Information Spectrum Approach to Strong Converse Theorems for Degraded Wiretap Channels

Vincent Y. F. Tan† and Matthieu R. Bloch‡

Abstract

We consider block codes for degraded wiretap channels in which the legitimate receiver decodes the message with an asymptotic error probability $\varepsilon$ but the leakage to the eavesdropper vanishes. For discrete memoryless and Gaussian wiretap channels, we show that the maximum rate of transmission does not depend on $\varepsilon \in [0, 1)$, i.e., such channels possess the partial strong converse property. Furthermore, we derive sufficient conditions for the partial strong converse property to hold for memoryless but non-stationary symmetric and degraded wiretap channels. Our proof techniques leverage the information spectrum method, which allows us to establish a necessary and sufficient condition for the partial strong converse to hold for general wiretap channels without any information stability assumptions.

Index Terms

Strong converse, Information spectrum method, Degraded wiretap channels, Information-theoretic security

I. INTRODUCTION

One of the most well-studied models in information-theoretic security [1], [2] is the wiretap channel [3]. This channel model was introduced by Wyner in 1975 [3] and consists of three parties—a sender $\mathcal{X}$, a legitimate receiver $Y$ and an eavesdropper $Z$. The task is to reliably communicate a uniformly distributed message $M \in \{1, \ldots, 2^nR\}$ from $\mathcal{X}$ to $Y$ while keeping $Z$ ignorant of $M$. The secrecy capacity for the wiretap channel $W : \mathcal{X} \rightarrow \mathcal{Y} \times \mathcal{Z}$ is the supremum of all rates $R$ for which there exists a code that is reliable, i.e., $Y$ can reconstruct $M$ with probability tending to one as the blocklength $n$ tends to infinity, and secure, i.e., that the normalized mutual information (leakage rate) of the message and the eavesdropper’s signal $I(M; Z^n)$ is arbitrarily small as $n$ grows. Wyner showed that the secrecy capacity of a degraded discrete memoryless wiretap channel is

$$\max_{P_X} I(X; Y) - I(X; Z) \quad \text{bits per channel use.} \quad (1)$$

This result was generalized by Csiszár and Körner [4] to non-degraded channels.

In this paper, we relax the reliability condition of the wiretap code. More precisely, we allow the wiretap code to be such that the legitimate receiver decodes the message $M$ with an asymptotic error probability $\varepsilon \in [0, 1)$. The wiretap code, however, must ensure that $M$ and $Z^n$ are asymptotically independent and just as in Bloch and Laneman [5], we consider six measures of asymptotic independence of varying strengths. We show that for many classes of degraded, memoryless wiretap channels, the $\varepsilon$-secrecy capacity (maximum code rate $R$ such that the error probability is asymptotically no larger than $\varepsilon$) does not depend on $\varepsilon$. In other words, the $\varepsilon$-secrecy capacity is not larger than the expression in (1), in which it is assumed that the error probability of decoding $M$ vanishes asymptotically. Because we still ask that the leakage rate vanishes with the blocklength, we say that a partial strong converse holds.

In the majority of the information-theoretic security literature [1], [2], only weak converse statements are established, typically using Fano’s inequality. However, some progress has been made in recent works to establish strong converses. For example, the authors of [6]–[8] proved strong converses for the multi-party secret key agreement problem and other related problems. In particular, the authors of [7], [8] proved that the secret key

† V. Y. F. Tan is with the Department of Electrical and Computer Engineering and the Department of Mathematics, National University of Singapore, Singapore 117583 (Email: vtan@nus.edu.sg).
‡ M. R. Bloch is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 and GT-CNRS UMI 2958, 2 rue Marconi, 57070 Metz, France (Email: matthieu.bloch@ece.gatech.edu).
In the present paper, prove a strong converse for ε ≤ δ for a concise comparison of the results. There is substantial motivation to prove strong converses because the channel method is also useful in establishing necessary and sufficient conditions for the strong converse to hold under the usual assumptions of memorylessness, stationarity, ergodicity and information stability. The information spectrum is a systematic and powerful method to characterize the fundamental limits of communication systems without the constraints, such as the wiretap channel.

Fig. 1. The strong converses for the degraded DM-WTC (Theorem 3) and G-WTC (Theorem 5) hold for (ε, δ) on the blue strip [0, 1) × {0} (i.e., partial strong converse). Here, ε denotes the error probability and δ the variation distance. Morgan and Winter’s [9, Thm. 14] pretty strong converse for the private capacity of degradable quantum channels holds for $\sqrt{\varepsilon} + 2\sqrt{\delta} < 1/2$, indicated by the region in red. Tyagi and Watanabe’s strong converse for the secret key capacity [8, Cor. 11] holds for $\varepsilon + \delta < 1$, indicated by the union of the cyan, red and blue regions. This result improves on Tyagi and Narayan’s strong converse for the same problem [6, Sec. VII] which holds for $(\varepsilon, \delta) \in [0, 1) \times \{0\}$. We caution that these information-theoretic security problems are different so the results are not directly comparable.

capacity does not depend on the error bound ε and the secrecy bound δ (measured according to the variation distance) as long as $\varepsilon + \delta < 1$. Another related work is the one by Morgan and Winter [9, Sec. VI] who used one-shot bounds in [10] to establish a so-called pretty strong converse for the private capacity of a degradable quantum channel [11]. Specifically, they prove that the private capacity does not depend on the error bound ε and the secrecy bound δ as long as $\sqrt{\varepsilon} + 2\sqrt{\delta} < 1/2$. In the present paper, prove a strong converse for $(\varepsilon, \delta) \in [0, 1) \times \{0\}$. We discuss connections of our approach to [6]–[9] and other related works in more detail in Section IV-B. Also see Fig. 1 for a concise comparison of the results. There is substantial motivation to prove strong converses because such statements indicate that there exists a sharp phase transition between rates that are achievable and those that are not. Codes with unachievable rates have error probabilities that tend to one (or a positive number strictly less than one for the pretty strong converse) as the blocklength grows. Unlike weak converses, the rates are not simply bounded away from zero. For point-to-point channel coding, Wolfowitz established the strong converse in the 1950s [13], but little attention has been paid to strong converses for information-theoretic problems with secrecy constraints, such as the wiretap channel.

In this work, we adopt the information spectrum method [14]–[16] to make strong converse statements for various classes of degraded wiretap channels. The information spectrum method, developed by Verdú and Han [14], [15], is a systematic and powerful method to characterize the fundamental limits of communication systems without the usual assumptions of memorylessness, stationarity, ergodicity and information stability. The information spectrum method is also useful in establishing necessary and sufficient conditions for the strong converse to hold [16, Sec. 3.5]. The latter is one of the reasons why we have adopted this approach. Other possible strong converse techniques and related works are discussed in Section IV. The use of the information spectrum approach for general wiretap channels was pioneered by Hayashi in [17] in which the use of channel resolvability [15] [16, Sec. 6.3] was shown to be a useful coding mechanism for secrecy.

A. Main Contributions

We summarize the three main contributions in this work.

1To be more precise, the authors in [9, Thm. 14] used results in [10] to prove that the private capacity does not depend on $\varepsilon'$ and $\delta'$, both measured in terms of the purified distance $d_{\text{pur}}(P, Q) := \left[1 - \left(\sum_x (P(x)Q(x))^{1/2}\right)^2\right]^{1/2}$, as long as $\varepsilon' + 2\delta' < 1/\sqrt{2}$. This can be translated to the true average error probability $\varepsilon$ and the variation distance $\delta$ using the bounds $\varepsilon' \leq \sqrt{2\varepsilon}$ and $\delta' \leq \sqrt{2\delta}$ (e.g. [12, Thm. 1]). Thus, one obtains the strong converse condition $\sqrt{\varepsilon} + 2\sqrt{\delta} < 1/2$ (albeit conservative) in terms of the error probability and variation distance.
First, we consider wiretap codes whose asymptotic error probabilities in decoding do not exceed $\varepsilon$. These wiretap codes satisfy the condition that the leakage, measured according to any one of the six secrecy metrics in Bloch and Laneman [5], tends to zero asymptotically. We derive general formulas for the $\varepsilon$-secrecy capacity (for $\varepsilon \in [0, 1]$) and the $\varepsilon$-optimistic secrecy capacity [18]–[20] (for $\varepsilon \in (0, 1]$) by extending the results in [5] and [17].

Second, and most importantly, we leverage on the general formulas for the secrecy capacity and its optimistic version to show that the degraded discrete memoryless (and stationary) wiretap channel (DM-WTC) admits a partial strong converse. This means that regardless of the permissible asymptotic error probability $\varepsilon \in [0, 1)$, if the leakage vanishes asymptotically, the maximum rate of transmission cannot exceed the secrecy capacity Wyner derived in (1). This contribution is a strengthening of Wyner’s seminal result [3].

Finally, we leverage on the proof technique for the partial strong converse of the DM-WTC to extend our results in two important directions. We consider discrete degraded memoryless but non-stationary wiretap channels whose main and eavesdropper’s channels are symmetric [1, Sec. 3.4]. We show that if the Cesàro means of the capacities of the main and eavesdropper’s channels converge, the partial strong converse holds. Next, we show that the Gaussian wiretap channel (G-WTC), which is degraded, also satisfies the partial strong converse, strengthening the capacity result by Leung-Yan-Cheong and Hellman [21].

B. Paper Organization

The rest of this paper is organized as follows: In Section II, we state the notational conventions, describe the system model and formally define the partial strong converse property for the wiretap channel. In Section III, we state our main results. In particular, after recapitulating some information spectrum quantities in Section III-A, in Section III-B, we state general formulas for the $\varepsilon$-secrecy capacity and its optimistic version. These are done for arbitrary wiretap channels where the legitimate receiver is allowed to make an error with probability not exceeding $\varepsilon \in [0, 1)$ but the leakage is required to tend to zero. The bulk of the contributions are contained in Section III-C where we present strong converse results for specific channel models such as the DM-WTC and the G-WTC. In Section IV, we discuss other possible proof techniques, related works, and fertile avenues for further research. The proofs of the statements are contained in Section V.

II. SYSTEM MODEL AND DEFINITIONS

In this section, we state our notation and the definitions of the various problems we consider in this paper.

A. Basic Notations

Random variables (e.g., $X$) and their realizations (e.g., $x$) are denoted by upper case and lower case serif font, respectively. Sets are denoted in calligraphic font (e.g., the alphabet of $X$ is $\mathcal{X}$). In this paper, a discrete set is a set with finite cardinality. We use the notation $X^n$ to denote a vector of random variables $(X_1, \ldots, X_n)$. In addition, $X = \{X^n\}_{n \in \mathbb{N}}$ is a general source in the sense that each member of the sequence $X^n = (X_1^{(n)}, \ldots, X_n^{(n)})$ is a random vector. The consistency condition, i.e., $X_i^{(n)} = X_i^{(m)}$, need not hold. A general broadcast channel $\mathbf{W} = \{W^n : \mathcal{X}^n \rightarrow \mathcal{Y}^n \times \mathcal{Z}^n\}_{n \in \mathbb{N}}$ is a sequence of stochastic mappings from $\mathcal{X}^n$ to $\mathcal{Y}^n \times \mathcal{Z}^n$. The set of all probability distributions with support on an alphabet $\mathcal{X}$ is denoted as $\mathcal{P}(\mathcal{X})$. We use the notation $X \sim P_X$ to mean that the distribution of $X$ is $P_X$. The joint distribution formed by the product of the input distribution $P_X \in \mathcal{P}(\mathcal{X})$ and the channel $W : \mathcal{X} \rightarrow \mathcal{Y}$ is denoted by $P_X \times W$. Information-theoretic quantities are denoted using the notations in Han’s book [16], e.g., $H(X)$ for entropy, $I(X; Y)$ for mutual information and $D(P||Q)$ for relative entropy. All logarithms are to an arbitrary base. We also use the discrete interval notation $[i : j] := \{i, \ldots, j\}$. The variation distance between two measures $P$ and $Q$ on the same measurable space $(\Omega, \mathfrak{F})$ is defined as

$$V(P, Q) := \sup_{A \in \mathfrak{F}} |P(A) - Q(A)|. \quad (2)$$

For an arbitrary sample space $\Omega$, the definition of the variation distance in (2) is equivalent to

$$V(P, Q) = \frac{1}{2} \int_{\Omega} |f_P - f_Q| \, d\nu \quad (3)$$
where \( \nu \) an arbitrary positive measure such that both \( P \) and \( Q \) are absolutely continuous with respect to it and where \( f_P \) and \( f_Q \) are, respectively, the Radon-Nikodym derivatives of \( P \) and \( Q \) with respect to \( \nu \). The probability density function of the normal distribution \( \mathcal{N}(y; \mu, \sigma^2) \) is defined as

\[
\mathcal{N}(y; \mu, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right).
\]  

(4)

B. System Model

We consider a general wiretap channel, which is simply a synonym for a general broadcast channel \( W = \{W^n : \mathcal{X}^n \rightarrow \mathcal{Y}^n \times \mathcal{Z}^n\}_{n \in \mathbb{N}} \). Terminal \( \mathcal{X} \) denotes the sender, terminal \( \mathcal{Y} \) denotes the legitimate receiver, and terminal \( \mathcal{Z} \) denotes the eavesdropper. We would like to reliably transmit a message \( M \) from the terminal \( \mathcal{X} \) to terminal \( \mathcal{Y} \), and at the same time, design the code such that terminal \( \mathcal{Z} \), the eavesdropper, obtains no information about \( M \). More precisely, the eavesdropper's signal or observation \( Z^n \) is required to be asymptotically independent of \( M \). There are various ways to quantify asymptotic independence. We adopt the methodology of Bloch and Laneman [5] and consider six metrics of varying strengths that quantify asymptotic independence.

