is, when the Transmitter has the knowledge of the state vector $\boldsymbol{S}$ at the beginning of the transmission, the following holds.

Theorem 13 ([16])**.** *If the secret key rate satisfies*

$$\begin{aligned}\frac{\log(L)}{n} > \frac{1}{2}\log\left(1+\frac{\left(P^*+\frac{P^*}{P^*+1}T^*\right)^2}{\left(P^*+\left(\frac{P^*}{P^*+1}\right)^2 T^*\right)(P^*+T^*+\sigma^2)+\left(P^*+\frac{P^*}{P^*+1}T^*\right)^2}\right) \\ -\frac{1}{2}\log\left(1+\frac{\left(P^*+\frac{P^*}{P^*+1}T^*\right)^2}{\left(P^*+\left(\frac{P^*}{P^*+1}\right)^2 T^*\right)(P^*+T^*+1)+\left(P^*+\frac{P^*}{P^*+1}T^*\right)^2}\right),\end{aligned} \tag{2.92}$$

then the covert capacity with non-causal channel state information at the Transmitter is lower-bounded as

$$\lim_{n\to\infty} \frac{\log\left(M^*_{\mathrm{KL}}(n,L,\epsilon,\delta)\right)}{n} = \frac{1}{2}\log\left(1+P^*\right), \tag{2.93}$$

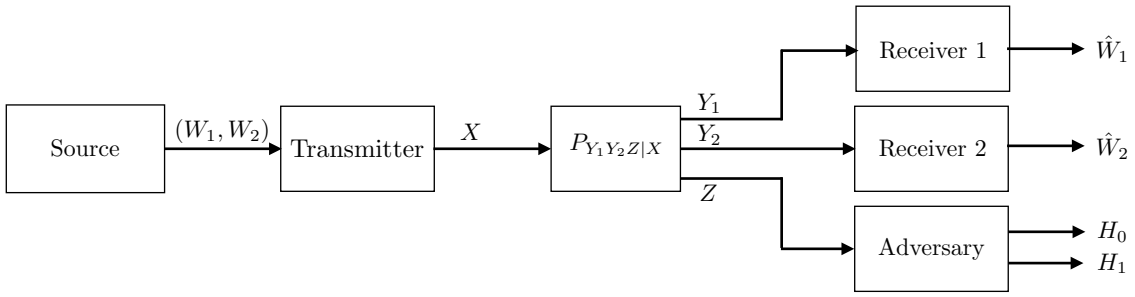

Figure 2.4.: Broadcast channel with a warden.

2.2.3. Broadcast Channels

This section reviews existing results on the broadcast channel. Several models can be used to study broadcast channels with covert messages. In particular two models are subsequently presented with the associated results in the next sections.

Broadcast Channels with covertness constraint

System Model

Consider a four-party communications system in which a transmitter broadcasts information to two legitimate receivers while a third receiver (the *warden*) observes the channel and aims to determine whether or not communication occurs between the legitimate parties. The noisy communication medium is described by a product random transformation

$$\left(\mathcal{X}^n, \mathcal{Y}_1^n, \mathcal{Y}_2^n, \mathcal{Z}^n, P_{\boldsymbol{Y}_1\boldsymbol{Y}_2\boldsymbol{Z}|\boldsymbol{X}}\right), \tag{2.94}$$

where $n \in \mathbb{N}$ is the block-length; $\mathcal{X}$ is the input alphabet with an "off" symbol x_0, $\mathcal{Y}_1$, $\mathcal{Y}_2$ and $\mathcal{Z}$ are the output alphabets. That is, given an input vector $\boldsymbol{x}$, the outputs $\boldsymbol{y}_k$, with $k \in \{1, 2\}$, and $\boldsymbol{z}$ are observed at Receiver k and at the warden, respectively, with probability

$$P_{\boldsymbol{Y}_1\boldsymbol{Y}_2\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{z}|\boldsymbol{x}) = \prod_{t=1}^{n} P_{Y_1Y_2Z|X}(y_{1t}, y_{2t}, z_t|x_t). \tag{2.95}$$

This channel is represented in Figure 2.4.

The message index pair to transmit is a realization of the random variable pair (W_1, W_2) that is uniformly distributed over $\mathcal{W}_1 \times \mathcal{W}_2 = \{1, 2, \ldots, M_1\} \times \{1, 2, \ldots, M_2\}$. The random variables W_1 and W_2 are assumed to be independent. To transmit a message index pair to the legitimate receivers, the Transmitter uses an (n, M_1, M_2, ϵ)-code defined as follows.

Definition 15 ((n, M_1, M_2, ϵ)-code)**.** *Given $(M_1, M_2, n) \in \mathbb{N}^3$ and $\epsilon \in [0, 1]$, an (n, M_1, M_2, ϵ)-code is a system*

$$\Big\{(\boldsymbol{u}(1,1), \mathcal{D}_1(1,1), \mathcal{D}_2(1,1)), (\boldsymbol{u}(1,2), \mathcal{D}_1(1,2), \mathcal{D}_2(1,2)), \ldots, (\boldsymbol{u}(M_1, M_2), \mathcal{D}_1(M_1, M_2), \mathcal{D}_2(M_1, M_2))\Big\}, \tag{2.96}$$

where for all $(i,j,k,l) \in (\mathcal{W}_1 \times \mathcal{W}_2)^2$ *with* $(i,j) \neq (k,l)$*, and all* $m \in \{1,2\}$

$$\boldsymbol{u}(i,j) = (u_1(i,j), u_2(i,j), \ldots, u_n(i,j)) \in \mathcal{X}^n, \tag{2.97}$$

$$\mathcal{D}_m(i,j) \cap \mathcal{D}_m(k,l) = \emptyset, \tag{2.98}$$

$$\bigcup_{p=1}^{M_1} \bigcup_{q=1}^{M_2} \mathcal{D}_m(p,q) \subseteq \mathcal{Y}_m^n, \text{ and} \tag{2.99}$$

$$\sum_{i=1}^{M_1} \sum_{j=1}^{M_2} \frac{\Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^c(i,j) \cup \boldsymbol{Y}_2 \in \mathcal{D}_2^c(i,j) | \boldsymbol{X} = \boldsymbol{u}(i,j)\right]}{M_1 M_2} \leqslant \epsilon. \tag{2.100}$$

The probability operator in (2.100) applies in with respect to the marginal $P_{\boldsymbol{Y}_1\boldsymbol{Y}_2|\boldsymbol{X}}$ of the probability distribution in (2.95).

Given an (n, M_1, M_2, ϵ)-code represented by the system in (2.96), the Transmitter sends the codeword $\boldsymbol{u}(i,j) \in \mathcal{X}^n$ to transmit the message index pair $(i,j) \in \mathcal{W}_1 \times \mathcal{W}_2$. In particular, at channel use $t \in \{1,2,\ldots,n\}$, the Transmitter inputs the symbol $u_t(i,j)$ to the channel. After n channel uses, Receiver k, with $k \in \{1,2\}$, observes the channel output vector $\boldsymbol{y}_k \in \mathcal{Y}_k^n$ and determines that the message index pair (i,j) was transmitted according if the following decoding rule holds:

$$\boldsymbol{y}_k \in \mathcal{D}_k(i,j). \tag{2.101}$$

The average decoding error probability denoted by $\lambda \in [0,1]$ is given in the left hand-side of (2.100).

Covert Codes

Let $Q_{\boldsymbol{Z}}$ and $R_{\boldsymbol{Z}}$ be respectively the probability distributions

$$Q_{\boldsymbol{Z}}(\boldsymbol{z}) = P_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x_0}), \text{ and} \tag{2.102}$$

$$R_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{M_1 M_2} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} P_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{u}(i,j)), \tag{2.103}$$

where $\boldsymbol{x_0} = (x_0, x_0, \ldots, x_0)$ denotes a vector that consists exclusively of "off" symbols. The warden aims at determining whether its observation of the channel is induced by the vector $\boldsymbol{x_0}$ or by a codeword from the (n, M_1, M_2, ϵ)-code. That is, the warden faces the following hypothesis test:

$$\begin{cases} H_0 : \boldsymbol{Z} \sim Q_{\boldsymbol{Z}} \\ H_1 : \boldsymbol{Z} \sim R_{\boldsymbol{Z}} \quad . \end{cases} \tag{2.104}$$

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

$$T(\boldsymbol{z}) \triangleq \begin{cases} 0 & \text{if } H_0 \text{ is accepted,} \\ 1 & \text{if } H_1 \text{ is accepted.} \end{cases} \tag{2.105}$$

That is,

$$\alpha \triangleq \Pr\left[T(\boldsymbol{Z}) = 1\right], \text{ and} \tag{2.106}$$
$$\beta \triangleq \Pr\left[T(\boldsymbol{Z}) = 0\right], \tag{2.107}$$

where the probability operator in (2.106) applies assuming that $\boldsymbol{Z} \sim Q_{\boldsymbol{Z}}$ and the probability operator in (2.107) applies assuming that $\boldsymbol{Z} \sim R_{\boldsymbol{Z}}$.

Definition 16 ($(n, M_1, M_2, \epsilon, \delta)_{\mathrm{KL}}$-covert code)**.** *Given $\delta \in [0, 1]$, an (n, M_1, M_2, ϵ)-code described by (2.96) is said to be an $(n, M_1, M_2, \epsilon, \delta)_{\mathrm{KL}}$-covert code if*

$$D(Q_{\boldsymbol{Z}} || R_{\boldsymbol{Z}}) \leqslant \delta, \tag{2.108}$$

where $Q_{\boldsymbol{Z}}$ and $R_{\boldsymbol{Z}}$ are respectively defined in (2.102) and (2.103).

To covertly communicate messages over the channel, the Transmitter and the legitimate receivers might share a secret key $A \in \mathcal{A} = \{1, 2, \ldots, L\}$ that is unknown to the warden.

Fundamental Limits

The information rate at which information can be covertly transmitted to Receiver k, $k \in \{1, 2\}$, using an $(n, M_1, M_2, \epsilon, \delta)$-covert code is $\frac{\log(M_k)}{n}$ bits per channel use. Thus, a fundamental limit on the rate at which information can be covertly transmitted is given by the largest possible M_k for which an $(n, M_1, M_2, \epsilon, \delta)$-covert code exists. For all M_k below the largest possible M_k, the ratio $\frac{\log(M_k)}{\sqrt{n}}$ is said to be achievable. This notion is formalized by the following definition.

Definition 17. *A tuple $(\rho_1, \rho_2, \rho_{\mathrm{key}})$ is said achievable if there exists a sequence of $(n, M_1, M_2, \epsilon, \delta)$-covert code such that for all $k \in \{1, 2\}$,*

$$\liminf_{n\to\infty} \frac{\log(M_k)}{\sqrt{n}} \geqslant \rho_k, \tag{2.109}$$
$$\limsup_{n\to\infty} \lambda_n \leqslant \epsilon, \text{ and} \tag{2.110}$$
$$\limsup_{n\to\infty} \frac{\log(K)}{\sqrt{n}} \leqslant \rho_{\mathrm{key}} \tag{2.111}$$

The covert-capacity region of the broadcast channel with a warden, denoted by $\mathcal{R}_\delta$ is the set of all achievable pairs $(\rho_1, \rho_2, \rho_{\mathrm{key}})$.

Main Results

In [30], it is shown that under covertness constraint, time division between the two legitimate receivers is an optimal communication scheme for a particular class of channels. Let ρ_k^*, with $k \in \{1, 2\}$ be the covert capacity of the point-to-point link between the Transmitter and Receiver k. Let also ρ_Z^* be the point-to-point covert capacity between the Transmitter and a Receiver that would observe the same channel output as the warden (see [12]). If $\rho_1^* \geqslant \rho_2^*$, then assume that

$$\max_{P_X} \frac{I(X; Y_1)}{I(X; Y_2)} \leqslant \frac{\rho_1^*}{\rho_2^*}, \tag{2.112a}$$

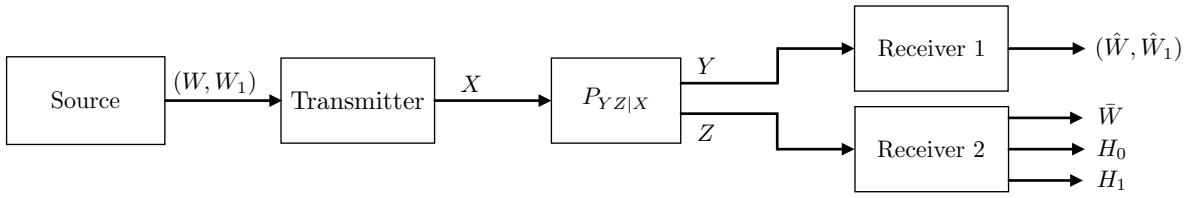

Figure 2.5.: Broadcast channel with covert messages.

otherwise, assume that

$$\max_{P_X} \frac{I(X;Y_2)}{I(X;Y_1)} \leqslant \frac{\rho_2^*}{\rho_1^*}. \tag{2.112b}$$

Theorem 14 ([30])**.** *Consider the random transformation in* (2.95) *such that* (2.112) *is satisfied. The tuple* $(\rho_1, \rho_2, \rho_{\text{key}}) \in \mathbb{R}_+^3$ *is achievable if and only if*

$$\frac{\rho_1}{\rho_1^*} + \frac{\rho_2}{\rho_2^*} \leqslant 1, \qquad \textit{and} \tag{2.113}$$

$$\rho_{\text{key}} \geqslant \left(\frac{\rho_1}{\rho_1^*} + \frac{\rho_2}{\rho_2^*}\right) \rho_Z^* - \rho_1 - \rho_2. \tag{2.114}$$

Theorem 14 shows that the square-root law also applies for the broadcast channel. In this case, time-sharing between the two users is the optimal transmissions strategy.

Embedding Covert Information into Broadcast Codes

System Model

Consider a three-party communications system in which a transmitter sends common information to two receivers and private information only to one receiver. The private information should be sent covertly with respect to the second receiver. That is, the second receiver acts as a warden regarding the private message and its goal is to determine if the codeword used is drawn from an innocent codebook or if it carries additional private information. The noisy communication medium is described by a product random transformation similar to that in (2.6). This channel is represented in Figure 2.5.

The common message index to transmit is a realization of the random variable W_1 that is uniformly distributed over $\mathcal{W}_1 = \{1, 2, \ldots, M_1\}$. The private message index to transmit is a realization of the random variable W_2 that is uniformly distributed over $\mathcal{W}_2 = \{1, 2, \ldots, M_2\}$, and independent of W_1. To transmit a message index to the legitimate Receiver, the Transmitter uses an (n, M_1, M_2, ϵ)-code defined as follows.

Definition 18 ((n, M_1, M_2, ϵ)-code)**.** *Given* $(M_1, M_2, n) \in \mathbb{N}^3$ *and* $\epsilon \in [0, 1]$*, an* (n, M_1, M_2, ϵ)*-code is a system*

$$\Big\{(\boldsymbol{u}(1,0), \mathcal{D}_1(1,0), \mathcal{D}_2(1)), (\boldsymbol{u}(1,1), \mathcal{D}_1(1,1), \mathcal{D}_2(1)), \ldots \\ , (\boldsymbol{u}(M_1, M_2), \mathcal{D}_1(M_1, M_2), \mathcal{D}_2(M_1))\Big\}, \tag{2.115}$$

where for all $(i,j,k,l) \in (\mathcal{W}_1 \times (\{0\} \cup \mathcal{W}_2))^2$ *with* $(i,j) \neq (k,l)$,

$$\boldsymbol{u}(i,j) = (u_1(i,j), u_2(i,j), \ldots, u_n(i,j)) \in \mathcal{X}^n, \tag{2.116}$$

$$\mathcal{D}_1(i,j) \cap \mathcal{D}_1(k,l) = \emptyset, \tag{2.117}$$

$$\mathcal{D}_2(i) \cap \mathcal{D}_2(k) = \emptyset, \tag{2.118}$$

$$\bigcup_{p=1}^{M_1} \bigcup_{q=1}^{M_2} \mathcal{D}_1(p,q) \subseteq \mathcal{Y}^n, \tag{2.119}$$

$$\bigcup_{p=1}^{M_1} \mathcal{D}_2(p) \subseteq \mathcal{Z}^n, \text{ and} \tag{2.120}$$

$$\frac{1}{M_1 M_2} \sum_{i=1}^{M_1} \sum_{j=0}^{M_2} \Pr\left[\boldsymbol{Y} \in \mathcal{D}_1^c(i,j) | \boldsymbol{X} = \boldsymbol{u}(i,j)\right] \leqslant \epsilon, \tag{2.121}$$

$$\frac{1}{M_1 M_2} \sum_{i=1}^{M_1} \sum_{j=0}^{M_2} \Pr\left[\boldsymbol{Z} \in \mathcal{D}_2^c(i) | \boldsymbol{X} = \boldsymbol{u}(i,j)\right] \leqslant \epsilon. \tag{2.122}$$

The probability in (2.121) is with respect to the marginal $P_{\boldsymbol{Y}|\boldsymbol{X}}$ of the probability distribution in (2.6). The probability in (2.122) is with respect to the marginal $P_{\boldsymbol{Z}|\boldsymbol{X}}$ of the probability distribution in (2.6).

The codewords $\boldsymbol{u}(i,0)$ with $i \in \mathcal{W}_1$ correspond to the innocent codewords whereas the codewords $\boldsymbol{u}(i,j)$ with $j \in \mathcal{W}_2$ correspond to the transmission of both common and private information.

Given an (n, M_1, M_2, ϵ)-code represented by the system in (2.115), the Transmitter sends the codeword $\boldsymbol{u}(i,j) \in \mathcal{X}^n$ to transmit the message index pair $(i,j) \in \mathcal{W}_1 \times \mathcal{W}_2$. In particular, at channel use $t \in \{1,2,\ldots,n\}$, the Transmitter inputs the symbol $u_t(i,j)$ to the channel. After n channel uses, Receiver 1 observes the channel output vector $\boldsymbol{y} \in \mathcal{Y}^n$ and determines that message index pair (i,j) was transmitted if the following decoding rule holds:

$$\boldsymbol{y} \in \mathcal{D}_1(i,j). \tag{2.123}$$

Receiver 2 on the other hand decides that the common message index i was transmitted according to the following decoding rule:

$$\boldsymbol{z} \in \mathcal{D}_2(i). \tag{2.124}$$

The average decoding error probabilities at Receiver 1 and Receiver 2 are given in the left hand-side of (2.121) and (2.122), respectively.

Covert Codes

Given $i \in \mathcal{W}_1$, let $Q_{\boldsymbol{Z}}^{(i)}$ and $R_{\boldsymbol{Z}}^{(i)}$ be respectively the probability distributions

$$Q_{\boldsymbol{Z}}^{(i)}(\boldsymbol{z}) = P_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{u}(i,0)), \text{ and} \tag{2.125}$$

$$R_{\boldsymbol{Z}}^{(i)}(\boldsymbol{z}) = \frac{1}{M_2} \sum_{j=1}^{M_2} P_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{u}(i,j)). \tag{2.126}$$

The warden aims at determining whether its observation of the channel is induced by the codeword $\boldsymbol{u}(i,0)$ or by a codeword $\boldsymbol{u}(i,j)$ with $j \in \mathcal{W}_2$. That is, the warden faces the following hypothesis test:

$$\begin{cases} H_0 : \boldsymbol{Z} \sim Q_{\boldsymbol{Z}}^{(i)} \\ H_1 : \boldsymbol{Z} \sim R_{\boldsymbol{Z}}^{(i)} \end{cases}. \tag{2.127}$$

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 $T_i : \mathcal{Z}^n \to \{0,1\}$ of the form

$$T_i(\boldsymbol{z}) \triangleq \begin{cases} 0 & \text{if } H_0 \text{ is accepted,} \\ 1 & \text{if } H_1 \text{ is accepted.} \end{cases} \tag{2.128}$$

That is,

$$\alpha_i \triangleq \Pr\left[T_i(\boldsymbol{Z}) = 1\right], \text{ and} \tag{2.129}$$
$$\beta_i \triangleq \Pr\left[T_i(\boldsymbol{Z}) = 0\right], \tag{2.130}$$

where the probability in (2.129) applies assuming that $\boldsymbol{Z} \sim Q_{\boldsymbol{Z}}^{(i)}$ and the probability operator in (2.130) applies assuming that $\boldsymbol{Z} \sim R_{\boldsymbol{Z}}^{(i)}$.

Definition 19 ($(n, M_1, M_2, \epsilon, \delta)_{\mathrm{KL}}$-covert code)**.** *Given $\delta \in [0,1]$, an (n, M_1, M_2, ϵ)-code $\mathcal{C}$ described by* (2.115) *is said to be an $(n, M_1, M_2, \epsilon, \delta)_{\mathrm{KL}}$-covert code if*

$$D\Big(Q_{\boldsymbol{Z}}^{(i)} || R_{\boldsymbol{Z}}^{(i)}\Big) \leqslant \delta, \tag{2.131}$$

where $Q_{\boldsymbol{Z}}^{(i)}$ and $R_{\boldsymbol{Z}}^{(i)}$ are respectively defined in (2.125) *and* (2.126)*.*

Fundamental Limits

The information rate at which information can be covertly transmitted to Receiver k, $k \in \{1,2\}$, using an $(n, M_1, M_2, \epsilon, \delta)$-covert code is $\frac{\log(M_k)}{n}$ bits per channel use. Thus, a fundamental limit on the rate at which information can be covertly transmitted is given by the largest possible M_k for which an $(n, M_1, M_2, \epsilon, \delta)$-covert code exists. This notion is formalized by the following definition.

Definition 20 (Largest covert code's size and smallest key size)**.** *Fix a pair $(\epsilon, \delta) \in [0,1]^2$. The largest covert code's size, denoted by $M_{1,\mathrm{KL}}^*(n, M_2, \epsilon, \delta)$ (resp. $M_{2,\mathrm{KL}}^*(n, M_1, \epsilon, \delta)$), is:*

$$M_{1,\mathrm{KL}}^*(n, M_2, \epsilon, \delta) \triangleq \max\{M_1 \in \mathbb{N} : \exists\, (n, M_1, M_2, L, \epsilon, \delta)_{\mathrm{KL}}\text{-covert code}\}, \text{ and} \tag{2.132}$$
$$M_{2,\mathrm{KL}}^*(n, M_1, \epsilon, \delta) \triangleq \max\{M_2 \in \mathbb{N} : \exists\, (n, M_1, M_2, L, \epsilon, \delta)_{\mathrm{KL}}\text{-covert code}\}. \tag{2.133}$$

Results

In [3], the number of covert bits that can be embedded in an innocent code that performs close to capacity is established for binary channels. Hence, assume that $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{0,1\}$. Assume also that the capacity achieving distribution of the channel between the Transmitter and Receiver 2 (the warden) P_X^* is of the form $P_X^*(1) = p^*$. Finally, assume that under the input distribution P_X^*, it holds that $I(X;Y) \geqslant I(X;Z)$.

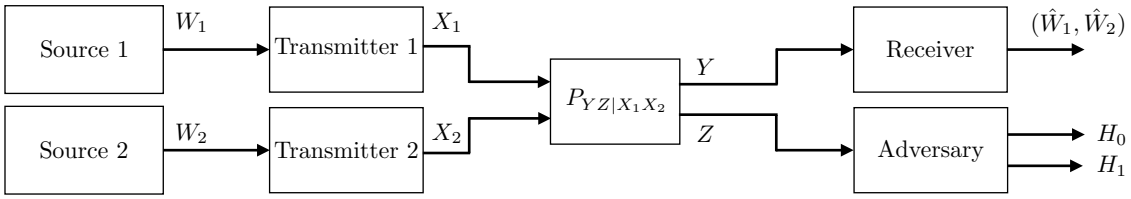

Figure 2.6.: Point-to-point channel with a warden.

Theorem 15. *Consider the random transformation in* (2.6) *such that* $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{0, 1\}$. *Then, if there exists* $\gamma \geqslant 0$ *such that*

$$\begin{aligned}(1-p^*)D\big(P_{Y|X=1}||P_{Y|X=0}\big) + p^*\gamma D\big(P_{Y|X=0}||P_{Y|X=1}\big) > \\ (1-p^*)D\big(P_{Z|X=1}||P_{Z|X=0}\big) + p^*\gamma D\big(P_{Z|X=0}||P_{Z|X=1}\big),\end{aligned} \tag{2.134}$$

there exists keyless covert communication schemes such that

$$\lim_{n\to\infty} \frac{\log\left(M_1^*(n, M_2, \epsilon, \delta)\right)}{n} = I(X;Y), \tag{2.135}$$

and for all $i \in \mathcal{W}_1$,

$$\begin{aligned}\lim_{n\to\infty} \frac{\log\left(M_2^*(n, M_1, \epsilon, \delta)\right)}{\sqrt{n}} = \\ \max_{\gamma\geqslant 0} \frac{\sqrt{2\delta}((1-p^*)D\big(P_{Y|X=1}||P_{Y|X=0}\big) + p^*\gamma D\big(P_{Y|X=0}||P_{Y|X=1}\big))}{\sqrt{(1-p^*)\chi_2(P_{Z|X=1}, P_{Z|X=0}) + p^*\gamma\chi_2(P_{Z|X=0}, P_{Z|X=1})}},\end{aligned} \tag{2.136}$$

$$\lim_{n\to\infty} \lambda_{1n} = \lim_{n\to\infty} \lambda_{2n} = \lim_{n\to\infty} D\Big(Q_{\boldsymbol{Z}}^{(i)}||T_{\boldsymbol{Z}}^{(i)}\Big) = 0. \tag{2.137}$$

Note that the error probability analysis in [3] is not performed using the average decoding error probabilities λ_{1n} and λ_{2n}. Nevertheless, both λ_{1n} and λ_{2n} are still guaranteed to tend to 0.

2.2.4. Multiple-Access Channels

In this section, covert communications over mutiple-access channels are reviewed. First, the channel model is described. Then, results for DMCs are subsequently exposed.

System Model

Consider a four-party communication system in which two transmitters send information to a legitimate receiver while a second receiver (the warden) observes the channel and aims to determine whether or not communication occurs between the legitimate parties (see Figure 2.6). The noisy communication medium is described by a product random transformation

$$\left(\mathcal{X}_1^n, \mathcal{X}_2^n, \mathcal{Y}^n, \mathcal{Z}^n, P_{\boldsymbol{YZ}|\boldsymbol{X}_1\boldsymbol{X}_2}\right), \tag{2.138a}$$

where $n \in \mathbb{N}$ is the block-length; $\mathcal{X}_1$ and $\mathcal{X}_2$ are the input alphabets, $\mathcal{Y}$ and $\mathcal{Z}$ are the output alphabets. That is, given two input vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, the outputs $\boldsymbol{y}$ and $\boldsymbol{z}$ are observed at the Receiver and the warden, respectively, with probability

$$P_{\boldsymbol{Y}\boldsymbol{Z}|\boldsymbol{X}_1\boldsymbol{X}_2}(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{x}_1,\boldsymbol{x}_2) = \prod_{t=1}^{n} P_{YZ|X_1X_2}(y_t,z_t|x_{1,t},x_{2,t}), \tag{2.138b}$$

where $P_{YZ|X_1X_2}$ is given parameter of the problem. That is, the channel is memoryless.

The message index to be sent from Transmitter k, with $k \in \{1,2\}$, to the Receiver is a realization of a random variable W_k that is uniformly distributed in the set

$$\mathcal{W}_k \triangleq \{1,2,\ldots,M_k\}, \tag{2.139}$$

with $M_k \in \mathbb{N}$. To send a message index within n channel uses, the Transmitter uses an (n,M_1,M_2,ϵ)-code.

Definition 21 ((n,M_1,M_2,ϵ)-code)**.** *Given $(M_1,M_2,n) \in \mathbb{N}^3$ and $\epsilon \in [0,1]$, an (n,M_1,M_2,ϵ)-code is a system*

$$\left\{(\boldsymbol{u}(1),\boldsymbol{v}(1),\mathcal{D}(1,1)),(\boldsymbol{u}(1),\boldsymbol{v}(2),\mathcal{D}(1,2)),\ldots,(\boldsymbol{u}(M_1),\boldsymbol{v}(M_2),\mathcal{D}(M_1,M_2))\right\}, \tag{2.140}$$

where for all $(i,k,j,l) \in \mathcal{W}_1^2 \times \mathcal{W}_2^2$ with $i \neq k$ and $j \neq l$,

$$\boldsymbol{u}(i) = (u_1(i),u_2(i),\ldots,u_n(i)) \in \mathcal{X}^n, \tag{2.141}$$

$$\boldsymbol{v}(i) = (v_1(i),v_2(i),\ldots,v_n(i)) \in \mathcal{X}^n, \tag{2.142}$$

$$\mathcal{D}(i,j) \cap \mathcal{D}(k,l) = \emptyset, \tag{2.143}$$

$$\bigcup_{p=1}^{M_1}\bigcup_{q=1}^{M_2} \mathcal{D}(p,q) \subseteq \mathcal{Y}^n, \text{ and} \tag{2.144}$$

$$\frac{1}{M_1M_2}\sum_{i=1}^{M_1}\sum_{j=1}^{M_2} \Pr\left[\boldsymbol{Y} \in \mathcal{D}^{\mathsf{c}}(i,j)\middle|\boldsymbol{X}_1 = \boldsymbol{u}(i), \boldsymbol{X}_2 = \boldsymbol{v}(j)\right] \leqslant \epsilon. \tag{2.145}$$

The probability operator in (2.145) applies with respect to the marginal $P_{\boldsymbol{Y}|\boldsymbol{X}_1\boldsymbol{X}_2}$ of the joint distribution in (2.138); and $\mathcal{D}^{\mathsf{c}}(i,j)$ in (2.145) represents the complement of $\mathcal{D}(i,j)$ with respect to $\mathcal{Y}^n$.

Given a code represented by the system in (2.140), the Transmitters use the codewords $\boldsymbol{u}(i)$ and $\boldsymbol{v}(j)$ to transmit the message index pair $(i,j) \in \mathcal{W}_1 \times \mathcal{W}_2$. At channel use t, with $t \in \{1,2,\ldots,n\}$, Transmitter 1 inputs the symbol $u_t(i)$ to the channel and Transmitter 2 inputs the symbol $v_t(j)$ to the channel. The Receiver observes the output $\boldsymbol{y} = (y_1,y_2,\ldots,y_n)$ after n channel uses and determines that the message index pair (i,j) was transmitted if the following decoding rule holds:

$$\boldsymbol{y} \in \mathcal{D}(i,j). \tag{2.146}$$

The average decoding error probability associated to the code at the Receiver is given in (2.145). This system is depicted in Figure 2.6.

In the remainder of this section on MACs, the focus will be on (n,M_1,M_2,ϵ)-codes that satisfy a covertness constraint, *i.e.*, the adversary or the warden must remain unaware of the transmission. These covert codes are formally described in the next section.

Covert Codes

Assume that the input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ contain an "off" symbol denoted by x_0. Let $Q_{\boldsymbol{Z}}$ and $R_{\boldsymbol{Z}}$ be respectively the probability distributions

$$Q_{\boldsymbol{Z}}(\boldsymbol{z}) = P_{\boldsymbol{Z}|\boldsymbol{X}_1\boldsymbol{X}_2}(\boldsymbol{z}|\boldsymbol{x_0}, \boldsymbol{x_0}), \text{ and} \tag{2.147}$$

$$R_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{M_1 M_2} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} P_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{u}(i), \boldsymbol{v}(j)), \tag{2.148}$$

where $\boldsymbol{x_0} = (x_0, x_0, \ldots, x_0)$ denotes an n-dimensional vector that consists exclusively of "off" symbols. Consider a hypothesis test in which the warden aims to determine whether the Transmitters are off (hypothesis H_0) or the (n, M_1, M_2, ϵ)-code (hypothesis H_1) is used upon the observation of the channel output $\boldsymbol{Z}$:

$$\begin{cases} H_0 : \boldsymbol{Z} \sim Q_{\boldsymbol{Z}} \\ H_1 : \boldsymbol{Z} \sim R_{\boldsymbol{Z}}, \end{cases} \tag{2.149}$$

where $Q_{\boldsymbol{Z}}$ and $R_{\boldsymbol{Z}}$ are respectively given in (2.147) and (2.148).

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

$$T(\boldsymbol{z}) \triangleq \begin{cases} 0 & \text{if } H_0 \text{ is accepted,} \\ 1 & \text{if } H_1 \text{ is accepted.} \end{cases} \tag{2.150}$$

That is,

$$\alpha \triangleq \Pr\left[T(\boldsymbol{Z}) = 1\right], \text{ and} \tag{2.151}$$

$$\beta \triangleq \Pr\left[T(\boldsymbol{Z}) = 0\right], \tag{2.152}$$

where the probability operator in (2.151) applies assuming that $\boldsymbol{Z} \sim Q_{\boldsymbol{Z}}$ and the probability operator in (2.152) applies assuming that $\boldsymbol{Z} \sim R_{\boldsymbol{Z}}$.

Definition 22 ($(n, M_1, M_2, \epsilon, \delta)_{\mathrm{KL}}$-covert code)**.** *Given $\delta \in [0, 1]$, an (n, M_1, M_2, ϵ)-code described by* (2.140) *is said to be an $(n, M_1, M_2, \epsilon, \delta)_{\mathrm{KL}}$-covert code if*

$$D(R_{\boldsymbol{Z}} || Q_{\boldsymbol{Z}}) \leqslant \delta, \tag{2.153}$$

where $Q_{\boldsymbol{Z}}$ and $R_{\boldsymbol{Z}}$ are respectively defined in (2.147) *and* (2.148)*.*

Fundamental Limits

The information rate at which information can be covertly transmitted to the Receiver using an $(n, M_1, M_2, \epsilon, \delta)$-covert code by Transmitter k with $k \in \{1, 2\}$, is $\frac{\log(M_k)}{n}$ bits per channel use. Thus, a fundamental limit on the rate at which information can be covertly transmitted is given by the covert capacity region. This notion is formalized by the following definition.

Definition 23 (Covert capacity region)**.** *Fix a pair $(\epsilon, \delta) \in [0, 1]^2$. A covert rate pair*

$(\rho_1, \rho_2) \in \mathbb{R}_+^2$ *is achievable if there exists a sequence of* $(n, M_1, M_2, \epsilon, \delta)$*-covert code such that*

$$\lim_{n\to\infty} \frac{\log(M_1)}{\sqrt{n}} = \rho_1, \tag{2.154}$$

$$\lim_{n\to\infty} \frac{\log(M_2)}{\sqrt{n}} = \rho_2, \tag{2.155}$$

$$\lim_{n\to\infty} \lambda = 0, \tag{2.156}$$

$$\lim_{n\to\infty} D(R_{\boldsymbol{Z}} || Q_{\boldsymbol{Z}}) = 0. \tag{2.157}$$

The covert capacity region is the set that consists of all achievable covert rate pairs (ρ_1, ρ_2)*.*

Results

In [31, 32], the number of covert bits that can be embedded in an innocent code that performs close to capacity is established for binary channels. The results presented here are for the 2-user MAC but can be extend to an arbitrary number of users (see [32]). Assume that $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{0, 1\}$.

Define

$$K_1(z) \triangleq P_{Z|X_1X_2}(z|1,0) - P_{Z|X_1X_2}(z|0,0), \tag{2.158}$$

$$K_2(z) \triangleq P_{Z|X_1X_2}(z|0,1) - P_{Z|X_1X_2}(z|0,0), \tag{2.159}$$

$$K_3(z) \triangleq P_{Z|X_1X_2}(z|0,0) + P_{Z|X_1X_2}(z|1,1) - P_{Z|X_1X_2}(z|0,1) - P_{Z|X_1X_2}(z|1,0), \tag{2.160}$$

Theorem 16. *Consider the random transformation in* (2.138) *such that* $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{0, 1\}$ *and let* $(\kappa_1, \kappa_2) \in [0, 1]^2$ *with* $\kappa_1 + \kappa_2 = 1$*. If the MACs satisfy*

$$D\big(P_{Y|X_1=1,X_2=0} || P_{Y|X_1=0,X_2=0}\big) > D\big(P_{Z|X_1=1,X_2=0} || P_{Z|X_1=0,X_2=0}\big), \quad \text{and} \tag{2.161}$$

$$D\big(P_{Y|X_1=0,X_2=1} || P_{Y|X_1=0,X_2=0}\big) > D\big(P_{Z|X_1=0,X_2=1} || P_{Z|X_1=0,X_2=0}\big), \tag{2.162}$$

then the covert capacity region consists of all rate pairs $(\rho_1, \rho_2) \in \mathbb{R}_+^2$ *that verify*

$$\rho_1 \leqslant \sqrt{2}\kappa_1 \sqrt{\frac{\delta}{\sum_{z\in\mathcal{Z}} \frac{(\kappa_1 K_1(z) + \kappa_2 K_2(z))^2}{P_{Z|X_1=0,X_2=0}(z)}}}\, D\big(P_{Y|X_1=1,X_2=0} || P_{Y|X_1=0,X_2=0}\big), \tag{2.163}$$

$$\rho_2 \leqslant \sqrt{2}\kappa_2 \sqrt{\frac{\delta}{\sum_{z\in\mathcal{Z}} \frac{(\kappa_1 K_1(z) + \kappa_2 K_2(z))^2}{P_{Z|X_1=0,X_2=0}(z)}}}\, D\big(P_{Y|X_1=0,X_2=1} || P_{Y|X_1=0,X_2=0}\big). \tag{2.164}$$

If (2.161) and (2.162) are satisfied, covert communications can be achieved using a keyless communication scheme.

2.3. Conclusion

This chapter reviewed existing results on covert communications with an emphasis on classical information theoretic channels. For most of the models it is shown that the square-root law

applies.

Despite the growing number of contributions in the context of covert communications, there are still open problems. Among those, two are of interest in the scope of this thesis. Namely, determining the maximum number of information bits that can be transmitted when the warden observes only a subset of the channel outputs and determining the maximum number of information bits that can be embedded in a given broadcast code that is designed to transmit a common message to two receivers. These two problems are presented and treated at least partially in the next chapters of this thesis.

