SECRECY AND COVERTNESS IN THE PRESENCE OF MULTI-CASTING, CHANNEL STATE INFORMATION, AND COOPERATIVE JAMMING

by

Hassan ZivariFard

APPROVED BY SUPERVISORY COMMITTEE:

Aria Nosratinia, Chair

Matthieu R. Bloch, Co-Chair

John P. Fonseka

Yiorgos Makris

Hlaing Minn

To my family.

SECRECY AND COVERTNESS IN THE PRESENCE OF MULTI-CASTING, CHANNEL STATE INFORMATION, AND COOPERATIVE JAMMING

by

HASSAN ZIVARIFARD, BS, MS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

December 2021

ACKNOWLEDGMENTS

I owe my sincere gratitude to my PhD advisors Professor Aria Nosratinia and Professor Matthieu R. Bloch for their wisdom and endless support. I appreciate the valuable remarks of my PhD committee members, Professors John P. Fonseka, Yiorgos Makris, Hlaing Minn, and Massimo Fischetti. I am thankful to my colleagues at Multimedia Communications Lab: Ahmed Hindy, Mohamed Fadel Shady, Ahmed Attia Abotabl, Hussein Saad, Heping Wan, Mohammad Esmaeili, Mehdi Karbalayghareh, Pinky Kapoor, Javad Zahedi Moghaddam, Negar Daryanavardan, Sameer Dhole, and also my friends at UTD tennis club. I appreciate Noha Helal and Fan Zhang for all the helpful discussions.

November 2021

SECRECY AND COVERTNESS IN THE PRESENCE OF MULTI-CASTING, CHANNEL STATE INFORMATION, AND COOPERATIVE JAMMING

Hassan ZivariFard, PhD
The University of Texas at Dallas, 2021

Supervising Professor: Aria Nosratinia, Chair

We study secret communication over multi-transmitter multicast problem in the presence of an eavesdropper, wherein weak and strong secrecy regimes are studied. For the weak secrecy regime, the method of Chia and El Gamal is extended to two transmitters. We show that the achievable region calculated for the weak secrecy regime in this channel configuration is no bigger than the one calculated under strong secrecy. Two examples are presented in which the inner and outer bounds of secrecy region meet. In the process, we also characterize the minimum amount of randomness necessary to achieve secrecy in the multiple-access wiretap channel.

We consider the problem of covert communication over a state-dependent channel when the Channel State Information (CSI) is available either non-causally, causally, or strictly causally, either at the transmitter alone, or at both transmitter and receiver. In contrast to previous work, we do not assume the availability of a large shared key at the transmitter and legitimate receiver. Instead, we only require a secret key with negligible rate to bootstrap the communication and our scheme extracts shared randomness from the CSI in a manner that keeps it secret from the warden, despite the influence of the CSI on the warden's output. When CSI is available at the transmitter and receiver, we derive the covert capacity region. When CSI is only available at the transmitter, we derive inner and outer bounds on the

covert capacity. We also provide examples for which the covert capacity is positive with knowledge of CSI but is zero without it.

We consider the problem of covert communication in the presence of a cooperative jammer. It is known that in general, a transmitter and a receiver can communicate only $O(\sqrt{n})$ covert bits over n channel uses, i.e., zero rate. Here, we show that a cooperative jammer can facilitate the communication of positive covert rates, subject to the presence of friendly jammer in the environment. We consider various scenarios in which it is possible to achieve positive rate for covert communication. For these scenarios, we derive inner and outer bounds on the covert capacity region, and also we characterize the covert capacity region for some of these scenarios.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Communication systems usually separate error correction and data encryption. The former is usually realized at the physical layer by transforming the noisy communication channel into a reliable channel. The data encryption is implemented in the higher layers of the Open System Interconnection (OSI) model by applying cryptographic principles. The encryption approach is based on computational limits of the eavesdropper. However, the looming prospect of quantum computers would boost the computational power, which makes some critical cryptosystems insecure and weakens others. Post quantum computer cryptography approaches offers partial solutions, which rely on larger keys, but this is an expensive resource and substantial efforts are made to avoid extravagant use of it. Nonetheless, cryptography will remain the main practical tool for securing data, at least for the time being.

Physical Layer Security, which is based on information-theoretic principles, is an alternative approach for provably secure communication, pioneerd by Wyner's celebrated paper on the wiretap channel (WTC) [1]. Essentially, Wyner's main idea was to exploit the channel noise together with proper physical layer coding to secure the communication against an eavesdropper with unlimited compuational power. This dissertation takes the latter approach.

1.1 Two-Multicast Channel with Confidential Messages

We study the multiuser secure multicast problem (Fig. 1.1), more specifically, when two transmitters multicast messages securely to two receivers in the presence of an eavesdropper. All senders, receivers, and eavesdropper are at different terminals. This problem is motivated in part by secure access of multiple users to data in a distributed cache [2, 3]. Another application of the considered model is a common situation in cellular networks, in which

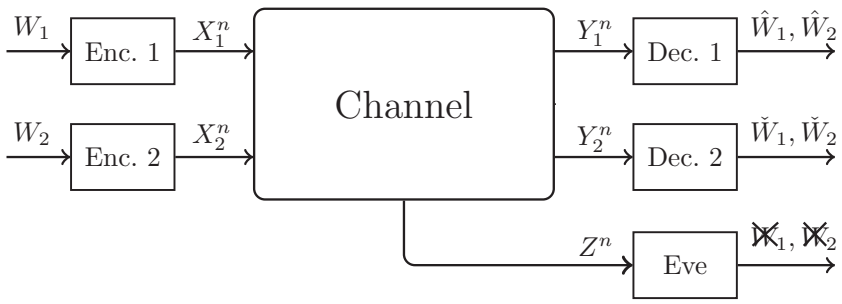

Figure 1.1. Two-sender, two-receiver channel with an eavesdropper

a user is in the coverage range of two different base stations [4, 5]. This problem is also equivalent to a one-transmitter, two-receiver compound channel with confidential messages with two different states [6]. It has been known [7] that problems involving compound channels have an equivalent multicast representation, in which the channel to each multicast receiver is equivalent to one of the states of the compound channel.[1]

In this dissertation, we take a two-pronged approach to the analysis of the network mentioned above, producing a number of new results and insights. In Section 2.3, we present an analysis inspired by the work of Chia and El Gamal [11], which uses Marton coding and indirect decoding (also known as non-unique decoding) [12] to achieve an improved secrecy rate for the transmission of *one* common message to two receivers that may experience different channel statistics. In extending the method of Chia and El Gamal to multiple transmitters, we introduce a two-level Marton-type coding with associated non-unique decoding.

In Section 2.6, we employ the method of output statistics of random binning (OSRB) [13] for analyzing the two-transmitter two-receiver problem (see also [14] for a related approach). OSRB analyzes channel coding problems by conversion to a related source coding problem, where it tests achievability by probability approximation rather than counting arguments on typical sets, followed by a reverse conversion to complete the analysis. OSRB is well suited

[1]The problem studied herein is the secrecy counterpart of the classical problem posed by Ahlswede [8], which proved highly influential for the MAC channel [9] and the interference channel [10].

for secrecy problems because secrecy is tightly related to probability approximation. OSRB encoding is purely by random binning and is enabled by (and named after) the following asymptotic result: apply two independent random binning schemes on the same set and take a random sample from the set. The two bin indices corresponding to the random sample are statistically independent as long as binning rates are sufficiently small [15, 13, 14]. We extend the tools and techniques of OSRB to match the requirements of the two-transmitter multicast problem.

The different parts of our results complement each other, producing a more complete picture in the understanding of the problem of multi-transmitter secure multicast. The extension of the method of Chia and El Gamal is utilized to highlight the minimal amount of randomness required to achieve secrecy rates over the multiple-access wiretap channel, and that therein channel prefixing can be replaced with superposition, in a manner reminiscent of Watanabe and Oohama [16] for minimizing the randomness resources for secrecy encoding. The analysis based on OSRB generates the strong secrecy, which interestingly has an expression that is a superset of the *achievable* region under weak secrecy calculated in the first part. Furthermore, the expression for the strong secrecy region can be greatly simplified via a constraint found in the weak secrecy analysis, highlighting the synergy between the two. More broadly, the developments in these two parts each offer techniques and insights that can potentially be useful in a wider class of problems.

Outer bounds for degraded and non-degraded channels are derived and shown to be tight against inner bounds in some special cases.

A brief outline of the related literature is as follows. Multicasting with common information in the presence of an eavesdropper has been studied in [17, 18], deriving inner bounds on the secrecy capacity, and in some special cases also deriving the secrecy capacity region. Salehkalaibar *et al.* [17] studied a one-receiver, two-eavesdropper broadcast channel with three degraded message sets. Ekrem and Ulukus [18] studied the transmission of public and

confidential messages to two legitimate users, in the presence of an eavesdropper. Benammar and Piantanida [19] calculated the secrecy capacity region of some classes of wiretap broadcast channels.

The MAC wiretap channel has been investigated in [20, 21, 22, 23, 24, 25, 26, 27]. In [20], a discrete memoryless MAC with confidential messages has been studied that consists of a MAC with generalized feedback [28] where each user's message must be kept confidential from the other. The multiple access wiretap channel [21, 22, 26] consists of a MAC with an additional channel output to an eavesdropper. In [21, 22], achievable rate regions for the secrecy capacity region have been derived. Secrecy in the interference channel and broadcast channel has been studied in [29], where inner and outer bounds for the broadcast channel with confidential messages and the interference channel with confidential messages have been compared.

1.2 Keyless Covert Communication via CSI

Covert communication is a mode of secrecy in which not just the content of communication, but also the act of communication, is kept secret from an adversary. More precisely, reliable communication over one channel must occur while simultaneously ensuring that another channel output, at a node called *the warden*, has a distribution identical to that induced by an innocent channel symbol [30, 31, 32, 33, 34]. It is known that in a DMC without state, the number of bits that can be reliably and covertly communicated over n channel transmissions scales at most as $O(\sqrt{n})$.[2] This result has motivated the study of other models in which positive rates are achievable [35, 36]. Of particular relevance to this dissertation, Lee *et al.* [37] have considered the problem of covert communication over a state-dependent channel in which the CSI is known either causally or non-causally to the transmitter but unknown

[2]Except for the special case when the output distribution (at the warden) induced by the innocent symbol is a convex combination of the output distributions generated by the other input symbols [32].

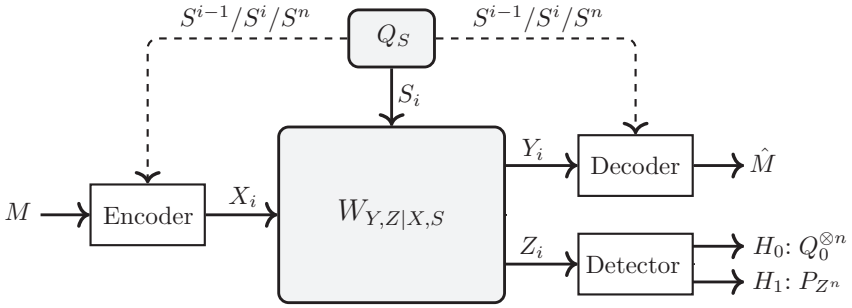

Figure 1.2. Model of covert communication over a state-dependent DMC with CSI available at both the transmitter and the receiver

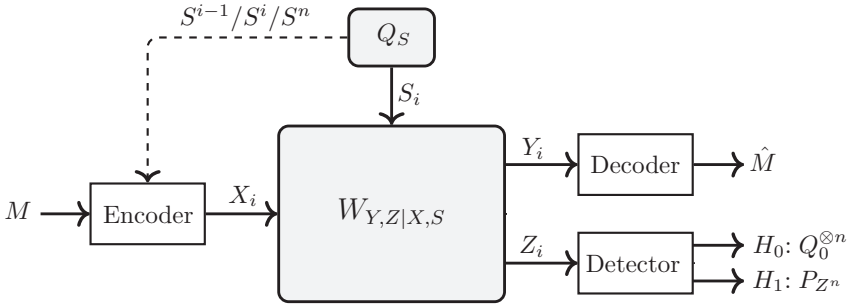

Figure 1.3. Model of covert communication over a state-dependent DMC with CSI only available at the transmitter

to the receiver and the warden. They derived the covert capacity when the transmitter and the receiver share a sufficiently long secret key, as well as a lower bound on the minimum secret key length needed to achieve the covert capacity. Since the presence of CSI provides a natural source of randomness from which to extract secret keys, one may wonder if covert communication with positive rate is possible without requiring an external secret key. This dissertation offers conclusive answers to this question in several scenarios.

The usefulness of exploiting CSI for secrecy has been extensively investigated in the context of state-dependent wiretap channels. A discrete memoryless wiretap channel with random states known non-causally at the transmitter was first studied by Chen and Vinck [38], who established a lower bound on the secrecy capacity based on a combination of wiretap coding with Gel'fand-Pinsker coding. Generally speaking, coding schemes with CSI outperform those without CSI because perfect knowledge of the CSI not only enables the

transmitter to align its signal toward the legitimate receiver but also provides a source of common randomness from which to generate a common secret key and enhance secrecy rates. Khisti *et al.* [39] studied the problem of secret key generation from non-causal CSI available at the transmitter and established inner and outer bounds on the secret key capacity. Chia and El Gamal [40] studied a wiretap channel in which the state information is available causally at both transmitter and receiver, proposing a scheme in which the transmitter and the receiver extract a weakly secret key from the state and protect the confidential message via a one-time-pad driven with the extracted key (see also [41] and [42]). Han and Sasaki [43] subsequently extended this result to strong secret keys. Goldfeld et al. [44] proposed a superposition coding scheme for the problem of transmitting a semantically secure message over a state-dependent channel with CSI available non-causally at the transmitter. In the context of covert communications, several works have demonstrated the benefits of exploiting common randomness and CSI to generate secret keys. For instance, stealth secret key generation from correlated sources was studied by Lin *et al.* [45, 46] and covert secret key generation was studied by Tahmasbi and Bloch [47, 48]. We note that covert communication over a compound channel was studied by Salehkalaibar *et al.* [49], although the objective therein is to mask the state of the compound channel and not to exploit CSI.

This dissertation studies covert communication over a state-dependent discrete memoryless channel with CSI available either non-causally, causally, or strictly causally, either at both the transmitter and the receiver or at the transmitter alone (see Fig. 1.2). One of the main contributions of this dissertation is to show that the CSI can be used to simultaneously and efficiently accomplish two necessary tasks: using the CSI for a Shannon strategy or Gel'fand-Pinsker coding, while also extracting a shared secret key at the two legitimate terminals to resolve the multiple codebooks that are necessary for covert communication. Secret key extraction from CSI replaces the external secret key in other models, thus potentially generalizing and expanding the applicability of covert communication. Our scheme requires

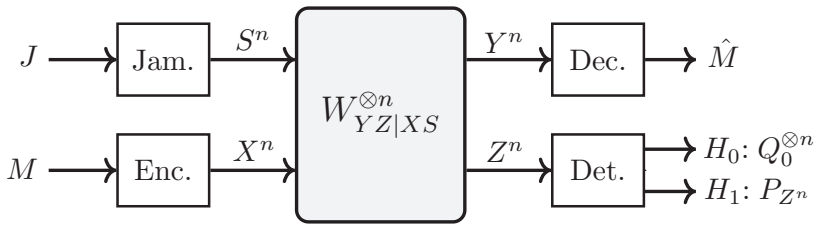

Figure 1.4. Covert Communications with a Cooperative Jammer

the transmitter and the receiver to share a secret key with negligible rate to bootstrap the communication. This bootstrapping is common in many security schemes, for instance in all schemes for secret communication based on seeded invertible extractors [50, 51, 52]. =With a slight abuse of terminology, we refer to our model as "keyless" instead of "asymptotically keyless."

Specifically, we characterize the exact covert capacity when CSI is available at both the transmitter and the receiver, and derive inner and outer bounds on the covert capacity when CSI is only available at the transmitter. For some channel models for which the covert capacity is zero without CSI, we show that the covert capacity is positive with CSI. The code constructions behind our proofs combine different coding mechanisms, including channel resolvability for covertness, channel randomness extraction for key generation, and Gel'fand-Pinsker coding for state-dependent channels. The key technical challenge consists in properly combining these mechanisms to ensure the overall covertness of the transmission through block-Markov chaining schemes.

1.3 Covert Communication via Cooperative Jamming

Next, we study the problem of covert communication over a DMC when a cooperative jammer [22] is present. Earlier results [53, 54, 35, 55, 56, 57, 58] have shown that it is possible to achieve positive covert rate when the warden has uncertainty about the power of noise or interference at its receiver.

In this dissertation, we consider four main jamming scenarios. First, we consider a scenario in which there is a jammer that has no cooperation with the transmitter and transmits codewords independent of the transmitter's codewords over the channel. In the second scenario, there is a shared secret key between all the legitimate parties (i.e., the transmitter, the receiver and the jammer). For this problem, we derive general inner and outer bounds on the covert capacity, and we characterize the covert capacity when the jammer has an unlimited source of local randomness. Third, we consider a scenario in which the jammer's output is available non-causally or causally at the transmitter and there is a shared secret key between the transmitter and the receiver. Since the jammer can simulate a random state (given sufficient resources), we expect to achieve a positive covert communication rate in the considered model. For each causal and non-causal case, an achievable rate region is calculated that highlights the relation between the covert communication rate, jammer's randomness (expressed as a rate), and rate of the needed shared secret key between transmitter and receiver. We also derive an upper bound for each of these cases. In the fourth jamming scenario, the transmitter's channel input is available non-causally, causally, or strictly causally available at the jammer. For each of these cases we characterize the covert capacity when the jammer has an unlimited source of local randomness and derive inner and outer bounds on the covert capacity otherwise.

Of particular relevance to this dissertation, arbitrarily varying wiretap channels under strong and semantic secrecy criterion have been studied in [59, 60, 61] and Covert communication over adversarially jammed channels has been studied in [62]. MAC with cribbing encoders was first studied by Willems and van der Meulen [63, 64] and channel resolvability and strong secrecy for a discrete memoryless multiple-access channel with cribbing has been studied in [65].

1.4 Preliminaries

Throughout this dissertation, random variables are denoted by capital letters and their realizations by lower case letters. The set of $\epsilon-$strongly jointly typical sequences of length n, according to $p_{X,Y}$, is denoted by $\mathcal{T}_\epsilon^{(n)}(p_{X,Y})$. For convenience in notation, whenever there is no danger of confusion, typicality will reference the random variables rather than the distribution, e.g., $\mathcal{T}_\epsilon^{(n)}(X,Y)$. The set of sequences $\{x^n : (x^n, y^n) \in T_\epsilon^{(n)}(X,Y)\}$ for a fixed y^n, when the fixed sequence y^n is clear from the context, is denoted with the shorthand notation $\mathcal{T}_\epsilon^{(n)}(X|Y)$. Superscripts denote the dimension of a vector, e.g., X^n. The integer set $\{1,\ldots,M\}$ is denoted by $[\![1,M]\!]$, and $X_{[i:j]}$ indicates the set $\{X_i, X_{i+1},\ldots,X_j\}$. The cardinality of a set is denoted by $|\cdot|$. Following Cuff [66] and [13, Remark 1], we use the concept of random Probability Mass Function (PMF) denoted by capital letters (e.g. P_X). $\mathbb{N}$ is the set of natural numbers, which does not include 0, while $\mathbb{R}$ denotes the set of real numbers. We define $\mathbb{R}_+ = \{x \in \mathbb{R} | x \geq 0\}$ and $\mathbb{R}_{++} = \mathbb{R}_+ \backslash \{0\}$. $\mathbb{E}_X(\cdot)$ is the expectation w.r.t. the random variable X and $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function. The cardinality of a set is denoted by $|\cdot|$. The total variation between PMF P and PMF Q is defined as, $||P-Q||_1 = \frac{1}{2}\sum_x |P(x)-Q(x)|$ and the Kullback-Leibler (KL) divergence between PMFs is defined as $\mathbb{D}(P||Q) = \sum_x p(x)\log\frac{P(x)}{Q(x)}$. The support of a probability distribution P is denoted by $\text{supp}(P)$. The n-fold product distribution constructed from the same distribution P is denoted $P^{\otimes n}$. Throughout the dissertation, log denotes the base 2 logarithm. For a set of random variables $\{X_i\}_{i\in\mathcal{A}}$ indexed over a countable set $\mathcal{A}$, $\mathbb{E}_{\backslash i}(\cdot)$ is the expectation with respect to all the random variables in $\mathcal{A}$ except the one with index $i \in \mathcal{A}$.

Finally, we recall a useful result about the relation between the total variation distance and the KL-divergence.

Lemma 1 (Reverse Pinsker's Inequality [67, eq. (323)])**.** *Pinsker's inequality indicates for two arbitrary distributions P and Q on the alphabet $\mathcal{A}$ we have,*

$$||P-Q||_1 \leq \sqrt{\frac{1}{2}\mathbb{D}(P||Q)}. \tag{1.1}$$

A reverse inequality is valid when the alphabet $\mathcal{A}$ is finite. Let P and Q be two arbitrary distributions on a finite alphabet set $\mathcal{A}$ such that P is absolutely continuous with respect to Q. If $\mu \triangleq \min_{a \in \mathcal{Q}: Q(a)>0} Q(a)$, we have,

$$\mathbb{D}(P||Q) \leq \log\left(\frac{1}{\mu}\right) ||P - Q||_1. \tag{1.2}$$

CHAPTER 2

TWO-MULTICAST CHANNEL WITH CONFIDENTIAL MESSAGES[1]

2.1 Introduction

In this chapter, we analyze secrecy rates for a channel in which two transmitters simultaneously multicast to two receivers in the presence of an eavesdropper. Achievable rates are calculated via extensions of a technique due to Chia and El Gamal and the method of output statistics of random binning. Outer bounds are derived for both the degraded and non-degraded versions of the channel, and examples are provided in which the inner and outer bounds meet. The inner bounds recover known results for the multiple-access wiretap channel, broadcast channel with confidential messages, and the compound MAC channel. An auxiliary result is also produced that derives an inner bound on the minimal randomness necessary to achieve secrecy in multiple-access wiretap channels.

2.2 Problem Statement

Definition 1. *A $(M_{1,n}, M_{2,n}, n)$ code for the considered model (Fig. 1.1) consists of the following:*

i) Two message sets $\mathcal{W}_i = [\![1, M_{i,n}]\!]$, $i = 1, 2$, from which independent messages W_1 and W_2 are drawn uniformly distributed over their respective sets.

ii) Stochastic encoders f_i, $i = 1, 2$, which are specified by conditional probability matrices $f_i(X_i^n|w_i)$, where $X_i^n \in \mathcal{X}_i^n$, $w_i \in \mathcal{W}_i$ are channel inputs and private messages, respectively, and $\sum_{x_i^n} f_i(x_i^n|w_i) = 1$. Here, $f_i(x_i^n|w_i)$ is the probability of the encoder producing the codeword x_i^n for the message w_i.

[1] ©2021 IEEE. Reprinted, with permission, from H. ZivariFard, M. R. Bloch, and A. Nosratinia, "Two multicast channel with confidential messages," 2021 IEEE Transactions on Information Forensics and Security, 2021, pp. 2743-2758

iii) A decoding function $\phi_1 : \mathcal{Y}_1^n \to \mathcal{W}_1 \times \mathcal{W}_2$ *that assigns* $(\hat{w}_1, \hat{w}_2) \in [\![1, M_{1,n}]\!] \times [\![1, M_{2,n}]\!]$ *to the received sequence* y_1^n.

iv) A decoding function $\phi_2 : \mathcal{Y}_2^n \to \mathcal{W}_1 \times \mathcal{W}_2$ *that assigns* $(\check{w}_1, \check{w}_2) \in [\![1, M_{1,n}]\!] \times [\![1, M_{2,n}]\!]$ *to the received sequence* y_2^n.

The probability of error is given by:

$$P_e \triangleq \mathbb{P}\big(\{(\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)\} \cup \{(\check{W}_1, \check{W}_2) \neq (W_1, W_2)\}\big).$$

Definition 2. *A rate pair* (R_1, R_2) *is said to be achievable if there exists a sequence of* $(M_{1,n}, M_{2,n}, n)$ *codes with* $M_{1,n} \geq 2^{nR_1}, M_{2,n} \geq 2^{nR_2}$, *so that* $P_e \underset{n\to\infty}{\longrightarrow} 0$ *and [68]*

$$\frac{1}{n}\mathbb{I}(W_1, W_2; Z^n) \underset{n\to\infty}{\longrightarrow} 0 \quad \textit{for the weak secrecy regime,} \tag{2.1}$$

$$\mathbb{I}(W_1, W_2; Z^n) \underset{n\to\infty}{\longrightarrow} 0 \quad \textit{for the strong secrecy regime.} \tag{2.2}$$

Definition 3. *For any PMFs* p_X *and* q_X *over* $\mathcal{X}$ *we denote* $\|p_X - q_X\|_1 < \epsilon$ *with* $p_X \approx_\epsilon q_X$. *Similarly, for any random PMFs* P_X *and* Q_X *over,* $\mathcal{X}$ *we denote* $\|P_X - Q_X\|_1 < \epsilon$ *with* $P_X \approx_\epsilon Q_X$. *The same notation applies for the sequential PMFs (i.e.* $p_{X^n} \approx_\epsilon q_{X^n}$ *if* $\|p_{X^n} - q_{X^n}\|_1 < \epsilon$*).*

2.3 Achievable Rate Region Under Weak Secrecy

We start with a lemma that fits Marton coding with indirect decoding in a MAC structure and produces an entropy bound needed in the secrecy analysis. Its basic idea can be highlighted as follows: given X^n, if we *independently* produce 2^{nR} random code vectors Y^n, we will have approximately $2^{nR-\mathbb{I}(X^n;Y^n)}$ *jointly* typical pairs, i.e., the "excess" rate will determine the number of jointly typical pairs. This lemma extends the basic idea of excess rate to multiple codebooks, multiple conditioning, and furthermore, a generalization is made from a counting argument to the entropy of the index of the codebook, which is essential for the subsequent secrecy analysis.

Lemma 2. *Consider random variables* $(Q, U_0, V_0, U_1, V_1, Z)$ *distributed according to* $p_Q p_{U_0,U_1|Q} p_{V_0,V_1|Q} p_{Z|U_0,U_1,V_0,V_1}$. *Draw random sequences* Q^n, U_0^n, V_0^n *according to* $\prod_{i=1}^n p_Q(q_i)\ p_{U_0|Q}(u_{0,i}|q_i)\ p_{V_0|Q}(v_{0,i}|q_i)$. *Conditioned on* U_0^n, *draw* 2^{nS} *i.i.d. copies of* U_1^n *according to* $\prod_{i=1}^n p_{U_1|U_0}(u_{1,i}|u_{0,i})$, *denoted* $U_1^n(\ell)$, $\ell \in [\![1, 2^{nS}]\!]$. *Similarly, conditioned on* V_0^n, *draw* 2^{nT} *i.i.d. copies of* V_1^n *according to* $\prod_{i=1}^n p_{V_1|V_0}(v_{1,i}|v_{0,i})$, *denoted* $V_1^n(k)$, $k \in [\![1, 2^{nT}]\!]$. *Let* $L \in [\![1, 2^{nS}]\!]$ *and* $K \in [\![1, 2^{nT}]\!]$ *be random variables with arbitrary PMF. If*

$$S > \mathbb{I}(U_1; Z|Q, U_0, V_0) + \delta_1(\epsilon)$$

$$T > \mathbb{I}(V_1; Z|Q, U_0, V_0) + \delta_1(\epsilon)$$

$$S + T > \mathbb{I}(U_1, V_1; Z|Q, U_0, V_0) + \delta_1(\epsilon)$$

for a positive $\delta_1(\epsilon)$ *and if for an arbitrary sequence* Z^n,

$$\mathbb{P}\big((Q^n, U_0^n, V_0^n, U_1^n(L), V_1^n(K), Z^n) \in \mathcal{T}_\epsilon^{(n)}\big) \underset{n\to\infty}{\longrightarrow} 1, \tag{2.3}$$

there exists a positive $\delta_2(\epsilon) \underset{\epsilon\to 0}{\longrightarrow} 0$, *such that for* n *sufficiently large*

$$\mathbb{H}(L, K|Q^n, U_0^n, V_0^n, Z^n, \mathcal{C}) \leq n(S + T - \mathbb{I}(U_1, V_1; Z|Q, U_0, V_0)) + n\delta_2(\epsilon), \tag{2.4}$$

where $\mathcal{C} = \{U_1^n(1), \ldots, U_1^n(2^{nS}), V_1^n(1), \ldots, V_1^n(2^{nT})\}$.

The proof is provided in Appendix A. This result is related to, and contains, [11, Lemma 1]. In particular, [11] considers a single-input channel and explores the properties of codebooks driven by this input, while observing an output Z. In contrast, this dissertation's Lemma 2 develops a corresponding result for a *multiple-access channel* with respect to Z, motivated by the two-transmitters present in the model of this dissertation. This accounts for the new features of our Lemma 2, namely three rate constraints instead of one, as well as monitoring the entropy of two index random variables instead of one. Furthermore, the present result has one additional layer of conditioning to allow for indirect decoding of multiple confidential messages in the sequel, while in [11] only one confidential message is decoded.

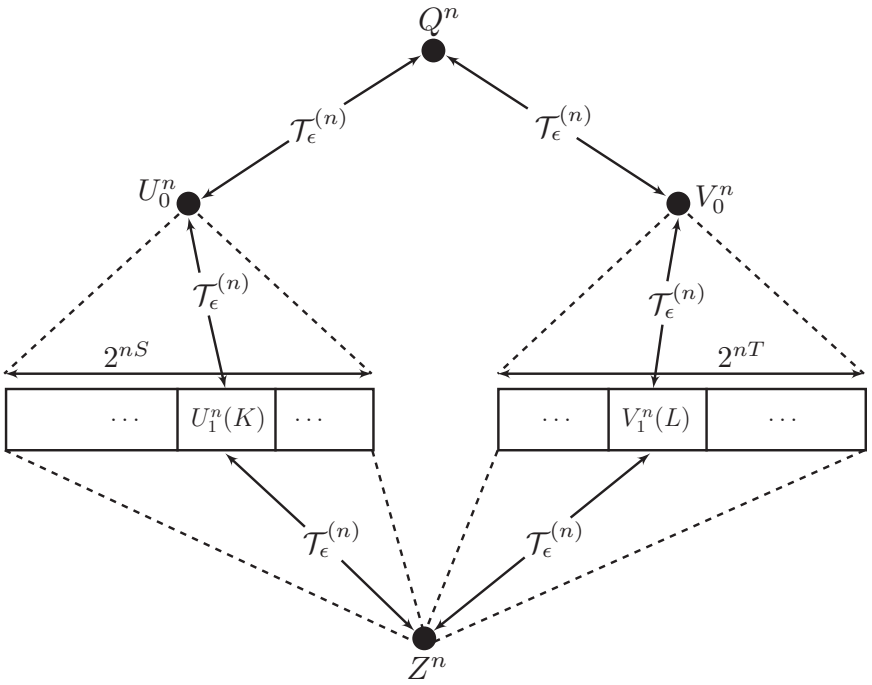

Figure 2.1. Structure of Lemma 2: subject to jointly typical sequences $(Q^n, U_0^n, V_0^n, U_1^n(K), V_1^n(L), Z^n)$, finding a bound on the conditional entropy of (K, L), thus implicitly bounding the number of sequence pairs that can be jointly typical with (Q^n, Z^n) from codebooks with certain size.

Remark 1. *In addition to establishing the main results of this dissertation, Lemma 2 also has broader implications on the necessity of prefixing in multi-transmitter secrecy problems [69] and deriving the minimum amount of randomness needed to achieve secrecy. Csiszár and Körner introduced prefixing in [70] to expand the achievable rate region of the non-degraded broadcast channel with confidential messages, a technique that was subsequently used in essentially the same manner in multi-transmitter settings. Subsequently, Chia and El Gamal showed that in a single-transmitter wiretap channel, prefixing can be replaced with superposition coding [11]. Appendix B extends this concept to a multi-transmitter setting and presents an achievability technique for the multiple access wiretap channel that utilizes minimal randomness and matches the best known achievable rates without prefixing.*

Theorem 1. *An inner bound on the secrecy capacity region of the two-transmitter two-receiver channel with confidential messages is given by the set of non-negative rate pairs* (R_1, R_2) *such that*

$$
\begin{aligned}
R_1 &< \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) - \mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_1; Z|U_0, V_0) \\
R_1 &< \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) - \mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_2; Z|U_0, V_0) \\
R_1 &< \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) - \mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_1, V_1; Z|U_0, V_0) \\
R_1 &< \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0) - \mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_2, V_2; Z|U_0, V_0) \\
R_2 &< \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) - \mathbb{I}(V_0; Z|Q) - \mathbb{I}(V_1; Z|U_0, V_0) \\
R_2 &< \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) - \mathbb{I}(V_0; Z|Q) - \mathbb{I}(V_2; Z|U_0, V_0) \\
R_2 &< \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) - \mathbb{I}(V_0; Z|Q) - \mathbb{I}(U_1, V_1; Z|U_0, V_0) \\
R_2 &< \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0) - \mathbb{I}(V_0; Z|Q) - \mathbb{I}(U_2, V_2; Z|U_0, V_0) \\
R_1 + R_2 &< \mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q) - \mathbb{I}(U_0, U_1, V_0, V_1; Z|Q) \\
R_1 + R_2 &< \mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q) - \mathbb{I}(U_0, U_2, V_0, V_2; Z|Q) \\
R_1 + R_2 &< \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) \\
&\quad - \mathbb{I}(U_0, U_1, V_0, V_1; Z|Q) - \mathbb{I}(U_1; Z|U_0, V_0) \\
R_1 + R_2 &< \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) \\
&\quad - \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1; Z|U_0, V_0) - \mathbb{I}(V_2; Z|U_0, V_0) \\
R_1 + R_2 &< \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0) \\
&\quad - \mathbb{I}(U_1; Z|U_0, V_0) - \mathbb{I}(U_0, U_2, V_0, V_2; Z|Q) \\
R_1 + R_2 &< \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) \\
&\quad - \mathbb{I}(U_0, U_1, V_0, V_1; Z|Q) - \mathbb{I}(V_1; Z|U_0, V_0) \\
R_1 + R_2 &< \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) \\
&\quad - \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(V_1; Z|U_0, V_0) - \mathbb{I}(U_2; Z|U_0, V_0)
\end{aligned}
$$

$$
\begin{aligned}
R_1 + R_2 &< \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0) \\
&\quad - \mathbb{I}(V_1; Z|U_0, V_0) - \mathbb{I}(U_0, U_2, V_0, V_2; Z|Q) \\
R_1 + R_2 &< \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) \\
&\quad - \mathbb{I}(U_0, U_1, V_0, V_1; Z|Q) - \mathbb{I}(U_2; Z|U_0, V_0) \\
R_1 + R_2 &< \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0) \\
&\quad - \mathbb{I}(U_0, U_2, V_0, V_2; Z|Q) - \mathbb{I}(U_2; Z|U_0, V_0) \\
R_1 + R_2 &< \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) + \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0) \\
&\quad - \mathbb{I}(U_0, U_2, V_0, V_2; Z|Q) - \mathbb{I}(V_2; Z|U_0, V_0) \\
R_1 + R_2 &< \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) + \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) \\
&\quad - \mathbb{I}(U_0, U_1, V_0, V_1; Z|Q) - \mathbb{I}(V_2; Z|U_0, V_0) \\
R_1 + R_2 &< \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) + \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) \\
&\quad - \mathbb{I}(U_0, V_0; Z|Q) - 2\mathbb{I}(U_1, V_1; Z|U_0, V_0) \\
R_1 + R_2 &< \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0) \\
&\quad - \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, V_1; Z|U_0, V_0) \\
&\quad - \mathbb{I}(U_2, V_2; Z|U_0, V_0) \\
R_1 + R_2 &< \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) + \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0) \\
&\quad - \mathbb{I}(U_0, U_1, V_0, V_1; Z|Q) - \mathbb{I}(U_2, V_2; Z|U_0, V_0) \\
R_1 + R_2 &< \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0) \\
&\quad - \mathbb{I}(U_0, V_0; Z|Q) - 2\mathbb{I}(U_2, V_2; Z|U_0, V_0)
\end{aligned}
$$

for some

$$
\begin{aligned}
&p(q)p(u_0|q)p(u_1, u_2|u_0)p(v_0|q)p(v_1, v_2|v_0) \\
&p(x_1|u_0, u_1, u_2)p(x_2|v_0, v_1, v_2)p(y_1, y_2, z|x_1, x_2),
\end{aligned} \tag{2.5}
$$

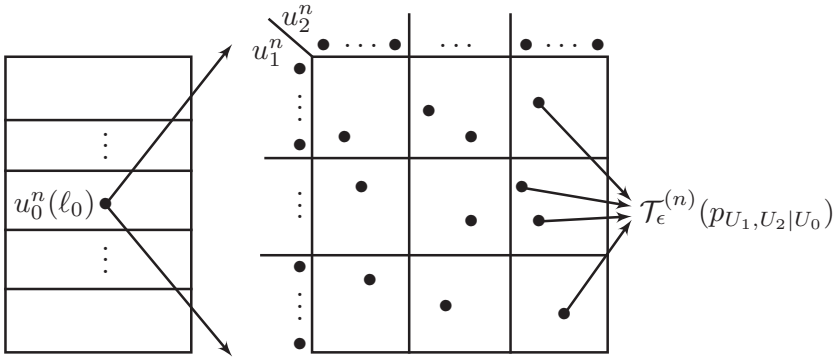

Figure 2.2. Coding scheme for the first transmitter

such that

$$\begin{aligned}\mathbb{I}(U_1,U_2,V_1,V_2;Z|U_0,V_0) &\le \mathbb{I}(U_1,V_1;Z|U_0,V_0)\\ &+\mathbb{I}(U_2,V_2;Z|U_0,V_0)-\mathbb{I}(U_1;U_2|U_0)-\mathbb{I}(V_1;V_2|V_0).\end{aligned}\tag{2.6}$$

The proof uses superposition coding, Wyner's wiretap coding, Marton coding, as well as indirect decoding. The details of the proof are provided in Appendix C.

This strategy was introduced in [11] to find an achievable rate region for broadcast channel with confidential messages. In this dissertation, this scheme has been extended to two transmitters. The codebook for the first transmitter is shown in Fig. 2.2. For the first transmitter, the message w_1 is represented by u_0^n codewords. Then we superimpose a Marton codebook consist of U_1^n and U_2^n codewords on this u_0^n codeword and select a jointly typical pair (u_1^n, u_2^n) at random from this codebook. The codebook structure for the second transmitter is the same and the codewords generated at the second transmitter are represented by v_0^n, v_1^n, and v_2^n. The receiver j, for $j = 1, 2$, decodes w_1 through (u_0^n, u_j^n), and decodes w_2 through (v_0^n, v_j^n). For decoding (w_1, w_2) at the receiver j it is not necessary to ensure that there is not any error in decoding (u_j^n, v_j^n), the efficacy of using (u_j^n, v_j^n) without decoding them has been illuminated in [11]. Here, a two-step secrecy analysis is necessary because the $(u_{[1:2]}^n, v_{[1:2]}^n)$ codewords should not leak any information about (u_0^n, v_0^n). Therefore, the secrecy constraints

for u_0^n and v_0^n codewords should be derived first, and then secrecy constraints for $(u_{[1:2]}^n, v_{[1:2]}^n)$ codewords should be derived, assuming that the eavesdropper has access to (u_0^n, v_0^n, z^n). This two-step secrecy can be seen in Theorem 1; for example in the first constraint on R_1 the first negative term stands for the security of u_0^n and the second negative term stands for the security of u_1^n assuming that eavesdropper has access to (u_0^n, v_0^n, z^n).

This result covers several known earlier results:

- By setting $Z = \emptyset$, $U_0 = U_1 = U_2 = X_1$, and $V_0 = V_1 = V_2 = X_2$, the result in Theorem 1 reduces to the capacity region of compound multiple access channel discussed in [8].

- By setting $Y_2 = \emptyset$ (or $Y_1 = \emptyset$), $U_0 = U_1 = U_2 = X_1$ and $V_0 = V_1 = V_2 = X_2$, the result in Theorem 1 reduces to the achievable rate region of multiple access wiretap channel without common message [21, 22, 23].

- By setting $X_2 = \emptyset$ (or $X_1 = \emptyset$), $U_0 = U_1 = U_2$, and $Y_2 = \emptyset$ (or $Y_1 = \emptyset$), the result in Theorem 1 reduces to the capacity region of broadcast channel with confidential message [70, Corollary 2].

- By setting $X_2 = \emptyset$ (or $X_1 = \emptyset$), the result in Theorem 1 reduces to the achievable rate region for two-receiver, one-eavesdropper wiretap channel presented in [11, Theorem 1].

Remark 2. *By doing some algebraic manipulation, we can show that the constraint in* (2.6) *holds only if,*

$$\mathbb{I}(U_1, V_1; U_2, V_2 | U_0, V_0, Z) = 0. \tag{2.7}$$

Intuitively speaking, (2.7) *shows that the Marton coding codebooks remain independent even if the eavesdropper has access to the cloud centers.*

Corollary 1. *An inner bound on the secrecy capacity region of degraded two-transmitter two-receiver channel with confidential messages (Definition 4) is given by the set of non-negative rate pairs* (R_1, R_2) *such that*

$$R_1 \leq \mathbb{I}(U_0; Y_2|V_0, Q) - \mathbb{I}(U_0; Z|Q) \tag{2.8}$$

$$R_2 \leq \mathbb{I}(V_0; Y_2|U_0, Q) - \mathbb{I}(V_0; Z|Q) \tag{2.9}$$

$$R_1 + R_2 \leq \mathbb{I}(U_0, V_0; Y_2|Q) - \mathbb{I}(U_0; Z|Q) - \mathbb{I}(V_0; Z|Q) \tag{2.10}$$

for some

$$p(q)p(u_0|q)p(v_0|q)p(x_1|u_0)p(x_2|v_0). \tag{2.11}$$

Proof. The proof follows from Theorem 1 by setting $U_0 = U_1 = U_2$ and $V_0 = V_1 = V_2$ and considering the fact that the channel is degraded. □

2.4 An Outer Bound for the Degraded Model

We develop an outer bound for the degraded version of the model and provide an example in which it meets the inner bound of Theorem 1.

Definition 4. *The degraded two-transmitter two-receiver channel with confidential messages obeys:*

$$p(y_1, y_2, z|x_1, x_2) = p(y_1|x_1, x_2)p(y_2|y_1)p(z|y_2). \tag{2.12}$$

Theorem 2. *The secrecy capacity region for the degraded two-transmitter two-receiver channel with confidential messages is included in the set of rate pairs* (R_1, R_2) *satisfying*

$$R_1 \leq \mathbb{I}(U_0; Y_2|Q) - \mathbb{I}(U_0; Z|Q), \tag{2.13}$$

$$R_2 \leq \mathbb{I}(V_0; Y_2|Q) - \mathbb{I}(V_0; Z|Q), \tag{2.14}$$

$$R_1 + R_2 \leq \mathbb{I}(U_0, V_0; Y_2|Q) - \mathbb{I}(U_0, V_0; Z|Q), \tag{2.15}$$

for some joint distribution

$$p(q)p(u_0, v_0|q)p(x_1|u_0)p(x_2|v_0). \tag{2.16}$$

The details of the proof are provided in Appendix D.

Example (Degraded Switch Model): We consider an example of the two-transmitter two-receiver channel where the first legitimate receiver has access to the noisy version of each of the two transmitted values in a time-sharing (switched) manner, without interference from the other transmitter (Fig. 2.3). The second legitimate receiver has access to a noisy version of the first receiver, and the eavesdropper has access to a noisy version of the second receiver. This example illustrates a situation where in cellular networks, a user is in the coverage range of two different base stations and an eavesdropper has access a noisy version of the receiver's signal. The switch channel state information is made available to all terminals. In this model, the channel outputs are as follows:

$$y_1' = (y_1, s), \tag{2.17}$$

$$y_2' = (y_2, s), \tag{2.18}$$

$$z' = (z, s). \tag{2.19}$$

This model consists of a channel with states that are causally available at both the encoders and decoders.

The statistics of the channel, conditioned on the switch state, are expressed as follows:

$$p(y_1', y_2', z|x_1, , x_2, s) = p(y_1|x_1, x_2, s)\, p(y_2|y_1, s)\, p(z|y_2, s). \tag{2.20}$$

The switch model describes, e.g., frequency hopping over two frequencies [29]. The state (switch) is a binary random variable that chooses between listening to the Transmitter 1, with probability τ, and listening to the Transmitter 2, with probability $1-\tau$, independently at each time slot. We further assume the state is i.i.d. across time,

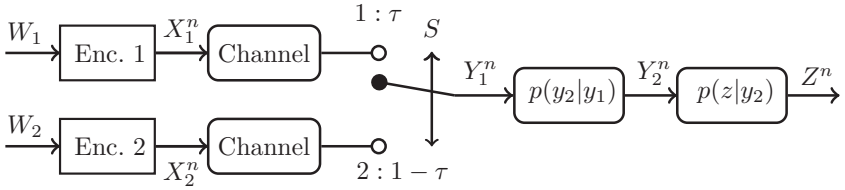

Figure 2.3. Degraded switch model

$$\begin{aligned} p(y_1|x_1,x_2,s) =& p(y_1|x_1)\mathbb{1}_{\{s=1\}} + p(y_1|x_2)\mathbb{1}_{\{s=2\}}, \\ =& p(y_1|x_s), \end{aligned} \tag{2.21}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function. Therefore, the channel model for degraded switch model is as follows

$$p(y_1,y_2,z|x,x,s) = p(y_1|x_s)p(y_2|y_1,s)p(z|y_2,s). \tag{2.22}$$

Theorem 3. *The secrecy capacity region for the degraded switch two-transmitter two-receiver channel with confidential messages, is given by the set of rate pairs* (R_1,R_2) *satisfying*

$$R_1 \leq \mathbb{I}(U_0;Y_2'|V_0,Q) - \mathbb{I}(U_0;Z'|Q), \tag{2.23}$$

$$R_2 \leq \mathbb{I}(V_0;Y_2'|U_0,Q) - \mathbb{I}(V_0;Z'|Q), \tag{2.24}$$

$$R_1+R_2 \leq \mathbb{I}(U_0,V_0;Y_2'|Q) - \mathbb{I}(U_0,V_0;Z'|Q), \tag{2.25}$$

for some joint distribution

$$p(q)p(u_0|q)p(v_0|q)p(x_1|u_0)p(x_2|v_0). \tag{2.26}$$

The details of the proof are provided in Appendix E.

2.5 A General Outer Bound

We now develop a general outer bound for the model of Fig. 1.1 and provide an example in which it meets the inner bound of Theorem 1.

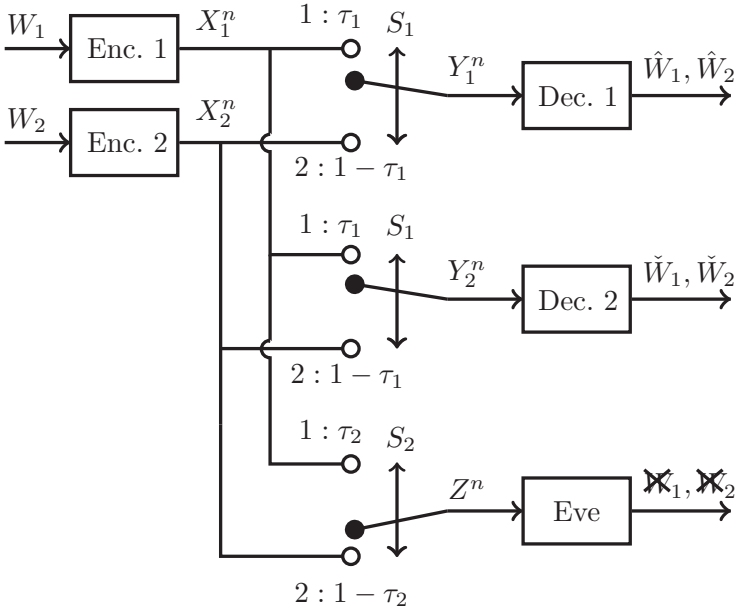

Figure 2.4. Noiseless switch model

Theorem 4. *The secrecy capacity region for the two-transmitter two-receiver channel with confidential messages is included in the set of rate pairs* (R_1, R_2) *satisfying*

$$R_1 \leq \mathbb{I}(U_0; Y_1, Y_2|Q) - \mathbb{I}(U_0; Z|Q), \tag{2.27}$$

$$R_2 \leq \mathbb{I}(V_0; Y_1, Y_2|Q) - \mathbb{I}(V_0; Z|Q), \tag{2.28}$$

$$R_1 + R_2 \leq \mathbb{I}(U_0, V_0; Y_1, Y_2|Q) - \mathbb{I}(U_0, V_0; Z|Q), \tag{2.29}$$

for some joint distribution

$$p(q)p(u_0, v_0|q)p(x_1|u_0)p(x_2|v_0). \tag{2.30}$$

The details of the proof are provided in Appendix F.

Example (Noiseless Switch Model): This example is motivated by two transmitters operating on different spectral bands, while the receiving terminals may receive adaptively on one band at a time [29]. The eavesdropper in our example has access to one noiseless, interference-free transmitted value at a time. Here, it is assumed that both legitimate receivers operate according to a common random switch s_1 that is connected to Transmitter 1

with probability τ_1 and to Transmitter 2 with probability $1-\tau_1$, and the eavesdropper operates according to another random switch s_2 that is connected to Transmitter 1 with probability τ_2 and to Transmitter 2 with probability $1-\tau_2$ (See Fig. 2.4). Aside from the switches, the channel is noiseless. Both receivers and the eavesdropper have access to their own switch state information. Therefore, the channel outputs are considered

$$y_1' = (y_1, s_1), \tag{2.31}$$

$$y_2' = (y_2, s_1), \tag{2.32}$$

$$z' = (z, s_2). \tag{2.33}$$

Since $y_1 = y_2$, we also have $y_1' = y_2'$.

Theorem 5. *The secrecy capacity region for the noiseless switch two-transmitter two-receiver channel with confidential messages is given by the set of rate pairs (R_1, R_2) satisfying*

$$R_1 \leq (\tau_1 - \tau_2)^+ \mathbb{H}(X_1), \tag{2.34}$$

$$R_2 \leq (\tau_2 - \tau_1)^+ \mathbb{H}(X_2), \tag{2.35}$$

where $(x)^+ = \max\{0, x\}$.

The details of the proof are provided in Appendix G. The capacity region in Theorem 5 shows that transmitters can securely communicate to receivers as long as $\tau_1 \neq \tau_2$.

2.6 Achievable Rate Region Under Strong Secrecy

Theorem 6. *An inner bound on the secrecy capacity region of the two-transmitter two-receiver channel with confidential messages is given by the set of non-negative rate pairs (R_1, R_2) such that*

$$R_1 < \mathbb{I}(U_0, U_1; Y_1 | Q, V_0, V_1) - \mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_1; Z|U_0, V_0)$$

$$R_1 < \mathbb{I}(U_0, U_2; Y_2 | Q, V_0, V_2) - \mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_2; Z|U_0, V_0)$$

$$R_1 < \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) - \mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_1, V_1; Z|U_0, V_0)$$

$$R_1 < \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0) - \mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_2, V_2; Z|U_0, V_0)$$

$$\begin{aligned} 2R_1 < {} & \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) \\ & - 2\mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_1, U_2; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0) \end{aligned}$$

$$\begin{aligned} 2R_1 < {} & \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0) \\ & - 2\mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_1, U_2, V_2; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0) \end{aligned}$$

$$\begin{aligned} 2R_1 < {} & \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) \\ & - 2\mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_1, U_2, V_1; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0) \end{aligned}$$

$$\begin{aligned} 2R_1 < {} & \mathbb{I}(U_0, U_1, V_1; Y_1|Q, U_0) + \mathbb{I}(U_0, U_2, V_2; Y_2|Q, U_0) \\ & - 2\mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0) \end{aligned}$$

$$R_2 < \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) - \mathbb{I}(V_0; Z|Q) - \mathbb{I}(V_1; Z|U_0, V_0)$$

$$R_2 < \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) - \mathbb{I}(V_0; Z|Q) - \mathbb{I}(V_2; Z|U_0, V_0)$$

$$R_2 < \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) - \mathbb{I}(V_0; Z|Q) - \mathbb{I}(U_1, V_1; Z|U_0, V_0)$$

$$R_2 < \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0) - \mathbb{I}(V_0; Z|Q) - \mathbb{I}(U_2, V_2; Z|U_0, V_0)$$

$$\begin{aligned} 2R_2 < {} & \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) \\ & - 2\mathbb{I}(V_0; Z|Q) - \mathbb{I}(V_1, V_2; Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0) \end{aligned}$$

$$\begin{aligned} 2R_2 < {} & \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0) \\ & - 2\mathbb{I}(V_0; Z|Q) - \mathbb{I}(U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0) \end{aligned}$$

$$\begin{aligned} 2R_2 < {} & \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) + \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) \\ & - 2\mathbb{I}(V_0; Z|Q) - \mathbb{I}(U_1, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0) \end{aligned}$$

$$\begin{aligned} 2R_2 < {} & \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0) \\ & - 2\mathbb{I}(V_0; Z|Q) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0) \end{aligned}$$

$$R_1 + R_2 < \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2)$$

$$- \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, V_2; Z|U_0, V_0)$$

$$R_1 + R_2 < \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0)$$
$$- \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; Z|U_0, V_0)$$

$$R_1 + R_2 < \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0)$$
$$- \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, U_2, V_2; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0)$$

$$R_1 + R_2 < \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1)$$
$$- \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_2, V_1; Z|U_0, V_0)$$

$$R_1 + R_2 < \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0)$$
$$- \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, U_2, V_1; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0)$$

$$R_1 + R_2 < \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0)$$
$$- \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_2; Z|U_0, V_0) - \mathbb{I}(U_2, V_2; Z|U_0, V_0)$$

$$R_1 + R_2 < \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0)$$
$$- \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(V_1; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; Z|U_0, V_0)$$

$$R_1 + R_2 < \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0)$$
$$- \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0)$$

$$R_1 + R_2 < \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) + \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0)$$
$$- \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0)$$

$$R_1 + R_2 < \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) + \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0)$$
$$- \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(V_2; Z|U_0, V_0) - \mathbb{I}(U_2, V_2; Z|U_0, V_0)$$

$$R_1 + R_2 < \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) + \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0)$$
$$- \mathbb{I}(U_0, V_0; Z|Q) - 2\mathbb{I}(U_1, V_1; Z|U_0, V_0)$$

$$R_1 + R_2 < \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0) - \mathbb{I}(U_0, V_0; Z|Q)$$
$$- \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$R_1 + R_2 < \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0) + \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) - \mathbb{I}(U_0, V_0; Z|Q)$$
$$- \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$
$$R_1 + R_2 < \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0)$$
$$- \mathbb{I}(U_0, V_0; Z|Q) - 2\mathbb{I}(U_2, V_2; Z|U_0, V_0)$$
$$R_1 + R_2 < \mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q) - \mathbb{I}(U_0, U_2, V_0, V_2; Z|Q)$$
$$R_1 + R_2 < \mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q) - \mathbb{I}(U_0, U_1, V_0, V_1; Z|Q)$$
$$2R_1 + R_2 < \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q)$$
$$- \mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, U_2, V_2; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0)$$
$$2R_1 + R_2 < \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + \mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q)$$
$$- \mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, U_2, V_1; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0)$$
$$2R_1 + R_2 < \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) + \mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q)$$
$$- \mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$
$$2R_1 + R_2 < \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0) + \mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q)$$
$$- \mathbb{I}(U_0; Z|Q) - \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$
$$R_1 + 2R_2 < \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q)$$
$$- \mathbb{I}(V_0; Z|Q) - \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0)$$
$$R_1 + 2R_2 < \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) + \mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q)$$
$$- \mathbb{I}(V_0; Z|Q) - \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0)$$
$$R_1 + 2R_2 < \mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0)$$
$$- \mathbb{I}(V_0; Z|Q) - \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$
$$R_1 + 2R_2 < \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) + \mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q)$$
$$- \mathbb{I}(V_0; Z|Q) - \mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$
$$2R_1 + 2R_2 < 2\mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) + \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - 2\mathbb{I}(U_1; Z|U_0, V_0) - \mathbb{I}(U_1, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0)$$

$$2R_1 + 2R_2 < 2\mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - 2\mathbb{I}(U_1; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2)$$

$$+ \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0) - 2\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, U_2; Z|U_0, V_0)$$

$$- \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + 2\mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) + \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - 2\mathbb{I}(V_2; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_2; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) + \mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1; Z|U_0, V_0) - \mathbb{I}(U_1, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) + \mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(V_2; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_2; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) + \mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0) + \mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + 2\mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - 2\mathbb{I}(V_1; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_1; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0)$$

$$2R_1 + 2R_2 < 2\mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - 2\mathbb{I}(U_2; Z|U_0, V_0) - \mathbb{I}(U_2, V_1, V_2 : Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0)$$

$$2R_1 + 2R_2 < 2\mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - 2\mathbb{I}(U_2; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(V_1; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_1; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_2; Z|U_0, V_0) - \mathbb{I}(U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) + \mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_2; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0) + \mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_2; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$2R_1 + 2R_2 < 2\mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) + \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - 2\mathbb{I}(V_1; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2)$$

$$+ \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) + \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0) - 2\mathbb{I}(U_0, V_0; Z|Q)$$

$$- \mathbb{I}(V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) + \mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(V_1; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0) + \mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(V_1; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$2R_1 + 2R_2 < 2\mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) + \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) + \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - 2\mathbb{I}(V_2; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) + \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) + \mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(V_2; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) + \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0) + \mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(V_2; Z|U_0, V_0) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$2R_1 + 2R_2 < \mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q) + \mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q)$$

$$- 2\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$3R_1 + 3R_2 < \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0)$$

$$+ 2\mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q) - 3\mathbb{I}(U_0, V_0; Z) - \mathbb{I}(U_1, U_2, V_2; Z|U_0, V_0)$$

$$- \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$3R_1 + 3R_2 < \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0)$$

$$+ 2\mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q) - 3\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, U_2, V_1; Z|U_0, V_0)$$

$$- \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$3R_1 + 3R_2 < \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0)$$

$$+ 2\mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q) - 3\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_2, V_1, V_2; Z|U_0, V_0)$$

$$- \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$3R_1 + 3R_2 < \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) + \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0)$$

$$+ 2\mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q) - 3\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, V_1, V_2; Z|U_0, V_0)$$

$$- \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0) - \mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$3R_1 + 3R_2 < \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) + \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0) + 2\mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q)$$

$$- 3\mathbb{I}(U_0, V_0; Z|Q) - 2\mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - 2\mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$3R_1 + 3R_2 < \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0) + \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0) + 2\mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q)$$

$$- 3\mathbb{I}(U_0, V_0; Z|Q) - 2\mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - 2\mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$4R_1 + 4R_2 < \mathbb{I}(U_0, U_1; Y_1|V_0, V_1) + \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2) + 2\mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0)$$

$$+ 2\mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q) - 4\mathbb{I}(U_0, V_0; Z|Q) - \mathbb{I}(U_1, U_2; Z|U_0, V_0)$$

$$- 2\mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0) - 2\mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$4R_1 + 4R_2 < \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1) + \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2)$$

$$+ 2\mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0) + 2\mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q) - 4\mathbb{I}(U_0, V_0; Z|Q)$$

$$- \mathbb{I}(U_1, U_2; Z|U_0, V_0) - 2\mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(U_1; U_2|U_0) - 2\mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)$$

$$
\begin{aligned}
4R_1 + 4R_2 &< \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) \\
&+ 2\mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0) + 2\mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q) - 4\mathbb{I}(U_0, V_0; Z|Q) \\
&- \mathbb{I}(V_1, V_2; Z|U_0, V_0) - 2\mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0) - 2\mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0) \\
4R_1 + 4R_2 &< \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1) + \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2) \\
&+ 2\mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0) + 2\mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q) - 4\mathbb{I}(U_0, V_0; Z|Q) \\
&- \mathbb{I}(V_1, V_2; Z|U_0, V_0) - 2\mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \mathbb{I}(V_1; V_2|V_0) - 2\mathbb{I}(U_1, V_1; U_2, V_2|U_0, V_0)
\end{aligned}
$$

for some distribution

$$
p(q)p(u_0, u_1, u_2|q)p(v_0, v_1, v_2|q)p(x_1|u_0, u_1, u_2)p(x_2|v_0, v_1, v_2)p(y_1, y_2, z|x_1, x_2), \tag{2.36}
$$

Proof. The achievability proof is inspired by [13], and is based on solving a dual secret key agreement problem in the source model that includes shared randomness at all terminals (see Fig. 2.5). In this dual model, rate constraints are derived so that the input and output distributions of the dual model approximate that of the original model while satisfying reliability and secrecy conditions in the dual model. The probability approximation then guarantees that reliability *and* secrecy conditions can be achieved in the original model. Finally, it is shown that there exists one realization of shared randomness for which the above-mentioned conditions are valid, thus removing the necessity for common randomness.

We begin by developing the encoding and decoding strategies for the source model and the original model, and derive and compare the joint probability distributions arising from these two strategies.

We begin with the multi-terminal secret key agreement problem in the source model as depicted in Fig. 2.5. Let $(U^n_{[0:2]}, V^n_{[0:2]}, X^n_1, X^n_2, Y^n_1, Y^n_2, Z^n)$ be i.i.d. and distributed according to

$$
p(u_{[0:2]}, x_1)p(v_{[0:2]}, x_2)p(y_1, y_2, z|x_1, x_2). \tag{2.37}
$$

Random Binning:

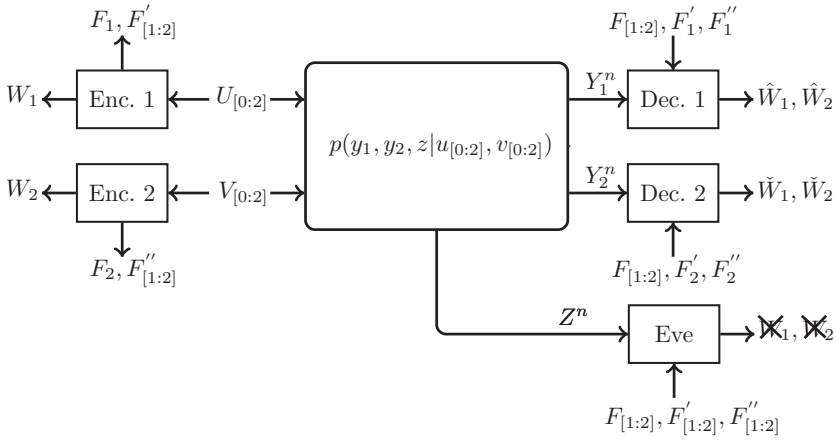

Figure 2.5. Dual secret key agreement problem in the source model for the original problem.

- To *each* u_0^n, uniformly and independently assign two random bin indices $w_1 \in [\![1, 2^{nR_1}]\!]$ and $f_1 \in [\![1, 2^{n\tilde{R}_1}]\!]$.

- To each pair (u_0^n, u_j^n) for $j = 1, 2$ uniformly and independently assign random bin index $f'_j \in [\![1, 2^{n\tilde{R}'_j}]\!]$.

- To each v_0^n uniformly and independently assign two random bin indices $w_2 \in [\![1, 2^{nR_2}]\!]$ and $f_2 \in [\![1, 2^{n\tilde{R}_2}]\!]$.

- To each pair (v_0^n, v_j^n) for $j = 1, 2$ uniformly and independently assign random bin index $f''_j \in [\![1, 2^{n\tilde{R}''_j}]\!]$.

- The random variables representing bin indices are:

$$W_{[1:2]}\ , \quad F_{[1:2]}\ , \quad F'_{[1:2]}\ , \quad F''_{[1:2]}. \tag{2.38}$$

- Decoder 1 is a Slepian-Wolf decoder observing $(y_1^n, f_{[1:2]}, f'_1, f''_1)$, and producing $(\hat{u}_0^n, \hat{u}_1^n)$ and $(\hat{v}_0^n, \hat{v}_1^n)$, thus declaring $\hat{w}_1 = W_1(\hat{u}_0^n)$ and $\hat{w}_2 = W_2(\hat{v}_0^n)$ to be the estimate of the pair (w_1, w_2).

- Decoder 2 is a Slepian-Wolf decoder observing $(y_2^n, f_{[1:2]}, f_2', f_2'')$, and producing $(\check{u}_0^n, \check{u}_2^n)$ and $(\check{v}_0^n, \check{v}_2^n)$, thus declaring the bin indices $\check{w}_1 = W_1(\check{u}_0^n)$ and $\check{w}_2 = W_2(\check{v}_0^n)$ as the estimate of the pair (w_1, w_2).

To condense the notation, we define the following variables:

$$\mathbf{f} \triangleq \big(f_{[1:2]}, f'_{[1:2]}, f''_{[1:2]}\big), \tag{2.39}$$

$$\hat{\mathbf{u}} \triangleq \big(\hat{u}_0^n, \check{u}_0^n, \hat{u}_1^n, \check{u}_2^n, \hat{v}_0^n, \check{v}_0^n, \hat{v}_1^n, \check{v}_2^n\big). \tag{2.40}$$

Since the binnings $\mathbf{f}$ are random the PMFs induced by $\mathbf{f}$ are random, therefore henceforth we use upper case letter for distributions when they depend on $\mathbf{f}$. The random PMF induced by random binning is then as follows:

$$\begin{aligned}
&P(u_{[0:2]}^n, v_{[0:2]}^n, x_{[1:2]}^n, y_1^n, y_2^n, z^n, w_{[1:2]}, \mathbf{f}, \hat{\mathbf{u}}) \\
&= p(u_{[0:2]}^n, v_{[0:2]}^n, x_{[1:2]}^n, y_1^n, y_2^n, z^n) P(w_{[1:2]}, f_{[1:2]} | u_0^n, v_0^n) P(f'_{[1:2]}, f''_{[1:2]} | u_{[0:2]}^n, v_{[0:2]}^n) \\
&\times P^{SW}(\hat{u}_0^n, \hat{u}_1^n, \hat{v}_0^n, \hat{v}_1^n | y_1^n, f_{[1:2]}, f_1', f_1'') P^{SW}(\check{u}_0^n, \check{u}_2^n, \check{v}_0^n, \check{v}_2^n | y_2^n, f_{[1:2]}, f_2', f_2'') \\
&= P(w_{[1:2]}, f_{[1:2]}, u_0^n, v_0^n) P(f'_{[1:2]}, f''_{[1:2]}, u_{[1:2]}^n, v_{[1:2]}^n | u_0^n, v_0^n) \\
&\times p(x_1^n | u_{[0:2]}^n) p(x_2^n | v_{[0:2]}^n) p(y_1^n, y_2^n, z^n | x_1^n, x_2^n) \\
&\times P^{SW}(\hat{u}_0^n, \hat{u}_1^n, \hat{v}_0^n, \hat{v}_1^n | y_1^n, f_{[1:2]}, f_1', f_1'') P^{SW}(\check{u}_0^n, \check{u}_2^n, \check{v}_0^n, \check{v}_2^n | y_2^n, f_{[1:2]}, f_2', f_2'') \\
&= P(w_{[1:2]}, f_{[1:2]}) P(u_0^n, v_0^n | w_{[1:2]}, f_{[1:2]}) P(f'_{[1:2]}, f''_{[1:2]} | u_0^n, v_0^n) P(u_{[1:2]}^n, v_{[1:2]}^n | u_0^n, v_0^n, f'_{[1:2]}, f''_{[1:2]}) \\
&\times p(x_1^n | u_{[0:2]}^n) p(x_2^n | v_{[0:2]}^n) p(y_1^n, y_2^n, z^n | x_1^n, x_2^n) \\
&\times P^{SW}(\hat{u}_0^n, \hat{u}_1^n, \hat{v}_0^n, \hat{v}_1^n | y_1^n, f_{[1:2]}, f_1', f_1'') P^{SW}(\check{u}_0^n, \check{u}_2^n, \check{v}_0^n, \check{v}_2^n | y_2^n, f_{[1:2]}, f_2', f_2''). \qquad (2.41)
\end{aligned}$$

Here, P^{SW} denotes the PMF of the output of the Slepian-Wolf decoder, which is a random PMF. $\hat{W}_1, \hat{W}_2$ and $\check{W}_1, \check{W}_2$ are omitted because they are functions of other random variables.

We now return to the original problem illustrated in Fig. 1.1 except that, in addition, a genie provides all terminals with shared randomness described by $(F_{[1:2]}, F'_{[1:2]}, F''_{[1:2]})$, whose distribution will be clarified in the sequel. In this augmented model:

- The messages W_1 and W_2 are mutually independent and uniformly distributed with rates R_1 and R_2 respectively. The shared randomness (F_1, F_2) is uniformly distributed over $[\![1, 2^{n\tilde{R}_1}]\!]$, $[\![1, 2^{n\tilde{R}_2}]\!]$, and independent of W_1, W_2.

- Encoder 1 and 2 are stochastic encoders producing codewords U_0^n, V_0^n according to distributions $P(u_0^n|w_{[1:2]}, f_{[1:2]})$ and $P(v_0^n|w_{[1:2]}, f_{[1:2]})$, respectively, which are the marginals of distribution $P(u_0^n, v_0^n|w_{[1:2]}, f_{[1:2]})$ appearing in (2.41). This choice of encoder establishes the connection between the two models.

- The four random variables $F'_{[1:2]}, F''_{[1:2]}$ are mutually independent and uniformly distributed over, respectively, $[\![1, 2^{n\tilde{R}'_1}]\!]$ and $[\![1, 2^{n\tilde{R}'_2}]\!]$, $[\![1, 2^{n\tilde{R}''_1}]\!]$ and $[\![1, 2^{n\tilde{R}''_2}]\!]$. They are also independent of (U_0^n, V_0^n) and therefore are independent of $(W_{[1:2]}, F_{[1:2]})$.

- Encoder 1 and 2 further generate $U_{[1:2]}^n$, $V_{[1:2]}^n$ according to $P(u_{[1:2]}^n|u_0^n, v_0^n, f'_{[1:2]}, f''_{[1:2]})$ and $P(v_{[1:2]}^n|u_0^n, v_0^n, f'_{[1:2]}, f''_{[1:2]})$, respectively, which are marginal distributions of $P(u_{[1:2]}^n, v_{[1:2]}^n|u_0^n, v_0^n, f'_{[1:2]}, f''_{[1:2]})$ from (2.41).

- Encoder 1 generates X_1^n i.i.d. according to $p(x_1|u_{[0:2]})$. Encoder 2 generates X_2^n i.i.d. according to $p(x_2|v_{[0:2]})$. X_1, X_2 are transmitted over the channel.

- Decoders 1 and 2 are Slepian-Wolf decoders inherited from the source model secret key agreement problem, observing respectively $(y_1^n, f_{[1:2]}, f'_1, f''_1)$ and $(y_2^n, f_{[1:2]}, f'_2, f''_2)$, and producing $(\hat{u}_0^n, \hat{u}_1^n, \hat{v}_0^n, \hat{v}_1^n)$ and $(\check{u}_0^n, \check{u}_2^n, \check{v}_0^n, \check{v}_2^n)$. Therefore, the following random PMFs for the decoder output distributions are inherited from the source model:

$$P^{SW}(\hat{u}_0^n, \hat{u}_1^n, \hat{v}_0^n, \hat{v}_1^n|y_1^n, f_{[1:2]}, f'_1, f''_1),$$
$$P^{SW}(\check{u}_0^n, \check{u}_2^n, \check{v}_0^n, \check{v}_2^n|y_2^n, f_{[1:2]}, f'_2, f''_2).$$

- Decoders 1 and 2 then produce estimates of (W_1, W_2), which are denoted $(\hat{W}_1, \hat{W}_2)$ and $(\check{W}_1, \check{W}_2)$ respectively.

The random PMF induced by the random binning and the encoding/decoding strategy is as follows:

$$\begin{aligned}
&\hat{P}(u^n_{[0:2]}, v^n_{[0:2]}, y^n_1, y^n_2, z^n, w_{[1:2]}, \mathbf{f}, \hat{\mathbf{u}}) \\
&= p^U(w_{[1:2]}) p^U(f_{[1:2]}) P(u^n_0, v^n_0 | w_{[1:2]}, f_{[1:2]}) p^U(f'_{[1:2]}) p^U(f''_{[1:2]}) P(u^n_{[1:2]}, v^n_{[1:2]} | u^n_0, v^n_0, f'_{[1:2]}, f''_{[1:2]}) \\
&\times p(x^n_1 | u^n_{[0:2]}) p(x^n_2 | v^n_{[0:2]}) p(y^n_1, y^n_2, z^n | x^n_1, x^n_2) \\
&\times P^{SW}(\hat{u}^n_0, \hat{u}^n_1, \hat{v}^n_0, \hat{v}^n_1 | y^n_1, f_{[1:2]}, f'_1, f''_1) P^{SW}(\check{u}^n_0, \check{u}^n_2, \check{v}^n_0, \check{v}^n_2 | y^n_2, f_{[1:2]}, f'_2, f''_2),
\end{aligned} \tag{2.42}$$

where $\mathbf{f}$ and $\hat{\mathbf{u}}$ are defined in (2.39) and (2.40), respectively, and p^U is the uniform distribution.

We now find constraints that ensure that the PMFs $\hat{P}$ and P are close in total variation distance. For the source model secret key agreement problem, substituting $X_1 = X_2 \leftarrow U_0$, and $X_3 = X_4 \leftarrow V_0$, in [13, Theorem 1] implies that $W_{[1:2]}$ is nearly independent of $F_{[1:2]}$ and Z^n, if

$$R_1 + \tilde{R}_1 < \mathbb{H}(U_0|Z), \tag{2.43}$$

$$R_2 + \tilde{R}_2 < \mathbb{H}(V_0|Z), \tag{2.44}$$

$$R_1 + \tilde{R}_1 + R_2 + \tilde{R}_2 < \mathbb{H}(U_0, V_0|Z), \tag{2.45}$$

note that [13, Theorem 1] returns a total of 15 inequalities, but the remaining are redundant because of (2.43)–(2.45). The above constraints imply that

$$P(z^n, w_{[1:2]}, f_{[1:2]}) \approx_\epsilon p(z^n) p^U(w_{[1:2]}) p^U(f_{[1:2]}).$$

Similarly, substituting $X_1 \leftarrow (U_0, U_1)$, $X_2 \leftarrow (U_0, U_2)$, $X_3 \leftarrow (V_0, V_1)$, $X_4 \leftarrow (V_0, V_2)$, and $Z \leftarrow (U_0, V_0, Z)$ in [13, Theorem 1] implies that $(f'_{[1:2]}, f''_{[1:2]})$ are nearly mutually independent and independent of (U_0, V_0, Z), therefore they are independent of $(w_{[1:2]}, f_{[1:2]})$, if

$$\tilde{R}'_j < \mathbb{H}(U_j|U_0, V_0, Z), \tag{2.46}$$

$$\tilde{R}''_j < \mathbb{H}(V_j|U_0, V_0, Z), \tag{2.47}$$

$$\tilde{R}'_1 + \tilde{R}''_j < \mathbb{H}(U_1, V_j|U_0, V_0, Z), \tag{2.48}$$

$$\tilde{R}'_2 + \tilde{R}''_j < \mathbb{H}(U_2, V_j|U_0, V_0, Z), \tag{2.49}$$

$$\tilde{R}'_1 + \tilde{R}'_2 < \mathbb{H}(U_1, U_2|U_0, V_0, Z), \tag{2.50}$$

$$\tilde{R}''_1 + \tilde{R}''_2 < \mathbb{H}(V_1, V_2|U_0, V_0, Z), \tag{2.51}$$

$$\tilde{R}'_1 + \tilde{R}'_2 + \tilde{R}''_j < \mathbb{H}(U_1, U_2, V_j|U_0, V_0, Z), \tag{2.52}$$

$$\tilde{R}'_j + \tilde{R}''_1 + \tilde{R}''_2 < \mathbb{H}(U_j, V_1, V_2|U_0, V_0, Z), \tag{2.53}$$

$$\tilde{R}'_1 + \tilde{R}'_2 + \tilde{R}''_1 + \tilde{R}''_2 < \mathbb{H}(U_1, U_2, V_1, V_2|U_0, V_0, Z), \tag{2.54}$$

for $j = 1, 2$. The above constraints imply

$$P(z^n, u_0^n, v_0^n, f'_{[1:2]}, f''_{[1:2]}) \approx_\epsilon p(z^n, u_0^n, v_0^n) p^U(f'_{[1:2]}) p^U(f''_{[1:2]}). \tag{2.55}$$

Hence,

$$P(w_{[1:2]}, f_{[1:2]}) \approx_\epsilon \hat{P}(w_{[1:2]}, f_{[1:2]}) = p^U(w_{[1:2]}) p^U(f_{[1:2]}), \tag{2.56}$$

$$P(f'_{[1:2]}, f''_{[1:2]}|u_0^n, v_0^n) \approx_\epsilon \hat{P}(f'_{[1:2]}, f''_{[1:2]}|u_0^n, v_0^n) = p^U(f'_{[1:2]}) p^U(f''_{[1:2]}). \tag{2.57}$$

In other words, the inequalities (2.43)–(2.45) and (2.46)–(2.54) imply that

$$P(z^n, w_{[1:2]}, f_{[1:2]}, f'_{[1:2]}, f''_{[1:2]}) \approx_\epsilon p(z^n) p^U(w_{[1:2]}) p^U(f_{[1:2]}) p^U(f'_{[1:2]}) p^U(f''_{[1:2]}). \tag{2.58}$$

Here, the PMF $P(z^n)$ is equal to $p(z^n)$ because the marginal distribution does not include random binning.

Therefore, the distributions in (2.41) and (2.42) are nearly equal, that is

$$P(u_{[0:2]}^n, v_{[0:2]}^n, y_1^n, y_2^n, z^n, w_{[1:2]}, \mathbf{f}, \hat{\mathbf{u}}) \approx_\epsilon \hat{P}(u_{[0:2]}^n, v_{[0:2]}^n, y_1^n, y_2^n, z^n, w_{[1:2]}, \mathbf{f}, \hat{\mathbf{u}}). \tag{2.59}$$

Similar to indirect decoding for channel coding, it is possible to use indirect decoding for source coding. More precisely, the first and the second decoders only need (u_0^n, v_0^n) to decode

(w_1, w_2). Decoder 1 and Decoder 2 can indirectly decode (u_0^n, v_0^n) from $(y_1^n, f_{[1:2]}, f_1', f_1'')$ and $(y_2^n, f_{[1:2]}, f_2', f_2'')$, respectively. From [13, Lemma 1] decoding is successful if

$$\tilde{R}_1 + \tilde{R}_j' > \mathbb{H}(U_0, U_j | V_0, V_j, Y_j), \tag{2.60}$$

$$\tilde{R}_2 + \tilde{R}_j'' > \mathbb{H}(V_0, V_j | U_0, U_j, Y_j), \tag{2.61}$$

$$\tilde{R}_1 + \tilde{R}_j' + \tilde{R}_j'' > \mathbb{H}(U_0, U_j, V_j | V_0, Y_j), \tag{2.62}$$

$$\tilde{R}_1 + \tilde{R}_2 + \tilde{R}_j'' > \mathbb{H}(V_0, V_j | U_0, U_j, Y_j), \tag{2.63}$$

$$\tilde{R}_j' + \tilde{R}_2 + \tilde{R}_j'' > \mathbb{H}(U_j, V_0, V_j | U_0, Y_j), \tag{2.64}$$

$$\tilde{R}_1 + \tilde{R}_j' + \tilde{R}_2 + \tilde{R}_j'' > \mathbb{H}(U_0, U_j, V_0, V_j | Y_j), \tag{2.65}$$

for $j = 1, 2$. Note that, inequality (2.63) is redundant because of (2.61). It yields

$$\begin{aligned} &P(u_{[0:2]}^n, v_{[0:2]}^n, y_1^n, y_2^n, z^n, w_{[1:2]}, \mathbf{f}, \hat{\mathbf{u}}) \approx_\epsilon P(u_{[0:2]}^n, v_{[0:2]}^n, y_1^n, y_2^n, z^n, w_{[1:2]}, \mathbf{f}) \\ &\times \mathbb{1}_{\left\{\hat{u}_0^n = \check{u}_0^n = u_0^n, \hat{u}_1^n = u_1^n, \check{u}_2^n = u_2^n\right\}} times \mathbb{1}_{\left\{\hat{v}_0^n = \check{v}_0^n = v_0^n, \hat{v}_1^n = v_1^n, \check{v}_2^n = v_2^n\right\}}. \end{aligned} \tag{2.66}$$

From equations (2.59), (2.66), and the triangle inequality,

$$\begin{aligned} &\hat{P}(u_{[0:2]}^n, v_{[0:2]}^n, y_1^n, y_2^n, z^n, w_{[1:2]}, \mathbf{f}, \hat{\mathbf{u}}) \approx_\epsilon P(u_{[0:2]}^n, v_{[0:2]}^n, y_1^n, y_2^n, z^n, w_{[1:2]}, \mathbf{f}) \\ &\times \mathbb{1}_{\left\{\hat{u}_0^n = \check{u}_0^n = u_0^n, \hat{u}_1^n = u_1^n, \check{u}_2^n = u_2^n\right\}} \mathbb{1}_{\left\{\hat{v}_0^n = \check{v}_0^n = v_0^n, \hat{v}_1^n = v_1^n, \check{v}_2^n = v_2^n\right\}}. \end{aligned} \tag{2.67}$$

For convenience, we reintroduce a lemma from [13]:

Lemma 3. *([13, Lemma 4]) Consider distributions p_{X^n}, $p_{Y^n|X^n}$, q_{X^n}, and $q_{Y^n|X^n}$ and random PMFs P_{X^n}, $P_{Y^n|X^n}$, Q_{X^n}, and $Q_{Y^n|X^n}$. Denoting asymptotic equality under total variation with $\approx_\epsilon$, we have:*

1.

$$P_{X^n} \approx_\epsilon Q_{X^n} \;\Rightarrow\; P_{X^n} P_{Y^n|X^n} \approx_\epsilon Q_{X^n} P_{Y^n|X^n}, \tag{2.68}$$

$$P_{X^n} P_{Y^n|X^n} \approx_\epsilon Q_{X^n} Q_{Y^n|X^n} \;\Rightarrow\; P_{X^n} \approx_\epsilon Q_{X^n}. \tag{2.69}$$

2. If $p_{X^n}p_{Y^n|X^n} \approx_\epsilon q_{X^n}q_{Y^n|X^n}$, then there exists a sequence $x^n \in \mathcal{X}^n$ such that

$$p_{Y^n|X^n=x^n} \approx_\epsilon q_{Y^n|X^n=x^n}. \tag{2.70}$$

3. If $P_{X^n} \approx_\epsilon Q_{X^n}$ and $P_{X^n}P_{Y^n|X^n} \approx_\epsilon P_{X^n}Q_{Y^n|X^n}$, then

$$P_{X^n}P_{Y^n|X^n} \approx_\epsilon Q_{X^n}Q_{Y^n|X^n}. \tag{2.71}$$

Using Lemma 3, Equation (2.69), the marginal distributions of the two sides of (2.67) are asymptotically equivalent, i.e.,

$$\begin{aligned}\hat{P}(u^n_{[0:2]}, v^n_{[0:2]}, z^n, w_{[1:2]}, \mathbf{f}, \hat{\mathbf{u}}) &\approx_\epsilon P(u^n_{[0:2]}, v^n_{[0:2]}, z^n, w_{[1:2]}, \mathbf{f}) \\ &\times \mathbb{1}_{\left\{\hat{u}^n_0=\check{u}^n_0=u^n_0, \hat{u}^n_1=u^n_1, \check{u}^n_2=u^n_2\right\}} \mathbb{1}_{\left\{\hat{v}^n_0=\check{v}^n_0=v^n_0, \hat{v}^n_1=v^n_1, \check{v}^n_2=v^n_2\right\}}.\end{aligned} \tag{2.72}$$

Using Lemma 3, Equation (2.68) we multiply the two sides of Equation (2.72) by the conditional distribution:

$$\hat{P}(\hat{w}_1, \check{w}_1, \hat{w}_2, \check{w}_2 | u^n_{[0:2]}, v^n_{[0:2]}, z^n, w_{[1:2]}, \mathbf{f}, \hat{\mathbf{u}}) = \mathbb{1}_{\left\{W_1(\hat{u}^n_0)=\hat{w}_1, W_1(\check{u}^n_0)=\check{w}_1\right\}} \mathbb{1}_{\left\{W_2(\hat{v}^n_0)=\hat{w}_2, W_2(\check{v}^n_0)=\check{w}_2\right\}},$$

to get:

$$\begin{aligned}&\hat{P}(u^n_{[0:2]}, v^n_{[0:2]}, z^n, w_{[1:2]}, \mathbf{f}, \hat{\mathbf{u}}, \hat{w}_1, \check{w}_1, \hat{w}_2, \check{w}_2) \approx_\epsilon P(u^n_{[0:2]}, v^n_{[0:2]}, z^n, w_{[1:2]}, \mathbf{f}) \\ &\times \mathbb{1}_{\left\{\hat{u}^n_0=\check{u}^n_0=u^n_0, \hat{u}^n_1=u^n_1, \check{u}^n_2=u^n_2\right\}} \mathbb{1}_{\left\{\hat{v}^n_0=\check{v}^n_0=v^n_0, \hat{v}^n_1=v^n_1, \check{v}^n_2=v^n_2\right\}} \\ &\times \mathbb{1}_{\left\{W_1(\hat{u}^n_0)=\hat{w}_1, W_1(\check{u}^n_0)=\check{w}_1\right\}} \mathbb{1}_{\left\{W_2(\hat{v}^n_0)=\hat{w}_2, W_2(\check{v}^n_0)=\check{w}_2\right\}} \\ &= P(u^n_{[0:2]}, v^n_{[0:2]}, z^n, w_{[1:2]}, \mathbf{f}) \\ &\times \mathbb{1}_{\left\{\hat{u}^n_0=\check{u}^n_0=u^n_0, \hat{u}^n_1=u^n_1, \check{u}^n_2=u^n_2\right\}} \mathbb{1}_{\left\{\hat{v}^n_0=\check{v}^n_0=v^n_0, \hat{v}^n_1=v^n_1, \check{v}^n_2=v^n_2\right\}} \mathbb{1}_{\left\{\hat{w}_1=\check{w}_1=w_1, \hat{w}_2=\check{w}_2=w_2\right\}},\end{aligned} \tag{2.73}$$

where $W_1(u^n_0) = \hat{w}_1$ and $W_2(v^n_0) = \hat{w}_2$ denote the bins assigned to u^n_0 and v^n_0, respectively. Using (2.73) and Lemma 3, Equation (2.68) leads to

$$\hat{P}(z^n, w_{[1:2]}, \mathbf{f}, \hat{w}_1, \check{w}_1, \hat{w}_2, \check{w}_2) \approx_\epsilon P(z^n, w_{[1:2]}, \mathbf{f}) \mathbb{1}_{\left\{\hat{w}_1=\check{w}_1=w_1, \hat{w}_2=\check{w}_2=w_2\right\}}. \tag{2.74}$$

Using equations (2.58) and (2.74) and Lemma 3, Equation (2.71) leads to

$$\hat{P}(z^n, w_{[1:2]}, \mathbf{f}, \hat{w}_1, \check{w}_1, \hat{w}_2, \check{w}_2) \approx_\epsilon p(z^n)p^U(w_{[1:2]}, f_{[1:2]})p^U(f'_{[1:2]}, f''_{[1:2]})\mathbb{1}_{\left\{\hat{w}_1=\check{w}_1=w_1, \hat{w}_2=\check{w}_2=w_2\right\}}. \tag{2.75}$$

We now eliminate the shared randomness $(F_{[1:2]}, F'_{[1:2]}, F''_{[1:2]})$ without affecting the secrecy and reliability requirements. By using Definition 3, Equation (2.75) ensures that there exists a fixed binning with corresponding PMF p that, if used in place of the random coding strategy P in (2.42), will induce the PMF $\hat{p}$ as follows:

$$\begin{aligned}&\hat{p}(z^n, w_{[1:2]}, f_{[1:2]}, f'_{[1:2]}, f''_{[1:2]}, \hat{w}_1, \check{w}_1, \hat{w}_2, \check{w}_2)\\ &\approx_\epsilon p(z^n)p^U(w_{[1:2]}, f_{[1:2]})p^U(f'_{[1:2]}, f''_{[1:2]})\mathbb{1}_{\left\{\hat{w}_1=\check{w}_1=w_1, \hat{w}_2=\check{w}_2=w_2\right\}}.\end{aligned} \tag{2.76}$$

Now, using Lemma 3, Equation (2.70) shows that there exists an instance of $(f_{[1:2]}, f'_{[1:2]}, f''_{[1:2]})$ such that:

$$\hat{p}(z^n, w_{[1:2]}, \hat{w}_1, \check{w}_1, \hat{w}_2, \check{w}_2|f_{[1:2]}, f'_{[1:2]}, f''_{[1:2]}) \approx_\epsilon p(z^n)p^U(w_1)p^U(w_2)\mathbb{1}_{\left\{\hat{w}_1=\check{w}_1=w_1, \hat{w}_2=\check{w}_2=w_2\right\}}. \tag{2.77}$$

This distribution satisfies the secrecy and reliability requirements as follows:

- Reliability: Using Lemma 3, Equation (2.69) leads to

$$\hat{p}(w_{[1:2]}, \hat{w}_{1,1}, \hat{w}_{1,2}, \hat{w}_{2,1}, \hat{w}_{2,2}|f_{[1:2]}, f'_{[1:2]}, f''_{[1:2]}) \approx_\epsilon \mathbb{1}_{\left\{\hat{w}_1=\check{w}_1=w_1, \hat{w}_2=\check{w}_2=w_2\right\}}, \tag{2.78}$$

  which is equivalent to:

$$\hat{p}\Big(\{(\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)\} \cup \{(\check{W}_1, \check{W}_2) \neq (W_1, W_2)\}\Big| f_{[1:2]}, f'_{[1:2]}, f''_{[1:2]}\Big) \to 0.$$

- Security: Again, using Lemma 3, Equation (2.69)

$$\hat{p}(z^n, w_{[1:2]}|f_{[1:2]}, f'_{[1:2]}, f''_{[1:2]}) \approx_\epsilon p(z^n)p^U(w_1)p^U(w_2). \tag{2.79}$$

Finally, we identify $p(x_1^n|w_1, f_1, f'_{[1:2]})$ and $p(x_2^n|w_2, f_2, f''_{[1:2]})$ (which is done by generating $u_{[0:2]}$ and $v_{[0:2]}$ first, respectively) as encoders and the Slepian-Wolf decoders as decoders for the channel coding problem. These encoders and decoders lead to reliable and secure encoders and decoders.

By applying a computer generated Fourier-Motzkin procedure [71] to (2.43)–(2.54), (2.60), (2.61), and (2.65) the achievable rate region for the strong secrecy regime in Theorem 6 is obtained. □

Remark 3. *If we assume that* (2.6)*, and therefore* (2.7)*, holds, the inequalities* (2.48) *for* $j = 2$*,* (2.49) *for* $j = 1$*, and* (2.50)–(2.54) *will be redundant and by applying the Fourier-Motzkin procedure [71, 72] to* (2.43)–(2.47)*,* (2.48) *for* $j = 1$*,* (2.49) *for* $j = 2$*,* (2.60)*,* (2.61)*, and* (2.65) *the region in Theorem 1 over the distribution* (2.36) *will be achieved. This shows that the region derived by OSRB is a superset of the region derived in the weak secrecy regime.*

Remark 4. *The random distributions* $P(u_0^n, v_0^n|w_{[1:2]}, f_{[1:2]})$ *and* $P(u_{[1:2]}^n, v_{[1:2]}^n|u_0^n, v_0^n, f'_{[1:2]}, f''_{[1:2]})$ *factorize as* $P(u_0^n|w_1, f_1)P(v_0^n|w_2, f_2)$ *and* $P(u_{[1:2]}^n|u_0^n, f'_{[1:2]})P(v_{[1:2]}^n|v_0^n, f''_{[1:2]})$*, respectively, which means that Encoders* 1 *and* 2 *are not using the common randomness and the message available at the other encoder to generate the common and private random variables. The common randomness* $(F_1, F'_{[1:2]})$ *represents the realization of Encoder 1's codebook and* $(F_2, F''_{[1:2]})$ *represents the realization of Encoder 2's codebook, which is available at all terminals, but the codebook at one encoder does not depend on the codebook of the other encoder.*

Remark 5. *The achievable region described in the proof of Theorem 6 was without time-sharing, i.e.,* $Q = \emptyset$*. One can incorporate this into the proof by generating i.i.d. copies of* Q*, and sharing it among all terminals and conditioning everything on it.*

CHAPTER 3

KEYLESS COVERT COMMUNICATION VIA CSI[1] [2]

3.1 Introduction

In this chapter, we consider the problem of covert communication over a state-dependent channel when the CSI is available either non-causally, causally, or strictly causally, either at the transmitter alone, or at both transmitter and receiver. Covert communication with respect to an adversary, called "warden," is one in which, despite communication over the channel, the warden's observation remains indistinguishable from an output induced by innocent channel-input symbols. Covert communication involves fooling an adversary in part by a proliferation of codebooks; for reliable decoding at the legitimate receiver, the codebook uncertainty is typically removed via a shared secret key that is unavailable to the warden. In contrast to previous work, we do not assume the availability of a large shared key at the transmitter and legitimate receiver. Instead, we only require a secret key with negligible rate to bootstrap the communication and our scheme extracts shared randomness from the CSI in a manner that keeps it secret from the warden, despite the influence of the CSI on the warden's output. When CSI is available at the transmitter and receiver, we derive the covert capacity region. When CSI is only available at the transmitter, we derive inner and outer bounds on the covert capacity. We also provide examples for which the covert capacity is positive with knowledge of CSI but is zero without it.

---

[1] ©2019 IEEE. Reprinted, with permission, from H. ZivariFard, M. R. Bloch, and A. Nosratinia, "Keyless covert communication in the presence of non-causal channel state information," 2019 IEEE Information Theory Workshop (ITW), 1-5

[2] ©2020 IEEE. Reprinted, with permission, from H. ZivariFard, M. R. Bloch, and A. Nosratinia, "Keyless Covert Communication in the Presence of Channel State Information," 2020 IEEE International Symposium on Information Theory (ISIT), 834-839

3.2 Channel model

Consider discrete memoryless state-dependent channels as shown in Fig. 1.2 or Fig. 1.3. The channel is characterized by input alphabet $\mathcal{X}$, legitimate output alphabet $\mathcal{Y}$, warden output alphabet $\mathcal{Z}$, state alphabet $\mathcal{S}$, and a transition probability $W_{YZ|XS}$. We assume that the CSI is independent and identically distributed (i.i.d.) and drawn according to Q_S and we let $x_0 \in \mathcal{X}$ be an "innocent" symbol corresponding to the absence of communication with the receiver. The distribution induced at the warden in the absence of communication is then,

$$Q_0(\cdot) = \sum_{s \in \mathcal{S}} Q_S(s) W_{Z|X,S}(\cdot|x_0, s), \tag{3.1}$$

and we let $Q_0^{\otimes n} = \prod_{i=1}^{n} Q_0$. The CSI may be available non-causally, causally, or strictly causally at the transmitter and may or may not be available at the receiver. Note that the exact causal or non-causal nature of CSI at the receiver is irrelevant because decoding is always done after transmission is completed. The warden is kept ignorant of the CSI.

Formally, a code with CSI available at both the transmitter and the receiver is defined as follows.

Definition 5. *A $(2^{nR}, n)$ code $\mathcal{C}_n$ with CSI available at both the transmitter and the receiver consists of:*

- *a message set $\mathcal{M} = [\![1, 2^{nR}]\!]$;*
- *when CSI is available non-causally at the transmitter, for each time slot $i \in [\![1, n]\!]$, a deterministic encoder $f_i : \mathcal{M} \times \mathcal{S}^n \mapsto \mathcal{X}_i$ that maps message and the entire CSI sequence to a channel input symbol x_i;*
- *when CSI is available causally at the transmitter, for each time slot $i \in [\![1, n]\!]$, a deterministic encoder $f_i : \mathcal{M} \times \mathcal{S}^i \mapsto \mathcal{X}_i$ that maps message and the past and current CSI samples to a channel input symbol x_i;*

- *when CSI is available strictly-causally at the transmitter, for each time slot* $i \in [\![1, n]\!]$*, a deterministic encoder* $f_i : \mathcal{M} \times \mathcal{S}^{i-1} \mapsto \mathcal{X}_i$ *that maps message and the past CSI samples to a channel input symbol* x_i*;*

- *a decoding function* $g : \mathcal{S}^n \times \mathcal{Y}^n \mapsto \mathcal{M} \cup \{?\}$ *that maps the channel observations and the CSI sequence to a message* $\hat{M} \in \mathcal{M}$ *or an error message* $?$*.*

A code with CSI available only at the transmitter is defined as follows.