**Definition 1.** Let \( \eta \) be a positive constant. Consider the following measures of independence (more correctly, dependence), also known as secrecy metrics:

\[
\begin{align*}
S_1(P_{MZ^n,M \times P_{Z^n}}) &:= D(P_{MZ^n} \parallel P_M \times P_{Z^n}) = I(M; Z^n), \\
S_2(P_{MZ^n,M \times P_{Z^n}}) &:= \mathbb{V}(P_{MZ^n}, P_M \times P_{Z^n}), \\
S_3^n(P_{MZ^n,M \times P_{Z^n}}) &:= \Pr \left( \log \frac{P_{MZ^n}(M, Z^n)}{P_M \times P_{Z^n}(M, Z^n)} > \eta \right), \\
S_4(P_{MZ^n,M \times P_{Z^n}}) &:= \frac{1}{n} D(P_{MZ^n} \parallel P_M \times P_{Z^n}) = \frac{1}{n} I(M; Z^n), \\
S_5(P_{MZ^n,M \times P_{Z^n}}) &:= \frac{1}{n} \mathbb{V}(P_{MZ^n}, P_M \times P_{Z^n}), \quad \text{and} \\
S_6^n(P_{MZ^n,M \times P_{Z^n}}) &:= \Pr \left( \frac{1}{n} \log \frac{P_{MZ^n}(M, Z^n)}{P_M \times P_{Z^n}(M, Z^n)} > \eta \right).
\end{align*}
\]  

(5)\(\text{ }\) (6)\(\text{ }\) (7)\(\text{ }\) (8)\(\text{ }\) (9)\(\text{ }\) (10)

Secrecy metrics \( S_1 \) and \( S_4 \) correspond respectively to strong [22], [23] and weak secrecy [3] in the information-theoretic security literature [1], [2]. We say that \( S_i \) dominates \( S_j \) if \( S_i(P_{MZ^n,M \times P_{Z^n}}) \rightarrow 0 \) implies that \( S_j(P_{MZ^n,M \times P_{Z^n}}) \rightarrow 0 \). This is denoted by \( S_i \geq S_j \). Bloch and Laneman [5, Prop. 1] showed that there exists an ordering of the above six secrecy metrics. In particular, for any \( \eta_1, \eta_2 > 0 \),

\[
S_1 \geq S_2 \geq S_3^n \geq S_4 \geq S_5 \geq S_6^n.
\]  

(11)

Given the wiretap channel \( W = \{W^n\}_{n \in \mathbb{N}} \), we define its \( \mathcal{Y}^n \)- and \( \mathcal{Z}^n \)-marginals as

\[
W^n_\mathcal{Y}(y|x) := \sum_{z \in \mathcal{Z}^n} W^n(y, z|x), \quad \text{and} \quad W^n_\mathcal{Z}(z|x) := \sum_{y \in \mathcal{Y}^n} W^n(y, z|x),
\]  

(12)

where \( (x, y, z) \in \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n \) is a tuple of vectors of length \( n \).

**Definition 2.** An \( (n, M_n, \varepsilon_n, \delta_n) \)-wiretap code for secrecy metric \( S_i, i \in [1 : 6] \) consists of

1) A message set \( \mathcal{M}_n = [1 : M_n] \);
2) A stochastic encoder \( Q_{X^n|M_n} : \mathcal{M}_n \rightarrow \mathcal{X}^n \) and
3) A decoder \( \varphi_n : \mathcal{Y}^n \rightarrow \mathcal{M}_n \)

such that the average error probability satisfies

\[
\frac{1}{M_n} \sum_{m \in \mathcal{M}_n} \sum_{x \in \mathcal{X}^n} Q_{X^n|M_n}(x|m) W^n_\mathcal{Y}(\mathcal{Y}^n \setminus \varphi^{-1}_n(m)|x) \leq \varepsilon_n
\]  

(13)

and the information leakage satisfies

\[
S_i(P_{MZ^n,M \times P_{Z^n}}) \leq \delta_n
\]  

(14)

where \( M \in \mathcal{M}_n \) is the message random variable which is assumed to be uniformly distributed over \( \mathcal{M}_n \).
We remark that secrecy metrics $S^\eta_0$ and $S^\eta_i$ depend on another parameter $\eta > 0$ but to simplify notation, we do not make the dependence of the code on $\eta$ explicit. This should not cause any confusion in the sequel.

We now define achievable rates and capacities for the general wiretap channel.

**Definition 3.** Let $\varepsilon \in [0, 1)$ and $i \in [1 : 6]$. Let $R \in \mathbb{R}$ be called an $(\varepsilon, i)$-achievable rate for the general wiretap channel $W$ if there exists a sequence of $(n, M_n, \varepsilon_n, \delta_n)$-wiretap codes for secrecy metric $i$ such that

$$\limsup_{n \to \infty} \varepsilon_n \leq \varepsilon, \quad \lim_{n \to \infty} \delta_n = 0, \quad \text{and} \quad \liminf_{n \to \infty} \frac{1}{n} \log M_n \geq R.$$  

Define the $(\varepsilon, i)$-secrecy capacity (or simply the $(\varepsilon, i)$-capacity) of the wiretap channel $W$ as

$$C^{(i)}_\varepsilon(W) := \sup \{ R \in \mathbb{R} : R \text{ is } (\varepsilon, i)\text{-achievable} \}.$$  

**Definition 4.** Let $\varepsilon \in (0, 1]$ and $i \in [1 : 6]$. Let $R \in \mathbb{R}$ be called an $(\varepsilon, i)$-optimistically achievable rate for the general wiretap channel $W$ if for all sequences of $(n, M_n, \varepsilon_n, \delta_n)$-wiretap codes for secrecy metric $i$ satisfying

$$\liminf_{n \to \infty} \frac{1}{n} \log M_n \geq R \quad \text{and} \quad \lim_{n \to \infty} \delta_n = 0,$$

we must also have

$$\liminf_{n \to \infty} \varepsilon_n \geq \varepsilon.$$  

Define the $(\varepsilon, i)$-optimistic secrecy capacity (or simply the $(\varepsilon, i)$-optimistic capacity) of the wiretap channel $W$ as

$$\overline{C}^{(i)}_{\varepsilon}(W) := \inf \{ R \in \mathbb{R} : R \text{ is } (\varepsilon, i)\text{-optimistically achievable} \}.$$  

**Definition 4.** Let $\varepsilon \in (0, 1)$ and $i \in [1 : 6]$. Let $R \in \mathbb{R}$ be called an $(\varepsilon, i)$-optimistically achievable rate for the general wiretap channel $W$ if for all sequences of $(n, M_n, \varepsilon_n, \delta_n)$-wiretap codes for secrecy metric $i$ satisfying

$$\liminf_{n \to \infty} \frac{1}{n} \log M_n \geq R \quad \text{and} \quad \lim_{n \to \infty} \delta_n = 0.$$  

we must also have

$$\liminf_{n \to \infty} \varepsilon_n \geq \varepsilon.$$  

Define the $(\varepsilon, i)$-optimistic secrecy capacity (or simply the $(\varepsilon, i)$-optimistic capacity) of the wiretap channel $W$ as

$$\overline{C}^{(i)}_{\varepsilon}(W) := \inf \{ R \in \mathbb{R} : R \text{ is } (\varepsilon, i)\text{-optimistically achievable} \}.$$  

Following [24], and by contrapositive, we note that the $(\varepsilon, i)$-optimistic capacity can equivalently be defined as the supremum of all numbers $R \in \mathbb{R}$ for which there exists a sequence of $(n, M_n, \varepsilon_n, \delta_n)$-wiretap codes for secrecy metric $i$ such that

$$\liminf_{n \to \infty} \varepsilon_n < \varepsilon, \quad \lim_{n \to \infty} \delta_n = 0, \quad \text{and} \quad \liminf_{n \to \infty} \frac{1}{n} \log M_n \geq R.$$  

The first condition in (23) explains the term optimistic. Indeed, by the definition of lim inf, roughly speaking, the error probability is only required to be smaller than $\varepsilon$ for infinitely many $n$ as opposed to for all sufficiently large $n$, implied by the first condition in (15) for the (pessimistic) capacity. Note that our definition of the optimistic capacity in Definition 4, or equivalently the conditions in (23), is slightly different from those in Chen and Alajaji [18, Def. 4.9] and Steinberg [19, Thm. 7]. Our definition has the advantage that it allows us to characterize the $(\varepsilon, i)$-optimistic secrecy capacity as an equality for all $\varepsilon \in (0, 1]$. See a discussion of this subtlety in [14, Sec. IV] and [16, Rmk. 1.6.3].

From the ordering of the secrecy metrics in (11), we know that for every $\varepsilon \in (0, 1]$, we have

$$\overline{C}^{(i)}_{\varepsilon}(W) \leq \overline{C}^{(j)}_{\varepsilon}(W), \quad \text{if} \quad i \leq j.$$  

It is also easily seen from the definitions that for all $i \in [1 : 6]$,

$$C^{(i)}(W) \leq \overline{C}^{(i)}(W).$$
Equality in (25) is particularly significant as can be seen from the following definition.

**Definition 5.** A wiretap channel $W$ is said to satisfy the partial strong converse under secrecy metric $i \in [1:6]$ if

$$C_i^{(i)}(W) = \bar{C}_i^{(i)}(W).$$  

(26)

The qualifier *partial* is used because we still insist that the information leakage, represented by $\delta_n$, tends to zero. The strong converse thus only pertains to the probability of decoding error in (13). This definition of the partial strong converse corresponds to that presented by Han [16, Sec. 3.7] and Hayashi and Nagaoka in [25]. Clearly, if $W$ satisfies the partial strong converse under secrecy metric $i$, both $C_i^{(i)}(W)$ and $\bar{C}_i^{(i)}(W)$ do not depend on $\varepsilon$. More precisely, the partial strong converse implies that

$$C_i^{(i)}(W) = C^{(i)}(W) \quad \forall \varepsilon \in [0, 1] \quad \text{and}$$

$$\bar{C}_i^{(i)}(W) = \bar{C}^{(i)}(W) \quad \forall \varepsilon \in (0, 1].$$  

(27) \hspace{1cm} (28)

However, as discussed in [16, Rmk. 3.5.1], Definition 5 implies (27)–(28) but not the other way round.

III. MAIN RESULTS

In this section, we state the main results in this paper. First, we generalize the results in [5], [17] and characterize $C_i^{(i)}(W)$ and $\bar{C}_i^{(i)}(W)$ for general wiretap channels. We then state our main result, namely that degraded DM-WTCs admit the partial strong converse. We also show that certain classes of non-stationary wiretap channels and the G-WTC possess the partial strong converse property.