— 3 —

Covert Communications Type II

Bash *et al.* established in [7, 8] that on the order of $\sqrt{n}$ bits of information can be covertly transmitted within n channel uses over point-to-point channels. Later, Wang, Wornell and Zheng considered memoryless channels in which the receiver and the steganalyst observe the same channel output, and derived in [12] the covert capacity of such channels. During the same year, Bloch considered point-to-point binary memoryless channels (BMC) under covertness constraints and characterized a first order approximation of the covert capacity in [11]. Together with Tahmasbi, he derived in [13] the second order asymptotic approximation of the optimal codebook size.

In the context of wiretap channels, Ozarow and Wyner introduced the problem of the *wiretap channel type II* [43], which is a traditional instance of the wiretap channel problem where the eavesdropper observes only a subset of the channel outputs. In this chapter, the problem of covert communications type II is introduced. Inspired from the wiretap channel type II, this problem is a traditional instance of the covert communications problem over point-to-point links where the warden observes only a fixed fraction of the channel outputs. The main contribution of this chapter is a second order achievable bound for covert communications type II.

The remainder of this chapter unfolds as follows. Section 3.1 details the model. Section 3.2 establishes an achievable bound for the problem of covert communications type II. Section 3.3 provides a discussion of this result. Finally, Section 3.4 concludes this work.

3.1. System Model

3.1.1. Channel Model

Consider a three-party communication system in which a transmitter sends information to a legitimate receiver while a second receiver (the warden) observes a fraction of the channel outputs and aims to determine whether or not communication occurs between the legitimate

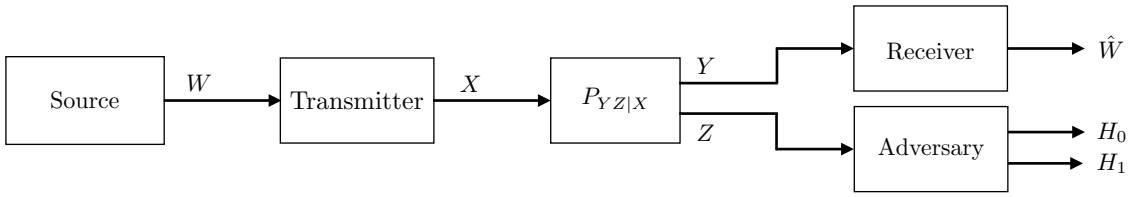

Figure 3.1.: Point-to-point channel with a warden.

parties. The noisy communication medium is described by a product random transformation

$$\left(\mathcal{X}^n, \mathcal{Y}^n, \mathcal{Z}^n, P_{\boldsymbol{YZ}|\boldsymbol{X}}\right), \tag{3.1a}$$

where $n \in \mathbb{N}$ is the block-length; $\mathcal{X}$ is the input alphabet, $\mathcal{Y}$ and $\mathcal{Z}$ are the output alphabets; and $\boldsymbol{X} = (X_1, X_2, \ldots, X_n) \in \mathcal{X}^n$, $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n) \in \mathcal{Y}^n$ and $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n) \in \mathcal{Z}^n$ are n-dimensional vectors of random variables. That is, given an input vector $\boldsymbol{x}$, the output $(\boldsymbol{y}, \boldsymbol{z})$ is observed with probability

$$P_{\boldsymbol{YZ}|\boldsymbol{X}}(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{x}) = \prod_{t=1}^{n} P_{YZ|X}(y_t, z_t|x_t), \tag{3.1b}$$

where $P_{YZ|X}$ is given parameter of the problem. That is, the channel is memoryless. This channel is represented in Figure 3.1.

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

$$\mathcal{W} \triangleq \{1, 2, \ldots, M\}, \tag{3.2}$$

with $M \in \mathbb{N}$. In order to transmit this message, the Transmitter and the Receiver might share a secret key $A \in \mathcal{A} = \{1, 2, \ldots, L\}$, with $L \in \mathbb{N}$. It is assumed that the warden knows the distribution P_A of the key, but not the actual realization of the key. Hence, to send a message index within n channel uses, the Transmitter uses an (n, M, L, ϵ)-code.

Definition 24 ((n, M, L, ϵ)-code)**.** *Given $(M, L, n) \in \mathbb{N}^3$ and $\epsilon \in [0, 1]$, an (n, M, L, ϵ)-code is a system*

$$\{(\boldsymbol{u}(1,1), \mathcal{D}(1,1)), (\boldsymbol{u}(1,2), \mathcal{D}(1,2)), \ldots, (\boldsymbol{u}(M,L), \mathcal{D}(M,L))\}, \tag{3.3}$$

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

$$\boldsymbol{u}(i,j) = (u_1(i,j), u_2(i,j), \ldots, u_n(i,j)) \in \mathcal{X}^n, \tag{3.4}$$

$$\mathcal{D}(i,j) \cap \mathcal{D}(k,j) = \emptyset, \tag{3.5}$$

$$\bigcup_{\ell=1}^{M} \mathcal{D}(\ell, j) \subseteq \mathcal{Y}^n, \text{ and} \tag{3.6}$$

$$\lambda \triangleq \frac{1}{ML} \sum_{i=1}^{M} \sum_{j=1}^{L} \Pr\left[\boldsymbol{Y} \in \mathcal{D}^{\mathsf{c}}(i,j) \middle| \boldsymbol{X} = \boldsymbol{u}(i,j)\right] \leqslant \epsilon. \tag{3.7}$$

The probability operator in (3.7) applies with respect to the marginal $P_{\boldsymbol{Y}|\boldsymbol{X}}$ of the joint

distribution in (3.1); and $\mathcal{D}^{\mathsf{c}}(i,j)$ in (3.7) represents the complement of $\mathcal{D}(i,j)$ with respect to $\mathcal{Y}^n$.

Given a code represented by the system in (3.3) and a key index $j \in \mathcal{A}$, the Transmitter uses the codeword $\boldsymbol{u}(i,j)$ 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 $u_t(i,j)$ to the channel. The Receiver, which knows the key index j, observes the output $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ after n channel uses and determines that the message index i was transmitted if it satisfies the decoding rule:

$$\boldsymbol{y} \in \mathcal{D}(i,j). \tag{3.8}$$

The average decoding error probability associated to the (n, M, L, ϵ)-code at the Receiver is given in (3.7). This system is depicted in Figure 3.1.

In the remainder of this chapter, the (n, M, L, ϵ)-codes of interest are those that satisfy a covertness constraint, *i.e.*, the adversary or the warden must remain unaware of the transmission. These covert codes are formally described in the next section.

3.1.2. Covert Codes

Assume that the input alphabet $\mathcal{X}$ contains an "off" symbol denoted by x_0. Let $\kappa \in [0,1]$ be fixed. Assume that the warden selects only $m = \lfloor \kappa n \rfloor$ components to determine whether or not communication takes place. Denote the set of indices chosen by the warden by $\mathcal{B} = \{b_1, b_2, \ldots, b_m\}$. Hence, the warden considers only the vector $\boldsymbol{Z}_{\mathcal{B}} = (Z_{b_1}, Z_{b_2}, \ldots, Z_{b_m})$ to perform its hypothesis test. Let $Q_{\boldsymbol{Z}_{\mathcal{B}}}$ and $R_{\boldsymbol{Z}_{\mathcal{B}}}$ be respectively the probability distributions

$$Q_{\boldsymbol{Z}_{\mathcal{B}}}(\boldsymbol{z}_{\mathcal{B}}) = P_{\boldsymbol{Z}_{\mathcal{B}}|\boldsymbol{X}_{\mathcal{B}}}(\boldsymbol{z}_{\mathcal{B}}|\boldsymbol{x_0}), \text{ and} \tag{3.9}$$

$$R_{\boldsymbol{Z}_{\mathcal{B}}}(\boldsymbol{z}_{\mathcal{B}}) = \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{L} P_A(j) P_{\boldsymbol{Z}_{\mathcal{B}}|\boldsymbol{X}_{\mathcal{B}}}(\boldsymbol{z}_{\mathcal{B}}|\boldsymbol{u}_{\mathcal{B}}(i,j)), \tag{3.10}$$

where $\boldsymbol{x_0} = (x_0, x_0, \ldots, x_0)$ denotes an m-dimensional vector that consists exclusively of "off" symbols. Consider a hypothesis test in which the warden aims to determine whether the transmitter is off (hypothesis H_0) or the (n, M, L, ϵ)-code (hypothesis H_1) is used upon the observation of the channel output $\boldsymbol{Z}_{\mathcal{B}}$:

$$\begin{cases} H_0 : \boldsymbol{Z} \sim Q_{\boldsymbol{Z}_{\mathcal{B}}} \\ H_1 : \boldsymbol{Z} \sim R_{\boldsymbol{Z}_{\mathcal{B}}}, \end{cases} \tag{3.11}$$

where $Q_{\boldsymbol{Z}_{\mathcal{B}}}$ and $R_{\boldsymbol{Z}_{\mathcal{B}}}$ are respectively given in (3.9) and (3.10).

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

$$T(\boldsymbol{z_B}) \triangleq \begin{cases} 0 & \text{if } H_0 \text{ is accepted,} \\ 1 & \text{if } H_1 \text{ is accepted.} \end{cases} \tag{3.12}$$

That is,

$$\alpha \triangleq \Pr\left[T(\boldsymbol{Z}_{\mathcal{B}}) = 1\right], \text{ and} \tag{3.13}$$

$$\beta \triangleq \Pr\left[T(\boldsymbol{Z}_{\mathcal{B}}) = 0\right], \tag{3.14}$$

where the probability operator in (3.13) applies assuming that $\boldsymbol{Z}_{\mathcal{B}} \sim Q_{\boldsymbol{Z}_{\mathcal{B}}}$ and the probability operator in (3.14) applies assuming that $\boldsymbol{Z}_{\mathcal{B}} \sim R_{\boldsymbol{Z}_{\mathcal{B}}}$.

Definition 25 ($(n, M, L, \epsilon, \delta, \kappa)$-covert code)**.** *Given $(\delta, \kappa) \in [0,1]^2$, an (n, M, L, ϵ)-code $\mathcal{C}$ described by* (3.3) *is said to be an $(n, M, L, \epsilon, \delta, \kappa)$-covert code if*

$$\max_{\substack{\mathcal{B}\subseteq\{1,2,\ldots,n\} \\ |\mathcal{B}|=\lfloor \kappa n \rfloor}} D(R_{\boldsymbol{Z}_{\mathcal{B}}} || Q_{\boldsymbol{Z}_{\mathcal{B}}}) \leqslant \delta, \tag{3.15}$$

where $Q_{\boldsymbol{Z}_{\mathcal{B}}}$ and $R_{\boldsymbol{Z}_{\mathcal{B}}}$ are respectively defined in (3.9) *and* (3.10)*.*

Note that the maximization in (3.15) takes into account the worst m-dimensional vector that can be chosen by the warden.

3.1.3. Fundamental Limits

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

Definition 26 (Largest covert code's size)**.** *Fix a triplet $(\epsilon, \delta, \kappa) \in [0,1]^3$ and $L \in \mathbb{N}$. The largest covert code's size is:*

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

3.2. An achievable bound for Binary Memoryless Channels

In this section, a lower bound on the largest code's size (Definition 26) is established for binary memoryless channels (BMCs) using techniques from [11] and [13]. Hence, it is assumed that $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{x_0, x_1\}$. It is also assumed that $Y = Z$ and that $P_{Y|X=x_0} \ll P_{Y|X=x_1}$. The next theorem introduces this lower bound.

Theorem 17. *Consider a BMC $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, P_{YZ|X})$. It holds that*

$$\begin{aligned} &\frac{\log\left(M^*(n, L, \epsilon, \delta, \kappa)\right)}{n} \geqslant \\ &\min\left\{ \max_{P_X} I(X;Y) - \sqrt{\frac{\mathbb{E}_{XY}\left[\log\left(\frac{P_{Y|X}(Y|X)}{P_Y(Y)}\right)^2\right]}{n}} Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right), \right. \\ &\qquad \left. \theta D\big(P_{Y|X=x_1} || P_{Y|X=x_0}\big) - \sqrt{\frac{\theta \mathbb{E}_{Y|X=x_1}\left[\log\left(\frac{P_{Y|X}(Y|x_1)}{P_{Y|X}(Y|x_0)}\right)^2\right]}{n}} Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right) \right\}, \end{aligned} \tag{3.16}$$

with

$$\theta = \sqrt{\frac{2\delta}{\lfloor \kappa n \rfloor \chi_2(P_{Y|X=x_1}, P_{Y|X=x_0})}}. \tag{3.17}$$

The construction of the lower bound is presented in three parts. In the first part, a probability mass function to randomly generate an (n, M, L, ϵ)-code is chosen. This distribution is expressed in terms of some parameters, which are referred to as the *generating parameters*. In the second part, the generating parameters are chosen in order to satisfy the covertness constraint in (3.15) for a fixed δ and fixed κ, which allows the construction of an $(n, M, L, \epsilon, \delta, \kappa)$-covert code. In the third part, the average of the decoding error probability (denoted by $\hat{\Lambda}$) of the covert code is upper-bounded. This upper-bound proves the existence of an $(n, M, L, \epsilon, \delta, \kappa)$-covert code whose number of codewords provides a lower bound on the largest code's size (Definition 26).

Part I: Generation of the Random Code

Consider the parameters $(M, L) \in \mathbb{N}^2$; and fix a probability mass function P_X such that

$$P_X(x_1) = 1 - P_X(x_0) = \theta, \tag{3.18}$$

with $\theta \in [0, 1]$. Often, the parameters M, L and θ are referred to as the *generating parameters*.

For all $j \in \{1, 2, \ldots, L\}$, generate M codewords

$$\boldsymbol{u}(1, j), \boldsymbol{u}(2, j), \ldots, \boldsymbol{u}(M, j) \tag{3.19}$$

to form a codebook. For all $i \in \{1, 2, \ldots, M\}$, the codeword $\boldsymbol{u}(i, j)$ is the realization of a random variable following the probability mass function $P_{\boldsymbol{X}}$ such that for all $\boldsymbol{x} \in \mathcal{X}^n$,

$$P_{\boldsymbol{X}}(\boldsymbol{x}) \triangleq \prod_{t=1}^{n} P_X(x_t). \tag{3.20}$$

In the following, the probability mass function $P_{\boldsymbol{X}}$ is referred to as the *generating distribution*. To complete the generation of the (n, M, L, ϵ)-code, the decoding sets must be specified. For all $j \in \{1, 2, \ldots, L\}$, the Receiver uses the decoding sets

$$\mathcal{D}(1, j), \mathcal{D}(2, j), \ldots, \mathcal{D}(M, j), \tag{3.21}$$

and the decoding rule in (3.8).

For all $j \in \{1, 2, \ldots, L\}$, upon the reception of the channel output $\boldsymbol{y} \in \mathcal{Y}^n$, the Receiver declares that the index pair $(i, j) \in \mathcal{W} \times \mathcal{A}$ was transmitted according to the decoding rule in (3.8), with

$$\mathcal{D}(i, j) = \left\{ \boldsymbol{y} \in \mathcal{Y}^n : \log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i, j))}{P_{\boldsymbol{Y}}(\boldsymbol{y})} \right) \geqslant n\eta \right\} \setminus \bigcup_{k<i} \mathcal{D}(k, j), \tag{3.22}$$

where $\eta \in \mathbb{R}$ is a parameter whose exact value is determined later. Note that the codewords in (3.19) and the decoding sets in (3.22) form an (n, M, L, ϵ)-code.

Finally, the distribution P_A of the key is chosen such that for all $j \in \mathcal{A}$,

$$P_A(j) = \prod_{i=1}^{M} P_{\boldsymbol{X}}(\boldsymbol{u}(i, j)). \tag{3.23}$$

Part II: Covertness Analysis

The next proposition establishes sufficient conditions on the generating parameters to ensure that the covertness constraint in (3.15) is satisfied.

Proposition 1. *To guarantee that* (3.15) *is verified, it is sufficient to satisfy*

$$\theta \leqslant \sqrt{\frac{2\delta}{\lfloor \kappa n \rfloor \chi_2(P_{Y|X=x_1}, P_{Y|X=x_0})}}. \tag{3.24}$$

The proof of Proposition 1 is presented in Appendix B.

Part III: Error Probability Analysis

Two cases are to be considered for the error probability analysis. Assume that the point-to-point capacity achieving distribution P_X^* (without any constraint) is such that

$$P_X^*(x_1) = 1 - P_X^*(x_0) = \theta^*, \tag{3.25}$$

with $\theta^* \in [0, 1]$.

Case 1 : $\theta^* \leqslant \sqrt{\frac{2\delta}{\lfloor \kappa n \rfloor \chi_2(P_{Y|X=x_1}, P_{Y|X=x_0})}}$

In this case, the constraint on θ is not active and the capacity achieving distribution P_X^* can be used to generate the codewords. That is, the capacity of the point-to-point link can be achieved. The corresponding proof is therefore provided in [44, Theorem 45].

Case 2 : $\theta^* > \sqrt{\frac{2\delta}{\lfloor \kappa n \rfloor \chi_2(P_{Y|X=x_1}, P_{Y|X=x_0})}}$

In this case, the existence of an $(n, M, L, \epsilon, \delta, \kappa)$-covert code is established in the next proposition.

Proposition 2. *There exists an* $(n, M, L, \epsilon, \delta, \kappa)$*-covert code such that*

$$\frac{\log(M)}{n} \geqslant \theta D\big(P_{Y|X=x_1} || P_{Y|X=x_0}\big) - \sqrt{\frac{\theta \mathbb{E}_{Y|X=x_1}\left[\log\left(\frac{P_{Y|X}(Y|x_1)}{P_{Y|X}(Y|x_0)}\right)^2\right]}{n}} Q^{-1}(\epsilon) + O(1), \tag{3.26}$$

where

$$\theta = \sqrt{\frac{2\delta}{\lfloor \kappa n \rfloor \chi_2(P_{Y|X=x_1}, P_{Y|X=x_0})}}. \tag{3.27}$$

The proof of Proposition 2 is presented in Appendix C.

Together, Proposition 1, Proposition 2 and [44, Theorem 45] prove Theorem 17. That is, there exists an $(n, M, L, \epsilon, \delta, \kappa)$-covert code that can simultaneously reliably and covertly transmit information at the rate given in (3.16).

3.3. Discussion

Several points are interesting to notice. First, when κ tends to 0, the capacity of the point-to-point link (without any constraint) is achievable. Hence, transmission can occur at a non-vanishing rate. This is the case when the warden does not observe enough channel outputs to take a reliable decision regardless of the rate at which information is sent.

Second, when κ equals 1, the expression in (3.16) is comparable to that in Theorem 5, with the difference that the latter is established for maximum decoding error probability. In case maximum error probability is considered, the proof technique of the achievability presented in [13] can be used to solve this case, even though this is not detailed here. In addition, the converse proof presented in [13] should also hold with codewords of weight $i = O\left(\sqrt{\frac{n}{\mu}}\right)$. As it is highlighted in [13], converses in finite-length for the covert communication problem are hard to establish when considering average error probability. This difficulty could not be overcome in this thesis and a converse bound is missing.

Finally, it is worth mentioning that the rate threshold observed in the finite block-length regime due to the constraint on the input distribution parameter θ cannot be observed in the asymptotic block-length regime. Indeed, in the latter case, the parameter θ must tend to 0 to satisfy the covertness constraint. That is, the covertness constraint on the parameter θ is always active and covert communications can only occur with a vanishing transmission rate.

Regarding the AWGN case, it was shown in [12] that the covertness constraint translates into a power constraint on the input. Hence, there are arguments to conjecture that a similar threshold can be observed in the AWGN channel with input power constraint in the finite block length regime. If κ is small enough, only the input power constraint will be active on the input. Otherwise, the covertness constraint will induce a more stringent constraint on the input. Nevertheless, this could not be formally established in this thesis.

3.4. Conclusion

In this chapter, an achievable second order bound was established for the problem of covert communications type II. It can be observed that there exists a threshold in the input distribution that determines the communication rate. That is, if the warden does not observe enough channel outputs, covert communications can occur at channel capacity. Otherwise, the transmission rate is expressed in terms of the fraction of channel outputs observed by the warden. So far, a converse bound is still missing for this problem. It is conjectured that a similar threshold can be observed in AWGN channels, but the formal treatment of these channels is left open.

Finally, it worth emphasizing that the type of constraint considered in this chapter was not studied before in any type of channels. It could be used for instance in multi-user channels such as the broadcast channel. The latter is the topic of the next chapter.

— 4 —

Embedding Covert Information into a Given Broadcast Code

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 [8, 9, 11, 12].

Two different covert communication problems have been studied within the context of broadcast channels [30, 2, 3]. In [30], the sender tries to send two covert messages to two receivers. In [2] and [3], 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 [2] and [3]. 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 [2] and [3] are:

- In [2] and [3], 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 [2] and [3] 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. Instead, in this work, the total variation distance is used by

[1] A technical condition is that the given non-covert code must have a positive error exponent; see (4.52).

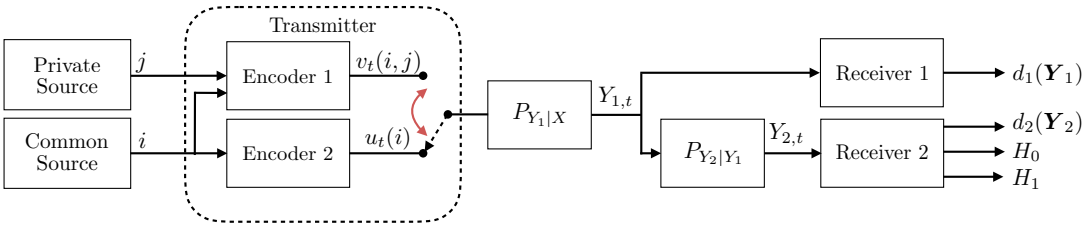

Figure 4.1.: Degraded broadcast channel with covert messages at channel use $t \in \{1, 2, \ldots, n\}$, where $d_1 : \mathcal{Y}_1^n \to \mathcal{W} \times \hat{\mathcal{W}}$ denotes the decoding function at Receiver 1 and $d_2 : \mathcal{Y}_2^n \to \mathcal{W}$ denotes the decoding function at Receiver 2.

adapting some techniques from [13]. 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 chapter is organized as follows. Section 4.1 presents the system model. Section 4.2 exposes examples in which covert communications can not be achieved. Section 4.3 presents an achievability scheme. Section 4.4 presents an impossibility (converse) result. Section 4.5 presents the main result of this work. Finally, Section 4.6 provides some examples and Section 4.7 concludes this work.

4.1. System Model

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

$$(\mathcal{X}^n, \mathcal{Y}_1^n \times \mathcal{Y}_2^n, P_{\boldsymbol{Y}_1\boldsymbol{Y}_2|\boldsymbol{X}}), \tag{4.1a}$$

where $n \in \mathbb{N}$ is the duration of the communication in channel uses (block-length) and the alphabets $\mathcal{X}$, $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are finite. Given a channel input $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, the channel 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 at Receiver k with probability:

$$P_{\boldsymbol{Y}_1\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}_1, \boldsymbol{y}_2|\boldsymbol{x}) \triangleq \prod_{t=1}^{n} P_{Y_1|X}(y_{1,t}|x_t) P_{Y_2|Y_1}(y_{2,t}|y_{1,t}). \tag{4.1b}$$

That is, the channel is degraded and memoryless.

Given the random transformation in (4.1a), the Transmitter uses a *broadcast code* (Encoder 2 in Figure 4.1) to transmit a message intended to both receivers at a fixed rate. Often, this message index is referred to as the *common message*. Section 4.1.1 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 4.1), it is possible to build

a new code (Encoder 1 in Figure 4.1) 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 4.1.2.
An induced code might satisfy some additional constraints on the transmission of the private message, e.g., a covertness constraint, a physical layer security constraint, a simultaneous information and energy transmission constraint, etc. This work focuses on a covertness constraint which 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 4.1.3.

The objective of this work is to determine the maximum amount of information that can be covertly transmitted given a broadcast code.

4.1.1. 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

$$\mathcal{W} \triangleq \{1, 2, \ldots, M\}, \tag{4.2}$$

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

Definition 27 ((n, M, ϵ)-broadcast code)**.** *Given $M \in \mathbb{N}$, $\epsilon \in [0, 1]$ and a block-length $n \in \mathbb{N}$, an (n, M, ϵ)-broadcast code for the random transformation in* (4.1) *is a system*

$$\left\{\left(\boldsymbol{u}(1), \mathcal{D}_1(1), \mathcal{D}_2(1)\right), \left(\boldsymbol{u}(2), \mathcal{D}_1(2), \mathcal{D}_2(2)\right), \ldots, \left(\boldsymbol{u}(M), \mathcal{D}_1(M), \mathcal{D}_2(M)\right)\right\}, \tag{4.3}$$

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

$$\boldsymbol{u}(i) \triangleq (u_1(i), u_2(i), \ldots, u_n(i)) \in \mathcal{X}^n, \tag{4.4a}$$

$$\mathcal{D}_k(i) \cap \mathcal{D}_k(j) = \emptyset, \tag{4.4b}$$

$$\bigcup_{l=1}^{M} \mathcal{D}_k(l) \subseteq \mathcal{Y}_k^n, \quad \text{and} \tag{4.4c}$$

$$\frac{1}{M} \sum_{i=1}^{M} \Pr\left[\boldsymbol{Y}_k \in \mathcal{D}_k^{\mathsf{c}}(i) \middle| \boldsymbol{X} = \boldsymbol{u}(i)\right] \leqslant \epsilon. \tag{4.4d}$$

The probability operator in (4.4d) applies with respect to the marginal $P_{\boldsymbol{Y}_k|\boldsymbol{X}}$ of the joint probability mass function in (4.1b); and $\mathcal{D}_k^{\mathsf{c}}(i)$ in (4.4d) represents the complement of $\mathcal{D}_k(i)$ with respect to $\mathcal{Y}_k^n$.

Given a broadcast code represented by the system in (4.3), 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. After n channel uses, Receiver k, with

$k \in \{1, 2\}$, observes the output $\boldsymbol{y}_k = (y_{k,1}, y_{k,2}, \ldots, y_{k,n})$ and determines that the message index i was transmitted if it satisfies the decoding rule:

$$\boldsymbol{y}_k \in \mathcal{D}_k(i). \tag{4.5}$$

The average decoding error probability associated to the given broadcast code at Receiver k, denoted by λ_k is given in the left hand-side of (4.4d).

4.1.2. Induced Codes

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

$$\hat{\mathcal{W}} \triangleq \{1, 2, \ldots, \hat{M}\}, \tag{4.6}$$

with $\hat{M} \in \mathbb{N}$. Assume that a broadcast code denoted by $\mathcal{C}$ is given and is represented by the system in (4.3). The transmitter uses an $(n, \mathcal{C}, \hat{M})$-induced code to transmit both the common and private message indices.

Definition 28 ($(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code)**.** *Given $\hat{M} \in \mathbb{N}$, $\hat{\epsilon} \in [0, 1]$, and an (n, M, ϵ)-broadcast code $\mathcal{C}$ described by (4.3), an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code is a system*

$$\Big\{ \left(\boldsymbol{v}(1,1), \mathcal{D}_1(1,1), \mathcal{D}_2(1)\right), \left(\boldsymbol{v}(1,2), \mathcal{D}_1(1,2), \mathcal{D}_2(1)\right), \ldots, \left(\boldsymbol{v}(M,\hat{M}), \mathcal{D}_1(M,\hat{M}), \mathcal{D}_2(M)\right) \Big\} \tag{4.7}$$

that satisfies for all $(i, k, j, l) \in \mathcal{W}^2 \times \hat{\mathcal{W}}^2$, with $(i, j) \neq (k, l)$:

$$\boldsymbol{v}(i,j) \triangleq (v_1(i,j), v_2(i,j), \ldots, v_n(i,j)) \in \mathcal{X}^n, \tag{4.8a}$$

$$\mathcal{D}_1(i,j) \cap \mathcal{D}_1(k,l) = \emptyset, \tag{4.8b}$$

$$\bigcup_{p=1}^{M} \bigcup_{q=1}^{\hat{M}} \mathcal{D}_1(p,q) \subseteq \mathcal{Y}_1^n, \tag{4.8c}$$

$$\frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i,j) | \boldsymbol{X} = \boldsymbol{v}(i,j)\right] \leqslant \hat{\epsilon}, \tag{4.8d}$$

$$\frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \Pr\left[\boldsymbol{Y}_2 \in \mathcal{D}_2^{\mathsf{c}}(i) | \boldsymbol{X} = \boldsymbol{v}(i,j)\right] \leqslant \hat{\epsilon}. \tag{4.8e}$$

The probability operators in (4.8d) and (4.8e) apply with respect to the conditional marginals $P_{\boldsymbol{Y}_1|\boldsymbol{X}}$ and $P_{\boldsymbol{Y}_2|\boldsymbol{X}}$ of the joint probability mass function in (4.1b), 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.

Given an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code denoted by $\hat{\mathcal{C}}$ and described by (4.7), 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:

$$\boldsymbol{y}_1 \in \mathcal{D}_1(i,j). \tag{4.9}$$

Receiver 2 determines that the message index i was transmitted if it satisfies (4.5), with $k = 2$, *i.e.*, Receiver 2 uses the same rule 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 $\hat{\lambda}_k$. The left hand-sides of (4.8d) and (4.8e) respectively define $\hat{\lambda}_1$ and $\hat{\lambda}_2$.

Remark 1. *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* (4.5)*, the inclusion in* (4.4c) *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* (4.4c) *holds with equality. Without any loss of generality, the inclusion in* (4.8c) *is assumed with equality for analogous reasons.*

4.1.3. Covert Codes

Consider an (n, M, ϵ)-broadcast code described by (4.3) 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 (4.7). For all $k \in \{1, 2\}$, let $Q_{\boldsymbol{Y}_k}$ and $R_{\boldsymbol{Y}_k}$ be respectively the probability mass 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 $\boldsymbol{y} \in \mathcal{Y}_k^n$,

$$Q_{\boldsymbol{Y}_k}(\boldsymbol{y}) \triangleq \frac{1}{M} \sum_{i=1}^{M} P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i)), \text{ and} \tag{4.10}$$

$$R_{\boldsymbol{Y}_k}(\boldsymbol{y}) \triangleq \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}(i,j)), \tag{4.11}$$

where $P_{\boldsymbol{Y}_k|\boldsymbol{X}}$ is the marginal obtained from (4.1b). 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 $\boldsymbol{Y}_2$:

$$\begin{cases} H_0 : \boldsymbol{Y}_2 \sim Q_{\boldsymbol{Y}_2} \\ H_1 : \boldsymbol{Y}_2 \sim R_{\boldsymbol{Y}_2}, \end{cases} \tag{4.12}$$

where $Q_{\boldsymbol{Y}_2}$ and $R_{\boldsymbol{Y}_2}$ are respectively given in (4.10) and (4.11).

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

$$T(\boldsymbol{y}) \triangleq \begin{cases} 0 & \text{if } H_0 \text{ is accepted,} \\ 1 & \text{if } H_1 \text{ is accepted.} \end{cases} \tag{4.13}$$

That is,

$$\alpha \triangleq \Pr\left[T(\boldsymbol{Y}_2) = 1\right], \text{ and} \tag{4.14}$$

$$\beta \triangleq \Pr\left[T(\boldsymbol{Y}_2) = 0\right], \tag{4.15}$$

where the probability operator in (4.14) applies assuming that $\boldsymbol{Y}_2 \sim Q_{\boldsymbol{Y}_2}$ and the probability operator in (4.15) applies assuming that $\boldsymbol{Y}_2 \sim R_{\boldsymbol{Y}_2}$.

From [37, Theorem 13.1.1], it holds that

$$\alpha + \beta \geqslant 1 - \|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}}\,, \tag{4.16}$$

with equality for the optimal test, and where α and β are respectively defined in (4.14) and (4.15), for all decision rules $T : \mathcal{Y}_2^n \to \{0, 1\}$ of the form (4.13).
Note that from (4.16), it follows that the smaller the parameter δ, the higher the probability of failing to determine whether the broadcast code or the covert code is used. Thus, a covert code is defined hereunder as follows.

Definition 29 ($(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code)**.** *Given $\delta \in [0, 1]$ and an (n, M, ϵ)-broadcast code $\mathcal{C}$ described by* (4.3)*, an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code described by* (4.7) *is said to be an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code if*

$$\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} \leqslant \delta, \tag{4.17}$$

where $Q_{\boldsymbol{Y}_2}$ and $R_{\boldsymbol{Y}_2}$ are respectively defined in (4.10) *and* (4.11)*.*

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 $\delta \in [0, 1]$. 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 probability mass function $Q_{\boldsymbol{Y}_2}$ 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 30 (Largest covert code's size)**.** *Fix a pair $(\hat{\epsilon}, \delta) \in [0, 1]^2$ and consider an (n, M, ϵ)-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:*

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

4.2. 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.

The next lemma establishes a lower-bound on the total variation $\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}}$ which shows that covert communications are not achievable for arbitrary values of δ when $R_{\boldsymbol{Y}_2} \not\ll Q_{\boldsymbol{Y}_2}$.

Lemma 3. *Consider the random transformation in* (4.1)*, an (n, M, ϵ)-broadcast code $\mathcal{C}$ and an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code respectively described in* (4.3) *and* (4.7) *such that $Q_{\boldsymbol{Y}_2} \not\ll R_{\boldsymbol{Y}_2}$. Then,*

$$\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} \geqslant \frac{1}{2}\left(1 - \Pr\left[\boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2}\right]\right), \tag{4.18}$$

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* (4.10) *and* (4.11)*.*

Proof: The proof of Lemma 3 is provided in Appendix E ■

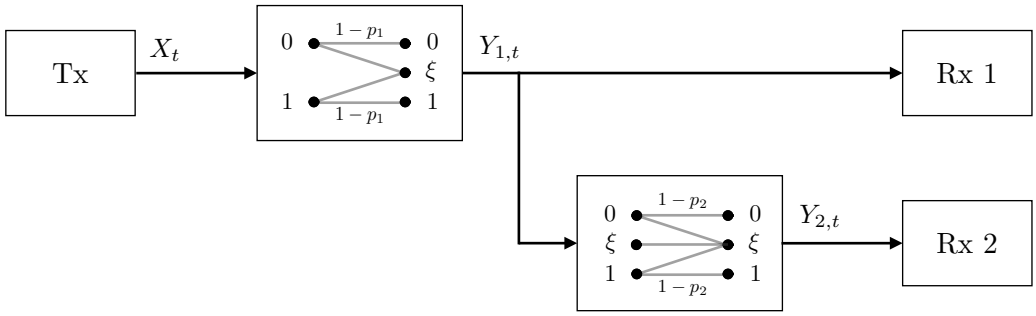

Figure 4.2.: Degraded erasure broadcast channel at channel use $t \in \{1, 2, \ldots, n\}$.

Example 1. *Assume that the random transformation in* (4.1) *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 mass functions* $P_{Y_1|X}$ *and* $P_{Y_2|Y_1}$ *respectively satisfy:*

$$P_{Y_1|X}(x|x) = 1 - P_{Y_1|X}(\xi|x) = 1 - p_1, \tag{4.19a}$$
$$P_{Y_1|X}(x|1-x) = 0, \tag{4.19b}$$

and

$$P_{Y_2|Y_1}(x|x) = 1 - P_{Y_2|Y_1}(\xi|x) = 1 - p_2, \tag{4.20a}$$
$$P_{Y_2|Y_1}(x|1-x) = 0, \tag{4.20b}$$
$$P_{Y_2|Y_1}(\xi|\xi) = 1, \tag{4.20c}$$

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

Figure 4.2 depicts the channel in Example 1. Note that the probability mass function $P_{Y_2|X}$ verifies that for all $x \in \mathcal{X}$:

$$\begin{aligned} P_{Y_2|X}(x|x) = 1 - P_{Y_1|X}(\xi|x) &= 1 - p_1 - p_2 + p_1 p_2 \\ &= 1 - p, \end{aligned} \tag{4.21}$$

with

$$p = p_1 + p_2 - p_1 p_2. \tag{4.22}$$

Given an (n, M, ϵ)-broadcast code $\mathcal{C}$ described by (4.3) and an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code described by (4.7), let $\mathcal{T}_{ij}$ and $\bar{\mathcal{T}}_{ij}$ be respectively defined for all $(i, j) \in \mathcal{W} \times \hat{\mathcal{W}}$ by

$$\mathcal{T}_{ij} = \{t \in \{1, 2, \ldots, n\} : u_t(i) \neq v_t(i, j)\}, \text{ and} \tag{4.23}$$
$$\bar{\mathcal{T}}_{ij} = \{t \in \{1, 2, \ldots, n\} : u_t(i) = v_t(i, j)\}. \tag{4.24}$$

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

$$|\mathcal{T}_{ij}| = \omega(i,j), \qquad \text{and} \tag{4.25}$$
$$\left|\bar{\mathcal{T}}_{ij}\right| = n - \omega(i,j). \tag{4.26}$$

Within this context, the term $\Pr[\boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2}]$ in (4.18) can be upper bounded as follows:

$$\begin{aligned}
\Pr[\boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2}] &= \sum_{\boldsymbol{y} \in \operatorname{supp} R_{\boldsymbol{Y}_2}} R_{\boldsymbol{Y}_2}(\boldsymbol{y}) \mathbb{1}_{\{\boldsymbol{y} \in \operatorname{supp} Q_{\boldsymbol{Y}_2}\}} \\
&= \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} R_{\boldsymbol{Y}_2}(\boldsymbol{y}) \mathbb{1}_{\{\boldsymbol{y} \in \operatorname{supp} Q_{\boldsymbol{Y}_2}\}} \mathbb{1}_{\{\boldsymbol{y} \in \operatorname{supp} R_{\boldsymbol{Y}_2}\}} \\
&= \Pr[\boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2} \cap \operatorname{supp} R_{\boldsymbol{Y}_2}] \\
&= \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{\substack{\boldsymbol{y} \in \operatorname{supp} Q_{\boldsymbol{Y}_2} \\ \cap \operatorname{supp} R_{\boldsymbol{Y}_2}}} \prod_{t=1}^{n} P_{Y_2|X}(y_t | v_t(i,j)) \\
&= \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{\substack{\boldsymbol{y} \in \operatorname{supp} Q_{\boldsymbol{Y}_2} \\ \cap \operatorname{supp} R_{\boldsymbol{Y}_2}}} \prod_{s \in \mathcal{T}_{ij}} P_{Y_2|X}(y_s | v_s(i,j)) \prod_{r \in \bar{\mathcal{T}}_{ij}} P_{Y_2|X}(y_r | u_r(i)) \\
&\leqslant \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{\substack{\boldsymbol{y} \in \operatorname{supp} Q_{\boldsymbol{Y}_2} \\ \cap \operatorname{supp} R_{\boldsymbol{Y}_2}}} \prod_{s \in \mathcal{T}_{ij}} P_{Y_2|X}(y_s | v_s(i,j)).
\end{aligned} \tag{4.27}$$

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 (4.27) that

$$\begin{aligned}
\Pr[\boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2}] &\leqslant \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} p^{|\mathcal{T}_{ij}|} \\
&\overset{(a)}{=} \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} p^{\omega(i,j)} \\
&\leqslant p^{\omega_{\min}},
\end{aligned} \tag{4.28}$$

where *(a)* follows from (4.25); and

$$\omega_{\min} \triangleq \min_{(i,j) \in \mathcal{W} \times \hat{\mathcal{W}}} \omega(i,j). \tag{4.29}$$

Finally, it follows from Lemma 3 that the total variation $\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}}$ verifies

$$\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} \geqslant \frac{1}{2}\left(1 - p^{\omega_{\min}}\right). \tag{4.30}$$

The above lower bound shows that the constraint in (4.17) can not be satisfied for values of $\delta \leqslant \frac{1}{2}\left(1 - p^{\omega_{\min}}\right)$.