Definition 6. *A* $(2^{nR}, n)$ *code* $\mathcal{C}_n$ *with CSI available only at the transmitter consists of*

- *a message set* $\mathcal{M} = [\![1, 2^{nR}]\!]$ *and a secret key set* $\mathcal{K} = [\![1, 2^{nR_K}]\!]$*;*

- *when CSI is available non-causally, for each time slot* $i \in [\![1, n]\!]$*, a stochastic encoder* $f_i : \mathcal{M} \times \mathcal{J} \times \mathcal{K} \times \mathcal{S}^n \mapsto \mathcal{X}_i$*, that maps message, local randomness, secret key, and the entire CSI sequence to a channel input symbol* x_i*;*

- *when CSI is available causally, for each time slot* $i \in [\![1, n]\!]$*, a stochastic encoder* $f_i : \mathcal{M} \times \mathcal{J} \times \mathcal{K} \times \mathcal{S}^i \mapsto \mathcal{X}_i$*, that maps message, local randomness, secret key, and the past and current CSI samples to a channel input symbol* x_i*;*

- *when CSI is available strictly-causally, for each time slot* $i \in [\![1, n]\!]$*, a stochastic encoder* $f_i : \mathcal{M} \times \mathcal{J} \times \mathcal{K} \times \mathcal{S}^{i-1} \mapsto \mathcal{X}_i$ *that maps message, local randomness, secret key, and the past CSI samples to a channel input symbol* x_i*;*

- *a decoding function* $g : \mathcal{Y}^n \times \mathcal{K} \mapsto \mathcal{M} \cup \{?\}$ *that maps the channel observations to a message* $\hat{M} \in \mathcal{M}$ *or an error message* $?$*.*

The reason for introducing a stochastic encoder when CSI is only available at the transmitter is that our achievability scheme then relies on a likelihood encoder [73]. The stochastic nature of the likelihood encoder greatly simplifies the covertness analysis by providing finer control over the statistics induced by the encoder.

The code is assumed known to all parties and the objective is to design a code that is reliable, covert, and keyless. Reliable means that the probability of error $P_e^{(n)} = \mathbb{P}(\hat{M} \neq M)$ vanishes when $n \to \infty$. Covert means that the warden cannot determine whether communication is happening (hypothesis H_1) or not (hypothesis H_0). Specifically, the probabilities of false alarm α_n (warden deciding H_1 when H_0 is true) and missed detection β_n (warden deciding H_0 when H_1 is true) satisfy $\alpha_n + \beta_n = 1$ for an uninformed warden making random decisions. When the channel carries communication, the warden's channel output distribution is P_{Z^n}, and the optimal hypothesis test by the warden satisfies $\alpha_n + \beta_n \geq 1 - \sqrt{\mathbb{D}(P_{Z^n}||Q_0^{\otimes n})}$ [74]. Therefore, we define a code as covert if $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n})$ vanishes when $n \to \infty$. We assume that $\text{supp}(Q_0) = \mathcal{Z}$ for otherwise $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n})$ diverges. Finally, keyless means that the rate of secret key $\frac{1}{n}\log|\mathcal{K}|$ vanishes as $n \to \infty$.

A rate R is achievable if there exists a sequence of reliable, covert, and keyless $(2^{nR}, n)$ codes and the covert capacity is the supremum of all achievable covert rates. We denote the covert capacity by $C_{\text{A-B}}$ where A $\in$ {NC,C,SC} indicates the non-causal, causal, or strictly causal nature of the CSI at the transmitter while B $\in$ {T,TR} indicates whether CSI is available only at the transmitter or both the transmitter and receiver. Hence, we are interested in characterizing $C_{\text{NC-TR}}$, $C_{\text{NC-T}}$, $C_{\text{C-TR}}$, $C_{\text{C-T}}$, $C_{\text{SC-TR}}$, $C_{\text{SC-T}}$.

3.3 Channel State Information Available at the Transmitter and the Receiver

Theorem 7. *Let*

$$\mathcal{A} \triangleq \left\{R \geq 0 : \exists P_{S,X,Y,Z} \in \mathcal{D} \text{ such that } R \leq \mathbb{I}(X;Y|S)\right\}, \tag{3.2a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,X,Y,Z} : \\ P_{S,X,Y,Z} = Q_S P_{X|S} W_{Y,Z|S,X} \\ P_Z = Q_0 \\ \mathbb{H}(S|Z) \geq \mathbb{I}(X;Z|S) - \mathbb{I}(X;Y|S) \end{array} \right\}. \tag{3.2b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with non-causal CSI at both the transmitter and the receiver is

$$C_{\text{NC-TR}} = \max\{a : a \in \mathcal{A}\}. \tag{3.3}$$

Theorem 7 suggests that the key rate $H(S|Z)$ extracted from CSI should exceed the difference between the capacity of the warden and the capacity of the legitimate receiver. The achievability is proved by superposition encoding and the complete proof is available in Appendix H.

Theorem 8. *Let*

$$\mathcal{A} \triangleq \big\{R \geq 0 : \exists P_{S,U,X,Y,Z} \in \mathcal{D} \text{ such that } R \leq \mathbb{I}(U;Y|S)\big\}, \tag{3.4a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,U,X,Y,Z} : \\ P_{S,U,X,Y,Z} = Q_S P_U \mathbb{1}_{\left\{X = X(U,S)\right\}} W_{Y,Z|S,X} \\ P_Z = Q_0 \\ \mathbb{H}(S|Z) \geq \mathbb{I}(U;Z|S) - \mathbb{I}(U;Y|S) \\ |\mathcal{U}| \leq |\mathcal{X}| + 1 \end{array} \right\}. \tag{3.4b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with causal CSI at both the transmitter and the receiver is

$$C_{\text{C-TR}} = \max\{a : a \in \mathcal{A}\}. \tag{3.5}$$

Again, Theorem 8 suggests that the key rate extracted from CSI should exceed the difference between the capacity of the warden and the capacity of the legitimate receiver. The achievability proof is based on block Markov encoding to combine a Shannon strategy for transmitting the message according to CSI with key generation and is available in Appendix I.

Theorem 9. *Let*

$$\mathcal{A} \triangleq \left\{ R \geq 0 : \exists P_{S,X,Y,Z} \in \mathcal{D} \text{ such that } R \leq \mathbb{I}(X;Y|S) \right\}, \tag{3.6a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,X,Y,Z} : \\ P_{S,X,Y,Z} = Q_S P_X W_{Y,Z|S,X} \\ P_Z = Q_0 \\ \mathbb{H}(S|Z) \geq \mathbb{I}(X;Z|S) - \mathbb{I}(X;Y|S) \end{array} \right\}. \tag{3.6b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with strictly causal CSI at both the transmitter and the receiver is

$$C_{\text{SC-TR}} = \max\{a : a \in \mathcal{A}\}. \tag{3.7}$$

Even though *strictly* causal CSI provides limited opportunities to enhance reliability, it is still useful, because it provides shared randomness from which to extract a secret key. The achievability proof merely uses a block Markov encoding scheme for key generation but not for data transmission and is available in Appendix J.

3.4 Examples of channels with CSI at transmitter and receiver

We now provide two examples of covert communication over state-dependent channels with CSI at the transmitter and receiver. A positive covert capacity is achieved without an external secret key, hence not subject to the square root law. The two examples explore additive

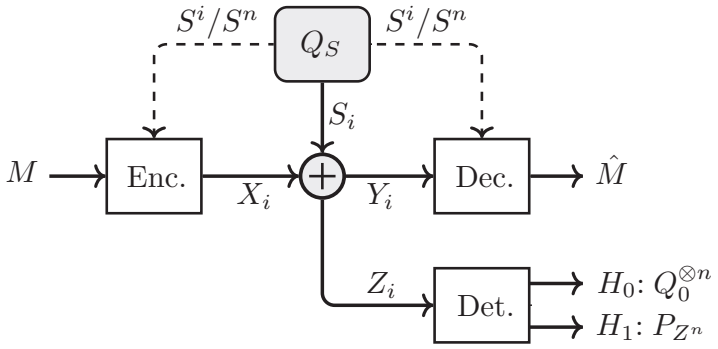

Figure 3.1. Binary symmetric channel with additive CSI at the transmitter and the receiver

and multiplicative CSI, respectively, with the former representing channels in which the channel state can in principle be cancelled and the latter representing fading-like channels.

Binary Additive State: Consider a channel in which X, Y, Z, and S are all binary, Q_S obeys a Bernoulli distribution with parameter $\zeta \in (0 : 0.5)$, and the innocent symbol is $x_0 = 0$. (See Fig. 3.1). The law of the channel is

$$Y = Z = X \oplus S, \tag{3.8}$$

so that $Q_0 = Q_S$.

Proposition 1. *The covert capacity of the DMC depicted in Fig. 3.1 with causal or non-causal CSI available at the transmitter and the receiver is*

$$C_{\text{NC-TR}} = C_{\text{C-TR}} = \mathbb{H}_b(\zeta) = \zeta \log \frac{1}{\zeta} + (1-\zeta) \log \frac{1}{1-\zeta}. \tag{3.9}$$

Intuitively, the encoder perfectly controls the warden's observations because it knows the CSI. By manipulating X, the encoder ensures that Z follows the statistics of S. In part, this means that the symbol $X = 1$ is associated half the time to $S = 0$ and half the time to $S = 1$ to ensure $P_Z = Q_S \sim \text{Bern}(\zeta)$. Further, since the transmitter and receiver share the CSI, the legitimate channel is error-free.

Table 3.1. Joint probability distribution of X, S

S \ X	0	1
0	α	β
1	$1-\alpha-\beta-\eta$	η

Proof. We first prove Proposition 1 when CSI is available non-causally at both the transmitter and the receiver. Substituting $Y = Z = X \oplus S$ in Theorem 7 results in

$$C_{\text{NC-TR}} = \max_{Q_S P_{X|S}} \mathbb{H}(X|S), \tag{3.10}$$

with the maximization subject to the constraint $P_Z = Q_0 = Q_S$. Let the joint distribution between X and S be according to Table 3.1, we have

$$P_Z(z=0) = P_{X,S}(x=0, s=0) + P_{X,S}(x=1, s=1) = \alpha + \eta, \tag{3.11}$$

$$Q_S(s=0) = P_{X,S}(x=0, s=0) + P_{X,S}(x=1, s=0) = \alpha + \beta. \tag{3.12}$$

Therefore $P_Z = Q_S$ implies that

$$Q_Z(z=0) = Q_S(s=0) \Rightarrow \alpha + \eta = \alpha + \beta \Rightarrow \eta = \beta. \tag{3.13}$$

Therefore,

$$\max_{Q_S P_{X|S}} \mathbb{H}(X|S) = \max_{\substack{(\alpha,\beta) \\ \zeta=\alpha+\beta}} \Big[-\alpha \log \frac{\alpha}{\alpha+\beta} - (1-\alpha-2\beta) \log \frac{1-\alpha-2\beta}{1-\alpha-\beta} - \beta \log \frac{\beta}{\alpha+\beta} - \beta \log \frac{\beta}{1-\alpha-\beta} \Big]. \tag{3.14}$$

Considering $Q_S(s=0) = \zeta = \alpha + \beta$ and substituting $\beta = \zeta - \alpha$ in (3.14) results in

$$\max_{Q_S P_{X|S}} \mathbb{H}(X|S) = \max_{\alpha} \Big[-\alpha \log \frac{\alpha}{\zeta} - (1+\alpha-2\zeta) \log \frac{1+\alpha-2\zeta}{1-\zeta} - (\zeta-\alpha) \log \frac{\zeta-\alpha}{\zeta} - (\zeta-\alpha) \log \frac{\zeta-\alpha}{1-\zeta} \Big]. \tag{3.15}$$

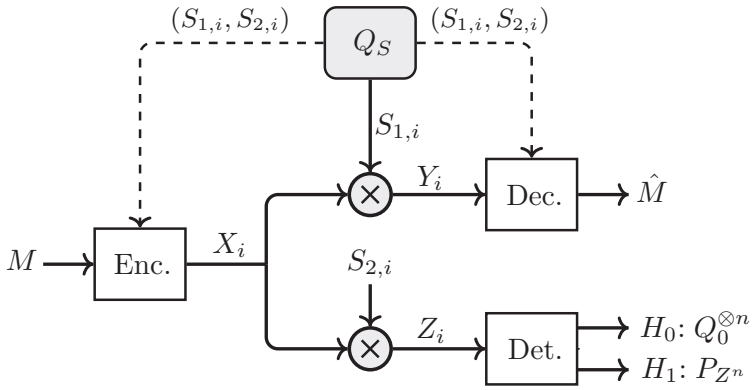

Figure 3.2. Binary symmetric channel with multiplicative CSI at the transmitter and the receiver

Since entropy is a continuous concave function, the maximizer of $\mathbb{H}(X|S)$ is found at the root of the first derivative of (3.15). This root is $\alpha = \zeta^2$, resulting in $\max \mathbb{H}(X|S) = \mathbb{H}(S)$. Since $Y = Z$, the condition $\mathbb{H}(S|Z) \geq \mathbb{I}(X;Z|S) - \mathbb{I}(X;Y|S)$ is automatically satisfied.

We now prove Proposition 1 when CSI is available causally at both the transmitter and the receiver. To prove achievability, we shall substitute specific choices of auxiliary random variables in Theorem 8. We choose U as a Bernoulli random variable with parameter $\eta \in (0 : 0.5)$ and independent of S, and we set $X = U \oplus S$. Therefore, $Y = Z = U$ and $\mathbb{I}(U;Y|S) = \mathbb{H}(U)$. Since $x_0 = 0$ we have $Q_0 = Q_S$ and the condition $Q_Z = Q_0$ results in $\eta = \zeta$ because

$$Q_S(z = 0) = \mathbb{P}(s = 0) = \zeta, \tag{3.16}$$

$$Q_Z(z = 0) = \mathbb{P}(u = 0) = \eta. \tag{3.17}$$

Since $Y = Z$, the condition $\mathbb{H}(S|Z) \geq \mathbb{I}(U;Z|S) - \mathbb{I}(U;Y|S)$ is automatically satisfied and the covert capacity is lower bounded by $\mathbb{H}_b(\zeta)$. The converse proof follows from the fact $C_{\text{C-TR}} \leq C_{\text{NC-TR}}$ by definition. □

Binary Multiplicative State: Consider a channel in which X, Y, Z, S_1, and S_2 are all binary and S_1 and S_2 have a joint distribution with parameters $P(S_1 = i, S_2 = j) = p_{i,j}$, for

$i, j \in \{0, 1\}$ and the innocent symbol is $x_0 = 0$ (See Fig. 3.2). The law of the channel is

$$Y = X \otimes S_1, \qquad Z = X \otimes S_2. \tag{3.18}$$

Proposition 2. *The covert capacity of the DMC depicted in Fig. 3.2 with causal or non-causal CSI available at the transmitter and the receiver is*

$$C_{\text{NC-TR}} = C_{\text{C-TR}} = p_{1,0}. \tag{3.19}$$

Intuitively, covert communication occurs when the warden's observation is impaired by a bad realization of CSI while the legitimate receiver simultaneously enjoys a good realization of the CSI. Since the receiver knows the CSI, the legitimate channel is effectively noise-free.

Proof. We prove Proposition 2 for the non-causal case, the proof for the causal case is similar and omitted for brevity. Substituting $S = (S_1, S_2)$ in Theorem 7, we obtain

$$\begin{aligned}
C_{\text{NC-TR}} &= \max_{\substack{P_{X|S_1,S_2}, \\ Q_Z = Q_0}} \big[\mathbb{I}(X; Y|S_1, S_2)\big] \\
&= \max_{\substack{P_{X|S_1,S_2}, \\ Q_Z = Q_0}} \Big[\sum_{i=0}^{1}\sum_{j=0}^{1} p_{i,j}\mathbb{I}(X; Y|S_1 = i, S_2 = j)\Big] \\
&\overset{(a)}{=} \max_{P_{X|S_1,S_2}} \Big[p_{1,0}\mathbb{I}(X; Y|S_1 = 1, S_2 = 0)\Big] \\
&= \max_{P_{X|S_1,S_2}} \Big[p_{1,0}\mathbb{H}(X|S_1 = 1, S_2 = 0)\Big] \\
&= p_{1,0},
\end{aligned} \tag{3.20}$$

where (a) holds because $Y = 0$ when $S_1 = 0$ so that $I(X; Y|S_1 = 0, S_2 = j) = 0$ and $Z = X$ when $S_2 = 1$ so that $P_X = P_Z = Q_0$ imposes $X = 0$, and $I(X; Y|S_1 = i, S_2 = 1) = 0$. Note that $Q_Z = Q_0$ implies that Z is always equal to zero, so that $\mathbb{I}(X; Z|S) \leq \mathbb{H}(Z) = 0$ and the condition $\mathbb{H}(S|Z) \geq \mathbb{I}(X; Z|S) - \mathbb{I}(X; Y|S)$ is automatically satisfied. □

3.5 Channel State Information Only Available at the Transmitter

We first recall the definitions of the following classes of broadcast channel, with channel state available only at the transmitter.

Definition 7 (Less Noisy Broadcast Channel With CSI available only at the transmitter)**.** *A discrete memoryless broadcast channel with CSI available only at the transmitter* $\left(\mathcal{X} \times \mathcal{S}, W_{Y,Z|X,S}, \mathcal{Y} \times \mathcal{Z}\right)$ *is said to be less noisy, if* $\mathbb{I}(U;Y) \geq \mathbb{I}(U;Z)$ *for all* $U-(X,S)-(Y,Z)$*. In this case, we say that* Y *is less noisy than* Z*.*

Definition 8 (More Capable Broadcast Channel With CSI available only at the transmitter)**.** *A discrete memoryless broadcast channel with CSI available only at the transmitter* $\left(\mathcal{X} \times \mathcal{S}, W_{Y,Z|X,S}, \mathcal{Y} \times \mathcal{Z}\right)$ *is said to be more capable, if* $\mathbb{I}(X;Y) \geq \mathbb{I}(X;Z)$ *for all* $P_{X,S}$*. In this case, we say that* Y *is more capable than* Z*.*

Theorem 10. *Let*

$$\mathcal{A} \triangleq \left\{ \begin{array}{l} R \geq 0 : \exists P_{U,V,S,X,Y,Z} \in \mathcal{D} : \\ R < \mathbb{I}(U;Y) - \max\left\{\mathbb{I}(U;S), \mathbb{I}(U,V;S) - \mathbb{I}(V;Y|U)\right\}\} \end{array} \right\}. \tag{3.21a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{U,V,S,X,Y,Z} : \\ P_{U,V,S,X,Y,Z} = P_U P_V P_{S|U,V} \mathbb{1}_{\left\{X = X(U,S)\right\}} W_{Y,Z|X,S} \\ Q_S(\cdot) = \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} P_U(u) P_V(v) P_{S|U,V}(\cdot|u,v) \\ P_Z = Q_0 \\ \mathbb{I}(V;Y|U) > \max\{\mathbb{I}(V;Z), \mathbb{I}(U,V;Z) - \mathbb{I}(U;Y)\} \\ |\mathcal{U}| \leq |\mathcal{X}| + 5 \\ |\mathcal{V}| \leq |\mathcal{X}| + 3 \end{array} \right\}. \tag{3.21b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with non-causal CSI at the transmitter is lower-bounded as

$$C_{\text{NC-T}} \geq \sup\{a : a \in \mathcal{A}\}. \tag{3.22}$$

The proof relies on block-Markov encoding to combine Gel'fand-Pinsker coding, for transmitting the message according to CSI [75], and Wyner-Ziv coding, for secret key generation [76]. The transmitter not only generates a key from S^n, but also selects its codeword according to S^n by using a likelihood encoder [77, 78, 79]. Instead of directly generating a secret key from the CSI, the transmitter relies on another random variable that is correlated with the CSI to help control the secret key rate. In particular, note that secret keys may not be needed, e.g, when the legitimate receiver's channel is a less noisy version of the warden's channel (see Corollary 2). Proof details are available in Appendix K.

A subset of rated in the region (3.21a) can be achieved without block-Markov coding or secret key generation. We provide these rates in Theorem 11 for reference. As shown in Section 3.6, however, secret key generation might be crucial to achieve positive covert rates.

Theorem 11. *Let*

$$\mathcal{A} \triangleq \left\{R \geq 0 : \exists P_{S,U,X,Y,Z} \in \mathcal{D} \text{ such that } R < \mathbb{I}(U;Y) - \mathbb{I}(U;S)\right\}, \tag{3.23a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,U,X,Y,Z} : \\ P_{S,U,X,Y,Z} = Q_S P_{U|S} \mathbb{1}_{\left\{X = X(U,S)\right\}} W_{Y,Z|X,S} \\ P_Z = Q_0 \\ \mathbb{I}(U;Y) > \mathbb{I}(U;Z) \\ |\mathcal{U}| \leq |\mathcal{X}| + 2 \end{array} \right\}. \tag{3.23b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with non-causal CSI at the transmitter is lower-bounded as

$$C_{\text{NC-T}} \geq \sup\{a : a \in \mathcal{A}\}. \tag{3.24}$$

Theorem 11 follows from Theorem 10 by choosing S independent of V, so that $P_{S|U,V} = P_{S|U}$. This choice ensures that $\mathbb{I}(V;S) = 0$ and $\mathbb{I}(V;U,Y) = 0$. Alternatively, Theorem 11 can be established with Gel'fand-Pinsker coding with a likelihood encoder but *without* block-Markov encoding or key generation from CSI. Details are omitted for brevity and are available online [80, Appendix E].

Theorem 12. *Let*

$$\mathcal{A} \triangleq \left\{ \begin{array}{l} R \geq 0 : \exists P_{S,U,V,X,Y,Z} \in \mathcal{D} : \\ R \leq \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\} \end{array} \right\}. \tag{3.25a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,U,V,X,Y,Z} : \\ P_{S,U,V,X,Y,Z} = Q_S P_{UV|S} \mathbb{1}_{\left\{X = X(U,S)\right\}} W_{Y,Z|X,S} \\ P_Z = Q_0 \\ \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\} \geq \mathbb{I}(V;Z) - \mathbb{I}(V;S) \\ \max\{|\mathcal{U}|, |\mathcal{V}|\} \leq |\mathcal{X}| + 3 \end{array} \right\}. \tag{3.25b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with non-causal CSI at the transmitter is upper-bounded as

$$C_{\text{NC-T}} \leq \max\{a : a \in \mathcal{A}\}. \tag{3.26}$$

Proof details are available in Appendix N.

Corollary 2. *Let*

$$\mathcal{A} \triangleq \left\{R \geq 0 : \exists P_{S,U,X,Y,Z} \in \mathcal{D} \text{ such that } R \leq \mathbb{I}(U;Y) - \mathbb{I}(U;S)\right\}, \tag{3.27a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,U,X,Y,Z} : \\ P_{S,U,X,Y,Z} = Q_S P_{U|S} \mathbb{1}_{\left\{X=X(U,S)\right\}} W_{Y,Z|X,S} \\ P_Z = Q_0 \\ |\mathcal{U}| \leq |\mathcal{X}| + 2 \end{array} \right\}. \tag{3.27b}$$

The covert capacity with CSI available non-causally only at the transmitter when the legitimate receiver's channel is less noisy than the warden's channel, is

$$C_{\text{NC-T}} = \max\{a : a \in \mathcal{A}\}. \tag{3.28}$$

Proof. The achievability follows from Theorem 11 and the less noisy property of the channel. We can also prove the achievability by using Theorem 10 while generating S independently of V (i.e. $P_{S|U,V} = P_{S|U}$) and the less noisy property of the channel. Furthermore, the converse proof follows from Theorem 12 and the less noisy property of the channel. □

Theorem 13. *Let*

$$\mathcal{A} \triangleq \left\{R \geq 0 : \exists P_{U,V,S,X,Y,Z} \in \mathcal{D} \text{ such that } R < \mathbb{I}(U;Y) + \min\left\{0, \mathbb{I}(V;Y|U) - \mathbb{I}(V;S)\right\}, \right. \tag{3.29a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{U,V,S,X,Y,Z} : \\ P_{U,V,S,X,Y,Z} = P_U P_V P_{S|V} \mathbb{1}_{\left\{X=X(U,S)\right\}} W_{Y,Z|X,S} \\ Q_S(\cdot) = \sum_{v \in \mathcal{V}} P_V(v) P_{S|V}(\cdot|v) \\ P_Z = Q_0 \\ \mathbb{I}(V;Y|U) > \max\{\mathbb{I}(V;Z), \mathbb{I}(U,V;Z) - \mathbb{I}(U;Y)\} \\ |\mathcal{U}| \leq |\mathcal{X}| + 2 \\ |\mathcal{V}| \leq |\mathcal{X}| + 3 \end{array} \right\}. \tag{3.29b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with causal CSI at the transmitter is lower-bounded as

$$C_{\text{C-T}} \geq \sup\{a : a \in \mathcal{A}\}. \tag{3.30}$$

Theorem 13 is proved using block-Markov encoding to combine a Shannon strategy for sending the message according to CSI and Wyner-Ziv coding for secret key generation. The details of the proof are available in Appendix O.

Theorem 14. *Let*

$$\mathcal{A} \triangleq \left\{ R \geq 0 : \exists P_{S,U,X,Y,Z} \in \mathcal{D} \text{ such that } R < \mathbb{I}(U;Y) \right\}, \tag{3.31a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,U,X,Y,Z} : \\ P_{S,U,X,Y,Z} = Q_S P_U \mathbb{1}_{\left\{X=X(U,S)\right\}} W_{Y,Z|X,S} \\ P_Z = Q_0 \\ \mathbb{I}(U;Y) > \mathbb{I}(U;Z) \\ |\mathcal{U}| \leq |\mathcal{X}| + 1 \end{array} \right\}. \tag{3.31b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with causal CSI at the transmitter is lower-bounded as

$$C_{\text{C-T}} \geq \sup\{a : a \in \mathcal{A}\}. \tag{3.32}$$

The proof is similar to the proof of Theorem 11, the details are omitted for brevity and are available online; please see [80, Appendix G].

Theorem 15. *Let*

$$\mathcal{A} \triangleq \left\{R \geq 0 : \exists P_{S,U,V,X,Y,Z} \in \mathcal{D} \text{ such that } R \leq \mathbb{I}(U;Y)\right\}, \tag{3.33a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,U,V,X,Y,Z} : \\ P_{S,U,V,X,Y,Z} = Q_S P_V P_{U|V} \mathbb{1}_{\left\{X = X(U,S)\right\}} W_{Y,Z|X,S} \\ P_Z = Q_0 \\ \mathbb{I}(U;Y) \geq \mathbb{I}(V;Z) \\ \max\{|\mathcal{U}|, |\mathcal{V}|\} \leq |\mathcal{X}| \end{array} \right\}. \tag{3.33b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with causal CSI at the transmitter is upper-bounded as

$$C_{\text{C-T}} \leq \max\{a : a \in \mathcal{A}\}. \tag{3.34}$$

Proof details are available in Appendix Q.

Corollary 3. *Let*

$$\mathcal{A} \triangleq \left\{R \geq 0 : \exists P_{S,U,X,Y,Z} \in \mathcal{D} \text{ such that } R \leq \mathbb{I}(U;Y)\right\}, \tag{3.35a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,U,X,Y,Z} : \\ P_{S,U,X,Y,Z} = Q_S P_U \mathbb{1}_{\left\{X=X(U,S)\right\}} W_{Y,Z|X,S} \\ P_Z = Q_0 \\ |\mathcal{U}| \leq |\mathcal{X}| + 1 \end{array} \right\}. \tag{3.35b}$$

The covert capacity with CSI available causally only at the transmitter when the legitimate receiver's channel is less noisy than the warden's channel is

$$C_{\text{C-T}} = \max\{a : a \in \mathcal{A}\}. \tag{3.36}$$

Proof. The achievability is proved by using Theorem 14 and the less noisy property of the channel. We can also prove the achievability by using Theorem 13 while generating S independently of V (i.e. $P_{S|V} = Q_S$) and the less noisy property of the channel. Furthermore, the converse proof follows from Theorem 15 and the less noisy property of the channel. □

Theorem 16. *Let*

$$\mathcal{A} \triangleq \left\{R \geq 0 : \exists P_{X,V,S,Y,Z} \in \mathcal{D} \text{ such that } R < \mathbb{I}(X;Y) + \min\left\{0, \mathbb{I}(V;Y|X) - \mathbb{I}(V;S)\right\}\right., \tag{3.37a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{X,V,X,Y,Z} : \\ P_{X,V,S,Y,Z} = P_X P_V P_{S|V} W_{Y,Z|X,S} \\ Q_S(\cdot) = \sum_{v \in \mathcal{V}} P_V(v) P_{S|V}(\cdot|v) \\ P_Z = Q_0 \\ \mathbb{I}(V;Y|X) > \max\{\mathbb{I}(V;Z), \mathbb{I}(X,V;Z) - \mathbb{I}(X;Y)\} \\ |\mathcal{V}| \leq |\mathcal{X}| + 3 \end{array} \right\}. \tag{3.37b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with strictly causal CSI at the transmitter is lower-bounded as

$$C_{\text{SC-T}} \geq \sup\{a : a \in \mathcal{A}\}. \tag{3.38}$$

The proof is similar to the proof of Theorem 13, and we only use the CSI for key generation and not for data transmission. The details are omitted for brevity and are available online; please see [80, Appendix H].

Remark 6. *In the proof of Theorem 10, Theorem 13, and Theorem 16, we assume that there exist a shared secret key for the first two transmission blocks to bootstrap the covert communication between the transmitter and the receiver. The overall rate of this secret key asymptotically amortizes to a negligible value as the number of transmission blocks $B \to \infty$.*

Theorem 17. *Let*

$$\mathcal{A} \triangleq \left\{R \geq 0 : \exists P_{S,X,Y,Z} \in \mathcal{D} \text{ such that } R < \mathbb{I}(X;Y)\right\}, \tag{3.39a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,X,Y,Z} : \\ P_{S,X,Y,Z} = Q_S P_X W_{Y,Z|X,S} \\ P_Z = Q_0 \\ \mathbb{I}(X;Y) > \mathbb{I}(X;Z) \end{array} \right\}. \tag{3.39b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with strictly causal CSI at the transmitter is lower-bounded as

$$C_{\text{SC-T}} \geq \sup\{a : a \in \mathcal{A}\}. \tag{3.40}$$

The proof is similar to the proof of Theorem 14, the details of the proof are omitted for brevity and are available online; please see [80, Appendix I]. We now present an upper bound on the covert capacity when the CSI is available strictly causally at the transmitter.

Theorem 18. *Let*

$$\mathcal{A} \triangleq \big\{R \geq 0 : \exists P_{S,V,X,Y,Z} \in \mathcal{D} \text{ such that } R \leq \mathbb{I}(X;Y)\big\}, \tag{3.41a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,V,X,Y,Z} : \\ P_{S,V,X,Y,Z} = Q_S P_V P_{X|V} W_{Y,Z|X,S} \\ P_Z = Q_0 \\ \mathbb{I}(X;Y) \geq \mathbb{I}(V;Z) \\ |\mathcal{V}| \leq |\mathcal{X}| \end{array} \right\}. \tag{3.41b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with strictly causal CSI at the transmitter is upper-bounded as

$$C_{\text{SC-T}} \leq \max\{a : a \in \mathcal{A}\}. \tag{3.42}$$

Proof details are available in Appendix T.

Corollary 4. *Let*

$$\mathcal{A} \triangleq \big\{R \geq 0 : \exists P_{S,X,Y,Z} \in \mathcal{D} \text{ such that } R \leq \mathbb{I}(X;Y)\big\}, \tag{3.43a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,X,Y,Z} : \\ P_{S,X,Y,Z} = Q_S P_X W_{Y,Z|X,S} \\ P_Z = Q_0 \end{array} \right\}. \tag{3.43b}$$

The covert capacity when CSI is available strictly causally at the transmitter and the legitimate receiver's channel is more capable than the warden's channel is,

$$C_{\text{SC-T}} = \max\{a : a \in \mathcal{A}\}. \tag{3.44}$$

Proof. The achievability is proved by using Theorem 17 and the more capable property of the channel. We can also prove the achievability by using Theorem 16 while generating S independently of V (i.e. $P_{S|V} = Q_S$) and the more capable property of the channel. Furthermore, the converse is proved by utilizing Theorem 18 and the more capable property of the channel. □

Remark 7 (Cardinality Bounds)**.** *The cardinality bounds on the auxiliary random variables in Theorems 8 to 16 follows by a standard application of the Eggleston-Fenchel-Carathéodory theorem [81, Theorem 18]. Details are omitted for brevity.*

Remark 8 (Do Stochastic Encoders Improve the Capacity Region?)**.** *We use deterministic encoders when the CSI is available at both of the legitimate terminals, while we use stochastic encoders when the CSI is only available at the transmitter. The use of stochastic encoders is merely motivated by technical convenience in our proof, and we could not conclude whether stochastic encoders outperform deterministic ones.*

3.6 Examples of Channels with CSI at transmitter

We provide two examples of covert communication over state-dependent channels with CSI at the transmitter alone, for which the covert capacity is positive. In both examples, the CSI is additive; however, in the first example the warden's channel is a degraded version of the legitimate receiver's channel while in the second example the legitimate receiver's channel is a degraded version of the warden's channel. The second example shows that our proposed coding scheme with block-Markov encoding and Wyner-Ziv encoding for secret key generation in Theorem 13, can outperform the simple approach for deriving the covert rates in Theorem 14.

Degraded Channel with Binary Additive State: Consider a channel in which X, Y, Z and $S = (S_1, S_2)$ are all binary, and let S_1 and S_2, be independent Bernoulli random variables

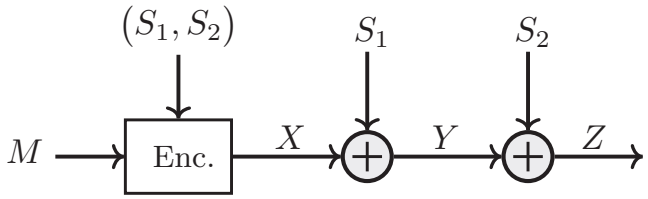

Figure 3.3. Degraded channel with binary additive CSI at the transmitter

with parameters $\alpha \in [0 : 0.5]$ and $\beta \in [0 : 0.5]$, respectively, and let $x_0 = 0$ (See Fig. 3.3). Here, S_1 and S_2 are the CSI of the legitimate receiver's channel and the warden's channel, respectively. The CSI is available causally at the Encoder and the law of the channel is as follows

$$Y = X \oplus S_1, \tag{3.45}$$

$$Z = Y \oplus S_2. \tag{3.46}$$

Proposition 3. *The covert capacity of the DMC depicted in Fig. 3.3 with causal CSI at the transmitter is*

$$C_{\text{C-T}} \overset{(a)}{=} \max_{\substack{P_U, \\ P_Z = Q_0}} \mathbb{H}(U) \overset{(b)}{=} \mathbb{H}_b(\alpha), \tag{3.47}$$

where $\mathbb{H}_b(\cdot)$ is binary entropy.

Proof. The achievability proof for (a) follows from the achievability part of Corollary 3 by considering U, which is the auxiliary random variable that represents the message, as a Bernoulli random variable independent of S_1 and S_2 with parameter $\lambda \in [0 : 0.5]$ and setting $X = U \oplus S_1$. The converse part of (a) follows from the converse part of Corollary 3 and the fact that $\mathbb{I}(U;Y) \leq \mathbb{H}(U)$. To prove (b) in Proposition 3, we have

$$\begin{aligned} Q_0(z=0) &= \mathbb{P}(s_1 \oplus s_2 = 0) \\ &= \mathbb{P}(s_1 = 0, s_2 = 0) + \mathbb{P}(s_1 = 1, s_2 = 1) \\ &= (1-\alpha)(1-\beta) + \alpha\beta. \end{aligned} \tag{3.48}$$

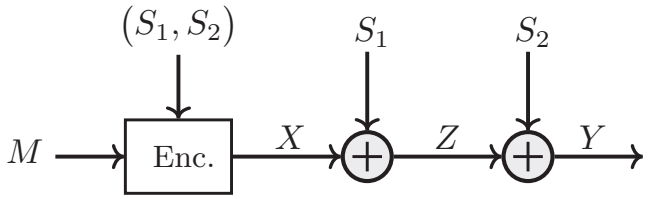

Figure 3.4. Reverse degraded channel with binary additive CSI at the transmitter

The distribution induced at the output of the warden when transmitting a codeword is

$$\begin{aligned} P_Z(z=0) &= \mathbb{P}(u \oplus s_2 = 0) \\ &= \mathbb{P}(u=0, s_2=0) + \mathbb{P}(u=1, s_2=1) \\ &= (1-\lambda)(1-\beta) + \lambda\beta. \end{aligned} \tag{3.49}$$

Therefore, the covertness constraint $P_Z = Q_0$ requires $\lambda = \alpha$. □

Reverse Degraded Channel with Binary Additive State: To show the benefits of the proposed scheme, we provide an example in which the region in Theorem 13 strictly improves the region in Theorem 14. Consider a channel in which X, Y, Z and $S = (S_1, S_2)$ are all binary, and let S_1, S_2 and U be independent Bernoulli random variables with parameters $\alpha \in (0 : 0.5]$, $\beta \in (0 : 0.5]$, and $\lambda \in (0 : 0.5]$, respectively, and let $x_0 = 0$ (See Fig. 3.4). Also, let V be a Bernoulli random variable. Here, S_1 and S_2 are the CSI of the warden's channel and the legitimate receiver's channel, respectively, U is an auxiliary random variable that represents the message, and V is an auxiliary random variable that represents a description of the CSI. The CSI is available causally at the Encoder and the law of the channel is as follows

$$Z = X \oplus S_1, \tag{3.50}$$

$$Y = Z \oplus S_2. \tag{3.51}$$

Since for this example $\mathbb{I}(U;Z) \geq \mathbb{I}(U;Y)$, the achievable rate region in Theorem 14 results in zero rate but Theorem 13 results in the following region.

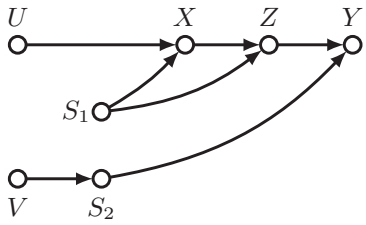

Figure 3.5. Chaining between the random variables for the reverse degraded channel with binary additive CSI

Proposition 4. *The covert capacity of the DMC depicted in Fig. 3.4 with causal CSI at the transmitter is lower bounded as*

$$C_{\text{C-T}} \geq \mathbb{H}_b(\eta) - \mathbb{H}_b(\beta), \tag{3.52}$$

where $\eta = \alpha\beta + (1-\alpha)(1-\beta)$.

Proof. Here we choose $X = U \oplus S_1$ therefore $Z = U$ and $Y = U \oplus S_2$. To prove the region in Proposition 4 by using Theorem 13, we start with covertness constraint $P_Z = Q_0$,

$$Q_0(z=0) = \mathbb{P}(s_1 = 0) = \alpha, \tag{3.53}$$

$$P_Z(z=0) = \mathbb{P}(u=0) = \lambda. \tag{3.54}$$

Therefore, the covertness constraint requires $\lambda = \alpha$. We also choose $V = S_2$ therefore,

$$P_{V|S} = P_{V|S_1,S_2} = P_{V|S_2} = \mathbb{1}_{\{V=S_2\}}. \tag{3.55}$$

The chaining between the random variables for this example is depicted in Fig. 3.5. Now we show that the fourth condition in Theorem 13 which includes the following conditions is satisfied,

$$\mathbb{I}(V;Y|U) > \mathbb{I}(V;Z), \tag{3.56}$$

$$\mathbb{I}(U,V;Y) > \mathbb{I}(U,V;Z). \tag{3.57}$$

We now have,

$$\begin{aligned}
\mathbb{I}(V;Y|U) &= \mathbb{H}(S_2|U) - \mathbb{H}(S_2|U,Y) \\
&= \mathbb{H}(S_2) - \mathbb{H}(S_2|U, U \oplus S_2) \\
&= \mathbb{H}(S_2) = \mathbb{H}_b(\beta), \\
\mathbb{I}(V;Z) &= \mathbb{H}(S_2) - \mathbb{H}(S_2|U) \overset{(a)}{=} 0.
\end{aligned}$$

where (a) follows since U and S_2 are independent. Therefore, the condition (3.56) is satisfied. For the condition in (3.57) we have,

$$\begin{aligned}
\mathbb{I}(U,V;Y) &= \mathbb{H}(U,S_2) - \mathbb{H}(U,S_2|Y) \\
&= \mathbb{H}(U) + \mathbb{H}(S_2) - \mathbb{H}(U,S_2|U \oplus S_2) \\
&= \mathbb{H}(U) + \mathbb{H}(S_2) - \mathbb{H}(S_2|U \oplus S_2), \\
\mathbb{I}(U,V;Z) &= \mathbb{I}(U,S_2;U) = \mathbb{H}(U)
\end{aligned}$$

since $\mathbb{H}(S_2) - \mathbb{H}(S_2|U \oplus S_2) > 0$ the condition in (3.57) is also satisfied. To calculate the covert rate (3.29a) in Theorem 13 we have,

$$\begin{aligned}
\mathbb{I}(V;Y|U) &= \mathbb{H}_b(\beta), \\
\mathbb{I}(V;S) &= \mathbb{I}(S_2;S_1,S_2) \\
&= \mathbb{H}(S_2) = \mathbb{H}_b(\beta), \\
\mathbb{I}(U;Y) &= \mathbb{H}(Y) - \mathbb{H}(Y|U) \\
&= \mathbb{H}(U \oplus S_2) - \mathbb{H}(U \oplus S_2|U) \\
&= \mathbb{H}(U \oplus S_2) - \mathbb{H}(S_2) \\
&= \mathbb{H}_b(\eta) - \mathbb{H}_b(\beta),
\end{aligned}$$

where $\eta = \alpha\beta + (1-\alpha)(1-\beta)$.

Remark 9 (Covertness vs. Security)**.** *This example also captures the difference between covertness and security. In this example, the warden has noiseless access to the transmitted sequence, and therefore it can decode the transmitted message, but since the transmitted sequence has the same statistics as the CSI it cannot prove that communication is happening.*

Remark 10 (Shared Key)**.** *In the examples provided in this section, the codebooks are generated with the same distribution as the CSI S_1 therefore the legitimate terminals need to have access to a shared secret key of negligible rate to discriminate the codewords from the CSI which is consistent with our code definition in Definition 6.*

□

CHAPTER 4

COVERT COMMUNICATION VIA COOPERATIVE JAMMING[1]

4.1 Introduction

In this chapter, we consider the problem of covert communication in the presence of a cooperative jammer. We show that a cooperative jammer can facilitate the communication of positive covert rates, subject to availability of a friendly jammer in the environment. We consider various scenarios and each of these scenarios have been defined, justified and studied in a different section of this dissertation.

4.2 Problem Definition

A discrete memoryless $(\mathcal{X}, \mathcal{S}, \mathcal{Y}, \mathcal{Z}, W_{Y,Z|X,S})$, illustrated in Fig 1.4, consists of channel input alphabet $\mathcal{X}$ at the transmitter, channel input alphabet $\mathcal{S}$ at the jammer, channel output alphabet $\mathcal{Y}$ at the receiver, and channel output alphabet $\mathcal{Z}$ at the warden. All alphabets are finite.

Let $x_0 \in \mathcal{X}$ be the innocent symbol which will be sent over the channel by the transmitter when no communication takes place. When the transmitter sends $x_0^n \in \mathcal{X}^n$, unlike other jamming problems, in this dissertation the jammer transmits a non-i.i.d. *coded* sequence S^n. Therefore, the distribution induced at the output of the channel when no communication takes place, denoted by $\Upsilon_{Z^n}^{(0)}$, is not necessarily i.i.d. The first reason that we use a coded jammer instead of a jammer that transmits an i.i.d. sequence for the no communication mode is that random numbers are a precious resource in practice, and we want to use this resource as little as possible. The second reason that we use a coded jammer is that it

[1] ©2021 IEEE. Reprinted, with permission, from H. ZivariFard, M. R. Bloch, and A. Nosratinia, "Covert Communication via Non-Causal Cribbing from a Cooperative Jammer," 2021 IEEE International Symposium on Information Theory (ISIT), 202-207

enables us to *design* the jammer's codebook in such a way that it helps the transmitter to communicate both covertly and reliably.

Here we consider four different jamming models. In the first model, the jammer and the receiver are assumed to have access to a rate limited and uniformly distributed shared secret key $K \in \mathcal{K}$, this helps the receiver cancel the interference caused by the randomness that the jammer interpolates into the channel. In this problem, there is no cooperation between the jammer and the transmitter and the transmitter and the jammer do not have access to any source of local randomness. Nevertheless, one can extend our results by removing the shared secret key between the jammer and the receiver and let the jammer use a source of local randomness.

In the second jamming model, the transmitter, the receiver, and the jammer are assumed to have access to a rate limited and uniformly distributed shared secret key $K \in \mathcal{K}$, this helps the transmitter to coordinate its channel input according to the jammer's channel input, and it also helps the receiver to cancel the interference caused by the jammer's channel input. In this problem, the transmitter and the jammer do not have access to any source of local randomness.

In the third model, the transmitter and the receiver are assumed to have access to a rate limited and uniformly distributed shared secret key $K \in \mathcal{K}$, this helps the transmitter communicate covertly with the receiver even if the channel from the transmitter to the receiver is noisier than the channel from the transmitter to the warden. Also, the jammer's channel input is assumed to be available non-causally or causally at the transmitter so that the transmitter can coordinate its channel input according to the jammer's codeword. In this problem, the transmitter does not have access to any source of local randomness, but the jammer has access to a limited amount of local randomness. Nevertheless, one can extend our results by removing the shared secret key between the transmitter and the receiver.

In the fourth jamming model, the jammer and the receiver are assumed to have access to a rate limited and uniformly distributed shared secret key $K \in \mathcal{K}$, this helps the receiver

to cancel the interference caused by the randomness that the jammer interpolates into the channel. Here, the transmitter's channel input is assumed to be available non-causally, causally, or strictly-causally at the jammer so that the jammer can coordinate its channel input according to the transmitter's codeword. The transmitter does not have access to any source of local randomness, but the jammer uses a rate limited source of local randomness, which is shared with the receiver. Nevertheless, one can extend our results by removing the shared secret key between the jammer and the receiver and let the jammer use a source of local randomness.

There are three main differences between the covert communication with jamming problems studied in this dissertation and the covert communication with jamming in the literature. First, the results for the covert communication with jamming in the literature are mainly based on the assumptions that the transmitter and the receiver share a secret codebook unknown to Willie and also there is a long secret key shared between the legitimate parties. But in the problems studied in this dissertation the codebook is a public knowledge and is available to all terminals including the warden. The second difference is that unlike the problems studied in this dissertation, in the jamming problems studied so far the jammer uses an unlimited source of local randomness, which is a very precious resource in practice. Therefore, the trade-off between the rate of the covert communication, the rate of the secret key needed between the legitimate terminals, and the rate of the local randomness used by jammer is missing in the literature. In this dissertation, we try to shed light on the interplay between the rate of the local randomness used by the jammer, the rate of the shared secret key between the legitimate terminals, and the rate of the covert communication. The third main difference is that, in this dissertation, unlike most of the works on covert communication with jamming in the literature, the warden is using a statistical detector instead of a power detector.

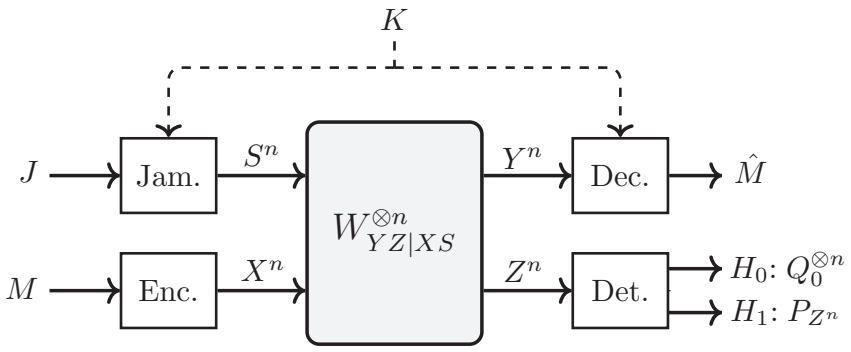

Figure 4.1. Covert communication with blind jammer

Definition 9. *$(R, R_K) \in \mathbb{R}_+^2$ is achievable if there exists a (n, R, R_K) code $\mathcal{C}_n$, such that*

$$\lim_{n\to\infty} P_e^{(n)} \to 0, \tag{4.1}$$

$$\lim_{n\to\infty} \mathbb{D}(P_{Z^n} || \Upsilon_{Z^n}) \to 0, \tag{4.2}$$

where

$$\Upsilon_{Z^n}(\cdot) = \frac{1}{|\mathcal{J}|} \sum_{J\in\mathcal{J}} W^{\otimes n}_{Z|X=x_0,S}\big(\,\cdot\,|x_0^n, S^n(J)\big). \tag{4.3}$$

The covert capacity region is the closure of the set of the all achievable regions.

4.3 Blind Jamming

In this section, we study a scenario in which a transmitter wishes to communicate a message $M \in \mathcal{M}$ covertly with a receiver while there is a friendly jammer in the environment. Here, the jammer and the receiver are assumed to have access to a rate limited and uniformly distributed shared secret key $K \in \mathcal{K}$, this helps the receiver cancel the interference caused by the randomness that the jammer interpolates into the channel. But, one can simply extend our scheme to the case that there is no shared secret key between the jammer and the receiver. We first study this problem when the transmitter and the jammer do not have a strategy. Then we discuss two different strategies for the jammer which can result to positive

rate for covert communication. The first strategy is that the jammer stays silent when the transmitter is communicating with the receiver and transmits when the transmitter is not communicating with the receiver. The second strategy is that the jammer transmits only when the transmitter is communicating. Note that having a strategy requires the legitimate terminals to share a secret key of negligible rate. The strategy described above requires the jammer to know in which blocks the transmitter is communicating. Then, we compare the results obtained for these three different cases in the context of three examples.

The following Theorem establishes an inner bound on the covert capacity when we do not force the jammer to use any specific strategy and the jammer always transmits a codeword regardless of the transmitter's action.

Theorem 19. *Let*

$$\mathcal{A} = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists \big(P_{SXYZ}, \Upsilon_{XSZY}\big) \in \mathcal{D} : \\ R < \mathbb{I}_P(X;Y|S) \\ R_K > \max\big\{\mathbb{I}_\Upsilon(S;Z), \mathbb{I}_P(S;Z), \mathbb{I}_P(X,S;Z) - \mathbb{I}_P(X;Y|S)\big\} \end{array} \right\}, \tag{4.4a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} \big(P_{SXYZ}, \Upsilon_{XSZY}\big) : \\ P_{SXYZ} = P_S P_X W_{YZ|XS} \\ \Upsilon_{SYZ} = P_S W_{YZ|S,X=x_0} \\ \mathbb{I}_P(X;Y|S) > \mathbb{I}_P(X;Z) \\ P_Z = \Upsilon_Z \end{array} \right\}. \tag{4.4b}$$

The covert capacity of the DMC $W_{YZ|XS}$ with a blind jammer is lower-bounded as

$$C_{\text{B-J}} \supseteq \text{conv}(\mathcal{A}), \tag{4.5}$$

where $\text{conv}(\mathcal{A})$ *is the convex hull of the set* $\mathcal{A}$*.*

Theorem 19 is proved in Appendix U.

Remark 11. *Theorem 19 results to positive rate for covert communication as long as the null-space of the legitimate receiver's channel is not a subset of the null-space of the warden's channel and the null-space of the warden's channel is not empty.*

We now provide an upper bound on the covert capacity. Note that in the problem described above the jammer is using a limited source of local randomness with rate R_K, which is shared between the jammer and the receiver, and the jammer does not know in which codeword blocks the transmitter is communicating with the receiver. To derive the upper bound, we use the fact that the covert capacity when the jammer knows in which blocks the transmitter is communicating with the receiver and the jammer uses an unlimited source of local randomness when the transmitter is not communicating with the receiver is not less than the covert capacity when the jammer does not know in which blocks the transmitter is communicating with the receiver and the jammer uses a limited source of local randomness. Hence, we derive an upper bound on the covert capacity when the jammer uses an unlimited source of local randomness for the no communication mode, by transmitting an i.i.d. sequence according to some distribution P_{S_2}, and therefore this upper bound is also an upper bound on the covert capacity when the jammer uses a limited amount of randomness, which is the problem that we study in this section. In this case, the distribution induced on the warden's observation is $Q_0^{\otimes n}$ where $Q_0(\cdot) = \sum_{s_2 \in \mathcal{S}_2} P_{S_2}(s_2) W_{Z|X=x_0,S}(\cdot|x_0, s_2)$.

Theorem 20. *Let*

$$\mathcal{A} = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists P_{QSXYZ} \in \mathcal{D} : \\ R \leq \mathbb{I}(X; Y|S, Q) \\ R_K \geq \max\left\{\mathbb{I}(X, S; Z|Q) - \mathbb{I}(X; Y|S, Q), \mathbb{I}(S; Z|Q)\right\} \end{array} \right\}, \tag{4.6a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{QSXYZ} : \\ P_{QSXYZ} = P_Q P_{S|Q} P_{X|Q} W_{YZ|XS} \\ \mathbb{I}(X;Y|S,Q) \geq \mathbb{I}(X;Z|Q) \\ P_Z = Q_0 \end{array} \right\}. \tag{4.6b}$$

The covert capacity of the DMC $W_{YZ|XS}$ with a blind jammer is upper-bounded as

$$C_{\text{B-J}} \subseteq \text{conv}(\mathcal{A}). \tag{4.7}$$

Theorem 20 is proved in Appendix V.

Now we provide two more achievable rate regions in which the jammer uses a strategy. The first strategy is that since the transmitter and the jammer share a secret key of negligible rate, the jammer knows in which blocks the transmitter is communicating with the receiver; the jammer stays silent when the transmitter is communicating and transmits otherwise [35, 58]. In this strategy, existence of shared secret key between the jammer and the receiver does not help the receiver because when the transmitter is communicating the jammer is silent and therefore its channel input does not interfere with the transmitter's signals. Hence, for this strategy, we assume that there is no shared secret key between the jammer and the receiver and the jammer has access to a source of local randomness with rate R_J. For this scenario, we can achieve the following inner bound on the covert capacity.

Theorem 21. *Let*

$$\mathcal{A} = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists \big(P_{SXYZ}, \Upsilon_{XSZY}\big) \in \mathcal{D} : \\ R < \mathbb{I}_P(X;Y) \\ R_J > \mathbb{I}_\Upsilon(S;Z) \end{array} \right\}, \tag{4.8a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} \big(P_{SXYZ}, \Upsilon_{XSZY}\big) : \\ P_{XYZ} = P_X W_{YZ|S=s_0,X} \\ \Upsilon_{SYZ} = P_S W_{YZ|S,X=x_0} \\ \mathbb{I}_P(X;Y) > \mathbb{I}_P(X;Z) \\ P_Z = \Upsilon_Z \end{array} \right\}. \tag{4.8b}$$

The covert capacity of the DMC $W_{YZ|XS}$ with a blind jammer is lower-bounded as

$$C_{\text{B-J}} \supseteq \text{conv}(\mathcal{A}). \tag{4.9}$$

Proof. Theorem 21 is proved in Appendix W. □

The second strategy is that, having the knowledge of the transmission block, the jammer transmits a codeword from its first codebook when the transmitter is communicating with the receiver and transmits a codeword from its second codebook otherwise. This strategy results to the following achievable region.

Theorem 22. *Let*

$$\mathcal{A} = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists \big(P_{QS_1XYZ}, \Upsilon_{QS_2ZY}\big) \in \mathcal{D} : \\ R < \mathbb{I}_P(X;Y|S_1,Q) \\ R_K > \max\big\{\mathbb{I}_P(X,S_1;Z|Q) - \mathbb{I}_P(X;Y|S_1,Q), \mathbb{I}_P(S_1;Z|Q), \mathbb{I}_\Upsilon(S_2;Z|Q)\big\} \end{array} \right\}, \tag{4.10a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} \big(P_{QS_1XYZ}, \Upsilon_{QS_2ZY}\big) : \\ P_{QS_1XYZ} = P_Q P_{S_1|Q} P_{X|Q} W_{YZ|XS} \\ \Upsilon_{QS_2YZ} = P_Q P_{S_2|Q} W_{YZ|S,X=x_0} \\ \mathbb{I}_P(X;Y|S_1,Q) > \mathbb{I}_P(X;Z|Q) \\ P_Z = \Upsilon_Z \end{array} \right\}. \tag{4.10b}$$

The covert capacity of the DMC $W_{YZ|XS}$ with a blind jammer is lower-bounded as

$$C_{\text{B-J}} \supseteq \text{conv}(\mathcal{A}). \tag{4.11}$$

Theorem 22 is proved in Appendix X.

Remark 12. *The region in Theorem 19 can be obtained from Theorem 22 by setting $P_{S_1} = P_{S_2}$. Also, the region in Theorem 21 can be obtained from Theorem 22 by setting $S_1 = \emptyset$.*

Remark 13 (When the Inner and Outer Bound Meet?)**.** *One can simply check that the achievability scheme in Theorem 22 meets the upper bound in Theorem 20 if the jammer has an unlimited source of local randomness and transmits an i.i.d. sequence when the transmitter is not communicating with the receiver and transmits a codeword from its codebook otherwise. This requires the transmitter and the jammer to share a secret key of negligible rate. In this case, the transmitter and the jammer can use the secret key of negligible rate to coordinate and use the strategy described above to achieve a higher covert rate.*

4.3.1 Examples

We compare the achievable rate regions in Theorem 19, Theorem 21, and Theorem 22 in the context of three examples and show that the scheme in Theorem 19 can perform better compared to when the jammer uses a strategy.

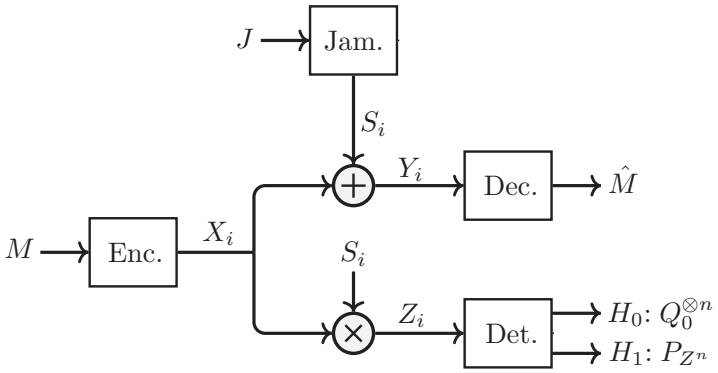

Figure 4.2. Noiseless binary Additive-multiplicative channel

Binary Additive-Multiplicative Channel

We first provide an example in which the schemes presented in Theorem 19, Theorem 21, and Theorem 22 results to the same positive rate for covert communications. Consider a scenario in which the channel inputs and outputs are all binary, the innocent symbols $x_0 = 0$ and $s_0 = 0$, and the channel rules are as follows (see Fig. 4.2),

$$Y = X \oplus S, \tag{4.12}$$

$$Z = X \otimes S. \tag{4.13}$$

Proposition 5. *The covert capacity for the example described above is,*

$$C_{\text{B-J}} = \Big\{(R, R_K) : \; R \leq 1, \; R_K \geq 0\Big\}. \tag{4.14}$$

Remark 14. *Intuitively speaking, in this channel, to enable covert communication between the transmitter and the receiver the jammer can force the warden's output to be always zero, without affecting the receiver's channel, by choosing $S = 0$.*

Proof. Here, we show that the region in Theorem 22, reduces to the region in Proposition 5. Showing that the region in Theorem 19 and Theorem 21 results to the same region is similar and is omitted for the sake of brevity.

Achievability Proof

Let the channel input X be a Bernoulli random variable with parameter $\alpha \in [\![0,1]\!]$ and the jammer's channel input S_1 when communication is happening be a Bernoulli random variable with parameter $\beta \in [\![0,1]\!]$ and the jammer's channel input S_2 when communication is not happening be a Bernoulli random variable with parameter $\eta \in [\![0,1]\!]$. To check the covertness constraint $P_Z = Q_0$ we have,

$$\begin{aligned} P_Z(z=0) &= \mathbb{P}(x=0,s=1) + \mathbb{P}(x=1,s=0) + \mathbb{P}(x=0,s=0) \\ &= \alpha(1-\beta) + \beta(1-\alpha) + \alpha\beta, \end{aligned} \tag{4.15}$$

$$Q_0(z=0) \overset{(a)}{=} 1, \tag{4.16}$$

where (a) follows by choosing $\eta = 1$. Therefore, the covertness constraint $P_Z = Q_0$, reduces to $\beta = 1$, which translates to $S = 0$. We now have,

$$\begin{aligned} \max_{P_X P_S} \mathbb{I}(X;Y|S) &= \max_{P_X P_S} \mathbb{H}(Y|S) \\ &= \max_{P_X P_S} \mathbb{H}(X \oplus S|S) \\ &= \max_{P_X P_S} \mathbb{H}(X|S) \\ &= \max_{P_X} \mathbb{H}(X) = 1. \end{aligned} \tag{4.17}$$

Also, since $S = 0$ and therefore $Z = 0$

$$\mathbb{I}(S;Z) = 0, \tag{4.18}$$

$$\mathbb{I}(X,S;Z) = 0. \tag{4.19}$$

Therefore, the constraint on the key rate in (4.4a) is also satisfied. Also, since $\mathbb{I}(X;Z) = 0$ the condition $\mathbb{I}(X;Y|S) > \mathbb{I}(X;Z)$ in (4.4b) is also satisfied.

Converse Proof

The converse proof is trivial, since it is not possible to communicate more than one bit for this channel and the covert capacity is always less than the capacity. □

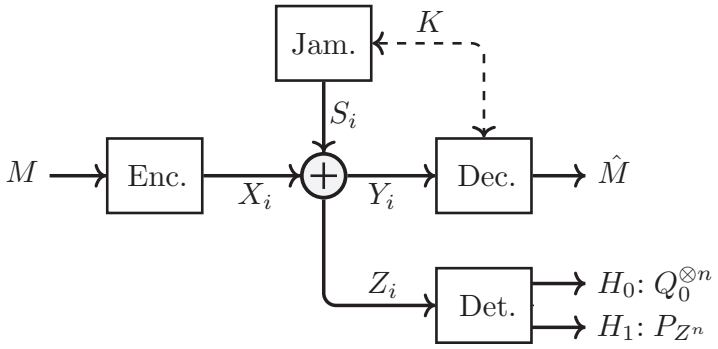

Figure 4.3. Noiseless binary Additive channel

Binary Additive Channel

We now provide an example in which the schemes presented in Theorem 19 and Theorem 22 perform better than the scheme in Theorem 21. Consider a scenario in which the channel inputs and outputs are all binary, the innocent symbols $x_0 = 0$ and $s_0 = 0$, and the channel rules are as follows (see Fig. 4.3),

$$Y = Z = X \oplus S. \tag{4.20}$$

Proposition 6. *The covert capacity for the example described above is lower-bounded as,*

$$C_{\text{B-J}} \supseteq \Big\{(R, R_K) : \, R < 1, \, R_K > 1\Big\}. \tag{4.21}$$

Proof. Let the channel input X be a Bernoulli random variable with parameter $\alpha \in [\![0, 0.5]\!]$, the jammer's input be a Bernoulli random variable with parameter $\beta \in [\![0, 0.5]\!]$. The covertness constraint $P_Z = \Upsilon_Z$ implies that,

$$P_Z(z = 0) = \mathbb{P}(x = 0, s = 0) + \mathbb{P}(x = 1, s = 1) = \alpha\beta + (1 - \alpha)(1 - \beta), \tag{4.22}$$

$$\Upsilon_Z(z = 0) = \mathbb{P}(s = 0) = \beta. \tag{4.23}$$

Therefore, the covertness constraint $P_Z = \Upsilon_Z$, reduces to $\beta = 0.5$. We now have,

$$\mathbb{I}_P(X; Y|S) = \mathbb{H}_P(Y|S)$$

$$= \mathbb{H}_P(X \oplus S|S)$$

$$= \mathbb{H}_P(X|S)$$

$$= \mathbb{H}_P(X) = \mathbb{H}_b(\alpha), \tag{4.24}$$

$$\mathbb{I}_\Upsilon(S;Z) = \mathbb{H}_\Upsilon(Z) - \mathbb{H}_\Upsilon(Z|S)$$

$$= \mathbb{H}_\Upsilon(x_0 \oplus S) - \mathbb{H}_\Upsilon(x_0 \oplus S|S)$$

$$= \mathbb{H}_\Upsilon(S) = 1, \tag{4.25}$$

$$\mathbb{I}_P(S;Z) = \mathbb{H}_P(Z) - \mathbb{H}_P(Z|S)$$

$$= \mathbb{H}_P(X \oplus S) - \mathbb{H}_P(X \oplus S|S) = 0, \tag{4.26}$$

$$\mathbb{I}_P(S;Z) = \mathbb{I}_P(S,X;Z) - \mathbb{I}_P(X;Y|S) = 0. \tag{4.27}$$

Therefore, the constraints in (4.4a) reduce to $R < 1$ and $R_K > 1$. □

We now show that the region in Theorem 21 does not result to positive covert rate for this example.

Proof. Note that in the strategy used in Theorem 21 in each block only one of the transmitter and the jammer uses the channel and the other user is silent. Let the channel input X be a Bernoulli random variable with parameter $\alpha \in [\![0, 0.5]\!]$, the jammer's input be a Bernoulli random variable with parameter $\beta \in [\![0, 0.5]\!]$. We now have,

$$\Upsilon_Z(z=0) = \mathbb{P}(s=0) = \beta, \tag{4.28}$$

$$P_Z(z=0) = \mathbb{P}(x=0) = \alpha. \tag{4.29}$$

Therefore, the covertness constraint $P_Z = \Upsilon_Z$, reduces to $\beta = \alpha$, which means that S and X should be generated with the same distribution. We now have,

$$\mathbb{I}_P(X;Y) = \mathbb{H}_P(Y) - \mathbb{H}_P(Y|X)$$

$$= \mathbb{H}_P(X \oplus 0)$$

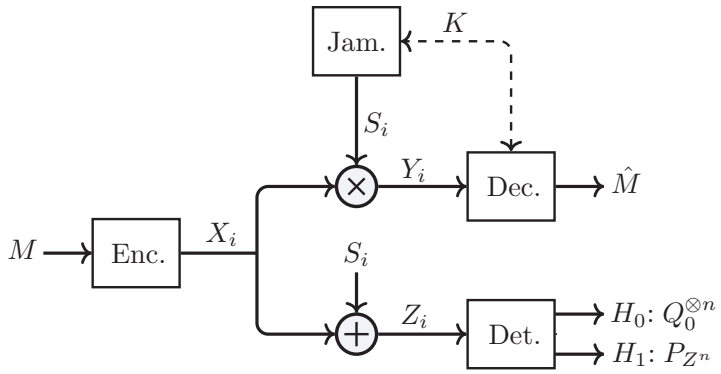

Figure 4.4. Noiseless binary Additive-multiplicative channel

$$= \mathbb{H}_b(\alpha). \tag{4.30}$$

We also have,

$$\begin{aligned} \mathbb{I}_\Upsilon(S;Z) &= \mathbb{H}_\Upsilon(Z) - \mathbb{H}_\Upsilon(Z|S) \\ &= \mathbb{H}_\Upsilon(S \oplus 0) \\ &= \mathbb{H}_b(\alpha). \end{aligned} \tag{4.31}$$

Similarly, one can show that,

$$\mathbb{I}_P(X;Z) = \mathbb{H}_b(\alpha). \tag{4.32}$$

Therefore, the condition $\mathbb{I}_P(X;Y) > \mathbb{I}_P(X;Z)$ in (4.8b) is *not* satisfied, and the strategy used in Theorem 21 does not result to positive rate for covert communication. □

Remark 15. *Similarly, one can show that the region in Proposition 6 can be achieved by using Theorem 22 by setting* $P_{S_1} = P_{S_2}$.

Binary Multiplicative-Additive Channel

Here we provide another example in which the schemes presented in Theorem 19 and Theorem 22 perform better than the scheme in Theorem 21. Consider a scenario in which the

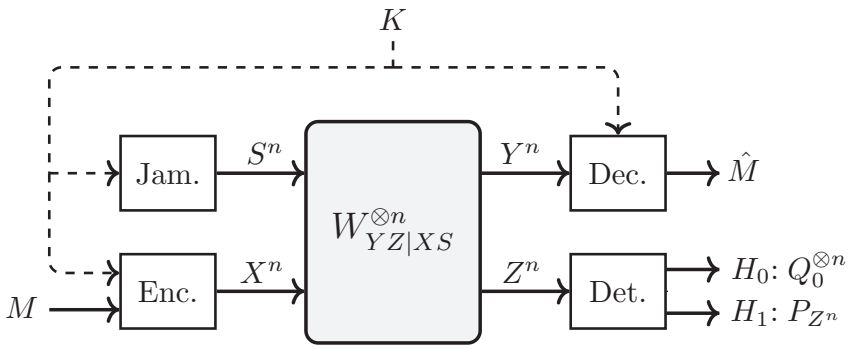

Figure 4.5. Covert communication in the presence of a jammer when there is a shared key between all the legitimate terminals

channel inputs and outputs are all binary, the innocent symbol $x_0 = 0$, and the channel rules are as follows (see Fig. 4.4),

$$Y = X \otimes S, \tag{4.33}$$

$$Z = X \oplus S. \tag{4.34}$$

The scheme in Theorem 19 results to the following achievable rate region.

Proposition 7. *The covert capacity for the example described above is lower-bounded as,*

$$C_{\text{B-J}} \supseteq \Big\{(R, R_K) : \; R < 0.5, \; R_K > 1\Big\}. \tag{4.35}$$

Proof. The proof is similar to the proof of the previous examples in this section and is omitted for the sake of brevity. □

Remark 16. *For this example one can also show that the regions in Theorem 21 and Theorem 22 results to zero rate.*

4.4 A Shared Key Between all the Legitimate Terminals

Now we study a problem in which there is a rate limited and uniformly distributed shared secret key $K \in \mathcal{K}$ between the legitimate terminals (i.e., the transmitter, the receiver, and

the jammer), this helps the transmitter to coordinate its channel input according to the jammer's channel input, and it also helps the receiver to cancel the interference caused by the jammer's channel input. In this problem, the transmitter and the jammer do not have access to any source of local randomness. Inner and outer bounds on the covert capacity of the DMC $W_{YZ|XS}$ when there is a shared secret key between the transmitter, the receiver and the jammer are established in the following theorems.

Theorem 23. *Let*

$$\mathcal{A} = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \left(P_{S_1XYZ}, \Upsilon_{S_2ZY}\right) \in \mathcal{D} : \\ R < \mathbb{I}_P(X; Y|S_1) \\ R_K > \max\{\mathbb{I}_P(X, S_1; Z) - \mathbb{I}_P(X; Y|S_1), \mathbb{I}_P(S_1; Z), \mathbb{I}_\Upsilon(S_2; Z)\} \end{array} \right\}, \tag{4.36a}$$

where

$$\mathcal{D} = \left\{ \begin{array}{l} \left(P_{S_1XYZ}, \Upsilon_{S_2ZY}\right) : \\ P_{S_1XYZ} = P_{S_1} P_{X|S_1} W_{YZ|XS_1} \\ \Upsilon_{S_2ZY} = P_{S_2} W_{YZ|X=x_0,S_2} \\ P_Z = \Upsilon_Z \end{array} \right\}. \tag{4.36b}$$

The covert capacity of the DMC $W_{YZ|XS}$ when there is a shared key between all the legitimate terminals is lower bounded as

$$C_{\text{F-K}} \supseteq \text{conv}(\mathcal{A}). \tag{4.37}$$

Theorem 23 is proved in Appendix Y.

We now provide an upper bound on the covert capacity when there is a shared secret key between the legitimate terminals. Note that in the problem described above the jammer is using a limited amount of randomness with rate R_K, which is shared between all the legitimate terminals, and therefore the legitimate terminals can coordinate that is the jammer

and the receiver know in which codeword blocks the transmitter is communicating with the receiver. Similar to the Theorem 20, to derive the upper bound we use the fact that the covert capacity when the jammer uses an unlimited amount of randomness when the transmitter is not communicating with the receiver is not less than the covert capacity when the jammer uses a limited amount of randomness. Hence, we derive an upper bound on the covert capacity when the jammer uses an unlimited amount of randomness for the no communication mode and therefore this upper bound is also an upper bound on the covert capacity when the jammer uses a limited amount of randomness, which is the problem that we study in this section.

Theorem 24. *Let*

$$\mathcal{A} = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists P_{SXYZ} \in \mathcal{D} : \\ R \leq \mathbb{I}(X; Y|S) \\ R_K \geq \max\{\mathbb{I}(X, S; Z) - \mathbb{I}(X; Y|S), \mathbb{I}(S; Z)\} \end{array} \right\}, \tag{4.38a}$$

where

$$\mathcal{D} = \left\{ \begin{array}{l} P_{SXYZ} : \\ P_{SXYZ} = P_S P_{X|S} W_{YZ|XS} \\ P_Z = Q_0 \end{array} \right\}. \tag{4.38b}$$

The covert capacity of the DMC $W_{YZ|XS}$ when there is a shared key between all the legitimate terminals is upper bounded by

$$C_{\text{F-K}} \subseteq \text{conv}(\mathcal{A}). \tag{4.39}$$

Theorem 24 is proved in Appendix Z.

Remark 17 (When the Inner and Outer Bound Meet?)**.** *Similar to Remark 13, the achievability scheme in Theorem 23 meets the upper bound in Theorem 24 if the jammer has*

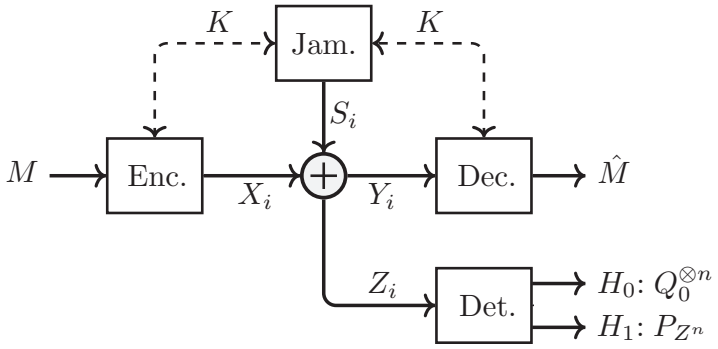

Figure 4.6. Binary Symmetric Additive Channel

unlimited source of local randomness. In this case, the jammer transmits an i.i.d. sequence when the transmitter is not communicating with the receiver and transmits a sequence from its codebook when communication is happening.

4.4.1 Examples

Here, we provide two examples in which Theorem 23 leads to a positive rate for covert communication.

Binary Additive Channel

Consider a scenario in which the channel inputs and outputs are all binary, the innocent channel input symbol $x_0 = 0$, and the channel rules are as follows, as it can be seen in Fig. 4.6,

$$Y = Z = X \oplus S. \tag{4.40}$$

Proposition 8. *The covert capacity for the example described above is,*

$$C_{\text{F-K}} \supseteq \text{conv} \left\{ \begin{array}{l} (R, R_K) : \alpha, \beta, \eta \in (0 : 0.5) \\ R < (\alpha+\beta)\mathbb{H}_b\left(\frac{\alpha}{\alpha+\beta}\right) + (1-\alpha-\beta)\mathbb{H}_b\left(\frac{\eta}{1-\alpha-\beta}\right) \\ R_K > \mathbb{H}_b(\alpha+\eta) \end{array} \right\}. \tag{4.41}$$

Table 4.1. Joint probability distribution between X and S

S_1 \ X	0	1
0	α	β
1	$1-\alpha-\beta-\eta$	η

Remark 18. *Intuitively speaking, in this channel, since the transmitter has access to the jammer's channel input through the shared key it chooses the channel input X^n such that after it adds up with the jammer's channel input the results look like it has been generated according to P_S. The receiver can recover X^n since it has access to S^n through the shared secret key, but the warden cannot distinguish it is output (i.e., $S^n \oplus X^n$) from S^n.*

Proof. Without loss of generality let's assume the joint probability distribution between X and S_1 is according to Table 4.1, and S_2 be a Bernoulli random variable independent of S_1 and X with parameter $\lambda \in [\![0,1]\!]$. We now have,

$$P_Z(z=0) = \mathbb{P}(x=0, s_1=0) + \mathbb{P}(x=1, s_1=1) = \alpha + \eta, \tag{4.42}$$

$$\Upsilon_Z(z=0) = \mathbb{P}(s_2=0) = \lambda. \tag{4.43}$$

Therefore, the covertness constraint $P_Z = \Upsilon_Z$, reduces to $\alpha + \eta = \lambda$. We now have,

$$\begin{aligned}
\mathbb{I}_P(X;Y|S_1) &= \mathbb{H}_P(Y|S_1) \\
&= \mathbb{H}_P(X \oplus S_1|S_1) \\
&= \mathbb{H}_P(X|S_1) \\
&= \mathbb{P}(S_1=0)\mathbb{H}_P(X|S_1=0) + \mathbb{P}(S_1=1)\mathbb{H}_P(X|S_1=1) \\
&= (\alpha+\beta)\mathbb{H}_b\left(\frac{\alpha}{\alpha+\beta}\right) + (1-\alpha-\beta)\mathbb{H}_b\left(\frac{\eta}{1-\alpha-\beta}\right), \qquad (4.44) \\
\mathbb{I}_\Upsilon(S_2;Z) &= \mathbb{H}_\Upsilon(Z) - \mathbb{H}_\Upsilon(Z|S_2) \\
&= \mathbb{H}_\Upsilon(x_0 \oplus S_2) - \mathbb{H}_\Upsilon(x_0 \oplus S_2|S_2) = \mathbb{H}(S_2) = \mathbb{H}_b(\alpha+\eta). \qquad (4.45)
\end{aligned}$$

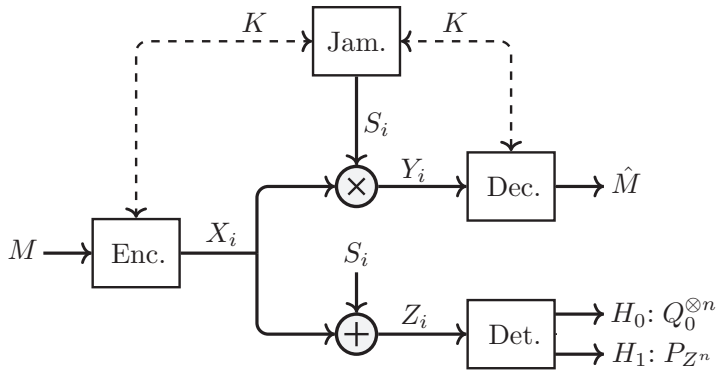

Figure 4.7. Binary Multiplicative-Additive Channel

Since $Y = Z$,

$$\begin{aligned}\mathbb{I}_P(X, S_1; Z) - \mathbb{I}_P(X; Y|S_1) &= \mathbb{I}_P(S_1; Z) \\ &= \mathbb{H}_P(Z) - \mathbb{H}_P(Z|S_1) \\ &= \mathbb{H}_P(\alpha + \eta) - \mathbb{H}_P(X|S_1). \end{aligned} \tag{4.46}$$

From (4.45) and (4.46), the constraint on the key rate in (4.36a) reduces to $R_K > \mathbb{H}_b(\alpha + \eta)$. □

Binary Multiplicative-Additive Channel

Consider a scenario in which the channel inputs and outputs are all binary, the innocent channel input symbol $x_0 = 0$, and the channel rules are as follows, as it can be seen in Fig. 4.7,

$$Y = X \otimes S, \tag{4.47}$$

$$Z = X \oplus S. \tag{4.48}$$

Table 4.2. Joint probability distribution between X and S

S_1 \ X	0	1
0	α	β
1	$1-\alpha-\beta-\eta$	η

Proposition 9. *The covert capacity for the example described above is,*

$$C_{\text{F-K}} \supseteq \text{conv}\left\{\begin{array}{l}(R, R_K) : \alpha, \beta \in (0 : 0.5) \\ R < (1-\alpha-\beta)\mathbb{H}_b\left(\frac{\eta}{1-\alpha-\beta}\right) \\ R_K > \mathbb{H}_b(\alpha+\eta)\end{array}\right\}. \tag{4.49}$$

Proof. Without loss of generality let's assume the joint probability distribution between X and S_1 is according to the Table 4.2 and S_2 be a Bernoulli random variable independent of S_1 and X with parameter $\lambda \in [\![0, 1]\!]$. We now have,

$$P_Z(z=0) = \mathbb{P}(x=0, s_1=0) + \mathbb{P}(x=1, s_1=1) = \alpha + \eta, \tag{4.50}$$

$$\Upsilon_Z(z=0) = \mathbb{P}(s_2=0) = \lambda. \tag{4.51}$$

Therefore, the covertness constraint $P_Z = \Upsilon_Z$, reduces to $\alpha + \eta = \lambda$. We now have,

$$\begin{aligned}\mathbb{I}_P(X;Y|S_1) &= \mathbb{H}_P(Y|S_1) \\ &= \mathbb{H}_P(X \otimes S_1|S_1) \\ &= \mathbb{P}(s_1=1)\mathbb{H}_P(X|s_1=1) \\ &= (1-\alpha-\beta)\mathbb{H}_b\left(\frac{\eta}{1-\alpha-\beta}\right).\end{aligned} \tag{4.52}$$

Also, choosing $\alpha = \beta = \eta = 0.25$ leads to,

$$\begin{aligned}\mathbb{I}_\Upsilon(S_2;Z) &= \mathbb{H}_\Upsilon(Z) - \mathbb{H}_\Upsilon(Z|S_2) \\ &= \mathbb{H}_\Upsilon(x_0 \oplus S_2) - \mathbb{H}_\Upsilon(x_0 \oplus S_2|S_2) = \mathbb{H}_\Upsilon(S_2) = \mathbb{H}_b(\alpha+\eta),\end{aligned} \tag{4.53}$$

$$\mathbb{I}_P(S_1;Z) = \mathbb{H}_P(Z) - \mathbb{H}_P(Z|S_1)$$

Table 4.3. Joint probability distribution between X and S

S_1 \ X	0	1
0	α	β
1	$1-\alpha-\beta-\eta$	η

$$= \mathbb{H}_b(\alpha+\eta) - \mathbb{H}_P(X|S_1), \tag{4.54}$$

$$\mathbb{I}_P(X, S_1; Z) = \mathbb{H}_P(Z) = \mathbb{H}_b(\alpha+\eta). \tag{4.55}$$

Therefore, the constraint on the key rate in (4.36a) should be $R_K > \mathbb{H}_b(\alpha+\eta)$. □

Binary Additive-Multiplicative Channel

Consider a scenario in which the channel inputs and outputs are all binary, the innocent channel input symbol $x_0 = 0$, and the channel rules are as follows,

$$Y = X \oplus S, \tag{4.56}$$

$$Z = X \otimes S. \tag{4.57}$$

Proposition 10. *The covert capacity for the example described above is lower bounded by,*

$$C_{\text{F-K}} = 1 \tag{4.58}$$

Remark 19. *Intuitively speaking, in this channel, since $x_0 = 0$ the jammer's output should always be equal to zero therefore the transmitter and the jammer should not transmit symbol* 1 *at the same time.*

Proof. Here we provide the achievability proof, the converse proof is trivial since it is not possible to communicate more than one bit for this channel.

Achievability Proof: Without loss of generality let's assume the joint probability distribution between X and S_1 is according to the Table 4.3, and S_2 be a Bernoulli random variable independent of S_1 and X with parameter $\lambda \in [\![0, 1]\!]$. To analyze the covertness we have,

$$P_Z(z=0) = \mathbb{P}(x=0, s_1=0) + \mathbb{P}(x=1, s_1=0) + \mathbb{P}(x=0, s_1=1) = 1-\eta, \tag{4.59}$$

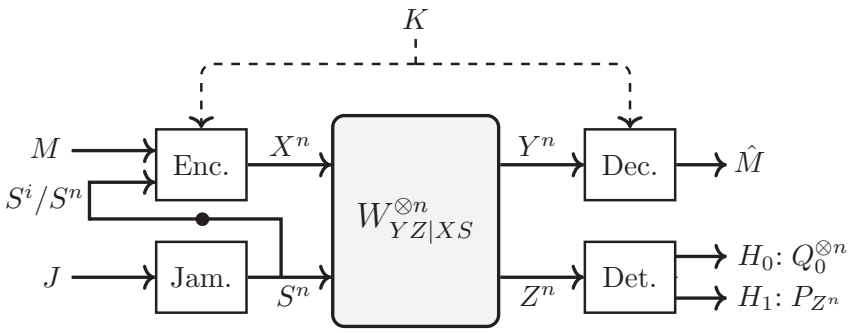

Figure 4.8. Covert communication by access to cooperative jammer's codeword

$$\Upsilon_Z(z=0) = \mathbb{P}(x_0 \otimes S_2 = 0) = 1. \tag{4.60}$$

Therefore, the covertness constraint $P_Z = \Upsilon_Z$*, reduces to* $\eta = 0$*, which means that the warden's output should always be zero. We now have,*

$$\begin{aligned}\mathbb{I}_P(X;Y|S_1) &= \mathbb{H}_P(Y|S_1) = \mathbb{H}(X \oplus S_1|S_1) = \mathbb{H}(X|S_1) \\ &= (\alpha+\beta)\mathbb{H}_b\left(\frac{\alpha}{\alpha+\beta}\right).\end{aligned} \tag{4.61}$$

Also, since we force the warden's output to be always zero we have,

$$\mathbb{I}_\Upsilon(S_2;Z) = \mathbb{H}_\Upsilon(Z) - \mathbb{H}_\Upsilon(Z|S_2) \overset{(a)}{=} 0, \tag{4.62}$$

$$\mathbb{I}_P(S_1;Z) \overset{(b)}{=} 0, \tag{4.63}$$

$$\mathbb{I}_P(X,S_1;Z) = \mathbb{H}_P(Z) \overset{(c)}{=} 0, \tag{4.64}$$

where (a) *follows since* $x_0 = 0$*, and* (b) *and* (c) *follow since* $\eta = 0$*. Therefore, the constraint on the key rate in* (4.36a) *reduces to* $R_K > 0$*. Also, by setting* $\alpha = \beta = 0.5$*, we have* $R < 1$*. We can also set* $\lambda = 1$ *so that the jammer does not use any source of local randomness when communication is not happening.* □

4.5 Jammer's Output Available at the Transmitter

In this section we study a case in which the jammer's output is available non-causally or causally at the transmitter so that the transmitter can coordinate its channel input according

to the jammer's codeword, as seen in Fig. 4.8. Here, the transmitter and the receiver are assumed to have access to a rate limited and uniformly distributed shared secret key $K \in \mathcal{K}$, this helps the transmitter to communicate covertly with the receiver even if the channel from the transmitter to the receiver is noisier than the channel from the transmitter to the warden.

4.5.1 Non-Causal Case

In this subsection, we study the problem described above when the transmitter has non-causal access to the jammer's output. Here, we assume that both the transmitter and the jammer have access to local randomness.

One-Sided MAC Resolvability Lemma

The achievable rate region when the transmitter has non-causal access to the jammer's output is based on Lemma 4 below. This lemma describes the rate required for a codebook exciting *one of* the inputs of a MAC so that the output distribution is indistinguishable from that arising from a random excitation of the same input, *while the other MAC input is being excited at the same time by a codebook with an arbitrary, prescribed rate.* A key distinction of this result from the usual resolvability results is that the target distribution may not be i.i.d..

We begin by characterizing the setup for this lemma. Consider a discrete memoryless MAC $(\mathcal{X}_1 \times \mathcal{X}_2, W_{Z|X_1X_2}, \mathcal{Z})$ over which two encoders transmit codewords as in Fig. 4.9. Let $C_i \triangleq \left\{X_i^n(m_i)\right\}_{m_i \in \mathcal{M}_i}$, where $\mathcal{M}_i = [\![1, 2^{nR_i}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_{X_i}^{\otimes n}$, for $i = 1, 2$. We denote a realization of C_i by $\mathcal{C}_i \triangleq \left\{x_i^n(m_i)\right\}_{m_i \in \mathcal{M}_i}$. The codebook construction described above induces the PMF λ for the codebooks.

$$\lambda(\mathcal{C}_1, \mathcal{C}_2) = \prod_{m_1 \in \mathcal{M}_1} \prod_{m_2 \in \mathcal{M}_2} P_{X_1}^{\otimes n}\big(x_1^n(m_1)\big) P_{X_2}^{\otimes n}\big(x_2^n(m_2)\big).$$

We now consider two scenarios, under *both* of which Transmitter 2 emits a codeword chosen randomly and uniformly from the random codebook C_2. In the first scenario, Transmitter 1 emits an i.i.d. sequence according to P_{X_1}. The distribution induced at the output of the channel is,

$$P_{Z^n|C_2}(z^n) \triangleq \frac{1}{2^{nR_2}} \sum_{m_2=1}^{2^{nR_2}} W_{Z|X_2}^{\otimes n}\big(z^n|X_2^n(m_2)\big), \tag{4.65}$$

where

$$W_{Z|X_2}(z|x_2) \triangleq \sum_{x_1 \in \mathcal{X}_1} P(x_1) W_{Z|X_1X_2}(z|x_1, x_2). \tag{4.66}$$

In the second scenario, Transmitter 1 emits a codeword uniformly at random from a random codebook C_1. The distribution induced at the channel output is,

$$P_{Z^n|C_1,C_2}(z^n) \triangleq \frac{1}{2^{n(R_1+R_2)}} \sum_{m_1=1}^{2^{nR_1}} \sum_{m_2=1}^{2^{nR_2}} W_{Z|X_1X_2}^{\otimes n}\big(z^n|X_1^n(m_1), X_2^n(m_2)\big). \tag{4.67}$$

We wish to find conditions under which the distributions induced at the channel output in the two scenarios are approximately equal. We call this problem *one-sided MAC resolvability.*

Definition 10. *A rate pair (R_1, R_2) is achievable for the one-sided resolvability of the discrete memoryless MAC $(\mathcal{X}_1 \times \mathcal{X}_2, W_{Z|X_1X_2}, \mathcal{Z})$ if for a given $W_{Z|X_1X_2}$ there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes such that $\mathbb{E}_{C_1,C_2}\Big[\mathbb{D}\big(P_{Z^n|C_1,C_2}||P_{Z^n|C_2}\big)\Big] \underset{n\to\infty}{\longrightarrow} 0$. The one-sided MAC resolvability region $\mathcal{R}$ is the convex hull of the set of all achievable rate pairs (R_1, R_2).*

The main difference between the resolvability region in Definition 10 and the standard resolvability defined in [82, 83, 77] is that in this problem the target distribution $P_{Z^n|C_2}$ at the output of channel is not necessarily i.i.d.. We now find sufficient conditions on the size of the two codebooks such that the distributions induced at the channel output in the two scenarios in Eq. (4.67) and (4.65) are approximately equal in terms of expected KL divergence.

Lemma 4. *For a discrete memoryless MAC, $W_{Z|X_1X_2}$ if (R_1, R_2) belongs to*

$$\bigcup_{P_{X_1}P_{X_2}} (\mathcal{R}_1 \cup \mathcal{R}_2), \tag{4.68}$$

where

$$\mathcal{R}_1 = \left\{ \begin{array}{l} (R_1, R_2) \in \mathbb{R}_+^2 : \\ R_1 > \mathbb{I}(X_1; Z) \\ R_2 > \mathbb{I}(X_2; Z) \\ R_1 + R_2 > \mathbb{I}(X_1, X_2; Z) \end{array} \right\}, \tag{4.69a}$$

$$\mathcal{R}_2 = \left\{ \; (R_1, R_2) \in \mathbb{R}_+^2 : \; R_1 > \mathbb{I}(X_1; Z|X_2) \; \right\}, \tag{4.69b}$$

then

$$\mathbb{E}_{C_1,C_2}\Big[\mathbb{D}\big(P_{Z^n|C_1,C_2}||P_{Z^n|C_2}\big)\Big] \underset{n\to\infty}{\longrightarrow} 0. \tag{4.70}$$

Lemma 4 is proved in Appendix AA.

Remark 20. *The region $\mathcal{R}_1$ is the channel resolvability region for MAC. Since X_1 and X_2 are independent therefore $\mathbb{I}(X_1; Z|X_2) = \mathbb{I}(X_1; X_2, Z)$, and the region $\mathcal{R}_2$ can be viewed as the resolvability region of a MAC against a wiretapper who has full access to the channel input X_2 while the first transmitter does not have access to X_2.*

Remark 21. *A related result [23, Theorem 3] states that if* (4.68) *holds then $\mathbb{E}_{C_1,C_2}\mathbb{V}\big(P_{Z^n|C_1,C_2}, P_{Z^n|C_2}\big) \underset{n\to\infty}{\longrightarrow} 0$. However, no proof is publicly available for [23, Theorem 3].*

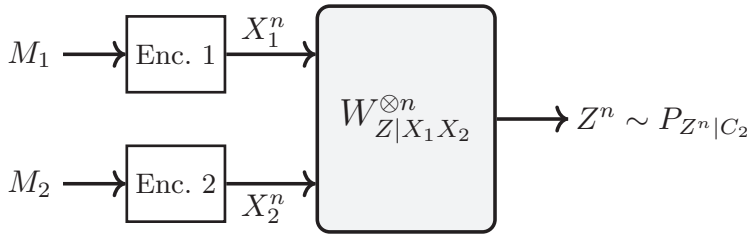

Figure 4.9. Distribution Approximation in MAC

Achievable Rate Region

Theorem 25. *Let*

$$\mathcal{A} \triangleq \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists\, (P_{SUVXYZ}, \Upsilon_{SYZ}) \in \mathcal{D} : \\ R < \mathbb{I}_P(U;Y) - \mathbb{I}_P(U;V|S) \\ R_K > \max\left\{\mathbb{I}_P(U;Z), \mathbb{I}_P(U,S;Z) - R_J\right\} - \mathbb{I}_P(U;Y) \\ R_J > \mathbb{I}_\Upsilon(S;Z) \end{array} \right\}, \tag{4.71a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} (P_{SUVXYZ}, \Upsilon_{SYZ}) : \\ P_{SUVXYZ} = P_S P_U P_{V|US} \mathbb{1}_{\left\{X = X(U,S)\right\}} W_{YZ|XS} \\ \Upsilon_{SYZ} = P_S W_{YZ|X=x_0,S} \\ P_Z = \Upsilon_Z \\ |\mathcal{U}| \leq |\mathcal{X}| + 5 \\ |\mathcal{V}| \leq |\mathcal{X}| + 3 \end{array} \right\}. \tag{4.71b}$$

The covert capacity of the DMC $W_{YZ|XS}$ when the transmitter has non-causal access to the jammer's output is lower-bounded as

$$C_{\text{CJ-NC}} \supseteq \text{conv}(\mathcal{A}). \tag{4.72}$$

Theorem 25 is proved in Appendix AB.

Remark 22. *The achievable rate region in Theorem 25 is still valid for the scenario in which the transmitter has access to the jammer's input, i.e., the dummy messages J.*

Remark 23. *Two extremal cases for the jammer rate are instructive and are considered next. First, if the jammer has maximal rate $R_J = H(S)$, the key rate requirements are the same as in the problem of covert communication over a state-dependent channel [37, 80]. Since U and S are independent, one can show that the constraint on the key rate in* (4.71a) *reduces to:*

$$R_K > \mathbb{I}_P(U;Z) - \mathbb{I}_P(U;Y). \tag{4.73}$$

This is the condition that has been derived in [37, Eq. (8)]. Second, if we set R_J as its minimum, i.e., $R_J = \mathbb{I}_P(S;Z) + \epsilon$, and since U and S are independent, one can show that the constraint on the key rate in (4.71a) *reduces to*

$$R_K > \mathbb{I}_P(U;Z|S) - \mathbb{I}_P(U;Y). \tag{4.74}$$

Since $\mathbb{I}_P(U;Z|S) \geq \mathbb{I}_P(U;Z)$, if the size of jammer's codebook is decreased, a higher rate is needed for the secret key shared between the legitimate terminals. From the constraint on the key rate in (4.71a)*, the smallest jammer codebook allowing minimal secret key rate is $R_J = \mathbb{I}_P(S;Z|U)$.*

Achievable Rate Region for Both Covert and Secure Communication

To prove Theorem 25 we employ Gel'fand-Pinsker encoding by using the likelihood encoder [77, 78, 79] to coordinate the transmitter's codeword according to the jammer's codeword; however unlike the channel's with i.i.d. state in this problem the jammer's codeword is not i.i.d. and the likelihood encoder is designed for i.i.d. sources. To overcome this issue, we randomize the jammer's codeword at the transmitter and then coordinate the transmitter's codeword according to the randomized version of the jammer's codeword by using the likelihood encoder. This shows up in the expression of the penalty term in the covert rate and

the distribution that we are optimizing over in Theorem 25. However, by adding the security constraint to make the communication both covert, $\lim_{n\to\infty}\mathbb{D}(P_{Z^n}||\Upsilon_{Z^n})\to 0$, and secure, $\lim_{n\to\infty}\mathbb{I}_P(M;Z^n)\to 0$, similar to [84, Theorem 1] one can use the likelihood encoder to align the transmitter's codeword according to the jammer's codeword. This is because the jammer's signal does not convey any information, and the security constraint is more restrictive than the covert constraint in this particular problem. The following theorem provides an achievable rate region for both covert and secure communication of the problem studied in Theorem 25.

Theorem 26. *Let*

$$\mathcal{A}\triangleq\left\{\begin{array}{l}(R,R_K)\geq 0:\exists\,(P_{SUXYZ},\Upsilon_{SYZ})\in\mathcal{D}:\\ R<\mathbb{I}_P(U;Y)-\mathbb{I}_P(U;S)\\ R_K>\mathbb{I}_P(U;S,Z)-\mathbb{I}_P(U;Y)\\ R_J>\mathbb{I}_\Upsilon(S;Z)\end{array}\right\},\tag{4.75a}$$

where,

$$\mathcal{D}\triangleq\left\{\begin{array}{l}(P_{SUXYZ},\Upsilon_{SYZ}):\\ P_{SUXYZ}=P_SP_{U|S}\mathbb{1}_{\left\{X=X(U,S)\right\}}W_{YZ|XS}\\ \Upsilon_{SYZ}=P_SW_{YZ|X=x_0,S}\\ P_Z=\Upsilon_Z\\ |\mathcal{U}|\leq|\mathcal{X}|+3\end{array}\right\}.\tag{4.75b}$$

The covert and secure capacity of the DMC $W_{YZ|XS}$ when the transmitter has non-causal access to the jammer's output is lower-bounded as

$$C_{\text{CJ-NC}}\supseteq\text{conv}(\mathcal{A}).\tag{4.76}$$

The proof for Theorem 26 is similar to that of [84, Theorem 1] and is available in Appendix AC.