A. Quantities in Information Spectrum Analysis

To state our results concisely, we recall some definitions from information spectrum analysis [14]–[16], [20]. For any general source pair $(V, Y)$ with joint distribution $P_{VY} := \{P_{V^nY^n}\}_{n \in \mathbb{N}}$, define the normalized information density random variable

$$i_{V^nY^n}(V^n; Y^n) := \frac{1}{n} \log P_{Y^n|V^n}(Y^n|V^n) P_{V^n}(Y^n).$$  

(29)

Further define for $\varepsilon \in [0, 1)$, the $\varepsilon$-$p$-lim inf of the sequence of random variables $\{i_{V^nY^n}(V^n; Y^n)\}_{n \in \mathbb{N}}$ as

$$L_\varepsilon(V; Y) := \sup \left\{ R \in \mathbb{R} : \limsup_{n \to \infty} \Pr(i_{V^nY^n}(V^n; Y^n) \leq R) \leq \varepsilon \right\};$$

(30)

and for $\varepsilon \in (0, 1]$, the $\varepsilon$-$p$-lim sup of the sequence of random variables $\{i_{V^nY^n}(V^n; Y^n)\}_{n \in \mathbb{N}}$ as

$$\bar{T}_\varepsilon(V; Y) := \sup \left\{ R \in \mathbb{R} : \liminf_{n \to \infty} \Pr(i_{V^nY^n}(V^n; Y^n) \leq R) < \varepsilon \right\}.$$  

(31)

Notice the strict inequality in (31), which differs from the non-strict inequality in (30). The properties of $L_\varepsilon(V; Y)$ and $\bar{T}_\varepsilon(V; Y)$ are described in [20, Sec. 2.4]. When $\varepsilon = 0$ in (30) and $\varepsilon = 1$ in (31), we leave out the subscripts, i.e., we define

$$I(V; Y) := L_0(V; Y), \quad \text{and} \quad \bar{T}(V; Y) := \bar{T}_1(V; Y).$$  

(32)

These respectively coincide with the usual $p$-lim inf and $p$-lim sup [16] of the sequence of random variables $\{i_{V^nY^n}(V^n; Y^n)\}_{n \in \mathbb{N}}$. In information spectrum analysis, $I(V; Y)$ and $\bar{T}(V; Y)$ are termed the spectral inf- and sup-mutual information rates respectively.
B. Capacity and Strong Converse Results for General Wiretap Channels

The following theorem is a straightforward extension of the results in [5], [17].

**Theorem 1 (General Formula).** For \( i \in [2 : 6] \), the \((\varepsilon, i)\)-capacity and the \((\varepsilon, i)\)-optimistic capacity of any general wiretap channel \( W \) are

\[
C^{(i)}_\varepsilon(W) = \sup_{V - X - (Y, Z)} \mathcal{L}(V; Y) - \overline{I}(V; Z), \quad \text{and}
\]

\[
\overline{C}^{(i)}_\varepsilon(W) = \sup_{V - X - (Y, Z)} \overline{I}(V; Y) - \overline{I}(V; Z).
\]

(33)

(34)

The suprema are over the set of all sequences of distributions \( P_{V=X} = \{ P_{V^n=X^n} \}_{n \in \mathbb{N}} \) or equivalently over all Markov chains \( V - X - (Y, Z) \) where the distribution of \( (Y, Z) \) given \( X \) corresponds to the wiretap channel \( W \).

For reference, a proof sketch of Theorem 1 is provided in Section V-A. Using the definition of the partial strong converse in Definition 5, we immediately obtain the following corollary.

**Corollary 2 (General Partial Strong Converse).** For any wiretap channel \( W \) and any secrecy metric \( S_i, i \in [2 : 6] \), the partial strong converse property holds if and only if

\[
\sup_{V - X - (Y, Z)} \mathcal{L}(V; Y) - \overline{I}(V; Z) = \sup_{V - X - (Y, Z)} \overline{I}(V; Y) - \overline{I}(V; Z).
\]

(35)

We now make things more concrete by providing an example.

**Example 1.** This example is analogous to [5, Prop. 4]. Let \( W_a \) and \( W_b \) be DM-WTCs formed from \( n \)-fold products of constituent channels \( W_a \) and \( W_b \). The channel \( W_j : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}, j \in \{a, b\} \) consists of two channels where the main channel is \( \text{BSC}(\delta_j) \), \( j \in \{a, b\} \) (binary symmetric channel (BSC) with crossover probability \( \delta_j \)) and the eavesdropper’s channel is \( \text{BSC}(\delta_a) \). Note that \( W_j \) can be regarded as a concatenation of two BSCs [1, Example 3.5]. We assume that the crossover parameters satisfy

\[
0 < \delta_a < \delta_b < \delta_e < \frac{1}{2}.
\]

(36)

Now define the mixed channel

\[
W = \alpha W_a + (1 - \alpha) W_b
\]

(37)

where \( \alpha \in (0, 1) \). Let \( h(q) := -q \log_2 q - (1 - q) \log_2 (1 - q) \) be the binary entropy. Since the capacity-achieving input distribution is uniform in this symmetric example [1, Lem. 3.2], the \((\varepsilon, i)\)-capacity and the \((\varepsilon, i)\)-optimistic capacity for \( i \in [2 : 6] \) can be computed to be respectively

\[
C^{(i)}_\varepsilon(W) = \begin{cases} 
  h(\delta_e) - h(\delta_b) & \varepsilon < \alpha \\
  h(\delta_e) - h(\delta_a) & \varepsilon \geq \alpha
\end{cases}, \quad \text{and}
\]

\[
\overline{C}^{(i)}_\varepsilon(W) = \begin{cases} 
  h(\delta_e) - h(\delta_b) & \varepsilon \leq \alpha \\
  h(\delta_e) - h(\delta_a) & \varepsilon > \alpha
\end{cases}.
\]

(38)

(39)

By definition, \( \varepsilon \mapsto C^{(i)}_\varepsilon(W) \) and \( \varepsilon \mapsto \overline{C}^{(i)}_\varepsilon(W) \) are upper and lower semi-continuous respectively. This is reflected in (38) and (39). Note that \( C^{(i)}_\varepsilon(W) \) and \( \overline{C}^{(i)}_\varepsilon(W) \) coincide at all points on \((0, 1)\) except the point of discontinuity \( \varepsilon = \alpha \). Since

\[
C^{(i)}(W) = h(\delta_e) - h(\delta_b) < \overline{C}^{(i)}(W) = h(\delta_e) - h(\delta_a),
\]

(40)

the partial strong converse does not hold under any secrecy metric \( i \in [2 : 6] \). However, if \( W_a = W_b \), so the channel \( W \) is the memoryless extension of a concatenation of two BSCs, the partial strong converse holds under secrecy metric \( i \in [2 : 6] \). A more general version of this result is presented in Theorem 3 to follow.

In fact, it can be shown (see Theorem 3) by using exponential-type concentration bounds (e.g., Chernoff bounds) that in fact, \( C^{(i)}(W) = \overline{C}^{(i)}(W) \) for all \( i \in [1 : 6] \) if \( W \) is a DM-WTC so the partial strong converse also holds under secrecy metric \( S_1 \). In fact, under many scenarios, \( S_1 \) decays exponentially fast as discussed by Csiszár [23], Kobayashi et al. [26, Sec. V] and Hayashi [17], [27].

2The notation \( A - B = C \) means that \( A^n - B^n = C^n \) forms a Markov chain for all \( n \in \mathbb{N} \).
C. Strong Converse Theorems for Specific Wiretap Channel Models

1) Degraded Discrete Memoryless Wiretap Channels: A physically degraded, or simply degraded, wiretap channel $\mathbf{W}$ is one in which for every $n \in \mathbb{N}$, and for every $(x, y, z) \in \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$,

$$W^n(y, z|x) = W^n_1(y|x)W^n_2(z|y) \quad (41)$$

for some channels $W^n_1: \mathcal{X}^n \rightarrow \mathcal{Y}^n$ and $W^n_2: \mathcal{Y}^n \rightarrow \mathcal{Z}^n$. In other words, $X^n - Y^n - Z^n$ forms a Markov chain for every $n \in \mathbb{N}$. A DM-WTC has alphabets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ that are finite sets and the channel is stationary and memoryless in the sense that

$$W^n_1(y|x) = \prod_{i=1}^{n} W_1(y_i|x_i), \quad \text{and} \quad W^n_2(z|y) = \prod_{i=1}^{n} W_2(z_i|y_i) \quad (42)$$

for every $(x, y, z) \in \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$. It is known from Wyner’s seminal work on the wiretap channel [3] that the capacity of a degraded DM-WTC $\mathbf{W}: \mathcal{X} \rightarrow \mathcal{Y} \times \mathcal{Z}$ (under the weak secrecy criterion $S_4$) is

$$C_{s}^{\text{DM}}(\mathbf{W}) := \max_{P_X \in P(\mathcal{X})} I(X; Y|Z) = \max_{P_X \in P(\mathcal{X})} I(X; Y) - I(X; Z), \quad (43)$$

where the mutual information quantities are calculated according to $\Pr(Y = y, Z = z|X = x) = W(y, z|x)$. The second equality in (43) follows from the fact that $X - Y - Z$ forms a Markov chain so $I(X; Y|Z) = I(X; Y) - I(X; Z)$. Wyner’s weak converse [3, Eq. (35) in Sec. IV] assumes that the probability of decoding error vanishes asymptotically. The first of our main results is a strengthening of Wyner’s seminal result.

**Theorem 3** (Degraded DM-WTCs). Any degraded DM-WTC $\mathbf{W}: \mathcal{X} \rightarrow \mathcal{Y} \times \mathcal{Z}$ satisfies the partial strong converse under any secrecy metric $S_i, i \in [1 : 6]$. Consequently, the $(i)$-capacities and $(i)$-optimistic capacities of $\mathbf{W} = \{W\}$ are equal to $C_{s}^{\text{DM}}(\mathbf{W})$ for all $i \in [1 : 6]$.

The proof of this theorem is provided in Section V-B. The basic idea is to lower bound the $(1)$-capacity $C^{(1)}(\mathbf{W})$ (capacity under secrecy metric $S_1$) with $C_{s}^{\text{DM}}(\mathbf{W})$ and to upper bound the $(6)$-optimistic capacity $\overline{C}^{(6)}(\mathbf{W})$ (optimistic capacity under secrecy metric $S_6$) with the same quantity $C_{s}^{\text{DM}}(\mathbf{W})$. This then allows us to assert that $C^{(1)}(\mathbf{W}) = \overline{C}^{(6)}(\mathbf{W})$ showing from (18), (24) and (25) that $C^{(i)}(\mathbf{W}) = \overline{C}^{(i)}(\mathbf{W})$ for all $i \in [1 : 6]$, i.e., the partial strong converse holds under all 6 secrecy metrics. The lower bound of $C^{(1)}(\mathbf{W})$ is straightforward and follows from using independent and identically distributed (i.i.d.) random codes and standard concentration bounds. This is already well known, see for example [5, Remark 3] or the papers by Hayashi on error exponents for the wiretap channel [17], [27]. The interesting part of the proof is in the upper bound of

$$\overline{C}^{(6)}(\mathbf{W}) = \sup_{\mathbf{V} \rightarrow \mathbf{X} \rightarrow \mathbf{Y}, \mathbf{Z}} \overline{T}({\mathbf{V}; {\mathbf{Y}}}) - \overline{T}({\mathbf{V}; \mathbf{Z}}). \quad (44)$$

The difficulty arises because we need to upper bound and subsequently single-letterize the supremum of the difference between two limit superiors in probability. To perform these tasks, we leverage the proof technique for [16, Thm. 3.7.2] and combine several known results and techniques from the information theory literature.

We remark that it is not difficult to make the same assertion as in Theorem 3 for degraded DM-WTCs with additive cost constraints on $\mathbf{X}$, e.g., $X^n = (X_1, \ldots, X_n)$ must satisfy $\sum_{i=1}^{n} b(X_i) \leq n\Gamma$ almost surely for some cost function $b: \mathcal{X} \rightarrow [0, \infty)$ and $\Gamma > 0$. This observation will be useful for the Gaussian case in Section III-C3.