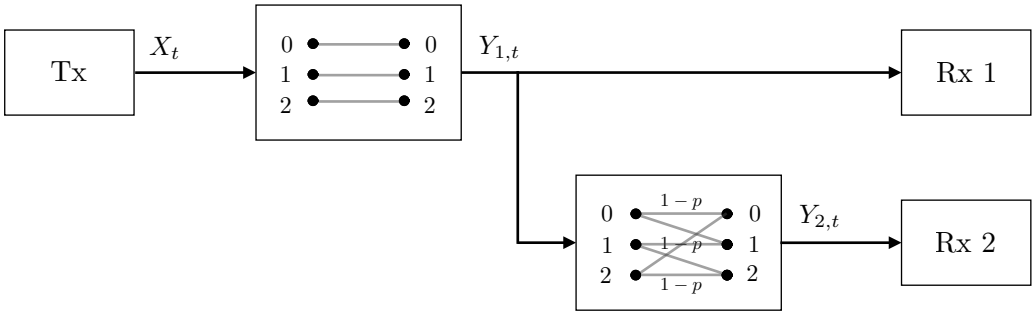

Figure 4.3.: Degraded typewriter broadcast channel at channel use $t \in \{1, 2, \ldots, n\}$.

Example 2. *Consider the random transformation in* (4.1) *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 mass functions* $P_{Y_1|X}$ *and* $P_{Y_2|Y_1}$ *respectively satisfy:*

$$P_{Y_1|X}(x|x) = 1, \tag{4.31}$$

and

$$P_{Y_2|Y_1}(x|x'') = 0, \tag{4.32a}$$
$$P_{Y_2|Y_1}(x|x) = 1 - P_{Y_2|Y_1}(x'|x) = 1 - p, \tag{4.32b}$$

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

Figure 4.3 depicts the channel in Example 3. Given that $P_{Y_1|X}(x|x) = 1$ for all $x \in \mathcal{X}$, it follows that $P_{Y_2|X=x} = P_{Y_2|Y_1=x}$.

Note that given an arbitrary (n, M, ϵ)-broadcast code $\mathcal{C}$ of the form in (4.3) and any $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code, the inequality in (4.28) holds, which implies

$$\Pr\left[\boldsymbol{Y}_2 \in \operatorname{supp} Q_{\boldsymbol{Y}_2}\right] \leqslant p^{\omega_{\min}}. \tag{4.33}$$

The above lower bound shows that the constraint in (4.17) can not be satisfied for values of $\delta \leqslant \frac{1}{2}\left(1 - p^{\omega_{\min}}\right)$.

4.3. Achievability of Covert Communications

In this section, a lower bound on the largest code's size (Definition 30) given an (n, M, ϵ)-broadcast code, denoted by $\mathcal{C}$, is established using techniques from [11] and [12]. The construction of this result is presented in three parts using a covert code (Definition 29). In the first part, a probability mass function to randomly generate an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code is chosen. Often, this probability mass function 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 generating parameters are chosen in order to satisfy the covertness constraint in (4.17) for a fixed δ, which allows the construction of an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code. In the third part, the average of the decoding error probabilities (denoted by $\hat{\Lambda}_k$, with $k \in \{1, 2\}$) of the covert code are upper-bounded. These upper-bounds

are expressed in terms of the generating parameters, which proves that the $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-random code satisfies $\hat{\Lambda}_k < \hat{\epsilon}$ for all $k \in \{1, 2\}$.

Part I: Generation of the Random Code

Consider an (n, M, ϵ)-broadcast code $\mathcal{C}$ for the random transformation in (4.1) described by the system in (4.3). Consider also the parameters $\hat{M} \in \mathbb{N}$; $K \in [0, \sqrt{n}]$; and a conditional probability mass function $\tilde{P}_{\hat{X}|X}$ such that, for all $x \in \mathcal{X}$,

$$\operatorname{supp} \tilde{P}_{\hat{X}|X=x} \subseteq \mathcal{X} \setminus \{x\}. \tag{4.34}$$

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

$$P_{\hat{X}|X}(\hat{x}|x) \triangleq (1 - \theta)\, \mathbb{1}_{\{x=\hat{x}\}} + \theta \tilde{P}_{\hat{X}|X}(\hat{x}|x), \tag{4.35}$$

with

$$\theta \triangleq \frac{K}{\sqrt{n}}. \tag{4.36}$$

Often, the parameters $\hat{M}, K$ 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

$$\boldsymbol{v}(i, 1), \boldsymbol{v}(i, 2), \ldots, \boldsymbol{v}(i, \hat{M}) \tag{4.37}$$

to form the codebook of an 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 mass function $P_{\hat{\boldsymbol{X}}|\boldsymbol{X}=\boldsymbol{u}(i)}$ such that for all $\hat{\boldsymbol{x}} \in \mathcal{X}^n$,

$$P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}\left(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)\right) \triangleq \prod_{t=1}^{n} P_{\hat{X}|X}\left(\hat{x}_t|u_t(i)\right), \tag{4.38}$$

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

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

$$\mathcal{D}_2(1), \mathcal{D}_2(2), \ldots, \mathcal{D}_2(M) \tag{4.39}$$

of the given broadcast code $\mathcal{C}$, and the decoding rule in (4.5), with $k = 2$.

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

$$\imath_k(\hat{\boldsymbol{x}}; \boldsymbol{y}|\boldsymbol{x}) \triangleq \log_2 \left(\frac{P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{\sum_{\boldsymbol{x}' \in \mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{x}'|\boldsymbol{x}) P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}')} \right). \tag{4.40}$$

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 (4.9), with

$$\mathcal{D}_1(i,j) = \Big\{ \boldsymbol{y} \in \mathcal{D}_1(i) : \imath_1\Big(\boldsymbol{v}(i,j), \boldsymbol{y} | \boldsymbol{u}(i)\Big) \geqslant n\eta \Big\} \setminus \bigcup_{k<j} \mathcal{D}_1(i,j), \tag{4.41}$$

where $\eta \in \mathbb{R}$ is a parameter whose exact value is determined later. Note that the codewords in (4.37), the decoding sets in (4.39) and the decoding sets in (4.41) form an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code.

Part II: Covertness Analysis

This part focuses on determining the conditions on the generating parameters to ensure that the $(n, \mathcal{C}, \hat{M}, L, \hat{\epsilon})$-random code generated is an $(n, \mathcal{C}, \hat{M}, L, \hat{\epsilon}, \delta)$-covert code.

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

$$Q_{W\boldsymbol{Y}_2}(i,\boldsymbol{y}) \triangleq \frac{1}{M} Q_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i), \text{ and} \tag{4.42}$$

$$S_{W\boldsymbol{Y}_2}(i,\boldsymbol{y}) \triangleq \frac{1}{M} S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i), \tag{4.43}$$

with

$$Q_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i) \triangleq \prod_{t=1}^{n} P_{Y_2|X}(y_t|u_t(i)), \text{ and} \tag{4.44}$$

$$S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i) \triangleq \prod_{t=1}^{n} \sum_{\hat{x} \in \mathcal{X}} P_{\hat{X}|X}(\hat{x}|u_t(i)) P_{Y_2|X}(y_t|\hat{x}). \tag{4.45}$$

Denote also by $\hat{\Lambda}_k$, with $k \in \{1,2\}$, the average of the decoding error probability at Receiver k over all possible codebooks. Using this notation, the following lemma establishes an upper-bound on the total variation $\|Q_{W\boldsymbol{Y}_2} - S_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}}$.

Lemma 4. *Given an (n, M, ϵ)-broadcast code $\mathcal{C}$ described by (4.3), it holds that*

$$\|Q_{W\boldsymbol{Y}_2} - S_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} \leqslant \|Q_{\boldsymbol{Y}_2} - S_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} + \epsilon + \hat{\Lambda}_2, \tag{4.46}$$

where the probability mass functions $Q_{\boldsymbol{Y}_2}$, $Q_{W\boldsymbol{Y}_2}$ and $S_{W\boldsymbol{Y}_2}$ are defined in (4.10), (4.42) and (4.43), respectively, and $S_{\boldsymbol{Y}_2}(\boldsymbol{y}) = \sum_{i=1}^{M} S_{W\boldsymbol{Y}_2}(i,\boldsymbol{y})$, for all $\boldsymbol{y} \in \mathcal{Y}_2^n$.

Proof: The proof of Lemma 4 is presented in Appendix F. ∎

The following proposition describes the conditions on the generating parameters to ensure that the $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code generated is an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code.

Proposition 3. *An $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code is an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code if*

$$\theta \leqslant \frac{2Q^{-1}\left(\frac{1-\delta-\epsilon-\hat{\epsilon}+\sqrt{c_n}-\frac{c}{\sqrt{n}}}{2}\right)}{\sqrt{n \sum_{x \in \mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x})}}, \tag{4.47}$$

where c is a positive constant and $c_n = 2^{-b\sqrt{n}} + 2n \log_2\left(\frac{2}{\mu_0}\right) \exp\left(-a\sqrt{n}\right)$ with a and b positive constants and $\mu_0 = \min_{(x,y) \in \mathcal{X} \times \mathcal{Y}_2} \sum_{\hat{x} \in \mathcal{X}} P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x})$.

Proof: The proof of Proposition 3 is presented in Appendix G and follows along the lines of the proof of [13, Lemma 8]. ■

Part III: Decoding Error Probability Analysis

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

$$\tilde{R}_{Y_k|X}(y|x) \triangleq \sum_{\hat{x}\in\mathcal{X}} \tilde{P}_{\hat{X}|X}(\hat{x}|x) P_{Y_k|X}(y|\hat{x}). \tag{4.48}$$

Let also $\bar{D}(\tilde{P}_{\hat{X}|X})$ and $\bar{\chi}_{2,k}(\tilde{P}_{\hat{X}|X})$, with $k \in \{1,2\}$, be respectively defined by

$$\bar{D}_1(\tilde{P}_{\hat{X}|X}) \triangleq \sum_{x\in\mathcal{X}} \sum_{\hat{x}\in\mathcal{X}} \bar{P}_X(x) \tilde{P}_{\hat{X}|X}(\hat{x}|x) D\big(P_{Y_1|X=\hat{x}} || P_{Y_1|X=x}\big), \tag{4.49}$$

and

$$\bar{\chi}_{2,k}(\tilde{P}_{\hat{X}|X}) \triangleq \sum_{x\in\mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_k|X=x}, P_{Y_k|X=x}). \tag{4.50}$$

Proposition 4. *Consider an (n,M,ϵ)-broadcast code $\mathcal{C}$ for the random transformation in (4.1). Then, there always exists an $(n,\mathcal{C},\hat{M},\hat{\epsilon},\delta)$-covert code that satisfies for all $\xi > 0$ arbitrarily small*

$$\begin{aligned}\frac{\log_2\left(\hat{M}\right)}{n} &\geqslant \max_{\theta,\tilde{P}_{\hat{X}|X}} (1-\xi)\theta \bar{D}_1(\tilde{P}_{\hat{X}|X}) \\ &= \max_{\tilde{P}_{\hat{X}|X}} (1-\xi) \frac{2Q^{-1}\left(\frac{1-\delta-\epsilon-\hat{\epsilon}+\sqrt{c_n}-\frac{c}{\sqrt{n}}}{2}\right)}{\sqrt{n\bar{\chi}_{2,2}(\tilde{P}_{\hat{X}|X})}} \bar{D}_1(\tilde{P}_{\hat{X}|X})\end{aligned} \tag{4.51}$$

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

In the asymptotic block-length regime, Proposition 4 leads to the following theorem.

Theorem 18. *Consider a sequence $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$, ..., of (n,M_n,ϵ_n)-broadcast codes for the random transformation in (4.1), with $n \in \{1,2,\ldots\}$ and*

$$\epsilon_n \leqslant \exp(-\zeta n), \tag{4.52}$$

for some fixed positive real ζ. 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 for all $\xi > 0$ arbitrarily small

$$\lim_{n\to\infty} \frac{\log_2\left(\hat{M}_n^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)}{\sqrt{n}} \geqslant \max_{\tilde{P}_{\hat{X}|X}} (1-\xi) \frac{2Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{\bar{\chi}_{2,2}(\tilde{P}_{\hat{X}|X})}} \bar{D}(\tilde{P}_{\hat{X}|X}). \tag{4.53}$$

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

$$K_n \triangleq \frac{2Q^{-1}\left(\frac{1-\delta-\epsilon_n-\hat{\epsilon}_n+\sqrt{c_n}-\frac{c}{\sqrt{n}}}{2}\right)}{\sqrt{\bar{\chi}_{2,2}(\tilde{P}_{\hat{X}|X})}}. \tag{4.54}$$

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

$$K_n < \frac{2Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{\bar{\chi}_{2,2}(\tilde{P}_{\hat{X}|X})}}. \tag{4.55}$$

Note that if ζ in (4.52) satisfies the following condition

$$\begin{aligned}\zeta > \max\Bigg\{ &\max_{\tilde{P}_{\hat{X}|X}} \ln\left(1 + \frac{2Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{n\bar{\chi}_{2,2}(\tilde{P}_{\hat{X}|X})}} \max_{(x,y)\in\mathcal{X}\times\mathcal{Y}_2} \frac{\tilde{R}_{Y_2|X}(y|x) - P_{Y_2|X}(y|x)}{P_{Y_2|X}(y|x)}\right), \\ &\max_{\tilde{P}_{\hat{X}|X}} \ln\left(1 + \frac{2Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{n\bar{\chi}_{2,2}(\tilde{P}_{\hat{X}|X})}} \max_{(x,y)\in\mathcal{X}\times\mathcal{Y}_2} \frac{\tilde{R}_{Y_1|X}(y|x) - P_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)}\right)\Bigg\}\end{aligned} \tag{4.56}$$

where the maximization is performed over all possible conditional probability mass functions $\tilde{P}_{\hat{X}|X}$, then it follows from Proposition 4, that for a fixed n, there always exists an $(n, \mathcal{C}_n, \hat{M}_n, \hat{\epsilon}_n, \delta)$-covert code such that

$$\frac{\log_2\left(\hat{M}_n\right)}{\sqrt{n}} \geqslant (1-\xi) K_n \bar{D}_1(\tilde{P}_{\hat{X}|X}). \tag{4.57}$$

In the asymptotic block-length regime, the condition in (4.56) holds for all $\zeta > 0$, which immediately implies that

$$\lim_{n\to\infty} \frac{\log_2\left(\hat{M}_n\right)}{\sqrt{n}} \geqslant (1-\xi) \frac{2Q^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{\bar{\chi}_{2,2}(\tilde{P}_{\hat{X}|X})}} \bar{D}_1(\tilde{P}_{\hat{X}|X}). \tag{4.58}$$

The proof is completed by optimizing the right-hand side of (4.58) over all possible conditional probability mass functions $\tilde{P}_{\hat{X}|X}$. ■

4.4. Impossibility of Covert Communications

Given an (n, M, ϵ)-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.

4.4.1. Auxiliary Results

One of the central parameters to characterize an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code $\hat{\mathcal{C}}$ described by (4.7) 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* $\boldsymbol{v}(i,j)$. Another parameter is number of times the symbol $x \in \mathcal{X}$ appears in the codewords from $\mathcal{C}$ and does not appear in the corresponding components of the codewords from $\hat{\mathcal{C}}$. This quantity is referred to as the *weight of the symbol* x.

Definition 31 (Weights)**.** *Given an (n, M, ϵ)-broadcast code $\mathcal{C}$ represented by the system in* (4.3)*, consider an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code $\hat{\mathcal{C}}$ represented by the system in* (4.7)*. 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:*

$$\omega(i,j) \triangleq \sum_{t=1}^{n} \mathbb{1}_{\{u_t(i) \neq v_t(i,j)\}}. \tag{4.59}$$

For all $x \in \mathcal{X}$, the weight of the symbol x, denoted by $\omega(x)$, is

$$\omega(x) \triangleq \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{t=1}^{n} \mathbb{1}_{\{u_t(i)=x\}} \mathbb{1}_{\{u_t(i) \neq v_t(i,j)\}}. \tag{4.60}$$

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 32 (Empirical Probability Distributions)**.** *Given an (n, M, ϵ)-broadcast code $\mathcal{C}$ represented by the system in* (4.3)*, consider an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code $\hat{\mathcal{C}}$ represented by the system in* (4.7)*. For all $(x, \hat{x}) \in \mathcal{X}^2$,*

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

$$\bar{P}_X(x) \triangleq \frac{1}{nM} \sum_{i=1}^{M} N(x|\boldsymbol{u}(i)); \tag{4.61}$$

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

$$\bar{P}_{X\hat{X}}(x, \hat{x}) \triangleq \frac{1}{nM\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} N(x, \hat{x}|\boldsymbol{u}(i), \boldsymbol{v}(i, j)); \tag{4.62}$$

- *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 $\hat{P}_{\hat{X}|X}$, is:*

$$\hat{P}_{\hat{X}|X}(\hat{x}|x) \triangleq \frac{\sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{t=1}^{n} \mathbb{1}_{\{x=u_t(i)\}} \mathbb{1}_{\{\hat{x}=v_t(i,j)\}} \mathbb{1}_{\{x \neq \hat{x}\}}}{\sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{t=1}^{n} \mathbb{1}_{\{x=u_t(i)\}} \mathbb{1}_{\{u_t(i) \neq v_t(i,j)\}}}, \tag{4.63}$$

  and

$$\operatorname{supp} \hat{P}_{\hat{X}|X=x} \subseteq \mathcal{X} \setminus \{x\}; \tag{4.64}$$

- *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*

$$\theta(x) \triangleq 1 - \bar{P}_{\hat{X}|X}(x|x), \tag{4.65}$$

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

$$\bar{P}_{\hat{X}X}(x,x) = \bar{P}_X(x)\bar{P}_{\hat{X}|X}(x|x). \tag{4.66}$$

The next lemma establishes relations between the empirical probability mass functions above and the weights.

Lemma 5. *Given an* (n, M, ϵ)*-broadcast code* $\mathcal{C}$ *represented by the system in* (4.3)*, consider an* $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$*-induced code* $\hat{\mathcal{C}}$ *represented by the system in* (4.7)*. For all* $(x, \hat{x}) \in \mathcal{X}^2$*, it holds that*

$$\bar{P}_{X\hat{X}}(x,\hat{x}) = \bar{P}_X(x)\left((1-\theta(x))\,\mathbb{1}_{\{x=\hat{x}\}} + \theta(x)\hat{P}_{\hat{X}|X}(\hat{x}|x)\right), \tag{4.67}$$

$$\omega(x) = n\bar{P}_X(x)\theta(x), \quad \textit{and} \tag{4.68}$$

$$n\sum_{x\in\mathcal{X}} \bar{P}_X(x)\theta(x) = \sum_{x\in\mathcal{X}} \omega(x) = \frac{1}{M\hat{M}}\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}} \omega(i,j). \tag{4.69}$$

Proof: The proof of Lemma 5 is presented in Appendix K. ∎

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

$$Q_{W\boldsymbol{Y}_2}(i,\boldsymbol{y}) \triangleq \frac{1}{M} Q_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i), \text{ and} \tag{4.70}$$

$$R_{W\boldsymbol{Y}_2}(i,\boldsymbol{y}) \triangleq \frac{1}{M} R_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i), \tag{4.71}$$

with

$$Q_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i) \triangleq \prod_{t=1}^{n} P_{Y_2|X}(y_t|u_t(i)), \text{ and} \tag{4.72}$$

$$R_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i) \triangleq \frac{1}{\hat{M}}\sum_{j=1}^{\hat{M}}\prod_{t=1}^{n} P_{Y_2|X}(y_t|v_t(i,j)). \tag{4.73}$$

Note that the marginal probability mass functions $Q_{\boldsymbol{Y}_2}$ and $R_{\boldsymbol{Y}_2}$ are respectively in (4.10) and (4.11).

Using this notation, the following lemma highlights that replacing the constraint $\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} < \delta$ in (4.17) by the constraint $\|Q_{W\boldsymbol{Y}_2} - R_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} < \delta$ is equivalent up to an additive constant.

Lemma 6. *Given an* (n, M, ϵ)*-broadcast code* $\mathcal{C}$ *described by* (4.3)*, any* $(n, \mathcal{C}, \hat{M}, L, \hat{\epsilon})$*-random code described by* (4.7) *satisfies*

$$\|Q_{W\boldsymbol{Y}_2} - R_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} \leqslant \|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} + \epsilon + \hat{\epsilon}, \tag{4.74}$$

where the probability mass functions $Q_{\boldsymbol{Y}_2}$*,* $R_{\boldsymbol{Y}_2}$*,* $Q_{W\boldsymbol{Y}_2}$ *and* $R_{W\boldsymbol{Y}_2}$ *are defined in* (4.10)*,* (4.11)*,* (4.70) *and* (4.71)*, respectively.*

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

4.4.2. Finite Block-length Results

For all $k \in \{1,2\}$ and all $(x,y) \in \mathcal{X} \times \mathcal{Y}_k$, define

$$\hat{R}_{Y_k|X}(y|x) = \sum_{\hat{x}\in\mathcal{X}} \hat{P}_{\hat{X}|X}(\hat{x}|x) P_{Y_k|X}(y|\hat{x}). \tag{4.75}$$

Using Fano's inequility [38], 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 5. *Consider an (n, M, ϵ)-broadcast code $\mathcal{C}$, described by the system in (4.3), for the random transformation in (4.1). Then, every $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code satisfies that*

$$\log_2\left(\hat{M}\right) \leqslant \frac{1}{1-\hat{\epsilon}}\Bigg(1 + \sum_{x\in\mathcal{X}} \bar{P}_X(x)\Bigg(n \sum_{\hat{x}\in\mathcal{X}} \theta(x)\hat{P}_{\hat{X}|X}(\hat{x}|x)$$
$$\cdot D\Big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\Big) + \frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x}, P_{Y_1|X=x})\Bigg)\Bigg). \tag{4.76}$$

Proof: The proof of Proposition 5 is presented in Appendix M and follows from Fano's inequality [38]. ∎

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 31) is bounded. More importantly, for a special class of channels, the cardinality of the set of upper-bounded-weight codewords can be lower-bounded. This result is presented by the following proposition, which is reminiscent of [13, Lemma 12].

Proposition 6. *Let $\eta > 0$ be arbitrarily small and assume that for all pairs $(x, x') \in \mathcal{X}^2$ such that $x \neq x'$, the random transformation in (4.1) satisfies the following conditions:*

$$\chi_2\left(P_{Y_2|X=x}, P_{Y_2|X=x'}\right) = d, \quad \text{and} \tag{4.77}$$
$$D\Big(P_{Y_1|X=x}||P_{Y_1|X=x'}\Big) = \ell, \tag{4.78}$$

where $(d, \ell) \in \mathbb{R}_+^2$. Consider an (n, M, ϵ)-broadcast code $\mathcal{C}$, described by the system in (4.3), for this random transformation. Then, every $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code described by the system in (4.7) can be formed by two sub-codes. One sub-code whose codewords are in the set

$$\tilde{\mathcal{W}} = \left\{\boldsymbol{v}(i,j) \colon \omega(i,j) < 2\sqrt{\frac{n}{d}}Q^{-1}\Big(\frac{1-\delta-\eta}{2}\Big), 1 \leqslant i \leqslant M, \text{ and } 1 \leqslant j \leqslant \hat{M}\right\}, \tag{4.79}$$

and another sub-code whose codewords are in the set

$$\tilde{\mathcal{W}}^{\mathsf{c}} = \left\{\boldsymbol{v}(i,j) \colon \omega(i,j) \geqslant 2\sqrt{\frac{n}{d}}Q^{-1}\Big(\frac{1-\delta-\eta}{2}\Big), 1 \leqslant i \leqslant M, \text{ and } 1 \leqslant j \leqslant \hat{M}\right\}. \tag{4.80}$$

Moreover,

$$\left|\tilde{\mathcal{W}}\right| > M\hat{M}\left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon - \hat{\epsilon}\right), \tag{4.81}$$

where c is a constant.

Proof: The proof of Proposition 6 is presented in Appendix N. ■

Note that the binary symmetric channel satisfies (4.77) and (4.78). Another example is presented in Section 4.6.

4.4.3. Asymptotic Result

The following theorem introduces the main result of this section.

Theorem 19. *Consider a sequence $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$, ..., of (n, M_n, ϵ_n)-broadcast codes for the random transformation in (4.1), with $\lim_{n\to\infty} \epsilon_n = 0$. Assume that the random transformation in (4.1) satisfies (4.77) and (4.78). 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*

$$\lim_{n\to\infty} \frac{\log_2\left(\hat{M}_n^*(n, \mathcal{C}_n, 1, \hat{\epsilon}_n, \delta)\right)}{\sqrt{n}} < \frac{2\ell}{\sqrt{d}} Q^{-1}\left(\frac{1-\delta-\eta}{2}\right), \tag{4.82}$$

with $\eta > 0$ arbitrarily small.

Proof: For all $n \in \mathbb{N}$, it follows from Proposition 6 that the covert sub-code of the covert code $\hat{\mathcal{C}}_n$ with codewords in the set

$$\tilde{\mathcal{W}}_n = \left\{ \boldsymbol{v}(i,j) \colon \omega(i,j) < 2\sqrt{\frac{n}{d}} Q^{-1}\Big(\frac{1-\delta-\eta}{2}\Big), 1 \leqslant i \leqslant M_n, \text{ and } 1 \leqslant j \leqslant \hat{M}_n \right\}, \tag{4.83}$$

satisfies

$$\left|\tilde{\mathcal{W}}_n\right| > M_n \hat{M}_n \left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon_n - \hat{\epsilon}_n\right). \tag{4.84}$$

Hence, from (4.84), it follows that for all index $i \in \mathcal{W}$, there are in average $\hat{M}_n \left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon_n - \hat{\epsilon}_n\right)$ codewords in the subcode. Thus, Proposition 5 applies, and it follows that

$$\begin{aligned} &\log_2\left(\hat{M}_n \left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon_n - \hat{\epsilon}_n\right)\right) \leqslant \\ &\qquad \frac{1}{1-\hat{\epsilon}_n}\left(1 + n\sum_{x\in\mathcal{X}} \ell \bar{P}_X(x)\theta(x) + \bar{P}_X(x)\frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x}, P_{Y_1|X=x})\right), \end{aligned} \tag{4.85}$$

which implies that

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

$$\overset{(b)}{\leqslant} \frac{1}{1-\hat{\epsilon}_n}\Bigg(1+2\frac{\ell\sqrt{n}}{\sqrt{d}}Q^{-1}\left(\frac{1-\delta-\eta}{2}\right)+\left(\frac{2\sqrt{n}}{\sqrt{d}}Q^{-1}\left(\frac{1-\delta-\eta}{2}\right)\right)^3$$
$$\cdot\sum_{x\in\mathcal{X}}\max_{x'\in\mathcal{X}}\frac{\chi_3(\hat{R}_{Y_1|X=x'},P_{Y_1|X=x'})}{6n^2\bar{P}_X(x')^2}\Bigg)-\log_2\left(\frac{\eta}{2}-\frac{c}{\sqrt{n}}-\epsilon_n-\hat{\epsilon}_n\right)$$
$$=\frac{1}{1-\hat{\epsilon}_n}\Bigg(1+2\frac{\ell\sqrt{n}}{\sqrt{d}}Q^{-1}\left(\frac{1-\delta-\eta}{2}\right)+\frac{4\left|\mathcal{X}\right|}{3\sqrt{n}\sqrt{d}^3}$$
$$\cdot Q^{-1}\left(\frac{1-\delta-\eta}{2}\right)^3\max_{x'\in\mathcal{X}}\frac{\chi_3(\hat{R}_{Y_1|X=x'},P_{Y_1|X=x'})}{\bar{P}_X(x')^2}\Bigg)-\log_2\left(\frac{\eta}{2}-\frac{c}{\sqrt{n}}-\epsilon_n-\hat{\epsilon}_n\right), \quad (4.86)$$

where c is a constant that depends only on the parameters of the random transformation in (4.1). Note that *(a)* follows from Lemma 5, and *(b)* follows from the fact that for all $x\in\mathcal{X}$,

$$\begin{aligned}\omega(x)&\leqslant\sum_{v\in\mathcal{X}}\omega(v)\\&=\sum_{i=1}^{M_n}\sum_{j=1}^{\hat{M}_n}\frac{\omega(i,j)}{M_n\hat{M}_n}\\&\leqslant 2\frac{\sqrt{n}}{\sqrt{d}}Q^{-1}\left(\frac{1-\delta-\eta}{2}\right).\end{aligned} \quad (4.87)$$

The proof is completed by dividing both hand-sides of (4.86) by $\sqrt{n}$ and taking the limit. ■

4.5. Main Result

Note that for channels satisfying (4.77) and (4.78), the right-hand side of (4.58) reduces to

$$(1-\xi)\frac{2\ell}{\sqrt{d}}Q^{-1}\left(\frac{1-\delta}{2}\right). \quad (4.88)$$

Recalling that ξ and η in (4.82) can be chosen arbitrarily small, it follows that, for such channels, the asymptotic bounds in Theorem 18 and Theorem 19 are tight, *i.e.*, (4.88) gives the optimal scaling constant for $\log_2\left(\hat{M}_n^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)$ with respect to $\sqrt{n}$.

Theorem 20. *Consider a sequence $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$, ..., of (n,M_n,ϵ_n)-broadcast codes for the random transformation in* (4.1) *such that* (4.77) *and* (4.78) *are verfied, with $n\in\{1,2,\ldots\}$ and*

$$\epsilon_n\leqslant\exp(-\zeta n), \quad (4.89)$$

for some fixed positive real ζ. Then,

$$\lim_{n\to\infty}\frac{\log_2\left(\hat{M}_n^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)}{\sqrt{n}}=\frac{2\ell}{\sqrt{d}}Q^{-1}\left(\frac{1-\delta}{2}\right). \quad (4.90)$$

4.6. Examples

This section presents examples to illustrate the results in Theorem 20.

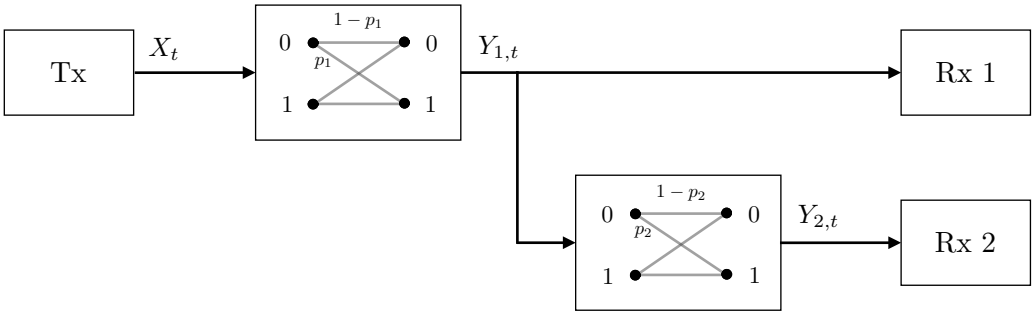

Figure 4.4.: Degraded broadcast channel satisfying (4.77) and (4.78) at channel use $t \in \{1, 2, \ldots, n\}$.

Example 3 (Binary Symmetric Channel)**.** *Consider the random transformation in* (4.1) *such that* $\mathcal{X} = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1\}$*, 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:*

$$P_{Y_1|X}(x|x) = 1 - P_{Y_1|X}(x'|x) = 1 - p_1, \text{ and} \tag{4.91}$$
$$P_{Y_2|Y_1}(x|x) = 1 - P_{Y_2|Y_1}(x'|x) = 1 - p_2, \tag{4.92}$$

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

Figure 4.4 depicts the channel in Example 3. The probability distribution $P_{Y_2|X}$ verifies that for all $(x, x') \in \mathcal{X}^2$ such that $x \neq x'$:

$$P_{Y_2|X}(x|x) = 1 - P_{Y_2|X}(x'|x) = 1 - p, \tag{4.93}$$

with

$$p = p_1 + p_2 - 2p_1p_2. \tag{4.94}$$

Hence, it follows that for any pair $(x, x') \in \mathcal{X}^2$ with $x \neq x'$,

$$\chi_2(P_{Y_2|X=x'}, P_{Y_2|X=x}) = \frac{(1-2p)^2}{p(1-p)}, \tag{4.95}$$
$$D\big(P_{Y_1|X=x'} || P_{Y_1|X=x}\big) = (1-2p_1)\log_2\left(\frac{1-p_1}{p_1}\right), \tag{4.96}$$

It follows as an immediate consequence of Theorem 20 that

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

with p in (4.94).

The above expression is plotted as a function of the probability p_1 and p_2 in Figure 4.5 and in Figure 4.6, respectively, with $\delta = 0.005$.

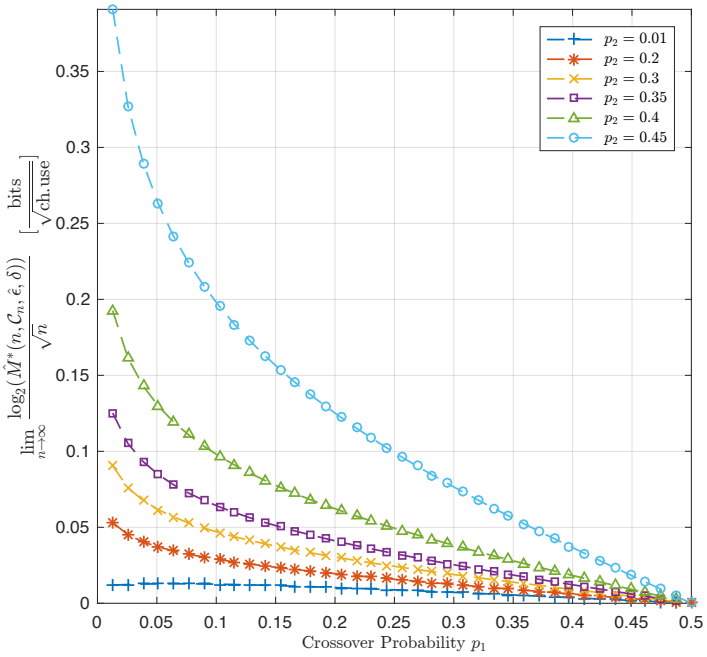

Figure 4.5.: Fundamental limit $\lim_{n\to\infty} \frac{\log_2\left(\hat{M}^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)}{\sqrt{n}}$ as a function of the crossover probability p_1, for $\delta = 0.005$.

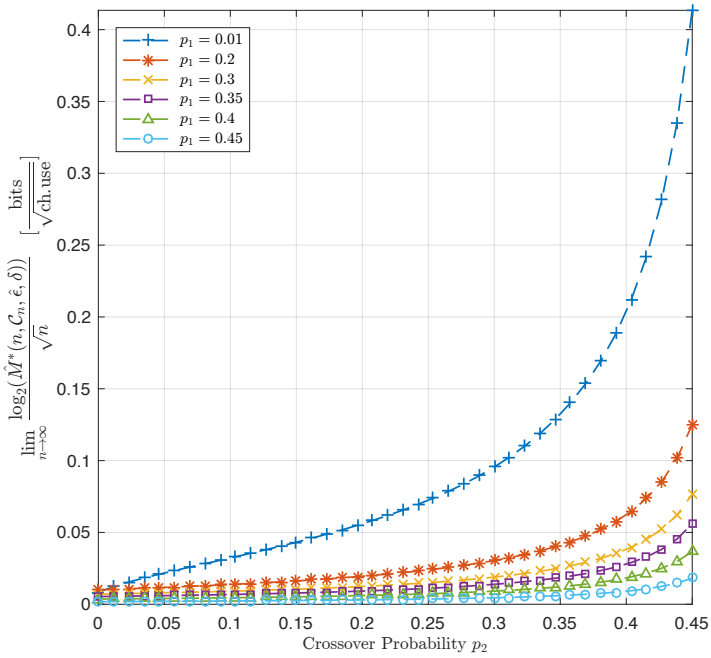

Figure 4.6.: Fundamental limit $\lim_{n\to\infty} \frac{\log_2\left(\hat{M}^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)}{\sqrt{n}}$ as a function of the crossover probability p_2, for $\delta = 0.005$.

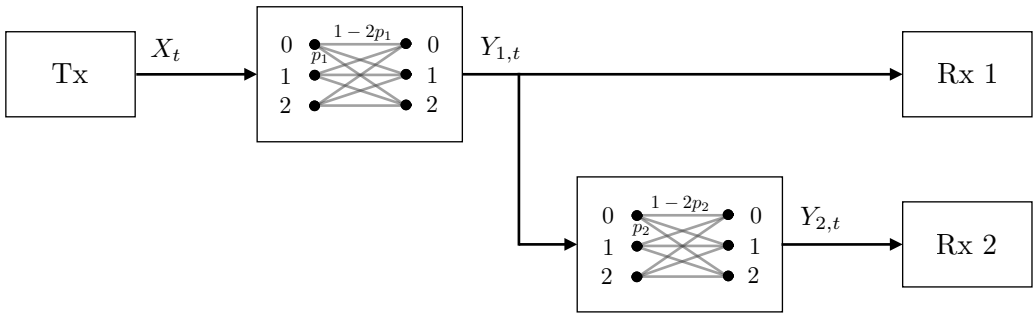

Figure 4.7.: Degraded broadcast channel satisfying (4.77) and (4.78) at channel use $t \in \{1, 2, \ldots, n\}$.