Upper Bound

We now provide an upper bound on the covert capacity when there is a shared secret key of negligible rate between the transmitter and the jammer so that they can coordinate. Similar to the upper bounds in the previous sections, we provide an upper bound on the covert capacity when the jammer knows in which blocks communication is happening, through the secret key, and uses an unlimited amount of local randomness when the transmitter is not communicating with the receiver. The covert capacity in this case is not less than the covert capacity when the jammer uses a limited amount of local randomness and does not know when the transmitter is communicating. Hence, this upper bound is also an upper bound on the covert capacity when the jammer uses a limited amount of randomness, which is the problem that we study in this section.

Theorem 27. *Let*

$$\mathcal{A} = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists P_{UVSXYZ} \in \mathcal{D} : \\ R \leq \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V) \\ R_K \geq \mathbb{I}(V;Z) - \mathbb{I}(V;S) - \mathbb{I}(U,V;Y) + \mathbb{I}(U;S|V) \\ R_K + R_J \geq \mathbb{I}(V;Z) - \mathbb{I}(U,V;Y) + \mathbb{I}(U;S|V) \\ R_J \geq \mathbb{I}(S;Z) \end{array} \right\}, \tag{4.77a}$$

where

$$\mathcal{D} = \left\{ \begin{array}{l} P_{UVSXYZ} : \\ P_{UVSXYZ} = P_{SUV} \mathbb{1}_{\left\{X = X(U,S)\right\}} W_{YZ|XS} \\ P_Z = Q_0 \\ \max\{|\mathcal{U}|, |\mathcal{V}|\} \leq |\mathcal{X}| + 3 \end{array} \right\}. \tag{4.77b}$$

The covert capacity of the DMC $W_{YZ|XS}$ when the transmitter has non-causal access to the jammer's output is upper-bounded as

$$C_{\text{IJ-NC}} \subseteq \text{conv}(\mathcal{A}). \tag{4.78}$$

Theorem 27 is proved in Appendix AD.

4.5.2 Causal Case

We now study the problem of covert communication when the transmitter has causal access to the jammer's codeword. Here, we assume that the transmitter does not have access to any source of local randomness, but the jammer has access to a limited amount of local randomness.

Achievable Rate Region

The following theorem establishes an achievable rate region for the problem described above.

Theorem 28. *Let*

$$\mathcal{A} \triangleq \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists\, (P_{SUXYZ}, \Upsilon_{SYZ}) \in \mathcal{D} : \\ R < \mathbb{I}_P(U;Y) \\ R_K > \max\left\{\mathbb{I}_P(U;Z), \mathbb{I}_P(U,S;Z) - R_J\right\} - \mathbb{I}_P(U;Y) \\ R_J > \mathbb{I}_\Upsilon(S;Z) \end{array} \right\}, \tag{4.79a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} (P_{SUXYZ}, \Upsilon_{SYZ}) : \\ P_{SUXYZ} = P_S P_U \mathbb{1}_{\left\{X = X(U,S)\right\}} W_{YZ|XS} \\ \Upsilon_{SYZ} = P_S W_{YZ|X=x_0,S} \\ P_Z = \Upsilon_Z \\ |\mathcal{U}| \leq |\mathcal{X}| + 5 \end{array} \right\}. \tag{4.79b}$$

The covert capacity of the DMC $W_{YZ|XS}$ when the transmitter has causal access to the jammer's output is lower-bounded as

$$C_{\text{IJ-C}} \supseteq \text{conv}(\mathcal{A}). \tag{4.80}$$

Theorem 28 is proved in Appendix AE.

Remark 24. *Two extremal cases for the jammer rate are instructive and are considered next. First, if the jammer has maximal rate $R_J = H(S)$, the region in Theorem 28 reduces to the region for the problem of covert communication over a state-dependent channel [37, 80]. Since U and S are independent, one can show that the constraint on the key rate in* (4.79a) *reduces to:*

$$R_K > \mathbb{I}_P(U;Z) - \mathbb{I}_P(U;Y). \tag{4.81}$$

This is the condition that has been derived in [37, Eq. (8)]. Second, if we set R_J as its minimum, i.e., $R_J = \mathbb{I}_P(S;Z) + \epsilon$, and since U and S are independent, one can show that the constraint on the key rate in (4.79a) *reduces to*

$$R_K > \mathbb{I}_P(U;Z|S) - \mathbb{I}_P(U;Y). \tag{4.82}$$

Since $\mathbb{I}_P(U;Z|S) \geq \mathbb{I}_P(U;Z)$, if the size of jammer's codebook is decreased, a higher rate is needed for the secret key shared between the legitimate terminals. From the constraint on the key rate in (4.79a)*, the smallest jammer codebook allowing minimal secret key rate is $R_J = \mathbb{I}_P(S;Z|U)$.*

Upper Bound

Similar to the previous sections, we provide an upper bound on the covert capacity when there is a shared secret key of negligible rate between the transmitter and the jammer so that they can coordinate, and the jammer uses an unlimited amount of local randomness when the transmitter is not communicating with the receiver. The covert capacity in this

case is not less than the covert capacity when the jammer uses a limited amount of local randomness and does not know when the transmitter is communicating. Hence, this upper bound is also an upper bound on the covert capacity when the jammer uses a limited amount of randomness, which is the problem that we study in this section.

Theorem 29. *Let*

$$\mathcal{A} = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists P_{UVSXYZ} \in \mathcal{D} : \\ R \leq \mathbb{I}(U;Y) \\ R_K \geq \mathbb{I}(V;Z) - \mathbb{I}(U;Y) \\ R_J \geq \mathbb{I}(S;Z) \end{array} \right\}, \tag{4.83a}$$

where

$$\mathcal{D} = \left\{ \begin{array}{l} P_{UVSXYZ} : \\ P_{UVSXYZ} = P_{SUV} \mathbb{1}_{\left\{X = X(U,S)\right\}} W_{YZ|XS} \\ P_Z = Q_0 \\ \max\{|\mathcal{U}|, |\mathcal{V}|\} \leq |\mathcal{X}| + 3 \end{array} \right\}. \tag{4.83b}$$

The covert capacity of the DMC $W_{YZ|XS}$ when the transmitter has causal access to the jammer's output is upper-bounded as

$$C_{\text{IJ-NC}} \subseteq \text{conv}(\mathcal{A}). \tag{4.84}$$

Theorem 29 is proved in Appendix AF.

4.5.3 Examples

Here, we provide an example in which Theorem 28 leads to a positive rate for covert communication. Consider a scenario in which the channel inputs and outputs are all binary,

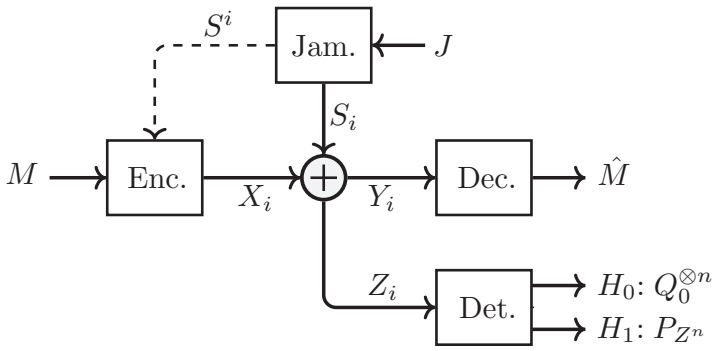

Figure 4.10. Binary Symmetric Additive Channel

the innocent channel input symbol $x_0 = 0$, the jammer's output is available causally at the transmitter, and the channel rules are as follows, as it can be seen in Fig. 4.10,

$$Y = Z = X \oplus S. \tag{4.85}$$

Proposition 11. *The covert capacity for the example described above is lower bounded as,*

$$C_{\text{F-K}} \supseteq \text{conv} \left\{ \begin{array}{l} (R, R_J, R_K) : \alpha \in [\![0,1]\!] \\ R < \mathbb{H}_b(\alpha) \\ R_J > \mathbb{H}_b(\alpha) \\ R_K > 0 \end{array} \right\}. \tag{4.86}$$

Remark 25. *Intuitively speaking, in this channel, since for the transmitter has access to the jammer's channel input it chooses the channel input X to be $U \oplus S$ therefore $Y = Z = U$ since the distribution of U is same as S the jammer cannot figure out whether communication is happening or not. But, since the transmitter and the legitimate receiver share a secret key, the receiver knows when communication is happening.*

Proof. Without loss of generality, let's assume the random variable U, which represent the message, is a Bernoulli random variable with parameter $\alpha \in [\![0,1]\!]$ and the random variable S, which represents the jammer's channel input and is independent of U, is a Bernoulli

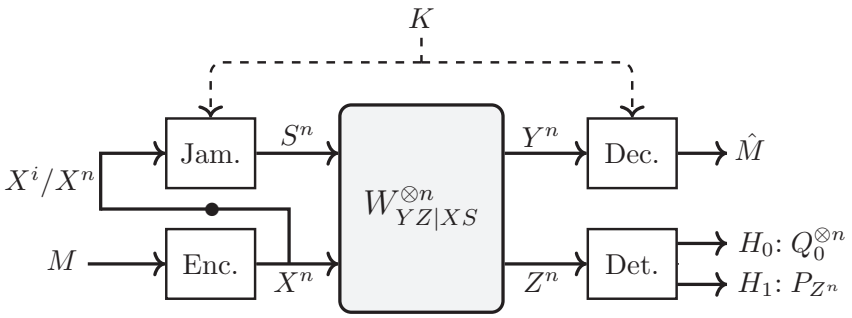

Figure 4.11. Model of covert communication with cooperative jamming

random variable with parameter $\beta \in [\![0,1]\!]$. We set $X = U \oplus S$, therefore

$$P_Z(z=0) = \mathbb{P}(x \oplus s = 0) = \mathbb{P}(u=0) = \alpha, \tag{4.87}$$

$$\Upsilon_Z(z=0) = \mathbb{P}(s=0) = \beta, \tag{4.88}$$

therefore the covertness constraint $P_Z = \Upsilon_Z$ implies that $\alpha = \beta$. We now have,

$$\mathbb{I}_P(U;Y) \overset{(a)}{=} \mathbb{H}_P(U) = \mathbb{H}_b(\alpha), \tag{4.89}$$

where, (a) follows since $X = U \oplus S$. Also,

$$\mathbb{I}_P(U;Z) = \mathbb{I}_P(U;Y) = \mathbb{H}_b(\alpha), \tag{4.90}$$

$$\mathbb{I}_P(U,S;Z) = \mathbb{H}_P(Z) - \mathbb{H}_P(Z|U,S) = \mathbb{H}_P(U) = \mathbb{H}_b(\alpha). \tag{4.91}$$

Therefore, the constraint on the key rate in (4.79b) reduces to $R_K > 0$. We also have,

$$\mathbb{I}_\Upsilon(S;Z) = \mathbb{H}(S) = \mathbb{H}_b(\alpha), \tag{4.92}$$

therefore $R_J > \mathbb{H}_b(\alpha)$. □

4.6 Transmitter's Output Available at Jammer

In this section we study a problem in which to transmit the covert message, denoted by $M \in \mathcal{M}$, the jammer and the receiver are assumed to share a rate limited and uniformly

distributed shared secret key $K \in \mathcal{K}$, this helps the receiver to cancel the interference caused by the randomness that the jammer interpolates into the channel. Here, the transmitter's channel input is assumed to be available non-causally, causally, or strictly-causally at the jammer so that the jammer can coordinate its channel input according to the transmitter's codeword. The transmitter does not use any source of local randomness, but the jammer uses a limited amount of local randomness, which is shared with the receiver as a shared secret key. This problem setup is illustrated in Fig. 4.11.

4.6.1 Strictly-Causal Case

In this subsection, we study the problem described above when the transmitter's channel input is available strictly causally at the jammer.

Theorem 30. *Let*

$$\mathcal{A} = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists \big(P_{UXS_1YZ}, \Upsilon_{S_2YZ} \big) \in \mathcal{D} : \\ R < \min\{\mathbb{H}_P(X|U), \mathbb{I}_P(X, S_1; Y)\} \\ R_K > \max\Big\{ \mathbb{I}_P(X, S_1; Z) - \min\{\mathbb{H}_P(X|U), \mathbb{I}_P(X, S_1; Y)\}, \mathbb{I}_\Upsilon(S_2; Z) \Big\} \end{array} \right\}, \quad (4.93a)$$

where

$$\mathcal{D} = \left\{ \begin{array}{l} \big(P_{UXS_1YZ}, \Upsilon_{S_2YZ} \big) : \\ P_{UXS_1YZ} = P_U P_{X|U} P_{S_1|U} W_{YZ|XS} \\ \Upsilon_{S_2YZ} = P_{S_2} W_{YZ|X=x_0,S} \\ \min\{\mathbb{H}_P(X|U), \mathbb{I}_P(X, S_1; Y)\} > \mathbb{I}_P(U, X; Z) \\ P_Z = \Upsilon_Z \\ |\mathcal{U}| \leq \min\{|\mathcal{X}|\,|\mathcal{S}| + 1, |\mathcal{Y}| + 2\} \end{array} \right\}. \quad (4.93b)$$

The covert capacity of the DMC $W_{YZ|XS}$ when the transmitter's codeword is available strictly-causally at the jammer is lower bounded by

$$C_{\text{CJ-SC}} \supseteq \text{conv}(\mathcal{A}), \tag{4.94}$$

where $\text{conv}(\mathcal{A})$ *is the convex hull of the set* $\mathcal{A}$.

To prove the achievable rate region in Theorem 30 we assume that the legitimate terminals share a secret key of negligible rate therefore the transmitter and the jammer can coordinate. The details of the proof is available in Appendix AG.

We now provide an upper bound on the covert capacity when there is a shared secret key of negligible rate between the transmitter and the jammer so that they can coordinate. Similar to the upper bounds in the previous sections, we provide an upper bound on the covert capacity for the case that the jammer knows in which blocks communication is happening and uses an unlimited source of local randomness when the transmitter is not communicating with the receiver by transmitting an i.i.d. sequence according to some distribution P_{S_2}. The covert capacity in this case is not less than the covert capacity when the jammer uses a limited amount of local randomness. Hence, this upper bound is also an upper bound on the covert capacity when the jammer uses a limited amount of randomness, which is the problem that we study in this section. In this case, the distribution induced on the warden's observation is $Q_0^{\otimes n}$ where $Q_0(\cdot) = \sum_{s_2 \in \mathcal{S}_2} P_{S_2}(s_2) W_{Z|X=x_0,S}(\cdot|x_0, s_2)$.

Theorem 31. *Let*

$$\mathcal{A} = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists P_{USXYZ} \in \mathcal{D} : \\ R \leq \min\{\mathbb{H}(X|U), \mathbb{I}(X,S;Y)\} \\ R_K \geq \mathbb{I}(X,S;Z) - \min\{\mathbb{H}(X|U), \mathbb{I}(X,S;Y)\} \end{array} \right\}, \tag{4.95a}$$

where

$$\mathcal{D} = \left\{ \begin{array}{l} P_{USXYZ} : \\ P_{USXYZ} = P_U P_{X|U} P_{S|U} W_{YZ|XS} \\ \min\{\mathbb{H}(X|U), \mathbb{I}(X,S;Y)\} \geq \mathbb{I}(U,X;Z) \\ P_Z = Q_0 \\ |\mathcal{U}| \leq \min\{|\mathcal{X}|\,|\mathcal{S}| + 1, |\mathcal{Y}| + 2\} \end{array} \right\}. \tag{4.95b}$$

The covert capacity of the DMC $W_{YZ|XS}$ when the transmitter's codeword is available strictly-causally at the jammer is upper bounded by

$$C_{\text{CJ-SC}} \subseteq \text{conv}(\mathcal{A}), \tag{4.96}$$

where $\text{conv}(\mathcal{A})$ *is the convex hull of the set* $\mathcal{A}$.

Theorem 31 is proved in Appendix AH.

Remark 26 (When the Inner and Outer Bound Meet?)**.** *Similar to Remark 13, one can simply check that the achievability scheme in Theorem 30 meets the upper bound in Theorem 31 if the jammer has an unlimited source of local randomness and transmits an i.i.d. sequence when the transmitter is not communicating with the receiver and transmits a codeword from its codebook otherwise. This is possible since our achievability scheme in Theorem 30 requires the transmitter, the receiver, and the jammer to share a secret key of negligible rate. In this case, the transmitter and the jammer can use the secret key of negligible rate to coordinate and use the strategy described above to achieve a higher covert rate.*

4.6.2 Non-Causal Case

In this subsection we study the problem described in Section 4.6 when the transmitter's channel input is available non-causally at the jammer.

Theorem 32. *Let*

$$\mathcal{A} = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists \big(P_{S_1XYZ}, \Upsilon_{S_2YZ}\big) \in \mathcal{D} : \\ R < \min\left\{\mathbb{I}_P(X, S_1; Y), \mathbb{H}_P(X)\right\} \\ R_K > \max\Big\{\mathbb{I}_P(X, S_1; Z) - \min\left\{\mathbb{I}_P(X, S_1; Y), \mathbb{H}_P(X)\right\}, \mathbb{I}_\Upsilon(S_2; Z)\Big\} \end{array} \right\}, \quad (4.97a)$$

where

$$\mathcal{D} = \left\{ \begin{array}{l} \big(P_{S_1XYZ}, \Upsilon_{S_2YZ}\big) : \\ P_{S_1XYZ} = P_X P_{S_1|X} W_{YZ|XS} \\ \Upsilon_{S_2YZ} = P_{S_2} W_{YZ|X=x_0,S} \\ \min\left\{\mathbb{I}_P(X, S_1; Y), \mathbb{H}_P(X)\right\} > \mathbb{I}_P(X; Z) \\ P_Z = \Upsilon_Z \end{array} \right\}. \quad (4.97b)$$

The covert capacity of the DMC $W_{YZ|XS}$ when the transmitter's codeword is available non-causally at the jammer is lower bounded by

$$C_{\text{CJ-NC}} \supseteq \text{conv}(\mathcal{A}), \quad (4.98)$$

where $\text{conv}(\mathcal{A})$ *is the convex hull of the set* $\mathcal{A}$.

Theorem 32 is proved in Appendix AI.

Similar to the upper bounds in the previous sections, we provide an upper bound on the covert capacity if the jammer has an unlimited source of local randomness and transmits an i.i.d. sequence, according to some distribution P_{S_2}, when the transmitter is not communicating with the receiver and transmits a sequence from its codebook otherwise. In this case, the distribution induced on the warden's observation is $Q_0^{\otimes n}$ where $Q_0(\cdot) = \sum_{s_2 \in \mathcal{S}_2} P_{S_2}(s_2) W_{Z|X=x_0,S}(\cdot|x_0, s_2)$.

Theorem 33. *Let*

$$\mathcal{A} = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists P_{XSYZ} \in \mathcal{D} : \\ R \leq \min\{\mathbb{I}(X, S; Y), \mathbb{H}(X)\} \\ R_K \geq \mathbb{I}(X, S; Z) - \min\{\mathbb{I}(X, S; Y), \mathbb{H}(X)\} \end{array} \right\}, \tag{4.99a}$$

where

$$\mathcal{D} = \left\{ \begin{array}{l} P_{XSYZ} : \\ P_{XSYZ} = P_X P_{S|X} W_{YZ|XS} \\ \min\{\mathbb{I}(X, S; Y), \mathbb{H}(X)\} \geq \mathbb{I}(X; Z) \\ P_Z = Q_0 \end{array} \right\}. \tag{4.99b}$$

The covert capacity of the DMC $W_{YZ|XS}$ when the transmitter's codeword is available non-causally at the jammer is upper bounded by

$$C_{\text{CJ-NC}} \subseteq \text{conv}(\mathcal{A}), \tag{4.100}$$

where $\text{conv}(\mathcal{A})$ *is the convex hull of the set* $\mathcal{A}$*.*

Theorem 33 is proved in Appendix AJ.

Remark 27 (When the Inner and Outer Bound Meet?)**.** *Similar to Remark 13, the achievability scheme in Theorem 32 meets the upper bound in Theorem 33 if the jammer has unlimited source of local randomness. In this case, the jammer transmits an i.i.d. sequence when the transmitter is not communicating with the receiver and transmits a sequence from its codebook when communication is happening.*

4.6.3 Causal Case

In this subsection we study the problem described in Section 4.6 when the transmitter's channel input is available causally at the jammer.

Theorem 34. *Let*

$$\mathcal{A} = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists \big(P_{S_1XYZ}, \Upsilon_{S_2YZ}\big) \in \mathcal{D} : \\ R < \min\{\mathbb{I}_P(X, S_1; Y), \mathbb{H}_P(X)\} \\ R_K > \max\Big\{\mathbb{I}_P(X, S_1; Z) - \min\{\mathbb{H}_P(X), \mathbb{I}_P(X, S_1; Y)\}, \mathbb{I}_\Upsilon(S_2; Z)\Big\} \end{array} \right\}, \quad (4.101a)$$

where

$$\mathcal{D} = \left\{ \begin{array}{l} \big(P_{S_1XYZ}, \Upsilon_{S_2YZ}\big) : \\ P_{S_1XYZ} = P_{XS_1} W_{YZ|XS_1} \\ \Upsilon_{S_2YZ} = P_{S_2} W_{YZ|X=x_0,S_2} \\ \min\{\mathbb{I}_P(X, S_1; Y), \mathbb{H}_P(X)\} > \mathbb{I}_P(X; Z) \\ P_Z = \Upsilon_Z \end{array} \right\}. \quad (4.101b)$$

The covert capacity of the DMC $W_{YZ|XS}$ when the transmitter's codeword is available causally at the jammer is lower bounded by

$$C_{\text{CJ-C}} \supseteq \text{conv}(\mathcal{A}), \quad (4.102)$$

where $\text{conv}(\mathcal{A})$ *is the convex hull of the set* $\mathcal{A}$*.*

Theorem 34 is proved in Appendix AK.

Since the covert capacity when the transmitter's codeword is available non-causally at the jammer is not less than the covert capacity when the transmitter's codeword is available causally at the jammer (i.e. $C_{\text{CJ-C}} \subseteq C_{\text{CJ-NC}}$), the upper bound provided in Theorem 33 also is an upper bound when the transmitter's codeword is available causally at the jammer.

Remark 28 (When the Inner and Outer Bound Meet?)**.** *Similar to Remark 13, one can simply check that the achievability scheme in Theorem 34 meets the upper bound in Theorem 33 if the jammer has an unlimited source of local randomness and transmits an i.i.d. sequence*

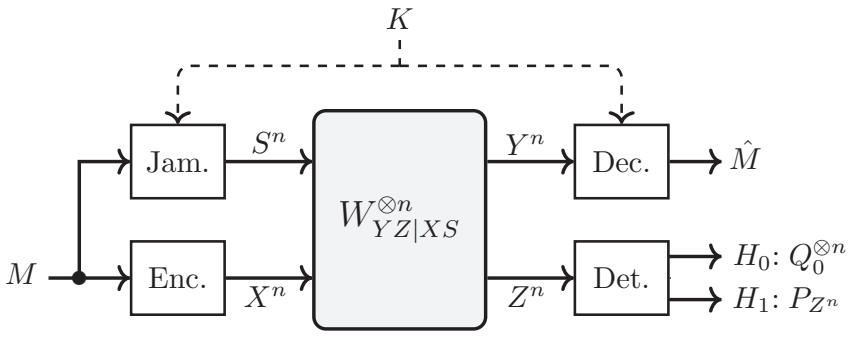

Figure 4.12. Model of covert communication with cooperative jamming

when the transmitter is not communicating with the receiver and transmits a codeword from its codebook otherwise. This is possible since our achievability scheme in Theorem 34 requires the transmitter, the receiver, and the jammer to share a secret key of negligible rate. In this case, the transmitter and the jammer can use the secret key of negligible rate to coordinate and use the strategy described above to achieve a higher covert rate.

4.6.4 Transmitter's Message available for the Jammer

In this section, we study a case in which the jammer has access to the transmitter's message (see Fig.4.12).

Theorem 35. *Let*

$$\mathcal{A} = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists \big(P_{XS_1YZ}, \Upsilon_{S_2YZ}\big) \in \mathcal{D} : \\ R < \mathbb{I}_P(X, S_1; Y) \\ R_K > \max\big\{\mathbb{I}_P(X, S_1; Z) - \mathbb{I}_P(X, S_1; Y), \mathbb{I}_\Upsilon(S_2; Z)\big\} \end{array} \right\}, \tag{4.103a}$$

where

$$\mathcal{D} = \left\{ \begin{array}{l} \left(P_{XS_1YZ}, \Upsilon_{S_2YZ}\right) : \\ P_{XS_1YZ} = P_X P_{S_1|X} W_{Y,Z|X,S_1} \\ \Upsilon_{S_2YZ} = P_{S_2} W_{YZ|X=x_0,S_2} \\ \mathbb{I}_P(X, S_1; Y) > \mathbb{I}_P(X; Z) \\ P_Z = \Upsilon_Z \end{array} \right\}. \tag{4.103b}$$

The covert capacity of the DMC $W_{YZ|XS}$ when the transmitter's message is available at the jammer is lower bounded by

$$C_{\text{CJ-DMS}} \supseteq \text{conv}(\mathcal{A}). \tag{4.104}$$

Theorem 35 is proved in Appendix AL.

Similar to the upper bounds in the previous sections, we provide an upper bound on the covert capacity for the case that the jammer knows in which blocks communication is happening and uses an unlimited source of local randomness, by transmitting an i.i.d. sequence according to some distribution P_{S_2}, when the transmitter is not communicating with the receiver. In this case, the distribution induced on the warden's observation is $Q_0^{\otimes n}$ where $Q_0(\cdot) = \sum_{s_2 \in \mathcal{S}_2} P_{S_2}(s_2) W_{Z|X=x_0,S}(\cdot|x_0, s_2)$.

Theorem 36. *Let*

$$\mathcal{A} = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists P_{SXYZ} \in \mathcal{D} : \\ R \leq \mathbb{I}(X, S; Y) \\ R_K \geq \mathbb{I}(X, S; Z) - \mathbb{I}(X, S; Y) \end{array} \right\}, \tag{4.105a}$$

where

$$\mathcal{D} = \left\{ \begin{array}{l} \left(P_{SXYZ}, \Upsilon_{SYZ}\right) : \\ P_{SXYZ} = P_X P_{S|X} W_{Y,Z|X,S} \\ \mathbb{I}(X,S;Y) \geq \mathbb{I}(X;Z) \\ P_Z = Q_0 \end{array} \right\}. \tag{4.105b}$$

The covert capacity of the DMC $W_{YZ|XS}$ when the transmitter's message is available at the jammer is upper bounded by

$$C_{\text{CJ-DMS}} \subseteq \text{conv}(\mathcal{A}). \tag{4.106}$$

Theorem 36 is proved in Appendix AM.

Remark 29 (When the Inner and Outer Bound Meet?)**.** *Similar to Remark 13, one can simply check that the achievability scheme in Theorem 35 meets the upper bound in Theorem 36 if the jammer has an unlimited source of local randomness and transmits an i.i.d. sequence when the transmitter is not communicating with the receiver and transmits a codeword from its codebook otherwise. This is possible since the jammer knows the transmitter's message.*

4.6.5 Examples

Binary Additive Channel

Here, we provide an example in which Theorem 32 and Theorem 34 leads to positive rate for covert communication. Consider a scenario in which the channel inputs and outputs are all binary, the innocent channel input symbol $x_0 = 0$, the transmitter's output is available non-causally or causally at the jammer, and the channel rules are as follows, as it can be seen in Fig. 4.13,

$$Y = Z = X \oplus S. \tag{4.107}$$

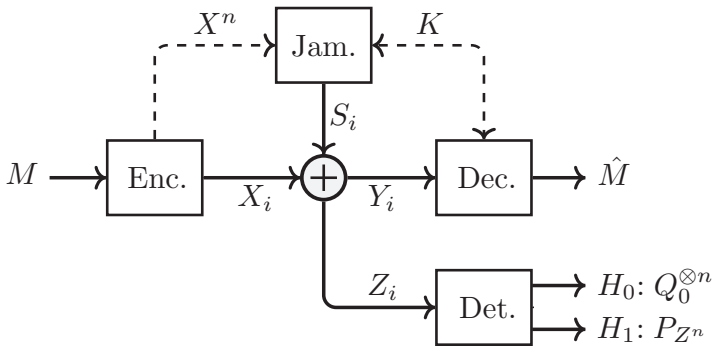

Figure 4.13. Additive channel with the transmitter's codeword available at the Jammer

Proposition 12. *The covert capacity of the DMC depicted in Fig. 4.13 with the transmitter's codeword available non-causally or causally at the jammer is lower bounded by*

$$\operatorname{conv}\left\{\begin{array}{l}(R,R_K):\alpha,\beta,\eta\in(0:0.5)\\ R<\min\left\{\mathbb{H}_b(\alpha+\eta),\mathbb{H}_b(\beta+\eta)\right\}\\ R_K>\mathbb{H}_b(\alpha+\eta)\\ \min\left\{\mathbb{H}_b(\alpha+\eta),\mathbb{H}_b(\beta+\eta)\right\}>\mathbb{H}_b(\beta+\eta)\\ \qquad -(1-\beta-\eta)\mathbb{H}_b\left(\frac{\alpha}{1-\beta-\eta}\right)-(\beta+\eta)\mathbb{H}_b\left(\frac{\beta}{\beta+\eta}\right)\end{array}\right\}. \tag{4.108}$$

Intuitively, because the jammer knows the transmitter's codeword, it can perfectly control the warden's observation. Here, the jammer by manipulating S_1 can ensure Z follows the statistics of S_2 when the transmitter is not communicating.

Proof. We prove Proposition 12 when the transmitter's codeword is available non-causally at the jammer the proof when the transmitter's codeword is available causally at the jammer is similar and is omitted for the sake of brevity.

Without loss of generality let's assume the joint probability distribution between X and S_1 is according to the Table 4.4, and S_2 be a Bernoulli random variable independent of S_1 and X with parameter $\lambda\in[\![0,1]\!]$. To analyze the covertness we have,

$$P_Z(z=0)=\mathbb{P}(x=0,s_1=0)+\mathbb{P}(x=1,s_1=1)=\alpha+\eta, \tag{4.109a}$$

Table 4.4. Joint probability distribution between X and S

S_1 \ X	0	1
0	α	β
1	$1-\alpha-\beta-\eta$	η

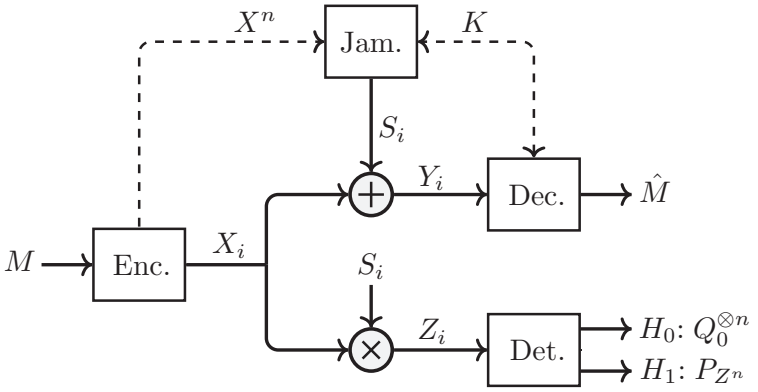

Figure 4.14. Noiseless binary channel with additive receiver's channel and multiplicative warden's channel

$$\Upsilon_Z(z=0) = \mathbb{P}(s_2 = 0) = \lambda. \tag{4.109b}$$

Therefore, $P_Z = \Upsilon_Z$ implies that $\alpha + \eta = \lambda$. Therefore,

$$\mathbb{I}_P(S_1, X; Y) = \mathbb{H}_P(Y) = \mathbb{H}_b(\alpha + \eta), \tag{4.110a}$$

$$\mathbb{H}_P(X) = \mathbb{H}_b(\beta + \eta), \tag{4.110b}$$

$$\begin{aligned}\mathbb{H}_P(Z|X) &= \mathbb{H}_P(X \oplus S_1|X) = \mathbb{H}_P(S_1|X) \\ &= (1-\beta-\eta)\mathbb{H}_b\left(\frac{\alpha}{1-\beta-\eta}\right) + (\beta+\eta)\mathbb{H}_b\left(\frac{\beta}{\beta+\eta}\right),\end{aligned} \tag{4.110c}$$

$$\mathbb{I}_\Upsilon(S_2; Z) = \mathbb{I}_\Upsilon(S_2; S_2) = \mathbb{H}_\Upsilon(S_2) = \mathbb{H}_b(\alpha + \eta). \tag{4.110d}$$

The region in Proposition 12 is achieved by substituting (4.110) in Theorem 32. □

Noiseless Binary Additive-Multiplicative Channel

Consider a channel in which X, Y, Z, and S are all binary and the innocent symbol is $x_0 = 0$ (See Fig. 4.14). The law of the channel is

$$Y = X \oplus S, \qquad Z = X \otimes S. \tag{4.111}$$

Proposition 13. *The covert capacity of the DMC depicted in Fig. 4.14 is*

$$C_{\text{CJ-NC}} = C_{\text{CJ-C}} = \max \mathbb{H}_b(X) = 1. \tag{4.112}$$

Intuitively speaking, to satisfy the condition $Q_Z = Q_0$ in this example the jammer can choose $S_1 = S_2 = 0$ therefore $Y = X$ and $Z = 0$ and the transmitter can communicate with rate $\max \mathbb{H}(X) = 1$ with the receiver.

Proof. We prove Proposition 13 when the transmitter's codeword is available non-causally at the jammer, the proof when the transmitter's codeword is available causally at the jammer is similar and is omitted for the sake of brevity.

Achievability Proof

Let the joint probability distribution between X and S_1 is according to the Table 4.4, and S_2 be a Bernoulli random variable independent of S_1 and X with parameter $\lambda \in [\![0, 1]\!]$. Therefore,

$$P_Z(z = 0) = P_{X,S_1}(x = 0, s_1 = 0) + P_{X,S_1}(x = 0, s_1 = 1) + P_{X,S_1}(x = 1, s_1 = 0) = 1 - \eta. \tag{4.113}$$

Also, since $x_0 = 0$ we have $\Upsilon_Z(z = 0) = 1$. Therefore, $P_Z = \Upsilon_Z$ implies $\eta = 0$. We now have,

$$\mathbb{I}_P(X, S_1; Y) = \mathbb{H}_P(Y) = -(\alpha + \eta) \log(\alpha + \eta) - (1 - \alpha - \eta) \log(1 - \alpha - \eta)$$

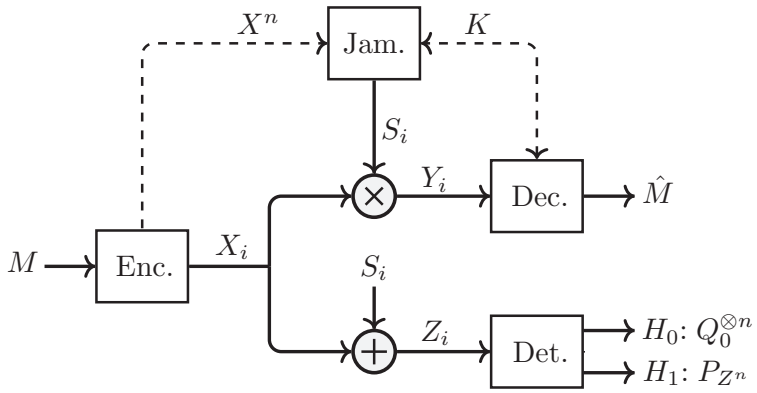

Figure 4.15. Noiseless binary channel with multiplicative receiver's channel and additive warden's channel

$$= -\alpha \log \alpha - (1-\alpha)\log(1-\alpha) = \mathbb{H}_b(\alpha), \tag{4.114a}$$

$$\mathbb{I}_P(S_1, X; Z) = \mathbb{H}_P(Z) \overset{(a)}{=} 0, \tag{4.114b}$$

$$\begin{aligned}\mathbb{H}_P(X) &= -(\beta+\eta)\log_2(\beta+\eta) - (1-\beta-\eta)\log_2(1-\beta-\eta),\\ &= -\beta\log_2\beta - (1-\beta)\log_2(1-\beta) = \mathbb{H}_b(\beta),\end{aligned} \tag{4.114c}$$

$$\mathbb{I}_P(X; Z) \overset{(b)}{=} 0, \tag{4.114d}$$

$$\mathbb{I}_\Upsilon(S_2; Z) \overset{(c)}{=} 0, \tag{4.114e}$$

where, (a) and (b) follow since $\eta = 0$ and (c) follows since $x_0 = 0$. The region in Proposition 13 is achieved by setting $(\alpha, \beta, \eta, 1-\alpha-\beta-\eta) = (0.5, 0.5, 0, 0)$ and substituting these choices of α, β, and η in (4.114). We can also set $\lambda = 1$ so that the jammer does not use any source of local randomness when communication is not happening.

Converse Proof

The converse proof is trivial, since it is not possible to communicate more than one bit in this channel. □

Table 4.5. Joint probability distribution between X and S

S_1 \ X	0	1
0	α	β
1	$1-\alpha-\beta-\eta$	η

Noiseless Binary Multiplicative-Additive Channel

Consider a channel in which X, Y, Z, and S are all binary and the innocent symbol is $x_0 = 0$ (See Fig. 4.15). The law of the channel is

$$Y = X \otimes S, \qquad Z = X \oplus S. \tag{4.115}$$

Proposition 14. *The covert capacity of the DMC depicted in Fig. 4.15 is lower bounded as*

$$C_{\text{CJ-NC}} = C_{\text{CJ-C}} = 1. \tag{4.116}$$

Intuitively speaking, to satisfy the condition $P_Z = \Upsilon_Z$ the jammer can choose $S = X$ therefore $Y = X$ and $Z = 0$ and the transmitter can communicate with rate $\max \mathbb{H}(X) = 1$ with the receiver.

Proof. We prove Proposition 14 when the transmitter's codeword is available non-causally at the jammer, the proof when the transmitter's codeword is available causally at the jammer is similar and is omitted for the sake of brevity.

Achievability Proof

Let the joint probability distribution between X and S_1 is according to the Table 4.5, and S_2 be a Bernoulli random variable independent of S_1 and X with parameter $\lambda \in [\![0, 1]\!]$. Therefore,

$$P_Z(z=0) = P_{X,S_1}(x=0, s_1=0) + P_{X,S_1}(x=1, s_1=1) = \alpha + \eta, \tag{4.117a}$$

$$\Upsilon_Z(z=0) = P_{S_2}(s=0) = \lambda. \tag{4.117b}$$

Therefore, $P_Z = \Upsilon_Z$ implies $\alpha + \eta = \lambda$. Hence,

$$\mathbb{I}_P(S_1, X; Y) = H_P(Y) = -\eta \log \eta - (1-\eta)\log(1-\eta) = \mathbb{H}_b(\eta), \tag{4.118a}$$

$$\begin{aligned}\mathbb{I}_P(S_1, X; Z) &= \mathbb{H}_P(Z) = -(\alpha+\eta)\log_2(\alpha+\eta) - (1-\alpha-\eta)\log_2(1-\alpha-\eta) \\ &= \mathbb{H}_b(\alpha+\eta),\end{aligned} \tag{4.118b}$$

$$\mathbb{H}_P(X) = \mathbb{H}_b(\beta+\eta), \tag{4.118c}$$

$$\begin{aligned}\mathbb{H}_P(Z|X) &= \mathbb{H}_P(S_1|X) = \mathbb{P}(X=0)\mathbb{H}_P(S_1|X=0) + \mathbb{P}(X=1)\mathbb{H}_P(S_1|X=1) \\ &= (1-\beta-\eta)\mathbb{H}_b\left(\frac{\alpha}{1-\beta-\eta}\right) + (\beta+\eta)\mathbb{H}_b\left(\frac{\beta}{\beta+\eta}\right),\end{aligned} \tag{4.118d}$$

$$\mathbb{I}_\Upsilon(S_2; Z) = \mathbb{H}_\Upsilon(S_2) = \mathbb{H}_b(\alpha+\eta). \tag{4.118e}$$

The region in Proposition 14 is achieved by setting $(\alpha, \beta, \eta, 1-\alpha-\beta-\eta) = (0.5, 0, 0.5, 0)$ and substituting these choices of α, β, and η in (4.118).

Converse Proof

The converse proof is trivial, since it is not possible to communicate more than one bit for this channel.

□

CHAPTER 5

CONCLUSION

We have studied three problems. First, the multi-transmitter multicast problem in the presence of an external eavesdropper. We have studied this problem under the weak and the strong secrecy regime. For the weak secrecy regime the method of Chia and El Gamal has been extended to multi-transmitter case, showing that by using this method one can find the minimum randomness necessary to achieve secrecy. For the strong secrecy regime, we used OSRB and showed that the region derived by this method is a super set of the achievable region derived for the weak secrecy regime. We have also provided some examples where these bounds are optimal.

Second, we have studied keyless covert communication over state dependent channels, when the CSI is available either at the transmitter alone, or at both the transmitter and receiver, but not to the adversary (warden). Our results show the feasibility of covertly communicating with a positive rate without an externally shared key between the transmitter and the receiver. This is in stark contrast with the known results showing that in the absence of CSI, covert communication without a shared key is impossible at positive rates.

Third, we have studied covert communication in the presence of a cooperative jammer. We show that a cooperative jammer can facilitate the communication of positive covert rates, when the transmitter have non-causal or causal access to the jammer's channel input, the jammer have non-causal, causal, strictly-causal access to the transmitter's channel input, there is a shared secret key between all the legitimate terminals, or when the jammer transmits codewords independent of transmitter's codewords.

APPENDIX A

PROOF OF LEMMA 2

Let $N(Q^n, U_0^n, V_0^n, Z^n) = |\{(k,\ell) \in [\![1, 2^{nS}]\!] \times [\![1, 2^{nT}]\!] : (Q^n, U_0^n, V_0^n, U_1^n(k), V_1^n(\ell), Z^n) \in \mathcal{T}_\epsilon^{(n)}\}|$. Next, let's define the following error events.

Let $E_1(Q^n, U_0^n, V_0^n, Z^n) = 1$ if $N(Q^n, U_0^n, V_0^n, Z^n) \geq (1+\delta_1(\epsilon))2^{n(S+T-\mathbb{I}(U_1,V_1;Z|Q,U_0,V_0)+\delta(\epsilon))}$ and $E_1 = 0$ otherwise.

Let $E = 0$ if $(Q^n, U_0^n, V_0^n, U_1^n(K), V_1^n(L), Z^n) \in \mathcal{T}_\epsilon^{(n)}$ and $E_1(Q^n, U_0^n, V_0^n, Z^n, K, L) = 0$, and $E = 1$ otherwise.

We now show that if $S \geq \mathbb{I}(U_1; Z|Q, U_0, V_0) + \delta(\epsilon)$, $T \geq \mathbb{I}(V_1; Z|Q, U_0, V_0) + \delta(\epsilon)$, and $S + T \geq \mathbb{I}(U_1, V_1; Z|Q, U_0, V_0) + \delta(\epsilon)$, then $\mathbb{P}(E = 1) \to 0$ as $n \to \infty$.

By the union bound we have

$$\mathbb{P}(E=1) \leq \mathbb{P}\big((Q^n, U_0^n, V_0^n, U_1^n(K), V_1^n(L), Z^n) \notin \mathcal{T}_\epsilon^{(n)}\big) + \mathbb{P}\big(E_1(Q^n, U_0^n, V_0^n, Z^n, K, L) = 1\big). \tag{A.1}$$

The first term tends to zero by the main assumption of the Lemma.

We then partition the event $\{E_1 = 1\}$ based on the composition of the typical sequences $(Q^n, U_0^n, V_0^n, U_1^n(k), V_1^n(\ell), Z^n) \in \mathcal{T}_\epsilon^{(n)}$:

- When all such typical sequences share the same $U_1^n(k)$, i.e., correspond to a single k.
- When all such typical sequences share the same $V_1^n(\ell)$, i.e., correspond to a single ℓ.
- Neither of the above

As usual, each of the three partitioned E_1 events gives rise to one rate constraint. We discuss the first in detail; the remaining two follow similarly. Define $A(Q^n, U_0^n, V_0^n, z^n)$ as the event $\{E_1(Q^n, U_0^n, V_0^n, Z^n) = 1\} \cap \{Z^n = z^n\}$,

$\mathbb{P}\big(E_1(Q^n, U_0^n, V_0^n, Z^n) = 1\big)$

$$= \sum_{(q^n,u_0^n,v_0^n)\in\mathcal{T}_\epsilon^{(n)}} \Big[p(q^n)p(u_0^n|q^n)p(v_0^n|q^n)\times$$
$$\mathbb{P}\Big((E_1(Q^n,U_0^n,V_0^n,Z^n)=1)|Q^n=q^n,U_0^n=u_0^n,V_0^n=v_0^n\Big)\Big]$$
$$= \sum_{\substack{(q^n,u_0^n,v_0^n)\in\mathcal{T}_\epsilon^{(n)}(Q,U_0,V_0)\\ z^n\in\mathcal{T}_\epsilon^{(n)}(Z|Q,U_0,V_0)}} p(q^n)p(u_0^n|q^n)p(v_0^n|q^n)\mathbb{P}\big(A(q^n,u_0^n,v_0^n,z^n)|Q^n=q^n,U_0^n=u_0^n,V_0^n=v_0^n\big)$$
$$\leq \sum_{(q^n,u_0^n,v_0^n)\in\mathcal{T}_\epsilon^{(n)}(Q,U_0,V_0)} p(q^n)p(u_0^n|q^n)p(v_0^n|q^n) \sum_{z^n\in\mathcal{T}_\epsilon^{(n)}(Z|Q,U_0,V_0)} \mathbb{P}\big((E_1(q^n,u_0^n,v_0^n,z^n)=1)$$
$$|Q^n=q^n,U_0^n=u_0^n,V_0^n=v_0^n\big). \tag{A.2}$$

Then,

$$\mathbb{P}\big(E_1(q^n,u_0^n,v_0^n,z^n)=1|Q^n=q^n,U_0^n=u_0^n,V_0^n=v_0^n\big)=$$
$$\mathbb{P}\big(N(q^n,u_0^n,v_0^n,z^n)\geq(1+\delta_1(\epsilon))2^{n(T-\mathbb{I}(V_1;Z|Q,U_0,V_0)+\delta(\epsilon))}\big).$$

Define $X_\ell=1$ if $(q^n,u_0^n,v_0^n,V_1^n(\ell),z^n)\in\mathcal{T}_\epsilon^{(n)}$ and 0 otherwise. Here, X_ℓ, $\ell\in[\![1,2^{nT}]\!]$, are i.i.d. Bernoulli-α random variables, where

$$2^{-n(\mathbb{I}(V_1;Z|Q,U_0,V_0)+\delta(\epsilon))}\leq\alpha\leq 2^{-n(\mathbb{I}(V_1;Z|Q,U_0,V_0)-\delta(\epsilon))}$$

Then

$$\mathbb{P}\bigg(N(q^n,u_0^n,v_0^n,z^n)\geq(1+\delta_1(\epsilon))2^{n(T-\mathbb{I}(V_1;Z|Q,U_0,V_0)+\delta(\epsilon))}\Big|Q^n=q^n,U_0^n=u_0^n,V_0^n=v_0^n\bigg)\leq$$
$$\mathbb{P}\bigg(\sum_{\ell=1}^{2^{nT}}X_\ell\geq(1+\delta_1(\epsilon))2^{nT}\alpha\Big|Q^n=q^n,U_0^n=u_0^n,V_0^n=v_0^n\bigg).$$

Applying the Chernoff Bound (e.g., see [85, Appendix B]), leads to

$$\mathbb{P}\bigg(\sum_{\ell=1}^{2^{nT}}X_\ell\geq(1+\delta_1(\epsilon))2^{nT}\alpha\Big|Q^n=q^n,U_0^n=u_0^n,V_0^n=v_0^n\bigg)$$
$$\leq\exp(-2^{nT}\alpha\delta_1^2(\epsilon)/4)$$
$$\leq\exp(-2^{n(T-\mathbb{I}(V_1;Z|Q,U_0,V_0)-\delta(\epsilon))}\delta_1^2(\epsilon)/4). \tag{A.3}$$

Therefore,

$$\begin{aligned}
&\mathbb{P}(E_1(Q^n, U_0^n, V_0^n, Z^n) = 1) \\
&\leq \sum_{(q^n, u_0^n, v_0^n) \in \mathcal{T}_\epsilon^{(n)}} p(q^n) p(u_0^n|q^n) p(v_0^n|q^n) \sum_{z^n \in \mathcal{T}_\epsilon^{(n)}(Z|Q,U_0,V_0)} \exp(-2^{n(T-\mathbb{I}(V_1;Z|Q,U_0,V_0)-\delta(\epsilon))} \delta_1^2(\epsilon)/4) \\
&\leq 2^{n \log |\mathcal{Z}|} \exp(-2^{n(T-\mathbb{I}(V_1;Z|Q,U_0,V_0)-\delta(\epsilon))} \delta_1^2(\epsilon)/4), \qquad \text{(A.4)}
\end{aligned}$$

which tends to zero as $n \to \infty$ if $T \geq \mathbb{I}(V_1; Z|Q, U_0, V_0) + \delta(\epsilon)$.

Similarly, the bounding of error probability for the second and third partition of E_1 (please see above) will give rise to the rate constraints $S \geq \mathbb{I}(U_1; Z|Q, U_0, V_0) + \delta(\epsilon)$, and $S + T \geq \mathbb{I}(U_1, V_1; Z|Q, U_0, V_0) + \delta(\epsilon)$, respectively. Details are omitted for brevity.

Finally, we bound $\mathbb{H}(L, K|Q^n, U_0^n, V_0^n, Z^n, \mathcal{C})$ as follows:

$$\begin{aligned}
&\mathbb{H}(L, K, E|Q^n, U_0^n, V_0^n, Z^n, C) \\
&\leq 1 + \mathbb{P}(E = 1)\mathbb{H}(L, K|E = 1, Q^n, U_0^n, V_0^n, Z^n, C) \\
&\qquad + \mathbb{P}(E = 0)\mathbb{H}(L, K|E = 0, Q^n, U_0^n, V_0^n, Z^n, C) \\
&\leq 1 + \mathbb{P}(E = 1) n(S + T) + \log\left((1 + \delta_1(\epsilon)) 2^{n(S+T-\mathbb{I}(U_1,V_1;Z|Q,U_0,V_0)+\delta(\epsilon))}\right) \\
&\leq n(S + T - \mathbb{I}(U_1, V_1; Z|Q, U_0, V_0) + \delta_2(\epsilon)). \qquad \text{(A.5)}
\end{aligned}$$

APPENDIX B

MAC-WTC UNDER RANDOMNESS CONSTRAINT

It is well-known that a stochastic encoding is required to avoid leaking information about the transmitted confidential messages to an eavesdropper. Here, a new achievability technique for characterizing the trade-off between the rate of the random number to realize the stochastic encoding and the communication rates in multiple access wiretap channel, by employing a variation of superposition coding, is presented.

Consider a MAC-WTC $(\mathcal{X}_1, \mathcal{X}_2, p(y, z|x_1, x_2), \mathcal{Y}, \mathcal{Z})$, in which $\mathcal{X}_1$, $\mathcal{X}_2$ are finite input alphabets and $\mathcal{Y}$ and $\mathcal{Z}$ are finite output alphabets at the legitimate receiver and the eavesdropper, respectively (as depicted in Fig. B.1). In this problem, each transmitter sends a confidential message which is supposed to be decoded by the legitimate receiver and must be kept secret from the eavesdropper. Furthermore, for stochastic encoding, Encoder 1 and Encoder 2 are allowed to use a limited amount of randomness. Thus, we are interested in the trade-off between the rate of randomness, and the rates of confidential messages.

Definition 11. *A $(M_{1,n}, M_{2,n}, n)$ code for the considered model (Fig. B.1) consists of the following:*

i) Two message sets $\mathcal{W}_i = [\![1, M_{i,n}]\!]$, $i = 1, 2$, from which independent messages W_1 and W_2 are drawn uniformly distributed over their respective sets. Also, Two dummy message sets $\mathcal{A}_i = [\![1, M'_{i,n}]\!]$, $i = 1, 2$, from which independent dummy messages A_1 and A_2 are drawn uniformly distributed over their respective sets.

ii) Deterministic encoders $f_{i,n}$, $i = 1, 2$, are defined by function $f_{i,n} : \mathcal{W}_i \times \mathcal{A}_i \to \mathcal{X}_i^n$.

iii) A decoding function $\phi : \mathcal{Y}^n \to \mathcal{W}_1 \times \mathcal{W}_2$ that assigns $(\hat{w}_1, \hat{w}_2) \in [\![1, M_{1,n}]\!] \times [\![1, M_{2,n}]\!]$ to the received sequence y^n.

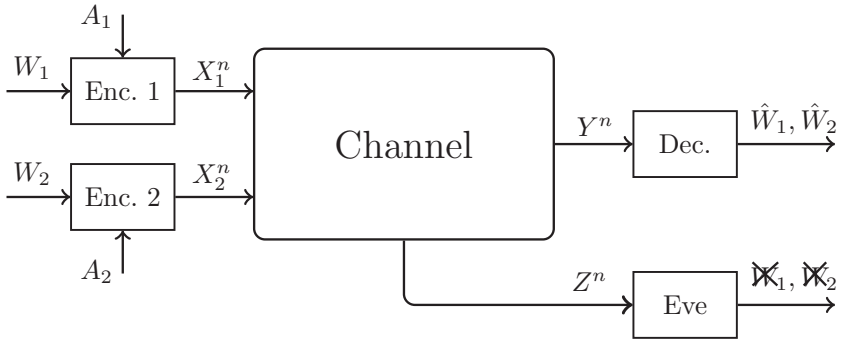

Figure B.1. Multiple access wiretap channel with deterministic encoders

The probability of error is given by:

$$P_e \triangleq \mathbb{P}\big(\{(\hat{W}_1, \hat{W}_2) \neq (w_1, w_2)\}\big). \tag{B.1}$$

Definition 12 ([68])**.** *A quadruple $(R_1, R_{d_1}, R_2, R_{d_2})$ is achievable under weak secrecy if there exists a sequence of $(M_{1,n}, M_{2,n}, M'_{1,n}, M'_{2,n}, n)$ codes with $M_{1,n} \geq 2^{nR_1}, M_{2,n} \geq 2^{nR_2}, M'_{1,n} \leq 2^{nR_{d_1}}, M'_{2,n} \leq 2^{nR_{d_2}}$, so that $P_e \underset{n\to\infty}{\longrightarrow} 0$ and*

$$\frac{1}{n}\mathbb{I}(W_1, W_2; Z^n) \underset{n\to\infty}{\longrightarrow} 0. \tag{B.2}$$

Theorem 37. *An inner bound on the secrecy capacity region of the multiple access wiretap channel is given by the set of non-negative quadruple $(R_1, R_{d_1}, R_2, R_{d_2})$ such that*

$$R_1 \leq \mathbb{I}(U; Y|Q, V) - \mathbb{I}(U; Z|Q), \tag{B.3}$$

$$R_2 \leq \mathbb{I}(V; Y|Q, U) - \mathbb{I}(V; Z|Q), \tag{B.4}$$

$$R_1 + R_2 \leq \mathbb{I}(U, V; Y|Q) - \mathbb{I}(U, V; Z|Q), \tag{B.5}$$

$$R_{d_1} \geq \mathbb{I}(U; Z|Q) + \mathbb{I}(X_1; Z|Q, U, V), \tag{B.6}$$

$$R_{d_2} \geq \mathbb{I}(V; Z|Q) + \mathbb{I}(X_2; Z|Q, U, V), \tag{B.7}$$

$$R_{d_1} + R_{d_2} \geq \mathbb{I}(U, V; Z|Q) + \mathbb{I}(X_1, X_2; Z|Q, U, V), \tag{B.8}$$

for some

$$p(q)p(u|q)p(v|q)p(x_1|u)p(x_2|v)p(y, z|x_1, x_2). \tag{B.9}$$

Remark 30. *By setting $U = X_1$, $V = X_2$, and by taking sufficiently large R_{d_1} and R_{d_2}, the result in Theorem 37 reduces to the achievable rate region of multiple access wiretap channel without common message [21, 22, 23].*

Remark 31. *By setting $X_2 = \emptyset$ and $V = \emptyset$ (or $X_1 = \emptyset$ and $U = \emptyset$), the result in Theorem 37 reduces to the capacity rate region of broadcast channel with confidential messages under randomness constraint in [16, Corollary 11].*

Proof. Rate Splitting: Divide the dummy message A_1 into independent dummy messages $A_{1,1} \in [\![1, 2^{nR_{1,1}}]\!]$ and $A_{1,2} \in [\![1, 2^{nR_{1,2}}]\!]$. Also, divide the dummy message A_2 into independent dummy messages $A_{2,1} \in [\![1, 2^{nR_{2,1}}]\!]$ and $A_{2,2} \in [\![1, 2^{nR_{2,2}}]\!]$. Therefore, $R_{d_1} = R_{1,1} + R_{1,2}$ and $R_{d_2} = R_{2,1} + R_{2,2}$.

Codebook Generation: Fix $p(q)$, $p(u|q)$, $p(v|q)$, $p(x_1|u)$, $p(x_2|v)$, and $\epsilon > 0$. Randomly and independently generate a typical sequence q^n according to $p(q^n) = \prod_{i=1}^{n} p(q_i)$. We suppose that all the terminals know q^n.

i) Generate $2^{n(R_1+R_{1,1})}$ sequences according to $\prod_{i=1}^{n} p_{U|Q}(u_i|q_i)$. Then, randomly bin these $2^{n(R_1+R_{1,1})}$ sequences into 2^{nR_1} bins. We index these sequences as $u^n(w_1, a_{1,1})$. For each $(w_1, a_{1,1})$, generate $2^{nR_{1,2}}$ codewords $x_1^n(w_1, a_{1,1}, a_{1,2})$ each according to $\prod_{i=1}^{n} p_{X_1|U}(x_{1,i}|u_i)$.

ii) Generate $2^{n(R_2+R_{2,1})}$ sequences according to $\prod_{i=1}^{n} p_{V|Q}(v_i|q_i)$. Then, randomly bin these $2^{n(R_2+R_{2,1})}$ sequences into 2^{nR_2} bins. We index these sequences as $v^n(w_2, a_{2,1})$. For each $(w_2, a_{2,1})$, generate $2^{nR_{2,2}}$ codewords $x_1^n(w_2, a_{2,1}, a_{2,2})$ each according to $\prod_{i=1}^{n} p_{X_2|V}(x_{2,i}|v_i)$.

Encoding: To send the message w_1, the Encoder 1 splits a_1 into $(a_{1,1}, a_{1,2})$, and chooses $u^n(w_1, a_{1,1})$. Then it chooses codeword $x_1^n(w_1, a_{1,1}, a_{1,2})$ and send it over the channel.

To send the message w_2, the Encoder 2 splits a_2 into $(a_{2,1}, a_{2,2})$, and chooses $v^n(w_2, a_{2,1})$. Then it chooses codeword $x_2^n(w_2, a_{2,1}, a_{2,2})$ and send it over the channel.

Decoding and Error Probability Analysis:

- Decoder decodes (w_1, w_2) by finding a unique pair (w_1, w_2) such that $(q^n, u^n(w_1, a_{1,1}), v^n(w_2, a_{2,1}), y^n) \in \mathcal{T}_\epsilon^{(n)}(p_{U,V,Y})$ for some $(a_{1,1}, a_{2,1})$. The probability of error for Receiver goes to zero as $n \to \infty$ if we choose [85]

$$R_1 + R_{1,1} \leq \mathbb{I}(U; Y|Q, V) - \epsilon, \tag{B.10}$$

$$R_2 + R_{2,1} \leq \mathbb{I}(V; Y|Q, U) - \epsilon, \tag{B.11}$$

$$R_1 + R_{1,1} + R_2 + R_{2,1} \leq \mathbb{I}(U, V; Y|Q) - \epsilon. \tag{B.12}$$

Equivocation Calculation: We analyze mutual information between (W_1, W_2) and Z^n, averaged over all random codebooks,

$$\begin{aligned}
&\mathbb{I}(W_1, W_2; Z^n|Q^n, \mathcal{C}) \\
&= \mathbb{I}(W_1, W_2, A_{1,1}, A_{1,2}, A_{2,1}, A_{2,2}; Z^n|Q^n, \mathcal{C}) - \mathbb{I}(A_{1,1}, A_{1,2}, A_{2,1}, A_{2,2}; Z^n|W_1, W_2, Q^n, \mathcal{C}) \\
&\overset{(a)}{=} \mathbb{I}(W_1, W_2, A_{1,1}, A_{1,2}, A_{2,1}, A_{2,2}, X_1^n, X_2^n; Z^n|Q^n, \mathcal{C}) - \mathbb{I}(A_{1,1}, A_{1,2}, A_{2,1}, A_{2,2}; Z^n|W_1, W_2, Q^n, \mathcal{C}) \\
&\overset{(b)}{=} \mathbb{I}(X_1^n, X_2^n; Z^n|Q^n, \mathcal{C}) - \mathbb{I}(A_{1,1}, A_{1,2}, A_{2,1}, A_{2,2}; Z^n|W_1, W_2, Q^n, \mathcal{C}) \\
&= \mathbb{I}(X_1^n, X_2^n; Z^n|Q^n, \mathcal{C}) - \mathbb{I}(A_{1,1}, A_{2,1}; Z^n|W_1, W_2, Q^n, \mathcal{C}) \\
&\quad - \mathbb{I}(A_{1,2}, A_{2,2}; Z^n|W_1, W_2, A_{1,1}, A_{1,2}, Q^n, \mathcal{C}) \\
&= \mathbb{I}(X_1^n, X_2^n; Z^n|Q^n, \mathcal{C}) - \mathbb{H}(A_{1,1}, A_{2,1}|W_1, W_2, Q^n, \mathcal{C}) + \mathbb{H}(A_{1,1}, A_{2,1}|W_1, W_2, Z^n, Q^n, \mathcal{C}) \\
&\quad - \mathbb{H}(A_{1,2}, A_{2,2}|W_1, W_2, A_{1,1}, A_{2,1}, Q^n, \mathcal{C}) + \mathbb{H}(A_{1,2}, A_{2,2}|W_1, W_2, A_{1,1}, A_{2,1}, Z^n, Q^n, \mathcal{C}),
\end{aligned} \tag{B.13}$$

where (a) is due to X_1^n and X_2^n are deterministic functions of $(W_1, A_{1,1}, A_{1,2})$ and $(W_2, A_{2,1}, A_{2,2})$, respectively. Also, (b) is due to the fact that, given X_1^n and X_2^n, the indices W_1, W_2, $A_{1,1}$, $A_{1,2}$,$A_{2,1}$, and $A_{2,2}$ are uniquely determined.

The first term in (B.13) is bounded as:

$$\mathbb{I}(X_1^n, X_2^n; Z^n|Q^n, \mathcal{C}) \leq n\mathbb{I}(X_1, X_2; Z|Q) + n\epsilon, \tag{B.14}$$

where $\epsilon \underset{n\to\infty}{\longrightarrow} 0$ similar to [85].

For the second term in (B.13) we have

$$\mathbb{H}(A_{1,1}, A_{2,1}|W_1, W_2, Q^n, \mathcal{C}) = n(R_{1,1} + R_{2,1}). \tag{B.15}$$

For the third term, substituting $U_0 \leftarrow Q$, $V_0 \leftarrow Q$, $U_1 \leftarrow U$, and $V_1 \leftarrow V$ in Lemma 2 result that if $\mathbb{P}\big((Q^n, U^n(W_1, A_{1,1}), V^n(W_2, A_{2,1}), Z^n) \in \mathcal{T}_\epsilon^{(n)}\big) \underset{n\to\infty}{\longrightarrow} 1$ and

$$R_{1,1} \geq \mathbb{I}(U; Z|Q) + \epsilon, \tag{B.16}$$

$$R_{2,1} \geq \mathbb{I}(V; Z|Q) + \epsilon, \tag{B.17}$$

$$R_{1,1} + R_{2,1} \geq \mathbb{I}(U, V; Z|Q) + \epsilon. \tag{B.18}$$

Then,

$$\mathbb{H}(A_{1,1}, A_{2,1}|W_1, W_2, Z^n, Q^n, \mathcal{C}) \leq n(R_{1,1} + R_{2,1} - \mathbb{I}(U, V; Z|Q) + \epsilon). \tag{B.19}$$

Here, this condition holds because

$$\mathbb{P}\big((Q^n, U^n(W_1, A_{1,1}), X_1^n(W_1, A_{1,1}, A_{1,2}), V^n(W_2, A_{2,1}), X_2^n(W_2, A_{2,1}, A_{2,2}), Z^n) \in \mathcal{T}_\epsilon^{(n)}\big) \underset{n\to\infty}{\longrightarrow} 1. \tag{B.20}$$

To bound the fourth term in (B.13), we have

$$\mathbb{H}(A_{1,2}, A_{2,2}|W_1, W_2, A_{1,1}, A_{2,1}, Q^n, \mathcal{C}) = n(R_{1,2} + R_{2,2}). \tag{B.21}$$

Now, we bound the last term in (B.13) by applying Lemma 2,

$$\mathbb{H}(A_{1,2}, A_{2,2}|W_1, W_2, A_{1,1}, A_{2,1}, Z^n, Q^n, \mathcal{C}) \leq n(R_{1,2} + R_{2,2} - \mathbb{I}(X_1, X_2; Z|Q, U, V) + \epsilon), \tag{B.22}$$

if (B.20) holds and

$$R_{1,2} \geq \mathbb{I}(X_1; Z|Q, U, V) + \epsilon, \tag{B.23}$$

$$R_{2,2} \geq \mathbb{I}(X_2; Z|Q, U, V) + \epsilon, \tag{B.24}$$

$$R_{1,2} + R_{2,2} \geq \mathbb{I}(X_1, X_2; Z|Q, U, V) + \epsilon. \tag{B.25}$$

Substituting (B.14), (B.15), (B.19), (B.21), and (B.22) into (B.13) yields

$$\begin{aligned}\mathbb{I}(W_1, W_2; Z^n|Q^n, \mathcal{C}) &\leq n\mathbb{I}(X_1, X_2; Z|Q) - n(R_{1,1} + R_{2,1}) \\ &+ n(R_{1,1} + R_{2,1} - \mathbb{I}(U, V; Z|Q) + \epsilon) - n(R_{1,2} + R_{2,2}) \\ &+ n(R_{1,2} + R_{2,2} - \mathbb{I}(X_1, X_2; Z|Q, U, V) + \epsilon).\end{aligned} \tag{B.26}$$

Therefore $\mathbb{I}(W_1, W_2; Z^n|Q^n, \mathcal{C}) \leq 2n\epsilon$. By applying the Fourier-Motzkin procedure [71] to (B.10)–(B.12), (B.16)–(B.18), (B.23)–(B.25), $R_{d_1} = R_{1,1} + R_{1,2}$, and $R_{d_2} = R_{2,1} + R_{2,2}$ we obtain the region in Theorem 37. □

APPENDIX C

PROOF OF THEOREM 1

The coding scheme is based on superposition coding, Wyner's random binning [1], Marton coding, and applying indirect decoding [11].

The random code generation is as follows:

Fix $p(q)$, $p(u_0|q)$, $p(u_1,u_2|u_0)$, $p(v_0|q)$, $p(v_1,v_2|v_0)$, $p(x_1|u_0,u_1,u_2)$, $p(x_2|v_0,v_1,v_2)$, $\epsilon_1 < \min\{\epsilon',\epsilon''\}$, and $\epsilon_2 < \min\{\epsilon',\epsilon''\}$.

Codebook Generation: Randomly and independently generate a typical sequence q^n according to $p(q^n) = \prod_{i=1}^{n} p(q_i)$. We suppose that all the terminals know q^n.

i) Generate $2^{n\tilde{R}_1}$ codewords $u_0^n(\ell_0)$ each according to $\prod_{i=1}^n p_{U_0|Q}(u_{0,i}|q_i)$. Then, randomly bin the $2^{n\tilde{R}_1}$ codewords into 2^{nR_1} bins, $\mathcal{B}(w_1)$, $w_1 \in [\![1, 2^{nR_1}]\!]$. For each ℓ_0, generate $2^{n\rho_1}$ codewords $u_1^n(\ell_0,t_1)$ each according to $\prod_{i=1}^n p_{U_1|U_0}(u_{1,i}|u_{0,i})$. Then, randomly bin the $2^{n\rho_1}$ codewords into $2^{n\rho_1'}$ bins, $\mathcal{B}(\ell_0,\ell_1)$, $\ell_1 \in [\![1, 2^{n\rho_1'}]\!]$. Similarly, for each ℓ_0, generate $2^{n\tilde{\rho}_1}$ codewords $u_2^n(\ell_0,t_2)$ each according to $\prod_{i=1}^n p_{U_2|U_0}(u_{2,i}|u_{0,i})$. Then, randomly bin the $2^{n\tilde{\rho}_1}$ codewords into $2^{n\tilde{\rho}_1'}$ bins, $\mathcal{B}(\ell_0,\ell_2)$, $\ell_2 \in [\![1, 2^{n\tilde{\rho}_1'}]\!]$.

ii) Similarly, generate $2^{n\tilde{R}_2}$ codewords $v_0^n(\ell_0')$ each according to $\prod_{i=1}^n p_{V_0|Q}(v_{0,i}|q_i)$. Then, randomly bin the $2^{n\tilde{R}_2}$ codewords into 2^{nR_2} bins, $\mathcal{B}(w_2)$, $w_2 \in [\![1, 2^{nR_2}]\!]$. For each ℓ_0', generate $2^{n\rho_2}$ codewords $v_1^n(\ell_0',s_1)$ each according to $\prod_{i=1}^n p_{V_1|V_0}(v_{1,i}|v_{0,i})$. Then, randomly bin the $2^{n\rho_2}$ codewords into $2^{n\rho_2'}$ bins, $\mathcal{B}(\ell_0',\ell_1')$, $\ell_1' \in [\![1, 2^{n\rho_2'}]\!]$. Similarly, for each ℓ_0', generate $2^{n\tilde{\rho}_2}$ codewords $v_2^n(\ell_0',s_2)$ each according to $\prod_{i=1}^n p_{V_2|V_0}(v_{2,i}|v_{0,i})$. Then, randomly bin the $2^{n\tilde{\rho}_2}$ codewords into $2^{n\tilde{\rho}_2'}$ bins, $\mathcal{B}(\ell_0',\ell_2')$, $\ell_2' \in [\![1, 2^{n\tilde{\rho}_2'}]\!]$.

Encoding: To send the message w_1, the encoder f_1 first uniformly chooses the index $L_0 \in \mathcal{B}(w_1)$. Then, it uniformly chooses a pair of indices (L_1,L_2) and selects a jointly typical sequence pair $(u_1^n(L_0,t_1(L_0,L_1)), u_2^n(L_0,t_2(L_0,L_1))) \in \mathcal{T}_{\epsilon_1}^{(n)}(U_1,U_2|U_0)$ in the product bin.

If the encoder f_1 finds more than one such pair, then it chooses one of them uniformly at random. We have an error if there is no such pair, in which the encoder f_1 uniformly at random chooses $t_1 \in \mathcal{B}(L_0, L_1)$, $t_2 \in \mathcal{B}(L_0, L_2)$. The error probability of the last event approaches to zero as $n \to \infty$, if [86]

$$\rho_1' + \tilde{\rho}_1' \leq \rho_1 + \tilde{\rho}_1 - \mathbb{I}(U_1; U_2|U_0) - \epsilon_1. \tag{C.1}$$

Finally, the encoder f_1 generates a sequence X_1^n at random according to $\prod_{i=1}^n p(x_{1,i}|u_{0,i}, u_{1,i}, u_{2,i})$. Encoder 2 proceeds similarly to encode w_2 and sends codeword X_2^n. The probability of not finding a jointly typical sequence pair $(v_1^n(L_0', s_1(L_0', L_1')), v_2^n(L_0', s_2(L_0', L_1'))) \in \mathcal{T}_{\epsilon_2}^{(n)}(V_1, V_2|V_0)$ in the product bin approaches to zero as $n \to \infty$, if [86]

$$\rho_2' + \tilde{\rho}_2' \leq \rho_2 + \tilde{\rho}_2 - \mathbb{I}(V_1; V_2|V_0) - \epsilon_2. \tag{C.2}$$

Decoding and Error Probability Analysis:

- Let (W_1, L_0, T_1) and (W_2, L_0', S_1) denote the transmitted indices by the first and second transmitter, respectively, and let $(\hat{W}_1, \hat{L}_0, \hat{T}_1)$ and $(\hat{W}_2, \hat{L}_0', \hat{S}_1)$ denote the corresponding decoded messages by the first receiver, respectively. Receiver 1 decodes (L_0, L_0') and therefore (w_1, w_2) indirectly by finding a unique pair $(\hat{\ell}_0, \hat{\ell}_0')$ such that $(q^n, u_0^n(\hat{\ell}_0), u_1^n(\hat{\ell}_0, t_1), v_0^n(\hat{\ell}_0'), v_1^n(\hat{\ell}_0', s_1), y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_0, U_1, V_0, V_1, Y_1)$ for some $t_1 \in [\![1, 2^{n\rho_1}]\!]$ and $s_1 \in [\![1, 2^{n\rho_2}]\!]$. The idea of indirect decoding for the situation that there is just one transmitter is shown in Fig. C.1. The error event $(\hat{W}_1, \hat{W}_1) \neq (W_1, W_1)$ occurs only if at least one of the following events occurs:

$$\mathcal{E}_1 = \Big\{ \big(Q^n, U_0^n(\ell_0), U_1^n(\ell_0, t_1), V_0^n(\ell_0'), V_1^n(\ell_0', s_1), Y_1^n\big) \notin \mathcal{T}_\epsilon^{(n)} \Big\}, \tag{C.3}$$

$$\mathcal{E}_2 = \Big\{ \big(Q^n, U_0^n(\hat{\ell}_0), U_1^n(\hat{\ell}_0, \hat{t}_1), V_0^n(\ell_0'), V_1^n(\ell_0', s_1), Y_1^n\big) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \hat{\ell}_0 \neq \ell_0, \hat{t}_1 \Big\}, \tag{C.4}$$

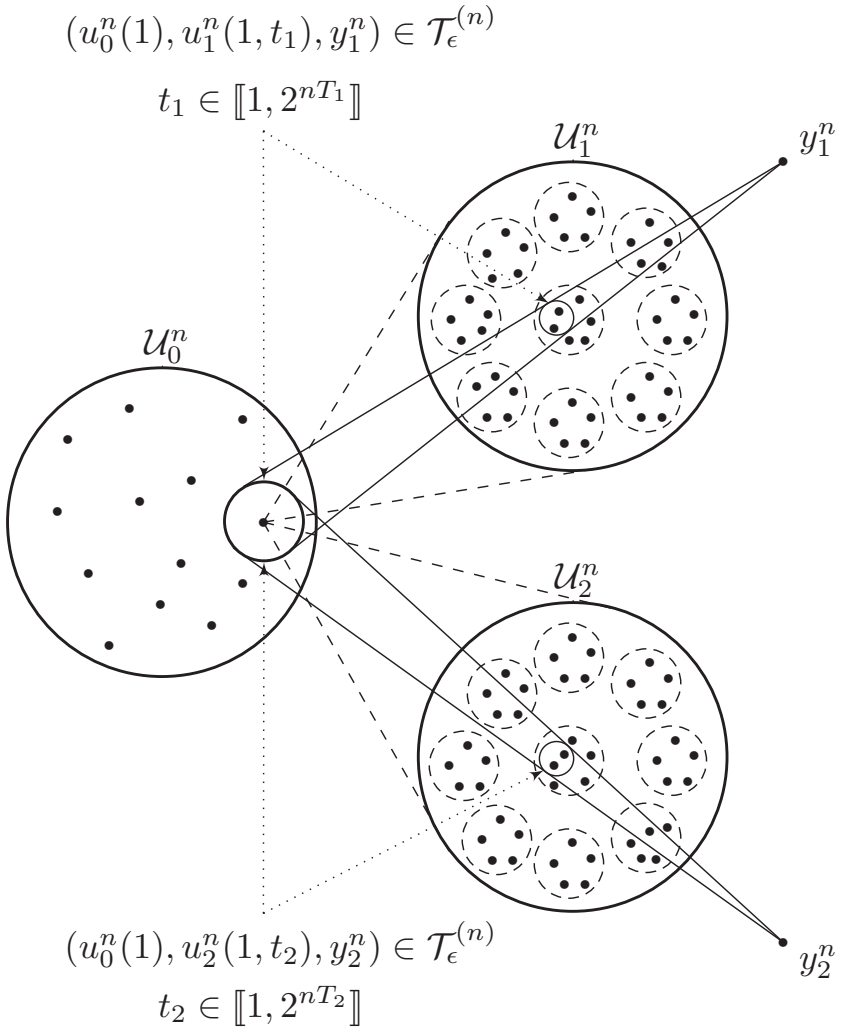

Figure C.1. Codebook structure and indirect decoding for $u_0^n(1)$ via $u_1^n(1,t_1)$ and $u_2^n(1,t_2)$ for the situation that there is just one transmitter.

$$\mathcal{E}_3 = \Big\{\big(Q^n, U_0^n(\ell_0), U_1^n(\ell_0, t_1), V_0^n(\hat{\ell}_0'), V_1^n(\hat{\ell}_0', \hat{s}_1), Y_1^n\big) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \hat{\ell}_0' \neq \ell_0', \hat{s}_1\Big\}, \tag{C.5}$$

$$\mathcal{E}_4 = \Big\{\big(Q^n, U_0^n(\hat{\ell}_0), U_1^n(\hat{\ell}_0, \hat{t}_1), V_0^n(\ell_0'), V_1^n(\ell_0', \hat{s}_1), Y_1^n\big) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \hat{\ell}_0 \neq \ell_0, \hat{t}_1, \hat{s}_1 \neq s_1\Big\}, \tag{C.6}$$

$$\mathcal{E}_5 = \Big\{\big(Q^n, U_0^n(\ell_0), U_1^n(\ell_0, \hat{t}_1), V_0^n(\hat{\ell}_0'), V_1^n(\hat{\ell}_0', \hat{s}_1), Y_1^n\big) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \hat{t}_1 \neq t_1, \hat{\ell}_0' \neq \ell_0', \hat{s}_1\Big\}, \tag{C.7}$$

$$\mathcal{E}_6 = \Big\{\big(Q^n, U_0^n(\hat{\ell}_0), U_1^n(\hat{\ell}_0, \hat{t}_1), V_0^n(\hat{\ell}_0'), V_1^n(\hat{\ell}_0', \hat{s}_1), Y_1^n\big) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \hat{\ell}_0 \neq \ell_0, \hat{t}_1, \hat{\ell}_0' \neq \ell_0', \hat{s}_1\Big\}. \tag{C.8}$$

Therefore, by Union Bound the average probability of error for decoder 1 is upper bounded as

$$P_{e_1} \leq \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2) + \mathbb{P}(\mathcal{E}_3) + \mathbb{P}(\mathcal{E}_4) + \mathbb{P}(\mathcal{E}_5) + \mathbb{P}(\mathcal{E}_6).$$

By law of large numbers, $\mathbb{P}(\mathcal{E}_1)$ tends to zero as $n \to \infty$. By packing lemma [85, Lemma 3.1] $\mathbb{P}(\mathcal{E}_2)$ to $\mathbb{P}(\mathcal{E}_6)$ respectively tend to zero as $n \to \infty$ if The probability of error for Receiver 1 goes to zero as $n \to \infty$ if we choose [85]

$$\tilde{R}_1 + \rho_1 < \mathbb{I}(U_0, U_1; Y_1|Q, V_0, V_1), \tag{C.9}$$

$$\tilde{R}_2 + \rho_2 < \mathbb{I}(V_0, V_1; Y_1|Q, U_0, U_1), \tag{C.10}$$

$$\tilde{R}_1 + \rho_1 + \rho_2 < \mathbb{I}(U_0, U_1, V_1; Y_1|Q, V_0), \tag{C.11}$$

$$\rho_1 + \tilde{R}_2 + \rho_2 < \mathbb{I}(U_1, V_0, V_1; Y_1|Q, U_0), \tag{C.12}$$

$$\tilde{R}_1 + \rho_1 + \tilde{R}_2 + \rho_2 < \mathbb{I}(U_0, U_1, V_0, V_1; Y_1|Q). \tag{C.13}$$

- Similarly, Receiver 2 decodes (L_0, L_0') and therefore (w_1, w_2) indirectly by finding a unique pair $(\check{\ell}_0, \check{\ell}_0')$ such that $(q^n, u_0^n(\check{\ell}_0), u_2^n(\check{\ell}_0, t_2), v_0^n(\check{\ell}_0'), v_2^n(\check{\ell}_0', s_2), y_2^n) \in \mathcal{T}_{\epsilon''}^{(n)}(U_0, U_2, V_0, V_2, Y_2)$ for some $t_2 \in [\![1, 2^{n\tilde{\rho}_1}]\!]$ and $s_2 \in [\![1, 2^{n\tilde{\rho}_2}]\!]$. The error analysis

for the second receiver is similar to the first receiver, and for the interest of brevity it is omitted here. Similar to Receiver 1 the probability of error for Receiver 2 goes to zero as $n \to \infty$ if we choose [85]

$$\tilde{R}_1 + \tilde{\rho}_1 < \mathbb{I}(U_0, U_2; Y_2|Q, V_0, V_2), \tag{C.14}$$

$$\tilde{R}_2 + \tilde{\rho}_2 < \mathbb{I}(V_0, V_2; Y_2|Q, U_0, U_2), \tag{C.15}$$

$$\tilde{R}_1 + \tilde{\rho}_1 + \tilde{\rho}_2 < \mathbb{I}(U_0, U_2, V_2; Y_2|Q, V_0), \tag{C.16}$$

$$\tilde{\rho}_1 + \tilde{R}_2 + \tilde{\rho}_2 < \mathbb{I}(U_2, V_0, V_2; Y_2|Q, U_0), \tag{C.17}$$

$$\tilde{R}_1 + \tilde{\rho}_1 + \tilde{R}_2 + \tilde{\rho}_2 < \mathbb{I}(U_0, U_2, V_0, V_2; Y_2|Q). \tag{C.18}$$

Equivocation Calculation: We analyze mutual information between (W_1, W_2) and Z^n, averaged over all random codebooks

$$\begin{aligned}
&\mathbb{I}(W_1, W_2; Z^n|Q^n, \mathcal{C}) \\
&= \mathbb{I}(W_1, W_2, L_0, T_1, T_2, L_0', S_1, S_2; Z^n|Q^n, \mathcal{C}) - \mathbb{I}(L_0, T_1, T_2, L_0', S_1, S_2; Z^n|W_1, W_2, Q^n, \mathcal{C}) \\
&\leq \mathbb{I}(U_0^n, U_1^n, U_2^n, V_0^n, V_1^n, V_2^n; Z^n|Q^n, \mathcal{C}) - \mathbb{I}(L_0, L_0'; Z^n|W_1, W_2, Q^n, \mathcal{C}) \\
&\quad - \mathbb{I}(T_1, T_2, S_1, S_2; Z^n|L_0, L_0', Q^n, \mathcal{C}) \\
&= \mathbb{I}(U_0^n, U_1^n, U_2^n, V_0^n, V_1^n, V_2^n; Z^n|Q^n, \mathcal{C}) - \mathbb{H}(L_0, L_0'|W_1, W_2, Q^n, \mathcal{C}) \\
&\quad + \mathbb{H}(L_0, L_0'|Z^n, W_1, W_2, Q^n, \mathcal{C}) - \mathbb{I}(T_1, T_2, S_1, S_2; Z^n|L_0, L_0', Q^n, \mathcal{C}),
\end{aligned} \tag{C.19}$$

where the inequality is due to the data processing inequality. Here, T_1, T_2, S_1, and S_2 are deterministic functions of (L_0, L_1), (L_0, L_2), (L_0', L_1'), and (L_0', L_2'), respectively.

The first term in (C.19) is bounded as:

$$\mathbb{I}(U_0^n, U_1^n, U_2^n, V_0^n, V_1^n, V_2^n; Z^n|Q^n, \mathcal{C}) \leq n\mathbb{I}(U_0, U_1, U_2, V_0, V_1, V_2; Z|Q) + n\epsilon, \tag{C.20}$$

as $n \to \infty$ where $\epsilon \to 0$ [85].

For the second term in (C.19) we have

$$\mathbb{H}(L_0, L_0'|W_1, W_2, Q^n, \mathcal{C}) = n(\tilde{R}_1 - R_1 + \tilde{R}_2 - R_2). \tag{C.21}$$

For the third term, substituting $U_0 \leftarrow Q$, $V_0 \leftarrow Q$, $U_1 \leftarrow U_0$, and $V_1 \leftarrow V_0$ in Lemma 2 result that,

$$\mathbb{H}(L_0, L_0'|Z^n, W_1, W_2, Q^n, \mathcal{C}) \leq n(\tilde{R}_1 - R_1 + \tilde{R}_2 - R_2 - \mathbb{I}(U_0, V_0; Z|Q) + \epsilon), \tag{C.22}$$

if $\mathbb{P}\big((Q^n, U_0^n(L_0), V_0^n(L_0'), Z^n) \in \mathcal{T}_\epsilon^{(n)}\big) \to 1$ as $n \to \infty$ and $\tilde{R}_1 - R_1 \geq \mathbb{I}(U_0; Z|Q) + \epsilon$, $\tilde{R}_2 - R_2 \geq \mathbb{I}(V_0; Z|Q) + \epsilon$, and $\tilde{R}_1 - R_1 + \tilde{R}_2 - R_2 \geq \mathbb{I}(U_0, V_0; Z|Q) + \epsilon$.

Here, the first condition holds because

$$\begin{aligned}\mathbb{P}\big((Q^n, U_0^n(L_0), U_1^n(L_0, t_1(L_0, L_1)), U_2^n(L_0, t_2(L_0, L_1)), V_0^n(L_0'), V_1^n(L_0', s_1(L_0', L_1'))\\ , V_2^n(L_0', s_2(L_0', L_1')), Z^n) \in \mathcal{T}_\epsilon^{(n)}\big) \to 1,\end{aligned} \tag{C.23}$$

as $n \to \infty$. Now, we bound the last term in (C.19)

$$\begin{aligned}&\mathbb{I}(T_1, T_2, S_1, S_2; Z^n|L_0, L_0', Q^n, \mathcal{C})\\ &= \mathbb{H}(T_1, T_2, S_1, S_2|L_0, L_0', Q^n, \mathcal{C}) - \mathbb{H}(T_1, T_2, S_1, S_2|Z^n, L_0, L_0', Q^n, \mathcal{C})\\ &\overset{(a)}{=} \mathbb{H}(T_1, T_2, S_1, S_2, L_1, L_2, L_1', L_2'|L_0, L_0', Q^n, \mathcal{C}) - \mathbb{H}(T_1, T_2, S_1, S_2|Z^n, L_0, L_0', Q^n, \mathcal{C})\\ &\geq \mathbb{H}(L_1, L_2, L_1', L_2'|L_0, L_0', Q^n, \mathcal{C}) - \mathbb{H}(T_1, S_1|Z^n, L_0, L_0', Q^n, \mathcal{C}) - \mathbb{H}(T_2, S_2|Z^n, L_0, L_0', Q^n, \mathcal{C})\\ &\overset{(b)}{=} \mathbb{H}(L_1, L_2|L_0, L_0', Q^n, \mathcal{C}) + \mathbb{H}(L_1', L_2'|L_0, L_0', Q^n, \mathcal{C})\\ &\quad - \mathbb{H}(T_1, S_1|Z^n, L_0, L_0', Q^n, \mathcal{C}) - \mathbb{H}(T_2, S_2|Z^n, L_0, L_0', Q^n, \mathcal{C}),\end{aligned} \tag{C.24}$$

where (a) follows since given the codebook $\mathcal{C}$ and (L_0, L_0'), (L_1, L_2, L_1', L_2') is a deterministic function of $(T_1(L_0, L_1), T_2(L_0, L_2), S_1(L_0', L_1'), S_2(L_0', L_2'))$, and (b) holds due to the fact that given $(L_0, L_0', Q^n, \mathcal{C})$, (L_1, L_2) and (L_1', L_2') are independent. Now,

$$\mathbb{H}(L_1, L_2|L_0, L_0', Q^n, \mathcal{C}) = n(\rho_1' + \tilde{\rho}_1'), \tag{C.25}$$

$$\mathbb{H}(L_1', L_2'|L_0, L_0', Q^n, \mathcal{C}) = n(\rho_2' + \tilde{\rho}_2'), \tag{C.26}$$

$$\mathbb{H}(T_1, S_1|Z^n, L_0, L_0', Q^n, \mathcal{C}) \overset{(a)}{\leq} n(\rho_1 + \rho_2 - \mathbb{I}(U_1, V_1; Z|Q, U_0, V_0) + \epsilon), \tag{C.27}$$

$$\mathbb{H}(T_2, S_2|Z^n, L_0, L_0', Q^n, \mathcal{C}) \overset{(b)}{\leq} n(\tilde{\rho}_1 + \tilde{\rho}_2 - \mathbb{I}(U_2, V_2; Z|Q, U_0, V_0) + \epsilon), \tag{C.28}$$

where (a) is due to the following. Consider,

$$\begin{aligned}\mathbb{H}(T_1, S_1|Z^n, L_0, L_0', Q^n, \mathcal{C}) &= \mathbb{H}(T_1, S_1|U_0^n(L_0), V_0^n(L_0'), Z^n, L_0, L_0', Q^n, \mathcal{C}) \\ &\leq \mathbb{H}(T_1, S_1|U_0^n(L_0), V_0^n(L_0'), Z^n, Q^n, \mathcal{C}).\end{aligned}$$

We now upper bound the term $\mathbb{H}(T_1, S_1|U_0^n(L_0), V_0^n(L_0'), Z^n, Q^n, \mathcal{C})$. From (C.23) we have $\mathbb{P}\big((Q^n, U_0^n(L_0), U_1^n(L_0, t_1(L_0, L_1)), V_0^n(L_0'),\ V_1^n(L_0', s_1(L_0', L_1')), Z^n) \in \mathcal{T}_\epsilon^{(n)}\big) \to 1$ as $n \to \infty$. Applying Lemma 2 leads to,

$$\mathbb{H}(T_1, S_1|U_0^n(L_0), V_0^n(L_0'), Z^n, Q^n, \mathcal{C}) \leq n(\rho_1 + \rho_2 - \mathbb{I}(U_1, V_1; Z|Q, U_0, V_0) + \epsilon), \tag{C.29}$$

if $\rho_1 \geq \mathbb{I}(U_1; Z|Q, U_0, V_0)+\epsilon$, $\rho_2 \geq \mathbb{I}(V_1; Z|Q, U_0, V_0)+\epsilon$, and $\rho_1+\rho_2 \geq \mathbb{I}(U_1, V_1; Z|Q, U_0, V_0)+\epsilon$. By the same argument the inequality (b) holds, if the following inequalities hold,

$$\begin{aligned}\tilde{\rho}_1 &\geq \mathbb{I}(U_2; Z|Q, U_0, V_0) + \epsilon, \\ \tilde{\rho}_2 &\geq \mathbb{I}(V_2; Z|Q, U_0, V_0) + \epsilon, \\ \tilde{\rho}_1 + \tilde{\rho}_2 &\geq \mathbb{I}(U_2, V_2; Z|Q, U_0, V_0) + \epsilon.\end{aligned}$$

Substituting (C.25)–(C.28) into (C.24) leads to,

$$\begin{aligned}\mathbb{I}(T_1, T_2, S_1, S_2; Z^n|L_0, L_0', Q^n, \mathcal{C}) \geq & n(\rho_1' + \tilde{\rho}_1') + n(\rho_2' + \tilde{\rho}_2') \\ & - n(\rho_1 + \rho_2 - \mathbb{I}(U_1, V_1; Z|Q, U_0, V_0) + \epsilon) \\ & - n(\tilde{\rho}_1 + \tilde{\rho}_2 - \mathbb{I}(U_2, V_2; Z|Q, U_0, V_0) + \epsilon).\end{aligned} \tag{C.30}$$

Substituting (C.20)–(C.22) and (C.30) into (C.19) yields

$$\mathbb{I}(W_1, W_2; Z^n|Q^n, \mathcal{C}) \leq n\mathbb{I}(U_0, U_1, U_2, V_0, V_1, V_2; Z|Q) - n(\tilde{R}_1 - R_1 + \tilde{R}_2 - R_2)$$

$$+ n(\tilde{R}_1 - R_1 + \tilde{R}_2 - R_2 - \mathbb{I}(U_0, V_0; Z|Q)) - n(\rho_1' + \tilde{\rho}_1') - n(\rho_2' + \tilde{\rho}_2')$$
$$+ n(\rho_1 + \rho_2 - \mathbb{I}(U_1, V_1; Z|Q, U_0, V_0) + \epsilon) + n(\tilde{\rho}_1 + \tilde{\rho}_2 - \mathbb{I}(U_2, V_2; Z|Q, U_0, V_0) + \epsilon). \quad \text{(C.31)}$$

Therefore $\mathbb{I}(W_1, W_2; Z^n|Q^n, \mathcal{C}) \leq n\epsilon$ if

$$\mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0) - \rho_1' - \tilde{\rho}_1' - \rho_2' - \tilde{\rho}_2' + \rho_1 + \rho_2 - \mathbb{I}(U_1, V_1; Z|Q, U_0, V_0)$$
$$+ \tilde{\rho}_1 + \tilde{\rho}_2 - \mathbb{I}(U_2, V_2; Z|Q, U_0, V_0) \leq \epsilon. \quad \text{(C.32)}$$

As a result, the rate constraints derived in equivocation analysis are

$$\tilde{R}_1 - R_1 + \tilde{R}_2 - R_2 \geq \mathbb{I}(U_0, V_0; Z|Q), \quad \text{(C.33)}$$
$$\tilde{R}_1 - R_1 \geq \mathbb{I}(U_0; Z|Q), \quad \text{(C.34)}$$
$$\tilde{R}_2 - R_2 \geq \mathbb{I}(V_0; Z|Q), \quad \text{(C.35)}$$
$$\rho_1 + \rho_2 \geq \mathbb{I}(U_1, V_1; Z|Q, U_0, V_0), \quad \text{(C.36)}$$
$$\rho_1 \geq \mathbb{I}(U_1; Z|Q, U_0, V_0), \quad \text{(C.37)}$$
$$\rho_2 \geq \mathbb{I}(V_1; Z|Q, U_0, V_0), \quad \text{(C.38)}$$
$$\tilde{\rho}_1 + \tilde{\rho}_2 \geq \mathbb{I}(U_2, V_2; Z|Q, U_0, V_0), \quad \text{(C.39)}$$
$$\tilde{\rho}_1 \geq \mathbb{I}(U_2; Z|Q, U_0, V_0), \quad \text{(C.40)}$$
$$\tilde{\rho}_2 \geq \mathbb{I}(V_2; Z|Q, U_0, V_0), \quad \text{(C.41)}$$
$$\rho_1 + \rho_2 + \tilde{\rho}_1 + \tilde{\rho}_2 - \rho_1' - \tilde{\rho}_1' - \rho_2' - \tilde{\rho}_2' \leq \mathbb{I}(U_1, V_1; Z|Q, U_0, V_0) + \mathbb{I}(U_2, V_2; Z|Q, U_0, V_0)$$
$$- \mathbb{I}(U_1, U_2, V_1, V_2; Z|U_0, V_0). \quad \text{(C.42)}$$

Finally, by applying the Fourier-Motzkin procedure [72] to (C.1), (C.2), (C.9)–(C.18), and (C.33)–(C.42) we obtain the inequalities in Theorem 1.

APPENDIX D

PROOF OF THEOREM 2

To prove Theorem 2, we first show that any achievable rate pairs (R_1, R_2) will satisfy (2.13)-(2.15) for some distribution factorized as (2.16).

Applying Fano's inequality [85] results in

$$H(W_1, W_2|Y_1^n) \leq n\varepsilon_1, \tag{D.1}$$

$$H(W_1, W_2|Y_2^n) \leq n\varepsilon_2, \tag{D.2}$$

where $\varepsilon_i \to 0,\ i = 1, 2$ as $P_e^n \to 0$.