2) Non-Stationary Wiretap Channels: The assumption of degradedness in Theorem 3 is rather strong but appears essential in the proof. We do not think that the assumption concerning memorylessness is critical (cf. [24, Cor. 3]), but we defer the study of wiretap channels with memory to future work. Instead we examine conditions under which the stationarity assumption may be relaxed. In this section, we assume that the wiretap channel is degraded in the sense of (41) but the components have the following non-stationary structure:

$$W^n_1(y|x) = \prod_{i=1}^{n} W_{1i}(y_i|x_i), \quad \text{and} \quad W^n_2(z|y) = \prod_{i=1}^{n} W_{2i}(z_i|y_i). \quad (45)$$
That is, the channels themselves may differ across time but the channel noises are nonetheless independent. We define the $i$-th wiretap channel as $W_i(y, z|x) := W_1(y|x)W_2(z|y)$. The main and eavesdropper’s channels are defined as
\[
W_{Y, i}(y|x) := W_1(y|x), \quad \text{and} \quad W_{Z, i}(z|x) := \sum_{y \in Y} W_1(y|x)W_2(z|y)
\]
respectively. These channels have Shannon capacities $C(W_{Y, i})$ and $C(W_{Z, i})$ respectively. We further assume that all component channels $\{W_{Y, i}\}_{i \in \mathbb{N}}$ and $\{W_{Z, i}\}_{i \in \mathbb{N}}$ are weakly symmetric [1, Def. 3.4]. Recall that a discrete memoryless channel $V : \mathcal{X} \rightarrow \mathcal{Y}$ is weakly symmetric if the rows of the channel transition probability matrix are permutations of each other and the column sums $\sum_{x \in \mathcal{X}} V(y|x)$ are independent of $y$. Under the condition that the channels are degraded and weakly symmetric, Leung-Yan-Cheong [28] (also see [1, Prop. 3.2]) showed that the secrecy capacity is the difference of the capacities of the main and eavesdropper’s channels, i.e.,
\[
C^\text{DM}(W_i) = C(W_{Y, i}) - C(W_{Z, i}).
\]

Note that (47) is a consequence of the fact that the (unique) capacity-achieving input distributions of the channels $W_{Y, i}$ and $W_{Z, i}$ are the same and, in particular, they are uniform on $\mathcal{X}$. See van Dijk [29]. With these preparations, we are in a position to state the following result:

**Theorem 4** (Non-Stationary Wiretap Channels). Consider the degraded, discrete, memoryless but non-stationary wiretap channel in (45)–(46). Assume that all $\{W_{Y, i}\}_{i \in \mathbb{N}}$ and $\{W_{Z, i}\}_{i \in \mathbb{N}}$ are weakly symmetric channels.

1. Under secrecy metrics $S_i, i \in [2 : 6]$, the partial strong converse holds for $W = \{W^n\}_{n \in \mathbb{N}}$ if
\[
\liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} C(W_{Y, i}) - \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} C(W_{Z, i}) = \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} C^\text{DM}(W_i).
\]

2. The condition in (48) holds if and only if sequences $\frac{1}{n} \sum_{i=1}^{n} C(W_{Y, i})$ and $\frac{1}{n} \sum_{i=1}^{n} C(W_{Z, i})$ converge.

3. If in addition to (48), the channels $\{W_{Z, i}\}_{i \in \mathbb{N}}$ are uniformly positive, i.e.,
\[
\kappa_{Z} := \inf_{i \in \mathbb{N}} \min_{x, z \in \mathcal{X} \times \mathcal{Z}} W_{Z, i}(z|x) > 0,
\]
the partial strong converse also holds under secrecy metric $S_1$.

The proof of this theorem, which builds on that of Theorem 3, is provided in Section V-C.

For the purposes of comparison, consider a point-to-point, discrete, memoryless and non-stationary channel $V^n(y|x) = \prod_{i=1}^{n} V_i(y_i|x_i)$. Let $P_{Y_i}$ be the unique [30, Cor. 2 to Thm. 4.5.2] capacity-achieving output distribution of $V_i$. It satisfies $P_{Y_i}(y) > 0$ for all $y \in \mathcal{Y}$ if all outputs are reachable [30, Cor. 1 to Thm. 4.5.2]. Further assume that
\[
\sup_{i \in \mathbb{N}} \max_{x \in \mathcal{X}} \text{Var} \left[ \log \frac{V_i(Y|x)}{P_{Y_i}(Y)} \right] < \infty.
\]

Then, it is easy to show from the strong converse theorem for general channels [16, Thm. 3.5.2] that the strong converse holds if and only if
\[
\liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} C(V_i) = \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} C(V_i).
\]

In other words, the sequence $\frac{1}{n} \sum_{i=1}^{n} C(V_i)$ converges. Indeed, the left-hand-side of (51) is the capacity [16, Rmk. 3.2.3] of the channel $V := \{V^n = \prod_{i=1}^{n} V_i\}_{n \in \mathbb{N}}$, while if we assume (50), the right-hand-side is the optimistic capacity (a statement generalizing [20, Example 5.14]). Thus, the equivalent condition in terms of the existence of the limits of the limits of the Cesàro means $\frac{1}{n} \sum_{i=1}^{n} C(W_{Y, i})$ and $\frac{1}{n} \sum_{i=1}^{n} C(W_{Z, i})$ in Theorem 4 is a generalization of channels without secrecy constraints to degraded (but weakly symmetric) wiretap channels.

Of course, (48) is only a sufficient condition for the partial strong converse to hold. It appears to be rather challenging to assert that it is also necessary, or to find an alternate and stronger characterization that is both necessary and sufficient. This is due, in part, to the fact that it is difficult to upper bound $C^{(2)}(W)$ by the left-hand-side of (48). The more capable condition for general wiretap channels introduced by Han et al. [31, Def. 2.3] is difficult to verify even for degraded DM-WTCs so we cannot replace the optimization over $\{V, X\}$ in the
expression for $\mathcal{C}^{(2)}(W)$ by an optimization over only $X$ (cf. [31, Thm. 2.2]). The uniform positivity assumption in (49) is used to prove a Hoeffding-type bound [32] on route to showing that the partial strong converse holds under $S_1$. The Gärtner-Ellis theorem does not apply in general because the non-stationarity of the channels results in the relevant sequence of cumulant generating functions to not converge. One way to remedy this is to resort to generalizations such as [33] but we leave this for future work.

3) Gaussian Wiretap Channels: We now demonstrate that the assumption of discreteness in Theorem 3 is not critical. In fact, we can make a partial strong converse statement for the (memoryless, stationary) G-WTC in which all the alphabets are the real line $\mathbb{R}$ and the channel laws are

$$W_{Y}(y|x) := \mathcal{N}(y; x, \sigma_1^2), \quad \text{and} \quad W_{Z}(z|x) := \mathcal{N}(z; x, \sigma_2^2),$$

and we assume that $\sigma_2 > \sigma_1$. Observe that by defining $W_1 = W_{Y}$ and

$$W_{Z}(z|y) := \mathcal{N}(z; y, \sigma_2^2 - \sigma_1^2),$$

we see that the G-WTC $W(y, z|x) = W_1(y|x)W_{Z}(z|y)$ is degraded. In fact, to be more correct, it is stochastically degraded but we will not differentiate between physical degradedness and stochastic degradedness since the capacities and optimistic capacities are identical, a direct consequence of [2, Lem. 2.1]. For every blocklength $n \in \mathbb{N}$, the input codeword $X^n$ is required to satisfy the almost sure power constraint

$$\Pr(X^n \in \mathcal{F}_n) = 1$$

(55)

where $\mathcal{F}_n := \{x \in \mathbb{R}^n : \|x\|^2 \leq nS\}$ is the $(n - 1)$-sphere with radius $\sqrt{nS}$ and $S > 0$ is the permissible power. Recall from Leung-Yan-Cheong and Hellman [21] that the capacity of the G-WTC, under secrecy metric $S_1$ (weak secrecy) and assuming that the probability of decoding error vanishes asymptotically, is

$$C_s(G; W; S) := \frac{1}{2} \log \left( 1 + \frac{S}{\sigma_1^2} \right) - \frac{1}{2} \log \left( 1 + \frac{S}{\sigma_2^2} \right).$$

(56)

Thus the capacity of the G-WTC is the difference between the Shannon capacities of the main and eavesdropper’s channels.

**Theorem 5** (Gaussian Wiretap Channels). The (memoryless, stationary) G-WTC satisfies the partial strong converse under any secrecy metric $S_i$, $i \in [1 : 6]$. Consequently, the $(i)$-capacities and $(i)$-optimistic capacities of $W = \{W\}$ under the cost constraint in (55) are equal to $C_s^G(W; S)$ for all $i \in [1 : 6]$.

The proof of this theorem, which builds on that of Theorem 3, is provided in Section V-D. One of the additional complications (vis-à-vis Theorem 3) we have to overcome is the need to carefully handle the almost sure cost constraint in (55) to ensure the statement holds for $S_1$. Similarly to the proof of Theorem 5, one can show, using the discretization procedure outlined in Han et al. [31, Sec. 6], that the degraded Poisson wiretap channel, studied by Laourine and Wagner [34], admits a partial strong converse.

**IV. DISCUSSION**

A. Other Possible Proof Techniques

We adopted the information spectrum approach in this work to establish a partial strong converse for various classes of wiretap channels. There are several other strong converse techniques in the information theory literature that we considered but eventually did not adopt, as it appeared that there were restrictions on their applicability to the wiretap channel. See [35, Sec. 5] for a comprehensive literature survey on techniques to prove strong converses for various channels, particularly discrete memoryless channels.

The blowing-up lemma [36] is a powerful tool for proving strong converses in multi-user information theory. For example, it was used in [36] to prove the strong converse of the “one-help-one” problem (lossless coding with coded side-information) and the broadcast channel with degraded message sets. In fact, it initially seemed to the authors to be the most natural approach for the problem at hand. For discrete memoryless channels, it allows us to assert that for every permissible (non-vanishing) maximum error probability $\varepsilon \in (0, 1)$, we have

$$H(M|Y^n) \leq n \cdot \zeta_n(\varepsilon)$$

(57)
where \( \zeta_n(\varepsilon) \) denotes some \( \varepsilon \)-dependent sequence that tends to zero that \( n \) tends to infinity. This key statement, combined with Fano’s inequality for list decoding, allows us to prove the strong converse for discrete memoryless channels. See [37, Sec. 3.6.1] for a modern treatment of this approach. Further if (57) were established for the wiretap channel, we would not need to assume degradedness. However, the blowing-up approach heavily relies on the fact that the encoder is a deterministic function from the message set \( M_n \) to the codeword space \( \mathcal{X}^n \). In the wiretap channel problem, the encoder is stochastic (cf. (13)) and so the blowing-up lemma does not seem to be directly applicable to prove (57). Finally, the blowing-up technique, to the best of the authors’ knowledge, is not readily applicable to systems with continuous alphabets such as the G-WTC (Theorem 5).

The strong converse for memoryless channels can also be proved by appealing to the properties of the Rényi-divergence (e.g., [38]), such as data processing. This was used implicitly by Arimoto [39] to determine the exponent of the probability of correct decoding at rates above capacity. However, it does not appear straightforward to use this technique to establish the strong converse property for the system we are interested in—the wiretap channel.

### B. Other Related Works

As mentioned in the Introduction, Morgan and Winter [9, Sec. VI] proved a pretty strong converse for the private capacity of degradable quantum channels [11]. They showed that the private capacity does not depend on the bound on the error \( \varepsilon \) and bound on the secrecy metric \( \delta \) (variation distance) as long as \( \sqrt{\varepsilon} + 2\sqrt{\delta} < 1/2 \) (also see footnote 1). The powerful min- and max-entropy calculus (see [10] for example) underlies the proofs in [9].

In the information-theoretic security context, this was used to prove the strong converse for the secret key agreement problem in [7], [8]. However, it does not seem to be applicable here due to the presence of the optimization over the input distribution in \( C_{x,y}^{\text{DM}}(W) \). There is no such optimization for the key agreement problem. Just as in [9], the advantage of the hypothesis testing approach is that both \( \varepsilon \) and \( \delta \) are allowed to be non-vanishing. In fact, \( \varepsilon + \delta < 1 \) suffices for the strong converse to hold in [7], [8]. See Fig. 1. This is in contrast to the strong converse result in Tyagi and Narayan [6, Sec. VII] in which, like this work, the error probability is non-vanishing but the leakage \( S_2 \) is required to vanish. A possible way to prove a pretty strong converse pertaining to both the error probability and the leakage in the context of the wiretap channel is to leverage on a recently developed non-asymptotic converse bound for channel resolvability [42, Lem. 3].

### C. Future Work

We discuss three promising avenues of further research.

First, in this paper, we were only concerned with the transmission of a single message from the sender to the legitimate receiver. Csiszár and Körner [4] considered the broadcast channel with confidential messages model in which two messages are to be sent, both to the legitimate receiver and only one to the eavesdropper. The eavesdropper’s signal is to be asymptotically independent of the non-intended message. It may be possible to prove a partial strong converse in this multi-terminal system but we note that the information spectrum technique does not extend in a straightforward manner to show that discrete memoryless multi-terminal systems, such as the multiple-access channel [43], admit the strong converse thus new techniques must be developed.

Second, as in [24], it may be possible to use the techniques contained herein to study wiretap channels with limited memory (such as channels with additive Markov noise) and show that they admit a partial strong converse. However, wiretap channels with Markov memory have not been studied extensively previously.