Example 4. *Consider the random transformation in* (4.1) *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 mass functions* $P_{Y_1|X}$ *and* $P_{Y_2|Y_1}$ *respectively satisfy:*

$$P_{Y_1|X}(x|x) = 1 - 2P_{Y_1|X}(x'|x) = 1 - 2p_1, \text{ and} \tag{4.97}$$
$$P_{Y_2|Y_1}(x|x) = 1 - 2P_{Y_2|Y_1}(x'|x) = 1 - 2p_2, \tag{4.98}$$

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

Figure 4.7 depicts the channel in Example 4. The probability mass function $P_{Y_2|X}$ verifies that for all $(x, x') \in \mathcal{X}^2$ such that $x \neq x'$:

$$\begin{aligned} P_{Y_2|X}(x|x) = 1 - 2P_{Y_2|X}(x'|x) &= 1 - 2\left(p_1 + p_2 - 3p_1p_2\right) \\ &= 1 - 2p, \end{aligned} \tag{4.99}$$

with

$$p = p_1 + p_2 - 3p_1p_2. \tag{4.100}$$

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

Lemma 7. *Consider Example 4. For all pairs* $(x, x') \in \mathcal{X}^2$, *with* $x \neq x'$, *it holds that*

$$\chi_2(P_{Y_2|X=x'}, P_{Y_2|X=x}) = \frac{(3p-1)^2(1-p)}{p(1-2p)}, \tag{4.101}$$
$$D\big(P_{Y_1|X=x'}||P_{Y_1|X=x}\big) = (1-3p_1)\log_2\left(\frac{1-2p_1}{p_1}\right), \tag{4.102}$$

where p *is defined in* (4.100).

Proof: The proof of Lemma 7 is presented in Appendix O. ∎

The following proposition follows immediately from Lemma 7 and Theorem 20.

Proposition 7. *Consider Example 4 and consider a sequence of* (n, M_n, ϵ_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,*

$$\lim_{n\to\infty} \frac{\log_2\left(\hat{M}_n^*(n, \mathcal{C}_n, L_n, \hat{\epsilon}_n, \delta)\right)}{\sqrt{n}} = 2Q^{-1}\left(\frac{1-\delta}{2}\right)\sqrt{\frac{p(1-2p)}{1-p}}\frac{1-3p_1}{1-3p}\log_2\left(\frac{1-2p_1}{p_1}\right) \quad (4.103)$$

The left hand side of (4.103) is plotted as a function of the probability p_1 and p_2 in Figure 4.8 and in Figure 4.9, respectively, with $\delta = 0.005$.

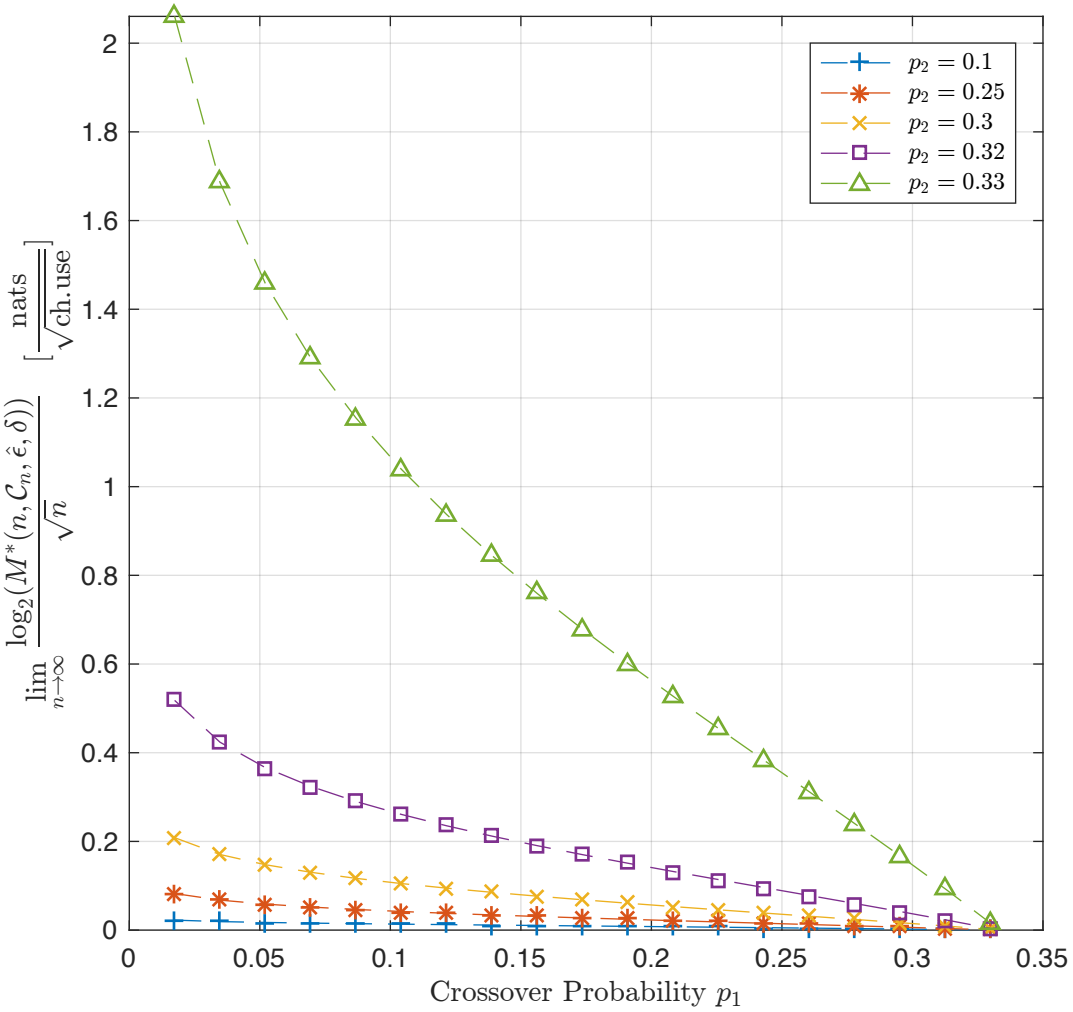

Figure 4.8.: Fundamental limit $\lim_{n\to\infty} \frac{\log_2\left(\hat{M}^*(n, \mathcal{C}_n, \hat{\epsilon}_n, \delta)\right)}{\sqrt{n}}$ as a function of the crossover probability p_1, for $\delta = 0.005$.

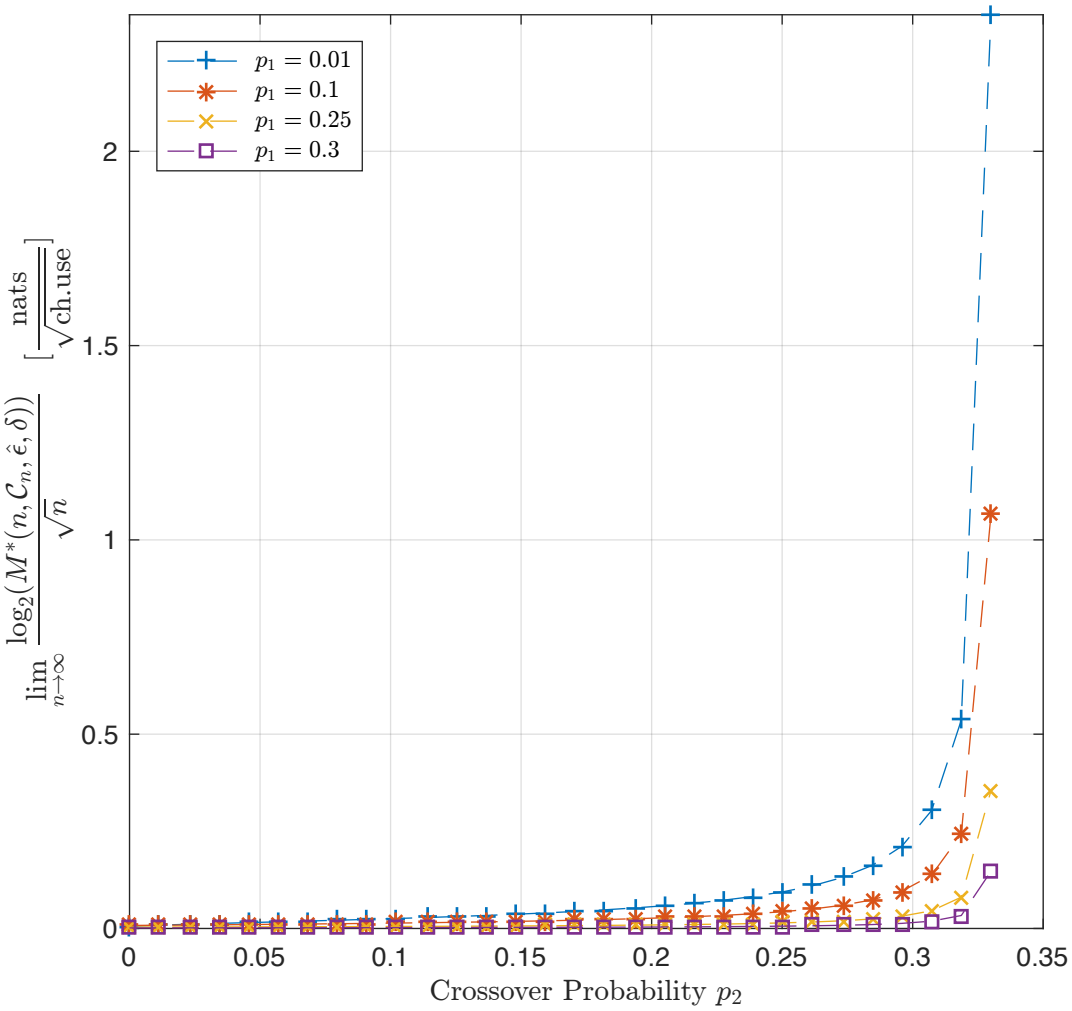

Figure 4.9.: Fundamental limit $\lim_{n\to\infty} \frac{\log_2\left(\hat{M}^*(n,\mathcal{C}_n,\hat{\epsilon}_n,\delta)\right)}{\sqrt{n}}$ as a function of the crossover probability p_1, for $\delta = 0.005$.

4.7. Conclusion

So far, a tight converse for general DM-BCs, i.e., those that do not necessarily satisfy the conditions in (4.77) and (4.78) 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 chapter 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.

— 5 —

Conclusion

In this thesis two new problems have been introduced. First, the problem of covertly transmitting information over a point-to-point channel when the warden observes only a fraction of the channel outputs, termed covert communications type II, was described. This model generalizes the problem of covert communications over point-to-point channels. An achievability bound in the finite block-length regime has been derived in Chapter 3 for this problem. This bound reveals two regimes of communications: one in which the point-to-point capacity of the channel is achievable and one in which the rate verifies the square-root law of covert communications. This work constitutes a first step towards resolving this general model, *i.e.*, determining the maximum rate at which information can be simultaneously reliably and covertly transmitted.

Second, the problem of embedding covert information on a given broadcast code was presented. In contrast with previous works, the broadcast code is assumed to be given which makes the achievability proof more difficult. An achievability and converse bound in the asymptotic block-length regime have been derived in Chapter 4 for a particular class of channels, *i.e.*, channels that satisfy a symmetry condition. Together these bounds characterize the maximum number of information bits that can be covertly embedded in a given broadcast code for symmetric channels. This work constitutes a first step towards resolving the problem for arbitrary DMCs.

These two contributions open a number of perspectives both in theory and application. From a theoretical standpoint, the contribution presented here on covert communications type II leaves open the problem of finding a tight converse in the finite block-length regime. Since the converse is not known, it might be possible that the achievability bound presented here can be improved, even though the first order term is optimal. In addition, the characterization of the optimal covert information rate under other constraints than the Kullback-Leibler divergence is left open. On the other hand, the contribution on broadcast channels leaves open the problem of finding both an achievability and converse bound for general DMCs. The characterization of the maximum number of information bits that can be covertly embedded under other types of constraints than the total variation constraint presented in this thesis is also left open. It is interesting to highlight that the problem of embedding covert information into a given broadcast code 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. In the two problems, only DMCs are considered, which leaves open the AWGN channel case. In addition, these bounds pave the way for coding theorists to develop codes that can meet these bounds. As discussed in [34, 35, 36], pulse position modulation and variations around this type of modulation seems to be a good candidate to achieve the optimal rate in the point-to-point case. This might extend to the broadcast setting as well.

In terms of applications, the problem of covert communications type II opens the way to the design of a variety of communications systems that can covertly transmit information. For instance, covert communications could be used to transmit control signals in the network which are low-rate signals that are often observed by attackers, leading to security issues. On the other hand, the problem of embedding covert information into a given broadcast code opens two main perspectives. First, this problem opens the way to re-designing existing broadcast communications systems in order to add a service, which is covertly transmitting an additional message. Second, the bounds derived for this problem show that there exist malware that can affect the transmitter so that the latter leaks information to some destination. In this case, the leakage can not be detected by monitoring the network but only by checking the code at the transmitter, which might in some cases be hard to do.

In addition, in both problems, friendly jamming could be used to improve the communication rate as discussed in [25]. Therefore, this makes full-duplex systems valuable candidates for the implementation of such covert communications systems.

— A —

Auxiliary Results

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

Theorem 21 (Berry-Esseen Theorem, [45, 46, 47])**.** *Let X_1, X_2, ..., X_n be independent random variables such that for all $t \in \{1, 2, \ldots, n\}$,*

$$\mu_t = \mathbb{E}_{X_t}\left[X_t\right], \tag{A.1}$$
$$\sigma_t^2 = \mathbb{E}_{X_t}\left[X_t^2\right] - \mu_t^2, \tag{A.2}$$
$$\phi_t = \mathbb{E}_{X_t}\left[|X_t - \mu_t|^3\right]. \tag{A.3}$$

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

$$\left|\Pr\left[\sum_{t=1}^{n} X_t - \mu_t \geqslant \sigma\lambda\right] - Q(\lambda)\right| \leqslant \frac{c_0\phi}{\sigma^3}, \tag{A.4}$$

where

$$\mu = \sum_{t=1}^{n}\mu_t, \quad \sigma^2 = \sum_{t=1}^{n}\sigma_t^2, \quad \text{and} \quad \phi = \sum_{t=1}^{n}\phi_t. \tag{A.5}$$

The best value of the constant c_0 is $c_0 = 0.4748$ [48].

Lemma 8. *Let P_X and P_Y be two probability mass functions on a common finite support $\mathcal{Z}$. Let also X and Y be two random variables following the probability mass functions P_X and P_Y, respectively. Then,*

$$\|P_X - P_Y\|_{\mathrm{TV}} = \Pr\left[P_X(X) \geqslant P_Y(X)\right] - \Pr\left[P_X(Y) \geqslant P_Y(Y)\right], \tag{A.6}$$

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

Lemma 8 is part of the proof of [13, Lemma 8]. The proof of the Lemma is provided here for sake of completeness.

Proof: The proof consists in the following algebraic manipulations:

$$\begin{aligned}
\|P_X - P_Y\|_{\mathrm{TV}} &= \frac{1}{2}\sum_{z\in\mathcal{Z}} |P_X(z) - P_Y(z)| \\
&= \frac{1}{2}\sum_{\substack{z\in\mathcal{Z}:\\ P_X(z)\geqslant P_Y(z)}} P_X(z) - P_Y(z) + \frac{1}{2}\sum_{\substack{z\in\mathcal{Z}:\\ P_X(z)\leqslant P_Y(z)}} P_Y(z) - P_X(z) \\
&= \frac{1}{2}\sum_{z\in\mathcal{Z}} P_X(z)\mathbb{1}_{\{P_X(z)\geqslant P_Y(z)\}} - \frac{1}{2}\sum_{z\in\mathcal{Z}} P_Y(z)\mathbb{1}_{\{P_X(z)\geqslant P_Y(z)\}} \\
&\quad + \frac{1}{2}\sum_{z\in\mathcal{Z}} P_Y(z)\mathbb{1}_{\{P_X(z)\leqslant P_Y(z)\}} - \frac{1}{2}\sum_{z\in\mathcal{Z}} P_X(z)\mathbb{1}_{\{P_X(z)\leqslant P_Y(z)\}} \\
&= \frac{1}{2}\Big(\Pr\left[P_X(X) \geqslant P_Y(X)\right] - \Pr\left[P_X(Y) \geqslant P_Y(Y)\right]) \\
&\quad + \Pr\left[P_X(Y) \leqslant P_Y(Y)\right] - \Pr\left[P_X(X) \leqslant P_Y(X)\right]\Big) \\
&= \Pr\left[P_X(X)\geqslant P_Y(X)\right] - \Pr\left[P_X(Y)\geqslant P_Y(Y)\right], \qquad \text{(A.7)}
\end{aligned}$$

and this completes the proof. ■

— B —

Proof of Proposition 1

This appendix proves Proposition 1.

Note that the probability mass function $R_{\boldsymbol{Z}_\mathcal{B}}$ verifies

$$
\begin{aligned}
R_{\boldsymbol{Z}_\mathcal{B}}(\boldsymbol{z}_\mathcal{B}) &= \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{L} P_A(j) P_{\boldsymbol{Z}_\mathcal{B}|\boldsymbol{X}_\mathcal{B}}(\boldsymbol{z}_\mathcal{B}|\boldsymbol{u}_\mathcal{B}(i,j)) \\
&= \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{L}\prod_{k=1}^{M} P_{\boldsymbol{X}}(\boldsymbol{u}(k,j)) P_{\boldsymbol{Z}_\mathcal{B}|\boldsymbol{X}_\mathcal{B}}(\boldsymbol{z}_\mathcal{B}|\boldsymbol{u}_\mathcal{B}(i,j)) \\
&= \frac{1}{M}\sum_{i=1}^{M}\sum_{\boldsymbol{u}_1\in\mathcal{X}^n}\sum_{\boldsymbol{u}_2\in\mathcal{X}^n}\cdots\sum_{\boldsymbol{u}_M\in\mathcal{X}^n}\prod_{k=1}^{M} P_{\boldsymbol{X}}(\boldsymbol{u}_k) P_{\boldsymbol{Z}_\mathcal{B}|\boldsymbol{X}_\mathcal{B}}(\boldsymbol{z}_\mathcal{B}|\boldsymbol{u}_{i,\mathcal{B}}) \\
&= \frac{1}{M}\sum_{i=1}^{M}\prod_{k=1}^{M}\sum_{\boldsymbol{u}_k\in\mathcal{X}^n} P_{\boldsymbol{X}}(\boldsymbol{u}_k) P_{\boldsymbol{Z}_\mathcal{B}|\boldsymbol{X}_\mathcal{B}}(\boldsymbol{z}_\mathcal{B}|\boldsymbol{u}_{i,\mathcal{B}}) \\
&= \frac{1}{M}\sum_{i=1}^{M}\sum_{\boldsymbol{u}_i\in\mathcal{X}^n} P_{\boldsymbol{X}}(\boldsymbol{u}_i) P_{\boldsymbol{Z}_\mathcal{B}|\boldsymbol{X}_\mathcal{B}}(\boldsymbol{z}_\mathcal{B}|\boldsymbol{u}_{i,\mathcal{B}}) \\
&= \frac{1}{M}\sum_{i=1}^{M}\sum_{\boldsymbol{x}_\mathcal{B}\in\mathcal{X}^m} P_{\boldsymbol{X}_\mathcal{B}}(\boldsymbol{x}_\mathcal{B}) P_{\boldsymbol{Z}_\mathcal{B}|\boldsymbol{X}_\mathcal{B}}(\boldsymbol{z}_\mathcal{B}|\boldsymbol{x}_\mathcal{B}) \\
&= \sum_{\boldsymbol{x}_\mathcal{B}\in\mathcal{X}^m} P_{\boldsymbol{X}_\mathcal{B}}(\boldsymbol{x}_\mathcal{B}) P_{\boldsymbol{Z}_\mathcal{B}|\boldsymbol{X}_\mathcal{B}}(\boldsymbol{z}_\mathcal{B}|\boldsymbol{x}_\mathcal{B}) \\
&= \sum_{\boldsymbol{x}_\mathcal{B}\in\mathcal{X}^m}\prod_{t=1}^{m} P_X(x_{b_t}) P_{Z|X}(z_{b_t}|x_{b_t}) \\
&= \prod_{t=1}^{m}\sum_{x\in\mathcal{X}} P_X(x) P_{Z|X}(z_{b_t}|x).
\end{aligned}
\tag{B.1}
$$

Therefore, it follows that

$$
\begin{aligned}
&\max_{\substack{\mathcal{B}\subseteq\{1,2,\ldots,n\}\\|\mathcal{B}|=\lfloor\kappa n\rfloor}} D(R_{\boldsymbol{Z}_\mathcal{B}}||Q_{\boldsymbol{Z}_\mathcal{B}}) \\
&= \max_{\substack{\mathcal{B}\subseteq\{1,2,\ldots,n\}\\|\mathcal{B}|=\lfloor\kappa n\rfloor}} D\big(R_{\boldsymbol{Z}_\mathcal{B}}||P_{\boldsymbol{Z}_\mathcal{B}|\boldsymbol{X}_\mathcal{B}=\boldsymbol{x}_0}\big) \\
&= \max_{\substack{\mathcal{B}\subseteq\{1,2,\ldots,n\}\\|\mathcal{B}|=\lfloor\kappa n\rfloor}} \sum_{\boldsymbol{z}_\mathcal{B}\in\mathcal{Z}^m} R_{\boldsymbol{Z}_\mathcal{B}}(\boldsymbol{z}_\mathcal{B}) \log\left(\frac{R_{\boldsymbol{Z}_\mathcal{B}}(\boldsymbol{z}_\mathcal{B})}{P_{\boldsymbol{Z}_\mathcal{B}|\boldsymbol{X}_\mathcal{B}}(\boldsymbol{z}_\mathcal{B}|\boldsymbol{x}_0)}\right) \\
&= \max_{\substack{\mathcal{B}\subseteq\{1,2,\ldots,n\}\\|\mathcal{B}|=\lfloor\kappa n\rfloor}} \sum_{\boldsymbol{z}_\mathcal{B}\in\mathcal{Z}^m} R_{\boldsymbol{Z}_\mathcal{B}}(\boldsymbol{z}_\mathcal{B}) \log\left(\prod_{t=1}^{m}\frac{\sum_{x\in\mathcal{X}} P_X(x)P_{Z|X}(z_{b_t}|x)}{P_{Z|X}(z_{b_t}|x_0)}\right) \\
&= \max_{\substack{\mathcal{B}\subseteq\{1,2,\ldots,n\}\\|\mathcal{B}|=\lfloor\kappa n\rfloor}} \sum_{\boldsymbol{z}_\mathcal{B}\in\mathcal{Z}^m} R_{\boldsymbol{Z}_\mathcal{B}}(\boldsymbol{z}_\mathcal{B}) \sum_{t=1}^{m}\log\left(\frac{\sum_{x\in\mathcal{X}} P_X(x)P_{Z|X}(z_{b_t}|x)}{P_{Z|X}(z_{b_t}|x_0)}\right) \\
&= \max_{\substack{\mathcal{B}\subseteq\{1,2,\ldots,n\}\\|\mathcal{B}|=\lfloor\kappa n\rfloor}} \sum_{t=1}^{m}\sum_{z_{b_t}\in\mathcal{Z}}\sum_{x'\in\mathcal{X}} P_X(x')P_{Z|X}(z_{b_t}|x')\log\left(\frac{\sum_{x\in\mathcal{X}} P_X(x)P_{Z|X}(z_{b_t}|x)}{P_{Z|X}(z_{b_t}|x_0)}\right) \\
&= m\sum_{z\in\mathcal{Z}}\sum_{x'\in\mathcal{X}} P_X(x')P_{Z|X}(z|x')\log\left(\frac{\sum_{x\in\mathcal{X}} P_X(x)P_{Z|X}(z|x)}{P_{Z|X}(z|x_0)}\right) \\
&= mD\big(R_Z||P_{Z|X=x_0}\big), \qquad \text{(B.2)}
\end{aligned}
$$

with

$$
R_Z(z) = \sum_{x\in\mathcal{X}} P_X(x)P_{Z|X}(z|x), \qquad \text{(B.3)}
$$

for all $z \in \mathcal{Z}$.

Therefore, to ensure covertness, it is sufficient to satisfy

$$
D\big(R_Z||P_{Z|X=x_0}\big) \leqslant \frac{\delta}{\lfloor\kappa n\rfloor}. \qquad \text{(B.4)}
$$

In addition, it holds that

$$
D\big(R_Z||P_{Z|X=x_0}\big) = \sum_{z\in\mathcal{Z}}\sum_{x'\in\mathcal{X}} P_X(x')P_{Z|X}(z|x')\log\left(\frac{\sum_{x\in\mathcal{X}} P_X(x)P_{Z|X}(z|x)}{P_{Z|X}(z|x_0)}\right)
$$

$$
\begin{aligned}
&= \sum_{z\in\mathcal{Z}} \Big((1-\theta)\, P_{Z|X}(z|x_0) + \theta P_{Z|X}(z|x_1)\Big) \log\left(\frac{\sum\limits_{x\in\mathcal{X}} P_X(x) P_{Z|X}(z|x)}{P_{Z|X}(z|x_0)}\right) \\
&= \sum_{z\in\mathcal{Z}} \Big((1-\theta)\, P_{Z|X}(z|x_0) + \theta P_{Z|X}(z|x_1)\Big) \log\left(1+\theta\frac{P_{Z|X}(z|x_1)-P_{Z|X}(z|x_0)}{P_{Z|X}(z|x_0)}\right) \\
&= \sum_{z\in\mathcal{Z}} \Big((1-\theta)\, P_{Z|X}(z|x_0) + \theta P_{Z|X}(z|x_1)\Big) \left(\theta\frac{P_{Z|X}(z|x_1)-P_{Z|X}(z|x_0)}{P_{Z|X}(z|x_0)}\right. \\
&\quad \left. -\frac{\theta^2}{2}\left(\frac{P_{Z|X}(z|x_1)-P_{Z|X}(z|x_0)}{P_{Z|X}(z|x_0)}\right)^2 + o(\theta^2)\right) \\
&= \theta \sum_{z\in\mathcal{Z}} P_{Z|X}(z|x_0) - P_{Z|X}(z|x_1) + \theta^2 \sum_{z\in\mathcal{Z}} \frac{(P_{Z|X}(z|x_1)-P_{Z|X}(z|x_0))^2}{P_{Z|X}(z|x_0)} \\
&\quad -\frac{\theta^2}{2}\sum_{z\in\mathcal{Z}} \frac{(P_{Z|X}(z|x_1)-P_{Z|X}(z|x_0))^2}{P_{Z|X}(z|x_0)} - \frac{\theta^3}{2}\sum_{z\in\mathcal{Z}} \frac{(P_{Z|X}(z|x_1)-P_{Z|X}(z|x_0))^3}{P_{Z|X}(z|x_0)^2} + o(\theta^2) \\
&\leqslant \frac{\theta^2}{2}\sum_{z\in\mathcal{Z}} \frac{(P_{Z|X}(z|x_1)-P_{Z|X}(z|x_0))^2}{P_{Z|X}(z|x_0)} + o(\theta^2) \\
&= \frac{\theta^2}{2}\chi_2(P_{Z|X=x_1}, P_{Z|X=x_0}) + o(\theta^2).
\end{aligned} \tag{B.5}
$$

Hence, it is sufficient to satisfy

$$
\theta \leqslant \sqrt{\frac{2\delta}{\lfloor \kappa n\rfloor \chi_2(P_{Z|X=x_1}, P_{Z|X=x_0})}} \tag{B.6}
$$

to ensure that the covertness criterion is verified. This completes the proof. ■

— C —

Proof of Proposition 2

This appendix proves Proposition 2.

Note that from [44, Lemma 19], it holds that

$$\lambda \leqslant \Pr\left[\log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})}{P_{\boldsymbol{Y}}(\boldsymbol{Y})}\right) \leqslant \eta\right] + \frac{M-1}{2}\Pr\left[\log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\bar{\boldsymbol{Y}}|\boldsymbol{X})}{P_{\boldsymbol{Y}}(\bar{\boldsymbol{Y}})}\right) > \eta\right], \tag{C.1}$$

where $P_{\boldsymbol{XY}}(\boldsymbol{x},\boldsymbol{y}) = P_{\boldsymbol{X}}(\boldsymbol{x})P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$ and $P_{\boldsymbol{X}\bar{\boldsymbol{Y}}}(\boldsymbol{x},\boldsymbol{y}) = P_{\boldsymbol{X}}(\boldsymbol{x})P_{\boldsymbol{Y}}(\boldsymbol{y})$.

Define the following moments:

$$\mu_t = \mathbb{E}_{XY}\left[\log\left(\frac{P_{Y|X}(Y_t|X_t)}{P_Y(Y_t)}\right)\right], \tag{C.2}$$

$$\sigma_t^2 = \mathbb{E}_{XY}\left[\log\left(\frac{P_{Y|X}(Y_t|X_t)}{P_Y(Y_t)}\right)^2 - \mu_t^2\right], \text{ and} \tag{C.3}$$

$$T_t = \mathbb{E}_{XY}\left[\left|\log\left(\frac{P_{Y|X}(Y_t|X_t)}{P_Y(Y_t)}\right) - \mu_t\right|^3\right]. \tag{C.4}$$

The next Lemma characterizes the three first moments defined above.

Lemma 9. *Given the input distribution $P_{\boldsymbol{X}}$ in* (3.20)*, it holds that*

$$\mu_t = \theta D\big(P_{Y|X=x_1}||P_{Y|X=x_0}\big) + O(\theta^2), \tag{C.5}$$

$$\sigma_t^2 = \theta\mathbb{E}_{Y|X=x_1}\left[\log\left(\frac{P_{Y|X}(Y|x_1)}{P_{Y|X}(Y|x_0)}\right)^2\right] + O(\theta^2), \text{ and} \tag{C.6}$$

$$T_t = \theta\mathbb{E}_{Y|X=x_1}\left[\left|\log\left(\frac{P_{Y|X}(Y|x_1)}{P_{Y|X}(Y|x_0)}\right)\right|^3\right] + O(\theta^2). \tag{C.7}$$

The proof of Lemma 9 is presented in Appendix D.

Define

$$\mu = \sum_{t=1}^{n} \mu_t, \tag{C.8}$$

$$\sigma^2 = \sum_{t=1}^{n} \sigma_t^2, \text{ and} \tag{C.9}$$

$$T = \sum_{t=1}^{n} T_t, \tag{C.10}$$

and note that

$$\Pr\left[\log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})}{P_{\boldsymbol{Y}}(\boldsymbol{Y})}\right) \leqslant \eta\right] = \Pr\left[\frac{1}{\sigma}\sum_{t=1}^{n} \log\left(\frac{P_{Y|X}(Y_t|X_t)}{P_Y(Y_t)}\right) - \mu_t \leqslant \frac{\eta-\mu}{\sigma}\right]. \tag{C.11}$$

Hence, from Berry-Esseen theorem (Theorem 21), it follows that

$$\left|\Pr\left[\log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})}{P_{\boldsymbol{Y}}(\boldsymbol{Y})}\right) \leqslant \eta\right] - Q\left(\frac{\eta-\mu}{\sigma}\right)\right| \leqslant \frac{6T}{\sigma^{\frac{3}{2}}}. \tag{C.12}$$

That is,

$$\Pr\left[\log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})}{P_{\boldsymbol{Y}}(\boldsymbol{Y})}\right) \leqslant \eta\right] \leqslant Q\left(\frac{\eta-\mu}{\sigma}\right) + \frac{6T}{\sigma^{\frac{3}{2}}}. \tag{C.13}$$

Note also that

$$\begin{aligned}\frac{6T}{\sigma^{\frac{3}{2}}} &= 6\frac{\sqrt{\omega n}\mathbb{E}_{Y|X=x_1}\left[\left|\log\left(\frac{P_{Y|X}(Y|x_1)}{P_{Y|X}(Y|x_0)}\right)\right|^3\right] + O(1)}{\left(\sqrt{\omega n}\mathbb{E}_{Y|X=x_1}\left[\log\left(\frac{P_{Y|X}(Y|x_1)}{P_{Y|X}(Y|x_0)}\right)^2\right] + O(1)\right)^{\frac{3}{2}}} \\ &= O(n^{-\frac{1}{4}}),\end{aligned} \tag{C.14}$$

with $\omega = \sqrt{\frac{2\delta}{\kappa\chi_2(P_{Z|X=x_1}, P_{Z|X=x_0})}}$.

Hence, it follows that

$$\Pr\left[\log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})}{P_{\boldsymbol{Y}}(\boldsymbol{Y})}\right) \leqslant \eta\right] \leqslant Q\left(\frac{\eta-\mu}{\sigma}\right) + O(n^{-\frac{1}{4}}). \tag{C.15}$$

In addition, for all $\boldsymbol{x} \in \mathcal{X}^n$ it holds that

$$\begin{aligned}\Pr\left[\log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\bar{\boldsymbol{Y}}|\boldsymbol{X})}{P_{\boldsymbol{Y}}(\bar{\boldsymbol{Y}})}\right) > \eta\right] &= \mathbb{E}_{\boldsymbol{Y}}\left[\mathbb{1}_{\left\{\log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{x})}{P_{\boldsymbol{Y}}(\boldsymbol{Y})}\right)>\eta\right\}}\right] \\ &= \sum_{\boldsymbol{y}\in\mathcal{Y}^n} P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}\mathbb{1}_{\left\{\log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\right)>\eta\right\}} \\ &= \sum_{\boldsymbol{y}\in\mathcal{Y}^n} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})\exp\left(-\log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\right)\right) \\ &\quad \cdot\mathbb{1}_{\left\{\log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\right)>\eta\right\}}\end{aligned}$$

$$= \mathbb{E}_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}\left[\exp\left(-\log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\right)\right)\mathbb{1}_{\left\{\log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\right)>\eta\right\}}\right]. \tag{C.16}$$

Hence, using [44, Lemma 47], it follows that

$$\begin{aligned}\Pr\left[\log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\bar{\boldsymbol{Y}}|\boldsymbol{X})}{P_{\boldsymbol{Y}}(\bar{\boldsymbol{Y}})}\right) > \eta\right] &\leqslant \frac{2}{\sigma\exp(\eta)}\left(\frac{\log(2)}{\sqrt{2\pi}} + \frac{12T}{\sigma^2}\right) \\ &= \frac{2}{\sigma\exp(\eta)}\left(\frac{\log(2)}{\sqrt{2\pi}} + O(1)\right).\end{aligned} \tag{C.17}$$

Thus, it holds that

$$\frac{M-1}{2}\Pr\left[\log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\bar{\boldsymbol{Y}}|\boldsymbol{X})}{P_{\boldsymbol{Y}}(\bar{\boldsymbol{Y}})}\right) > \eta\right] \leqslant \frac{M}{\sigma\exp(\eta)}\left(\frac{\log(2)}{\sqrt{2\pi}} + O(1)\right). \tag{C.18}$$

Therefore, it follows that

$$\begin{aligned}\lambda &\leqslant Q\left(\frac{\eta-\mu}{\sigma}\right) + O(n^{-\frac{1}{4}}) + \frac{M}{\sigma\exp(\eta)}\left(\frac{\log(2)}{\sqrt{2\pi}} + O(1)\right) \\ &\overset{(a)}{=} Q\left(\frac{\eta-\mu}{\sigma}\right) + O(n^{-\frac{1}{4}}) + \frac{1}{\sigma}\left(\frac{\log(2)}{\sqrt{2\pi}} + O(1)\right) \\ &= Q\left(\frac{\eta-\mu}{\sigma}\right) + O(n^{-\frac{1}{2}}),\end{aligned} \tag{C.19}$$

where *(a)* follows after choosing $\log(M) = \eta$.

Now, choosing

$$\begin{aligned}\eta &= \mu - \sigma Q^{-1}(\epsilon) \\ &= n\theta D\big(P_{Y|X=x_1}||P_{Y|X=x_0}\big) - \sqrt{n\theta\mathbb{E}_{Y|X=x_1}\left[\log\left(\frac{P_{Y|X}(Y|x_1)}{P_{Y|X}(Y|x_0)}\right)^2\right]}Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right),\end{aligned} \tag{C.20}$$

yields

$$\lambda \leqslant \epsilon + O(n^{-\frac{1}{2}}). \tag{C.21}$$

This proves the existence of a code whose rate is in (3.26). This completes the proof. ■

— D —

Proof of Lemma 9

This appendix presents the proof of Lemma 9.