We first derive the bound on R_1. Note that the secrecy condition (2.1) implies that

$$nR_1 - n\delta \leq H(W_1|Z^n), \tag{D.3}$$

$$nR_2 - n\delta \leq H(W_2|Z^n). \tag{D.4}$$

We first define

$$Q_i = (Z_{i+1}^n, Y_2^{i-1}), \tag{D.5}$$

$$U_{0,i} = (W_1, Q_i), \tag{D.6}$$

$$V_{0,i} = (W_2, Q_i). \tag{D.7}$$

From (D.3) we have,

$$\begin{aligned}
nR_1 &\leq H(W_1|Z^n) + n\delta \\
&= H(W_1) - I(W_1; Z^n) + n\delta \\
&\overset{(a)}{\leq} H(W_1) - H(W_1|Y_2^n) - I(W_1; Z^n) + n(\varepsilon_2 + \delta) \\
&\overset{(b)}{=} I(W_1; Y_2^n) - I(W_1; Z^n) + n\varepsilon \\
&= \sum_{i=1}^{n} [I(W_1; Y_{2,i}|Y_2^{i-1}) - I(W_1; Z_i|Z_{i+1}^n)] + n\varepsilon
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} [I(W_1, Z_{i+1}^n; Y_{2,i}|Y_2^{i-1}) - I(Z_{i+1}^n; Y_{2,i}|W_1, Y_2^{i-1}) \\
&\quad - I(W_1, Y_2^{i-1}; Z_i|Z_{i+1}^n) + I(Y_2^{i-1}; Z_i|W_1, Z_{i+1}^n)] + n\varepsilon \\
&\overset{(c)}{=} \sum_{i=1}^{n} [I(W_1, Z_{i+1}^n; Y_{2,i}|Y_2^{i-1}) - I(W_1, Y_2^{i-1}; Z_i|Z_{i+1}^n)] + n\varepsilon \\
&= \sum_{i=1}^{n} [I(Z_{i+1}^n; Y_{2,i}|Y_2^{i-1}) + I(W_1; Y_{2,i}|Z_{i+1}^n, Y_2^{i-1}) \\
&\quad - I(Y_2^{i-1}; Z_i|Z_{i+1}^n) - I(W_1; Z_i|Z_{i+1}^n, Y_2^{i-1})] + n\varepsilon \\
&\overset{(d)}{=} \sum_{i=1}^{n} [I(W_1; Y_{2,i}|Z_{i+1}^n, Y_2^{i-1}) - I(W_1; Z_i|Z_{i+1}^n, Y_2^{i-1})] + n\varepsilon \\
&\overset{(e)}{=} \sum_{i=1}^{n} [I(U_{0,i}; Y_{2,i}|Q_i) - I(U_{0,i}; Z_i|Q_i)] + n\varepsilon, \qquad \text{(D.8)}
\end{aligned}
$$

where (a) follows from Fano's inequality, (b) follows by setting $\varepsilon = \varepsilon_2 + \delta$. Equalities in (c) and (d) result from Csiszár's sum identity [70] where we have

$$\sum_{i=1}^{n} I(Z_{i+1}^n; Y_{2,i}|W_1, Y_2^{i-1}) = \sum_{i=1}^{n} I(Y_2^{i-1}; Z_i|W_1, Z_{i+1}^n), \qquad \text{(D.9)}$$

$$\sum_{i=1}^{n} I(Z_{i+1}^n; Y_{2,i}|Y_2^{i-1}) = \sum_{i=1}^{n} I(Y_2^{i-1}; Z_i|Z_{i+1}^n). \qquad \text{(D.10)}$$

The equality (e) follows from definition of random variables in (D.5)-(D.7).

Now, based on (D.8) we have:

$$
\begin{aligned}
nR_1 &\leq n \sum_{i=1}^{n} \frac{1}{n} [I(U_{0,K}; Y_{2,K}|Q_K, K = i) - I(U_{0,K}; Z_K|Q_K, K = i)] + n\varepsilon \\
&= n \sum_{i=1}^{n} p(K = i)[I(U_{0,K}; Y_{2,K}|Q_K, K = i) - I(U_{0,K}; Z_K|Q_K, K = i)] + n\varepsilon \\
&= n[I(U_{0,K}; Y_{2,K}|Q_K, K) - I(U_{0,K}; Z_K|Q_K, K)] + n\varepsilon \\
&= n[I(U_0; Y_2|Q) - I(U_0; Z|Q)] + n\varepsilon \qquad \text{(D.11)}
\end{aligned}
$$

where $U_{0,K} = U_0$, $Y_{2,K} = Y_2$, $Z_K = Z$, $(Q_K, K) = Q$ and K has a uniform distribution over $\{1, 2, ..., n\}$ outcomes.

The bounds on R_2 and $R_1 + R_2$ can be proven similar to the bound on R_1 by substitution of W_1 by W_2 and W_1 by (W_1, W_2), respectively. We omit the details for brevity.

APPENDIX E

PROOF OF THEOREM 3

The proof of achievability follows from Theorem 1 by setting $U_0 = U_1 = U_2$ and $V_0 = V_1 = V_2$ and considering the fact that the channel is degraded. Now, we show that for the degraded switch model the outer bound in Theorem 2 will reduce to the region in Theorem 3. We need to show that the outer bound distribution for the degraded switch case is equal to (2.26). Therefore, we need to show that given Q, U_0 and V_0 are independent, i.e.,

$$I(U_0; V_0|Q) = 0. \tag{E.1}$$

Moreover, we have to show that

$$I(U_0; Y_2'|V_0, Q) = I(U_0; Y_2'|Q), \tag{E.2}$$

$$I(V_0; Y_2'|U_0, Q) = I(V_0; Y_2'|Q). \tag{E.3}$$

To prove (E.2) and (E.3) we need to show that

$$I(U_0; V_0|Y_2', Q) = 0, \tag{E.4}$$

because if this equation holds we have

$$\begin{aligned} I(U_0; Y_2'|Q) &= I(U_0; V_0|Q) + I(U_0; Y_2'|V_0, Q) - I(U_0; V_0|Y_2', Q) \\ &= I(U_0; Y_2'|V_0, Q), \end{aligned} \tag{E.5}$$

$$\begin{aligned} I(V_0; Y_2'|Q) &= I(V_0; U_0|Q) + I(V_0; Y_2'|U_0, Q) - I(V_0; U_0|Y_2', Q) \\ &= I(V_0; Y_2'|U_0, Q). \end{aligned} \tag{E.6}$$

From (D.5)-(D.7) and (2.17)-(2.19) the equations in (E.1) and (E.4) are equal to the following equalities, respectively,

$$I(W_1; W_2|Z_{i+1}^n, S_{i+1}^n, Y_2^{i-1}, S^{i-1}) = 0, \tag{E.7}$$

$$I(W_1; W_2 | Y_{2,i}, S_i, Z_{i+1}^n, S_{i+1}^n, Y_2^{i-1}, S^{i-1}) = 0. \tag{E.8}$$

First, we prove (E.7),

$$\begin{aligned}
&I(W_1; W_2 | Z_{i+1}^n, S_{i+1}^n, Y_2^{i-1}, S^{i-1}) \\
&= \sum_{s_{i+1}^n} \sum_{s^{i-1}} p(S_{i+1}^n = s_{i+1}^n, S^{i-1} = s^{i-1}) I(W_1; W_2 | Z_{i+1}^n, S_{i+1}^n = s_{i+1}^n, Y_2^{i-1}, S^{i-1} = s^{i-1}) \\
&= \sum_{s_{i+1}^n} \sum_{s^{i-1}} \prod_{\substack{j=1 \\ j \neq i}}^{n} [p(S_j = s_j)] I(W_1; W_2 | Z_{i+1}^n, S_{i+1}^n = s_{i+1}^n, Y_2^{i-1}, S^{i-1} = s^{i-1}).
\end{aligned} \tag{E.9}$$

For a given s_i, (2.22) implies that $y_{1,i}$ and therefore $y_{2,i}$ and z_i only depend on the channel input $x_{s_i,i}$. By using functional dependence graphs [87], one can show that

$$I(W_1; W_2 | Z_{i+1}^n, S_{i+1}^n = s_{i+1}^n, Y_2^{i-1}, S^{i-1} = s^{i-1}) = 0, \tag{E.10}$$

for fixed switch state information s^{i-1} and s_{i+1}^n. This completes the proof of the equality (E.1). By following the same approach, we can also prove (E.8).

APPENDIX F

PROOF OF THEOREM 4

To prove Theorem 4, we first show that any achievable rate pairs (R_1, R_2) will satisfy (2.27)-(2.29) for some distribution factorized as (2.30).

Applying Fano's inequality [85] results in

$$H(W_1, W_2|Y_1^n) \leq n\varepsilon_1 \tag{F.1}$$

$$H(W_1, W_2|Y_2^n) \leq n\varepsilon_2 \tag{F.2}$$

where $\varepsilon_i \to 0,\ i = 1, 2$ as $P_e^n \to 0$.

We first derive the bound on R_1. Note that the perfect secrecy (2.1) implies that

$$nR_1 - n\delta \leq H(W_1|Z^n) \tag{F.3}$$

$$nR_2 - n\delta \leq H(W_2|Z^n). \tag{F.4}$$

Define,

$$Q_i = (Z_{i+1}^n, Y_1^{i-1}, Y_2^{i-1}), \tag{F.5}$$

$$U_{0,i} = (W_1, Q_i), \tag{F.6}$$

$$V_{1,i} = (W_2, Q_i), \tag{F.7}$$

From (F.3) we have,

$$\begin{aligned}
nR_1 &\leq H(W_1|Z^n) + n\delta \\
&= H(W_1) - I(W_1; Z^n) + n\delta \\
&\overset{(a)}{\leq} H(W_1) - H(W_1|Y_1^n, Y_2^n) - I(W_1; Z^n) + n(\varepsilon_2 + \delta) \\
&\overset{(b)}{=} I(W_1; Y_1^n, Y_2^n) - I(W_1; Z^n) + n\varepsilon \\
&= \sum_{i=1}^{n} [I(W_1; Y_{1,i}, Y_{2,i}|Y_1^{i-1}, Y_2^{i-1}) - I(W_1; Z_i|Z_{i+1}^n)] + n\varepsilon
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} [I(W_1, Z_{i+1}^n; Y_{1,i}, Y_{2,i}|Y_1^{i-1}, Y_2^{i-1}) - I(Z_{i+1}^n; Y_{1,i}, Y_{2,i}|W_1, Y_1^{i-1}, Y_2^{i-1}) \\
&\quad - I(W_1, Y_1^{i-1}, Y_2^{i-1}; Z_i|Z_{i+1}^n) + I(Y_1^{i-1}, Y_2^{i-1}; Z_i|W_1, Z_{i+1}^n)] + n\varepsilon \\
&\overset{(c)}{=} \sum_{i=1}^{n} [I(W_1, Z_{i+1}^n; Y_{1,i}, Y_{2,i}|Y_1^{i-1}, Y_2^{i-1}) - I(W_1, Y_1^{i-1}, Y_2^{i-1}; Z_i|Z_{i+1}^n)] + n\varepsilon \\
&= \sum_{i=1}^{n} [I(Z_{i+1}^n; Y_{1,i}, Y_{2,i}|Y_1^{i-1}, Y_2^{i-1}) + I(W_1; Y_{1,i}, Y_{2,i}|Z_{i+1}^n, Y_1^{i-1}, Y_2^{i-1}) \\
&\quad - I(Y_1^{i-1}, Y_2^{i-1}; Z_i|Z_{i+1}^n) - I(W_1; Z_i|Z_{i+1}^n, Y_1^{i-1}, Y_2^{i-1})] + n\varepsilon \\
&\overset{(d)}{=} \sum_{i=1}^{n} [I(W_1; Y_{1,i}, Y_{2,i}|Z_{i+1}^n, Y_1^{i-1}, Y_2^{i-1}) - I(W_1; Z_i|Z_{i+1}^n, Y_1^{i-1}, Y_2^{i-1})] + n\varepsilon \\
&\overset{(e)}{=} \sum_{i=1}^{n} [I(U_{0,i}; Y_{1,i}, Y_{2,i}|Q_i) - I(U_{0,i}; Z_i|Q_i)] + n\varepsilon
\end{aligned} \tag{F.8}
$$

where (a) follows from Fano's inequality, (b) follows by setting $\varepsilon = \varepsilon_2 + \delta$. Equalities in (c) and (d) result from Csiszár's sum identity [70] where we have

$$
\sum_{i=1}^{n} I(Z_{i+1}^n; Y_{1,i}, Y_{2,i}|W_1, Y_1^{i-1}, Y_2^{i-1}) = \sum_{i=1}^{n} I(Y_1^{i-1}, Y_2^{i-1}; Z_i|W_1, Z_{i+1}^n), \tag{F.9}
$$

$$
\sum_{i=1}^{n} I(Z_{i+1}^n; Y_{1,i}, Y_{2,i}|Y_1^{i-1}, Y_2^{i-1}) = \sum_{i=1}^{n} I(Y_1^{i-1}, Y_2^{i-1}; Z_i|Z_{i+1}^n). \tag{F.10}
$$

The equality (e) follows from definition of random variables in (F.5)-(F.7).

Now, by applying the same time-sharing strategy as (D.11) we have

$$
R_1 \leq I(U_0; Y_1, Y_2|Q) - I(U_0; Z|Q) + n\varepsilon. \tag{F.11}
$$

The bounds on R_2 and $R_1 + R_2$ can be proven similar to the bound on R_1 by substitution of W_1 by W_2 and W_1 by (W_1, W_2), respectively. We omit the details for brevity.

APPENDIX G

PROOF OF THEOREM 5

First, we rewrite the achievable rate region in Theorem 1 and the outer bound in Theorem 4 for the considered model in this subsection, and then we show that these two bounds are equal, and they are equal to the rate region in Theorem 5.

Corollary 5. *By setting $U_0 = U_1 = U_2$ and $V_0 = V_1 = V_2$ and considering the fact that $Y_1 = Y_2$, and therefore $Y_1' = Y_2'$, the achievable rate region in Theorem 1 will reduce to the set of non-negative rate pair (R_1, R_2) such that*

$$R_1 \leq I(U_0; Y_1'|Q, V_0) - I(U_0; Z|Q) \tag{G.1}$$

$$R_2 \leq I(V_0; Y_1'|Q, U_0) - I(V_0; Z|Q) \tag{G.2}$$

$$R_1 + R_2 \leq I(U_0, V_0; Y_1'|Q) - I(U_0, V_0; Z|Q) \tag{G.3}$$

for some

$$p(q)p(u_0|q)p(v_0|q)p(x_1|u_0)p(x_2|v_0). \tag{G.4}$$

Corollary 6. *By considering the fact that Y_1' is equal to Y_2' the outer bound in Theorem 4 will reduce to the set of couple rates (R_1, R_2) satisfying*

$$R_1 \leq I(U_0; Y_1'|Q) - I(U_0; Z|Q) \tag{G.5}$$

$$R_2 \leq I(V_0; Y_1'|Q) - I(V_0; Z|Q) \tag{G.6}$$

$$R_1 + R_2 \leq I(U_0, V_0; Y_1'|Q) - I(U_0, V_0; Z|Q) \tag{G.7}$$

for some joint distribution

$$p(q)p(u_0, v_0|q)p(x_1|u_0)p(x_2|v_0). \tag{G.8}$$

By using a similar approach to the proof of Theorem 4 one can show that for the outer bound we have

$$I(U_0; V_0|Q) = 0, \tag{G.9}$$

$$I(U_0; V_0|Q, Y_1') = 0. \tag{G.10}$$

Therefore, the achievable rate region in Corollary 5 and the outer bound in Corollary 6 meet. By setting $Q = \emptyset, U_0 = X_1$, and $V_0 = X_2$ and considering the fact that the channel is noiseless one can verify the region in Theorem 5.

APPENDIX H

PROOF OF THEOREM 7

Achievability Proof: Fix $P_{X|S}(x|s)$ and $\epsilon_1 > \epsilon_2 > 0$ such that, $P_Z = Q_0$.

Codebook Generation: For every $s^n \in \mathcal{S}^n$ let $C_n \triangleq \{X^n(s^n, m)\}_{(s^n,m)\in\mathcal{S}^n\times\mathcal{M}}$, where $\mathcal{M} \triangleq [\![1, 2^{nR}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_{X|S}^{\otimes n}(\cdot|s_i)$. We denote a realization of C_n by $\mathcal{C}_n \triangleq \{x^n(s^n, m)\}_{(s^n,m)\in\mathcal{S}^n\times\mathcal{M}}$.

Encoding: Given the CSI s^n, to send the message m, the transmitter computes $x^n(s^n, m)$ and transmits it over the channel. For a fixed codebook $\mathcal{C}_n$, the induced joint distribution is

$$P^{(\mathcal{C}_n)}_{S^n,M,X^n,Z^n}(s^n, m, \tilde{x}^n, z^n) = Q_S^{\otimes n}(s^n)2^{-nR}\mathbb{1}_{\{\tilde{x}^n=x^n(s^n,m)\}}W^{\otimes n}_{Z|S,X}(z^n|s^n, \tilde{x}^n). \tag{H.1}$$

Covert Analysis: We now show $\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n})] \underset{n\to\infty}{\longrightarrow} 0$, where

$$Q_Z(\cdot) = \sum_{s\in\mathcal{S}}\sum_{x\in\mathcal{X}} Q_S(s)P_{X|S}(x|s)W_{Z|X,S}(\cdot|x, s). \tag{H.2}$$

Then we choose $P_{X|S}$ such that it satisfies $Q_Z = Q_0$. Henceforth, we denote by $P^{(\mathcal{C}_n)}$ the distributions induced by a fixed codebook $\mathcal{C}_n$, and by $P_{\cdot|C_n}$ the distributions induced by a random codebook C_n. First, consider the following marginal from (H.1),

$$P_{Z^n|C_n}(z^n) = \sum_{s^n}\sum_{m} Q_S^{\otimes n}(s^n)2^{-nR}W^{\otimes n}_{Z|S,X}\big(z^n|s^n, X^n(s^n, m)\big). \tag{H.3}$$

We now have,

$$\begin{aligned}
&\mathbb{E}_{C_n}\big[\mathbb{D}\big(P_{Z^n|C_n}||Q_Z^{\otimes n}\big)\big] = \mathbb{E}_{C_n}\Big[\sum_{z^n} P_{Z^n|C_n}(z^n)\log\Big(\frac{P_{Z^n|C_n}(z^n)}{Q_Z^{\otimes n}(z^n)}\Big)\Big] \\
&= \mathbb{E}_{C_n}\Bigg[\sum_{z^n}\sum_{s^n}\sum_{m}\frac{1}{2^{nR}}Q_S^{\otimes n}(s^n)W^{\otimes n}_{Z|S,X}\big(z^n|s^n, X^n(s^n, m)\big) \\
&\qquad\times\log\Bigg(\frac{\sum\limits_{(\tilde{s}^n,\tilde{m})} Q_S^{\otimes n}(\tilde{s}^n)W^{\otimes n}_{Z|S,X}\big(z^n|\tilde{s}^n, X^n(\tilde{s}^n, \tilde{m})\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}\Bigg)\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\leq} \sum_{z^n}\sum_{s^n}\sum_{m}\frac{1}{2^{nR}}\sum_{x^n(s^n,m)} P_{S,X,Z}^{\otimes n}\big(s^n,x^n(s^n,m),z^n\big)\\
&\qquad\times\log\mathbb{E}_{\setminus(s^n,m)}\left[\frac{\sum\limits_{(\tilde{s}^n,\tilde{m})} Q_S^{\otimes n}(\tilde{s}^n)W_{Z|S,X}^{\otimes n}\big(z^n|\tilde{s}^n,X^n(\tilde{s}^n,\tilde{m})\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}\right]\\
&=\sum_{z^n}\sum_{s^n}\sum_{m}\frac{1}{2^{nR}}\sum_{x^n(s^n,m)} P_{S,X,Z}^{\otimes n}\big(s^n,x^n(s^n,m),z^n\big)\\
&\qquad\times\log\left(\frac{Q_S^{\otimes n}(s^n)W_{Z|S,X}^{\otimes n}\big(z^n|s^n,x^n(s^n,m)\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}\right.\\
&\qquad\left.+\,\mathbb{E}_{\setminus(s^n,m)}\left[\frac{\sum\limits_{(\tilde{s}^n,\tilde{m})\neq(s^n,m)} Q_S^{\otimes n}(\tilde{s}^n)W_{Z|S,X}^{\otimes n}\big(z^n|\tilde{s}^n,X^n(\tilde{s}^n,\tilde{m})\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}\right]\right)\\
&\overset{(b)}{\leq}\sum_{z^n}\sum_{s^n}\sum_{m}\frac{1}{2^{nR}}\sum_{x^n(s^n,m)} P_{S,X,Z}^{\otimes n}\big(s^n,x^n(s^n,m),z^n\big)\\
&\qquad\times\log\left(\frac{Q_S^{\otimes n}(s^n)W_{Z|S,X}^{\otimes n}\big(z^n|s^n,x^n(s^n,m)\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}\right.\\
&\qquad\left.+\,\mathbb{E}_{\setminus s^n}\left[\sum_{\tilde{m}\neq m}\frac{\mathbb{E}_{\setminus m}\sum\limits_{\tilde{s}^n}\Big(Q_S^{\otimes n}(\tilde{s}^n)W_{Z|S,X}^{\otimes n}\big(z^n|\tilde{s}^n,X^n(\tilde{s}^n,\tilde{m})\big)\Big)}{2^{nR}Q_Z^{\otimes n}(z^n)}\right]\right)\\
&=\sum_{z^n}\sum_{s^n}\sum_{m}\frac{1}{2^{nR}}\sum_{x^n(s^n,m)} P_{S,X,Z}^{\otimes n}\big(s^n,x^n(s^n,m),z^n\big)\\
&\qquad\times\log\left(\frac{Q_S^{\otimes n}(s^n)W_{Z|S,X}^{\otimes n}\big(z^n|s^n,x^n(s^n,m)\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}+\mathbb{E}_{\setminus s^n}\left[\sum_{\tilde{m}\neq m}\frac{Q_Z^{\otimes n}\big(z^n\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}\right]\right)\\
&\leq\sum_{z^n}\sum_{s^n}\sum_{m}\frac{1}{2^{nR}}\sum_{x^n(s^n,m)} P_{S,X,Z}^{\otimes n}\big(s^n,x^n(s^n,m),z^n\big)\\
&\qquad\times\log\left(\frac{Q_S^{\otimes n}(s^n)W_{Z|S,X}^{\otimes n}\big(z^n|s^n,x^n(s^n,m)\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}+1\right)\\
&\triangleq\Psi_1+\Psi_2,
\end{aligned}
\tag{H.4}
$$

where (a) follows from Jensen's inequality and (b) follows by adding some terms to the nominator of the second term in the argument of the log function. We define Ψ_1 and Ψ_2 as

$$\begin{aligned}
\Psi_1 &= \sum_m \frac{1}{2^{nR}} \sum_{(s^n,x^n(s^n,m),z^n)\in\mathcal{T}_\epsilon^{(n)}} P_{S,X,Z}^{\otimes n}\big(s^n, x^n(s^n,m), z^n\big) \\
&\quad \times \log\left(\frac{Q_S^{\otimes n}(s^n) W_{Z|S,X}^{\otimes n}\big(z^n|s^n, x^n(s^n,m)\big)}{2^{nR} Q_Z^{\otimes n}(z^n)} + 1\right) \\
&\leq \log\left(\frac{2^{-n(1-\epsilon)\left(\mathbb{H}(S)+\mathbb{H}(Z|S,X)\right)}}{2^{nR} 2^{-n(1+\epsilon)\mathbb{H}(Z)}} + 1\right) \qquad \text{(H.5)} \\
\Psi_2 &= \sum_m \frac{1}{2^{nR}} \sum_{(s^n,x^n(s^n,m),z^n)\notin\mathcal{T}_\epsilon^{(n)}} P_{S,X,Z}^{\otimes n}\big(s^n, x^n(s^n,m), z^n\big) \\
&\quad \times \log\left(\frac{Q_S^{\otimes n}(s^n) W_{Z|S,X}^{\otimes n}\big(z^n|s^n, x^n(s^n,m)\big)}{2^{nR} Q_Z^{\otimes n}(z^n)} + 1\right) \\
&\leq 2|S||X||Z| e^{-n\epsilon^2\mu_{S,X,Z}} n \log\left(\frac{3}{\mu_Z} + 1\right) \qquad \text{(H.6)}
\end{aligned}$$

where $\mu_{S,X,Z} = \min_{(s,x,z)\in(\mathcal{S},\mathcal{X},\mathcal{Z})} P_{S,X,Z}(s,x,z)$ and $\mu_Z = \min_{z\in\mathcal{Z}} P_Z(z)$. When $n\to\infty$ then $\Psi_2 \to 0$; and $\Psi_1 \to 0$ when $n\to\infty$ if,

$$R > \mathbb{I}(S,X;Z) - \mathbb{H}(S). \qquad \text{(H.7)}$$

Basic information identities yield:

$$\mathbb{I}(X,S;Z) = \mathbb{I}(X;S,Z) + \mathbb{I}(S;Z) - \mathbb{I}(X;S). \qquad \text{(H.8)}$$

Substituting (H.8) into (H.7) leads to

$$R > \mathbb{I}(X;Z|S) - \mathbb{H}(S|Z). \qquad \text{(H.9)}$$

Decoding and Error Probability Analysis: By access to the CSI s^n, the receiver declares that $\hat{m} = m$ if there exists a unique index $\hat{m}$ such that $(x^n(s^n,\hat{m}), y^n, s^n) \in \mathcal{T}_\epsilon^{(n)}(X,Y,S)$. According to the law of large numbers and the packing lemma the probability of error goes to zero as $n\to\infty$ if [85],

$$R < \mathbb{I}(X;Y|S). \qquad \text{(H.10)}$$

The region in Theorem 7 is derived by combining (H.9) and (H.10).

Converse Proof: We now develop an upper bound for the non-causal side information. Consider any sequence of length-n codes for a state-dependent channel with CSI available non-causally at both the transmitter and the receiver, such that $P_e^{(n)} \leq \epsilon_n$ and $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \delta$ with $\lim_{n\to\infty} \epsilon_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish.

Epsilon Rate Region: We first define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ that expands the region defined in (3.2) as follows.

$$\mathcal{A}_\epsilon \triangleq \left\{R \geq 0 : \exists P_{S,X,Y,Z} \in \mathcal{D}_\epsilon : R \leq \mathbb{I}(X;Y|S) + \epsilon\right\}, \tag{H.11a}$$

where

$$\mathcal{D}_\epsilon = \left\{ \begin{array}{l} P_{S,X,Y,Z} : \\ P_{S,X,Y,Z} = Q_S P_{X|S} W_{Y,Z|X,S} \\ \mathbb{D}\left(P_Z \| Q_0\right) \leq \epsilon \\ \mathbb{H}(S|Z) \geq \mathbb{I}(X;Z|S) - \mathbb{I}(X;Y|S) - 2\epsilon \end{array} \right\}. \tag{H.11b}$$

We next show that if a rate R is achievable, then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$ and $\nu > 0$, we start by upper bounding nR using standard techniques,

$$\begin{aligned} nR &= \mathbb{H}(M) \\ &\overset{(a)}{\leq} \mathbb{H}(M|S^n) - \mathbb{H}(M|Y^n, S^n) + n\epsilon_n \\ &= \mathbb{I}(M; Y^n|S^n) + n\epsilon_n \\ &= \sum_{i=1}^{n} \mathbb{I}(M; Y_i|Y^{i-1}, S^n) + n\epsilon_n \\ &= \sum_{i=1}^{n} [\mathbb{H}(Y_i|Y^{i-1}, S^n) - \mathbb{H}(Y_i|M, Y^{i-1}, S^n)] + n\epsilon_n \\ &\overset{(b)}{\leq} \sum_{i=1}^{n} [\mathbb{H}(Y_i|S_i) - \mathbb{H}(Y_i|M, Y^{i-1}, S^n, X^n)] + n\epsilon_n \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} [\mathbb{H}(Y_i|S_i) - \mathbb{H}(Y_i|S_i, X_i)] + n\epsilon_n \\
&= \sum_{i=1}^{n} \mathbb{I}(X_i; Y_i|S_i) + n\epsilon_n \\
&\overset{(c)}{\leq} n\mathbb{I}(\tilde{X}; \tilde{Y}|\tilde{S}) + n\epsilon_n \\
&\overset{(d)}{\leq} n\mathbb{I}(\tilde{X}; \tilde{Y}|\tilde{S}) + n\epsilon \\
&\overset{(e)}{=} n\mathbb{I}(X; Y|S) + n\epsilon, \qquad \text{(H.12)}
\end{aligned}
$$

where

(a) follows from Fano's inequality and since M is independent of S^n;

(b) holds because conditioning does not increase entropy;

(c) follows from the concavity of mutual information, with the resulting random variables $\tilde{X}$, $\tilde{S}$, $\tilde{Y}$, and $\tilde{Z}$ having the following distributions

$$\tilde{P}_{X,S}(x, s) \triangleq \frac{1}{n} \sum_{i=1}^{n} P_{X_i,S_i}(x, s), \qquad \text{(H.13a)}$$

$$\tilde{P}_{X,S,Y,Z}(x, s, y, z) \triangleq \tilde{P}_{X,S}(x, s) W_{Y,Z|X,S}(y, z|x, s); \qquad \text{(H.13b)}$$

(d) follows by defining $\epsilon \triangleq \max\{\epsilon_n, \nu\}$, where we choose n large enough such that $\nu \geq \frac{\delta}{n}$;

(e) follows by defining $X = \tilde{X}$, $Y = \tilde{Y}$, and $S = \tilde{S}$.

We also have,

$$
\begin{aligned}
nR &= \mathbb{H}(M) \\
&= \mathbb{H}(M|S^n) \\
&\geq \mathbb{I}(M; Z^n|S^n) \\
&\overset{(a)}{=} \mathbb{I}(M, X^n; Z^n|S^n)
\end{aligned}
$$

$$
\begin{aligned}
&\geq \mathbb{I}(X^n; Z^n|S^n) \\
&= \mathbb{I}(X^n, S^n; Z^n) - \mathbb{I}(S^n; Z^n) \\
&= \sum_{x^n}\sum_{s^n}\sum_{z^n} P(x^n, s^n, z^n) \log \frac{W_{Z|X,S}^{\otimes n}(z^n|x^n, s^n)}{P(z^n)} - \mathbb{H}(S^n) + \mathbb{H}(S^n|Z^n) \\
&\geq \sum_{x^n}\sum_{s^n}\sum_{z^n} P(x^n, s^n, z^n) \log \frac{W_{Z|X,S}^{\otimes n}(z^n|x^n, s^n)}{P(z^n)} + \mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) - \mathbb{H}(S^n) - \delta \\
&\geq \sum_{i=1}^{n}\sum_{x_i}\sum_{s_i}\sum_{z_i} P(x_i, s_i, z_i) \log \frac{W_{Z|X,S}(z_i|x_i, s_i)}{Q_0(z_i)} - \sum_{i=1}^{n}\mathbb{H}(S_i) - \delta \\
&= \sum_{i=1}^{n} \mathbb{D}(P_{X_i,S_i,Z_i}||P_{X_i,S_i}Q_0) - \sum_{i=1}^{n}\mathbb{H}(S_i) - \delta \\
&\overset{(b)}{\geq} n\mathbb{D}(\tilde{P}_{X,S,Z}||\tilde{P}_{X,S}Q_0) - n\mathbb{H}(\tilde{S}) - \delta \\
&= n\mathbb{D}(\tilde{P}_{X,S,Z}||\tilde{P}_{X,S}\tilde{P}_Z) + \mathbb{D}(\tilde{P}_Z||Q_0) - n\mathbb{H}(\tilde{S}) - \delta \\
&\overset{(c)}{\geq} n\mathbb{I}(\tilde{X}, \tilde{S}; \tilde{Z}) - n\mathbb{H}(\tilde{S}) - \delta \\
&\overset{(d)}{=} n\mathbb{I}(X, S; Z) - n\mathbb{H}(S) - \delta,
\end{aligned} \tag{H.14}
$$

where

(a) follows because X^n is a function of (M, S^n);

(b) follows from Jensen's inequality, the convexity of $\mathbb{D}(\cdot||\cdot)$, and concavity of $\mathbb{H}(\cdot)$;

(c) follows from the positivity of the KL-divergence and the definition of random variables $\tilde{X}$, $\tilde{S}$, $\tilde{Y}$, and $\tilde{Z}$ in (H.13);

(d) follows by defining $X = \tilde{X}$, $Z = \tilde{Z}$, and $S = \tilde{S}$.

For any $\nu > 0$, by choosing n large enough and substituting (H.8) into (H.14) ensures that

$$
\begin{aligned}
R &\geq \mathbb{I}(X; Z|S) - \mathbb{H}(S|Z) - \nu, \\
&\geq \mathbb{I}(X; Z|S) - \mathbb{H}(S|Z) - \epsilon,
\end{aligned} \tag{H.15}
$$

where the last inequality follows from the definition of $\epsilon \triangleq \max\{\epsilon_n, \nu\}$. To show that $\mathbb{D}(P_Z||Q_0) \leq \epsilon$, note that for n large enough,

$$\begin{aligned}\mathbb{D}(P_Z||Q_0) = \mathbb{D}(P_{\tilde{Z}}||Q_0) &= \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} P_{Z_i}\middle|\middle|Q_0\right) \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \leq \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \frac{\delta}{n} \leq \nu \leq \epsilon. \end{aligned} \tag{H.16}$$

Combining (H.12) and (H.15) shows that $\forall \epsilon_n, \nu > 0$, $R \leq \max\{a : a \in \mathcal{A}_\epsilon\}$. Therefore,

$$C_{\text{NC-TR}} = \max\left\{a : a \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon\right\}. \tag{H.17}$$

Continuity at Zero: One can prove the continuity at zero of $\mathcal{A}_\epsilon$ by substituting $\min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\}$ with $\mathbb{I}(X;Y|S)$ and $\mathbb{I}(V;Z) - \mathbb{I}(V;S)$ with $\mathbb{I}(X;Z|S) - \mathbb{H}(S|Z)$ in the continuity at zero proof in Appendix N and following the exact same arguments.

APPENDIX I

PROOF OF THEOREM 8

Achievability Proof: To prove the achievability of Theorem 8 it is convenient to introduce an associated channel $W_{Y,Z|U,S}$ as follows: Let $U \in \mathcal{U}$ be an arbitrary auxiliary random variable which is independent of the state S, and let $x : \mathcal{U} \times \mathcal{S} \mapsto \mathcal{X}$ be a deterministic mapping subject to $\mathbb{1}_{\{x=x(s,u)\}}$. According to the Shannon strategy [88], we define the $W_{Y,Z|U,S}$ as a channel specified by

$$W_{Y,Z|U,S} = \sum_{x \in \mathcal{X}} \mathbb{1}_{\{x=x(s,u)\}} W_{Y,Z|X,S}(y, z|x, s), \tag{I.1}$$

which results in a channel with input U, outputs Y, Z, and state S. Therefore, we only focus on the coding problem for the channel $W_{Y,Z|U,S}$ for the achievability proof.

We use block-Markov coding in which B independent messages are transmitted over B channel blocks, each of length r, therefore the overall codeword length is $n = rB$ symbols. The warden's observation Z^n can be described in terms of observations in individual block-Markov blocks $Z^n = (Z_1^r, \ldots, Z_B^r)$. The distribution of the warden's observation, induced by the block-Markov coding, is $P_{Z^n} \triangleq P_{Z_1^r,\ldots,Z_B^r}$ and the target output distribution is $Q_0^{\otimes n} = \prod_{j=1}^{B} Q_0^{\otimes r}$. Therefore,

$$\begin{aligned}
\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) &= \mathbb{D}(P_{Z_1^r \ldots Z_B^r}||Q_0^{\otimes rB}) \\
&= \sum_{j=1}^{B} \mathbb{D}(P_{Z_j^r|Z_{j+1}^{B,r}}||Q_0^{\otimes r}|P_{Z_{j+1}^{B,r}}) \\
&= \sum_{j=1}^{B} [\mathbb{D}(P_{Z_j^r}||Q_0^{\otimes r}) + \mathbb{D}(P_{Z_j^r|Z_{j+1}^{B,r}}||P_{Z_j^r}|P_{Z_{j+1}^{B,r}})] \\
&= \sum_{j=1}^{B} [\mathbb{D}(P_{Z_j^r}||Q_0^{\otimes r}) + \mathbb{I}(Z_j^r; Z_{j+1}^{B,r})],
\end{aligned} \tag{I.2}$$

where $Z_{j+1}^{B,r} = \{Z_{j+1}^r, \ldots Z_B^r\}$. Hence, $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \underset{n\to\infty}{\longrightarrow} 0$, is equivalent to;

$$\mathbb{D}(P_{Z_j^r}||Q_0^{\otimes r}) \underset{r\to\infty}{\longrightarrow} 0, \qquad \mathbb{I}(Z_j^r; Z_{j+1}^{B,r}) \underset{r\to\infty}{\longrightarrow} 0, \qquad \forall j \in [\![1, B]\!]. \tag{I.3}$$

This requires constructing a code that approximates $Q_0^{\otimes r}$ in each block, while eliminating the dependencies across blocks created by block-Markov coding. The random code generation is as follows.

Fix $P_U(u)$, $x = x(s,u)$, and $\epsilon_1 > \epsilon_2 > 0$ such that, $P_Z = Q_0$.

Codebook Generation for Keys: For each block $j \in [\![1, B]\!]$, create a function $\Phi : S_j^r \mapsto [\![1, 2^{rR_K}]\!]$ through random binning by choosing the value of $\Phi(s_j^r)$ independently and uniformly at random for every $s_j^r \in \mathcal{S}^r$. The key $k_j = \Phi(s_j^r)$ obtained in the block $j \in [\![1, B]\!]$ from the state sequence s_j^r is used to assist the encoder in the next block.

Codebook Generation for Messages: For each block $j \in [\![1, B]\!]$, let $C_r \triangleq \{U^r(m_j, k_{j-1})\}_{(m_j,k_{j-1})\in\mathcal{M}\times\mathcal{K}}$, where $\mathcal{M} \triangleq [\![1, 2^{rR}]\!]$ and $\mathcal{K} \triangleq [\![1, 2^{rR_k}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_U^{\otimes r}$. We denote a realization of C_r by $\mathcal{C}_r \triangleq \{u^r(m_j, k_{j-1})\}_{(m_j,k_{j-1})\in\mathcal{M}\times\mathcal{K}}$.

Encoding: For the first block, we assume that the transmitter and the receiver have access to a shared secret key k_0, in this block to transmit m_1 the encoder computes $u^r(m_1, k_0)$ and transmits codeword x^r, where $x_i = x(u_i(m_1, k_0), s_i)$. At the end of the first block, the encoder generates a key from CSI s_1^r to be used in Block 2.

For block $j \in [\![2, B]\!]$, to send the message m_j according to the generated key k_{j-1} from the previous block, the encoder computes $u^r(m_j, k_{j-1})$ and transmits codeword x^r, where $x_i = x(u_i(m_j, k_{j-1}), s_i)$. Also, at the end of each block $j \in [\![2, B]\!]$, the encoder generates a key from CSI s_j^r to be used in the next block.

Define

$$\Upsilon^{(\mathcal{C}_r)}_{M_j,K_{j-1},U^r,S_j^r,Z_j^r,K_j}(m_j, k_{j-1}, \tilde{u}^r, s_j^r, z_j^r, k_j) \triangleq 2^{-r(R_k+R)} \mathbb{1}_{\{\tilde{u}^r = u^r(m_j,k_{j-1})\}} Q_S^{\otimes r}(s_j^r) \times W_{Z|U,S}^{\otimes r}(z_j^r|\tilde{u}^r, s_j^r) \mathbb{1}_{\{k_j = \Phi(\tilde{s}_j^r)\}}. \tag{I.4}$$

For a fixed codebook $\mathcal{C}_r$, the induced joint distribution by our code design (i.e. $P^{(\mathcal{C}_r)}$) satisfies

$$\mathbb{D}\Big(P^{(\mathcal{C}_r)}_{M_j,K_{j-1},U^r,S_j^r,Z_j^r,K_j} || \Upsilon^{(\mathcal{C}_r)}_{M_j,K_{j-1},U^r,S_j^r,Z_j^r,K_j}\Big) \leq \epsilon. \tag{I.5}$$

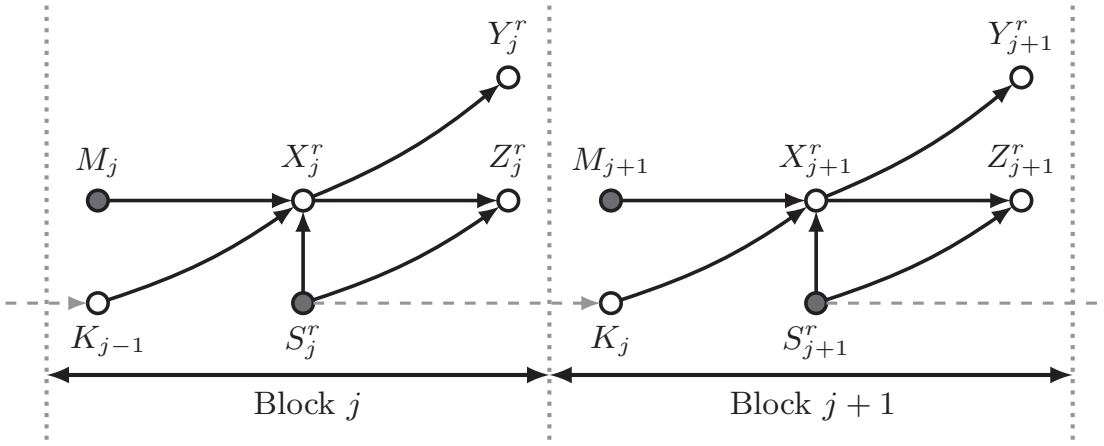

Figure I.1. Functional dependence graph for the block-Markov encoding scheme

This intermediate distribution $\Upsilon^{(\mathcal{C}_r)}$ approximates the true distribution $P^{(\mathcal{C}_r)}$ and will be used in the sequel for bounding purposes. Expression (I.5) holds because the main difference between $P^{(\mathcal{C}_r)}$ and $\Upsilon^{(\mathcal{C}_r)}$ is that the key K_{j-1} is assumed to be uniformly distributed in $\Upsilon^{(\mathcal{C}_r)}$, which is made (arbitrarily) nearly uniform in $P^{(\mathcal{C}_r)}$ with appropriate control of rate as in (I.17).

Covert Analysis: We now show $\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n})] \underset{n\to\infty}{\longrightarrow} 0$, where C_n is the set of all codebooks from all blocks, and

$$Q_Z(\cdot) = \sum_{s\in\mathcal{S}}\sum_{u\in\mathcal{U}}\sum_{x\in\mathcal{X}} Q_S(s)P_U(u)\mathbb{1}_{\left\{X=X(u,s)\right\}}W_{Z|X,S}(\cdot|x,s). \tag{I.6}$$

Then we choose P_U and $X(U,S)$ such that it satisfies $Q_Z = Q_0$. From the expansion in (I.2), by substituting Q_0 with Q_Z, for every block $j \in [\![2, B]\!]$,

$$\begin{aligned}\mathbb{I}(Z_j^r; Z_{j+1}^{B,r}) &\leq \mathbb{I}(Z_j^r; K_j, Z_{j+1}^{B,r}) \\ &\overset{(a)}{=} \mathbb{I}(Z_j^r; K_j),\end{aligned} \tag{I.7}$$

where (a) holds because $Z_j^r - K_j - Z_{j+1}^{B,r}$ forms a Markov chain, as seen in the functional dependence graph depicted in Fig. I.1. Also,

$$\mathbb{I}(Z_j^r; K_j) = \mathbb{D}(P_{Z_j^r,K_j}^{(\mathcal{C}_n)}||P_{Z_j^r}^{(\mathcal{C}_n)}P_{K_j}^{(\mathcal{C}_n)})$$

$$\overset{(b)}{\leq} \mathbb{D}(P^{(\mathcal{C}_n)}_{Z_j^r,K_j}||Q_Z^{\otimes r}Q_{K_j}), \tag{I.8}$$

where Q_{K_j} is the uniform distribution over $[\![1, 2^{rR_K}]\!]$ and (b) follows from the positivity of relative entropy and

$$\mathbb{D}(P_{Z_j^r,K_j}||P_{Z_j^r}P_{K_j}) = \mathbb{D}(P_{Z_j^r,K_j}||Q_Z^{\otimes r}Q_{K_j}) - \mathbb{D}(P_{K_j}||Q_{K_j}) - \mathbb{D}(P_{Z_j^r}||Q_Z^{\otimes r}). \tag{I.9}$$

Therefore by combining (I.2), (I.8), and (I.9),

$$\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n}) \leq 2\sum_{j=1}^{B} \mathbb{D}(P_{Z_j^r,K_j|C_r}||Q_Z^{\otimes r}Q_{K_j}). \tag{I.10}$$

We now proceed to bound the right-hand side of (I.10). First, consider the following marginal from (I.4),

$$\Upsilon_{Z_j^r,K_j|C_r}(z_j^r, k_j) = \sum_{m_j}\sum_{k_{j-1}}\sum_{s_j^r} \frac{1}{2^{r(R+R_k)}} Q_S^{\otimes r}(s_j^r) W_{Z|U,S}^{\otimes r}\big(z_j^r|U^r(m_j, k_{j-1}), s_j^r\big)\mathbb{1}_{\{k_j=\Phi(s_j^r)\}}. \tag{I.11}$$

From (I.5) and the monotonicity of KL-divergence we have,

$$\mathbb{D}\big(\Upsilon_{Z_j^r,K_j|C_r}||P_{Z_j^r,K_j|C_r}\big) \leq \epsilon. \tag{I.12}$$

To bound the Right Hand Side (RHS) of (I.10) by using Lemma 1 and the triangle inequality we have,

$$\mathbb{E}_{C_r}||P_{Z_j^r,K_j|C_r} - Q_Z^{\otimes r}Q_{K_j}||_1 \leq \mathbb{E}_{C_r}||P_{Z_j^r,K_j|C_r} - \Upsilon_{Z_j^r,K_j|C_r}||_1 + \mathbb{E}_{C_r}||\Upsilon_{Z_j^r,K_j|C_r} - Q_Z^{\otimes r}Q_{K_j}||_1. \tag{I.13}$$

From Lemma 1 and (I.12) the first term on the RHS of (I.13) vanishes as r grows; to bound the second term on the RHS of (I.13) by using Lemma 1 we have,

$$\mathbb{E}_{C_r}[\mathbb{D}(\Upsilon_{Z_j^r,K_j|C_r}||Q_Z^{\otimes r}Q_{K_j})] = \mathbb{E}_{C_r}\Big[\sum_{z_j^r,k_j} \Upsilon_{Z_j^r,K_j|C_r}(z_j^r, k_j)\log\Big(\frac{\Upsilon_{Z_j^r,K_j|C_r}(z_j^r, k_j)}{Q_Z^{\otimes r}(z_j^r)Q_{K_j}(k_j)}\Big)\Big]$$

$$
\begin{aligned}
&= \mathbb{E}_{C_r}\Bigg[\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k)}}\sum_{s_j^r}Q_S^{\otimes r}(s_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|U^r(m_j,k_{j-1}),s_j^r\big)\mathbb{1}_{\{k_j=\Phi(s_j^r)\}}\\
&\qquad\times\log\left(\frac{\sum_{\tilde{m}_j}\sum_{\tilde{k}_{j-1}}\sum_{\tilde{s}_j^r}Q_S^{\otimes r}(\tilde{s}_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{k}_{j-1}),\tilde{s}_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\right)\Bigg]\\
&\overset{(a)}{\leq}\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k)}}\sum_{s_j^r}\sum_{u^r(m_j,k_{j-1})}\Upsilon_{U^r,S^r,Z^r}^{\otimes r}\big(u^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\mathbb{E}_{\Phi(s_j^r)}\big[\mathbb{1}_{\{k_j=\Phi(s_j^r)\}}\big]\\
&\qquad\times\log\mathbb{E}_{\backslash((m_j,k_{j-1}),\Phi(s_j^r))}\left[\frac{\sum_{\tilde{m}_j}\sum_{\tilde{k}_{j-1}}\sum_{\tilde{s}_j^r}Q_S^{\otimes r}(\tilde{s}_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{k}_{j-1}),\tilde{s}_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\right]\\
&\overset{(b)}{\leq}\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k)}}\sum_{s_j^r}\sum_{u^r(m_j,k_{j-1})}\Upsilon_{U^r,S^r,Z^r}^{\otimes r}\big(u^r(m_j,k_{j-1},s_j^r,z_j^r)\big)\frac{1}{2^{rR_K}}\\
&\qquad\times\log\frac{1}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\Bigg(Q_S^{\otimes r}(s_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|u^r(m_j,k_{j-1}),s_j^r\big)\\
&\qquad+\mathbb{E}_{\backslash(m_j,k_{j-1})}\left[\sum_{(\tilde{m}_j,\tilde{k}_{j-1})\neq(m_j,k_{j-1})}Q_S^{\otimes r}(s_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{k}_{j-1}),s_j^r\big)\right]\\
&\qquad+\mathbb{E}_{\backslash\Phi(s_j^r)}\left[\sum_{\tilde{s}_j^r}Q_S^{\otimes r}(\tilde{s}_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|u^r(m_j,k_{j-1}),\tilde{s}_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}\right]\\
&\qquad+\mathbb{E}_{\backslash((m_j,k_{j-1}),\Phi(s_j^r))}\left[\sum_{\tilde{s}_j^r}\sum_{(\tilde{m}_j,\tilde{k}_{j-1})\neq(m_j,k_{j-1})}Q_S^{\otimes r}(\tilde{s}_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{k}_{j-1}),\tilde{s}_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}\right]\Bigg)\\
&\leq\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k+R_K)}}\sum_{s_j^r}\sum_{u^r(m_j,k_{j-1})}\Upsilon_{U^r,S^r,Z^r}^{\otimes r}\big(u^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\\
&\qquad\times\log\left(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|u^r(m_j,k_{j-1}),s_j^r\big)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}+\sum_{(\tilde{k}_{j-1},\tilde{m}_j)\neq(m_j,k_{j-1})}\frac{Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\right.\\
&\qquad\left.+\frac{W_{Z|U}^{\otimes r}\big(z_j^r|u^r(m_j,k_{j-1})\big)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\right)\\
&\leq\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k+R_K)}}\sum_{s_j^r}\sum_{u^r(m_j,k_{j-1})}\Upsilon_{U^r,S^r,Z^r}^{\otimes r}\big(u^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\\
&\times\log\left(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|u^r(m_j,k_{j-1}),s_j^r\big)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}+\frac{2^{rR_K}Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|U}^{\otimes r}\big(z_j^r|u^r(m_j,k_{j-1})\big)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\right)
\end{aligned}
$$

$$\triangleq \Psi_1 + \Psi_2, \tag{I.14}$$

where (a) follows from Jensen's inequality and (b) holds because $\mathbb{1}_{\{\cdot\}} \leq 1$. We define Ψ_1 and Ψ_2 as

$$\begin{aligned}
\Psi_1 = &\sum_{k_j}\sum_{m_j}\sum_{k_{j-1}} \frac{1}{2^{r(R+R_k+R_K)}} \sum_{\left(u^r(m_j,k_{j-1}),s_j^r,z_j^r\right)\in\mathcal{T}_\epsilon^{(n)}} \Upsilon_{U^r,S^r,Z^r}^{\otimes r}\left(u^r(m_j,k_{j-1}),s_j^r,z_j^r\right) \\
&\times \log\left(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|U,S}^{\otimes r}\left(z_j^r|u^r(m_j,k_{j-1}),s_j^r\right)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\right. \\
&\left.+\frac{2^{rR_K}Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|U}^{\otimes r}\left(z_j^r|u^r(m_j,k_{j-1})\right)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\right) \\
\leq &\log\left(\frac{2^{rR_K}2^{-r(1-\epsilon)(\mathbb{H}(S)+\mathbb{H}(Z|U,S))}}{2^{r(R+R_k)}2^{-r(1+\epsilon)\mathbb{H}(Z)}}+\frac{2^{rR_K}2^{-r(1-\epsilon)\mathbb{H}(S,Z)}}{2^{-r(1+\epsilon)\mathbb{H}(Z)}}+\frac{2^{-r(1-\epsilon)\mathbb{H}(Z|U)}}{2^{r(R+R_k)}2^{-r(1+\epsilon)\mathbb{H}(Z)}}+1\right)
\end{aligned} \tag{I.15}$$

$$\begin{aligned}
\Psi_2 = &\sum_{k_j}\sum_{m_j}\sum_{k_{j-1}} \frac{1}{2^{r(R+R_k+R_K)}} \sum_{\left(u^r(m_j,k_{j-1}),s_j^r,z_j^r\right)\notin\mathcal{T}_\epsilon^{(n)}} \Upsilon_{U^r,S^r,Z^r}^{\otimes r}\left(u^r(m_j,k_{j-1}),s_j^r,z_j^r\right) \\
&\times \log\left(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|U,S}^{\otimes r}\left(z_j^r|u^r(m_j,k_{j-1}),s_j^r\right)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\right. \\
&\left.+\frac{2^{rR_K}Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|U}^{\otimes r}\left(z_j^r|u^r(m_j,k_{j-1})\right)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\right) \\
\leq &\, 2|U||S||Z|e^{-r\epsilon^2\mu_{U,S,Z}}r\log\left(\frac{3}{\mu_Z}+1\right),
\end{aligned} \tag{I.16}$$

where $\mu_{U,S,Z} = \min_{(u,s,z)\in(\mathcal{U},\mathcal{S},\mathcal{Z})} P_{U,S,Z}(u,s,z)$ and $\mu_Z = \min_{z\in\mathcal{Z}} P_Z(z)$. When $r\to\infty$ then $\Psi_2 \to 0$, by choosing $R_K = \mathbb{H}(S|Z) - \epsilon$, Ψ_1 vanishes when r grows if,

$$R + R_k > \mathbb{I}(U;Z|S), \tag{I.17a}$$

$$R + R_k > \mathbb{I}(U;Z). \tag{I.17b}$$

Since U and S are independent, (I.17b) is redundant because of (I.17a).

Decoding and Error Probability Analysis: At the end of the block $j \in [\![1, B]\!]$, using its knowledge of the CSI s_j^r of the current block and the key k_{j-1} generated from the previous

block, the receiver finds a unique $\hat{m}_j$ such that $\big(u^r(\hat{m}_j, k_{j-1}), s_j^r, y_j^r\big) \in \mathcal{T}_\epsilon^{(r)}$. To analyze the probability of error, we define the following error events for $j \in [\![1, B]\!]$

$$\mathcal{E} = \big\{\hat{M} \neq M\big\}, \tag{I.18a}$$

$$\mathcal{E}_j = \big\{\hat{M}_j \neq M_j\big\}, \tag{I.18b}$$

$$\mathcal{E}_{1,j} = \big\{\big(U^r(M_j, K_{j-1}), S_j^r\big) \notin \mathcal{T}_{\epsilon_1}^{(r)}(Q_S P_U)\big\}, \tag{I.18c}$$

$$\mathcal{E}_{2,j} = \big\{\big(U^r(M_j, K_{j-1}), S_j^r, Y_j^r\big) \notin \mathcal{T}_{\epsilon_2}^{(r)}(Q_S P_U W_{Y|U,S})\big\}, \tag{I.18d}$$

$$\mathcal{E}_{3,j} = \big\{\big(U^r(k_{j-1}, \hat{m}_j), S_j^r, Y_j^r\big) \in \mathcal{T}_{\epsilon_2}^{(r)}, \quad \text{for some} \quad \hat{m}_j \neq M_j\big\}, \tag{I.18e}$$

where $\epsilon_2 > \epsilon_1 > \epsilon > 0$. The probability of error is upper bounded as follows,

$$\mathbb{P}(\mathcal{E}) \leq \mathbb{P}\Big\{\bigcup\nolimits_{j=1}^{B} \mathcal{E}_j\Big\} \leq \sum_{j=1}^{B} \mathbb{P}(\mathcal{E}_j). \tag{I.19}$$

Now we bound $\mathbb{P}(\mathcal{E}_j)$ by using union bound

$$\mathbb{P}(\mathcal{E}_j) \leq \mathbb{P}(\mathcal{E}_{1,j}) + \mathbb{P}(\mathcal{E}_{1,j}^c \cap \mathcal{E}_{2,j}) + \mathbb{P}(\mathcal{E}_{2,j}^c \cap \mathcal{E}_{3,j}). \tag{I.20}$$

By the law of large numbers the first and second term on RHS of (I.20) vanishes when r grows. According to the law of large numbers and the packing lemma, the last term on RHS of (I.20) vanishes when r grows if [85],

$$R < \mathbb{I}(U; S, Y) = \mathbb{I}(U; Y|S). \tag{I.21}$$

Furthermore, this scheme requires that

$$R_k \leq R_K = \mathbb{H}(S|Z) - \epsilon, \tag{I.22}$$

The region in Theorem 8 is obtained by applying Fourier-Motzkin to (I.17a), (I.21), and (I.22).

Converse Proof: We now develop an upper bound when CSI is available causally at both of the legitimate terminals. Consider any sequence of length-n codes for a state-dependent channel with CSI available causally at both the transmitter and the receiver

such that $P_e^{(n)} \leq \epsilon_n$ and $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \delta$ with $\lim_{n\to\infty} \epsilon_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish.

Epsilon Rate Region: We first define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ that expands the region defined in (3.4) as follows,

$$\mathcal{A}_\epsilon \triangleq \left\{R \geq 0 : \exists P_{S,U,X,Y,Z} \in \mathcal{D}_\epsilon : R \leq \mathbb{I}(U;Y|S) + \epsilon\right\}, \tag{I.23a}$$

where

$$\mathcal{D}_\epsilon = \left\{ \begin{array}{l} P_{S,U,X,Y,Z} : \\ P_{S,U,X,Y,Z} = Q_S P_U \mathbb{1}_{\left\{X=X(U,S)\right\}} W_{Y,Z|X,S} \\ \mathbb{D}\left(P_Z \| Q_0\right) \leq \epsilon \\ \mathbb{H}(S|Z) \geq \mathbb{I}(U;Z|S) - \mathbb{I}(U;Y|S) - 2\epsilon \\ |\mathcal{U}| \leq |\mathcal{X}| + 1 \end{array} \right\}. \tag{I.23b}$$

We next show that if a rate R is achievable then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$ and $\nu > 0$, we start by upper bounding nR using standard techniques,

$$\begin{aligned} nR &= \mathbb{H}(M) \\ &\overset{(a)}{\leq} \mathbb{H}(M|S^n) - \mathbb{H}(M|Y^n, S^n) + n\epsilon_n \\ &= \mathbb{I}(M; Y^n|S^n) + n\epsilon_n \\ &= \sum_{i=1}^n \mathbb{I}(M; Y_i|Y^{i-1}, S^n) + n\epsilon_n \\ &= \sum_{i=1}^n [\mathbb{H}(Y_i|Y^{i-1}, S^n) - \mathbb{H}(Y_i|M, Y^{i-1}, S^n)] + n\epsilon_n \\ &\overset{(b)}{\leq} \sum_{i=1}^n [\mathbb{H}(Y_i|S_i) - \mathbb{H}(Y_i|U_i, S_i)] + n\epsilon_n \\ &= \sum_{i=1}^n \mathbb{I}(U_i; Y_i|S_i) + n\epsilon_n \end{aligned}$$

$$\begin{aligned}
&\overset{(c)}{\leq} n\mathbb{I}(\tilde{U};\tilde{Y}|\tilde{S}) + n\epsilon_n \\
&\overset{(d)}{\leq} n\mathbb{I}(\tilde{U};\tilde{Y}|\tilde{S}) + n\epsilon \\
&\overset{(e)}{=} n\mathbb{I}(U;Y|S) + n\epsilon
\end{aligned} \tag{I.24}$$

where

(a) follows from Fano's inequality and since M is independent of S^n;

(b) holds because conditioning does not increase entropy and $U_i = (M, Y^{i-1}, S^n_{\sim i})$;

(c) follows from the concavity of mutual information, with the resulting random variables $\tilde{U}$, $\tilde{S}$, and $\tilde{Y}$ having the following distributions

$$\tilde{P}_{U,S,X}(u,s,x) \triangleq \frac{1}{n}\sum_{i=1}^{n} P_{U_i,S_i,X_i}(u,s,x), \tag{I.25a}$$

$$\tilde{P}_{U,S,X,Y,Z}(u,s,x,y,z) \triangleq \tilde{P}_{U,S,X}(u,s,x) W_{Y,Z|X,S}(y,z|x,s); \tag{I.25b}$$

(d) follows by defining $\epsilon \triangleq \max\{\epsilon_n, \nu\}$, where we choose n large enough such that $\nu \geq \frac{\delta}{n}$;

(e) follows by defining $U = \tilde{U}$, $Y = \tilde{Y}$, and $S = \tilde{S}$.

We now have,

$$\begin{aligned}
nR &\overset{(a)}{\geq} n\mathbb{I}(\tilde{X},\tilde{S};\tilde{Z}) - n\mathbb{H}(\tilde{S}) - \epsilon \\
&\overset{(b)}{\geq} n\mathbb{I}(\tilde{U},\tilde{S};\tilde{Z}) - n\mathbb{H}(\tilde{S}) - \epsilon \\
&\overset{(c)}{=} \mathbb{I}(U,S;Z) - \mathbb{H}(S) - \epsilon,
\end{aligned} \tag{I.26}$$

where

(a) follows from the exact same steps as in (H.15);

(b) follows from the Markov chain $U - (X,S) - Z$ and from the definition of random variables $\tilde{U}$, $\tilde{X}$, $\tilde{S}$, $\tilde{Y}$, and $\tilde{Z}$ in (I.25);

(c) follows by defining $U = \tilde{U}$, $Z = \tilde{Z}$, and $S = \tilde{S}$.

Rewriting the bound in (I.26) by using the basic property in (H.8) leads to

$$R \geq \mathbb{I}(U; Z|S) - \mathbb{H}(S|Z) - \epsilon. \tag{I.27}$$

To show that $\mathbb{D}(P_Z||Q_0) \leq \epsilon$, note that for n large enough,

$$\begin{aligned}\mathbb{D}(P_Z||Q_0) = \mathbb{D}(P_{\tilde{Z}}||Q_0) &= \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} P_{Z_i}\middle|\middle|Q_0\right) \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \leq \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \frac{\delta}{n} \leq \nu \leq \epsilon.\end{aligned} \tag{I.28}$$

Combining (I.24) and (I.27) shows that $\forall \epsilon_n, \nu > 0$, $R \leq \max\{a : a \in \mathcal{A}_\epsilon\}$. Therefore,

$$C_{\text{C-TR}} = \max\left\{a : a \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon\right\}. \tag{I.29}$$

Continuity at Zero: Continuity at zero for $\mathcal{A}_\epsilon$ is established by substituting $\min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\}$ with $\mathbb{I}(U;Y|S)$ and $\mathbb{I}(V;Z) - \mathbb{I}(V;S)$ with $\mathbb{I}(U;Z|S) - \mathbb{H}(S|Z)$ in the continuity at zero proof in Appendix N and following the same arguments.

APPENDIX J

PROOF OF THEOREM 9

Achievability Proof: We adopt a block-Markov encoding scheme in which B independent messages are transmitted over B channel blocks each of length r, such that $n = rB$. The warden's observation is $Z^n = (Z_1^r, \ldots, Z_B^r)$, the target output distribution is $Q_0^{\otimes n}$, and Equation (I.2), describing the distance between the two distributions, continues to hold. The random code generation is as follows.

Fix P_X and $\epsilon_1 > \epsilon_2 > 0$ such that, $P_Z = Q_0$.

Codebook Generation for Keys: For each block $j \in [\![1, B]\!]$, create a function $\Phi : S_j^r \mapsto [\![1, 2^{rR_K}]\!]$ through random binning by choosing the value of $\Phi(s_j^r)$ independently and uniformly at random for every $s_j^r \in \mathcal{S}^r$. The key $k_j = \Phi(s_j^r)$ obtained in the block $j \in [\![1, B]\!]$ from the state sequence s_j^r is used to assist the encoder in the next block.

Codebook Generation for Messages: For each block $j \in [\![1, B]\!]$, let $C_r \triangleq \{X^r(m_j, k_{j-1})\}_{(m_j, k_{j-1}) \in \mathcal{M} \times \mathcal{K}}$, where $\mathcal{M} \triangleq [\![1, 2^{rR}]\!]$ and $\mathcal{K} \triangleq [\![1, 2^{rR_k}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_X^{\otimes r}$. We denote a realization of C_r by $\mathcal{C}_r \triangleq \{x^r(m_j, k_{j-1})\}_{(m_j, k_{j-1}) \in \mathcal{M} \times \mathcal{K}}$.

Encoding: For the first block, we assume that the transmitter and the receiver have access to a shared secret key k_0, in this block to transmit m_1 the encoder computes $x^r(m_1, k_0)$ and transmits it over the channel. At the end of the first block, the encoder generates a key from CSI s_1^r to be used in Block 2.

For block $j \in [\![2, B]\!]$, to send the message m_j according to the generated key k_{j-1} from the previous block, the encoder computes $x^r(m_j, k_{j-1})$ and transmits it over the channel. Also, at the end of the block $j \in [\![2, B]\!]$, the encoder generates a key from CSI s_j^r to be used in the next block.

Define

$$\Upsilon^{(\mathcal{C}_r)}_{M_j, K_{j-1}, X^r, S_j^r, Z_j^r, K_j}(m_j, k_{j-1}, \tilde{x}^r, s_j^r, z_j^r, k_j) \triangleq 2^{-r(R+R_k)} \mathbb{1}_{\{\tilde{x}^r = x^r(m_j, k_{j-1})\}} Q_S^{\otimes r}(s_j^r)$$

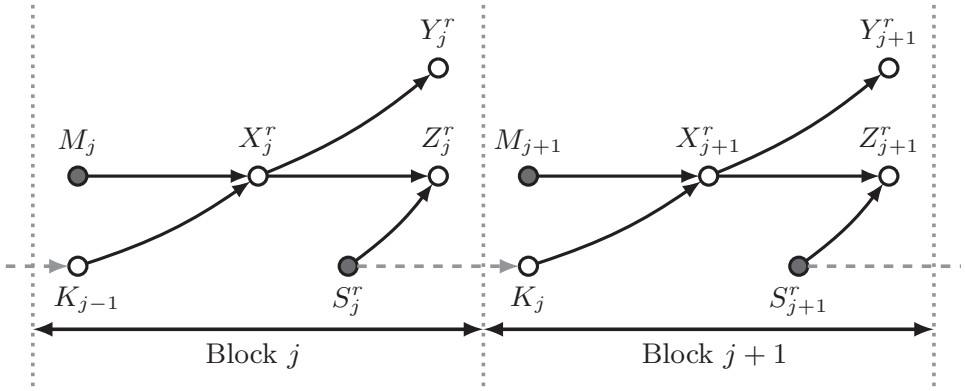

Figure J.1. Functional dependence graph for the block-Markov encoding scheme

$$\times W_{Z|X,S}^{\otimes r}\big(z_j^r|\tilde{x}^r, s_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}. \tag{J.1}$$

For a fixed codebook $\mathcal{C}_r$, the induced joint distribution by our code design (i.e. $P^{(\mathcal{C}_r)}$) satisfies

$$\mathbb{D}\Big(P^{(\mathcal{C}_r)}_{M_j,K_{j-1},X^r,S_j^r,Z_j^r,K_j}||\Upsilon^{(\mathcal{C}_r)}_{M_j,K_{j-1},X^r,S_j^r,Z_j^r,K_j}\Big) \leq \epsilon. \tag{J.2}$$

This intermediate distribution $\Upsilon^{(\mathcal{C}_r)}$ approximates the true distribution $P^{(\mathcal{C}_r)}$ and will be used in the sequel for bounding purposes. Expression (J.2) holds because the main difference between $P^{(\mathcal{C}_r)}$ and $\Upsilon^{(\mathcal{C}_r)}$ is that the key K_{j-1} is assumed to be uniformly distributed in $\Upsilon^{(\mathcal{C}_r)}$, which is made (arbitrarily) nearly uniform in $P^{(\mathcal{C}_r)}$ with appropriate control of rate as in (J.10).

Covert Analysis: We now show that this coding scheme guarantees that $\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n})] \underset{n\to\infty}{\longrightarrow} 0$, where C_n is the set of all the codebooks for all blocks, and

$$Q_Z(\cdot) = \sum_{s\in\mathcal{S}}\sum_{x\in\mathcal{X}} Q_S(s)P_X(x)W_{Z|X,S}(\cdot|x,s). \tag{J.3}$$

Then we choose P_X such that it satisfies $Q_Z = Q_0$. Similar to (I.10), by using the functional dependence graph depicted in Fig. J.1,

$$\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n}) \leq 2\sum_{j=1}^{B}\mathbb{D}(P_{Z_j^r,K_j|C_r}||Q_Z^{\otimes r}Q_{K_j}). \tag{J.4}$$

We now proceed to bound the RHS of (J.4). First, consider the following marginal from (J.1),

$$\Upsilon_{Z_j^r,K_j|C_r}(z_j^r,k_j)=\sum_{m_j}\sum_{k_{j-1}}\sum_{s_j^r}\frac{1}{2^{r(R+R_k)}}Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|X^r(m_j,k_{j-1}),s_j^r\big)\mathbb{1}_{\{k_j=\Phi(s_j^r)\}}. \tag{J.5}$$

To bound the RHS of (J.4) by using Lemma 1 and the triangle inequality we have,

$$\mathbb{E}_{C_r}||P_{Z_j^r,K_j|C_r}-Q_Z^{\otimes r}Q_{K_j}||_1\leq\mathbb{E}_{C_r}||P_{Z_j^r,K_j|C_r}-\Upsilon_{Z_j^r,K_j|C_r}||_1+\mathbb{E}_{C_r}||\Upsilon_{Z_j^r,K_j|C_r}-Q_Z^{\otimes r}Q_{K_j}||_1. \tag{J.6}$$

From Lemma 1 and (J.2) the first term on the RHS of (J.6) vanishes as r grows; to bound the second term by using Lemma 1 we have,

$$\mathbb{E}_{C_r}[\mathbb{D}(\Upsilon_{Z_j^r,K_j|C_r}||Q_Z^{\otimes r}Q_{K_j})]=\mathbb{E}_{C_r}\bigg[\sum_{(z_j^r,k_j)}\Upsilon_{Z_j^r,K_j|C_r}(z_j^r,k_j)\log\bigg(\frac{\Upsilon_{Z_j^r,K_j|C_r}(z_j^r,k_j)}{Q_Z^{\otimes r}(z_j^r)Q_{K_j}(k_j)}\bigg)\bigg]$$

$$=\mathbb{E}_{C_r}\Bigg[\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k)}}\sum_{s_j^r}Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|X^r(m_j,k_{j-1}),s_j^r\big)\mathbb{1}_{\{k_j=\Phi(s_j^r)\}}$$
$$\times\log\Bigg(\frac{\sum_{\tilde{m}_j}\sum_{\tilde{k}_{j-1}}\sum_{\tilde{s}_j^r}Q_S^{\otimes r}(\tilde{s}_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|X^r(\tilde{m}_j,\tilde{k}_{j-1}),\tilde{s}_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\Bigg)\Bigg]$$

$$\overset{(a)}{\leq}\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k)}}\sum_{s_j^r}\sum_{x^r(m_j,k_{j-1})}\Upsilon_{X^r,S^r,Z^r}^{\otimes r}\big(x^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\times\mathbb{E}_{\Phi(s_j^r)}\big[\mathbb{1}_{\{k_j=\Phi(s_j^r)\}}\big]$$
$$\times\log\mathbb{E}_{\backslash((m_j,k_{j-1}),\Phi(s_j^r))}\Bigg[\frac{\sum_{\tilde{m}_j}\sum_{\tilde{k}_{j-1}}\sum_{\tilde{s}_j^r}Q_S^{\otimes r}(\tilde{s}_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|X^r(\tilde{m}_j,\tilde{k}_{j-1}),\tilde{s}_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\Bigg]$$

$$\overset{(b)}{\leq}\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k)}}\sum_{s_j^r}\sum_{x^r(m_j,k_{j-1})}\Upsilon_{X^r,S^r,Z^r}^{\otimes r}\big(x^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\times\frac{1}{2^{rR_K}}$$
$$\times\log\frac{1}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\Bigg(Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1}),s_j^r\big)$$
$$+\mathbb{E}_{\backslash(m_j,k_{j-1})}\Bigg[\sum_{(\tilde{m}_j,\tilde{k}_{j-1})\neq(m_j,k_{j-1})}Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|X^r(\tilde{m}_j,\tilde{k}_{j-1}),s_j^r\big)\Bigg]$$

$$
\begin{aligned}
&+\mathbb{E}_{\setminus\Phi(s_j^r)}\left[\sum_{\tilde{s}_j^r\neq s_j^r} Q_S^{\otimes r}(\tilde{s}_j^r)\times W_{Z|X,S}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1}),\tilde{s}_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}\right]\\
&+\mathbb{E}_{\setminus((m_j,k_{j-1}),\Phi(s_j^r))}\left[\sum_{\tilde{s}_j^r\neq s_j^r}\sum_{(\tilde{m}_j,\tilde{k}_{j-1})\neq(m_j,k_{j-1})} Q_S^{\otimes r}(\tilde{s}_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|X^r(\tilde{m}_j,\tilde{k}_{j-1}),\tilde{s}_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}\right]\Bigg)\\
&\overset{(c)}{\leq}\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k+R_K)}}\sum_{s_j^r}\sum_{x^r(m_j,k_{j-1})}\Upsilon_{X^r,S^r,Z^r}^{\otimes r}\big(x^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\\
&\quad\times\log\Bigg(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1}),s_j^r\big)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}+\sum_{(\tilde{m}_j,\tilde{k}_{j-1})\neq(m_j,k_{j-1})}\frac{Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&\quad+\sum_{\tilde{s}_j^r\neq s_j^r}\frac{Q_S^{\otimes r}(\tilde{s}_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1}),\tilde{s}_j^r\big)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\Bigg)\\
&\leq\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k+R_K)}}\sum_{s_j^r}\sum_{x^r(m_j,k_{j-1})}\Upsilon_{X^r,S^r,Z^r}^{\otimes r}\big(x^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\\
&\times\log\left(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1}),s_j^r\big)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}+\frac{2^{rR_K}Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|X}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1})\big)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\right)\\
&\triangleq\Psi_1+\Psi_2,
\end{aligned}
\tag{J.7}
$$

where (a) follows from Jensen's inequality, (b) and (c) are because $\mathbb{1}_{\{.\}}\leq 1$, and the last term in the RHS of (b) is smaller than 1. We define Ψ_1 and Ψ_2 as

$$
\begin{aligned}
\Psi_1=&\sum_{k_j}\sum_{k_{j-1}}\sum_{m_j}\frac{1}{2^{r(R+R_k+R_K)}}\sum_{\left(x^r(m_j,k_{j-1}),s_j^r,z_j^r\right)\in\mathcal{T}_\epsilon^{(n)}}\Upsilon_{X^r,S^r,Z^r}^{\otimes r}\big(x^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\\
&\times\log\Bigg(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1}),s_j^r\big)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&+\frac{2^{rR_K}Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|X}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1})\big)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\Bigg)\\
\leq&\log\left(\frac{2^{rR_K}2^{-r(1-\epsilon)\left(\mathbb{H}(S)+\mathbb{H}(Z|X,S)\right)}}{2^{r(R+R_k)}2^{-r(1+\epsilon)\mathbb{H}(Z)}}+\frac{2^{rR_K}2^{-r(1-\epsilon)\mathbb{H}(S,Z)}}{2^{-r(1+\epsilon)\mathbb{H}(Z)}}+\frac{2^{-r(1-\epsilon)\mathbb{H}(Z|X)}}{2^{r(R+R_k)}2^{-r(1+\epsilon)\mathbb{H}(Z)}}+1\right)
\end{aligned}
\tag{J.8}
$$

$$
\Psi_2=\sum_{k_j}\sum_{k_{j-1}}\sum_{m_j}\frac{1}{2^{r(R+R_k+R_K)}}\sum_{\left(x^r(m_j,k_{j-1}),s_j^r,z_j^r\right)\notin\mathcal{T}_\epsilon^{(n)}}\Upsilon_{X^r,S^r,Z^r}^{\otimes r}\big(x^r(m_j,k_{j-1}),s_j^r,z_j^r\big)
$$

$$\times \log\left(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1}),s_j^r\big)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\right.$$
$$\left.+\frac{2^{rR_K}Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|X}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1})\big)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\right)$$
$$\leq 2|X||S||Z|e^{-r\epsilon^2\mu_{X,S,Z}}r\log\Big(\frac{3}{\mu_Z}+1\Big), \tag{J.9}$$

where $\mu_{X,S,Z}=\min_{(x,s,z)\in(\mathcal{X},\mathcal{S},\mathcal{Z})}P_{X,S,Z}(x,s,z)$ and $\mu_Z=\min_{z\in\mathcal{Z}}P_Z(z)$. When $r\to\infty$ then $\Psi_2\to 0$, by choosing $R_K=\mathbb{H}(S|Z)-\epsilon$, Ψ_1 vanishes when r grows if,

$$R+R_k>\mathbb{I}(X;Z|S), \tag{J.10a}$$
$$R+R_k>\mathbb{I}(X;Z). \tag{J.10b}$$

Since X and S are independent, (J.10b) is redundant because of (J.10a).

Decoding and Error Probability Analysis: At the end of the block $j\in[\![1,B]\!]$, using its knowledge of the CSI s_j^r of the current block and the key k_{j-1} generated from the previous block, the receiver finds a unique $\hat{m}_j$ such that $\big(u^r(\hat{m}_j,k_{j-1}),s_j^r,y_j^r\big)\in\mathcal{T}_\epsilon^{(r)}$. To analyze the probability of error, we define the following error events for $j\in[\![1,B]\!]$,

$$\mathcal{E}=\big\{\hat{M}\neq M\big\}, \tag{J.11a}$$
$$\mathcal{E}_j=\big\{\hat{M}_j\neq M_j\big\}, \tag{J.11b}$$
$$\mathcal{E}_{1,j}=\big\{\big(X^r(M_j,K_{j-1}),S_j^r\big)\notin\mathcal{T}_{\epsilon_1}^{(r)}(Q_SP_X)\big\}, \tag{J.11c}$$
$$\mathcal{E}_{2,j}=\big\{\big(X^r(M_j,K_{j-1}),S_j^r,Y_j^r\big)\notin\mathcal{T}_{\epsilon_2}^{(r)}(Q_SP_XW_{Y|X,S})\big\}, \tag{J.11d}$$
$$\mathcal{E}_{3,j}=\big\{\big(X^r(k_{j-1},\hat{m}_j),S_j^r,Y_j^r\big)\in\mathcal{T}_{\epsilon_2}^{(r)},\quad\text{for some}\quad\hat{m}_j\neq M_j\big\}, \tag{J.11e}$$

where $\epsilon_2>\epsilon_1>\epsilon>0$. The probability of error is upper bounded as follows,

$$\mathbb{P}(\mathcal{E})\leq\mathbb{P}\Big\{\bigcup_{j=1}^B\mathcal{E}_j\Big\}\leq\sum_{j=1}^B\mathbb{P}(\mathcal{E}_j). \tag{J.12}$$

Now we bound $\mathbb{P}(\mathcal{E}_j)$ by using union bound

$$\mathbb{P}(\mathcal{E}_j)\leq\mathbb{P}(\mathcal{E}_{1,j})+\mathbb{P}(\mathcal{E}_{1,j}^c\cap\mathcal{E}_{2,j})+\mathbb{P}(\mathcal{E}_{2,j}^c\cap\mathcal{E}_{3,j}). \tag{J.13}$$

By the law of large numbers the first and second term on RHS of (J.13) vanishes when r grows. According to the law of large numbers and the packing lemma, the last term on RHS of (J.13) vanishes when r grows if [85],

$$R < \mathbb{I}(X; S, Y) = \mathbb{I}(X; Y|S). \tag{J.14}$$

Furthermore, this scheme requires that,

$$R_k \leq R_K = \mathbb{H}(S|Z) - \epsilon. \tag{J.15}$$

The region in Theorem 9 is obtained by applying Fourier-Motzkin to (J.10a), (J.14), and (J.15).

Remark 32. *In the achievability proof of Theorem 8 and Theorem 9 we transmit B messages over B blocks. We assume that there exists a shared secret key between the transmitter and the receiver that is used in the first block to bootstrap the covert communication. Consequently, the shared secret key rate is negligible. However, to eliminate the need for this secret key, similar to the block Markov encoding schemes in [89, 64] we can transmit $B-1$ messages over B blocks and remove the decodability condition of the message of the first block, this results in a slight rate loss in the first block, which becomes asymptotically negligible as the number of blocks $B \to \infty$.*

Converse Proof: To establish the upper bound, consider any sequence of length-n codes for a state-dependent channel with CSI available strictly causally at both the transmitter and the receiver, such that $P_e^{(n)} \leq \epsilon_n$ and $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \delta$ with $\lim_{n\to\infty} \epsilon_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish.

Epsilon Rate Region: We first define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ that expands the region defined in (3.6) as follows,

$$\mathcal{A}_\epsilon \triangleq \big\{R \geq 0 : \exists P_{S,X,Y,Z} \in \mathcal{D}_\epsilon : R \leq \mathbb{I}(X; Y|S) + \epsilon\big\}, \tag{J.16a}$$

where

$$\mathcal{D}_\epsilon = \left\{ \begin{array}{l} P_{S,X,Y,Z} : \\ P_{S,X,Y,Z} = Q_S P_X W_{Y,Z|X,S} \\ \mathbb{D}\left(P_Z \| Q_0\right) \leq \epsilon \\ \mathbb{H}(S|Z) \geq \mathbb{I}(X;Z|S) - \mathbb{I}(X;Y|S) - 2\epsilon \end{array} \right\}. \tag{J.16b}$$

We next show that if a rate R is achievable then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$ and $\nu > 0$, we start by upper bounding nR using standard techniques.

$$\begin{aligned} nR &= \mathbb{H}(M) \\ &\overset{(a)}{\leq} \mathbb{H}(M|S^n) - \mathbb{H}(M|Y^n, S^n) + n\epsilon_n \\ &= \mathbb{I}(M; Y^n|S^n) + n\epsilon_n \\ &= \sum_{i=1}^{n} \mathbb{I}(M; Y_i|Y^{i-1}, S^n) + n\epsilon_n \\ &\leq \sum_{i=1}^{n} \left[\mathbb{H}(Y_i|Y^{i-1}, S^n) - \mathbb{H}(Y_i|Y^{i-1}, S^n, X_i, M)\right] + n\epsilon_n \\ &\overset{(b)}{\leq} \sum_{i=1}^{n} \mathbb{I}(X_i; Y_i|S_i) + n\epsilon_n \\ &\overset{(c)}{\leq} n\mathbb{I}(\tilde{X}; \tilde{Y}|\tilde{S}) + n\epsilon_n \\ &\overset{(d)}{\leq} n\mathbb{I}(\tilde{X}; \tilde{Y}|\tilde{S}) + n\epsilon \\ &\overset{(e)}{=} n\mathbb{I}(X; Y|S) + n\epsilon, \end{aligned} \tag{J.17}$$

where

(a) follows from Fano's inequality and since M is independent of S^n;

(b) holds because conditioning does not increase entropy and $(M, Y^{i-1}, S^n_{\sim i}, X^n_{\sim i}) - (X_i, S_i) - Y_i$ forms a Markov chain;

(c) follows from concavity of mutual information, with respect to the input distribution, with the random variables $\tilde{X}$, $\tilde{S}$, $\tilde{Y}$, and $\tilde{Z}$ having the following distributions

$$\tilde{P}_{X,S}(x,s) \triangleq \frac{1}{n}\sum_{i=1}^{n} P_{X_i,S_i}(x,s), \tag{J.18a}$$

$$\tilde{P}_{X,S,Y,Z}(x,s,y,z) \triangleq \tilde{P}_{X,S}(x,s) W_{Y,Z|X,S}(y,z|x,s); \tag{J.18b}$$

(d) follows by defining $\epsilon \triangleq \max\{\epsilon_n, \nu\}$, where we choose n large enough such that $\nu \geq \frac{\delta}{n}$;

(e) follows by defining $U = \tilde{U}$, $Y = \tilde{Y}$, and $S = \tilde{S}$.

By following the same steps as in (H.15) we also have,

$$nR \geq n\mathbb{I}(\tilde{X},\tilde{S};\tilde{Z}) - n\mathbb{H}(\tilde{S}) - \epsilon, \tag{J.19}$$

where the random variables $\tilde{X}$, $\tilde{S}$, $\tilde{Y}$, and $\tilde{Z}$ have been defined in (J.18). Substituting (H.8) into (J.19) leads to

$$\begin{aligned} R &\geq \mathbb{I}(\tilde{X};\tilde{Z}|\tilde{S}) - \mathbb{H}(\tilde{S}|\tilde{Z}) - \epsilon \\ &= \mathbb{I}(X;Z|S) - \mathbb{H}(S|Z) - \epsilon, \end{aligned} \tag{J.20}$$

where the last equality follows by defining $U = \tilde{U}$, $Z = \tilde{Z}$, and $S = \tilde{S}$. To show that $\mathbb{D}(P_Z||Q_0) \leq \epsilon$, note that for n large enough,

$$\begin{aligned} \mathbb{D}(P_Z||Q_0) = \mathbb{D}(P_{\tilde{Z}}||Q_0) &= \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} P_{Z_i}\middle\|Q_0\right) \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \leq \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \frac{\delta}{n} \leq \nu \leq \epsilon. \end{aligned} \tag{J.21}$$

Combining (J.17) and (J.20) shows that $\forall \epsilon_n, \nu > 0$, $R \leq \max\{a : a \in \mathcal{A}_\epsilon\}$. Therefore,

$$C_{\text{SC-TR}} = \max\left\{a : a \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon\right\}. \tag{J.22}$$

Continuity at Zero: One can prove the continuity at zero of $\mathcal{A}_\epsilon$ by substituting $\min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\}$ with $\mathbb{I}(X;Y|S)$ and $\mathbb{I}(V;Z) - \mathbb{I}(V;S)$ with $\mathbb{I}(X;Z|S) - \mathbb{H}(S|Z)$ in the continuity at zero proof in Appendix N and following the exact same arguments.

APPENDIX K

PROOF OF THEOREM 10

We adopt a block-Markov encoding scheme in which B independent messages are transmitted over B channel blocks each of length r, such that $n = rB$. The warden's observation is $Z^n = (Z_1^r, \ldots, Z_B^r)$, the distribution induced at the output of the warden is P_{Z^n}, the target output distribution is $Q_0^{\otimes n}$, and Equation (I.2), describing the distance between the two distributions, continues to hold. The random code generation is as follows.

Fix $P_{U|S}(u|s)$, $P_{V|S}(v|s)$, $x(u,s)$, and $\epsilon_1 > \epsilon_2 > 0$ such that, $P_Z = Q_0$.

Codebook Generation for Keys: For each block $j \in [\![1, B]\!]$, let $C_1^{(r)} \triangleq \left\{V^r(a_j)\right\}_{a_j \in \mathcal{A}}$, where $\mathcal{A} \triangleq [\![1, 2^{r\tilde{R}}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_V^{\otimes r}$, where $P_V = \sum_{s \in \mathcal{S}} Q_S(s) P_{V|S}(v|s)$. We denote a realization of $C_1^{(r)}$ by $\mathcal{C}_1^{(r)} \triangleq \left\{v^r(a_j)\right\}_{a_j \in \mathcal{A}}$. Partition the set of indices $a_j \in [\![1, 2^{r\tilde{R}}]\!]$ into bins $\mathcal{B}(t)$, $t \in [\![1, 2^{rR_T}]\!]$ by using function $\varphi : V^r(a_j) \mapsto [\![1, 2^{rR_T}]\!]$ through random binning by choosing the value of $\varphi(v^r(a_j))$ independently and uniformly at random for every $v^r(a_j) \in \mathcal{V}^r$. For each block $j \in [\![1, B]\!]$, create a function $\Phi : V^r(a_j) \mapsto [\![1, 2^{rR_K}]\!]$ through random binning by choosing the value of $\Phi(v^r(a_j))$ independently and uniformly at random for every $v^r(a_j) \in \mathcal{V}^r$. The key $k_j = \Phi(v^r(a_j))$ obtained in block $j \in [\![1, B]\!]$ from the description of the CSI sequence $v^r(a_j)$ is used to assist the encoder in block $j + 2$.

Codebook Generation for Messages: For each block $j \in [\![1, B]\!]$, let $C_2^{(r)} \triangleq \left\{U^r(m_j, t_{j-1}, k_{j-2}, \ell_j)\right\}_{(m_j, t_{j-1}, k_{j-2}, \ell_j) \in \mathcal{M} \times \mathcal{T} \times \mathcal{K} \times \mathcal{L}}$, where $\mathcal{M} \triangleq [\![1, 2^{rR}]\!]$, $\mathcal{T} \triangleq [\![1, 2^{rR_t}]\!]$, $\mathcal{K} \triangleq [\![1, 2^{rR_k}]\!]$, and $\mathcal{L} \triangleq [\![1, 2^{rR'}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_U^{\otimes r}$. We denote a realization of $C_2^{(r)}$ by $\mathcal{C}_2^{(r)} \triangleq \left\{u^r(m_j, t_{j-1}, k_{j-2}, \ell_j)\right\}_{(m_j, t_{j-1}, k_{j-2}, \ell_j) \in \mathcal{M} \times \mathcal{T} \times \mathcal{K} \times \mathcal{L}}$. Let, $C_r = \left\{C_1^{(r)}, C_2^{(r)}\right\}$ and $\mathcal{C}_r = \left\{\mathcal{C}_1^{(r)}, \mathcal{C}_2^{(r)}\right\}$. The indices $(m_j, t_{j-1}, k_{j-2}, \ell_j)$ can be viewed as a three layer binning. We define an ideal PMF for codebook $\mathcal{C}_r$, as an approximate distribution to facilitate the analysis

$$\Gamma^{(\mathcal{C}_r)}_{M_j, T_{j-1}, K_{j-2}, L_j, A_j, U^r, V^r, S_j^r, Z_j^r, K_{j-1}, T_j, K_j}(m_j, t_{j-1}, k_{j-2}, \ell_j, a_j, \tilde{u}^r, \tilde{v}_j^r, s_j^r, z_j^r, k_{j-1}, t_j, k_j)$$

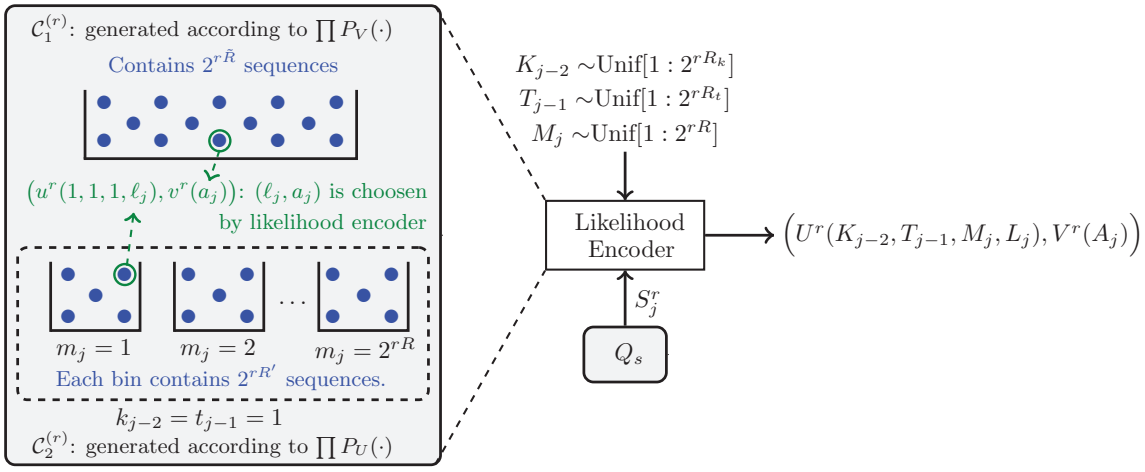

Figure K.1. Proposed coding scheme for the dual use of CSI

$$= 2^{-r(R+R_t+R_k+R'+\tilde{R})} \mathbb{1}_{\{\tilde{u}^r = u^r(m_j, t_{j-1}, k_{j-2}, \ell_j)\}} \mathbb{1}_{\{\tilde{v}^r = v^r(a_j)\}} P_{S|U,V}^{\otimes r}(s_j^r | \tilde{u}^r, \tilde{v}^r)$$
$$\times W_{Z|U,S}^{\otimes r}(z_j^r | \tilde{u}^r, s_j^r) 2^{-rR_k} \mathbb{1}_{\{t_j = \sigma(\tilde{v}^r)\}} \mathbb{1}_{\{k_j = \Phi(\tilde{v}^r)\}}, \tag{K.1}$$

where $W_{Z|U,S}$ is the marginal distribution $W_{Z|U,S} = \sum_{x \in \mathcal{X}} \mathbb{1}_{\{x = x(u,s)\}} W_{Z|X,S}$ and $P_{S|U,V}$ is defined as follows

$$P_{S|U,V}(s|u,v) \triangleq \frac{P_{S,U,V}(s,u,v)}{P_{U,V}(u,v)} = \frac{Q_S(s) P_{U|S}(u|s) P_{V|S}(v|s)}{\sum_{s \in \mathcal{S}} Q_S(s) P_{U|S}(u|s) P_{V|S}(v|s)}. \tag{K.2}$$

Encoding: We assume that the transmitter and the receiver have access to shared secret keys k_{-1} and k_0 for the first two blocks, but after the first two blocks they use the key that they generate from the CSI.

In the first block, to send the message m_1 according to k_{-1}, the encoder generates the index t_0 uniformly at random and then generates the indices ℓ_1 and a_1 according to the following distribution with $j = 1$,

$$f(\ell_j, a_j | s_j^r, m_j, t_{j-1}, k_{j-2}) = \frac{P_{S|U,V}^{\otimes r}\left(s_j^r | u^r(m_j, t_{j-1}, k_{j-2}, \ell_j), v^r(a_j)\right)}{\sum_{\ell_j' \in [\![1, 2^{rR'}]\!]} \sum_{a_j' \in [\![1, 2^{r\tilde{R}}]\!]} P_{S|U,V}^{\otimes r}\left(s_j^r | u^r(m_j, t_{j-1}, k_{j-2}, \ell_j'), v^r(a_j')\right)}, \tag{K.3}$$

where $P_{S|U,V}$ is defined in (K.2). Based on these indices, the encoder computes $u^r(m_1, t_0, k_{-1}, \ell_1)$ and $v^r(a_1)$ and transmits codeword x^r, where $x_i = x(u_i(m_1, t_0, k_{-1}, \ell_1), s_i)$. Note that, the index t_0 does not convey any useful information. Simultaneously, it uses the description of the CSI $v^r(a_1)$ to generate a reconciliation index t_1 and a key k_1 to be used in the second and the third blocks, respectively.

In the second block, to send the message m_2 and reconciliation index t_1 according to k_0, the encoder generates the indices ℓ_2 and a_2 according to the likelihood encoder described in (K.3) with $j = 2$. Based on these indices, the encoder computes $u^r(m_2, t_1, k_0, \ell_2)$ and $v^r(a_2)$ and transmits codeword x^r, where $x_i = x(u_i(m_2, t_1, k_0, \ell_1), s_i)$. Simultaneously, it uses the description of the CSI $v^r(a_2)$ to generate a reconciliation index t_2 and a key k_2 to be used in the third and the fourth block, respectively.

In block $j \in [\![3, B]\!]$, to send the message m_j and the reconciliation index t_{j-1}, generated in the previous block, according to the key k_{j-2}, generated in the block $j-2$, and the CSI of the current block, the encoder generates indices ℓ_j and a_j from the bin (m_j, t_{j-1}, k_{j-2}) according to the likelihood encoder described in (K.3). The encoder then transmits the codeword x^r, where each coordinate of the transmitted signal is a function of the CSI, as well as the corresponding sample of the transmitter's codeword u_i, i.e., $x_i = x(u_i(m_j, t_{j-1}, k_{j-2}, \ell_j), s_i)$. Simultaneously, the encoder uses the description of the CSI $v^r(a_j)$ to generate a reconciliation index t_j and a key k_j to be used in the block $j+1$ and the block $j+2$, respectively. The encoding scheme in block $j \in [\![3, B]\!]$ is depicted in Fig. K.1.

Define

$$\begin{aligned}
&\Upsilon^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j}(m_j, t_{j-1}, k_{j-2}, s_j^r, \ell_j, a_j, \tilde{u}^r, \tilde{v}^r, z_j^r, k_{j-1}, t_j, k_j) \\
&\quad \triangleq 2^{-r(R+R_t+R_k)} Q_S^{\otimes r}(s_j^r) f(\ell_j, a_j | s_j^r, m_j, t_{j-1}, k_{j-2}) \mathbb{1}_{\{\tilde{u}^r = u^r(m_j, t_{j-1}, k_{j-2}, \ell_j)\}} \mathbb{1}_{\{\tilde{v}^r = v^r(a_j)\}} \\
&\qquad \times W_{Z|U,S}^{\otimes r}(z_j^r | \tilde{u}^r, s_j^r) 2^{-rR_k} \mathbb{1}_{\{t_j = \sigma(\tilde{v}^r)\}} \mathbb{1}_{\{k_j = \Phi(\tilde{v}^r)\}}. \qquad \text{(K.4)}
\end{aligned}$$

For a given codebook $\mathcal{C}_r$, the induced joint distribution over the codebook (i.e. $P^{(\mathcal{C}_r)}$) satisfies

$$\mathbb{D}\Big(P^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j} || \Upsilon^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j}\Big) \leq \epsilon. \quad \text{(K.5)}$$

This intermediate distribution $\Upsilon^{(\mathcal{C}_r)}$ approximates the true distribution $P^{(\mathcal{C}_r)}$ and will be used in the sequel for bounding purposes. Expression (K.5) holds because the main difference between $P^{(\mathcal{C}_r)}$ and $\Upsilon^{(\mathcal{C}_r)}$ is that the keys K_{j-2}, K_{j-1} and the reconciliation index T_{j-1} are assumed to be uniformly distributed in $\Upsilon^{(\mathcal{C}_r)}$, which are made (arbitrarily) nearly uniform in $P^{(\mathcal{C}_r)}$ with appropriate control of rate as in (K.14) and (K.20).