Finally, and most ambitiously, it would be interesting to study whether a full (and not partial or pretty) strong converse holds for some classes of wiretap channels, i.e., whether the capacity depends on \( (\varepsilon, \delta) \) for \( \varepsilon + \delta < 1 \).
However, this appears to require general capacity formula with non-vanishing error probability and non-vanishing leakage, which in turn requires the evaluation a convenient non-asymptotic converse bound for channel resolvability. Initial work on refinements of non-asymptotic and asymptotic channel resolvability bounds has been conducted by Watanabe and Hayashi [42]. On a separate note, one-shot (non-asymptotic) bounds on the wiretap capacity for non-zero $(\varepsilon, \delta)$ were proved by Renes and Renner [10] using min- and max-entropy calculus.

V. PROOFS

A. Proof Sketch of Theorem 1

Proof: We prove the achievability statement for the strongest secrecy metric $S_2$ and the converse statement for the weakest secrecy metric $S_6$.

For achievability, fix a sequence of input distributions $P_{VX} = \{P_{V^n,X^n}\}_{n\in\mathbb{N}}$. For each message $m \in [1 : M_n]$, generate a subcodebook $C(m)$ consisting of $M_n/M_n$ randomly and independently generated sequences $v(l)$, $l \in [1 + (m - 1)M_n/M_n : mM_n/M_n]$, each according to $P_{V^n}$. The codebook is revealed to all parties including the eavesdropper. Given $m \in [M_n]$, the encoder chooses an index $I$ uniformly at random from $[1 + (m - 1)M_n/M_n : mM_n/M_n]$ and generates $x(m) \sim P_{X^n|V^n}(\cdot|v(L))$ as the channel input.

Let $\gamma > 0$. Given $y \in Y^n$, the legitimate receiver finds the unique message $\hat{m}$ such that $(v(l), y) \in T^{(n)}_\gamma$ for some $v(l) \in C(\hat{m})$, where

$$T^{(n)}_\gamma := \left\{ (v, y) \in Y^n \times Y^n : \frac{1}{n} \log \frac{P_{Y^n|V^n}(y|v)}{P_{Y^n}(y)} \geq \frac{1}{n} \log \bar{M}_n + \gamma \right\}.$$  

(58)

Let $\varepsilon_n$ be the average error probability of the legitimate receiver (over the random message and the random code) given by (13). By a standard calculation, we have

$$\varepsilon_n \leq P_{V^n,Y^n}( (Y^n \times Y^n) \setminus T^{(n)}_\gamma ) + \exp(-n\gamma).$$  

(59)

From $\varepsilon$-capacity [16, Sec. 3.4] and $\varepsilon$-optimistic capacity analysis [18, Thm. 4.3] (or simply by using the definitions of $T^{(n)}_\gamma$, $L_\varepsilon$ and $\bar{T}_\varepsilon$ in (59)), we know that if $\bar{M}_n$ is chosen such that

$$\frac{1}{n} \log \bar{M}_n \leq L_\varepsilon(V; Y) - 2\gamma,$$  

(60)

then $\limsup_{n \to \infty} E[\varepsilon_n] \leq \varepsilon$, where the expectation is over the random code. Similarly if $\bar{M}_n$ is chosen such that

$$\frac{1}{n} \log \bar{M}_n \leq \bar{T}_\varepsilon(V; Y) - 2\gamma,$$  

(61)

then $\liminf_{n \to \infty} E[\varepsilon_n] < \varepsilon$. From the secrecy from resolvability condition in [5, Lem. 2], we know that if

$$\frac{1}{n} \log \bar{M}_n - \frac{1}{n} \log M_n \geq \bar{T}(V; Z) + 2\gamma,$$  

(62)

then $\liminf_{n \to \infty} E[S_2] = 0$. Now because averaged over the random code, $S_2$ tends to zero, by a Markov inequality argument (see proof of [44, Thm. 1] for example), there exists a sequence of codes such that both the reliability and security conditions are satisfied. This completes the direct part of Theorem 1 upon eliminating $M_n$ from the above inequalities, taking $\liminf_{n \to \infty}$, and finally taking $\gamma \downarrow 0$.

For the converse, by using the Verdú-Han lemma [14, Lem. 4] we know that if $\limsup_{n \to \infty} \varepsilon_n \leq \varepsilon$, for every $\gamma > 0$, we must have that

$$\frac{1}{n} \log M_n \leq L_\varepsilon(V; Y) + \gamma,$$  

(63)

for some chain $V \to X \to (Y, Z)$ and all $n$ sufficiently large (depending on $\gamma$). The auxiliary random process $V$ represents the sequence of messages which are uniform on the message sets $\{M_n\}_{n \in \mathbb{N}}$. Similarly, if $\liminf_{n \to \infty} \varepsilon_n < \varepsilon$, we must have that

$$\frac{1}{n} \log M_n \leq \bar{T}_\varepsilon(V; Y) + \gamma.$$  

(64)

Furthermore, [45, Lem. 4] tells us that if $S_6 \to 0$, we must have that

$$\bar{T}(V; Z) = 0.$$  

(65)

This follows directly from the definitions of $S_6$ in (10) and the spectral sup-mutual information rate. Subtracting $\bar{T}(V; Z)$ from (63) and (64), maximizing over all chains $V \to X \to (Y, Z)$ to make the bound code-independent, and finally taking $\liminf_{n \to \infty}$ and $\gamma \downarrow 0$ completes the converse proof of Theorem 1. ■
B. Proof of Theorem 3

Proof: Here we prove that any degraded DM-WTC $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ satisfies the partial strong converse for any secrecy metric $i \in [1 : 6]$. We proceed in two steps. First, we show that $C_s^{(1)}(W) \geq C_s^{DM}(W)$ (where $W$ is the stationary, memoryless channel induced by $W$) and second, we show that $C_s^{(6)}(W) \leq C_s^{DM}(W)$.

To show that $C_s^{(1)}(W) \geq C_s^{DM}(W)$, we adopt the strategy in [5, Sec. V.C], which is, by now, standard but we reiterate here for the sake of completeness. Particularize the supremum over $P_{Y|X}$ by choosing $V = X$ and also choosing $P_X$ to be a sequence of product distributions induced by any $P_X \in \arg \max_{P_X \in \mathcal{P}(\mathcal{X})} I(X; Y|Z)$, i.e., any capacity-achieving input distribution in the definition of $C_s^{DM}(W)$ in (43). Then, it suffices to appeal to [5, Rmk. 3] which says that if

\[ q_n := \Pr \left( \frac{1}{n} \log \frac{W_Z^n(Z^n|X^n)}{P_{Z^n}(Z^n)} \geq \frac{1}{n} \log \frac{M_n}{M_n} - \gamma \right) \quad (66) \]

decays exponentially fast in $n$ then $S_1 \to 0$ also exponentially fast. This remark was also made by Kobayashi et al. [26, Sec. V], and is a simple consequence of a bound presented by Csiszár in [23, Lem. 1] relating mutual information to variation distance. Choose $M_n$ to be the smallest integer exceeding $\exp[\min\{I(X; Y) - 2\gamma\}]$ (so the decoding error probability tends to zero per (60)), and choose $\bar{M}_n$ to be the largest integer smaller than $\exp[\min\{I(X; Y) - I(X; Z) - 4\gamma\}] = \exp[\min\{C_s^{DM}(W) - 4\gamma\}]$. The mutual informations are computed with respect to the distribution $P_X \times W$. Now, by substituting the choices of $M_n$ and $\bar{M}_n$, $q_n$ can be upper bounded as

\[ q_n \leq \Pr \left( \frac{1}{n} \log \frac{W_Z^n(Z^n|X^n)}{P_{Z^n}(Z^n)} \geq I(X; Z) + \gamma \right) \quad (67) \]

\[ = \Pr \left( \frac{1}{n} \sum_{i=1}^{n} \log \frac{W_Z(Z_i|X_i)}{P_{Z}(Z_i)} \geq I(X; Z) + \gamma \right). \quad (68) \]

Since the mean of the information density random variable $\log W_Z(Z|X) - \log P_Z(Z)$ is $I(X; Z)$, by the Chernoff bound [5, Lem. 6], $q_n$ indeed decays exponentially fast and so $C_s^{(1)}(W) \geq C_s^{DM}(W) - 4\gamma$. Finally, let $\gamma \downarrow 0$.

Now, we prove that $C_s^{(6)}(W) \leq C_s^{DM}(W)$. Starting from (44) and fixing a chain $V - X - (Y, Z)$, we have

\[ \overline{T}(V; Y) - \overline{T}(V; Z) \leq \overline{T}(V; Y, Z) - \overline{T}(V; Z) \quad (69) \]

\[ = \limsup_{n \to \infty} \frac{1}{n} \log \frac{P_{Y^n|V^n}(Y^n, Z^n|V^n)}{P_{Y^n}(Y^n, Z^n)} - \limsup_{n \to \infty} \frac{1}{n} \log \frac{P_{Z^n|V^n}(Z^n|V^n)}{P_{Z^n}(Z^n)} \quad (70) \]

\[ \leq \limsup_{n \to \infty} \left( \frac{1}{n} \log \frac{P_{Y^n|V^n}(Y^n, Z^n|V^n)}{P_{Y^n|Z^n}(Y^n, Z^n)} - \frac{1}{n} \log \frac{P_{Z^n|V^n}(Z^n|V^n)}{P_{Z^n}(Z^n)} \right) \quad (71) \]

\[ = \limsup_{n \to \infty} \frac{1}{n} \log \frac{P_{Y^n|Z^n}(Y^n, Z^n)}{P_{Y^n}(Y^n, Z^n)} = \overline{T}(V; Y|Z) \quad (72) \]

where (69) follows from the sup-version of [14, Thm. 8(f)], (71) is from the sub-additivity of $\limsup$, i.e., that

\[ \limsup_{n \to \infty} (A_n + B_n) \leq \limsup_{n \to \infty} A_n + \limsup_{n \to \infty} B_n, \quad (73) \]

and (72) is by Bayes rule and the definition of the spectral sup-conditional mutual information rate. We further upper bound $\overline{T}(V; Y|Z)$ in (72). By a conditional version of the data processing inequality (sup version) in [14, Thm. 9], we have

\[ \overline{T}(V; Y|Z) \leq \overline{T}(X; Y|Z) \quad (74) \]

because $V - X - (Y, Z)$ forms a Markov chain. Thus, it holds that

\[ \overline{C}^{(6)}(W) \leq \sup_X \overline{T}(X; Y|Z) \quad (75) \]

for any general wiretap channel $W = \{W^n\}_{n \in \mathbb{N}}$. Now, it suffices to simplify the spectral sup-conditional mutual information rate in (75) and, in particular, to show that

\[ \sup_X \overline{T}(X; Y|Z) \leq C_s^{DM}(W), \quad (76) \]
where \( C_{s}^{DM}(W) \) is the capacity of the degraded DM-WTC defined in (43). At this point, we note that Koga and Sato [46] argued (without proof) that \( \sup_{\mathcal{X}} L(X; Y|Z) \leq C_{s}^{DM}(W) \) for degraded DM-WTCs, but (76) is a stronger statement because we are optimizing the spectral sup- (instead of the spectral inf-) conditional mutual information. Hence, an immediate corollary of (76) is Koga and Sato’s claim [46]. For this purpose, define the conditional channel
\[
W_{Y|Z}(y|x, z) := \frac{W(y, z|x)}{\sum_{y \in \mathcal{Y}} W(y, z|x)}.
\]
(77)

We proceed to show (76) by first considering the sequence of random variables
\[
\nu_{n}(X^n; Y^n|Z^n) := \frac{1}{n} \log \frac{W^n(Y^n|X^n, Z^n)}{P_{Y^n|Z^n}(Y^n|Z^n)}
\]
(78)

where \( X = \{X^n\}_{n \in \mathbb{N}} \) is an arbitrary input that induces the output random variables \( (Y, Z) = \{(Y^n, Z^n)\}_{n \in \mathbb{N}} \). Let \( P_{Y^n Z^n} \in \mathcal{P}(\mathcal{Y} \times \mathcal{Z}) \) be a single-letter capacity-achieving output distribution, i.e., a distribution on \( \mathcal{Y} \times \mathcal{Z} \) such that
\[
P_{Y^n Z^n}(y, z) := \sum_{x \in \mathcal{X}} P_{X}(x)W(y, z|x)
\]
(79)

for some \( P_{X} \in \mathcal{P}(\mathcal{X}) \) that achieves the max in (43). By the same logic as [30, Cor. 2 to Thm. 4.5.2], \( P_{Y^n Z^n} \) is unique. In contrast, \( P_{Z} \) is not necessarily unique but, as we will see, this is inconsequential for the subsequent derivations.

Since the p-lim sup is sub-additive as in (73),
\[
\text{p-lim sup}_{n \to \infty} \nu_{n}(X^n; Y^n|Z^n) = \text{p-lim sup}_{n \to \infty} \left\{ \frac{1}{n} \log \frac{W^n(Y^n|X^n, Z^n)}{P_{Y^n|Z^n}(Y^n|Z^n)} - \log \frac{P_{Y^n|Z^n}(Y^n|Z^n)}{P_{Y^n|Z^n}(Y^n|Z^n)} \right\}
\]
(80)
\[
\leq \text{p-lim sup}_{n \to \infty} \frac{1}{n} \log \frac{W^n(Y^n|X^n, Z^n)}{P_{Y^n|Z^n}(Y^n|Z^n)} - \log \frac{P_{Y^n|Z^n}(Y^n|Z^n)}{P_{Y^n|Z^n}(Y^n|Z^n)}
\]
(81)

Now, we argue that the final term is non-negative following the steps of the proof of [16, Lem. 3.2.1]. Fix \( \gamma > 0 \) and consider the probability
\[
\text{Pr}\left( \frac{1}{n} \log \frac{P_{Y^n|Z^n}(Y^n|Z^n)}{P_{Y^n|Z^n}(Y^n|Z^n)} \leq -\gamma \right) = \sum_{z} P_{Z^n}(z) \sum_{y} P_{Y^n|Z^n}(y|z) 1\left\{ \frac{1}{n} \log \frac{P_{Y^n|Z^n}(y|z)}{P_{Y^n|Z^n}(y|z)} \leq -\gamma \right\}
\]
(82)
\[
\leq \sum_{z} P_{Z^n}(z) \sum_{y} P_{Y^n|Z^n}(y|z) \exp(-n\gamma) 1\left\{ \frac{1}{n} \log \frac{P_{Y^n|Z^n}(y|z)}{P_{Y^n|Z^n}(y|z)} \leq -\gamma \right\}
\]
(83)
\[
\leq \exp(-n\gamma).
\]
(84)

Since \( \exp(-n\gamma) \to 0 \), by the definition of p-lim inf, we know that
\[
\text{p-lim inf}_{n \to \infty} \frac{1}{n} \log \frac{P_{Y^n|Z^n}(Y^n|Z^n)}{P_{Y^n|Z^n}(Y^n|Z^n)} \geq -\gamma.
\]
(85)

Since \( \gamma > 0 \) is arbitrary, we may take \( \gamma \downarrow 0 \) to assert that the final term in (81) is non-negative and hence
\[
\text{p-lim sup}_{n \to \infty} \nu_{n}(X^n; Y^n|Z^n) \leq \text{p-lim sup}_{n \to \infty} \frac{1}{n} \log \frac{W^n(Y^n|X^n, Z^n)}{P_{Y^n|Z^n}(Y^n|Z^n)}.
\]
(86)

Now let \( X^n = (X_1^n, \ldots, X_n^n) \), \( Y^n = (Y_1^n, \ldots, Y_n^n) \) and \( Z^n = (Z_1^n, \ldots, Z_n^n) \) for each blocklength \( n \in \mathbb{N} \). Since the channel \( W^n \) and the conditional capacity-achieving output measure \( P_{Y^n|Z^n} \) are memoryless,
\[
\text{p-lim sup}_{n \to \infty} \nu_{n}(X^n; Y^n|Z^n) \leq \text{p-lim sup}_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \log \frac{W_{Y|Z}(Y_i^n|X_i^n, Z_i^n)}{P_{Y|Z}(Y_i^n|Z_i^n)}.
\]
(87)
Now we establish that for every $x \in \mathcal{X}$,
\[
\mathbb{E} \left[ \log \frac{W_{Y|Z}(Y|x, Z)}{P_{Y|Z}(Y|Z)} \right] \leq C_{s}^{\text{DM}}(W) \tag{88}
\]
where $(Y, Z)\{X = x\} \sim W(\cdot, \cdot | x)$. In fact, this bound was already mentioned in Yasui et al. [47, Lem. 1] but we present the details for completeness. We have
\[
\mathbb{E} \left[ \log \frac{W_{Y|Z}(Y|x, Z)}{P_{Y|Z}(Y|Z)} \right] = \mathbb{E} \left[ \log \frac{W(Y, Z|x)}{P_{Y|Z}(Y|Z)} \right] - \mathbb{E} \left[ \log \frac{W(Z|x)}{P_{Z}(Z)} \right] = D(W(\cdot, \cdot | x)\| P_{Y|Z}) - D(W_{Z}(\cdot|x)\| P_{Z}). \tag{89}
\]
In the first equality above, we introduced any capacity-achieving $P_{Z}$, i.e., the $Z$-marginal of the joint distribution in (79). Note that $P_{X} \rightarrow I(X; Y|Z) = I(X; YZ) - I(X; Z)$ is concave function (e.g., van Dijk [29] or Khisti et al. [48, App. A]). Thus, by applying the Karush-Kuhn-Tucker (KKT) conditions to the optimization problem in the definition of $C_{s}^{\text{DM}}(W)$ in (43), and straightforward differentiation of the mutual information functionals $P_{X} \rightarrow I(X; YZ)$ and $P_{X} \rightarrow I(X; Z)$ (see Gallager’s book for example [30, Thms. 4.4.1 and 4.5.1]), we know that for any $x \in \mathcal{X}$,
\[
D(W(\cdot, \cdot | x)\| P_{Y|Z}) - D(W_{Z}(\cdot|x)\| P_{Z}) \leq C_{s}^{\text{DM}}(W). \tag{90}
\]
Uniting (89) and (90), we deduce that (88) is true. Note that we used the fact that $W$ is degraded to establish (88) since $X - Y - Z$. From (88), for every $x = (x_{1}, \ldots, x_{n}) \in \mathcal{X}^{n}$, we have
\[
\mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{n} \log \frac{W_{Y|Z}(Y^{(n)}_{i}|x_{i}, Z^{(n)}_{i})}{P_{Y|Z}(Y^{(n)}_{i}|Z^{(n)}_{i})} \right] \leq C_{s}^{\text{DM}}(W). \tag{91}
\]
Because we fixed a deterministic $x \in \mathcal{X}^{n}$ and the channel is memoryless, the component random variables $(Y^{(n)}_{i}, Z^{(n)}_{i}), i = 1, \ldots, n$ in the sum in (91) are independent under the conditional probability distribution $W^{n}(\cdot, \cdot | x)$. By memorylessness and Chebyshev’s inequality, for every $\gamma > 0$,
\[
\Pr \left( \frac{1}{n} \sum_{i=1}^{n} \log \frac{W_{Y|Z}(Y^{(n)}_{i}|x_{i}, Z^{(n)}_{i})}{P_{Y|Z}(Y^{(n)}_{i}|Z^{(n)}_{i})} \geq C_{s}^{\text{DM}}(W) + \gamma \bigg| X^{n} = x \right) \leq \frac{\sigma_{0}^{2}}{n \gamma^{2}}, \tag{92}
\]
where the constant $\sigma_{0}^{2}$ is defined as
\[
\sigma_{0}^{2} := \max_{x \in \mathcal{X}} \text{Var} \left[ \log \frac{W_{Y|Z}(Y|x, Z)}{P_{Y|Z}(Y|Z)} \right]. \tag{93}
\]
The constant $\sigma_{0}^{2}$ is finite because $P_{Y|Z}(y|z)$ is positive for all $(y, z)$ in view of (88) and the finiteness of $C_{s}^{\text{DM}}(W) \leq \min\{ \log |\mathcal{X}|, \log |\mathcal{Y}| \}$. Since (92) is true uniformly over every $x \in \mathcal{X}^{n}$, we may average it over $x$ to obtain
\[
\Pr \left( \frac{1}{n} \sum_{i=1}^{n} \log \frac{W_{Y|Z}(Y^{(n)}_{i}|x^{(n)}_{i}, Z^{(n)}_{i})}{P_{Y|Z}(Y^{(n)}_{i}|Z^{(n)}_{i})} \geq C_{s}^{\text{DM}}(W) + \gamma \bigg| X^{n} \right) \leq \frac{\sigma_{0}^{2}}{n \gamma^{2}}. \tag{94}
\]
The upper bound $\sigma_{0}^{2}/(n \gamma^{2})$ clearly tends to zero as $n$ tends to infinity. From the definition of $\text{p-lim sup}$, we have
\[
\text{p-lim sup}_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \log \frac{W_{Y|Z}(Y^{(n)}_{i}|x^{(n)}_{i}, Z^{(n)}_{i})}{P_{Y|Z}(Y^{(n)}_{i}|Z^{(n)}_{i})} \leq C_{s}^{\text{DM}}(W) + \gamma. \tag{95}
\]
Consequently, from (86), this proves that
\[
\sup_{X} \mathcal{T}(X; Y|Z) \leq C_{s}^{\text{DM}}(W) + \gamma. \tag{96}
\]
Since $\gamma$ is arbitrary, we may take $\gamma \downarrow 0$. That is, we have proved the claim in (76), completing the proof of the partial strong converse for degraded DM-WTCs.
C. Proof of Theorem 4

Proof: For the first statement (i.e., that in (48)), it suffices to establish

\[
C^{(2)}(W) \geq \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} C(W_{Y,i}) - \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} C(W_{Z,i})
\]

as well as

\[
\overline{C}^{(6)}(W) \leq \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} C_{DM}^{(6)}(W_i).
\]

Indeed, if the right-hand-sides of (97) and (98) are equal, so are \(C^{(2)}(W)\) and \(\overline{C}^{(6)}(W)\) which implies that \(C^{(i)}(W) = \overline{C}^{(i)}(W)\) for all \(i \in [2 : 6]\), i.e., the partial strong converse holds for secrecy metrics \(S_i, i \in [2 : 6]\).

For inequality (97), we follow the steps in the proof of [49, Cor. 3] for the non-stationary Gel’fand-Pinsker channel. Particularize the optimization over \(V - X - (Y, Z)\) to \(V = X\) being uniform on \(\mathcal{X}^n\) for every \(n \in \mathbb{N}\). Invoking Theorem 1, we then find

\[
C^{(2)}(W) \geq \text{p-lim inf}_{n \to \infty} \frac{1}{n} \log \frac{W^n(Y^n|X^n)}{P^n(Y^n)} - \text{p-lim sup}_{n \to \infty} \frac{1}{n} \log \frac{W^n(Z^n|X^n)}{P^n(Z^n)}
\]

\[
= \text{p-lim inf}_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \log \frac{W_{Y,i}(Y_i|X_i)}{P_{Y,i}(Y_i)} - \text{p-lim sup}_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \log \frac{W_{Z,i}(Z_i|X_i)}{P_{Z,i}(Z_i)}
\]

\[
= \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} C(W_{Y,i}) - \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} C(W_{Z,i})
\]

where (100) follows from memorylessness and (101) follows from Chebyshev’s inequality and the fact that the alphabets are finite. See [16, Eq. (3.2.15)] for a similar statement.

Now we prove inequality (98). By using (88), we know that for every \(x \in \mathcal{X}^n\),

\[
\mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{n} \log \frac{W_{Y_iZ,i}(Y_i^{(n)}|x_i, Z_i^{(n)})}{P_{Y_iZ,i}(Y_i^{(n)}|Z_i^{(n)})} \right] \leq \frac{1}{n} \sum_{i=1}^{n} C_{DM}^{(6)}(W_i),
\]

where \(W_{Y_iZ,i} : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}\) and \(P_{Y_iZ,i} : \mathcal{Z} \to \mathcal{Y}\) are induced by \(W_i : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}\). Note that we leveraged on the degradedness of the channels \(\{W_i\}_{i \in \mathbb{N}}\) to arrive at (102). Define

\[
C^4 := \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} C_{DM}^{(6)}(W_i).
\]

By the definition of \(\limsup\), for every \(\gamma > 0\), there exists an integer \(N_{\gamma}\) such that for all \(n > N_{\gamma}\), we have

\[
\frac{1}{n} \sum_{i=1}^{n} C_{DM}^{(6)}(W_i) \leq C^4 + \gamma.
\]

Uniting (102) and (104), we obtain

\[
\mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{n} \log \frac{W_{Y_iZ,i}(Y_i^{(n)}|x_i, Z_i^{(n)})}{P_{Y_iZ,i}(Y_i^{(n)}|Z_i^{(n)})} \right] \leq C^4 + \gamma,
\]

for all \(n > N_{\gamma}\). Let \(\sigma_0^2\), analogously to (93), be defined as

\[
\sigma_0^2 := \sup_{x \in \mathcal{X}} \text{Var}_{i \in \mathbb{N}} \left[ \log \frac{W_{Y|Z,i}(Y|x, Z)}{P_{Y|Z,i}(Y|Z)} \right].
\]

By the symmetry of the channels, \(P_{Y|Z,i}(\cdot|z)\) is uniform on \(\mathcal{Y}\) for each \(z\) so each variance in (106) depends only on \(|\mathcal{Y}|\). Thus, \(\sigma_0^2\) also depends only on \(|\mathcal{Y}|\) and is finite. By Chebyshev’s inequality and (105) (the same logic that led to (92)), we have

\[
\text{Pr} \left( \frac{1}{n} \sum_{i=1}^{n} \log \frac{W_{Y_iZ,i}(Y_i^{(n)}|X_i^{(n)}, Z_i^{(n)})}{P_{Y_iZ,i}(Y_i^{(n)}|Z_i^{(n)})} \geq C^4 + 2\gamma \bigg| X^n = x \right) \leq \frac{\sigma_0^2}{n\gamma^2}.
\]


for all \( n > N_\gamma \) and all \( x \in \mathcal{X}^n \). It is also true that

\[
\Pr \left( \frac{1}{n} \sum_{i=1}^{n} \log \frac{W_{Y_i,Z_i}(Y_i^{(n)}|X_i^{(n)},Z_i^{(n)})}{P_{Y_i,Z_i}(Y_i^{(n)}|Z_i^{(n)})} \geq C^\dagger + 2\gamma \right) \leq \sigma_n^2 \frac{n}{\gamma^2}, \tag{108}\]

holds for all \( n > N_\gamma \). By the definition of \( \text{p-lim sup} \),

\[
\text{p-lim sup}_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \log \frac{W_{Y_i,Z_i}(Y_i^{(n)}|X_i^{(n)},Z_i^{(n)})}{P_{Y_i,Z_i}(Y_i^{(n)}|Z_i^{(n)})} \leq C^\dagger + 2\gamma. \tag{109}\]

Finally, from (86), we have

\[
\bar{C}^{(6)}(\mathbf{W}) \leq \sup_{\mathbf{X}} \bar{T}(\mathbf{X}; \mathbf{Y}|\mathbf{Z}) \leq C^\dagger + 2\gamma. \tag{110}\]

Since this holds for all \( \gamma > 0 \), we may take \( \gamma \downarrow 0 \) to complete the proof of (98).

For the second statement of the theorem, we first show that (48) implies that both \( a_n := \frac{1}{n} \sum_{i=1}^{n} C(W_{Y_i,i}) \) and \( b_n := \frac{1}{n} \sum_{i=1}^{n} C(W_{Z,i}) \) must converge. Note from (47) that \( a_n - b_n = \frac{1}{n} \sum_{i=1}^{n} C_{DM}(W_i) \). Suppose, to the contrary, that at least one of \( a_n \) and \( b_n \) does not converge. For the sake of concreteness, let us assume that

\[
\liminf_{n \to \infty} a_n < \limsup_{n \to \infty} b_n. \tag{111}\]

which contradicts (48) so \( a_n \) must converge. The same argument goes through \textit{mutatis mutandis} if \( b_n \) does not converge. It is evident that if both \( a_n \) and \( b_n \) converge, the \( \liminf \)‘s and \( \limsup \)’s in (48) are limits and equality there holds.

For the third statement, we verify that \( q_n \) in (66) converges to zero exponentially fast. Choose \( \mathbf{X} = \{X^n\}_{n \in \mathbb{N}} \) such that \( X^n \) is uniform on \( \mathcal{X}^n \) for each \( n \). In addition, if we choose \( M_n \) in (66) to be the smallest integer exceeding \( \exp[n(L(\mathbf{X}; \mathbf{Y}) - 2\gamma)] \), and \( M_n \) to be the largest integer smaller than \( \exp[n(L(\mathbf{X}; \mathbf{Y}) - 7(\mathbf{X}; \mathbf{Z}) - 4\gamma)] \), we have

\[
q_n \leq \Pr \left( \frac{1}{n} \log \frac{W_{Z}(Z^n|X^n)}{P_{Z^n}(Z^n)} \geq \bar{T}(\mathbf{X}; \mathbf{Z}) + \gamma \right). \tag{112}\]

Furthermore, by the same argument that led to (101), we notice that with \( X^n \) uniform on \( \mathcal{X}^n \),

\[
\bar{T}(\mathbf{X}; \mathbf{Z}) = \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} C(W_{Z,i}). \tag{113}\]

Thus, for every \( \gamma > 0 \), there exists an integer \( N'_\gamma \) such that for all \( n > N'_\gamma \),

\[
\bar{T}(\mathbf{X}; \mathbf{Z}) + \frac{\gamma}{2} \geq \frac{1}{n} \sum_{i=1}^{n} C(W_{Z,i}). \tag{114}\]

Uniting (112) and (114) and invoking the memorylessness of \( W^n_{Z} \), we have

\[
q_n \leq \Pr \left( \frac{1}{n} \sum_{i=1}^{n} \left( \log \frac{W_{Z,i}(Z_i|X_i)}{P_{Z,i}(Z_i)} - C(W_{Z,i}) \right) \geq \frac{\gamma}{2} \right). \tag{115}\]

Now, we argue that the random variable \( \log W_{Z,i}(Z_i|X_i) - \log P_{Z,i}(Z_i) \) is uniformly bounded in \( i \) by the assumptions of symmetry and (49). First, because the column sums of the transition matrices \( W_{Z,i} \) are independent of \( z \) and the capacity-achieving input distribution is uniform, \( P_{Z,i} \) is uniform on \( Z \). Second by (49), \( W_{Z,i}(z|x) \geq \kappa_z \) for \( (x,z) \in \mathcal{X} \times \mathcal{Z} \) and \( i \in \mathbb{N} \). Thus, with probability one, for all \( i \in \mathbb{N} \),

\[
\log (\kappa_{z}|Z|) \leq \log \frac{W_{Z,i}(Z_i|X_i)}{P_{Z,i}(Z_i)} \leq \log |Z|. \tag{116}\]

The exponential decay of \( q_n \) immediately follows from Hoeffding’s inequality for bounded and independent (but not necessarily identically distributed) random variables [32, Thm. 2].
D. Proof of Theorem 5

Proof: Similarly to the proof of Theorem 3, we show that \( C^{(1)}(W) \geq C^G_s(W; S) \) and \( C^{(6)}(W) \leq C^G_s(W; S) \). Note, however, that the form of the optimistic capacity \( C^{(6)}(W) \) in (44) has to be modified to take into account the cost constraint \( \Pr(X^n \in \mathcal{F}_n) = 1 \) in (55). The optimization over the chain \( V - X - (Y, Z) \) has to be further constrained to all distributions \( P_{VX} \) satisfying \( X^n \in \mathcal{F}_n \) for all \( n \in \mathbb{N} \).

For the lower bound, \( C^{(1)}(W) \geq C^G_s(W; S) \), we need to show that \( q_n, \) defined in (66), decays exponentially fast for an appropriate choice of input distribution. This argument is adapted from the proofs of Lemmas 2 and 5 in He and Yener [50]. Fix a constant \( \delta > 0 \) and define the product distribution

\[
P_{\tilde{X}^n}(\mathbf{x}) := \prod_{i=1}^{n} \mathcal{N}(x_i; 0, S - \delta).
\]  

(117)

Now define the input distribution to be

\[
P_{X^n}(\mathbf{x}) := \frac{P_{\tilde{X}^n}(\mathbf{x})}{\mu_n} 1\{\mathbf{x} \in \mathcal{F}_n\}
\]

(118)

where \( \mu_n \) is the normalizing constant that ensures that \( \int P_{X^n}(\mathbf{x}) \, d\mathbf{x} = 1 \). This is simply a truncated version of the jointly Gaussian distribution \( P_{\tilde{X}^n} \) in (117). Because of the constant backoff \( \delta > 0 \) from the permissible power \( S \) in (117), it can be seen from Cramer’s large deviations theorem that \( \mu_n := P_{\tilde{X}^n}(\mathcal{F}_n) \) tends to 1 exponentially fast, i.e.,

\[
\mu_n = \Pr \left( \frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i^2 \leq S \right) \geq 1 - \exp(-n\eta_1)
\]

(119)

for some \( \eta_1 > 0 \) depending on \( \delta \). By the construction of the input distribution in (118), \( X^n \in \mathcal{F}_n \) with probability one, satisfying the almost sure power constraint in (55). Using (3), the variation distance between \( P_{\tilde{X}^n} \) and \( P_{X^n} \) can be estimated as

\[
\mathbb{V}(P_{\tilde{X}^n}, P_{X^n}) = \frac{1}{2} \int_{\mathbb{R}^n} |P_{\tilde{X}^n}(\mathbf{x}) - P_{X^n}(\mathbf{x})| \, d\mathbf{x}
\]

(120)

\[
= \frac{1}{2} \int_{\mathcal{F}_n} |P_{\tilde{X}^n}(\mathbf{x}) - P_{X^n}(\mathbf{x})| \, d\mathbf{x} + \frac{1}{2} \int_{\mathcal{F}_n^c} |P_{\tilde{X}^n}(\mathbf{x}) - P_{X^n}(\mathbf{x})| \, d\mathbf{x}
\]

(121)

\[
= \frac{1}{2} \int_{\mathcal{F}_n} P_{X^n}(\mathbf{x})|\mu_n - 1| \, d\mathbf{x} + \frac{1}{2} P_{\tilde{X}^n}(\mathcal{F}_n^c)
\]

(122)

\[
\leq \frac{1}{2} \exp(-n\eta_1) + \frac{1}{2} \exp(-n\eta_1) = \exp(-n\eta_1)
\]

(123)

where (122) follows from the definition of \( P_{X^n}(\mathbf{x}) \), and (123) follows from (119). Consequently,

\[
\mathbb{V}(P_{\tilde{X}^n} \times W^n_z, P_{X^n} \times W^n_z) = \mathbb{V}(P_{\tilde{X}^n}, P_{X^n}) \leq \exp(-n\eta_1).
\]

(124)

Let \( (\tilde{X}, \tilde{Z}) \sim P_{\tilde{X}^n} \times W_z, (\tilde{X}^n, \tilde{Z}^n) \sim P_{\tilde{X}^n} \times W^n_z \) and \( (X^n, Z^n) \sim P_{X^n} \times W^n_z \). By using the characterization of the variation distance in (2) as well as the bound in (124), we deduce that for any \( \beta \in \mathbb{R} \),

\[
\left| \Pr \left( \frac{1}{n} \log \frac{W^n_z(Z^n|X^n)}{P^n_z(Z^n)} \geq \beta \right) - \Pr \left( \frac{1}{n} \log \frac{W^n_z(Z^n|\tilde{X}^n)}{P^n_z(Z^n)} \geq \beta \right) \right| \leq \exp(-n\eta_1).
\]

(125)

Define \( \alpha := \frac{1}{n} \log(\tilde{M}_n/M_n) - \gamma \). Let \( \eta_2 > 0 \) be an arbitrary constant for now. The probability \( q_n \) in (66) can be
written and bounded as

\[ q_n = \Pr \left( \frac{1}{n} \log \frac{W(Z|X^n)}{P_z(Z^n)} \geq \alpha \right) \]

(126)

\[ = \Pr \left( \frac{1}{n} \log \frac{W(Z|X^n)}{P_z(Z^n)} - \frac{1}{n} \log P_z(Z^n) \geq \alpha \right) \]

(127)

\[ \leq \Pr \left( \frac{1}{n} \log \frac{W(Z|X^n)}{P_z(Z^n)} - \frac{1}{n} \log P_z(Z^n) \geq \alpha \right) \]

(128)

\[ + \Pr \left( \frac{1}{n} \log P_z(Z^n) \leq -\eta_2 \right) \]

(129)

\[ \leq \Pr \left( \frac{1}{n} \log \frac{W(Z|X^n)}{P_z(Z^n)} \geq \alpha - \eta_2 \right) + \exp(-n\eta_2) \]

(130)

\[ \leq \Pr \left( \frac{1}{n} \log \frac{W(Z|X^n)}{P_z(Z^n)} \geq \alpha - \eta_2 \right) + \exp(-n\eta_2) + \exp(-n\eta_1) \]

(131)

\[ = \Pr \left( \frac{1}{n} \sum_{i=1}^{n} \log \frac{W(Z_i|X_i)}{P_z(Z_i)} \geq \alpha - \eta_2 \right) + \exp(-n\eta_2) + \exp(-n\eta_1) \]

(132)

where (130) uses the bound in (125) with the identification \( \beta = \alpha - \eta_2 \).

Choose \( M_n \) to be the smallest integer exceeding \( \exp[n(\frac{1}{2} \log(1+S/\sigma_1^2)) - 2\gamma] \). It can be shown using a change of output measure argument (cf. proof of direct part of [16, Thm. 3.6.2]) that \( q_n \) as the input distribution in (118) and with \( \delta \) set to \( \gamma/2 \), the decoding error probability \( 1 - (P_{X^n} \times W(Z^n))(T^{(n)}_\gamma) \) tends to zero, where \( T^{(n)}_\gamma \) is the set defined in (58) with \( V^n = X^n \). Choose \( M_n \) to be the largest integer smaller than \( \exp[n(\frac{1}{2} \log(1+S/\sigma_1^2) - \frac{1}{2} \log(1+S/\sigma_2^2) - 4\gamma)] = \exp[n(C_s^G(W;S) - 4\gamma)] \) and \( \eta_2 = \gamma/2 \). Thus, \( \alpha - \eta_2 \geq \frac{1}{2} \log(1+S/\sigma_2^2) + \gamma/2 \). With these choices,

\[ \mathbb{E} \left[ \log \frac{W(Z|X)}{P_z(Z)} \right] = \frac{1}{2} \log \left( 1 + \frac{S - \gamma/2}{\sigma_2^2} \right), \]

(133)

and from (131),

\[ q_n \leq \Pr \left( \frac{1}{n} \sum_{i=1}^{n} \log \frac{W(Z_i|X_i)}{P_z(Z_i)} \geq \frac{1}{2} \log \left( 1 + \frac{S}{\sigma_2^2} + \frac{\gamma}{2} \right) + \exp(-n\eta_2) + \exp(-n\eta_1) \right). \]

(134)

By the Chernoff bound [5, Lem. 6], the upper bound in (133) tends to zero exponentially fast. Thus, \( q_n \) tends to zero exponentially fast, proving the lower bound \( C^{(1)}(W) \geq C_s^G(W;S) - 4\gamma \). Now take \( \gamma \downarrow 0 \) to complete the proof.

For the upper bound, \( \overline{C}^{(6)}(W) \leq C_s^G(W;S) \), we emulate the proof of Theorem 3 with the (now) unique capacity-achieving output distribution \( P_{YZ} \) being

\[ P_{YZ}(y,z) = N(y;0,S + \sigma_2^2)N(z;0,\sigma_2^2 - \sigma_1^2). \]

(135)

The derivation up to (87) holds verbatim. So we simply have to check the condition in (91) (with \( C_s^G(W;S) \) in place of \( C_s^{DM}(W) \)) and the behavior of the variance corresponding to (93). We first fix an arbitrary sequence \( x = (x_1, \ldots, x_n) \in \mathcal{F}_n \) and study the first two moments of the following information density random variable

\[ J_n(x) := \frac{1}{n} \sum_{i=1}^{n} \log \frac{W(Y_{i|X_i}|X_i, Z_{i|X_i})}{P_{Y_i|X_i}(Y_{i|X_i}|Z_{i|X_i})}. \]

(136)

We would like to show that \( \mathbb{E}[J_n(x)] \leq C_s^G(W;S) \) and that the variance of \( J_n(x) \) is \( O(n^{-1}) \) uniform on \( \mathcal{F}_n \). For this task, let \( N_j^n \) (for \( j = 1, 2 \)) be a sequence of i.i.d. zero-mean Gaussian random variables with variance \( \sigma_j^2 \). Using the constraint \( \|x\|_2^2 \leq nS \), the fact that \( \sigma_2 > \sigma_1 \) (as assumed), and the form of the output distributions in (134), it
can easily be seen that $J_n(x)$ can be upper bounded as

$$J_n(x) \leq C_s^G(W; S) + \frac{\log e}{2(1 + \frac{S}{\sigma})} \left( \frac{S}{\sigma_1} \left( \frac{1 - \|N_1^n\|_2^2}{n\sigma_1^2} \right) + \frac{2\langle N_1^n, x \rangle}{n\sigma_1^2} \right)$$

$$- \frac{\log e}{2(1 + \frac{S}{\sigma})} \left( \frac{S}{\sigma_2} \left( \frac{1 - \|N_2^n\|_2^2}{n\sigma_2^2} \right) + \frac{2\langle N_2^n, x \rangle}{n\sigma_2^2} \right).$$

(136)

Since $N_1^n$ and $N_2^n$ have zero means and covariances $\sigma_1^2 \cdot I_{n \times n}$ and $\sigma_2^2 \cdot I_{n \times n}$ respectively, the expectation of $J_n(x)$ is bounded above by $C_s^G(W; S)$. The variance of $J_n(x)$ can be written and bounded as

$$\text{Var}[J_n(x)] = \text{Var} \left[ \frac{1}{n} \log \frac{W^n_p(Y^n|x)}{P^n_p(Y^n)} \right] \leq 2 \text{Var} \left[ \frac{1}{n} \log \frac{W^n_p(Y^n|x)}{P^n_p(Y^n)} \right] + 2 \text{Var} \left[ \frac{1}{n} \log \frac{W^n_p(Z^n|x)}{P^n_p(Z^n)} \right]$$

$$\leq 2 \log^2 e \left( \frac{9S^2}{4n(S + \sigma)} + \frac{\sigma^2 S}{n(S + \sigma_1)} \right) + 2 \log^2 e \left( \frac{9S^2}{4n(S + \sigma_2)} + \frac{\sigma^2 S}{n(S + \sigma_2)} \right)$$

(139)

where (139) follows from direct calculations per [16, Eq. (3.7.24)] and the fact that $x \in F_n$. We conclude that uniform over all $x \in F_n$, the variance of $J_n(x)$ is of the order $O(n^{-1})$ (depending only on $S, \sigma_1^2, \sigma_2^2$) and hence the Chebyshev argument at the end of the proof of Theorem 3 holds, yielding $\overline{C}^{(6)}(W) \leq C_s^G(W; S)$ as desired.

Acknowledgements: The authors would like to acknowledge several helpful discussions with Mark M. Wilde, and especially for bringing our attention to [9] and [10]. Vincent Tan’s work is supported by NUS startup grants WBS R-263-000-A98-750 (FoE) and WBS R-263-000-A98-133 (ODPRT). Matthieu Bloch’s work is supported by NSF with grant CCF1320298, and by ANR with grant 13-BS03-0008.

REFERENCES

[1] M. Bloch and J. Barros. Physical Layer Security: From Information Theory to Security Engineering. Cambridge University Press, 2011.

[2] Y. Liang, H. V. Poor, and S. Shamai (Shitz). Information Theoretic Security, volume 5 of Foundations and Trends in Communications and Information Theory. Now Publishers Inc, 2009.

[3] A. D. Wyner. The wire-tap channel. The Bell System Technical Journal, 54:1355–1387, 1975.

[4] I. Csiszár and J. Körner. Broadcast channels with confidential messages. IEEE Trans. on Inform. Th., 24(3):339–348, 1978.

[5] M. Bloch and J. N. Laneman. Strong secrecy from channel resolvability. IEEE Trans. on Inform. Th., 59(12):8077–8098, 2013.

[6] H. Tyagi and P. Narayan. How many queries will resolve common randomness? IEEE Trans. on Inform. Th., 59(9):5363–5378, 2013.

[7] M. Hayashi, H. Tyagi, and S. Watanabe. Secret key agreement: General capacity and second-order asymptotics. In Proc. of Intl. Symp. on Inform. Th., 2014.

[8] H. Tyagi and S. Watanabe. A bound for multiparty secret key agreement and implications for a problem of secure computing. In Proc. of EUROCRYPT, 2014. arXiv:1404.5715 [cs.IT].

[9] C. Morgan and A. Winter. “Pretty strong” converse for the quantum capacity of degradable channels. IEEE Trans. on Inform. Th., 60(1):317–333, 2014.

[10] J. M. Renes and R. Renner. Noisy channel coding and privacy amplification and information reconciliation. IEEE Trans. on Inform. Th., 57(11):7377–7385, 2011.

[11] G. Smith. The private classical capacity with a symmetric side channel and its application to quantum cryptography. Physical Review A, 78(022306), 2008.

[12] C. A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. IEEE Trans. on Inform. Th., 45(4):1216–1227, 1999.

[13] J. Wolfowitz. The coding of messages subject to chance errors. Illinois J. Math., 1:591–606, 1957.

[14] S. Verdú and T. S. Han. A general formula for channel capacity. IEEE Trans. on Inform. Th., 40(4):1147–1157, 1994.

[15] T. S. Han and S. Verdú. Approximation theory of output statistics. IEEE Trans. on Inform. Th., 39(3):752–772, 1993.

[16] T. S. Han. Information-Spectrum Methods in Information Theory. Springer Berlin Heidelberg, Feb 2003.

[17] M. Hayashi. General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel. IEEE Trans. on Inform. Th., 52(4):1562–1575, Apr 2006.

[18] P.-N. Chen and F. Alajaji. Optimistic Shannon coding theorems for arbitrary single-user systems. IEEE Trans. on Inform. Th., 45(7):2623–2629, 1999.

[19] Y. Steinberg. New converses in the theory of identification via channels. IEEE Trans. on Inform. Th., 44(3):984–998, 1998.

[20] P.-N. Chen and F. Alajaji. Lecture Notes in Information Theory; Volume II. http://shannon.cm.nctu.edu.tw/itvol22004.pdf.
[21] S. Leung-Yan-Cheong and M. E. Hellman. The Gaussian wire-tap channel. *IEEE Trans. on Inform. Th.*, 24(4):451–456, 1975.
[22] U. Maurer and S. Wolf. Information-theoretic key agreement: From weak to strong secrecy for free. In *Proc. 19th Intl. Conf. Theory Appl. Cryptographic Tech.*, EUROCRYPT'00, pages 351–368. Springer-Verlag, 2000.
[23] I. Csiszár. Almost independence and secrecy capacity. *Problems of Information Transmission*, 32(1):40–47, 1996.
[24] M. Tomamichel and V. Y. F. Tan. Second-order coding rates for channels with state. *IEEE Trans. on Inform. Th.*, 2014. arXiv:1305.6789 [cs.IT].
[25] M. Hayashi and H. Nagaoka. General formulas for capacity of classical-quantum channels. *IEEE Trans. on Inform. Th.*, 49(7):1753–1768, 2003.
[26] D. Kobayashi, H. Yamamoto, and T. Ogawa. Secure multiplex coding attaining channel capacity in wiretap channels. *IEEE Trans. on Inform. Th.*, 59(12):8131–8143, 2013.
[27] M. Hayashi. Exponential decreasing rate of leaked information in universal random privacy amplification. *IEEE Trans. on Inform. Th.*, 57(6):3989–4001, 2011.
[28] S. Leung-Yan-Cheong. On a special class of wiretap channels. *IEEE Trans. on Inform. Th.*, 23(5):625–627, 1977.
[29] M. van Dijk. On a special class of broadcast channels with confidential messages. *IEEE Trans. on Inform. Th.*, 43(3):712–714, 1997.
[30] R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, 1968.
[31] T. S. Han, H. Endo, and M. Sasaki. Reliability and security functions of the wiretap channel under cost constraint. arXiv, 2013. arXiv:1307.0608 [cs.IT].
[32] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, Mar 1936.
[33] P.-N. Chen. Generalization of Gartner-Ellis theorem. *IEEE Trans. on Inform. Th.*, 46(7):2752–2760, 2000.
[34] A. Laourine and A. B. Wagner. The degraded Poisson wiretap channel. *IEEE Trans. on Inform. Th.*, 58(12):7073–7085, 2012.
[35] M. Tomamichel and V. Y. F. Tan. A tight upper bound for the third-order asymptotics of most discrete memoryless channels. *IEEE Trans. on Inform. Th.*, 59(11):7041–7051, 2013.
[36] S. Watanabe and M. Hayashi. Strong converse and second-order asymptotics of channel resolvability. In *Proc. of Intl. Symp. on Inform. Th.*, pages 818–825, 2008.