Note that

$$
\begin{aligned}
\mu_t =& \mathbb{E}_{XY}\left[\log\left(\frac{P_{Y|X}(Y_t|X_t)}{P_Y(Y_t)}\right)\right] \\
=& \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{XY}(x,y)\log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right) \\
=& \sum_{y\in\mathcal{Y}}(1-\theta)P_{Y|X}(y|x_0)\log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right) + \theta P_{Y|X}(y|x_1)\log\left(\frac{P_{Y|X}(y|x_1)}{P_Y(y)}\right) \\
=& \sum_{y\in\mathcal{Y}}(1-\theta)P_{Y|X}(y|x_0)\log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right) + \theta P_{Y|X}(y|x_1)\log\left(\frac{P_{Y|X}(y|x_1)P_{Y|X}(y|x_0)}{P_Y(y)P_{Y|X}(y|x_0)}\right) \\
=& \sum_{y\in\mathcal{Y}}\theta P_{Y|X}(y|x_1)\log\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)}\right) + \Big((1-\theta)P_{Y|X}(y|x_0) + \theta P_{Y|X}(y|x_1)\Big) \\
& \cdot\log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right) \\
=& \theta D\Big(P_{Y|X=x_1}||P_{Y|X=x_0}\Big) - \sum_{y\in\mathcal{Y}} P_Y(y)\log\left(\frac{P_Y(y)}{P_{Y|X}(y|x_0)}\right) \\
=& \theta D\Big(P_{Y|X=x_1}||P_{Y|X=x_0}\Big) + \sum_{y\in\mathcal{Y}} P_Y(y)\log\left(1+\theta\frac{P_{Y|X}(y|x_1)-P_{Y|X}(y|x_0)}{P_{Y|X}(y|x_0)}\right) \\
=& \theta D\Big(P_{Y|X=x_1}||P_{Y|X=x_0}\Big) + \sum_{y\in\mathcal{Y}} P_Y(y)\left(\theta\frac{P_{Y|X}(y|x_1)-P_{Y|X}(y|x_0)}{P_{Y|X}(y|x_0)} + O(\theta^2)\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \theta D\big(P_{Y|X=x_1}||P_{Y|X=x_0}\big) + \sum_{y\in\mathcal{Y}} \theta(1-\theta) P_{Y|X}(y|x_0) \frac{P_{Y|X}(y|x_1) - P_{Y|X}(y|x_0)}{P_{Y|X}(y|x_0)} \\
&\quad + \theta^2 \frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)} \big(P_{Y|X}(y|x_1) - P_{Y|X}(y|x_0)\big) + O(\theta^2) \\
&= \theta D\big(P_{Y|X=x_1}||P_{Y|X=x_0}\big) + O(\theta^2). \qquad \text{(D.1)}
\end{aligned}
$$

Note also that

$$
\begin{aligned}
&\mathbb{E}_{XY}\left[\log\left(\frac{P_{Y|X}(Y_t|X_t)}{P_Y(Y_t)}\right)^2\right] \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{XY}(x,y) \log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right)^2 \\
&= \sum_{y\in\mathcal{Y}} (1-\theta) P_{Y|X}(y|x_0) \log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right)^2 + \theta P_{Y|X}(y|x_1) \log\left(\frac{P_{Y|X}(y|x_1)}{P_Y(y)}\right)^2 \\
&= \sum_{y\in\mathcal{Y}} (1-\theta) P_{Y|X}(y|x_0) \log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right)^2 + \theta P_{Y|X}(y|x_1) \log\left(\frac{P_{Y|X}(y|x_1)P_{Y|X}(y|x_0)}{P_Y(y)P_{Y|X}(y|x_0)}\right)^2 \\
&= \sum_{y\in\mathcal{Y}} (1-\theta) P_{Y|X}(y|x_0) \log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right)^2 + \theta P_{Y|X}(y|x_1) \\
&\quad \cdot \left(\log\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)}\right) + \log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right)\right)^2 \\
&= \sum_{y\in\mathcal{Y}} (1-\theta) P_{Y|X}(y|x_0) \log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right)^2 + \theta P_{Y|X}(y|x_1) \\
&\quad \cdot \left(\log\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)}\right)^2 + 2\log\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)}\right)\log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right) + \log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right)^2\right) \\
&= \sum_{y\in\mathcal{Y}} \theta P_{Y|X}(y|x_1) \log\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)}\right)^2 + 2\theta P_{Y|X}(y|x_1) \log\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)}\right)\log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right) \\
&\quad + P_Y(y) \log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right)^2 \\
&= \theta\mathbb{E}_{Y|X=x_1}\left[\log\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)}\right)^2\right] - \sum_{y\in\mathcal{Y}} 2\theta P_{Y|X}(y|x_1) \log\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)}\right)\log\left(\frac{P_Y(y)}{P_{Y|X}(y|x_0)}\right) \\
&\quad + P_Y(y) \log\left(\frac{P_Y(y)}{P_{Y|X}(y|x_0)}\right)^2 \\
&= \theta\mathbb{E}_{Y|X=x_1}\left[\log\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)}\right)^2\right] - \sum_{y\in\mathcal{Y}} 2\theta P_{Y|X}(y|x_1) \log\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)}\right) \\
&\quad \cdot \log\left(1 + \theta\frac{P_{Y|X}(y|x_1) - P_{Y|X}(y|x_0)}{P_{Y|X}(y|x_0)}\right) + P_Y(y)\log\left(1 + \theta\frac{P_{Y|X}(y|x_1) - P_{Y|X}(y|x_0)}{P_{Y|X}(y|x_0)}\right)^2
\end{aligned}
$$

$$
\begin{aligned}
&= \theta \mathbb{E}_{Y|X=x_1}\left[\log\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)}\right)^2\right] - \sum_{y\in\mathcal{Y}} 2\theta P_{Y|X}(y|x_1)\log\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)}\right) \\
&\quad \cdot \left(\theta\frac{P_{Y|X}(y|x_1)-P_{Y|X}(y|x_0)}{P_{Y|X}(y|x_0)} + O(\theta^2)\right) + P_Y(y)\left(\theta\frac{P_{Y|X}(y|x_1)-P_{Y|X}(y|x_0)}{P_{Y|X}(y|x_0)} + O(\theta^2)\right)^2 \\
&= \theta \mathbb{E}_{Y|X=x_1}\left[\log\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)}\right)^2\right] + O(\theta^2). \qquad \text{(D.2)}
\end{aligned}
$$

Therefore, it follows that

$$
\begin{aligned}
\sigma_t^2 &= \mathbb{E}_{XY}\left[\log\left(\frac{P_{Y|X}(Y_t|X_t)}{P_Y(Y_t)}\right)^2\right] - \mu_t^2 \\
&= \theta \mathbb{E}_{Y|X=x_1}\left[\log\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)}\right)^2\right] + O(\theta^2). \qquad \text{(D.3)}
\end{aligned}
$$

Finally, it also holds that

$$
\begin{aligned}
T_t &= \mathbb{E}_{XY}\left[\left|\log\left(\frac{P_{Y|X}(Y_t|X_t)}{P_Y(Y_t)}\right) - \mu_t\right|^3\right] \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{XY}(x,y)\left|\log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right) - \mu_t\right|^3 \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{XY}(x,y)\left|\log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right) - \mu_t\right|\left(\log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right) - \mu_t\right)^2 \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{XY}(x,y)\left|\log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right) - \mu_t\right| \\
&\quad \cdot \left(\log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right)^2 - 2\mu_t\log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right) + \mu_t^2\right) \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{XY}(x,y)\left(\left|\log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right)^3 - \mu_t\log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right)^2\right|\right. \\
&\quad \left. - \log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right)\left|2\mu_t\log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right) - 2\mu_t^2\right| + \left|\mu_t^2\log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right) - \mu_t^3\right|\right) \\
&= \sum_{y\in\mathcal{Y}}(1-\theta)P_{Y|X}(y|x_0)\left(\left|\log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right)^3 - \mu_t\log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right)^2\right|\right. \\
&\quad \left. - \log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right)\left|2\mu_t\log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right) - 2\mu_t^2\right| + \left|\mu_t^2\log\left(\frac{P_{Y|X}(y|x_0)}{P_Y(y)}\right) - \mu_t^3\right|\right) \\
&\quad + \theta P_{Y|X}(y|x_1)\left(\left|\log\left(\frac{P_{Y|X}(y|x_1)}{P_Y(y)}\right)^3 - \mu_t\log\left(\frac{P_{Y|X}(y|x_1)}{P_Y(y)}\right)^2\right|\right. \\
&\quad \left. - \log\left(\frac{P_{Y|X}(y|x_1)}{P_Y(y)}\right)\left|2\mu_t\log\left(\frac{P_{Y|X}(y|x_1)}{P_Y(y)}\right) - 2\mu_t^2\right| + \left|\mu_t^2\log\left(\frac{P_{Y|X}(y|x_1)}{P_Y(y)}\right) - \mu_t^3\right|\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{y\in\mathcal{Y}} (1-\theta) P_{Y|X}(y|x_0) O(\theta^3) + \theta P_{Y|X}(y|x_1) \left| \log \left(\frac{P_{Y|X}(y|x_1)}{P_Y(y)} \right)^3 \right| + O(\theta^2) \\
&= \sum_{y\in\mathcal{Y}} \theta P_{Y|X}(y|x_1) \left| \log \left(\frac{P_{Y|X}(y|x_1)}{P_Y(y)} \right)^3 \right| + O(\theta^2) \\
&= \sum_{y\in\mathcal{Y}} \theta P_{Y|X}(y|x_1) \left| \log \left(\frac{P_{Y|X}(y|x_1) P_{Y|X}(y|x_0)}{P_Y(y) P_{Y|X}(y|x_0)} \right)^3 \right| + O(\theta^2) \\
&= \sum_{y\in\mathcal{Y}} \theta P_{Y|X}(y|x_1) \left| \log \left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)} \right)^3 \right| + O(\theta^2). \qquad \text{(D.4)}
\end{aligned}
$$

This completes the proof. ■

— E —

Proof of Lemma 3

This appendix presents the proof of Lemma 3.

Note that from the triangle inequality, it follows that

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

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

— F —

Proof of Lemma 4

This appendix presents the 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 probability mass functions $Q_{\bar{W}\boldsymbol{Y}_2}$ and $S_{\bar{W}\boldsymbol{Y}_2}$ such that, for all pairs $(i, \boldsymbol{y}) \in \mathcal{W} \times \mathcal{Y}_2^n$,

$$Q_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) = Q_{\boldsymbol{Y}_2}(\boldsymbol{y}) Q_{\bar{W}|\boldsymbol{Y}_2}(i|\boldsymbol{y}), \tag{F.1}$$

$$\text{and} \quad S_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) = S_{\boldsymbol{Y}_2}(\boldsymbol{y}) S_{\bar{W}|\boldsymbol{Y}_2}(i|\boldsymbol{y}), \tag{F.2}$$

where $Q_{\boldsymbol{Y}_2}$ and $S_{\boldsymbol{Y}_2}$ are the marginal channel output probability mass functions, and

$$Q_{\bar{W}|\boldsymbol{Y}_2}(i|\boldsymbol{y}) = S_{\bar{W}|\boldsymbol{Y}_2}(i|\boldsymbol{y}) = \mathbb{1}_{\{\boldsymbol{y} \in \mathcal{D}_2(i)\}}. \tag{F.3}$$

Consider also the joint probability mass functions $Q_{W\boldsymbol{Y}_2}$ and $S_{W\boldsymbol{Y}_2}$ respectively in (4.42) and (4.43). Note that

$$\begin{aligned} \|S_{W\boldsymbol{Y}_2} - Q_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} = \frac{1}{2} \sum_{i=1}^{M} \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} \Big| & S_{W\boldsymbol{Y}_2}(i, \boldsymbol{y}) + S_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) - S_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) \\ & - Q_{W\boldsymbol{Y}_2}(i, \boldsymbol{y}) + Q_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) - Q_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) \Big| \\ \leqslant & \left\| S_{\bar{W}\boldsymbol{Y}_2} - Q_{\bar{W}\boldsymbol{Y}_2} \right\|_{\mathrm{TV}} + \left\| Q_{W\boldsymbol{Y}_2} - Q_{\bar{W}\boldsymbol{Y}_2} \right\|_{\mathrm{TV}} + \left\| S_{W\boldsymbol{Y}_2} - S_{\bar{W}\boldsymbol{Y}_2} \right\|_{\mathrm{TV}}, \end{aligned} \tag{F.4}$$

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 (L.4).

First, note that

$$\begin{aligned}\left\|S_{\bar{W}\boldsymbol{Y}_2}-Q_{\bar{W}\boldsymbol{Y}_2}\right\|_{\mathrm{TV}} &= \frac{1}{2M}\sum_{i=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2(i)\}}\left|S_{\boldsymbol{Y}_2}(\boldsymbol{y})-Q_{\boldsymbol{Y}_2}(\boldsymbol{y})\right| \\ &\overset{(a)}{=} \frac{1}{2}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\left|S_{\boldsymbol{Y}_2}(\boldsymbol{y})-Q_{\boldsymbol{Y}_2}(\boldsymbol{y})\right| \\ &= \left\|S_{\boldsymbol{Y}_2}-Q_{\boldsymbol{Y}_2}\right\|_{\mathrm{TV}}, \end{aligned} \tag{F.5}$$

where *(a)* holds since (4.4c) is assumed with equality.

Note also that

$$\begin{aligned}&\left\|Q_{W\boldsymbol{Y}_2}-Q_{\bar{W}\boldsymbol{Y}_2}\right\|_{\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}) \\ &\quad\cdot\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^{\mathrm{c}}(i)\}}\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}) \\ &\quad\cdot\left(1-\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2^{\mathrm{c}}(i)\}}\right)+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{F.6}$$

where *(a)* holds since (4.4c) holds with equality. Note that since (4.8c) is assumed with equality, the equality in (L.3) ensures that the same steps can be followed with the total variation $\left\|S_{W\boldsymbol{Y}_2}-S_{\bar{W}\boldsymbol{Y}_2}\right\|_{\mathrm{TV}}$. This yields

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

$$
\begin{aligned}
&= \frac{1}{2}\sum_{i=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n} S_{\boldsymbol{Y}_2}(\boldsymbol{y})\Big(\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2(i)\}}\big(1-S_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y})\big)+\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2^{\mathrm{c}}(i)\}}S_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y})\Big)\\
&= \frac{1}{2}\sum_{i=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\Big(S_{\boldsymbol{Y}_2}(\boldsymbol{y})\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2(i)\}}-S_{\boldsymbol{Y}_2}(\boldsymbol{y})S_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y})\\
&\quad\cdot\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2(i)\}}+S_{\boldsymbol{Y}_2}(\boldsymbol{y})S_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y})\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2^{\mathrm{c}}(i)\}}\Big)\\
&\overset{(a)}{=} \frac{1}{2}\left(1-1+2\sum_{i=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n} S_{\boldsymbol{Y}_2}(\boldsymbol{y})S_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y})\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2^{\mathrm{c}}(i)\}}\right)\\
&\leqslant \hat{\Lambda}_2, \qquad\qquad (\mathrm{F}.7)
\end{aligned}
$$

Plugging (L.5)–(L.7) into (L.4) completes the proof. ■

— G —

Proof of Proposition 3

This appendix presents the proof of Proposition G.

Let $S_{\boldsymbol{Y}_2}$ be a probability mass function such that, for all $(\boldsymbol{y}) \in \mathcal{Y}_2^n$,

$$S_{\boldsymbol{Y}_2}(\boldsymbol{y}) = \sum_{i=1}^{M} \frac{1}{M} \sum_{\hat{\boldsymbol{x}} \in \mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)) P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}}) \tag{G.1}$$

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

$$S_{W\boldsymbol{Y}_2}(i, \boldsymbol{y}) \triangleq \frac{1}{M} S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i), \tag{G.2}$$

with

$$\begin{aligned} S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i) &\triangleq \sum_{\hat{\boldsymbol{x}} \in \mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)) P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}}) \\ &= \sum_{\hat{\boldsymbol{x}} \in \mathcal{X}^n} \prod_{t=1}^{n} P_{\hat{X}|X}(\hat{x}_t|u_t(i)) P_{Y_2|X}(y_t|\hat{x}_t) \\ &= \prod_{t=1}^{n} \sum_{\hat{x} \in \mathcal{X}} P_{\hat{X}|X}(\hat{x}|u_t(i)) P_{Y_2|X}(y_t|\hat{x}). \end{aligned} \tag{G.3}$$

Note that from the triangle inequality, it follows that

$$\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} \leqslant \|Q_{\boldsymbol{Y}_2} - S_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} + \|S_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}}\,. \tag{G.4}$$

Note also that

$$\begin{aligned} \|S_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} &\overset{(a)}{\leqslant} \|S_{W\boldsymbol{Y}_2} - R_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} \\ &= \sum_{i=1}^{M} \frac{1}{M} \left\|S_{\boldsymbol{Y}_2|W=i} - R_{\boldsymbol{Y}_2|W=i}\right\|_{\mathrm{TV}} \end{aligned}$$

$$= \sum_{i=1}^{M} \frac{1}{M} \sqrt{D\big(S_{\boldsymbol{Y}_2|W=i} || R_{\boldsymbol{Y}_2|W=i}\big)}$$
$$\overset{(b)}{\leqslant} \sqrt{\sum_{i=1}^{M} \frac{1}{M} D\big(S_{\boldsymbol{Y}_2|W=i} || R_{\boldsymbol{Y}_2|W=i}\big)}, \tag{G.5}$$

where *(a)* follows from the triangle inequality, and *(b)* follows from Jensen's inequality.

It holds that the expectation over all the codebooks of $\sum_{i=1}^{M} \frac{1}{M} D\big(S_{\boldsymbol{Y}_2|W=i} || R_{\boldsymbol{Y}_2|W=i}\big)$ satisfies

$$\mathbb{E}_{\hat{\mathcal{C}}|\mathcal{C}} \left[\sum_{i=1}^{M} \frac{1}{M} D\big(S_{\boldsymbol{Y}_2|W=i} || R_{\boldsymbol{Y}_2|W=i}\big) \right]$$
$$= \mathbb{E}_{\hat{\mathcal{C}}|\mathcal{C}} \left[\sum_{i=1}^{M} \frac{1}{M} \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} R_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i) \log_2 \left(\frac{R_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)}{S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)} \right) \right]$$
$$= \mathbb{E}_{\hat{\mathcal{C}}|\mathcal{C}} \left[\sum_{i=1}^{M} \frac{1}{M} \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} \sum_{j=1}^{\hat{M}} \frac{1}{\hat{M}} P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{V}(i,j)) \log_2 \left(\frac{\sum_{k=1}^{\hat{M}} P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{V}(i,k))}{\hat{M} S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)} \right) \right]$$
$$= \sum_{i=1}^{M} \frac{1}{M} \sum_{\boldsymbol{v}_{i1} \in \mathcal{X}^n} \sum_{\boldsymbol{v}_{i2} \in \mathcal{X}^n} \dots \sum_{\boldsymbol{v}_{i\hat{M}} \in \mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{v}_{i1}|\boldsymbol{u}(i)) P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{v}_{i2}|\boldsymbol{u}(i)) \dots P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{v}_{i\hat{M}}|\boldsymbol{u}(i))$$
$$\cdot \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} \sum_{j=1}^{\hat{M}} \frac{1}{\hat{M}} P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}_{ij}) \log_2 \left(\frac{\sum_{k=1}^{\hat{M}} P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}_{ik})}{\hat{M} S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)} \right)$$
$$= \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \frac{1}{M\hat{M}} \sum_{\boldsymbol{v}_{ij} \in \mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{v}_{ij}|\boldsymbol{u}(i)) P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}_{ij}) \prod_{k \neq j} \sum_{\boldsymbol{v}_{ik} \in \mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{v}_{ik}|\boldsymbol{u}(i))$$
$$\cdot \log_2 \left(\frac{\sum_{l=1}^{\hat{M}} P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}_{il})}{\hat{M} S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)} \right)$$
$$\overset{(a)}{\leqslant} \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \frac{1}{M\hat{M}} \sum_{\boldsymbol{v}_{ij} \in \mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{v}_{ij}|\boldsymbol{u}(i)) P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}_{ij})$$
$$\cdot \log_2 \left(\prod_{k \neq j} \sum_{\boldsymbol{v}_{ik} \in \mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{v}_{ik}|\boldsymbol{u}(i)) \frac{\sum_{l=1}^{\hat{M}} P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}_{il})}{\hat{M} S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)} \right)$$

$$
\begin{aligned}
&= \sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\frac{1}{M\hat{M}}\sum_{\boldsymbol{v}_{ij}\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{v}_{ij}|\boldsymbol{u}(i))P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}_{ij}) \\
&\quad\cdot\log_2\left(\frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}_{ij})}{\hat{M}S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)}+\frac{\sum\limits_{l\neq j}\sum\limits_{\boldsymbol{v}_{il}\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{v}_{il}|\boldsymbol{u}(i))P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}_{il})}{\hat{M}S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)}\right) \\
&= \sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\frac{1}{M\hat{M}}\sum_{\boldsymbol{v}_{ij}\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{v}_{ij}|\boldsymbol{u}(i))P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}_{ij}) \\
&\quad\cdot\log_2\left(\frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}_{ij})}{\hat{M}S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)}+\frac{\sum\limits_{k\neq j}\sum\limits_{\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}})}{\hat{M}S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)}\right) \\
&= \sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\frac{1}{M\hat{M}}\sum_{\boldsymbol{v}_{ij}\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{v}_{ij}|\boldsymbol{u}(i))P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}_{ij})\log_2\left(\frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}_{ij})}{\hat{M}S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)}+\frac{\hat{M}-1}{\hat{M}}\right) \\
&\leqslant \sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\sum_{i=1}^{M}\frac{1}{M}\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}})\log_2\left(1+\frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{\hat{M}S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)}\right) \\
&= \sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\sum_{i=1}^{M}\frac{1}{M}\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}})\log_2\left(1+\frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{\hat{M}S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)}\right)\mathbb{1}_{\left\{\log_2\left(\frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}\right)<\tau\right\}} \\
&+ \sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\sum_{i=1}^{M}\frac{1}{M}\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}})\log_2\left(1+\frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{\hat{M}S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)}\right)\mathbb{1}_{\left\{\log_2\left(\frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}\right)\geqslant\tau\right\}} \\
&\leqslant \sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\sum_{i=1}^{M}\frac{1}{M}\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}})\log_2\left(1+\frac{2^\tau P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}{\hat{M}S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)}\right)\mathbb{1}_{\left\{\log_2\left(\frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}\right)<\tau\right\}} \\
&+ \sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\sum_{i=1}^{M}\frac{1}{M}\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}})\log_2\left(1+\frac{1}{S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)}\right)\mathbb{1}_{\left\{\log_2\left(\frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}\right)\geqslant\tau\right\}} \\
&\leqslant \sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\sum_{i=1}^{M}\frac{1}{M}\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}})\frac{2^\tau P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}{\hat{M}S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)}\mathbb{1}_{\left\{\log_2\left(\frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}\right)<\tau\right\}} \\
&+ \sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\sum_{i=1}^{M}\frac{1}{M}\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}})\log_2\left(\frac{2}{S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)}\right)\mathbb{1}_{\left\{\log_2\left(\frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}\right)\geqslant\tau\right\}} \\
&= \frac{2^\tau}{\hat{M}}+\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\sum_{i=1}^{M}\frac{1}{M}\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}})\log_2\left(\frac{2}{S_{\boldsymbol{Y}_2|W}(\boldsymbol{y}|i)}\right)\mathbb{1}_{\left\{\log_2\left(\frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}\right)\geqslant\tau\right\}} \\
&\overset{(b)}{\leqslant} \frac{2^\tau}{\hat{M}}+\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\sum_{i=1}^{M}\frac{1}{M}\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}})\log_2\left(\frac{2}{\mu_0^n}\right)\mathbb{1}_{\left\{\log_2\left(\frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}\right)\geqslant\tau\right\}} \\
&\leqslant \frac{2^\tau}{\hat{M}}+n\log_2\left(\frac{2}{\mu_0}\right)\Pr\left[\log_2\left(\frac{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{Y}_2|\hat{\boldsymbol{X}})}{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{Y}_2|\boldsymbol{u}(W))}\right)\geqslant\tau\right] \\
&= \frac{2^\tau}{\hat{M}}+n\log_2\left(\frac{2}{\mu_0}\right)\Pr\left[\sum_{t=1}^{n}\log_2\left(\frac{P_{Y_2|X}(Y_{2t}|\hat{X}_t)}{P_{Y_2|X}(Y_{2t}|u_t(W))}\right)\geqslant\tau\right], \qquad \text{(G.6)}
\end{aligned}
$$

where *(a)* follows from Jensen's inequality, and *(b)* follows with $\mu_0 = \min_{(x,y)\in\mathcal{X}\times\mathcal{Y}_2} \sum_{\hat{x}\in\mathcal{X}} P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x})$.

Let $Z_t = \log_2\left(\frac{P_{Y_2|X}(Y_{2t}|\hat{X}_t)}{P_{Y_2|X}(Y_{2t}|u_t(W))}\right)$ and let $z_m = \min_{t\in\{1,2,\dots,n\}} Z_t$, and $z_M = \max_{t\in\{1,2,\dots,n\}} Z_t$. Note that if $\hat{X} = u_t(W)$, then $\log_2\left(\frac{P_{Y_2|X}(Y_{2t}|\hat{X}_t)}{P_{Y_2|X}(Y_{2t}|u_t(W))}\right) = 0$. Otherwise, if $\hat{X} \neq u_t(W)$, then $\log_2\left(\frac{P_{Y_2|X}(Y_{2t}|\hat{X}_t)}{P_{Y_2|X}(Y_{2t}|u_t(W))}\right) \neq 0$.

Let L be a random variable defined as

$$L \triangleq \sum_{t=1}^{n} \mathbb{1}_{\left\{\hat{X}_t \neq u_t(W)\right\}}. \tag{G.7}$$

Define also

$$\bar{D}_2 \triangleq \sum_{x\in\mathcal{X}} \sum_{\hat{x}\in\mathcal{X}} \bar{P}_X(x) \tilde{P}_{\hat{X}|X}(\hat{x}|x) D\big(P_{Y_2|X=\hat{x}} || P_{Y_2|X=x}\big), \tag{G.8}$$

$$\tau = (1+\mu)(1+\nu) K\sqrt{n}\bar{D}_2, \quad \text{and} \tag{G.9}$$

$$\mathcal{D}_\mu \triangleq \left\{\ell \in \mathbb{N}^* : \left|\ell - K\sqrt{n}\right| < \mu K\sqrt{n}\right\}, \tag{G.10}$$

with $(\mu,\nu) \in]0,1[^2$ arbitrarily small.

Then, it follows that

$$\begin{aligned}
&\Pr\left[\sum_{t=1}^{n} \log_2\left(\frac{P_{Y_2|X}(Y_{2t}|\hat{X}_t)}{P_{Y_2|X}(Y_{2t}|u_t(W))}\right) \geqslant \tau\right] \\
&= \sum_{\ell\in\mathcal{D}_\mu} \Pr[L=\ell] \Pr\left[\sum_{t=1}^{n} \log_2\left(\frac{P_{Y_2|X}(Y_{2t}|\hat{X}_t)}{P_{Y_2|X}(Y_{2t}|u_t(W))}\right) \geqslant \tau | L=\ell\right] \\
&\quad + \sum_{\ell\notin\mathcal{D}_\mu} \Pr[L=\ell] \Pr\left[\sum_{t=1}^{n} \log_2\left(\frac{P_{Y_2|X}(Y_{2t}|\hat{X}_t)}{P_{Y_2|X}(Y_{2t}|u_t(W))}\right) \geqslant \tau | L=\ell\right] \\
&\leqslant \sum_{\ell\in\mathcal{D}_\mu} \Pr[L=\ell] \Pr\left[\sum_{t=1}^{\ell} \log_2\left(\frac{P_{Y_2|X}(Y_{2t}|\hat{X}_t)}{P_{Y_2|X}(Y_{2t}|u_t(W))}\right) \geqslant \tau | L=\ell\right] + \Pr[L\notin\mathcal{D}_\mu],
\end{aligned} \tag{G.11}$$

where the probability operator applies with respect to the probability mass function $P_{W\hat{\boldsymbol{X}}_{(1:L)}\boldsymbol{Y}_{2,(1:L)}|L}$ such that for all $(i,\ell,\hat{\boldsymbol{x}}_{(1:\ell)},\boldsymbol{y}_{(1:\ell)}) \in \mathcal{W}\times\mathcal{D}_\mu\times\mathcal{X}^\ell\times\mathcal{Y}_2^\ell$,

$$P_{W\hat{\boldsymbol{X}}_{(1:L)}\boldsymbol{Y}_{2,(1:L)}|L}(i,\hat{\boldsymbol{x}}_{(1:\ell)},\boldsymbol{y}_{(1:\ell)}|\ell) = \prod_{t=1}^{\ell} \tilde{P}_{\hat{X}|X}(\hat{x}_t|u_t(i)) P_{Y_2|X}(y_t|\hat{x}_t). \tag{G.12}$$

Note that

$$\begin{aligned}
\Pr[L\notin\mathcal{D}_\mu] &= \Pr\left[\left|L - K\sqrt{n}\right| \geqslant \mu K\sqrt{n}\right] \\
&\overset{(a)}{\leqslant} 2\exp\left(-\frac{\mu^2 K\sqrt{n}}{3}\right),
\end{aligned} \tag{G.13}$$

where *(a)* follows using a Chernoff bound (see [49, Corollary 5]).

It remains to bound the first term in (G.11). Note that

$$
\begin{aligned}
&\mathbb{E}_{W\hat{\boldsymbol{X}}_{(1:L)}\boldsymbol{Y}_{2,(1:L)}|L=\ell}\left[\sum_{t=1}^{\ell}\log_2\left(\frac{P_{Y_2|X}(Y_{2t}|\hat{X}_t)}{P_{Y_2|X}(Y_{2t}|u_t(W))}\right)\right]\\
&=\sum_{t=1}^{\ell}\sum_{\boldsymbol{y}_{(1:\ell)}\in\mathcal{Y}_2^{\ell}}\sum_{\hat{\boldsymbol{x}}_{(1:\ell)}\in\mathcal{X}^{\ell}}\sum_{i=1}^{M}\frac{1}{M}\prod_{k=1}^{\ell}\tilde{P}_{\hat{X}|X}(\hat{x}_k|u_k(i))P_{Y_2|X}(y_k|\hat{x}_k)\log_2\left(\frac{P_{Y_2|X}(y_t|\hat{x}_t)}{P_{Y_2|X}(y_t|u_t(i))}\right)\\
&=\sum_{t=1}^{\ell}\sum_{i=1}^{M}\frac{1}{M}\prod_{k=1}^{\ell}\sum_{y\in\mathcal{Y}_2}\sum_{\hat{x}\in\mathcal{X}}\tilde{P}_{\hat{X}|X}(\hat{x}|u_k(i))P_{Y_2|X}(y|\hat{x})\log_2\left(\frac{P_{Y_2|X}(y|\hat{x})}{P_{Y_2|X}(y|u_t(i))}\right)\\
&=\sum_{t=1}^{\ell}\sum_{i=1}^{M}\frac{1}{M}\sum_{y\in\mathcal{Y}_2}\sum_{\hat{x}\in\mathcal{X}}\tilde{P}_{\hat{X}|X}(\hat{x}|u_t(i))P_{Y_2|X}(y|\hat{x})\log_2\left(\frac{P_{Y_2|X}(y|\hat{x})}{P_{Y_2|X}(y|u_t(i))}\right)\\
&=\sum_{t=1}^{\ell}\sum_{i=1}^{M}\frac{1}{M}\sum_{x\in\mathcal{X}}\mathbb{1}_{\{x=u_t(i)\}}\sum_{y\in\mathcal{Y}_2}\sum_{\hat{x}\in\mathcal{X}}\tilde{P}_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})\log_2\left(\frac{P_{Y_2|X}(y|\hat{x})}{P_{Y_2|X}(y|x)}\right)\\
&=\sum_{t=1}^{\ell}\sum_{i=1}^{M}\frac{1}{M}\sum_{x\in\mathcal{X}}\mathbb{1}_{\{x=u_t(i)\}}\sum_{\hat{x}\in\mathcal{X}}\tilde{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_2|X=\hat{x}}||P_{Y_2|X=x}\big)\\
&=\ell\sum_{i=1}^{M}\sum_{t=1}^{\ell}\sum_{x\in\mathcal{X}}\frac{\mathbb{1}_{\{x=u_t(i)\}}}{\ell M}\sum_{\hat{x}\in\mathcal{X}}\tilde{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_2|X=\hat{x}}||P_{Y_2|X=x}\big)\\
&=\ell\sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}}\bar{P}_X(x)\tilde{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_2|X=\hat{x}}||P_{Y_2|X=x}\big)\\
&=\ell\bar{D}_2.
\end{aligned}
\tag{G.14}
$$

Note also that for $\ell\in\mathcal{D}_\mu$, it holds that

$$
(1+\mu)K\sqrt{n}>\ell>(1-\mu)K\sqrt{n}. \tag{G.15}
$$

Therefore, it follows that

$$
\begin{aligned}
\tau-\ell\bar{D}_2&=(1+\mu)(1+\nu)K\sqrt{n}\bar{D}_2-\ell\bar{D}_2\\
&>(1+\nu)\ell\bar{D}_2-\ell\bar{D}_2\\
&=\nu\ell\bar{D}_2.
\end{aligned}
\tag{G.16}
$$

Hence, it follows that

$$
\begin{aligned}
&\Pr\left[\sum_{t=1}^{\ell}\log_2\left(\frac{P_{Y_2|X}(Y_{2t}|\hat{X}_t)}{P_{Y_2|X}(Y_{2t}|u_t(W))}\right)\geqslant\tau|L=\ell\right]\\
&=\Pr\left[\sum_{t=1}^{\ell}\log_2\left(\frac{P_{Y_2|X}(Y_{2t}|\hat{X}_t)}{P_{Y_2|X}(Y_{2t}|u_t(W))}\right)-\ell\bar{D}_2\geqslant\tau-\ell\bar{D}_2|L=\ell\right]\\
&\leqslant\Pr\left[\sum_{t=1}^{\ell}\log_2\left(\frac{P_{Y_2|X}(Y_{2t}|\hat{X}_t)}{P_{Y_2|X}(Y_{2t}|u_t(W))}\right)-\ell\bar{D}_2\geqslant\nu\ell\bar{D}_2|L=\ell\right]
\end{aligned}
$$

$$\begin{aligned}
&\overset{(a)}{\leqslant} 2\exp\left(-\frac{2\ell^2\nu^2\bar{D}_2^2}{\ell(z_M-z_m)^2}\right)\\
&= 2\exp\left(-\frac{2\nu^2\bar{D}_2^2}{(z_M-z_m)^2}\ell\right)\\
&= 2\exp\left(-\frac{2\nu^2\bar{D}_2^2}{(z_M-z_m)^2}(1-\mu)K\sqrt{n}\right),
\end{aligned} \tag{G.17}$$

where *(a)* follows from Hoeffding's inequality.

Combining (G.11), (G.13) and (G.17), it follows that

$$\begin{aligned}
&\Pr\left[\sum_{t=1}^{\ell}\log_2\left(\frac{P_{Y_2|X}(Y_{2t}|\hat{X}_t)}{P_{Y_2|X}(Y_{2t}|u_t(W))}\right)\geqslant\tau\right]\\
&\leqslant 2\exp\left(-\frac{2\nu^2\bar{D}_2^2}{(z_M-z_m)^2}(1-\mu)K\sqrt{n}\right)+2\exp\left(-\frac{\mu^2K\sqrt{n}}{3}\right)\\
&\leqslant 2\exp(-a\sqrt{n}),
\end{aligned} \tag{G.18}$$

for some $a>0$.

Combining (**??**) and (G.18), it follows that

$$\mathbb{E}_{\hat{\mathcal{C}}|\mathcal{C}}\left[\sum_{i=1}^{M}\frac{1}{M}D\big(S_{\boldsymbol{Y}_2|W=i}||R_{\boldsymbol{Y}_2|W=i}\big)\right]\leqslant\frac{2^\tau}{\hat{M}}+2n\log_2\left(\frac{2}{\mu_0}\right)\exp\left(-a\sqrt{n}\right) \tag{G.19}$$

Thus, if $\log_2\left(\hat{M}\right)>\tau$, it follows that there exists a code for which

$$\|R_{\boldsymbol{Y}_2}-S_{\boldsymbol{Y}_2}\|_{\mathrm{TV}}\leqslant\sqrt{2^{-b\sqrt{n}}+2n\log_2\left(\frac{2}{\mu_0}\right)\exp\left(-a\sqrt{n}\right)}, \tag{G.20}$$

for some $b>0$. Let $c_n=2^{-b\sqrt{n}}+2n\log_2\left(\frac{2}{\mu_0}\right)\exp\left(-a\sqrt{n}\right)$. Then

$$\|R_{\boldsymbol{Y}_2}-S_{\boldsymbol{Y}_2}\|_{\mathrm{TV}}\leqslant\sqrt{c_n}. \tag{G.21}$$

Consider the probability mass function $Q_{W\boldsymbol{Y}_2}$ in (4.42). From Lemma 8 (in Appendix **??**), it follows that the total variation $\|S_{W\boldsymbol{Y}_2}-Q_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}}$ verifies

$$\begin{aligned}
&\|S_{W\boldsymbol{Y}_2}-Q_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}}\\
&\quad=\Pr\left[S_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2S})\geqslant Q_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2S})\right]-\Pr\left[S_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2Q})\geqslant Q_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2Q})\right],
\end{aligned} \tag{G.22}$$

where the first probability operator in the left hand-side of (G.22) applies assuming that $(W,\boldsymbol{Y}_{2S})$ follows the joint probability mass function $S_{W\boldsymbol{Y}_2}$; and the second applies assuming that $(W,\boldsymbol{Y}_{2Q})$ follows the joint probability mass function $Q_{W\boldsymbol{Y}_2}$.