Covert Analysis: We now show $\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n})] \underset{n\to\infty}{\longrightarrow} 0$, where C_n is the set of all the codebooks for all blocks and,

$$Q_Z(\cdot) = \sum_{u\in\mathcal{U}}\sum_{v\in\mathcal{V}}\sum_{s\in\mathcal{S}}\sum_{x\in\mathcal{X}} P_U(u)P_V(v)P_{S|U,V}(s|u,v)\mathbb{1}_{\left\{X=X(u,s)\right\}}W_{Z|X,S}(\cdot|x,s), \tag{K.6}$$

such that $\sum_{u\in\mathcal{U}}\sum_{v\in\mathcal{V}} P_U(u)P_V(v)P_{S|U,V}(\cdot|u,v) = Q_S(\cdot)$. Then we choose P_U, P_V, $P_{S|U,V}$, and $x(u,s)$ such that it satisfies $Q_Z = Q_0$. For every $j \in [\![2, B]\!]$,

$$\begin{aligned}\mathbb{I}(Z_j^r; Z_{j+1}^{B,r}) &\leq \mathbb{I}(Z_j^r; K_{j-1}, T_j, K_j, Z_{j+1}^{B,r}) \\ &\overset{(a)}{=} \mathbb{I}(Z_j^r; K_{j-1}, T_j, K_j),\end{aligned} \tag{K.7}$$

where (a) holds because $Z_j^r - (K_{j-1}, T_j, K_j) - Z_{j+1}^{B,r}$ forms a Markov chain, as seen in the functional dependence graph depicted in Fig. K.2. Also,

$$\begin{aligned}\mathbb{I}(Z_j^r; K_{j-1}, T_j, K_j) &= \mathbb{D}(P^{(\mathcal{C}_r)}_{Z_j^r, K_{j-1}, T_j, K_j}||P_{Z_j^r}P_{K_{j-1},T_j,K_j}) \\ &\overset{(b)}{\leq} \mathbb{D}(P^{(\mathcal{C}_r)}_{Z_j^r, K_{j-1}, T_j, K_j}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}),\end{aligned} \tag{K.8}$$

where $Q_{K_{j-1}}Q_{K_j}Q_{T_j}$ is the uniform distribution over $[\![1, 2^{rR_k}]\!] \times [\![1, 2^{rR_K}]\!] \times [\![1, 2^{rR_T}]\!]$ and (b) follows from

$$\begin{aligned}\mathbb{D}(P^{(\mathcal{C}_r)}_{Z_j^r, K_{j-1}, T_j, K_j}||P^{(\mathcal{C}_r)}_{Z_j^r}P^{(\mathcal{C}_r)}_{K_{j-1},T_j,K_j}) &= \mathbb{D}(P^{(\mathcal{C}_r)}_{Z_j^r, K_{j-1}, T_j, K_j}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}) \\ &\quad - \mathbb{D}(P^{(\mathcal{C}_r)}_{Z_j^r}||Q_Z^{\otimes r}) - \mathbb{D}(P^{(\mathcal{C}_r)}_{K_{j-1},T_j,K_j}||Q_{K_{j-1}}Q_{T_j}Q_{K_j}).\end{aligned} \tag{K.9}$$

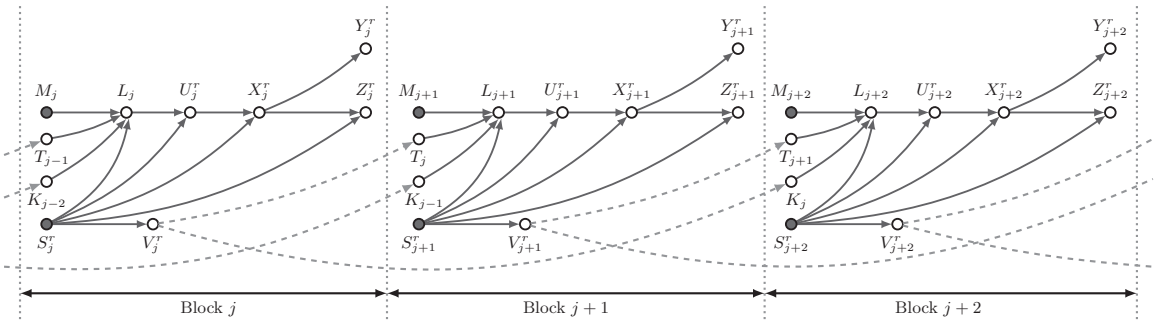

Figure K.2. Functional dependence graph for the block-Markov encoding scheme

Therefore, from the expansion in (I.2), by substituting Q_0 with Q_Z, and also from (K.8) and (K.9),

$$\mathbb{D}(P_{Z^n}^{(\mathcal{C}_r)}||Q_Z^{\otimes n}) \leq 2\sum_{j=1}^{B}\mathbb{D}(P_{Z_j^r,K_{j-1},T_j,K_j}^{(\mathcal{C}_r)}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}). \tag{K.10}$$

To bound the RHS of (K.10) by using Lemma 1 and the triangle inequality we have,

$$\begin{aligned}
\mathbb{E}_{C_r}||P_{Z_j^r,K_{j-1},T_j,K_j|C_r} - Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}||_1 &\leq \mathbb{E}_{C_r}||P_{Z_j^r,K_{j-1},T_j,K_j|C_r} - \Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||_1 \\
&+ \mathbb{E}_{C_r}||\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r} - Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}||_1 \\
\leq \mathbb{E}_{C_r}||P_{Z_j^r,K_{j-1},T_j,K_j|C_r} - \Upsilon_{Z_j^r,K_{j-1},T_j,K_j|C_r}||_1 &+ \mathbb{E}_{C_r}||\Upsilon_{Z_j^r,K_{j-1},T_j,K_j|C_r} - \Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||_1 \\
&+ \mathbb{E}_{C_r}||\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r} - Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}||_1.
\end{aligned} \tag{K.11}$$

From (K.5) and the monotonicity of KL-divergence the first term on the RHS of (K.11) vanishes when r grows. To bound the second term on the RHS of (K.11) for a fixed codebook $\mathcal{C}_r$, we have,

$$\Gamma_{M_j,T_{j-1},K_{j-2}}^{(\mathcal{C}_r)} = 2^{-r(R+R_t+R_k)} = \Upsilon_{M_j,T_{j-1},K_{j-2}}^{(\mathcal{C}_r)}, \tag{K.12a}$$

$$\Gamma_{L_j,A_j|M_j,T_{j-1},K_{j-2},S_j^r}^{(\mathcal{C}_r)} = f(\ell_j,a_j|s_j^r,m_j,t_{j-1},k_{j-2}) = \Upsilon_{L_j,A_j|M_j,T_{j-1},K_{j-2},S_j^r}^{(\mathcal{C}_r)}, \tag{K.12b}$$

$$\Gamma_{U^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j}^{(\mathcal{C}_r)} = \mathbb{1}_{\{\tilde{u}^r=u^r(m_j,t_{j-1},k_{j-2},\ell_j)\}} = \Upsilon_{U^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j}^{(\mathcal{C}_r)}, \tag{K.12c}$$

$$\Gamma^{(\mathcal{C}_r)}_{V^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r} = \mathbb{1}_{\{\tilde{v}^r=v^r(a_j)\}} = \Upsilon^{(\mathcal{C}_r)}_{V^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r}, \tag{K.12d}$$

$$\Gamma^{(\mathcal{C}_r)}_{Z_j^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r} = W^{\otimes r}_{Z|U,S} = \Upsilon^{(\mathcal{C}_r)}_{Z_j^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r}, \tag{K.12e}$$

$$\Gamma^{(\mathcal{C}_r)}_{K_{j-1}|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r} = 2^{-rR_k} = \Upsilon^{(\mathcal{C}_r)}_{K_{j-1}|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r}, \tag{K.12f}$$

$$\Gamma^{(\mathcal{C}_r)}_{T_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1}} = \mathbb{1}_{\{t_j=\sigma(v^r)\}} = \Upsilon^{(\mathcal{C}_r)}_{T_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1}}, \tag{K.12g}$$

$$\Gamma^{(\mathcal{C}_r)}_{K_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j} = \mathbb{1}_{\{k_j=\Phi(v^r)\}} = \Upsilon^{(\mathcal{C}_r)}_{K_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j}, \tag{K.12h}$$

where (K.12b) follows from (K.3). Hence,

$$\begin{aligned}
&\mathbb{E}_{C_r}||\Upsilon_{Z_j^r,K_{j-1},T_j,K_j|C_r} - \Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||_1 \\
&\leq \mathbb{E}_{C_r}||\Upsilon_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j|C_r} - \Gamma_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j|C_r}||_1 \\
&\overset{(a)}{=} \mathbb{E}_{C_r}||\Upsilon_{M_j,T_{j-1},K_{j-2},S_j^r|C_r} - \Gamma_{M_j,T_{j-1},K_{j-2},S_j^r|C_r}||_1 \\
&\overset{(b)}{=} \mathbb{E}_{C_r}||Q_S^{\otimes r} - \Gamma_{S_j^r|M_j=1,T_{j-1}=1,K_{j-2}=1,C_r}||_1,
\end{aligned} \tag{K.13}$$

where (a) follows from (K.12b)-(K.12h) and (b) follows from the symmetry of the codebook construction with respect to M_j, T_{j-1}, and K_{j-2} and (K.12a). Based on [90] or [23, Theorem 2] the RHS of (K.13) vanishes if

$$R' > \mathbb{I}(U;S), \tag{K.14a}$$

$$\tilde{R} > \mathbb{I}(V;S), \tag{K.14b}$$

$$R' + \tilde{R} > \mathbb{I}(U,V;S). \tag{K.14c}$$

We now proceed to bound the third term on the RHS of (K.11). First, consider the following marginal from (K.1),

$$\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}(z_j^r,k_{j-1},t_j,k_j) = \sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j}\sum_{s_j^r}\frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R})}}$$

$$\times P_{S|U,V}^{\otimes r}\big(s_j^r|U^r(m_j,t_{j-1},k_{j-2},\ell_j),V^r(a_j)\big)W_{Z|U,S}^{\otimes r}\big(z_j^r|U^r(m_j,t_{j-1},k_{j-2},\ell_j),s_j^r\big)$$

$$\times \mathbb{1}_{\{t_j=\sigma(V^r(a_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(a_j))\}} \tag{K.15}$$

$$= \sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j}\frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R})}}W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(m_j,t_{j-1},k_{j-2},\ell_j),V^r(a_j)\big)$$

$$\times \mathbb{1}_{\{t_j=\sigma(V^r(a_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(a_j))\}}, \tag{K.16}$$

where $W_{Z|U,V}(z|u,v) = \sum_{s\in\mathcal{S}}P_{S|U,V}(s|u,v)W_{Z|U,S}(z|u,s)$. To bound the third term on the RHS of (K.11) by using Pinsker's inequality, it is sufficient to bound $\mathbb{E}_{C_r}[\mathbb{D}(\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j})]$ as follows,

$$\mathbb{E}_{C_r}[\mathbb{D}(\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j})]$$

$$= \mathbb{E}_{C_r}\Big[\sum_{(z_j^r,k_{j-1},t_j,k_j)}\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}(z_j^r,k_{j-1},t_j,k_j)\log\Big(\frac{\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}(z_j^r,k_{j-1},t_j,k_j)}{Q_Z^{\otimes r}(z_j^r)Q_{K_{j-1}}(k_{j-1})Q_{T_j}(t_j)Q_{K_j}(k_j)}\Big)\Big]$$

$$= \mathbb{E}_{C_r}\Bigg[\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j}\frac{W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(m_j,t_{j-1},k_{j-2},\ell_j),V^r(a_j)\big)}{2^{r(R+R_t+2R_k+R'+\tilde{R})}}$$

$$\times \mathbb{1}_{\{t_j=\sigma(V^r(a_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(a_j))\}}$$

$$\times \log\Bigg(\frac{\sum_{\tilde{m}_j}\sum_{\tilde{t}_{j-1}}\sum_{\tilde{k}_{j-2}}\sum_{\tilde{\ell}_j}\sum_{\tilde{a}_j}W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2},\tilde{\ell}_j),V^r(\tilde{a}_j)\big)\mathbb{1}_{\{t_j=\sigma(V^r(\tilde{a}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{a}_j))\}}}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\Bigg)\Bigg]$$

$$\overset{(a)}{\leq} \sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j}\frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R})}}$$

$$\times \sum_{(u^r,v^r)}\Gamma_{U^r,V^r,Z^r}^{\otimes r}\big(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r\big)$$

$$\times \mathbb{E}_{\sigma(v^r(a_j))}\big[\mathbb{1}_{\{t_j=\sigma(v^r(a_j))\}}\big]\times\mathbb{E}_{\Phi(v^r(a_j))}\big[\mathbb{1}_{\{k_j=\Phi(v^r(a_j))\}}\big]$$

$$\times \log\mathbb{E}_{\substack{\setminus(m_j,t_{j-1},k_{j-2},\ell_j,a_j),\\ \setminus(\sigma(v^r(a_j)),\Phi(v^r(a_j)))}}\Bigg[\frac{1}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}$$

$$\times \sum_{\tilde{m}_j}\sum_{\tilde{t}_{j-1}}\sum_{\tilde{k}_{j-2}}\sum_{\tilde{\ell}_j}\sum_{\tilde{a}_j}W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(,\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2},\tilde{\ell}_j),V^r(\tilde{a}_j)\big)\mathbb{1}_{\{t_j=\sigma(V^r(\tilde{a}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{a}_j))\}}\Bigg]$$

$$\overset{(b)}{\leq} \sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j}\frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R}+R_T+R_K)}}$$

$$
\begin{aligned}
&\times \sum_{(u^r,v^r)} \Gamma^{\otimes r}_{U^r,V^r,Z^r}\big(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r\big)\\
&\times \log \frac{1}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\Bigg(W^{\otimes r}_{Z|U,V}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j)\big)\\
&+\mathbb{E}_{\setminus(m_j,t_{j-1},k_{j-2},\ell_j)}\Bigg[\sum_{(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2},\tilde{\ell}_j)\neq(m_j,t_{j-1},k_{j-2},\ell_j)} W^{\otimes r}_{Z|U,V}\big(z_j^r|U^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2},\tilde{\ell}_j),v^r(a_j)\big)\Bigg]\\
&+\mathbb{E}_{\setminus(a_j,\sigma(v^r(a_j)),\Phi(v^r(a_j)))}\Bigg[\sum_{\tilde{a}_j\neq a_j} W^{\otimes r}_{Z|U,V}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j),V^r(\tilde{a}_j)\big)\\
&\times \mathbb{1}_{\{t_j=\sigma(V^r(\tilde{a}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{a}_j))\}}\Bigg]\\
&+\mathbb{E}_{\substack{\setminus(m_j,t_{j-1},k_{j-2},\ell_j,a_j),\\ \setminus(\sigma(v^r(a_j)),\Phi(v^r(a_j)))}}\Bigg[\sum_{(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2},\tilde{\ell}_j)\neq(m_j,t_{j-1},k_{j-2},\ell_j)}\sum_{\tilde{a}_j\neq a_j} W^{\otimes r}_{Z|U,V}\big(z_j^r|U^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2},\tilde{\ell}_j),V^r(\tilde{a}_j)\big)\\
&\times \mathbb{1}_{\{t_j=\sigma(V^r(\tilde{a}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{a}_j))\}}\Bigg]\Bigg)\\
&\overset{(c)}{\leq} \sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j}\frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R}+R_T+R_K)}}\\
&\times \sum_{(u^r,v^r)} \Gamma^{\otimes r}_{U^r,V^r,Z^r}\big(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r\big)\\
&\times \log\Bigg(\frac{W^{\otimes r}_{Z|U,V}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j)\big)}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&+\sum_{(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2},\tilde{\ell}_j)\neq(m_j,t_{j-1},k_{j-2},\ell_j)}\frac{W^{\otimes r}_{Z|V}\big(z_j^r|v^r(a_j)\big)}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&+\sum_{\tilde{a}_j\neq a_j}\frac{W^{\otimes r}_{Z|U}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j)\big)}{2^{r(R+R_t+R_k+R'+\tilde{R})}Q_Z^{\otimes r}(z_j^r)}+1\Bigg)\\
&\leq \sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j}\frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R}+R_T+R_K)}}\\
&\times \sum_{(u^r,v^r)} \Gamma^{\otimes r}_{U^r,V^r,Z^r}\big(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r\big)\\
&\times \log\Bigg(\frac{W^{\otimes r}_{Z|U,V}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j)\big)}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}
\end{aligned}
$$

$$+\frac{W_{Z|V}^{\otimes r}\big(z_j^r|v^r(a_j)\big)}{2^{r(\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|U}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j)\big)}{2^{r(R+R_t+R_k+R')}Q_Z^{\otimes r}(z_j^r)}+1\Bigg)$$

$$\triangleq \Psi_1+\Psi_2, \tag{K.17}$$

where (a) follows from Jensen's inequality, (b) and (c) hold because $\mathbb{1}_{\{.\}}\leq 1$. We define Ψ_1 and Ψ_2 as

$$\begin{aligned}
\Psi_1 &= \sum_{(k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j}\frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R}+R_T+R_K)}}\\
&\times \sum_{(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r)\in\mathcal{T}_\epsilon^{(n)}}\Gamma_{U^r,V^r,Z^r}^{\otimes r}\big(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r\big)\\
&\times\log\Bigg(\frac{W_{Z|U,V}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j)\big)}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|V}^{\otimes r}\big(z_j^r|v^r(a_j)\big)}{2^{r(\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&+\frac{W_{Z|U}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j)\big)}{2^{r(R+R_t+R_k+R')}Q_Z^{\otimes r}(z_j^r)}+1\Bigg)\\
&\leq\log\Bigg(\frac{2^{r(R_T+R_K)}\times 2^{-r(1-\epsilon)\mathbb{H}(Z|U,V)}}{2^{r(R+R_t+R_k+R'+\tilde{R})}\times 2^{-r(1+\epsilon)\mathbb{H}(Z)}}+\frac{2^{r(R_T+R_K)}\times 2^{-r(1-\epsilon)\mathbb{H}(Z|V)}}{2^{r\tilde{R}}\times 2^{-r(1+\epsilon)\mathbb{H}(Z)}}\\
&+\frac{2^{-r(1-\epsilon)\mathbb{H}(Z|U)}}{2^{r(R+R_t+R_k+R')}\times 2^{-r(1+\epsilon)\mathbb{H}(Z)}}+1\Bigg)
\end{aligned}\tag{K.18}$$

$$\begin{aligned}
\Psi_2 &= \sum_{(k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j}\frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R}+R_T+R_K)}}\\
&\times \sum_{(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r)\notin\mathcal{T}_\epsilon^{(n)}}\Gamma_{U^r,V^r,Z^r}^{\otimes r}\big(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r\big)\\
&\times\log\Bigg(\frac{W_{Z|U,V}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j)\big)}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|V}^{\otimes r}\big(z_j^r|v^r(a_j)\big)}{2^{r(\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&+\frac{W_{Z|U}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j)\big)}{2^{r(R+R_t+R_k+R')}Q_Z^{\otimes r}(z_j^r)}+1\Bigg)\\
&\leq 2|V||U||Z|e^{-r\epsilon^2\mu_{S,V,U,Z}}r\log\Big(\frac{3}{\mu_Z}+1\Big).
\end{aligned}\tag{K.19}$$

In (K.19) $\mu_{V,U,Z} = \min_{(v,u,z)\in(\mathcal{V},\mathcal{U},\mathcal{Z})} P_{V,U,Z}(v,u,z)$ and $\mu_Z = \min_{z\in\mathcal{Z}} P_Z(z)$. When $r \to \infty$ then $\Psi_2 \to 0$ and Ψ_1 goes to zero when r grows if

$$R + R_t + R_k + R' + \tilde{R} - R_T - R_K > \mathbb{I}(U,V;Z), \tag{K.20a}$$

$$\tilde{R} - R_T - R_K > \mathbb{I}(V;Z), \tag{K.20b}$$

$$R + R_t + R_k + R' > \mathbb{I}(U;Z). \tag{K.20c}$$

Decoding and Error Probability Analysis: At the end of the block $j \in [\![1, B]\!]$, using its knowledge of the key k_{j-2} generated from the block $j-2$, the receiver finds a unique triple $(\hat{m}_j, \hat{t}_{j-1}, \hat{\ell}_j)$ such that $\big(u^r(\hat{m}_j, \hat{t}_{j-1}, k_{j-2}, \hat{\ell}_j), y_j^r\big) \in \mathcal{T}_\epsilon^{(r)}$. To bound the probability of error at the encoder and the decoder, we use the following lemma.

Lemma 5 (Typical With High Probability)**.** *If* $(R', \tilde{R}) \in \mathbb{R}_+^2$ *satisfies* (K.14)*, then for any* $(m_j, t_{j-1}, k_{j-2}) \in (\mathcal{M}, \mathcal{T}, \mathcal{K})$ *and* $\epsilon > 0$*, we have*

$$\mathbb{E}_{C_r}\mathbb{P}_P\Big(\big(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j), S_j^r\big) \notin \mathcal{T}_\epsilon^{(r)} | C_r\Big) \underset{r\to\infty}{\longrightarrow} 0, \tag{K.21}$$

where P *is the induced distribution over the codebook defined in* (K.5)*.*

The proof of Lemma 5 is given in Appendix L.

To analyze the probability of error, we define the following error events for $j \in [\![1, B]\!]$,

$$\mathcal{E} \triangleq \big\{M \neq \hat{M}\big\}, \tag{K.22a}$$

$$\mathcal{E}_j \triangleq \big\{M_j \neq \hat{M}_j\big\}, \tag{K.22b}$$

$$\mathcal{E}_{1,j} \triangleq \big\{\big(U^r(M_j, T_{j-1}, K_{j-2}, L_j), S_j^r\big) \notin \mathcal{T}_{\epsilon_1}^{(r)}(U,S)\big\}, \tag{K.22c}$$

$$\mathcal{E}_{2,j} \triangleq \big\{\big(U^r(M_j, T_{j-1}, K_{j-2}, L_j), Y_j^r\big) \notin \mathcal{T}_{\epsilon_2}^{(r)}(U,Y)\big\}, \tag{K.22d}$$

$$\mathcal{E}_{3,j} \triangleq \big\{\big(U^r(M_j, T_{j-1}, K_{j-2}, L_j), Y_j^r\big) \in \mathcal{T}_{\epsilon_2}^{(r)}(U,Y) \text{ for some } m_j \neq M_j \text{ and } \ell_j \in [1 : 2^{rR'}]\big\}. \tag{K.22e}$$

where $\epsilon_2 > \epsilon_1 > \epsilon > 0$. The probability of error is upper bounded as follows,

$$\mathbb{P}(\mathcal{E}) = \mathbb{P}\Big\{ \bigcup_{j=1}^{B} \mathcal{E}_j \Big\} \leq \sum_{j=1}^{B} \mathbb{P}(\mathcal{E}_j). \tag{K.23}$$

Now we bound $\mathbb{P}(\mathcal{E}_j)$ by using union bound,

$$\mathbb{P}(\mathcal{E}_j) \leq \mathbb{P}(\mathcal{E}_{1,j}) + \mathbb{P}(\mathcal{E}_{1,j}^c \cap \mathcal{E}_{2,j}) + \mathbb{P}(\mathcal{E}_{2,j}^c \cap \mathcal{E}_{3,j}). \tag{K.24}$$

According to Lemma 5 the first term on the RHS of (K.24) vanishes when r grows, and by the law of large numbers the second term on the RHS of (K.24) vanishes when r grows. Also, according to the law of large numbers and the packing lemma, the last term on the RHS of (K.24) vanishes when r grows if [85],

$$R + R_t + R' \leq \mathbb{I}(U;Y). \tag{K.25}$$

We now analyze the probability of error at the encoder and the decoder for key generation. Let (A_{j-1}, T_{j-1}) denote the chosen indices at the encoder and $\hat{A}_{j-1}$ and $\hat{T}_{j-1}$ be the estimates of the indices A_{j-1} and T_{j-1} at the decoder. At the end of block j, by decoding U_j^r, the decoder knows $\hat{T}_{j-1}$. To find A_{j-1} we define the error event,

$$\mathcal{E}' = \Big\{ \Big(V_{j-1}^r(\hat{A}_{j-1}), S_{j-1}^r, U_{j-1}^r, Y_{j-1}^r \Big) \notin \mathcal{T}_\epsilon^{(r)} \Big\}. \tag{K.26}$$

Also, consider the error events,

$$\mathcal{E}_1' = \Big\{ \Big(V_{j-1}^r(a_{j-1}), S_{j-1}^r \Big) \notin \mathcal{T}_{\epsilon'}^{(r)} \text{ for all } a_{j-1} \in [\![1, 2^{r\tilde{R}}]\!] \Big\}, \tag{K.27a}$$

$$\mathcal{E}_2' = \Big\{ \Big(V_{j-1}^r(A_{j-1}), S_{j-1}^r, U_{j-1}^r, Y_{j-1}^r \Big) \notin \mathcal{T}_\epsilon^{(r)} \Big\}, \tag{K.27b}$$

$$\mathcal{E}_3' = \Big\{ \Big(V_{j-1}^r(\tilde{a}_{j-1}), U_{j-1}^r, Y_{j-1}^r \Big) \in \mathcal{T}_\epsilon^{(r)} \text{ for some } \tilde{a}_{j-1} \in \mathcal{B}(\hat{T}_{j-1}), \tilde{a}_{j-1} \neq A_{j-1} \Big\}, \tag{K.27c}$$

where $\epsilon > \epsilon' > 0$. By the union bound we have,

$$P(\mathcal{E}') \leq P(\mathcal{E}_1') + P(\mathcal{E}_1'^c \cap \mathcal{E}_2') + P(\mathcal{E}_3'). \tag{K.28}$$

According to Lemma 5 the first term on the RHS of (K.28) vanishes when r grows if we have (K.14). Following the steps in [85, Sec. 11.3.1], the last two terms on the RHS of (K.28) go to zero when r grows if,

$$\tilde{R} > \mathbb{I}(V;S), \tag{K.29a}$$

$$\tilde{R} - R_t < \mathbb{I}(V;U,Y). \tag{K.29b}$$

The region in Theorem 10 is derived by remarking that the scheme requires $R_K + R_T \geq R_k + R_t$ and applying Fourier-Motzkin to (K.14) and (K.20), (K.25), and (K.29).

APPENDIX L

PROOF OF LEMMA 5

For a fix $\epsilon > 0$, consider the PMF Γ defined in (K.1). For the random experiment described by Γ; since $U^r(m_j, t_{j-1}, k_{j-2}, L_j) \sim P_U^{\otimes r}$, for every $(m_j, t_{j-1}, k_{j-2}) \in \mathcal{M} \times \mathcal{T} \times \mathcal{K}$) and $V^r(A_j) \sim P_V^{\otimes r}$, for every $a_j \in \mathcal{A}$), and S_j^r is derived by passing $\big(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j)\big)$ through the DMC $P_{S|U,V}^{\otimes r}$ by the weak law of large numbers we have

$$\mathbb{E}_{C_r}\mathbb{P}_\Gamma\Big(\big(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j), S_j^r\big) \notin \mathcal{T}_\epsilon^{(r)}|C_r\Big) \underset{r\to\infty}{\longrightarrow} 0. \tag{L.1}$$

We also have

$$\begin{aligned}
&\mathbb{E}_{C_r}||P_{U^r,V^r,S_j^r|C_r} - \Gamma_{U^r,V^r,S_j^r|C_r}||_1 \\
&\leq \mathbb{E}_{C_r}||P_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j|C_r} \\
&\qquad - \Gamma_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j|C_r}||_1 \underset{r\to\infty}{\longrightarrow} 0,
\end{aligned} \tag{L.2}$$

where the RHS of (L.2) vanishes when r grows because of (K.13). We now define $g_r : \mathcal{U}^r \times \mathcal{V}^r \times \mathcal{S}_j^r \mapsto \mathbb{R}$ as $g_r(u^r, v^r, s_j^r) \triangleq \mathbb{1}_{\{(u^r,v^r,s_j^r)\notin\mathcal{T}_\epsilon^{(r)}\}}$. We now have,

$$\begin{aligned}
&\mathbb{E}_{C_r}\mathbb{P}_P\Big(\big(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j), S_j^r\big) \notin \mathcal{T}_\epsilon^{(r)}|C_r\Big) \\
&= \mathbb{E}_{C_r}\mathbb{E}_P\Big[g_r(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j), S_j^r)|C_r\Big] \\
&\leq \mathbb{E}_{C_r}\mathbb{E}_\Gamma\Big[g_r(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j), S_j^r)|C_r\Big] \\
&+ \mathbb{E}_{C_r}\Big|\mathbb{E}_P\Big[g_r(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j), S_j^r)|C_r\Big] \\
&- \mathbb{E}_\Gamma\Big[g_r(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j), S_j^r)|C_r\Big]\Big| \\
&\overset{(a)}{\leq} \mathbb{E}_{C_r}\mathbb{E}_\Gamma\Big[g_r(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j), S_j^r)|C_r\Big] + \mathbb{E}_{C_r}||P_{U^r,V^r,S_j^r|C_r} - \Gamma_{U^r,V^r,S_j^r|C_r}||_1,
\end{aligned} \tag{L.3}$$

where (a) follows from [91, Property 1] for g_r being bounded by 1. From (L.1) and (L.2) the RHS of (L.3) vanishes when r grows.

APPENDIX M

PROOF OF THEOREM 11

Fix $P_{U|S}(u|s)$, $x(u,s)$, and $\epsilon_1 > \epsilon_2 > 0$ such that, $P_Z = Q_0$.

Codebook Generation: Let $C_n \triangleq \{U^n(m,\ell)\}_{(m,\ell)\in\mathcal{M}\times\mathcal{L}}$, where $\mathcal{M} \triangleq [\![1, 2^{nR}]\!]$ and $\mathcal{L} \triangleq [\![1, 2^{nR'}]\!]$, be a random codebook consisting of independent random sequences each generated according to $\prod_{i=1}^n P_U(u_i)$. We denote a realization of C_n by $\mathcal{C}_n \triangleq \{u^n(m,\ell)\}_{(m,\ell)\in\mathcal{M}\times\mathcal{L}}$. We define an ideal PMF for codebook $\mathcal{C}_n$, as an approximate distribution to facilitate the analysis

$$\Gamma^{(\mathcal{C}_n)}_{M,L,U^n,S^n,Z^n}(m,\ell,\tilde{u}^n,s^n,z^n) = 2^{-n(R+R')}\mathbb{1}_{\{\tilde{u}^n=u^n(m,\ell)\}}P^{\otimes n}_{S|U}(s^n|\tilde{u}^n)W^{\otimes n}_{Z|U,S}(z^n|\tilde{u}^n,s^n), \quad \text{(M.1)}$$

where $W_{Z|U,S}$ is the marginal distribution $W_{Z|U,S} = \sum_{x\in\mathcal{X}} \mathbb{1}_{\{x=x(u,s)\}}W_{Z|X,S}$ and $P_{S|U}$ is defined as follows,

$$P_{S|U}(s|u) \triangleq \frac{P_{S,U}(s,u)}{P_U(u)} = \frac{Q_S(s)P_{U|S}(u|s)}{\sum_{s\in\mathcal{S}} Q_S(s)P_{U|S}(u|s)}. \quad \text{(M.2)}$$

Encoding: To send the message m the encoder generates ℓ according to

$$f(\ell|s^n,m) = \frac{P^{\otimes n}_{S|U}\big(s^n|u^n(m,\ell)\big)}{\sum\limits_{\ell'\in[\![1,2^{nR'}]\!]} P^{\otimes n}_{S|U}\big(s^n_j|u^n(m,\ell')\big)}, \quad \text{(M.3)}$$

where $P_{S|U}$ is defined in (M.2). Based on (m,ℓ), the encoder computes $u^n(m,\ell)$ and transmits codeword x^n, where $x_i = x(u_i(m,\ell),s_i)$. For a fixed codebook $\mathcal{C}_n$, the induced joint distribution over the codebook is as follows

$$P^{(\mathcal{C}_n)}_{M,S^n,L,U^n,Z^n}(m,s^n,\ell,\tilde{u}^n,z^n) = 2^{-nR}Q^{\otimes n}_S(s^n)f(\ell|s^n,m)\mathbb{1}_{\{\tilde{u}^n=u^n(m,\ell)\}}W^{\otimes n}_{Z|U,S}(z^n|\tilde{u}^n,s^n). \quad \text{(M.4)}$$

Covert Analysis: We now show that this coding scheme guarantees that

$$\mathbb{E}_{C_n}\big[\mathbb{D}(P_{Z^n|C_n}||Q^{\otimes n}_Z)\big] \underset{n\to\infty}{\longrightarrow} 0, \quad \text{(M.5)}$$

where

$$Q_Z(\cdot) = \sum_{s\in\mathcal{S}} \sum_{u\in\mathcal{U}} \sum_{x\in\mathcal{X}} Q_S(s) P_{U|S}(u|s) \mathbb{1}_{\left\{X=X(U,S)\right\}} W_{Z|X,S}(\cdot|x,s). \tag{M.6}$$

Then we choose $P_{U|S}$ and $x(u,s)$ such that it satisfies $Q_Z = Q_0$. By Combining Lemma 1 and the triangle inequality a sufficient condition for (M.5) is to show that the RHS of the following inequality vanishes when n grows,

$$\mathbb{E}_{C_n} ||P_{Z^n|C_n} - Q_Z^{\otimes n}||_1 \leq \mathbb{E}_{C_n} ||P_{Z^n|C_n} - \Gamma_{Z^n|C_n}||_1 + \mathbb{E}_{C_n} ||\Gamma_{Z^n|C_n} - Q_Z^{\otimes n}||_1. \tag{M.7}$$

By [77, Corollary VII.5] the second term on the RHS of (M.7) vanishes when n grows if

$$R + R' > \mathbb{I}(U;Z). \tag{M.8}$$

To bound the first term on the RHS of (M.7) we have,

$$\Gamma_M^{(\mathcal{C}_n)} = 2^{-nR} = P_M^{(\mathcal{C}_n)}, \tag{M.9a}$$

$$\Gamma_{L|M,S^n}^{(\mathcal{C}_n)} = f(\ell|s^n, m) = P_{L|M,S^n}^{(\mathcal{C}_n)}, \tag{M.9b}$$

$$\Gamma_{U^n|M,S^n,L}^{(\mathcal{C}_n)} = \mathbb{1}_{\{\tilde{u}^n = u^n(m,\ell)\}} = P_{U^n|M,S^n,L}^{(\mathcal{C}_n)}, \tag{M.9c}$$

$$\Gamma_{Z^n|M,S^n,L,U^n}^{(\mathcal{C}_n)} = W_{Z|U,S}^{\otimes n} = P_{Z^n|M,S^n,L,U^n}^{(\mathcal{C}_n)}, \tag{M.9d}$$

where (M.9b) follows from (M.3). Hence,

$$\begin{aligned}
\mathbb{E}_{C_n} ||P_{Z^n|C_n} - \Gamma_{Z^n|C_n}||_1 &\leq \mathbb{E}_{C_n} ||P_{M,S^n,L,U^n,Z^n|C_n} - \Gamma_{M,S^n,L,U^n,Z^n|C_n}||_1 \\
&\overset{(a)}{=} \mathbb{E}_{C_n} ||P_{S^n,L,U^n,Z^n|M=1,C_n} - \Gamma_{S^n,L,U^n,Z^n|M=1,C_n}||_1 \\
&\overset{(b)}{=} \mathbb{E}_{C_n} ||Q_S^{\otimes n} - \Gamma_{S^n|M=1,C_n}||_1,
\end{aligned} \tag{M.10}$$

where (a) follows from (M.9b)-(M.9d) and (b) follows from the symmetry of the codebook construction with respect to M and (M.9a). Based on the soft covering lemma [77, Corollary VII.5] the RHS of (M.10) vanishes if

$$R' > \mathbb{I}(U;S). \tag{M.11}$$

Decoding and Error Probability Analysis: To decode the message m, the receiver finds a unique pair $(\hat{m}, \hat{\ell})$ such that $\left(u^n(\hat{m}, \hat{\ell}), y^n\right) \in \mathcal{T}_\epsilon^{(n)}$. To analyze the probability of error, we define the following error events,

$$\mathcal{E} \triangleq \left\{M \neq \hat{M}\right\}, \tag{M.12a}$$

$$\mathcal{E}_1 \triangleq \left\{\left(U^n(M, L), S^n\right) \notin \mathcal{T}_{\epsilon_1}^{(n)}(U, S)\right\}, \tag{M.12b}$$

$$\mathcal{E}_2 \triangleq \left\{\left(U^n(M, L), Y^n\right) \notin \mathcal{T}_{\epsilon_2}^{(n)}(U, Y)\right\}, \tag{M.12c}$$

$$\mathcal{E}_3 \triangleq \left\{\left(U^n(M, L), Y^n\right) \in \mathcal{T}_{\epsilon_2}^{(n)}(U, Y) \text{ for some } m \neq M \text{ and } \ell \in [1 : 2^{nR'}]\right\}, \tag{M.12d}$$

where $\epsilon_2 > \epsilon_1 > 0$. Now we bound $\mathbb{P}(\mathcal{E})$ by using union bound,

$$\mathbb{P}(\mathcal{E}) \leq \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_1^c \cap \mathcal{E}_2) + \mathbb{P}(\mathcal{E}_2^c \cap \mathcal{E}_3). \tag{M.13}$$

Similar to Lemma 5 one can show that the first term on the RHS of (M.13) vanishes when n grows, and by the law of large numbers the second term on the RHS of (M.13) vanishes when n grows. Also, according to the law of large numbers and the packing lemma, the last term on the RHS of (M.13) vanishes when n grows if [85],

$$R + R' < \mathbb{I}(U; Y). \tag{M.14}$$

The region in Theorem 11 is derived by applying Fourier-Motzkin to (M.8), (M.11), and (M.14).

APPENDIX N

PROOF OF THEOREM 12

Consider any sequence of length-n codes for a state-dependent channel with CSI available non-causally only at the transmitter such that $P_e^{(n)} \leq \epsilon_n$, $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \delta$, and $R_K/n \leq \lambda_n$ with $\lim_{n\to\infty} \epsilon_n = \lim_{n\to\infty} \lambda_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish. The following lemma, a version of which with variational distance can be found in [77, Lemma VI.3], will prove useful.

Lemma 6. *If $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \delta$, then $\sum_{i=1}^n \mathbb{I}(Z_i; Z^{i-1}) \leq \delta$ and $\sum_{i=1}^n \mathbb{I}(Z_i; Z_{i+1}^n) \leq \delta$. In addition, if $T \in [\![1, n]\!]$ is an independent variable uniformly distributed, then $\mathbb{I}(T; Z_T) \leq \nu$, where $\nu \triangleq \frac{\delta}{n}$.*

Note that Lemma 6 is slightly different from [77, Lemma VI.3], as the upper bounds are tighter and do not include a factor of n. This is a consequence of using a constraint based on relative entropy instead of total variation. This is crucial in what follows, as we do no necessarily require $\delta \to 0$.

Proof. First note that,

$$\begin{aligned}
\sum_{i=1}^n \mathbb{I}(Z_i; Z^{i-1}) &= \sum_{i=1}^n [\mathbb{H}(Z_i) - \mathbb{H}(Z_i|Z^{i-1})] \\
&= \sum_{i=1}^n \mathbb{H}(Z_i) - \mathbb{H}(Z^n) \\
&= -\sum_{i=1}^n \sum_z P_{Z_i}(z) \log P_{Z_i}(z) + \sum_{z^n} P_{Z^n}(z^n) \log P_{Z^n}(z^n) \\
&= -\sum_{i=1}^n \sum_z P_{Z_i}(z) \log P_{Z_i}(z) + \sum_{i=1}^n \sum_z P_{Z_i}(z) \log Q_0(z) \\
&\quad - \sum_{i=1}^n \sum_z P_{Z_i}(z) \log Q_0(z) + \sum_{z^n} P_{Z^n}(z^n) \log P_{Z^n}(z^n) \\
&= -\sum_{i=1}^n \mathbb{D}(P_{Z_i}||Q_0) - \sum_{z^n} P_{Z^n}(z^n) \log Q_0^{\otimes n}(z^n) + \sum_{z^n} P_{Z^n}(z^n) \log P_{Z^n}(z^n)
\end{aligned}$$

$$\leq \mathbb{D}(P_{Z^n}||Q_0^{\otimes n})$$

$$\leq \delta.$$

Similarly, one can prove $\sum_{i=1}^{n} \mathbb{I}(Z_i; Z_{i+1}^n) \leq \delta$. Next,

$$\begin{aligned}
\mathbb{I}(T; Z_T) &= \mathbb{H}(Z_T) - \mathbb{H}(Z_T|T) \\
&= -\sum_z \frac{1}{n}\sum_{i=1}^{n} P_{Z_i}(z) \log \frac{1}{n}\sum_{j=1}^{n} P_{Z_j}(z) + \frac{1}{n}\sum_{i=1}^{n}\sum_z P_{Z_i}(z)\log P_{Z_i}(z) \\
&= -\sum_z \frac{1}{n}\sum_{i=1}^{n} P_{Z_i}(z) \log \frac{1}{n}\sum_{j=1}^{n} P_{Z_j}(z) + \sum_z \frac{1}{n}\sum_{i=1}^{n} P_{Z_i}(z)\log Q_0(z) \\
&\quad - \sum_z \frac{1}{n}\sum_{i=1}^{n} P_{Z_i}(z) \log Q_0(z) + \frac{1}{n}\sum_{i=1}^{n}\sum_z P_{Z_i}(z)\log P_{Z_i}(z) \\
&= -\mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} P_{Z_i}\middle|\middle|Q_0\right) + \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \\
&\leq \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \\
&\leq \frac{\delta}{n}.
\end{aligned} \tag{N.1}$$

□

Epsilon Rate Region: We first define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ that expands the region defined in (3.25) as follows,

$$\mathcal{A}_\epsilon \triangleq \big\{R \geq 0 : \exists P_{U,V,S,X,Y,Z} \in \mathcal{D}_\epsilon : R \leq \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\} + \epsilon\big\}, \tag{N.2a}$$

where

$$\mathcal{D}_\epsilon = \left\{ \begin{array}{l} P_{U,V,S,X,Y,Z} : \\ P_{U,V,S,X,Y,Z} = Q_S P_{UV|S} \mathbb{1}_{\left\{X=X(U,S)\right\}} W_{Y,Z|X,S} \\ \mathbb{D}\left(P_Z \| Q_0\right) \leq \epsilon \\ \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\} \geq \mathbb{I}(V;Z) - \mathbb{I}(V;S) - 4\epsilon \\ \max\{|\mathcal{U}|, |\mathcal{V}|\} \leq |\mathcal{X}| + 3 \end{array} \right\}. \tag{N.2b}$$

We next show that if a rate R is achievable then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$ and $\nu > 0$, we start by upper bounding nR using standard techniques,

$$\begin{aligned} nR &= \mathbb{H}(M) \\ &\overset{(a)}{\leq} \mathbb{I}(M; Y^n) + n\epsilon_n \\ &= \sum_{i=1}^n \mathbb{I}(M; Y_i | Y^{i-1}) + n\epsilon_n \\ &\leq \sum_{i=1}^n \mathbb{I}(M, Y^{i-1}; Y_i) + n\epsilon_n \\ &= \sum_{i=1}^n \left[\mathbb{I}(M, Y^{i-1}, S_{i+1}^n; Y_i) - \mathbb{I}(S_{i+1}^n; Y_i | M, Y^{i-1})\right] + n\epsilon_n \\ &\overset{(b)}{=} \sum_{i=1}^n \left[\mathbb{I}(M, Y^{i-1}, S_{i+1}^n; Y_i) - \mathbb{I}(Y^{i-1}; S_i | M, S_{i+1}^n)\right] + n\epsilon_n \\ &\overset{(c)}{=} \sum_{i=1}^n \left[\mathbb{I}(M, Y^{i-1}, S_{i+1}^n; Y_i) - \mathbb{I}(M, Y^{i-1}, S_{i+1}^n; S_i)\right] + n\epsilon_n \\ &\overset{(d)}{=} \sum_{i=1}^n \left[\mathbb{I}(U_i; Y_i) - \mathbb{I}(U_i; S_i)\right] + n\epsilon_n \\ &= n \sum_{i=1}^n \frac{1}{n} \left[\mathbb{I}(U_i; Y_i | T = i) - \mathbb{I}(U_i; S_i | T = i)\right] + n\epsilon_n \\ &= n \sum_{i=1}^n \mathbb{P}(T = i) \left[\mathbb{I}(U_i; Y_i | T = i) - \mathbb{I}(U_i; S_i | T = i)\right] + n\epsilon_n \end{aligned}$$

$$
\begin{aligned}
&= n\big[\mathbb{I}(U_T;Y_T|T) - \mathbb{I}(U_T;S_T|T)\big] + n\epsilon_n \\
&\overset{(e)}{=} n\big[\mathbb{I}(U_T;Y_T|T) - \mathbb{I}(U_T,T;S_T)\big] + n\epsilon_n \\
&\leq \big[\mathbb{I}(U_T,T;Y_T) - \mathbb{I}(U_T,T;S_T)\big] + n\epsilon_n \\
&\overset{(f)}{=} n\big[\mathbb{I}(U;Y) - \mathbb{I}(U;S)\big] + n\epsilon_n \\
&\overset{(g)}{\leq} n\big[\mathbb{I}(U;Y) - \mathbb{I}(U;S)\big] + n\epsilon
\end{aligned} \tag{N.3}
$$

where

(a) follows from Fano's inequality for n large enough;

(b) follows from Csiszár-Körner sum identity [70, Lemma 7];

(c) follows since S_i is independent of (M, S_{i+1}^n);

(d) follows by defining $U_i \triangleq (M, Y^{i-1}, S_{i+1}^n)$;

(e) follows from the independence of S_T and T;

(f) follows by defining $U = (U_T, T)$, $Y = Y_T$, and $S = S_T$;

(g) follows by defining $\epsilon \triangleq \max\{\epsilon_n, \lambda_n, \nu\}$, where we choose n large enough such that $\nu \geq \frac{\delta}{n}$.

We also have,

$$
\begin{aligned}
nR &= \mathbb{H}(M) \\
&= \mathbb{H}(M|K) \\
&\overset{(a)}{\leq} \mathbb{I}(M;Y^n|K) + n\epsilon_n \\
&= \sum_{i=1}^{n} \mathbb{I}(M;Y_i|Y^{i-1},K) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} \mathbb{I}(M,K,Y^{i-1},Z^{i-1};Y_i) + n\epsilon_n
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \big[\mathbb{I}(M, K, Y^{i-1}, Z^{i-1}, S_{i+1}^{n}; Y_i) - \mathbb{I}(S_{i+1}^{n}; Y_i | M, K, Y^{i-1}, Z^{i-1})\big] + n\epsilon_n \\
&\overset{(b)}{=} \sum_{i=1}^{n} \big[\mathbb{I}(M, K, Y^{i-1}, Z^{i-1}, S_{i+1}^{n}; Y_i) - \mathbb{I}(Y^{i-1}; S_i | M, K, S_{i+1}^{n}, Z^{i-1})\big] + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} [\mathbb{I}(U_i, V_i; Y_i) - \mathbb{I}(U_i; S_i | V_i)] + n\epsilon_n \\
&= n \sum_{i=1}^{n} \frac{1}{n} \big[\mathbb{I}(U_T, V_T; Y_T | T = i) - \mathbb{I}(U_T; S_T | V_T, T = i)\big] + n\epsilon_n \\
&= n \sum_{i=1}^{n} \mathbb{P}(T = i) \big[\mathbb{I}(U_T, V_T; Y_T | T = i) - \mathbb{I}(U_T; S_T | V_T, T = i)\big] + n\epsilon_n \\
&= n \big[\mathbb{I}(U_T, V_T; Y_T | T) - \mathbb{I}(U_T; S_T | V_T, T)\big] + n\epsilon_n \\
&\leq \big[\mathbb{I}(U_T, V_T, T; Y_T) - \mathbb{I}(U_T; S_T | V_T, T)\big] + n\epsilon_n \\
&\overset{(d)}{=} n \big[\mathbb{I}(U, V; Y) - \mathbb{I}(U; S | V)\big] + n\epsilon_n \\
&\overset{(e)}{\leq} n \big[\mathbb{I}(U, V; Y) - \mathbb{I}(U; S | V)\big] + n\epsilon,
\end{aligned} \tag{N.4}
$$

where

- (a) follows from Fano's inequality for n large enough and the fact that conditioning does not increase entropy;
- (b) follows from Csiszár-Körner sum identity [70, Lemma 7];
- (c) follows by defining $U_i \triangleq (M, Y^{i-1}, S_{i+1}^{n})$ and $V_i \triangleq (M, K, Z^{i-1}, S_{i+1}^{n})$;
- (d) follows by defining $U = (U_T, T)$, $V = (V_T, T)$, $Y = Y_T$, and $S = S_T$;
- (e) follows from definition $\epsilon \triangleq \max\{\epsilon_n, \lambda_n, \nu\}$.

Next, we lower bound nR as follows,

$$
\begin{aligned}
nR + R_K &\geq \mathbb{H}(M, K) \\
&\geq \mathbb{I}(M, K; Z^n)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \mathbb{I}(M, K; Z_i | Z^{i-1}) \\
&= \sum_{i=1}^{n} \big[\mathbb{I}(M, K, S_{i+1}^n; Z_i | Z^{i-1}) - \mathbb{I}(S_{i+1}^n; Z_i | M, K, Z^{i-1}) \big] \\
&\overset{(a)}{=} \sum_{i=1}^{n} \big[\mathbb{I}(M, K, S_{i+1}^n; Z_i | Z^{i-1}) - \mathbb{I}(Z^{i-1}; S_i | M, K, S_{i+1}^n) \big] \\
&\overset{(b)}{\geq} \sum_{i=1}^{n} \big[\mathbb{I}(M, K, S_{i+1}^n, Z^{i-1}; Z_i) - \mathbb{I}(Z^{i-1}; S_i | M, K, S_{i+1}^n) \big] - \delta \\
&\overset{(c)}{=} \sum_{i=1}^{n} \big[\mathbb{I}(M, K, S_{i+1}^n, Z^{i-1}; Z_i) - \mathbb{I}(M, K, S_{i+1}^n, Z^{i-1}; S_i) \big] - \delta \\
&\overset{(d)}{=} \sum_{i=1}^{n} \big[\mathbb{I}(V_i; Z_i) - \mathbb{I}(V_i; S_i) \big] - \delta \\
&= n \sum_{i=1}^{n} \frac{1}{n} \big[\mathbb{I}(V_T; Z_T | T = i) - \mathbb{I}(V_T; S_T | T = i) \big] - \delta \\
&= n \sum_{i=1}^{n} \mathbb{P}(T = i) \big[\mathbb{I}(V_T; Z_T | T = i) - \mathbb{I}(V_T; S_T | T = i) \big] - \delta \\
&= n \big[\mathbb{I}(V_T; Z_T | T) - \mathbb{I}(V_T; S_T | T) \big] - \delta \\
&\overset{(e)}{=} n \big[\mathbb{I}(V_T; Z_T | T) - \mathbb{I}(V_T, T; S_T) \big] - \delta \\
&\overset{(f)}{\geq} n \big[\mathbb{I}(V_T, T; Z_T) - \mathbb{I}(V_T, T; S_T) \big] - 2\delta \\
&\overset{(g)}{=} n \big[\mathbb{I}(V; Z) - \mathbb{I}(V; S) \big] - 2\delta
\end{aligned} \tag{N.5}
$$

where

(a) follows from Csiszár-Körner sum identity [70, Lemma 7];

(b) follows from Lemma 6;

(c) follows since S_i is independent of (M, K, S_{i+1}^n);

(d) follows by defining $V_i \triangleq (M, K, S_{i+1}^n, Z^{i-1})$;

(e) follows from the independence of S_T and T;

(f) follows from Lemma 6;

(g) follows by defining $V = (V_T, T)$, $Z = Z_T$, and $S = S_T$.

For any $\nu > 0$, choosing n large enough ensures that,

$$R + \frac{R_K}{n} \geq \mathbb{I}(V;Z) - \mathbb{I}(V;S) - 2\nu. \tag{N.6}$$

Therefore,

$$\begin{aligned} R &\geq \mathbb{I}(V;Z) - \mathbb{I}(V;S) - 2\nu - \frac{R_K}{n} \\ &\geq \mathbb{I}(V;Z) - \mathbb{I}(V;S) - 2\nu - \lambda_n \\ &\geq \mathbb{I}(V;Z) - \mathbb{I}(V;S) - 3\epsilon, \end{aligned} \tag{N.7}$$

where the last inequality follows since $\epsilon \triangleq \max\{\epsilon_n, \lambda_n, \nu\}$. To show that $\mathbb{D}(P_Z||Q_0) \leq \epsilon$, note that for n large enough

$$\begin{aligned} \mathbb{D}(P_Z||Q_0) = \mathbb{D}(P_{Z_T}||Q_0) &= \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} P_{Z_i}\middle|\middle|Q_0\right) \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \leq \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \frac{\delta}{n} \leq \nu \leq \epsilon. \end{aligned} \tag{N.8}$$

Combining (N.3), (N.4), (N.7), and (N.8) shows that $\forall \epsilon_n, \lambda_n, \nu > 0$, $R \leq \max\{a : a \in \mathcal{A}_\epsilon\}$. Therefore,

$$R \leq \max\left\{a : a \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon\right\}. \tag{N.9}$$

Continuity at Zero: Our objective is to show that the capacity region is included in the region defined in (3.25). The challenge, first highlighted in [77, Section VI.D], is that our converse arguments only establish that the capacity region is included in the region $\bigcap_{\epsilon>0}\mathcal{A}_\epsilon$ where $\mathcal{A}_\epsilon$ is defined in (N.2). In the sequel, the continuity of the slackness in the mutual information inequality (N.2b) will assume some importance, hence for ease of expression we define and refer to $g(\epsilon) \triangleq 3\epsilon$. As ϵ vanishes, *both* the region $\mathcal{A}_\epsilon$ and the set of distributions $\mathcal{D}_\epsilon$ shrink, so that proving the continuity at $\epsilon = 0$ is not completely straightforward. We carefully lay out the arguments leading to the result in a series of lemmas.

Lemma 7. *For all $\epsilon > 0$, the set $\mathcal{D}_\epsilon$ is closed and bounded, hence compact.*

Proof. We need to check that every constraint defining the set $\mathcal{D}_\epsilon$ defines a closed set of distributions, so that $\mathcal{D}_\epsilon$ is an intersection of closed sets and remains closed. First note that:

- the function that outputs the marginal P_Z of $P_{U,V,S,X,Y,Z}$ is continuous in P_Z;
- Q_0 has support $\mathcal{Z}$ so that the divergence $\mathbb{D}(P_Z||Q_0)$ is a continuous function of P_Z;
- mutual information, viewed as a function of the joint distribution of the random variables involved, is continuous;
- all the constraints in the definition of $\mathcal{D}_\epsilon$ are non-strict inequalities.

Consequently, the pre-images of the closed sets defined by the inequalities are pre-images of closed sets through continuous functions, hence closed. $\mathcal{D}_\epsilon$ is bounded because it is a subset of the probability simplex, hence it is compact. □

Lemma 8. *For all $\epsilon > 0$, the set $\mathcal{A}_\epsilon$ is non-empty, closed, and bounded.*

Proof. The set of Pareto optimal points in $\mathcal{A}_\epsilon$ is the image of $\mathcal{D}_\epsilon$ through a continuous function. Since $\mathcal{D}_\epsilon$ is compact, the set of Pareto optimal points is compact. In $\mathbb{R}$, compact sets are closed, hence the set of Pareto optimal points is closed and $\mathcal{A}_\epsilon$ itself is closed by definition. $\mathcal{A}_\epsilon$ is also non-empty because it contains 0. $\mathcal{A}_\epsilon$ is bounded because we can upper bound R by $2\log|\mathcal{X}| + \epsilon$. □

Now define the set

$$\mathcal{A}'_\epsilon \triangleq \left\{ R \geq 0 : \exists P_{U,V,S,X,Y,Z} \in \mathcal{D}_\epsilon : R \leq \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\} \right\}. \tag{N.10}$$

Note that $\mathcal{A}'_\epsilon$ differs from $\mathcal{A}_\epsilon$ in the absence of ϵ in the rate constraint.

Lemma 9. *For all $\epsilon > 0$, the set $\mathcal{A}'_\epsilon$ is closed and bounded.*

Proof. The proof is similar to the proof of Lemma 8. □

Lemma 10.

$$\bigcap_{\epsilon>0} \mathcal{A}'_\epsilon = \bigcap_{\epsilon>0} \mathcal{A}_\epsilon \tag{N.11}$$

Proof. First, note that $\bigcap_{\epsilon>0} \mathcal{A}'_\epsilon$ is closed, since it is an intersection of closed sets, and bounded, since the sets $\mathcal{A}'_\epsilon$ are nested and bounded. Hence, $\bigcap_{\epsilon>0} \mathcal{A}'_\epsilon$ is compact. Consequently, there exists a maximal element, $r^\dagger$. Consider any $r \in [0, r^\dagger]$. Then $\forall \epsilon > 0$ $\exists P_{U,V,S,X,Y,Z} \in \mathcal{D}_\epsilon$ $r \leq r^\dagger \leq \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\}$ and $r \in \bigcap_{\epsilon>0} \mathcal{A}'_\epsilon$.

We now want to show that $\bigcap_{\epsilon>0} \mathcal{A}'_\epsilon = \bigcap_{\epsilon>0} \mathcal{A}_\epsilon$. The hard part is showing that, $\bigcap_{\epsilon>0} \mathcal{A}_\epsilon \subset \bigcap_{\epsilon>0} \mathcal{A}'_\epsilon$ since the other direction follows by the definition of $\mathcal{A}_\epsilon$ and $\mathcal{A}'_\epsilon$. We proceed by contradiction. Assume $\exists r^* \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon$ such that $r^* \notin \bigcap_{\epsilon>0} \mathcal{A}'_\epsilon$. It must be that $r^\dagger < r^*$ for otherwise $r^* \in \bigcap_{\epsilon>0} \mathcal{A}'_\epsilon$ as noted earlier.

Set $r_0 \triangleq \frac{1}{2}(r^\dagger + r^*)$, which is such that $r_0 > r^\dagger$ and therefore $r_0 \notin \bigcap_{\epsilon>0} \mathcal{A}'_\epsilon$. Set $\epsilon' > 0$ such that $\forall \epsilon \leq \epsilon'$ $g(\epsilon) < \frac{r^* - r^\dagger}{2}$, which exists by the assumptions on g. Assume that $\forall \epsilon \in (0; \epsilon']$ $r_0 \in \mathcal{A}'_\epsilon$. Then, $r_0 \in \bigcap_{\epsilon>0} \mathcal{A}'_\epsilon$ which contradicts our assumption. Hence, there exists $0 < \epsilon_0 \leq \epsilon'$ such that $r_0 \notin \mathcal{A}'_{\epsilon_0}$. Hence $\forall P_{U,V,S,X,Y,Z} \in \mathcal{D}_{\epsilon_0}$ $r_0 > \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\}$. Then $\forall P_{U,V,S,X,Y,Z} \in \mathcal{D}_{\epsilon_0}$

$$r_0 > \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\} \tag{N.12}$$

$$\Rightarrow \frac{r^* + r^\dagger}{2} > \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\} \tag{N.13}$$

$$\Rightarrow \frac{r^* + r^\dagger}{2} + \frac{r^* - r^\dagger}{2} > \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\} + \frac{r^* - r^\dagger}{2} \tag{N.14}$$

$$\Rightarrow r^* > \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\} + g(\epsilon_0) \tag{N.15}$$

Since $r^* \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon$, we have $\forall \epsilon > 0 \ \exists P_{U,V,S,X,Y,Z} \in \mathcal{D}_\epsilon$ such that $r^* \leq \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\} + g(\epsilon)$. Hence, there is a contradiction, and we must have $r^* \in \bigcap_{\epsilon>0} \mathcal{A}'_\epsilon$. □

To conclude, one can prove that $\bigcap_{\epsilon>0} \mathcal{A}'_\epsilon = \mathcal{A}_0$, following the exact same arguments as in [77, Section IV.C].

APPENDIX O

PROOF OF THEOREM 13

We adopt a block-Markov encoding scheme in which B independent messages are transmitted over B channel blocks each of length r, such that $n = rB$. The warden's observation is $Z^n = (Z_1^r, \ldots, Z_B^r)$, the distribution induced at the output of the warden is P_{Z^n}, the target output distribution is $Q_0^{\otimes n}$, and Equation (I.2), describing the distance between the two distributions, continues to hold. The random code generation is as follows.

Fix $P_U(u)$, $P_{V|S}(v|s)$, $x(u,s)$, and $\epsilon_1 > \epsilon_2 > 0$ such that, $P_Z = Q_0$.

Codebook Generation for Keys: For each block $j \in [\![1, B]\!]$, let $C_1^{(r)} \triangleq \left\{V^r(\ell_j)\right\}_{\ell_j \in \mathcal{L}}$, where $\mathcal{L} \triangleq [\![1, 2^{r\tilde{R}}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_V^{\otimes r}$, where $P_V = \sum_{s\in\mathcal{S}} Q_S(s) P_{V|S}(v|s)$. We denote a realization of $C_1^{(r)}$ by $\mathcal{C}_1^{(r)} \triangleq \left\{v^r(\ell_j)\right\}_{\ell_j \in \mathcal{L}}$. Partition the set of indices $\ell_j \in [\![1, 2^{r\tilde{R}}]\!]$ into bins $\mathcal{B}(t)$, $t \in [\![1, 2^{rR_T}]\!]$ by using function $\varphi : V^r(\ell_j) \mapsto [\![1, 2^{rR_T}]\!]$ through random binning by choosing the value of $\varphi(v^r(\ell_j))$ independently and uniformly at random for every $v^r(\ell_j) \in \mathcal{V}^r$. For each block $j \in [\![1, B]\!]$, create a function $\Phi : V^r(\ell_j) \mapsto [\![1, 2^{rR_K}]\!]$ through random binning by choosing the value of $\Phi(v^r(\ell_j))$ independently and uniformly at random for every $v^r(\ell_j) \in \mathcal{V}^r$. The key $k_j = \Phi(v^r(\ell_j))$ obtained in block $j \in [\![1, B]\!]$ from the description of the CSI sequence $v^r(\ell_j)$ is used to assist the encoder in block $j + 2$.

Codebook Generation for Messages: For each block $j \in [\![1, B]\!]$, let $C_2^{(r)} \triangleq \left\{U^r(m_j, t_{j-1}, k_{j-2})\right\}_{(m_j, t_{j-1}, k_{j-2}) \in \mathcal{M}\times\mathcal{T}\times\mathcal{K}}$, where $\mathcal{M} \triangleq [\![1, 2^{rR}]\!]$, $\mathcal{T} \triangleq [\![1, 2^{rR_t}]\!]$, and $\mathcal{K} \triangleq [\![1, 2^{rR_k}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_U^{\otimes r}$. We denote a realization of $C_2^{(r)}$ by $\mathcal{C}_2^{(r)} \triangleq \left\{u^r(m_j, t_{j-1}, k_{j-2})\right\}_{(m_j, t_{j-1}, k_{j-2}) \in \mathcal{M}\times\mathcal{T}\times\mathcal{K}}$. Also, let $C_r = \left\{C_1^{(r)}, C_2^{(r)}\right\}$ and $\mathcal{C}_r = \left\{\mathcal{C}_1^{(r)}, \mathcal{C}_2^{(r)}\right\}$. The indices (m_j, t_{j-1}, k_{j-2}) can be viewed as a two layer binning. We define an ideal PMF for codebook $\mathcal{C}_n$, as an approximate distribution to facilitate the analysis as follows

$$\Gamma^{(\mathcal{C}_n)}_{M_j, T_{j-1}, K_{j-2}, L_j, U^r, V^r, S_j^r, Z_j^r, K_{j-1}, T_j, K_j}(m_j, t_{j-1}, k_{j-2}, \ell_j, \tilde{u}^r, \tilde{v}^r, s_j^r, z_j^r, k_{j-1}, t_j, k_j)$$

$$= 2^{-r(R+R_t+R_k+\tilde{R})} \mathbb{1}_{\{\tilde{u}^r = u^r(m_j, t_{j-1}, k_{j-2})\}} \mathbb{1}_{\{\tilde{v}^r = v^r(\ell_j)\}} P_{S|V}^{\otimes r}(s_j^r|\tilde{v}^r) W_{Z|U,S}^{\otimes r}(z_j^r|\tilde{u}^r, s_j^r)$$
$$\times 2^{-rR_k} \mathbb{1}_{\{t_j = \varphi(\tilde{v}^r)\}} \mathbb{1}_{\{k_j = \Phi(\tilde{v}^r)\}}, \tag{O.1}$$

where $W_{Z|U,S}$ is the marginal distribution of $W_{Z|U,S} = \sum_{x \in \mathcal{X}} \mathbb{1}_{\{x = x(u,s)\}} W_{Z|X,S}$ and

$$P_{S|V} = \frac{P_{S,V}(s,v)}{P_V(v)} = \frac{Q_S(s) P_{V|S}(v|s)}{\sum_{s \in \mathcal{S}} Q_S(s) P_{V|S}(v|s)}. \tag{O.2}$$

Encoding: We assume that the transmitter and the receiver have access to the shared secret keys k_{-1} and k_0 for the first two blocks, but after the first two blocks they use the key that they generate from the CSI.

In the first block, to send the message m_1 according to k_{-1}, the encoder generates the index t_0 uniformly at random and computes $u^r(m_1, t_0, k_{-1})$ and transmits a codeword x^r, where $x_i = x(u_i(m_1, t_0, k_{-1}), s_{1,i})$. Note that, the index t_0 does not convey any useful information. At the end of the first block, to generate a secret key shared between the transmitter and the receiver, the encoder generates the index ℓ_1 according to the following distribution with $j = 1$,

$$f\big(\ell_j | s_j^r\big) = \frac{P_{S|V}^{\otimes r}\big(s_j^r | v^r(\ell_j)\big)}{\sum\limits_{\ell' \in [\![1, 2^{r\tilde{R}}]\!]} P_{S|V}^{\otimes r}\big(s_j^r | v^r(\ell_j')\big)}, \tag{O.3}$$

where $P_{S|V}$ is defined in (O.2). Then generates the reconciliation index $t_1 = \varphi(v^n(\ell_1))$; simultaneously, the transmitter generates a key $k_1 = \Phi(v^r(\ell_1))$ from the description of its CSI of the first block $v^r(\ell_1)$ to be used in Block 3.

In the second block, to transmit the message m_2 and the reconciliation index t_1 according to the key k_0, the encoder computes $u^r(m_2, t_1, k_0)$ and transmits a codeword x^r, where $x_i = x(u_i(m_2, t_1, k_0), s_{2,i})$. At the end of the second block, to generate a secret key shared between the transmitter and the receiver, the encoder generates the index ℓ_2 based s_2^r by using the likelihood encoder described in (O.3) with $j = 2$. Then generates the reconciliation index $t_2 = \varphi(v^n(\ell_2))$; simultaneously, the transmitter generates a key $k_2 = \Phi(v^r(\ell_2))$ from the description of its CSI of the second block $v^r(\ell_1)$ to be used in Block 4.

In block $j \in [\![3, B]\!]$, to send the message m_j and the reconciliation index t_{j-1} according to the generated key k_{j-2} from the previous blocks and the CSI of the current block s_j^r, the encoder computes $u^r(m_j, t_{j-1}, k_{j-2})$ and transmits a codeword x^r, where each coordinate of the transmitted signal is a function of the current state s_j^r as well as the corresponding sample of the transmitter's codeword u_i, i.e., $x_i = x(u_i(m_j, t_{j-1}, k_{j-2}), s_{j,i})$. At the end of this block, the encoder first selects the index ℓ_j based s_j^r by using the likelihood encoder described in (O.3) and then generates the reconciliation index $t_j = \varphi(v^r(\ell_j))$; simultaneously the encoder generates a key $k_j = \Phi(v^r(\ell_j))$ from the description of its CSI of the block j, $v^r(\ell_j)$, to be used in the Block $j+2$.

Define

$$\begin{aligned}
&\Upsilon^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2},U^r,S_j^r,L_j,V^r,Z_j^r,K_{j-1},T_j,K_j}(m_j, t_{j-1}, k_{j-2}, \tilde{u}^r, s_j^r, \ell_j, \tilde{v}^r, z_j^r, k_{j-1}, t_j, k_j) \\
&\quad \triangleq 2^{-r(R+R_t+R_k)} \mathbb{1}_{\{\tilde{u}^r = u^r(m_j, t_{j-1}, k_{j-2})\}} Q_S^{\otimes r}(s_j^r) f(\ell_j | s_j^r) \mathbb{1}_{\{\tilde{v}^r = v^r(\ell_j)\}} \\
&\qquad \times W_{Z|U,S}^{\otimes r}(z_j^r | \tilde{u}^r, s_j^r) 2^{-rR_k} \mathbb{1}_{\{t_j = \varphi(\tilde{v}^r)\}} \mathbb{1}_{\{k_j = \Phi(\tilde{v}^r)\}}.
\end{aligned} \tag{O.4}$$

For a fixed codebook $\mathcal{C}_r$, the induced joint distribution by our code design (i.e. $P^{(\mathcal{C}_r)}$) satisfies

$$\mathbb{D}\Big(P^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2},U^r,S_j^r,L_j,V^r,Z_j^r,K_{j-1},T_j,K_j} || \Upsilon^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2},U^r,S_j^r,L_j,V^r,Z_j^r,K_{j-1},T_j,K_j}\Big) \leq \epsilon. \tag{O.5}$$

This intermediate distribution $\Upsilon^{(\mathcal{C}_r)}$ approximates the true distribution $P^{(\mathcal{C}_r)}$ and will be used in the sequel for bounding purposes. Expression (O.5) holds because the main difference between $P^{(\mathcal{C}_r)}$ and $\Upsilon^{(\mathcal{C}_r)}$ is that the keys K_{j-2}, K_{j-1} and the reconciliation index T_{j-1} are assumed to be uniformly distributed in $\Upsilon^{(\mathcal{C}_r)}$, which are made (arbitrarily) nearly uniform in $P^{(\mathcal{C}_r)}$ with appropriate control of rate as in (O.11) and (O.16).

Covert Analysis: We now show that this coding scheme guarantees that $\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n} || Q_Z^{\otimes n})] \underset{n\to\infty}{\longrightarrow} 0$, where C_n is the set of all the codebooks for all blocks, and

$$Q_Z(\cdot) = \sum_{u\in\mathcal{U}} \sum_{v\in\mathcal{V}} \sum_{s\in\mathcal{S}} \sum_{x\in\mathcal{X}} P_U(u) P_V(v) P_{S|V}(s|v) \mathbb{1}_{\left\{X = X(u,s)\right\}} W_{Z|X,S}(\cdot|x,s), \tag{O.6}$$

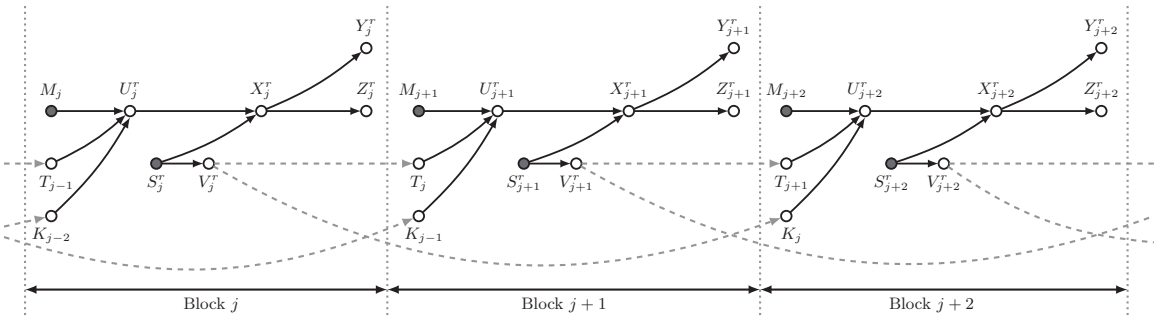

Figure O.1. Functional dependence graph for the block-Markov encoding scheme

such that $\sum_{v\in\mathcal{V}} P_V(v)P_{S|V}(\cdot|v) = Q_S(\cdot)$. Then, we choose P_U, P_V, $P_{S|V}$, and $x(u,s)$ such that it satisfies $Q_Z = Q_0$. Similar to (K.10) using the functional dependence graph depicted in Fig. O.1 it follows that,

$$\mathbb{D}(P_{Z^n}^{(\mathcal{C}_r)}||Q_Z^{\otimes n}) \leq 2\sum_{j=1}^{B}\mathbb{D}(P_{Z_j^r,K_{j-1},T_j,K_j}^{(\mathcal{C}_r)}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}). \tag{O.7}$$

To bound the RHS of (O.7) by using Lemma 1 and the triangle inequality we have,

$$\begin{aligned}
&\mathbb{E}_{C_r}||P_{Z_j^r,K_{j-1},T_j,K_j|C_r} - Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}||_1 \\
&\leq \mathbb{E}_{C_r}||P_{Z_j^r,K_{j-1},T_j,K_j|C_r} - \Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||_1 + \mathbb{E}_{C_r}||\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r} - Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}||_1 \\
&\leq \mathbb{E}_{C_r}||P_{Z_j^r,K_{j-1},T_j,K_j|C_r} - \Upsilon_{Z_j^r,K_{j-1},T_j,K_j|C_r}||_1 + \mathbb{E}_{C_r}||\Upsilon_{Z_j^r,K_{j-1},T_j,K_j|C_r} - \Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||_1 \\
&\quad + \mathbb{E}_{C_r}||\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r} - Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}||_1.
\end{aligned} \tag{O.8}$$

From (O.5) and the monotonicity of KL-divergence the first term on the RHS of (O.8) goes to zero when r grows. To bound the second term on the RHS of (O.8) for a fixed codebook $\mathcal{C}_r$, we have,

$$\Gamma_{M_j,T_{j-1},K_{j-2}}^{(\mathcal{C}_r)} = 2^{-r(R+R_t+R_k)} = \Upsilon_{M_j,T_{j-1},K_{j-2}}^{(\mathcal{C}_r)}, \tag{O.9a}$$

$$\Gamma_{U^r|M_j,T_{j-1},K_{j-2},S_j^r}^{(\mathcal{C}_r)} = \mathbb{1}_{\{\tilde{u}^r=u^r(m_j,t_{j-1},k_{j-2})\}} = \Upsilon_{U^r|M_j,T_{j-1},K_{j-2},S_j^r}^{(\mathcal{C}_r)}, \tag{O.9b}$$

$$\Gamma_{L_j|M_j,T_{j-1},K_{j-2},S_j^r,U^r}^{(\mathcal{C}_r)} = f(\ell_j|s_j^r) = \Upsilon_{L_j|M_j,T_{j-1},K_{j-2},S_j^r,U^r}^{(\mathcal{C}_r)}, \tag{O.9c}$$

$$\Gamma_{V^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r}^{(\mathcal{C}_r)} = \mathbb{1}_{\{\tilde{v}^r=v^r(\ell_j)\}} = \Upsilon_{V^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r}^{(\mathcal{C}_r)}, \tag{O.9d}$$

$$\Gamma^{(\mathcal{C}_r)}_{Z_j^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r} = W^{\otimes r}_{Z|U,S} = \Upsilon^{(\mathcal{C}_r)}_{Z_j^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r}, \tag{O.9e}$$

$$\Gamma^{(\mathcal{C}_r)}_{K_{j-1}|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r} = 2^{-rR_k} = \Upsilon^{(\mathcal{C}_r)}_{K_{j-1}|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r}, \tag{O.9f}$$

$$\Gamma^{(\mathcal{C}_r)}_{T_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r,K_{j-1}} = \mathbb{1}_{\{t_j=\sigma(v^r(\ell_j))\}} = \Upsilon^{(\mathcal{C}_r)}_{T_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r,K_{j-1}}, \tag{O.9g}$$

$$\Gamma^{(\mathcal{C}_r)}_{K_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r,K_{j-1},T_j} = \mathbb{1}_{\{k_j=\Phi(v^r(\ell_j))\}} = \Upsilon^{(\mathcal{C}_r)}_{K_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r,K_{j-1},T_j}, \tag{O.9h}$$

where (O.9c) follows from (O.3). Hence,

$$\begin{aligned}
&\mathbb{E}_{C_r}||\Upsilon_{Z_j^r,K_{j-1},T_j,K_j|C_r} - \Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||_1 \\
&\leq \mathbb{E}_{C_r}||\Upsilon_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j|C_r} - \Gamma_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j|C_r}||_1 \\
&\overset{(a)}{=} \mathbb{E}_{C_r}||\Upsilon_{M_j,T_{j-1},K_{j-2},S_j^r|C_r} - \Gamma_{M_j,T_{j-1},K_{j-2},S_j^r|C_r}||_1 \\
&\overset{(b)}{=} \mathbb{E}_{C_r}||Q_S^{\otimes r} - \Gamma_{S_j^r|M_j=1,T_{j-1}=1,K_{j-2}=1,C_r}||_1,
\end{aligned} \tag{O.10}$$

where (a) follows from (O.9b)-(O.9h) and (b) follows from the symmetry of the codebook construction with respect to M_j, T_{j-1}, and K_{j-2} and (O.9a). Based on the soft covering lemma [77, Corollary VII.5] the RHS of (O.10) vanishes when r grows if

$$\tilde{R} > \mathbb{I}(S;V). \tag{O.11}$$

We now proceed to bound the third term on the RHS of (O.8). First, consider the following marginal from (O.1),

$$\begin{aligned}
&\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}(z_j^r,k_{j-1},t_j,k_j) \\
&= \sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{s_j^r} \frac{1}{2^{r(R+R_t+2R_k+\tilde{R})}} P_{S|V}^{\otimes r}\big(s_j^r|V^r(\ell_j)\big) \\
&\quad \times W_{Z|U,S}^{\otimes r}\big(z_j^r|U^r(m_j,t_{j-1},k_{j-2}),s_j^r\big)\mathbb{1}_{\{t_j=\varphi(V^r(\ell_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\ell_j))\}} \\
&= \sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j} \frac{1}{2^{r(R+R_t+2R_k+\tilde{R})}} W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(m_j,t_{j-1},k_{j-2}),V^r(\ell_j)\big)
\end{aligned}$$

$$\times \mathbb{1}_{\{t_j=\varphi(V^r(\ell_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\ell_j))\}}, \tag{O.12}$$

where $W_{Z|U,V}(z|u,v) = \sum_{s\in\mathcal{S}} P_{S|V}(s|v)W_{Z|U,S}(z|u,s)$. To bound the third term on the RHS of (O.8) by using Pinsker's inequality, it is sufficient to bound $\mathbb{E}_{C_r}[\mathbb{D}(\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j})]$ as follows,

$$\begin{aligned}
&\mathbb{E}_{C_r}\left[\mathbb{D}(\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j})\right]\\
&= \mathbb{E}_{C_r}\left[\sum_{z_j^r,k_{j-1},t_j,k_j}\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}(z_j^r,k_{j-1},t_j,k_j)\log\left(\frac{\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}(z_j^r,k_{j-1},t_j,k_j)}{Q_Z^{\otimes r}(z_j^r)Q_{K_{j-1}}(k_{j-1})Q_{T_j}(t_j)Q_{K_j}(k_j)}\right)\right]\\
&= \mathbb{E}_{C_r}\Bigg[\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{1}{2^{r(R+R_t+2R_k+\tilde{R})}}\\
&\times W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(m_j,t_{j-1},k_{j-2}),V^r(\ell_j)\big)\mathbb{1}_{\{t_j=\varphi(V^r(\ell_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\ell_j))\}}\\
&\times\log\left(\frac{\sum_{\tilde{m}_j}\sum_{\tilde{t}_{j-1}}\sum_{\tilde{k}_{j-2}}\sum_{\tilde{\ell}_j}W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2}),V^r(\tilde{\ell}_j)\big)\mathbb{1}_{\{t_j=\varphi(V^r(\tilde{\ell}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{\ell}_j))\}}}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\right)\Bigg]\\
&\overset{(a)}{\leq}\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{1}{2^{r(R+R_t+2R_k+\tilde{R})}}\\
&\times\sum_{\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}\Gamma_{U^r,V^r,Z_j^r}^{\otimes r}\big(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\big)\\
&\times\mathbb{E}_{\varphi(v^r(\ell_j))}\left[\mathbb{1}_{\{t_j=\varphi(v^r(\ell_j))\}}\right]\times\mathbb{E}_{\Phi(v^r(\ell_j))}\left[\mathbb{1}_{\{k_j=\Phi(v^r(\ell_j))\}}\right]\\
&\times\log\mathbb{E}_{\substack{\setminus(m_j,t_{j-1},k_{j-2},\ell_j),\\ \setminus(\varphi(v^r(\ell_j)),\Phi(v^r(\ell_j)))}}\Bigg[\frac{1}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&\times\sum_{\tilde{m}_j}\sum_{\tilde{t}_{j-1}}\sum_{\tilde{k}_{j-2}}\sum_{\tilde{\ell}_j}W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2}),V^r(\tilde{\ell}_j)\big)\mathbb{1}_{\{t_j=\varphi(V^r(\tilde{\ell}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{\ell}_j))\}}\Bigg]\\
&\overset{(b)}{\leq}\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{1}{2^{r(R+R_t+2R_k+\tilde{R}+R_T+R_K)}}\\
&\times\sum_{\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}\Gamma_{U^r,V^r,Z_j^r}^{\otimes r}\big(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\big)\\
&\times\log\frac{1}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\times\Bigg(W_{Z|U,V}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\big)
\end{aligned}$$

$$
\begin{aligned}
&+\mathbb{E}_{\backslash(m_j,t_{j-1},k_{j-2})}\Bigg[\sum_{(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2})\neq(m_j,t_{j-1},k_{j-2})} W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2}),v^r(\ell_j)\big)\Bigg]\\
&+\mathbb{E}_{\substack{\backslash\ell_j,\\ \backslash(\varphi(v^r(\ell_j)),\Phi(v^r(\ell_j)))}}\Bigg[\sum_{\tilde{\ell}_j\neq\ell_j} W_{Z|U,V}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2}),V^r(\tilde{\ell}_j)\big)\mathbb{1}_{\{t_j=\varphi(V^r(\tilde{\ell}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{\ell}_j))\}}\Bigg]\\
&+\mathbb{E}_{\substack{\backslash(m_j,t_{j-1},k_{j-2},\ell_j),\\ \backslash(\varphi(v^r(\ell_j)),\Phi(v^r(\ell_j)))}}\Bigg[\sum_{\tilde{\ell}_j\neq\ell_j}\sum_{(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2})\neq(m_j,t_{j-1},k_{j-2})} W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2}),V^r(\tilde{\ell}_j)\big)\\
&\times\mathbb{1}_{\{t_j=\varphi(V^r(\tilde{\ell}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{\ell}_j))\}}\Bigg]\Bigg)\\
&\overset{(c)}{\leq}\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{1}{2^{r(R+R_t+2R_k+\tilde{R}+R_T+R_K)}}\\
&\times\sum_{\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}\Gamma_{U^r,V^r,Z_j^r}^{\otimes r}\big(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\big)\\
&\times\log\Bigg(\frac{W_{Z|U,V}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\big)}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&+\sum_{(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2})\neq(m_j,t_{j-1},k_{j-2})}\frac{W_{Z|V}^{\otimes r}\big(z_j^r|v^r(\ell_j)\big)}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&+\sum_{\tilde{\ell}_j\neq\ell_j}\frac{W_{Z|U}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2})\big)}{2^{r(R+R_t+R_k+\tilde{R})}Q_Z^{\otimes r}(z_j^r)}+1\Bigg)\\
&\leq\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{1}{2^{r(R+R_t+2R_k+\tilde{R}+R_T+R_K)}}\\
&\times\sum_{\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}\Gamma_{U^r,V^r,Z_j^r}^{\otimes r}\big(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\big)\\
&\times\log\Bigg(\frac{W_{Z|U,V}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\big)}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|V}^{\otimes r}\big(z_j^r|v^r(\ell_j)\big)}{2^{r(\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&+\frac{W_{Z|U}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2})\big)}{2^{r(R+R_t+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\Bigg)\\
&\triangleq\Psi_1+\Psi_2,
\end{aligned}
\tag{O.13}
$$

where (a) follows from Jensen's inequality, (b) and (c) hold because $\mathbb{1}_{\{\cdot\}} \leq 1$. We define Ψ_1 and Ψ_2 as

$$
\begin{aligned}
\Psi_1 = & \sum_{(k_{j-1},t_j,k_j)} \sum_{m_j} \sum_{t_{j-1}} \sum_{k_{j-2}} \sum_{\ell_j} \frac{1}{2^{r(R+R_t+2R_k+\tilde{R}+R_T+R_K)}} \\
& \times \sum_{\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\right)\in\mathcal{T}_\epsilon^{(r)}} \Gamma_{U^r,V^r,Z_j^r}^{\otimes r}\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\right) \\
& \times \log\left(\frac{W_{Z|U,V}^{\otimes r}\left(z_j^r|u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)} + \frac{W_{Z|V}^{\otimes r}\left(z_j^r|v^r(\ell_j)\right)}{2^{r(\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\right. \\
& \left. + \frac{W_{Z|U}^{\otimes r}\left(z_j^r|u^r(m_j,t_{j-1},k_{j-2})\right)}{2^{r(R+R_t+R_k)}Q_Z^{\otimes r}(z_j^r)} + 1\right) \\
& \leq \log\left(\frac{2^{r(R_T+R_K)}\times 2^{-r(1-\epsilon)\mathbb{H}(Z|U,V)}}{2^{r(R+R_t+R_k+\tilde{R})}\times 2^{-r(1+\epsilon)\mathbb{H}(Z)}} + \frac{2^{r(R_T+R_K)}\times 2^{-r(1-\epsilon)\mathbb{H}(Z|V)}}{2^{r\tilde{R}}\times 2^{-r(1+\epsilon)\mathbb{H}(Z)}}\right. \\
& \left. + \frac{2^{-r(1-\epsilon)\mathbb{H}(Z|U)}}{2^{r(R+R_t+R_k)}\times 2^{-r(1+\epsilon)\mathbb{H}(Z)}} + 1\right) \qquad \text{(O.14)} \\
\Psi_2 = & \sum_{(k_{j-1},t_j,k_j)} \sum_{m_j} \sum_{t_{j-1}} \sum_{k_{j-2}} \sum_{\ell_j} \frac{1}{2^{r(R+R_t+2R_k+\tilde{R}+R_T+R_K)}} \\
& \times \sum_{\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\right)\notin\mathcal{T}_\epsilon^{(r)}} \Gamma_{U^r,V^r,Z_j^r}^{\otimes r}\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\right) \\
& \times \log\left(\frac{W_{Z|U,V}^{\otimes r}\left(z_j^r|u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)} + \frac{W_{Z|V}^{\otimes r}\left(z_j^r|v^r(\ell_j)\right)}{2^{r(\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\right. \\
& \left. + \frac{W_{Z|U}^{\otimes r}\left(z_j^r|u^r(m_j,t_{j-1},k_{j-2})\right)}{2^{r(R+R_t+R_k)}Q_Z^{\otimes r}(z_j^r)} + 1\right) \\
& \leq 2|V||U||Z|e^{-r\epsilon^2\mu_{V,U,Z}} r\log\left(\frac{3}{\mu_Z}+1\right). \qquad \text{(O.15)}
\end{aligned}
$$

In (O.15) $\mu_{V,U,Z} = \min_{(v,u,z)\in(\mathcal{V},\mathcal{U},\mathcal{Z})} P_{V,U,Z}(v,u,z)$ and $\mu_Z = \min_{z\in\mathcal{Z}} P_Z(z)$. When $r\to\infty$ then $\Psi_2 \to 0$ and Ψ_1 goes to zero when r grows if

$$R + R_t + R_k + \tilde{R} - R_T - R_K > \mathbb{I}(U,V;Z), \qquad \text{(O.16a)}$$

$$\tilde{R} - R_T - R_K > \mathbb{I}(V;Z), \qquad \text{(O.16b)}$$

$$R + R_t + R_k > \mathbb{I}(U;Z). \qquad \text{(O.16c)}$$

Decoding and Error Probability Analysis: At the end of the block $j \in [\![1, B]\!]$, using its knowledge of the key k_{j-2} generated from the block $j-2$, the receiver finds a unique pair $(\hat{m}_j, \hat{t}_{j-1})$ such that $\left(u^r(\hat{m}_j, \hat{t}_{j-1}, k_{j-2}), y_j^r\right) \in \mathcal{T}_\epsilon^{(r)}$. According to the law of large numbers and the packing lemma probability of error vanishes when r grows if [85],

$$R + R_t < \mathbb{I}(U; Y). \tag{O.17}$$

We now analyze the probability of error at the encoder and the decoder for key generation. Let (L_{j-1}, T_{j-1}) denote the chosen indices at the encoder and $\hat{L}_{j-1}$ and $\hat{T}_{j-1}$ be the estimate of the index L_{j-1} and T_{j-1} at the decoder. At the end of block j, by decoding U_j^r, the receiver knows T_{j-1} and to find L_{j-1} we define the error event,

$$\mathcal{E} = \left\{ \left(V_{j-1}^r(\hat{L}_{j-1}), S_{j-1}^r, U_{j-1}^r, Y_{j-1}^r\right) \notin \mathcal{T}_\epsilon^{(r)} \right\}. \tag{O.18}$$

Also, consider the error events,

$$\mathcal{E}_1 = \left\{ \left(V_{j-1}^r(\ell_{j-1}), S_{j-1}^r\right) \notin \mathcal{T}_{\epsilon'}^{(r)} \text{ for all } \ell_{j-1} \in [\![1, 2^{r\tilde{R}}]\!] \right\}, \tag{O.19a}$$

$$\mathcal{E}_2 = \left\{ \left(V_{j-1}^r(L_{j-1}), S_{j-1}^r, U_{j-1}^r, Y_{j-1}^r\right) \notin \mathcal{T}_\epsilon^{(r)} \right\}, \tag{O.19b}$$

$$\mathcal{E}_3 = \left\{ \left(V_{j-1}^r(\tilde{\ell}_{j-1}), U_{j-1}^r, Y_{j-1}^r\right) \in \mathcal{T}_\epsilon^{(r)} \text{ for some } \ell_{j-1} \in \mathcal{B}(T_{j-1}), \tilde{\ell}_{j-1} \neq \ell_{j-1} \right\}, \tag{O.19c}$$

where $\epsilon > \epsilon' > 0$. By the union bound we have,

$$P(\mathcal{E}) \leq P(\mathcal{E}_1) + P(\mathcal{E}_1^c \cap \mathcal{E}_2) + P(\mathcal{E}_3). \tag{O.20}$$

Similar to the proof of Lemma 5 one can show that the first term on the RHS of (O.20) vanishes when r grows if we have (O.11). Following the steps in [85, Sec. 11.3.1], the last two terms on the RHS of (O.20) go to zero when r grows if,

$$\tilde{R} > \mathbb{I}(S; V), \tag{O.21a}$$

$$\tilde{R} - R_t < \mathbb{I}(V; U, Y). \tag{O.21b}$$

Applying Fourier-Motzkin to (O.11), (O.16), (O.17), and (O.21) and remarking that the scheme requires $R_t + R_k \leq R_T + R_K$ results in the achievable region in Theorem 13.

APPENDIX P

PROOF OF THEOREM 14

Fix $P_U(u)$, $x(u,s)$, and $\epsilon_1 > \epsilon_2 > 0$ such that, $P_Z = Q_0$.

Codebook Generation: Let $C_n \triangleq \{U^n(m)\}_{m\in\mathcal{M}}$, where $\mathcal{M} \triangleq [\![1, 2^{nR}]\!]$, be a random codebook consisting of independent random sequences each generated according to $\prod_{i=1}^n P_U(u_i)$. We denote a realization of C_n by $\mathcal{C}_n \triangleq \{u^n(m)\}_{m\in\mathcal{M}}$.

Encoding: To send the message m, the encoder computes $u^n(m)$ and transmits codeword x^n, where $x_i = x(u_i(m), s_i)$. For a fixed codebook $\mathcal{C}_n$, the induced joint distribution over the codebook is as follows

$$P^{(\mathcal{C}_n)}_{M,S^n,U^n,Z^n}(m, s^n, \tilde{u}^n, z^n) = 2^{-nR} Q_S^{\otimes n}(s^n) \mathbb{1}_{\{\tilde{u}^n = u^n(m)\}} W^{\otimes n}_{Z|U,S}(z^n|\tilde{u}^n, s^n). \tag{P.1}$$

Covert Analysis: We now show that this coding scheme guarantees that

$$\mathbb{E}_{C_n}\left[\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n})\right] \underset{n\to\infty}{\longrightarrow} 0, \tag{P.2}$$

where

$$Q_Z(\cdot) = \sum_{s\in\mathcal{S}} \sum_{u\in\mathcal{U}} \sum_{x\in\mathcal{X}} Q_S(s) P_U(u) \mathbb{1}_{\left\{X = X(U,S)\right\}} W_{Z|X,S}(\cdot|x,s). \tag{P.3}$$

Then we choose P_U and $x(u,s)$ such that it satisfies $Q_Z = Q_0$. First, consider the following marginal from (P.1)

$$P_{Z^n|C_n}(z^n) = \sum_m \sum_{s^n} 2^{-nR} Q_S^{\otimes n}(s^n) W^{\otimes n}_{Z|U,S}\big(z^n|u^n(m), s^n\big) \tag{P.4}$$

$$= \sum_m 2^{-nR} W^{\otimes n}_{Z|U}\big(z^n|u^n(m)\big), \tag{P.5}$$

where $W_{Z|U} = \sum_{s\in\mathcal{S}} Q_S(s) W_{Z|U,S}(z|u,s)$. By [92, Theorem 1] one can show that (P.2) holds if,

$$R > \mathbb{I}(U;Z). \tag{P.6}$$

Decoding and Error Probability Analysis: Upon receiving y^n, the receiver finds a unique message $\hat{m}$ such that $\left(u^n(\hat{m}), y^n\right) \in \mathcal{T}_\epsilon^{(n)}$. According to the law of large numbers and the packing lemma, probability of error vanishes when n grows if [85],

$$R < \mathbb{I}(U;Y). \tag{P.7}$$

APPENDIX Q

PROOF OF THEOREM 15

Consider any sequence of length-n codes for a state-dependent channel with CSI available causally only at the transmitter such that $P_e^{(n)} \leq \epsilon_n$, $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \delta$, and $R_K/n \leq \lambda_n$ with $\lim_{n\to\infty} \epsilon_n = \lim_{n\to\infty} \lambda_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish.

Epsilon Rate Region: We first define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ that expands the region defined in (3.33) as follows,

$$\mathcal{A}_\epsilon \triangleq \left\{R \geq 0 : \exists P_{U,V,S,X,Y,Z} \in \mathcal{D}_\epsilon : R \leq \mathbb{I}(U;Y) + \epsilon\right\}, \tag{Q.1a}$$

where

$$\mathcal{D}_\epsilon = \left\{ \begin{array}{l} P_{U,V,S,X,Y,Z} : \\ P_{U,V,S,X,Y,Z} = Q_S P_V P_{U|V} \mathbb{1}_{\left\{X=X(U,S)\right\}} W_{Y,Z|X,S} \\ \mathbb{D}\left(P_Z \| Q_0\right) \leq \epsilon \\ \mathbb{I}(U;Y) \geq \mathbb{I}(V;Z) - 4\epsilon \\ \max\{|\mathcal{U}|, |\mathcal{V}|\} \leq |\mathcal{X}| \end{array} \right\}. \tag{Q.1b}$$

We next show that if a rate R is achievable then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$ and $\nu > 0$, we start by upper bounding nR using standard techniques,

$$\begin{aligned} nR &= \mathbb{H}(M) \\ &= \mathbb{H}(M|K) \\ &\overset{(a)}{\leq} \mathbb{I}(M;Y^n|K) + n\epsilon_n \\ &= \sum_{i=1}^{n} \mathbb{I}(M;Y_i|Y^{i-1},K) + n\epsilon_n \\ &\leq \sum_{i=1}^{n} \mathbb{I}(M,K,Y^{i-1};Y_i) + n\epsilon_n \end{aligned}$$

$$
\begin{aligned}
&\overset{(b)}{\leq} \sum_{i=1}^{n} \mathbb{I}(M, K, S^{i-1}; Y_i) + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} \mathbb{I}(U_i; Y_i) + n\epsilon_n \\
&= n \sum_{i=1}^{n} \frac{1}{n} \mathbb{I}(U_i; Y_i) + n\epsilon_n \\
&= n \sum_{i=1}^{n} \mathbb{P}(T = i) \mathbb{I}(U_T; Y_T | T = i) + n\epsilon_n \\
&= n\mathbb{I}(U_T; Y_T | T) + n\epsilon_n \\
&\leq n\mathbb{I}(U_T, T; Y_T) + n\epsilon_n \\
&\overset{(d)}{=} n\mathbb{I}(U; Y) + n\epsilon_n \\
&\overset{(e)}{\leq} n\mathbb{I}(U; Y) + n\epsilon, \qquad \text{(Q.2)}
\end{aligned}
$$

where

- (a) follows from Fano's inequality;
- (b) follows since $(M, K, Y^{i-1}) - (M, K, S^{i-1}) - Y_i$, note that we also have $V_i - (M, K, S^{i-1}) - Y_i$, where $V_i \triangleq (M, K, Z^{i-1})$;
- (c) follows by defining $U_i \triangleq (M, K, S^{i-1})$;
- (d) follows by defining $U = (U_T, T)$ and $Y = Y_T$;
- (e) follows by defining $\epsilon \triangleq \max\{\epsilon_n, \lambda_n, \nu\}$, where we choose n large enough such that $\nu \geq \frac{\delta}{n}$.