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

$$B(x,y)\triangleq\log_2\left(1+\theta C(x,y)\right), \tag{G.23}$$

where

$$C(x,y) \triangleq \frac{\tilde{R}_{Y_2|X}(y|x) - P_{Y_2|X}(y|x)}{P_{Y_2|X}(y|x)}. \tag{G.24}$$

Then, note that

$$\begin{aligned}
&\Pr\left[S_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2S}) \geqslant Q_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2S})\right] \\
&= \Pr\left[\frac{S_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2S})}{Q_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2S})} \geqslant 1\right] \\
&= \Pr\left[\log_2\left(\frac{S_{\boldsymbol{Y}_2|W}(\boldsymbol{Y}_{2S}|W)}{Q_{\boldsymbol{Y}_2|W}(\boldsymbol{Y}_{2S}|W)}\right) \geqslant 0\right] \\
&= \Pr\left[\log_2\left(\frac{\sum\limits_{\hat{\boldsymbol{x}}\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(W))P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{Y}_{2S}|\hat{\boldsymbol{x}})}{P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{Y}_{2S}|\boldsymbol{u}(W))}\right) \geqslant 0\right] \\
&= \Pr\left[\sum_{t=1}^{n}\log_2\left(\frac{\sum\limits_{\hat{x}\in\mathcal{X}} P_{\hat{X}|X}(\hat{x}|u_t(W))P_{Y_2|X}(Y_{2S,t}|\hat{x})}{P_{Y_2|X}(Y_{2S,t}|u_t(W))}\right) \geqslant 0\right] \\
&= \Pr\left[\sum_{t=1}^{n}\log\left(\frac{(1-\theta)P_{Y_2|X}(Y_{2S,t}|u_t(W))}{P_{Y_2|X}(Y_{2S,t}|u_t(W))} + \frac{\theta\sum_{\hat{x}\in\mathcal{X}}\tilde{P}_{\hat{X}|X}(\hat{x}|u_t(W))P_{Y_2|X}(Y_{2S,t}|\hat{x})}{P_{Y_2|X}(Y_{2S,t}|u_t(W))}\right) \geqslant 0\right] \\
&= \Pr\left[\sum_{t=1}^{n}\log\left(1 + \frac{\theta\left(\tilde{R}_{Y_2|X}(Y_{2S,t}|u_t(W)) - P_{Y_2|X}(Y_{2S,t}|u_t(W))\right)}{P_{Y_2|X}(Y_{2S,t}|u_t(W))}\right) \geqslant 0\right] \\
&= \Pr\left[\sum_{t=1}^{n}\log_2\left(1+\theta C(u_t(W),Y_{2S,t})\right) \geqslant 0\right] \\
&= \Pr\left[\sum_{t=1}^{n} B(u_t(W),Y_{2S,t}) \geqslant 0\right].
\end{aligned} \tag{G.25}$$

Following similar steps, it can be shown that

$$\begin{aligned}
&\Pr\left[S_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2Q}) \geqslant Q_{W\boldsymbol{Y}_2}(W,\boldsymbol{Y}_{2Q})\right] \\
&= \Pr\left[\sum_{t=1}^{n} B(u_t(W),Y_{2Q,t}) \geqslant 0\right].
\end{aligned} \tag{G.26}$$

Plugging (G.25) and (G.26) into (G.22) yields

$$\|S_{W\boldsymbol{Y}_2} - Q_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} = \Pr\left[\sum_{t=1}^{n} B(u_t(W),Y_{2S,t}) \geqslant 0\right] = \Pr\left[\sum_{t=1}^{n} B(u_t(W),Y_{2Q,t}) \geqslant 0\right] \tag{G.27}$$

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 (G.27), respectively.
Consider the first term in the right hand-side of (G.27). 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

$$B(u_t(W),Y_{2S,t}) = \log_2\left(1+\theta C(u_t(W),Y_{2S,t})\right). \tag{G.28}$$

That is,

$$\hat{\mu}_t = \mathbb{E}_{WY_{2S,t}}\left[B(u_t(W), Y_{2S,t})\right], \tag{G.29}$$

$$\hat{\sigma}_t^2 = \mathbb{E}_{WY_{2S,t}}\left[B(u_t(W), Y_{2S,t})^2\right] - \hat{\mu}_t^2, \quad \text{and} \tag{G.30}$$

$$\hat{\phi}_t = \mathbb{E}_{WY_{2S,t}}\left[\left|B(u_t(W), Y_{2S,t}) - \hat{\mu}_t\right|^3\right]. \tag{G.31}$$

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

$$\hat{\mu} \triangleq \sum_{t=1}^{n} \hat{\mu}_t, \quad \hat{\sigma}^2 \triangleq \sum_{t=1}^{n} \hat{\sigma}_t^2, \text{ and } \hat{\phi} \triangleq \sum_{t=1}^{n} \hat{\phi}_t. \tag{G.32}$$

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

Lemma 10. *The terms $\hat{\mu}$, $\hat{\sigma}^2$ and $\hat{\phi}$ in* (G.32) *satisfy*

$$\hat{\mu} \geqslant \frac{K^2}{2} \sum_{x \in \mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}} |c_1|, \tag{G.33}$$

$$\hat{\sigma}^2 \leqslant K^2 \sum_{x \in \mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}} |c_2|, \tag{G.34}$$

$$\hat{\sigma}^2 \geqslant K^2 \sum_{x \in \mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}} |c_3|, \tag{G.35}$$

$$\hat{\phi} \leqslant \frac{K^3}{\sqrt{n}} c_4. \tag{G.36}$$

where c_1, c_2, c_3 and c_4 are constants that depend only on the random transformation in (4.1).

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

From Lemma 10, it follows that

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

$$\overset{(c)}{\geqslant} 1 - \frac{c_{14}}{\sqrt{n}} - Q\left(\frac{\frac{K^2}{2}\sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_1|}{\sqrt{K^2\sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_2|}}\right) \tag{G.37}$$

where *(a)* follows from the Berry-Esseen Theorem (Theorem 21); *(b)* follows from Lemma 10; and *(c)* follows with

$$c_{14} \triangleq \max c_0 c_4 \left(\sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2\big(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}\big) - \frac{K}{\sqrt{n}}|c_3|\right)^{-\frac{3}{2}}, \tag{G.38}$$

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

$$\sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2\big(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}\big) - \frac{K}{\sqrt{n}}|c_3| > 0. \tag{G.39}$$

Note that c_{14} depends only on the random transformation in (4.1). Consider the second term in the right hand-side of (G.27). For all $t \in \{1, 2, \ldots, n\}$, let μ_t, σ_t^2 and ϕ_t be the first, second and third absolute moments of the random variable

$$B(u_t(W), Y_{2Q,t}) = \log_2\left(1 + \theta C(u_t(W), Y_{2Q,t})\right). \tag{G.40}$$

That is,

$$\mu_t \triangleq \mathbb{E}_{WY_{2Q,t}}\left[B(u_t(W), Y_{2Q,t})\right], \tag{G.41}$$
$$\sigma_t^2 \triangleq \mathbb{E}_{WY_{2Q,t}}\left[B(u_t(W), Y_{2Q,t})^2\right] - \mu_t^2, \quad \text{and} \tag{G.42}$$
$$\phi_t \triangleq \mathbb{E}_{WY_{2Q,t}}\left[\left|B(u_t(W), Y_{2Q,t}) - \mu_t\right|^3\right]. \tag{G.43}$$

Using this notation, let μ, σ^2 and ϕ be

$$\mu \triangleq \sum_{t=1}^{n} \mu_t, \quad \sigma^2 \triangleq \sum_{t=1}^{n} \sigma_t^2, \text{ and } \phi \triangleq \sum_{t=1}^{n} \phi_t. \tag{G.44}$$

The following lemma characterizes μ, σ^2 and ϕ.

Lemma 11. *The terms μ, σ^2 and ϕ in* (G.44) *satisfy*

$$\mu \leqslant \frac{-K^2}{2}\sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_5|, \tag{G.45}$$

$$\sigma^2 \leqslant K^2\sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_6|, \tag{G.46}$$

$$\sigma^2 \geqslant K^2\sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_7|, \tag{G.47}$$

$$\phi \leqslant \frac{K^3}{\sqrt{n}}c_8, \tag{G.48}$$

where c_5, c_6, c_7 and c_8 are constants that depend only on the random transformation in (4.1).

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

From Lemma 11, it follows that

$$\begin{aligned}
&\Pr\left[\sum_{t=1}^{n} B(u_t(W), Y_{2Q,t}) \geqslant 0\right] \\
&= \Pr\left[\sum_{t=1}^{n} B(u_t(W), Y_{2Q,t}) - \mu \geqslant -\sigma\frac{\mu}{\sigma}\right] \\
&\overset{(a)}{\leqslant} Q\left(\frac{-\mu}{\sigma}\right) + c_0\frac{\phi}{\sigma^3} \\
&\overset{(b)}{\leqslant} Q\left(\frac{\frac{K^2}{2}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_5|}{\sqrt{K^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_7|}}\right) \\
&+ \frac{\frac{K^3}{\sqrt{n}}c_0c_8}{\sqrt{K^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_7|}^3} \\
&\overset{(c)}{\leqslant} Q\left(\frac{\frac{K^2}{2}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_5|}{\sqrt{K^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_7|}}\right) + \frac{c_{10}}{\sqrt{n}},
\end{aligned} \tag{G.49}$$

where *(a)* follows from the Berry-Esseen Theorem (Theorem 21); and *(b)* follows from Lemma 11; and *(c)* follows with

$$c_{10} \triangleq \max c_0c_8\left(\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2\big(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}\big) - \frac{K}{\sqrt{n}}|c_7|\right)^{-\frac{3}{2}}, \tag{G.50}$$

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

$$\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2\big(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}\big) - \frac{K}{\sqrt{n}}|c_7| > 0. \tag{G.51}$$

Note that c_{10} depends only on the random transformation in (4.1).

Combining (G.27), (G.49), and (G.37) yields

$$\begin{aligned}
&\|S_{W\boldsymbol{Y}_2} - Q_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} \geqslant 1 - \frac{c_{10}+c_{14}}{\sqrt{n}} \\
&-Q\left(\frac{\frac{K^2}{2}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_1|}{\sqrt{K^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_2|}}\right) \\
&-Q\left(\frac{\frac{K^2}{2}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_5|}{\sqrt{K^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_7|}}\right) \\
&\geqslant 1 - \frac{c_{10}+c_{14}}{\sqrt{n}} - 2Q\left(\frac{\frac{K^2}{2}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}|c_1|}{\sqrt{K^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_2|}}\right). \qquad \text{(G.52)}
\end{aligned}$$

Note that for all $(x,a,b) \in \mathbb{R}_+^3$ and $n \in \mathbb{N}$,

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

$$\leqslant Q\left(\frac{\sqrt{x}}{2}\right)+\frac{2a+b}{4\sqrt{2\pi n\left(x+\frac{b}{\sqrt{n}}\right)}}$$
$$\leqslant Q\left(\frac{\sqrt{x}}{2}\right)+\frac{2a+b}{4\sqrt{2\pi n\left(x+b\right)}}, \tag{G.53}$$

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 $0\leqslant y\leqslant x$.

To guarantee that the code is covert, it is sufficient to verify

$$\begin{aligned}
\delta &\geqslant \|Q_{\boldsymbol{Y}_2}-S_{\boldsymbol{Y}_2}\|_{\mathrm{TV}}+\sqrt{c_n}\\
&\overset{(a)}{\geqslant} \|S_{W\boldsymbol{Y}_2}-Q_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}}-\epsilon-\hat{\Lambda}_2+\sqrt{c_n}\\
&\overset{(b)}{\geqslant} 1-\frac{c_{10}+c_{14}}{\sqrt{n}}-\epsilon-\hat{\Lambda}_2+\sqrt{c_n}\\
&-2Q\left(\frac{\frac{K^2}{2}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})-\frac{K^3}{\sqrt{n}}|c_1|}{\sqrt{K^2\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})+\frac{K^3}{\sqrt{n}}|c_2|}}\right)\\
&\geqslant 1-\frac{c_{10}+c_{14}}{\sqrt{n}}-\epsilon-\hat{\Lambda}_2+\sqrt{c_n}-2Q\left(\frac{K}{2}\sqrt{\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2\left(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x}\right)}\right)\\
&-\frac{\frac{K^3}{\sqrt{n}}\left(4|c_1|+|c_2|\right)}{4\sqrt{2\pi n\left(\sum_{x\in\mathcal{X}}K^2\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})+\frac{K^3}{\sqrt{n}}|c_2|\right)}}\\
&=1-\frac{c_{10}+c_{14}}{\sqrt{n}}-\epsilon-\hat{\Lambda}_2+\sqrt{c_n}-2Q\left(\frac{K}{2}\sqrt{\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2\left(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x}\right)}\right)\\
&-\frac{K^2\left(4|c_1|+|c_2|\right)}{4n\sqrt{2\pi\left(\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})+\frac{K}{\sqrt{n}}|c_2|\right)}}\\
&\overset{(c)}{\geqslant} 1-\frac{c_{10}+c_{14}}{\sqrt{n}}-\frac{c_{15}}{n}-\epsilon-\hat{\Lambda}_2+\sqrt{c_n}-2Q\left(\frac{K}{2}\sqrt{\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2\left(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x}\right)}\right)\\
&\overset{(d)}{\geqslant} 1-\frac{c}{\sqrt{n}}-\epsilon-\hat{\Lambda}_2+\sqrt{c_n}-2Q\left(\frac{K}{2}\sqrt{\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2\left(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x}\right)}\right),
\end{aligned} \tag{G.54}$$

where *(a)* is due to Lemma 4; *(b)* is due to (G.52); *(c)* follows from (G.53), with

$$c_{15} \triangleq \frac{K^2\left(4\left|c_1\right|+\left|c_2\right|\right)}{4\sqrt{2\pi\left(\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})+\frac{K}{\sqrt{n}}\left|c_2\right|\right)}}; \tag{G.55}$$

and *(d)* follows with

$$c \triangleq c_{10}+c_{14}+c_{15}. \tag{G.56}$$

From (G.54), it follows that

$$\frac{K}{2}\sqrt{\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2\left(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x}\right)} \leqslant Q^{-1}\left(\frac{1-\delta-\epsilon-\hat{\Lambda}_2+\sqrt{c_n}-\frac{c}{\sqrt{n}}}{2}\right), \tag{G.57}$$

that is,

$$K \leqslant \frac{2Q^{-1}\left(\frac{1-\delta-\epsilon-\hat{\Lambda}_2+\sqrt{c_n}-\frac{c}{\sqrt{n}}}{2}\right)}{\sqrt{\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})}}. \tag{G.58}$$

This completes the proof. ■

— H —

Proof of Lemma 10

This appendix presents the proof of Lemma 10.

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

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

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

and

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

Thus, it follows that

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

$$
\begin{aligned}
&= n \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \log_2 \left(1+\theta C(x,y)\right) \\
&\quad +\theta(\tilde{R}_{Y_2|X}(y|x) - P_{Y_2|X}(y|x)) \log_2 \left(1+\theta C(x,y)\right)\Big) \\
&= n \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} C(x,y)^k \\
&\quad +\theta(\tilde{R}_{Y_2|X}(y|x) - P_{Y_2|X}(y|x)) \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} C(x,y)^k \\
&= n \sum_{x\in\mathcal{X}} \bar{P}_X(x) \sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \\
&\quad +n \sum_{x\in\mathcal{X}} \bar{P}_X(x) \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\theta^{k+1}}{k} \chi_{k+1}(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \\
&= n \sum_{x\in\mathcal{X}} \bar{P}_X(x) \sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \\
&\quad +n \sum_{x\in\mathcal{X}} \bar{P}_X(x) \sum_{k=2}^{\infty} \frac{(-1)^{k}\theta^k}{k-1} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \\
&= n \sum_{x\in\mathcal{X}} \bar{P}_X(x) \sum_{k=2}^{\infty} \frac{(-1)^{k}\theta^k}{k(k-1)} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \\
&= n\frac{\theta^2}{2} \sum_{x\in\mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 \sum_{x\in\mathcal{X}} \bar{P}_X(x) \sum_{k=3}^{\infty} \frac{(-1)^{k}\theta^{k-3}}{k(k-1)} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}).
\end{aligned}
\tag{H.4}
$$

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 (H.4) is also finite. Let c_1 be defined as

$$
c_1 \triangleq \min \sum_{x\in\mathcal{X}} \bar{P}_X(x) \sum_{k=3}^{\infty} \frac{(-1)^{k}\theta^{k-3}}{k(k-1)} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}),
\tag{H.5}
$$

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

$$
\begin{aligned}
\hat{\mu} &\geqslant n\frac{\theta^2}{2} \sum_{x\in\mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 c_1 \\
&\geqslant n\frac{\theta^2}{2} \sum_{x\in\mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - n\theta^3 \left|c_1\right| \\
&= \frac{K^2}{2} \sum_{x\in\mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}} \left|c_1\right|.
\end{aligned}
\tag{H.6}
$$

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) \\
&= 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^2 C(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} \\
&\quad \bar{P}_X(x) 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) \\
&= 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} \\
&\quad 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-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).
\end{aligned}
\tag{H.7}
$$

Note that since $B(x,y)$ is bounded, the upper-bound in (H.7) is finite. Thus, the terms in

(H.7) are also finite. Let c_2 be defined by

$$c_2 \triangleq \max \sum_{x\in\mathcal{X}} \bar{P}_X(x)\Bigg(\chi_3(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2} P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x}) \cdot\Bigg(2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-2}}{k}C(x,y)^{k+1} + \Bigg(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k}C(x,y)^{k}\Bigg)^2\Bigg), \tag{H.8}$$

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

$$\begin{aligned}\hat{\sigma}^2 &\leqslant n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 c_2\\ &\leqslant n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 \left|c_2\right|\\ &= K^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}\left|c_2\right|.\end{aligned} \tag{H.9}$$

It also holds that

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

$$
\begin{aligned}
&= 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(\left(P_{Y_2|X}(y|x)+\theta\Big(\tilde{R}_{Y_2|X}(y|x)-P_{Y_2|X}(y|x)\Big)\right)\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\\
&= 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)\left(\chi_3(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})\right.\\
&\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})\left(2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-2}}{k}C(x,y)^{k+1}\right.\\
&\quad \left.\left.+\left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k}C(x,y)^k\right)^2\right)\right) - 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)\left(\chi_3(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})\right.\\
&\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})\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 \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)\right). \qquad \text{(H.10)}
\end{aligned}
$$

Note that since $B(x,y)$ is bounded, $\hat{\sigma}^2$ is finite. Thus all the terms in (H.10) are also finite. Let c_3 be defined by

$$
\begin{aligned}
c_3 = \min &\sum_{x\in\mathcal{X}}\bar{P}_X(x)\left(\chi_3(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x}) + \sum_{\hat{x}\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}P_{\hat{X}|X}(\hat{x}|x)P_{Y_2|X}(y|\hat{x})\right.\\
&\cdot\left(2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-2}}{k}C(x,y)^{k+1}+\left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k}C(x,y)^k\right)^2\right.\\
&\left.\left.-\Big(\sum_{k=2}^{\infty}\frac{(-1)^{k}\theta^{k-\frac{3}{2}}}{k(k-1)}\chi_k(\tilde{R}_{Y_2|X=x},P_{Y_2|X=x})\Big)^2\right)\right), \qquad \text{(H.11)}
\end{aligned}
$$

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

probability mass functions $\tilde{P}_{\hat{X}|X}$. Plugging (H.11) into (H.10) yields

$$\begin{aligned}\hat{\sigma}^2 &\geqslant n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 c_3 \\ &\geqslant n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - n\theta^3 \left|c_3\right| \\ &= K^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}}\left|c_3\right|. \end{aligned} \tag{H.12}$$

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}) \left|\log_2\left(1+\theta C(x,y)\right)\right|^3 \qquad \text{(H.13)} \\ &= 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|\log_2\left(1+\theta C(x,y)\right)\right|^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}) \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_{\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-1}}{k}C(x,y)^k\right|^3. \qquad \text{(H.14)}\end{aligned}$$

Note that the upper-bound in (H.13) is finite since $B(x,y)$ is bounded. Thus, the expression

in (H.14) is also finite. Let c_4 be defined by

$$c_4 \triangleq \max \sum_{x\in\mathcal{X}} \sum_{\hat{x}\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \left| \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\theta^{k-1}}{k} C(x,y)^k \right|^3, \quad \text{(H.15)}$$

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

$$\begin{aligned} \hat{\phi} &\leqslant n\theta^3 c_4 \\ &= \frac{K^3}{\sqrt{n}} c_4. \end{aligned} \quad \text{(H.16)}$$

This completes the proof. ■

I

Proof of Lemma 11

This appendix presents the proof of Lemma 11.

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

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

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

and

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

Thus, it follows that

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

$$= -n\frac{\theta^2}{2}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x})$$
$$+n\theta^3\sum_{x\in\mathcal{X}}\bar{P}_X(x)\sum_{k=3}^{\infty}\frac{(-1)^{k+1}\theta^{k-3}}{k}\chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}). \tag{I.4}$$

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 (I.4) is also finite. Let c_5 be defined as

$$c_5 \triangleq \max\sum_{x\in\mathcal{X}}\bar{P}_X(x)\sum_{k=3}^{\infty}\frac{(-1)^{k}\theta^{k-3}}{k(k-1)}\chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}), \tag{I.5}$$

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

$$\mu \leqslant -n\frac{\theta^2}{2}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3\left|c_5\right|$$
$$= -\frac{K^2}{2}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}\left|c_5\right|. \tag{I.6}$$

Similarly, it holds that

$$\begin{aligned}
\sigma^2 &= \sum_{t=1}^{n}\sigma_t^2 \\
&= \frac{1}{M}\sum_{t=1}^{n}\sum_{i=1}^{M}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\mathbb{1}_{\{x=u_t(i)\}}P_{Y_2|X}(y|u_t(i))B(u_t(i),y)^2 - \mu_t^2 \\
&\leqslant \frac{1}{M}\sum_{t=1}^{n}\sum_{i=1}^{M}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\mathbb{1}_{\{x=u_t(i)\}}P_{Y_2|X}(y|x)B(x,y)^2 \qquad \text{(I.7)} \\
&= n\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{Y_2|X}(y|x)\log_2\left(1+\theta C(x,y)\right)^2 \\
&= n\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{Y_2|X}(y|x)\left(\sum_{k=1}^{\infty}\frac{(-1)^{k+1}\theta^k}{k}C(x,y)^k\right)^2 \\
&= n\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{Y_2|X}(y|x)\Bigg(\theta^2C(x,y)^2 + 2\theta C(x,y)\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^k}{k}C(x,y)^k \\
&\quad + \left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^k}{k}C(x,y)^k\right)^2\Bigg) \\
&= n\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2}\bar{P}_X(x)P_{Y_2|X}(y|x)\Bigg(\theta^2C(x,y)^2 + 2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k+1}}{k}C(x,y)^{k+1} \\
&\quad + \left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^k}{k}C(x,y)^k\right)^2\Bigg)
\end{aligned}$$

$$= n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2} \bar{P}_X(x)P_{Y_2|X}(y|x)$$
$$\cdot\left(2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k+1}}{k}C(x,y)^{k-2} + \left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k}C(x,y)^{k}\right)^2\right). \tag{I.8}$$

Note that since $B(x,y)$ is bounded, the upper-bound in (I.7) is finite. Hence, the terms in (I.8) are also finite. Let c_6 be defined by

$$c_6 \triangleq \max \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_2} \bar{P}_X(x)P_{Y_2|X}(y|x)\left(2\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k+1}}{k}C(x,y)^{k-2}\right.$$
$$\left. + \left(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k}C(x,y)^{k}\right)^2\right), \tag{I.9}$$

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

$$\begin{aligned}\sigma^2 &\leqslant n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 c_6\\ &\leqslant n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 |c_6|\\ &\leqslant K^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x)\chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \frac{K^3}{\sqrt{n}}|c_6|.\end{aligned} \tag{I.10}$$

On the other hand, it also holds that

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

$$
\begin{aligned}
&= n \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) B(x,y)^2 \\
&\quad -n \sum_{x\in\mathcal{X}} \bar{P}_X(x) \Bigg(\sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|x) \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} C(x,y)^k \Bigg)^2 \\
&= n \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) B(x,y)^2 - n \sum_{x\in\mathcal{X}} \bar{P}_X(x) \Bigg(\sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \Bigg)^2 \\
&= n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \\
&\quad \cdot \Bigg(2 \sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^{k-2}}{k} C(x,y)^{k+1} + \Bigg(\sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k} C(x,y)^k \Bigg)^2 \Bigg) \\
&\quad -n \sum_{x\in\mathcal{X}} \bar{P}_X(x) \Bigg(\sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^k}{k} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \Bigg)^2 \\
&= n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \\
&\quad \cdot \Bigg(2 \sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^{k-2}}{k} C(x,y)^{k+1} + \Bigg(\sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k} C(x,y)^k \Bigg)^2 \\
&\quad - \Bigg(\sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \Bigg)^2 \Bigg). \qquad \text{(I.11)}
\end{aligned}
$$

Note that since $B(x,y)$ is bounded, σ^2 is finite. Thus all the terms in (I.11) are also finite. Let c_7 be defined by

$$
\begin{aligned}
c_7 \triangleq \min \sum_{x\in\mathcal{X}} \bar{P}_X(x) \Bigg(& \chi_3(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + \sum_{\hat{x}\in\mathcal{X}} \sum_{y\in\mathcal{Y}_2} P_{\hat{X}|X}(\hat{x}|x) P_{Y_2|X}(y|\hat{x}) \\
& \cdot \Bigg(2 \sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^{k-2}}{k} C(x,y)^{k+1} + \Bigg(\sum_{k=2}^{\infty} \frac{(-1)^{k+1}\theta^{k-\frac{3}{2}}}{k} C(x,y)^k \Bigg)^2 \\
& - \Big(\sum_{k=2}^{\infty} \frac{(-1)^{k}\theta^{k-\frac{3}{2}}}{k(k-1)} \chi_k(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) \Big)^2 \Bigg), \qquad \text{(I.12)}
\end{aligned}
$$

where the minimization is over all possible values of $\theta \in (0,1)$ and all possible conditional probability mass functions $\tilde{P}_{\hat{X}|X}$. Plugging (I.12) into (I.11) yields

$$
\begin{aligned}
\sigma^2 &\geqslant n\theta^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) + n\theta^3 c_7 \\
&\geqslant K^2 \sum_{x\in\mathcal{X}} \bar{P}_X(x) \chi_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x}) - \frac{K^3}{\sqrt{n}} |c_7| . \qquad \text{(I.13)}
\end{aligned}
$$

Finally,

$$\begin{aligned}
\phi &= \sum_{t=1}^{n} \phi_t \\
&= \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \mathbb{1}_{\{x=u_t(i)\}} P_{Y_2|X}(y|x) |B(x,y) - \mu_t|^3 \\
&= \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \mathbb{1}_{\{x=u_t(i)\}} P_{Y_2|X}(y|x) \Bigg| B(x,y) - \frac{1}{M} \\
&\quad \cdot \sum_{i'=1}^{M} \sum_{y' \in \mathcal{Y}_2} P_{Y_2|X}(y'|u_t(i')) \log_2 \left(1 + \theta C(u_t(i'), y')\right) \Bigg|^3 \\
&\leqslant \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \mathbb{1}_{\{x=u_t(i)\}} P_{Y_2|X}(y|x) \Bigg| B(x,y) - \\
&\quad \cdot \log_2 \left(1 + \theta \frac{1}{M} \sum_{i'=1}^{M} \sum_{y' \in \mathcal{Y}_2} P_{Y_2|X}(y'|u_t(i')) C(u_t(i'), y')\right) \Bigg|^3 \\
&\leqslant \frac{1}{M} \sum_{t=1}^{n} \sum_{i=1}^{M} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \mathbb{1}_{\{x=u_t(i)\}} P_{Y_2|X}(y|x) |B(x,y)|^3 \qquad \text{(I.14)} \\
&= n \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) |\log_2 (1 + \theta C(x,y))|^3 \\
&= n \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \left| \sum_{k=1}^{\infty} \frac{(-1)^{k+1} \theta^k}{k} C(x,y)^k \right|^3 \\
&= n\theta^3 \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \left| \sum_{k=1}^{\infty} \frac{(-1)^{k+1} \theta^{k-1}}{k} C(x,y)^k \right|^3 . \qquad \text{(I.15)}
\end{aligned}$$

Note that the upper-bound in (I.14) is finite since $B(x,y)$ is bounded. Thus, the expression in (I.15) is also finite. Let c_8 be defined by

$$c_8 \triangleq \max \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}_2} \bar{P}_X(x) P_{Y_2|X}(y|x) \left| \sum_{k=1}^{\infty} \frac{(-1)^{k+1} \theta^{k-1}}{k} C(x,y)^k \right|^3 , \qquad \text{(I.16)}$$

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

$$\begin{aligned}
\phi &\leqslant n\theta^3 c_8 \\
&= \frac{K^3}{\sqrt{n}} c_8 . \qquad \text{(I.17)}
\end{aligned}$$

This completes the proof. ■

— J —

Proof of Proposition 4

This appendix presents the proof of Proposition 4. The proof consists in two steps. First, an upper-bound on the average error probability at each receiver is established. Given these upper-bounds, the second step consists in determining the maximum number of messages that can be transmitted while keeping the error probabilities upper-bounded.

Step 1:

Consider $\hat{\lambda}_1$ and $\hat{\lambda}_2$ respectively given in (4.8d) and (4.8e). Denote their expected value over the possible codebooks generated by $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$, respectively.

From (4.8e), it follows that

$$\begin{aligned}
\hat{\Lambda}_2 &= \sum_{i=1}^{M}\sum_{j=1}^{\hat{M}} \frac{1}{M\hat{M}} \sum_{\boldsymbol{v}_{11}\in\mathcal{X}^n} \sum_{\boldsymbol{v}_{12}\in\mathcal{X}^n} \dots \sum_{\boldsymbol{v}_{M\hat{M}}\in\mathcal{X}^n} \prod_{i'=1}^{M}\prod_{j'=1}^{\hat{M}} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{v}_{i'j'}|\boldsymbol{u}(i'))\Pr\left[\boldsymbol{Y}_2 \in \mathcal{D}_2^{\mathsf{c}}(i)|\hat{\boldsymbol{X}}=\boldsymbol{v}_{ij}\right] \\
&= \sum_{i=1}^{M}\sum_{j=1}^{\hat{M}} \frac{1}{M\hat{M}} \prod_{i'=1}^{M}\prod_{j'=1}^{\hat{M}} \sum_{\boldsymbol{v}_{i'j'}\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{v}_{i'j'}|\boldsymbol{u}(i'))\Pr\left[\boldsymbol{Y}_2 \in \mathcal{D}_2^{\mathsf{c}}(i)|\hat{\boldsymbol{X}}=\boldsymbol{v}_{ij}\right] \\
&= \sum_{i=1}^{M}\sum_{j=1}^{\hat{M}} \frac{1}{M\hat{M}} \sum_{\boldsymbol{v}_{ij}\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{v}_{ij}|\boldsymbol{u}(i))\Pr\left[\boldsymbol{Y}_2 \in \mathcal{D}_2^{\mathsf{c}}(i)|\hat{\boldsymbol{X}}=\boldsymbol{v}_{ij}\right] \\
&\quad \cdot \prod_{(i',j')\neq(i,j)} \sum_{\boldsymbol{v}_{i'j'}\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{v}_{i'j'}|\boldsymbol{u}(i')) \\
&= \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)\}}
\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}_{\{\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) \cdots \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))} \\
&= \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} \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}_{\{\boldsymbol{y}\in\mathcal{D}_2^{\mathsf{c}}(i)\}} \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 \text{(J.1)}
\end{aligned}
$$

From (4.8d), it follows that

$$
\begin{aligned}
\hat{\lambda}_1 &= \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \frac{1}{M\hat{M}} \Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i,j) | \hat{\boldsymbol{X}} = \boldsymbol{v}(i,j) \right] \\
&= \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \frac{1}{M\hat{M}} \Big(\Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i) | \hat{\boldsymbol{X}} = \boldsymbol{v}(i,j) \right] \Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i,j) | \hat{\boldsymbol{X}} = \boldsymbol{v}(i,j), \boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i) \right] \\
&\quad + \Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i) | \hat{\boldsymbol{X}} = \boldsymbol{v}(i,j) \right] \Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i,j) | \hat{\boldsymbol{X}} = \boldsymbol{v}(i,j), \boldsymbol{Y}_1 \in \mathcal{D}_1(i) \right] \Big) \\
&\leqslant \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \frac{1}{M\hat{M}} \Big(\Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i) | \hat{\boldsymbol{X}} = \boldsymbol{v}(i,j) \right] + \Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i,j) | \hat{\boldsymbol{X}} = \boldsymbol{v}(i,j), \boldsymbol{Y}_1 \in \mathcal{D}_1(i) \right] \Big). \qquad \text{(J.2)}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\hat{\Lambda}_1 \leqslant\ & \mathbb{E}_{\hat{\mathcal{C}}|\mathcal{C}} \left[\sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \frac{1}{M\hat{M}} \Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i) | \hat{\boldsymbol{X}} = \boldsymbol{v}(i,j) \right] \right] \\
& + \mathbb{E}_{\hat{\mathcal{C}}|\mathcal{C}} \left[\sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \frac{1}{M\hat{M}} \Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i,j) | \hat{\boldsymbol{X}} = \boldsymbol{v}(i,j), \boldsymbol{Y}_1 \in \mathcal{D}_1(i) \right] \right]
\end{aligned}
$$

$$
\begin{aligned}
&\leqslant \mathbb{E}_{\hat{\mathcal{C}}|\mathcal{C}}\left[\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\frac{1}{M\hat{M}}\Pr\left[\boldsymbol{Y}_1\in\mathcal{D}_1^{\mathsf{c}}(i)|\hat{\boldsymbol{X}}=\boldsymbol{v}(i,j)\right]\right]\\
&\quad+\mathbb{E}_{\hat{\mathcal{C}}|\mathcal{C}}\left[\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\frac{1}{M\hat{M}}\Pr\left[\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{v}(i,j))}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(i))}\right)<\eta|\hat{\boldsymbol{X}}=\boldsymbol{v}(i,j),\boldsymbol{Y}_1\in\mathcal{D}_1(i)\right]\right]\\
&\quad+\mathbb{E}_{\hat{\mathcal{C}}|\mathcal{C}}\left[\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\frac{1}{M\hat{M}}\Pr\left[\exists k\neq j:\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{v}(i,k))}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(i))}\right)\geqslant\eta|\hat{\boldsymbol{X}}=\boldsymbol{v}(i,j),\boldsymbol{Y}_1\in\mathcal{D}_1(i)\right]\right].
\end{aligned}
\tag{J.3}
$$

Note that

$$
\begin{aligned}
&\mathbb{E}_{\hat{\mathcal{C}}|\mathcal{C}}\left[\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\frac{1}{M\hat{M}}\Pr\left[\boldsymbol{Y}_1\in\mathcal{D}_1^{\mathsf{c}}(i)|\hat{\boldsymbol{X}}=\boldsymbol{v}(i,j)\right]\right]\\
&=\sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n}\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\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)|\hat{\boldsymbol{X}}=\hat{\boldsymbol{x}}\right]\\
&=\sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n}\sum_{\boldsymbol{y}\in\mathcal{Y}_1^n}\sum_{i=1}^{M}\frac{P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i))}{M}P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_1^{\mathsf{c}}(i)\}}\\
&=\sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n}\sum_{\boldsymbol{y}\in\mathcal{Y}_1^n}\sum_{i=1}^{M}\frac{P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i))}{M}P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_1^{\mathsf{c}}(i)\}}\\
&=\sum_{\boldsymbol{y}\in\mathcal{Y}_1^n}\sum_{i=1}^{M}\frac{1}{M}P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_1^{\mathsf{c}}(i)\}}\sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n}\prod_{t=1}^{n}P_{\hat{X}|X}(\hat{x}_t|u_t(i))\frac{P_{Y_1|X}(y_t|\hat{x}_t)}{P_{Y_1|X}(y_t|u_t(i))}\\
&=\sum_{\boldsymbol{y}\in\mathcal{Y}_1^n}\sum_{i=1}^{M}\frac{1}{M}P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_1^{\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_1|X}(y_t|\hat{x})}{P_{Y_1|X}(y_t|u_t(i))}\\
&=\sum_{\boldsymbol{y}\in\mathcal{Y}_1^n}\sum_{i=1}^{M}\frac{\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_1^{\mathsf{c}}(i)\}}}{M}P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(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}|x)\right)\frac{P_{Y_1|X}(y_t|\hat{x})}{P_{Y_1|X}(y_t|u_t(i))}\\
&=\sum_{\boldsymbol{y}\in\mathcal{Y}_1^n}\sum_{i=1}^{M}\frac{\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_1^{\mathsf{c}}(i)\}}}{M}P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))\prod_{t=1}^{n}\left(1+\theta\frac{\tilde{R}_{Y_1|X}(y_t|u_t(i))-P_{Y_1|X}(y_t|u_t(i))}{P_{Y_1|X}(y_t|u_t(i))}\right)\\
&\leqslant\sum_{\boldsymbol{y}\in\mathcal{Y}_1^n}\sum_{i=1}^{M}\frac{\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_1^{\mathsf{c}}(i)\}}}{M}P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))\prod_{t=1}^{n}\left(1+\theta\max_{(x,y)\in\mathcal{X}\times\mathcal{Y}_1}\frac{\tilde{R}_{Y_1|X}(y|x)-P_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)}\right)\\
&\leqslant\epsilon\left(1+\theta\max_{(x,y)\in\mathcal{X}\times\mathcal{Y}_1}\frac{\tilde{R}_{Y_1|X}(y|x)-P_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)}\right)^n.
\end{aligned}
\tag{J.4}
$$

It holds that

$$
\begin{aligned}
&\mathbb{E}_{\hat{\mathcal{C}}|\mathcal{C}}\left[\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}}\frac{1}{M\hat{M}}\Pr\left[\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{v}(i,j))}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(i))}\right)<\eta|\hat{\boldsymbol{X}}=\boldsymbol{v}(i,j),\boldsymbol{Y}_1\in\mathcal{D}_1(i)\right]\right]\\
&=\sum_{i=1}^{M}\sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n}\frac{1}{M}P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i))\Pr\left[\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\hat{\boldsymbol{x}})}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(i))}\right)<\eta|\hat{\boldsymbol{X}}=\hat{\boldsymbol{x}},\boldsymbol{Y}_1\in\mathcal{D}_1(i)\right]
\end{aligned}
$$

$$\leqslant \sum_{i=1}^{M} \sum_{\hat{\boldsymbol{x}} \in \mathcal{X}^n} \frac{1}{M} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)) \Pr\left[\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\hat{\boldsymbol{x}})}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(i))}\right) < \eta | \hat{\boldsymbol{X}} = \hat{\boldsymbol{x}}\right]$$
$$= \Pr\left[\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\hat{\boldsymbol{X}})}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(W))}\right) < \eta\right]. \tag{J.5}$$

It also holds that

$$\mathbb{E}_{\hat{\mathcal{C}}|\mathcal{C}}\left[\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}} \frac{1}{M\hat{M}} \Pr\left[\exists k \neq j : \log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{v}(i,k))}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(i))}\right) \geqslant \eta | \hat{\boldsymbol{X}} = \boldsymbol{v}(i,j), \boldsymbol{Y}_1 \in \mathcal{D}_1(i)\right]\right]$$
$$\leqslant \mathbb{E}_{\hat{\mathcal{C}}|\mathcal{C}}\left[\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}} \frac{1}{M\hat{M}} \sum_{k\neq j} \Pr\left[\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{v}(i,k))}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(i))}\right) \geqslant \eta | \hat{\boldsymbol{X}} = \boldsymbol{v}(i,j), \boldsymbol{Y}_1 \in \mathcal{D}_1(i)\right]\right]$$
$$\leqslant \mathbb{E}_{\hat{\mathcal{C}}|\mathcal{C}}\left[\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}} \frac{1}{M\hat{M}} \sum_{k\neq j} \Pr\left[\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{v}(i,k))}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(i))}\right) \geqslant \eta | \hat{\boldsymbol{X}} = \boldsymbol{v}(i,j)\right]\right]$$
$$\leqslant \mathbb{E}_{\hat{\mathcal{C}}|\mathcal{C}}\left[\sum_{i=1}^{M}\sum_{j=1}^{\hat{M}} \frac{1}{M\hat{M}} \sum_{k=1}^{\hat{M}} \Pr\left[\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{v}(i,k))}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(i))}\right) \geqslant \eta | \hat{\boldsymbol{X}} = \boldsymbol{v}(i,j)\right]\right]$$
$$= \sum_{k=1}^{\hat{M}}\sum_{i=1}^{M} \frac{1}{M} \sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n} \sum_{\hat{\boldsymbol{x}}'\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)) P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}'|\boldsymbol{u}(i)) \Pr\left[\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\hat{\boldsymbol{x}}')}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(i))}\right) \geqslant \eta | \hat{\boldsymbol{X}} = \hat{\boldsymbol{x}}\right]$$
$$= \hat{M}\sum_{i=1}^{M} \frac{1}{M} \sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n} \sum_{\hat{\boldsymbol{x}}'\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)) P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}'|\boldsymbol{u}(i)) \sum_{\boldsymbol{y}\in\mathcal{Y}_1^n} P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}}) \mathbb{1}_{\left\{\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\hat{\boldsymbol{x}}')}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(i))}\right)\geqslant\eta\right\}}$$
$$= \sum_{i=1}^{M} \frac{\hat{M}}{M} \sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n} \sum_{\hat{\boldsymbol{x}}'\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)) P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}'|\boldsymbol{u}(i)) \sum_{\boldsymbol{y}\in\mathcal{Y}_1^n} P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}}) \frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}$$
$$\cdot \mathbb{1}_{\left\{\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\hat{\boldsymbol{x}}')}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(i))}\right)\geqslant\eta\right\}}$$
$$\leqslant \sum_{i=1}^{M} \frac{\hat{M}}{M} \sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n} \sum_{\hat{\boldsymbol{x}}'\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)) P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}'|\boldsymbol{u}(i)) \sum_{\boldsymbol{y}\in\mathcal{Y}_1^n} P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}}) \frac{2^{-\eta} P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}}')}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}$$
$$\cdot \mathbb{1}_{\left\{\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\hat{\boldsymbol{x}}')}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(i))}\right)\geqslant\eta\right\}}$$
$$\leqslant \hat{M}2^{-\eta}\sum_{i=1}^{M} \frac{1}{M} \sum_{\boldsymbol{y}\in\mathcal{Y}_1^n} \sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n} \sum_{\hat{\boldsymbol{x}}'\in\mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)) P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}}) \frac{P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\hat{\boldsymbol{x}}'|\boldsymbol{u}(i)) P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}}')}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))}$$
$$= \hat{M}2^{-\eta}\sum_{i=1}^{M} \frac{1}{M} \sum_{\boldsymbol{y}\in\mathcal{Y}_1^n} \sum_{\hat{\boldsymbol{x}}\in\mathcal{X}^n} \sum_{\hat{\boldsymbol{x}}'\in\mathcal{X}^n} \prod_{t=1}^{n} P_{\hat{X}|X}(\hat{x}_t|u_t(i)) P_{Y_1|X}(y_t|\hat{x}_t) \frac{P_{\hat{X}|X}(\hat{x}'_t|u_t(i)) P_{Y_1|X}(y_t|\hat{x}'_t)}{P_{Y_1|X}(y_t|u_t(i))}$$
$$= \hat{M}2^{-\eta}\sum_{i=1}^{M} \frac{1}{M} \prod_{t=1}^{n} \sum_{y\in\mathcal{Y}_1} \sum_{\hat{x}\in\mathcal{X}} \sum_{\hat{x}'\in\mathcal{X}} P_{\hat{X}|X}(\hat{x}|u_t(i)) P_{Y_1|X}(y|\hat{x}) \frac{P_{\hat{X}|X}(\hat{x}'|u_t(i)) P_{Y_1|X}(y|\hat{x}')}{P_{Y_1|X}(y|u_t(i))}$$
$$= \hat{M}2^{-\eta}\sum_{i=1}^{M} \frac{1}{M} \prod_{t=1}^{n} \sum_{y\in\mathcal{Y}_1} \Big((1-\theta) P_{Y_1|X}(y|u_t(i)) + \theta \tilde{R}_{Y_1|X}(y|u_t(i))\Big)$$
$$\cdot \frac{(1-\theta) P_{Y_1|X}(y|u_t(i)) + \theta \tilde{R}_{Y_1|X}(y|u_t(i))}{P_{Y_1|X}(y|u_t(i))}$$

$$
\begin{aligned}
&= \hat{M}2^{-\eta}\sum_{i=1}^{M}\frac{1}{M}\prod_{t=1}^{n}\sum_{y\in\mathcal{Y}_1}\left((1-\theta)P_{Y_1|X}(y|u_t(i))+\theta\tilde{R}_{Y_1|X}(y|u_t(i))\right)\left(1-\theta+\theta\frac{\tilde{R}_{Y_1|X}(y|u_t(i))}{P_{Y_1|X}(y|u_t(i))}\right)\\
&= \hat{M}2^{-\eta}\sum_{i=1}^{M}\frac{1}{M}\prod_{t=1}^{n}1-\theta+\theta(1-\theta)\sum_{y\in\mathcal{Y}_1}\tilde{R}_{Y_1|X}(y|u_t(i))+\theta^2\sum_{y\in\mathcal{Y}_1}\frac{\tilde{R}_{Y_1|X}(y|u_t(i))^2}{P_{Y_1|X}(y|u_t(i))}\\
&= \hat{M}2^{-\eta}\sum_{i=1}^{M}\frac{1}{M}\prod_{t=1}^{n}1-\theta^2+\theta^2\sum_{y\in\mathcal{Y}_1}\frac{\tilde{R}_{Y_1|X}(y|u_t(i))^2}{P_{Y_1|X}(y|u_t(i))}\\
&\leqslant \hat{M}2^{-\eta}\sum_{i=1}^{M}\frac{1}{M}\prod_{t=1}^{n}1-\theta^2+\theta^2\max_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\frac{\tilde{R}_{Y_1|X}(y|x)^2}{P_{Y_1|X}(y|x)}\\
&= \hat{M}2^{-\eta}\sum_{i=1}^{M}\frac{1}{M}\left(1-\theta^2+\theta^2\max_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\frac{\tilde{R}_{Y_1|X}(y|x)^2}{P_{Y_1|X}(y|x)}\right)^n\\
&= \hat{M}2^{-\eta}\left(1+\theta^2\left(\max_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\frac{\tilde{R}_{Y_1|X}(y|x)^2}{P_{Y_1|X}(y|x)}-1\right)\right)^n\\
&= \hat{M}2^{-\eta}\exp\left(n\log\left(1+\theta^2\left(\max_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\frac{\tilde{R}_{Y_1|X}(y|x)^2}{P_{Y_1|X}(y|x)}-1\right)\right)\right)\\
&\leqslant \hat{M}2^{-\eta}\exp\left(n\theta^2\left(\max_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\frac{\tilde{R}_{Y_1|X}(y|x)^2}{P_{Y_1|X}(y|x)}-1\right)\right)\\
&\leqslant \hat{M}2^{-\eta}\exp\left(K^2\left(\max_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\frac{\tilde{R}_{Y_1|X}(y|x)^2}{P_{Y_1|X}(y|x)}-1\right)\right). \qquad \text{(J.6)}
\end{aligned}
$$

Thus, combining (J.3), (J.4), (J.5) and (J.6), it follows that

$$
\begin{aligned}
\hat{\Lambda}_1 \leqslant\ & \epsilon\left(1+\theta\max_{(x,y)\in\mathcal{X}\times\mathcal{Y}_1}\frac{\tilde{R}_{Y_1|X}(y|x)-P_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)}\right)^n+\Pr\left[\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\hat{\boldsymbol{X}})}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(W))}\right)<\eta\right]\\
&+\hat{M}2^{-\eta}\exp\left(K^2\left(\max_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\frac{\tilde{R}_{Y_1|X}(y|x)^2}{P_{Y_1|X}(y|x)}-1\right)\right). \qquad \text{(J.7)}
\end{aligned}
$$

Step 2:

Note that

$$
\Pr\left[\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\hat{\boldsymbol{X}})}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(W))}\right)<\eta\right]=\Pr\left[\sum_{t=1}^{n}\log_2\left(\frac{P_{Y_1|X}(Y_{1,t}|\hat{X}_t)}{P_{Y_1|X}(Y_{1,t}|u_t(W))}\right)<\eta\right]. \qquad \text{(J.8)}
$$

Note also that if $\hat{X}_t=u_t(W)$, then $\log_2\left(\frac{P_{Y_1|X}(Y_{1,t}|\hat{X}_t)}{P_{Y_1|X}(Y_{1,t}|u_t(W))}\right)=0$. Otherwise, if $\hat{X}_t\neq u_t(W)$,

then $\log_2\left(\frac{P_{Y_1|X}(Y_{1,t}|\hat{X}_t)}{P_{Y_1|X}(Y_{1,t}|u_t(W))}\right) \neq 0$. Define the random variable L as

$$L \triangleq \sum_{t=1}^{n} \mathbb{1}_{\{\hat{X}_t \neq u_t(W)\}}. \tag{J.9}$$

Define also

$$\bar{D}_1 = \sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}} \bar{P}_X(x)\tilde{P}_{\hat{X}|X}(\hat{x}|x) D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big), \tag{J.10}$$

$$\eta = (1-\mu)(1-\nu)K\sqrt{n}\bar{D}_1, \tag{J.11}$$

$$\text{and} \quad \mathcal{E}_\mu = \left\{\ell \in \{1,2,\ldots,n\} : \ell > (1-\mu)K\sqrt{n}\right\}. \tag{J.12}$$

Then, it follows that

$$\begin{aligned}
&\Pr\left[\sum_{t=1}^{n}\log_2\left(\frac{P_{Y_1|X}(Y_{1,t}|\hat{X}_t)}{P_{Y_1|X}(Y_{1,t}|u_t(W))}\right) < \eta\right] \\
&= \sum_{\ell\in\mathcal{E}_\mu}\Pr\left[L=\ell\right]\Pr\left[\sum_{t=1}^{n}\log_2\left(\frac{P_{Y_1|X}(Y_{1,t}|\hat{X}_t)}{P_{Y_1|X}(Y_{1,t}|u_t(W))}\right) < \eta | L=\ell\right] + \sum_{\ell\notin\mathcal{E}_\mu}\Pr\left[L=\ell\right] \\
&\quad \cdot\Pr\left[\sum_{t=1}^{n}\log_2\left(\frac{P_{Y_1|X}(Y_{1,t}|\hat{X}_t)}{P_{Y_1|X}(Y_{1,t}|u_t(W))}\right) < \eta | L=\ell\right] \\
&\leqslant \sum_{\ell\in\mathcal{E}_\mu}\Pr\left[L=\ell\right]\Pr\left[\sum_{t=1}^{\ell}\log_2\left(\frac{P_{Y_1|X}(Y_{1,t}|\hat{X}_t)}{P_{Y_1|X}(Y_{1,t}|u_t(W))}\right) < \eta | L=\ell\right] + \Pr\left[L\notin\mathcal{E}_\mu\right]
\end{aligned} \tag{J.13}$$

Using a Chernoff bound, it holds that

$$\begin{aligned}
\Pr\left[L\notin\mathcal{E}_\mu\right] &= \Pr\left[L \leqslant (1-\mu)K\sqrt{n}\right] \\
&\leqslant \exp\left(-\frac{1}{2}\mu^2 K\sqrt{n}\right).
\end{aligned} \tag{J.14}$$

It remains to establish an upper-bound on the first term in (J.13). Note that

$$\begin{aligned}
&\mathbb{E}_{W\hat{\boldsymbol{X}}_{(1:L)}\boldsymbol{Y}_{1,(1:L)}|L=\ell}\left[\sum_{t=1}^{\ell}\log_2\left(\frac{P_{Y_1|X}(Y_{1,t}|\hat{X}_t)}{P_{Y_1|X}(Y_{1,t}|u_t(W))}\right)\right] \\
&= \sum_{t=1}^{\ell}\sum_{\boldsymbol{y}_{(1:\ell)}\in\mathcal{Y}_1^\ell}\sum_{\hat{\boldsymbol{x}}_{(1:\ell)}\in\mathcal{X}^\ell}\sum_{i=1}^{M}\frac{1}{M}\prod_{k=1}^{\ell}\tilde{P}_{\hat{X}|X}(\hat{x}_k|u_k(i))P_{Y_1|X}(y_k|\hat{x}_k)\log_2\left(\frac{P_{Y_1|X}(y_t|\hat{x}_t)}{P_{Y_1|X}(y_t|u_t(i))}\right) \\
&= \sum_{t=1}^{\ell}\sum_{y\in\mathcal{Y}_1}\sum_{\hat{x}\in\mathcal{X}}\sum_{i=1}^{M}\frac{1}{M}\tilde{P}_{\hat{X}|X}(\hat{x}|u_t(i))P_{Y_1|X}(y|\hat{x})\log_2\left(\frac{P_{Y_1|X}(y|\hat{x})}{P_{Y_1|X}(y|u_t(i))}\right) \\
&= \sum_{t=1}^{\ell}\sum_{y\in\mathcal{Y}_1}\sum_{\hat{x}\in\mathcal{X}}\sum_{i=1}^{M}\sum_{x\in\mathcal{X}}\frac{\mathbb{1}_{\{x=u_t(i)\}}}{M}\tilde{P}_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})\log_2\left(\frac{P_{Y_1|X}(y|\hat{x})}{P_{Y_1|X}(y|x)}\right) \\
&= \ell\sum_{\hat{x}\in\mathcal{X}}\sum_{x\in\mathcal{X}}\bar{P}_X(x)\tilde{P}_{\hat{X}|X}(\hat{x}|x)D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big) \\
&= \ell\bar{D}_1.
\end{aligned} \tag{J.15}$$

Note also that for $\ell \in \mathcal{E}_\mu$,

$$\ell > (1-\mu)K\sqrt{n}. \tag{J.16}$$

Hence, it holds that

$$\begin{aligned}\eta - \ell\bar{D}_1 &= (1-\mu)(1-\nu)K\sqrt{n}\bar{D}_1 - \ell\bar{D}_1 \\ &< (1-\nu)\ell\bar{D}_1 - \ell\bar{D}_1 \\ &= -\nu\ell\bar{D}_1.\end{aligned} \tag{J.17}$$

Therefore, it follows that

$$\begin{aligned}&\Pr\left[\sum_{t=1}^{\ell}\log_2\left(\frac{P_{Y_1|X}(Y_{1,t}|\hat{X}_t)}{P_{Y_1|X}(Y_{1,t}|u_t(W))}\right) < \eta | L=\ell\right] \\ &= \Pr\left[\sum_{t=1}^{\ell}\log_2\left(\frac{P_{Y_1|X}(Y_{1,t}|\hat{X}_t)}{P_{Y_1|X}(Y_{1,t}|u_t(W))}\right) - \ell\bar{D}_1 < \eta - \ell\bar{D}_1 | L=\ell\right] \\ &\leqslant \Pr\left[\sum_{t=1}^{\ell}\log_2\left(\frac{P_{Y_1|X}(Y_{1,t}|\hat{X}_t)}{P_{Y_1|X}(Y_{1,t}|u_t(W))}\right) - \ell\bar{D}_1 < -\nu\ell\bar{D}_1 | L=\ell\right] \\ &\overset{(a)}{\leqslant} 2\exp\left(-\frac{\nu^2\ell^2\bar{D}_1^2}{\sum_{t=1}^{\ell}(a_M-a_m)^2}\right) \\ &= 2\exp\left(-\frac{\nu^2\bar{D}_1^2}{(a_M-a_m)^2}\ell\right) \\ &< 2\exp\left(-\frac{\nu^2\bar{D}_1^2}{(a_M-a_m)^2}(1-\mu)K\sqrt{n}\right),\end{aligned} \tag{J.18}$$

where *(a)* follows from Hoeffding's inequality with $a_M = \max_{(x,\hat{x},y)\in\mathcal{X}^2\times\mathcal{Y}_1}\log_2\left(\frac{P_{Y_1|X}(y|\hat{x})}{P_{Y_1|X}(y|x)}\right)$ and $a_m = \min_{(x,\hat{x},y)\in\mathcal{X}^2\times\mathcal{Y}_1}\log_2\left(\frac{P_{Y_1|X}(y|\hat{x})}{P_{Y_1|X}(y|x)}\right)$.

Therefore, it follows by combining (J.13), (J.14) and (J.18) that

$$\Pr\left[\log_2\left(\frac{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\hat{\boldsymbol{X}})}{P_{\boldsymbol{Y}_1|\boldsymbol{X}}(\boldsymbol{Y}_1|\boldsymbol{u}(W))}\right) < \eta\right] \leqslant \exp\left(-\frac{1}{2}\mu^2K\sqrt{n}\right) + 2\exp\left(-\frac{\nu^2\bar{D}_1^2}{(a_M-a_m)^2}(1-\mu)K\sqrt{n}\right). \tag{J.19}$$

Further, combining (J.7), (J.11) and (J.19), it follows that

$$\begin{aligned}\hat{\Lambda}_1 \leqslant\ &\epsilon\left(1+\theta\max_{(x,y)\in\mathcal{X}\times\mathcal{Y}_1}\frac{\tilde{R}_{Y_1|X}(y|x)-P_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)}\right)^n + 2\exp\left(-\frac{\nu^2\bar{D}_1^2}{(a_M-a_m)^2}(1-\mu)K\sqrt{n}\right) \\ &+\exp\left(-\frac{1}{2}\mu^2K\sqrt{n}\right) + \hat{M}2^{-(1-\mu)(1-\nu)K\sqrt{n}\bar{D}_1}\exp\left(K^2\left(\max_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1}\frac{\tilde{R}_{Y_1|X}(y|x)^2}{P_{Y_1|X}(y|x)}-1\right)\right).\end{aligned} \tag{J.20}$$

Thus, $\hat{\Lambda}_1$ can be made arbitrarily small by choosing

$$\log_2\left(\hat{M}\right) = (1-\zeta)(1-\mu)(1-\nu)K\sqrt{n}\bar{D}_1, \tag{J.21}$$

with $(\zeta, \mu, \nu) \in]0,1[^3$ arbitrarily small. ■

— K —

Proof of Lemma 5

This appendix presents the proof of Lemma 5. 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 $x \in \mathcal{X}$:

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

Developping the right hand-side of (K.1) and dividing the two hand-sides by the block-length n yields

$$\frac{N(x|\boldsymbol{u}(i))}{n} = \frac{N(x, x|\boldsymbol{u}(i), \boldsymbol{v}(i,j))}{n} + \sum_{t=1}^{n} \frac{\mathbb{1}_{\{x=u_t(i)\}} \mathbb{1}_{\{x\neq v_t(i,j)\}}}{n}. \tag{K.2}$$

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

$$\begin{aligned} \bar{P}_X(x) &= \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \frac{N(x, x|\boldsymbol{u}(i), \boldsymbol{v}(i,j))}{nM\hat{M}} + \sum_{t=1}^{n} \frac{\mathbb{1}_{\{x=u_t(i)\}} \mathbb{1}_{\{x\neq v_t(i,j)\}}}{nM\hat{M}}. \\ &= \bar{P}_{X\hat{X}}(x, x) + \frac{\omega(x)}{n}. \end{aligned} \tag{K.3}$$

Thus, it follows that

$$\begin{aligned} \omega(x) &= n \left(\bar{P}_X(x) - \bar{P}_{X\hat{X}}(x, x) \right) \\ &= n\bar{P}_X(x) \left(1 - \bar{P}_{\hat{X}|X}(x|x) \right) \\ &= n\bar{P}_X(x)\theta(x), \end{aligned} \tag{K.4}$$

where the last equality follows from the definition of $\theta(x)$ in (4.65).

Note also that from (4.62), 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),
\end{aligned}\tag{K.5}$$

where *(a)* follows from (4.63); *(b)* follows from (4.60); and *(c)* follows from (K.4).

This completes the proof. ■

— L —

Proof of Lemma 6

This appendix presents the proof of Lemma 6. Let $\bar{W} \in \mathcal{W}$ be a random variable that represents the decoded message index at Receiver 2. Consider the joint probability mass functions $Q_{\bar{W}\boldsymbol{Y}_2}$ and $R_{\bar{W}\boldsymbol{Y}_2}$ such that, for all pairs $(i, \boldsymbol{y}) \in \mathcal{W} \times \mathcal{Y}_2^n$,

$$Q_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) = Q_{\boldsymbol{Y}_2}(\boldsymbol{y}) Q_{\bar{W}|\boldsymbol{Y}_2}(i|\boldsymbol{y}), \tag{L.1}$$

$$\text{and} \quad R_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) = R_{\boldsymbol{Y}_2}(\boldsymbol{y}) R_{\bar{W}|\boldsymbol{Y}_2}(i|\boldsymbol{y}), \tag{L.2}$$

where $Q_{\boldsymbol{Y}_2}$ and $R_{\boldsymbol{Y}_2}$ are the marginal channel output probability mass functions respectively in (4.10) and (4.11), and

$$Q_{\bar{W}|\boldsymbol{Y}_2}(i|\boldsymbol{y}) = R_{\bar{W}|\boldsymbol{Y}_2}(i|\boldsymbol{y}) = \mathbb{1}_{\{\boldsymbol{y} \in \mathcal{D}_2(i)\}}. \tag{L.3}$$

Consider also the joint probability mass functions $Q_{W\boldsymbol{Y}_2}$ and $R_{W\boldsymbol{Y}_2}$ respectively in (4.70) and (4.71). Note that

$$\begin{aligned} \|R_{W\boldsymbol{Y}_2} - Q_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} = \frac{1}{2} \sum_{i=1}^{M} \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} \Big| & R_{W\boldsymbol{Y}_2}(i, \boldsymbol{y}) + R_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) - R_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) \\ & - Q_{W\boldsymbol{Y}_2}(i, \boldsymbol{y}) + Q_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) - Q_{\bar{W}\boldsymbol{Y}_2}(i, \boldsymbol{y}) \Big| \\ \leqslant & \left\| R_{\bar{W}\boldsymbol{Y}_2} - Q_{\bar{W}\boldsymbol{Y}_2} \right\|_{\mathrm{TV}} + \left\| Q_{W\boldsymbol{Y}_2} - Q_{\bar{W}\boldsymbol{Y}_2} \right\|_{\mathrm{TV}} + \left\| R_{W\boldsymbol{Y}_2} - R_{\bar{W}\boldsymbol{Y}_2} \right\|_{\mathrm{TV}}, \end{aligned} \tag{L.4}$$

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 (L.4).

First, note that

$$\begin{aligned}\left\|R_{\bar{W}\boldsymbol{Y}_2}-Q_{\bar{W}\boldsymbol{Y}_2}\right\|_{\mathrm{TV}} &= \frac{1}{2M}\sum_{i=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2(i)\}}\left|R_{\boldsymbol{Y}_2}(\boldsymbol{y})-Q_{\boldsymbol{Y}_2}(\boldsymbol{y})\right| \\ &\overset{(a)}{=} \frac{1}{2}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\left|R_{\boldsymbol{Y}_2}(\boldsymbol{y})-Q_{\boldsymbol{Y}_2}(\boldsymbol{y})\right| \\ &= \left\|R_{\boldsymbol{Y}_2}-Q_{\boldsymbol{Y}_2}\right\|_{\mathrm{TV}}, \end{aligned} \tag{L.5}$$

where *(a)* holds since (4.4c) is assumed with equality.

Note also that

$$\begin{aligned}&\left\|Q_{W\boldsymbol{Y}_2}-Q_{\bar{W}\boldsymbol{Y}_2}\right\|_{\mathrm{TV}} \\ &= \frac{1}{2}\sum_{i=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}Q_{\boldsymbol{Y}_2}(\boldsymbol{y})\left|Q_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y})-Q_{\bar{W}|\boldsymbol{Y}_2}(i|\boldsymbol{y})\right| \\ &= \frac{1}{2}\sum_{i=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}Q_{\boldsymbol{Y}_2}(\boldsymbol{y})\left|Q_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y})-\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2(i)\}}\right| \\ &= \frac{1}{2}\sum_{i=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}Q_{\boldsymbol{Y}_2}(\boldsymbol{y})\Big(\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2(i)\}}\big(1-Q_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y})\big)+\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2^{\mathrm{c}}(i)\}}Q_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y})\Big) \\ &= \frac{1}{2}\sum_{i=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\Big(Q_{\boldsymbol{Y}_2}(\boldsymbol{y})\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2(i)\}}-Q_{\boldsymbol{Y}_2}(\boldsymbol{y})Q_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y}) \\ &\quad\cdot\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2(i)\}}+Q_{\boldsymbol{Y}_2}(\boldsymbol{y})Q_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y})\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2^{\mathrm{c}}(i)\}}\Big) \\ &\overset{(a)}{=} \frac{1}{2}\left(1-1+2\sum_{i=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}Q_{\boldsymbol{Y}_2}(\boldsymbol{y})Q_{W|\boldsymbol{Y}_2}(i|\boldsymbol{y})\mathbb{1}_{\{\boldsymbol{y}\in\mathcal{D}_2^{\mathrm{c}}(i)\}}\right) \\ &\leqslant \epsilon, \end{aligned} \tag{L.6}$$

where *(a)* holds since (4.4c) holds with equality. Note that since (4.8c) is assumed with equality, the equality in (L.3) 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

$$\left\|R_{W\boldsymbol{Y}_2}-R_{\bar{W}\boldsymbol{Y}_2}\right\|_{\mathrm{TV}}\leqslant\hat{\epsilon}. \tag{L.7}$$

Plugging (L.5)–(L.7) into (L.4) completes the proof. ■

— M —

Proof of Proposition 5

This appendix presents the proof of Proposition 5.

Note that

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

where *(a)* follows from the independence between W and $\hat{W}$; *(b)* follows from Fano's inequality [38]; *(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 (M.1) is computed with respect to a joint probability mass function $Q_{X\hat{X}Y_1}$, where for all triplets $(x,\hat{x},y) \in \mathcal{X}^2 \times \mathcal{Y}_1$,

$$
Q_{X\hat{X}Y_1}(x,\hat{x},y) \triangleq \bar{P}_X(x)\bar{P}_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x}); \qquad \text{(M.2)}
$$

the empirical conditional probability mass function $\bar{P}_{\hat{X}|X}$ is obtained from both (4.61) and (4.62); and $P_{Y_1|X}$ is the marginal of the joint probability mass function in (4.1b).

In order to calculate the mutual information in (M.1), let $Q_{Y_1|X}$ be the conditional marginal of the joint probability mass function $Q_{X\hat{X}Y_1}$ in (M.2), that is, for all pairs $(x, y) \in \mathcal{X} \times \mathcal{Y}_1$:

$$\begin{aligned} Q_{Y_1|X}(y|x) &= \sum_{\hat{x}\in\mathcal{X}} \bar{P}_{\hat{X}|X}(\hat{x}|x) P_{Y_1|X}(y|\hat{x}) \\ &\overset{(a)}{=} (1-\theta(x))\, P_{Y_1|X}(y|x) + \theta(x) \sum_{\hat{x}\in\mathcal{X}} \hat{P}_{\hat{X}|X}(\hat{x}|x) P_{Y_1|X}(y|\hat{x}) \\ &\overset{(b)}{=} (1-\theta(x))\, P_{Y_1|X}(y|x) + \theta(x) \hat{R}_{Y_1|X}(y|x), \end{aligned} \tag{M.3}$$

where *(a)* follows from Lemma 5 and *(b)* follows with $\hat{R}_{Y_1|X}(y|x)$ in (N.6).
Using (M.2) and (M.3), the mutual information in (M.1) satisfies

$$\begin{aligned} &I(\hat{X}; Y_1|X) \\ &= \sum_{x\in\mathcal{X}} \sum_{\hat{x}\in\mathcal{X}} \sum_{y\in\mathcal{Y}_1} \bar{P}_X(x) \bar{P}_{\hat{X}|X}(\hat{x}|x) P_{Y_1|X}(y|\hat{x}) \cdot \log_2\left(\frac{P_{Y_1|X}(y|\hat{x})}{Q_{Y_1|X}(y|x)}\right) \\ &= \sum_{x\in\mathcal{X}} \sum_{\hat{x}\in\mathcal{X}} \sum_{y\in\mathcal{Y}_1} \bar{P}_X(x) \bar{P}_{\hat{X}|X}(\hat{x}|x) P_{Y_1|X}(y|\hat{x}) \cdot \log_2\left(\frac{P_{Y_1|X}(y|\hat{x}) P_{Y_1|X}(y|x)}{Q_{Y_1|X}(y|x) P_{Y_1|X}(y|x)}\right) \\ &= \sum_{x\in\mathcal{X}} \sum_{\hat{x}\in\mathcal{X}} \sum_{y\in\mathcal{Y}_1} \bar{P}_X(x) \bar{P}_{\hat{X}|X}(\hat{x}|x) P_{Y_1|X}(y|\hat{x}) \cdot \left(\log_2\left(\frac{P_{Y_1|X}(y|\hat{x})}{P_{Y_1|X}(y|x)}\right) - \log_2\left(\frac{Q_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)}\right)\right) \\ &= \sum_{x\in\mathcal{X}} \sum_{\hat{x}\in\mathcal{X}} \sum_{y\in\mathcal{Y}_1} \bar{P}_X(x) \bar{P}_{\hat{X}|X}(\hat{x}|x) P_{Y_1|X}(y|\hat{x}) \\ &\quad \cdot \left(\log_2\left(\frac{P_{Y_1|X}(y|\hat{x})}{P_{Y_1|X}(y|x)}\right) - \log_2\left(\frac{(1-\theta(x))\, P_{Y_1|X}(y|x) + \theta(x)\hat{R}_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)}\right)\right) \\ &= \sum_{x\in\mathcal{X}} \sum_{\hat{x}\in\mathcal{X}} \bar{P}_X(x) \bar{P}_{\hat{X}|X}(\hat{x}|x) \Bigg(D\big(P_{Y_1|X=\hat{x}} || P_{Y_1|X=x}\big) \\ &\quad - \sum_{y\in\mathcal{Y}_1} P_{Y_1|X}(y|\hat{x}) \log_2\left(1 + \theta(x) \frac{\hat{R}_{Y_1|X}(y|x) - P_{Y_1|X}(y|x)}{P_{Y_1|X}(y|x)}\right)\Bigg). \end{aligned} \tag{M.4}$$

The last term in (M.4) 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

$$A_k(x, y) = \frac{\hat{R}_{Y_k|X}(y|x) - P_{Y_k|X}(y|x)}{P_{Y_k|X}(y|x)}. \tag{M.5}$$

Then, the second term in the right hand-side of (M.4) can be written as follows:

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

$$
\begin{aligned}
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1} \bar{P}_X(x)\Bigg(P_{Y_1|X}(y|\hat{x})\log_2\left(1+\theta(x)A_1(x,y)\right)\\
&\quad +\theta(x)\left(\sum_{\hat{x}\in\mathcal{X}}\hat{P}_{\hat{X}|X}(\hat{x}|x)P_{Y_1|X}(y|\hat{x})-P_{Y_1|X}(y|x)\right)\log_2\left(1+\theta(x)A_1(x,y)\right)\Bigg)\\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1} \bar{P}_X(x)\Bigg(P_{Y_1|X}(y|\hat{x})\log_2\left(1+\theta(x)A_1(x,y)\right)\\
&\quad +\theta(x)\left(\hat{R}_{Y_1|X}(y|x)-P_{Y_1|X}(y|x)\right)\log_2\left(1+\theta(x)A_1(x,y)\right)\Bigg)\\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_1} \bar{P}_X(x)\Bigg(P_{Y_1|X}(y|\hat{x})\sum_{k=1}^{\infty}\frac{(-1)^{k+1}\theta(x)^k}{k}A_1(x,y)^k\\
&\quad +\theta(x)\left(\hat{R}_{Y_1|X}(y|x)-P_{Y_1|X}(y|x)\right)\sum_{k'=1}^{\infty}\frac{(-1)^{k'+1}\theta(x)^{k'}}{k'}A_1(x,y)^{k'}\Bigg)\\
&= \sum_{x\in\mathcal{X}}\bar{P}_X(x)\Bigg(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta(x)^k}{k}\chi_k(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\\
&\quad +\sum_{k'=1}^{\infty}\frac{(-1)^{k'+1}\theta(x)^{k'+1}}{k'}\chi_{k'+1}(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\Bigg)\\
&= \sum_{x\in\mathcal{X}}\bar{P}_X(x)\Bigg(\sum_{k=2}^{\infty}\frac{(-1)^{k+1}\theta(x)^k}{k}\chi_k(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\\
&\quad +\sum_{k'=2}^{\infty}\frac{(-1)^{k'}\theta(x)^{k'}}{k'-1}\chi_{k'}(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\Bigg)\\
&= \sum_{x\in\mathcal{X}}\bar{P}_X(x)\sum_{k=2}^{\infty}\frac{(-1)^{k}\theta(x)^k}{k(k-1)}\chi_k(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\\
&= \sum_{x\in\mathcal{X}}\bar{P}_X(x)\Bigg(\frac{\theta(x)^2}{2}\chi_2(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})+\sum_{k=3}^{\infty}\frac{(-1)^{k}\theta(x)^k}{k(k-1)}\chi_k(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\Bigg)\\
&\geqslant \sum_{x\in\mathcal{X}}\bar{P}_X(x)\Bigg(\frac{\theta(x)^2}{2}\chi_2(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})-\frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x},P_{Y_1|X=x})\Bigg)\\
&\geqslant -\sum_{x\in\mathcal{X}}\bar{P}_X(x)\frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x},P_{Y_1|X=x}).
\end{aligned}
\tag{M.6}
$$

Therefore, from (M.4) and (M.6) it follows that

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

$$= \sum_{x\in\mathcal{X}}\sum_{\hat{x}\in\mathcal{X}} \bar{P}_X(x)\theta(x)\hat{P}_{\hat{X}|X}(\hat{x}|x) D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big) + \bar{P}_X(x)\frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x}, P_{Y_1|X=x}). \tag{M.7}$$

Finally, from (M.1) and (M.7), it follows that

$$\log_2\big(\hat{M}\big) \leqslant \frac{1}{1-\hat{\epsilon}}\Bigg(1 + \sum_{x\in\mathcal{X}}\bar{P}_X(x)\Bigg(n\theta(x)\sum_{\hat{x}\in\mathcal{X}}\hat{P}_{\hat{X}|X}(\hat{x}|x) \cdot D\big(P_{Y_1|X=\hat{x}}||P_{Y_1|X=x}\big) + \frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x}, P_{Y_1|X=x})\Bigg)\Bigg). \tag{M.8}$$

This completes the proof. ■

— N —

Proof of Proposition 6

This appendix presents the proof of Proposition 6. The proof is established in two steps. First, upper-bounds on the type-I and type-II error probabilities are established for a particular hypothesis test. Then, these upper-bounds are used to derive a lower-bound on the cardinality of the set $\tilde{\mathcal{W}}$.

Step 1:

Upper bounds 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. This step is reminiscent of [13, Lemma 11].

Under these assumptions, the hypothesis test run by Receiver 2 to determine whether or not a covert communication occurs is the following:

$$\begin{cases} H_0 : \boldsymbol{Y}_2 \sim Q_{\boldsymbol{Y}_2|W=i}, \\ H_1 : \boldsymbol{Y}_2 \sim R_{\boldsymbol{Y}_2|W=i}, \end{cases} \tag{N.1}$$

where the probability mass functions $Q_{\boldsymbol{Y}_2|W=i}$ and $R_{\boldsymbol{Y}_2|W=i}$ are respectively defined in (4.72) and (4.73).

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

$$T_i(\boldsymbol{y}) \triangleq \begin{cases} 0 & \text{if } H_0 \text{ is accepted,} \\ 1 & \text{if } H_1 \text{ is accepted.} \end{cases} \tag{N.2}$$

That is,

$$\alpha_i \triangleq \Pr\left[T_i(\boldsymbol{Y}_2) = 1\right], \text{ and} \tag{N.3}$$

$$\beta_i \triangleq \Pr\left[T_i(\boldsymbol{Y}_2) = 0\right], \tag{N.4}$$

where the probability operator in (N.3) applies assuming that $\boldsymbol{Y}_2 \sim Q_{\boldsymbol{Y}_2|W=i}$ and the probability operator in (N.4) applies assuming that $\boldsymbol{Y}_2 \sim R_{\boldsymbol{Y}_2|W=i}$. The next proposition establishes upper-bounds on α_i in (N.3) and β_i in (N.4), under certain conditions.

Given a fixed block-length $n \in \mathbb{N}$, an (n, M, ϵ)-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

$$\hat{\mathcal{W}}_i = \left\{ j \in \hat{\mathcal{W}} : \omega(i,j) \geqslant \nu \right\}, \tag{N.5}$$

where ν will be specified later.

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

$$\hat{R}_{Y_k|X}(y|x) = \sum_{\hat{x} \in \mathcal{X}} \hat{P}_{\hat{X}|X}(\hat{x}|x) P_{Y_k|X}(y|x), \tag{N.6}$$

with $\hat{P}_{\hat{X}|X}$ is in (4.63).