Next, we lower bound nR as follows,

$$
\begin{aligned}
nR + R_K &\geq \mathbb{H}(M, K) \\
&\geq \mathbb{I}(M, K; Z^n)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \mathbb{I}(M, K; Z_i | Z^{i-1}) \\
&\overset{(a)}{\geq} \sum_{i=1}^{n} \mathbb{I}(M, K, Z^{i-1}; Z_i) - \delta \\
&\overset{(b)}{=} \sum_{i=1}^{n} \mathbb{I}(V_i; Z_i) - \delta \\
&= n \sum_{i=1}^{n} \mathbb{P}(T = i) \mathbb{I}(V_T; Z_T | T = i) - \delta \\
&= n \mathbb{I}(V_T; Z_T | T) - \delta \\
&\overset{(c)}{\geq} n \mathbb{I}(V_T, T; Z_T) - 2\delta \\
&\overset{(d)}{=} n \mathbb{I}(V; Z) - 2\delta
\end{aligned} \tag{Q.3}
$$

where

(a) and (c) follow from Lemma 6;

(b) follows from the definition of $V_i \triangleq (M, K, Z^{i-1})$, which is defined in the process of deriving (Q.2);

(d) follows by defining $V = (V_T, T)$ and $Z = Z_T$.

For any $\nu > 0$, choosing n large enough ensures that,

$$
R + \frac{R_K}{n} \geq \mathbb{I}(V; Z) - 2\nu. \tag{Q.4}
$$

Therefore,

$$
\begin{aligned}
R &\geq \mathbb{I}(V; Z) - 2\nu - \frac{R_K}{n} \\
&\geq \mathbb{I}(V; Z) - 2\nu - \lambda_n \\
&\geq \mathbb{I}(V; Z) - 3\epsilon,
\end{aligned} \tag{Q.5}
$$

where the last inequality follows since $\epsilon \triangleq \max\{\epsilon_n, \lambda_n, \nu\}$. To show that $\mathbb{D}(P_Z||Q_0) \leq \epsilon$, note that for n large enough

$$\begin{aligned}\mathbb{D}(P_Z||Q_0) = \mathbb{D}(P_{Z_T}||Q_0) &= \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} P_{Z_i}\middle|\middle|Q_0\right) \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \leq \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \frac{\delta}{n} \leq \nu \leq \epsilon. \end{aligned} \tag{Q.6}$$

Combining (Q.2) and (Q.5), and (Q.6) shows that $\forall \epsilon_n, \lambda_n, \nu > 0$, $R \leq \max\{a : a \in \mathcal{A}_\epsilon\}$. Therefore,

$$R \leq \max\left\{a : a \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon\right\}. \tag{Q.7}$$

Continuity at Zero: One can prove the continuity at zero of $\mathcal{A}_\epsilon$ by substituting $\min\{\mathbb{I}(U;Y)-\mathbb{I}(U;S), \mathbb{I}(U,V;Y)-\mathbb{I}(U;S|V)\}$ with $\mathbb{I}(U;Y)$ and $\mathbb{I}(V;Z)-\mathbb{I}(V;S)$ with $\mathbb{I}(V;Z)$ in the continuity at zero proof in Appendix N and following the exact same arguments.

APPENDIX R

PROOF OF THEOREM 16

We adopt a block-Markov encoding scheme in which B independent messages are transmitted over B channel blocks each of length r, such that $n = rB$. The warden's observation is $Z^n = (Z_1^r, \ldots, Z_B^r)$, the distribution induced at the output of the warden is P_{Z^n}, the target output distribution is $Q_0^{\otimes n}$, and Equation (I.2), describing the distance between the two distributions, continues to hold. The random code generation is as follows.

Fix $P_X(x)$, $P_{V|S}(v|s)$, and $\epsilon_1 > \epsilon_2 > 0$ such that, $P_Z = Q_0$.

Codebook Generation for Keys: For each block $j \in [\![1, B]\!]$, let $C_1^{(r)} \triangleq \left\{V^r(\ell_j)\right\}_{\ell_j \in \mathcal{L}}$, where $\mathcal{L} \triangleq [\![1, 2^{r\tilde{R}}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_V^{\otimes r}$, where $P_V = \sum_{s \in \mathcal{S}} Q_S(s) P_{V|S}(v|s)$. We denote a realization of $C_1^{(r)}$ by $\mathcal{C}_1^{(r)} \triangleq \left\{v^r(\ell_j)\right\}_{\ell_j \in \mathcal{L}}$. Partition the set of indices $\ell_j \in [\![1, 2^{r\tilde{R}}]\!]$ into bins $\mathcal{B}(t)$, $t \in [\![1, 2^{rR_T}]\!]$ by using function $\varphi : V^r(\ell_j) \mapsto [\![1, 2^{rR_T}]\!]$ through random binning by choosing the value of $\varphi(v^r(\ell_j))$ independently and uniformly at random for every $v^r(\ell_j) \in \mathcal{V}^r$. For each block $j \in [\![1, B]\!]$, create a function $\Phi : V^r(\ell_j) \mapsto [\![1, 2^{rR_K}]\!]$ through random binning by choosing the value of $\Phi(v^r(\ell_j))$ independently and uniformly at random for every $v^r(\ell_j) \in \mathcal{V}^r$. The key $k_j = \Phi(v^r(\ell_j))$ obtained in block $j \in [\![1, B]\!]$ from the description of the CSI sequence $v^r(\ell_j)$ is used to assist the encoder in block $j + 2$.

Codebook Generation for Messages: For each block $j \in [\![1, B]\!]$, let $C_2^{(r)} \triangleq \left\{X^r(m_j, t_{j-1}, k_{j-2})\right\}_{(m_j, t_{j-1}, k_{j-2}) \in \mathcal{M} \times \mathcal{T} \times \mathcal{K}}$, where $\mathcal{M} \triangleq [\![1, 2^{rR}]\!]$, $\mathcal{T} \triangleq [\![1, 2^{rR_t}]\!]$, and $\mathcal{K} \triangleq [\![1, 2^{rR_k}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_X^{\otimes r}$. We denote a realization of $C_2^{(r)}$ by $\mathcal{C}_2^{(r)} \triangleq \left\{x^r(m_j, t_{j-1}, k_{j-2})\right\}_{(m_j, t_{j-1}, k_{j-2}) \in \mathcal{M} \times \mathcal{T} \times \mathcal{K}}$. Let, $C_r = \left\{C_1^{(r)}, C_2^{(r)}\right\}$ and $\mathcal{C}_r = \left\{\mathcal{C}_1^{(r)}, \mathcal{C}_2^{(r)}\right\}$. The indices (m_j, t_{j-1}, k_{j-2}) can be viewed as a two layer binning. We define an ideal PMF for codebook $\mathcal{C}_r$, as an approximate distribution to facilitate the analysis

$$\Gamma^{(\mathcal{C}_r)}_{M_j, T_{j-1}, K_{j-2}, L_j, X^r, V^r, S_j^r, Z_j^r, K_{j-1}, T_j, K_j}(m_j, t_{j-1}, k_{j-2}, \ell_j, \tilde{x}^r, \tilde{v}^r, s_j^r, z_j^r, k_{j-1}, t_j, k_j)$$

$$= 2^{-r(R+R_t+R_k+\tilde{R})} \mathbb{1}_{\{\tilde{x}^r = x^r(m_j, t_{j-1}, k_{j-2})\}} \mathbb{1}_{\{\tilde{v}^r = v^r(\ell_j)\}} P_{S|V}^{\otimes r}(s_j^r|\tilde{v}^r) W_{Z|X,S}^{\otimes r}(z_j^r|\tilde{x}^r, s_j^r)$$
$$\times 2^{-rR_k} \mathbb{1}_{\{t_j = \varphi(\tilde{v}^r)\}} \mathbb{1}_{\{k_j = \Phi(\tilde{v}^r)\}}, \tag{R.1}$$

where $W_{Z|X,S}$ is the marginal distribution of $W_{Y,Z|X,S}$ defined in Theorem 16 and

$$P_{S|V} = \frac{P_{S,V}(s,v)}{P_V(v)} = \frac{Q_S(s)P_{V|S}(v|s)}{\sum_{s\in\mathcal{S}} Q_S(s)P_{V|S}(v|s)}. \tag{R.2}$$

Encoding: We assume that the transmitter and the receiver have access to shared secret keys k_{-1} and k_0 for the first two blocks, but after the first two blocks they use the key that they generate from the CSI.

In the first block, to send the message m_1 according to k_{-1}, the encoder generates the index t_0 uniformly at random and computes $x^r(m_1, t_0, k_{-1})$ and transmits it over the channel. Note that, the index t_0 does not convey any useful information.

At the beginning of the second block, to generate a secret key shared between the transmitter and the receiver, the encoder generates the index ℓ_1 according to the following distribution with $j = 1$,

$$f(\ell_j|s_j^r) = \frac{P_{S|V}^{\otimes r}\big(s_j^r|v^r(\ell_j)\big)}{\sum\limits_{\ell'\in[\![1,2^{r\tilde{R}}]\!]} P_{S|V}^{\otimes r}\big(s_j^r|v^r(\ell'_j)\big)}, \tag{R.3}$$

where $P_{S|V}$ is defined in (R.2). Then generates the reconciliation index $t_1 = \varphi\big(v^n(\ell_1)\big)$; simultaneously, the transmitter generates a key $k_1 = \Phi\big(v^r(\ell_1)\big)$ from the description of the CSI of the first block $v^r(\ell_1)$ to be used in the next block. To transmit the message m_2 and the reconciliation index t_1 according to the key k_0, the encoder computes $x^r(m_2, t_1, k_0)$ and transmits it over the channel.

In block $j \in [\![3, B]\!]$, the encoder first selects the index ℓ_{j-1} based s_{j-1}^r by using the likelihood encoder described in (R.3) and then generates the reconciliation index $t_{j-1} = \varphi\big(v^r(\ell_{j-1})\big)$; simultaneously the transmitter generates a key $k_{j-1} = \Phi\big(v^r(\ell_{j-1})\big)$ from the description of the CSI of the block $j-1$, $v^r(\ell_{j-1})$, to be used in the next block. Then to

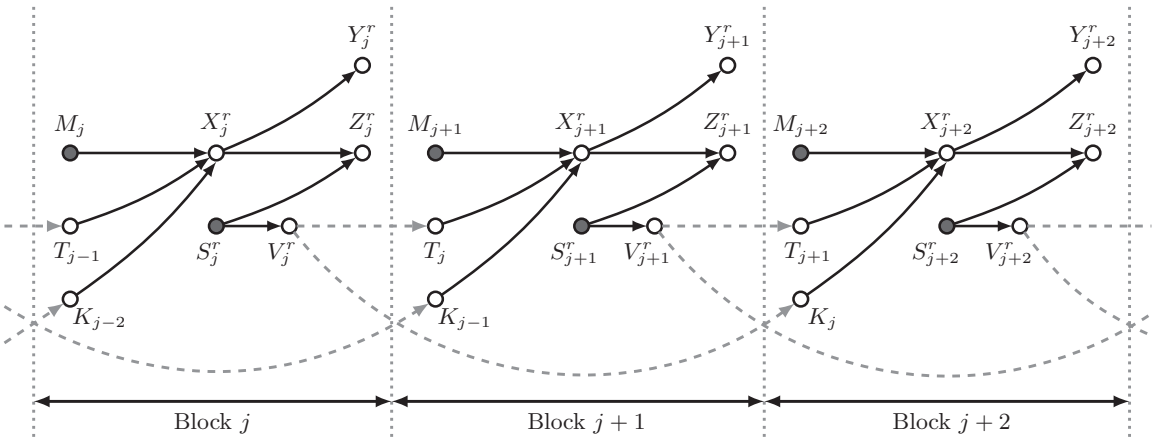

Figure R.1. Functional dependence graph for the block-Markov encoding scheme

send the message m_j and the reconciliation index t_{j-1} according to the generated key k_{j-2} from the previous block, the encoder computes $x^r(m_j, t_{j-1}, k_{j-2})$ and transmits it over the channel.

Define

$$\begin{aligned}
&\Upsilon^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2},X^r,S_j^r,L_j,V^r,Z_j^r,K_{j-1},T_j,K_j}(m_j, t_{j-1}, k_{j-2}, \tilde{x}^r, s_j^r, \ell_j, v^r, z_j^r, k_{j-1}, t_j, k_j) \\
&\triangleq 2^{-r(R+R_t+R_k)} \mathbb{1}_{\{\tilde{x}^r = x^r(m_j, t_{j-1}, k_{j-2})\}} Q_S^{\otimes r}(s_j^r) f(\ell_j | s_j^r) \mathbb{1}_{\{\tilde{v}^r = v^r(\ell_j)\}} \\
&\quad \times W_{Z|X,S}^{\otimes r}(z_j^r | \tilde{x}^r, s_j^r) 2^{-rR_k} \mathbb{1}_{\{t_j = \sigma(v^r(\ell_j))\}} \mathbb{1}_{\{k_j = \Phi(v^r(\ell_j))\}}. \qquad \text{(R.4)}
\end{aligned}$$

For a fixed codebook $\mathcal{C}_r$, the induced joint distribution by our code design (i.e. $P^{(\mathcal{C}_r)}$) satisfies

$$\mathbb{D}\Big(P^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2},X^r,S_j^r,L_j,V^r,Z_j^r,K_{j-1},T_j,K_j} || \Upsilon^{(\mathcal{C}_n)}_{M_j,T_{j-1},K_{j-2},X^r,S_j^r,L_j,V^r,Z_j^r,K_{j-1},T_j,K_j}\Big) \leq \epsilon. \qquad \text{(R.5)}$$

This intermediate distribution $\Upsilon^{(\mathcal{C}_r)}$ approximates the true distribution $P^{(\mathcal{C}_r)}$ and will be used in the sequel for bounding purposes. Expression (R.5) holds because the main difference between $P^{(\mathcal{C}_r)}$ and $\Upsilon^{(\mathcal{C}_r)}$ is that the keys K_{j-2}, K_{j-1} and the reconciliation index T_{j-1} are assumed to be uniformly distributed in $\Upsilon^{(\mathcal{C}_r)}$, which are made (arbitrarily) nearly uniform in $P^{(\mathcal{C}_r)}$ with appropriate control of rate as in (R.11) and (R.16).

Covert Analysis: We now show that this coding scheme guarantees that $\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n})] \underset{n\to\infty}{\longrightarrow} 0$, where C_n is the set of all the codebooks for all blocks, and

$$Q_Z(\cdot) = \sum_{x\in\mathcal{X}}\sum_{v\in\mathcal{V}}\sum_{s\in\mathcal{S}} P_X(x)P_V(v)P_{S|V}(s|v)W_{Z|X,S}(\cdot|x,s), \tag{R.6}$$

such that $Q_S(\cdot) = \sum_{v\in\mathcal{V}} P_V(v)P_{S|V}(\cdot|v)$. Then, we choose P_X, P_V, and $P_{S|V}$ such that it satisfies $Q_Z = Q_0$. Similar to (K.10) by using functional dependence graph depicted in Fig. R.1 one can show that,

$$\mathbb{D}(P_{Z^n}^{(\mathcal{C}_r)}||Q_Z^{\otimes n}) \le 2\sum_{j=1}^{B}\mathbb{D}(P_{Z_j^r,K_{j-1},T_j,K_j}^{(\mathcal{C}_r)}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}). \tag{R.7}$$

To bound the RHS of (R.7) by using Lemma 1 and the triangle inequality we have,

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}_r}||P_{Z_j^r,K_{j-1},T_j,K_j|C_r} - Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}||_1\\
&\le \mathbb{E}_{C_r}||P_{Z_j^r,K_{j-1},T_j,K_j|C_r} - \Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||_1 + \mathbb{E}_{C_r}||\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r} - Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}||_1\\
&\le \mathbb{E}_{C_r}||P_{Z_j^r,K_{j-1},T_j,K_j|C_r} - \Upsilon_{Z_j^r,K_{j-1},T_j,K_j|C_r}||_1 + \mathbb{E}_{C_r}||\Upsilon_{Z_j^r,K_{j-1},T_j,K_j|C_r} - \Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||_1\\
&\quad + \mathbb{E}_{C_r}||\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r} - Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}||_1.
\end{aligned} \tag{R.8}$$

From (R.5) and the monotonicity of KL-divergence the first term on the RHS of (R.8) goes to zero when n grows. To bound the second term on the RHS of (R.8) for a fixed codebook $\mathcal{C}_r$, we have,

$$\Gamma_{M_j,T_{j-1},K_{j-2}}^{(\mathcal{C}_r)} = 2^{-r(R+R_t+R_k)} = \Upsilon_{M_j,T_{j-1},K_{j-2}}^{(\mathcal{C}_r)}, \tag{R.9a}$$

$$\Gamma_{X^r|M_j,T_{j-1},K_{j-2},S_j^r}^{(\mathcal{C}_r)} = \mathbb{1}_{\{\tilde{x}^r=x^r(m_j,t_{j-1},k_{j-2})\}} = \Upsilon_{X^r|M_j,T_{j-1},K_{j-2},S_j^r}^{(\mathcal{C}_r)}, \tag{R.9b}$$

$$\Gamma_{L_j|M_j,T_{j-1},K_{j-2},S_j^r,X^r}^{(\mathcal{C}_r)} = f(\ell_j|s_j^r) = \Upsilon_{L_j|M_j,T_{j-1},K_{j-2},S_j^r,X^r}^{(\mathcal{C}_r)}, \tag{R.9c}$$

$$\Gamma_{V^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,X^r}^{(\mathcal{C}_r)} = \mathbb{1}_{\{\tilde{v}^r=v^r(\ell_j)\}} = \Upsilon_{V^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,X^r}^{(\mathcal{C}_r)}, \tag{R.9d}$$

$$\Gamma_{Z_j^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,X^r,V^r}^{(\mathcal{C}_r)} = W_{Z|X,S}^{\otimes r} = \Upsilon_{Z_j^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,X^r,V^r}^{(\mathcal{C}_r)}, \tag{R.9e}$$

$$\Gamma_{K_{j-1}|M_j,T_{j-1},K_{j-2},S_j^r,L_j,X^r,V^r,Z_j^r}^{(\mathcal{C}_r)} = 2^{-rR_k} = \Upsilon_{K_{j-1}|M_j,T_{j-1},K_{j-2},S_j^r,L_j,X^r,V^r,Z_j^r}^{(\mathcal{C}_r)}, \tag{R.9f}$$

$$\Gamma^{(\mathcal{C}_r)}_{T_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,X^r,V^r,Z_j^r,K_{j-1}} = \mathbb{1}_{\{t_j=\sigma(v^r(\ell_j))\}} = \Upsilon^{(\mathcal{C}_r)}_{T_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,X^r,V^r,Z_j^r,K_{j-1}}, \tag{R.9g}$$

$$\Gamma^{(\mathcal{C}_r)}_{K_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,X^r,V^r,Z_j^r,K_{j-1},T_j} = \mathbb{1}_{\{k_j=\Phi(v^r(\ell_j))\}} = \Upsilon^{(\mathcal{C}_r)}_{K_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,X^r,V^r,Z_j^r,K_{j-1},T_j}, \tag{R.9h}$$

where (R.9c) follows from (R.3). Hence,

$$\begin{aligned}
&\mathbb{E}_{C_r}||\Upsilon_{Z_j^r,K_{j-1},T_j,K_j|C_r} - \Gamma_{Z_j^r,K_{j-1},T_j,K_j}||_1 \\
&\leq \mathbb{E}_{C_r}||\Upsilon_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,X^r,V^r,Z_j^r,K_{j-1},T_j,K_j|C_r} - \Gamma_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,X^r,V^r,Z_j^r,K_{j-1},T_j,K_j|C_r}||_1 \\
&\overset{(a)}{=} \mathbb{E}_{C_r}||\Upsilon_{M_j,T_{j-1},K_{j-2},S_j^r|C_r} - \Gamma_{M_j,T_{j-1},K_{j-2},S_j^r|C_r}||_1 \\
&\overset{(b)}{=} \mathbb{E}_{C_r}||Q_S^{\otimes r} - \Gamma_{S_j^r|M_j=1,T_{j-1}=1,K_{j-2}=1,C_r}||_1,
\end{aligned} \tag{R.10}$$

where (a) follows from (R.9b)-(R.9h) and (b) follows from the symmetry of the codebook construction with respect to M_j, T_{j-1}, and K_{j-2} and (R.9a). Based on the soft covering lemma [77, Corollary VII.5] the RHS of (R.10) vanishes if

$$\tilde{R} > \mathbb{I}(S;V). \tag{R.11}$$

We now proceed to bound the third term on the RHS of (R.8). First, consider the marginal

$$\begin{aligned}
\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}(z_j^r,k_{j-1},t_j,k_j) &= \sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{s_j^r}\frac{1}{2^{r(R+R_t+R_k+\tilde{R})}}P_{S|V}^{\otimes r}\big(s_j^r|V^r(\ell_j)\big) \\
&\quad\times W_{Z|X,S}^{\otimes r}\big(z_j^r|X^r(m_j,t_{j-1},k_{j-2}),s_j^r\big)2^{-rR_k}\mathbb{1}_{\{t_j=\sigma(V^r(\ell_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\ell_j))\}} \\
&= \sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{1}{2^{r(R+R_t+R_k+\tilde{R})}}W_{Z|X,V}^{\otimes r}\big(z_j^r|X^r(m_j,t_{j-1},k_{j-2}),V^r(\ell_j)\big)2^{-rR_k} \\
&\quad\times \mathbb{1}_{\{t_j=\sigma(V^r(\ell_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\ell_j))\}},
\end{aligned} \tag{R.12}$$

where $W_{Z|X,V}(z|x,v) = \sum_{s\in\mathcal{S}}P_{S|V}(s|v)W_{Z|X,V}(z|x,v)$. To bound the third term on the RHS of (R.8) by using Pinsker's inequality, it is sufficient to bound $\mathbb{E}_{C_r}[\mathbb{D}(\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j})]$ as follows,

$$\mathbb{E}_{C_r}[\mathbb{D}(\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j})]$$

$$= \mathbb{E}_{C_r}\Big[\sum_{z_j^r,k_{j-1},t_j,k_j} \Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}(z_j^r,k_{j-1},t_j,k_j) \log\left(\frac{\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}(z_j^r,k_{j-1},t_j,k_j)}{Q_Z^{\otimes r}(z_j^r)Q_{K_{j-1}}(k_{j-1})Q_{T_j}(t_j)Q_{K_j}(k_j)}\right)\Big]$$

$$= \mathbb{E}_{C_r}\Bigg[\sum_{(z_j^r,k_{j-1},t_j,k_j)} \sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j} \frac{1}{2^{r(R+R_t+2R_k+\tilde{R})}} W_{Z|X,V}^{\otimes r}\big(z_j^r|X^r(m_j,t_{j-1},k_{j-2}),V^r(\ell_j)\big)$$

$$\times \mathbb{1}_{\{t_j=\varphi(V^r(\ell_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\ell_j))\}}$$

$$\times \log\left(\frac{\sum_{\tilde{m}_j}\sum_{\tilde{t}_{j-1}}\sum_{\tilde{k}_{j-2}}\sum_{\tilde{\ell}_j} W_{Z|X,V}^{\otimes r}\big(z_j^r|X^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2}),V^r(\tilde{\ell}_j)\big)\mathbb{1}_{\{t_j=\varphi(V^r(\tilde{\ell}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{\ell}_j))\}}}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\right)\Bigg]$$

$$\overset{(a)}{\leq} \sum_{(z_j^r,k_{j-1},t_j,k_j)} \sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j} \frac{1}{2^{r(R+R_t+2R_k+\tilde{R})}}$$

$$\times \sum_{\left(x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)} \Gamma_{X^r,V^r,Z_j^r}^{\otimes r}\big(x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\big)$$

$$\times \mathbb{E}_{\varphi(v^r(\ell_j))}\big[\mathbb{1}_{\{t_j=\varphi(v^r(\ell_j))\}}\big] \times \mathbb{E}_{\Phi(v^r(\ell_j))}\big[\mathbb{1}_{\{k_j=\Phi(v^r(\ell_j))\}}\big]$$

$$\times \log \mathbb{E}_{\substack{\setminus(m_j,t_{j-1},k_{j-2},\ell_j),\\ \setminus(\varphi(v^r(\ell_j)),\Phi(v^r(\ell_j)))}}\Bigg[\frac{1}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}$$

$$\times \sum_{\tilde{m}_j}\sum_{\tilde{t}_{j-1}}\sum_{\tilde{k}_{j-2}}\sum_{\tilde{\ell}_j} W_{Z|X,V}^{\otimes r}\big(z_j^r|X^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2}),V^r(\tilde{\ell}_j)\big)\mathbb{1}_{\{t_j=\varphi(V^r(\tilde{\ell}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{\ell}_j))\}}\Bigg]$$

$$\overset{(b)}{\leq} \sum_{(z_j^r,k_{j-1},t_j,k_j)} \sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j} \frac{1}{2^{r(R+R_t+2R_k+\tilde{R})}}$$

$$\times \sum_{\left(x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)} \Gamma_{X^r,V^r,Z_j^r}^{\otimes r}\big(x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\big)$$

$$\times \frac{1}{2^{r(R_T+R_K)}} \log \frac{1}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\Bigg(W_{Z|X,V}^{\otimes r}\big(z_j^r|x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\big)$$

$$+ \mathbb{E}_{\setminus(m_j,t_{j-1},k_{j-2})}\Bigg[\sum_{(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2})\neq(m_j,t_{j-1},k_{j-2})} W_{Z|X,V}^{\otimes r}\big(z_j^r|X^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2}),v^r(\ell_j)\big)\Bigg]$$

$$+ \mathbb{E}_{\substack{\setminus\ell_j,\\ \setminus(\varphi(v^r(\ell_j)),\Phi(v^r(\ell_j)))}}\Bigg[\sum_{\tilde{\ell}_j\neq\ell_j} W_{Z|X,V}^{\otimes r}\big(z_j^r|x^r(m_j,t_{j-1},k_{j-2}),V^r(\tilde{\ell}_j)\big)\mathbb{1}_{\{t_j=\varphi(V^r(\tilde{\ell}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{\ell}_j))\}}\Bigg]$$

$$
\begin{aligned}
&+\mathbb{E}_{\substack{\setminus(m_j,t_{j-1},k_{j-2},\ell_j),\\ \setminus(\varphi(v^r(\ell_j)),\Phi(v^r(\ell_j)))}}\Bigg[\sum_{\tilde{\ell}_j\neq\ell_j}\sum_{(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2})\neq(m_j,t_{j-1},k_{j-2})} W_{Z|X,V}^{\otimes r}\big(z_j^r|X^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2}),V^r(\tilde{\ell}_j)\big)\\
&\times \mathbb{1}_{\{t_j=\varphi(V^r(\tilde{\ell}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{\ell}_j))\}}\Bigg]\Bigg)\\
&\overset{(c)}{\leq}\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{1}{2^{r(R+R_t+2R_k+\tilde{R})}}\\
&\times\sum_{\left(x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}\Gamma_{X^r,V^r,Z_j^r}^{\otimes r}\big(x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\big)\\
&\times\frac{1}{2^{r(R_T+R_K)}}\log\Bigg(\frac{W_{Z|X,V}^{\otimes r}\big(z_j^r|x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\big)}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&+\sum_{(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2})\neq(m_j,t_{j-1},k_{j-2})}\frac{W_{Z|V}^{\otimes r}\big(z_j^r|v^r(\ell_j)\big)}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&+\sum_{\tilde{\ell}_j\neq\ell_j}\frac{W_{Z|X}^{\otimes r}\big(z_j^r|x^r(m_j,t_{j-1},k_{j-2})\big)}{2^{r(R+R_t+R_k+\tilde{R})}\times Q_Z^{\otimes r}(z_j^r)}+1\Bigg)\\
&\leq\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{1}{2^{r(R+R_t+2R_k+\tilde{R}+R_T+R_K)}}\\
&\times\sum_{\left(x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}\Gamma_{X^r,V^r,Z_j^r}^{\otimes r}\big(x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\big)\\
&\times\log\Bigg(\frac{W_{Z|X,V}^{\otimes r}\big(z_j^r|x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\big)}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|V}^{\otimes r}\big(z_j^r|v^r(\ell_j)\big)}{2^{r(\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&+\frac{W_{Z|X}^{\otimes r}\big(z_j^r|x^r(m_j,t_{j-1},k_{j-2})\big)}{2^{r(R+R_t+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\Bigg)\\
&\triangleq\Psi_1+\Psi_2,
\end{aligned}
\tag{R.13}
$$

where (a) follows from Jensen's inequality, (b) and (c) hold because $\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}\leq 1$. We define Ψ_1 and Ψ_2 as

$$
\Psi_1=\sum_{(k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{1}{2^{r(R+R_t+2R_k+R_T+R_K)}}
$$

$$
\begin{aligned}
&\times \sum_{\left(x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\right)\in\mathcal{T}_\epsilon^{(r)}} \Gamma^{\otimes r}_{X^r,V^r,Z_j^r}(x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r)\\
&\times \log\left(\frac{W^{\otimes r}_{Z|X,V}\left(z_j^r|x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}+\frac{W^{\otimes r}_{Z|V}\left(z_j^r|v^r(\ell_j)\right)}{2^{r(\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\right.\\
&\quad\left.+\frac{W^{\otimes r}_{Z|X}\left(z_j^r|x^r(m_j,t_{j-1},k_{j-2})\right)}{2^{r(R+R_t+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\right)\\
&\le \log\left(\frac{2^{r(R_T+R_K)}\times 2^{-r(1-\epsilon)\mathbb{H}(Z|X,V)}}{2^{r(R+R_t+R_k+\tilde{R})}\times 2^{-r(1+\epsilon)\mathbb{H}(Z)}}\right.\\
&\quad\left.+\frac{2^{r(R_T+R_K)}\times 2^{-r(1-\epsilon)\mathbb{H}(Z|V)}}{2^{r\tilde{R}}\times 2^{-r(1+\epsilon)\mathbb{H}(Z)}}+\frac{2^{-r(1-\epsilon)\mathbb{H}(Z|X)}}{2^{r(R+R_t+R_k)}\times 2^{-r(1+\epsilon)\mathbb{H}(Z)}}+1\right) \qquad \text{(R.14)}
\end{aligned}
$$

$$
\begin{aligned}
\Psi_2 = &\sum_{(k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{1}{2^{r(R+R_t+2R_k+R_T+R_K)}}\\
&\times \sum_{\left(x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\right)\notin\mathcal{T}_\epsilon^{(r)}} \Gamma^{\otimes r}_{X^r,V^r,Z_j^r}(x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r)\\
&\times \log\left(\frac{W^{\otimes r}_{Z|X,V}\left(z_j^r|x^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}+\frac{W^{\otimes r}_{Z|V}\left(z_j^r|v^r(\ell_j)\right)}{2^{r(\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\right.\\
&\quad\left.+\frac{W^{\otimes r}_{Z|X}\left(z_j^r|x^r(m_j,t_{j-1},k_{j-2})\right)}{2^{r(R+R_t+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\right)\\
&\le 2|V||X||Z|e^{-r\epsilon^2\mu_{V,X,Z}}r\log\Big(\frac{3}{\mu_Z}+1\Big). \qquad \text{(R.15)}
\end{aligned}
$$

In (R.15) $\mu_{X,V,Z}=\min\limits_{(x,v,z)\in(\mathcal{X},\mathcal{V},\mathcal{Z})}P_{X,V,Z}(x,v,z)$ and $\mu_Z=\min\limits_{z\in\mathcal{Z}}P_Z(z)$. When $r\to\infty$ then $\Psi_2\to 0$ and Ψ_1 goes to zero when r grows if

$$R+R_t+R_k+\tilde{R}-R_T-R_K>\mathbb{I}(X,V;Z), \qquad \text{(R.16a)}$$

$$\tilde{R}-R_T-R_K>\mathbb{I}(V;Z), \qquad \text{(R.16b)}$$

$$R+R_t+R_k>\mathbb{I}(X;Z). \qquad \text{(R.16c)}$$

Decoding and Error Probability Analysis: At the end of the block $j\in[\![1,B]\!]$, using its knowledge of the key k_{j-2} generated from the block $j-2$, the receiver finds a unique pair $(\hat{m}_j,\hat{t}_{j-1})$ such that $\left(x^r(\hat{m}_j,\hat{t}_{j-1},k_{j-2}),y_j^r\right)\in\mathcal{T}_\epsilon^{(r)}$. According to the law of large numbers

and the packing lemma probability of error vanishes when r grows if [85],

$$R + R_t < \mathbb{I}(X;Y). \tag{R.17}$$

We now analyze the probability of error at the encoder and the decoder for key generation. Let (L_{j-1}, T_{j-1}) denote the chosen indices at the encoder and $\hat{L}_{j-1}$ and $\hat{T}_{j-1}$ be the estimate of the index L_{j-1} and T_{j-1} at the decoder. At the end of block j, by decoding U_j^r, the receiver knows T_{j-1} and to find L_{j-1} we define the error event,

$$\mathcal{E} = \Big\{ \big(V_{j-1}^r(\hat{L}_{j-1}), S_{j-1}^r, X_{j-1}^r, Y_{j-1}^r\big) \notin \mathcal{T}_\epsilon^{(r)} \Big\}. \tag{R.18}$$

Also, consider the events

$$\mathcal{E}_1 = \Big\{ \big(V_{j-1}^r(\ell_{j-1}), S_{j-1}^r\big) \notin \mathcal{T}_\epsilon^{(r)} \text{ for all } \ell_{j-1} \in [\![1, 2^{r\tilde{R}}]\!] \Big\}, \tag{R.19a}$$

$$\mathcal{E}_2 = \Big\{ \big(V_{j-1}^r(L_{j-1}), S_{j-1}^r, X_{j-1}^r, Y_{j-1}^r\big) \notin \mathcal{T}_\epsilon^{(r)} \Big\}, \tag{R.19b}$$

$$\mathcal{E}_3 = \Big\{ \big(V_{j-1}^r(\tilde{\ell}_{j-1}), X_{j-1}^r, Y_{j-1}^r\big) \in \mathcal{T}_\epsilon^{(r)} \text{ for some } \ell_{j-1} \in \mathcal{B}(T_{j-1}), \tilde{\ell}_{j-1} \neq \ell_{j-1} \Big\}. \tag{R.19c}$$

By the union bound we have

$$P(\mathcal{E}) \leq P(\mathcal{E}_1) + P(\mathcal{E}_1^c \cap \mathcal{E}_2) + P(\mathcal{E}_3). \tag{R.20}$$

According to [84, Lemma 2] the first term on the RHS of (R.20) vanishes when r grows if we have (R.11). Following the steps in [85, Sec. 11.3.1], the last two terms on the RHS of (R.20) go to zero when r grows if we have

$$\tilde{R} > \mathbb{I}(V;S), \tag{R.21a}$$

$$\tilde{R} - R_t < \mathbb{I}(V;X,Y). \tag{R.21b}$$

Applying Fourier-Motzkin to (R.11), (R.16), (R.17), and (R.21) and remarking that the scheme requires $R_t + R_k \leq R_T + R_K$ results in the region in Theorem 16.

APPENDIX S

PROOF OF THEOREM 17

Fix $P_X(x)$ and $\epsilon_1 > \epsilon_2 > 0$ such that, $P_Z = Q_0$.

Codebook Generation: Let $C_n \triangleq \{X^n(m)\}_{m\in\mathcal{M}}$, where $\mathcal{M} \triangleq [\![1, 2^{nR}]\!]$, be a random codebook consisting of independent random sequences each generated according to $\prod_{i=1}^{n} P_X(x_i)$. We denote a realization of C_n by $\mathcal{C}_n \triangleq \{x^n(m)\}_{m\in\mathcal{M}}$.

Encoding: To send the message m, the encoder computes $x^n(m)$ and transmits it over the channel. For a fixed codebook $\mathcal{C}_n$, the induced joint distribution over the codebook is as follows

$$P^{(\mathcal{C}_n)}_{M,S^n,X^n,Z^n}(m, s^n, \tilde{x}^n, z^n) = 2^{-nR} Q_S^{\otimes n}(s^n) \mathbb{1}_{\{\tilde{x}^n = x^n(m)\}} W^{\otimes n}_{Z|X,S}(z^n|\tilde{x}^n, s^n). \tag{S.1}$$

Covert Analysis: We now show that this coding scheme guarantees that

$$\mathbb{E}_{C_n}\left[\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n})\right] \underset{n\to\infty}{\longrightarrow} 0, \tag{S.2}$$

where

$$Q_Z(\cdot) = \sum_{s\in\mathcal{S}} \sum_{x\in\mathcal{X}} Q_S(s) P_X(x) W_{Z|X,S}(\cdot|x, s). \tag{S.3}$$

Then we choose P_X such that it satisfies $Q_Z = Q_0$. Here, $P_{Z^n|C_n}$ is the marginal distribution of the joint distribution induced by our code design defined in (S.1) and is as follows,

$$P_{Z^n|C_n}(z^n) = \sum_m \sum_{s^n} 2^{-nR} Q_S^{\otimes n}(s^n) W^{\otimes n}_{Z|X,S}\big(z^n|x^n(m), s^n\big) \tag{S.4}$$

$$= \sum_m 2^{-nR} W^{\otimes n}_{Z|X}\big(z^n|x^n(m)\big), \tag{S.5}$$

where $W_{Z|X} = \sum_{s\in\mathcal{S}} Q_S(s) W_{Z|X,S}(z|x, s)$. By [92, Theorem 1] one can show that (S.2) holds if

$$R > \mathbb{I}(X; Z). \tag{S.6}$$

Decoding and Error Probability Analysis: Upon receiving y^n, the receiver finds a unique message $\hat{m}$ such that $\left(x^n(\hat{m}), y^n\right) \in \mathcal{T}_\epsilon^{(n)}$. According to the law of large numbers and the packing lemma, probability of error vanishes when n grows if [85],

$$R < \mathbb{I}(X;Y). \tag{S.7}$$

APPENDIX T

PROOF OF THEOREM 18

Consider any sequence of length-n codes for a state-dependent channel with CSI available strictly causally only at the transmitter such that $P_e^{(n)} \leq \epsilon_n$, $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \delta$, and $R_K/n \leq \lambda_n$ with $\lim_{n\to\infty} \epsilon_n = \lim_{n\to\infty} \lambda_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish.

Epsilon Rate Region: We first define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ that expands the region defined in (3.41) as follows,

$$\mathcal{A}_\epsilon \triangleq \left\{R \geq 0 : \exists P_{V,S,X,Y,Z} \in \mathcal{D}_\epsilon : R \leq \mathbb{I}(X;Y) + \epsilon\right\}, \tag{T.1a}$$

where

$$\mathcal{D}_\epsilon = \left\{ \begin{array}{l} P_{V,S,X,Y,Z} : \\ P_{V,S,X,Y,Z} = Q_S P_V P_{X|V} W_{Y,Z|X,S} \\ \mathbb{D}\left(P_Z \| Q_0\right) \leq \epsilon \\ \mathbb{I}(X;Y) \geq \mathbb{I}(V;Z) - 4\epsilon \\ |\mathcal{V}| \leq |\mathcal{X}| \end{array} \right\}. \tag{T.1b}$$

We next show that if a rate R is achievable then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$ and $\nu > 0$, we start by upper bounding nR using standard techniques.

$$\begin{aligned} nR &= \mathbb{H}(M) \\ &= \mathbb{H}(M|K) \\ &\overset{(a)}{\leq} \mathbb{I}(M;Y^n|K) + n\epsilon_n \\ &= \sum_{i=1}^{n} \mathbb{I}(M;Y_i|Y^{i-1},K) + n\epsilon_n \\ &\leq \sum_{i=1}^{n} \mathbb{I}(M,K,Y^{i-1};Y_i) + n\epsilon_n \end{aligned}$$

$$
\begin{aligned}
&\overset{(b)}{\leq} \sum_{i=1}^{n} \mathbb{I}(M, K, S^{i-1}; Y_i) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} \mathbb{I}(M, K, S^{i-1}, Z^{i-1}; Y_i) + n\epsilon_n \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} \mathbb{I}(X_i; Y_i) + n\epsilon_n \qquad \text{(T.2)} \\
&= n \sum_{i=1}^{n} \frac{1}{n} \mathbb{I}(X_i; Y_i) + n\epsilon_n \\
&= n \sum_{i=1}^{n} \mathbb{P}(T = i) \mathbb{I}(X_i; Y_i | T = i) + n\epsilon_n \\
&= n\mathbb{I}(X_T; Y_T | T) + n\epsilon_n \\
&\leq n\mathbb{I}(X_T, T; Y_T) + n\epsilon_n \\
&\overset{(d)}{=} n\mathbb{I}(X; Y) + n\epsilon_n \\
&\overset{(e)}{\leq} n\mathbb{I}(X; Y) + n\epsilon, \qquad \text{(T.3)}
\end{aligned}
$$

where

(a) follows from Fano's inequality;

(b) follows from the Markov chain $(M, K, Y^{i-1}) - (M, K, S^{i-1}) - Y_i$;

(c) follows since S_i is independent of $\big(M, K, S^{i-1}, Z^{i-1}, X_i(M, S^{i-1})\big)$ and therefore

$$\mathbb{I}(M, K, S^{i-1}, Z^{i-1}; Y_i | X_i) \leq \mathbb{I}(M, K, S^{i-1}, Z^{i-1}; Y_i | X_i, S_i) = 0, \qquad \text{(T.4)}$$

that is $(M, K, S^{i-1}, Z^{i-1}) - X_i - Y_i$ forms a Markov chain, which implies $V_i - X_i - Y_i$, where $V_i \triangleq (M, K, Z^{i-1})$, also forms a Markov chain;

(d) follows by defining $X = (X_T, T)$ and $Y = Y_T$.

(e) follows by defining $\epsilon \triangleq \max\{\epsilon_n, \lambda_n, \nu\}$, where we choose n large enough such that $\nu \geq \frac{\delta}{n}$.

Next, we lower bound nR as follows,

$$\begin{aligned} nR + R_K &\geq \mathbb{H}(M, K) \\ &\geq \mathbb{I}(M, K; Z^n) \\ &= \sum_{i=1}^{n} \mathbb{I}(M, K; Z_i | Z^{i-1}) \\ &\overset{(a)}{\geq} \sum_{i=1}^{n} \mathbb{I}(M, K, Z^{i-1}; Z_i) - \delta \\ &\overset{(b)}{=} \sum_{i=1}^{n} \big[\mathbb{I}(V_i; Z_i)\big] - \delta \\ &= n\mathbb{I}(V_T; Z_T | T) - \delta \\ &\overset{(c)}{\geq} n\mathbb{I}(V_T, T; Z_T) - 2\delta \\ &\overset{(d)}{=} n\mathbb{I}(V; Z) - 2\delta \end{aligned} \tag{T.5}$$

where

(a) and (c) follow from Lemma 6;

(b) follows by defining $V_i \triangleq (M, K, Z^{i-1})$ which is defined in the process of deriving (T.3);

(d) follows by defining $V = (V_T, T)$ and $Z = Z_T$.

For any $\nu > 0$, choosing n large enough ensures that,

$$R + \frac{R_K}{n} \geq \mathbb{I}(V; Z) - 2\nu. \tag{T.6}$$

Therefore,

$$\begin{aligned} R &\geq \mathbb{I}(V; Z) - 2\nu - \frac{R_K}{n} \\ &\geq \mathbb{I}(V; Z) - 2\nu - \lambda_n \\ &\geq \mathbb{I}(V; Z) - 3\epsilon, \end{aligned} \tag{T.7}$$

where the last inequality follows since $\epsilon \triangleq \max\{\epsilon_n, \lambda_n, \nu\}$. To show that $\mathbb{D}(P_Z||Q_0) \leq \epsilon$, note that for n large enough

$$\begin{aligned}\mathbb{D}(P_Z||Q_0) = \mathbb{D}(P_{Z_T}||Q_0) &= \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} P_{Z_i}\middle|\middle|Q_0\right) \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \leq \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \frac{\delta}{n} \leq \nu \leq \epsilon. \end{aligned} \tag{T.8}$$

Combining (T.3) and (T.7), and (T.8) shows that $\forall \epsilon_n, \nu > 0$, $R \leq \max\{a : a \in \mathcal{A}_\epsilon\}$. Therefore,

$$R \leq \max\left\{a : a \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon\right\}. \tag{T.9}$$

Continuity at Zero: One can prove the continuity at zero of $\mathcal{A}_\epsilon$ by substituting $\min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\}$ with $\mathbb{I}(X;Y)$ and $\mathbb{I}(V;Z) - \mathbb{I}(V;S)$ with $\mathbb{I}(V;Z)$ in the continuity at zero proof in Appendix N and following the exact same arguments.

APPENDIX U

PROOF OF THEOREM 19

Fix $P_S(s)$, $P_X(x)$, and $\epsilon > 0$ such that, $Q_Z = Q_0$.

Codebook Generation: Let $C_1^{(n)} \triangleq \left\{X^n(m)\right\}_{m\in\mathcal{M}_n}$, where $\mathcal{M}_n = [\![1, 2^{nR}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_X^{\otimes n}$. We denote a realization of $C_1^{(n)}$ by $\mathcal{C}_1^{(n)} \triangleq \left\{x^n(m)\right\}_{m\in\mathcal{M}_n}$.

Let $C_2^{(n)} \triangleq \left\{S^n(k)\right\}_{k\in\mathcal{K}_n}$, where $\mathcal{K}_n = [\![1, 2^{nR_K}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_S^{\otimes n}$. We denote a realization of $C_2^{(n)}$ by $\mathcal{C}_2^{(n)} \triangleq \left\{s^n(k)\right\}_{k\in\mathcal{K}_n}$.

Also, $C_n = \left\{C_1^{(n)}, C_2^{(n)}\right\}$ denotes a random codebook and $\mathcal{C}_n = \left\{\mathcal{C}_1^{(n)}, \mathcal{C}_2^{(n)}\right\}$ denotes a fixed codebook. The set of all possible values of C_n is denoted by $\mathfrak{C}_n$. The codebook construction described above induces the PMF $\lambda \in \mathcal{P}(\mathfrak{C}_n)$ over the codebook ensemble. For each $\mathcal{C}_n \in \mathfrak{C}_n$

$$\lambda(\mathcal{C}_n) = \prod_{m\in\mathcal{M}_n} P_X^{\otimes n}\big(x^n(m)\big) \prod_{k\in\mathcal{K}_n} P_S^{\otimes n}\big(s^n(k)\big). \tag{U.1}$$

Encoding: To send the message m, the transmitter computes $x^n(m)$ and transmits it over the channel. Also, given the key k, the jammer computes $s^n(k)$ and transmits it over the channel.

For a fixed codebook $\mathcal{C}_n$, the induced joint distribution is

$$P_{KMS^nX^nZ^n}^{(\mathcal{C}_n)}(k, m, \tilde{s}^n, \tilde{x}^n, z^n) = 2^{-n(R_K+R)} \mathbb{1}_{\left\{\tilde{s}^n = s^n(k)\right\} \cap \left\{\tilde{x}^n = x^n(m)\right\}} W_{Z|XS}^{\otimes n}(z^n|\tilde{x}^n, \tilde{s}^n). \tag{U.2}$$

Considering the random codebook generation, we have

$$P(\mathcal{C}_n, k, m, \tilde{s}^n, \tilde{x}^n, z^n) = \lambda(\mathcal{C}_n) P^{(\mathcal{C}_n)}(k, m, \tilde{s}^n, \tilde{x}^n, z^n), \tag{U.3}$$

where $\lambda \in \mathcal{P}(\mathfrak{C}_n)$ is defined in (U.1).

Covert Analysis: Henceforth, we use $P^{(\mathcal{C}_n)}$ when the codebook is fixed, and we use $P_{|C_n}$ when the codebook is random. Consider a scenario in which the jammer selects a codeword

from its codebook (i.e. $\mathcal{C}_2^{(n)}$) according to the key k, but the transmitter chooses the innocent sequence x_0^n as the channel input. For this scenario for a fixed codebook $\mathcal{C}_2^{(n)}$, the induced joint distribution is as follows

$$\Upsilon^{(\mathcal{C}_n)}_{KS^nZ^n}(k,\tilde{s}^n,z^n)=\frac{1}{2^{nR_K}}\mathbb{1}_{\{\tilde{s}^n=s^n(k)\}}W^{\otimes n}_{Z|X=x_0,S}\big(z^n|x_0^n,\tilde{s}^n\big). \tag{U.4}$$

Therefore, the distribution induced at the output of the warden is

$$\Upsilon^{(\mathcal{C}_n)}_{Z^n}(z^n)=\frac{1}{2^{nR_K}}\sum_{j=1}^{2^{nR_K}}W^{\otimes n}_{Z|X=x_0,S}\big(z^n|x_0^n,s^n(k)\big). \tag{U.5}$$

For this scenario if $R_K>\mathbb{I}_\Upsilon(S;Z)$ then according to soft covering lemma [83, Theorem 4] or [77, Corollary VII.4], we have

$$\mathbb{E}_{C_n}\mathbb{V}\big(\Upsilon_{Z^n|C_n},Q_0^{\otimes n}\big)\underset{n\to\infty}{\longrightarrow}0, \tag{U.6}$$

where

$$Q_0(\cdot)=\sum_{s\in\mathcal{S}}P_S(s)W_{Z|XS}\big(\,\cdot\,|x_0,s\big). \tag{U.7}$$

Note that if $R_K<\mathbb{I}_\Upsilon(S;Z)$ according to Shannon's channel coding theorem, the warden might be able to decode J, which reduces the problem to the point to point channel for which the covert rate will be zero.

From (U.2) the distribution induced at the output of the channel by our code design is as follows

$$P_{Z^n|C_n}(z^n)=\sum_{k=1}^{2^{nR_K}}\sum_{m=1}^{2^{nR}}2^{-n(R_K+R)}W^{\otimes n}_{Z|XS}(z^n|X^n(m),S^n(k)). \tag{U.8}$$

Let $Q_Z^{\otimes n}=\prod_{i=1}^{n}Q_Z$ where

$$Q_Z(\cdot)=\sum_{s\in\mathcal{S}}\sum_{x\in\mathcal{X}}P_S(s)P_{X|S}(x|s)W_{Z|XS}\big(\,\cdot\,|x,s\big). \tag{U.9}$$

One can show that $\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n})] \underset{n\to\infty}{\longrightarrow} 0$ if

$$R_K > \mathbb{I}_P(S;Z), \tag{U.10a}$$

$$R > \mathbb{I}_P(X;Z), \tag{U.10b}$$

$$R_K + R > \mathbb{I}_P(X,S;Z). \tag{U.10c}$$

Now, by using the triangle inequality we have

$$\mathbb{E}_{C_n}\mathbb{V}\big(P_{Z^n|C_n}, \Upsilon_{Z^n|C_n}\big) \leq \mathbb{E}_{C_n}\mathbb{V}\big(P_{Z^n|C_n}, Q_0^{\otimes n}\big) + \mathbb{E}_{C_n}\mathbb{V}\big(\Upsilon_{Z^n|C_n}, Q_0^{\otimes n}\big). \tag{U.11}$$

Using Pinsker inequality the first term on the RHS of (U.11) vanishes when n grows if (U.10) holds, and we choose P_S and P_X such that $Q_Z = Q_0$. Also, from (U.6) the second term on the RHS of (U.11) vanishes when n grows. Therefore, from (U.11) and using Lemma 1 we have $\mathbb{E}_{C_n}\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n}) \underset{n\to\infty}{\longrightarrow} 0$.

Decoding and Error Probability Analysis: By following the same steps as in [37], the probability of error vanishes when n grows if

$$R < \mathbb{I}_P(X;Y|S). \tag{U.12}$$

APPENDIX V

PROOF OF THEOREM 20

To prove the upper bound for the case that the jammer knows in which blocks the transmitter is communicating with the receiver and has an unlimited source of local randomness and transmits an i.i.d. sequence when communication is not happening, consider any sequence of codes with length n such that $P_e^{(n)} \leq \epsilon_n$ and $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \delta$ where $\lim_{n\to\infty} \epsilon_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish.

Epsilon Rate Region: We first define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ that expands the region defined in (4.6) *as follows.*

$$\mathcal{A}_\epsilon = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists P_{QSXYZ} \in \mathcal{D}_\epsilon : \\ R \leq \mathbb{I}(X;Y|S,Q) + \epsilon \\ R_K \geq \max\{\mathbb{I}(S;Z|Q), \mathbb{I}(X,S;Z|Q) - \mathbb{I}(X;Y|S,Q)\} - 2\epsilon \end{array} \right\}, \tag{V.1a}$$

where

$$\mathcal{D}_\epsilon = \left\{ \begin{array}{l} P_{QSXYZ} : \\ P_{QSXYZ} = P_Q P_{S|Q} P_{X|Q} W_{YZ|XS} \\ \mathbb{I}(X;Y|S,Q) \geq \mathbb{I}(X;Z|Q) - 2\epsilon \\ \mathbb{D}\left(P_Z \| Q_0\right) \leq \epsilon \end{array} \right\}. \tag{V.1b}$$

We next show that if a rate R is achievable, then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$, we start by upper bounding nR using standard techniques.

$$\begin{aligned} nR &= \mathbb{H}(M) \\ &\overset{(a)}{\leq} \mathbb{H}(M|S^n) - \mathbb{H}(M|Y^n, S^n) + n\epsilon_n \\ &= \mathbb{I}(M; Y^n|S^n) + n\epsilon_n \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \mathbb{I}(M; Y_i | Y^{i-1}, S^n) + n\epsilon_n \\
&= \sum_{i=1}^{n} [\mathbb{H}(Y_i | Y^{i-1}, S^n) - \mathbb{H}(Y_i | M, Y^{i-1}, S^n)] + n\epsilon_n \\
&\overset{(b)}{\leq} \sum_{i=1}^{n} [\mathbb{H}(Y_i | S_i) - \mathbb{H}(Y_i | M, Y^{i-1}, S^n, X^n)] + n\epsilon_n \\
&= \sum_{i=1}^{n} [\mathbb{H}(Y_i | S_i) - \mathbb{H}(Y_i | S_i, X_i)] + n\epsilon_n \\
&= \sum_{i=1}^{n} \mathbb{I}(X_i; Y_i | S_i) + n\epsilon_n \\
&= n \sum_{i=1}^{n} \frac{1}{n} \mathbb{I}(X_i; Y_i | S_i) + n\epsilon_n \\
&= n \sum_{i=1}^{n} \mathbb{P}(Q = i) \mathbb{I}(X_Q; Y_Q | S_Q, Q = i) + n\epsilon_n \\
&= n\mathbb{I}(X_Q; Y_Q | S_Q, Q) + n\epsilon_n \\
&\overset{(c)}{=} n\mathbb{I}(X; Y | S, Q) + n\epsilon_n \qquad \text{(V.2)} \\
&\overset{(d)}{\leq} n\mathbb{I}(X; Y | S, Q) + n\epsilon \qquad \text{(V.3)}
\end{aligned}
$$

where

(a) *follows from Fano's inequality and independence of* M *from* S^n*;*

(b) *holds because conditioning does not increase entropy;*

(c) *follows by defining* $X_Q = X$*,* $Y_Q = Y$*, and* $S_Q = S$*;*

(d) *follows by defining* $\epsilon \triangleq \max\{\epsilon_n, \nu \geq \frac{\delta}{n}\}$.

We also have,

$$
\begin{aligned}
n(R + R_K) &\geq \mathbb{H}(M, K) \\
&= \mathbb{H}(M, K)
\end{aligned}
$$

$$
\begin{aligned}
&\geq \mathbb{I}(M, K; Z^n) \\
&\overset{(a)}{=} \mathbb{I}(M, K, X^n, S^n; Z^n) \\
&\geq \mathbb{I}(X^n, S^n; Z^n) \\
&= \sum_{i=1}^{n} [\mathbb{H}(Z_i|Z^{i-1}) - \mathbb{H}(Z_i|Z^{i-1}, X^n, S^n)] \\
&\overset{(b)}{\geq} \sum_{i=1}^{n} [\mathbb{H}(Z_i) - \mathbb{H}(Z_i|X_i, S_i)] - \delta \\
&= \sum_{i=1}^{n} \mathbb{I}(X_i, S_i; Z_i) - \delta \\
&= n \sum_{i=1}^{n} \frac{1}{n} \mathbb{I}(X_i, S_i; Z_i) - \delta \\
&= n \sum_{i=1}^{n} \mathbb{P}(Q = i) \mathbb{I}(X_Q, S_Q; Z_Q|Q = i) - \delta \\
&= n\mathbb{I}(X_Q, S_Q; Z_Q|Q) - \delta \\
&\overset{(c)}{=} n\mathbb{I}(X, S; Z|Q) - \delta, \qquad \text{(V.4)}
\end{aligned}
$$

where

(a) *follows because* X^n *is a function of* M *and* S^n *is a function of* K*;*

(b) *follows from [80, Lemma 3];*

(c) *follows by defining* $X_Q = X$*,* $Z_Q = Z$*, and* $S_Q = S$*.*

From (V.4) *for any* $\nu > 0$*, choosing* n *large enough ensures that*

$$
\begin{aligned}
R + R_K &\geq \mathbb{I}(X, S; Z|Q) - \nu, \\
&\geq \mathbb{I}(X, S; Z|Q) - \epsilon, \qquad \text{(V.5)}
\end{aligned}
$$

where the last equality follows from the definition of $\epsilon \triangleq \max\{\epsilon_n, \nu\}$*. We now have,*

$$nR_K \geq \mathbb{H}(K)$$

$$
\begin{aligned}
&\geq \mathbb{I}(K; Z^n) \\
&\overset{(a)}{=} \mathbb{I}(K, S^n; Z^n) \\
&\geq \mathbb{I}(S^n; Z^n) \\
&= \sum_{i=1}^{n} [\mathbb{H}(Z_i|Z^{i-1}) - \mathbb{H}(Z_i|Z^{i-1}, S^n)] \\
&\overset{(b)}{\geq} \sum_{i=1}^{n} [\mathbb{H}(Z_i) - \mathbb{H}(Z_i|S_i)] - \delta \\
&= \sum_{i=1}^{n} \mathbb{I}(S_i; Z_i) - \delta \\
&= n \sum_{i=1}^{n} \frac{1}{n} \mathbb{I}(S_i; Z_i) - \delta \\
&= n \sum_{i=1}^{n} \mathbb{P}(Q = i) \mathbb{I}(S_Q; Z_Q|Q = i) - \delta \\
&= n\mathbb{I}(S_Q; Z_Q|Q) - \delta \\
&\overset{(d)}{=} n\mathbb{I}(S; Z|Q) - \delta
\end{aligned} \tag{V.6}
$$

where

(*a*) *follows because* S^n *is a function of* J*;*

(*b*) *follows from [80, Lemma 3] and the fact that conditioning does not increase the entropy;*

(*c*) *follows by defining* $Z_Q = Z$ *and* $S_Q = S$*.*

From (V.6) *for any* $\nu > 0$*, choosing* n *large enough ensures that*

$$
\begin{aligned}
R_K &\geq \mathbb{I}(S; Z|Q) - \nu, \\
&\geq \mathbb{I}(S; Z|Q) - \epsilon,
\end{aligned} \tag{V.7}
$$

where the last equality follows from the definition of $\epsilon \triangleq \max\{\epsilon_n, \nu\}$*. Similarly, one can show that,*

$$
R \geq \mathbb{I}(X; Z|Q) - \epsilon, \tag{V.8}
$$

To show that $\mathbb{D}(P_Z||Q_0) \leq \epsilon$*, note that for* n *large enough*

$$\begin{aligned}\mathbb{D}(P_Z||Q_0) = \mathbb{D}(P_{Z_Q}||Q_0) &= \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} P_{Z_i}\middle|\middle|Q_0\right) \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \leq \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \frac{\delta}{n} \leq \nu \leq \epsilon. \end{aligned} \tag{V.9}$$

Continuity at Zero: The proof for continuity at zero is similar to that of Appendix N.

Combining (V.3), (V.5), (V.7), and (V.8) shows that $\forall \epsilon_n, \nu > 0$, $C_{\text{B-J}} \subseteq \mathcal{A}_\epsilon$.

APPENDIX W

PROOF OF THEOREM 21

Fix $P_S(s)$, $P_X(x)$, and $\epsilon > 0$ such that, $Q_Z = Q_0$.

Codebook Generation: Let $C_1^{(n)} \triangleq \left\{X^n(m)\right\}_{m\in\mathcal{M}_n}$, where $\mathcal{M}_n = [\![1, 2^{nR}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_X^{\otimes n}$. We denote a realization of $C_1^{(n)}$ by $\mathcal{C}_1^{(n)} \triangleq \left\{x^n(m)\right\}_{m\in\mathcal{M}_n}$.

Let $C_2^{(n)} \triangleq \left\{S^n(j)\right\}_{j\in\mathcal{J}_n}$, where $\mathcal{J}_n = [\![1, 2^{nR_J}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_S^{\otimes n}$. We denote a realization of $C_2^{(n)}$ by $\mathcal{C}_2^{(n)} \triangleq \left\{s^n(j)\right\}_{j\in\mathcal{J}_n}$.

Also, $C_n = \left\{C_1^{(n)}, C_2^{(n)}\right\}$ denotes a random codebook and $\mathcal{C}_n = \left\{\mathcal{C}_1^{(n)}, \mathcal{C}_2^{(n)}\right\}$ denotes a fixed codebook. The set of all possible values of C_n is denoted by $\mathfrak{C}_n$. The codebook construction described above induces the PMF $\lambda \in \mathcal{P}(\mathfrak{C}_n)$ over the codebook ensemble. For each $\mathcal{C}_n \in \mathfrak{C}_n$

$$\lambda(\mathcal{C}_n) = \prod_{m\in\mathcal{M}_n} P_X^{\otimes n}\big(x^n(m)\big) \prod_{j\in\mathcal{J}_n} P_S^{\otimes n}\big(s^n(j)\big). \tag{W.1}$$

Encoding: To send the message m, the transmitter computes $x^n(m)$ and transmits it over the channel when the transmitter is communicating the jammer transmits the innocent symbol s_0^n. When the transmitter is not communicating with the receiver, and it transmits the innocent sequence x_0^n, the jammer computes a codeword $s^n(j)$ from its codebook according to the local randomness j and transmits it over the channel.

For a fixed codebook $\mathcal{C}_n$, the induced joint distribution when the transmitter is communicating with the receiver is

$$P_{MX^nZ^n}^{(\mathcal{C}_n)}(m, \tilde{x}^n, z^n) = 2^{-nR} \mathbb{1}_{\left\{\tilde{x}^n = x^n(m)\right\}} W_{Z|XS=s_0}^{\otimes n}(z^n|\tilde{x}^n, s_0^n). \tag{W.2}$$

Considering the random codebook generation, we have

$$P(\mathcal{C}_n, k, m, \tilde{s}^n, \tilde{x}^n, z^n) = \lambda(\mathcal{C}_n) P^{(\mathcal{C}_n)}(k, m, \tilde{s}^n, \tilde{x}^n, z^n), \tag{W.3}$$

where $\lambda \in \mathcal{P}(\mathfrak{C}_n)$ is defined in (W.1).

Covert Analysis: Henceforth, we use $P^{(\mathcal{C}_n)}$ when the codebook is fixed, and we use $P_{\cdot|C_n}$ when the codebook is random. Consider a scenario in which the jammer selects a codeword from its codebook (i.e. $\mathcal{C}_2^{(n)}$) according to the local randomness j, but the transmitter chooses the innocent sequence x_0^n as the channel input. For this scenario for a fixed codebook $\mathcal{C}_2^{(n)}$, the induced joint distribution is as follows

$$\Upsilon^{(\mathcal{C}_n)}_{JS^nZ^n}(j,\tilde{s}^n,z^n) = 2^{-nR_J}\mathbb{1}_{\left\{\tilde{s}^n=s^n(j)\right\}}W^{\otimes n}_{Z|X=x_0,S}(z^n|x_0^n,\tilde{s}^n). \tag{W.4}$$

Therefore, the distribution induced at the output of the warden is

$$\Upsilon^{(\mathcal{C}_n)}_{Z^n}(z^n) = \frac{1}{2^{nR_J}}\sum_{j=1}^{2^{nR_J}} W^{\otimes n}_{Z|X=x_0,S}\big(z^n|x_0^n,s^n(j)\big). \tag{W.5}$$

Also, from (W.2) the distribution induced at the output of the channel by our code design is as follows

$$P_{Z^n|C_n}(z^n) = \sum_{m=1}^{2^{nR}} 2^{-nR} W^{\otimes n}_{Z|XS=s_0}(z^n|X^n(m),s_0^n). \tag{W.6}$$

Here, the goal is to show that $\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||\Upsilon_{Z^n|C_n})] \underset{n\to\infty}{\longrightarrow} 0$. Now, by using the triangle inequality and Lemma 1 we have

$$\mathbb{E}_{C_n}\mathbb{V}\big(P_{Z^n|C_n},\Upsilon_{Z^n|C_2^{(n)}}\big) \le \mathbb{E}_{C_n}\mathbb{V}\big(P_{Z^n|C_n},Q_1^{\otimes n}\big) + \mathbb{E}_{C_n}\mathbb{V}\big(\Upsilon_{Z^n|C_2^{(n)}},Q_2^{\otimes n}\big), \tag{W.7}$$

where

$$Q_1(\cdot) = \sum_{x\in\mathcal{X}} P_X(x)W_{Z|XS}\big(\,\cdot\,|x,s_0\big), \tag{W.8}$$

$$Q_2(\cdot) = \sum_{s\in\mathcal{S}} P_S(s)W_{Z|XS}\big(\,\cdot\,|x_0,s\big). \tag{W.9}$$

Therefore, we should have $Q_1 = Q_2$. According to the soft covering lemma [83, Theorem 4] or [77, Corollary VII.4] the first and the second term on the RHS of (W.7) vanishes when n grows if

$$R > \mathbb{I}_P(X;Z), \tag{W.10a}$$

$$R_J > \mathbb{I}_{\Upsilon}(S;Z). \tag{W.10b}$$

Decoding and Error Probability Analysis: By following the same steps as in [37], the probability of error vanishes when n grows if

$$R < \mathbb{I}_P(X;Y). \tag{W.11}$$

Combining (W.10) and (W.11) and noting that $Q_1 = Q_2$ completes the proof of Theorem 21.

APPENDIX X

PROOF OF THEOREM 22

For simplicity, we consider the case where the time-sharing random variable Q is constant. One can incorporate this into our proof by generating its i.i.d. copies, and sharing it among all parties and conditioning everything on it.

Fix P_{S_1}, P_X, P_{S_2}, and $\epsilon > 0$ such that, $Q_Z = Q_0$.

Random Codebook Generation for Communication Mode:

- Let $C_1^{(n)} \triangleq \big\{X^n(m)\big\}_{m\in\mathcal{M}_n}$, where $\mathcal{M}_n = [\![1, 2^{nR}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_X^{\otimes n}$. We denote a realization of $C_1^{(n)}$ by $\mathcal{C}_1^{(n)} \triangleq \big\{x^n(m)\big\}_{m\in\mathcal{M}_n}$.

- Let $C_2^{(n)} \triangleq \big\{S_1^n(k)\big\}_{k\in\mathcal{K}_n}$, where $\mathcal{K}_n = [\![1, 2^{nR_K}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_{S_1}^{\otimes n}$. We denote a realization of $C_2^{(n)}$ by $\mathcal{C}_2^{(n)} \triangleq \big\{s_1^n(k)\big\}_{k\in\mathcal{K}_n}$.

Random Codebook Generation for No-Communication Mode:

- Let $C_3^{(n)} \triangleq \big\{S_2^n(k)\big\}_{k\in\mathcal{K}}$ be a random codebook consisting of independent random sequences, each generated according to $P_{S_2}^{\otimes n}$. We denote a realization of $C_3^{(n)}$ by $\mathcal{C}_3^{(n)} \triangleq \big\{s_2^n(k)\big\}_{k\in\mathcal{K}}$.

Also, $C_n = \big\{C_1^{(n)}, C_2^{(n)}, C_3^{(n)}\big\}$ denotes a random codebook and $\mathcal{C}_n = \big\{\mathcal{C}_1^{(n)}, \mathcal{C}_2^{(n)}, \mathcal{C}_3^{(n)}\big\}$ denotes a fixed codebook. The set of all possible values of C_n is denoted by $\mathfrak{C}_n$. The codebook construction described above induces the PMF $\lambda \in \mathcal{P}(\mathfrak{C}_n)$ over the codebook ensemble. For each $\mathcal{C}_n \in \mathfrak{C}_n$

$$\lambda(\mathcal{C}_n) = \prod_{m\in\mathcal{M}_n} P_X^{\otimes n}\big(x^n(m)\big) \prod_{k'\in\mathcal{K}_n} P_S^{\otimes n}\big(s^n(k')\big) \prod_{k''\in\mathcal{K}_n} P_S^{\otimes n}\big(s^n(k'')\big). \tag{X.1}$$

Encoding for Communication Mode: To send the message m, the transmitter computes $x^n(m)$ and transmits it over the channel. Also, given the key k, the jammer computes $s_1^n(k)$ and transmits it over the channel.

For a fixed codebook $\mathcal{C}_n$, the induced joint distribution is

$$P^{(\mathcal{C}_n)}_{KMS_1^nX^nZ^n}(k,m,\tilde{s}_1^n,\tilde{x}^n,z^n) = 2^{-n(R_K+R)}\mathbb{1}_{\left\{\tilde{s}_1^n=s_1^n(k)\right\}\cap\left\{\tilde{x}^n=x^n(m)\right\}}W^{\otimes n}_{Z|XS}(z^n|\tilde{x}^n,\tilde{s}_1^n). \quad \text{(X.2)}$$

Considering the random codebook generation, we have

$$P(\mathcal{C}_n,k,m,\tilde{s}^n,\tilde{x}^n,z^n) = \lambda(\mathcal{C}_n)P^{(\mathcal{C}_n)}(k,m,\tilde{s}^n,\tilde{x}^n,z^n), \quad \text{(X.3)}$$

where $\lambda \in \mathcal{P}(\mathfrak{C}_n)$ is defined in (X.1).

Encoding for No-Communication Mode: When the transmitter is not communicating with the receiver, and therefore it transmits x_0^n, the jammer computes a sequence $s_2^n(k)$ according to the key k, and transmits it over the channel. For this scenario for a fixed codebook $\mathcal{C}_n$, the induced joint distribution is as follows

$$\Upsilon^{(\mathcal{C}_n)}_{KS_2^nZ^n}(k,\tilde{s}_2^n,z^n) = \frac{1}{2^{nR_K}}\mathbb{1}_{\{\tilde{s}_2^n=s_2^n(k)\}}W^{\otimes n}_{Z|X=x_0,S}\big(z^n|x_0^n,\tilde{s}_2^n\big). \quad \text{(X.4)}$$

Therefore, the distribution induced at the output of the warden is

$$\Upsilon^{(\mathcal{C}_n)}_{Z^n}(z^n) = \frac{1}{2^{nR_K}}\sum_{j=1}^{2^{nR_K}} W^{\otimes n}_{Z|X=x_0,S}\big(z^n|x_0^n,s_2^n(k)\big). \quad \text{(X.5)}$$

For this scenario if $R_K > \mathbb{I}_\Upsilon(S_2;Z)$ then according to soft covering lemma [83, Theorem 4] or [77, Corollary VII.4], we have

$$\mathbb{E}_{C_n}\mathbb{V}\big(\Upsilon_{Z^n|C_n},Q_0^{\otimes n}\big) \underset{n\to\infty}{\longrightarrow} 0, \quad \text{(X.6)}$$

where

$$Q_0(\cdot) = \sum_{s_2\in\mathcal{S}_2} P_{S_2}(s_2)W_{Z|X=0,S_2}\big(\,\cdot\,|x_0,s_2\big). \quad \text{(X.7)}$$

Note that if $R_K < \mathbb{I}_\Upsilon(S_2; Z)$ according to Shannon's channel coding theorem, the warden might be able to decode J, which reduces the problem to the point to point channel for which the covert rate will be zero.

Covert Analysis: Henceforth, we use $P^{(\mathcal{C}_n)}$ when the codebook is fixed, and we use $P_{\cdot|C_n}$ when the codebook is random. Our goal is to show that the coding scheme described above guarantees that

$$\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||\Upsilon_{Z^n|C_n})] \underset{n\to\infty}{\longrightarrow} 0. \tag{X.8}$$

To show that (X.8) holds by using Lemma 1 and the triangle inequality we have

$$\mathbb{E}_{C_n}\mathbb{V}\big(P_{Z^n|C_n}, \Upsilon_{Z^n|C_n}\big) \leq \mathbb{E}_{C_n}\mathbb{V}\big(P_{Z^n|C_n}, Q_0^{\otimes n}\big) + \mathbb{E}_{C_n}\mathbb{V}\big(\Upsilon_{Z^n|C_n}, Q_0^{\otimes n}\big). \tag{X.9}$$

From (X.6) the second term on the RHS of (X.9) vanishes when n grows. To bound the first term on the RHS of (X.9) we first show that the coding scheme described above guarantees that

$$\mathbb{E}_{C_n}\big[\mathbb{D}\big(P_{Z^n|C_n}||Q_Z^{\otimes n}\big)\big] \underset{n\to\infty}{\longrightarrow} 0, \tag{X.10}$$

where

$$P_{Z^n|C_n}(z^n) = \sum_{k=1}^{2^{nR_K}} \sum_{m=1}^{2^{nR}} 2^{-n(R_K+R)} W_{Z|XS}^{\otimes n}(z^n|X^n(m), S_1^n(k)), \tag{X.11}$$

$$Q_Z(\cdot) = \sum_{s_1\in\mathcal{S}_1} \sum_{x\in\mathcal{X}} P_{S_1}(s_1) P_X(x) W_{Z|XS}\big(\,\cdot\,|x, s_1\big). \tag{X.12}$$

Then, we choose P_{S_1}, P_X, and P_{S_2} such that $Q_Z = Q_0$. One can show that $\mathbb{E}_{C_n}\big[\mathbb{D}\big(P_{Z^n|C_n}||Q_Z^{\otimes n}\big)\big] \underset{n\to\infty}{\longrightarrow} 0$ if

$$R_K > \mathbb{I}_P(S_1; Z), \tag{X.13a}$$

$$R > \mathbb{I}_P(X; Z), \tag{X.13b}$$

$$R_K + R > \mathbb{I}_P(X, S_1; Z). \tag{X.13c}$$

Therefore, from (X.9) and using Lemma 1 we have $\mathbb{E}_{C_n}\left[\mathbb{D}\left(P_{Z^n|C_n}||Q_Z^{\otimes n}\right)\right] \underset{n\to\infty}{\longrightarrow} 0$.

Decoding and Error Probability Analysis: By following the same steps as in [37], the probability of error vanishes when n grows if

$$R < \mathbb{I}_P(X;Y|S_1). \tag{X.14}$$

APPENDIX Y

PROOF OF THEOREM 23

Fix P_{S_1}, $P_{X|S_1}$, P_{S_2}, and $\epsilon > 0$ such that, $Q_Z = Q_0$.

Random Codebook Generation for Communication Mode:

- Let $C_1^{(n)} \triangleq \left\{S_1^n(k)\right\}_{k\in\mathcal{K}_n}$, where $\mathcal{K}_n = [\![1, 2^{nR_K}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_{S_1}^{\otimes n}$. We denote a realization of $C_1^{(n)}$ by $\mathcal{C}_1^{(n)} \triangleq \left\{s_1^n(k)\right\}_{k\in\mathcal{K}_n}$.

- Fix $\mathcal{C}_1^{(n)}$ and for every $k \in \mathcal{K}_n$ let $C_2^{(n)} \triangleq \left\{X^n(k,m)\right\}_{m\in\mathcal{M}_n}$, where $\mathcal{M}_n = [\![1, 2^{nR}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_{X|S_1=s_1(k)}^{\otimes n}$. We denote a realization of $C_2^{(n)}$ by $\mathcal{C}_2^{(n)} \triangleq \left\{x^n(k,m)\right\}_{m\in\mathcal{M}}$.

Random Codebook Generation for No-Communication Mode:

- Let $C_3^{(n)} \triangleq \left\{S_2^n(k)\right\}_{k\in\mathcal{K}}$ be a random codebook consisting of independent random sequences, each generated according to $P_{S_2}^{\otimes n}$. We denote a realization of $C_3^{(n)}$ by $\mathcal{C}_3^{(n)} \triangleq \left\{s_2^n(k)\right\}_{k\in\mathcal{K}}$.

Also, $C_n = \left\{C_1^{(n)}, C_2^{(n)}, C_3^{(n)}\right\}$ denotes a random codebook and $\mathcal{C}_n = \left\{\mathcal{C}_1^{(n)}, \mathcal{C}_2^{(n)}, \mathcal{C}_3^{(n)}\right\}$ denotes a fixed codebook. The set of all possible values of C_n is denoted by $\mathfrak{C}_n$. The codebook construction described above induces the PMF $\lambda \in \mathcal{P}(\mathfrak{C}_n)$ over the codebook ensemble. For each $\mathcal{C}_n \in \mathfrak{C}_n$

$$\lambda(\mathcal{C}_n) = \prod_{k\in\mathcal{K}_n} P_S^{\otimes n}\big(s^n(k)\big) \prod_{(k',m)\in\mathcal{K}_n\times\mathcal{M}_n} P_{X|S}^{\otimes n}\Big(x^n\big(k',m\big)\big|s^n(k')\Big) \prod_{k''\in\mathcal{K}_n} P_S^{\otimes n}\big(s^n(k'')\big). \quad \text{(Y.1)}$$

Encoding for Communication Mode: The jammer computes $s^n(k)$ according to the shared key k and transmits it over the channel. Given the key k, the encoder computes $x^n(k,m)$ according to the message m and transmits it over the channel.

For a fixed codebook $\mathcal{C}_n$, the induced joint distribution is

$$P^{(\mathcal{C}_n)}_{K,M,S^n,X^n,Z^n}(k,m,\tilde{s}^n,\tilde{x}^n,z^n) = 2^{-n(R_K+R)}\mathbb{1}_{\left\{\tilde{s}^n=s^n(k)\right\}\cap\left\{\tilde{x}^n=x^n(k,m)\right\}}W^{\otimes n}_{Z|XS}(z^n|\tilde{x}^n,\tilde{s}^n). \tag{Y.2}$$

Considering the random codebook generation, we have

$$P(\mathcal{C}_n,k,m,\tilde{s}^n,\tilde{x}^n,z^n) = \lambda(\mathcal{C}_n)P^{(\mathcal{C}_n)}(k,m,\tilde{s}^n,\tilde{x}^n,z^n), \tag{Y.3}$$

where $\lambda \in \mathcal{P}(\mathfrak{C}_n)$ is defined in (Y.1).

Encoding for No-Communication Mode: When the transmitter is not communicating with the receiver, and therefore it transmits x_0^n, the jammer computes a sequence $s_2^n(k)$ according to the key k, and transmits it over the channel. For this scenario for a fixed codebook $\mathcal{C}_n$, the induced joint distribution is as follows

$$\Upsilon^{(\mathcal{C}_n)}_{KS_2^nZ^n}(k,\tilde{s}_2^n,z^n) = \frac{1}{2^{nR_K}}\mathbb{1}_{\{s_2^n=s_2^n(k)\}}W^{\otimes n}_{Z|XS}\big(z^n|x_0^n,\tilde{s}_2^n\big). \tag{Y.4}$$

Therefore, the distribution induced at the output of the warden for a random codebook is

$$\Upsilon_{Z^n|C_n}(z^n) = 2^{-nR_K}\sum_{k=1}^{2^{nR_K}} W^{\otimes n}_{Z|XS}\big(z^n|x_0^n,S_2^n(k)\big). \tag{Y.5}$$

For this scenario if $R_K > \mathbb{I}_\Upsilon(S_2;Z)$ according to soft covering lemma [83, Theorem 4] or [77, Corollary VII.4], we have

$$\mathbb{E}_{C_n}\left[\mathbb{V}\big(\Upsilon_{Z^n|C_n},Q_0^{\otimes n}\big)\right] \underset{n\to\infty}{\longrightarrow} 0, \tag{Y.6}$$

where

$$Q_0(\cdot) = \sum_{s_2\in\mathcal{S}_2} P_{S_2}(s_2)W_{Z|XS}\big(\,\cdot\,|x_0,s_2\big). \tag{Y.7}$$

Note that if $R_K < \mathbb{I}_\Upsilon(S_2;Z)$ according to Shannon's channel coding theorem, the warden might be able to decode K, which reduces the problem to the point to point channel for which the covert rate will be zero.

Covert Analysis: Our goal is to show that the coding scheme described above guarantees that

$$\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||\Upsilon_{Z^n|C_n})] \underset{n\to\infty}{\longrightarrow} 0. \tag{Y.8}$$

To show that (Y.8) holds by using Lemma 1 and the triangle inequality we have

$$\mathbb{E}_{C_n}\big[\mathbb{V}\big(P_{Z^n|C_n}, \Upsilon_{Z^n|C_n}\big)\big] \leq \mathbb{E}_{C_n}\big[\mathbb{V}\big(P_{Z^n|C_n}, Q_0^{\otimes n}\big)\big] + \mathbb{E}_{C_n}\big[\mathbb{V}\big(\Upsilon_{Z^n|C_n}, Q_0^{\otimes n}\big)\big]. \tag{Y.9}$$

From (Y.6) the second term on the RHS of (Y.9) vanishes when n grows. To bound the first term on the RHS of (Y.9) we first show that the coding scheme described above guarantees that

$$\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n})] \underset{n\to\infty}{\longrightarrow} 0, \tag{Y.10}$$

where $Q_Z^{\otimes n} = \prod_{i=1}^{n} Q_Z$ and

$$Q_Z(\cdot) = \sum_{s_1\in\mathcal{S}_1}\sum_{x\in\mathcal{X}} P_{S_1}(s_1)P_{X|S_1}(x|s_1)W_{Z|XS}\big(\,\cdot\,|x, s_1\big). \tag{Y.11}$$

Then, we choose P_{S_1}, $P_{X|S_1}$, and P_{S_2} such that $Q_Z = Q_0$. From (Y.2) we have,

$$P_{Z^n|C_n}(z^n) = 2^{-n(R_K+R)} \sum_{k=1}^{2^{nR_K}} \sum_{m=1}^{2^{nR}} W_{Z|XS}^{\otimes n}\big(z^n|X^n(k,m), S_1^n(k)\big). \tag{Y.12}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}_{C_r}\big[\mathbb{D}\big(P_{Z^n|C_r}||Q_Z^{\otimes n}\big)\big] \\
&= \mathbb{E}_{C_r}\left[\sum_{z^n} P_{Z^n|C_r}(z^n)\log\frac{P_{Z^n|C_r}(z^n)}{Q_Z^{\otimes n}(z^n)}\right] \\
&= \mathbb{E}_{C_r}\Bigg[\sum_{z^n}\frac{1}{2^{n(R_K+R)}}\sum_{(k,m)} W_{Z|XS}^{\otimes n}\big(z^n|X^n(k,m), S_1^n(k)\big) \\
&\times \log\frac{\frac{1}{2^{n(R_K+R)}}\sum\limits_{(\tilde{k},\tilde{m})} W_{Z|XS}^{\otimes n}\big(z^n|X^n(\tilde{k},\tilde{m}), S_1^n(\tilde{k})\big)}{Q_Z^{\otimes n}(z^n)}\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\leq} \frac{1}{2^{n(R_K+R)}} \sum_{(k,m)} \sum_{(x^n,s_1^n,z^n)} P_{XS_1Z}^{\otimes n}\big(x^n(k,m), s_1^n(k), z^n\big) \\
&\times \log \mathbb{E}_{\setminus(k,m)}\left[\frac{\sum\limits_{(\tilde{k},\tilde{m})} W_{Z|XS}^{\otimes n}\big(z^n|X^n(\tilde{k},\tilde{m}), S_1^n(\tilde{k})\big)}{2^{n(R_K+R)}Q_Z^{\otimes n}(z^n)}\right] \\
&= \frac{1}{2^{n(R_K+R)}} \sum_{(k,m)} \sum_{(x^n,s_1^n,z^n)} P_{XS_1Z}^{\otimes n}\big(x^n(k,m), s_1^n(k), z^n\big) \log \mathbb{E}_{\setminus(k,m)}\Bigg[\frac{W_{Z|XS}^{\otimes n}\big(z^n|x^n(k,m), s_1^n(k)\big)}{2^{n(R_K+R)}Q_Z^{\otimes n}(z^n)} \\
&\quad + \frac{\sum\limits_{\tilde{m}\neq m} W_{Z|XS}^{\otimes n}\big(z^n|X^n(k,\tilde{m}), s_1^n(k)\big)}{2^{n(R_K+R)}Q_Z^{\otimes n}(z^n)} + \frac{\sum\limits_{\tilde{k}\neq k}\sum\limits_{\tilde{m}} W_{Z|XS}^{\otimes n}\big(z^n|X^n(\tilde{k},\tilde{m}), S_1^n(\tilde{k})\big)}{2^{n(R_K+R)}Q_Z^{\otimes n}(z^n)}\Bigg] \\
&\overset{(b)}{\leq} \frac{1}{2^{n(R_K+R)}} \sum_{(k,m)} \sum_{(x^n,s_1^n,z^n)} P_{X,S_1,Z}^{\otimes n}\big(x^n(k,m), s_1^n(k), z^n\big) \log \Bigg[\frac{W_{Z|XS}^{\otimes n}\big(z^n|x^n(k,m), s_1^n(k)\big)}{2^{n(R_K+R)}Q_Z^{\otimes n}(z^n)} \\
&\quad + \frac{\sum\limits_{\tilde{m}\neq m} W_{Z|S}^{\otimes n}\big(z^n|s_1^n(k)\big)}{2^{n(R_K+R)}Q_Z^{\otimes n}(z^n)} + 1\Bigg] \\
&\leq \frac{1}{2^{n(R_K+R)}} \sum_{(k,m)} \sum_{(x^n,s_1^n,z^n)} P_{XS_1Z}^{\otimes n}\big(x^n(k,m), s_1^n(k), z^n\big) \log \Bigg[\frac{W_{Z|XS}^{\otimes n}\big(z^n|x^n(k,m), s_1^n(k)\big)}{2^{n(R_K+R)}Q_Z^{\otimes n}(z^n)} \\
&\quad + \frac{W_{Z|S}^{\otimes n}\big(z^n|s_1^n(k)\big)}{2^{nR_K}Q_Z^{\otimes n}(z^n)} + 1\Bigg] \\
&\triangleq \Psi_1 + \Psi_2
\end{aligned} \tag{Y.13}
$$

where (a) follows from Jensen's inequality and (b) follows by taking expectation with respect to $\setminus(k,m)$ and by removing some terms from the denominator of the first term in the log function and adding one term to the nominator of the second term in the log function. We defined Ψ_1 and Ψ_2 as

$$
\begin{aligned}
\Psi_1 &= \frac{1}{2^{n(R_K+R)}} \sum_{(k,m)} \sum_{(x^n,s_1^n,z^n)\in\mathcal{T}_\epsilon^{(r)}} P_{XS_1Z}^{\otimes n}\big(x^n(k,m), s_1^n(k), z^n\big) \\
&\qquad \times \log\left[\frac{W_{Z|XS}^{\otimes n}\big(z^n|x^n(k,m), s_1^n(k)\big)}{2^{n(R_K+R)}Q_Z^{\otimes n}(z^n)} + \frac{W_{Z|S}^{\otimes n}\big(z^n|s_1^n(k)\big)}{2^{nR_K}Q_Z^{\otimes n}(z^n)} + 1\right] \\
&\leq \log\left(\frac{2^{-r(1-\epsilon)\mathbb{H}(Z|X,S_1)}}{2^{n(R_K+R)}2^{-r(1+\epsilon)\mathbb{H}(Z)}} + \frac{2^{-r(1-\epsilon)\mathbb{H}(Z|S_1)}}{2^{nR_K}2^{-r(1+\epsilon)\mathbb{H}(Z)}} + 1\right)
\end{aligned} \tag{Y.14}
$$

and

$$\Psi_2 = \frac{1}{2^{n(R_K+R)}} \sum_{(k,m)} \sum_{(x^n,s_1^n,z^n)\notin\mathcal{T}_\epsilon^{(r)}} P_{XS_1Z}^{\otimes n}\big(x^n(k,m), s_1^n(k), z^n\big)$$
$$\times \log\left[\frac{W_{Z|XS}^{\otimes n}\big(z^n|x^n(k,m), s_1^n(k)\big)}{2^{n(R_K+R)}Q_Z^{\otimes n}(z^n)} + \frac{W_{Z|S}^{\otimes n}\big(z^n|s_1^n(k)\big)}{2^{nR_K}Q_Z^{\otimes n}(z^n)} + 1\right]$$
$$\leq 2|\mathcal{X}||\mathcal{S}_1||\mathcal{Z}|e^{-n\epsilon^2\mu_{X,S_1,Z}} n\log(\frac{2}{\mu_Z}+1). \tag{Y.15}$$

where

$$\mu_Z = \min_{\substack{z\in\mathcal{Z} \\ \text{s.t. } Q_z>0}} Q(z) \tag{Y.16}$$

$$\mu_{X,S,Z} = \min_{\substack{(x,s_1,z)\in(\mathcal{X},\mathcal{S}_1,\mathcal{Z}) \\ \text{s.t. } Q(x,s_1,z)>0}} Q(x,s_1,z) \tag{Y.17}$$

When $n\to\infty$ then $\Psi_2\to 0$ and $\Psi_1\to 0$ when n grows if

$$R_K > \mathbb{I}_P(S_1;Z), \tag{Y.18a}$$

$$R_K + R > \mathbb{I}_P(X,S_1;Z). \tag{Y.18b}$$

Decoding and Error Probability Analysis: By following the same steps as in [37], the probability of error vanishes when n grows if

$$R < \mathbb{I}_P(X;Y|S_1). \tag{Y.19}$$

APPENDIX Z

PROOF OF THEOREM 24

To prove the upper bound for the case that the jammer has an unlimited source of local randomness and transmits an i.i.d. sequence when communication is not happening, consider any sequence of codes with length n such that $P_e^{(n)} \leq \epsilon_n$ and $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \delta$ where $\lim_{n\to\infty} \epsilon_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish.

Epsilon Rate Region: We first define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ that expands the region defined in (4.38) as follows.

$$\mathcal{A}_\epsilon = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists P_{SXYZ} \in \mathcal{D}_\epsilon : \\ R \leq \mathbb{I}(X;Y|S) + \epsilon \\ R_K \geq \max\{\mathbb{I}(X,S;Z) - \mathbb{I}(X;Y|S), \mathbb{I}(S;Z)\} - 3\epsilon \end{array} \right\}, \tag{Z.1a}$$

where

$$\mathcal{D}_\epsilon = \left\{ \begin{array}{l} P_{SXYZ} : \\ P_{SXYZ} = P_S P_{X|S} W_{YZ|XS} \\ \mathbb{D}\left(P_Z \| Q_0\right) \leq \epsilon \end{array} \right\}. \tag{Z.1b}$$

We next show that if a rate R is achievable, then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$, we start by upper bounding nR using standard techniques.

$$\begin{aligned} nR &= \mathbb{H}(M) \\ &\overset{(a)}{\leq} \mathbb{H}(M|S^n) - \mathbb{H}(M|Y^n, S^n) + n\epsilon_n \\ &= \mathbb{I}(M; Y^n|S^n) + n\epsilon_n \\ &= \sum_{i=1}^{n} \mathbb{I}(M; Y_i|Y^{i-1}, S^n) + n\epsilon_n \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} [\mathbb{H}(Y_i|Y^{i-1}, S^n) - \mathbb{H}(Y_i|M, Y^{i-1}, S^n)] + n\epsilon_n \\
&\overset{(b)}{\leq} \sum_{i=1}^{n} [\mathbb{H}(Y_i|S_i) - \mathbb{H}(Y_i|M, Y^{i-1}, S^n, X^n)] + n\epsilon_n \\
&= \sum_{i=1}^{n} [\mathbb{H}(Y_i|S_i) - \mathbb{H}(Y_i|S_i, X_i)] + n\epsilon_n \\
&= \sum_{i=1}^{n} \mathbb{I}(X_i; Y_i|S_i) + n\epsilon_n \\
&\overset{(c)}{\leq} n\mathbb{I}(\tilde{X}; \tilde{Y}|\tilde{S}) + n\epsilon_n \qquad \text{(Z.2)} \\
&\overset{(d)}{\leq} n\mathbb{I}(\tilde{X}; \tilde{Y}|\tilde{S}) + n\epsilon \qquad \text{(Z.3)}
\end{aligned}
$$

where

(a) follows from Fano's inequality and independence of M from S^n;

(b) holds because conditioning does not increase entropy;

(c) follows from the concavity of mutual information, with the resulting random variables $\tilde{X}$, $\tilde{S}$, $\tilde{Y}$, and $\tilde{Z}$ to having the following distributions

$$\tilde{P}_{X,S}(x, s) \triangleq \frac{1}{n} \sum_{i=1}^{n} P_{X_i,S_i}(x, s), \qquad \text{(Z.4a)}$$

$$\tilde{P}_{X,S,Y,Z}(x, s, y, z) \triangleq \tilde{P}_{X,S}(x, s) W_{YZ|XS}(y, z|x, s); \qquad \text{(Z.4b)}$$

(d) follows by defining $\epsilon \triangleq \max\{\epsilon_n, \nu \geq \frac{\delta}{n}\}$.

We also have,

$$
\begin{aligned}
n(R + R_K) &\geq \mathbb{H}(M, K) \\
&= \mathbb{H}(M, K) \\
&\geq \mathbb{I}(M, K; Z^n) \\
&\overset{(a)}{=} \mathbb{I}(M, K, X^n, S^n; Z^n)
\end{aligned}
$$

$$
\begin{aligned}
&\geq \mathbb{I}(X^n, S^n; Z^n) \\
&= \sum_{x^n} \sum_{s^n} \sum_{z^n} P(x^n, s^n, z^n) \log \frac{W_{Z|XS}^{\otimes n}(z^n|x^n, s^n)}{P(z^n)} \\
&\geq \sum_{x^n} \sum_{s^n} \sum_{z^n} P(x^n, s^n, z^n) \log \frac{W_{Z|XS}^{\otimes n}(z^n|x^n, s^n)}{P(z^n)} + \mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) - \delta \\
&\geq \sum_{i=1}^{n} \sum_{x_i} \sum_{s_i} \sum_{z_i} P(x_i, s_i, z_i) \log \frac{W_{Z|XS}(z_i|x_i, s_i)}{Q_0(z_i)} - \delta \\
&= \sum_{i=1}^{n} \mathbb{D}(P_{X_i,S_i,Z_i}||P_{X_i,S_i}Q_0) - \delta \\
&\overset{(b)}{\geq} n\mathbb{D}(\tilde{P}_{X,S,Z}||\tilde{P}_{X,S}Q_0) - \delta \\
&= n\mathbb{D}(\tilde{P}_{X,S,Z}||\tilde{P}_{X,S}\tilde{P}_Z) + \mathbb{D}(\tilde{P}_Z||Q_0) - \delta \\
&\overset{(c)}{\geq} n\mathbb{I}(\tilde{X}, \tilde{S}; \tilde{Z}) - 2\delta
\end{aligned} \tag{Z.5}
$$

where

(a) follows because X^n is a function of (M, K, S^n) and S^n is a function of K;

(b) follows from Jensen's inequality, the convexity of $\mathbb{D}(\cdot||\cdot)$, and concavity of $\mathbb{H}(\cdot)$;

(c) follows from the definition of random variables $\tilde{X}$, $\tilde{S}$, $\tilde{Y}$, and $\tilde{Z}$ in (Z.4).

From (Z.5) for any $\nu > 0$, choosing n large enough ensures that

$$
\begin{aligned}
R + R_K &\geq \mathbb{I}(\tilde{X}, \tilde{S}; \tilde{Z}) - 2\nu, \\
&\geq \mathbb{I}(\tilde{X}, \tilde{S}; \tilde{Z}) - 2\epsilon,
\end{aligned} \tag{Z.6}
$$

where the last equality follows from the definition of $\epsilon \triangleq \max\{\epsilon_n, \nu\}$. We now have,

$$
\begin{aligned}
nR_K &\geq \mathbb{H}(K) \\
&\geq \mathbb{I}(K; Z^n) \\
&\overset{(a)}{=} \mathbb{I}(K, S^n; Z^n)
\end{aligned}
$$

$$
\begin{aligned}
&\geq \mathbb{I}(S^n; Z^n) \\
&= \mathbb{I}(X^n, S^n; Z^n) - \mathbb{I}(X^n; Z^n|S^n) \\
&\overset{(b)}{\geq} \sum_{x^n}\sum_{s^n}\sum_{z^n} P(x^n, s^n, z^n) \log \frac{W_{Z|XS}^{\otimes n}(z^n|x^n, s^n)}{P(z^n)} - \sum_{i=1}^{n} \mathbb{I}(X_i; Z_i|S_i) \\
&\geq \sum_{x^n}\sum_{s^n}\sum_{z^n} P(x^n, s^n, z^n) \log \frac{W_{Z|XS}^{\otimes n}(z^n|x^n, s^n)}{P(z^n)} + \mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) - \sum_{i=1}^{n} \mathbb{I}(X_i; Z_i|S_i) - \delta \\
&\geq \sum_{i=1}^{n}\sum_{x_i}\sum_{s_i}\sum_{z_i} P(x_i, s_i, z_i) \log \frac{W_{Z|XS}(z_i|x_i, s_i)}{Q_0(z_i)} - \sum_{i=1}^{n} \mathbb{I}(X_i; Z_i|S_i) - \delta \\
&= \sum_{i=1}^{n}\sum_{x_i}\sum_{s_i}\sum_{z_i} P(x_i, s_i, z_i) \log \frac{W_{Z|XS}(z_i|x_i, s_i)}{Q_0(z_i)} \\
&\quad - \sum_{i=1}^{n}\sum_{x_i}\sum_{s_i}\sum_{z_i} P(x_i, s_i, z_i) \log \frac{W_{Z|XS}(z_i|x_i, s_i)}{P_{Z|S}(z_i|s_i)} - \delta \\
&= \sum_{i=1}^{n}\sum_{x_i}\sum_{s_i}\sum_{z_i} P(x_i, s_i, z_i) \Big[\log \frac{W_{Z|XS}(z_i|x_i, s_i)}{Q_0(z_i)} - \log \frac{W_{Z|XS}(z_i|x_i, s_i)}{P_{Z|S}(z_i|s_i)} \Big] - \delta \\
&= \sum_{i=1}^{n}\sum_{x_i}\sum_{s_i}\sum_{z_i} P(x_i, s_i, z_i) \log \frac{P_{Z|S}(z_i|s_i)}{Q_0(z_i)} - \delta \\
&= \sum_{i=1}^{n} \mathbb{D}(P_{S_i,Z_i}||P_{S_i}Q_0) - \delta \\
&\overset{(c)}{\geq} n\mathbb{D}(\tilde{P}_{S,Z}||\tilde{P}_S Q_0) - \delta \\
&= n\mathbb{D}(\tilde{P}_{S,Z}||\tilde{P}_S\tilde{P}_Z) + \mathbb{D}(\tilde{P}_Z||Q_0) - \delta \\
&\geq n\mathbb{I}(\tilde{S}; \tilde{Z}) - 2\delta
\end{aligned} \tag{Z.7}
$$

where

(a) follows because S^n is a function of J;

(b) follows from Jensen's inequality, the convexity of $\mathbb{D}(\cdot||\cdot)$, and concavity of $\mathbb{H}(\cdot)$;

(c) follows from the definition of random variables $\tilde{X}$, $\tilde{S}$, $\tilde{Y}$, and $\tilde{Z}$ in (Z.4).

From (Z.7) for any $\nu > 0$, choosing n large enough ensures that

$$R_K \geq \mathbb{I}(\tilde{S}; \tilde{Z}) - 2\nu,$$

$$\geq \mathbb{I}(\tilde{S};\tilde{Z}) - 2\epsilon, \tag{Z.8}$$

where the last equality follows from the definition of $\epsilon \triangleq \max\{\epsilon_n, \nu\}$. To show that $\mathbb{D}(P_Z||Q_0) \leq \epsilon$, note that for n large enough

$$\begin{aligned}\mathbb{D}(P_Z||Q_0) = \mathbb{D}(P_{\tilde{Z}}||Q_0) &= \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} P_{Z_i}\middle\|Q_0\right) \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \leq \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \frac{\delta}{n} \leq \nu \leq \epsilon.\end{aligned} \tag{Z.9}$$

Combining (Z.3), (Z.6), and (Z.8) shows that $\forall \epsilon_n, \nu > 0$, $C_{\text{FK}} \subseteq \mathcal{A}_\epsilon$.

Continuity at Zero: The proof for continuity at zero is similar to that of Appendix N.

APPENDIX AA

PROOF OF LEMMA 4

We Prove Lemma 4 in two different cases. First when $R_2 > \mathbb{I}(X_2; Z)$, for this case we have

$$\begin{aligned}
&\mathbb{E}_{C_1,C_2}\Big[\mathbb{D}\big(P_{Z^n|C_1,C_2}||P_{Z^n|C_2}\big)\Big] \\
&\overset{(a)}{\leq} \sqrt{2\ln 2}\log\left(\frac{1}{\mu_Q}\right)\mathbb{E}_{C_1,C_2}\Big[\sqrt{\mathbb{D}\big(P_{Z^n|C_1,C_2}||Q_Z^{\otimes n}\big)}+\sqrt{\mathbb{D}\big(P_{Z^n|C_2}||Q_Z^{\otimes n}\big)}\Big], \\
&\overset{(b)}{\leq} \sqrt{2\ln 2}\log\left(\frac{1}{\mu_Q}\right)\Big[\sqrt{\mathbb{E}_{C_1,C_2}\left(\mathbb{D}\big(P_{Z^n|C_1,C_2}||Q_Z^{\otimes n}\big)\right)}+\sqrt{\mathbb{E}_{C_1,C_2}\left(\mathbb{D}\big(P_{Z^n|C_2}||Q_Z^{\otimes n}\big)\right)}\Big], \quad \text{(AA.1)}
\end{aligned}$$

where (a) follows from [93, Lemma 38], for finite output alphabet $\mathcal{Z}$, $Q_Z(\cdot) = \sum_{x_1}\sum_{x_2} P(x_1)P(x_2)W_{Z|X_1X_2}(\cdot|x_1,x_2)$, and $\mu_Q \triangleq \min\limits_{z\in\mathcal{Z}:\,Q_Z>0} Q_Z(z)$, and (b) follows from Jensen's inequality. For the first term on the RHS of (AA.1) we have

$$\begin{aligned}
&\mathbb{E}_{C_1,C_2}\Big[\mathbb{D}\Big(P_{Z^n|C_1,C_2}||Q_Z^{\otimes n}\Big)\Big] = \mathbb{E}_{C_1,C_2}\left[\sum_{z^n} P_{Z^n|C_1,C_2}(z^n)\log\frac{P_{Z^n|C_1,C_2}(z^n)}{Q_Z^{\otimes n}(z^n)}\right] \\
&= \mathbb{E}_{C_1,C_2}\Bigg[\sum_{z^n}\frac{1}{2^{n(R_1+R_2)}}\sum_{(m_1,m_2)} W_{Z|X_1X_2}^{\otimes n}\big(z^n|X_1^n(m_1),X_2^n(m_2)\big) \\
&\qquad\times\log\frac{\frac{1}{2^{n(R_1+R_2)}}\sum\limits_{(\tilde{m}_1,\tilde{m}_2)} W_{Z|X_1X_2}^{\otimes n}\big(z^n|X_1^n(\tilde{m}_1),X_2^n(\tilde{m}_2)\big)}{Q_Z^{\otimes n}(z^n)}\Bigg] \\
&\overset{(a)}{\leq}\frac{1}{2^{n(R_1+R_2)}}\sum_{(m_1,m_2)}\sum_{(x_1^n,x_2^n,z^n)} P_{X_1,X_2,Z}^{\otimes n}\big(x_1^n(m_1),x_2^n(m_2),z^n\big) \\
&\qquad\times\log\mathbb{E}_{\setminus(m_1,m_2)}\left[\frac{\sum\limits_{(\tilde{m}_1,\tilde{m}_2)} W_{Z|X_1X_2}^{\otimes n}\big(z^n|X_1^n(\tilde{m}_1),X_2^n(\tilde{m}_2)\big)}{2^{n(R_1+R_2)}Q_Z^{\otimes n}(z^n)}\right] \\
&=\frac{1}{2^{n(R_1+R_2)}}\sum_{(m_1,m_2)}\sum_{(x_1^n,x_2^n)} P_{X_1,X_2,Z}^{\otimes n}\big(x_1^n(m_1),x_2^n(m_2),z^n\big) \\
&\qquad\times\log\mathbb{E}_{\setminus(m_1,m_2)}\Bigg[\frac{W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),x_2^n(m_2)\big)}{2^{n(R_1+R_2)}Q_Z^{\otimes n}(z^n)}+\frac{\sum\limits_{\tilde{m}_2\neq m_2} W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),X_2^n(\tilde{m}_2)\big)}{2^{n(R_1+R_2)}Q_Z^{\otimes n}(z^n)} \\
&\qquad+\frac{\sum\limits_{\tilde{m}_1\neq m_1} W_{Z|X_1X_2}^{\otimes n}\big(z^n|X_1^n(\tilde{m}_1),x_2^n(m_2)\big)}{2^{n(R_1+R_2)}Q_Z^{\otimes n}(z^n)}+\frac{\sum\limits_{\tilde{m}_1\neq m_1}\sum\limits_{\tilde{m}_2\neq m_2} W_{Z|X_1X_2}^{\otimes n}\big(z^n|X_1^n(\tilde{m}_1),X_2^n(\tilde{m}_2)\big)}{2^{n(R_1+R_2)}Q_Z^{\otimes n}(z^n)}\Bigg]
\end{aligned}$$

$$= \frac{1}{2^{n(R_1+R_2)}} \sum_{(m_1,m_2)} \sum_{(x_1^n,x_2^n,z^n)} P_{X_1,X_2,Z}^{\otimes n}\big(x_1^n(m_1),x_2^n(m_2),z^n\big) \log \Bigg[\frac{W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),x_2^n(m_2)\big)}{2^{n(R_1+R_2)}Q_Z^{\otimes n}(z^n)} + \sum_{\tilde{m}_2\neq m_2} \frac{W_{Z|X_1}^{\otimes n}\big(z^n|x_1^n(m_1)\big)}{2^{n(R_1+R_2)}Q_Z^{\otimes n}(z^n)} + \sum_{\tilde{m}_1\neq m_1} \frac{W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)}{2^{n(R_1+R_2)}Q_Z^{\otimes n}(z^n)} + \sum_{\tilde{m}_1\neq m_1}\sum_{\tilde{m}_2\neq m_2} \frac{Q_Z^{\otimes n}\big(z^n\big)}{2^{n(R_1+R_2)}Q_Z^{\otimes n}(z^n)} \Bigg]$$

$$\leq \frac{1}{2^{n(R_1+R_2)}} \sum_{(m_1,m_2)} \sum_{(x_1^n,x_2^n,z^n)} P_{X_1,X_2,Z}^{\otimes n}\big(x_1^n(m_1),x_2^n(m_2),z^n\big) \log \Bigg[\frac{W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),x_2^n(m_2)\big)}{2^{n(R_1+R_2)}Q_Z^{\otimes n}(z^n)} + \frac{W_{Z|X_1}^{\otimes n}\big(z^n|x_1^n(m_1)\big)}{2^{nR_1}Q_Z^{\otimes n}(z^n)} + \frac{W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)}{2^{nR_2}Q_Z^{\otimes n}(z^n)} + 1 \Bigg]$$

$$\triangleq \Psi_1 + \Psi_2 \tag{AA.2}$$

where (a) follows from Jensen's inequality. We defined Ψ_1 and Ψ_2 as

$$\Psi_1 = \frac{1}{2^{n(R_1+R_2)}} \sum_{(m_1,m_2)} \sum_{(x_1^n,x_2^n,z^n)\in\mathcal{T}_\epsilon^{(n)}} P_{X_1,X_2,Z}^{\otimes n}\big(x_1^n(m_1),x_2^n(m_2),z^n\big) \times \log \Bigg[\frac{W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),x_2^n(m_2)\big)}{2^{n(R_1+R_2)}Q_Z^{\otimes n}(z^n)} + \frac{W_{Z|X_1}^{\otimes n}\big(z^n|x_1^n(m_1)\big)}{2^{nR_1}Q_Z^{\otimes n}(z^n)} + \frac{W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)}{2^{nR_2}Q_Z^{\otimes n}(z^n)} + 1 \Bigg]$$

$$\leq \log\Big(\frac{2^{-r(1-\epsilon)\mathbb{H}(Z|X_1,X_2)}}{2^{n(R_1+R_2)}2^{-r(1+\epsilon)\mathbb{H}(Z)}} + \frac{2^{-r(1-\epsilon)\mathbb{H}(Z|X_1)}}{2^{nR_1}2^{-r(1+\epsilon)\mathbb{H}(Z)}} + \frac{2^{-r(1-\epsilon)\mathbb{H}(Z|X_2)}}{2^{nR_2}2^{-r(1+\epsilon)\mathbb{H}(Z)}} + 1 \Big) \tag{AA.3}$$

and

$$\Psi_2 = \frac{1}{2^{n(R_1+R_2)}} \sum_{(m_1,m_2)} \sum_{(x_1^n,x_2^n,z^n)\notin\mathcal{T}_\epsilon^{(n)}} P_{X_1,X_2,Z}^{\otimes n}\big(x_1^n(m_1),x_2^n(m_2),z^n\big) \times \log \Bigg[\frac{W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),x_2^n(m_2)\big)}{2^{n(R_1+R_2)}Q_Z^{\otimes n}(z^n)} + \frac{W_{Z|X_1}^{\otimes n}\big(z^n|x_1^n(m_1)\big)}{2^{nR_1}Q_Z^{\otimes n}(z^n)} + \frac{W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)}{2^{nR_2}Q_Z^{\otimes n}(z^n)} + 1 \Bigg]$$

$$\leq 2|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Z}|e^{-n\epsilon^2\mu_{X_1,X_2,Z}} n \log(\frac{2}{\mu_Z}+1). \tag{AA.4}$$

where

$$\mu_Z = \min_{\substack{z\in\mathcal{Z} \\ \text{s.t. } Q_z>0}} Q(z) \tag{AA.5}$$

$$\mu_{X_1,X_2,Z} = \min_{\substack{(x_1,x_2,z)\in(\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}) \\ \text{s.t. } Q(x_1,x_2,z)>0}} Q(x_1,x_2,z). \tag{AA.6}$$

When $n \to \infty$ then $\Psi_2 \to 0$ and $\Psi_1 \to 0$ when n grows if[1]

$$R_1 > \mathbb{I}(X_1; Z), \tag{AA.7a}$$

$$R_2 > \mathbb{I}(X_2; Z), \tag{AA.7b}$$

$$R_1 + R_2 > \mathbb{I}(X_1, X_2; Z). \tag{AA.7c}$$

Also, since $P_{Z^n|C_2}$ is for a scenario where the first transmitter, transmits an i.i.d. sequence and the second transmitter, transmits a codeword from C_2 by using (AA.7) one can show that the second term on the RHS of (AA.1) vanishes when n grows if $R_2 > \mathbb{I}(X_2; Z)$. This results to the region in $\mathcal{R}_1$.

When $R_2 \leq \mathbb{I}(X_2; Z)$ we have

$$\mathbb{E}_{C_1,C_2}\Big[\mathbb{D}\Big(P_{Z^n|C_1,C_2}||P_{Z^n|C_2}\Big)\Big] = \mathbb{E}_{C_1,C_2}\left[\sum_{z^n} P_{Z^n|C_1,C_2}(z^n)\log\frac{P_{Z^n|C_1,C_2}(z^n)}{P_{Z^n|C_2}(z^n)}\right]$$

$$= \mathbb{E}_{C_1,C_2}\Bigg[\sum_{z^n}\frac{1}{2^{n(R_1+R_2)}}\sum_{(m_1,m_2)} W_{Z|X_1X_2}^{\otimes n}\big(z^n|X_1^n(m_1),X_2^n(m_2)\big)$$
$$\times\log\frac{\frac{1}{2^{n(R_1+R_2)}}\sum\limits_{(\tilde{m}_1,\tilde{m}_2)} W_{Z|X_1X_2}^{\otimes n}\big(z^n|X_1^n(\tilde{m}_1),X_2^n(\tilde{m}_2)\big)}{\frac{1}{2^{nR_2}}\sum\limits_{m_2'} W_{Z|X_2}^{\otimes n}\big(z^n|X_2^n(m_2')\big)}\Bigg]$$

$$\overset{(a)}{\leq}\frac{1}{2^{n(R_1+R_2)}}\sum_{(m_1,m_2)}\sum_{(x_1^n,x_2^n,z^n)} P_{X_1,X_2,Z}^{\otimes n}\big(x_1^n(m_1),x_2^n(m_2),z^n\big)$$
$$\times\log\mathbb{E}_{\setminus(m_1,m_2)}\left[\frac{\sum\limits_{(\tilde{m}_1,\tilde{m}_2)} W_{Z|X_1X_2}^{\otimes n}\big(z^n|X_1^n(\tilde{m}_1),X_2^n(\tilde{m}_2)\big)}{2^{nR_1}\sum\limits_{m_2'} W_{Z|X_2}^{\otimes n}\big(z^n|X_2^n(m_2')\big)}\right]$$

$$= \frac{1}{2^{n(R_1+R_2)}}\sum_{(m_1,m_2)}\sum_{(x_1^n,x_2^n)} P_{X_1,X_2,Z}^{\otimes n}\big(x_1^n(m_1),x_2^n(m_2),z^n\big)$$
$$\times\log\mathbb{E}_{\setminus(m_1,m_2)}\left[\frac{W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),x_2^n(m_2)\big)}{2^{nR_1}\sum\limits_{m_2'} W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2')\big)} + \frac{\sum\limits_{\tilde{m}_2\neq m_2} W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),X_2^n(\tilde{m}_2)\big)}{2^{nR_1}\sum\limits_{m_2'} W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2')\big)}\right.$$

[1]In [23] and [94] it has been shown that the first term on the RHS of (AA.1) under total variation distance vanishes when n grows.

$$
\begin{aligned}
&+\frac{\sum\limits_{\tilde{m}_1\neq m_1} W_{Z|X_1X_2}^{\otimes n}\big(z^n|X_1^n(\tilde{m}_1),x_2^n(m_2)\big)}{2^{nR_1}\sum\limits_{m_2'} W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2')\big)}+\frac{\sum\limits_{\tilde{m}_1\neq m_1}\sum\limits_{\tilde{m}_2\neq m_2} W_{Z|X_1X_2}^{\otimes n}\big(z^n|X_1^n(\tilde{m}_1),X_2^n(\tilde{m}_2)\big)}{2^{nR_1}\sum\limits_{m_2'} W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2')\big)}\Bigg]\\
&\overset{(b)}{\leq}\frac{1}{2^{n(R_1+R_2)}}\sum_{(m_1,m_2)}\sum_{(x_1^n,x_2^n,z^n)} P_{X_1,X_2,Z}^{\otimes n}\big(x_1^n(m_1),x_2^n(m_2),z^n\big)\log\Bigg[\frac{W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),x_2^n(m_2)\big)}{2^{nR_1}W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)}\\
&+\mathbb{E}_{\setminus(m_1,m_2)}\left(\frac{\sum\limits_{\tilde{m}_2\neq m_2} W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),X_2^n(\tilde{m}_2)\big)}{2^{nR_1}W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)}\right)\\
&+\mathbb{E}_{\setminus m_2}\left(\frac{\sum\limits_{\tilde{m}_1} W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)}{2^{nR_1}\sum\limits_{m_2'} W_{Z|X_2}^{\otimes n}\big(z^n|X_2^n(m_2')\big)}+\frac{\sum\limits_{\tilde{m}_1}\sum\limits_{\tilde{m}_2\neq m_2} W_{Z|X_2}^{\otimes n}\big(z^n|X_2^n(\tilde{m}_2)\big)}{2^{nR_1}\sum\limits_{m_2'} W_{Z|X_2}^{\otimes n}\big(z^n|X_2^n(m_2')\big)}\right)\Bigg]\\
&=\frac{1}{2^{n(R_1+R_2)}}\sum_{(m_1,m_2)}\sum_{(x_1^n,x_2^n,z^n)} P_{X_1,X_2,Z}^{\otimes n}\big(x_1^n(m_1),x_2^n(m_2),z^n\big)\log\Bigg[\frac{W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),x_2^n(m_2)\big)}{2^{nR_1}W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)}\\
&+\mathbb{E}_{\setminus m_2}\left(\frac{\sum\limits_{\tilde{m}_2\neq m_2} W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),X_2^n(\tilde{m}_2)\big)}{2^{nR_1}W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)}\right)\\
&+\mathbb{E}_{\setminus m_2}\left(\frac{2^{nR_1}}{2^{nR_1}}\times\frac{W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)}{\sum\limits_{m_2'} W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2')\big)}+\frac{2^{nR_1}}{2^{nR_1}}\times\frac{\sum\limits_{\tilde{m}_2\neq m_2} W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(\tilde{m}_2)\big)}{\sum\limits_{m_2'} W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2')\big)}\right)\Bigg]\\
&=\frac{1}{2^{n(R_1+R_2)}}\sum_{(m_1,m_2)}\sum_{(x_1^n,x_2^n,z^n)} P_{X_1,X_2,Z}^{\otimes n}\big(x_1^n(m_1),x_2^n(m_2),z^n\big)\log\Bigg[\frac{W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),x_2^n(m_2)\big)}{2^{nR_1}W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)}\\
&+\frac{\sum\limits_{\tilde{m}_2\neq m_2} W_{Z|X_1}^{\otimes n}\big(z^n|x_1^n(m_1)\big)}{2^{nR_1}W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)}+1\Bigg]\\
&\leq\frac{1}{2^{n(R_1+R_2)}}\sum_{(m_1,m_2)}\sum_{(x_1^n,x_2^n,z^n)} P_{X_1,X_2,Z}^{\otimes n}\big(x_1^n(m_1),x_2^n(m_2),z^n\big)\log\Bigg[\frac{W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),x_2^n(m_2)\big)}{2^{nR_1}W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)}\\
&+\frac{2^{nR_2}W_{Z|X_1}^{\otimes n}\big(z^n|x_1^n(m_1)\big)}{2^{nR_1}W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)}+1\Bigg]\\
&\triangleq\Psi_1+\Psi_2
\end{aligned}
\tag{AA.8}
$$

where (a) follows from Jensen's inequality and (b) follows by taking expectation with respect to $\setminus m_1$, removing some terms from the denominator of the first and the second term in the log function, and adding one term to the nominator of the third and the fourth terms in the

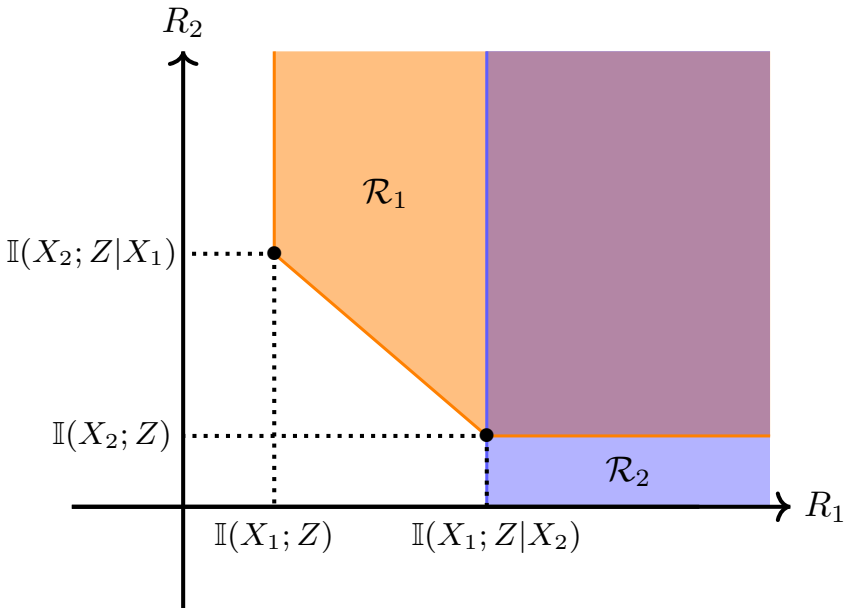

Figure AA.1. The orange region depicts $\mathcal{R}_1$ and the blue region depicts $\mathcal{R}_2$.

log function. We defined Ψ_1 and Ψ_2 as

$$\begin{aligned}\Psi_1 &= \frac{1}{2^{n(R_1+R_2)}} \sum_{(m_1,m_2)} \sum_{(x_1^n,x_2^n,z^n)\in\mathcal{T}_\epsilon^{(r)}} P_{X_1,X_2,Z}^{\otimes n}\big(x_1^n(m_1),x_2^n(m_2),z^n\big) \\ &\quad\times \log\left[\frac{W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),x_2^n(m_2)\big)}{2^{nR_1}W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)} + \frac{2^{nR_2}W_{Z|X_1}^{\otimes n}\big(z^n|x_1^n(m_1)\big)}{2^{nR_1}W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)} + 1\right] \\ &\le \log\left(\frac{2^{-n(1-\epsilon)\mathbb{H}(Z|X_1,X_2)}}{2^{nR_1}2^{-n(1+\epsilon)\mathbb{H}(Z|X_2)}} + \frac{2^{nR_2}2^{-n(1-\epsilon)\mathbb{H}(Z|X_1)}}{2^{nR_1}2^{-n(1+\epsilon)\mathbb{H}(Z|X_2)}} + 1\right) \end{aligned} \tag{AA.9}$$

and

$$\begin{aligned}\Psi_2 &= \frac{1}{2^{n(R_1+R_2)}} \sum_{(m_1,m_2)} \sum_{(x_1^n,x_2^n,z^n)\notin\mathcal{T}_\epsilon^{(r)}} P_{X_1,X_2,Z}^{\otimes n}\big(x_1^n(m_1),x_2^n(m_2),z^n\big) \\ &\quad\times \log\left[\frac{W_{Z|X_1X_2}^{\otimes n}\big(z^n|x_1^n(m_1),x_2^n(m_2)\big)}{2^{nR_1}W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)} + \frac{2^{nR_2}W_{Z|X_1}^{\otimes n}\big(z^n|x_1^n(m_1)\big)}{2^{nR_1}W_{Z|X_2}^{\otimes n}\big(z^n|x_2^n(m_2)\big)} + 1\right] \\ &\le 2|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Z}|e^{-n\epsilon^2\mu_{X_1,X_2,Z}}n\log(\frac{3}{\mu_Z}+1), \end{aligned} \tag{AA.10}$$

where

$$\mu_Z = \min_{\substack{z\in\mathcal{Z}\\ \text{s.t. } Q_z>0}} Q(z), \tag{AA.11}$$

$$\mu_{X_1,X_2,Z} = \min_{\substack{(x_1,x_2,z)\in(\mathcal{X}_1,\mathcal{X}_2,\mathcal{Z}) \\ \text{s.t. } Q(x_1,x_2,z)>0}} Q(x_1,x_2,z). \tag{AA.12}$$

When $r \to \infty$ then $\Psi_2 \to 0$ and $\Psi_1 \to 0$ when n grows if

$$R_1 > \mathbb{I}(X_1; Z|X_2), \tag{AA.13a}$$

$$R_1 > R_2 + \mathbb{I}(X_1; Z) - \mathbb{I}(X_2; Z). \tag{AA.13b}$$

When $R_2 \leq \mathbb{I}(X_2; Z)$, (AA.13b) is redundant because of (AA.13a). This is because the corner points $\big(R_1 = \mathbb{I}(X_1; Z|X_2), R_2 = 0\big)$ and $\big(R_1 = \mathbb{I}(X_1; Z|X_2), R_2 = \mathbb{I}(X_2; Z)\big)$ satisfy (AA.13b). As it can be seen from Fig. AA.1, $R_1 > \mathbb{I}(X_1; Z|X_2)$ is also achievable when $R_2 > \mathbb{I}(X_2; Z)$. This completes the achievability proof for the region $\mathcal{R}_2$.

APPENDIX AB

PROOF OF THEOREM 25

Fix $P_S(s)$, $P_U(u)$, $P_{V|US}(v|u,s)$, $x(u,s)$, and $\epsilon > 0$ subject to the conditions $P_Z = \Upsilon_Z$.

Codebook Generation: Let $C_1^{(n)} \triangleq \big\{U^n(k,m,\ell)\big\}_{(k,m,\ell)\in\mathcal{K}\times\mathcal{M}\times\mathcal{L}}$, where $\mathcal{K} = [\![1, 2^{nR_K}]\!]$, $\mathcal{M} = [\![1, 2^{nR}]\!]$, and $\mathcal{L} = [\![1, 2^{nR'}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_U^{\otimes n}$. We denote a realization of $C_1^{(n)}$ by $\mathcal{C}_1^{(n)} \triangleq \big\{u^n(k,m,\ell)\big\}_{(k,m,\ell)\in\mathcal{K}\times\mathcal{M}\times\mathcal{L}}$. The indices (k,m,ℓ) can be viewed as a two-layer binning.

Let $C_2^{(n)} \triangleq \big\{S^n(j)\big\}_{j\in\mathcal{J}}$, where $\mathcal{J} = [\![1, 2^{nR_J}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_S^{\otimes n}$. We denote a realization of $C_2^{(n)}$ by $\mathcal{C}_2^{(n)} \triangleq \big\{s^n(j)\big\}_{j\in\mathcal{J}}$.

Let, $C_n = \big\{C_1^{(n)}, C_2^{(n)}\big\}$ and $\mathcal{C}_n = \big\{\mathcal{C}_1^{(n)}, \mathcal{C}_2^{(n)}\big\}$. The set of all possible values of C_n is denoted by $\mathfrak{C}_n$. The codebook construction described above induces the PMF $\lambda(\mathcal{C}_n)$ for the codebooks,

$$\lambda(\mathcal{C}_n) = \prod_{j\in\mathcal{J}} P_S^{\otimes n}\big(s^n(j)\big) \prod_{(k,m,\ell)\in\mathcal{K}\times\mathcal{M}\times\mathcal{L}} P_U^{\otimes n}\Big(u^n\big(k,m,\ell\big)\Big). \tag{AB.1}$$

To facilitate the analysis, we define a so-called ideal joint PMF for all input, output, message, key, and auxiliary variables, conditioned on the choice of codebooks $\mathcal{C}_n$,

$$\begin{aligned}\Gamma^{(\mathcal{C}_n)}_{KMJLS^nU^nV^nZ^n}(k,m,j,\ell,\tilde{s}^n,\tilde{u}^n,v^n,z^n) &= 2^{-n(R_K+R+R_J+R')}\mathbb{1}_{\left\{\tilde{s}^n=s^n(j)\right\}\bigcap\left\{\tilde{u}^n=u^n(k,m,\ell)\right\}} \\ &\quad\times P_{V|S,U}^{\otimes n}(v^n|\tilde{s}^n,\tilde{u}^n)W_{Z|US}^{\otimes n}(z^n|\tilde{u}^n,\tilde{s}^n),\end{aligned} \tag{AB.2}$$

where $P_{V|S,U}$ is a test channel and $W_{Z|US}$ is the marginal distribution of $W_{Z,Y|U,S}$ defined in Theorem 25.

Encoding: The jammer selects an index j uniformly at random and transmits $s^n(j)$. The encoder cribs this s^n and, conditioned on it, generates a sequence v^n i.i.d. according to $P_{V|S}^{\otimes n}$. To do so, the encoder employs local randomness in a manner reminiscent of Csiszár and

Körner's stochastic encoder [70]. Then, given v^n as well as the cribbed signal $s^n(j)$, the key k, and the message m, the encoder chooses the index ℓ via a likelihood encoder [77, 78, 79], according to the following distribution:

$$f_{\mathrm{LE}}^{(\mathcal{C}_n)}(\ell|k,m,j,v^n) = \frac{P_{V|US}^{\otimes n}\big(v^n|u^n(k,m,\ell),s^n(j)\big)}{\sum\limits_{\ell'\in[\![1,2^{nR'}]\!]} P_{V|US}^{\otimes n}\big(v^n|u^n(k,m,\ell'),s^n(j)\big)}. \tag{AB.3}$$

Using the resulting index ℓ as well as the key k and message m, the encoder computes $u^n(k,m,\ell)$ and transmits codeword x^n, where $x_i = x(u_i(k,m,\ell),s_i)$. For a fixed codebook $\mathcal{C}_n$, the induced joint distribution is

$$\begin{aligned} P_{KMJS^nV^nLU^nZ^n}^{(\mathcal{C}_n)}(k,m,j,\tilde{s}^n,v^n,\ell,\tilde{u}^n,z^n) &= 2^{-n(R_K+R+R_J)}\mathbb{1}_{\left\{\tilde{s}^n=s^n(j)\right\}}P_{V|S}^{\otimes n}(v^n|\tilde{s}^n) \\ &\times f_{\mathrm{LE}}^{(\mathcal{C}_n)}(\ell|k,m,j,v^n)\mathbb{1}_{\left\{\tilde{u}^n=u^n(k,m,\ell)\right\}}W_{Z|US}^{\otimes n}(z^n|\tilde{u}^n,\tilde{s}^n). \end{aligned} \tag{AB.4}$$

Considering the random codebook generation, we have

$$P(\mathcal{C}_n,k,m,j,\tilde{s}^n,\ell,\tilde{u}^n,z^n) = \lambda(\mathcal{C}_n)P^{(\mathcal{C}_n)}(k,m,j,\tilde{s}^n,\ell,\tilde{u}^n,z^n), \tag{AB.5}$$

where $\lambda \in \mathcal{P}$ is defined in (AB.1).

Covert Analysis: We denote by $P^{(\mathcal{C}_n)}$ the distributions induced by a fixed codebook $\mathcal{C}_n$, and by $P_{\cdot|C_n}$ the distributions induced by a random codebook C_n. Consider a scenario in which the jammer selects a codeword from its codebook uniformly at random, and the transmitter chooses the innocent sequence x_0^n. Under a fixed codebook $\mathcal{C}_n$, the induced joint distribution is as follows

$$\Upsilon_{JS^nZ^n}^{(\mathcal{C}_n)}(j,s^n,z^n) = \frac{1}{2^{nR_J}}\mathbb{1}_{\{s^n=s^n(j)\}}W_{Z|X=x_0,S}^{\otimes n}\big(z^n|x_0^n,s^n\big).$$

Therefore, the distribution induced on the warden's observation by a random codebook is

$$\Upsilon_{Z^n|C_n}(z^n) = \frac{1}{2^{nR_J}}\sum_{j=1}^{2^{nR_J}} W_{Z|X=x_0,S}^{\otimes n}\big(z^n|x_0^n,S^n(j)\big). \tag{AB.6}$$

If $R_J > \mathbb{I}(S;Z)$ then according to the soft covering lemma [83, Theorem 4] or [77, Corollary VII.4],

$$\mathbb{E}_{C_n}\mathbb{V}\left(\Upsilon_{Z^n|C_n}, Q_0^{\otimes n}\right) \underset{n\to\infty}{\longrightarrow} 0, \tag{AB.7}$$

where

$$Q_0(\cdot) = \sum_{s\in\mathcal{S}} P_S(s) W_{Z|X=x_0,S}\big(\,\cdot\,|x_0, s\big). \tag{AB.8}$$

Note that if $R_K < \mathbb{I}_\Upsilon(S;Z)$ according to Shannon's channel coding theorem, the warden might be able to decode J, which reduces the problem to the point to point channel for which the covert rate will be zero. We aim to show that the coding scheme described above guarantees

$$\mathbb{E}_{C_n}\mathbb{D}\left(P_{Z^n|C_n}||\Upsilon_{Z^n|C_n}\right) \underset{n\to\infty}{\longrightarrow} 0. \tag{AB.9}$$

To show that (AB.9) holds by using Lemma 1 and the triangle inequality we have

$$\mathbb{E}_{C_n}\mathbb{V}\left(P_{Z^n|C_n}, \Upsilon_{Z^n|C_n}\right) \le \mathbb{E}_{C_n}\mathbb{V}\big(P_{Z^n|C_n}, Q_0^{\otimes n}\big) + \mathbb{E}_{C_n}\mathbb{V}\big(\Upsilon_{Z^n|C_n}, Q_0^{\otimes n}\big). \tag{AB.10}$$

According to the soft covering lemma [83, Theorem 4] or [77, Corollary VII.4], the second term on the RHS of (AB.10) vanishes when n grows if

$$R_J > \mathbb{I}_\Upsilon(S;Z). \tag{AB.11}$$

To bound the first term on the RHS of (AB.10) we first show

$$\mathbb{E}_{C_n}\mathbb{V}\left(P_{Z^n|C_n}, Q_Z^{\otimes n}\right) \underset{n\to\infty}{\longrightarrow} 0, \tag{AB.12}$$

and then we choose P_S, P_U, $P_{V|US}$ and $x(u,s)$ such that $P_Z = \Upsilon_Z$. To prove (AB.12) by using the triangle inequality,

$$\mathbb{E}_{C_n}\mathbb{V}\left(P_{Z^n|C_n}, Q_Z^{\otimes n}\right) \le \mathbb{E}_{C_n}\mathbb{V}\left(P_{Z^n|C_n}, \Gamma_{Z^n|C_n}\right) + \mathbb{E}_{C_n}\mathbb{V}\left(\Gamma_{Z^n|C_n}, Q_Z^{\otimes n}\right). \tag{AB.13}$$

We proceed to bound the first term on the RHS of (AB.13). For every codebook $\mathcal{C}_n$,

$$\Gamma^{(\mathcal{C}_n)}_{KMJ} = 2^{-n(R_K+R+R_J)} = P^{(\mathcal{C}_n)}_{KMJ}, \tag{AB.14a}$$

$$\Gamma^{(\mathcal{C}_n)}_{S^n|KMJ} = \mathbb{1}_{\left\{\tilde{s}^n = s^n(j)\right\}} = P^{(\mathcal{C}_n)}_{S^n|KMJ}, \tag{AB.14b}$$

$$\Gamma^{(\mathcal{C}_n)}_{L|KMJS^nV^n} = f^{(\mathcal{C}_n)}_{\mathrm{LE}}(\ell|k,m,j,v^n) = P^{(\mathcal{C}_n)}_{L|KMJS^nV^n}, \tag{AB.14c}$$

$$\Gamma^{(\mathcal{C}_n)}_{U^n|KMJS^nV^nL} = \mathbb{1}_{\left\{\tilde{u}^n = u^n(k,m,\ell)\right\}} = P^{(\mathcal{C}_n)}_{U^n|KMJS^nV^nL}, \tag{AB.14d}$$

$$\Gamma^{(\mathcal{C}_n)}_{Z^n|KMJS^nV^nLU^n} = W^{\otimes n}_{Z|US}(z^n|\tilde{u}^n,\tilde{s}^n) = P^{(\mathcal{C}_n)}_{Z^n|KMJS^nV^nLU^n}, \tag{AB.14e}$$

where (AB.14a)-(AB.14b) and (AB.14d)-(AB.14e) follow directly from (AB.2) and (AB.4) and (AB.14c) follow since for every codebook $\mathcal{C}_n$,

$$\begin{aligned}
\Gamma^{(\mathcal{C}_n)}_{L|KMJV^n}(\ell|k,m,j,v^n) &= \frac{\Gamma^{(\mathcal{C}_n)}_{KMLJV^n}(k,m,\ell,j,v^n)}{\Gamma^{(\mathcal{C}_n)}_{KMJV^n}(k,m,j,v^n)} \\
&= \frac{\sum\limits_{\tilde{u}^n} 2^{-n(R_K+R+R'+R_J)} \mathbb{1}_{\left\{\tilde{s}^n = s^n(j), \tilde{u}^n = u^n(k,m,\ell)\right\}} P^{\otimes n}_{V|SU}(v^n|\tilde{s}^n,\tilde{u}^n)}{\sum\limits_{\tilde{u}^n}\sum\limits_{\ell'} 2^{-n(R_K+R+R'+R_J)} \mathbb{1}_{\left\{\tilde{s}^n = s^n(j), \tilde{u}^n = u^n(k,m,\ell)\right\}} P^{\otimes n}_{V|SU}(v^n|\tilde{s}^n,\tilde{u}^n)} \\
&= \frac{P^{\otimes n}_{V|US}\big(v^n|u^n(k,m,\ell),s^n(j)\big)}{\sum\limits_{\ell' \in [\![1,2^{nR'}]\!]} P^{\otimes n}_{V|US}\big(v^n|u^n(k,m,\ell'),s^n(j)\big)} \\
&= f^{(\mathcal{C}_n)}_{\mathrm{LE}}(\ell|k,m,j,v^n).
\end{aligned} \tag{AB.15}$$

Thus, the first term on the RHS of (AB.13) is bounded as

$$\begin{aligned}
\mathbb{E}_{C_n}\mathbb{V}\big(P_{Z^n|C_n}, \Gamma_{Z^n|C_n}\big) &\le \mathbb{E}_{C_n}\mathbb{V}\big(P_{KMJS^nV^nLU^nZ^n|C_n}, \Gamma_{KMJS^nV^nLU^nZ^n|C_n}\big) \\
&\overset{(a)}{=} \mathbb{E}_{C_n}\mathbb{V}\big(P_{S^nV^nLU^nZ^n|K=1,M=1,J=1,C_n}, \Gamma_{S^nV^nLU^nZ^n|K=1,M=1,J=1,C_n}\big) \\
&\overset{(b)}{=} \mathbb{E}_{C_n}\mathbb{V}\big(P^{\otimes n}_{V|S}\big(\,\cdot\,|S^n(1)\big), \Gamma_{V^n|K=1,M=1,J=1,C_n}\big),
\end{aligned} \tag{AB.16}$$

where (a) follows from (AB.14a), the independence of (K,M,J) and C_n, and symmetry of codebook construction with respect to (K,M,J); and (b) follows from (AB.14b)-(AB.14e). According to Lemma 4 the RHS of (AB.16) vanishes when n grows if

$$R' > \mathbb{I}_P(U;V|S). \tag{AB.17}$$

This follows since conditioning on $M_2 = 1$ the distribution in (4.65) reduces to $P_{Z^n|M_2=1}(z^n) = W_{Z|X_2}^{\otimes n}(z^n|X_2^n(1))$ and the distribution in (4.67) reduces to

$$P_{Z^n|C_1,M_2=1}(z^n) \triangleq \sum_{m_1=1}^{2^{nR_1}} \frac{1}{2^{nR_1}} W_{Z|X_1X_2}^{\otimes n}\big(z^n|X_1^n(m_1), X_2^n(1)\big).$$

Also, according to [90, 23] the second term on the RHS of (AB.13) vanishes when n grows if

$$R_J > \mathbb{I}_P(S;Z), \tag{AB.18a}$$

$$R_K + R + R' > \mathbb{I}_P(U;Z), \tag{AB.18b}$$

$$R_K + R + R' + R_J > \mathbb{I}_P(U,S;Z). \tag{AB.18c}$$

Decoding and Error Probability Analysis: We show that the average probability of error can be made arbitrarily small. By access to the key K, the receiver declares that $\hat{M} = M$ if there exists a unique index $\hat{M}$ such that $(U^n(K,\hat{M},\ell), Y^n) \in \mathcal{T}_\epsilon^{(n)}(U,Y)$. Then the error event $(\hat{M} \neq M)$ occurs only if one or more of the following error events occur:

$$\mathcal{E}_1 \triangleq \{(U^n(K,M,L), V^n) \notin \mathcal{T}_{\epsilon'}^{(n)}(U,V)\}, \tag{AB.19a}$$

$$\mathcal{E}_2 \triangleq \{(U^n(K,M,L), Y^n) \notin \mathcal{T}_{\epsilon}^{(n)}(U,Y)\}, \tag{AB.19b}$$

$$\mathcal{E}_3 \triangleq \{(U^n(K,m,\ell), Y^n) \in \mathcal{T}_{\epsilon}^{(n)}(U,Y) \text{ for some } m \neq M \text{ and } \ell \in [1:2^{rR'}]\}. \tag{AB.19c}$$

Therefore, from the union bound we can bound the probability of error as follows,

$$\mathbb{P}(\hat{M} \neq M) \leq \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_1^c \cap \mathcal{E}_2) + \mathbb{P}(\mathcal{E}_3). \tag{AB.20}$$

According to the law of large numbers the second term on the RHS of (AB.20) goes to zero as $n \to \infty$ [85]. According to the law of large numbers the third term on the RHS of (AB.20) goes to zero as $n \to \infty$ if [85],

$$R + R' < \mathbb{I}_P(U;Y). \tag{AB.21}$$

We now show that the first term on the RHS of (AB.20) also vanishes as $n \to \infty$. For a fix $\epsilon > 0$, consider the PMF Γ defined in (AB.2). With respect to the random experiment described by Γ, we have

$$\mathbb{E}_{C_n}\mathbb{P}_{\Gamma}\Big(\big(U^n(m,k,L),V^n,S^n(j)\big) \notin \mathcal{T}_{\epsilon'}^{(n)}|C_n\Big) \underset{n\to\infty}{\longrightarrow} 0, \tag{AB.22}$$

this follows because V^n is derived by passing $U^n(k,m,L) \sim P_U^{\otimes n}$, for every $(m,k) \in (\mathcal{M},\mathcal{K})$, and $S^n(j) \sim P_S^{\otimes n}$, for every $j \in \mathcal{J}$, through the DMC $P_{V|US}^{\otimes n}$. Therefore, (AB.22) holds by weak law of large numbers. We also have

$$\begin{aligned}&\mathbb{E}_{C_n}\mathbb{V}\big(P_{U^nS^nV^n|C_n},\Gamma_{U^nS^nV^n|C_n}\big)\\ &\leq \mathbb{E}_{C_n}\mathbb{V}\big(P_{JKMS^nLU^nV^nZ^n|C_n},\Gamma_{JKMS^nLU^nV^nZ^n|C_n}\big),\end{aligned} \tag{AB.23}$$

where based on (AB.16) the RHS of (AB.23) vanishes when n grows.

We now define $g_n : \mathcal{U}^n \times \mathcal{V}^n \times \mathcal{S}^n \to \mathbb{R}$ as $g_n(u^n,s^n,v^n) \triangleq \mathbb{1}_{\{(u^n,s^n,v^n)\notin\mathcal{T}_{\epsilon'}^{(n)}\}}$. We now have

$$\begin{aligned}&\mathbb{E}_{C_n}\mathbb{P}_P\Big(\big(U^n(k,m,L),S^n(j),V^n\big) \notin \mathcal{T}_{\epsilon'}^{(n)}|C_n\Big)\\ &= \mathbb{E}_{C_n}\mathbb{E}_P\Big[g_n(U^n(k,m,L),S^n(j),V^n)|C_n\Big]\\ &\leq \mathbb{E}_{C_n}\mathbb{E}_{\Gamma}\Big[g_n(U^n(k,m,L),S^n(j),V^n)|C_n\Big] + \mathbb{E}_{C_n}\Big|\mathbb{E}_P\Big[g_n(U^n(k,m,L),S^n(j),V^n)|C_n\Big]\\ &\quad - \mathbb{E}_{\Gamma}\Big[g_n(U^n(k,m,L),S^n(j),V^n)|C_n\Big]\Big|\\ &\overset{(a)}{\leq} \mathbb{E}_{C_n}\mathbb{E}_{\Gamma}\Big[g_n(U^n(k,m,L),S^n(j),V^n)|C_n\Big] + \mathbb{E}_{C_n}\mathbb{V}\big(P_{U^nV^nS_j^n|C_n},\Gamma_{U^nV^nS_j^n|C_n}\big),\end{aligned} \tag{AB.24}$$

where (a) follows from [91, Property 1] for g_n being bounded by 1. From (AB.22) and (AB.23) the RHS of (AB.24) vanishes when n grows.

The region in Theorem 25 is derived by applying Fourier-Motzkin [71] to (AB.11), (AB.17), (AB.18), and (AB.21).

APPENDIX AC

PROOF OF THEOREM 26

Fix $P_S(s)$, $P_{U|S}(u|s)$, $x(u,s)$, and $\epsilon > 0$ such that, $P_Z = \Upsilon_Z$.

Codebook Generation: Let $C_1^{(n)} \triangleq \left\{U^n(k,m,\ell)\right\}_{(k,m,\ell)\in\mathcal{K}\times\mathcal{M}\times\mathcal{L}}$ be a random codebook consisting of independent random sequences each generated according to $P_U^{\otimes n}$. We denote a realization of $C_1^{(n)}$ by $\mathcal{C}_1^{(n)} \triangleq \left\{u^n(k,m,\ell)\right\}_{(k,m,\ell)\in\mathcal{K}\times\mathcal{M}\times\mathcal{L}}$. The indices (k,m,ℓ) can be viewed as a two layer binning.

Let $C_2^{(n)} \triangleq \left\{S^n(j)\right\}_{j\in\mathcal{J}}$ be a random codebook consisting of independent random sequences each generated according to $P_S^{\otimes n}$. We denote a realization of $C_2^{(n)}$ by $\mathcal{C}_2^{(n)} \triangleq \left\{s^n(j)\right\}_{j\in\mathcal{J}}$.

Also, $C_n = \left\{C_1^{(n)}, C_2^{(n)}\right\}$ denotes a random codebook and $\mathcal{C}_n = \left\{\mathcal{C}_1^{(n)}, \mathcal{C}_2^{(n)}\right\}$ denotes a fixed codebook. The set of all possible values of C_n is denoted by $\mathfrak{C}_n$. The codebook construction described above induces the PMF $\lambda \in \mathcal{P}(\mathfrak{C}_n)$ over the codebook ensemble. For each $\mathcal{C}_n \in \mathfrak{C}_n$

$$\lambda(\mathcal{C}_n) = \prod_{j\in\mathcal{J}} P_S^{\otimes n}\big(s^n(j)\big) \prod_{(k,m,\ell)\in\mathcal{K}\times\mathcal{M}\times\mathcal{L}} P_U^{\otimes n}\Big(u^n\big(k,m,\ell\big)\Big). \tag{AC.1}$$

Encoding: Given the jammer's channel input, the message m, and the key k, the encoder chooses the index ℓ according to

$$f^{(\mathcal{C}_n)}(\ell|s^n,k,m) = \frac{P_{S|U}^{\otimes n}\left(s^n|u^n(k,m,\ell)\right)}{\sum\limits_{\ell'\in[\![1,2^{nR'}]\!]} P_{S|U}^{\otimes n}\left(s^n|u^n(k,m,\ell')\right)}. \tag{AC.2}$$

Based on these indices, the encoder computes $u^n(k,m,\ell)$ and transmits codeword x^n, where $x_i = x(u_i(k,m,\ell), s_i)$. For a fixed codebook $\mathcal{C}_n$, the induced joint distribution is

$$\begin{aligned} P_{KMJS^nLU^nZ^n}^{(\mathcal{C}_n)}(k,m,j,\tilde{s}^n,\ell,\tilde{u}^n,z^n) \triangleq\ & 2^{-n(R_K+R+R_J)}\mathbb{1}_{\{\tilde{s}^n=s^n(j)\}} f(\ell|s^n(j),k,m,j) \\ & \times \mathbb{1}_{\{\tilde{u}^n=u^n(k,m,\ell)\}} W_{Z|US}^{\otimes n}(z^n|\tilde{u}^n,\tilde{s}^n). \end{aligned} \tag{AC.3}$$

Considering the random codebook generation, we have

$$P(k,m,j,\tilde{s}^n,\ell,\tilde{u}^n,z^n) = \lambda(\mathcal{C}_n) P^{(\mathcal{C}_n)}(k,m,j,\tilde{s}^n,\ell,\tilde{u}^n,z^n), \tag{AC.4}$$

where $\lambda \in \mathcal{P}(\mathfrak{C}_n)$ is defined in (AC.1).

Covert and Security Analysis: Throughout the proof, we use $P^{(\mathcal{C}_n)}$ when the codebook is fixed, and we use $P_{\cdot|C_n}$ when the codebook is random. Consider a scenario in which the jammer selects a codeword from its codebook (i.e. $\mathcal{C}_2^{(n)}$) uniformly at random, but the transmitter chooses the innocent sequence x_0^n as the channel input. For this scenario for a fixed codebook $\mathcal{C}_2^{(n)}$, the induced joint distribution is as follows

$$\Upsilon^{(\mathcal{C}_n)}_{JS^nZ^n}(j, s^n, z^n) = \frac{1}{2^{nR_J}} \mathbb{1}_{\{s^n = s^n(j)\}} W^{\otimes n}_{Z|XS}\big(z^n | x_0^n, s^n(j)\big). \tag{AC.5}$$

Therefore, the distribution induced at the output of the warden is

$$\Upsilon^{(\mathcal{C}_n)}_{Z^n}(z^n) = \frac{1}{2^{nR_J}} \sum_{j=1}^{2^{nR_J}} W^{\otimes n}_{Z|XS}\big(z^n | x_0^n, s^n(j)\big). \tag{AC.6}$$

For this scenario if

$$R_J > \mathbb{I}_{\Upsilon}(S; Z). \tag{AC.7}$$

Then according to soft covering lemma [83, Theorem 4] or [77, Corollary VII.4] and Pinsker's inequality, we have

$$\mathbb{E}_{C_n} \mathbb{V}\left(\Upsilon_{Z^n|C_n}, Q_0^{\otimes n}\right) \underset{n\to\infty}{\longrightarrow} 0, \tag{AC.8}$$

where $Q_0^{\otimes n} = \prod_{i=1}^n Q_0$ and

$$Q_0(\cdot) = \sum_{s\in\mathcal{S}} P_S(s) W_{Z|X=x_0,S}\big(\,\cdot\,|x_0, s\big). \tag{AC.9}$$

Note that if $R_K < \mathbb{I}_{\Upsilon}(S; Z)$ according to Shannon's channel coding theorem, the warden might be able to decode J, which reduces the problem to the point to point channel for which the covert rate will be zero.

We now show that this coding scheme guarantees both covert and secure communication, i.e., $\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||Q_0^{\otimes n})] \underset{n\to\infty}{\longrightarrow} 0$ and $\mathbb{I}_P(M; Z^n) \underset{n\to\infty}{\longrightarrow} 0$. For every codebook $\mathcal{C}_n \in \mathfrak{C}_n$ we have

$$\mathbb{D}\Big(P^{(\mathcal{C}_n)}_{Z^n}\Big|\Big|Q_Z^{\otimes n}\Big) \overset{(a)}{\leq} \mathbb{D}\Big(P^{(\mathcal{C}_n)}_{Z^nMJS^n}\Big|\Big|P^{(\mathcal{C}_n)}_{MJS^n} Q_{Z|S}^{\otimes n}\Big),$$

$$= \mathbb{D}\Big(P^{(\mathcal{C}_n)}_{Z^n|MJS^n}\Big|\Big|Q^{\otimes n}_{Z|S}\Big|P^{(\mathcal{C}_n)}_{MJS^n}\Big), \tag{AC.10}$$

where (a) follows from the monotonicity of KL-divergence and

$$Q_Z(z) = \sum_{s\in\mathcal{S}}\sum_{u\in\mathcal{U}}\sum_{x\in\mathcal{X}} P_S(s)P_{U|S}(u|s)\mathbb{1}_{\{x=x(u,s)\}}W_{Z|XS}\big(z|x,s\big), \tag{AC.11a}$$

$$Q_{Z|S}(z|s) = \sum_{u\in\mathcal{U}}\sum_{x\in\mathcal{X}} P_{U|S}(u|s)\mathbb{1}_{\{x=x(u,s)\}}W_{Z|XS}(z|x,s). \tag{AC.11b}$$

Also, we have

$$\begin{aligned}
\mathbb{I}(M;Z^n) &\leq \mathbb{I}(M;J,S^n,Z^n),\\
&= \mathbb{I}(M;Z^n|J,S^n),\\
&= \mathbb{D}\Big(P^{(\mathcal{C}_n)}_{Z^n|MJS^n}\Big|\Big|P^{(\mathcal{C}_n)}_{Z^n|JS^n}\Big|P^{(\mathcal{C}_n)}_{MJS^n}\Big)\\
&\overset{(a)}{\leq} \mathbb{D}\Big(P^{(\mathcal{C}_n)}_{Z^n|MJS^n}\Big|\Big|Q^{\otimes n}_{Z|S}\Big|P^{(\mathcal{C}_n)}_{MJS^n}\Big),
\end{aligned} \tag{AC.12}$$

where (a) follows from

$$\mathbb{D}\Big(P^{(\mathcal{C}_n)}_{Z^n|MJS^n}\Big|\Big|P^{(\mathcal{C}_n)}_{Z^n|JS^n}\Big|P^{(\mathcal{C}_n)}_{MJS^n}\Big) = \mathbb{D}\Big(P^{(\mathcal{C}_n)}_{Z^n|MJS^n}\Big|\Big|Q^{\otimes n}_{Z|S}\Big|P^{(\mathcal{C}_n)}_{MJS^n}\Big) - \mathbb{D}\Big(P^{(\mathcal{C}_n)}_{Z^n|JS^n}\Big|\Big|Q^{\otimes n}_{Z|S}\Big|P^{(\mathcal{C}_n)}_{JS^n}\Big). \tag{AC.13}$$

From (AC.10) and (AC.12) it follows

$$\mathbb{I}_P(M;Z^n) + \mathbb{D}\Big(P^{(\mathcal{C}_n)}_{Z^n}\Big|\Big|Q^{\otimes n}_{Z}\Big) \leq 2\mathbb{D}\Big(P^{(\mathcal{C}_n)}_{Z^n|MJS^n}\Big|\Big|Q^{\otimes n}_{Z|S}\Big|P^{(\mathcal{C}_n)}_{MJS^n}\Big). \tag{AC.14}$$

Taking the expectation of the RHS of (AC.14) with respect to the ensemble of codebooks results to

$$\begin{aligned}
&\mathbb{E}_{C_n}\Big[\mathbb{D}\Big(P_{Z^n|MJS^n(J)C_n}\Big|\Big|Q^{\otimes n}_{Z|S}\Big|P_{MJS^n(J)C_n}\Big)\Big]\\
&= \mathbb{E}_{C_n}\Big[\sum_{(m,j,s^n)} 2^{-n(R+R_J)}\mathbb{1}_{\{S^n(j)=s^n\}}\mathbb{D}\Big(P_{Z^n|M=m,J=j,S^n(j)=s^n,C_n}\Big|\Big|Q^{\otimes n}_{Z|S}(\cdot|s^n)\Big)\Big]\\
&\overset{(a)}{=} \sum_{s^n}\mathbb{E}_{C_n}\Big[\mathbb{1}_{\{S^n(1)=s^n\}}\mathbb{D}\Big(P_{Z^n|M=1,J=1,S^n(1)=s^n,C_n}\Big|\Big|Q^{\otimes n}_{Z|S}(\cdot|s^n)\Big)\Big]
\end{aligned}$$

$$\overset{(b)}{=} \sum_{s^n} \mathbb{E}_{C_2^{(n)}}\Big[\mathbb{1}_{\{S^n(1)=s^n\}} \mathbb{E}_{C_1^{(n)}|C_2^{(n)}}\Big[\mathbb{D}\Big(P_{Z^n|M=1,J=1,S^n(1)=s^n,C_n}\Big|\Big|Q_{Z|S}^{\otimes n}(\cdot|s^n)\Big)\Big]\Big] \tag{AC.15}$$

where (a) is due to the symmetry of the codebook with respect to J; and (b) follows by the law of total expectation. Thus far, when $\mathcal{C}_n \in \mathfrak{C}_n$ is fixed $P^{(\mathcal{C}_n)}_{Z^n|M=1,J=1,S^n(1)=s^n}$ is defined only when $s^n = s^n(1)$. For any other s^n, we can set this conditional PMF to any arbitrary PMF on $\mathcal{Z}^n$. Therefore, when $s^n \neq s^n(1)$, we define

$$P^{(\mathcal{C}_n)}_{Z^n|M=1,J=1,S^n(1)=s^n} = Q_{Z|S}^{\otimes n}(\cdot|s^n). \tag{AC.16}$$

For any $C_2^{(n)} = \mathcal{C}_2^{(n)}$ and $s^n \in \mathcal{S}^n$ we have

$$\begin{aligned}
&\mathbb{E}_{C_1^{(n)}|C_2^{(n)}=\mathcal{C}_2^{(n)}}\Big[\mathbb{D}\Big(P_{Z^n|M=1,J=1,S^n(1)=s^n,C_n}\Big|\Big|Q_{Z|S}^{\otimes n}(\cdot|s^n)\Big)\Big] \\
&= \mathbb{E}_{C_1^{(n)}|C_2^{(n)}=\mathcal{C}_2^{(n)}}\Big[\mathbb{1}_{\{s^n(1)=s^n\}}\mathbb{D}\Big(P_{Z^n|M=1,J=1,S^n(1)=s^n,C_n}\Big|\Big|Q_{Z|S}^{\otimes n}(\cdot|s^n)\Big) \\
&\quad + \mathbb{1}_{\{s^n(1)\neq s^n\}}\mathbb{D}\Big(P_{Z^n|M=1,J=1,S^n(1)=s^n,C_n}\Big|\Big|Q_{Z|S}^{\otimes n}(\cdot|s^n)\Big)\Big] \\
&\overset{(a)}{=} \mathbb{E}_{C_1^{(n)}|S^n(1)=s^n(1)}\Big[\mathbb{1}_{\{s^n(1)=s^n\}}\mathbb{D}\Big(P_{Z^n|M=1,J=1,S^n(1)=s^n,C_1^{(n)}}\Big|\Big|Q_{Z|S}^{\otimes n}(\cdot|s^n)\Big)\Big] \\
&\overset{(b)}{\leq} \mathbb{E}_{C_1^{(n)}|S^n(1)=s^n(1)}\Big[\mathbb{D}\Big(P_{Z^n|M=1,J=1,S^n(1)=s^n,C_1^{(n)}}\Big|\Big|Q_{Z|S}^{\otimes n}(\cdot|s^n)\Big)\Big]
\end{aligned} \tag{AC.17}$$

where (a) follows from (AC.16) and because conditioned on $S^n(1)$, $P_{Z^n|J=1,S^n(1)=s^n}$ is independent of all the other codewords in $C_2^{(n)}$ and (b) follows by $\mathbb{1}_{\{\cdot\}} \leq 1$.

We now show that this problem falls within the framework of [84, Lemma 1]. Let $\tilde{C}_n \triangleq \{\tilde{U}^n(k,\ell)\}_{(k,\ell)\in\mathcal{K}\times\mathcal{L}}$ be a set of i.i.d. sequences distributed according to $P_U^{\otimes n}(\cdot)$. The set $\tilde{C}_n$ is independent of C_n and is distributed according to

$$\tilde{\lambda}(\tilde{\mathcal{C}}_n) = \prod_{(k,\ell)\in\mathcal{K}\times\mathcal{L}} P_U^{\otimes n}\big(\tilde{u}^n(k,\ell)\big). \tag{AC.18}$$

where $\tilde{\mathcal{C}}_n \triangleq \{\tilde{u}^n(k,\ell)\}_{(k,\ell)\in\mathcal{K}\times\mathcal{L}}$ is a realization of $\tilde{C}_n$. For each $s^n \in \mathcal{S}^n$ let

$$\tilde{P}^{(\tilde{\mathcal{C}}_n)}(k,\ell,\tilde{u}^n,z^n|s^n) = 2^{-nR_K}\tilde{f}^{(\tilde{\mathcal{C}}_n)}(\ell|k,s^n)\mathbb{1}_{\{\tilde{u}^n=\tilde{u}^n(k,\ell)\}}W_{Z|US}^n(z^n|\tilde{u}^n,s^n), \tag{AC.19}$$

where $f^{(\tilde{\mathcal{C}}_n)}(\ell|k,s^n)$ is defined as

$$f^{(\tilde{\mathcal{C}}_n)}(\ell|s^n,k,m)=\frac{P_{S|U}^{\otimes n}\left(s^n|u^n(k,m,\ell)\right)}{\sum\limits_{\ell'\in[\![1,2^{nR'}]\!]}P_{S|U}^{\otimes n}\left(s^n|u^n(k,m,\ell')\right)}, \tag{AC.20}$$

Also, let

$$\tilde{P}(k,\ell,\tilde{u}^n,z^n|s^n)=\tilde{\lambda}(\tilde{\mathcal{C}}_n)\tilde{P}^{(\tilde{\mathcal{C}}_n)}(k,\ell,\tilde{u}^n,z^n|s^n). \tag{AC.21}$$

Now for any $s^n\in\mathcal{S}^n$ we further bound the RHS of (AC.17) by

$$\mathbb{E}_{\tilde{C}_n}\Big[\mathbb{D}\Big(\tilde{P}_{Z^n|S^n=s^n,\tilde{C}_n}\Big|\Big|Q_{Z|S}^{\otimes n}(\cdot|s^n)\Big)\Big]. \tag{AC.22}$$

This is because when $\mathcal{C}_1^{(n)}=\tilde{\mathcal{C}}_n$ the distribution $P_{Z^n|M=1,J=1,S^n(1)=s^n,C_1^{(n)}=\mathcal{C}_1^{(n)}}$ and $\tilde{P}_{Z^n|S^n=s^n,\tilde{C}_n=\tilde{\mathcal{C}}_n}$ are equal as PMFs on $\mathcal{Z}^n$. Substituting (AC.22) in (AC.15) we have

$$\begin{aligned}
&\mathbb{E}_{C_n}\Big[\mathbb{D}\Big(P_{Z^n|MJS^n(J)C_n}\Big|\Big|Q_{Z|S}^{\otimes n}\Big|P_{MJS^n(J)C_n}\Big)\Big]\\
&\leq\sum_{s^n}\mathbb{E}_{C_2^{(n)}}\Big[\mathbb{1}_{\{S^n(1)=s^n\}}\Big]\mathbb{E}_{\tilde{C}_n}\Big[\mathbb{D}\Big(\tilde{P}_{Z^n|S^n=s^n,\tilde{C}_n}\Big|\Big|Q_{Z|S}^{\otimes n}(\cdot|s^n)\Big)\Big]\\
&=\sum_{s^n}P_S^{\otimes n}(s^n)\mathbb{E}_{\tilde{C}_n}\Big[\mathbb{D}\Big(\tilde{P}_{Z^n|S^n=s^n,\tilde{C}_n}\Big|\Big|Q_{Z|S}^{\otimes n}(\cdot|s^n)\Big)\Big]\\
&=\mathbb{E}_{\tilde{C}_n}\Big[\mathbb{D}\Big(\tilde{P}_{Z^n|S^n\tilde{C}_n}\Big|\Big|Q_{Z|S}^{\otimes n}\Big|P_S^{\otimes n}\Big)\Big].
\end{aligned} \tag{AC.23}$$

Substituting (AC.23) back into (AC.14) yields to

$$\mathbb{E}_{C_n}\Big[\mathbb{I}(M;Z^n|C_n)+\mathbb{D}\Big(P_{Z^n|C_n}\Big|\Big|Q_Z^{\otimes n}\Big)\Big]\leq 2\mathbb{E}_{\tilde{C}_n}\Big[\mathbb{D}\Big(\tilde{P}_{Z^n|S^n\tilde{C}_n}\Big|\Big|Q_{Z|S}^{\otimes n}\Big|P_S^n\Big)\Big]. \tag{AC.24}$$

From [84, Lemma 1], while seeing $W_{Z|US}$ as a DMC with state from U to Z with state S, the RHS of (AC.24), and therefore the LHS of (AC.24), vanishes when n grows if

$$R'>\mathbb{I}_P(U;S) \tag{AC.25a}$$

$$R_K+R+R'>\mathbb{I}_P(U;S,Z). \tag{AC.25b}$$

To proof covertness $\mathbb{E}_{C_n}\mathbb{D}\left(P_{Z^n|C_n}||\Upsilon_{Z^n|C_n}\right) \underset{n\to\infty}{\longrightarrow} 0$, by using Lemma 1 and the triangle inequality we have,

$$\mathbb{E}_{C_n}\mathbb{V}\left(P_{Z^n|C_n}, \Upsilon_{Z^n|C_n}\right) \leq \mathbb{E}_{C_n}\mathbb{V}\left(P_{Z^n|C_n}, Q_0^{\otimes n}\right) + \mathbb{E}_{C_n}\mathbb{V}\left(\Upsilon_{Z^n|C_n}, Q_0^{\otimes n}\right). \tag{AC.26}$$

Using Pinsker inequality and (AC.24) the first term on the RHS of (AC.26) vanishes when n grows if we choose $P_S(s)$, $P_{U|S}(u|s)$, and $x(u,s)$ such that $Q_Z = Q_0$ and (AC.25) holds. Also, from (AC.8) the second term on the RHS of (AC.26) vanishes when n grows.

Decoding and Error Probability Analysis: By following the same steps as in [37], the probability of error vanishes when n grows if

$$R + R' < \mathbb{I}_P(U;Y). \tag{AC.27}$$

The region in Theorem 26 is derived by applying Fourier-Motzkin to (AC.7), (AC.25), and (AC.27).

APPENDIX AD

PROOF OF THEOREM 27

To prove the upper bound for the case that the jammer knows in which blocks the transmitter is communicating with the receiver and has an unlimited source of local randomness and transmits an i.i.d. sequence when communication is not happening. Consider any sequence of codes with length n such that $P_e^{(n)} \leq \epsilon_n$ and $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \delta$ where $\lim_{n\to\infty} \epsilon_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish.

Epsilon Rate Region: First we define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ which extends the region defined in (4.77) *as follows.*

$$\mathcal{A}_\epsilon = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists P_{UVSXYZ} \in \mathcal{D}_\epsilon : \\ R \leq \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V) + \epsilon \\ R_K \geq \mathbb{I}(V;Z) - \mathbb{I}(V;S) - \mathbb{I}(U,V;Y) + \mathbb{I}(U;S|V) - 3\epsilon \\ R_K + R_J \geq \mathbb{I}(V;Z) - \mathbb{I}(U,V;Y) + \mathbb{I}(U;S|V) - 3\epsilon \\ R_J \geq \mathbb{I}(S;Z) - 2\epsilon \end{array} \right\}, \tag{AD.1a}$$

where

$$\mathcal{D}_\epsilon = \left\{ \begin{array}{l} P_{UVSXYZ} : \\ P_{UVSXYZ} = P_{SUV} \mathbb{1}_{\left\{X = X(U,S)\right\}} W_{YZ|XS} \\ \mathbb{D}\left(P_Z \| Q_0\right) \leq \epsilon \\ \max\{|\mathcal{U}|, |\mathcal{V}|\} \leq |\mathcal{X}| + 3 \end{array} \right\}. \tag{AD.1b}$$

We next show that if a rate R is achievable, then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$, we start by upper bounding nR using standard techniques,

$$nR = \mathbb{H}(M)$$

$$
\begin{aligned}
&= \mathbb{H}(M|K) \\
&\overset{(a)}{\leq} \mathbb{I}(M; Y^n|K) + n\epsilon_n \\
&= \sum_{i=1}^{n} \mathbb{I}(M; Y_i|K, Y^{i-1}) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} \mathbb{I}(M, K, Y^{i-1}, Z^{i-1}; Y_i) + n\epsilon_n \\
&= \sum_{i=1}^{n} \big[\mathbb{I}(M, K, Y^{i-1}, Z^{i-1}, S_{i+1}^n; Y_i) - \mathbb{I}(S_{i+1}^n; Y_i|M, K, Y^{i-1}, Z^{i-1})\big] + n\epsilon_n \\
&\overset{(b)}{=} \sum_{i=1}^{n} \big[\mathbb{I}(M, K, Y^{i-1}, Z^{i-1}, S_{i+1}^n; Y_i) - \mathbb{I}(Y^{i-1}; S_i|M, K, S_{i+1}^n, Z^{i-1})\big] + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} [\mathbb{I}(U_i, V_i; Y_i) - \mathbb{I}(U_i; S_i|V_i)] + n\epsilon_n \\
&= n \sum_{i=1}^{n} \frac{1}{n} \big[\mathbb{I}(U_i, V_i; Y_i|T = i) - \mathbb{I}(U_i; S_i|V_i, T = i)\big] + n\epsilon_n \\
&= n \sum_{i=1}^{n} \mathbb{P}(T = i) \big[\mathbb{I}(U_i, V_i; Y_i|T = i) - \mathbb{I}(U_i; S_i|V_i, T = i)\big] + n\epsilon_n \\
&= n\big[\mathbb{I}(U_T, V_T; Y_T|T) - \mathbb{I}(U_T; S_T|V_T, T)\big] + n\epsilon_n \\
&\leq \big[\mathbb{I}(U_T, V_T, T; Y_T) - \mathbb{I}(U_T; S_T|V_T, T)\big] + n\epsilon_n \\
&\overset{(d)}{=} n\big[\mathbb{I}(U, V; Y) - \mathbb{I}(U; S|V)\big] + n\epsilon_n \\
&\overset{(e)}{\leq} n\big[\mathbb{I}(U, V; Y) - \mathbb{I}(U; S|V)\big] + n\epsilon,
\end{aligned}
\tag{AD.2}
$$

where

(a) *follows from Fano's inequality and the entropy function property that conditioning does not increase entropy;*

(b) *follows from Csiszár-Körner sum identity [70, Lemma 7];*

(c) *follows by defining* $U_i \triangleq (M, K, Y^{i-1}, S_{i+1}^n)$ *and* $V_i \triangleq (M, K, Z^{i-1}, S_{i+1}^n)$*;*

(d) *follows by defining* $U = (U_T, T)$, $V = (V_T, T)$, $Y = Y_T$, *and* $S = (S_T, T)$*;*

(e) *follows from definition* $\epsilon \triangleq \max\{\epsilon_n, \nu\}$.

Next, we lower bound $n(R+R_K)$ *as follows,*

$$\begin{aligned}
n(R+R_K) &\geq \mathbb{H}(M,K) \\
&\geq \mathbb{I}(M,K;Z^n) \\
&= \sum_{i=1}^{n} \mathbb{I}(M,K;Z_i|Z^{i-1}) \\
&= \sum_{i=1}^{n} \big[\mathbb{I}(M,K,S_{i+1}^n;Z_i|Z^{i-1}) - \mathbb{I}(S_{i+1}^n;Z_i|M,K,Z^{i-1})\big] \\
&\overset{(a)}{=} \sum_{i=1}^{n} \big[\mathbb{I}(M,K,S_{i+1}^n;Z_i|Z^{i-1}) - \mathbb{I}(Z^{i-1};S_i|M,K,S_{i+1}^n)\big] \\
&\overset{(b)}{\geq} \sum_{i=1}^{n} \big[\mathbb{I}(M,K,S_{i+1}^n,Z^{i-1};Z_i) - \mathbb{I}(Z^{i-1};S_i|M,K,S_{i+1}^n)\big] - \delta \\
&\overset{(c)}{\geq} \sum_{i=1}^{n} \big[\mathbb{I}(M,K,S_{i+1}^n,Z^{i-1};Z_i) - \mathbb{I}(M,K,S_{i+1}^n,Z^{i-1};S_i)\big] - \delta \\
&\overset{(d)}{=} \sum_{i=1}^{n} \big[\mathbb{I}(V_i;Z_i) - \mathbb{I}(V_i;S_i)\big] - \delta \\
&= n\sum_{i=1}^{n} \frac{1}{n}\big[\mathbb{I}(V_T;Z_T|T=i) - \mathbb{I}(V_T;S_T|T=i)\big] - \delta \\
&= n\sum_{i=1}^{n} \mathbb{P}(T=i)\big[\mathbb{I}(V_T;Z_T|T=i) - \mathbb{I}(V_T;S_T|T=i)\big] - \delta \\
&= n\big[\mathbb{I}(V_T;Z_T|T) - \mathbb{I}(V_T;S_T|T)\big] - \delta \\
&\geq n\big[\mathbb{I}(V_T;Z_T|T) - \mathbb{I}(V_T,T;S_T)\big] - \delta \\
&\overset{(e)}{\geq} n\big[\mathbb{I}(V_T,T;Z_T) - \mathbb{I}(V_T,T;S_T)\big] - 2\delta \\
&\overset{(f)}{=} n\big[\mathbb{I}(V;Z) - \mathbb{I}(V;S)\big] - 2\delta
\end{aligned} \tag{AD.3}$$

where

(a) *follows from Csiszár-Körner sum identity [70, Lemma 7];*

(b) *follows from [80, Lemma 3];*

(c) *follows since* S_i *is independent of* (M, K, S_{i+1}^n)*;*

(d) *follows by defining* $V_i \triangleq (M, K, S_{i+1}^n, Z^{i-1})$*;*

(e) *follows from [80, Lemma 3];*

(f) *follows by defining* $V = (V_T, T)$*,* $Z = Z_T$*, and* $S = (S_T, T)$*.*

For any $\nu > 0$*, choosing* n *large enough ensures that*

$$\begin{aligned} R + R_K &\geq \mathbb{I}(V; Z) - \mathbb{I}(V; S) - 2\nu \\ &\geq \mathbb{I}(V; Z) - \mathbb{I}(V; S) - 2\epsilon, \end{aligned} \tag{AD.4}$$

where the last inequality follows from definition $\epsilon \triangleq \max\{\epsilon_n, \nu\}$*. Next, we have*

$$\begin{aligned} n(R + R_K + R_J) &\geq \mathbb{H}(M, K, J) \\ &\geq \mathbb{I}(M, K, J; Z^n) \\ &\overset{(a)}{=} \mathbb{I}(M, K, J, S^n; Z^n) \\ &\geq \mathbb{I}(M, K, S^n; Z^n) \\ &= \sum_{i=1}^{n} [\mathbb{H}(Z_i|Z^{i-1}) - \mathbb{H}(Z_i|M, K, Z^{i-1}, S^n)] \\ &\overset{(b)}{\geq} \sum_{i=1}^{n} [\mathbb{H}(Z_i) - \mathbb{H}(Z_i|M, K, Z^{i-1}, S^n)] - \delta \\ &\geq \sum_{i=1}^{n} [\mathbb{H}(Z_i) - \mathbb{H}(Z_i|M, K, Z^{i-1}, S_{i+1}^n)] - \delta \\ &\overset{(c)}{=} \sum_{i=1}^{n} \mathbb{I}(V_i; Z_i) - \delta \\ &= n \sum_{i=1}^{n} \frac{1}{n} \mathbb{I}(V_i; Z_i|T = i) - \delta \\ &= n \sum_{i=1}^{n} \mathbb{P}(T = i) \mathbb{I}(V_i; Z_i|T = i) - \delta \\ &= n\mathbb{I}(V_T; Z_T|T) - \delta \end{aligned}$$

$$\begin{aligned}
&\overset{(d)}{\geq} n\mathbb{I}(V_T, T; Z_T) - 2\delta \\
&\overset{(e)}{=} n\mathbb{I}(V; Z) - 2\delta
\end{aligned} \tag{AD.5}$$

where

(a) *follows because* S^n *is a function of* J*;*

(b) *and* (d) *follow from [80, Lemma 3];*

(c) *follows by defining* $V_i \triangleq (M, K, S_{i+1}^n, Z^{i-1})$*;*

(e) *follows by defining* $V = (V_T, T)$ *and* $Z = Z_T$*.*

For any $\nu > 0$*, choosing* n *large enough ensures that*

$$\begin{aligned}
R + R_K + R_J &\geq \mathbb{I}(V; Z) - 2\nu \\
&\geq \mathbb{I}(V; Z) - 2\epsilon,
\end{aligned} \tag{AD.6}$$

where the last inequality follows from definition $\epsilon \triangleq \max\{\epsilon_n, \nu\}$*. We now have,*

$$\begin{aligned}
nR_J &\geq \mathbb{H}(J) \\
&\geq \mathbb{I}(J; Z^n) \\
&\overset{(a)}{=} \mathbb{I}(J, S^n; Z^n) \\
&\geq \mathbb{I}(S^n; Z^n) \\
&= \sum_{i=1}^{n} \left[\mathbb{H}(Z_i|Z^{i-1}) - \mathbb{H}(Z_i|S^n, Z^{i-1})\right] \\
&\overset{(b)}{\geq} \sum_{i=1}^{n} \left[\mathbb{H}(Z_i) - \mathbb{H}(Z_i|S^n, Z^{i-1})\right] - \delta \\
&\geq \sum_{i=1}^{n} \left[\mathbb{H}(Z_i) - \mathbb{H}(Z_i|S_i)\right] - \delta \\
&= n\sum_{i=1}^{n} \frac{1}{n}\mathbb{I}(S_i; Z_i) - \delta
\end{aligned}$$

$$
\begin{aligned}
&= n \sum_{i=1}^{n} \mathbb{P}(T=i)\mathbb{I}(S_T; Z_T|T=i) - \delta \\
&= n\mathbb{I}(S_T; Z_T|T) - \delta \\
&\overset{(c)}{\geq} n\mathbb{I}(S_T, T; Z_T) - 2\delta \\
&\overset{(d)}{=} n\mathbb{I}(S; Z) - 2\delta
\end{aligned} \tag{AD.7}
$$

where

(*a*) *follows because* S^n *is a function of* J*;*

(*b*) *and* (*c*) *follows from [80, Lemma 3];*

(*d*) *follows by defining* $S = (S_T, T)$ *and* $Z = Z_T$*.*

From (AD.7) *for any* $\nu > 0$*, choosing* n *large enough ensures that*

$$
\begin{aligned}
R &\geq \mathbb{I}(S; Z) - 2\nu, \\
&\geq \mathbb{I}(S; Z) - 2\epsilon,
\end{aligned} \tag{AD.8}
$$

where the last equality follows from the definition of $\epsilon \triangleq \max\{\epsilon_n, \nu\}$*. To prove Covertness (i.e.,* $\mathbb{D}(P_Z||Q_0) \leq \epsilon$*), note that for* n *large enough*

$$
\begin{aligned}
\mathbb{D}(P_Z||Q_0) = \mathbb{D}(P_{Z_T}||Q_0) &= \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} P_{Z_i}\middle|\middle|Q_0\right) \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \mathbb{D}(P_{Z_i}||Q_0) \leq \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \frac{\delta}{n} \leq \nu \leq \epsilon.
\end{aligned} \tag{AD.9}
$$

Combining (AD.2), (AD.4), (AD.6), (AD.8), *and* (AD.9) *shows that* $\forall \epsilon_n, \nu > 0$, $R \leq \max\{x : x \in \mathcal{A}_\epsilon\}$*. Therefore,*

$$
C_{\text{IJ-NC}} \leq \max\left\{x : x \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon\right\}. \tag{AD.10}
$$

Continuity at Zero: The proof for continuity at zero of $\mathcal{A}_\epsilon$ is similar to that of Appendix Y and is omitted for the sake of brevity.

APPENDIX AE

PROOF OF THEOREM 28

Fix P_S, P_U, $x(u,s)$, and $\epsilon > 0$ such that, $Q_Z = Q_0$.

Codebook Generation: Let $C_1^{(n)} \triangleq \big\{U^n(k,m)\big\}_{(k,m)\in\mathcal{K}\times\mathcal{M}}$ be a random codebook consisting of independent random sequences, each generated according to $P_U^{\otimes n}$. We denote a realization of $C_1^{(n)}$ by $\mathcal{C}_1^{(n)} \triangleq \big\{u^n(k,m)\big\}_{(k,m)\in\mathcal{K}\times\mathcal{M}}$. The indices (k,m) can be viewed as a one layer binning.

Let $C_2^{(n)} \triangleq \big\{S^n(j)\big\}_{j\in\mathcal{J}}$ be a random codebook consisting of independent random sequences, each generated according to $P_S^{\otimes n}$. We denote a realization of $C_2^{(n)}$ by $\mathcal{C}_2^{(n)} \triangleq \big\{s^n(j)\big\}_{j\in\mathcal{J}}$.

Also, $C_n = \big\{C_1^{(n)}, C_2^{(n)}\big\}$ denotes a random codebook and $\mathcal{C}_n = \big\{\mathcal{C}_1^{(n)}, \mathcal{C}_2^{(n)}\big\}$ denotes a fixed codebook. The set of all possible values of C_n is denoted by $\mathfrak{C}_n$. The codebook construction described above induces the PMF $\lambda \in \mathcal{P}(\mathfrak{C}_n)$ over the codebook ensemble. For each $\mathcal{C}_n \in \mathfrak{C}_n$

$$\lambda(\mathcal{C}_n) = \prod_{j\in\mathcal{J}} P_S^{\otimes n}\big(s^n(j)\big) \prod_{(k,m)\in\mathcal{K}\times\mathcal{M}} P_U^{\otimes n}\big(u^n(k,m)\big). \tag{AE.1}$$

Encoding: To send the message m according to the key k, the encoder computes $u^n(k,m)$ from the codebook and given the jammer's channel input it transmits codeword x^n, where $x_i = x(u_i(k,m), s_i)$. For a fixed codebook $\mathcal{C}_n$, the induced joint distribution is

$$P_{KMU^nJS^nZ^n}^{(\mathcal{C}_n)}(k,m,j,\tilde{s}^n,\tilde{u}^n,z^n) = 2^{-n(R_K+R+R_J)} \mathbb{1}_{\left\{\tilde{u}^n = u^n(k,m)\right\} \cap \left\{\tilde{s}^n = s^n(j)\right\}} W_{Z|US}^{\otimes n}(z^n|\tilde{u}^n,\tilde{s}^n). \tag{AE.2}$$

Considering the random codebook generation, we have

$$P(\mathcal{C}_n,k,m,\tilde{u}^n,j,\tilde{s}^n,z^n) = \lambda(\mathcal{C}_n) P^{(\mathcal{C}_n)}(k,m,\tilde{u}^n,j,\tilde{s}^n,z^n), \tag{AE.3}$$

where $\lambda \in \mathcal{P}(\mathfrak{C}_n)$ is defined in (AE.1). From (AE.2) we have

$$P_{Z^n|C_n}(z^n) = \frac{1}{2^{n(R_K+R+R_J)}} \sum_{(k,m,j)} W_{Z|US}^{\otimes n}\big(z^n|U^n(k,m), S^n(j)\big). \tag{AE.4}$$

Covert Analysis: Consider a scenario in which the jammer selects a codeword from its codebook (i.e. $\mathcal{C}_2^{(n)}$) uniformly at random, but the transmitter chooses the innocent sequence x_0^n as the channel input. For this scenario for a fixed codebook $\mathcal{C}_2^{(n)}$, the induced joint distribution is as follows

$$\Upsilon^{(\mathcal{C}_n)}_{JS^nZ^n}(j, s^n, z^n) = \frac{1}{2^{nR_J}} \mathbb{1}_{\{s^n = s^n(j)\}} W^{\otimes n}_{Z|XS}\big(z^n | x_0^n, s^n(j)\big). \tag{AE.5}$$

Therefore, the distribution induced at the output of the warden for a random codebook is

$$\Upsilon_{Z^n|C_n}(z^n) = \frac{1}{2^{nR_J}} \sum_{f=1}^{2^{nR_J}} W^{\otimes n}_{Z|XS}\big(z^n | x_0^n, S^n(j)\big). \tag{AE.6}$$

For this scenario if $R_J > \mathbb{I}_\Upsilon(S;Z)$ then according to soft covering lemma [83, Theorem 4] or [77, Corollary VII.4] and Pinsker's inequality, we have

$$\mathbb{E}_{C_n} \mathbb{V}\big(\Upsilon_{Z^n|C_n}, Q_0^{\otimes n}\big) \underset{n\to\infty}{\longrightarrow} 0, \tag{AE.7}$$

where

$$Q_0(\cdot) = \sum_{s\in\mathcal{S}} P_S(s) W_{Z|XS}\big(\,\cdot\,|x_0, s\big). \tag{AE.8}$$

Note that if $R_K < \mathbb{I}_\Upsilon(S;Z)$ according to Shannon's channel coding theorem, the warden might be able to decode J, which reduces the problem to the point to point channel for which the covert rate will be zero. We aim to show that the coding scheme described above guarantees

$$\mathbb{E}_{C_n} \mathbb{D}(P_{Z^n|C_n} || \Upsilon_{Z^n|C_n}) \underset{n\to\infty}{\longrightarrow} 0. \tag{AE.9}$$

To show that (AE.9) holds by using Lemma 1 and the triangle inequality we have

$$\mathbb{E}_{C_n} \mathbb{V}\big(P_{Z^n|C_n}, \Upsilon_{Z^n|C_n}\big) \leq \mathbb{E}_{C_n} \mathbb{V}\big(P_{Z^n|C_n}, Q_0^{\otimes n}\big) + \mathbb{E}_{C_n} \mathbb{V}\big(\Upsilon_{Z^n|C_n}, Q_0^{\otimes n}\big). \tag{AE.10}$$

According to the soft covering lemma [83, Theorem 4] or [77, Corollary VII.4], the second term on the RHS of (AE.10) vanishes when n grows if

$$R_J > \mathbb{I}_{\Upsilon}(S;Z). \tag{AE.11}$$

To bound the first term on the RHS of (AE.10) by using Lemma 1 we first show

$$\mathbb{E}_{C_n}\mathbb{D}\left(P_{Z^n|C_n}||Q_Z^{\otimes n}\right) \underset{n\to\infty}{\longrightarrow} 0, \tag{AE.12}$$

where

$$Q_Z(\cdot) = \sum_{s\in\mathcal{S}}\sum_{u\in\mathcal{U}}\sum_{x\in\mathcal{X}} P_S(s)P_U(u)\mathbb{1}_{\{x=x(u,s)\}}W_{Z|XS}\big(\cdot\,|x,s\big). \tag{AE.13}$$

Then we choose P_S, P_U, and $x(u,s)$ such that $Q_Z = Q_0$. We now have,

$$\begin{aligned}
&\mathbb{E}_{C_n}\big[\mathbb{D}\big(P_{Z^n|C_n}||Q_Z^{\otimes n}\big)\big] \\
&= \mathbb{E}_{C_n}\left[\sum_{z^n} P_{Z^n|C_n}(z^n)\log\frac{P_{Z^n|C_n}(z^n)}{Q_Z^{\otimes n}(z^n)}\right] \\
&= \mathbb{E}_{C_n}\Bigg[\sum_{z^n}\frac{1}{2^{n(R_K+R+R_J)}}\sum_{(k,m,j)} W_{Z|US}^{\otimes n}\big(z^n|U^n(k,m),S^n(j)\big) \\
&\quad \log\frac{\sum\limits_{(\tilde{k},\tilde{m},\tilde{j})} W_{Z|US}^{\otimes n}\big(z^n|U^n(\tilde{k},\tilde{m}),S^n(\tilde{j})\big)}{2^{n(R_K+R+R_J)}Q_Z^{\otimes n}(z^n)}\Bigg] \\
&\overset{(a)}{\leq} \frac{1}{2^{n(R_K+R+R_J)}}\sum_{(k,m,j)}\sum_{(u^n,s^n,z^n)} P_{USZ}^{\otimes n}\big(u^n(k,m),s^n(j),z^n\big) \\
&\quad \log\mathbb{E}_{\setminus(k,m,j)}\left[\frac{\sum\limits_{(\tilde{k},\tilde{m},\tilde{j})} W_{Z|US}^{\otimes n}\big(z^n|U^n(\tilde{k},\tilde{m}),S^n(\tilde{j})\big)}{2^{n(R_K+R+R_J)}Q_Z^{\otimes n}(z^n)}\right] \\
&= \frac{1}{2^{n(R_K+R+R_J)}}\sum_{(k,m,j)}\sum_{(u^n,s^n,z^n)} P_{USZ}^{\otimes n}\big(u^n(k,m),s^n(j),z^n\big) \\
&\quad \log\mathbb{E}_{\setminus(k,m,j)}\Bigg[\frac{W_{Z|US}^{\otimes n}\big(z^n|u^n(k,m),s^n(j)\big)}{2^{n(R_K+R+R_J)}Q_Z^{\otimes n}(z^n)} \\
&\qquad + \frac{\sum\limits_{(\tilde{k},\tilde{m})\neq(k,m)} W_{Z|US}^{\otimes n}\big(z^n|U^n(\tilde{k},\tilde{m}),s^n(j)\big)}{2^{n(R_K+R+R_J)}Q_Z^{\otimes n}(z^n)} + \frac{\sum\limits_{\tilde{j}\neq j} W_{Z|US}^{\otimes n}\big(z^n|u^n(k,m),S^n(\tilde{j})\big)}{2^{n(R_K+R+R_J)}Q_Z^{\otimes n}(z^n)}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{\sum_{(\tilde{k},\tilde{m})}\sum_{\tilde{j}\neq j} W_{Z|US}^{\otimes n}\big(z^n|U^n(\tilde{k},\tilde{m}),S^n(\tilde{j})\big)}{2^{n(R_K+R+R_J)}Q_Z^{\otimes n}(z^n)}\Bigg]\\
&\overset{(b)}{\leq}\frac{1}{2^{n(R_K+R+R_J)}}\sum_{(k,m,j)}\sum_{(u^n,s^n,z^n)}P_{USZ}^{\otimes n}\big(u^n(k,m),s^n(j),z^n\big)\log\Bigg[\frac{W_{Z|US}^{\otimes n}\big(z^n|u^n(k,m),s^n(j)\big)}{2^{n(R_K+R+R_J)}Q_Z^{\otimes n}\big(z^n\big)}\\
&\quad+\frac{\sum_{(\tilde{k},\tilde{m})\neq(k,m)}W_{Z|S}^{\otimes n}\big(z^n|s^n(j)\big)}{2^{n(R_K+R+R_J)}Q_Z^{\otimes n}(z^n)}+\frac{\sum_{\tilde{j}\neq j}W_{Z|U}^{\otimes n}\big(z^n|u^n(k,m)\big)}{2^{n(R_K+R+R_J)}Q_Z^{\otimes n}(z^n)}+1\Bigg]\\
&\leq\frac{1}{2^{n(R_K+R+R_J)}}\sum_{(k,m,j)}\sum_{(u^n,s^n,z^n)}P_{USZ}^{\otimes n}\big(u^n(k,m),s^n(j),z^n\big)\\
&\quad\times\log\Bigg[\frac{W_{Z|US}^{\otimes n}\big(z^n|u^n(k,m),s^n(j)\big)}{2^{n(R_K+R+R_J)}Q_Z^{\otimes n}\big(z^n\big)}+\frac{W_{Z|S}^{\otimes n}\big(z^n|s^n(j)\big)}{2^{nR_J}Q_Z^{\otimes n}(z^n)}+\frac{W_{Z|U}^{\otimes n}\big(z^n|u^n(k,m)\big)}{2^{n(R_K+R)}Q_Z^{\otimes n}(z^n)}+1\Bigg]\\
&\triangleq\Psi_1+\Psi_2
\end{aligned}
\tag{AE.14}
$$

where (a) follows from Jensen's inequality and (b) follows by taking expectation with respect to $\backslash(k,m)$ and by removing some terms from the denominator of the first term in the log function and adding one term to the nominator of the second term in the log function. We defined Ψ_1 and Ψ_2 as

$$
\begin{aligned}
\Psi_1&\leq\frac{1}{2^{n(R_K+R+R_J)}}\sum_{(k,m,j)}\sum_{(u^n,s^n,z^n)\in\mathcal{T}_\epsilon^{(n)}}P_{USZ}^{\otimes n}\big(u^n(k,m),s^n(j),z^n\big)\\
&\quad\times\log\Bigg[\frac{W_{Z|US}^{\otimes n}\big(z^n|u^n(k,m),s^n(j)\big)}{2^{n(R_K+R+R_J)}Q_Z^{\otimes n}(z^n)}+\frac{W_{Z|S}^{\otimes n}\big(z^n|s^n(j)\big)}{2^{nR_J}Q_Z^{\otimes n}(z^n)}+\frac{W_{Z|U}^{\otimes n}\big(z^n|u^n(k,m)\big)}{2^{n(R_K+R)}Q_Z^{\otimes n}(z^n)}+1\Bigg]\\
&\leq\log\left(\frac{2^{-n(1-\epsilon)\mathbb{H}(Z|U,S)}}{2^{n(R_K+R+R_J)}2^{-n(1+\epsilon)\mathbb{H}(Z)}}+\frac{2^{-n(1-\epsilon)\mathbb{H}(Z|S)}}{2^{nR_J}2^{-n(1+\epsilon)\mathbb{H}(Z)}}+\frac{2^{-n(1-\epsilon)\mathbb{H}(Z|U)}}{2^{n(R_K+R)}2^{-n(1+\epsilon)\mathbb{H}(Z)}}+1\right)
\end{aligned}
\tag{AE.15}
$$

and

$$
\begin{aligned}
\Psi_2&\leq\frac{1}{2^{n(R_K+R+R_J)}}\sum_{(k,m,j)}\sum_{(u^n,s^n,z^n)\notin\mathcal{T}_\epsilon^{(n)}}P_{USZ}^{\otimes n}\big(u^n(k,m),s^n(j),z^n\big)\\
&\quad\times\log\Bigg[\frac{W_{Z|US}^{\otimes n}\big(z^n|u^n(k,m),s^n(j)\big)}{2^{n(R_K+R+R_J)}Q_Z^{\otimes n}(z^n)}+\frac{W_{Z|S}^{\otimes n}\big(z^n|s^n(j)\big)}{2^{nR_J}Q_Z^{\otimes n}(z^n)}+\frac{W_{Z|U}^{\otimes n}\big(z^n|u^n(k,m)\big)}{2^{n(R_K+R)}Q_Z^{\otimes n}(z^n)}+1\Bigg]
\end{aligned}
$$

$$\leq 2|\mathcal{U}||\mathcal{S}||\mathcal{Z}|e^{-n\epsilon^2\mu_{U,S,Z}}n\log(\frac{2}{\mu_Z}+1). \tag{AE.16}$$

where

$$\mu_Z = \min_{\substack{z\in\mathcal{Z} \\ \text{s.t. } Q_z>0}} Q(z) \tag{AE.17}$$

$$\mu_{U,S,Z} = \min_{\substack{(u,s,z)\in(\mathcal{U},\mathcal{S},\mathcal{Z}) \\ \text{s.t. } Q(u,s,z)>0}} Q(u,s,z) \tag{AE.18}$$

When $n\to\infty$ then $\Psi_2\to 0$ and $\Psi_1\to 0$ when n grows if

$$R_J > \mathbb{I}_P(S;Z), \tag{AE.19a}$$

$$R_K + R > \mathbb{I}_P(U;Z), \tag{AE.19b}$$

$$R_K + R + R_J > \mathbb{I}_P(U,S;Z). \tag{AE.19c}$$

Decoding and Error Probability Analysis: By following the standard error analysis one can show that probability of error vanishes when n grows if,

$$R < \mathbb{I}_P(U;Y). \tag{AE.20}$$

The region in Theorem 28 is derived by combining (AE.11), (AE.19), and (AE.20).

APPENDIX AF

PROOF OF THEOREM 29

To prove the upper bound for the case that the jammer knows in which blocks the transmitter is communicating with the receiver and has an unlimited source of local randomness and transmits an i.i.d. sequence when communication is not happening. Consider any sequence of codes with length n such that $P_e^{(n)} \leq \epsilon_n$ and $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \delta$ where $\lim_{n\to\infty} \epsilon_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish.

Epsilon Rate Region: First we define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ which extends the region defined in (4.83) *as follows.*

$$\mathcal{A}_\epsilon = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists P_{UVSXYZ} \in \mathcal{D}_\epsilon : \\ R \leq \mathbb{I}(U;Y) + \epsilon \\ R + R_K \geq \mathbb{I}(V;Z) - 3\epsilon \\ R_J \geq \mathbb{I}(S;Z) - 2\epsilon \end{array} \right\}, \tag{AF.1a}$$

where

$$\mathcal{D}_\epsilon = \left\{ \begin{array}{l} P_{UVSXYZ} : \\ P_{UVSXYZ} = P_{SUV} \mathbb{1}_{\left\{X = X(U,S)\right\}} W_{YZ|XS} \\ \mathbb{D}\left(P_Z \| Q_0\right) \leq \epsilon \\ \max\{|\mathcal{U}|, |\mathcal{V}|\} \leq |\mathcal{X}| + 3 \end{array} \right\}, \tag{AF.1b}$$

Next, we prove that if a rate R is achievable, then $R \in \mathcal{A}_\epsilon$ for $\forall \epsilon > 0$. For any $\epsilon_n > 0$, we start by upper bounding nR using standard techniques.

$$\begin{aligned} nR &= \mathbb{H}(M) \\ &= \mathbb{H}(M|K) \end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\leq} \mathbb{I}(M; Y^n|K) + n\epsilon_n \\
&= \sum_{i=1}^{n} \mathbb{I}(M; Y_i|K, Y^{i-1}) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} \mathbb{I}(M, K, Y^{i-1}; Y_i) + n\epsilon_n \\
&\overset{(b)}{\leq} \sum_{i=1}^{n} \mathbb{I}(M, K, S^{i-1}; Y_i) + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} \mathbb{I}(U_i; Y_i) + n\epsilon_n \\
&= n \sum_{i=1}^{n} \frac{1}{n} \mathbb{I}(U_i; Y_i) + n\epsilon_n \\
&= n \sum_{i=1}^{n} \mathbb{P}(T = i) \mathbb{I}(U_i; Y_i|T = i) + n\epsilon_n \\
&= n\mathbb{I}(U_T; Y_T|T) + n\epsilon_n \\
&\leq n\mathbb{I}(U_T, T; Y_T) + n\epsilon_n \\
&\overset{(d)}{=} n\mathbb{I}(U; Y) + n\epsilon_n \\
&\overset{(e)}{\leq} n\mathbb{I}(U; Y) + n\epsilon,
\end{aligned}
\tag{AF.2}
$$

where

(a) *follows from Fano's inequality;*

(b) *follows because* $(M, K, Y^{i-1}) - (M, K, S^{i-1}) - Y_i$*, note that we also have* $V_i - (M, K, S^{i-1}) - Y_i$*, where* $V_i \triangleq (M, K, Z^{i-1})$*;*

(c) *follows by defining* $U_i \triangleq (M, K, S^{i-1})$*;*

(d) *follows by defining* $U = (U_T, T)$ *and* $Y = Y_T$*;*

(e) *follows by defining* $\epsilon \triangleq \max\{\epsilon_n, \nu \geq \frac{\delta}{n}\}$.

Next, we lower bound nR as follows,

$$\begin{aligned} n(R+R_K) &\geq \mathbb{H}(M,K) \\ &\geq \mathbb{I}(M,K;Z^n) \\ &= \sum_{i=1}^{n} \mathbb{I}(M,K;Z_i|Z^{i-1}) \\ &\overset{(a)}{\geq} \sum_{i=1}^{n} \mathbb{I}(M,K,Z^{i-1};Z_i) - \delta \\ &\overset{(b)}{=} \sum_{i=1}^{n} \mathbb{I}(V_i;Z_i) - \delta \\ &= n\sum_{i=1}^{n} \mathbb{P}(T=i)\mathbb{I}(V_i;Z_i|T=i) - \delta \\ &= n\mathbb{I}(V_T;Z_T|T) - \delta \\ &\overset{(c)}{\geq} n\mathbb{I}(V_T,T;Z_T) - 2\delta \\ &\overset{(d)}{=} n\mathbb{I}(V;Z) - 2\delta \end{aligned} \tag{AF.3}$$

where

(a) *follows from [80, Lemma 3];*

(b) *follows from the definition of* $V_i \triangleq (M,K,Z^{i-1})$*, that has been defined in the process of the derivation of* (AF.2)*;*

(c) *follows from [80, Lemma 3];*

(d) *follows by defining* $V=(V_T,T)$ *and* $Z=Z_T$*.*

For any $\nu>0$*, choosing* n *large enough ensures that*

$$\begin{aligned} R+R_K &\geq \mathbb{I}(V;Z) - 2\nu \\ &\geq \mathbb{I}(V;Z) - 2\epsilon, \end{aligned} \tag{AF.4}$$

where the last equality follows from the definition $\epsilon \triangleq \max\{\epsilon_n, \nu\}$. *Also, similar to* (AD.8) *one can show that,*

$$R_J \geq \mathbb{I}(S;Z) - 2\epsilon. \tag{AF.5}$$

To prove Covertness (i.e., $\mathbb{D}(P_Z||Q_0) \leq \epsilon$*), for* n *large enough*

$$\begin{aligned}\mathbb{D}(P_Z||Q_0) = \mathbb{D}(P_{Z_T}||Q_0) &= \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} P_{Z_i}\middle|\middle|Q_0\right) \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \leq \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \frac{\delta}{n} \leq \nu \leq \epsilon.\end{aligned} \tag{AF.6}$$

Combining (AF.2), (AF.4), (AF.5), *and* (AF.6) *shows that* $\forall \epsilon_n, \nu > 0$, $R \leq \max\{x : x \in \mathcal{A}_\epsilon\}$. *Therefore,*

$$R \leq \max\left\{x : x \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon\right\}. \tag{AF.7}$$

Continuity at Zero: The proof for continuity at zero of $\mathcal{A}_\epsilon$ is similar to that of Appendix Y and is omitted for the sake of brevity.

APPENDIX AG

PROOF OF THEOREM 30

Our achievability scheme is based on a block-Markov encoding scheme in which $B-1$ independent messages will be transmitted over B channel blocks, each of length r, therefore the overall codeword length is $n = rB$ symbols. The warden's observation Z^n will be described in terms of observations in each block $Z^n = (Z_1^r, \ldots, Z_B^r)$. The distribution induced on the warden's observation, by the block-Markov encoding, is $P_{Z^n} \triangleq P_{Z_1^r,\ldots,Z_B^r}$ and the target distribution on the warden's observation is $Q_0^{\otimes n} = \prod_{j=1}^{B} Q_0^{\otimes r}$. Therefore,

$$\begin{aligned}
\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) &= \mathbb{D}(P_{Z_1^r \ldots Z_B^r}||Q_0^{\otimes rB}) \\
&= \sum_{b=1}^{B} \mathbb{D}(P_{Z_b^r|Z_{b+1}^{B,r}}||Q_0^{\otimes r}|P_{Z_{b+1}^{B,r}}) \\
&= \sum_{b=1}^{B} [\mathbb{D}(P_{Z_b^r}||Q_0^{\otimes r}) + \mathbb{D}(P_{Z_b^r|Z_{b+1}^{B,r}}||P_{Z_b^r}|P_{Z_{b+1}^{B,r}})] \\
&= \sum_{b=1}^{B} [\mathbb{D}(P_{Z_b^r}||Q_0^{\otimes r}) + \mathbb{I}(Z_b^r; Z_{b+1}^{B,r})], \qquad \text{(AG.1)}
\end{aligned}$$

where $Z_{b+1}^{B,r} = \{Z_{b+1}^r, \ldots Z_B^r\}$. Consequently, $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \underset{n\to\infty}{\longrightarrow} 0$, is equivalent to $\forall b \in [\![1, B]\!]$;

$$\mathbb{D}(P_{Z_b^r}||Q_0^{\otimes r}) \underset{r\to\infty}{\longrightarrow} 0 \qquad \text{(AG.2a)}$$

$$\mathbb{I}(Z_b^r; Z_{b+1}^{B,r}) \underset{r\to\infty}{\longrightarrow} 0. \qquad \text{(AG.2b)}$$

This suggests that our code design should induce $Q_0^{\otimes r}$ on the warden's observation in each block, while the dependencies across blocks created by block-Markov coding should be eliminated. The random code scheme generation is described as follows:

Fix P_U, $P_{X|U}$, $P_{S_1|U}$, P_{S_2} and $\epsilon > 0$ such that, $Q_Z = Q_0$.

Random Codebook Generation for Communication Mode: For each block $b \in [\![1, B]\!]$:

- Let $C_0^{(r)} \triangleq \big\{U^r(m_0^{(b)})\big\}_{m_0^{(b)} \in \mathcal{M}_0}$, where $\mathcal{M}_0 = [\![1, 2^{rR}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_U^{\otimes r}$. We denote a realization of $C_0^{(r)}$ by $\mathcal{C}_0^{(r)} \triangleq \big\{u^r(m_0^{(b)})\big\}_{m_0^{(b)} \in \mathcal{M}_0}$.

- For every $m_0^{(b)}$, let $C_1^{(r)} \triangleq \big\{X^r(m_0^{(b)}, m^{(b)})\big\}_{m^{(b)} \in \mathcal{M}}$, where $\mathcal{M} = [\![1, 2^{rR}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_{X|U}^{\otimes r}\big(\,\cdot\,|u_i(m_0^{(b)})\big)$. We denote a realization of $C_1^{(r)}$ by $\mathcal{C}_1^{(r)} \triangleq \big\{x^r(m_0^{(b)}, m^{(b)})\big\}_{(m^{(b)}) \in \mathcal{M}}$.

- For every $m_0^{(b)}$, let $C_2^{(r)} \triangleq \big\{S_1^r(m_0^{(b)}, k^{(b)})\big\}_{k^{(b)} \in \mathcal{K}}$, where $\mathcal{K} = [\![1, 2^{rR_K}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_{S_1|U}^{\otimes r}\big(\,\cdot\,|u_i(m_0^{(b)})\big)$. We denote a realization of $C_2^{(r)}$ by $\mathcal{C}_2^{(r)} \triangleq \big\{s_1^r(m_0^{(b)}, k^{(b)})\big\}_{k^{(b)} \in \mathcal{K}}$.

$C_r = \big\{C_0^{(r)}, C_1^{(r)}, C_2^{(r)}\big\}$ denotes a random codebook, $\mathcal{C}_r = \big\{\mathcal{C}_0^{(r)}, \mathcal{C}_1^{(r)}, \mathcal{C}_2^{(r)}\big\}$ denotes a fixed codebook, and the set of all possible values of C_r is denoted by $\mathfrak{C}_r$.

Random Codebook Generation for No-Communication Mode:

- Let $C_3^{(n)} \triangleq \big\{S_2^n(k)\big\}_{k \in \mathcal{K}}$ be a random codebook consisting of independent random sequences, each generated according to $P_{S_2}^{\otimes n}$. We denote a realization of $C_3^{(n)}$ by $\mathcal{C}_3^{(n)} \triangleq \big\{s_2^n(k)\big\}_{k \in \mathcal{K}}$.

The set of all possible values of $C_3^{(n)}$ is denoted by $\mathfrak{C}_3^{(n)}$; and the set of all C_r codebooks from all blocks and the codebook $C_3^{(n)}$ is denoted by C_n.

This codebook construction induces the PMFs $\lambda_1 \in \mathcal{P}(\mathfrak{C}_r)$ and $\lambda_2 \in \mathcal{P}(\mathfrak{C}_3^{(n)})$ over the ensemble of codebooks. For every $\mathcal{C}_r \in \mathfrak{C}_r$ and $\mathcal{C}_3^{(n)} \in \mathfrak{C}_3^{(n)}$

$$\lambda_1(\mathcal{C}_r) = \prod_{m_0^{(b)} \in \mathcal{M}_0} P_U^{\otimes r}\big(u^r(m_0^{(b)})\big) \prod_{(m_0'^{(b)}, m^{(b)}) \in \mathcal{M}_0 \times \mathcal{M}} P_{X|U}^{\otimes r}\big(x^r(m_0'^{(b)}, m_1^{(b)})|u^r(m_0'^{(b)})\big) \times \prod_{(m_0''^{(b)}, k'^{(b)}) \in \mathcal{M}_0 \times \mathcal{K}} P_{S_1|U}^{\otimes r}\big(s_1^r(m_0''^{(b)}, k'^{(b)})|u^r(m_0''^{(b)})\big), \tag{AG.3}$$

$$\lambda_2(\mathcal{C}_3^{(n)}) = \prod_{k^{(b)} \in \mathcal{K}} P_{S_2}^{\otimes n}\big(s_2^n(k^{(b)})\big). \tag{AG.4}$$

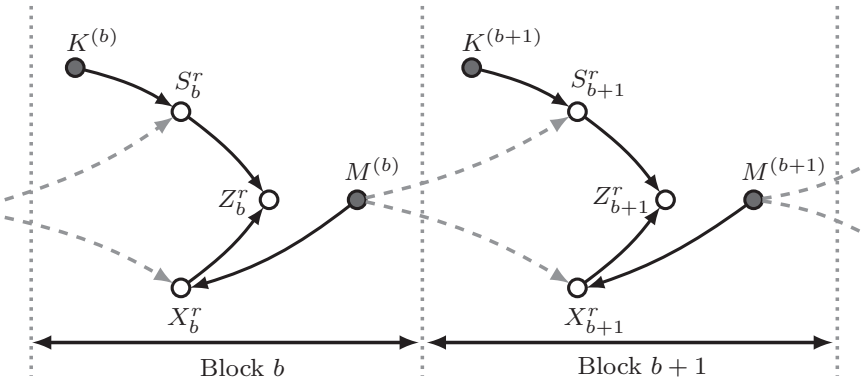

Figure AG.1. Functional dependence graph for the block-Markov encoding scheme

Encoding for Communication Mode: We assume that the transmitter, the jammer, and the receiver have access to common randomness $M_0^{(1)}$ and $M^{(B)}$. As a result of cribbing, the jammer has access to x_{b-1}^r at the end of the block $b-1$, therefore it finds an index $\hat{m}^{(b-1)}$ such that

$$\left(u^r(\hat{m}_0^{(b-1)}), x_{b-1}^r(\hat{m}_0^{(b-1)}, \hat{m}^{(b-1)}), x_{b-1}^r\right) \in \mathcal{T}_\epsilon^{(r)}(P_{U,X}), \tag{AG.5}$$

where $\hat{m}^{(b-1)}$ is an estimate of $m^{(b-1)}$. After decoding $\hat{m}^{(b-1)}$, the jammer sets $\hat{m}_0^{(b)} = \hat{m}^{(b-1)}$ to be used in the next block.

In the first block, to send the message $m^{(1)}$ by access to $m_0^{(1)}$, the transmitter computes $x_b^r(m_0^{(1)}, m^{(1)})$ and transmits it over the channel. Also, by access to $m_0^{(1)}$, the jammer computes $s_{1,b}^r(\hat{m}_0^{(1)}, k^{(1)})$ according to $k^{(1)}$ and transmits it over the channel.

In block $b \in [\![1, B]\!]$, to send the message $m^{(b)}$ by access to $m_0^{(b)} = m^{(b-1)}$, the transmitter computes $x_b^r(m_0^{(b)}, m^{(b)})$ and transmits it over the channel. Also, the jammer computes $s_{1,b}^r(\hat{m}_0^{(b)}, k^{(b)})$ and transmits it over the channel. Here, $k^{(b)}$ is the shared key between the jammer and the receiver and $\hat{m}_0^{(b)} = \hat{m}^{(b-1)}$ is the jammer's estimate of the message $m^{(b-1)}$ from the previous block.

The encoding procedure ensures that at the end of block b the transmitter and the jammer know $M_0^{(b)}$ with high probability, therefore the transmitter and the jammer can coordinate

their channel inputs. Also, the dependencies across the blocks is created through $M^{(b)}$ (see Fig. AG.1). The dependencies can be hidden from the warden by transmitting $M^{(b)}$ securely over a conceptual wiretap channel.

For a fixed codebook $\mathcal{C}_r$ and for each block b, let P denote the induced joint distribution by our code design when the transmitter uses $M_0^{(b)}$ and the jammer uses an estimate $\hat{M}_0^{(b)}$, which is derived from its estimate $\hat{M}^{(b-1)}$, at the end of the block $b-1$. Also, let $\bar{P}$ denote the induced joint distribution by our code design when both the transmitter and the jammer use $M_0^{(b)}$,

$$P^{(\mathcal{C}_r)}_{M_0^{(b)}\hat{M}_0^{(b)}M^{(b)}K^{(b)}U_b^rX_b^rS_{1,b}^rZ_b^r}(m_0^{(b)},\hat{m}_0^{(b)},m^{(b)},k^{(b)},\tilde{u}_b^r,\tilde{x}_b^r,\tilde{s}_{1,b}^r,z_b^r)=2^{-r(3R+R_K)}$$
$$\times\,\mathbb{1}_{\left\{\tilde{u}_b^r=u_b^r(m_0^{(b)})\right\}\cap\left\{\tilde{u}_b^r=u_b^r(\hat{m}_0^{(b)})\right\}\cap\left\{\tilde{x}_b^r=x_b^r(m_0^{(b)},m^{(b)})\right\}\cap\left\{\tilde{s}_{1,b}^r=s_{1,b}^r(\hat{m}_0^{(b)},k^{(b)})\right\}}W_{Z|UXS}^{\otimes r}(z_b^r|\tilde{u}_b^r,\tilde{x}_b^r,\tilde{s}_{1,b}^r). \tag{AG.6a}$$

$$\bar{P}^{(\mathcal{C}_r)}_{M_0^{(b)}M^{(b)}K^{(b)}U_b^rX_b^rS_{1,b}^rZ_b^r}(m_0^{(b)},\hat{m}_0^{(b)},m^{(b)},k^{(b)},\tilde{u}_b^r,\tilde{x}_b^r,\tilde{s}_{1,b}^r,z_b^r)=2^{-r(2R+R_K)}$$
$$\times\,\mathbb{1}_{\left\{\tilde{u}_b^r=u_b^r(m_0^{(b)})\right\}\cap\left\{\tilde{u}_b^r=u_b^r(\hat{m}_0^{(b)})\right\}\cap\left\{\tilde{x}_b^r=x_b^r(m_0^{(b)},m^{(b)})\right\}\cap\left\{\tilde{s}_{1,b}^r=s_{1,b}^r(\hat{m}_0^{(b)},k^{(b)})\right\}}W_{Z|UXS}^{\otimes r}(z_b^r|\tilde{u}_b^r,\tilde{x}_b^r,\tilde{s}_{1,b}^r). \tag{AG.6b}$$

Considering the random codebook generation, we have

$$P(\mathcal{C}_r,m_0,m,k,\tilde{s}_1^r,\tilde{u}^r,z_b^r)=\lambda_1(\mathcal{C}_r)P^{(\mathcal{C}_r)}(m_0,m,k,\tilde{s}_1^r,\tilde{u}^r,z_b^r), \tag{AG.7}$$

where $\lambda\in\mathcal{P}$ is defined in (AG.3).

Encoding for No-Communication Mode: When the transmitter is not communicating with the receiver, and therefore it transmits the innocent sequence x_0^n, the jammer computes $s_2^n(k)$ according to the key k and transmits it aver the channel. For this scenario for a fixed codebook $\mathcal{C}_2^{(n)}$, the induced joint distribution is as follows

$$\Upsilon^{(\mathcal{C}_n)}_{KS_2^nZ^n}(k,\tilde{s}_2^n,z^n)=\frac{1}{2^{nR_K}}\mathbb{1}_{\{\tilde{s}_2^n=s_2^n(k)\}}W_{Z|XS}^{\otimes n}\big(z^n|x_0^n,\tilde{s}_2^n\big). \tag{AG.8}$$

Therefore, the distribution induced at the output of the warden for a random codebook is

$$\Upsilon_{Z^n|C_n}(z^n) = \frac{1}{2^{nR_K}} \sum_{j=1}^{2^{nR_K}} W_{Z|XS}^{\otimes n}\big(z^n|x_0^n, S_2^n(k)\big). \tag{AG.9}$$

Decoding: The legitimate receiver waits until the transmission of the block B is complete and starts decoding each message in the sub-blocks going backwards $b \in [B, B-1, ..., 1]$. By access to the key and $m^{(B)}$ the decoder first finds $\hat{\hat{m}}_0^{(B)}$ such that

$$\Big(u^r(\hat{\hat{m}}_0^{(B)}), x^r(\hat{\hat{m}}_0^{(B)}, m^{(B)}), s_1^r(\hat{\hat{m}}_0^{(B)}, k^{(B)}), y_B^r\Big) \in \mathcal{T}_\epsilon^{(r)}(P_{UXS_1Y}). \tag{AG.10}$$

In block $b \in [\![1, B]\!]$, assuming that the decoder has successfully decoded $\hat{\hat{m}}_0^{(B)}$, $\hat{\hat{m}}_0^{(B-1)}$, ..., $\hat{\hat{m}}_0^{(b+1)}$ it sets $\hat{\hat{m}}_0^{(b+1)} = \hat{\hat{m}}^{(b)}$ and finds $\hat{\hat{m}}_0^{(b)}$ such that

$$\Big(u^r(\hat{\hat{m}}_0^{(b)}), x^r(\hat{\hat{m}}_0^{(b)}, \hat{\hat{m}}^{(b)}), s_1^r(\hat{\hat{m}}_0^{(B)}, k^{(b)}), y_b^r\Big) \in \mathcal{T}_\epsilon^{(r)}(P_{UXS_1Y}). \tag{AG.11}$$

Using the error analysis arguments in [95, Lemma 4], the probability of error in each block decreases exponentially with r and in turn vanishes across blocks if

$$R \leq \mathbb{H}_P(X|U), \tag{AG.12a}$$

$$R \leq \mathbb{I}_P(X, S_1; Y). \tag{AG.12b}$$

Covert Analysis: We aim to show that the coding scheme described above guarantees

$$\mathbb{E}_{C_n}\mathbb{D}\big(P_{Z^n|C_n}||\Upsilon_{Z^n|C_n}\big) \underset{n\to\infty}{\longrightarrow} 0. \tag{AG.13}$$

To show that (AG.13) holds by using Lemma 1 and the triangle inequality we have

$$\mathbb{E}_{C_n}\mathbb{V}\big(P_{Z^n|C_n}, \Upsilon_{Z^n|C_n}\big) \leq \mathbb{E}_{C_n}\mathbb{V}\big(P_{Z^n|C_n}, Q_0^{\otimes n}\big) + \mathbb{E}_{C_n}\mathbb{V}\big(\Upsilon_{Z^n|C_n}, Q_0^{\otimes n}\big), \tag{AG.14}$$

where

$$Q_0(\cdot) = \sum_{s_2 \in \mathcal{S}_2} P_{S_2}(s_2) W_{Z|X=x_0,S}\big(\,\cdot\,|x_0, s_2\big). \tag{AG.15}$$

According to soft covering lemma [83, Theorem 4] or [77, Corollary VII.4] the second term on the RHS of (AG.14) vanishes when n grows if

$$R_K > \mathbb{I}_\Upsilon(S_2; Z). \tag{AG.16}$$

Note that if $R_K < \mathbb{I}_\Upsilon(S_2; Z)$, according to Shannon's channel coding theorem, the warden might be able to decode J, which reduces the problem to the point to point channel for which the covert rate will be zero. To bound the first term on the RHS of (AG.14) by using Pinsker's inequality in Lemma 1 we first show $\mathbb{E}_{C_n}\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n}) \underset{n\to\infty}{\longrightarrow} 0$, where

$$Q_Z(\cdot) = \sum_{u\in\mathcal{U}}\sum_{x\in\mathcal{X}}\sum_{s_1\in\mathcal{S}_1} P_U(u)P_{X|U}(x|u)P_{S_1|U}(s_1|u)W_{Z|XS}(\cdot|x,s_1). \tag{AG.17}$$

Then we choose P_U, $P_{X|U}$, $P_{S_1|U}$, and P_{S_2} such that it satisfies $Q_Z = Q_0$. From the expansion in (AG.1), for every block $b \in [\![2, B]\!]$, by substituting Q_0 with Q_Z,

$$\begin{aligned}\mathbb{I}(Z_b^r; Z_{b+1}^{B,r}) &\le \mathbb{I}(Z_b^r; M^{(b)}, Z_{b+1}^{B,r}) \\ &\overset{(a)}{=} \mathbb{I}(Z_b^r; M^{(b)}),\end{aligned} \tag{AG.18}$$

where (a) holds because $Z_b^r - M^{(b)} - Z_{b+1}^{B,r}$ forms a Markov chain, as seen in the functional dependence graph depicted in Fig. AG.1. Also,

$$\begin{aligned}\mathbb{I}(Z_b^r; M^{(b)}) &= \mathbb{D}\Big(P^{(\mathcal{C}_n)}_{Z_b^r M^{(b)}}||P^{(\mathcal{C}_n)}_{Z_b^r}P^{(\mathcal{C}_n)}_{M^{(b)}}\Big) \\ &\overset{(b)}{\le} \mathbb{D}\Big(P^{(\mathcal{C}_n)}_{Z_b^r M^{(b)}}||Q_Z^{\otimes r}Q_{M^{(b)}}\Big),\end{aligned} \tag{AG.19}$$

where $Q_{M^{(b)}}$ is the uniform distribution on $[\![1, 2^{rR}]\!]$ and (b) follows from the positivity of relative entropy and

$$\mathbb{D}\big(P_{Z_b^r M^{(b)}}||P_{Z_b^r}P_{M^{(b)}}\big) = \mathbb{D}\big(P_{Z_b^r M^{(b)}}||Q_Z^{\otimes r}Q_{M^{(b)}}\big) - \mathbb{D}\big(P_{M^{(b)}}||Q_{M^{(b)}}\big) - \mathbb{D}\big(P_{Z_b^r}||Q_Z^{\otimes r}\big). \tag{AG.20}$$

Therefore by combining (AG.1), (AG.19), and (AG.20)

$$\mathbb{D}\big(P_{Z^n|C_n}||Q_Z^{\otimes n}\big) \leq 2\sum_{b=1}^{B}\mathbb{D}\big(P_{Z_b^r M^{(b)}|C_r}||Q_Z^{\otimes r}Q_{M^{(b)}}\big). \tag{AG.21}$$

We now proceed to bound the RHS of (AG.21). For this purpose, we first show that our coding scheme guarantees that

$$\mathbb{E}_{C_r}\Big[\mathbb{D}\Big(\bar{P}_{Z_b^r M^{(b)}|C_r}\Big||Q_Z^{\otimes r}Q_{M^{(b)}}\Big)\Big] \underset{r\to\infty}{\longrightarrow} 0, \tag{AG.22}$$

and then show that (AG.22) implies

$$\mathbb{E}_{C_r}\Big[\mathbb{D}\Big(P_{Z_b^r M^{(b)}|C_r}\Big||Q_Z^{\otimes r}Q_{M^{(b)}}\Big)\Big] \underset{r\to\infty}{\longrightarrow} 0. \tag{AG.23}$$

To bound (AG.22) from (AG.6b) we have,

$$\bar{P}_{Z_b^r M^{(b)}|C_r}(z_b^r, m^{(b)}) = \sum_{\left(m_0^{(b)},k^{(b)}\right)} 2^{-r(2R+R_K)} W_{Z|UXS}^{\otimes r}\Big(z_b^r|U_b^r\big(m_0^{(b)}\big), X_b^r\big(m_0^{(b)}, m^{(b)}\big), S_{1,b}^r\big(m_0^{(b)}, k^{(b)}\big)\Big). \tag{AG.24}$$

We now have,

$$\mathbb{E}_{C_r}\Big[\mathbb{D}\Big(\bar{P}_{Z_b^r M^{(b)}|C_r}\Big||Q_Z^{\otimes r}Q_{M^{(b)}}\Big)\Big] = \mathbb{E}_{C_r}\Big[\sum_{\left(z_b^r,m^{(b)}\right)} \bar{P}\big(z_b^r, m^{(b)}|C_r\big)\log\left(\frac{\bar{P}\big(z_b^r, m^{(b)}|C_r\big)}{Q_Z^{\otimes r}(z_b^r)Q_{M^{(b)}}(m^{(b)})}\right)\Big]$$

$$= \mathbb{E}_{C_r}\Bigg[\sum_{\left(z_b^r,m^{(b)}\right)}\sum_{(i,j)} 2^{-r(2R+R_K)} W_{Z|UXS}^{\otimes r}\Big(z_b^r|U_b^r(i), X_b^r\big(i, m^{(b)}\big), S_{1,b}^r\big(i,j\big)\Big)$$

$$\times\log\left(\frac{\sum\limits_{(\tilde{i},\tilde{j})} 2^{-r(2R+R_K)} W_{Z|UXS}^{\otimes r}\Big(z_b^r|U_b^r(\tilde{i}), X_b^r(\tilde{i}, m^{(b)}), S_{1,b}^r(\tilde{i},\tilde{j})\Big)}{2^{-rR}Q_Z^{\otimes r}(z_b^r)}\right)\Bigg]$$

$$\overset{(a)}{\leq} \sum_{\left(z_b^r,m^{(b)}\right)}\sum_{(i,j)}\frac{1}{2^{r(2R+R_K)}}\sum_{u_b^r(i)}\sum_{x_b^r(i,m^{(b)})}\sum_{s_{1,b}^r(i,j)} \bar{P}_{U^r X^r S_1^r Z_b^r}^{\otimes r}\big(u_b^r(i), x_b^r(i,m^{(b)}), s_{1,b}^r(i,j), z_b^r\big)$$

$$\times\log\mathbb{E}_{\setminus(i,j)}\left[\frac{\sum\limits_{(\tilde{i},\tilde{j})} W_{Z|UXS}^{\otimes r}\Big(z_b^r|U_b^r(\tilde{i}), X_b^r(\tilde{i}, m^{(b)}), S_{1,b}^r(\tilde{i},\tilde{j})\Big)}{2^{r(R+R_K)}Q_Z^{\otimes r}(z_b^r)}\right]$$

$$
\begin{aligned}
&\overset{(b)}{=} \sum_{\left(z_b^r, m^{(b)}\right)} \sum_{(i,j)} \frac{1}{2^{r(2R+R_K)}} \sum_{u_b^r(i)} \sum_{x_b^r(i,m^{(b)})} \sum_{s_{1,b}^r(i,j)} \bar{P}_{U^r X^r S_1^r Z_b^r}^{\otimes r}\left(u_b^r(i), x_b^r(i,m^{(b)}), s_{1,b}^r(i,j), z_b^r\right) \\
&\quad \times \log \Bigg(\frac{W_{Z|UXS}^{\otimes r}\Big(z_b^r | u_b^r(i), x_b^r(i,m^{(b)}), s_{1,b}^r(i,j)\Big)}{2^{r(R+R_K)} Q_Z^{\otimes r}(z_b^r)} \\
&\quad + \mathbb{E}_{\setminus j}\left[\frac{\sum_{\tilde{j}\neq j} W_{Z|UXS}^{\otimes r}\Big(z_b^r | u_b^r(i), x_b^r(i,m^{(b)}), S_{1,b}^r(i,\tilde{j})\Big)}{2^{r(R+R_K)} Q_Z^{\otimes r}(z_b^r)} \right] \\
&\quad + \mathbb{E}_{\setminus (i,j)}\left[\frac{\sum_{\tilde{i}\neq i} \sum_{\tilde{j}} W_{Z|UXS}^{\otimes r}\Big(z_b^r | U_b^r(\tilde{i}), X_b^r(\tilde{i}, m^{(b)}), S_{1,b}^r(\tilde{i},\tilde{j})\Big)}{2^{r(R+R_K)} Q_Z^{\otimes r}(z_b^r)} \right] \Bigg) \\
&= \sum_{\left(z_b^r, m^{(b)}\right)} \sum_{(i,j)} \frac{1}{2^{r(2R+R_K)}} \sum_{u_b^r(i)} \sum_{x_b^r(i,m^{(b)})} \sum_{s_{1,b}^r(i,j)} \bar{P}_{U^r X^r S_1^r Z_b^r}^{\otimes r}\left(u_b^r(i), x_b^r(i,m^{(b)}), s_{1,b}^r(i,j), z_b^r\right) \\
&\quad \times \log \left(\frac{W_{Z|UXS}^{\otimes r}\Big(z_b^r | u_b^r(i), x_b^r(i,m^{(b)}), s_{1,b}^r(i,j)\Big)}{2^{r(R+R_K)} Q_Z^{\otimes r}(z_b^r)} + \sum_{\tilde{j}\neq j} \frac{W_{Z|UX}^{\otimes r}\Big(z_b^r | u_b^r(i), x_b^r(i,m^{(b)})\Big)}{2^{r(R+R_K)} Q_Z^{\otimes r}(z_b^r)} + 1 \right) \\
&\leq \sum_{\left(z_b^r, m^{(b)}\right)} \sum_{(i,j)} \frac{1}{2^{r(2R+R_K)}} \sum_{u_b^r(i)} \sum_{x_b^r(i,m^{(b)})} \sum_{s_{1,b}^r(i,j)} \bar{P}_{U^r X^r S_1^r Z_b^r}^{\otimes r}\left(u_b^r(i), x_b^r(i,m^{(b)}), s_{1,b}^r(i,j), z_b^r\right) \\
&\quad \times \log \left(\frac{W_{Z|UXS}^{\otimes r}\Big(z_b^r | u_b^r(i), x_b^r(i,m^{(b)}), s_{1,b}^r(i,j)\Big)}{2^{r(R+R_K)} Q_Z^{\otimes r}(z_b^r)} + \frac{W_{Z|UX}^{\otimes r}\big(z_b^r | u_b^r(i), x_b^r(i,m^{(b)})\big)}{2^{rR} Q_Z^{\otimes r}(z_b^r)} + 1 \right) \\
&\triangleq \Psi_1 + \Psi_2, \qquad \text{(AG.25)}
\end{aligned}
$$

where (a) follows from Jensen's inequality, (b) is because $\mathbb{1}_{\{\cdot\}} \leq 1$, and the last term in the RHS of (b) is smaller than 1. We defined Ψ_1 and Ψ_2 as

$$
\begin{aligned}
\Psi_1 = &\sum_{(i,m^{(b)},j)} \frac{1}{2^{r(2R+R_K)}} \sum_{\left(u_b^r(i), x_b^r(i,m^{(b)}), s_{1,b}^r(i,j), z_b^r\right) \in \mathcal{T}_\epsilon^{(r)}} \bar{P}_{U^r X^r S_1^r Z_b^r}^{\otimes r}\left(u_b^r(i), x_b^r(i,m^{(b)}), s_{1,b}^r(i,j), z_b^r\right) \\
&\times \log \left(\frac{W_{Z|UXS}^{\otimes r}\Big(z_b^r | u_b^r(i), x_b^r(i,m^{(b)}), s_{1,b}^r(i,j)\Big)}{2^{r(R+R_K)} Q_Z^{\otimes r}(z_b^r)} + \frac{W_{Z|UX}^{\otimes r}\big(z_b^r | u_b^r(i), x_b^r(i,m^{(b)})\big)}{2^{rR} Q_Z^{\otimes r}(z_b^r)} + 1 \right) \\
\leq &\log \left(\frac{2^{-r(1-\epsilon)(\mathbb{H}(Z|X,S_1))}}{2^{r(R+R_K)} 2^{-r(1+\epsilon)\mathbb{H}(Z)}} + \frac{2^{-r(1-\epsilon)\mathbb{H}(Z|U,X)}}{2^{rR} 2^{-r(1+\epsilon)\mathbb{H}(Z)}} + 1 \right) \qquad \text{(AG.26)}
\end{aligned}
$$

$$\Psi_2 = \sum_{(i,m^{(b)},j)} \frac{1}{2^{r(2R+R_K)}} \sum_{\left(u_b^r(i),x_b^r(i,m^{(b)}),s_{1,b}^r(i,j),z_b^r\right)\notin\mathcal{T}_\epsilon^{(r)}} \bar{P}_{U^rX^rS_1^rZ_b^r}^{\otimes r}\left(u_b^r(i),x_b^r(i,m^{(b)}),s_{1,b}^r(i,j),z_b^r\right)$$
$$\times \log\left(\frac{W_{Z|UXS}^{\otimes r}\left(z_b^r|u_b^r(i),x_b^r(i,m^{(b)}),s_{1,b}^r(i,j)\right)}{2^{r(R+R_K)}Q_Z^{\otimes r}(z_b^r)} + \frac{W_{Z|UX}^{\otimes r}\left(z_b^r|u_b^r(i),x_b^r(i,m^{(b)})\right)}{2^{rR}Q_Z^{\otimes r}(z_b^r)} + 1\right)$$
$$\leq 2|U||S_1||X||Z|e^{-r\epsilon\mu_{U,S_1,X,Z}} r\log(\frac{4}{\mu_Z}+1), \tag{AG.27}$$

Ψ_2 goes to zero when $r\to\infty$ and Ψ_1 goes to zero when $r\to\infty$ if,

$$R + R_K > \mathbb{I}_P(X,S_1;Z), \tag{AG.28a}$$
$$R > \mathbb{I}_P(U,X;Z). \tag{AG.28b}$$

Finally, we show that (AG.22) implies (AG.23). Using the properties of total variation distance results to

$$\begin{aligned}&\mathbb{V}\left(P_{Z_b^rM^{(b)}},\bar{P}_{Z_b^rM^{(b)}}\right)\\ &\quad\leq \mathbb{V}\left(P_{Z_b^rM_0^{(b)}M^{(b)}\hat{M}_0^{(b)}K^{(b)}},\bar{P}_{Z_b^rM_0^{(b)}M^{(b)}M_0^{(b)}K^{(b)}}\right)\\ &\quad= 2\mathbb{P}\left(M_0^{(b)}\neq\hat{M}_0^{(b)}\right).\end{aligned} \tag{AG.29}$$

Since $P_{M^{(b)}} = \bar{P}_{M^{(b)}} = Q_{M^{(b)}}$,

$$\begin{aligned}&\mathbb{V}\left(P_{Z_b^rM^{(b)}},Q_Z^{\otimes r}Q_{M^{(b)}}\right)\\ &\quad\leq \mathbb{V}\left(P_{Z_b^rM^{(b)}},\bar{P}_{Z_b^rM^{(b)}}\right)+\mathbb{V}\left(\bar{P}_{Z_b^rM^{(b)}},Q_Z^{\otimes r}Q_{M^{(b)}}\right).\end{aligned} \tag{AG.30}$$

Since the probability of error vanishes as r grows and using (AG.29) the first term on the RHS of (AG.30) vanishes as r grows. Also, Pinsker's inequality ensures that the second term on the RHS of (AG.30) vanishes as r grows if we have (AG.28). Using (AG.30) together with Lemma ? ensures that (AG.23) holds. The region in Theorem 30 is derived by applying Fourier-Motzkin to (AG.12), (AG.16), and (AG.28).

APPENDIX AH

PROOF OF THEOREM 31

To prove the upper bound for the case that the transmitter's codeword is available strictly-causally at the jammer and the jammer has unlimited source of local randomness and transmits an i.i.d. sequence when communication is not happening, consider any sequence of codes with length n such that $P_e^{(n)} \leq \epsilon_n$ and $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \delta$ where $\lim_{n\to\infty} \epsilon_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish.

Epsilon Rate Region: We first define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ that expands the region defined in (4.95) *as follows.*

$$\mathcal{A}_\epsilon = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists P_{USXYZ} \in \mathcal{D}_\epsilon : \\ R \leq \min\{\mathbb{H}(X|U), \mathbb{I}(X,S;Y)\} + \epsilon \\ R_K \geq \mathbb{I}(X,S;Z) - \min\{\mathbb{H}(X|U), \mathbb{I}(X,S;Y)\} - 3\epsilon \end{array} \right\}, \tag{AH.1a}$$

where

$$\mathcal{D}_\epsilon = \left\{ \begin{array}{l} P_{USXYZ} : \\ P_{USXYZ} = P_U P_{X|U} P_{S|U} W_{YZ|XS} \\ \min\{\mathbb{H}(X|U), \mathbb{I}(X,S;Y)\} \geq \mathbb{I}(U,X;Z) - 3\epsilon \\ \mathbb{D}(P_Z||Q_0) \leq \epsilon \\ |\mathcal{U}| \leq \min\{|\mathcal{X}|\,|\mathcal{S}| + 1, |\mathcal{Y}| + 2\} \end{array} \right\}. \tag{AH.1b}$$

We next show that for any $\epsilon > 0$ and for any achievable rate pair (R, R_K) we have $(R, R_K) \in \mathcal{A}_\epsilon$. For any $\epsilon_n > 0$, using standard techniques, we start by upper bounding nR.

$$\begin{aligned} nR &= \mathbb{H}(M) \\ &= \mathbb{H}(M|K) \end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\leq} \mathbb{I}(M;Y^n|K) + n\epsilon_n \\
&\leq \mathbb{I}(M,K,X^n,S^n;Y^n) + n\epsilon_n \\
&\overset{(b)}{=} \mathbb{I}(X^n,S^n;Y^n) + n\epsilon_n \\
&= \sum_{i=1}^{n} \left[\mathbb{H}(Y_i|Y^{i-1}) - \mathbb{H}(Y_i|X^n,S^n,Y^{i-1})\right] + n\epsilon_n \\
&\leq \sum_{i=1}^{n} \left[\mathbb{H}(Y_i) - \mathbb{H}(Y_i|X^n,S^n,Y^{i-1})\right] + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} \mathbb{I}(X_i,S_i;Y_i) + n\epsilon_n \\
&= n\sum_{i=1}^{n} \frac{1}{n}\mathbb{I}(X_i,S_i;Y_i) + n\epsilon_n \\
&= n\sum_{i=1}^{n} \mathbb{P}(T=i)\mathbb{I}(X_T,S_T;Y_T|T=i) + n\epsilon_n \\
&= n\mathbb{I}(X_T,S_T;Y_T|T) + n\epsilon_n \\
&\leq n\mathbb{I}(X_T,S_T,T;Y_T) + n\epsilon_n \\
&\overset{(d)}{=} n\mathbb{I}(X,S;Y) + n\epsilon_n \\
&\overset{(e)}{=} n\mathbb{I}(X,S;Y) + n\epsilon,
\end{aligned}
$$

where

(a) *follows from Fano's inequality;*

(b) *holds because of the Markov chain* $(M,K) - (X^n,S^n) - Y^n$*;*

(c) *follows because the channel is memoryless;*

(d) *follows by defining* $X = (X_T,T)$*,* $S = S_T$*, and* $Y = Y_T$*;*

(e) *follows by defining* $\epsilon \triangleq \max\{\epsilon_n, \nu \geq \frac{\delta}{n}\}$.

We also have,

$$
\begin{aligned}
nR &= \mathbb{H}(M) \\
&\overset{(a)}{\leq} \mathbb{I}(M; Y^n) + n\epsilon_n \\
&\leq \mathbb{I}(M, X^n; Y^n) + n\epsilon_n \\
&\overset{(b)}{\leq} \mathbb{I}(X^n; Y^n) + n\epsilon_n \\
&\leq \mathbb{H}(X^n) + n\epsilon_n \\
&= \sum_{i=1}^{n} \mathbb{H}(X_i|X^{i-1}) + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} \mathbb{H}(X_i|U_i) + n\epsilon_n \\
&= n\sum_{i=1}^{n} \frac{1}{n}\mathbb{H}(X_i|U_i) + n\epsilon_n \\
&= n\sum_{i=1}^{n} \mathbb{P}(T=i)\mathbb{H}(X_T|U_T, T=i) + n\epsilon_n \\
&= n\mathbb{H}(X_T|U_T, T) + n\epsilon_n \\
&\overset{(d)}{\leq} n\mathbb{H}(X|U) + n\epsilon_n \\
&\overset{(e)}{\leq} n\mathbb{H}(X|U) + n\epsilon
\end{aligned}
\tag{AH.2}
$$

where

(a) *follows from Fano's inequality;*

(b) *holds because of the Markov chain* $M - X^n - Y^n$*;*

(c) *follows by defining* $U_i \triangleq X^{i-1}$*;*

(d) *follows by defining* $X = (X_T, T)$ *and* $U = (U_T, T)$*;*

(e) *follows by defining* $\epsilon \triangleq \max\{\epsilon_n, \nu \geq \frac{\delta}{n}\}$.

We now have,

$$
\begin{aligned}
nR &= \mathbb{H}(M) \\
&\geq \mathbb{I}(M; Z^n) \\
&\overset{(a)}{=} \mathbb{I}(M, X^n; Z^n) \\
&\geq \mathbb{I}(X^n; Z^n) \\
&= \sum_{i=1}^{n} \big[\mathbb{H}(Z_i|Z^{i-1}) - \mathbb{H}(Z_i|Z^{i-1}, X^n)\big] \\
&\overset{(b)}{\geq} \sum_{i=1}^{n} \big[\mathbb{H}(Z_i) - \mathbb{H}(Z_i|Z^{i-1}, X^n)\big] - \delta \\
&\geq \sum_{i=1}^{n} \big[\mathbb{H}(Z_i) - \mathbb{H}(Z_i|X_i, X^{i-1})\big] - \delta \\
&\overset{(c)}{=} \sum_{i=1}^{n} \mathbb{I}(U_i, X_i; Z_i) - \delta \\
&= n \sum_{i=1}^{n} \frac{1}{n} \big[\mathbb{I}(U_i, X_i; Z_i)\big] - \delta \\
&= n \sum_{i=1}^{n} \mathbb{P}(T = i) \mathbb{I}(U_T, X_T; Z_T|T = i) - \delta \\
&= n\mathbb{I}(U_T, X_T; Z_T|T) - \delta \\
&\overset{(d)}{\geq} n\mathbb{I}(U_T, X_T, T; Z_T) - 2\delta \\
&\overset{(e)}{=} n\mathbb{I}(U, X; Z) - 2\delta
\end{aligned}
$$

where

(a) *follows since* X^n *is a deterministic function of* M*;*

(b) *and* (d) *follows from [80, Lemma 3];*

(c) *follows by defining* $U_i \triangleq X^{i-1}$*;*

(e) *follows by defining* $X = (X_T, T)$*,* $U = (U_T, T)$*, and* $Z = Z_T$*.*

For any $\nu > 0$, choosing n large enough ensures that

$$\begin{aligned} R &\geq \mathbb{I}(U, X; Z) - 2\nu \\ &\geq \mathbb{I}(U, X; Z) - 2\epsilon, \end{aligned} \tag{AH.3}$$

where the last inequality follows from definition $\epsilon \triangleq \max\{\epsilon_n, \nu\}$. We now have,

$$\begin{aligned} n(R + R_K) &= \mathbb{H}(M, K) \\ &\geq \mathbb{I}(M, K; Z^n) \\ &\overset{(a)}{=} \mathbb{I}(M, K, X^n, S^n; Z^n) \\ &\geq \mathbb{I}(X^n, S^n; Z^n) \\ &= \sum_{i=1}^{n} \big[\mathbb{H}(Z_i|Z^{i-1}) - \mathbb{H}(Z_i|Z^{i-1}, X^n, S^n)\big] \\ &\overset{(b)}{\geq} \sum_{i=1}^{n} \big[\mathbb{H}(Z_i) - \mathbb{H}(Z_i|Z^{i-1}, X^n, S^n)\big] - \delta \\ &\geq \sum_{i=1}^{n} \big[\mathbb{H}(Z_i) - \mathbb{H}(Z_i|X_i, S_i)\big] - \delta \\ &= \sum_{i=1}^{n} \mathbb{I}(X_i, S_i; Z_i) - \delta \\ &= n \sum_{i=1}^{n} \frac{1}{n} \big[\mathbb{I}(X_i, S_i; Z_i)\big] - \delta \\ &= n \sum_{i=1}^{n} \mathbb{P}(T = i)\mathbb{I}(X_T, S_T; Z_T|T = i) - \delta \\ &= n\mathbb{I}(X_T, S_T; Z_T|T) - \delta \\ &\overset{(c)}{\geq} n\mathbb{I}(X_T, S_T, T; Z_T) - 2\delta \\ &\overset{(d)}{=} n\mathbb{I}(X, S; Z) - 2\delta \end{aligned}$$

where

(a) *follows since X^n is a deterministic function of M and S^n is a deterministic function of (K, X^{n-1});*

(*b*) *and* (*c*) *follow from [80, Lemma 3];*

(*d*) *follows by defining* $X = (X_T, T)$, $S = S_T$, *and* $Z = Z_T$.

For any $\nu > 0$, *choosing* n *large enough ensures that*

$$\begin{aligned} R + R_K &\geq \mathbb{I}(X, S; Z) - 2\nu \\ &\geq \mathbb{I}(X, S; Z) - 2\epsilon, \end{aligned} \tag{AH.4}$$

where the last inequality follows from definition $\epsilon \triangleq \max\{\epsilon_n, \nu\}$.

To show that $\mathbb{D}(P_Z||Q_0) \leq \epsilon$, *note that for* n *large enough*

$$\begin{aligned} \mathbb{D}(P_Z||Q_0) = \mathbb{D}(P_{Z_T}||Q_0) &= \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} P_{Z_i}\middle|\middle|Q_0\right) \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \leq \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \frac{\delta}{n} \leq \nu \leq \epsilon. \end{aligned} \tag{AH.5}$$

Continuity at zero: The proof for continuity at zero is similar to that of Appendix N.

APPENDIX AI

PROOF OF THEOREM 32

Fix P_X, $P_{S_1|X}$, P_{S_2} and $\epsilon > 0$ such that, $Q_Z = Q_0$.

Random Codebook Generation for Communication Mode:

- Let $C_1^{(n)} \triangleq \left\{X^n(m)\right\}_{m\in\mathcal{M}}$ be a random codebook consisting of independent random sequences, each generated according to $P_X^{\otimes n}$. We denote a realization of $C_1^{(n)}$ by $\mathcal{C}_1^{(n)} \triangleq \left\{x^n(m)\right\}_{m\in\mathcal{M}}$.

- For every $m \in \mathcal{M}$, let $C_2^{(n)} \triangleq \left\{S_1^n(X^n(m),k)\right\}_{(m,k)\in\mathcal{M}\times\mathcal{K}}$ be a random codebook consisting of independent random sequences, each generated according to $P_{S_1|X}^{\otimes n}$. We denote a realization of $C_2^{(n)}$ by $\mathcal{C}_2^{(n)} \triangleq \left\{s_1^n(x^n(m),k)\right\}_{(m,k)\in\mathcal{M}\times\mathcal{K}}$.

Random Codebook Generation for No-Communication Mode:

- Let $C_3^{(n)} \triangleq \left\{S_2^n(k)\right\}_{k\in\mathcal{K}}$ be a random codebook consisting of independent random sequences, each generated according to $P_{S_2}^{\otimes n}$. We denote a realization of $C_3^{(n)}$ by $\mathcal{C}_3^{(n)} \triangleq \left\{s_2^n(k)\right\}_{k\in\mathcal{K}}$.

Also, $C_n = \left\{C_1^{(n)}, C_2^{(n)}, C_3^{(n)}\right\}$ denotes a random codebook and $\mathcal{C}_n = \left\{\mathcal{C}_1^{(n)}, \mathcal{C}_2^{(n)}, \mathcal{C}_3^{(n)}\right\}$ denotes a fixed codebook. The set of all possible values of C_n is denoted by $\mathfrak{C}_n$.

The codebook construction described above induces the PMF $\lambda \in \mathcal{P}(\mathfrak{C}_n)$ over the codebook ensemble. For each $\mathcal{C}_n \in \mathfrak{C}_n$

$$\lambda(\mathcal{C}_n) = \prod_{m\in\mathcal{M}} P_X^{\otimes n}\big(x^n(m)\big) \prod_{(x^n(m),k)\in\mathcal{X}^n\times\mathcal{K}} P_{S_1|X}^{\otimes n}\big(s_1^n\big(x^n(m),k\big)|x^n(m)\big) \prod_{k'\in\mathcal{K}} P_{S_2}^{\otimes n}\big(s_2^n(k')\big). \tag{AI.1}$$

Encoding for Communication Mode: To send the message m, the transmitter chooses the codeword $x^n(m)$. Also, given the codeword $x^n(m)$ and the key k, the jammer computes $s_1^n(x^n(m),k)$ and transmits it over the channel. As a result of cribbing, the jammer has access

to x^n in advance, therefore before transmission, it finds an index m such that $\big(x^n(\hat{m}), x^n\big) \in \mathcal{T}_\epsilon^{(n)}(P_X)$ where $\hat{m}$ is an estimate of m. According to the law of large numbers and the packing lemma, as $n \to \infty$ the jammer with a vanishing probability of error can find m uniquely if

$$R \leq \mathbb{H}(X). \tag{AI.2}$$

For a fixed codebook $\mathcal{C}_n \in \mathfrak{C}_n$, the induced joint distribution is,

$$P^{(\mathcal{C}_n)}(m, k, \tilde{x}^n, \tilde{s}_1^n, z^n) \triangleq 2^{-n(R+R_K)} \mathbb{1}_{\left\{\tilde{x}^n = x^n(m)\right\} \bigcap \left\{\tilde{s}_1^n = s_1^n(\tilde{x}^n, k)\right\}} W_{Z|XS}^{\otimes n}(z^n|\tilde{x}^n, \tilde{s}_1^n). \tag{AI.3}$$

Therefore, the distribution induced on the warden's observation by our code design is

$$P_{Z^n}^{(\mathcal{C}_n)}(z^n) = \frac{1}{2^{n(R+R_K)}} \sum_{m=1}^{2^{nR}} \sum_{k=1}^{2^{nR_K}} W_{Z|XS}^{\otimes n}\big(z^n|x^n(m), s_1^n\big(x^n(m), k\big)\big). \tag{AI.4}$$

Encoding for No-Communication Mode: When the transmitter is not communicating with the receiver, and therefore it transmits x_0^n, the jammer computes a sequence $s_2^n(k)$ according to the key k, and transmits it over the channel. For a fixed codebook $\mathcal{C}_n$, the induced joint distribution for this case is,

$$\Upsilon_{KS_2^nZ^n}^{(\mathcal{C}_n)}(m, k, \tilde{x}^n, \tilde{s}^n, z^n) = \frac{1}{2^{n(R+R_K)}} \mathbb{1}_{\left\{\tilde{s}_2^n = s_2^n(k)\right\}} W_{Z|X=x_0,S}^{\otimes n}\big(z^n|x_0^n, \tilde{s}_2^n\big).$$

Therefore, the distribution induced on the warden's observation is

$$\Upsilon_{Z^n}^{(\mathcal{C}_n)}(z^n) = \frac{1}{2^{nR_K}} \sum_{k=1}^{2^{nR_K}} W_{Z|X=x_0,S}^{\otimes n}\big(z^n|x_0^n, s_2^n(k)\big). \tag{AI.5}$$

Decoding and Error Probability Analysis: Upon receiving y^n by access to k the decoder finds a unique m such that

$$\Big(x^n(m), s_1^n\big(x^n(m), k\big), y^n\Big) \in \mathcal{T}_\epsilon^{(n)}(P_{X,S_1,Y}). \tag{AI.6}$$

According to the law of large numbers and the packing lemma, the probability of error goes to zero as $n \to \infty$ if [85],

$$R < \mathbb{I}_P(X, S_1; Y). \tag{AI.7}$$

Covert Analysis: We aim to show that the coding scheme described above guarantees

$$\mathbb{E}_{C_n} \mathbb{D}\big(P_{Z^n|C_n} || \Upsilon_{Z^n|C_n}\big) \underset{n\to\infty}{\longrightarrow} 0. \tag{AI.8}$$

To prove (AI.8) by using Lemma 1 and the triangle inequality we have

$$\mathbb{E}_{C_n} \mathbb{V}\big(P_{Z^n|C_n}, \Upsilon_{Z^n|C_n}\big) \leq \mathbb{E}_{C_n} \mathbb{V}\big(P_{Z^n|C_n}, Q_0^{\otimes n}\big) + \mathbb{E}_{C_n} \mathbb{V}\big(\Upsilon_{Z^n|C_n}, Q_0^{\otimes n}\big), \tag{AI.9}$$

where

$$Q_0(\cdot) = \sum_{s_2 \in \mathcal{S}_2} P_{S_2}(s_2) W_{Z|X=x_0,S}(\cdot|x_0, s_2). \tag{AI.10}$$

To bound the first term on the RHS of (AI.9) by using Pinsker's inequality we first show

$$\mathbb{E}_{C_n} \mathbb{D}\big(P_{Z^n|C_n} || Q_Z^{\otimes n}\big) \underset{n\to\infty}{\longrightarrow} 0, \tag{AI.11}$$

where

$$Q_Z(\cdot) = \sum_{x \in \mathcal{X}} \sum_{s_1 \in \mathcal{S}_1} P_X(x) P_{S_1|X}(s_1|x) W_{Z|XS}\big(\,\cdot\,|x, s_1\big). \tag{AI.12}$$

Then we choose P_X, $P_{S_1|X}$, and P_{S_2} such that $Q_Z = Q_0$. From [65, Theorem 3], (AI.11) is satisfied if

$$R > \mathbb{I}_P(X; Z), \tag{AI.13a}$$

$$R_K > \mathbb{I}_P(X, S_1; Z) - \mathbb{H}(X), \tag{AI.13b}$$

$$R + R_K > \mathbb{I}_P(X, S_1; Z). \tag{AI.13c}$$

Also, according to the soft covering lemma [83, Theorem 4] or [77, Corollary VII.4], the second term on the RHS of (AI.9) vanishes when n grows if

$$R_K > \mathbb{I}(S_2; Z). \tag{AI.14}$$

Combining (AI.2), (AI.7), (AI.13), and (AI.14) completes the achievability proof of Theorem 32.

APPENDIX AJ

PROOF OF THEOREM 33

To prove the upper bound for the case that the transmitter's codeword is available non-causally at the jammer and the jammer has unlimited source of local randomness and transmits an i.i.d. sequence when communication is not happening, consider any sequence of codes with length n such that $P_e^{(n)} \leq \epsilon_n$ and $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \delta$ where $\lim_{n\to\infty} \epsilon_n = 0$. Note this assumption is consistent with the problem setup and does *not* require δ to vanish.

Epsilon Rate Region: First we define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ which extends the region defined in (4.97) *as follows.*

$$\mathcal{A}_\epsilon = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists (P_{SXYZ}, \Upsilon_{SZ}) \in \mathcal{D}_\epsilon : \\ R \leq \min\{\mathbb{I}_P(X, S; Y), \mathbb{H}_P(X)\} + \epsilon \\ R_K \geq \mathbb{I}_P(X, S; Y) - \mathbb{H}_P(X) - 3\epsilon \\ R + R_K \geq \mathbb{I}_\Upsilon(S; Z) - 2\epsilon \end{array} \right\}, \tag{AJ.1a}$$

where

$$\mathcal{D}_\epsilon = \left\{ \begin{array}{l} (P_{SXYZ}, \Upsilon_{SZ}) : \\ P_{SXYZ} = P_X P_{S|X} \mathbb{1}_{\left\{X = X(U,S)\right\}} W_{YZ|XS} \\ \Upsilon_{SZ} = P_{S|X=x_0} W_{Z|X=x_0,S} \\ \min\{\mathbb{I}(X, S; Y), \mathbb{H}(X)\} \geq \mathbb{I}(X; Z) - 3\epsilon \\ \mathbb{D}\left(P_Z \| \Upsilon_Z\right) \leq \epsilon \end{array} \right\}. \tag{AJ.1b}$$

We next show that if a rate R is achievable, then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$, we start by upper bounding nR using standard techniques.

$$nR = \mathbb{H}(M)$$

$$
\begin{aligned}
&= \mathbb{H}(M|K) \\
&= \mathbb{I}(M; Y^n|K) + \mathbb{H}(M|Y^n, K) \\
&\overset{(a)}{\leq} \mathbb{I}(M; Y^n|K) + n\epsilon_n \\
&= \sum_{i=1}^{n} \mathbb{I}(M; Y_i|K, Y^{i-1}) + n\epsilon_n \\
&= \sum_{i=1}^{n} \left[\mathbb{H}(Y_i|K, Y^{i-1}) - \mathbb{H}(Y_i|K, M, Y^{i-1})\right] + n\epsilon_n \\
&\leq \sum_{i=1}^{n} \left[\mathbb{H}(Y_i) - \mathbb{H}(Y_i|K, M, Y^{i-1})\right] + n\epsilon_n \\
&\overset{(b)}{=} \sum_{i=1}^{n} \left[\mathbb{H}(Y_i) - \mathbb{H}(Y_i|K, M, Y^{i-1}, X^n, S^n)\right] + n\epsilon_n \\
&= \sum_{i=1}^{n} \left[\mathbb{H}(Y_i) - \mathbb{H}(Y_i|X_i, S_i)\right] + n\epsilon_n \\
&= \sum_{i=1}^{n} \mathbb{I}(X_i, S_i; Y_i) + n\epsilon_n \\
&\overset{(c)}{\leq} n\mathbb{I}(\tilde{X}, \tilde{S}; \tilde{Y}) + n\epsilon_n \\
&\overset{(d)}{\leq} n\mathbb{I}(\tilde{X}, \tilde{S}; \tilde{Y}) + n\epsilon
\end{aligned}
\tag{AJ.2}
$$

where

(a) *follows from Fano's inequality;*

(b) *follows because* X^n *is a function of* M *and* S^n *is a function of* (K, X^n)*;*

(c) *follows from the concavity of mutual information, with the resulting random variables* $\tilde{X}$*,* $\tilde{S}$*, and* $\tilde{Y}$ *having the following distributions*

$$P_{\tilde{X},\tilde{S},\tilde{Y}}(x) \triangleq \frac{1}{n}\sum_{i=1}^{n} P_{X_i,S_i,Y_i}(x) \tag{AJ.3a}$$

$$P_{\tilde{X},\tilde{S},\tilde{Y}}(x, s, y) \triangleq P_{\tilde{X},\tilde{S}}(x, s)W_{Y|X,S}(y|x, s); \tag{AJ.3b}$$

(c) *follows from the definition of* $\epsilon \triangleq \max\{\epsilon_n, \nu \geq \frac{\delta}{n}\}$.

We also have,

$$\begin{aligned} nR &= \mathbb{H}(M) \\ &\overset{(a)}{\leq} \mathbb{I}(M; Y^n) + n\epsilon_n \\ &\leq \mathbb{I}(M, X^n; Y^n) + n\epsilon_n \\ &\overset{(b)}{=} \mathbb{I}(X^n; Y^n) + n\epsilon_n \\ &\leq \mathbb{H}(X^n) + n\epsilon_n \\ &= \sum_{i=1}^{n} \mathbb{H}(X_i|X^{i-1}) + n\epsilon_n \\ &\leq \sum_{i=1}^{n} \mathbb{H}(X_i) + n\epsilon_n \\ &\overset{(c)}{\leq} n\mathbb{H}(\tilde{X}) + n\epsilon_n \\ &\overset{(d)}{\leq} n\mathbb{H}(\tilde{X}) + n\epsilon, \end{aligned} \tag{AJ.4}$$

where

(a) *follows from Fano's inequality;*

(b) *holds because of the Markov chain* $M - X^n - Y^n$*;*

(c) *follows from the concavity of the entropy function, with the resulting random variable* $\tilde{X}$ *having the distribution defined in* (AJ.3)*;*

(d) *follows by defining* $\epsilon \triangleq \max\{\epsilon_n, \nu \geq \frac{\delta}{n}\}$.

Also, by following the same steps as [65, Theorem 5] one can prove the following upper bounds,

$$R > \mathbb{I}(\tilde{X}; \tilde{Z}), \tag{AJ.5a}$$

$$R_K > \mathbb{I}(\tilde{X}, \tilde{S}; \tilde{Z}) - \mathbb{H}(\tilde{X}), \tag{AJ.5b}$$

$$R + R_K > \mathbb{I}(\tilde{X}, \tilde{S}; \tilde{Z}). \tag{AJ.5c}$$

Combining the conditions (AJ.2), (AJ.4), *and* (AJ.5) *results to the region in* (AJ.1).

Continuity at zero: The proof for continuity at zero is similar to that of Appendix N.

APPENDIX AK

PROOF OF THEOREM 34

The achievability of the discrete memoryless channel with a cooperative jammer described in Section 4.6 when the jammer has causal access to the transmitter's codeword is similar to that of with strictly causal case in Section 4.6.1; except that here instead of generating codewords for S_1 we use Shannon strategy [64]. We denote the set of all strategies (functions) by $\mathscr{T} \triangleq \mathcal{S}_1^{\mathcal{X}}$ which map $\mathcal{X}$ to $\mathcal{S}_1$; and for $t \in \mathcal{T}$ we denote the image of $x \in \mathcal{X}$ by $t(x) \in \mathcal{S}_1$. The channel induced by the Shannon strategy is denoted by $(\mathcal{X} \times \mathscr{T}, W^{+}_{YZ|XS}, \mathcal{Y} \times \mathcal{Z})$ where $W^{+}_{YZ|XS_1} \triangleq W_{YZ|X,S_1=T(X)}$. From Theorem 30, rate pairs (R, R_K) that satisfy the following conditions are achievable when the jammer has strictly causal access to the transmitter's signals,

$$R < \mathbb{H}_P(X|U), \tag{AK.1a}$$

$$R < \mathbb{I}_P(X, T; Y), \tag{AK.1b}$$

$$R + R_K > \mathbb{I}_P(X, T; Z), \tag{AK.1c}$$

$$R > \mathbb{I}_P(U, X; Z), \tag{AK.1d}$$

$$R + R_K > \mathbb{I}_{\Upsilon}(T; Z), \tag{AK.1e}$$

for any joint probability distributions $P_{UXSYZ} = P_U P_{X|U} P_{S_1|U} W^{+}_{YZ|XS_1}$ and $\Upsilon_{S_2YZ} = P_{S_2} W_{YZ|X=x_0,S_2}$ such that $P_Z = \Upsilon_Z$. Therefore, the rate pair (R, R_K) in (AK.1) are also achievable when the jammer has causal access to the transmitter's signal. Restricting the joint distributions to be $P_{UXS_1YZ} = P_U P_X P_{S_1} W^{+}_{YZ|XS}$ and $\Upsilon_{S_2YZ} = P_{S_2} W_{YZ|X=x_0,S_2}$ results to,

$$\mathbb{H}_P(X|U) = \mathbb{H}_P(X), \tag{AK.2a}$$

$$\mathbb{I}_P(U, X; Z) = \mathbb{I}_P(X; Z), \tag{AK.2b}$$

$$\mathbb{I}_P(X, T; Z) = \mathbb{I}_P(X, S_1, T; Z) = \mathbb{I}_P(X, S_1; Z), \tag{AK.2c}$$

where $P_{XS_1YZ}(x, s_1, y, z) = P_X(x) \sum_{t:t(x)=s_1} P_T(t) W_{YZ|XS_1}(y, z|x, s_1)$ and $\Upsilon_{S_2YZ}(s_2, y, z) = P_{S_2}(s_2) W_{YZ|X=x_0,S_2}(y, z|x_0, s_2)$. This is possible since for an arbitrary joint distribution $P^*_{XS_1}$, always there exist a product distribution $P_{X,T} = P_X P_T$ such that $P^*_{XS_1}(x, s_1) = P_X(x) \sum_{t:t(x)=s_1} P_T(t)$. This is done by choosing [64, Eq. (44)],

$$P_X(x) = \sum_{s_1} P^*_{XS_1}(x, s_1), \tag{AK.3a}$$

$$P_T(t) = \prod_x \frac{P^*_{XS_1}(x, s_1 = t(x))}{P_X(x)}. \tag{AK.3b}$$

Therefore, from the arguments above all the rate pairs (R, R_K) satisfying the following conditions are achievable when the jammer has causal access to the transmitter's signal,

$$R < \mathbb{H}_P(X), \tag{AK.4a}$$

$$R < \mathbb{I}_P(X, S_1; Y), \tag{AK.4b}$$

$$R + R_K > \mathbb{I}_P(X, S_1; Z), \tag{AK.4c}$$

$$R > \mathbb{I}_P(X; Z), \tag{AK.4d}$$

$$R_K > \mathbb{I}_\Upsilon(S_2; Z), \tag{AK.4e}$$

for any joint probability distributions $P_{XS_1YZ} = P_{XS_1} W^+_{YZ|XS}$ and $\Upsilon_{S_2YZ} = P_{S_2} W_{YZ|X=x_0,S_2}$ such that $P_Z = \Upsilon_Z$.

APPENDIX AL

PROOF OF THEOREM 35

First we prove the achievability of Theorem 35. Fix P_X, $P_{S_1|X}$, P_{S_2} and $\epsilon > 0$ such that, $Q_Z = Q_0$.

Random Codebook Generation for Communication Mode:

- Let $C_1^{(n)} \triangleq \big\{X^n(m)\big\}_{m\in\mathcal{M}}$ be a random codebook consisting of independent random sequences, each generated according to $P_X^{\otimes n}$. We denote a realization of $C_1^{(n)}$ by $\mathcal{C}_1^{(n)} \triangleq \big\{x^n(m)\big\}_{m\in\mathcal{M}}$.
- For every $m \in \mathcal{M}$, let $C_2^{(n)} \triangleq \big\{S_1^n(m,k)\big\}_{(m,k)\in\mathcal{M}\times\mathcal{K}}$ be a random codebook consisting of independent random sequences, each generated according to $P_{S_1|X}^{\otimes n}\big(s_1^n|x^n(m)\big)$. We denote a realization of $C_2^{(n)}$ by $\mathcal{C}_2^{(n)} \triangleq \big\{s_1^n(m,k)\big\}_{(m,k)\in\mathcal{M}\times\mathcal{K}}$.

Random Codebook Generation for No-Communication Mode:

- Let $C_3^{(n)} \triangleq \big\{S_2^n(k)\big\}_{k\in\mathcal{K}}$ be a random codebook consisting of independent random sequences, each generated according to $P_{S_2}^{\otimes n}$. We denote a realization of $C_3^{(n)}$ by $\mathcal{C}_3^{(n)} \triangleq \big\{s_2^n(k)\big\}_{k\in\mathcal{K}}$.

Also, $C_n = \big\{C_1^{(n)}, C_2^{(n)}, C_3^{(n)}\big\}$ denotes a random codebook and $\mathcal{C}_n = \big\{\mathcal{C}_1^{(n)}, \mathcal{C}_2^{(n)}, \mathcal{C}_3^{(n)}\big\}$ denotes a fixed codebook. The set of all possible values of C_n is denoted by $\mathfrak{C}_n$.

The codebook construction described above induces the PMF $\lambda \in \mathcal{P}(\mathfrak{C}_n)$ over the codebook ensemble. For each $\mathcal{C}_n \in \mathfrak{C}_n$

$$\lambda(\mathcal{C}_n) = \prod_{m'\in\mathcal{M}} P_X^{\otimes n}\big(x^n(m')\big) \prod_{(m'',k')\in\mathcal{M}\times\mathcal{K}} P_{S|X}^{\otimes n}\big(s_1^n|x^n(m'')\big) \prod_{k''\in\mathcal{K}} P_{S_2}^{\otimes n}\big(s_2^n(k'')\big). \tag{AL.1}$$

Encoding for Communication Mode: To send the message m, the transmitter computes the codeword $x^n(m)$ and transmits it over the channel. Also, given the message m and the key k, the jammer computes $s_1^n(m,k)$ and transmits it over the channel.

For a fixed codebook $\mathcal{C}_n \in \mathfrak{C}_n$, the induced joint distribution is,

$$P^{(\mathcal{C}_n)}(m,k,\tilde{x}^n,\tilde{s}_1^n,z^n) \triangleq 2^{-n(R+R_K)} \times \mathbb{1}_{\left\{\tilde{x}^n=x^n(m)\right\}\bigcap\left\{\tilde{s}_1^n=s_1^n(m,k)\right\}} \times W_{Z|XS}^{\otimes n}(z^n|\tilde{x}^n,\tilde{s}_1^n). \tag{AL.2}$$

Therefore, the distribution induced on the warden's observation by our code design is

$$P_{Z^n}^{(\mathcal{C}_n)}(z^n) = \frac{1}{2^{n(R+R_K)}} \sum_{m=1}^{2^{nR}} \sum_{k=1}^{2^{nR_K}} W_{Z|XS}^{\otimes n}\big(z^n|x^n(m),s^n\big(m,k\big)\big). \tag{AL.3}$$

Encoding for No-Communication Mode: When the transmitter is not communicating with the receiver, and therefore it transmits x_0^n the jammer computes $s_2^n(k)$, according to the key k, and transmits it over the channel. For a fixed codebook $\mathcal{C}_n$, the induced joint distribution for this case is,

$$\Upsilon_{KS_2^nZ^n}^{(\mathcal{C}_n)}(k,s_2^n,z^n) = \frac{1}{2^{nR_K}} \mathbb{1}_{\left\{\tilde{s}_2^n=s_2^n(k)\right\}} W_{Z|X=x_0,S}^{\otimes n}\big(z^n|x_0^n,s_2^n\big).$$

Therefore, the distribution induced on the warden's observation is

$$\Upsilon_{Z^n}^{(\mathcal{C}_n)}(z^n) = \frac{1}{2^{nR_K}} \sum_{k=1}^{2^{nR_K}} W_{Z|X=x_0,S}^{\otimes n}\big(z^n|x_0^n,s_2^n(k)\big). \tag{AL.4}$$

Decoding and Error Probability Analysis: Upon receiving y^n by access to k the decoder finds a unique m such that

$$\Big(x^n(m),s_1^n\big(m,k\big),y^n\Big) \in \mathcal{T}_\epsilon^{(n)}(P_{XS_1Y}). \tag{AL.5}$$

According to the law of large numbers and the packing lemma, the probability of error goes to zero as $n \to \infty$ if [85],

$$R < \mathbb{I}_P(X,S_1;Y). \tag{AL.6}$$

Covert Analysis: We aim to show that the coding scheme described above guarantees

$$\mathbb{E}_{C_n}\mathbb{D}\big(P_{Z^n|C_n}||\Upsilon_{Z^n|C_n}\big)\big] \underset{n\to\infty}{\longrightarrow} 0. \tag{AL.7}$$

To prove (AL.7) by using Lemma 1 and the triangle inequality we have

$$\mathbb{E}_{C_n}\mathbb{V}\big(P_{Z^n|C_n}, \Upsilon_{Z^n|C_n}\big) \leq \mathbb{E}_{C_n}\mathbb{V}\big(P_{Z^n|C_n}, Q_0^{\otimes n}\big) + \mathbb{E}_{C_n}\mathbb{V}\big(\Upsilon_{Z^n|C_n}, Q_0^{\otimes n}\big), \tag{AL.8}$$

where

$$Q_0(\cdot) = \sum_{s_2 \in \mathcal{S}_2} P_{S_2}(s_2) W_{Z|X=x_0,S}\big(\,\cdot\,|x_0, s_2\big). \tag{AL.9}$$

According to the soft covering lemma [83, Theorem 4] or [77, Corollary VII.4], the second term on the RHS of (AL.8) vanishes when n grows if

$$R_K > \mathbb{I}_\Upsilon(S_2; Z). \tag{AL.10}$$

To bound the first term on the RHS of (AL.8) we first show

$$\mathbb{E}_{C_n}\mathbb{V}\big(P_{Z^n|C_n}, Q_Z^{\otimes n}\big) \underset{n\to\infty}{\longrightarrow} 0, \tag{AL.11}$$

where

$$Q_Z(\cdot) = \sum_{x \in \mathcal{X}} \sum_{s_1 \in \mathcal{S}_1} P_X(x) P_{S_1|X}(s_1|x) W_{Z|XS}\big(\,\cdot\,|x, s_1\big). \tag{AL.12}$$

Then we choose P_X, $P_{S_1|X}$, and P_{S_2} such that $Q_Z = Q_0$. From [65, Theorem 6], $\mathbb{E}_{C_n}\big[\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n})\big] \underset{n\to\infty}{\longrightarrow} 0$ and therefore (AL.11) is satisfied if

$$R > \mathbb{I}_P(X; Z), \tag{AL.13a}$$

$$R + R_K > \mathbb{I}_P(X, S_1; Z). \tag{AL.13b}$$

Combining (AL.6), (AL.10), and (AL.13) completes the achievability proof of Theorem 35.

APPENDIX AM

PROOF OF THEOREM 36

To prove the upper bound for the case that the jammer knows the transmitter's message and the jammer has unlimited source of local randomness and transmits an i.i.d. sequence when communication is not happening, consider any sequence of codes with length n such that $P_e^{(n)} \leq \epsilon_n$ and $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leq \delta$ where $\lim_{n\to\infty} \epsilon_n = 0$. Note this assumption is consistent with the problem setup and does *not* require δ to vanish.

Epsilon Rate Region: First we define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ which extends the region defined in (4.103) *as follows.*

$$\mathcal{A}_\epsilon = \left\{ \begin{array}{l} (R, R_K) \geq 0 : \exists \big(P_{SXYZ}, \Upsilon_{SYZ}\big) \in \mathcal{D}_\epsilon : \\ R \leq \mathbb{I}_P(X, S; Y) + \epsilon \\ R_K > \max\big\{\mathbb{I}_P(X, S; Z) - \mathbb{I}_P(X, S; Y), \mathbb{I}_P(X, S; Z) - \mathbb{H}_P(X)\big\} - 3\epsilon \\ R + R_K \geq \mathbb{I}_\Upsilon(S; Z) - 2\epsilon \end{array} \right\}, \quad \text{(AM.1a)}$$

where

$$\mathcal{D}_\epsilon = \left\{ \begin{array}{l} \big(P_{SXYZ}, \Upsilon_{SYZ}\big) : \\ P_{SXYZ} = P_X P_{S|X} W_{Y,Z|X,S} \\ \Upsilon_{SYZ} = P_S W_{YZ|X=x_0,S} \\ \mathbb{I}_P(X, S; Y) \geq \mathbb{I}_P(X; Z) - 3\epsilon \\ \mathbb{D}\left(P_Z \| \Upsilon_Z\right) \leq \epsilon \end{array} \right\}. \quad \text{(AM.1b)}$$

Next, we prove that if a rate R is achievable, then $R \in \mathcal{A}_\epsilon$ for $\forall \epsilon > 0$. For any $\epsilon_n > 0$, we start by upper bounding nR using standard techniques. By using standard techniques,

$$nR = \mathbb{H}(M)$$

$$
\begin{aligned}
&= \mathbb{H}(M|K) \\
&= \mathbb{I}(M; Y^n|K) + \mathbb{H}(M|Y^n, K) \\
&\overset{(a)}{\leq} \mathbb{I}(M; Y^n|K) + n\epsilon \\
&= \sum_{i=1}^{n} \mathbb{I}(M; Y_i|K, Y^{i-1}) + n\epsilon \\
&= \sum_{i=1}^{n} \big[\mathbb{H}(Y_i|K, Y^{i-1}) - \mathbb{H}(Y_i|K, M, Y^{i-1})\big] + n\epsilon \\
&\leq \sum_{i=1}^{n} \big[\mathbb{H}(Y_i) - \mathbb{H}(Y_i|K, M, Y^{i-1})\big] + n\epsilon \\
&\overset{(b)}{=} \sum_{i=1}^{n} \big[\mathbb{H}(Y_i) - \mathbb{H}(Y_i|K, M, Y^{i-1}, X^n, S^n)\big] + n\epsilon \\
&= \sum_{i=1}^{n} \big[\mathbb{H}(Y_i) - \mathbb{H}(Y_i|X_i, S_i)\big] + n\epsilon \\
&= \sum_{i=1}^{n} \mathbb{I}(X_i, S_i; Y_i) + n\epsilon \qquad \text{(AM.2)} \\
&\overset{(c)}{\leq} n\mathbb{I}(\tilde{X}, \tilde{S}; \tilde{Y}) + n\epsilon \qquad \text{(AM.3)}
\end{aligned}
$$

where

(a) *follows from Fano's inequality;*

(b) *follows because* X^n *is a function of* M *and* S^n *is a function of* (K, X^n)*;*

(c) *follows from the concavity of mutual information, with the resulting random variables* $\tilde{X}$*,* $\tilde{S}$*, and* $\tilde{Y}$ *having the following distributions*

$$P_{\tilde{X},\tilde{S},\tilde{Y}}(x) \triangleq \frac{1}{n}\sum_{i=1}^{n} P_{X_i,S_i,Y_i}(x) \qquad \text{(AM.4a)}$$

$$P_{\tilde{X},\tilde{S},\tilde{Y}}(x, s, y) \triangleq P_{\tilde{X},\tilde{S}}(x, s) W_{Y|X,S}(y|x, s). \qquad \text{(AM.4b)}$$

Also, by following the same steps as [65, Theorem 6] one can prove the following upper bounds,

$$R > \mathbb{I}(\tilde{X}; \tilde{Z}), \tag{AM.5a}$$

$$R_K > \mathbb{I}(\tilde{X}, \tilde{S}; \tilde{Z}) - \mathbb{H}(\tilde{X}), \tag{AM.5b}$$

$$R + R_K > \mathbb{I}(\tilde{X}, \tilde{S}; \tilde{Z}). \tag{AM.5c}$$

Combining the conditions (AM.3) *and* (AM.5) *results to the region in* (AM.1).

Continuity at zero: The proof for continuity at zero is similar to that of Appendix N.

REFERENCES

[1] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 57, no. 8, pp. 1355–1367, Oct. 1975.

[2] X. Song, H. Li, M. Yuan, and Y. Huang, "Coverage performance analysis of wireless caching networks with non-orthogonal multiple access-based multicasting," *IEEE Access*, vol. 7, pp. 164 009–164 020, Nov. 2019.

[3] Z. Zhao, M. Xu, Y. Li, and M. Peng, "A non-orthogonal multiple access-based multicast scheme in wireless content caching networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2723–2735, Dec. 2017.

[4] O. Tervo, L.-M. Tran, H. Pennanen, S. Chatzinotas, B. Ottersten, and M. Juntti, "Energy-efficient multi-cell multigroup multicasting with joint beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4904–4919, Sep. 2018.

[5] A. Cohen, A. Cohen, M. Médard, and O. Gurewitz, "Secure multi-source multicast," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 708–723, Jan. 2019.

[6] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacities of certain channel classes under random coding," *Ann. Math. Statist.*, vol. 31, no. 3, pp. 558–567, Sep. 1960.

[7] Y. Liang, G. Kramer, H. V. Poor, and S. Shamai, "Compound wiretap channels," in *Proc. 45th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep. 2007, pp. 136–143.

[8] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Ann. Probab.*, vol. 2, no. 5, pp. 805–814, 1974.

[9] S. Verdú, "Fifty years of Shannon theory," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2057–2078, Oct. 1998.

[10] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.

[11] Y. K. Chia and A. El Gamal, "Three-receiver broadcast channel with common and confidential messages," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 2748–2765, May 2012.

[12] C. Nair and A. El Gamal, "The capacity region of a class of three-receiver broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4479–4493, Oct. 2009.

[13] M. H. Yassaee, M. R. Aref, and A. A. Gohari, "Achievability proof via output statistics of random binning," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6786–6760, Nov. 2014.

[14] J. M. Renes and R. Renner, "Noisy channel coding via privacy amplification and information reconciliation," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7377–7385, Nov. 2011.

[15] I. Csiszár, "Almost independence and secrecy capacity," *IEEE Trans. Inf. Theory*, vol. 32, no. 1, pp. 40–47, Jan. 1996.

[16] S. Watanabe and Y. Oohama, "The optimal use of rate-limited randomness in broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 983–995, Feb 2015.

[17] S. Salehkalaibar, M. Mirmohseni, and M. R. Aref, "One-receiver two-eavesdropper broadcast channel with degraded message sets," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 12, pp. 1961–1974, Dec. 2013.

[18] E. Ekrem and S. Ulukus, "Multi-receiver wiretap channel with public and confidential messages," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2165–2177, Apr. 2013.

[19] M. Benammar and P. Piantanida, "Secrecy capacity region of some classes of wiretap broadcast channels," *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5564–5582, Oct. 2015.

[20] L. Liang and H. V. Poor, "Multiple-access channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 3, no. 3, pp. 976–1002, Mar. 2008.

[21] E. Tekin and A. Yener, "The Gaussian multiple access wire-tap channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5747–5755, Dec. 2008.

[22] ——, "The general Gaussian multiple-access and two-way wiretap channels: Achievable rates and cooperative jamming," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2735–2751, Jun. 2008.

[23] M. H. Yassaee and M. R. Aref, "Multiple access wiretap channels with strong secrecy," in *Proc. IEEE Info. Theory Workshop (ITW)*, Dublin, Ireland, Sep. 2010, pp. 1–5.

[24] M. Wiese and H. Boche, "Strong secrecy for multiple access channels," in *Information Theory, Combinatorics, and Search Theory*, Springer, 2013, pp. 71–122.

[25] A. J. Pierrot and M. R. Bloch, "Strongly secure communications over the two-way wiretap channel," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 595–605, Sep. 2011.

[26] H. Zivari-Fard, B. Akhbari, M. Ahmadian-Attari, and M. R. Aref, "Imperfect and perfect secrecy in compound multiple access channel with confidential message," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 6, pp. 1239 –1251, Jun. 2016.

[27] A. A. Chou and A. Yener, "Polar coding for the multiple access wiretap channel via rate-splitting and cooperative jamming," *IEEE Trans. Inf. Theory*, vol. 64, no. 12, pp. 7903–7921, Dec. 2018.

[28] A. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 841–850, Nov. 1982.

[29] R. Liu, I. Marić, P. Spasojević, and R. D. Yates, "Discrete memoryless interference and broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2493–2507, Jun. 2008.

[30] A. B. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1921–1930, Sep. 2013.

[31] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 2945–2949.

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

[33] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2334–2354, May 2016.

[34] M. Tahmasbi and M. R. Bloch, "First- and second-order asymptotics in covert communication," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2190–2212, Apr. 2019.

[35] T. V. Sobers, A. B. Bash, S. Guha, D. Towsley, and D. Goeckel, "Covert communication in the presence of an uninformed jammer," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6193–6206, Sep. 2017.

[36] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, "Reliable deniable communication with channel uncertainty," in *Proc. IEEE Info. Theory Workshop (ITW)*, Hobart, TAS, Australia, Nov. 2014, pp. 30–34.

[37] S.-H. Lee, L. Wang, A. Khisti, and G. W. Wornell, "Covert communication with channel-state information at the transmitter," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 9, pp. 2310–2319, Sep. 2018.

[38] Y. Chen and A. J. Han Vinck, "Wiretap channel with side information," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 395–402, Jan. 2008.

[39] A. Khisti, S. N. Diggavi, and G. W. Wornell, "Secret key agreement using asymmetry in channel state knowledge," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, Seoul, South Korea, Jul. 2009, pp. 2286–2290.

[40] Y.-K. Chia and A. El Gamal, "Wiretap channel with causal state information," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2838–2849, May 2012.

[41] H. Fujita, "On the secrecy capacity of wiretap channels with side information at the transmitter," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 11, pp. 2441–2452, Nov. 2016.

[42] A. Sonee and G. A. Hodtani, "Wiretap channel with strictly causal side information at encoder," in *Proc. Iran Workshop on commun. and Info. Theory (IWCIT)*, Tehran, Iran, May 2014, pp. 1–6.

[43] T. S. Han and M. Sasaki, "Wiretap channels with causal state information: Strong secrecy," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6750–6765, Oct. 2019.

[44] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Wiretap channels with random states non-causally available at the encoder," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1497–1519, Mar. 2020.

[45] P.-H. Lin, C. R. Janda, and E. A. Jorswieck, "Stealthy secret key generation," in *Proc. IEEE Global Conf. on Signal and Info. Processing (GlobalSIP)*, Montreal, QC, Canada, Mar. 2017, pp. 492–496.

[46] P.-H. Lin, C. R. Janda, E. A. Jorswieck, and R. F. Schaefer, "Stealthy keyless secret key generation from degraded sources," in *51st Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Apr. 2018, pp. 14–18.

[47] M. Tahmasbi and M. R. Bloch, "Covert secret key generation," in *Proc. IEEE Conf. on Commun. and Network Security (CNS)*, Las Vegas, NV, USA, Dec. 2017, pp. 540–544.

[48] ——, "Framework for covert and secret key expansion over classical-quantum channels," *Phys. Rev. A*, vol. 99, p. 052329, May 2019.

[49] S. Salehkalaibar, M. H. Yassaee, V. Y. F. Tan, and M. Ahmadipour, "State masking over a two-state compound channel," *IEEE Trans. Inf. Theory*, vol. 67, no. 9, pp. 5651–5673, Sep. 2021.

[50] M. Cheraghchi, F. Didier, and A. Shokrollahi, "Invertible extractors and wiretap protocols," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 1254–1274, Feb. 2012.

[51] M. Bellare, S. Tessaro, and A. Vardy, "Semantic security for the wiretap channel," in *Advances in Cryptology & CRYPTO 2012*, ser. Lecture Notes in Computer Science, R. Safavi-Naini and R. Canetti, Eds., vol. 7417. Springer Berlin Heidelberg, 2012, pp. 294–311, hard-copy.

[52] M. R. Bloch, M. Hayashi, and A. Thangaraj, "Error-control coding for physical-layer secrecy," *Proceedings of IEEE*, vol. 103, no. 10, pp. 1725–1746, Oct. 2015.

[53] S. Lee, R. J. Baxley, M. A. Weitnauer, and B. Walkenhorst, "Achieving undetectable communication," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 7, pp. 1195–1205, Oct. 2015.

[54] B. He, S. Yan, X. Zhou, and V. K. N. Lau, "On covert communication with noise uncertainty," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 941–944, Apr. 2017.

[55] R. Soltani, D. Goeckel, D. Towsley, A. B. Bash, and S. Guha, "Covert wireless communication with artificial noise generation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7252–7267, Nov. 2018.

[56] K. Shahzad, X. Zhou, S. Yan, J. Hu, F. Shu, and J. Li, "Achieving covert wireless communications using a full-duplex receiver," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8517–8530, Dec. 2018.

[57] O. Shmuel, A. Cohen, O. Gurewitz, and A. Cohen, "Multi-antenna jamming in covert communication," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 987–991.

[58] W. Xiong, Y. Yao, X. Fu, and S. Li, "Covert communication with cognitive jammer," *IEEE Commun. Lett.*, vol. 9, no. 10, pp. 1753–1757, Oct. 2020.

[59] J. Nötzel, M. Wiese, and H. Boche, "The arbitrarily varying wiretap channel—randomness, stability, and super-activation," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3504–3531, Jun. 2016.

[60] M. Wiese, J. Nötzel, and H. Boche, "A channel under simultaneous jamming and eavesdropping attack—correlated random coding capacities under strong secrecy criteria," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3844–3862, Jul. 2016.

[61] Z. Goldfeld, G. Kramer, H. H. Permuter, and P. Cuff, "Arbitrarily varying wiretap channels with type constrained states," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7216–7244, Sep. 2016.

[62] Q. Zhang, M. Bakshi, and S. Jaggi, "Covert communication over adversarially jammed channels," in *Proc. IEEE Info. Theory Workshop (ITW)*, Guangzhou, China, Nov. 2018, pp. 1–5.

[63] E. C. van der Meulen, "A survey of multi-way channels in information theory: 1961-1976," *IEEE Trans. Inf. Theory*, vol. 23, no. 1, pp. 1–37, Jan. 1977.

[64] F. M. J. Willems and E. C. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 313–327, May 1985.

[65] N. Helal, M. R. Bloch, and A. Nosratinia, "Cooperative resolvability and secrecy in the cribbing multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 66, no. 9, pp. 5429–5447, Sep. 2020.

[66] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.

[67] I. Sason and S. Verdú, "f-divergence inequalities," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 5973–6006, Nov. 2016.

[68] M. R. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*, 1st ed. Cambridge, U.K: Cambridge University Press, 2011.

[69] Y. Liang, H. V. Poor, and S. Shamai, *Information Theoretic Security*, 1st ed. Hanover, MA, USA: Now Publishers Inc., 2009.

[70] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[71] I. B. Gattegno, Z. Goldfeld, and H. H. Permuter, "Fourier-Motzkin elimination software for information theoretic inequalities," in *http://www.ee.bgu.ac.il/~fmeit/*, 2016.

[72] F. S. Chaharsooghi, M. J. Emadi, M. Zamanighomi, and M. R. Aref, "A new method for variable elimination in systems of inequations," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, St. Petersburg, Russia, Jul. 2011, pp. 1215–1219.

[73] C. S. Song, P. Cuff, and H. V. Poor, "The likelihood encoder for lossy compression," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1836–1849, Apr. 2016.

[74] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. New York, NY, USA: Springer-Verlag, 2005.

[75] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problem Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, Jan. 1980.

[76] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[77] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.

[78] M. H. Yassaee, M. R. Aref, and A. A. Gohari, "A technique for deriving one-shot achievability results in network information theory," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 1287–1291.

[79] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, "Nonasymptotic and second-order achievability bounds for coding with side-information," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1574–1605, Apr. 2015.

[80] H. ZivariFard, M. R. Bloch, and A. Nosratinia, "Keyless covert communication via channel state information," *available at https://arxiv.org/abs/2003.03308*, Mar. 2020.

[81] H. G. Eggleston, *Convexity*, 6th ed. Cambridge, U.K: Cambridge University Press, 1958.

[82] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, Mar. 1975.

[83] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[84] Z. Goldfeld, G. Kramer, H. H. Permuter, and P. Cuff, "Strong secrecy for cooperative broadcast channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 19, pp. 469–495, Jan. 2017.

[85] A. El Gamal and Y.-H. Kim, *Network Information Theory*, 1st ed. Cambridge, U.K: Cambridge University Press, 2012.

[86] A. El Gamal and E. C. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 120–122, Jan. 1981.

[87] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 4–21, Jan. 2003.

[88] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, no. 4, pp. 289–293, Oct. 1958.

[89] T. M. Cover and E. G. .A, "Capacity theorems for the relay channels," *IEEE Trans. Inf. Theory*, vol. 25, no. 6, pp. 572–584, Sep. 1979.

[90] Y. Steinberg, "Resolvability theory for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 472–487, Mar. 1998.

[91] E. C. Song, P. Cuff, and H. V. Poor, "A rate-distortion based secrecy system with side information at the decoders," in *Proc. 52th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep. 2014, pp. 755–762.

[92] J. Hou and G. Kramer, "Informational divergence approximations to product distributions," in *Proc. 13th Can. Workshop Inf. Theory (CWIT)*, Toronto, ON, Canada, Jun. 2013, pp. 76–81.

[93] R. A. Chou, R. M. Bloch, and J. Kliewer, "Empirical and strong coordination via soft covering with polar codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5087–5100, Jul. 2018.

[94] M. Frey, I. Bjelakovic, and S. Stanczak, "The mac resolvability region, semantic security and its operational implications," *available at http://arxiv.org/abs/1710.02342*, Aug. 2016.

[95] H. Asnani and H. H. Permuter, "Multiple-access channel with partial and controlled cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2252–2266, Apr. 2013.

BIOGRAPHICAL SKETCH

Hassan ZivariFard received his BSc degree in Electrical Engineering from Amirkabir University of Technology - Tafresh Branch, Tehran, Iran in 2009 and his MSc degree in Electrical Engineering from K. N. Toosi University of Technology, Tehran, Iran in 2012. Since January 2017 he has been working as a research assistant in the Multimedia Communications Laboratory at The University of Texas at Dallas where he is pursuing his PhD under the supervision of Professor Aria Nosratinia. His research interests are communication and information theory.

CURRICULUM VITAE

Hassan ZivariFard

November 10, 2020

Contact Information:

Department of Electrical Engineering
The University of Texas at Dallas
800 W. Campbell Rd.
Richardson, TX 75080-3021, U.S.A.

Voice: (972) 883-6433
Fax: (972) 883-6433
Email: hassan@utdallas.edu

Educational History:

BSc Electrical Engineering, Amirkabir University of Technology - Tafresh Branch, 2009
MSc Electrical Engineering, K. N. Toosi University of Technology, 2012
PhD Electrical Engineering, The University of Texas at Dallas, In-Progress

Lower and upper bounds on the capacity region of the key agreement models
MSc Dissertation
Electrical Engineering Department, K. N. Toosi University of Technology
Advisors: Prof. Mohammad-Reza Aref and Prof. Mahmoud Ahmadian-Attari

Designing and Manufacturing of a Rangefinder by Image Processing
BSc Dissertation
Electrical Engineering Department, Amirkabir University of Technology - Tafresh Branch
Advisors: Prof. Reza Anvari

Publication:

Journal Papers:

Hassan ZivariFard, Matthieu Bloch, Aria Nosratinia, "*Covert Communication via Cooperative Jamming,*" To be submitted to IEEE Transactions on Information Theory.

Hassan ZivariFard, Matthieu Bloch, Aria Nosratinia, "*Keyless Covert Communication via Channel State Information,*" Submitted to IEEE Transactions on Information Theory.

Hassan ZivariFard, Matthieu Bloch, Aria Nosratinia, "*Two-Multicast Channel with Confidential Messages,*" IEEE Transactions on Information Forensics and Security, Vol. 16, pp. 2743 - 2758, Jan., 2021.

Conference Papers:

Hassan ZivariFard, Matthieu Bloch, Aria Nosratinia, "*Covert Communication via Non-Causal Cribbing from a Cooperative Jammer,*" in Proc. IEEE International Symposium on Information Theory (ISIT), Victoria, Australia, Jul. 2021, pp. 202 - 207.

Hassan ZivariFard, Matthieu Bloch, Aria Nosratinia, "*Keyless Covert Communication in the Presence of Channel State Information,*" in Proc. IEEE International Symposium on Information Theory (ISIT), LA, USA, Jun. 2020, pp. 834 - 839.

Hassan ZivariFard, Matthieu Bloch, Aria Nosratinia, "*Keyless Covert Communication in the Presence of Non-causal Channel State Information,*" in Proc. IEEE Information Theory Workshop (ITW), Visby, Sweden, Aug. 2019, pp. 1 - 5.

Hassan ZivariFard, Matthieu Bloch, Aria Nosratinia, "*Two-Transmitter Two-Receiver Channel with Confidential Messages,*" in Proc. 55th Annual Allerton Conference on Communication, Control, and Computing, Monticello,IL, Oct 2017, pp. 103–110.