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

$$B(\boldsymbol{y}) = \sum_{t=1}^{n} A(u_t(i), v_t(i, j^*), y_t), \tag{N.7}$$

with

$$A(x, \hat{x}, y) = \frac{P_{Y_2|X}(y|\hat{x}) - P_{Y_2|X}(y|x)}{P_{Y_2|X}(y|x)}. \tag{N.8}$$

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

Consider the decision rule $T : \mathcal{Y}_2^n \to \{0, 1\}$ of the form in (4.13) such that for all $\boldsymbol{y} \in \mathcal{Y}_2^n$,

$$T(\boldsymbol{y}) = \mathbb{1}_{\{B(\boldsymbol{y}) \geqslant \tau\}}, \tag{N.9}$$

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

Consider first the type-I error probability α_i in (N.3). It follows from the choice of T in (N.9) that

$$\begin{aligned} \alpha_i &= \Pr\left[B(\boldsymbol{Y}_2) \geqslant \tau \right] \\ &= \sum_{\boldsymbol{y} \in \mathcal{Y}_2^n} \prod_{t=1}^{n} P_{Y_2|X}(y_t|u_t(i)) \mathbb{1}_{\{B(\boldsymbol{y}) \geqslant \tau\}}. \end{aligned} \tag{N.10}$$

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

$$Z_{it} = A(u_t(i), v_t(i, j^*), Y), \tag{N.11}$$

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

Then, it follows from (N.10) and the definition of the random variable Z_{it} in (N.11) that

$$\alpha_i = \Pr\left[\sum_{t=1}^{n} Z_{it} \geqslant \tau \right]. \tag{N.12}$$

Denote by μ_{it}, σ_{it} and ϕ_{it} the first, second and third absolute moments of the random

variable Z_{it}, respectively, *i.e.*,

$$\mu_{it} = \sum_{y \in \mathcal{Y}_2} P_{Y_2|X}(y|u_t(i)) A(u_t(i), v_t(i, j^*), y), \tag{N.13}$$

$$\sigma_{it}^2 = \sum_{y \in \mathcal{Y}_2} P_{Y_2|X}(y|u_t(i)) A(u_t(i), v_t(i, j^*), y)^2 - \mu_{it}^2, \tag{N.14}$$

and

$$\phi_{it} = \sum_{y \in \mathcal{Y}_2} P_{Y_2|X}(y|u_t(i)) \left| A(u_t(i), v_t(i, j^*), y) - \mu_{it} \right|^3. \tag{N.15}$$

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

$$\begin{aligned}
\mu_{it} &= \sum_{y \in \mathcal{Y}_2} P_{Y_2|X}(y|u_t(i)) A(u_t(i), v_t(i, j^*), y) \\
&= \sum_{y \in \mathcal{Y}_2} P_{Y_2|X}(y|v_t(i, j^*)) - P_{Y_2|X}(y|u_t(i)) \\
&= 0,
\end{aligned} \tag{N.16}$$

$$\begin{aligned}
\sigma_{it}^2 &= \sum_{y \in \mathcal{Y}_2} P_{Y_2|X}(y|u_t(i)) A(u_t(i), v_t(i, j^*), y)^2 - \mu_{it}^2 \\
&= \sum_{y \in \mathcal{Y}_2} \frac{\left(P_{Y_2|X}(y|v_t(i, j^*)) - P_{Y_2|X}(y|u_t(i))\right)^2}{P_{Y_2|X}(y|u_t(i))} \\
&\overset{(a)}{=} d,
\end{aligned} \tag{N.17}$$

where *(a)* follows from (4.77). Finally,

$$\begin{aligned}
\phi_{it} &= \sum_{y \in \mathcal{Y}_2} P_{Y_2|X}(y|u_t(i)) \left| A(u_t(i), v_t(i, j^*), y) - \mu_{it} \right|^3 \\
&= \sum_{y \in \mathcal{Y}_2} \frac{\left| P_{Y_2|X}(y|v_t(i, j^*)) - P_{Y_2|X}(y|u_t(i)) \right|^3}{P_{Y_2|X}(y|u_t(i))^2} \\
&\leqslant \phi_i^*,
\end{aligned} \tag{N.18}$$

with

$$\phi_i^* \triangleq \max_{t \in \{1, 2, \ldots, n\}} \phi_{it}. \tag{N.19}$$

Therefore, it follows that

$$\mu_i = \sum_{t=1}^{n} \mu_{it} = 0, \tag{N.20}$$

$$\sigma_i^2 = \sum_{t=1}^{n} \sigma_{it}^2 = nd, \tag{N.21}$$

$$\text{and} \quad \phi_i = \sum_{t=1}^{n} \phi_{it} \leqslant n\phi_i^*, \tag{N.22}$$

with d in (4.77) and ϕ_i^* in (N.19).

It follows from (N.12) that

$$\begin{aligned}\alpha_i &= \Pr\left[\sum_{t=1}^{n} Z_{it} - \mu_i \geqslant \sigma_i \frac{\tau - \mu_i}{\sigma_i}\right] \\ &\overset{(a)}{\leqslant} Q\left(\frac{\tau - \mu_i}{\sigma_i}\right) + c_0 \frac{\phi_i}{\sigma_i^3} \\ &\overset{(b)}{\leqslant} Q\left(\frac{\tau}{\sqrt{nd}}\right) + \frac{n c_0 \phi_i^*}{\sqrt{nd}^3} \\ &\overset{(c)}{=} Q\left(\frac{\tau}{\sqrt{nd}}\right) + \frac{c_3}{\sqrt{n}},\end{aligned} \tag{N.23}$$

where *(a)* follows from the Berry-Esseen Theorem (Theorem 21); *(b)* follows from (N.20)-(N.22); and *(c)* follows with

$$c_3 \triangleq c_0 \phi_i^* d^{-\frac{3}{2}}. \tag{N.24}$$

Consider now the type-II error probability β_i in (N.4). It follows from the choice of T in (N.9) that

$$\begin{aligned}\beta_i &= \Pr\left[B(\boldsymbol{Y}_2) \leqslant \tau\right] \\ &= \frac{1}{|\hat{\mathcal{W}}_i|} \sum_{j \in \hat{\mathcal{W}}_i} \Pr\left[B(\boldsymbol{Y}_2) < \tau \middle| \hat{W} = j\right] \\ &\leqslant \max_{j \in \hat{\mathcal{W}}_i} \Pr\left[B(\boldsymbol{Y}_2) < \tau \middle| \hat{W} = j\right].\end{aligned} \tag{N.25}$$

Let j^* be

$$j^* \in \operatorname*{argmax}_{j \in \hat{\mathcal{W}}_i} \Pr\left[B(\boldsymbol{Y}_2) < \tau \middle| \hat{W} = j\right], \tag{N.26}$$

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

$$\hat{Z}_{it} = A(u_t(i), v_t(i, j^*), \hat{Y}), \tag{N.27}$$

where $\hat{Y}$ is distributed according to $P_{Y_2|X=v_t(i,j^*)}$. Then, it follows from (N.25) and the definition of the random variable $\hat{Z}_{it}$ in (N.27) that

$$\beta_i \leqslant \Pr\left[\sum_{t=1}^{n} \hat{Z}_{it} < \tau\right]. \tag{N.28}$$

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.*,

$$\hat{\mu}_{it} = \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))A(u_t(i),v_t(i,j^*),y), \tag{N.29}$$

$$\hat{\sigma}^2_{it} = \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))A(u_t(i),v_t(i,j^*),y)^2 - \hat{\mu}^2_{it}, \tag{N.30}$$

and

$$\hat{\phi}_{it} = \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))\left|A(u_t(i),v_t(i,j^*),y) - \hat{\mu}_{it}\right|^3. \tag{N.31}$$

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

$$\hat{\mu}_{it} = \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))A(u_t(i),v_t(i,j^*),y), \tag{N.32}$$

$$\hat{\sigma}^2_{it} = \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))A(u_t(i),v_t(i,j^*),y)^2 - \hat{\mu}^2_{it}, \tag{N.33}$$

and

$$\begin{aligned}\hat{\phi}_{it} &= \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))\left|A(u_t(i),v_t(i,j^*),y) - \hat{\mu}_{it}\right|^3 \\ &\leqslant \hat{\phi}^*_i,\end{aligned} \tag{N.34}$$

with

$$\hat{\phi}^*_i \triangleq \max_{t\in\{1,2,\ldots,n\}} \hat{\phi}_{it}, \tag{N.35}$$

Therefore, it follows that

$$\begin{aligned}\hat{\mu}_i &= \sum_{t=1}^n \hat{\mu}_{it} \\ &= \sum_{t=1}^n \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*))A(u_t(i),v_t(i,j^*),y) \\ &= \sum_{t=1}^n \sum_{y\in\mathcal{Y}_2} \mathbb{1}_{\{u_t(i)\neq v_t(i,j^*)\}} P_{Y_2|X}(y|v_t(i,j^*))A(u_t(i),v_t(i,j^*),y) \\ &= \sum_{t=1}^n \sum_{y\in\mathcal{Y}_2} \mathbb{1}_{\{u_t(i)\neq v_t(i,j^*)\}} A(u_t(i),v_t(i,j^*),y)\left(P_{Y_2|X}(y|v_t(i,j^*)) - P_{Y_2|X}(y|u_t(i))\right) \\ &= \sum_{t=1}^n \sum_{y\in\mathcal{Y}_2} \frac{\left(P_{Y_2|X}(y|v_t(i,j^*)) - P_{Y_2|X}(y|u_t(i))\right)^2}{P_{Y_2|X}(y|u_t(i))} \mathbb{1}_{\{u_t(i)\neq v_t(i,j^*)\}} \\ &\overset{(a)}{=} \omega(i,j^*)d,\end{aligned} \tag{N.36}$$

where *(a)* follows from (4.77). It also follows that,

$$\begin{aligned}
\hat{\sigma}_i^2 &= \sum_{t=1}^{n} \hat{\sigma}_{it}^2 \\
&= \sum_{t=1}^{n} \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*)) \left(A(u_t(i),v_t(i,j^*),y)^2 - \hat{\mu}_{it}^2\right) \\
&\leqslant \sum_{t=1}^{n} \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*)) A(u_t(i),v_t(i,j^*),y)^2 \\
&= \sum_{t=1}^{n} \mathbb{1}_{\{u_t(i)=v_t(i,j)\}} \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|u_t(i)) A(u_t(i),v_t(i,j^*),y)^2 \\
&+ \sum_{t=1}^{n} \mathbb{1}_{\{u_t(i)\neq v_t(i,j)\}} \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|v_t(i,j^*)) A(u_t(i),v_t(i,j^*),y)^2 \\
&= \sum_{t=1}^{n} \sum_{y\in\mathcal{Y}_2} P_{Y_2|X}(y|u_t(i)) A(u_t(i),v_t(i,j^*),y)^2 + A(u_t(i),v_t(i,j^*),y)^2 \mathbb{1}_{\{u_t(i)\neq v_t(i,j)\}} \\
&\quad \cdot (P_{Y_2|X}(y|v_t(i,j^*)) - P_{Y_2|X}(y|u_t(i))) \\
&\overset{(a)}{\leqslant} nd + \omega(i,j^*)d' \\
&\leqslant nd + \omega(i,j^*)\left|d'\right|,
\end{aligned} \tag{N.37}$$

where *(a)* follows from (4.77), with

$$d' \triangleq \max_{x\in\mathcal{X}} \chi_3(P_{Y_2|X=x}, P_{Y_2|X=u_t(i)}). \tag{N.38}$$

In addition, $\hat{\sigma}_i^2$ also satisfies:

$$\begin{aligned}
\hat{\sigma}_i^2 &= \sum_{t=1}^{n} \hat{\sigma}_{it}^2 \\
&\geqslant nd + \omega(i,j^*)d'' - \sum_{t=1}^{n} \hat{\mu}_{it}^2 \\
&\geqslant n(d - \hat{\mu}_i^*) + \omega(i,j^*)d'',
\end{aligned} \tag{N.39}$$

with

$$\hat{\mu}_i^* \triangleq \max_{t\in\{1,2,\ldots,n\}} \hat{\mu}_{it}^2, \tag{N.40}$$

and

$$d'' \triangleq \min_{x\in\mathcal{X}} \chi_3(P_{Y_2|X=x}, P_{Y_2|X=u_t(i)}). \tag{N.41}$$

Finally, it also holds that

$$\hat{\phi}_i = \sum_{t=1}^{n} \hat{\phi}_{it} \leqslant n\hat{\phi}_i^*. \tag{N.42}$$

It follows from (N.28) that

$$\begin{aligned}
\beta_i &\leqslant \Pr\left[\sum_{t=1}^{n} \hat{Z}_{it} - \hat{\mu}_i < \hat{\sigma}_i \frac{\tau - \hat{\mu}_i}{\hat{\sigma}_i}\right] \\
&= 1 - \Pr\left[\sum_{t=1}^{n} \hat{Z}_{it} - \hat{\mu}_i \geqslant \hat{\sigma}_i \frac{\tau - \hat{\mu}_i}{\hat{\sigma}_i}\right] \\
&\overset{(a)}{\leqslant} Q\left(-\frac{\tau - \hat{\mu}_i}{\hat{\sigma}_i}\right) + c_0 \frac{\hat{\phi}_i}{\hat{\sigma}_i^3} \\
&= Q\left(\frac{\hat{\mu}_i - \tau}{\hat{\sigma}_i}\right) + c_0 \frac{\hat{\phi}_i}{\hat{\sigma}_i^3} \\
&\overset{(b)}{\leqslant} Q\left(\frac{\omega(i,j^*)d - \tau}{\sqrt{nd + \omega(i,j^*)\,|d'|}}\right) + \frac{n c_0 \hat{\phi}_i^*}{\sqrt{n\,(d - \hat{\mu}_i^*) + \omega(i,j^*)\,|d'|}^3} \\
&\overset{(c)}{\leqslant} Q\left(\frac{\omega(i,j^*)d - \tau}{\sqrt{nd + \omega(i,j^*)\,|d'|}}\right) + \frac{c_4}{\sqrt{n}},
\end{aligned} \tag{N.43}$$

where d' is in (N.38), *(a)* follows from the Berry-Esseen Theorem (Theorem 21); *(b)* follows from (N.36), (N.37), (N.39), (N.42); and *(c)* follows with c_4 a positive constant

$$c_4 \triangleq c_0 \hat{\phi}^* \left(d - \hat{\mu}_i^*\right)^{-\frac{3}{2}}. \tag{N.44}$$

Finally, choosing τ such that

$$\tau = \frac{\nu d}{2}, \tag{N.45}$$

and plugging it into (N.23) and (N.43) yields respectively

$$\alpha_i \leqslant Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right) + \frac{c_3}{\sqrt{n}}, \tag{N.46}$$

and

$$\begin{aligned}
\beta_i &\leqslant Q\left(\frac{\omega(i,j^*)d - \frac{\nu d}{2}}{\sqrt{nd + \omega(i,j^*)\,|d'|}}\right) + \frac{c_4}{\sqrt{n}} \\
&\leqslant Q\left(\frac{\nu d}{2\sqrt{nd + \omega(i,j^*)\,|d'|}}\right) + \frac{c_4}{\sqrt{n}},
\end{aligned} \tag{N.47}$$

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

$$\begin{aligned}
Q\left(\frac{\nu d}{2\sqrt{nd+\omega(i,j^*)\,|d'|}}\right) &= Q\left(\frac{\nu d}{2\sqrt{n(d+\frac{\omega(i,j^*)}{n}\,|d'|)}}\right)\\
&= Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\frac{1}{\sqrt{1+\frac{\omega(i,j^*)|d'|}{nd}}}\right)\\
&\overset{(a)}{\leqslant} Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\left(1-\frac{\omega(i,j^*)\,|d'|}{2nd}\right)\right)\\
&\overset{(b)}{\leqslant} Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right)+\frac{\nu\omega(i,j^*)\,|d'|}{4n\sqrt{2\pi d}},\\
&\leqslant Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right)+\frac{\nu^2\sqrt{n}\,|d'|}{4n\sqrt{2\pi d}},\\
&= Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right)+\frac{\nu^2\,|d'|}{4\sqrt{n}\sqrt{2\pi d}},
\end{aligned}\tag{N.48}$$

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

$$c_5 \triangleq c_4 + \frac{\nu^2\,|d'|}{4\sqrt{2\pi d}},\tag{N.49}$$

it follows that

$$\beta_i \leqslant Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right)+\frac{c_5}{\sqrt{n}}.\tag{N.50}$$

Step 2:

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

$$\begin{aligned}
\delta &\geqslant \|R_{\boldsymbol{Y}_2} - Q_{\boldsymbol{Y}_2}\|_{\mathrm{TV}}\\
&\geqslant \|R_{W\boldsymbol{Y}_2} - Q_{W\boldsymbol{Y}_2}\|_{\mathrm{TV}} - \epsilon - \hat{\epsilon}\\
&= -\epsilon-\hat{\epsilon}+\frac{1}{2M}\sum_{i=1}^{M}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\left|\frac{1}{\hat{M}}\sum_{j=1}^{\hat{M}}P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}(i,j))-P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i)\right|\\
&= \frac{1}{M}\sum_{i=1}^{M}\left\|R_{\boldsymbol{Y}_2|W=i}-Q_{\boldsymbol{Y}_2|W=i}\right\|_{\mathrm{TV}}-\epsilon-\hat{\epsilon},
\end{aligned}\tag{N.51}$$

where the probability mass functions $Q_{\boldsymbol{Y}_2|W=i}$ and $R_{\boldsymbol{Y}_2|W=i}$ are respectively defined in (4.72) and (4.73). For all message indices $i \in \mathcal{W}$, consider the set $\hat{\mathcal{W}}_i$ in (N.5). Note that $\hat{\mathcal{W}}_i$ and

$\hat{\mathcal{W}}_i^{\mathrm{c}}$ form a partition of the set $\hat{\mathcal{W}}$. Let $R_{\boldsymbol{Y}_2|W}^{(\hat{\mathcal{W}}_i)}$ and $R_{\boldsymbol{Y}_2|W}^{(\hat{\mathcal{W}}_i^{\mathrm{c}})}$ be respectively defined by

$$R_{\boldsymbol{Y}_2|W}^{(\hat{\mathcal{W}}_i)}(\boldsymbol{y}|i) \triangleq \frac{1}{\left|\hat{\mathcal{W}}_i\right|} \sum_{j \in \hat{\mathcal{W}}_i} \prod_{t=1}^{n} P_{Y_2|X}(y_t|v_t(i,j)), \text{ and} \tag{N.52}$$

$$R_{\boldsymbol{Y}_2|W}^{(\hat{\mathcal{W}}_i^{\mathrm{c}})}(\boldsymbol{y}|i) \triangleq \frac{1}{\left|\hat{\mathcal{W}}_i^{\mathrm{c}}\right|} \sum_{j \in \hat{\mathcal{W}}_i^{\mathrm{c}}} \prod_{t=1}^{n} P_{Y_2|X}(y_t|v_t(i,j)). \tag{N.53}$$

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

$$\begin{cases} H_0 : \boldsymbol{Y}_2 \sim Q_{\boldsymbol{Y}_2|W=i}, \\ H_1 : \boldsymbol{Y}_2 \sim R_{\boldsymbol{Y}_2|W=i}^{(\hat{\mathcal{W}}_i)}, \end{cases} \tag{N.54}$$

where the probability mass functions $Q_{\boldsymbol{Y}_2|W=i}$ and $R_{\boldsymbol{Y}_2|W=i}^{(\hat{\mathcal{W}}_i)}$ are respectively defined in (4.72) and (N.52).
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

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

That is,

$$\hat{\alpha}_i \triangleq \Pr\left[T_i(\boldsymbol{Y}_2) = 1\right], \text{ and} \tag{N.56}$$

$$\hat{\beta}_i \triangleq \Pr\left[T_i(\boldsymbol{Y}_2) = 0\right], \tag{N.57}$$

where the probability operator in (N.56) applies assuming that $\boldsymbol{Y}_2 \sim Q_{\boldsymbol{Y}_2|W=i}$ and the probability operator in (N.57) 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

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

$$
\begin{aligned}
&\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}\left(P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i))-P_{\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}(i,j))\right)\right| \\
&\quad -\left|\frac{1}{\hat{M}}\sum_{j\in\hat{\mathcal{W}}_i^{\mathsf{c}}}\left(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| \\
&= \frac{1}{2}\sum_{\boldsymbol{y}\in\mathcal{Y}_2^n}\Bigg(\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| \\
&\quad -\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|\Bigg) \\
&\overset{(b)}{=} \frac{|\hat{\mathcal{W}}_i|}{\hat{M}}\left\|R_{\boldsymbol{Y}_2|W=i}^{(\hat{\mathcal{W}}_i)}-Q_{\boldsymbol{Y}_2|W=i}\right\|_{\mathrm{TV}}-\frac{|\hat{\mathcal{W}}_i^{\mathsf{c}}|}{\hat{M}}\left\|R_{\boldsymbol{Y}_2|W=i}^{(\hat{\mathcal{W}}_i^{\mathsf{c}})}-Q_{\boldsymbol{Y}_2|W=i}\right\|_{\mathrm{TV}} \\
&\overset{(c)}{\geqslant} \frac{|\hat{\mathcal{W}}_i|}{\hat{M}}\left\|R_{\boldsymbol{Y}_2|W=i}^{(\hat{\mathcal{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{N.58}
$$

where ν will be specified later, c_{15} is a constant, *(a)* is a consequence of the triangle inequality; *(b)* follows from the definition of $R_{\boldsymbol{Y}_2|W}^{(\hat{\mathcal{W}}_i)}$ and $R_{\boldsymbol{Y}_2|W}^{(\hat{\mathcal{W}}_i^{\mathsf{c}})}$ in (N.52) and (N.53) respectively; *(c)* follows since $\left\|R_{\boldsymbol{Y}_2|W=i}^{(\hat{\mathcal{W}}_i^{\mathsf{c}})}-Q_{\boldsymbol{Y}_2|W=i}\right\|_{\mathrm{TV}}\leqslant 1$; *(d)* follows since $\alpha_i+\beta_i\geqslant 1-\left\|R_{\boldsymbol{Y}_2|W=i}^{(\hat{\mathcal{W}}_i)}-Q_{\boldsymbol{Y}_2|W=i}\right\|_{\mathrm{TV}}$ [37, Theorem 13.1.1]; and *(e)* follows from (N.46) and (N.50).

Plugging (N.51) into (N.58) yields

$$
\delta\geqslant 1-2Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right)-\frac{c_{15}}{\sqrt{n}}-2\sum_{i=1}^{M}\frac{|\hat{\mathcal{W}}_i^{\mathsf{c}}|}{M\hat{M}}-\epsilon-\hat{\epsilon}. \tag{N.59}
$$

Choosing ν in (N.5) such that

$$
\nu=\frac{2\sqrt{n}}{\sqrt{d}}Q^{-1}\left(\frac{1-\delta-\eta}{2}\right), \tag{N.60}
$$

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

$$\begin{aligned} 2\sum_{i=1}^{M} \frac{|\hat{\mathcal{W}}_i^{\mathsf{c}}|}{M\hat{M}} &\geqslant 1 - 2Q\left(\frac{\nu\sqrt{d}}{2\sqrt{n}}\right) - \frac{c_{15}}{\sqrt{n}} - \delta - \epsilon - \hat{\epsilon} \\ &= 1 - 1 + \delta + \eta - \frac{c_{15}}{\sqrt{n}} - \delta - \epsilon - \hat{\epsilon} \\ &= \eta - \frac{c_{15}}{\sqrt{n}} - \epsilon - \hat{\epsilon}, \end{aligned} \tag{N.61}$$

which implies

$$\frac{\left|\tilde{\mathcal{W}}\right|}{M} = \sum_{i=1}^{M} \frac{|\hat{\mathcal{W}}_i^{\mathsf{c}}|}{M} \geqslant \hat{M}\left(\frac{\eta}{2} - \frac{c_6}{\sqrt{n}} - \epsilon - \hat{\epsilon}\right), \tag{N.62}$$

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

— O —

Proof of Lemma 7

This appendix presents the proof of Lemma 7. Note that the probability mass function $P_{Y_2|X}$ verifies for all $(x, x') \in \mathcal{X}^2$ such that $x \neq x'$:

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

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 $x \neq x'$:

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

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

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

This completes the proof. ■

Bibliography

[1] C. E. Shannon. Communication theory of secrecy systems. *The Bell System Technical Journal*, 28(4):656–715, Oct 1949.

[2] K. S. K. Arumugam and M. R. Bloch. Covert communication over broadcast channels. In *2017 IEEE Information Theory Workshop (ITW)*, pages 299–303, Kaohsiung, Taiwan, Nov. 2017.

[3] Keerthi Suria Kumar Arumugam and Matthieu R. Bloch. Embedding Covert Information in Broadcast Communications. *ArXiv e-prints*, page arXiv:1808.09556, Aug. 2018.

[4] A. D. Wyner. The wire-tap channel. *The Bell System Technical Journal*, 54(8):1355–1387, Oct 1975.

[5] I. Csiszár. Almost independence and secrecy capacity. *Problems Inform. Transmission*, 32(1):40–47, 1996.

[6] J. Hou and G. Kramer. Effective secrecy: Reliability, confusion and stealth. In *IEEE International Symposium on Information Theory (ISIT)*, pages 601–605, Honolulu, Hawaii, Jun. 2014.

[7] B. A. Bash, D. Goeckel, and D. Towsley. Square root law for communication with low probability of detection on AWGN channels. In *IEEE International Symposium on Information Theory Proceedings*, pages 448–452, Cambridge, MA, USA, Jul. 2012.

[8] 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*, 31(9):1921–1930, Sep. 2013.

[9] 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)*, pages 2945–2949, Istanbul, Turkey, Jul. 2013.

[10] P. H. Che, M. Bakshi, and S. Jaggi. Reliable deniable communication: Hiding messages in noise. *CoRR*, abs/1304.6693, 2013.

[11] M. Bloch. Covert communications over noisy channels: A resolvability perspective. *IEEE Transactions on Information Theory*, 62(5):2334–2354, May 2016.

[12] L. Wang, G. Wornell, and L. Zheng. Fundamental limits of communication with low probability of detection. *IEEE Trans. Inf. Theory*, 62(6):3493–3503, Jun. 2016.

[13] M. Tahmasbi and M. R. Bloch. First and second order asymptotics in covert communication. *IEEE Transactions on Information Theory*, 65(4):2190 –2212, Apr. 2019.

[14] L. Wang. On covert communication over infinite-bandwidth gaussian channels. In *IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5, Kalamata, Greece, Jun. 2018.

[15] S. Lee, L. Wang, A. Khisti, and G. W. Womell. Covert communication with non-causal channel-state information at the transmitter. In *IEEE International Symposium on Information Theory (ISIT)*, pages 2830–2834, Aachen, Germany, Jun. 2017.

[16] S. Lee, L. Wang, A. Khisti, and G. W. Wornell. Covert communication with channel-state information at the transmitter. *IEEE Transactions on Information Forensics and Security*, 13(9):2310–2319, Sep. 2018.

[17] Mehrdad Tahmasbi, Anne Savard, and Matthieu R Bloch. Covert capacity of non-coherent rayleigh-fading channels. submitted to *IEEE Transactions on Information Theory*, Oct. 2018.

[18] Ramin Soltani, Dennis Goeckel, Don Towsley, and Amir Houmansadr. Covert communications on poisson packet channels. *CoRR*, abs/1610.00381, 2016.

[19] Ramin Soltani, Dennis Goeckel, Don Towsley, and Amir Houmansadr. Covert communications on renewal packet channels. *CoRR*, abs/1610.00368, 2016.

[20] Ramin Soltani, Dennis Goeckel, Don Towsley, and Amir Houmansadr. Fundamental limits of covert packet insertion. *CoRR*, abs/1903.11640, 2019.

[21] Ramin Soltani, Dennis Goeckel, Don Towsley, and Amir Houmansadr. Fundamental limits of covert bit insertion in packets. *CoRR*, abs/1810.03510, 2018.

[22] L. Wang. The continuous-time poisson channel has infinite covert communication capacity. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 756–760, Vail, CO, USA, Jun. 2018.

[23] M. Tahmasbi, M. R. Bloch, and V. Y. F. Tan. Error exponent for covert communications over discrete memoryless channels. In *2017 IEEE Information Theory Workshop (ITW)*, pages 304–308, Kaohsiung, Taiwan, Nov. 2017.

[24] M. Tahmasbi and M. R. Bloch. Covert secret key generation. In *2017 IEEE Conference on Communications and Network Security (CNS)*, pages 540–544, Las Vegas, NV, USA, Oct. 2017.

[25] R. Soltani, B. A. Bash, D. Goeckel, S. Guha, and D. Towsley. Covert wireless communication with artificial noise generation. In *Allerton Conference*, pages 1078–1085, Oct. 2014.

[26] R. Soltani, D. Goeckel, D. Towsley, B. A. Bash, and S. Guha. Covert wireless communication with artificial noise generation. *IEEE Transactions on Wireless Communications*, pages 1078–1085, Oct. 2014.

[27] Q. E. Zhang, M. Bakshi, and S. Jaggi. Covert communication over adversarially jammed channels. In *IEEE Information Theory Workshop (ITW)*, pages 1–5, Guangzhou, China, Nov. 2018.

[28] Qiaosheng Eric Zhang, Mayank Bakshi, and Sidharth Jaggi. Covert communication over adversarially jammed channels. *CoRR*, abs/1805.02426, 2018.

[29] T. V. Sobers, B. A. Bash, S. Guha, D. Towsley, and D. Goeckel. Covert communication in the presence of an uninformed jammer. *IEEE Transactions on Wireless Communications*, 16(9):6193–6206, Sep. 2017.

[30] V. Y. F. Tan and S. Lee. Time-division is optimal for covert communication over some broadcast channels. *IEEE Transactions on Information Forensics and Security*, 14(5):1377–1389, May 2019.

[31] K. S. K. Arumugam and M. R. Bloch. Keyless covert communication over multiple-access channels. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 2229–2233, Jul. 2016.

[32] Keerthi Suria Kumar Arumugam and Matthieu R Bloch. Covert communication over a k-user multiple access channel. submitted to *IEEE Transactions on Information Theory*, Mar. 2018.

[33] K. S. Kumar Arumugam, M. R. Bloch, and L. Wang. Covert communication over a physically degraded relay channel with non-colluding wardens. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 766–770, Vail, CO, USA, Jun. 2018.

[34] Matthieu R Bloch and Saikat Guha. Optimal covert communications using pulse-position modulation. In *Proc. of IEEE International Symposium on Information Theory*, pages 2835–2839, Aachen, Germany, Jun. 2017.

[35] Ishaque Ashar Kadampot, Mehrdad Tahmasbi, and Matthieu R. Bloch. Multilevel-coded pulse position modulation for covert communications. In *Proc. of IEEE International Symposium on Information Theory*, pages 1864–1868, Vail, CO, USA, Jun. 2018.

[36] Ishaque Ashar Kadampot, Mehrdad Tahmasbi, and Matthieu R Bloch. Multilevel-coded pulse-position modulation for covert communications over binary-input discrete memoryless channels. submitted to *IEEE Transactions on Information Theory*, Nov. 2018.

[37] E.L. Lehmann and J.P. Romano. *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer, 3rd edition, 2005.

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

[39] M. Bellare, S. Tessaro, and A. Vardy. A cryptographic treatment of the wiretap channel. *CoRR*, abs/1201.2205, 2012.

[40] T. S. Han and S. Verdú. Approximation theory of output statistics. *IEEE Trans. Inf. Theory*, 39(3):752–772, May 1993.

[41] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, Hoboken, NJ, USA, 1991.

[42] A. Kerckhoffs. La cryptographie militaire. *Journal des Sciences Militaires*, IX:5–38, Feb. 1883.

[43] L. H. Ozarow and A. D. Wyner. Wire-tap channel II. *AT&T Bell Laboratories Technical Journal*, 63(10):2135–2157, Dec. 1984.

[44] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory*, 56(5):2307–2359, May 2010.

[45] A. C. Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, 49(1):122–136, May 1941.

[46] C.-G. Esseen. On the liapunoff limit of error in the theory of probability. *Arkiv för Matematik, Astronomi och Fysik*, A28:1–19, 1942.

[47] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. John Wiley and Sons, New York, NY, USA, 2nd edition, 1971.

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

[49] M. Goemans. Chernoff bounds, and some applications.

FOLIO ADMINISTRATIF

THESE DE L'UNIVERSITE DE LYON OPEREE AU SEIN DE L'INSA LYON

NOM : KIBLOFF
(avec précision du nom de jeune fille, le cas échéant)

DATE de SOUTENANCE : 26/09/2019

Prénoms : David Antoine Michel

TITRE : Contributions théoriques sur les communications furtives

NATURE : Doctorat

Numéro d'ordre : AAAALYSEIXXXX

Ecole doctorale : Electronique, Electrotechnique et Automatique

Spécialité : Génie Electrique

RESUME :

L'étude des communications furtives, aussi connues sous le nom de communications avec faible probabilité de détection, a connu un regain d'intérêt dans la communauté Théorie de l'Information dans les années passées. Depuis que Bash et. al. ont montré en 2012 que les communications point-à-point sous contrainte de furtivité obéissent à une loi en racine carrée, le nombre de contributions dans ce domaine n'a cessé de croître. Dans cette thèse, deux nouveaux problèmes de communications furtives sont présentés. Premièrement, les communications furtives sur les liens point-à-point sont étudiées quand l'adversaire observe uniquement une fraction des sorties de canal pour essayer de détecter la communication. Une borne de faisabilité pour une longueur finie de blocs est obtenue pour ce problème. Deuxièmement, le problème d'introduction d'information furtive dans un code de broadcast existant est présenté. Etant donné un code de broadcast pour transmettre de l'information à deux récepteurs, le but de cette étude est de déterminer le nombre maximum de bits d'information qui peuvent être envoyés de manière fiable à l'un des récepteurs tout en étant furtifs pour l'autre récepteur. Pour ce problème, une borne de faisabilité et une borne d'impossibilité sont obtenues dans le régime asymptotique pour une classe particulière de canaux, i.e., les canaux symétriques. Ces deux bornes caractérisent le nombre maximal de bits d'information qui peuvent être introduits de manière furtive dans le code de broadcast donné pour des canaux symétriques.

MOTS-CLÉS : Communications furtives, faible probabilité de détection, sécurité de la couche physique, théorie de l'information

Laboratoire (s) de recherche : CITI

Directeur de thèse: Guillaume VILLEMAUD

Président de jury :

Composition du jury : Prof. Inbar FIJALKOW, Dr. Aline ROUMY, Prof. Albert GUILLEN I FABREGAS, Prof. Laurent CLAVIER, Prof. Jean-Marie GORCE, Dr. Ligong WANG, Dr. Samir PERLAZA, Dr. Guillaume VILLEMAUD

Département FEDORA – INSA Lyon - Ecoles Doctorales – Quinquennal 2016-2020

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	**CHIMIE DE LYON** **http://www.edchimie-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage **secretariat@edchimie-lyon.fr** INSA : R. GOURDON	M. Stéphane DANIELE Institut de recherches sur la catalyse et l'environnement de Lyon IRCELYON-UMR 5256 Équipe CDFA 2 Avenue Albert EINSTEIN 69 626 Villeurbanne CEDEX **directeur@edchimie-lyon.fr**
E.E.A.	**ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE** **http://edeea.ec-lyon.fr** Sec. : M.C. HAVGOUDOUKIAN **ecole-doctorale.eea@ec-lyon.fr**	M. Gérard SCORLETTI École Centrale de Lyon 36 Avenue Guy DE COLLONGUE 69 134 Écully Tél : 04.72.18.60.97 Fax 04.78.43.37.17 **gerard.scorletti@ec-lyon.fr**
E2M2	**ÉVOLUTION, ÉCOSYSTÈME, MICROBIOLOGIE, MODÉLISATION** **http://e2m2.universite-lyon.fr** Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 INSA : H. CHARLES **secretariat.e2m2@univ-lyon1.fr**	M. Philippe NORMAND UMR 5557 Lab. d'Ecologie Microbienne Université Claude Bernard Lyon 1 Bâtiment Mendel 43, boulevard du 11 Novembre 1918 69 622 Villeurbanne CEDEX **philippe.normand@univ-lyon1.fr**
EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTÉ** **http://www.ediss-lyon.fr** Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 INSA : M. LAGARDE **secretariat.ediss@univ-lyon1.fr**	Mme Emmanuelle CANET-SOULAS INSERM U1060, CarMeN lab, Univ. Lyon 1 Bâtiment IMBL 11 Avenue Jean CAPELLE INSA de Lyon 69 621 Villeurbanne Tél : 04.72.68.49.09 Fax : 04.72.68.49.16 **emmanuelle.canet@univ-lyon1.fr**
INFOMATHS	**INFORMATIQUE ET MATHÉMATIQUES** **http://edinfomaths.universite-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage Tél : 04.72.43.80.46 **infomaths@univ-lyon1.fr**	M. Luca ZAMBONI Bât. Braconnier 43 Boulevard du 11 novembre 1918 69 622 Villeurbanne CEDEX Tél : 04.26.23.45.52 **zamboni@maths.univ-lyon1.fr**
Matériaux	**MATÉRIAUX DE LYON** **http://ed34.universite-lyon.fr** Sec. : Stéphanie CAUVIN Tél : 04.72.43.71.70 Bât. Direction **ed.materiaux@insa-lyon.fr**	M. Jean-Yves BUFFIÈRE INSA de Lyon MATEIS - Bât. Saint-Exupéry 7 Avenue Jean CAPELLE 69 621 Villeurbanne CEDEX Tél : 04.72.43.71.70 Fax : 04.72.43.85.28 **jean-yves.buffiere@insa-lyon.fr**
MEGA	**MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE** **http://edmega.universite-lyon.fr** Sec. : Stéphanie CAUVIN Tél : 04.72.43.71.70 Bât. Direction **mega@insa-lyon.fr**	M. Jocelyn BONJOUR INSA de Lyon Laboratoire CETHIL Bâtiment Sadi-Carnot 9, rue de la Physique 69 621 Villeurbanne CEDEX **jocelyn.bonjour@insa-lyon.fr**
ScSo	**ScSo*** **http://ed483.univ-lyon2.fr** Sec. : Véronique GUICHARD INSA : J.Y. TOUSSAINT Tél : 04.78.69.72.76 **veronique.cervantes@univ-lyon2.fr**	M. Christian MONTES Université Lyon 2 86 Rue Pasteur 69 365 Lyon CEDEX 07 **christian.montes@univ-lyon2.fr**

*ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie