Algorithms and Architectures for Multiuser, Multi-terminal, Multi-layer Information Theoretic Security

by

Ashish Khisti

B.A.Sc., University of Toronto (2002)
S.M., Massachusetts Institute of Technology (2004)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2009

Author .. ..
Department of Electrical Engineering and Computer Science
09/15/2008

Certified by. .....
Gregory W. Wornell
Professor
Thesis Supervisor

Accepted by ....
Terry P. Orlando
Chairman, Department Committee on Graduate Theses

Algorithms and Architectures for Multiuser, Multi-terminal, Multi-layer Information Theoretic Security

by

Ashish Khisti

Submitted to the Department of Electrical Engineering and Computer Science
on 09/12/2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

As modern infrastructure systems become increasingly more complex, we are faced with many new challenges in the area of information security. In this thesis we examine some approaches to security based on ideas from information theory. The protocols considered in this thesis, build upon the "wiretap channel," a model for physical layer security proposed by A. Wyner in 1975. At a higher level, the protocols considered here can strengthen existing mechanisms for security by providing a new location based approach at the physical layer.

In the first part of this thesis, we extend the wiretap channel model to the case when there are multiple receivers, each experiencing a time varying fading channel. Both the scenario when each legitimate receiver wants a common message as well as the scenario when they all want separate messages are studied and capacity results are established in several special cases. When each receiver wants a separate independent message, an opportunistic scheme that transmits to the strongest user at each time, and uses Gaussian codebooks is shown to achieve the sum secrecy capacity in the limit of many users. When each receiver wants a common message, a lower bound to the capacity is provided, independent of the number of receivers.

In the second part of the thesis the role of multiple antennas for secure communication studied. We establish the secrecy capacity of the multi antenna wiretap channel (MIMOME channel), when the channel matrices of the legitimate receiver and eavesdropper are fixed and known to all the terminals. To establish the capacity, a new computable upper bound on the secrecy capacity of the wiretap channel is developed, which may be of independent interest. It is shown that Gaussian codebooks suffice to attain the capacity for this problem. For the case when the legitimate receiver has a single antenna (MISOME channel) a rank one transmission scheme is shown to attain the capacity. In the high signal-to-noise ratio (SNR) regime, it is shown that a capacity achieving scheme involves simultaneous diagonalization of the channel matrices using the generalized singular value decomposition and independently coding accross the resulting parallel channels. Furthermore a semi-blind masked beamforming scheme is studied, which transmits signal of interest in the subspace of

the legitimate receiver's channel and synthetic noise in the orthogonal subspace. It is shown that this scheme is nearly optimal in the high SNR regime for the MISOME case and the performance penalty for the MIMOME channel is evaluated in terms of the generalized singular values. The behavior of the secrecy capacity in the limit of many antennas is also studied. When the channel matrices have i.i.d. $\mathcal{C}N(0,1)$ entries, we show that (1) the secrecy capacity for the MISOME channel converges (almost surely) to zero if and only if the eavesdropper increases its antennas at a rate twice as fast as the sender (2) when a total of $T \gg 1$ antennas have to be allocated betweeen the sender and the receiver, the optimal allocation, which maximizes the number of eavesdropping antennas for zero secrecy capacity is $2:1$.

In the final part of the thesis, we consider a variation of the wiretap channel where the sender and legitimate receiver also have access to correlated source sequences. They use both the sources and the structure of the underlying channel to extract secret keys. We provide general upper and lower bounds on the secret key rate and establish the capacity for the reversely degraded case.

Thesis Supervisor: Gregory W. Wornell
Title: Professor

This thesis is dedicated to[1]*:*

The women who have shaped my life:

My Mother, My Sister and My Grandmother

[1]The author was pleasantly surprised to see a similar dedication in [42].

Acknowledgments

First and foremost, I thank God for his supreme kindness and generosity by blessing me with courage, talents and an upbringing that has taken me this far in my career. I thank my thesis advisor, Professor Greg Wornell, for accepting me into his group six years ago and supporting my endeavors throughout graduate school. Greg leads a very simulating research group at MIT that attracts highly talented students from all over the world. I feel extremely fortunate to have been a part of his group. Greg is in many ways a complete advisor. Not only did he provide astute advice in helping me choose research topics to work on, but also provided me with very valuable advice regarding careers beyond graduate school. I am simply amazed at his ability to convey the most intricate technical details in an extremely simple manner, a skill I experienced first hand while writing journal papers with him, and can only hope to cultivate these skills as I progress beyond graduate school.

I also thank my thesis readers — Lizhong Zheng and Uri Erez. Both have been my role models and have deeply influenced my way of thinking about research problems. Lizhong's quest for finding an intuition to explain every theorem, a principle I learned while taking courses and discussing research problems with him, has been my mantra for doing graduate research. Uri Erez, as a co-advisor of my master's thesis, had a strong influence on me during my formative years in graduate school.

During the course of my graduate school I had a chance to collaborate with many great individuals. In particular, I deeply enjoyed working with Aggelos Bletsas, Suhas Diggavi, and Amos Lapidoth for their never ending enthusiasm and high ethical standards that made these long-distance collaborations both fun and productive. In addition, Dave Forney, Vivek Goyal, Bob Gallager, Dina Katabi, Muriel Medard, Sanjoy Mitter, David Staelin, Devavrat Shah, Mitchell Trott, Emre Telatar, Moe Win, and Ram Zamir provided valuable advice during various stages of my graduate research. Besides research, I was presented an opportunity to TA several courses at MIT and enjoyed working closely with my course instructors — Dave Forney, Bob Gallager, Asuman Ozdaglar, Greg Wornell, and Lizhong Zheng.

I would like to express my deepest gratitude towards Tricia O'Donnell for being a great friend and an excellent administrative assistant of our group. I also thank Eric Strattman and Cindy Leblanc, whom I could freely approach when Tricia was not around. I also thank Janet Fischer from the EECS Graduate office and Danielle Ashbrook and Maria Brennan from the international student office for their crucial help during my stay at MIT.

An important element of my grad school experience was my interactions with several students both at MIT and elsewhere. It is a pleasure to thank them as I complete this thesis. In particular, I feel deeply privileged to have known Emmanuel Abbe, Anthony Accardi, Mukul Agarwal, Shashi Borade, Manish Bharadwaj, Albert Chan, Venkat Chandar, Todd Coleman, Carlos Coelho, Vijay Divi, Stark Draper, Sanket Dusad, Krishnan Eswaran, Carlos Gomez, James Geraci, Saikat Guha, Ying-Zong Huang, Everest Huang, Sheng Jing, Tobias Koch, Nicholas Laneman, Tie Liu, Yingbin (Grace) Liang, Desmond Lun, Emin Martinian, Dmitry Malioutov, Natalia Miliou, Baris Nakiboglu, Bobak Nazer, Urs Niesen, Vinod Prabhakaran, Etienne Perron,

Tal Philosof, Tony Quek, Sibi Raj, Sujay Sanghavi, Prasad Santhanam, Anand Sarwate, Charles Sestok, Maryam Shanechi, Anand Srinivas, Jay-kumar Sundararajan, Watcharapan Suwansantisuk, Charles Swannack, Aslan Tchamkerten, Ender Tekin, Elif Uysal, Lav Varshaney, Kush Varshany, Yonggang Wen, and Huan Yao, Chen-Pang Yeang and Murtaza Zafer.

Most importantly I would like to thank my family — my mother, my sister and my grandmother, and dedicate this thesis to them. Without their genuine love, never ending patience in understanding my situations (even when I lacked the same in explaining them), and sincere advice in balancing life and work, I would never have had the strength to spend six years in graduate school and complete this dissertation. They are not only my relatives, but in fact the best friends I have ever had in the three countries that I have lived in.

Contents

Chapter 1

Introduction

Traditional approaches to secure communication require that the legitimate parties share secret keys which are not available to adversaries. The physical-layer provides a reliable communication bit-pipe, while the encryption and decryption operations are performed at the higher layers. Thus there is a separation between layers that implement secure communication and reliable communication. In contrast, this thesis considers protocols that jointly perform both reliable and secure communication at the physical layer. See Fig. 1.

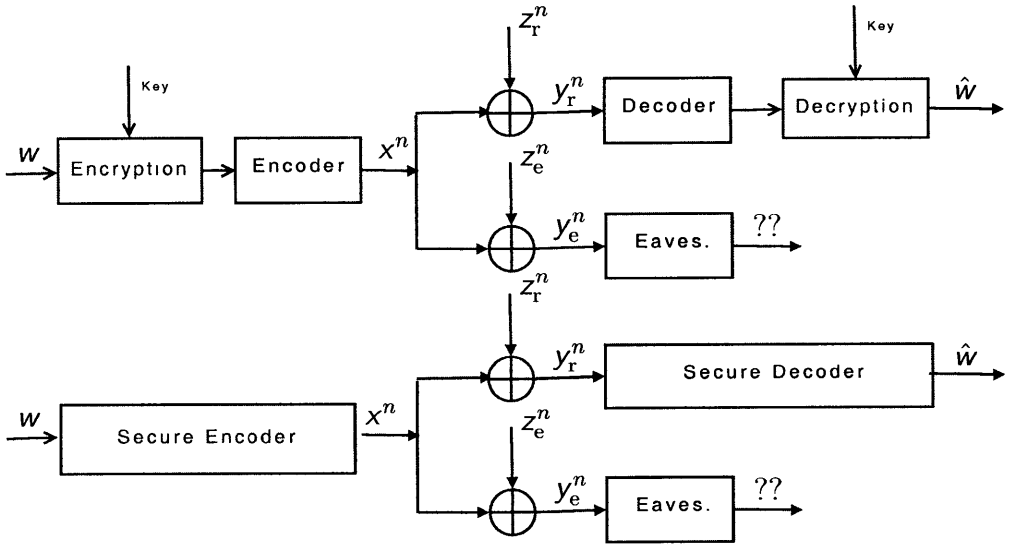


Figure 1-1: The top figure shows traditional approaches to secure communication. The figure below shows approaches using secure communication.

Our protocols are motivated by an information theoretic problem — the wiretap channel. This setup is described in Section 1.1. Our motivation for studying these protocols comes from the Pay-TV application.

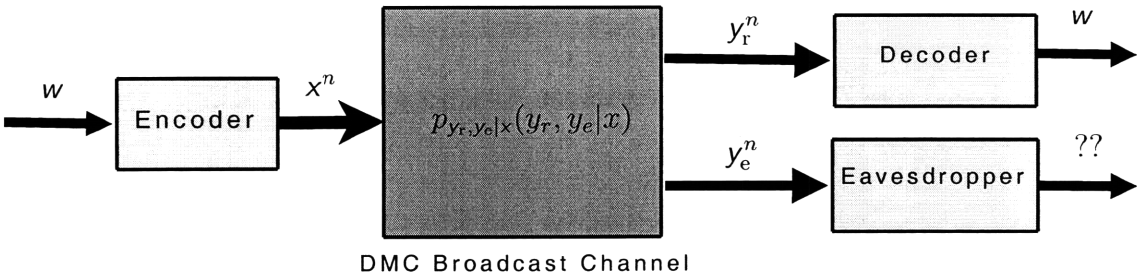


Figure 1-2: The wiretap channel model with one sender, one receiver and one eavesdropper. The (legitimate) receiver reliably decodes the message, while the eavesdroppers' channel produces a certain level of equivocation.

Example — Pay TV systems

A content provider wants to distribute programming content to a subset of receivers that have subscribed to the program. Traditional cryptographic techniques for this application suffer from piracy based attacks [16]. In these approaches, each user has a unique private key, which can be used for decryption of the programming content when it is subscribed. If any of the receivers' key gets leaked in public, all the users can use this key to decrypt the program subscribed by the user [5] resulting in serious revenue losses to the content provider.

This thesis develops a class of protocols that can provide a different approach — distribute secret keys online using physical layer techniques. We examine conditions on the physical channels under which such transmission is possible. In particular, we show how diversity techniques at the physical layer, which have been studied to improve reliability, can also enhance physical layer security.

1.1 Wiretap Channel

The wiretap cannel model, introduced by Wyner [53] is shown in Fig. 1-2.In this setup, there are three terminals — one sender, one receiver and one eavesdropper. As shown in Fig. 1-2, the sender has a message w, that it wishes to communicate reliably to the legitimate receiver, while keeping it secret from an eavesdropper. The communication link is a broadcast channel described by the transition probability the $p_{y_r,y_e|x}(\cdot)$ i.e., each channel use accepts an input symbol x and produces two output symbols — y_r at the legitimate receiver and y_e at the eavesdropper, according to this distribution. The alphabets of the input and output are denoted via $\mathcal{X}$, $\mathcal{Y}_r$ and $\mathcal{Y}_e$ respectively. The sender transmits a sequence x^n over n channel uses of and the legitimate receiver and the eavesdropper observe y_r^n and y_e^n according to the transition law,

$$\Pr(y_r^n, y_e^n \mid x^n) = \prod_{i=1}^{n} p_{y_r,y_e|x}(y_r, y_e|x). \tag{1.1}$$

A length n, rate R *wiretap code* consists of,

1. A set $\mathcal{W} \triangleq \{1, 2, \ldots, 2^{nR}\}$, and the message w uniformly distributed over this set.

2. An encoding function $f : \mathcal{W} \to \mathcal{X}^n$

3. A decoding function $g : \mathcal{Y}_r^n \to \mathcal{W}$.

A rate R, equivocation E is achieved by a wiretap code, if for some non-negative sequence ε_n that vanishes to zero, there exists a sequence of rate $R - \varepsilon_n$ codes such that $\Pr(e) \triangleq \Pr(g(y_r^n) \neq w) \to 0$ as $n \to \infty$ and

$$\frac{1}{n} H(w|y_e^n) \geq E - \varepsilon_n. \tag{1.2}$$

A sequence of rate R wiretap codes, that achieve an equivocation level of $E = R$ achieve (asymptotically) perfect secrecy. In this situation a negligible fraction of information bits are leaked to the eavesdropper. The supremum of all rates that can be achieved by perfect-secrecy wiretap codes is called the *secrecy capacity* of the wiretap channel.

1.1.1 Secrecy Notion

The notion of secrecy (1.2) is an information theoretic notion of security. It is interesting to compare this notion with other notions of secrecy. First, note that the cryptographic approaches use a computational notion of security. These approaches do not guarantee secrecy if the eavesdropper has sufficient computational power. In the information theoretic literature, the notion of perfect secrecy is first introduced by Shannon [46]. This notion requires that $H(w|y_e^n) = H(w)$ i.e., message be statistically independent of the observation sequence at the eavesdropper. Unfortunately this notion is too strong a notion in practice. Wyner's notion (1.2) is clearly a relaxation of this notion. A wiretap code that satisfies Wyner's notion guarantees that, asymptotically in n, the fraction of information that gets leaked to an eavesdropper is zero, but the number of information bits that get leaked can be arbitrarily large. Stronger versions of the secrecy notion are discussed in works by Maurer and Wolf [38] and Csisz'ar[7]. Another notion that measures the number of guesses required by an eavesdropper to learn the message is introduced in [39].

1.1.2 Rate-Equivocation Region

We summarize the main results of the wiretap channel in the literature.

For the discrete memoryless channel model, a single-letter expression for the trade-off between rate and equivocation is obtained in [8].

Fact 1 (I. Csiszár and J. Körner: [8]) *Let v and u be auxiliary random variables and the joint distribution $p_{vuxy_ry_e}(\cdot)$ satisfy $v \to u \to x \to (y_r, y_e)$. The rate-equivocation region is obtained by taking the convex hull of all the union of all rate*

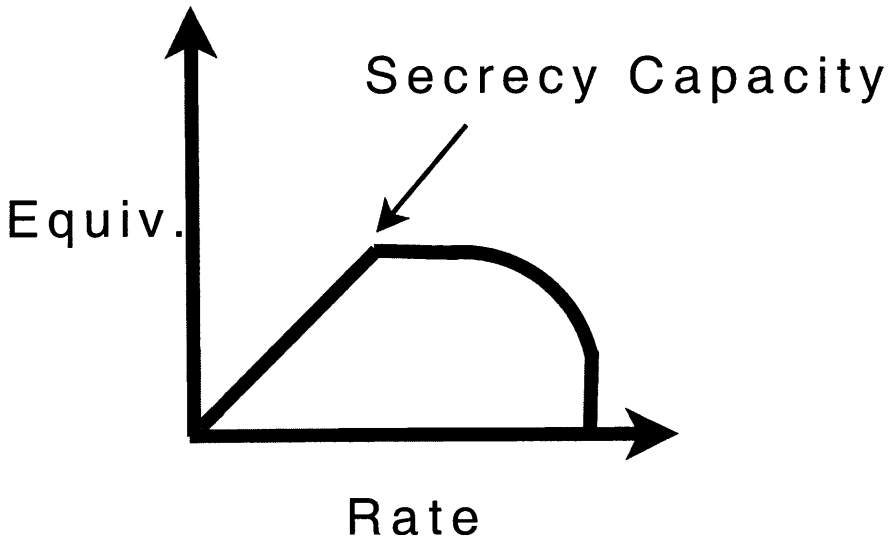


Figure 1-3: The structure of rate-equivocation tradeoff for the wiretap channel. For rates below the secrecy capacity, the maximum equivocation equals the transmission rate. For rates above the secrecy capacity, the equivocation does not increase.

pairs (R, R_{eq}) *over such joint distributions*

$$R_{\text{eq}} \leq I(u; y_{\text{r}}|v) - I(u; y_{\text{e}}|v) \tag{1.3}$$
$$R \leq I(u; y_{\text{r}}). \tag{1.4}$$

and it suffices to consider random variables u *and* v *with cardinality* $\mathcal{U} \leq |\mathcal{X}|^2 + 4|\mathcal{X}| + 3$ *and* $|\mathcal{V}| \leq |\mathcal{U}| + 1$.

A typical structure of the rate-equivocation region is provided in Fig. 1-3. Of particular interest on the rate equivocation region is the secrecy capacity.

Fact 2 (I. Csiszár and J. Körner: [8]) *The secrecy capacity of the discrete memoryles wiretap channel is*

$$C = \max_{p_u p_{x|u}} I(u; y_{\text{r}}) - I(u; y_{\text{e}}), \tag{1.5}$$

where the maximization is over random variables $u \rightarrow x \rightarrow (y_{\text{r}}, y_{\text{e}})$ *and* $|\mathcal{U}| \leq |\mathcal{X}|^2 + 4|\mathcal{X}| + 3$.

To establish these results, the authors provide an achievability scheme and a converse. We focus on achieving the perfect secrecy. The achievability scheme is based on a "random binning technique", a technique[1] used to establish capacity results in many multi-terminal source and channel coding problems.

1. Generate $\approx 2^{nI(u;y_{\text{r}})}$ codewords i.i.d. from a distribution $p_u(\cdot)$.

[1]A more structured coding approach is discussed later on in this chapter.

2. Partition the space of codewords by randomly assigning to the messages, so that there are $\approx 2^{nI(u;y_e)}$ codewords per message.

3. Given a message, select one of the candidate codewords uniformly at random, and transit x^n over the channel, generated from u^n according to $p_{x|u}(\cdot)$.

The legitimate receiver decodes the message with a vanishingly small error probability, by joint typical decoding. An eavesdropper upon observing y_e^n finds a typical codeword sequence in each bin and hence remains in (asymptotically) perfect equivocation.

The converse is established using a chain of mutual information inequalities reminiscent of the converse in many multiuser information theory problems such as channel coding with side information [17] and the broadcast channel with degraded message sets [26].

The bounds on the cardinality of alphabets are obtained via the Caratheodory's theorem.

When the underlying broadcast channel has a degraded structure i.e., $x \to y_r \to y_e$ holds, the secrecy capacity expression (5.6) has a simpler form. This result is due to Wyner [53].

$$C = \max_{p_x} I(x; y_r|y_e) \tag{1.6}$$

The achievability follows by setting $u = x$ in (5.6) and noting that since $x \to y_r \to y_e$, we have that

$$I(x; y_r) - I(x; y_e) = I(x; y_r|y_e).$$

For the converse, we need to show that it suffices to optimize (5.6) over input distributions with $u = x$.

$$\begin{aligned} I(u; y_r) - I(u; y_e) &= I(u; y_r, y_e) - I(u; y_e) && (1.7)\\ &= I(u; y_r|y_e) \\ &= H(y_r|y_e) - H(y_r|y_e, u) \\ &\le H(y_r|y_e) - H(y_r|y_e, u, x) \\ &= H(y_r|y_e) - H(y_r|y_e, x) && (1.8)\\ &= I(x; y_r|y_e) \end{aligned}$$

where both (1.7) and (1.8) follow from the Markov condition $u \to x \to y_r \to y_e$.

An explicit expression for the secrecy capacity of the Gaussian wiretap channel has been obtained by Leung-Yan-Cheong and M. Hellman in 1978 [27]. The authors consider a model

$$\begin{aligned} y_r &= x + z_r \\ y_e &= x + z_e, \end{aligned} \tag{1.9}$$

where $z_r \sim \mathcal{N}(0, N_r)$ and $z_e \sim \mathcal{N}(0, N_e)$ are additive white Gaussian noise random variables, with $N_e > N_r$ and the input sequence satisfies an average power constraint

$E[X^2] \leq P$. In this case, the secrecy capacity is

$$C(P, N_r, N_e) = \frac{1}{2}\log\left(1 + \frac{P}{N_r}\right) - \frac{1}{2}\log\left(1 + \frac{P}{N_e}\right). \tag{1.10}$$

The characterization of the secrecy capacity of the Gaussian wiretap channel does not specify the joint correlation between (z_r, z_e). In fact, it follows from the definition of the capacity that the secrecy capacity does not depend on the joint distribution of these variables. Only the marginal distributions matter.

The achievability in (1.10) follows by setting $u = x \sim \mathcal{N}(0, P)$. For the converse, it suffices to show that a Gaussian input achieves the capacity. In their paper, Hellman and Leung-Yan-Cheong [27] use entropy power inequality to establish this result. Nevertheless a simpler proof follows via (1.6), since the Gaussian wiretap channel is degraded. We can assume, without loss of generality that $z_e = z_r + \Delta z$, where $\Delta z \sim \mathcal{N}(0, N_e - N_r)$ is independent of z_r. Now note that

$$\begin{aligned}
I(x; y_r|y_e) &= h(y_r|y_e) - h(z_r|z_e) \\
&= h(y_r - \alpha y_e|y_e) - h(z_r|z_e) && (1.11) \\
&\leq h(y_r - \alpha y_e) - h(z_r|z_e) \\
&\leq \frac{1}{2}\log\left(\frac{(P+N_r)\Delta N}{P+N_e}\right) - \frac{1}{2}\log\left(\frac{N_r \Delta N}{N_e}\right) && (1.12) \\
&= \frac{1}{2}\log\left(1 + \frac{P}{N_r}\right) - \frac{1}{2}\log\left(1 + \frac{P}{N_e}\right)
\end{aligned}$$

where α in (1.11) is the linear minimum mean squared estimate coefficient in estimating y_r from y_e and $\Delta N \triangleq N_r - N_e$ in (1.12).

We now provide a few remarks about the wiretap channel.

1. What point should one operate on the rate-equivocation region? The secrecy capacity is a natural operating point, if the application involves transmission of secret keys. We will investigate a scenario of joint source channel coding where the operating point depends on the correlation of sources and capacity of the channels.

2. The secrecy capacity of the wiretap channel depends on the channel $p_{y_r, y_e|x}$. In practice the eavesdropper's channel is not known to the legitimate terminals. So the wiretap code could be designed for the worst case assumption on the eavesdropper's channel model.

3. For the Gaussian case, the secrecy capacity is zero if $N_r \geq N_e$. So the scheme is only applicable in situations where the eavesdropper is guaranteed to have a degraded channel. We will explore the use of diversity techniques in the physical layer of wireless systems to put an eavesdropper at a significant disadvantage.

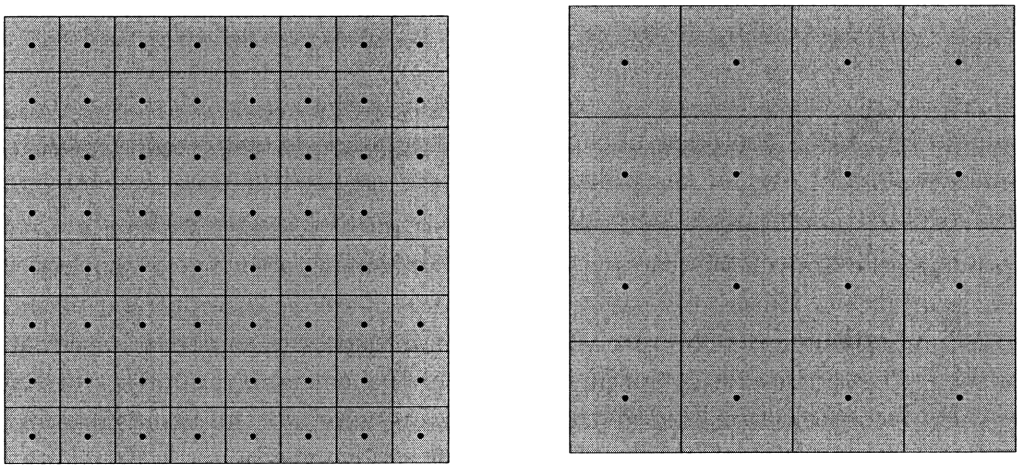
1.1.3 Code construction

The design of structured codes for the wiretap channel will not be explored in this thesis. In this section, we provide some insights into the design of structured codes by studying the uniform-additive-noise model in (1.9). We make the following assumptions

1. The sender uses quadrature-amplitude-modulation i.e., x^n is a sequence of n-QAM symbols. See Fig. 1-4.

2. The additive noise z_{r} is uniformly distributed on $[-1/2, 1/2] \times [-1/2, 1/2]$ and z_{e} is uniformly distributed on $[-1, 1] \times [-1, 1]$.

Fig. 1-4 shows QAM constellations on the legitimate receiver and eavesdropper's chan-

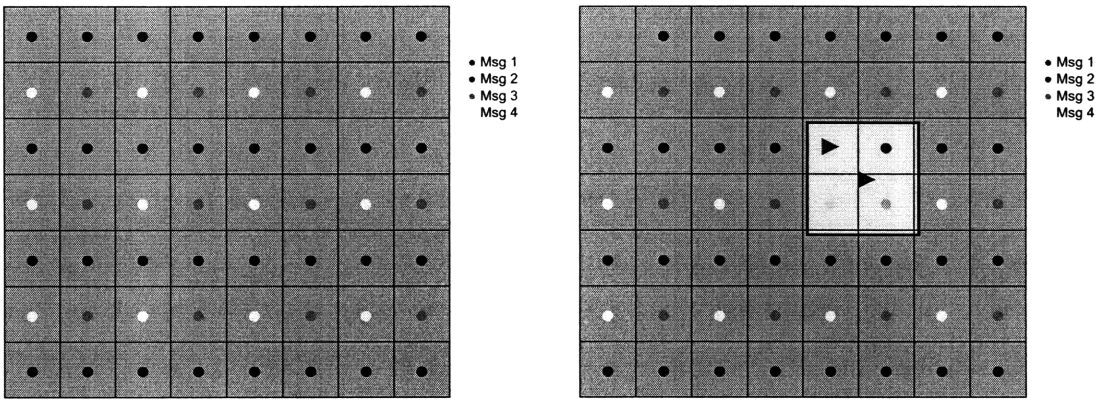


(a) Legitimate Receiver's Constellation (b) Eavesdropper's Constellation

Figure 1-4: Standard QAM constellation for legitimate receiver and eavesdropper's channel.

nel respectively. The minimum distance between two points is governed by the noise on the respective receiver's channel. Accordingly, the eavesdropper's constellation is a sparser 16-QAM while the legitimate receiver's constellation is 64-QAM.

Fig. 1-5 shows how one can transmit at a rate of 2 bits/symbol in a secure manner to the legitimate receiver, while keeping the eavesdropper in near perfect equivocation. Each of the four messages is represented by a separate color. There are 16 points assigned to each message, i.e., every fourth point is assigned to the same message. To transmit a message, the sender selects one of the sixteen points uniformly at random from the constellation. The legitimate receiver's channel perturbs the transmitted point in the smaller square and hence the receiver correctly identifies the transmitted point and declares the corresponding color to be the transmitted message. The eavesdropper's noise is sufficiently large to create a confusion as to which of the four messages is transmitted. As shown in Fig. 1-5(b), the eavesdropper upon receiving



(a) 16 candidate points for each of the 4 messages (b) The eavesdropper receives the black point and cannot decide on which color was selected.

Figure 1-5: Secure QAM constellation.

its observation, draws the noise uncertainty ball around this point and all four colors appear in this ball. Any of the points could have resulted in this received point, so the eavesdropper does not get any information regarding the message.

Note that the above construction does leak information to the eavesdropper — if one of the point on the boundary is transmitted, the eavesdropper can eliminate a subset of points whenever the received signal is outside the constellation area. This difficulty can be addressed by adding a (public) dither signal uniformly distributed on the QAM constellation and reducing the sum, modulo the constellation. This approach effectively "folds" the constellation to remove the boundary effects.

This approach of mapping multiple transmission points to a single message is known as binning. Higher dimensional version of binning techniques can be used to attain the capacity of a variety of wiretap channel models. Also note that the encoder used is a stochastic encoder — given the message the transmitted signal is chosen at random for a candidate set. This randomization is not known apriori to the receiver.

1.2 Secret Key Generation Using Correlated Sources

In this setup, two remote terminals A and B, observe a pair of correlated sources u^N, and v^N as shown in Fig. 1-6. The sources are sampled i.i.d. from a joint distribution $p_{u,v}(\cdot,\cdot)$. They also have access to a noiseless public channel of unlimited capacity. They can exchange any amount of information over this channel, but this communication happens in the clear and is observed by a wiretapper. The legitimate terminals distill a common key that needs to be concealed from the wiretapper who observes the public communications but does not have access to the correlated source sequence.

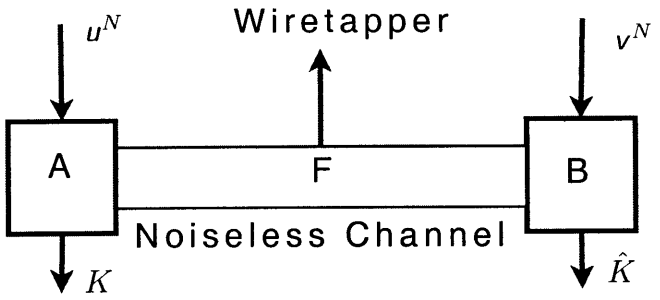


Figure 1-6: The problem of secret key generation using correlated sources. Terminals A and B observe a pair

1.2.1 One-Shot Secret Key Generation

In the *one-shot* protocol, terminal A sends $F = f(u^N)$ to terminal B and produces a key $K = K_A(u^N)$. Terminal B, upon receiving F, produces $\hat{K} = K_B(v^N, F)$. These functions must satisfy

1. $\Pr(e) = \Pr\left(K_A(u^N) \neq K_B(v^N, f(u^N))\right) \to 0$,
2. $\frac{1}{N} I\left(f(u^N); K_A(u^N)\right) \to 0$,

as $N \to \infty$. Here $I(u; v)$ denotes the mutual information function and is defined as $I(u; v) = H(u) - H(u|v)$. The quantity of interest is the secret key rate, defined as $\frac{1}{N} H(K)$ and the secrecy capacity is the supremum of all achievable secret key rates.

This setup was studied in [2] and [37] where the authors show that the secret key capacity equals $I(u; v)$. The main steps of their coding theorem are listed below.

1. Assign each sequence u^N to one of the $M \approx 2^{NH(u|v)}$ bins randomly. There are $\approx 2^{NI(u;v)}$ sequences in each bin.
2. All typical sequences u^N in a given bin are ordered. The secret key is the number assigned to a sequence in this bin.
3. Given a sequence u^N, find the bin index i it is assigned to and transmit $F = i$ over the public noiseless channel.
4. The receiver upon observing i, searches all typical sequences that are assigned to this bin index that are also typical with v^N. It recovers the sequence u^N with high probability and hence recovers the secret key.
5. The wiretapper upon observing F, knows the bin index, but no information about the secret key is revealed to the wiretapper.

The converse shows that the secret key rate cannot be higher than what is achieved by the one-shot protocol above.

Since the key is obtained from u^N and (v^N, F), we have that[2]

$$H(K|u^N) = 0, \quad H(K|v^N, F) = 0 \tag{1.13}$$

Now note that

$$\begin{aligned} NR &= H(K) = H(K|F) + I(K;F) \\ &= H(K|F) + No_N(1) && (1.14) \\ &= H(K|F, v^N) + I(K; v^N|F) + No_N(1) \\ &= I(K; v^N|F) + No_N(1) && (1.15) \\ &\leq I(K, F; v^N) + No_N(1) \\ &\leq I(u^N, K, F; v^N) + No_N(1) \\ &= I(u^N; v^N) + No_N(1) && (1.16) \\ &= N(I(u;v) + o_N(1)) \end{aligned}$$

where (1.14) follows from the definition of the secret key rate that $I(K;F) = No_N(1)$, and (1.15) follows from (1.13) and (1.16) follows from the fact that $(K, F) \to u^N \to v^N$ form a Markov chain.

Note that in this setup no constraint is imposed on the rate of transmission over the public channel. An extension to the case when the noiseless channel has a rate constraint is provided in [10].

The setup of secret key generation has found applications in diverse problems such as wireless channels with reciprocity constraints [52] as well as secure biometrics [14].

A natural scenario to further investigate is secret key generation in a joint source and channel coding setup which we will investigate in this thesis.

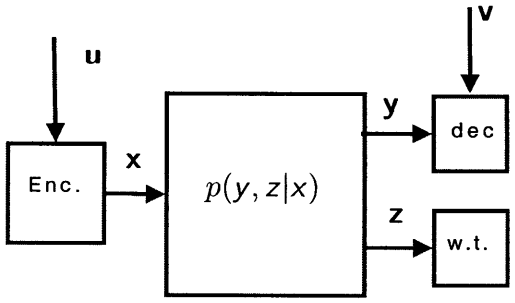


Figure 1-7: Secret-key-generation setup using correlated sources and channels.

[2]In our definition, a small error probability is allowed. In this case, we can use Fano's inequality to replace the 0 in the right hand side of (1.13) with a term that goes to zero as $N \to \infty$.

Chapter 2

Parallel Broadcast Channels

In this chapter, we study an extension of the basic wiretap channel. We study the case when there is one sender, multiple legitimate receivers, and one eavesdropper. We consider two scenarios:

- A common message needs to be delivered to all legitimate receivers
- An individual message needs to be delivered to each legitimate receiver

Further we restrict our attention to the case when there are multiple parallel and independent channels. Often we will assume that each of these channels is degraded in a certain order, but the overall system may not be degraded. Such channel models are referred to as reversely degraded broadcast channels [15]. See Fig. 2-1.

For the common message scenario we first derive upper and lower bounds on the common-message secrecy capacity. These bounds coincide when the channels are reversely-degraded thus establishing the secrecy capacity in this case.

For the case of independent messages we establish the secrecy sum-capacity for the reversely degraded case. The capacity-achieving scheme is simple: transmit to the strongest receiver on each channel and use independent codebooks across the sub-channels.

We note that the problem of transmitting common and independent messages to multiple receivers over parallel channels is first considered in [15]. Our results generalize [15], by considering the secrecy constraint. Interestingly, however, the specializations of our capacity-achieving schemes to the case of no eavesdropper are different from those in [15].

2.1 Problem Model

We formulate the problems of interest as extensions of the wiretap channel model introduced by Wyner [53] for studying reliable and secure communication in an information-theoretic framework. As such, we emphasize that in our models there is no prior key shared between the sender and legitimate receivers, and both the encoding and decoding functions, and the codebook itself, are public.

In this broadcast model, there are M parallel subchannels connecting a single sender to each of K legitimate receivers and an eavesdropper, where M and K are parameters.

Definition 1 *A* product *broadcast channel is one in which the constituent sub-channels have finite input and output alphabets, are memoryless and independent of each other, and are characterized by their transition probabilities*

$$\Pr\left(\{y_{1m}^n, \ldots, y_{Km}^n, y_{\mathrm{e}m}^n\}_{m=1,\ldots,M} \mid \{x_m^n\}_{m=1,\ldots,M}\right) = \prod_{m=1}^{M} \prod_{t=1}^{n} \Pr(y_{1m}(t), \ldots, y_{Km}(t), y_{\mathrm{e}m}(t) \mid x_m(t)), \quad (2.1)$$

where $x_m^n = (x_m(1), x_m(2), \ldots, x_m(n))$ denotes the sequence of symbols transmitted on subchannel m, where $y_{km}^n = (y_{km}(1), y_{km}(2), \ldots, y_{km}(n))$ denotes the sequence of symbols obtained by receiver k on subchannel m, and where $y_{\mathrm{e}m}^n = (y_{\mathrm{e}m}(1), y_{\mathrm{e}m}(2), \ldots, y_{\mathrm{e}m}(n))$ denotes the sequence of symbols received by the eavesdropper on subchannel m. The alphabet of x_m is $\mathcal{X}$, and the alphabet for both y_{km} and $y_{\mathrm{e}m}$ is $\mathcal{Y}$.

A special class of product broadcast channels, known as the reversely degraded broadcast channel [15] are of particular interest.

Definition 2 *A product broadcast channel is* reversely-degraded *when each of the M constituent subchannels is degraded in a prescribed order. In particular, for each subchannel m, there exists a permutation $\{\pi_m(1), \pi_m(2), \ldots, \pi_m(K+1)\}$ of the set $\{1, 2, \ldots, K, \mathrm{e}\}$ such that the following Markov chain is satisfied, i.e.,*

$$x_m \to y_{\pi_m(1)} \to y_{\pi_m(2)} \to \cdots \to y_{\pi_m(K+1)}.$$

With this definition, $y_{\pi_m(1)}, y_{\pi_m(2)}, \ldots, y_{\pi_m(K+1)}$ is an ordering of the receivers from strongest to weakest in the mth subchannel, and we will at times find it convenient to adopt the additional notation $\pi_m \triangleq \pi_m(1)$. Also, we stress that in Definition 2 the order of degradation need not be the same for all subchannels, so the overall channel need not be degraded. An example of reversely-degraded parallel broadcast channel is depicted in Fig. 2-1.

We also emphasize that in any subchannel the K receivers and eavesdropper are *physically* degraded. Our capacity results, however, only depend on the marginal distribution of receivers in each subchannel. Accordingly, our results in fact hold for the larger class of channels in which there is only stochastic degradation in the subchannels.

Finally, we obtain further results when the channel is Gaussian.

Definition 3 *A reversely-degraded product broadcast channel is* Gaussian *when it*

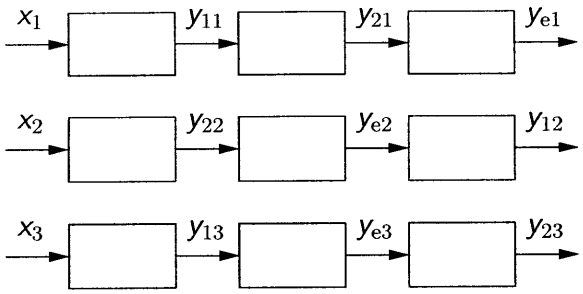


Figure 2-1: An example of reversely-degraded parallel broadcast channel, in which there are $M = 3$ subchannels connecting a single sender to each of $K = 2$ legitimate receivers and an eavesdropper. The input symbols to the subchannels are (x_1, x_2, x_3). The output symbols at the kth intended receiver are (y_{k1}, y_{k2}, y_{k3}), and at the eavesdropper are (y_{e1}, y_{e2}, y_{e3}). Note that the order of degradation is not the same for for all subchannels.

takes the form

$$\begin{aligned} y_{km} &= x_m + z_{km}, \quad m = 1, \ldots, M, \quad k = 1, \ldots, K \\ y_{em} &= x_m + z_{em}, \end{aligned} \tag{2.2}$$

where the noise variables are all mutually independent, and $z_{km} \sim \mathcal{CN}(0, \sigma_{km}^2)$ *and* $z_{em} \sim \mathcal{CN}(0, \sigma_{em}^2)$. *For this channel, there is also an average power constraint*

$$E\left[\sum_{m=1}^{M} x_m^2\right] \leq P.$$

We now provide the formal definitions of the common-message secrecy capacity and the sum-secrecy capacity for independent messages.

Definition 4 *A* $(n, 2^{nR})$ *code consists of a message set* $\mathcal{W} = \{1, 2, \ldots 2^{nR}\}$, *a (possibly stochastic) encoder* $\omega_n : \mathcal{W} \to \mathcal{X}^n \times \mathcal{X}^n \times \overset{(M\text{-fold})}{\ldots\ldots} \times \mathcal{X}^n$ *mapping the message set to the codewords for the* M *subchannels, and a decoder* $\Phi_{k,n} : \mathcal{Y}^n \times \mathcal{Y}^n \times \overset{(M\text{-fold})}{\ldots\ldots} \times \mathcal{Y}^n \to \mathcal{W}$ *for* $k = 1, 2, \ldots, K$ *at each receiver. Using* $\hat{w}_k$ *to denote message estimate at decoder* k, *a common-message-secrecy-rate* R *is said to be achievable if, for any* $\varepsilon > 0$, *there exists a length* n *code such that* $\Pr(w \neq \hat{w}_k) \leq \varepsilon$ *for* $k = 1, 2, \ldots, K$, *while*

$$\frac{1}{n} H(w | y_{e1}^n, y_{e2}^n, \ldots, y_{eK}^n) \geq R - \varepsilon. \tag{2.3}$$

The common-message secrecy capacity is the supremum over all achievable rates.

Definition 5 *A* $(2^{nR_1}, 2^{nR_2}, \ldots, 2^{nR_K}, n)$ *code for the product broadcast channel in Definition 1 consists of a message set* $\mathcal{W}_k = \{1, 2, \ldots 2^{nR_k)}\}$, *for* $k = 1, 2, \ldots, K$, *an encoder* $\omega_n : \mathcal{W}_1 \times \mathcal{W}_2 \times \cdots \times \mathcal{W}_K \to \mathcal{X}^n \times \mathcal{X}^n \overset{(M\text{-fold})}{\ldots\ldots} \times \mathcal{X}^n$ *mapping the messages for the* K *receivers to the* M *subchannel inputs, and* K *decoding functions* $\phi_{k,n} : \mathcal{Y}^n \times \mathcal{Y}^n \times \overset{(M\text{-fold})}{\ldots\ldots} \times \mathcal{Y}^n \to \mathcal{W}_k$, *one at each legitimate receiver. We denote*

the message estimate at decoder k by $\hat{w}_k$. A secrecy rate-tuple $(R_1, R_2, \ldots, R_K)$ is achievable if, for every $\varepsilon > 0$, there is a code of length n such that $\Pr(w_k \neq \hat{w}_k) \leq \varepsilon$ for all $k = 1, 2, \ldots, K$, and such that

$$\frac{1}{n} H(w_k | w_1, \ldots, w_{k-1}, w_{k+1}, \ldots, w_K, y_{e1}^n, \ldots, y_{eM}^n) \geq \frac{1}{n} H(w_k) - \varepsilon, \qquad k = 1, 2, \ldots, K. \quad (2.4)$$

The secrecy sum-capacity is the supremum of $R_1 + R_2 + \cdots + R_K$ over the achievable rate tuples $(R_1, R_2, \ldots, R_K)$.

We remark that our constraint (2.4) provides perfect equivocation for each message, even if all the other messages are revealed to the eavesdropper. It may be possible to increase the secrecy rate by exploiting the fact that the eavesdropper does not have access to other messages. This weaker notion of secrecy is not considered here.

2.2 Capacity results

We summarize the capacity results in this section. First, we will provide the capacity results when there is a single legitimate receiver. The results here have been derived in [53, 8] and [33]. Subsequently we provide the common message secrecy capacity and the sum secrecy capacity with independent messages for the case of reversely degraded parallel channels. Bounds on the capacity are provided when the channels are not reversely degraded. These results have been published in [23].

2.2.1 Single User Case

First note that the case when $K = 1$ and $M = 1$ is the single user wiretap channel.

Theorem 1 (I. Csiszár and J. Körner: [8]) *The secrecy capacity with $M = 1$ channel and $K = 1$ receiver is given by*

$$C = \max_{p_u p_{x|u}} I(u; y_r) - I(u; y_e), \quad (2.5)$$

where the maximization is over random variables $u \to x \to (y_r, y_e)$ and $|\mathcal{U}| \leq |\mathcal{X}|^2 + 4|\mathcal{X}| + 3$.

The secrecy capacity with M parallel independent channels is provided in [33].

Theorem 2 (Liang et.al. [33]) *The secrecy capacity for the case of M parallel channels and $K = 1$ receiver and one eavesdropper is*

$$C = \sum_{i=1}^{M} \max_{p_{u_i} p_{x_i|u_i}} I(u_i; y_i) - I(u_i; y_{ei}), \quad (2.6)$$

where the maximization is over random variables $u_i \rightarrow x_i \rightarrow (y_i, y_{ei})$ *with appropriate cardinality constraints.*

This result establishes that independent codebooks across the parallel channels achieve the secrecy capacity. When the broadcast channel is reversely degraded, the secrecy capacity expression simplifies as follows —

Corollary 1 *The secrecy capacity for the reversely degraded broadcast channels with* $K = 1$ *receiver is,*

$$C = \sum_{i=1}^{M} \max_{p_{x_i}} \{I(x_i; y_i) - I(x_i; y_{ei})\}. \tag{2.7}$$

The single user results are reminiscent of the well known fact (see e.g., [6]) that (in absence of the secrecy constraint) independent coding across parallel channels achieves the capacity in the single user case.

2.2.2 Common Message

We have the following upper and lower bounds on the common-message secrecy capacity for the product broadcast channel of Definition 1.

Proposition 1 *For the product broadcast channel model, an upper bound on the secrecy capacity is given by*

$$\bar{C}_{K,M} \leq \bar{R}^{+}_{K,M} \triangleq \min_{\mathcal{P}} \max_{\prod_{m=1}^{M} p(x_m)} \min_{k \in \{1,\ldots,K\}} \sum_{m=1}^{M} I(x_m; y_{km} | y_{em}) \tag{2.8}$$

where the set $\mathcal{P} = \mathcal{P}_1 \times \cdots \times \mathcal{P}_M$ *is a cartesian product of the sets* $\{\mathcal{P}_m\}_{m=1}^{M}$*, and where each* $\mathcal{P}_m$ *is the collection of all joint distributions* $p'(y_{1m}, \ldots, y_{Km}, y_{em} | x_m)$ *having the same marginal distribution as* $p(y_{1m}|x_m), \ldots, p(y_{Km}|x_m)$ *and* $p(y_{em}|x_m)$*, and where the maximum is over all marginal distributions* $p(x_1), \ldots, p(x_M)$*.*

Proposition 2 *A lower bound on the secrecy capacity for the product broadcast channel model is given by*

$$\bar{C}_{K,M} \geq \bar{R}^{-}_{K,M} = \max_{\substack{\{p(u_m)\}_{m=1}^{M} \\ \{x_m = f_m(u_m)\}_{m=1}^{M}}} \min_{k \in \{1,\ldots,K\}} \sum_{m=1}^{M} \{I(u_m; y_{km}) - I(u_m; y_{em})\}^{+}, \tag{2.9}$$

where the random variables $u_1, \ldots, u_M$ *are independent over some alphabet* $\mathcal{U}$*, and each* $f_m(\cdot)$ *for* $m = 1, \ldots, M$ *is a mapping from* $\mathcal{U}$ *to* $\mathcal{X}$*.*

For the special case of a product broadcast channel that is reversely-degraded, our upper and lower bounds above coincide, yielding the following common-message secrecy capacity.

Theorem 3 *The common-message secrecy capacity for the reversely-degraded channel model is*

$$\bar{C}_{K,M} = \max_{\prod_{m=1}^{M} p(x_m)} \min_{k\in\{1,2,\ldots,K\}} \sum_{m=1}^{M} I(x_m; y_{km}|y_{em}). \tag{2.10}$$

Finally, for the Gaussian parallel channel model of Definition 3, we have the following straightforward extension of Theorem 3.

Corollary 2 *The common-message secrecy capacity for the Gaussian parallel broadcast channel is*

$$\bar{C}^{\mathrm{G}}_{K,M} = \max_{(P_1,\ldots,P_M)\in\mathcal{F}} \min_{1\le k\le K} \sum_{m=1}^{M} \left\{\log\left(\frac{1+P_m/\sigma_{km}^2}{1+P_m/\sigma_{em}^2}\right)\right\}^+, \tag{2.11}$$

where $\mathcal{F}$ is the set of all feasible power allocations, i.e.,

$$\mathcal{F} = \left\{(P_1,\ldots,P_M) \;\middle|\; P_m \ge 0, \sum_{m=1}^{M} P_m \le P\right\}. \tag{2.12}$$

2.2.3 Independent Message

In absence of the secrecy constraint, the sum capacity for the reversely degraded broadcast channel is maximized when only the strongest user on each parallel channel is served [15, 48]. We show that the same scheme is also optimal with the secrecy constraint.

Theorem 4 *Let π_j denote the strongest user on channel j. The secrecy-sum-capacity for the reversely broadcast channel is given by*

$$C^{\mathrm{sum}}_{K,M} = \max_{p(x_1)p(x_2)\ldots p(x_M)} \sum_{j=1}^{M} I(x_j; y_{\pi_j}|y_{ej}). \tag{2.13}$$

Furthermore, the expression in (2.13) *is an upper bound on the secrecy-sum-capacity when only the legitimate users are reversely degraded — but the set of receivers together with the eavesdropper is not degraded.*

2.2.4 Specializing to no-secrecy constraint

El Gamal [15] has studied this setup in absence of a secrecy constraint for reversely degraded channels. It is interesting to specialize our schemes when there is no eavesdropper to compare with [15].

1. For the case of a common message [15] shows that independent coding across the sub-channels is sub-optimal. He proposes a scheme where a single vector

codebook is used across the channels to achieve the capacity. In contrast our scheme yields another technique to achieve the common message capacity. We use a separate codebook on each of the channels at a rate of the common message capacity. The receivers are required to jointly decode accross the codebooks.

2. For the case of independent messages, we show that the sum-secrecy-capacity can be achieved by transmitting to the strongest receiver on each channel and independently coding across the channels, analogous to the case of no-secrecy in [15].

2.3 Parallel Channels - Common Message

In this section we provide proofs for the capacity results in section 2.2.2. In sections 2.3.1 and 2.3.2 we derive Propositions 1 and 2 respectively. These bounds are shown to coincide the the case of reversely-degraded channels in section 2.3.3 and the Gaussian case is studied in section 2.3.4.

2.3.1 Upper Bound

We first state a few facts that will be used in this sequel. In Defn. 4, both the error probability as well as the equivocation depend only on the marginal distribution of the channels. Hence we have the following.

Fact 3 *The common-message-secrecy-capacity for the wiretap channel depends on the joint distribution* $p(y_{1j}, \ldots, y_{Kj}, p(y_{ej})|x_j)$ *only via the marginal distributions* $p(y_{1j}|x_j), p(y_{2j}|x_j), \ldots, p(y_{ej}|x_j)$ *in* (2.1) *for each* $j = 1, 2, \ldots, M$.

We establish the following in Appendix A.

Fact 4 *For any random variables* x, y, *and* z *the quantity* $I(x; y|z)$ *is concave in* $p(x)$.

We use these facts in the proof of the upper bound.

Proof. [Lemma 1]

Suppose there exists a sequence of $(n, 2^{nR})$ codes such that, for every $\varepsilon > 0$, as $n \to \infty$

$$\begin{aligned} &\Pr(w \neq \hat{w}_i) \leq \varepsilon, \quad i = 1, 2, \ldots K \\ &\frac{1}{n} I(w; y_{e1}^n, \ldots, y_{eM}^n) \leq \varepsilon. \end{aligned} \tag{2.14}$$

We first note that from Fano's inequality we have

$$\frac{1}{n} H(w|y_{i1}^n, y_{i2}^n, \ldots, y_{iM}^n) \leq \frac{1}{n} + \varepsilon R \quad i = 1, 2, \ldots K. \tag{2.15}$$

Combining (2.14) and (2.15) we have, for all $i = 1, 2, \ldots K$ and $\varepsilon' = \varepsilon + \frac{1}{n} + \varepsilon R$,

$$
\begin{aligned}
nR &\leq I(w; y_{i1}^n, \ldots, y_{iM}^n) - I(w; y_{e1}^n, \ldots, y_{eM}^n) + n\varepsilon' \\
&\leq I(w; y_{i1}^n, \ldots, y_{iM}^n | y_{e1}^n, \ldots, y_{eM}^n) + n\varepsilon' \\
&= h(y_{i1}^n, \ldots, y_{iM}^n | y_{e1}^n, \ldots, y_{eM}^n) - h(y_{i1}^n, \ldots, y_{iM}^n | y_{e1}^n, \ldots, y_{eM}^n, w) \\
&\leq h(y_{i1}^n, \ldots, y_{iM}^n | y_{e1}^n, \ldots, y_{eM}^n) - h(y_{i1}^n, \ldots, y_{iM}^n | y_{e1}^n, \ldots, y_{eM}^n, x_1^n, \ldots, x_M^n, w) \\
&= h(y_{i1}^n, \ldots, y_{iM}^n | y_{e1}^n, \ldots, y_{eM}^n) - h(y_{i1}^n, \ldots, y_{iM}^n | y_{e1}^n, \ldots, y_{eM}^n, x_1^n, \ldots, x_M^n) \qquad (2.16) \\
&= h(y_{i1}^n, \ldots, y_{iM}^n | y_{e1}^n, \ldots, y_{eM}^n) - \sum_{j=1}^{M} h(y_{ij}^n | x_j^n, y_{ej}^n) + n\varepsilon' \qquad (2.17) \\
&\leq \sum_{j=1}^{M} h(y_{ij}^n | y_{ej^n}) - \sum_{j=1}^{M} h(y_{ij}^n | x_j^n, y_{ej}^n) + n\varepsilon' \\
&\leq \sum_{j=1}^{M} I(x_j^n; y_{ij}^n | y_{ej}^n) + n\varepsilon', \qquad (2.18)
\end{aligned}
$$

where (2.16) follows from the fact that $w \to (x_1^n, \ldots x_M^n, y_{e1}^n, \ldots, y_{eM}^n) \to (y_{i1}^n, \ldots, y_{iM}^n)$ form a Markov chain, and (2.17) holds because the parallel channels are mutually independent in (2.1) so that

$$
h(y_{i1}^n, \ldots, y_{iM}^n | y_{e1}^n, \ldots, y_{eM}^n, x_1^n, \ldots, x_M^n) = \sum_{j=1}^{M} h(y_{ij}^n | x_j^n, y_{ej}^n) \; .
$$

We now upper bound each term in the summation (2.18). We have

$$
\begin{aligned}
I(x_j^n; y_{ij}^n | y_{ej}^n) &\leq \sum_{k=1}^{n} I(x_j(k); y_{ij}(k) | y_{ej}(k)) \qquad (2.19) \\
&= \sum_{k=1}^{n} I(x_j(k); y_{ij}(k), y_{ej}(k)) - I(x_j(k); y_{ej}(k)) \qquad (2.20) \\
&= nI(x_j; y_{ij}, y_{ej} | q) - nI(x_j; y_{ej} | q) \qquad (2.21) \\
&= nI(x_j; y_{ij} | y_{ej}, q) \\
&\leq nI(x_j; y_{ij} | y_{ej}), \qquad (2.22)
\end{aligned}
$$

where (2.19) follows from the fact that the channel is memoryless,and (2.21) is obtained by defining q to be a (time-sharing) random variable uniformly distributed over $\{1, 2, \ldots, n\}$ independent of everything else. The random variables (x_j, y_{ij}, y_{ej}) are such that, conditioned on $q = k$, they have the same joint distribution as $(x_j(k), y_{ij}(k), y_{ej}(k))$. Finally (2.22) follows from the fact that the mutual information is concave with respect to the input distribution $p(x_j)$ as stated in Fact 4.

Combining (2.22) and (2.17) we have

$$R \leq \sum_{j=1}^{M} I(x_j; y_{ij}|y_{ej}) + \varepsilon', \quad i = 1, 2, \ldots K$$

$$= \min_{1\leq i\leq K} \sum_{j=1}^{M} I(x_j; y_{ij}|y_{ej}) + \varepsilon' \tag{2.23}$$

$$\leq \max_{\prod_{j=1}^{M} p(x_j)} \min_{1\leq i\leq K} \sum_{j=1}^{M} I(x_j; y_{ij}|y_{ej}) + \varepsilon' \,. \tag{2.24}$$

The last step follows from that fact that for any input distribution $p(x_1, x_2, \ldots, x_M)$, the objective function $\min_{1\leq i\leq K} \sum_{j=1}^{M} I(x_j; y_{ij}|y_{ej})$ only depends on the marginal distributions $p(x_1), \ldots, p(x_M)$. Accordingly it suffices to take $x_1, x_2, \ldots, x_M$ as mutually independent random variables. Finally note that (2.24) depends on the joint distribution across the channels. Accordingly, we tighten the upper bound by considering the worst distribution in $\mathcal{P} = \mathcal{P}_1 \times \mathcal{P}_2 \times \ldots \times \mathcal{P}_M$ which gives

$$R_{\text{KM}}^{+} \leq \min_{\mathcal{P}} \max_{\prod_{j=1}^{M} p(x_j)} \min_{1\leq i\leq K} \sum_{j=1}^{M} I(x_j; y_{ij}|y_{ej}) + \varepsilon' \,. \tag{2.25}$$

■

2.3.2 Lower Bound

We now present a coding scheme that achieves the our lower bound.

We first discuss the structure of the coding scheme informally. We construct M independent random codebooks $\mathcal{C}_1, \ldots, \mathcal{C}_M$, one for each subchannel. Codebook $\mathcal{C}_m$ has nearly $2^{n(R+I(u_m;y_{em}))}$ codewords, randomly partitioned into 2^{nR} bins, one for each possible message. Hence, there are nearly $Q_m = 2^{nI(u_m;y_{em})}$ codewords per bin. Given a particular message $W \in \{1, 2, \ldots, 2^{nR}\}$ to be sent, the encoder selects M codewords, one for each subchannel. Specifically, if the message is w, then for each subchannel m the encoder randomly selects for transmission one of the Q_m codewords from the wth bin in $\mathcal{C}_m$. This bin structure of the codebooks is depicted in Fig. 2-2 for the case of $M = 2$ subchannels.

To decode, each legitimate receiver attempts to find a message that is jointly typical with its set of M received sequences. As we now show, the rate R of the code can be chosen arbitrarily close to $\bar{R}_{K,M}^{-}$ as defined in (2.9) and guarantees both successful decoding with high probability for each legitimate receiver, and near-perfect equivocation at the eavesdropper.

Before presenting our proof, we make some remarks. As mentioned earlier, when specialized to the case in which there is no eavesdropper (and hence no secrecy constraint), our construction is different from that developed by El Gamal [15] for such product broadcast channels. In particular, as illustrated in Fig. 2-3 for the case of $M = 3$ subchannels, our construction has the distinguishing feature that independent

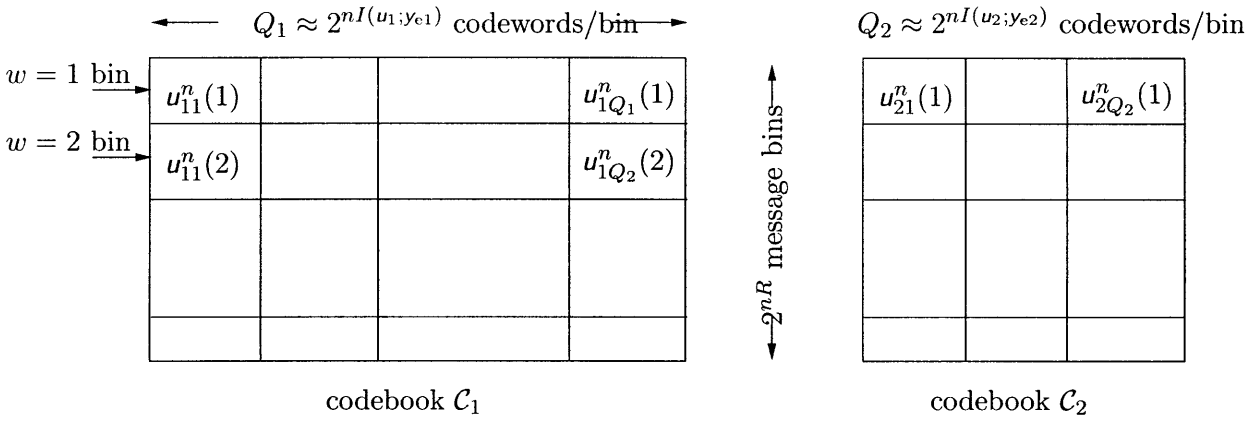


Figure 2-2: Binning encoder for the secure product broadcast channel, for the case of $M = 2$ subchannels. The set of codewords for representing a particular message $w \in \{1, \ldots, 2^{nR}\}$ in the mth subchannel are denoted by $u_{m1}^n(w), \ldots, u_{mQ_m}^n(w)$. To encode a particular message w, the encoder randomly selects one of the Q_m codewords in the associated bin for transmission in the mth subchannel, for $m = 1, \ldots, M$.

codebooks are used for the different subchannels. By comparison, with the scheme in [15], each message is mapped to a $M \times n$ dimensional codeword and the mth component of the codeword is transmitted on subchannel m. This corresponds to a single-codebook scheme. By extending this scheme to provide secrecy by incorporating random binning, one can achieve, again for the reversely-degraded channel,

$$R^{\text{single}} = \max_{p(x_1,\ldots,x_M)} \min_{k\in\{1,\ldots,K\}} \left\{ I(x_1, x_2 \ldots, x_K; y_{k1}, \ldots, y_{kK}) - I(x_1, x_2 \ldots, x_K; y_{e1}, \ldots, y_{eK}) \right\}, \tag{2.26}$$

which we observe is in general smaller than that achieved by our construction, viz., (2.10). Ultimately, allowing the sizes of bins to depend on the mutual information at the eavesdropper on each particular subchannel makes it possible to confuse the eavesdropper on each subchannel, and thereby achieve higher secrecy rates than (2.26).

We now provide the formal details and analysis of the coding scheme.

Proof. [Proof of Proposition 2] First, fix the distributions $p(u_1), p(u_2), \ldots, p(u_M)$ and the (possibly stochastic) functions $f_1(\cdot), \ldots, f_M(\cdot)$. Let η_2 and η_1 be positive constants, to be quantified later. With respect to these quantities, define

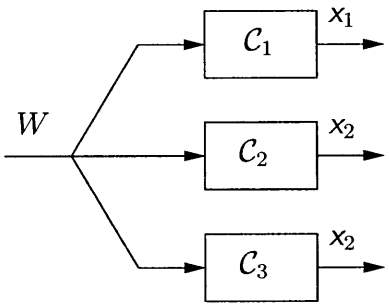
$$R = \min_{1\le k\le K} \sum_{m=1}^{M} \{I(u_m; y_{km}) - I(u_m; y_{em})\}^+ - \eta_1 \tag{2.27}$$

and

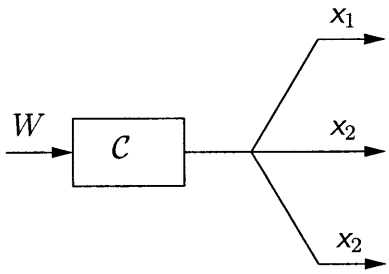
$$R_{em} = I(u_m; y_{em}) - \eta_2, \quad m = 1, 2, \ldots, M. \tag{2.28}$$

The set $T(u_m)$ denotes the set of all sequences that are typical[1] with respect to distribution $p(u_m)$ and the set $T(x_m, u_m)$ denotes the set of all jointly typical sequences

[1]Throughout our development, we mean typicality in the ε-weak sense; see, e.g., [6, Chapter 3].



(a) Secrecy capacity-achieving code structure.



(b) Nonsecrecy capacity-achieving code structure of [15].

Figure 2-3: Structure of two coding schemes for common message transmission over reversely degraded product broadcast channels, for the case of $K = 2$ legitimate receivers and one eavesdropper. To obtain secrecy, separate codebooks are required for each subchannel, so that separate binning can be performed on each. A single codebook is sufficient when there is no secrecy requirement.

(x_m^n, u_m^n) with respect to the distribution $p(x_m, u_m)$. In turn, $T_{u_m^n}(x_m|u_m)$ denotes the set of all sequences x_m^n conditionally typical with respect to a given sequence u_m^n according to $p(x_m|u_m)$.

The details of our construction are as follows.

Codebook Generation

- Codebook $\mathcal{C}_m$ for $m = 1, 2, \ldots, M$ has a total of $M_m = 2^{n(R+R_{em})}$ length n codeword sequences. Each sequence is selected uniformly and independently from the set $T(u_m)$.
- We randomly partition the M_m sequences into 2^{nR} message bins so that there are $Q_m = 2^{nR_{em}}$ codewords per bin.
- The set of codewords associated with bin w in codebook $\mathcal{C}_m$ is denoted as

$$\mathcal{C}_m(w) = \{u_{m1}^n(w), u_{m2}^n(w), \ldots, u_{mQ_m}^n(w)\}, \tag{2.29}$$

  for $w = 1, 2, \ldots, 2^{nR}$ and $m = 1, 2, \ldots, M$. Note that $\mathcal{C}_m = \bigcup_{w=1}^{2^{nR}} \mathcal{C}_m(w)$ is the codebook on subchannel m.

Encoding

To encode message w, the encoder randomly and uniformly selects a codeword in the set $\mathcal{C}_m(w)$ for all $1 \leq m \leq M$. Specifically,

- Select M integers $q_1, q_2, \ldots, q_M$, where q_m is selected independently and uniformly from the set $\{1, 2, \ldots Q_m\}$.
- Given a message w, select a codeword $u_{mq_m}^n(w)$ from codebook $\mathcal{C}_m(w)$ for $m = 1, 2, \ldots, M$.
- The transmitted sequence on subchannel m is denoted by $x_m^n = x_m(1), x_m(2), \ldots, x_m(n)$. The symbol $x_m(t)$ is obtained by taking the (possibly stochastic) function $f_m(\cdot)$ of each element of the codeword $u_{mq_m}^n(w)$.

Decoding

Receiver k, based on its observations $(y_{k1}^n, y_{k2}^n, \ldots, y_{kM}^n)$ from the M parallel subchannels, declares message w according to the following rule:

- Let $\mathcal{S}_k = \{m|1 \leq m \leq M, I(u_m; y_{km}) > I(u_m; y_{em})\}$ denote the set of subchannels where receiver k has larger mutual information than the eavesdropper. The receiver only considers the outputs y_{km}^n from these subchannels.
- Receiver k searches for a message w such that, for each $m \in \mathcal{S}_k$, there is an index l_m such that $(u_{ml_m}^n(w), y_{km}^n) \in T(u_m, y_{km})$. If a unique w has this property, the receiver declares it as the transmitted message. Otherwise, the receiver declares an arbitrary message.

We now analyze the properties of this code.

Error Probability

We show that, averaged over the ensemble of codebooks, the error probability is smaller than a constant ε' (to be specified). This demonstrates the existence of a codebook with error probability less than ε'. We do the analysis for receiver k and, without loss of generality, assume that message w_1 is transmitted.

We first analyze the false-reject event. Let $\mathcal{E}_{1m}^c$ be the event $\{(u_{mq_m}^n(w_1), y_{km}^n) \notin T(u_m, y_{km})\}$. Since $u_{mq_m}^n \in T(u_m)$ by construction and y_{km} is obtained by passing u_m through a discrete memoryless channel, it follows that [6, Page 72, Theorem 3.1.2], $\Pr(\mathcal{E}_{1m}^c) \leq \varepsilon$. Accordingly, if $\mathcal{E}_1^c$ denotes the event that message w_1 does not appear typical, then we have

$$\Pr(\mathcal{E}_1^c) = \Pr\left(\bigcup_{m=1}^{M} \mathcal{E}_{1m}^c\right) \leq M\varepsilon. \tag{2.30}$$

We next analyze the false-accept event. As before, let $\mathcal{S}_k \subseteq \{1, 2, \ldots, M\}$ denote the subset of subchannels for which $I(u_m; y_{km}) > I(u_m; y_{em})$. In what follows, the index m refers only to subchannels in $\mathcal{S}_k$.

For each $m \in \mathcal{S}_k$, let $\mathcal{E}_{im}$ denote the event that there is a codeword in the set $\mathcal{C}_m(w_i)$ $(i > 1)$ typical with y_{km}^n. Then

$$\begin{aligned}
\Pr(\mathcal{E}_{im}) &= \Pr\big(\exists l \in \{1, \ldots, Q_m\} : (u_{ml}^n(w_i), y_{km}^n) \in T(u_m, y_{km})\big) \\
&\leq \sum_{l=1}^{Q_m} \Pr((u_{ml}^n(w_i), y_{km}^n) \in T(u_m, y_{km})) \\
&\leq \sum_{l=1}^{Q_m} 2^{-n(I(u_m;y_{km})-3\varepsilon)} \qquad (2.31) \\
&\leq 2^{-n(I(u_m;y_{km})-I(u_m;y_{em})-3\varepsilon+\eta_2)}, \qquad (2.32)
\end{aligned}$$

where (2.31) follows from the fact that since the sequences $(u_{ml}^n(w_i), y_{km}^n)$ are drawn independently, the results in [6, Page 216, Theorem 8.6.1] apply and (2.32) follows by noting that $Q_m = 2^{n(I(u_m;y_{em})-\eta_2)}$.

In turn, let $\mathcal{E}_i$ denote the event that message w_i has a codeword typical on every subchannel. Then

$$\begin{aligned}
\Pr(\mathcal{E}_i) &= \Pr\left(\bigcap_{m\in\mathcal{S}_k} \mathcal{E}_{im}\right) \\
&= \prod_{m\in\mathcal{S}_k} \Pr(\mathcal{E}_{im}) \qquad (2.33) \\
&= 2^{-n\sum_{m\in\mathcal{S}_k}(I(u_m;y_{km})-I(u_m;y_{em})-3\varepsilon+\eta_2)} \\
&= 2^{-n\sum_{m=1}^{M}(\{I(u_m;y_{km})-I(u_m;y_{em})\}^+-3\varepsilon+\eta_2)},
\end{aligned}$$

where (2.33) follows by independence of codebooks and sub-channels.

Finally, the probability of false accept event $\mathcal{E}_{\mathrm{F}}$ is given by

$$\begin{aligned}\Pr(\mathcal{E}_{\mathrm{F}}) &= \Pr\left(\bigcup_{i=2}^{2^{nR}} \mathcal{E}_i\right) \\ &\leq 2^{nR} 2^{-n\sum_{m=1}^{M}(\{I(u_m;y_{km})-I(u_m;y_{em})\}^+ - 3\varepsilon+\eta_2)} \\ &\leq 2^{-n(-3M\varepsilon+M\eta_2+\eta_1)},\end{aligned}$$

which vanishes with increasing n by selecting the code-parameters such that $\eta_1 + M\eta_2 - 3M\varepsilon > 0$.

Thus, the probability of error averaged over the ensemble of codebooks is less than

$$\varepsilon' = \max\left(M\varepsilon, 2^{-n(-3M\varepsilon+M\eta_2+\eta_1)}\right),$$

which demonstrates the existence of a codebook with error probability less than ε'.

Secrecy Analysis

We now show that for any typical code in the ensemble the normalized mutual information between the message and the output of the eavesdropper is vanishing in the block-length. We establish this in two steps. First, our construction of codebooks is such that an eavesdropper who observes only the output of channel m satisfies $(1/n)I(w; y_{em}^n) \in o_n(1)$.[2] Secondly, as we show below, the eavesdropper's mutual information only increases by a factor of M even when all the channel outputs are observed:

$$\begin{aligned}\frac{1}{n} I(w; y_{\mathrm{e}1}^n, \ldots, y_{\mathrm{e}M}^n) &= \frac{1}{n} h(y_{\mathrm{e}1}^n, \ldots, y_{\mathrm{e}M}^n) - \frac{1}{n} h(y_{\mathrm{e}1}^n, \ldots, y_{\mathrm{e}M}^n | w) \\ &= \frac{1}{n} h(y_{\mathrm{e}1}^n, \ldots, y_{\mathrm{e}M}^n) - \sum_{m=1}^{M} \frac{1}{n} h(y_{em}^n | w) \qquad (2.34) \\ &\leq \sum_{m=1}^{M} \frac{1}{n} I(w; y_{em}^n) \in o_n(1) \qquad (2.35)\end{aligned}$$

where (2.34) follows from the fact that the codewords in the sets $\mathcal{C}_1(w), \mathcal{C}_2(w), \ldots, \mathcal{C}_M(w)$ are independently selected.

We now show that for all $m = 1, \ldots, M$,

$$\frac{1}{n} I(w; y_{em}^n) \in o_n(1), . \qquad (2.36)$$

Since there are there are $Q_m = 2^{n(I(u_m;y_{em})-\eta_2)}$ codewords in each codebook $\mathcal{C}_m(w)$

[2] We will use $o_n(1)$ to refer to a function that approaches zero as $n \to \infty$.

we have that,

$$\frac{1}{n}H(u_m^n|w) = I(u_m; y_{em}) - \eta_2, \tag{2.37}$$

$$\frac{1}{n}H(u_m^n|w, y_{em}^n) \leq \gamma \triangleq \frac{1}{n} + \eta_2 \, R_{em}, \tag{2.38}$$

where, (2.37) follows from the fact that the codewords in each bin are selected uniformly, while (2.38) follows from the fact that a typical codebook $\mathcal{C}_m(w)$ satisfies Fano's inequality. Furthermore, following [53], we can show that for our codebook $\mathcal{C}_m$, all of whose codewords are equally likely to be transmitted, we have that

$$\frac{1}{n}I(u_m^n; y_{em}^n) \leq I(u_m; y_{em}) + |\mathcal{U}| \Pr(u_m^n \notin T(u)) + o_n(1). \tag{2.39}$$

The equivocation at the eavesdropper can then be lower bounded using (2.37)-(2.39).

$$\begin{aligned}
H(w|y_{em}^n) &= H(w, u_m^n|y_{em}^n) - H(u_m^n|w, y_{em}^n) \\
&\geq H(u_m^n|y_{em}^n) - n\gamma && (2.40) \\
&= H(u_m^n) - I(u_m^n; y_{em}^n) - n\gamma \\
&= H(u_m^n, w) - I(u_m^n; y_{em}^n) - n\gamma && (2.41) \\
&= H(w) + H(u_m^n|w) - I(u_m^n; y_{em}^n) - n\gamma \\
&\geq H(w) + nI(u_m; y_{em}) - I(u_m^n; y_{em}^n) - n\gamma - n\eta_2, && (2.42) \\
&\geq H(w) - n\gamma - n\eta_2 - no_n(1) - n|\mathcal{U}|\varepsilon \, , && (2.43)
\end{aligned}$$

where (2.40) follows from substituting (2.38), where (2.41) follows from the fact that w is deterministic given u_m^n, and where (2.42) and (2.43) follow from substituting (2.37) and (2.39) respectively, and the fact that $\Pr(u_m^n \notin T(u)) \leq \varepsilon$. Since γ, η_2 and ε can be selected to be arbitrarily small, provided n is sufficiently large, we establish (2.36).

■

2.3.3 Capacity for Reversely degraded channels

We observe that the upper and lower bounds in Proposition 1 and 2 respectively, coincide when the underlying channel is reversely degraded.

Proof. [Proof of Theorem 3] By selecting $u_m = x_m$ for each $m = 1, 2, \ldots, M$, in the achievable rate expression (2.9) in Prop. 2, we have that

$$\bar{R}_{K,M}^{-} = \min_{k \in \{1,\ldots,K\}} \sum_{m=1}^{M} \{I(x_m; y_{km}) - I(x_m; y_{em})\}^+,$$

is an achievable rate. For the reversely degraded channel, for each $k = 1, 2, \ldots, K$, and $m = 1, 2, \ldots, M$, we have that either $x_m \to y_{km} \to y_{em}$ or $x_m \to y_{em} \to y_{km}$ holds.

In either case, note that

$$\{I(x_m; y_{km}) - I(x_m; y_{em})\}^+ = I(x_m; y_{km}|y_{em}),$$

holds, and hence the lower bound above coincides with (2.8) in Prop. 1. ∎

2.3.4 Gaussian Channel Capacity

We extend the secrecy capacity in Theorem 3 to Gaussian parallel channels. Since the extension is based on standard techniques, we will only sketch the key steps in the proof.

Proof. [Proof of Corollary 2] Note that the channel of Definition 3 has the same capacity as another (M, K) reversely-degraded broadcast channel in which the sequence obtained at receiver $\pi_m(k+1)$ on subchannel m is

$$\hat{y}_{\pi_m(k+1)m} = \hat{y}_{\pi_m(k)m} + \hat{z}_{\pi_m(k)m}, \quad k = 0, 1, \ldots, K,$$

where $\pi_m(1), \ldots, \pi_m(K+1)$ denotes the ordering of the eavesdropper and legitimate receivers from strongest to weakest, where $\hat{y}_{\pi_m(0)m} \triangleq x_m$ and $\sigma^2_{\pi_m(0)m} \triangleq 0$, and where the noises $\hat{z}_{\pi_m(k)m} \sim \mathcal{CN}(0, \sigma^2_{\pi_m(k+1)m} - \sigma^2_{\pi_m(k)m})$ are mutually independent.

With the appropriate Fano's inequality, the converse for Theorem 3 extends to continuous alphabets. The achievability argument relies on weak typicality and also extends to the Gaussian case. Furthermore, the power constraint can be incorporated in the capacity expression, since the objective function is concave in the input distribution, whence

$$\bar{C}_{K,M}(P) = \max_{\substack{\prod_{m=1}^M p(x_m), \\ E\left[\sum_{m=1}^M x_m^2\right] \le P}} \min_{k \in \{1,\ldots,K\}} \sum_{m=1}^M I(x_m; \hat{y}_{km}|\hat{y}_{em}). \tag{2.44}$$

Next observe that

$$\max_{p(x_m), E[x_m^2] \le P_m} I(x_m; \hat{y}_{km}|\hat{y}_{em})$$

denotes the capacity of a Gaussian wiretap channel [27]. Accordingly, for each $m = 1, 2, \ldots, M$,

$$\max_{p(x_m), E[x_m^2] \le P_m} I(x_m; \hat{y}_{km}|\hat{y}_{em}) = \left\{ \log\left(\frac{1 + P_m/\sigma^2_{km}}{1 + P_m/\sigma^2_{em}} \right) \right\}^+. \tag{2.45}$$

Now if $(P_1^*, \ldots, P_M^*)$ denotes an optimal power allocation in (2.44), then via (2.45), we have that

$$\bar{C}_{K,M}(P) = \min_{k \in \{1,\ldots,K\}} \sum_{m=1}^M \left\{ \log\left(\frac{1 + P_m^*/\sigma^2_{km}}{1 + P_m^*/\sigma^2_{em}} \right) \right\}^+,$$

whence (2.11) follows. ∎

2.4 Parallel Channels — Independent Messages

We establish Theorem 4 by providing a converse and achievability result.

2.4.1 Converse Theorem 4

We establish the upper bound in Theorem 4. Suppose a genie provides the output of the strongest receiver, π_j, to all other receivers on each channel, i.e., on channel j the output $y^n_{\pi_j}$ is made available to all the receivers. Because of degradation, we may assume, without loss of generality, that each receiver only observes $(y^n_{\pi_1}, \ldots, y^n_{\pi_M})$. Clearly, such a genie aided channel can only have a sum capacity larger than the original channel. Since all receivers are identical, to compute the sum capacity it suffices to consider the situation with one sender, one receiver, and one eavesdropper.

Lemma 1 *The secrecy-sum-capacity in Theorem 4 is upper bounded by the secrecy capacity of the genie aided channel, i.e.,* $C^{\text{sum}}_{K,M} \leq C^{\text{GenieAided}}$.

Proof. Suppose that a secrecy rate point $(R_1, R_2, \ldots R_K)$ is achievable for the K user channel in Theorem 4 and let the messages be denoted as $(w_1, w_2, \ldots w_K)$. This implies that, for any $\varepsilon > 0$ and n large enough, there is a length n code such that $\Pr(\hat{w}_i \neq w_i) \leq \varepsilon$ for $i = 1, 2, \ldots, K$, and such that

$$\frac{1}{n} H(w_i | w_1, \ldots w_{i-1}, w_{i+1}, \ldots w_K, y^n_{e1}, y^n_{e2}, \ldots, y^n_{eM}) \geq R_i - \varepsilon \,. \tag{2.46}$$

We now show that a rate of $(\sum_{i=1}^K R_i, \underbrace{0, \ldots, 0}_{K-1})$ is achievable on the genie aided channel. First, note that any message that is correctly decoded on the original channel is also correctly decoded by user 1 on the genie aided channel. It remains to bound the equivocation on the genie aided channel when the message to receiver 1 is $w = (w_1, w_2, \ldots, w_K)$. We have

$$\begin{aligned}
\frac{1}{n} H(w | y^n_{e1}, y^n_{e2}, \ldots, y^n_{eM}) &= \frac{1}{n} H(w_1, w_2, \ldots, w_K | y^n_{e1}, y^n_{e2}, \ldots, y^n_{eM}) \\
&\geq \sum_{i=1}^K \frac{1}{n} H(w_i | w_1, \ldots w_{i-1}, w_{i+1}, \ldots w_K, y^n_{e1}, y^n_{e2}, \ldots, y^n_{eM}) \\
&\geq \sum_{i=1}^K R_i - K\varepsilon
\end{aligned}$$

where the last step follows from (2.46). Since ε is arbitrary, this establishes the claim. ∎

Lemma 2 *The secrecy capacity of the genie aided channel is*

$$C^{\text{GenieAided}} = \max_{p(x_1)p(x_2)\ldots p(x_M)} \sum_{j=1}^M I(x_j; y_{\pi_j} | y_{ej}). \tag{2.47}$$

Proof. Since all receivers are identical on the genie aided channel, this setup reduces to the case of $K = 1$ receivers and Corollary 1 applies. ∎

Remark 1 *The upper bound continues to hold even if the eavesdroppers channel is not ordered with respect to the legitimate receivers. In general, following Lemma 1, the upper bound can be tightened by considering, for all $1 \leq j \leq M$, the worst joint distribution $p'(y_{\pi_j}, y_{ej}|x_j)$ among all joint distributions with the same marginal distribution as $p(y_{\pi_j}|x_j)$ and $p(y_{ej}|x_j)$, yielding*

$$C_{K,M}^{\text{sum}} \leq \min_{\prod_{j=1}^{M} p'(y_{\pi_j}, y_{ej}|x_j)} \max_{\prod_{j=1}^{M} p(x_j)} \sum_{j=1}^{M} I(x_j; y_{\pi_j}|y_{ej}). \tag{2.48}$$

2.4.2 Achievability for Theorem 4

The achievability scheme for Theorem 4 is as follows: we only send information intended to the strongest user, i.e., only user π_j on channel j can decode. It follows from the result of the wiretap channel [53] that a rate of $R_j = \max_{p(x_j)} I(x_j; y_{\pi_j}|y_{e_j})$ is achievable on channel j. Accordingly the total sum rate of $\sum_j R_j$ is achievable which is the capacity expression.

2.4.3 Gaussian Channels

Theorem 4 can be extended to the case of Gaussian parallel channels. Let $\sigma_{\pi_j}^2$ denote the noise variance of the strongest user on channel j. Then the secrecy-sum-capacity is given by

$$C_{K,M}^{\text{sum,Gaussian}}(P) = \max_{(P_1,P_2,\ldots P_M)} \sum_{j=1}^{M} \left\{ \frac{1}{2} \log \left(1 + \frac{P_j}{\sigma_{\pi_j}^2} \right) - \frac{1}{2} \log \left(1 + \frac{P_j}{\sigma_{ej}^2} \right) \right\}^+ \tag{2.49}$$

where the maximization is over all power allocations satisfying $\sum_{j=1}^{M} P_j \leq P$. The achievability follows by using independent Gaussian wiretap codebooks on each channel and only considering the strongest user on each channel. For the upper bound we have to show that Gaussian inputs are optimal in the capacity expression in Theorem 4. The justifications are the same as in the common message case in Section 2.3.4.

2.5 Conclusions

This chapter studies an extension of the wiretap channel when there are multiple legitimate receivers. We examine two scenarios — when all receivers want a common message and when each of the receivers wants an independent message. The notion of secrecy capacities are appropriately extended in these cases and upper and lower bounds are derived on the capacities in several cases. The common-message-secrecy-capacity is established for the case of reversely degraded parallel channels. For the case of independent messages, we show that a scheme that transmits to the strongest

user on each channel achieves the sum-secrecy-capacity. The results on parallel channels provide important insights into the scenario of fading channels. The case of fading channels is studied in the next chapter.

Chapter 3

Fading Channels

In this chapter, we extend the schemes for the parallel channels discussed in the previous chapter to the case of fading channels. In this setup, we assume that the channel state information (CSI) of legitimate receivers is revealed to all communicating parties (including the eavesdropper), while the eavesdropper's channel gains are revealed only to her. The sender and receiver know the statistical properties of the eavesdropper channel and this knowledge is used to characterize the secrecy capacity.

We first examine the case when a common message needs to be delivered to all legitimate receivers in the presence of potential eavesdroppers and present a scheme that achieves a rate that does not decay to zero with increasing number of receivers. Interestingly, without the secrecy constraint the problem of multicasting a common message to several receivers over ergodic fading channels has received little attention. Indeed transmitter CSI appears to be of little value in these situations. Some thought can convince that the capacity appears to be not too far from the maximum achievable rate with a flat power allocation scheme.

In contrast, with the secrecy constraint a flat power allocation may not provide any rate unless the eavesdropper's channel is weaker on average than the legitimate receiver. In this sense, the secrecy constraint adds a new dimension to the multicasting problem. It requires us to consider protocols that exploit transmitter CSI in an efficient manner. Note that there is a tension between receivers when the channels undergo independent fading. Not all receivers will experience a strong channel simultaneously and the proposed multicasting protocols resolve this tension in an efficient manner and achieve a rate that does not vanish with the number of receivers.

When there are independent messages, we propose an opportunistic scheme that selects the user with the strongest channel at each time. With Gaussian wiretap codebooks for each legitimate receiver, we show that this scheme achieves the sum capacity in the limit of large number of receivers. Note that the analogous results without secrecy constraints are established in [49].

3.1 Problem Model

Definition 6 *Our fast-fading broadcast model of interest has the following properties. The received sequences $y_1^n, y_2^n, \ldots, y_K^n$ and y_e^n at the legitimate receivers and eavesdropper, respectively, are of the form*

$$\begin{aligned} y_k(t) &= h_k(t)\,x(t) + z_k(t), \quad k = 1, 2, \ldots, K, \\ y_e(t) &= h_e(t)\,x(t) + z_e(t), \end{aligned} \tag{3.1}$$

where x^n is the transmitted sequence, and $z_k(t) \sim \mathcal{CN}(0,1)$. The channel gains and noises among all receivers (including the eavesdropper) are all mutually independent of one another, and all vary in an independently, identically-distributed (i.i.d.) manner with time, corresponding to fast-fading.[1] *Finally, the input must satisfy an average power constraint $E[|x(t)|^2] \le P$.*

Furthermore we will assume the case that $h_k(t) \sim \mathcal{CN}(0,1)$ and $h_e(t) \sim \mathcal{CN}(0,1)$, although many of our results can extend to more general fading models.

In addition, in our model the $h_1(t), \ldots, h_K(t)$ are revealed to the transmitter, the K legitimate receivers and the eavesdropper in a causal manner. Implicitly we assume that there is an authenticated public feedback link from the receivers to the transmitter. The channel coefficients of the eavesdropper $\{h_e^n\}$ are known only to the eavesdropper, but the transmitter and the legitimate receivers know the probability distribution of the eavesdropper's channel gains.

We now provide the formal definitions of the common-message secrecy capacity and the sum-secrecy capacity for independent messages.

Definition 7 *A $(n, 2^{nR})$ code for the channel consists of an encoding function that maps from the message $w \in \{1, 2, \ldots, 2^{nR}\}$ into transmitted symbols $x(t) = f_t(w; h_1^t, h_2^t, \ldots, h_K^t)$ for $t = 1, 2, \ldots, n$, and a decoding function $\hat{w}_k = \phi_k(y_k^n; h_1^n, h_2^n, \ldots, h_K^n)$ at each receiver k. A rate R is achievable if, for every $\varepsilon > 0$, there exists a sequence of length n codes such that $\Pr(\hat{w}_k \neq w) \le \varepsilon$ for any $k = 1, 2, \ldots, K$ such that*

$$\frac{1}{n} H\left(w \mid y_e^n, h_e^n, h_1^n, \ldots, h_K^n\right) \ge R - \varepsilon. \tag{3.2}$$

Definition 8 *A $(n, 2^{nR_1}, \ldots, 2^{nR_K})$ code consists of an encoding function from the messages $w_1, \ldots, w_K$ with $w_k \in \{1, 2, \ldots, 2^{nR_k}\}$ to transmitted symbols*

$$x(t) = f_t(w_1, w_2, \ldots, w_K; h_1^t, h_2^t, \ldots, h_K^t)$$

for $t = 1, 2, \ldots, n$, and a decoding function at each receiver $\hat{w}_k = \phi_k(y_k^n; h_1^n, h_2^n, \ldots, h_K^n)$. A secrecy rate-tuple $(R_1, R_2, \ldots, R_K)$ is achievable if, for any $\varepsilon > 0$, there exists a length n code such that, for each $k = 1, 2, \ldots, K$, with w_k uniformly distributed over

[1] In practice, the fast fading model (3.1) applies when the codebooks are interleaved so that each symbol sees an independent fade.

$\{1, 2, \ldots, 2^{nR_k}\}$, *we have* $\Pr(\hat{w}_k \neq w_k) \leq \varepsilon$ *and*

$$\frac{1}{n} H\left(w_k \mid w_1, \ldots, w_{k-1}, w_{k+1}, \ldots, w_K, y_e^n, h_e^n, h_1^n, \ldots, h_K^n\right) \geq R_k - \varepsilon. \tag{3.3}$$

The secrecy sum-capacity is the supremum value of $R_1 + R_2 + \ldots + R_K$ *among all achievable rate tuples.*

Note that the entropy term in both (3.2) and (3.3) is conditioned on $h_1^n, \ldots, h_K^n$ as these channel gains of the K receivers are assumed to be known to the eavesdropper. However, the encoding and decoding functions do not depend on h_e^n as this realization is not known to the sender and the receivers.

An immediate consequence of this formulation is that the secrecy capacity depends only on the distribution of $h_e(t)$ and not on the actual realized sequence of these eavesdropper gains. Indeed, since the transmitter and the legitimate receivers do not have the eavesdropper's CSI, the encoding and decoding functions cannot depend on this information. From this perspective, in our formulation a message that is secure with respect to any given eavesdropper is also secure against any statistically equivalent eavesdropper. Thus, the assumption of only a single eavesdropper in our model is purely one of convenience.

3.2 Capacity Results

In this section, we summarize the capacity results.

We first study the case of a single eavesdropper and present bounds on the secrecy capacity. To the best of our knowledge, the secrecy capacity for this scenario remains open. Upper and lower bounds for this problem were first provided in [22]. Similar results subsequently appeared in [31].

We note that the secrecy capacity has been resolved in the following variations of this setup which will not be discussed in this thesis.

1. When the eavesdropper's channel coefficients $h_e(t)$ are known to the sender, the setup is mapped to the case of parallel independent channels and the capacity is resolved by Liang et. al. [32]

2. When the eavesdropper's channel coefficients are not known, but the coherence period goes to infinity, the secrecy capacity is obtained in [20].

After discussing the bounds on capacity for the single user case, we discuss the case of many users. These results have also been published in [22, 23].

3.2.1 Single User Case

We first consider the case when there is only one receiver in this section.

$$\begin{aligned} y(t) &= h(t)x(t) + z(t) \\ y_e(t) &= h_e(t)x(t) + z_e(t). \end{aligned} \tag{3.4}$$

Note that here we denote the channel of legitimate receiver with $h(t)$ instead of $h_1(t)$ as in (3.1).

Proposition 3 *For the single user fast-fading channel, the secrecy sum-capacity is bounded by*

$$R^-(P) \leq C_K(P) \leq R^+(P), \tag{3.5}$$

where

$$R^+(P) = \max_{\substack{\rho(h): \\ E[\rho(h)] \leq P}} E\left[\left\{\log\left(\frac{1+|h|^2\rho(h)}{1+|h_e|^2\rho(h)}\right)\right\}^+\right] \tag{3.6a}$$

and

$$R^-(P) = \max_{\substack{\rho(h): \\ E[\rho(h)] \leq P}} E\left[\log\left(\frac{1+|h|^2\rho(h)}{1+|h_e|^2\rho(h)}\right)\right]. \tag{3.6b}$$

Note that the difference between the upper and lower bounds is the $\{\cdot\}^+$ function inside the expectation. Evaluating these bounds for i.i.d. Rayleigh fading channels, with $E[|h|^2] = E[|h_e|^2] = 1$, in the SNR regime yields,

$$\lim_{P\to\infty} R^-(P) = E\left[\left\{\log|h|^2 + \frac{\gamma}{\log 2}\right\}^+\right] = 0.7089 \text{ b/s/Hz} \tag{3.7a}$$

$$\lim_{P\to\infty} R^+(P) = E\left[\left\{\log\frac{|h|^2}{|h_e|^2}\right\}^+\right] = 1 \text{ b/s/Hz}. \tag{3.7b}$$

The lower bound can be further improved by transmitting synthetic noise. This improvement results in

$$\lim_{P\to\infty} R^-_{\text{SN}}(P) = 0.7479 \text{ b/s/Hz},$$

which is still far from the upper bound.

The achievability scheme corresponding to the lower bound (3.6b) involves mapping the fading channel to a set of parallel independent channels and using independent codebooks accross these channels. The upper bound (3.6a) was first provided by Gopala et. al. [20] and is included here for completeness.

3.2.2 Common Message

The common message constraint requires us to simultaneously adapt rate and power to the channel gains of several legitimate users. How efficiently can this be done as the number of receivers increases? Somewhat surprisingly, we observe that it is possible to broadcast at a rate independent of the number of legitimate users.

Theorem 5 *The common-message-secrecy-rate for the fast-fading broadcast channel is bounded by*

$$\bar{R}^{-}(P) \leq \bar{C}_K(P) \leq \bar{R}^{+}(P), \tag{3.8}$$

where

$$\bar{R}^{-}(P) = \min_{1\leq k\leq K} E_{h_k}\left[\left\{\log\left(\frac{1+|h_k|^2P}{\exp\left\{E_{h_e}[\log(1+|h_e|^2P)]\right\}}\right)\right\}^{+}\right] \tag{3.9a}$$

and

$$\bar{R}^{+}(P) = \min_{1\leq k\leq K} \max_{\substack{\rho(h_k):\\ E[\rho(h_k)]\leq P}} E\left[\left\{\log\left(\frac{1+|h_k|^2\rho(h_k)}{1+|h_e|^2\rho(h_k)}\right)\right\}^{+}\right]. \tag{3.9b}$$

When the channel gains h_k are identically distributed across the users, note that both lower and upper bounds in (3.9) are independent of the number of receivers K. The fact that the common-message secrecy-capacity does not vanish with the number of users is surprising. Simple schemes such as transmitting when all the users have a channel gain above a threshold or time-sharing between the users only achieve a rate that vanishes with the number of users. In contrast our lower bound is achieved by a scheme that simultaneously adapts to the time variations of all the legitimate users.

In the high signal-to-noise ratio (SNR) regime, the bounds Theorem 5 specialize as follows.

Corollary 3 *When the channel gains of all the receivers are distributed as* $\mathcal{CN}(0,1)$*, the bounds in* (3.9) *are, asymptotically,*

$$\lim_{P\to\infty} \bar{R}^{+}(P) = E\left[\left\{\log\frac{|h|^2}{|h_e|^2}\right\}^{+}\right] = 1 \ b/s/Hz \tag{3.10a}$$

$$\lim_{P\to\infty} \bar{R}^{-}(P) = E\left[\left\{\log|h|^2 + \frac{\gamma}{\log 2}\right\}^{+}\right] = 0.7089 \ b/s/Hz, \tag{3.10b}$$

where γ *is the Euler-Gamma constant* ($\gamma \approx 0.5772$).

While our proposed scheme achieves a rate independent of the number of users (and hence the best possible scaling with the number of users), the optimality of the scheme remains open.

3.2.3 Independent Messages

The problem of broadcasting independent messages to multiple receivers over ergodic fading channels has been well studied when there is no security constraint; see. e.g., [48]. For such scenarios, an opportunistic transmission scheme is shown to attain the largest sum-capacity. We establish the following analogous result for secure transmission.

Proposition 4 *For the fast-fading broadcast channel, the secrecy sum-capacity is bounded by*

$$R_K^-(P) \leq C_K(P) \leq R_K^+(P), \tag{3.11}$$

where

$$R^+(P) = \max_{\substack{\rho(h_{\max}):\\ E[\rho(h_{\max})]\leq P}} E\left[\left\{\log\left(\frac{1+|h_{\max}|^2\rho(h_{\max})}{1+|h_e|^2\rho(h_{\max})}\right)\right\}^+\right] \tag{3.12a}$$

and

$$R_K^-(P) = \max_{\substack{\rho(h_{\max}):\\ E[\rho(h_{\max})]\leq P}} E\left[\log\left(\frac{1+|h_{\max}|^2\rho(h_{\max})}{1+|h_e|^2\rho(h_{\max})}\right)\right], \tag{3.12b}$$

with $h_{\max}$ denoting the gain of the strongest of the K legitimate receivers (at any instant).

Our upper and lower bounds in (3.12) are distinguished by the inclusion of the operator $\{\cdot\}^+$ is inside the expectation of the former. Hence, the arguments of the expectation differ whenever $|h_{\max}|^2 \leq |h_e|^2$, and so an upper bound on the rate gap is

$$R_K^+(P) - R_K^-(P) \leq \Pr(|h_e|^2 \geq |h_{\max}|^2) E\left[\log\left(|h_e|^2/|h_{\max}|^2\right) \mid |h_e|^2 \geq |h_{\max}|^2\right]. \tag{3.13}$$

As the number of legitimate receivers grows the event $\{|h_{\max}|^2 \leq |h_e|^2\}$ happens increasingly rarely and for the case of identical Rayleigh distributed fading, the gap between the bounds vanishes. As a result, we obtain the following theorem.

Theorem 6 *For the fast-fading broadcast channel with identical Rayleigh distributed fading and large K, the secrecy capacity scales according to*

$$C_K(P) = \max_{\substack{\rho(h_{\max}):\\ E[\rho(h_{\max})]\leq P}} E\left[\log\left(\frac{1+|h_{\max}|^2\rho(h_{\max})}{1+|h_e|^2\rho(h_{\max})}\right)\right] + o(1). \tag{3.14}$$

where we use $o(1)$ to denote terms that approach zero as $K \to \infty$.

Theorem 6 establishes that an architecture that uses single-user Gaussian wiretap base codes in conjunction with opportunistic transmission achieves the secrecy sum-capacity in the limit of a large number of receivers.

For finite values of K, incorporating synthesized noise into the transmission as a masking technique yields still higher rates [22]. However, even with such refinements, there remains a gap between the upper and lower bounds. Fig. 3-1 illustrates the upper and lower bounds in (3.12) in the high SNR regime for identically distributed Rayleigh fading distribution. We note that even for a moderate number of users, these bounds are nearly tight and further improvements will only provide diminishing gains in this regime.

We also remark that Theorem 6 more generally guarantees an arbitrarily small gap between upper and lower bounds on the secrecy sum-capacity for Rayleigh fading channels of fixed coherence time, provided the number of receivers is large enough.

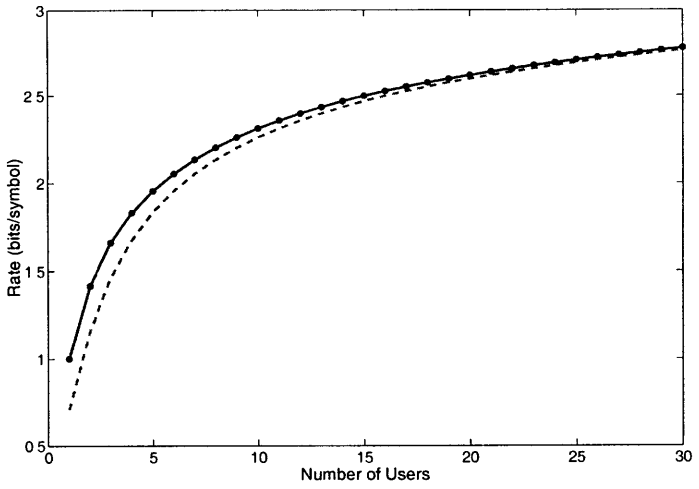


Figure 3-1: Upper and lower bounds on the secrecy sum-capacity in (3.12) for the broadcasting of independent messages in Rayleigh fast-fading environments in the high SNR regime, as a function of the number of legitimate receivers.

In [20] *variable-rate* and *fixed-rate* schemes are developed for the case of a single receiver in a slow fading environment. Straightforward extensions of these schemes for multiple receivers reveals the following insights. The variable-rate scheme achieves our upper bound (3.12a), whereas the fixed-rate scheme achieves our lower bound (3.12b). Since these two expressions coincide as the number of receivers tends to infinity, it follows that the gains of variable-rate schemes become negligible in this limit.

As a final remark, we comment on collusion attacks. As noted earlier, any number of statistically equivalent eavesdroppers does not affect our capacity—as long as they do not collude. However, if the eavesdroppers collude, they can combine the received signals and attempt to decode the message. In such scenarios, the upper and lower bounds in Proposition 4 can be extended by replacing the term $|h_e|^2$ with $\|h_e\|^2$, where h_e is the vector of channel gains of the colluding eavesdroppers. One interesting implication of the resulting bounds is that the secrecy capacity is positive unless the colluding eavesdropper population grows as $\log K$.

3.3 Single User

We first consider the case when there is only one receiver, and study in turn a lower bound and an upper bound on the secrecy capacity.

3.3.1 Achievability

We can view the model (3.4) as a set of parallel channels in Fig. 3-2 indexed by the channel gain h of the intended receiver, which is known globally. Thus in each parallel

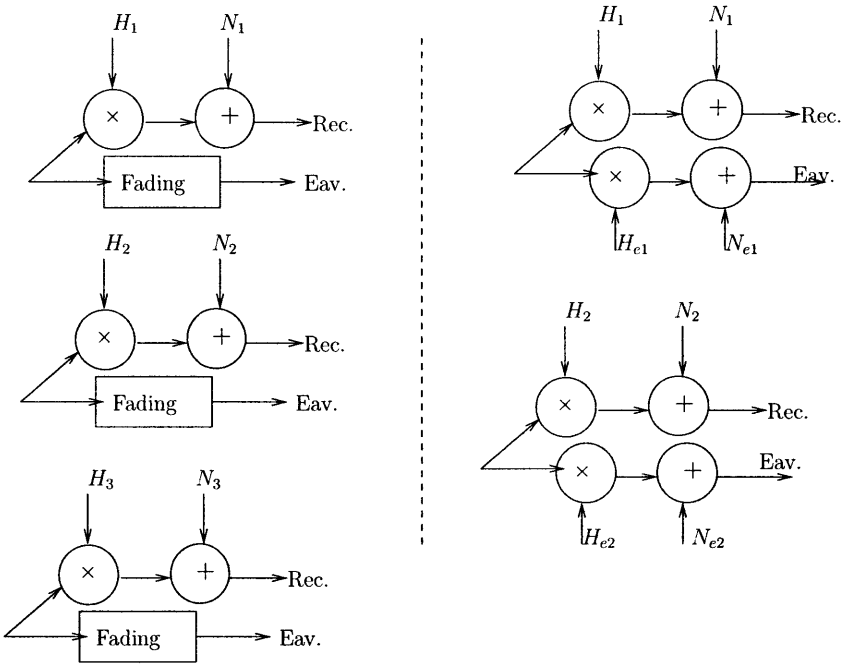


Figure 3-2: Parallel channel decomposition of the fading channel with one receiver and one eavesdropper. The decomposition on the left is used in the achievability scheme when the channel coefficients of the intended receiver are known to the sender, the receiver and the eavesdropper. This contrasts with the decomposition on the right when both the intended receiver and the eavesdropper are known to all the nodes.

channel the intended receiver's channel is complex Gaussian while the eavesdropper's channel is a fading channel. We use an independent Gaussian codebook on each parallel channel.

Consider a particular sub-channel where the intended receivers experiences a gain of a (i.e. $|h|^2 = a$). Generate an i.i.d. Gaussian wiretap codebook [27] with power P_a and rate $R^I(a, P_a)$. The power P_a is selected to satisfy the average power constraint $E[P_a] = P$. The achievable rate is:

$$\begin{aligned} R^I(a, P_a) &= I(x; y_r) - I(x; y_e, h_e) \\ &= \left\{\log(1 + aP_a) - E[\log(1 + |h_e|^2 P_a)]\right\}. \end{aligned} \tag{3.15}$$

From the expression (3.15), it is clear that our achievable rate $R^I(a, P_a)$ is increasing in a. It is possible to show that if a is fixed and greater than $T \triangleq \exp(-\gamma)$, where $\gamma = 0.5772$ is the Euler's constant, the supremum of $R^I(a, P_a)$ is obtained in the limit $P_a \to \infty$. On the other hand if $a < T$, then $\sup_{P_a > 0} R^I(a, P_a) = 0$. Thus for the proposed scheme, the transmitter will not transmit whenever $a < T$.

The expression (3.6b) in Proposition 3 follows by taking expectation with respect to the fading states.

Synthetic Noise Transmission

It is possible to improve upon the proposed rate in (3.15) by transmitting artificial noise in addition to the intended codeword. We split the available power P_a into two parts. Generate an i.i.d. Gaussian wiretap codebook with power P_u. Before transmission of a codeword u^n, generate an i.i.d Gaussian noise sequence v^n with power P_v, independent of everything else and not known to the receiver. Our choice of the powers satisfy $P_u + P_v = P_a$. We transmit $x^n = u^n + v^n$. The received symbols at the intended receiver and eavesdropper are

$$\begin{aligned} y(i) &= hu(i) + hv(i) + z(i) \\ y_{\mathrm{e}}(i) &= h_{\mathrm{e}}(i)u(i) + h_{\mathrm{e}}(i)v(i) + z(i) \end{aligned} \tag{3.16}$$

Our expression for the achievable rate is given by,

$$\begin{aligned} R^{II}(a, P_a) &= I(u; u_{\mathrm{r}}) - I(u; y_{\mathrm{e}}, h_{\mathrm{e}}) \\ &= \left\{ \log\left(1 + \frac{aP_u}{1 + aP_v}\right) - E\left[\log\left(1 + \frac{|h_{\mathrm{e}}|^2 P_u}{1 + |h_{\mathrm{e}}|^2 P_v}\right)\right] \right\} \end{aligned} \tag{3.17}$$

We optimize over the choice of P_u and P_v. It can be shown that for any $a > 0$, we have that $\sup_{P_a} R^{II}(a, P_a) > 0$. Thus secret communication is possible for every choice of $a > 0$, provided the available power is sufficiently large. Note that the gain from artificial noise should not be very surprising. As seen in (3.17), the artificial noise gets amplified by the channel gain of the receivers and hence there is a net gain if the channel gain to the intended receiver is small. The optimal value of P_v is positive only if $a < 1$. Thus if the channel gain of the intended receiver is greater than one, our scheme reduces to the previous one in (3.15).

Numerical evaluation in the high SNR limit yields

$$\begin{aligned} \lim_{P \to \infty} R^{-}(P) &= 0.7089 \text{ bits/symbol}, \\ \lim_{P \to \infty} R^{-}_{\mathrm{SN}}(P) &= 0.7479 \text{ bits/symbol}. \end{aligned} \tag{3.18}$$

As a final remark, we note that even though our proposed scheme uses an independent codeword for each parallel channel, this is not necessary. In particular, the rate can also be obtained by using a *single* Gaussian wiretap codebook generated i.i.d. $\mathcal{CN}(0, 1)$ and scaling each transmitted symbol by the transmit power P_a depending on the channel state. This reduces the complexity of encoding and decoding significantly.

3.3.2 Single User: Upper Bound

Our upper bounding technique follows closely [20], where a similar setup for large coherence periods is studied. The derivation below is provided for completeness.

First note that the joint distribution of the noise variables $(z(t), z_{\mathrm{e}}(t))$ is selected so that if $|h_{\mathrm{e}}(t)| \leq |h(t)|$ we have the Markov chain $x(t) \to y(t) \to y_{\mathrm{e}}(t)$; otherwise we have the chain $x(t) \to y_{\mathrm{e}}(t) \to y(t)$. We show that for any sequence of length n,

rate R codes as in Def. 8 the upper bound (3.12a) holds. Recall that the encoding function has the form

$$x(t) = f_t(w, h^t), \quad t = 1, 2, \ldots, n, \tag{3.19}$$

and for every $\varepsilon > 0$, and sufficiently large n, we have, via Fano's inequality and the secrecy condition,

$$\frac{1}{n} H(w|h^n, y^n) \leq \varepsilon \tag{3.20}$$

$$\frac{1}{n} I(w; y_e^n, h_e^n|h^n) \leq \varepsilon. \tag{3.21}$$

An upper bound on the rate is as follows,

$$\begin{aligned}
nR &= H(w|h^n) \\
&\leq I(w; y^n|h^n) - I(w; y_e^n, h_e^n|h^n) + 2n\varepsilon && (3.22)\\
&\leq I(x^n; y^n|h^n, h_e^n, y_e^n) + 2n\varepsilon && (3.23)\\
&= h(y^n|h^n, h_e^n, y_e^n) - h(y^n|h^n, h_e^n, y_e^n, x^n) + 2n\varepsilon \\
&= h(y^n|h^n, h_e^n, y_e^n) - \sum_{t=1}^{n} h(y(t)|h(t), h_e(t), y_e(t), x(t)) + 2n\varepsilon && (3.24)\\
&\leq h(y^n|h^n, h_e^n, y_e^n) - \sum_{t=1}^{n} h(y(t)|h^t, h_e(t), y_e(t), x(t)) + 2n\varepsilon \\
&\leq \sum_{t=1}^{n} I(x(t); y(t)|y_e(t), h^t, h_e(t)) + 2n\varepsilon && (3.25)
\end{aligned}$$

where (3.22) follows by substituting (3.20) and (3.21), (3.23) follows from the Markov chain $w \to (x^n, y_e^n, h^n, h_e^n) \to y^n$, where (3.24) follows from the fact that the channel is memoryless.

From the capacity of the Gaussian wiretap channel [27], we have that,

$$I(x(t); y(t)|y_e(t), h^t, h_e(t)) \leq E_{h^t, h_e(t)}\left[\left\{\log \frac{1 + |h(t)|^2 E[|x(t)|^2]}{1 + |h_e(t)|^2 E[|x(t)|^2]}\right\}^+\right] \tag{3.26}$$

with equality if $x(t)$ is conditionally Gaussian given $(h^t, h_e(t))$. Since a Gaussian distribution depends only on its mean and variance and $x(t)$ is independent of $h_e(t)$, we can write without loss of generality[2] that

$$x(t) \sim \mathcal{CN}\left(0, \sqrt{\rho_t(h^t)}\right), \tag{3.27}$$

for some sequence of functions $\rho_t(\cdot)$ that satisfy the average power constraint $\frac{1}{n}\sum_{t=1}^{n} E[\rho_t(h^t)] \leq$

[2]Analogous approach is taken in [4, Section IV, Proposition 3] for establishing the capacity of fading channels with side information at the transmitter.

P. With this substitution, we have from (3.25) that

$$nR \leq \sum_{t=1}^{n} E_{h^t,h_e(t)} \left[\left\{ \log \frac{1+|h(t)|^2\rho_t(h^t)}{1+|h_e(t)|^2\rho_t(h^t)} \right\}^+ \right] + 2n\varepsilon. \tag{3.28}$$

As shown below, that the right hand side in (3.28) is maximized, for each t, by a function $\gamma_t(\cdot)$ that only depends on h^t via $h(t)$. The upper bound expression in (3.12a) then follows, since from (3.28),

$$\begin{aligned} nR - 2n\varepsilon &\leq \sum_{t=1}^{n} E_{h,h_e} \left[\left\{ \log \frac{1+|h|^2\gamma_t(h)}{1+|h_e|^2\gamma_t(h)} \right\}^+ \right] \\ &\leq nE_{h,h_e} \left[\left\{ \log \frac{1+|h|^2 \frac{1}{n}\sum_{t=1}^{n}\gamma_t(h)}{1+|h_e|^2 \frac{1}{n}\sum_{t=1}^{n}\gamma_t(h)} \right\}^+ \right] \qquad (3.29) \\ &= nE_{h,h_e} \left[\left\{ \log \frac{1+|h|^2\gamma(h)}{1+|h_e|^2\gamma(h)} \right\}^+ \right], \qquad (3.30) \end{aligned}$$

where (3.29) follows from the fact $\{\log(1+ax)/(1+bx)\}^+$ is concave in $x > 0$ for fixed a and b, so Jensen's inequality can be applied and where (3.30) follows by defining $\gamma(h) = \frac{1}{n}\sum_{t=1}^{n}\gamma_t(h)$. Note that the power constraint $E[\gamma(h)] \leq P$ naturally follows from the definition of $\gamma(\cdot)$.

It remains to establish the existence of $\gamma_t(\cdot)$ as we now do.

In particular, for any sequence of functions $\rho_t(\cdot)$, we define $\gamma_t(\cdot)$ according to,

$$\gamma_t(h(t)) \triangleq E_{h^{t-1}}[\rho_t(h^t)|h(t)],$$

and show below that each term in the summation in (3.28) only increases if we replace $\rho_t(\cdot)$ by $\gamma_t(\cdot)$.

$$\begin{aligned} & E_{h^t,h_e(t)} \left[\left\{ \log \frac{1+|h(t)|^2\rho_t(h^t)}{1+|h_e(t)|^2\rho_t(h^t)} \right\}^+ \right] \qquad (3.31) \\ &= E_{h(t),h_e(t)} \left[E_{h^{t-1}} \left[\left\{ \log \frac{1+|h(t)|^2\rho_t(h^t)}{1+|h_e(t)|^2\rho_t(h^t)} \right\}^+ \right] \right] \\ &\leq E_{h(t),h_e(t)} \left[\left\{ \log \frac{1+|h(t)|^2 E_{h^{t-1}}[\rho_t(h^t)|h(t)]}{1+|h_e(t)|^2 E_{h^{t-1}}[\rho_t(h^t)|h(t)]} \right\}^+ \right] \qquad (3.32) \\ &= E_{h(t),h_e(t)} \left[\left\{ \log \frac{1+|h(t)|^2\gamma_t(h(t))}{1+|h_e(t)|^2\gamma_t(h(t))} \right\}^+ \right] \qquad (3.33) \\ &= E_{h,h_e} \left[\left\{ \log \frac{1+|h|^2\gamma_t(h)}{1+|h_e|^2\gamma_t(h)} \right\}^+ \right], \qquad (3.34) \end{aligned}$$

where (3.32) follows from Jensen's inequality. This completes the proof.

Finally, the high SNR upper bound in (3.7b) follows by noting that for each $P \geq 0$, we have,

$$\left\{\log \frac{1+P|h|^2}{1+P|h_e|^2}\right\}^+ \leq \left\{\log \frac{|h|^2}{|h_e|^2}\right\}^+.$$

3.4 Common Message

We establish, in order, the upper and lower capacity bounds in (3.8).

3.4.1 Upper Bound

To obtain our upper bound, suppose that we only need to transmit the message to receiver k. An upper bound on the secrecy capacity for this single-user channel is obtained via Proposition 3.

$$\bar{R}^+(P) \leq \max_{\substack{\rho(h_k): \\ E[\rho(h_k)] \leq P}} E\left[\left\{\log\left(\frac{1+|h_k|^2\rho(h_k)}{1+|h_e|^2\rho(h_k)}\right)\right\}^+\right], \tag{3.35}$$

and since k is arbitrary, we tighten the upper bound (3.35) by minimizing over k, yielding (3.9b).

3.4.2 Lower Bound

Next, we establish the lower bound (3.9a) by considering the following probabilistic extension of the parallel broadcast channel [29]. At each time, only one of the subchannels operates, and subchannel m is selected with a probability p_m, independent of the selection at all other times. Also, suppose that there is a total power constraint P on the input.

In this case, a straightforward extension of Proposition 2 provides the following achievable rate

$$\bar{R}_{K,M}(P) \triangleq \max \min_{k\in\{1,\ldots,K\}} \sum_{m=1}^{M} p_m \left\{I(u_m; y_{km}) - I(u_m; y_{em})\right\}^+, \tag{3.36}$$

where $u_1, u_2, \ldots, u_M$ are auxiliary random variables and the maximum is over the product distribution $p(u_1)p(u_2)\ldots p(u_M)$ and the stochastic mappings $x_m = f_m(u_m)$ that satisfy $\sum_{m=1}^{M} p_m E[x_m^2] \leq P$.

To simplify the exposition, we focus on the case of $K = 2$ receivers. The extension to $K > 2$ receivers is analogous and straightforward.

To start, we fix a threshold $T > 0$ and decompose the system into four states as shown in Fig. 3-3. The transmission takes place over a block of length n, and we

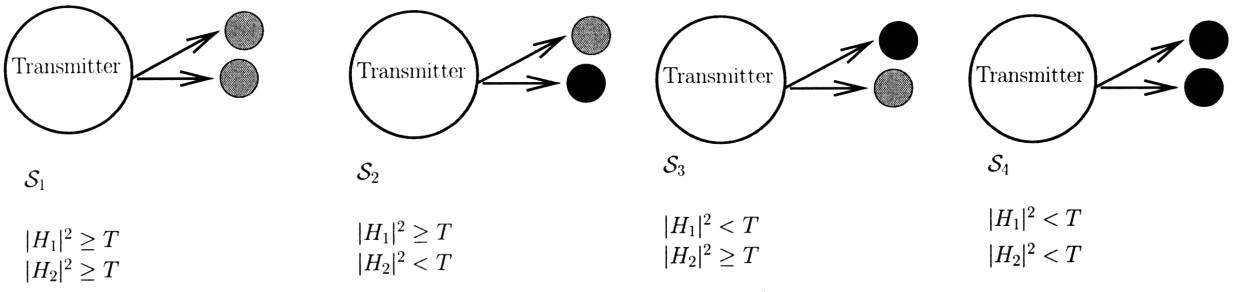


Figure 3-3: Decomposition of the system with $K = 2$ receivers into four states, as a function of their channel gains relative to a threshold T. The darkly and lightly shaded circles, respectively, indicate that a channel gain is, respectively, below and above the threshold.

classify $t = 1, 2, \ldots, n$ according to

$$\begin{aligned}
\mathcal{S}_1 &= \left\{t \in \{1,n\} \;\middle|\; |h_1(t)|^2 \geq T, |h_2(t)|^2 \geq T\right\} \\
\mathcal{S}_2 &= \left\{t \in \{1,n\} \;\middle|\; |h_1(t)|^2 \geq T, |h_2(t)|^2 < T\right\} \\
\mathcal{S}_3 &= \left\{t \in \{1,n\} \;\middle|\; |h_1(t)|^2 < T, |h_2(t)|^2 \geq T\right\} \\
\mathcal{S}_4 &= \left\{t \in \{1,n\} \;\middle|\; |h_1(t)|^2 < T, |h_2(t)|^2 < T\right\}.
\end{aligned} \tag{3.37}$$

The resulting channel is a probabilistic parallel channel with probabilities of the four channels are then given by

$$\begin{aligned}
p(\mathcal{S}_1) &= \Pr\left(|h_1|^2 \geq T, |h_2|^2 \geq T\right) \\
p(\mathcal{S}_2) &= \Pr\left(|h_1|^2 \geq T, |h_2|^2 < T\right) \\
p(\mathcal{S}_3) &= \Pr\left(|h_1|^2 < T, |h_2|^2 \geq T\right) \\
p(\mathcal{S}_4) &= \Pr\left(|h_1|^2 < T, |h_2|^2 < T\right).
\end{aligned}$$

In turn, with $x_m = u_m \sim \mathcal{CN}(0, P)$ in (3.36) the achievable rate expression is

$$\bar{R}^-(P) = \min_{k \in \{1,2\}} \left\{ \Pr(|h_k|^2 \geq T) E\left[\log\left(\frac{1 + |h_k|^2 P}{1 + |h_e|^2 P} \right) \;\middle|\; |h_k|^2 \geq T \right] \right\}. \tag{3.38}$$

Finally, optimizing (3.38) over the threshold, we obtain (3.9a) as follows (for the

case $K = 2$):

$$\begin{aligned}
\bar{R}^{-}(P) &= \max_{T>0} \min_{k\in\{1,2\}} \left\{ \Pr(|h_k|^2 \geq T) E\left[\log\left(\frac{1+|h_k|^2 P}{1+|h_e|^2 P} \right) \,\middle|\, |h_k|^2 \geq T \right] \right\} \\
&= \max_{T>0} \min_{k\in\{1,2\}} \left\{ \int_T^\infty \log\left(\frac{1+xP}{\exp\{E_{h_e}[\log(1+|h_e|^2 P)]\}} \right) p_k(x)\, dx \right\} \\
&\geq \min_{k\in\{1,2\}} \int_{T^*}^\infty \log\left(\frac{1+xP}{\exp\{E_{h_e}[\log(1+|h_e|^2 P)]\}} \right) p_k(x)\, dx \qquad (3.39) \\
&= \min_{k\in\{1,2\}} E_{h_k}\left[\left\{ \log\left(\frac{1+|h_k|^2 P}{\exp\{E_{h_e}[\log(1+|h_e|^2 P)]\}} \right) \right\}^+ \right], \qquad (3.40)
\end{aligned}$$

where T^* in (3.39) is obtained via

$$\log(1+T^*P) - E_{h_e}[\log(1+|h_e|^2 P] = 0.$$

For $K > 2$ receivers, we use the straightforward generalization of this scheme to a construction with 2^K states, where each state specifies the subset of receivers that are above the threshold T^*.

3.5 Independent Messages

In this section we establish the upper and lower bounds in Proposition 4.

3.5.1 Upper Bound

The upper bound is based on introducing a single-user genie-aided channel i.e., we consider the following channel with one receiver and one eavesdropper:

$$\begin{aligned}
y(t) &= h_{\max}(t)x(t) + z(t) \\
y_e(t) &= h_e(t)x(t) + z_e(t).
\end{aligned} \qquad (3.41)$$

Following the reasoning analogous to section 2.2.3), we note that the sum-capacity of the channel (3.1) is upper bounded by the secrecy capacity of the genie-aided-channel (3.41). Finally (3.12a) follows via the single user upper bound in Prop. 3, (see also [20]).

3.5.2 Lower Bound

The lower bound (3.12b) is achieved by a scheme that, at each time, transmits only to the receiver with the best instantaneous channel gain. Accordingly the sum rate is given by the achievable rate for the single user channel (3.41), and the expression (3.12b) follows via the lower bound in Prop. 3.

3.5.3 Scaling Laws

We now show that the upper and lower bounds on the sum secrecy capacity coincide as the number of users goes to infinity and obtain the capacity in Theorem 6.

Letting $\rho^*(h_{\max})$ denote the power allocation that maximizes $R_K^+(P)$ in (3.12a), we obtain

$$R_K^+(P) - R_K^-(P) \tag{3.42}$$

$$\leq E\left[\left\{\log\left(\frac{1+|h_{\max}|^2\rho^*(h_{\max})}{1+|h_e|^2\rho^*(h_{\max})}\right)\right\}^+\right] - E\left[\log\left(\frac{1+|h_{\max}|^2\rho(h_{\max})}{1+|h_e|^2\rho(h_{\max})}\right)\right] \tag{3.43}$$

$$= \Pr(|h_e|^2 \geq |h_{\max}|^2) E\left[\log\frac{1+|h_e|^2\rho^*(h_{\max})}{1+|h_{\max}|^2\rho^*(h_{\max})} \;\middle|\; |h_e|^2 \geq |h_{\max}|^2\right]$$

$$\leq \Pr(|h_e|^2 \geq |h_{\max}|^2) E\left[\log\frac{|h_e|^2}{|h_{\max}|^2} \;\middle|\; |h_e|^2 \geq |h_{\max}|^2\right] \tag{3.44}$$

$$\leq \frac{2\log 2}{K+1}, \tag{3.45}$$

where (6.50) follows from substituting the bounds in Proposition 4, where (3.44) follows from the fact that $\log((1+|h_e|^2a)/(1+|h_{\max}|^2a))$ is increasing in a for $|h_e|^2 \geq |h_{\max}|^2$, and where (3.45) follows from the fact that $\Pr(|h_e|^2 \geq |h_{\max}|^2) = 1/(1+K)$, since we assumed the channel coefficients to be i.i.d., and from the following "helper" lemma.

Lemma 3 *If $h_1, h_2, \ldots, h_K, h_e$ are i.i.d. unit-mean exponentials, then for $K \geq 2$ we have*

$$E\left[\log\frac{|h_e|^2}{|h_{\max}|^2} \;\middle|\; |h_e|^2 \geq |h_{\max}|^2\right] \leq 2\log 2 \tag{3.46}$$

A proof immediately follows.

Proof. [Proof of Lemma 3] First, we use the following:

Fact 5 ([12]) *Let $v_1, v_2, \ldots, v_K, v_{K+1}$ be i.i.d. exponentially distributed random variables with mean λ, and let $v_{\max}(K+1)$ and $v_{\max}(K)$ respectively denote the largest and second-largest of these random variables. Then the joint distribution of $(v_{\max}(K), v_{\max}(K+1))$ satisfies*

$$v_{\max}(K+1) = v_{\max}(K) + y, \tag{3.47}$$

where y is an exponentially distributed random variable with mean λ that is independent of $v_{\max}(K)$.

Proceeding, we have

$$E\left[\log\frac{|h_e|^2}{|h_{\max}|^2}\,\middle|\,|h_e|^2 \geq |h_{\max}|^2\right] \tag{3.48}$$

$$= E\left[\log\frac{|h_{\max}|^2+y}{|h_{\max}|^2}\right] \tag{3.49}$$

$$\leq E\left[\frac{y}{|h_{\max}|^2}\right] \tag{3.50}$$

$$= E[y]E\left[\frac{1}{|h_{\max}|^2}\right] \tag{3.51}$$

$$= E\left[\frac{1}{|h_{\max}|^2}\right], \tag{3.52}$$

where (3.50) follows from the identity $\log(1+x) \leq x$ for $x > 0$, where (3.51) follows from the independence of y and $h_{\max}$, and where (3.52) from the fact that $E[y] = 1$. Since $|h_{\max}|^2 \geq \max(|h_1|^2, |h_2|^2)$ we obtain

$$E\left[\frac{1}{|h_{\max}|^2}\right] \leq E\left[\frac{1}{\max(|h_1|^2, |h_2|^2)}\right] = 2\log 2,$$

whence (3.46) ∎

3.6 Conclusions

In this chapter we developed some techniques for secure communication over fading channels. The basic strategy was to map the fading channel into a set of parallel independent channels and then code across these channels. The transmission of a common message to several receivers requires us to simultaneously adapt the transmit power to multiple receivers, and this creates a tension if the receivers experience independent fading. It was shown that one can achieve a rate, independently of the number intended receiver, in this scenario — thus establishing that the secrecy capacity does not vanish with the number of receivers. For the case of a independent messages, we showed that an opportunistic transmission scheme achieves the sum-secrecy-capacity in the limit of large number of receivers.

Chapter 4

Multiple Antennas — MISOME Channel

In the present and the next chapter we study the gains from multiple antennas for confidentiality of data at the physical layer. As in the preceding chapters, these gains will be quantified within the framework of the wiretap channel. Multiple antennas have been an active area of research in the last decade or so. The primary goal has been to improve the throughput and reliability at the physical layer. In contrast we develop insights into the gains from multiple antennas for security at the physical layer.

In the present chapter, we restrict our attention to the case when the sender and eavesdropper have multiple antennas, but the intended receiver has a single antenna. We refer to this configuration as the multi-input, single-output, multi-eavesdropper (MISOME) case. It is worth emphasizing that the multiple eavesdropper antennas can correspond to a physical multiple-element antenna array at a single eavesdropper, a collection of geographically dispersed but perfectly colluding single-antenna eavesdroppers, or related variations. The MIMOME case will be treated in the next chapter. For the MISOME case, the secrecy capacity can be expressed in a closed form and it is analytically tractable. Hence it is worth treating this case separately from the general MIMOME case.

We first develop the secrecy capacity when the channel gains are fixed and known to all the terminals. Note that the multiple antenna wiretap channel is a non-degraded broadcast channel. A characterization of the secrecy capacity for non-degraded broadcast channels, when the channel alphabets are discrete and memoryless is provided in [8] as discussed in Chapter 1. However their characterization is not computable when the channel inputs are continuous valued, as is the case with multi-antenna channels. Our approach is to provide a new upper bound on the secrecy capacity for the wiretap channel and to show that this bound is in fact the true capacity. Our result thus indirectly establishes the optimum choice of auxiliary random variable in the secrecy capacity expression of [8].

While the capacity achieving scheme generally requires that the sender and the intended receiver have knowledge of the eavesdropper's channel (and thus number of antennas as well)—which is often not practical—we futher show that performance is

not strongly sensitive to this knowledge. Specifically, we show that a simple *masked beamforming* scheme described in [41, 18] that does not require knowledge of the eavesdropper's channel is close to optimal in the high SNR regime.

In addition, we examine the degree to which the eavesdropper can drive the secrecy capacity of the channel to zero, thereby effectively blocking secure communication between sender and (intended) receiver. In particular, for Rayleigh fading in the large antenna array limit, we use random matrix theory to characterize the secrecy capacity (and the rate achievable by masked beamforming) as a function of the ratio of the number of antennas at the eavesdropper to that at the sender. Among other results in this scenario, we show that 1) to defeat the security in the transmission it is sufficient for the eavesdropper to use at least twice as many antennas as the sender; and 2) an eavesdropper with significantly fewer antennas than the transmitter is not particularly effective.

Our results extend to the case of time-varying channels. We focus on the case of fast (ergodic, Rayleigh) fading, where the message is transmitted over a block that is long compared to the coherence time of the fading. In our model the state of the channel to the receiver is known by all three parties (sender, receiver, and eavesdropper), but the state of the channel to the eavesdropper is known only to the eavesdropper. Using techniques from the previous chapter, we develop upper and lower bounds on the secrecy capacity both for finitely many antennas and in the large antenna limit.

4.1 Preliminaries: Generalized Eigenvalues

Many of our results arise out of generalized eigenvalue analysis. We summarize the properties of generalized eigenvalues and eigenvectors we require in the sequel. For more extensive developments of the topic, see, e.g., [19, 1].

Definition 9 (Generalized eigenvalues) *For a Hermitian matrix* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *and positive definite matrix* $\mathbf{B} \in \mathbb{C}^{n\times n}$, *we refer to* $(\lambda, \boldsymbol{\psi})$ *as a generalized eigenvalue-eigenvector pair of* $(\mathbf{A}, \mathbf{B})$ *if* $(\lambda, \boldsymbol{\psi})$ *satisfy*

$$\mathbf{A}\boldsymbol{\psi} = \lambda \mathbf{B}\boldsymbol{\psi}. \tag{4.1}$$

Since $\mathbf{B}$ in Definition 9 is invertible, first note that generalized eigenvalues and eigenvectors can be readily expressed in terms of regular ones. Specifically,

Fact 6 *The generalized eigenvalues and eigenvectors of the pair* $(\mathbf{A}, \mathbf{B})$ *are the regular eigenvalues and eigenvectors of the matrix* $\mathbf{B}^{-1}\mathbf{A}$.

Other characterizations reveal more useful properties for our development. For example, we have the following:

Fact 7 (Variational Characterization) *The generalized eigenvectors of* $(\mathbf{A}, \mathbf{B})$ *are the stationary point solution to a particular Rayleigh quotient. Specifically, the largest*

generalized eigenvalue is the maximum of the Rayleigh quotient[1]

$$\lambda_{\max}(\mathbf{A}, \mathbf{B}) = \max_{\psi \in \mathbb{C}^n} \frac{\psi^\dagger \mathbf{A} \psi}{\psi^\dagger \mathbf{B} \psi}, \tag{4.2}$$

and the optimum is attained by the eigenvector corresponding to $\lambda_{\max}(\mathbf{A}, \mathbf{B})$.

The case when $\mathbf{A}$ has rank one is of special interest to us. In this case, the generalized eigenvalue admits a particularly simple expression:

Fact 8 (Quadratic Form) *When* $\mathbf{A}$ *in Definition 9 has rank one, i.e.,* $\mathbf{A} = \mathbf{a}\mathbf{a}^\dagger$ *for some* $\mathbf{a} \in \mathbb{C}^n$*, then*

$$\lambda_{\max}(\mathbf{a}\mathbf{a}^\dagger, \mathbf{B}) = \mathbf{a}^\dagger \mathbf{B}^{-1} \mathbf{a}. \tag{4.3}$$

4.2 Channel and System Model

The MISOME channel and system model is as follows. We use n_t and n_e to denote the number of sender and eavesdropper antennas, respectively; the (intended) receiver has a single antenna. The signals observed at the receiver and eavesdropper, respectively, are, for $t = 1, 2, \ldots$,

$$\begin{aligned} y_r(t) &= \mathbf{h}_r^\dagger \mathbf{x}(t) + z_r(t) \\ \mathbf{y}_e(t) &= \mathbf{H}_e \mathbf{x}(t) + \mathbf{z}_e(t), \end{aligned} \tag{4.4}$$

where $\mathbf{x}(t) \in \mathbb{C}^{n_t}$ is the transmitted signal vector, $\mathbf{h}_r \in \mathbb{C}^{n_t}$ and $\mathbf{H}_e \in \mathbb{C}^{n_e \times n_t}$ are complex channel gains, and $z_r(t)$ and $\mathbf{z}_e(t)$ are independent identically-distributed (i.i.d.) circularly-symmetric complex-valued Gaussian noises: $z_r(t) \sim \mathcal{CN}(0, 1)$ and $\mathbf{z}_e(t) \sim \mathcal{CN}(0, \mathbf{I})$. Moreover, the noises are independent, and the input satisfies an average power constraint of P, i.e.,

$$E\left[\frac{1}{n}\sum_{t=1}^{n} \|\mathbf{x}(t)\|^2\right] \leq P. \tag{4.5}$$

Finally, except when otherwise indicated, all channel gains are fixed throughout the entire transmission period, and are known to all the terminals. Communication takes place at a rate R in bits per channel use over a transmission interval of length n. Specifically, a $(2^{nR}, n)$ code for the channel consists of a message w uniformly distributed over the index set $\mathcal{W}_n = \{1, 2, \ldots, 2^{nR}\}$, an encoder $\mu_n : \mathcal{W}_n \to \mathbb{C}^{n_t \times n}$ that maps the message w to the transmitted (vector) sequence $\{\mathbf{x}(t)\}_{t=1}^n$, and a decoding function $\nu_n : \mathbb{C}^n \to \mathcal{W}_n$ that maps the received sequence $\{y_r(t)\}_{t=1}^n$ to a message estimate $\hat{w}$. The error event is $\mathcal{E}_n = \{\nu_n(\mu_n(w)) \neq w\}$, and the amount of information obtained by the eavesdropper from the transmission is measured via the equivocation $I(w; \mathbf{y}_e^n)$.

[1]Throughout the paper we use $\lambda_{\max}$ to denote the largest eigenvalue. Whether this is a regular or generalized eigenvalue will be clear from context, and when there is a need to be explicit, the relevant matrix or matrices will be indicated as arguments.

Definition 10 (Secrecy Capacity) *A secrecy rate R is achievable if there exists a sequence of $(2^{nR}, n)$ codes such that $\Pr(\mathcal{E}_n) \to 0$ and $I(w; \mathbf{y}_e^n)/n \to 0$ as $n \to \infty$. The* secrecy capacity *is the supremum of all achievable secrecy-rates.*

4.3 Main Results

The MISOME wiretap channel is a nondegraded broadcast channel. In Csiszár and Körner [8], the secrecy capacity of the nondegraded discrete memoryless broadcast channel $p_{y_r, \mathbf{y}_e|x}$ is expressed in the form

$$C = \max_{p_u, p_{x|u}} I(u; y_r) - I(u; y_e), \tag{4.6}$$

where u is an auxiliary random variable over a certain alphabet that satisfies the Markov relation $u \leftrightarrow x \leftrightarrow (y_r, y_e)$. Moreover, the secrecy capacity (5.6) readily extends to the continuous alphabet case with a power constraint, so it also gives a characterization of the MISOME channel capacity.

Rather than attempting to solve for the optimal choice of u and $p_{x|u}$ in (5.6) directly to evaluate this capacity,[2] we consider an indirect approach based on a useful upper bound as the converse, which we describe next.

4.3.1 Upper Bound on Achievable Rates

A key result is the following upper bound, which we derive in Section 6.4.

Theorem 7 *An upper bound on the secrecy capacity for the MISOME channel model is*

$$R_+ = \min_{\mathbf{K}_\phi \in \mathcal{K}_\phi} \max_{\mathbf{K}_P \in \mathcal{K}_P} R_+(\mathbf{K}_P, \mathbf{K}_\phi), \tag{4.7}$$

where $R_+(\mathbf{K}_P, \mathbf{K}_\phi) = I(\mathbf{x}; y_r|\mathbf{y}_e)$ with $\mathbf{x} \sim \mathcal{CN}(0, \mathbf{K}_P)$ and

$$\mathcal{K}_P \triangleq \left\{ \mathbf{K}_P \;\middle|\; \mathbf{K}_P \succeq 0, \quad \mathrm{tr}(\mathbf{K}_P) \leq P \right\}, \tag{4.8}$$

and where

$$\begin{bmatrix} z_r \\ \mathbf{z}_e \end{bmatrix} \sim \mathcal{CN}(0, \mathbf{K}_\phi) \tag{4.9}$$

[2]The direct approach is explored in, e.g., [30] and [45], where the difficulty of performing this optimization is reported even when restricting $p_{x|u}$ to be singular (a deterministic mapping) and/or the input distribution to be Gaussian.

with

$$\begin{aligned}\mathcal{K}_{\boldsymbol{\phi}} &\triangleq \left\{ \mathbf{K}_{\boldsymbol{\phi}} \;\middle|\; \mathbf{K}_{\boldsymbol{\phi}} = \begin{bmatrix} 1 & \boldsymbol{\phi}^\dagger \\ \boldsymbol{\phi} & \mathbf{I} \end{bmatrix}, \quad \mathbf{K}_{\boldsymbol{\phi}} \succeq 0 \right\} \\ &= \left\{ \mathbf{K}_{\boldsymbol{\phi}} \;\middle|\; \mathbf{K}_{\boldsymbol{\phi}} = \begin{bmatrix} 1 & \boldsymbol{\phi}^\dagger \\ \boldsymbol{\phi} & \mathbf{I} \end{bmatrix}, \quad \|\boldsymbol{\phi}\| \leq 1 \right\}.\end{aligned} \tag{4.10}$$

To obtain this bound, we consider a genie-aided channel in which the eavesdropper observes $\mathbf{y}_\mathrm{e}$ but the receiver observes *both* y_r and $\mathbf{y}_\mathrm{e}$. Such a channel clearly has a capacity larger than the original channel. Moreover, since it is a degraded broadcast channel, the secrecy capacity of the genie-aided channel can be easily derived and is given by (cf. [53]) $\max I(\mathbf{x}; y_\mathrm{r}|\mathbf{y}_\mathrm{e})$ where the maximum is over the choice of input distributions. As we will see, it is straightforward to establish that the maximizing input distribution is Gaussian (in contrast to the original channel).

Next, while the secrecy capacity of the original channel depends only on the marginal distributions $p_{y_\mathrm{r}|\mathbf{x}}$ and $p_{\mathbf{y}_\mathrm{e}|\mathbf{x}}$ (see, e.g., [8]), mutual information $I(\mathbf{x}; y_\mathrm{r}|\mathbf{y}_\mathrm{e})$ for the genie-aided channel depends on the joint distribution $p_{y_\mathrm{r},\mathbf{y}_\mathrm{e}|\mathbf{x}}$. Accordingly we obtain the tightest such upper bound by finding the joint distribution (having the required marginal distributions), whence (4.7).

The optimization (4.7) can be carried out analytically, yielding an explicit expression, as we now develop.

4.3.2 MISOME Secrecy Capacity

The upper bound described in the preceding section is achievable, yielding the MISOME channel capacity. Specifically, we have the following theorem, which we prove in Section 4.5.1.

Theorem 8 *The secrecy capacity of the channel* (4.4) *is*

$$C(P) = \left\{ \log \lambda_{\max} \left(\mathbf{I} + P\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}^\dagger, \mathbf{I} + P\mathbf{H}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e} \right) \right\}^+, \tag{4.11}$$

with $\lambda_{\max}$ *denoting the largest generalized eigenvalue of its argument pair. Furthermore, the capacity is obtained by beamforming (i.e., signaling with rank one covariance) along the direction* $\boldsymbol{\psi}_{\max}$ *of the*[3] *generalized eigenvector corresponding to* $\lambda_{\max}$ *with an encoding of the message using a code for the scalar Gaussian wiretap channel.*

We emphasize that the beamforming direction in Theorem 8 for achieving capacity will in general depend on all of the target receiver's channel $\mathbf{h}_\mathrm{r}$, the eavesdropper's channel $\mathbf{H}_\mathrm{e}$, and the SNR (P).

In the high SNR regime, the MISOME capacity (4.11) exhibits one of two possible behaviors, corresponding to whether

$$\lim_{P\to\infty} C(P) = \left\{ \log \lambda_{\max} \left(\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}^\dagger, \mathbf{H}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e} \right) \right\}^+, \tag{4.12}$$

[3]If there is more than one generalized eigenvector for $\lambda_{\max}$, we choose any one of them.

is finite or infinite, which depends on whether or not $\mathbf{h}_r$ has a component in the null space of $\mathbf{H}_e$. Specifically, we have the following corollary, which we prove in Section 4.5.2.

Corollary 4 *The high SNR asymptote of the secrecy capacity* (4.11) *takes the form*

$$\lim_{P\to\infty} C(P) = \{\log \lambda_{\max}(\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{H}_e^\dagger\mathbf{H}_e)\}^+ < \infty \quad \textit{if } \mathbf{H}_e^\perp\mathbf{h}_r = \mathbf{0}, \tag{4.13a}$$

$$\lim_{P\to\infty} [C(P) - \log P] = \log \|\mathbf{H}_e^\perp\mathbf{h}_r\|^2 \quad \textit{if } \mathbf{H}_e^\perp\mathbf{h}_r \neq \mathbf{0}, \tag{4.13b}$$

where $\mathbf{H}_e^\perp$ *denotes the projection matrix onto the null space of* $\mathbf{H}_e$.[4]

This behavior can be understood rather intuitively. In particular, when $\mathbf{H}_e^\perp\mathbf{h}_r = \mathbf{0}$, as is typically the case when the eavesdropper uses enough antennas ($n_e \geq n_t$) or the intended receiver has an otherwise unfortunate channel, the secrecy capacity is SNR-limited. In essence, while more transmit power is advantageous to communication to the intended receiver, it is also advantageous to the eavesdropper, resulting in diminishing returns.

By contrast, when $\mathbf{H}_e^\perp\mathbf{h}_r \neq \mathbf{0}$, as is typically the case when, e.g., the eavesdropper uses insufficiently many antennas ($n_e < n_t$) unless the eavesdropper has an otherwise unfortunate channel, the transmitter is able to steer a null to the eavesdropper without simultaneously nulling the receiver and thus capacity grows by 1 b/s/Hz with every 3 dB increase in transmit power as it would if there were no eavesdropper to contend with.

The MISOME capacity (4.11) is also readily specialized to the low SNR regime, as we develop in Section 4.5.3, and takes the following form.

Corollary 5 *The low SNR asymptote of the secrecy capacity is*

$$\lim_{P\to 0} \frac{C(P)}{P} = \frac{1}{\ln 2}\{\lambda_{\max}(\mathbf{h}_r\mathbf{h}_r^\dagger - \mathbf{H}_e^\dagger\mathbf{H}_e)\}^+. \tag{4.14}$$

In this low SNR regime, the direction of optimal beamforming vector approaches the (regular) eigenvector corresponding to the largest (regular) eigenvalue of $\mathbf{h}_r\mathbf{h}_r^\dagger - \mathbf{H}_e^\dagger\mathbf{H}_e$. Note that the optimal direction is in general not along $\mathbf{h}_r$. Thus, ignoring the eavesdropper is in general not an optimal strategy even at low SNR.

4.3.3 Eavesdropper-Ignorant Coding: Masked Beamforming

In our basic model the channel gains are fixed and known to all the terminals. Our capacity-achieving scheme in Theorem 8 uses the knowledge of $\mathbf{H}_e$ for selecting the beamforming direction. However, in many applications it may be difficult to know the eavesdropper's channel. Accordingly, in this section we analyze a simple alternative scheme that uses only knowledge of $\mathbf{h}_r$ in choosing the transmit directions, yet achieves near-optimal performance in the high SNR regime.

[4]That is, the columns of $\mathbf{H}_e^\perp$ constitute an orthogonal basis for the null space of $\mathbf{H}_e$.

The scheme we analyze is a masked beamforming scheme described in [41, 18]. In this scheme, the transmitter signals isotropically (i.e., with a covariance that is a scaled identity matrix), and as such can be naturally viewed as a "secure space-time code." More specifically, it simultaneously transmits the message (encoded using a scalar Gaussian wiretap code) in the direction corresponding to the intended receiver's channel $\mathbf{h}_{\mathrm{r}}$ while transmitting synthesized spatio-temporal white noise in the orthogonal subspace (i.e., all other directions).

The performance of masked beamforming is given by the following proposition, which is proved in Section 4.6.1.

Proposition 5 (Masked Beamforming Secrecy Rate) *A rate achievable by the masked beamforming scheme for the MISOME channel is*

$$R_{\mathrm{MB}}(P) = \left\{ \log \lambda_{\max} \left(\frac{P}{n_{\mathrm{t}}} \mathbf{h}_{\mathrm{r}} \mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I} + \frac{P}{n_{\mathrm{t}}} \mathbf{H}_{\mathrm{e}}^{\dagger} \mathbf{H}_{\mathrm{e}} \right) + \log \left(1 + \frac{n_{\mathrm{t}}}{P \|\mathbf{h}_{\mathrm{r}}\|^2} \right) \right\}^{+}. \tag{4.15}$$

The rate (4.15) is, in general, suboptimal. We characterize the loss with respect to the capacity achieving scheme below.

Theorem 9 *The rate $R_{\mathrm{MB}}(P)$ achievable by masked beamforming scheme for the MISOME case [cf. (4.15)] satisfies*

$$\lim_{P \to \infty} \left[C\left(\frac{P}{n_{\mathrm{t}}} \right) - R_{\mathrm{MB}}(P) \right] = 0. \tag{4.16}$$

From the relation in (4.16) we note that, in the high SNR regime, the masked beamforming scheme achieves a rate of $C(P/n_{\mathrm{t}})$, where n_{t} is the number of transmit antennas. Combining (4.16) with (4.13), we see that the asymptotic masked beamforming loss is at most $\log n_{\mathrm{t}}$ b/s/Hz, or equivalently $10 \log_{10} n_{\mathrm{t}}$ dB in SNR. Specifically,

$$\lim_{P \to \infty} \left[C(P) - R_{\mathrm{MB}}(P) \right] = \begin{cases} \log n_{\mathrm{t}}, & \mathbf{H}_{\mathrm{e}}^{\perp} \mathbf{h}_{\mathrm{r}} \neq \mathbf{0} \\ 0, & \mathbf{H}_{\mathrm{e}}^{\perp} \mathbf{h}_{\mathrm{r}} = \mathbf{0}. \end{cases} \tag{4.17}$$

That at least some loss (if vanishing) is associated with the masked beamforming scheme is expected, since the capacity-achieving scheme performs beamforming to concentrate the transmission along the optimal direction, whereas the masked beamforming scheme uses isotropic inputs.

As one final comment, note that although the covariance structure of the masked beamforming transmission does not depend on the eavesdropper's channel, the rate of the base (scalar Gaussian wiretap) code does, as (4.15) reflects. In practice, the selection of this rate determines an insecurity zone around the sender, whereby the transmission is secure from eavesdroppers outside this zone, but insecure from ones inside.

4.3.4 Example

In this section, we illustrate the preceding results for a typical MISOME channel. In our example, there are $n_t = 2$ transmit antennas, and $n_e = 2$ eavesdropper antennas. The channel to the receiver is

$$\mathbf{h}_r = \begin{bmatrix} 0.0991 + j0.8676 & 1.0814 - j1.1281 \end{bmatrix}^T,$$

while the channel to the eavesdropper is

$$\mathbf{H}_{e,1} = \begin{bmatrix} 0.3880 + j1.2024 & -0.9825 + j0.5914 \\ 0.4709 - j0.3073 & 0.6815 - j0.2125 \end{bmatrix}, \tag{4.18}$$

where $j = \sqrt{-1}$.

Fig. 4-1 depicts communication rate as a function of SNR. The upper and lower solid curves depict the secrecy capacity (4.11) when the eavesdropper is using one or both its antennas, respectively.[5] As the curves reflect, when the eavesdropper has only a single antenna, the transmitter can securely communicate at any desired rate to its intended receiver by using enough power. However, by using both its antennas, the eavesdropper caps the rate at which the transmitter can communicate securely regardless of how much power it has available. Note that the lower and upper curves are representative of the cases where $\mathbf{H}_e^{\perp}\mathbf{h}_r$ is, and is not $\mathbf{0}$, respectively.

Fig. 4-1 also shows other curves of interest. In particular, using dotted curves we superimpose the secrecy capacity high-SNR asymptotes as given by (4.13). As is apparent, these asymptotes can be quite accurate approximations even for moderate values of SNR. Finally, using dashed curves we show the rate (4.15) achievable by the masked beamforming coding scheme, which doesn't use knowledge of the eavesdropper channel. Consistent with (4.17), the loss in performance at high SNR approaches 3 dB when the eavesdropper uses only one of its antennas, and 0 dB when it uses both. Again, these are good estimates of the performance loss even at moderate SNR. Thus the penalty for ignorance of the eavesdropper's channel can be quite small in practice.

4.3.5 Scaling Laws in the Large System Limit

Our analysis in Section 4.3.2 of the scaling behavior of capacity with SNR in the high SNR limit with a fixed number of antennas in the system yielded several useful insights into secure space-time coding systems. In this section, we develop equally valuable insights from a complementary scaling. In particular, we consider the scaling behavior of capacity with the number of antennas in the large system limit at a fixed SNR.

One convenient feature of such analysis is that for many large ensembles of channel gains, almost all randomly drawn realizations produce the same capacity asymptotes. For our analysis, we restrict our attention to an ensemble corresponding to Rayleigh

[5] When a single eavesdropper antenna is in use, the relevant channel corresponds to the first row of (4.18).

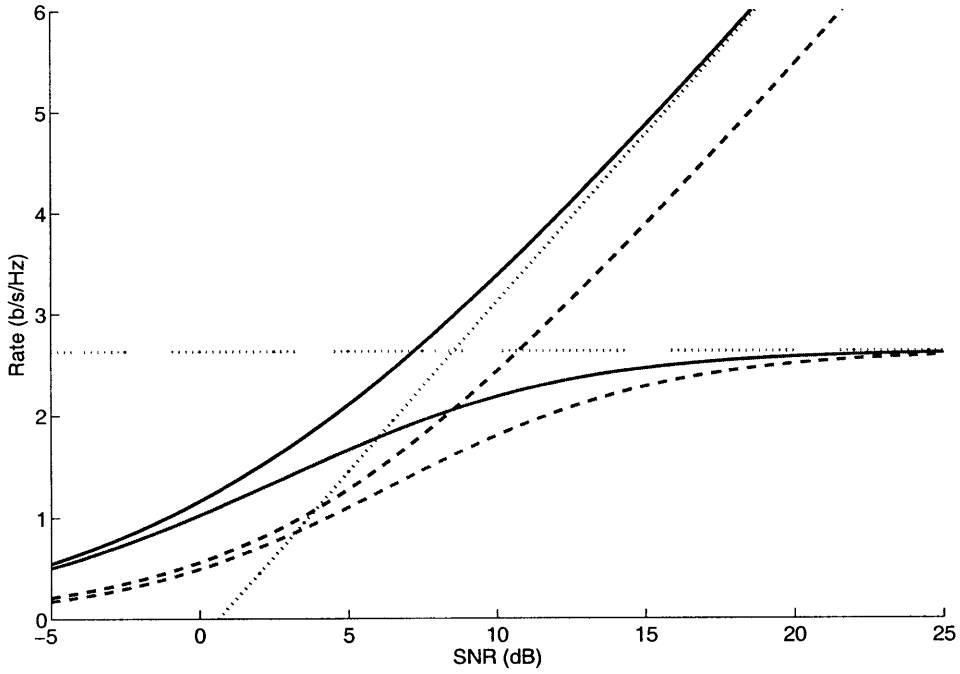


Figure 4-1: Performance over an example MISOME channel with $n_t = 2$ transmit antennas. The successively lower solid curves give the secrecy capacity for $n_e = 1$ and $n_e = 2$ eavesdropper antennas, respectively and the dotted curves indicat the corresponding high-SNR asymptote. The dashed curves give the corresponding rates achievable by masked beamforming, which does not require the transmitter to have knowledge of the eavesdropper's channel.

fading in which $\mathbf{h}_r$ and $\mathbf{H}_e$ are independent, and each has i.i.d. $\mathcal{CN}(0,1)$ entries. The realization from the ensemble is known to all terminals prior to communication.

In anticipation of our analysis, we make the dependency of secrecy rates on the number of transmit and eavesdropper antennas explicit in our notation (but leave the dependency on the realization of $\mathbf{h}_r$ and $\mathbf{H}_e$ implicit). Specifically, we now use $C(P, n_t, n_e)$ to denote the secrecy capacity, and $R_{MB}(P, n_t, n_e)$ to denote the rate of the masked beamforming scheme. With this notation, the scaled rates of interest are

$$\tilde{C}(\gamma,\beta) = \lim_{n_t\to\infty} C\left(P{=}\gamma/n_t, n_t, n_e{=}\beta n_t\right), \tag{4.19a}$$

and

$$\tilde{R}_{MB}(\gamma,\beta) = \lim_{n_t\to\infty} R_{MB}(P{=}\gamma, n_t, n_e{=}\beta n_t). \tag{4.19b}$$

Our choice of scalings ensures that the $\tilde{C}(\gamma,\beta)$ and $\tilde{R}_{MB}(\gamma,\beta)$ are not degenerate. In particular, note that the capacity scaling (4.19a) involves an SNR normalization. In particular, the transmitted power P is reduced as the number of transmitter antennas n_t grows so as to keep the *received* SNR remains fixed (at specified value γ) independent of n_t. However, the scaling (4.19b) is not SNR normalized in this way. This is because the masked beamforming already suffers a nominal factor of n_t SNR loss [cf. (4.16)] relative to a capacity-achieving system.

In what follows, we do not attempt an exact evaluation of the secrecy rates for our chosen scalings. Rather we find compact lower and upper bounds that are tight in the high SNR limit.

We begin with our lower bound, which is derived in Section 4.7.2.

Theorem 10 (Scaling Laws) *The asymptotic secrecy capacity satisfies*

$$\tilde{C}(\gamma,\beta) \overset{a.s.}{\geq} \{\log \xi(\gamma,\beta)\}^+, \tag{4.20}$$

where

$$\xi(\gamma,\beta) = \gamma - \frac{1}{4}\left[\sqrt{1{+}\gamma\left(1{+}\sqrt{\beta}\right)^2} - \sqrt{1{+}\gamma\left(1{-}\sqrt{\beta}\right)^2}\right]^2. \tag{4.21}$$

Furthermore, the same bound holds for the corresponding asymptotic masked beamforming rate, i.e.,

$$\tilde{R}_{MB}(\gamma,\beta) \overset{a.s.}{\geq} \{\log \xi(\gamma,\beta)\}^+. \tag{4.22}$$

Since the secrecy rates increase monotonically with SNR, the infinite-SNR rates constitute a useful upper bound. As derived in Section 4.7.3, this bound is as follows.

Theorem 11 *The asymptotic secrecy capacity satisfies*

$$\begin{aligned}\tilde{C}(\gamma,\beta) &\leq \lim_{n_t\to\infty}\lim_{P\to\infty} C(P,n_t,\beta n_t)\\ &\overset{a.s.}{=} \tilde{C}(\infty,\beta) \triangleq \begin{cases} 0 & \beta \geq 2 \\ -\log(\beta-1) & 1<\beta<2 \\ \infty & \beta \leq 1. \end{cases}\end{aligned} \tag{4.23}$$

Furthermore, the right hand side of (4.23) *is also an upper bound on* $\tilde{R}_{\mathrm{MB}}(\gamma,\beta)$*, i.e.,*

$$\begin{aligned}\tilde{R}_{\mathrm{MB}}(\gamma,\beta) &\leq \lim_{n_t\to\infty}\lim_{P\to\infty} R_{\mathrm{MB}}(P,n_t,\beta n_t)\\ &\overset{a.s.}{=} \tilde{C}(\infty,\beta)\end{aligned} \tag{4.24}$$

Note that it is straightforward to verify that the lower bound (4.20) is tight at high SNR, i.e., that, for all β,

$$\{\log \xi(\infty,\beta)\}^+ = \tilde{C}(\infty,\beta). \tag{4.25}$$

The same argment confirms the corresponding behavior for masked beamforming.

Our lower and upper bounds of Theorem 10 and Theorem 11, respectively, are depicted in Fig. 4-2. In particular, we plot rate as a function of the antenna ratio β for various values of the SNR γ.

As Fig. 4-2 reflects, there are essentially three main regions of behavior, the boundaries between which are increasingly sharp with increasing SNR. First, for $\beta < 1$ the eavesdropper has proportionally fewer antennas than the sender, and thus is effectively thrwarted. It is in this regime that the transmitter can steer a null to the eavesdropper and achieve any desired rate to the receiver by using enough power.

Second, for $1 \leq \beta < 2$ the eavesdropper has proportionally more antennas than the sender, and thus can cap the secure rate achievable to the receiver regardless of how much power the transmitter has available. For instance, when the transmitter has 50% more antennas than the eavesdropper ($\beta = 1.5$), the sender is constrained to a maximum secure rate no more than 1 b/s/Hz. Moreover, if the sender is sufficiently limited in power that the received SNR is at most, say, 10 dB, the maximum rate is less than 1/2 b/s/Hz.

We emphasize that these results imply the eavesdropper is at a substantial disadvantage compared to the intended receiver when the number of tranmitter antennas is chosen to be large. Indeed, the intended receiver needs only a single antenna to decode the message, while the eavesdropper needs a large number of antennas to constrain the transmission.

Finally, for $\beta \geq 2$ the eavesdropper is able to entirely prevent secure communication (drive the secrecy capacity to zero) even if the transmitter has unlimited power available. Useful intuition for this phenomenon is obtained from consideration of the masked beamforming scheme, in which the sender transmits the signal of interest in the direction of $\mathbf{h}_r$ and synthesized noise in the $n_t - 1$ directions orthogonal to

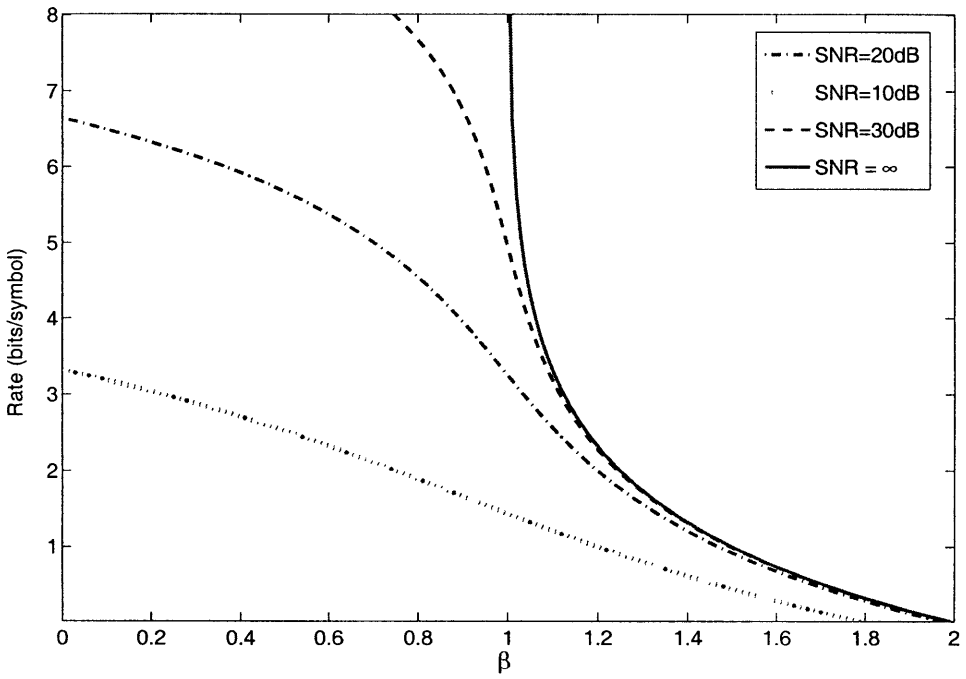


Figure 4-2: Secrecy capacity bounds in the large system limit. The solid red curve is the high SNR secrecy capacity, which is an upper bound on the for finite SNR. The progressively lower dashed curves are lower bounds on the asymptotic secrecy capacity (and masked beamforming secrecy rate). The channel realizations are fixed but drawn at random according to Gaussian distribution.

$\mathbf{h}_r$. With such a transmission, the intended receiver experiences a channel gain of $\|\mathbf{h}_r\|^2 P/n_t$. In the high SNR regime, the eavesdropper must cancel the synthesized noise, which requires at least $n_t - 1$ receive antennas. Moreover, after canceling the noise it must have the "beamforming gain" of n_t so its channel quality is of the same order as that of the intended receiver. This requires having at least n_t more antennas. Thus at least $2n_t - 1$ antennas are required by the eavesdropper to guarantee successful interception of the transmission irrespective of the power used, which corresponds to $\beta \geq 2$ as $n_t \to \infty$.

4.3.6 Capacity Bounds in Fading

Thus far we have focused on the scenarios where the receiver and eavesdropper channels are fixed for the duration n of the message transmission. In this section, we briefly turn our attention to the case of time-varying channels—specifically, the case of fast fading where there are many channel fluctuations during the course of transmission. In particular, we consider a model in which $\mathbf{h}_r(t)$ and $\mathbf{H}_e(t)$ are temporally and spatially i.i.d. sequences that are independent of one another and have $\mathcal{CN}(0,1)$ elements, corresponding to Rayleigh fading.

In our model, $\mathbf{h}_r(t)$ is known (in a causal manner) to all the three terminals, but only the eavesdropper has knowledge of $\mathbf{H}_e(t)$. Accordingly, the channel model is, for $t = 1, 2, \ldots,$

$$\begin{aligned} y_r(t) &= \mathbf{h}_r^\dagger(t)\mathbf{x}(t) + z_r(t) \\ \mathbf{y}_e(t) &= \mathbf{H}_e(t)\mathbf{x}(t) + \mathbf{z}_e(t). \end{aligned} \tag{4.26}$$

The definition of the secrecy rate and capacity is as in Definition 10, with the exception that the equivocation $I(w; \mathbf{y}_e^n)$ is replaced with $I(w; \mathbf{y}_e^n, \mathbf{H}_e^n | \mathbf{h}_r^n)$, which takes into account the channel state information at the different terminals.

For this model we have the following nontrivial upper and lower bounds on the secrecy capacity, which are developed in Section 4.8. The upper bound is developed via the same genie-aided channel analysis used in the proof of Theorem 8, but with modifications to account for the presence of fading. The lower bound is achieved by the adaptive version of masked beamforming described in [41].

Theorem 12 *The secrecy capacity for the MISOME fast fading channel* (4.26) *is bounded by*

$$C_{\mathrm{FF}}(P, n_t, n_e) \geq \max_{\rho(\cdot) \in \mathcal{P}_{\mathrm{FF}}} E\left[R_{\mathrm{FF},-}(\mathbf{h}_r, \mathbf{H}_e, \rho(\cdot))\right], \tag{4.27a}$$

$$C_{\mathrm{FF}}(P, n_t, n_e) \leq \max_{\rho(\cdot) \in \mathcal{P}_{\mathrm{FF}}} E\left[R_{\mathrm{FF},+}(\mathbf{h}_r, \mathbf{H}_e, \rho(\cdot))\right], \tag{4.27b}$$

where $\mathcal{P}_{\mathrm{FF}}$ *is the set of all valid power allocations, i.e.,*

$$\mathcal{P}_{\mathrm{FF}} = \left\{\rho(\cdot) \;\middle|\; \rho(\cdot) \geq 0,\ E[\rho(\mathbf{h}_r)] \leq P\right\}, \tag{4.28}$$

and

$$R_{\mathrm{FF},-}(\mathbf{h}_{\mathrm{r}},\mathbf{H}_{\mathrm{e}},\rho(\cdot)) \triangleq \log\left(\frac{\rho(\mathbf{h}_{\mathrm{r}})}{n_{\mathrm{t}}}\mathbf{h}_{\mathrm{r}}^{\dagger}\left[\mathbf{I}+\frac{\rho(\mathbf{h}_{\mathrm{r}})}{n_{\mathrm{t}}}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}\right]^{-1}\mathbf{h}_{\mathrm{r}}\right)+\log\left(1+\frac{n_{\mathrm{t}}}{\rho(\mathbf{h}_{\mathrm{r}})\|\mathbf{h}_{\mathrm{r}}\|^2}\right). \tag{4.29a}$$

$$R_{\mathrm{FF},+}(\mathbf{h}_{\mathrm{r}},\mathbf{H}_{\mathrm{e}},\rho(\cdot)) \triangleq \left\{\log\lambda_{\max}(\mathbf{I}+\rho(\mathbf{h}_{\mathrm{r}})\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger},\mathbf{I}+\rho(\mathbf{h}_{\mathrm{r}})\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})\right\}^{+}, \tag{4.29b}$$

In general, our upper and lower bounds do not coincide. Indeed, even in the case of single antennas at all terminals ($n_{\mathrm{t}} = n_{\mathrm{e}} = 1$), the secrecy capacity for the fading channel is unknown, except in the case of large coherence period [20].

However, based on our scaling analysis in Section 4.3.5, there is one regime in which the capacity can be calculated: in the limit of both high SNR and a large system. Indeed, since (4.22) and (4.23) hold for almost every channel realization, we have the following proposition, whose proof is provided in Section 4.8.3.

Proposition 6 *The secrecy capacity of the fast fading channel satisfies*

$$\lim_{n_{\mathrm{t}}\to\infty} C_{\mathrm{FF}}(P=\gamma, n_{\mathrm{t}}, n_{\mathrm{e}}=\beta n_{\mathrm{t}}) \geq \left\{\log\xi(\gamma,\beta)\right\}^{+}, \tag{4.30}$$

where $\xi(\cdot,\cdot)$ is as defined in (4.21)*, and*

$$\lim_{n_{\mathrm{t}}\to\infty} C_{\mathrm{FF}}(P=\gamma, n_{\mathrm{t}}, n_{\mathrm{e}}=\beta n_{\mathrm{t}}) \leq \tilde{C}(\infty,\beta) \tag{4.31}$$

with the $\tilde{C}(\infty,\beta)$ as given in (4.23)*.*

Finally, via (4.25) we see that (4.30) and (4.31) converge as $\gamma \to \infty$.

This concludes our statement of the main results. The following sections are devoted to the proofs of these results and some further discussion.

4.4 Upper Bound Derivation

In this section we prove Theorem 7. We begin with the following lemma, which establishes that the capacity of genie-aided channel is an upper bound on the channel of interest. A proof is provided in Appendix B, and closely follows the general converse of Wyner [53], but differs in that the latter was for discrete channels and thus did not incorporate a power constraint.

Lemma 4 *An upper bound on the secrecy capacity of the MISOME wiretap channel is*

$$C \leq \max_{p_{\mathbf{x}}\in\mathcal{P}} I(\mathbf{x}; y_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}}), \tag{4.32}$$

where $\mathcal{P}$ is the set of all probability distributions that satisfy $E[\|\mathbf{x}\|^2] \leq P$.

Among all such bounds, we can choose that corresponding to the noises $(z_r, \mathbf{z}_e)$ being jointly Gaussian (they are already constrained to be marginally Gaussian) with a covariance making the bound as small as possible. Then, provided the maximizing distribution in (4.32) is Gaussian, we can express the final bound in the form (4.7)

It thus remains only to show that the maximizing distribution is Gaussian.

Lemma 5 *For each $\mathbf{K}_\phi \in \mathcal{K}_\phi$, the distribution $p_\mathbf{x}$ maximizing $I(\mathbf{x}; y_r|\mathbf{y}_e)$ is Gaussian.*

Proof. Since

$$I(\mathbf{x}; y_r|\mathbf{y}_e) = h(y_r|\mathbf{y}_e) - h(z_r|\mathbf{z}_e),$$

and the second term does not depend on $p_\mathbf{x}$, it suffices to establish that $h(y_r|\mathbf{y}_e)$ is maximized when $\mathbf{x}$ is Gaussian.

To this end, let $\boldsymbol{\alpha}_{\mathrm{LMMSE}}\mathbf{y}_e$ denote the linear minimum mean-square error (MMSE) estimator of y_r from $\mathbf{y}_e$, and λ_{LMMSE} the corresponding mean-square estimation error. Recall that

$$\boldsymbol{\alpha}_{\mathrm{LMMSE}} = (\mathbf{h}_r^\dagger \mathbf{K}_P \mathbf{H}_e^\dagger + \boldsymbol{\phi}^\dagger)(\mathbf{I} + \mathbf{H}_e \mathbf{K}_P \mathbf{H}_e^\dagger)^{-1}, \tag{4.33}$$

$$\begin{aligned}\lambda_{\mathrm{LMMSE}} &= 1 + \mathbf{h}_r^\dagger \mathbf{K}_P \mathbf{h}_r \\ &\quad - (\mathbf{h}_r^\dagger \mathbf{K}_P \mathbf{H}_e^\dagger + \boldsymbol{\phi}^\dagger)(\mathbf{I} + \mathbf{H}_e \mathbf{K}_P \mathbf{H}_e^\dagger)^{-1}(\boldsymbol{\phi} + \mathbf{H}_e \mathbf{K}_P \mathbf{h}_r)\end{aligned} \tag{4.34}$$

depend on the input and noise distributions only through their (joint) second-moment characterization, i.e.,

$$\mathbf{K}_P = \operatorname{cov} \mathbf{x}, \qquad \mathbf{K}_\phi = \begin{bmatrix} 1 & \boldsymbol{\phi}^\dagger \\ \boldsymbol{\phi} & \mathbf{I} \end{bmatrix} = \operatorname{cov} \begin{bmatrix} z_r \\ \mathbf{z}_e \end{bmatrix}. \tag{4.35}$$

Proceeding, we have

$$\begin{aligned} h(y_r|\mathbf{y}_e) &= h(y_r - \boldsymbol{\alpha}_{\mathrm{LMMSE}}\mathbf{y}_e|\mathbf{y}_e) && (4.36)\\ &\le h(y_r - \boldsymbol{\alpha}_{\mathrm{LMMSE}}\mathbf{y}_e) && (4.37)\\ &\le \log 2\pi e \lambda_{\mathrm{LMMSE}}, && (4.38)\end{aligned}$$

where (4.36) holds because adding a constant doesn't change entropy, (4.37) holds because conditioning only reduces differential entropy, and (4.38) is the maximum entropy bound on differential entropy expressed in terms of

$$\operatorname{var} e = \lambda_{\mathrm{LMMSE}}, \tag{4.39}$$

where e is the estimation error

$$e = (y_r - \boldsymbol{\alpha}_{\mathrm{LMMSE}}\mathbf{y}_e). \tag{4.40}$$

It remains only to verify that the above inequalities are tight for a Gaussian distribution. To see this, note that (4.37) holds with equality when $\mathbf{x}$ is Gaussian (and thus $(y_r, \mathbf{y}_e)$ are jointly Gaussian) since in this case e is the (unconstrained) MMSE estimation error and is therefore independent of the "data" $\mathbf{y}_e$. Furthermore, note

that in this case (4.38) holds with equality since the Gaussian distribution maximizes differential entropy subject to a variance constraint.

■

4.5 MISOME Secrecy Capacity Derivation

In this section we derive the MISOME capacity and its high and low SNR asymptotes.

4.5.1 Proof of Theorem 8

Achievability of (4.11) follows from evaluating (5.6) with the particular choices

$$u \sim \mathcal{CN}(0, P), \quad \mathbf{x} = \boldsymbol{\psi}_{\max} u, \tag{4.41}$$

where $\boldsymbol{\psi}_{\max}$ is as defined in Theorem 8. With this choice of parameters,

$$\begin{aligned} I(u;&y_{\mathrm{r}}) - I(u;\mathbf{y}_{\mathrm{e}}) \\ &= I(\mathbf{x};y_{\mathrm{r}}) - I(\mathbf{x};\mathbf{y}_{\mathrm{e}}), && (4.42)\\ &= \log\left(1 + P|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}_{\max}|^2\right) - \log\left(1 + P\|\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}_{\max}\|^2\right) && (4.43)\\ &= \log \frac{\boldsymbol{\psi}_{\max}^{\dagger}(\mathbf{I} + P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger})\boldsymbol{\psi}_{\max}}{\boldsymbol{\psi}_{\max}^{\dagger}(\mathbf{I} + P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})\boldsymbol{\psi}_{\max}} \\ &= \log \lambda_{\max}(\mathbf{I} + P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I} + P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}), && (4.44) \end{aligned}$$

where (4.42) follows from the fact that $\mathbf{x}$ is a deterministic function of u, (4.43) follows from the choice of $\mathbf{x}$ and u in (4.41), and (4.44) follows from the variational characterization of generalized eigenvalues (4.2).

We next show a converse—that rates greater than (4.11) are not achievable using our upper bound. Specifically, we show that (4.11) corresponds to our upper bound expression (4.7) in Theorem 7.

It suffices to show that a particular choice of $\boldsymbol{\phi}$ that is admissible (i.e., such that $\mathbf{K}_{\boldsymbol{\phi}} \in \mathcal{K}_{\boldsymbol{\phi}}$) minimizes (4.7). We can do this by showing that

$$\max_{\mathbf{K}_P \in \mathcal{K}_P} R_+(\mathbf{K}_P, \mathbf{K}_{\boldsymbol{\phi}}) \tag{4.45}$$

with the chosen $\boldsymbol{\phi}$ corresponds to (4.11).

Since only the first term on the right hand side of

$$R_+(\mathbf{K}_P, \mathbf{K}_{\boldsymbol{\phi}}) = I(\mathbf{x}; y_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}}) = h(y_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}}) - h(z_{\mathrm{r}}|\mathbf{z}_{\mathrm{e}})$$

depends on $\mathbf{K}_P$, we can restrict our attention to maximizing this first term with respect to $\mathbf{K}_P$.

Proceeding, exploiting that all variables are jointly Gaussian, we express this first

term in the form of the optimization

$$\begin{aligned} h(y_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}}) &= \min_{\boldsymbol{\theta}\in\mathbb{C}^{n_{\mathrm{e}}}} h(y_{\mathrm{r}} - \boldsymbol{\theta}^{\dagger}\mathbf{y}_{\mathrm{e}}) \\ &= \min_{\boldsymbol{\theta}\in\mathbb{C}^{n_{\mathrm{e}}}} h((\mathbf{h}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}}^{\dagger}\boldsymbol{\theta})^{\dagger}\mathbf{x} + z_{\mathrm{r}} - \boldsymbol{\theta}^{\dagger}\mathbf{z}_{\mathrm{e}}) \\ &= \min_{\boldsymbol{\theta}\in\mathbb{C}^{n_{\mathrm{e}}}} \log\big[(\mathbf{h}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}}^{\dagger}\boldsymbol{\theta})^{\dagger}\mathbf{K}_P(\mathbf{h}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}}^{\dagger}\boldsymbol{\theta}) \\ &\qquad\qquad + 1 + \|\boldsymbol{\theta}\|^2 - 2\,\mathrm{Re}\{\boldsymbol{\theta}^{\dagger}\boldsymbol{\phi}\}\big], \end{aligned} \tag{4.46}$$

and bound its maximum over $\mathbf{K}_P$ according to

$$\begin{aligned} &\max_{\mathbf{K}_P\in\mathcal{K}_P} h(y_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}}) \\ &\quad= \max_{\mathbf{K}_P\in\mathcal{K}_P} \min_{\boldsymbol{\theta}\in\mathbb{C}^{n_{\mathrm{e}}}} \log\big[(\mathbf{h}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}}^{\dagger}\boldsymbol{\theta})^{\dagger}\mathbf{K}_P(\mathbf{h}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}}^{\dagger}\boldsymbol{\theta}) \\ &\qquad\qquad + 1 + \|\boldsymbol{\theta}\|^2 - 2\,\mathrm{Re}\{\boldsymbol{\theta}^{\dagger}\boldsymbol{\phi}\}\big] \\ &\quad\le \min_{\boldsymbol{\theta}\in\mathbb{C}^{n_{\mathrm{e}}}} \max_{\mathbf{K}_P\in\mathcal{K}_P} \log\big[(\mathbf{h}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}}^{\dagger}\boldsymbol{\theta})^{\dagger}\mathbf{K}_P(\mathbf{h}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}}^{\dagger}\boldsymbol{\theta}) \\ &\qquad\qquad + 1 + \|\boldsymbol{\theta}\|^2 - 2\,\mathrm{Re}\{\boldsymbol{\theta}^{\dagger}\boldsymbol{\phi}\}\big] \\ &\quad= \min_{\boldsymbol{\theta}\in\mathbb{C}^{n_{\mathrm{e}}}} \log\big[P\|\mathbf{h}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}}^{\dagger}\boldsymbol{\theta}\|^2 + 1 + \|\boldsymbol{\theta}\|^2 - 2\,\mathrm{Re}\{\boldsymbol{\theta}^{\dagger}\boldsymbol{\phi}\}\big], \end{aligned} \tag{4.47}$$

where (4.47) follows by observing that a rank one $\mathbf{K}_P$ maximizes the quadratic form $(\mathbf{h}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}}^{\dagger}\boldsymbol{\theta})^{\dagger}\mathbf{K}_P(\mathbf{h}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}}^{\dagger}\boldsymbol{\theta})$.

Note that directly verifying that rank one covariance maximizes the term $h(y_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}})$ appears difficult. The above elegant derivation between (4.46) and (4.47) was suggested to us by Yonina C. Eldar and Ami Wiesel. In the literature, this line of reasoning has been used in deriving an extremal characterization of the Schur complement of a matrix (see e.g., [35, Chapter 20],[28]).

We now separately consider the cases $\lambda_{\max} > 1$ and $\lambda_{\max} \le 1$.

Case: $\lambda_{\max} > 1$

We show that the choice

$$\boldsymbol{\phi} = \frac{\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}_{\max}}{\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}_{\max}} \tag{4.48}$$

in (4.45) yields (4.11), i.e., $\log\lambda_{\max}$.

We begin by noting that since $\lambda_{\max} > 1$, the variational characterization (4.2) establishes that $\|\boldsymbol{\phi}\| < 1$ and thus $\mathbf{K}_{\boldsymbol{\phi}} \in \mathcal{K}_{\boldsymbol{\phi}}$ as defined in (5.4).

Then, provided that, with $\boldsymbol{\phi}$ as given in (4.48), the right hand side of (4.47) evaluates to

$$\min_{\boldsymbol{\theta}\in\mathbb{C}^{n_{\mathrm{e}}}} \log\big[P\|\mathbf{h}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}}^{\dagger}\boldsymbol{\theta}\|^2 + 1 + \|\boldsymbol{\theta}\|^2 - 2\,\mathrm{Re}\{\boldsymbol{\theta}^{\dagger}\boldsymbol{\phi}\}\big] = \log\left(\lambda_{\max}\cdot(1 - \|\boldsymbol{\phi}\|^2)\right), \tag{4.49}$$

we have

$$\begin{aligned} R_+ &\leq \max_{\mathbf{K}_P \in \mathcal{K}_P} R_+(\mathbf{K}_P, K_{\boldsymbol{\phi}}) \\ &= \max_{\mathbf{K}_P \in \mathcal{K}_P} h(y_\mathrm{r}|\mathbf{y}_\mathrm{e}) - h(z_\mathrm{r}|\mathbf{z}_\mathrm{e}) \\ &\leq \log(\lambda_{\max} \cdot (1 - \|\boldsymbol{\phi}\|^2)) - \log(1 - \|\boldsymbol{\phi}\|^2) \\ &= \log(\lambda_{\max}), \end{aligned}$$

i.e., (4.11), as required. Verifying (4.49) with (4.48) is a straightforward computation, the details of which are provided in Appendix B.1.

Case: $\lambda_{\max} \leq 1$, $\mathbf{H}_\mathrm{e}$ full column rank

We show that the choice

$$\boldsymbol{\phi} = \mathbf{H}_\mathrm{e}(\mathbf{H}_\mathrm{e}^\dagger \mathbf{H}_\mathrm{e})^{-1}\mathbf{h}_\mathrm{r} \tag{4.50}$$

in (4.45) yields (4.11), i.e., zero.

To verify that $\|\boldsymbol{\phi}\| \leq 1$, first note that since $\lambda_{\max} \leq 1$, it follows from (4.2) that

$$\lambda_{\max}(\mathbf{I} + P\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}^\dagger, \mathbf{I} + P\mathbf{H}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e}) \leq 1 \Leftrightarrow \lambda_{\max}(\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}^\dagger, \mathbf{H}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e}) \leq 1, \tag{4.51}$$

so that for any choice of $\boldsymbol{\psi}$,

$$\boldsymbol{\psi}^\dagger \mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}^\dagger \boldsymbol{\psi} \leq \boldsymbol{\psi}^\dagger \mathbf{H}_\mathrm{e}^\dagger \mathbf{H}_\mathrm{e} \boldsymbol{\psi}. \tag{4.52}$$

Choosing $\boldsymbol{\psi} = (\mathbf{H}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e})^{-1}\mathbf{h}_\mathrm{r}$ in (4.52) yields $\|\boldsymbol{\phi}\|^2 \leq \|\boldsymbol{\phi}\|$, i.e., $\|\boldsymbol{\phi}\| \leq 1$, as required.

Next, note that (4.47) is further upper bounded by choosing any particular choice of $\boldsymbol{\theta}$. Choosing $\boldsymbol{\theta} = \boldsymbol{\phi}$ yields

$$R_+ \leq \log\left(\frac{P\|\mathbf{h}_\mathrm{r} - \mathbf{H}_\mathrm{e}^\dagger\boldsymbol{\phi}\|^2}{1 - \|\boldsymbol{\phi}\|^2} + 1\right) \tag{4.53}$$

which with the choice (4.50) for $\boldsymbol{\phi}$ is zero.

Case: $\lambda_{\max} \leq 1$, $\mathbf{H}_\mathrm{e}$ not full column rank

Consider a new MISOME channel with $n_\mathrm{t}' < n_\mathrm{t}$ transmit antennas, where n_t' is the column rank of $\mathbf{H}_\mathrm{e}$, where the intended receiver and eavesdropper channel gains are given by

$$\mathbf{g}_\mathrm{r} = \mathbf{Q}^\dagger \mathbf{h}_\mathrm{r}, \qquad \mathbf{G}_\mathrm{e} = \mathbf{H}_\mathrm{e}\mathbf{Q}, \tag{4.54}$$

and where $\mathbf{Q}$ is a matrix whose columns constitute an orthogonal basis for the column space of $\mathbf{H}_\mathrm{e}^\dagger$, so that in this new channel $\mathbf{G}_\mathrm{e}$ has full rank.

Then provided the new channel (4.54) has the same capacity as the original channel, it follows by the analysis of the previous case that the capacity of both channels is zero. Thus it remains only to show the following.

Claim 1 *The MISOME channel* $(\mathbf{g}_{\mathrm{r}}, \mathbf{G}_{\mathrm{e}})$ *corresponding to* (4.54) *has the same secrecy capacity as that corresponding to* $(\mathbf{h}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}})$.

Proof. First we show that the new channel capacity is no larger than the original one. In particular, we have

$$\begin{aligned}
&\lambda_{\max}(\mathbf{I} + P\mathbf{g}_{\mathrm{r}}\mathbf{g}_{\mathrm{r}}^{\dagger}, \mathbf{I} + P\mathbf{G}_{\mathrm{e}}^{\dagger}\mathbf{G}_{\mathrm{e}}) \\
&\qquad = \max_{\{\boldsymbol{\psi}':\|\boldsymbol{\psi}'\|=1\}} \left\{ \frac{1 + P|\mathbf{g}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}'|^2}{1 + P\|\mathbf{G}_{\mathrm{e}}\boldsymbol{\psi}'\|^2} \right\} && (4.55) \\
&\qquad = \max_{\{\boldsymbol{\psi}':\|\boldsymbol{\psi}'\|=1\}} \frac{1 + P|\mathbf{h}_{\mathrm{r}}^{\dagger}\mathbf{Q}\boldsymbol{\psi}'|^2}{1 + P\|\mathbf{H}_{\mathrm{e}}\mathbf{Q}\boldsymbol{\psi}'\|^2} && (4.56) \\
&\qquad = \max_{\{\boldsymbol{\psi}:\boldsymbol{\psi}=\mathbf{Q}\boldsymbol{\psi}', \|\boldsymbol{\psi}\|=1\}} \frac{1 + P|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}|^2}{1 + P\|\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}\|^2} && (4.57) \\
&\qquad \leq \max_{\{\boldsymbol{\psi}:\|\boldsymbol{\psi}\|=1\}} \left\{ \frac{1 + P|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}|^2}{1 + P\|\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}\|^2} \right\} && (4.58) \\
&\qquad = \lambda_{\max}(\mathbf{I} + P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I} + P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}), && (4.59)
\end{aligned}$$

where to obtain (4.55) we have used (4.2) for the new channel, to obtain (4.56) we have used (4.54), to obtain (4.57) we have used that $\mathbf{Q}^{\dagger}\mathbf{Q} = \mathbf{I}$, to obtain (4.58) we have used that we are maximizing over a larger set, and to obtain (4.59) we have used (4.2) for the original channel. Thus,

$$\left\{\lambda_{\max}(\mathbf{I} + P\mathbf{g}_{\mathrm{r}}\mathbf{g}_{\mathrm{r}}^{\dagger}, \mathbf{I} + P\mathbf{G}_{\mathrm{e}}^{\dagger}\mathbf{G}_{\mathrm{e}})\right\}^{+} \leq \left\{\lambda_{\max}(\mathbf{I} + P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I} + P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})\right\}^{+}, \qquad (4.60)$$

Next, we show the new channel capacity is no smaller than the original one. To begin, note that

$$\mathrm{Null}(\mathbf{H}_{\mathrm{e}}) \subseteq \mathrm{Null}(\mathbf{h}_{\mathrm{r}}^{\dagger}), \qquad (4.61)$$

since if $\mathrm{Null}(\mathbf{H}_{\mathrm{e}}) \nsubseteq \mathrm{Null}(\mathbf{h}_{\mathrm{r}}^{\dagger})$, then $\lambda_{\max}(\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}) = \infty$, which would violate (4.51).

Proceeding, every $\mathbf{x} \in \mathbb{C}^{n_{\mathrm{t}}}$ can we written as

$$\mathbf{x} = \mathbf{Q}\mathbf{x}' + \tilde{\mathbf{x}}, \qquad (4.62)$$

where $\mathbf{H}_{\mathrm{e}}\tilde{\mathbf{x}} = \mathbf{0}$ and thus, via (4.61), $\mathbf{h}_{\mathrm{r}}^{\dagger}\tilde{\mathbf{x}} = 0$ as well. Hence, we have that $\mathbf{h}_{\mathrm{r}}^{\dagger}\mathbf{x} = \mathbf{g}_{\mathrm{r}}^{\dagger}\mathbf{x}'$, $\mathbf{H}_{\mathrm{e}}\mathbf{x} = \mathbf{G}_{\mathrm{e}}\mathbf{x}'$, and $\|\mathbf{x}'\|^2 \leq \|\mathbf{x}\|^2$, so any rate achieved by $p_{\mathbf{x}}$ on the channel $(\mathbf{h}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}})$ is also achieved by $p_{\mathbf{x}'}$ on the channel $(\mathbf{g}_{\mathrm{r}}, \mathbf{G}_{\mathrm{e}})$, with $p_{\mathbf{x}'}$ derived from $p_{\mathbf{x}}$ via (4.62), whence

$$\left\{\lambda_{\max}(\mathbf{I} + P\mathbf{g}_{\mathrm{r}}\mathbf{g}_{\mathrm{r}}^{\dagger}, \mathbf{I} + P\mathbf{G}_{\mathrm{e}}^{\dagger}\mathbf{G}_{\mathrm{e}})\right\}^{+} \quad \geq \quad \left\{\lambda_{\max}(\mathbf{I} + P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I} + P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})\right\}^{+}. \qquad (4.63)$$

Combining (4.63) and (4.60) establishes our claim. ∎

4.5.2 High SNR Analysis

We restrict our attention to the case $\lambda_{\max} > 1$ where the capacity is nonzero. In this case, since, via (4.2),

$$\lambda_{\max}(\mathbf{I} + P\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + P\mathbf{H}_e^\dagger\mathbf{H}_e) = \frac{1 + P|\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}(P)|^2}{1 + P\|\mathbf{H}_e\boldsymbol{\psi}_{\max}(P)\|^2} > 1, \tag{4.64}$$

where

$$\boldsymbol{\psi}_{\max}(P) \triangleq \arg\max_{\{\boldsymbol{\psi}:\|\boldsymbol{\psi}\|=1\}} \frac{1 + P|\mathbf{h}_r^\dagger\boldsymbol{\psi}|^2}{1 + P\|\mathbf{H}_e\boldsymbol{\psi}\|^2}, \tag{4.65}$$

we have

$$|\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}(P)| > \|\mathbf{H}_e\boldsymbol{\psi}_{\max}(P)\| \tag{4.66}$$

for all $P > 0$.

To obtain an upper bound note that, for all $P > 0$,

$$\begin{aligned}
&\lambda_{\max}(\mathbf{I} + P\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + P\mathbf{H}_e^\dagger\mathbf{H}_e) \\
&\qquad \le \frac{|\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}(P)|^2}{\|\mathbf{H}_e\boldsymbol{\psi}_{\max}(P)\|^2} &(4.67)\\
&\qquad \le \lambda_{\max}(\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{H}_e^\dagger\mathbf{H}_e), &(4.68)
\end{aligned}$$

where (4.67) follows from the Rayleigh quotient expansion (4.64) and the fact that, due to (4.66), the right hand side of (4.64) is increasing in P, and where (4.68) follows from (4.2). Thus, since the right hand side of (4.68) is independent of P we have

$$\lim_{P\to\infty} \lambda_{\max}(\mathbf{I} + P\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + P\mathbf{H}_e^\dagger\mathbf{H}_e) \le \lambda_{\max}(\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{H}_e^\dagger\mathbf{H}_e). \tag{4.69}$$

Next, defining

$$\boldsymbol{\psi}_{\max}(\infty) \triangleq \arg\max_{\boldsymbol{\psi}} \frac{|\mathbf{h}_r^\dagger\boldsymbol{\psi}|^2}{\|\mathbf{H}_e\boldsymbol{\psi}\|^2}, \tag{4.70}$$

we have the lower bound

$$\begin{aligned}
&\lim_{P\to\infty} \lambda_{\max}(\mathbf{I} + P\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + P\mathbf{H}_e^\dagger\mathbf{H}_e) \\
&\qquad \ge \lim_{P\to\infty} \frac{1/P + |\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}(\infty)|^2}{1/P + \|\mathbf{H}_e\boldsymbol{\psi}_{\max}(\infty)\|^2} &(4.71)\\
&\qquad = \lambda_{\max}(\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{H}_e^\dagger\mathbf{H}_e) &(4.72)
\end{aligned}$$

where (4.71) follows from (4.2) and (4.72) follows from (4.70).

Since (4.69) and (4.72) coincide we obtain (4.12). Thus, to obtain the remainder of (4.13a) we need only verify the following.

Claim 2 *The high SNR capacity is finite, i.e.,* $\lambda_{\max}(\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{H}_e^\dagger\mathbf{H}_e) < \infty$, *when* $\mathbf{H}_e^\perp\mathbf{h}_r = \mathbf{0}$.

Proof. We argue by contradiction. Suppose $\lambda_{\max}(\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{H}_e^\dagger\mathbf{H}_e) = \infty$. Then there must exist a sequence $\boldsymbol{\psi}_k$ such that $\|\mathbf{H}_e\boldsymbol{\psi}_k\| > 0$ for each $k = 1, 2, \ldots$, but $\|\mathbf{H}_e\boldsymbol{\psi}_k\| \to 0$ as $k \to \infty$. But then the hypothesis cannot be true, because, as we now show, $|\mathbf{h}_r^\dagger\boldsymbol{\psi}|^2/\|\mathbf{H}_e\boldsymbol{\psi}\|^2$, when viewed as a function of $\boldsymbol{\psi}$, is bounded whenever the denominator is nonzero.

Let $\boldsymbol{\psi}$ be any vector such that $\|\mathbf{H}_e\boldsymbol{\psi}\| \triangleq \delta > 0$. It suffices to show that

$$\frac{|\mathbf{h}_r^\dagger\boldsymbol{\psi}|^2}{\|\mathbf{H}_e\boldsymbol{\psi}\|^2} \le \frac{\|\mathbf{h}_r\|^2}{\sigma^2}, \tag{4.73}$$

where σ^2 is the smallest *nonzero* singular value of $\mathbf{H}_e$.

To verify (4.73), we first express $\boldsymbol{\psi}$ in the form

$$\boldsymbol{\psi} = c\boldsymbol{\psi}' + d\tilde{\boldsymbol{\psi}}, \tag{4.74}$$

where $\boldsymbol{\psi}'$ and $\tilde{\boldsymbol{\psi}}$ are unit vectors, c and d are real and nonnegative, $d\tilde{\boldsymbol{\psi}}$ is the projection of $\boldsymbol{\psi}$ onto the null space of $\mathbf{H}_e$, and $c\boldsymbol{\psi}'$ is the projection of $\boldsymbol{\psi}$ onto the orthogonal complement of this null space.

Next, we note that $\delta = \|\mathbf{H}_e\boldsymbol{\psi}\| = c\|\mathbf{H}_e\boldsymbol{\psi}'\| \ge c\sigma$, whence

$$c \le \frac{\delta}{\sigma}. \tag{4.75}$$

But since $\mathbf{H}_e^\perp\mathbf{h}_r = 0$ it follows that $\mathbf{h}_r^\dagger\tilde{\boldsymbol{\psi}} = 0$, so

$$|\mathbf{h}_r^\dagger\boldsymbol{\psi}|^2 = c^2|\mathbf{h}_r^\dagger\boldsymbol{\psi}'|^2 \le c^2\|\mathbf{h}_r\|^2 \le \frac{\delta^2}{\sigma^2}\|\mathbf{h}_r\|^2, \tag{4.76}$$

where the first inequality follows from the Cauchy-Schwarz inequality, and the second inequality is a simple substitution from (4.75). Dividing through by $\|\mathbf{H}_e\boldsymbol{\psi}\|^2 = \delta^2$ in (4.76) yields (4.73).

■

We now develop (4.13b) for the case where $\mathbf{H}_e^\perp\mathbf{h}_r \neq \mathbf{0}$.

First, defining

$$\mathcal{S}_\infty = \{\boldsymbol{\psi} : \|\boldsymbol{\psi}\| = 1, \|\mathbf{H}_e\boldsymbol{\psi}\| = 0\} \tag{4.77}$$

we obtain the lower bound

$$\begin{aligned}
&\frac{1}{P}\lambda_{\max}(\mathbf{I} + P\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + P\mathbf{H}_e^\dagger\mathbf{H}_e) \\
&\ge \max_{\boldsymbol{\psi}\in\mathcal{S}_\infty} \frac{1/P + |\mathbf{h}_r^\dagger\boldsymbol{\psi}|^2}{1 + P\|\mathbf{H}_e\boldsymbol{\psi}\|^2} \\
&= \max_{\boldsymbol{\psi}\in\mathcal{S}_\infty} \frac{1}{P} + |\mathbf{h}_r^\dagger\boldsymbol{\psi}|^2 \qquad (4.78) \\
&= \frac{1}{P} + \|\mathbf{H}_e^\perp\mathbf{h}_r\|^2, \qquad (4.79)
\end{aligned}$$

where to obtain (4.79) we have used,

$$\max_{\{\boldsymbol{\psi}:\|\boldsymbol{\psi}\|=1,\mathbf{H}_e\boldsymbol{\psi}=0\}} \left|\mathbf{h}_r^\dagger\boldsymbol{\psi}\right|^2 = \left\|\mathbf{H}_e^\perp\mathbf{h}_r\right\|^2. \tag{4.80}$$

Next we develop an upper bound. We first establish the following.

Claim 3 *If* $\mathbf{H}_e^\perp\mathbf{h}_r \neq \mathbf{0}$ *then there is a function* $\varepsilon(P)$ *such that* $\varepsilon(P) \to 0$ *as* $P \to \infty$, *and*

$$\|\mathbf{H}_e\boldsymbol{\psi}_{\max}(P)\| \leq \varepsilon(P).$$

Proof. We have

$$\begin{aligned}
\frac{1+P\|\mathbf{h}_r\|^2}{1+P\|\mathbf{H}_e\boldsymbol{\psi}_{\max}(P)\|^2} &\geq \frac{1+P|\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}(P)|^2}{1+P\|\mathbf{H}_e\boldsymbol{\psi}_{\max}(P)\|^2} && (4.81)\\
&\geq \max_{\{\boldsymbol{\psi}:\mathbf{H}_e\boldsymbol{\psi}=\mathbf{0},\|\boldsymbol{\psi}\|=1\}} \frac{1+P|\mathbf{h}_r^\dagger\boldsymbol{\psi}|^2}{1+P\|\mathbf{H}_e\boldsymbol{\psi}\|^2} && (4.82)\\
&= \max_{\{\boldsymbol{\psi}:\mathbf{H}_e\boldsymbol{\psi}=\mathbf{0},\|\boldsymbol{\psi}\|=1\}} \left(1+P|\mathbf{h}_r^\dagger\boldsymbol{\psi}|^2\right)\\
&= 1+P\|\mathbf{H}_e^\perp\mathbf{h}_r\|^2 && (4.83)
\end{aligned}$$

where to obtain (4.81) we have used the Cauchy-Schwarz inequality $|\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}(P)|^2 \leq \|\mathbf{h}_r\|^2$, to obtain (4.82) we have used (4.65), and to obtain (4.83) we have used (4.80).

Rearranging (4.83) then gives

$$\|\mathbf{H}_e\boldsymbol{\psi}_{\max}(P)\|^2 \leq \frac{1}{P}\left(\frac{1+P\|\mathbf{h}_r\|^2}{1+P\|\mathbf{H}_e^\perp\mathbf{h}_r\|^2} - 1\right) \triangleq \varepsilon^2(P).$$

as desired. ∎

Thus with $\mathcal{S}_P = \{\boldsymbol{\psi} : \|\boldsymbol{\psi}\| = 1, \|\mathbf{H}_e\boldsymbol{\psi}\| \leq \varepsilon(P)\}$ we have

$$\begin{aligned}
&\frac{1}{P}\lambda_{\max}(\mathbf{I}+P\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I}+P\mathbf{H}_e^\dagger\mathbf{H}_e)\\
&= \max_{\boldsymbol{\psi}\in\mathcal{S}_P} \frac{1/P+|\mathbf{h}_r^\dagger\boldsymbol{\psi}|^2}{1+P\|\mathbf{H}_e\boldsymbol{\psi}\|^2} && (4.84)\\
&\leq \max_{\boldsymbol{\psi}\in\mathcal{S}_P} \frac{1}{P}+|\mathbf{h}_r^\dagger\boldsymbol{\psi}|^2, && (4.85)
\end{aligned}$$

where (4.84) follows from (4.2) and Claim 3 that the maximizing $\boldsymbol{\psi}_{\max}$ lies in $\mathcal{S}_P$.

Now, as we will show,

$$\max_{\boldsymbol{\psi}\in\mathcal{S}_P} |\mathbf{h}_r^\dagger\boldsymbol{\psi}|^2 \leq \|\mathbf{H}_e^\perp\mathbf{h}_r\|^2 + \frac{\varepsilon^2(P)}{\sigma^2}\|\mathbf{h}_r\|^2. \tag{4.86}$$

so using (4.86) in (4.85) we obtain

$$\frac{1}{P}\lambda_{\max}(\mathbf{I}+P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I}+P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}) \leq \|\mathbf{H}_{\mathrm{e}}^{\perp}\mathbf{h}_{\mathrm{r}}\|^2 + \frac{\varepsilon^2(P)}{\sigma^2}\|\mathbf{h}_{\mathrm{r}}\|^2 + \frac{1}{P} \tag{4.87}$$

Finally, combining (4.87) and (4.79) we obtain

$$\lim_{P\to\infty}\frac{1}{P}\lambda_{\max}(\mathbf{I}+P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I}+P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}) = \|\mathbf{H}_{\mathrm{e}}^{\perp}\mathbf{h}_{\mathrm{r}}\|^2,$$

whence (4.13b).

Thus, it remains only to verify (4.86), which we do now.

We start by expressing $\boldsymbol{\psi} \in \mathcal{S}_P$ in the form [cf. (4.74)]

$$\boldsymbol{\psi} = c\boldsymbol{\psi}' + d\tilde{\boldsymbol{\psi}}, \tag{4.88}$$

where $\boldsymbol{\psi}'$ and $\tilde{\boldsymbol{\psi}}$ are unit vectors, c, d are real valued scalars in $[0,1]$, $d\tilde{\boldsymbol{\psi}}$ is the projection of $\boldsymbol{\psi}$ onto the null space of $\mathbf{H}_{\mathrm{e}}$, and $c\boldsymbol{\psi}'$ is the projection of $\boldsymbol{\psi}$ onto the orthogonal complement of this null space.

With these definitions we have,

$$\varepsilon(P) \geq \|\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}\| = c\|\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}'\| \geq c\sigma \tag{4.89}$$

since $\mathbf{H}_{\mathrm{e}}\tilde{\boldsymbol{\psi}} = \mathbf{0}$ and $\|\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}'\| \geq \sigma$.

Finally,

$$\begin{aligned}
|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}|^2 &= |d\mathbf{h}_{\mathrm{r}}^{\dagger}\tilde{\boldsymbol{\psi}} + c\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}'|^2 && (4.90)\\
&= d^2|\mathbf{h}_{\mathrm{r}}^{\dagger}\tilde{\boldsymbol{\psi}}|^2 + c^2|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}'|^2 && (4.91)\\
&\leq |\mathbf{h}_{\mathrm{r}}^{\dagger}\tilde{\boldsymbol{\psi}}|^2 + \frac{\varepsilon(P)^2}{\sigma^2}|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}'|^2 && (4.92)\\
&\leq |\mathbf{h}_{\mathrm{r}}^{\dagger}\tilde{\boldsymbol{\psi}}|^2 + \frac{\varepsilon(P)^2}{\sigma^2}\|\mathbf{h}_{\mathrm{r}}\|^2 && (4.93)\\
&\leq \|\mathbf{H}_{\mathrm{e}}^{\perp}\mathbf{h}_{\mathrm{r}}\|^2 + \frac{\varepsilon(P)^2}{\sigma^2}\|\mathbf{h}_{\mathrm{r}}\|^2, && (4.94)
\end{aligned}$$

where (4.90) follows from substituting (4.88), (4.91) follows from the fact that $\boldsymbol{\psi}'$ and $\tilde{\boldsymbol{\psi}}$ are orthogonal, (4.92) follows from using (4.89) to bound c^2, and (4.94) follows from the fact that $\mathbf{H}_{\mathrm{e}}\tilde{\boldsymbol{\psi}} = \mathbf{0}$ and (4.80).

4.5.3 Low SNR Analysis

We consider the limit $P \to 0$. In the following steps, the order notation $\mathcal{O}(P)$ means that $\mathcal{O}(P)/P \to 0$ as $P \to 0$.

$$\begin{aligned}
&\lambda_{\max}(\mathbf{I} + P\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + P\mathbf{H}_e^\dagger\mathbf{H}_e) && (4.95)\\
&\quad = \lambda_{\max}\left((\mathbf{I} + P\mathbf{H}_e^\dagger\mathbf{H}_e)^{-1}(\mathbf{I} + P\mathbf{h}_r\mathbf{h}_r^\dagger)\right) && (4.96)\\
&\quad = \lambda_{\max}\left((\mathbf{I} - P\mathbf{H}_e^\dagger\mathbf{H}_e + \mathcal{O}(P))(\mathbf{I} + P\mathbf{h}_r\mathbf{h}_r^\dagger)\right) && (4.97)\\
&\quad = \lambda_{\max}\left((\mathbf{I} - P\mathbf{H}_e^\dagger\mathbf{H}_e)(\mathbf{I} + P\mathbf{h}_r\mathbf{h}_r^\dagger)\right) + \mathcal{O}(P) && (4.98)\\
&\quad = \lambda_{\max}\left(\mathbf{I} + P(\mathbf{h}_r\mathbf{h}_r^\dagger - \mathbf{H}_e^\dagger\mathbf{H}_e)\right) + \mathcal{O}(P) && (4.99)\\
&\quad = 1 + P\lambda_{\max}(\mathbf{h}_r\mathbf{h}_r^\dagger - \mathbf{H}_e^\dagger\mathbf{H}_e) + \mathcal{O}(P), && (4.100)
\end{aligned}$$

where (4.96) follows from the definition of generalized eigenvalue, (4.97) follows from the Taylor series expansion of $(\mathbf{I} + P\mathbf{H}_e^\dagger\mathbf{H}_e)^{-1}$, where we have assumed that P is sufficiently small so that all eigenvalues of $P\mathbf{H}_e^\dagger\mathbf{H}_e$ are less than unity, (4.98) and (4.99) follow from the continuity of the eigenvalue function in its arguments and (4.100) follows from the property of eigenvalue function that $\lambda(\mathbf{I} + \mathbf{A}) = 1 + \lambda(\mathbf{A})$.

In turn, we have,

$$\begin{aligned}
\frac{C(P)}{P} &= \frac{\log(1 + P\lambda_{\max}(\mathbf{h}_r\mathbf{h}_r^\dagger - \mathbf{H}_e^\dagger\mathbf{H}_e) + \mathcal{O}(P))}{P} && (4.101)\\
&= \frac{\lambda_{\max}(\mathbf{h}_r\mathbf{h}_r^\dagger - \mathbf{H}_e^\dagger\mathbf{H}_e)}{\ln 2} + \frac{\mathcal{O}(P)}{P}, && (4.102)
\end{aligned}$$

where to obtain (4.101) we have used (4.100) in (4.11), and to obtain (4.102) we have used Taylor Series expansion of the $\ln(\cdot)$ function.

Finally, taking the limit $P \to 0$ in (4.102) yields (4.14) as desired.

4.6 Masked Beamforming Scheme Analysis

From Csiszár-Körner [8], secrecy rate $R = I(u; y_r) - I(u; \mathbf{y}_e)$ is achievable for any choice of p_u and $p_{\mathbf{x}|u}$ that satisfy the power constraint $E[|x|^2] \leq P$. While a capacity-achieving scheme corresponds to maximizing this rate over the choice of p_u and $p_{\mathbf{x}|u}$ (cf. (5.6)), the masked beamforming scheme corresponds to different (suboptimal) choice of these distributions. In particular, we choose

$$p_u = \mathcal{CN}(0, \tilde{P}) \quad \text{and} \quad p_{\mathbf{x}|u} = \mathcal{CN}(u\tilde{\mathbf{h}}_r, \tilde{P}(\mathbf{I} - \tilde{\mathbf{h}}_r\tilde{\mathbf{h}}_r^\dagger), \tag{4.103}$$

where we have chosen the convenient normalizations

$$\tilde{P} = \frac{P}{n_t} \tag{4.104}$$

and

$$\tilde{\mathbf{h}}_{\mathrm{r}} = \frac{\mathbf{h}_{\mathrm{r}}}{\|\mathbf{h}_{\mathrm{r}}\|}. \tag{4.105}$$

In this form, the secrecy rate of masked beamforming is readily obtained, as we now show

4.6.1 Rate Analysis

With p_u and $p_{\mathbf{x}|u}$ as in (4.103), we evaluate (5.6). To this end, first we have

$$I(u; y_{\mathrm{r}}) = \log(1 + \tilde{P}\|\mathbf{h}_{\mathrm{r}}\|^2) \tag{4.106}$$

Then, to evaluate $I(u; \mathbf{y}_{\mathrm{e}})$, note that

$$\begin{aligned} h(\mathbf{y}_{\mathrm{e}}) &= \log\det(\mathbf{I} + \tilde{P}\mathbf{H}_{\mathrm{e}}\mathbf{H}_{\mathrm{e}}^{\dagger}) \\ h(\mathbf{y}_{\mathrm{e}}|u) &= \log\det(\mathbf{I} + \tilde{P}\mathbf{H}_{\mathrm{e}}(\mathbf{I} - \tilde{\mathbf{h}}_{\mathrm{r}}\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger})\mathbf{H}_{\mathrm{e}}^{\dagger}) \end{aligned}$$

so

$$\begin{aligned} &I(u;\mathbf{y}_{\mathrm{e}}) \\ &\quad = h(\mathbf{y}_{\mathrm{e}}) - h(\mathbf{y}_{\mathrm{e}}|u) \\ &\quad = \log\det(\mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}\mathbf{H}_{\mathrm{e}}^{\dagger}) - \log\det(\mathbf{I} + \tilde{P}\mathbf{H}_{\mathrm{e}}(\mathbf{I}-\tilde{\mathbf{h}}_{\mathrm{r}}\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger})\mathbf{H}_{\mathrm{e}}^{\dagger}) \\ &\quad = \log\det(\mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}) - \log\det(\mathbf{I} + \tilde{P}(\mathbf{I}-\tilde{\mathbf{h}}_{\mathrm{r}}\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger})\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}) \\ &\quad = \log\det(\mathbf{I} + \tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}) \\ &\qquad\quad - \log\det(\mathbf{I} + \tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}} - \tilde{P}\tilde{\mathbf{h}}_{\mathrm{r}}\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}) \\ &\quad = -\log\det\left(\mathbf{I} - \tilde{P}\tilde{\mathbf{h}}_{\mathrm{r}}\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}(\mathbf{I} + \tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\right) \\ &\quad = -\log\left(1 - \tilde{P}\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}(\mathbf{I} + \tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\tilde{\mathbf{h}}_{\mathrm{r}}\right) \\ &\quad = -\log\left(\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger}(\mathbf{I} + \tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\tilde{\mathbf{h}}_{\mathrm{r}}\right), \end{aligned} \tag{4.107}$$

where we have repeatedly used the matrix identity $\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA})$ valid for any $\mathbf{A}$ and $\mathbf{B}$ with compatible dimensions.

Thus, combining (4.106) and (4.107) we obtain (4.15) as desired:

$$\begin{aligned} &R_{\mathrm{MB}}(P) \\ &\quad = I(u; y_{\mathrm{r}}) - I(u; \mathbf{y}_{\mathrm{e}}) \\ &\quad = \log(1 + \tilde{P}\|\mathbf{h}_{\mathrm{r}}\|^2) + \log(\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger}(\mathbf{I} + \tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\tilde{\mathbf{h}}_{\mathrm{r}}) \\ &\quad = \log\left(1 + \frac{1}{\tilde{P}\|\mathbf{h}_{\mathrm{r}}\|^2}\right) + \log(\tilde{P}\mathbf{h}_{\mathrm{r}}^{\dagger}(\mathbf{I} + \tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{h}_{\mathrm{r}}) \\ &\quad = \log\left(1 + \frac{1}{\tilde{P}\|\mathbf{h}_{\mathrm{r}}\|^2}\right) + \log(\lambda_{\max}(\tilde{P}\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I} + \tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})), \end{aligned}$$

where to obtain the last equality we have used the special form (4.3) for the largest generalized eigenvalue.

4.6.2 Comparison with capacity achieving scheme

In this section we provide a proof of Theorem 9 First, from Theorem 8 and Proposition 5 we have, with again $\tilde{P}$ as in (4.104) for convenience,

$$C\left(\frac{P}{n_t}\right) - R_{\mathrm{MB}}(P) \leq \log \frac{\lambda_{\max}(\mathbf{I} + \tilde{P}\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + \tilde{P}\mathbf{H}_e^\dagger\mathbf{H}_e)}{\lambda_{\max}(\tilde{P}\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + \tilde{P}\mathbf{H}_e^\dagger\mathbf{H}_e)}. \tag{4.108}$$

Next, with $\boldsymbol{\psi}_{\max}$ denoting the generalized eigenvector corresponding to $\lambda_{\max}(\mathbf{I} + \tilde{P}\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + \tilde{P}\mathbf{H}_e^\dagger\mathbf{H}_e)$, we have

$$\lambda_{\max}(\mathbf{I} + \tilde{P}\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + \tilde{P}\mathbf{H}_e^\dagger\mathbf{H}_e) = \frac{1 + \tilde{P}|\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}|^2}{1 + \tilde{P}\|\mathbf{H}_e\boldsymbol{\psi}_{\max}\|^2} \tag{4.109}$$

$$\lambda_{\max}(\tilde{P}\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + \tilde{P}\mathbf{H}_e^\dagger\mathbf{H}_e) \geq \frac{\tilde{P}|\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}|^2}{1 + \tilde{P}\|\mathbf{H}_e\boldsymbol{\psi}_{\max}\|^2} \tag{4.110}$$

$$\tag{4.111}$$

Finally, substituting (4.109) and (4.110) into (4.108), we obtain

$$0 \leq C\left(\frac{P}{n_t}\right) - R_{\mathrm{MB}}(P) \leq \log\left(1 + \frac{n_t}{P|\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}|^2}\right), \tag{4.112}$$

the right hand side of which approaches zero as $P \to \infty$, whence (4.16) as desired.

4.7 Scaling Laws Development

We begin by summarizing a few well-known results from random matrix theory that will be useful in our scaling laws; for further details, see, e.g., [50].

4.7.1 Some Random Matrix Properties

Three basic facts will suffice for our purposes.

Fact 9 *Suppose that* $\mathbf{v}$ *is a random length-*n *complex vector with independent, zero-mean, variance-*$1/n$ *elements, and that* $\mathbf{B}$ *is a random* $n \times n$ *complex positive semi-definite matrix distributed independently of* $\mathbf{v}$*. Then if the spectrum of* $\mathbf{B}$ *converges we have*

$$\lim_{n\to\infty} \mathbf{v}^\dagger(\mathbf{I} + \gamma\mathbf{B})^{-1}\mathbf{v} \overset{a.s.}{=} \eta_{\mathbf{B}}(\gamma), \tag{4.113}$$

where $\eta_{\mathbf{B}}(\gamma)$ *is the* η*-transform [50] of the matrix* $\mathbf{B}$*.*

Of particular interest to us is the η-transform of a special class of matrices below.

Fact 10 *Suppose that* $\mathbf{H} \in \mathbb{C}^{K\times N}$ *is random matrix whose entries are i.i.d. with variance* $1/N$. *As* $K, N \to \infty$ *with the ratio* $K/N \triangleq \beta$ *fixed, the* η*-transform of* $\mathbf{B} = \mathbf{H}^\dagger\mathbf{H}$ *is given by*

$$\eta_{\mathbf{H}^\dagger\mathbf{H}}(\gamma) = \frac{\xi(\gamma,\beta)}{\gamma}, \tag{4.114}$$

where $\xi(\cdot,\cdot)$ *is as defined in* (4.21).

The distribution of generalized eigenvalues of the pair $(\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}^\dagger, \mathbf{H}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e})$ is also known [21, 40]. For our purposes, the following is sufficient.

Fact 11 *Suppose that* $\mathbf{h}_\mathrm{r}$ *and* $\mathbf{H}_\mathrm{e}$ *have i.i.d.* $\mathcal{CN}(0,1)$ *entries, and* $n_\mathrm{e} > n_\mathrm{t}$. *Then*

$$\lambda_{\max}(\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}^\dagger, \mathbf{H}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e}) \sim \frac{2n_\mathrm{t}}{2n_\mathrm{e}-2n_\mathrm{t}+1} F_{2n_\mathrm{t},2n_\mathrm{e}-2n_\mathrm{t}+1}, \tag{4.115}$$

where $F_{2n_\mathrm{t},2n_\mathrm{e}-2n_\mathrm{t}+1}$ *is the F-distribution with* $2n_\mathrm{t}$ *and* $2n_\mathrm{e}-2n_\mathrm{t}+1$ *degrees of freedom, i.e.,*

$$F_{2n_\mathrm{t},2n_\mathrm{e}-2n_\mathrm{t}+1} \overset{\mathrm{d}}{=} \frac{v_1/(2n_\mathrm{t})}{v_2/(2n_\mathrm{e}-2n_\mathrm{t}+1)}, \tag{4.116}$$

where $\overset{\mathrm{d}}{=}$ *denote equality in distribution, and where* v_1 *and* v_2 *are independent chi-squared random variables with* $2n_\mathrm{t}$ *and* $2n_\mathrm{e}-2n_\mathrm{t}+1$ *degrees of freedom, respectively.*

Using Fact 11 it follows that with $\beta = n_\mathrm{e}/n_\mathrm{t}$ fixed,

$$\lim_{n_\mathrm{t}\to\infty} \lambda_{\max}(\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}^\dagger, \mathbf{H}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e}) \overset{\text{a.s.}}{=} \frac{1}{\beta-1}, \quad \text{when } \beta > 1. \tag{4.117}$$

Indeed, from the strong law of large numbers we have that the random variables v_1 and v_2 in (4.116) satisfy, for $\beta > 1$,

$$\lim_{n_\mathrm{t}\to\infty} \frac{v_1}{2n_\mathrm{t}} \overset{\text{a.s.}}{=} 1, \quad \text{and} \quad \lim_{n_\mathrm{t}\to\infty} \frac{v_2}{2n_\mathrm{t}(\beta-1)+1} \overset{\text{a.s.}}{=} 1 \tag{4.118}$$

Combining (4.118) with (4.116) yields (4.117).

4.7.2 Asymptotic rate analysis

We provide a proof of Theorem 10. First, from Theorem 8 we have that

$$\begin{aligned} C(P, n_\mathrm{t}, n_\mathrm{e}) &= \left\{\log \lambda_{\max}(\mathbf{I} + P\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}^\dagger, \mathbf{I} + P\mathbf{H}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e})\right\}^+ \\ &\geq \left\{\log \lambda_{\max}(P\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}^\dagger, \mathbf{I} + P\mathbf{H}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e})\right\}^+ \\ &= \left\{\log\left(P\mathbf{h}_\mathrm{r}^\dagger(\mathbf{I} + P\mathbf{H}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e})^{-1}\mathbf{h}_\mathrm{r}\right)\right\}^+, \end{aligned} \tag{4.119}$$

where (4.119) follows from the quadratic form representation (4.3) of the generalized eigenvalue.

Rewriting (4.119) using the notation

$$\tilde{\mathbf{h}}_{\mathrm{r}} = \frac{1}{\sqrt{n_{\mathrm{t}}}}\mathbf{h}_{\mathrm{r}}, \quad \text{and} \quad \tilde{\mathbf{H}}_{\mathrm{e}} = \frac{1}{\sqrt{n_{\mathrm{t}}}}\mathbf{H}_{\mathrm{e}}, \tag{4.120}$$

we then obtain (4.20) as desired:

$$\begin{aligned}\tilde{C}(\gamma,\beta) &= C(\gamma/n_{\mathrm{t}}, n_{\mathrm{t}}, \beta n_{\mathrm{t}}) \\ &\geq \left\{\log\left(\gamma\tilde{\mathbf{h}}_{\mathrm{r}}(\mathbf{I}+\gamma\tilde{\mathbf{H}}_{\mathrm{e}}^{\dagger}\tilde{\mathbf{H}}_{\mathrm{e}})^{-1}\tilde{\mathbf{h}}_{\mathrm{r}}\right)\right\}^{+} \\ &\xrightarrow{\text{a.s.}} \{\log\xi(\gamma,\beta)\}^{+} \quad \text{as } n_{\mathrm{t}}\to\infty,\end{aligned} \tag{4.121}$$

where to obtain (4.121) we have applied (4.113) and (4.114).

The derivation of the scaling law (4.22) for the masked beamforming scheme is analogous. Indeed, from Proposition 5 we have

$$\begin{aligned}R_{\mathrm{MB}}(\gamma, n_{\mathrm{t}}, \beta n_{\mathrm{t}}) &\geq \left\{\log\lambda_{\max}(\gamma\tilde{\mathbf{h}}_{\mathrm{r}}\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger}, \mathbf{I}+\gamma\tilde{\mathbf{H}}_{\mathrm{e}}^{\dagger}\tilde{\mathbf{H}}_{\mathrm{e}})\right\}^{+} \\ &= \left\{\log\left(\gamma\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger}(\mathbf{I}+\gamma\tilde{\mathbf{H}}_{\mathrm{e}}^{\dagger}\tilde{\mathbf{H}}_{\mathrm{e}})^{-1}\tilde{\mathbf{h}}_{\mathrm{r}}\right)\right\}^{+} \\ &\xrightarrow{\text{a.s.}} \{\log\xi(\gamma,\beta)\}^{+} \quad \text{as } n_{\mathrm{t}}\to\infty,\end{aligned}$$

where as above the last line comes from applying (4.113) and (4.114).

4.7.3 High SNR Scaling analysis

We provide a proof of Theorem 11

When $\beta < 1$ (i.e., $n_{\mathrm{e}} < n_{\mathrm{t}}$), we have $\mathbf{H}_{\mathrm{e}}^{\perp}\mathbf{h}_{\mathrm{r}} \neq \mathbf{0}$ almost surely, so (4.13b) holds, i.e.,

$$\lim_{P\to\infty} C(P) = \infty \tag{4.122}$$

as (4.23) reflects.

When $\beta \geq 1$ (i.e., $n_{\mathrm{e}} > n_{\mathrm{t}}$) $\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}$ is nonsingular almost surely, (4.13a) holds, i.e.,

$$\lim_{P\to\infty} C(P) = \left\{\log\lambda(\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})\right\}^{+}.$$

Taking the limit $n_{\mathrm{e}}, n_{\mathrm{t}} \to \infty$ with $n_{\mathrm{e}}/n_{\mathrm{t}} = \beta$ fixed, and using (4.117), we obtain

$$\lim_{n_{\mathrm{t}}\to\infty}\lim_{P\to\infty} C(P) = \{-\log(\beta-1)\}^{+}$$

as (4.23) asserts.

Furthermore, via (4.16) we have that

$$\lim_{P\to\infty} R_{\mathrm{MB}}(P) = \left\{\log\lambda(\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})\right\}^{+} = \lim_{P\to\infty} C(P),$$

whence (4.24).

4.8 Fading Channel Analysis

We prove the lower and upper bounds of Theorem 12 separately.

4.8.1 Proof of Lower bound

We establish (4.27a) in this section. By viewing the fading channel as a set of parallel channels indexed by the channel gain $\mathbf{h}_r$ of the intended receiver[6] and the eavesdropper's observation as $(\mathbf{y}_e, \mathbf{H}_e)$, the rate

$$R = I(u; y_r \mid \mathbf{h}_r) - I(u; \mathbf{y}_e, \mathbf{H}_e \mid \mathbf{h}_r). \tag{4.123}$$

is achievable for any choice of $p_{u|\mathbf{h}_r}$ and $p_{\mathbf{x}|u,\mathbf{h}_r}$ that satisfies the power constraint $E[\rho(\mathbf{h}_r)] \leq P$. We choose distributions corresponding to an adaptive version of masked beamforming, i.e., [cf. (4.103)]

$$p_{u|\mathbf{h}_r} = \mathcal{CN}(0, \tilde{\rho}(\mathbf{h}_r)), \quad p_{\mathbf{x}|u,\mathbf{h}_r} = \mathcal{CN}\left(u\tilde{\mathbf{h}}_r, \tilde{\rho}(\mathbf{h}_r)(\mathbf{I} - \tilde{\mathbf{h}}_r\tilde{\mathbf{h}}_r^\dagger)\right), \tag{4.124}$$

where we have chosen the convenient normalizations [cf. (4.104) and (4.105)]

$$\tilde{\rho}(\mathbf{h}_r) \triangleq \frac{\rho(\mathbf{h}_r)}{n_t} \tag{4.125}$$

and

$$\tilde{\mathbf{h}}_r = \frac{\mathbf{h}_r}{\|\mathbf{h}_r\|}. \tag{4.126}$$

Evaluating (4.123) with the distributions (4.124) yields (4.27a) with (4.29a):

$$\begin{aligned} I(u;&y_r \mid \mathbf{h}_r) - I(u; \mathbf{y}_e, \mathbf{H}_e \mid \mathbf{h}_r) && (4.127)\\ &= E[\log(1 + \tilde{\rho}(\mathbf{h}_r)\|\mathbf{h}_r\|^2)] \\ &\quad + E[\log(\tilde{\mathbf{h}}_r^\dagger(\mathbf{I} + \tilde{\rho}(\mathbf{h}_r)\mathbf{H}_e^\dagger\mathbf{H}_e)^{-1}\tilde{\mathbf{h}}_r)] && (4.128)\\ &= E\left[\log\left(1 + \frac{1}{\tilde{\rho}(\mathbf{h}_r)\|\mathbf{h}_r\|^2}\right)\right] \\ &\quad + E\left[\log\left(\tilde{\rho}(\mathbf{h}_r)\mathbf{h}_r^\dagger(\mathbf{I} + \tilde{\rho}(\mathbf{h}_r)\mathbf{H}_e^\dagger\mathbf{H}_e)^{-1}\mathbf{h}_r\right)\right], && (4.129) \end{aligned}$$

where the steps leading to (4.128) are analogous to those used in Section 4.6.1 for the nonfading case and hence have been omitted.

[6]Since the fading coefficients are continuous valued, one has to discretize these coefficients before mapping to parallel channels. By choosing appropriately fine quantization levels one can approach the rate as closely as possible. See e.g., [23] for a discussion.

4.8.2 Proof of upper bound

We provide a proof of (4.27b). Suppose that there is a sequence of $(2^{nR}, n)$ codes such that for a sequence ε_n (with $\varepsilon_n \to 0$ as $n \to \infty$),

$$\begin{aligned} &\frac{1}{n}H(w) - \frac{1}{n}H(w|\mathbf{y}_{\mathrm{e}}^n, \mathbf{H}_{\mathrm{e}}^n, \mathbf{h}_{\mathrm{r}}^n) \le \varepsilon_n, \\ &\Pr(\hat{w} \neq w) \le \varepsilon_n. \end{aligned} \tag{4.130}$$

An auxiliary channel

We now introduce another channel for which the noise vaiables $\mathsf{z}_{\mathrm{r}}(t)$ and $\mathbf{z}_{\mathrm{e}}(t)$ are correlated, but the conditions in (4.130) still hold. Hence any rate achievable on the original channel is also achievable on this new channel. In what follows, we will upper bound the rate achievable for this new channel instead of the original channel.

We begin by introducing some notation. Let,

$$\rho_t(\mathbf{h}_{\mathrm{r}}^t) \triangleq E\big[\|\mathbf{x}(t)\|^2 \mid \mathbf{h}_{\mathrm{r}}^t = \mathbf{h}_{\mathrm{r}}^t\big] \tag{4.131}$$

denote the transmitted power at time t, when the channel realization of the intended receiver from time 1 to t is $\mathbf{h}_{\mathrm{r}}^t$. Note that $\rho_t(\cdot)$ satisfies the long term average power constraint i.e.,

$$E_{\mathbf{h}_{\mathrm{r}}^n}\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t(\mathbf{h}_{\mathrm{r}}^t)\right] \le P. \tag{4.132}$$

Next, let, $p_{\mathbf{h}_{\mathrm{r}}}$ and $p_{\mathbf{H}_{\mathrm{e}}}$ denote the density functions of $\mathbf{h}_{\mathrm{r}}$ and $\mathbf{H}_{\mathrm{e}}$, respectively, and let $p_{z_{\mathrm{r}}}$ and $p_{\mathbf{z}_{\mathrm{e}}}$ denote the density function of the noise random variables in our channel model (4.26).

Observe that the constraints in (4.130) (and hence the capacity) depend only on the distributions $p_{\mathbf{z}_{\mathrm{e}}^n,\mathbf{h}_{\mathrm{r}}^n,\mathbf{H}_{\mathrm{e}}^n}(\mathbf{z}_e^n, \mathbf{h}_{\mathrm{r}}^n, \mathbf{H}_{\mathrm{e}}^n)$ and $p_{z_{\mathrm{r}}^n,\mathbf{h}_{\mathrm{r}}^n}(z_r^n, \mathbf{h}_{\mathrm{r}}^n)$. Furthermore since the channel model (4.26) is memoryless and $(\mathbf{h}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}})$ are i.i.d. and mutually independent, we have

$$p_{\mathbf{z}_{\mathrm{e}}^n,\mathbf{h}_{\mathrm{r}}^n,\mathbf{H}_{\mathrm{e}}^n}(\mathbf{z}_e^n, \mathbf{h}_{\mathrm{r}}^n, \mathbf{H}_{\mathrm{e}}^n) = \prod_{t=1}^{n} p_{\mathbf{z}_{\mathrm{e}}}(\mathbf{z}_e(t)) p_{\mathbf{h}_{\mathrm{r}}}(\mathbf{h}_{\mathrm{r}}(t)) p_{\mathbf{H}_{\mathrm{e}}}(\mathbf{H}_{\mathrm{e}}(t)), \tag{4.133}$$

$$p_{z_{\mathrm{r}}^n,\mathbf{h}_{\mathrm{r}}^n}(z_r^n, \mathbf{h}_{\mathrm{r}}^n) = \prod_{t=1}^{n} p_{z_{\mathrm{r}}}(z_r(t)) p_{\mathbf{h}_{\mathrm{r}}}(\mathbf{h}_{\mathrm{r}}(t)). \tag{4.134}$$

Let $\mathcal{P}_t$ denote the set of conditional-joint distributions $p_{\mathsf{z}_{\mathrm{r}}(t),\mathbf{z}_{\mathrm{e}}(t)|\mathbf{h}_{\mathrm{r}}^n,\mathbf{H}_{\mathrm{e}}^n}$ with fixed conditional-marginals, i.e.,

$$\begin{aligned} \mathcal{P}_t = \big\{ & p_{\mathsf{z}_{\mathrm{r}}(t),\mathbf{z}_{\mathrm{e}}(t)|\mathbf{h}_{\mathrm{r}}^n,\mathbf{H}_{\mathrm{e}}^n}(z_r, \mathbf{z}_e \mid \mathbf{h}_{\mathrm{r}}^n, \mathbf{H}_{\mathrm{e}}^n) \mid \\ & p_{\mathsf{z}_{\mathrm{r}}(t)|\mathbf{h}_{\mathrm{r}}^n,\mathbf{H}_{\mathrm{e}}^n}(z_r \mid \mathbf{h}_{\mathrm{r}}^n, \mathbf{H}_{\mathrm{e}}^n) = p_{z_{\mathrm{r}}}(z_r), p_{\mathbf{z}_{\mathrm{e}}(t)|\mathbf{h}_{\mathrm{r}}^n,\mathbf{H}_{\mathrm{e}}^n}(\mathbf{z}_e \mid \mathbf{h}_{\mathrm{r}}^n, \mathbf{H}_{\mathrm{e}}^n) = p_{\mathbf{z}_{\mathrm{e}}}(\mathbf{z}_e) \big\}. \end{aligned} \tag{4.135}$$

Suppose that for each $t = 1, 2, \ldots, n$ we select a distribution $p_{\mathsf{z}_{\mathrm{r}}(t),\mathbf{z}_{\mathrm{e}}(t)|\mathbf{h}_{\mathrm{r}}^n,\mathbf{H}_{\mathrm{e}}^n} \in \mathcal{P}_t$

and consider a channel with distribution

$$p_{\mathbf{z}_\mathrm{r}^n,\mathbf{z}_\mathrm{e}^n,\mathbf{h}_\mathrm{r}^n,\mathbf{H}_\mathrm{e}^n}(z_r^n,\mathbf{z}_e^n,\mathbf{h}_\mathrm{r}^n,\mathbf{H}_\mathrm{e}^n) = \prod_{t=1}^{n} p_{\mathbf{z}_\mathrm{r}(t)\mathbf{z}_\mathrm{e}(t)|\mathbf{h}_\mathrm{r}^n,\mathbf{H}_\mathrm{e}^n}(z_r(t),\mathbf{z}_e(t)|\mathbf{h}_\mathrm{r}^n,\mathbf{H}_\mathrm{e}^n) p_{\mathbf{h}_\mathrm{r}}(\mathbf{h}_\mathrm{r}(t)) p_{\mathbf{H}_\mathrm{e}}(\mathbf{H}_\mathrm{e}(t)). \quad (4.136)$$

This new channel distribution has noise variables $(\mathbf{z}_\mathrm{r}(t), \mathbf{z}_\mathrm{e}(t))$ correlated, where the correlation is possibly time-dependent, but from (4.135) and (4.136), note that $\mathbf{z}_\mathrm{r}^n$ and $\mathbf{z}_\mathrm{e}^n$ are marginally Gaussian and i.i.d., and satisfy (4.133) and (4.134). Hence the conditions in (4.130) are satisfied for this channel and the rate R is achievable.

In the sequel we select $p_{\mathbf{z}_\mathrm{r}(t),\mathbf{z}_\mathrm{e}(t)|\mathbf{h}_\mathrm{r}^n,\mathbf{H}_\mathrm{e}^n}(z_r, \mathbf{z}_e \mid \mathbf{h}_\mathrm{r}^n, \mathbf{H}_\mathrm{e}^n)$ to be the worst case noise distribution for the Gaussian channel with gains $\mathbf{h}_\mathrm{r}(t)$, and, $\mathbf{H}_\mathrm{e}(t)$, and power of $\rho_t(\mathbf{h}_\mathrm{r}^t)$ in Theorem 8 i.e., if $\boldsymbol{\psi}_t$ is the eigenvector corresponding to the largest generalized eigenvalue $\lambda_{\max}(\mathbf{I} + \rho_t(\mathbf{h}_\mathrm{r}^t)\mathbf{h}_\mathrm{r}(t)\mathbf{h}_\mathrm{r}(t)^\dagger, \mathbf{I} + \rho_t(\mathbf{h}_\mathrm{r}^t)\mathbf{H}_\mathrm{e}^\dagger(t)\mathbf{H}_\mathrm{e}(t))$,

$$p_{\mathbf{z}_\mathrm{r}(t),\mathbf{z}_\mathrm{e}(t)|\mathbf{h}_\mathrm{r}^n,\mathbf{H}_\mathrm{e}^n} = \mathcal{CN}\left(0, \begin{bmatrix} 1 & \boldsymbol{\phi}_t^\dagger \\ \boldsymbol{\phi}_t & \mathbf{I} \end{bmatrix}\right), \text{ where} \quad (4.137)$$

$$\boldsymbol{\phi}_t = \begin{cases} \frac{1}{\mathbf{h}_\mathrm{r}^\dagger(t)\boldsymbol{\psi}_t}(\mathbf{H}_\mathrm{e}(t)\boldsymbol{\psi}_t), & \lambda_{\max} \geq 1, \\ \mathbf{G}_\mathrm{e}(t)(\mathbf{G}_\mathrm{e}^\dagger(t)\mathbf{G}_\mathrm{e}(t))^{-1}\mathbf{g}_\mathrm{r}(t), & \lambda_{\max} < 1, \end{cases}$$

and where $\mathbf{G}_\mathrm{e}(t)$ and $\mathbf{g}_\mathrm{r}(t)$ are related to $\mathbf{H}_\mathrm{e}(t)$ and $\mathbf{h}_\mathrm{r}(t)$ as in (4.54). Our choice of $p_{\mathbf{z}_\mathrm{r}(t),\mathbf{z}_\mathrm{e}(t)|\mathbf{h}_\mathrm{r}^n,\mathbf{H}_\mathrm{e}^n}$ is such that $(\mathbf{z}_\mathrm{r}(t), \mathbf{z}_\mathrm{e}(t))$ only depend on the $(\mathbf{H}_\mathrm{e}(t), \mathbf{h}_\mathrm{r}(t), \rho_t(\mathbf{h}_\mathrm{r}^t))$ i.e.,

$$(\mathbf{H}_\mathrm{e}^n, \mathbf{h}_\mathrm{r}^n) \to (\rho(\mathbf{h}_\mathrm{r}^t), \mathbf{h}_\mathrm{r}(t), \mathbf{H}_\mathrm{e}(t)) \to (\mathbf{z}_\mathrm{r}(t), \mathbf{z}_\mathrm{e}(t)) \quad (4.138)$$

forms a Markov chain.

Upper bound on the auxiliary channel

We now upper bound the secrecy rate for the channel (4.136). Note that this also upper bounds the rate on the original channel.

From Fano's inequality, that there exists a sequence ε'_n such that $\varepsilon'_n \to 0$ as $n \to \infty$, and,

$$\frac{1}{n} H(w|y_{\mathrm{r}}^n, \mathbf{h}_{\mathrm{r}}^n) \leq \varepsilon'_n.$$

$$\begin{aligned}
nR = H(w) &= I(w; y_{\mathrm{r}}^n \mid \mathbf{h}_{\mathrm{r}}^n) + n\varepsilon'_n \\
&= I(w; y_{\mathrm{r}}^n \mid \mathbf{h}_{\mathrm{r}}^n) - I(w; \mathbf{y}_{\mathrm{e}}^n, \mathbf{H}_{\mathrm{e}}^n \mid \mathbf{h}_{\mathrm{r}}^n) + n(\varepsilon_n + \varepsilon'_n) && (4.139) \\
&\leq I(w; y_{\mathrm{r}}^n \mid \mathbf{h}_{\mathrm{r}}^n, \mathbf{H}_{\mathrm{e}}^n, \mathbf{y}_{\mathrm{e}}^n) + n(\varepsilon_n + \varepsilon'_n) \\
&\leq I(\mathbf{x}^n; y_{\mathrm{r}}^n \mid \mathbf{h}_{\mathrm{r}}^n, \mathbf{H}_{\mathrm{e}}^n, \mathbf{y}_{\mathrm{e}}^n) + n(\varepsilon_n + \varepsilon'_n) && (4.140) \\
&\leq \sum_{t=1}^{n} I(\mathbf{x}(t); y_{\mathrm{r}}(t) \mid \mathbf{H}_{\mathrm{e}}^n, \mathbf{h}_{\mathrm{r}}^n, \mathbf{y}_{\mathrm{e}}(t)) + n(\varepsilon_n + \varepsilon'_n), && (4.141)
\end{aligned}$$

where (4.139) follows from the secrecy condition (c.f. (4.130)), and (4.140) follows from the Markov relation $w \leftrightarrow (\mathbf{x}^n, \mathbf{y}_{\mathrm{e}}^n, \mathbf{h}_{\mathrm{r}}^n, \mathbf{H}_{\mathrm{e}}^n) \leftrightarrow y_{\mathrm{r}}^n$, and (4.141) holds because for the channel (4.136) we have

$$h(y_{\mathrm{r}}^n|\mathbf{y}_{\mathrm{e}}^n, \mathbf{H}_{\mathrm{e}}^n, \mathbf{h}_{\mathrm{r}}^n, \mathbf{x}^n) = \sum_{t=1}^{n} h(y_{\mathrm{r}}(t)|\mathbf{y}_{\mathrm{e}}(t), \mathbf{h}_{\mathrm{r}}^n, \mathbf{H}_{\mathrm{e}}^n, \mathbf{x}(t)).$$

We next upper bound the term $I(\mathbf{x}(t); y_{\mathrm{r}}(t) \mid \mathbf{y}_{\mathrm{e}}(t), \mathbf{H}_{\mathrm{e}}^n, \mathbf{h}_{\mathrm{r}}^n)$ in (4.141) for each $t = 1, 2, \ldots, n$.

$$\begin{aligned}
&I(\mathbf{x}(t); y_{\mathrm{r}}(t) \mid \mathbf{y}_{\mathrm{e}}(t), \mathbf{H}_{\mathrm{e}}^n, \mathbf{h}_{\mathrm{r}}^n) \\
&\leq I(\mathbf{x}(t); y_{\mathrm{r}}(t) \mid \mathbf{y}_{\mathrm{e}}(t), \mathbf{H}_{\mathrm{e}}(t), \mathbf{h}_{\mathrm{r}}(t), \rho_t(\mathbf{h}_{\mathrm{r}}^t)) && (4.142) \\
&\leq E[\{\log \lambda_{\max}(\mathbf{I} + \rho_t(\mathbf{h}_{\mathrm{r}}^t)\mathbf{h}_{\mathrm{r}}(t)\mathbf{h}_{\mathrm{r}}^\dagger(t), \mathbf{I} + \rho_t(\mathbf{h}_{\mathrm{r}}^t)\mathbf{H}_{\mathrm{e}}^\dagger(t)\mathbf{H}_{\mathrm{e}}(t))\}^+], && (4.143)
\end{aligned}$$

where (4.142) follows from the fact that (c.f. (4.138)),

$$(\mathbf{H}_{\mathrm{e}}^n, \mathbf{h}_{\mathrm{r}}^n) \to (\mathbf{x}(t), \rho_t(\mathbf{h}_{\mathrm{r}}^t), \mathbf{h}_{\mathrm{r}}(t), \mathbf{H}_{\mathrm{e}}(t)) \to (y_{\mathrm{r}}(t), \mathbf{y}_{\mathrm{e}}(t))$$

forms a Markov chain and (4.143) follows since our choice of the noise distribution in (4.137) is the worst case noise in (4.7) for the Gaussian channel with gains $\mathbf{h}_{\mathrm{r}}(t)$, $\mathbf{H}_{\mathrm{e}}(t)$ and power $\rho_t(\mathbf{h}_{\mathrm{r}}^t)$, hence the derivation in Theorem 8 applies.

Substituting (4.143) into (4.141) we have,

$$nR - n(\varepsilon_n + \varepsilon_n') \tag{4.144}$$

$$= \sum_{t=1}^{n} E_{\mathbf{H}_e(t),\mathbf{h}_r^t}\left[\left\{\log \lambda_{\max}(\mathbf{I} + \rho_t(\mathbf{h}_r^t)\mathbf{h}_r(t)\mathbf{h}_r^\dagger(t), \mathbf{I} + \rho_t(\mathbf{h}_r^t)\mathbf{H}_e^\dagger(t)\mathbf{H}_e(t))\right\}^+\right] \tag{4.145}$$

$$\leq \sum_{t=1}^{n} E_{\mathbf{H}_e(t),\mathbf{h}_r(t)}\left[\left\{\log \lambda_{\max}(\mathbf{I} + E_{\mathbf{h}_r^{t-1}}[\rho_t(\mathbf{h}_r^t)]\mathbf{h}_r(t)\mathbf{h}_r^\dagger(t), \mathbf{I} + E_{\mathbf{h}_r^{t-1}}[\rho_t(\mathbf{h}_r^t)]\mathbf{H}_e^\dagger(t)\mathbf{H}_e(t))\right\}^+\right] \tag{4.146}$$

$$= \sum_{t=1}^{n} E_{\mathbf{H}_e(t),\mathbf{h}_r(t)}\left[\left\{\log \lambda_{\max}(\mathbf{I} + \hat{\rho}_t(\mathbf{h}_r(t))\mathbf{h}_r(t)\mathbf{h}_r^\dagger(t), \mathbf{I} + \hat{\rho}_t(\mathbf{h}_r(t))\mathbf{H}_e^\dagger(t)\mathbf{H}_e(t))\right\}^+\right] \tag{4.147}$$

$$= \sum_{t=1}^{n} E_{\mathbf{H}_e,\mathbf{h}_r}[\{\log \lambda_{\max}(\mathbf{I} + \hat{\rho}_t(\mathbf{h}_r)\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + \hat{\rho}_t(\mathbf{h}_r)\mathbf{H}_e^\dagger\mathbf{H}_e)\}^+] \tag{4.148}$$

$$\leq nE_{\mathbf{H}_e,\mathbf{h}_r}\left[\left\{\log \lambda_{\max}(\mathbf{I} + \sum_{t=1}^{n}\frac{1}{n}\hat{\rho}_t(\mathbf{h}_r)\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + \sum_{t=1}^{n}\frac{1}{n}\hat{\rho}_t(\mathbf{h}_r)\mathbf{H}_e^\dagger\mathbf{H}_e)\right\}^+\right] \tag{4.149}$$

$$= nE_{\mathbf{H}_e,\mathbf{h}_r}[\{\log \lambda_{\max}(\mathbf{I} + \rho(\mathbf{h}_r)\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + \rho(\mathbf{h}_r)\mathbf{H}_e^\dagger\mathbf{H}_e)\}^+] \tag{4.150}$$

where (4.146) and (4.149) follow from Jensen's inequality since $C(P) = \{\log \lambda_{\max}(\mathbf{I} + P\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + P\mathbf{H}_e^\dagger\mathbf{H}_e)\}^+$ is a capacity and therefore concave in P, (4.147) follows by defining

$$\hat{\rho}_t(\mathbf{h}_r) = E_{\mathbf{h}_r^{t-1}}[\rho_t(\mathbf{h}_r^t)], \tag{4.151}$$

(4.148) follows from the fact that the distribution of both $\mathbf{h}_r$ and $\mathbf{H}_e$ does not depend on t, and (4.150) follows by defining $\rho(\mathbf{h}_r) = \frac{1}{n}\sum_{t=1}^{n}\hat{\rho}_t(\mathbf{h}_r)$.

To complete the proof, note that

$$E_{\mathbf{h}_r}[\rho(\mathbf{h}_r)] = \frac{1}{n}\sum_{t=1}^{n} E_{\mathbf{h}_r}[\hat{\rho}_t(\mathbf{h}_r)]$$

$$= \frac{1}{n}\sum_{t=1}^{n} E_{\mathbf{h}_r^t}[\rho_t(\mathbf{h}_r^t)] \tag{4.152}$$

$$= \frac{1}{n}\sum_{t=1}^{n} E_{\mathbf{h}_r^n}[\rho_t(\mathbf{h}_r^t)] \leq P, \tag{4.153}$$

where (4.152) follows from (4.151) and the fact that the channel gains are i.i.d., and (4.153) follows from (4.132).

4.8.3 Proof of Proposition 6

The proof is immediate from Theorems 10, 11 and 12.

For the lower bound, we only consider the case when $\log \xi(P, \beta) > 0$, since otherwise the rate is zero. We select $\rho(\mathbf{h}_r) = P$ to be fixed for each $\mathbf{h}_r$. Then we have

from Theorem 10 that

$$R_{\mathrm{FF},-}(\mathbf{h}_\mathrm{r}, \mathbf{H}_\mathrm{e}, P) \xrightarrow{\text{a.s.}} \log \xi(P, \beta).$$

Finally since almost-sure convergence implies convergence in expectation,

$$\lim_{n_\mathrm{t}\to\infty} E[R_{\mathrm{FF},-}(\mathbf{h}_\mathrm{r}, \mathbf{H}_\mathrm{e}, P)] = \log \xi(P, \beta),$$

which establishes the lower bound (4.30). For the upper bound, since

$$R_{\mathrm{FF},+}(\mathbf{h}_\mathrm{r}, \mathbf{H}_\mathrm{e}, P) = \left\{\log\lambda_{\max}(\mathbf{I}+P\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I}+P\mathbf{H}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e})\right\}^+,$$

we have from Theorem 11 that

$$\lim_{n_\mathrm{t}\to\infty} R_{\mathrm{FF},+}(\mathbf{h}_\mathrm{r}, \mathbf{H}_\mathrm{e}, P) \overset{\text{a.s.}}{\leq} \tilde{C}(\infty, \beta), \tag{4.154}$$

and hence

$$\begin{aligned}\lim_{n_\mathrm{t}\to\infty} C_{FF}(P = \gamma, n_\mathrm{t}, n_\mathrm{e} = \beta n_\mathrm{t}) &\leq \lim_{n_\mathrm{t}\to\infty} E[R_{FF,+}(\mathbf{h}_\mathrm{r}, \mathbf{H}_\mathrm{e}, \gamma)] \\ &\leq \tilde{C}(\infty, \beta),\end{aligned}$$

where we again use the fact that almost sure convergence implies convergence in expectation.

4.9 Concluding Remarks

The present chapter characterizes the key performance characteristics and tradeoffs inherent in communication over the MISOME channel. In the next chapter, we will see analogous results for the general MIMOME channel. However, unlike the MISOME channel, we do not have a closed form solution for the secrecy capacity for the MIMOME channel, so it is less amenable to analysis.

Chapter 5

MIMOME Channel

In this chapter, we study the case when all the three terminals — the sender, the receiver and the eavesdropper have multiple antennas and establish the secrecy capacity. Our approach to establish the secrecy capacity is analogous to the MISOME case. We start with the upper bound established in the previous chapter and show that it is tight for the MIMOME channel. Unlike the MISOME channel however, the secrecy capacity does not admit a closed form expression. So it is expressed as a solution to an optimization problem that can be computed numerically.

We further study the capacity in the high signal-to-noise-ratio (SNR) regime. In this regime, to achieve the capacity, it suffices to simultaneously diagonalize both the channel matrices, using the generalized singular value decomposition, and use independent codebooks across the resulting parallel channels. A necessary and sufficient condition under which capacity is zero is also provided.

In addition to the capacity achieving scheme, a synthetic noise transmission scheme is analyzed. This scheme is semi-blind — it selects the transmit directions only based on the channel of the legitimate receiver, but needs the knowledge of the eavesdropper's channel for selecting the rate.

Finally, we study the scaling laws for the zero capacity condition. Suppose there are a total of $T \gg 1$ antennas that need to be allocated between the sender and the receiver. It is well known that the optimal allocation that maximizes both the rate and the diversity is to set $n_{\mathrm{t}} = n_{\mathrm{r}} = \frac{1}{2}T$. However from a secrecy point of view this allocation may not be optimal. Indeed, we show that the optimal allocation (for the zero capacity condition) is $n_{\mathrm{t}} = \frac{2}{3}T$ and $n_{\mathrm{r}} = \frac{1}{3}T$.

5.1 Channel Model

We denote the number of antennas at the sender, the receiver and the eavesdropper by n_{t}, n_{r} and n_{e} respectively.

$$\begin{aligned} \mathbf{y}_{\mathrm{r}}(t) &= \mathbf{H}_{\mathrm{r}}\mathbf{x}(t) + \mathbf{z}_{\mathrm{r}}(t) \\ \mathbf{y}_{\mathrm{e}}(t) &= \mathbf{H}_{\mathrm{e}}\mathbf{x}(t) + \mathbf{z}_{\mathrm{e}}(t), \end{aligned} \tag{5.1}$$

where $\mathbf{H}_{\mathrm{r}} \in \mathbb{C}^{n_{\mathrm{r}} \times n_{\mathrm{t}}}$ and $\mathbf{H}_{\mathrm{e}} \in \mathbb{C}^{n_{\mathrm{e}} \times n_{\mathrm{t}}}$ are channel matrices associated with the receiver and the eavesdropper. The channel matrices are fixed for the entire transmission period and known to all the three terminals. The additive noise $\mathbf{z}_{\mathrm{r}}(t)$ and $\mathbf{z}_{\mathrm{e}}(t)$ are circularly-symmetric and complex-valued Gaussian random variables. The input satisfies a power constraint $E\left[\frac{1}{n}\sum_{t=1}^{n} ||\mathbf{x}(t)||^2\right] \leq P$.

The definition for the secrecy capacity is analogous to the case of the MISOME channel in the previous chapter and will be omitted.

5.2 Main Results

We summarize the main results in this chapter in this section.

5.2.1 Secrecy Capacity of the MIMOME Channel

The secrecy capacity of the MIMOME channel is stated in the theorem below.

Theorem 13 *The secrecy capacity of the MIMOME wiretap channel is*

$$C = \min_{\mathbf{K}_{\mathbf{\Phi}} \in \mathcal{K}_{\mathbf{\Phi}}} \max_{\mathbf{K}_{\mathrm{P}} \in \mathcal{K}_{\mathrm{P}}} R_{+}(\mathbf{K}_{\mathrm{P}}, \mathbf{K}_{\mathbf{\Phi}}), \tag{5.2}$$

where $R_{+}(\mathbf{K}_{\mathrm{P}}, \mathbf{K}_{\mathbf{\Phi}}) = I(\mathbf{x}; \mathbf{y}_{\mathrm{r}} \mid \mathbf{y}_{\mathrm{e}})$ *with* $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_{\mathrm{P}})$ *and*

$$\mathcal{K}_{\mathrm{P}} \triangleq \left\{ \mathbf{K}_{\mathrm{P}} \;\middle|\; \mathbf{K}_{\mathrm{P}} \succeq \mathbf{0}, \quad \mathrm{tr}(\mathbf{K}_{\mathrm{P}}) \leq P \right\}, \tag{5.3}$$

and where $[\mathbf{z}_{\mathrm{r}}^{\dagger}, \mathbf{z}_{\mathrm{e}}^{\dagger}]^{\dagger} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_{\mathbf{\Phi}})$, *with*

$$\begin{aligned} \mathcal{K}_{\mathbf{\Phi}} &\triangleq \left\{ \mathbf{K}_{\mathbf{\Phi}} \;\middle|\; \mathbf{K}_{\mathbf{\Phi}} = \begin{bmatrix} \mathbf{I}_{n_{\mathrm{r}}} & \mathbf{\Phi} \\ \mathbf{\Phi}^{\dagger} & \mathbf{I}_{n_{\mathrm{e}}} \end{bmatrix}, \quad \mathbf{K}_{\mathbf{\Phi}} \succeq \mathbf{0} \right\} \\ &= \left\{ \mathbf{K}_{\mathbf{\Phi}} \;\middle|\; \mathbf{K}_{\mathbf{\Phi}} = \begin{bmatrix} \mathbf{I}_{n_{\mathrm{r}}} & \mathbf{\Phi} \\ \mathbf{\Phi}^{\dagger} & \mathbf{I}_{n_{\mathrm{e}}} \end{bmatrix}, \quad \sigma_{\max}(\mathbf{\Phi}) \leq 1 \right\}. \end{aligned} \tag{5.4}$$

Furthermore, the minimax problem in (5.2) *has a saddle point solution* $(\bar{\mathbf{K}}_{\mathrm{P}}, \bar{\mathbf{K}}_{\mathbf{\Phi}})$ *and the secrecy capacity can also be expressed as,*

$$C = R_{+}(\bar{\mathbf{K}}_{\mathrm{P}}, \bar{\mathbf{K}}_{\mathbf{\Phi}}) = \log \frac{\det(\mathbf{I} + \mathbf{H}_{\mathrm{r}} \bar{\mathbf{K}}_{\mathrm{P}} \mathbf{H}_{\mathrm{r}}^{\dagger})}{\det(\mathbf{I} + \mathbf{H}_{\mathrm{e}} \bar{\mathbf{K}}_{\mathrm{P}} \mathbf{H}_{\mathrm{e}}^{\dagger})}. \tag{5.5}$$

Connection with Csiszár and Körner Capacity

A characterization of the secrecy capacity for the non-degraded discrete memoryless broadcast channel $p_{y_r,y_e|x}$ is provided by Csiszár and Körner [8],

$$C = \max_{p_u, p_{x|u}} I(u; y_r) - I(u; y_e), \tag{5.6}$$

where u is an auxiliary random variable (over a certain alphabet with bounded cardinality) that satisfies $u \to x \to (y_r, y_e)$. As remarked in [8], the secrecy capacity (5.6) can be extended in principle to incorporate continuous-valued inputs. However, directly identifying the optimal u for the MIMOME case is not straightforward.

Theorem 13 indirectly establishes an optimal choice of u in (5.6). Suppose that $(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_\Phi)$ is a saddle point solution to the minimax problem in (5.2). From (5.5) we have

$$R_+(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_\Phi) = R_-(\bar{\mathbf{K}}_P), \tag{5.7}$$

where

$$R_-(\bar{\mathbf{K}}_P) \triangleq \log \frac{\det(\mathbf{I} + \mathbf{H}_r \bar{\mathbf{K}}_P \mathbf{H}_r^\dagger)}{\det(\mathbf{I} + \mathbf{H}_e \bar{\mathbf{K}}_P \mathbf{H}_e^\dagger)}$$

is the achievable rate obtained by evaluating (5.6) for $u = x \sim \mathcal{CN}(\mathbf{0}, \bar{\mathbf{K}}_P)$. This choice of $p_u, p_{\mathbf{x}|u}$ thus maximizes (5.6). Furthermore note that

$$\bar{\mathbf{K}}_P \in \arg\max_{\mathbf{K}_P \in \mathcal{K}_P} \log \frac{\det(\mathbf{I} + \mathbf{H}_r \mathbf{K}_P \mathbf{H}_r^\dagger)}{\det(\mathbf{I} + \mathbf{H}_e \mathbf{K}_P \mathbf{H}_e^\dagger)} \tag{5.8}$$

where the set $\mathcal{K}_P$ is defined in (5.3). Unlike the minimax problem (5.2) the maximization problem (5.8) is not a convex optimization problem since the objective function is not a concave function of $\mathbf{K}_P$. Even if one verifies that $\bar{\mathbf{K}}_P$ satisfies the optimality conditions associated with (5.8), this will only establish that $\bar{\mathbf{K}}_P$ is a locally optimal solution. The capacity expression (5.2) provides a convex reformulation of (5.8) and establishes that $\bar{\mathbf{K}}_P$ is a globally optimal solution in (5.8).[1]

Structure of the optimal solution

The saddle point solution $(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_\Phi)$ satisfies a certain necessary condition that admits an intuitive interpretation. In particular, in the proof of Theorem 13, we show the following: Let $\mathbf{S}$ be any matrix that has a full column rank matrix and satisfies $\bar{\mathbf{K}}_P = \mathbf{S}\mathbf{S}^\dagger$ and let $\bar{\mathbf{\Phi}}$ be the cross-covariance matrix between the noise random variables in (5.2), (c.f. (5.4)), then

$$\mathbf{H}_e \mathbf{S} = \bar{\mathbf{\Phi}}^\dagger \mathbf{H}_r \mathbf{S}. \tag{5.9}$$

Note that $\bar{\mathbf{\Phi}}$ is a contraction matrix i.e., all its singular values are less than or equal to unity. The column space of $\mathbf{S}$ is the subspace in which the sender transmits

[1]The "high SNR" case of this problem i.e., $\max_{\mathbf{K} \in \mathcal{K}_\infty} \log \frac{\det(\mathbf{H}_r \mathbf{K} \mathbf{H}_r^\dagger)}{\det(\mathbf{H}_e \mathbf{K} \mathbf{H}_e^\dagger)}$ is known as the multiple-discriminant-function in multivariate statistics and is well-studied; see, e.g., [51].

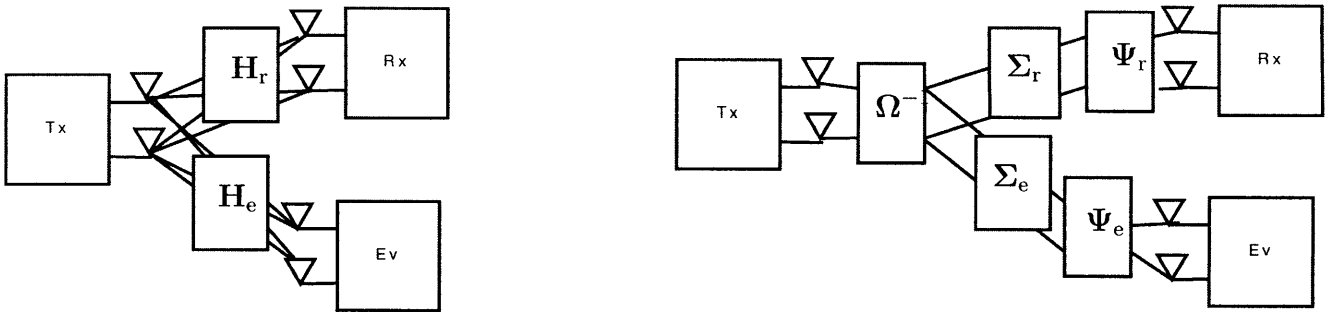


Figure 5-1: Simultaneous diagonalization via the GSVD transform. The left figure show the original channel model with 2×2 channel matrices $\mathbf{H}_r$ and $\mathbf{H}_e$. The right figure shows the GSVD transform applied to the channel matrices i.e., $\mathbf{H}_r = \mathbf{\Psi}_r \mathbf{\Sigma}_r \mathbf{\Omega}^{-1}$ and $\mathbf{H}_e = \mathbf{\Psi}_e \mathbf{\Sigma}_e \mathbf{\Omega}^{-1}$, where $\mathbf{\Psi}_r$ and $\mathbf{\Psi}_e$ are unitary matrices and $\mathbf{\Sigma}_r$ and $\mathbf{\Sigma}_e$ are diagonal matrices.

information. So (5.9) states that no information is transmitted along any direction where the eavesdropper observes a stronger signal than the intended receiver. The effective channel of the eavesdropper, $\mathbf{H}_e\mathbf{S}$, is a degraded version of the effective channel of the intended receiver, $\mathbf{H}_r\mathbf{S}$ even though the channel matrices may not be ordered a-priori. This condition explains why the genie upper bound, which provides $\mathbf{y}_e$ to the legitimate receiver (c.f. Lemma 7) does not increase the capacity of the fictitious channel.

5.2.2 Capacity analysis in the High SNR Regime

While the capacity expression in Theorem 13 can be computed numerically, it does not admit a closed form solution. In this section, we develop a closed form expression for the capacity in the high signal-to-noise-ratio (SNR) regime, in terms of the generalized singular values of the channel matrices $\mathbf{H}_r$ and $\mathbf{H}_e$. The main message here is that in the high SNR regime, an optimal scheme involves simultaneously diagonalizing the channel matrices $\mathbf{H}_r$ and $\mathbf{H}_e$ using the generalized singular value decomposition (GSVD) transform. This creates a set of parallel channels independent channels between the sender and the receivers, and it suffices to use independent Gaussian codebooks across these channels. This architecture for the case of $2 \times 2 \times 2$ channel is shown in Fig. 5-1.

The high-SNR secrecy capacity is stated below.

Theorem 14 *Let, $\sigma_1 \leq \sigma_2 \leq \ldots \leq \sigma_s$, be the generalized singular values of the channel matrices $\mathbf{H}_r$ and $\mathbf{H}_e$ as defined in (5.52). The high SNR secrecy capacity is given as follows. If*

$$\text{Null}(\mathbf{H}_e) \cap \text{Null}(\mathbf{H}_r)^{\perp} = \{\cdot\} \tag{5.10}$$

then

$$\lim_{P \to \infty} C(P) = \sum_{j:\sigma_j \geq 1} \log \sigma_j^2, \tag{5.11}$$

else,

$$C(P) = \sum_{j:\sigma_j \geq 1} \log \sigma_j^2 + \log\det\left(\mathbf{I} + \frac{P}{p}\mathbf{H}_{\mathrm{r}}\mathbf{H}_{\mathrm{e}}^{\perp}\mathbf{H}_{\mathrm{r}}^{\dagger}\right) - o_P(1), \tag{5.12}$$

where p is defined via (5.50)*, and $o_P(1) \to 0$ as $P \to \infty$, and $\mathbf{H}_{\mathrm{e}}^{\perp} \in \mathbb{C}^{n_{\mathrm{t}} \times n_{\mathrm{t}}}$ is the projection matrix (see* (5.59)*) onto the null space of $\mathbf{H}_{\mathrm{e}}$.*

We also consider a sub-optimal synthetic noise transmission strategy analogous to the masked beamforming strategy described in the previous section. Note that for this strategy the allocated rate depends on both $(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}})$, so the scheme is only semi-blind. For simplicity we focus on the case when $\mathrm{rank}(\mathbf{H}_{\mathrm{r}}) = n_{\mathrm{r}}$ and $\mathrm{rank}(\mathbf{H}_{\mathrm{e}}) = n_{\mathrm{t}}$.

Corollary 6 *In the high SNR regime the rate expression* (5.102)*, can be expressed in terms of the generalized singular values of $(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}})$. In particular,*

$$\lim_{P\to\infty} R_{\mathrm{SN}}(P) = \sum_{j=1}^{n_t} \log \sigma_j^2 \tag{5.13}$$

It is interesting to compare the expression (5.13) with the high SNR capacity expression (5.11). While the capacity expression involves summation over only those generalized singular values that exceed unity, the synthetic noise transmission scheme involves summation over all the singular values and hence is sub-optimal. Rather surprisingly, both the capacity achieving scheme and the synthetic noise scheme can be characterized using just the generalized singular values of $(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}})$ in the high SNR regime.

5.2.3 Zero Capacity Condition and Scaling Laws

The conditions on $\mathbf{H}_{\mathrm{r}}$ and $\mathbf{H}_{\mathrm{e}}$ for which the secrecy capacity is zero have a simple form.

Lemma 6 *The secrecy capacity of the MIMOME channel is zero if and only if*

$$\sigma_{\max}(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}}) \triangleq \sup_{\mathbf{v}\in\mathbb{C}^{n_{\mathrm{t}}}} \frac{||\mathbf{H}_{\mathrm{r}}\mathbf{v}||}{||\mathbf{H}_{\mathrm{e}}\mathbf{v}||} \leq 1. \tag{5.14}$$

Analysis of the zero-capacity condition in the limit of large number of antennas provides some useful insights we develop below.

Corollary 7 *Suppose that $\mathbf{H}_{\mathrm{r}}$ and $\mathbf{H}_{\mathrm{e}}$ have i.i.d. $CN(0,1)$ entries. Suppose that $n_{\mathrm{r}}, n_{\mathrm{e}}, n_{\mathrm{t}} \to \infty$, while keeping $n_{\mathrm{r}}/n_{\mathrm{e}} = \gamma$ and $n_{\mathrm{t}}/n_{\mathrm{e}} = \beta$ fixed. The secrecy capacity*[2]

[2]We assume that the channels are sampled once, then stay fixed for the entire period of transmission, and are revealed to all the terminals.

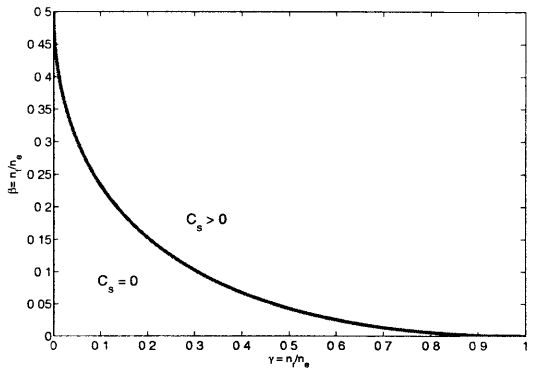


Figure 5-2: Zero-capacity condition in the (γ, β) plane. The capacity is zero for any point below the curve, i.e., the eavesdropper has sufficiently many antennas to get non-vanishing fraction of the message, even when the sender and receiver fully exploit the knowledge of $\mathbf{H}_{\mathrm{e}}$.

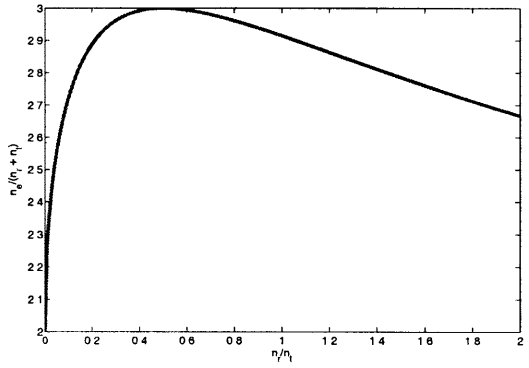


Figure 5-3: The minimum number of eavesdropping antennas per sender plus receiver antenna for the secrecy capacity to be zero, plotted as a function of $n_{\mathrm{r}}/n_{\mathrm{t}}$.

$C(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}})$ *converges almost surely to zero if and only if* $0 \leq \beta \leq 1/2$, $0 \leq \gamma \leq 1$, *and*

$$\gamma \leq (1 - \sqrt{2\beta})^2. \tag{5.15}$$

Figs. 5-2 and 5-3 provide further insight into the asymptotic analysis for the capacity achieving scheme. In Fig. 5-2, we show the values of (γ, β) where the secrecy rate is zero. If the eavesdropper increases its antennas at a sufficiently high rate so that the point (γ, β) lies below the solid curve, then secrecy capacity is zero. The MISOME case corresponds to the vertical intercept of this plot. The secrecy capacity is zero, if $\beta \leq 1/2$, i.e., the eavesdropper has at least twice the number of antennas as the sender. The single transmit antenna (SIMOME) case corresponds to the horizontal intercept. In this case the secrecy capacity is zero if $\gamma \leq 1$, i.e., the eavesdropper has more antennas than the receiver.

In Fig. 5-3, we consider the scenario where a total of $T \gg 1$ antennas are divided between the sender and the receiver. The horizontal axis plots the ratio $n_{\mathrm{r}}/n_{\mathrm{t}}$, while the vertical axis plots the minimum number of antennas at the eavesdropper (normalized by T) for the secrecy capacity to be zero. We note that the optimal allocation of antennas, that maximizes the number of eavesdropper antennas happens at $n_{\mathrm{r}}/n_{\mathrm{t}} = 1/2$. This can be explicitly obtained from the following minimization

$$\begin{aligned} &\text{minimize } \beta + \gamma \\ &\text{subject to, } \gamma \geq (1 - \sqrt{2\beta})^2,\ \beta \geq 0,\ \gamma \geq 0. \end{aligned} \tag{5.16}$$

The optimal solution can be easily verified to be $(\beta^*, \gamma^*) = (2/9, 1/9)$. In this case, the eavesdropper needs $\approx 3T$ antennas for the secrecy capacity to be zero. We remark that the objective function in (5.16) is not sensitive to variations in the optimal solution. If fact even if we allocate equal number of antennas to the sender and the receiver, the eavesdropper needs $\frac{(3+2\sqrt{2})}{2}T \approx 2.9142 \times T$ antennas for the secrecy capacity to be zero.

5.3 Derivation of the Secrecy Capacity

Our proof involves two main parts. First we note that the right hand side in (5.2) is an upper bound on the secrecy capacity. Then we examine the optimality conditions associated with the saddle point solution to establish (5.7), which completes the proof since

$$C \leq R_+(\bar{\mathbf{K}}_{\mathrm{P}}, \bar{\mathbf{K}}_{\mathbf{\Phi}}) = R_-(\bar{\mathbf{K}}_{\mathrm{P}}) \leq C.$$

We begin with an upper bound on the secrecy capacity of the multi-antenna wiretap established in the previous chapter.

Lemma 7 *An upper bound on the secrecy capacity is given by*

$$C(P) \leq R_{\mathrm{UB}}(P) = \min_{\mathbf{K}_{\mathbf{\Phi}} \in \mathcal{K}_{\mathbf{\Phi}}} \max_{\mathbf{K}_{\mathrm{P}} \in \mathcal{K}_{\mathrm{P}}} R_+(\mathbf{K}_{\mathrm{P}}, \mathbf{K}_{\mathbf{\Phi}}), \tag{5.17}$$

where

$$R_+(\mathbf{K}_{\mathrm{P}}, \mathbf{K}_{\mathbf{\Phi}}) \triangleq I(\mathbf{x}; \mathbf{y}_{\mathrm{r}} \mid \mathbf{y}_{\mathrm{e}}) \tag{5.18}$$

is the conditional mutual information expression evaluated with $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_{\mathrm{P}})$, *and* $[\mathbf{z}_{\mathrm{r}}^\dagger, \mathbf{z}_{\mathrm{e}}^\dagger]^\dagger \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_{\mathbf{\Phi}})$, *and the domain sets* $\mathcal{K}_{\mathrm{P}}$ *and* $\mathcal{K}_{\mathbf{\Phi}}$ *are defined via* (5.3) *and* (5.4) *respectively.*

It remains to establish that this upper bound expression satisfies (5.7), which we do in the remainder of this section. We divide the proof into several steps, which are outlined in Fig. 5-4.

Lemma 8 (Existence of a saddle point solution) *The function* $R_+(\mathbf{K}_{\mathrm{P}}, \mathbf{K}_{\mathbf{\Phi}})$ *in* (5.18) *has the following properties:*

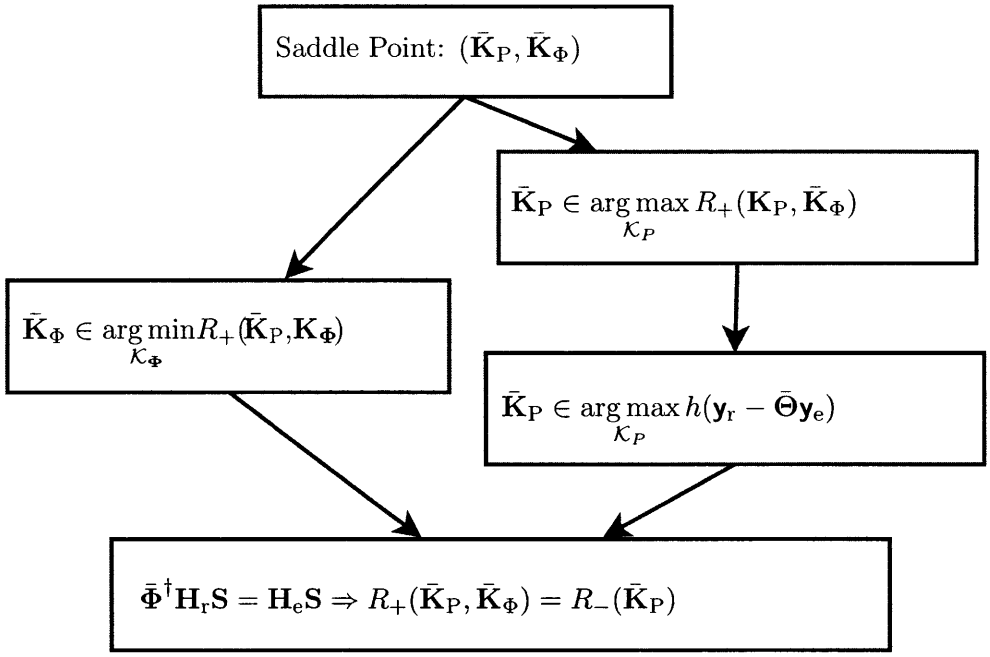


Figure 5-4: Key steps in the Proof of Theorem 1. The existence of a saddle point $(\bar{\mathbf{K}}_{\mathrm{P}}, \bar{\mathbf{K}}_{\mathbf{\Phi}})$ is first established. Thereafter the KKT conditions associated with the minimax expressions are used to simplify the saddle value and show that it matches the lower bound.

1. *For each fixed* $\mathbf{K}_{\mathbf{\Phi}} \in \mathcal{K}_{\mathbf{\Phi}}$, *the function* $R_+(\cdot, \mathbf{K}_{\mathbf{\Phi}})$ *is concave* $(\cap)$ *in the variable* $\mathbf{K}_{\mathrm{P}} \in \mathcal{K}_{\mathrm{P}}$.

2. *For each fixed* $\mathbf{K}_{\mathrm{P}} \in \mathcal{K}_{\mathrm{P}}$, *the function* $R_+(\mathbf{K}_{\mathrm{P}}, \cdot)$ *is convex* $(\cup)$ *in the variable* $\mathbf{K}_{\mathbf{\Phi}} \in \mathcal{K}_{\mathbf{\Phi}}$.

3. *There exists a saddle point solution to* (5.17) *i.e.,* $\exists \bar{\mathbf{K}}_{\mathrm{P}} \in \mathcal{K}_{\mathrm{P}}$ *and* $\exists \bar{\mathbf{K}}_{\mathbf{\Phi}} \in \mathcal{K}_{\mathbf{\Phi}}$, *such that*
$$R_+(\mathbf{K}_{\mathrm{P}}, \bar{\mathbf{K}}_{\mathbf{\Phi}}) \leq R_+(\bar{\mathbf{K}}_{\mathrm{P}}, \bar{\mathbf{K}}_{\mathbf{\Phi}}) \leq R_+(\bar{\mathbf{K}}_{\mathrm{P}}, \mathbf{K}_{\mathbf{\Phi}}) \tag{5.19}$$
holds for each $\mathbf{K}_{\mathrm{P}} \in \mathcal{K}_{\mathrm{P}}$, *and each* $\mathbf{K}_{\mathbf{\Phi}} \in \mathcal{K}_{\mathbf{\Phi}}$.

Proof.

To establish 1) above, with a slight abuse in notation, let us define $R_+(p_{\mathbf{x}}, \mathbf{K}_{\mathbf{\Phi}}) = I(\mathbf{x}; \mathbf{y}_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}})$, to be the conditional mutual information evaluated when the noise random variables are jointly Gaussian random variables with a covariance $\mathbf{K}_{\mathbf{\Phi}}$, and with input distribution of $p_{\mathbf{x}}$. As before, $R_+(\mathbf{Q}, \mathbf{K}_{\mathbf{\Phi}})$ denotes the conditional mutual information, evaluated when the noise random variables are jointly Gaussian with covariance $\mathbf{K}_{\mathbf{\Phi}}$ and the input distribution is Gaussian with a covariance $\mathbf{Q}$. Let $p_{\mathbf{x}}^1 = \mathcal{CN}(0, \mathbf{Q}_1)$, $p_{\mathbf{x}}^2 = \mathcal{CN}(0, \mathbf{Q}_2)$ and $p_{\mathbf{x}}^{\theta} = \theta p_{\mathbf{x}}^1 + (1-\theta)p_{\mathbf{x}}^2$, $\mathbf{Q}^{\theta} = \theta\mathbf{Q}_1 + (1-\theta)\mathbf{Q}_2$, for some $\theta \in [0,1]$ and $p_{\mathbf{x}}^G = \mathcal{CN}(0, \mathbf{Q}^{\theta})$. It suffices to show that

$$R_+(\mathbf{Q}^{\theta}, \mathbf{K}_{\mathbf{\Phi}}) \geq \theta R_+(\mathbf{Q}_1, \mathbf{K}_{\mathbf{\Phi}}) + (1-\theta)R_+(\mathbf{Q}_2, \mathbf{K}_{\mathbf{\Phi}}),$$

which we do below:

$$\begin{aligned} R_+(\mathbf{Q}^\theta, \mathbf{K}_\Phi) &= R_+(p_{\mathbf{x}}^G, \mathbf{K}_\Phi) \\ &\geq R_+(p_{\mathbf{x}}^\theta, \mathbf{K}_\Phi) \qquad (5.20) \\ &\geq \theta R_+(p_{\mathbf{x}}^1, \mathbf{K}_\Phi) + (1-\theta) R_+(p_{\mathbf{x}}^2, \mathbf{K}_\Phi) \qquad (5.21) \\ &= \theta R_+(\mathbf{Q}_1, \mathbf{K}_\Phi) + (1-\theta) R_+(\mathbf{Q}_2, \mathbf{K}_\Phi), \end{aligned}$$

where (5.20) follows from the fact that, as shown in Appendix C.1, a Gaussian distribution maximizes function $R_+(p_{\mathbf{x}}^\theta, \mathbf{K}_\Phi)$, among all distributions with a fixed covariance, and (5.21) from the fact that for each fixed $p_{\mathbf{y}_\mathrm{r},\mathbf{y}_\mathrm{e}|\mathbf{x}}$, the function $I(\mathbf{x}; \mathbf{y}_\mathrm{r}|\mathbf{y}_\mathrm{e})$ is a concave function in the input distribution (see e.g., [23, Appendix I]).

To establish the 2), we note that for each $\mathbf{x} \sim \mathcal{CN}(0, \mathbf{K}_\mathrm{P})$, the function $I(\mathbf{x}; \mathbf{y}_\mathrm{r}, \mathbf{y}_\mathrm{e})$ is convex in the noise covariance $\mathbf{K}_\Phi$ (see e.g., [13, Lemma II-3, pg. 3076] for an information theoretic proof).

The existence of a saddle point $(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi)$ as stated in 3) follows from 1) and 2) and the fact that the domain sets $\mathcal{K}_\mathrm{P}$ and $\mathcal{K}_\Phi$ are convex and compact.

■

In the sequel, let $(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi)$ denote a saddle point solution in (5.17), and define $\bar{\mathbf{\Phi}}$ and $\bar{\mathbf{\Theta}}$ via,

$$\bar{\mathbf{K}}_\Phi = \begin{bmatrix} \mathbf{I}_{n_\mathrm{r}} & \bar{\mathbf{\Phi}} \\ \bar{\mathbf{\Phi}}^\dagger & \mathbf{I}_{n_\mathrm{e}} \end{bmatrix}, \qquad (5.22)$$

$$\bar{\mathbf{\Theta}} = (\mathbf{H}_\mathrm{r}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger + \bar{\mathbf{\Phi}})(\mathbf{I} + \mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger)^{-1}. \qquad (5.23)$$

Lemma 9 (Properties of saddle-point) *The saddle point solution* $(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi)$ *to* (5.17) *satisfies the following*

1.
$$(\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e})\bar{\mathbf{K}}_\mathrm{P}(\bar{\mathbf{\Phi}}^\dagger\mathbf{H}_\mathrm{r} - \mathbf{H}_\mathrm{e})^\dagger = \mathbf{0} \qquad (5.24)$$

2. *Suppose that* $\mathbf{S}$ *is a full rank square root matrix of* $\bar{\mathbf{K}}_\mathrm{P}$, *i.e.,* $\bar{\mathbf{K}}_\mathrm{P} = \mathbf{S}\mathbf{S}^\dagger$ *and* $\mathbf{S}$ *has a full column rank. Then provided* $\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e} \neq \mathbf{0}$, *the matrix*
$$\mathbf{M} = (\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e})\mathbf{S} \qquad (5.25)$$
has a full column rank[3].

Proof. The conditions 1) and 2) are established by examining the optimality conditions satisfied by the saddle-point in (5.17) i.e.,

$$\bar{\mathbf{K}}_\Phi \in \arg\min_{\mathbf{K}_\Phi \in \mathcal{K}_\Phi} R_+(\bar{\mathbf{K}}_\mathrm{P}, \mathbf{K}_\Phi) \qquad (5.26)$$

and

$$\bar{\mathbf{K}}_\mathrm{P} \in \arg\max_{\mathbf{K}_\mathrm{P} \in \mathcal{K}_\mathrm{P}} R_+(\mathbf{K}_\mathrm{P}, \bar{\mathbf{K}}_\Phi). \qquad (5.27)$$

[3]A matrix $\mathbf{M}$ has a full column rank if, for any vector $\mathbf{a}$, $\mathbf{M}\mathbf{a} = \mathbf{0}$ if and only if $\mathbf{a} = \mathbf{0}$.

We first consider the optimality condition in (5.26) and establish (5.24). The derivation is most direct when $\bar{\mathbf{K}}_{\mathbf{\Phi}}$ is non-singular. The extension to the case when $\bar{\mathbf{K}}_{\mathbf{\Phi}}$ is singular is provided in Appendix C.3. The Lagrangian associated with the minimization (5.26) is

$$\mathcal{L}_{\mathbf{\Phi}}(\mathbf{K}_{\mathbf{\Phi}}, \mathbf{\Upsilon}) = R_+(\bar{\mathbf{K}}_{\mathrm{P}}, \mathbf{K}_{\mathbf{\Phi}}) + \mathrm{tr}(\mathbf{\Upsilon}\mathbf{K}_{\mathbf{\Phi}}), \tag{5.28}$$

where the dual variable

$$\mathbf{\Upsilon} = \begin{array}{c} \\ n_{\mathrm{r}} \\ n_{\mathrm{e}} \end{array} \begin{array}{c} \begin{array}{cc} n_{\mathrm{r}} & n_{\mathrm{e}} \end{array} \\ \begin{bmatrix} \mathbf{\Upsilon}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Upsilon}_2 \end{bmatrix} \end{array} \tag{5.29}$$

is a block diagonal matrix corresponding to the constraint that the noise covariance $\mathbf{K}_{\mathbf{\Phi}}$ must have identity matrices on its diagonal. The associated Kuhn-Tucker (KKT) conditions yield

$$\begin{aligned} &\nabla_{\mathbf{K}_{\mathbf{\Phi}}} \mathcal{L}_{\mathbf{\Phi}}(\mathbf{K}_{\mathbf{\Phi}}, \mathbf{\Upsilon})\big|_{\bar{\mathbf{K}}_{\mathbf{\Phi}}} \\ =&\nabla_{\mathbf{K}_{\mathbf{\Phi}}} R_+(\bar{\mathbf{K}}_{\mathrm{P}}, \mathbf{K}_{\mathbf{\Phi}})\big|_{\bar{\mathbf{K}}_{\mathbf{\Phi}}} + \mathbf{\Upsilon} = \mathbf{0}, \end{aligned} \tag{5.30}$$

where,

$$\begin{aligned} &\nabla_{\mathbf{K}_{\mathbf{\Phi}}} R_+(\bar{\mathbf{K}}_{\mathrm{P}}, \mathbf{K}_{\mathbf{\Phi}})\big|_{\bar{\mathbf{K}}_{\mathbf{\Phi}}} && (5.31) \\ &\quad = \nabla_{\mathbf{K}_{\mathbf{\Phi}}} \left[\log\det(\mathbf{K}_{\mathbf{\Phi}} + \mathbf{H}_{\mathrm{t}}\bar{\mathbf{K}}_{\mathrm{P}}\mathbf{H}_{\mathrm{t}}^{\dagger}) - \log\det(\mathbf{K}_{\mathbf{\Phi}})\right]\Big|_{\bar{\mathbf{K}}_{\mathbf{\Phi}}} \\ &\quad = (\bar{\mathbf{K}}_{\mathbf{\Phi}} + \mathbf{H}_{\mathrm{t}}\bar{\mathbf{K}}_{\mathrm{P}}\mathbf{H}_{\mathrm{t}}^{\dagger})^{-1} - \bar{\mathbf{K}}_{\mathbf{\Phi}}^{-1} && (5.32) \end{aligned}$$

and where we have used

$$\mathbf{H}_{\mathrm{t}} = \begin{bmatrix} \mathbf{H}_{\mathrm{r}} \\ \mathbf{H}_{\mathrm{e}} \end{bmatrix}. \tag{5.33}$$

Substituting (5.32) in (5.30), and simplifying, we obtain,

$$\mathbf{H}_{\mathrm{t}}\bar{\mathbf{K}}_{\mathrm{P}}\mathbf{H}_{\mathrm{t}}^{\dagger} = \bar{\mathbf{K}}_{\mathbf{\Phi}}\mathbf{\Upsilon}(\bar{\mathbf{K}}_{\mathbf{\Phi}} + \mathbf{H}_{\mathrm{t}}\bar{\mathbf{K}}_{\mathrm{P}}\mathbf{H}_{\mathrm{t}}^{\dagger}), \tag{5.34}$$

and the relation in (5.24) follows from (5.34) through a straightforward computation as shown in Appendix C.2.

To establish 2) above, we use the optimality condition associated with $\bar{\mathbf{K}}_{\mathrm{P}}$ i.e., (5.27) As in establishing 1), the proof is most direct when $\bar{\mathbf{K}}_{\mathbf{\Phi}}$ is non-singular. Hence this case is treated first, while the case when $\bar{\mathbf{K}}_{\mathbf{\Phi}}$ is singular is treated in Appendix C.6.

$$\begin{aligned} \bar{\mathbf{K}}_{\mathrm{P}} &\in \underset{\mathbf{K}_{\mathrm{P}} \in \mathcal{K}_{\mathrm{P}}}{\arg\max}\, R_+(\mathbf{K}_{\mathrm{P}}, \bar{\mathbf{K}}_{\mathbf{\Phi}}) \\ &= \underset{\mathbf{K}_{\mathrm{P}} \in \mathcal{K}_{\mathrm{P}}}{\arg\max}\, h(\mathbf{y}_{\mathrm{r}} \mid \mathbf{y}_{\mathrm{e}}) \\ &= \underset{\mathbf{K}_{\mathrm{P}} \in \mathcal{K}_{\mathrm{P}}}{\arg\max}\, h\left(\mathbf{y}_{\mathrm{r}} - \Theta(\mathbf{K}_{\mathrm{P}})\mathbf{y}_{\mathrm{e}}\right), \end{aligned} \tag{5.35}$$

where $\Theta(\mathbf{K}_{\mathrm{P}}) = (\mathbf{H}_{\mathrm{r}}\mathbf{K}_{\mathrm{P}}\mathbf{H}_{\mathrm{e}}^{\dagger} + \bar{\mathbf{\Phi}})(\mathbf{H}_{\mathrm{e}}\mathbf{K}_{\mathrm{P}}\mathbf{H}_{\mathrm{e}}^{\dagger} + \mathbf{I})^{-1}$ is the linear minimum mean squared estimation coefficient of $\mathbf{y}_{\mathrm{r}}$ given $\mathbf{y}_{\mathrm{e}}$. Directly working with the Kuhn-Tucker conditions associated with (5.35) appears difficult. Nevertheless it turns out that we can

replace the objecive function above, with a simpler objective function as described below. First, note that since $\bar{\mathbf{K}}_{\mathrm{P}}$ is an optimum solution to (5.35), in general

$$\arg\max_{\mathbf{K}_{\mathrm{P}}\in\mathcal{K}_{\mathrm{P}}} h\left(\mathbf{y}_{\mathrm{r}}-\bar{\mathbf{\Theta}}\mathbf{y}_{\mathrm{e}}\right) \geq \arg\max_{\mathbf{K}_{\mathrm{P}}\in\mathcal{K}_{\mathrm{P}}} h\left(\mathbf{y}_{\mathrm{r}}-\Theta(\mathbf{K}_{\mathrm{P}})\mathbf{y}_{\mathrm{e}}\right) \tag{5.36}$$

holds, since substituting $\mathbf{K}_{\mathrm{P}}=\bar{\mathbf{K}}_{\mathrm{P}}$ in the objective function on the left hand side, attains the maximum on the right hand side. Somewhat surprisingly, it turns out that the inequality above is in fact an equality, i.e., the left hand side also attains the maximum when $\mathbf{K}_{\mathrm{P}}=\bar{\mathbf{K}}_{\mathrm{P}}$. This observation is stated formally below, and allows us to replace the objective function in (5.35) with a simpler objective function on the left hand side in (5.36).

Claim 4 *Suppose that* $\bar{\mathbf{K}}_{\Phi}\succ\mathbf{0}$ *and define*

$$\mathcal{H}(\mathbf{K}_{\mathrm{P}}) \triangleq h(\mathbf{y}_{\mathrm{r}}-\bar{\mathbf{\Theta}}\mathbf{y}_{\mathrm{e}}). \tag{5.37}$$

Then,

$$\bar{\mathbf{K}}_{\mathrm{P}} \in \arg\max_{\mathbf{K}_{\mathrm{P}}\in\mathcal{K}_{\mathrm{P}}} \mathcal{H}(\mathbf{K}_{\mathrm{P}}). \tag{5.38}$$

The proof involves showing that $\bar{\mathbf{K}}_{\mathrm{P}}$, satisfies the Kuhn-Tucker conditions which we do in Appendix C.4.

Finally, to establish 2), we note that,

$$\bar{\mathbf{K}}_{\mathrm{P}} \in \arg\max_{\mathbf{K}_{\mathrm{P}}\in\mathcal{K}_{\mathrm{P}}} \mathcal{H}(\mathbf{K}_{\mathrm{P}}) \tag{5.39}$$

$$= \arg\max_{\mathbf{K}_{\mathrm{P}}\in\mathcal{K}_{\mathrm{P}}} \log\det(\mathbf{I}+\mathbf{J}^{-\frac{1}{2}}(\mathbf{H}_{\mathrm{r}}-\bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})\mathbf{K}_{\mathrm{P}}(\mathbf{H}_{\mathrm{r}}-\bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})^{\dagger}\mathbf{J}^{-\frac{1}{2}}), \tag{5.40}$$

where

$$\mathbf{J} \triangleq \mathbf{I}+\bar{\mathbf{\Theta}}\bar{\mathbf{\Theta}}^{\dagger}-\bar{\mathbf{\Theta}}\bar{\mathbf{\Phi}}^{\dagger}-\bar{\mathbf{\Phi}}\bar{\mathbf{\Theta}}^{\dagger}\succ\mathbf{0}$$

is an invertible matrix. We can interpret (5.40) as stating that $\bar{\mathbf{K}}_{\mathrm{P}}$ is an optimal input covariance for a MIMO channel with white noise and matrix $\mathbf{H}_{\mathrm{eff}} \triangleq \mathbf{J}^{-\frac{1}{2}}(\mathbf{H}_{\mathrm{r}}-\bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})$. The fact that $\mathbf{H}_{\mathrm{eff}}\mathbf{S}$ is a full rank matrix, then a consequence of the so called "water-filling" conditions. The proof is provided in Appendix C.5.

■

The conditions in Lemma 9 can be used in turn to establish the tightness of the upper bound in (5.17).

Lemma 10 *The saddle value in* (5.17) *can be expressed as follows,*

$$R_{\mathrm{UB}}(P) = \begin{cases} 0, & (\mathbf{H}_{\mathrm{r}}-\bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})=\mathbf{0}, \\ R_{-}(\bar{\mathbf{K}}_{\mathrm{P}}), & \textit{otherwise}, \end{cases} \tag{5.41}$$

where,

$$R_{-}(\bar{\mathbf{K}}_{\mathrm{P}}) \triangleq \log\det(\mathbf{I}+\mathbf{H}_{\mathrm{r}}\bar{\mathbf{K}}_{\mathrm{P}}\mathbf{H}_{\mathrm{r}}^{\dagger}) - \log\det(\mathbf{I}+\mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_{\mathrm{P}}\mathbf{H}_{\mathrm{e}}^{\dagger}). \tag{5.42}$$

Proof. The proof is most direct when we assume that the saddle point solution is such that $\bar{\mathbf{K}}_\Phi \succ \mathbf{0}$ i.e., when $||\mathbf{\Phi}||_2 < 1$. The extension when $\bar{\mathbf{K}}_\Phi$ is singular is provided in Appendix C.7.

First consider the case when $\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e} = \mathbf{0}$. From (5.23), it follows that $\bar{\mathbf{\Theta}} = \bar{\mathbf{\Phi}}$, using which one can establish the first part in (5.41):

$$\begin{aligned} R_+(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi) &= I(\mathbf{x}; \mathbf{y}_\mathrm{r}|\mathbf{y}_\mathrm{e}) && (5.43)\\ &= h(\mathbf{y}_\mathrm{r}|\mathbf{y}_\mathrm{e}) - h(\mathbf{z}_\mathrm{r}|\mathbf{z}_\mathrm{e}) \\ &= h(\mathbf{y}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{y}_\mathrm{e}) - h(\mathbf{z}_\mathrm{r} - \bar{\mathbf{\Phi}}\mathbf{z}_\mathrm{e}) && (5.44)\\ &= h(\mathbf{z}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{z}_\mathrm{e}) - h(\mathbf{z}_\mathrm{r} - \bar{\mathbf{\Phi}}\mathbf{z}_\mathrm{e}) && (5.45)\\ &= 0, \end{aligned}$$

where (5.44) follows from the fact that $\bar{\mathbf{\Theta}}$ in (5.23) is the linear minimum mean squared estimation (LMMSE) coefficient in estimation $\mathbf{y}_\mathrm{r}$ given $\mathbf{y}_\mathrm{e}$ and $\bar{\mathbf{\Phi}}$ is the LMMSE coefficient in estimating $\mathbf{z}_\mathrm{r}$ given $\mathbf{z}_\mathrm{e}$ and (5.45) follows via the relation $\mathbf{H}_\mathrm{r} = \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e}$, so that, $\mathbf{y}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{y}_\mathrm{e} = \mathbf{z}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{z}_\mathrm{e}$.

When $\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e} \neq \mathbf{0}$, combining parts (1) and (2) in Lemma 9, it follows that,

$$\bar{\mathbf{\Phi}}^\dagger \mathbf{H}_\mathrm{r} \mathbf{S} = \mathbf{H}_\mathrm{e} \mathbf{S}, \tag{5.46}$$

which can be used to establish the second case in (5.41) as we now do. In particular, we show that

$$\begin{aligned} \Delta R \triangleq &= R_+(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi) - R_-(\bar{\mathbf{K}}_\mathrm{P}) \\ &= I(\mathbf{x}; \mathbf{y}_\mathrm{r}|\mathbf{y}_\mathrm{e}) - \{I(\mathbf{x}; \mathbf{y}_\mathrm{r}) - I(\mathbf{x}; \mathbf{y}_\mathrm{e})\} \\ &= I(\mathbf{x}; \mathbf{y}_\mathrm{e}|\mathbf{y}_\mathrm{r}) \\ &= h(\mathbf{y}_\mathrm{e}|\mathbf{y}_\mathrm{r}) - h(\mathbf{z}_\mathrm{e}|\mathbf{z}_\mathrm{r}), \end{aligned}$$

equals zero. Indeed,

$$\begin{aligned} &h(\mathbf{y}_\mathrm{e} \mid \mathbf{y}_\mathrm{r}) \\ &= \log\det(\mathbf{I} + \mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger - \\ &\quad (\mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger + \bar{\mathbf{\Phi}}^\dagger)(\mathbf{H}_\mathrm{r}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger + \mathbf{I})^{-1}(\mathbf{H}_\mathrm{r}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger + \bar{\mathbf{\Phi}})) \\ &= \log\det(\mathbf{I} + \mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger - \bar{\mathbf{\Phi}}^\dagger(\mathbf{H}_\mathrm{r}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger + \mathbf{I})\bar{\mathbf{\Phi}}) \\ &= \log\det(\mathbf{I} - \bar{\mathbf{\Phi}}^\dagger\bar{\mathbf{\Phi}}) = h(\mathbf{z}_\mathrm{e} \mid \mathbf{z}_\mathrm{r}), \end{aligned} \tag{5.47}$$

where we have used the relation (5.46) in simplifying (5.47). This establishes the second half of (5.41). ∎

The proof of Theorem 13 is a direct consequence of Lemma 10. If $R_+(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi) = 0$, the capacity is zero, otherwise $R_+(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi) = R_-(\bar{\mathbf{K}}_\mathrm{P})$, and the latter expression is an achievable rate as can be seen by setting $p_u = p_\mathbf{x} = \mathcal{CN}(0, \bar{\mathbf{K}}_\mathrm{P})$ in the Csiszár-Körner

expression (5.6).

5.4 GSVD transform and High SNR Capacity

We begin with a definition of the generalized singular value decomposition [43, 34].

Definition 11 (GSVD Transform) *Given two matrices* $\mathbf{H}_{\mathrm{r}} \in \mathbb{C}^{n_r \times n_t}$ *and* $\mathbf{H}_{\mathrm{e}} \in \mathbb{C}^{n_e \times n_t}$, *there exist unitary matrices* $\mathbf{\Psi}_{\mathrm{r}} \in \mathbb{C}^{n_r \times n_r}$, $\mathbf{\Psi}_{\mathrm{e}} \in \mathbb{C}^{n_e \times n_e}$ *and* $\mathbf{\Psi}_{\mathrm{t}} \in \mathbb{C}^{n_t \times n_t}$, *a non-singular, lower triangular matrix* $\mathbf{\Omega} \in \mathbb{C}^{k \times k}$, *and two matrices* $\mathbf{\Sigma}_{\mathrm{r}} \in \mathbb{R}^{n_r \times k}$ *and* $\mathbf{\Sigma}_{\mathrm{e}} \in \mathbb{R}^{n_e \times k}$, *such that*

$$\mathbf{\Psi}_{\mathrm{r}}^{\dagger} \mathbf{H}_{\mathrm{r}} \mathbf{\Psi}_{\mathrm{t}} = \mathbf{\Sigma}_{\mathrm{r}} \left[\mathbf{\Omega}^{-1}, \mathbf{0}_{k \times n_t - k} \right], \tag{5.48a}$$

$$\mathbf{\Psi}_{\mathrm{e}}^{\dagger} \mathbf{H}_{\mathrm{e}} \mathbf{\Psi}_{\mathrm{t}} = \mathbf{\Sigma}_{\mathrm{e}} \left[\mathbf{\Omega}^{-1}, \mathbf{0}_{k \times n_t - k} \right], \tag{5.48b}$$

where the matrices $\mathbf{\Sigma}_{\mathrm{r}}$ *and* $\mathbf{\Sigma}_{\mathrm{e}}$ *have the following structure,*

$$\mathbf{\Sigma}_{\mathrm{r}} = \begin{array}{c} \\ n_{\mathrm{r}}-p-s \\ s \\ p \end{array} \begin{array}{c} \begin{array}{ccc} k-p-s & s & p \end{array} \\ \begin{bmatrix} \mathbf{0} & & \\ & \mathbf{D}_{\mathrm{r}} & \\ & & \mathbf{I} \end{bmatrix} \end{array}, \tag{5.49a}$$

$$\mathbf{\Sigma}_{\mathrm{e}} = \begin{array}{c} \\ k-p-s \\ s \\ n_{\mathrm{e}}+p-k \end{array} \begin{array}{c} \begin{array}{ccc} k-p-s & s & p \end{array} \\ \begin{bmatrix} \mathbf{I} & & \\ & \mathbf{D}_{\mathrm{e}} & \\ & & \mathbf{0} \end{bmatrix} \end{array}, \tag{5.49b}$$

and the constants

$$k = \mathrm{rank}\left(\begin{bmatrix} \mathbf{H}_{\mathrm{r}} \\ \mathbf{H}_{\mathrm{e}} \end{bmatrix} \right), p = \dim\left(\mathrm{Null}(\mathbf{H}_{\mathrm{e}}) \bigcap \mathrm{Null}(\mathbf{H}_{\mathrm{r}})^{\perp} \right), \tag{5.50}$$

and s *depend on the matrices* $\mathbf{H}_{\mathrm{r}}$ *and* $\mathbf{H}_{\mathrm{e}}$. *The matrices*

$$\mathbf{D}_{\mathrm{r}} = \mathrm{diag}\{r_1, \ldots, r_s\}, \quad \mathbf{D}_{\mathrm{e}} = \mathrm{diag}\{e_1, \ldots, e_s\}, \tag{5.51}$$

are diagonal matrices with strictly positive entries, and the generalized singular values are given by

$$\sigma_i = \frac{r_i}{e_i}, \quad i = 1, 2, \ldots, s. \tag{5.52}$$

We provide a few properties of the GSVD-transform that are used in the sequel.

1. The GSVD transform provides a characterization of the null space of $\mathbf{H}_{\mathrm{e}}$. Let

$$\mathbf{\Psi}_{\mathrm{t}} = [\psi_1, \ldots, \psi_{n_t}], \tag{5.53}$$

where $\mathbf{\Psi}_{\mathrm{t}}$ is defined via (5.48). Then

$$\mathcal{S}_{\mathrm{n}} = \mathrm{Null}(\mathbf{H}_{\mathrm{e}}) \bigcap \mathrm{Null}(\mathbf{H}_{\mathrm{r}}) = \mathrm{span}\{\psi_{k+1}, \ldots, \psi_{n_t}\} \tag{5.54a}$$

$$\mathcal{S}_{\mathrm{z}} = \mathrm{Null}(\mathbf{H}_{\mathrm{e}}) \bigcap \mathrm{Null}(\mathbf{H}_{\mathrm{r}})^{\perp} = \mathrm{span}\{\psi_{k-p+1}, \ldots, \psi_k\} \tag{5.54b}$$

Indeed, it can be readily verified from (5.48) that

$$\mathbf{H}_{\mathrm{r}}\psi_j = \mathbf{H}_{\mathrm{e}}\psi_j = \mathbf{0}, \quad j = k+1, \ldots, n_t, \tag{5.55}$$

which establishes (5.54a). To establish (5.54b), we will show that for each j such that $k-p+1 \leq j \leq k$, $\mathbf{H}_{\mathrm{e}}\psi_j = \mathbf{0}$ and $\{\mathbf{H}_{\mathrm{r}}\psi_j\}$ are linearly independent. It suffices to show that the last p columns of $\mathbf{\Sigma}_{\mathrm{r}}\mathbf{\Omega}^{-1}$ are linearly independent and the last p columns of $\mathbf{\Sigma}_{\mathrm{e}}\mathbf{\Omega}^{-1}$ are zero. Note that since $\mathbf{\Omega}^{-1}$ in (5.48) is a lower triangular matrix, we can express it as

$$\mathbf{\Omega}^{-1} = \begin{array}{c} \\ k-p-s \\ s \\ p \end{array} \begin{array}{c} \begin{array}{ccc} k-p-s & s & p \end{array} \\ \begin{bmatrix} \mathbf{\Omega}_1^{-1} & & \\ \mathbf{T}_{21} & \mathbf{\Omega}_2^{-1} & \\ \mathbf{T}_{31} & \mathbf{T}_{32} & \mathbf{\Omega}_3^{-1} \end{bmatrix} \end{array}. \tag{5.56}$$

By direct block multiplication with (5.49a) and (5.49b), we have,

$$\mathbf{\Sigma}_{\mathrm{r}}\mathbf{\Omega}^{-1} = \begin{array}{c} \\ n_r-s-p \\ s \\ p \end{array} \begin{array}{c} \begin{array}{ccc} k-s-p & s & p \end{array} \\ \begin{bmatrix} \mathbf{0} & & \\ \mathbf{D}_{\mathrm{r}}\mathbf{T}_{21} & \mathbf{D}_{\mathrm{r}}\mathbf{\Omega}_2^{-1} & \\ \mathbf{T}_{31} & \mathbf{T}_{32} & \mathbf{\Omega}_3^{-1} \end{bmatrix} \end{array} \tag{5.57a}$$

$$\mathbf{\Sigma}_{\mathrm{e}}\mathbf{\Omega}^{-1} = \begin{array}{c} \\ k-p-s \\ s \\ n_e+p-k \end{array} \begin{array}{c} \begin{array}{ccc} k-s-p & s & p \end{array} \\ \begin{bmatrix} \mathbf{\Omega}_1^{-1} & & \\ \mathbf{D}_{\mathrm{e}}\mathbf{T}_{21} & \mathbf{D}_{\mathrm{e}}\mathbf{\Omega}_2^{-1} & \\ & & \mathbf{0} \end{bmatrix} \end{array} \tag{5.57b}$$

Since $\mathbf{\Omega}_3$ is invertible, the last p columns of $\mathbf{\Sigma}_{\mathrm{r}}\mathbf{\Omega}^{-1}$ are linearly independent and clearly the last p columns of $\mathbf{\Sigma}_{\mathrm{e}}\mathbf{\Omega}^{-1}$ are zero establishing (5.54b).

Furthermore,

$$\mathrm{Null}(\mathbf{H}_{\mathrm{e}}) = \mathrm{span}\{\psi_{k-p+1}, \ldots, \psi_{n_t}\}. \tag{5.58}$$

Hence if $\mathbf{\Psi}_{\mathrm{ne}} = [\psi_{k-p+1}, \ldots, \psi_{n_t}]$, then the projection matrix on to the $\mathrm{Null}(\mathbf{H}_{\mathrm{e}})$,

$$\mathbf{H}_{\mathrm{e}}^{\perp} = \mathbf{\Psi}_{\mathrm{ne}}\mathbf{\Psi}_{\mathrm{ne}}^{\dagger}. \tag{5.59}$$

Also from (5.48) and (5.57a), note that

$$\mathbf{H}_{\mathrm{r}}\mathbf{\Psi}_{\mathrm{ne}} = \mathbf{\Psi}_{\mathrm{r}} \left\{ \begin{array}{c} \\ n_r-p \\ p \\ \\ \end{array} \begin{array}{c} \begin{array}{cc} p & n_t-k \end{array} \\ \begin{bmatrix} \mathbf{0} & \\ \mathbf{\Omega}_3^{-1} & \mathbf{0} \end{bmatrix} \end{array} \right\}, \tag{5.60}$$

and hence,

$$\mathbf{H}_{\mathrm{r}}\mathbf{H}_{\mathrm{e}}^{\perp}\mathbf{H}_{\mathrm{r}}^{\dagger} = \mathbf{\Psi}_{\mathrm{r}} \left\{ \begin{array}{c} \\ n_r-p \\ p \\ \\ \end{array} \overset{\begin{array}{cc} n_r-p & p \end{array}}{\left[\begin{array}{cc} \mathbf{0} & \\ & \mathbf{\Omega}_3^{-1}\mathbf{\Omega}_3^{-\dagger} \end{array} \right]} \right\} \mathbf{\Psi}_{\mathrm{r}}^{\dagger}, \tag{5.61}$$

denotes the projection of $\mathbf{H}_{\mathrm{r}}$ onto the null space of $\mathbf{H}_{\mathrm{e}}$.

2. The GSVD definition simplifies considerably when the matrix $\mathbf{H}_{\mathrm{e}}$ has a full column rank. In this case note from (5.50) that $p = 0$ and $k = n_t$. Defining, $\mathbf{A} = \mathbf{\Psi}_{\mathrm{t}}\mathbf{\Omega}$, we note from (5.48) that

$$\mathbf{\Psi}_{\mathrm{r}}^{\dagger}\mathbf{H}_{\mathrm{r}}\mathbf{A} = \mathbf{\Sigma}_{\mathrm{r}}, \quad \mathbf{\Psi}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}\mathbf{A} = \mathbf{\Sigma}_{\mathrm{e}}, \tag{5.62}$$

where $\mathbf{\Sigma}_{\mathrm{r}}$ and $\mathbf{\Sigma}_{\mathrm{e}}$ have the form:

$$\mathbf{\Sigma}_{\mathrm{r}} = \begin{array}{c} \\ n_{\mathrm{r}}-s \\ s \\ \end{array} \overset{\begin{array}{cc} n_{\mathrm{t}}-s & s \end{array}}{\left[\begin{array}{cc} \mathbf{0} & \\ & \mathbf{D}_{\mathrm{r}} \end{array} \right]}, \mathbf{\Sigma}_{\mathrm{e}} = \begin{array}{c} n_{\mathrm{t}}-s \\ s \\ n_{\mathrm{e}}-n_{\mathrm{t}} \\ \end{array} \overset{\begin{array}{cc} n_{\mathrm{t}}-s & s \end{array}}{\left[\begin{array}{cc} \mathbf{I} & \\ & \mathbf{D}_{\mathrm{e}} \\ & \mathbf{0} \end{array} \right]}, \tag{5.63}$$

and $\mathbf{D}_{\mathrm{r}}$ and $\mathbf{D}_{\mathrm{e}}$ are diagonal matrices with positive entries (c.f. (5.51)).

Also if $\mathbf{H}_{\mathrm{e}}^{\ddagger}$ denotes the Moore-Penrose pseudo-inverse of $\mathbf{H}_{\mathrm{e}}$,

$$\mathbf{H}_{\mathrm{e}}^{\ddagger} = \mathbf{A} \left\{ \begin{array}{c} \\ n_t-s \\ s \\ \\ \end{array} \overset{\begin{array}{ccc} n_t-s & s & n_e-n_t \end{array}}{\left[\begin{array}{ccc} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\mathrm{e}}^{-1} & \mathbf{0} \end{array} \right]} \right\} \mathbf{\Psi}_{\mathrm{e}}^{\dagger} \tag{5.64}$$

and $\mathbf{H}_{\mathrm{e}}^{\ddagger}\mathbf{H}_{\mathrm{e}} = \mathbf{I}$. From (5.62), (5.63) and (5.64),

$$\mathbf{H}_{\mathrm{r}}\mathbf{H}_{\mathrm{e}}^{\ddagger} = \mathbf{\Psi}_{\mathrm{r}} \left\{ \begin{array}{c} \\ n_{\mathrm{r}}-s \\ s \\ \\ \end{array} \overset{\begin{array}{ccc} n_{\mathrm{t}}-s & s & n_{\mathrm{e}}-n_{\mathrm{t}} \end{array}}{\left[\begin{array}{ccc} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\mathrm{r}}\mathbf{D}_{\mathrm{e}}^{-1} & \mathbf{0} \end{array} \right]} \right\} \mathbf{\Psi}_{\mathrm{e}}^{\dagger}, \tag{5.65}$$

i.e., the generalized singular values of $(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}})$ in (5.52) are also the (ordinary) singular values of $\mathbf{H}_{\mathrm{r}}\mathbf{H}_{\mathrm{e}}^{\ddagger}$.

5.4.1 Derivation of the High SNR Capacity Expression

For simplicity we first consider the case when $\mathbf{H}_{\mathrm{e}}$ has a full column rank. In this case, it is clear that the condition in (5.10) is satisfied and accordingly we establish (5.11).

The achievability part follows by simultaneously diagonalizing the channel matrices $\mathbf{H}_{\mathrm{r}}$ and $\mathbf{H}_{\mathrm{e}}$ using the GSVD transform. This reduces the system into a set

of parallel independent channels and independent codebooks are used across these channels. More specifically, recall that in the case of interest, the transform is given in (5.62). Let $\sigma_1 \leq \sigma_2 \leq \ldots \leq \sigma_s$ be the ordered set of singular values and suppose that $\sigma_i > 1$ for $i \geq \nu$. We select the following choices for $\mathbf{x}$ and $\mathbf{u}$ in the Csiszár and Körner expression (5.6)

$$\mathbf{x} = \mathbf{A}\begin{bmatrix}\mathbf{0}_{n_t-s}\\ \mathbf{u}\end{bmatrix}, \mathbf{u} = [0,\ldots,0,u_\nu,u_{\nu+1},\ldots,u_s], \tag{5.66}$$

and the random variables u_i are sampled i.i.d. according to $\mathcal{CN}(0,\alpha P)$. Here $\alpha = \frac{1}{n_t\sigma_{\max}(\mathbf{A})}$ is selected so that the average power constraint is satisfied. Substituting (5.66) and (5.62) into the channel model (5.1) yields,

$$\mathbf{y}_r = \mathbf{\Psi}_r\begin{bmatrix}\mathbf{0}_{n_t-s}\\ \mathbf{D}_r\mathbf{u}\end{bmatrix} + \mathbf{z}_r, \quad \mathbf{y}_e = \mathbf{\Psi}_e\begin{bmatrix}\mathbf{0}_{n_t-s}\\ \mathbf{D}_e\mathbf{u}\\ \mathbf{0}_{n_e-n_t}\end{bmatrix} + \mathbf{z}_e. \tag{5.67}$$

Since $\mathbf{\Psi}_r$ and $\mathbf{\Psi}_e$ are unitary, and $\mathbf{D}_r$ and $\mathbf{D}_e$ are diagonal, the system of equations (5.67) indeed represents a parallel channel model. See Fig. 5-1 for an illustration of the 2-2-2 case. The achievable rate obtained by substituting (5.67) and (5.66) into (5.6), is

$$\begin{aligned} R &= I(\mathbf{u};\mathbf{y}_r) - I(\mathbf{u};\mathbf{y}_e) && (5.68)\\ &= \sum_{j=\nu}^{n_t}\log\frac{1+\alpha P r_j^2}{1+\alpha P e_j^2}\\ &= \sum_{j:\sigma_j>1}\log\sigma_j^2 - o_P(1), && (5.69)\end{aligned}$$

where $o_P(1) \to 0$ as $P \to \infty$.

For the converse we begin with a more convenient upper bound expression to the secrecy capacity (5.17),

$$\begin{aligned} R_{\text{UB}} &= \min_{\substack{\mathbf{\Phi}:\|\mathbf{\Phi}\|_2\leq 1\\ \mathbf{\Theta}\in\mathbb{C}^{n_r\times n_t}}}\max_{\tilde{\mathbf{K}}_P\in\mathcal{K}_P} R_{++}(\mathbf{K}_P,\mathbf{\Theta},\mathbf{\Phi})\\ R_{++} &= \log\frac{\det(\mathbf{H}_{\text{eff}}\mathbf{K}_P\mathbf{H}_{\text{eff}}^\dagger + \mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^\dagger - \mathbf{\Theta}\mathbf{\Phi}^\dagger - \mathbf{\Phi}\mathbf{\Theta}^\dagger)}{\det(\mathbf{I}-\mathbf{\Phi}\mathbf{\Phi}^\dagger)}, && (5.70)\\ \mathbf{H}_{\text{eff}} &= \mathbf{H}_r - \mathbf{\Theta}\mathbf{H}_e. \end{aligned}$$

This expression, as an upper bound, was suggested to us by Y. Eldar and A. Wiesel and was first used in establishing the secrecy capacity of the MISOME channel in [25]. To establish (5.70), first note that the objective function $R_+(\mathbf{K}_P,\mathbf{K}_\Phi)$ in (5.17) can

be upper bounded as follows:

$$\begin{aligned}
&R_+(\mathbf{K}_\mathrm{P}, \mathbf{K}_\Phi) \\
&= I(\mathbf{x}; \mathbf{y}_\mathrm{r}|\mathbf{y}_\mathrm{e}) \\
&= h(\mathbf{y}_\mathrm{r}|\mathbf{y}_\mathrm{e}) - h(\mathbf{z}_\mathrm{r}|\mathbf{z}_\mathrm{e}) \\
&= h(\mathbf{y}_\mathrm{r}|\mathbf{y}_\mathrm{e}) - \log\det(\mathbf{I} - \mathbf{\Phi}\mathbf{\Phi}^\dagger) \\
&= \min_{\mathbf{\Theta}} h(\mathbf{y}_\mathrm{r} - \mathbf{\Theta}\mathbf{y}_\mathrm{e}) - \log\det(\mathbf{I} - \mathbf{\Phi}\mathbf{\Phi}^\dagger) \\
&= \min_{\mathbf{\Theta}} R_{++}(\mathbf{K}_\mathrm{P}, \mathbf{\Theta}, \mathbf{\Phi}).
\end{aligned}$$

Thus, we have from (5.17) that

$$R_+(P) = \min_{\mathcal{K}_\Phi} \max_{\mathcal{K}_\mathrm{P}} R_+(\mathbf{K}_\mathrm{P}, \mathbf{K}_\Phi) \tag{5.71}$$

$$= \min_{\mathcal{K}_\Phi} \max_{\mathcal{K}_\mathrm{P}} \min_{\mathbf{\Theta}} R_{++}(\mathbf{K}_\mathrm{P}, \mathbf{\Theta}, \mathbf{\Phi}) \tag{5.72}$$

$$\leq \min_{\mathcal{K}_\Phi} \min_{\mathbf{\Theta}} \max_{\mathcal{K}_\mathrm{P}} R_{++}(\mathbf{K}_\mathrm{P}, \mathbf{\Theta}, \mathbf{\Phi}), \tag{5.73}$$

as required.

To establish the capacity, we show that the upper bound in (5.70) above, reduces to the capacity expression (5.11), for a specific choice of $\mathbf{\Theta}$ and $\mathbf{\Phi}$ as stated below.

Our choice of parameters in the minimization of (5.70) is as follows

$$\mathbf{\Theta} = \mathbf{H}_\mathrm{r}\mathbf{H}_\mathrm{e}^\ddagger, \quad \mathbf{\Phi} = \mathbf{\Psi}_\mathrm{r} \begin{Bmatrix} \\ n_\mathrm{r}-s \\ s \\ \\ \end{Bmatrix} \overset{\begin{matrix} n_\mathrm{t}-s & s & n_\mathrm{e}-n_\mathrm{t} \end{matrix}}{\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Delta} & \mathbf{0} \end{bmatrix}} \mathbf{\Psi}_\mathrm{e}^\dagger, \tag{5.74}$$

where,

$$\mathbf{\Delta} = \mathrm{diag}\{\delta_1, \delta_2, \ldots, \delta_s\}, \qquad \delta_i = \min\left(\sigma_i, \frac{1}{\sigma_i}\right), \tag{5.75}$$

and $\mathbf{H}_\mathrm{e}^\ddagger$ denotes the Moore-Penrose pesudo-inverse of $\mathbf{H}_\mathrm{e}$ (c.f. 5.64). Note that with these choice of parameters, $\mathbf{H}_\mathrm{eff} = \mathbf{0}$. So the maximization over $\mathbf{K}_\mathrm{P}$ in (5.70) is not effective. Simplifying (5.70) with these choice of parameters the upper bound expression reduces to

$$R_{++} \leq \log \frac{\det(\mathbf{I} + (\mathbf{D}_\mathrm{r}\mathbf{D}_\mathrm{e}^{-1})^2 - 2\mathbf{D}_\mathrm{r}\mathbf{D}_\mathrm{e}^{-1}\mathbf{\Delta})}{\det(\mathbf{I} - \mathbf{\Delta}^2)} = \sum_{j:\sigma_j>1} \log \sigma_j^2.$$

as in (5.11).

When $\mathbf{H}_\mathrm{e}$ does not have a full column rank, the capacity result in (5.12) will now

be established. To show the achievability, we identify the subspaces

$$\begin{aligned}\mathcal{S}_{\mathrm{z}} &= \mathrm{Null}(\mathbf{H}_{\mathrm{e}}) \bigcap \mathrm{Null}(\mathbf{H}_{\mathrm{r}})^{\perp} = \mathrm{span}\{\psi_{k-p+1}, \ldots, \psi_k\} \\ \mathcal{S}_{\mathrm{s}} &= \mathrm{Null}(\mathbf{H}_{\mathrm{e}})^{\perp} \bigcap \mathrm{Null}(\mathbf{H}_{\mathrm{r}})^{\perp} = \mathrm{span}\{\psi_{k-p-s+1}, \ldots, \psi_{k-p}\}.\end{aligned} \tag{5.76}$$

We will use most of the power for transmission in the subspace $\mathcal{S}_{\mathrm{z}}$ and a small fraction of power for transmissoin in the subspace $\mathcal{S}_{\mathrm{s}}$. More specifically, by selecting,

$$\mathbf{x} = \mathbf{\Psi}_{\mathrm{t}} \begin{bmatrix} \mathbf{0}_{k-p-s} \\ \mathbf{\Omega}_2 \mathbf{u} \\ \mathbf{v} \\ \mathbf{0}_{n_t-k} \end{bmatrix}, \tag{5.77}$$

we have,

$$\begin{aligned}\mathbf{y}_{\mathrm{r}} &= \mathbf{\Psi}_{\mathrm{r}} \begin{bmatrix} \mathbf{0}_{n_r-p-s} \\ \mathbf{D}_{\mathrm{r}} \mathbf{u} \\ \mathbf{T}_{32} \mathbf{\Omega}_2^{-1} \mathbf{u} + \mathbf{\Omega}_3^{-1} \mathbf{v} \end{bmatrix} + \mathbf{z}_{\mathrm{r}}, \\ \mathbf{y}_{\mathrm{e}} &= \mathbf{\Psi}_{\mathrm{e}} \begin{bmatrix} \mathbf{0}_{\mathrm{k-p-s}} \\ \mathbf{D}_{\mathrm{e}} \mathbf{u} \\ \mathbf{0}_{n_e+p-k} \end{bmatrix} + \mathbf{z}_{\mathrm{e}}.\end{aligned} \tag{5.78}$$

In (5.77), we select $\mathbf{v} = [v_1, v_2, \ldots, v_p]^T$ to be a vector of i.i.d. Gaussian random variables with a distribution $\mathcal{CN}\left(0, \frac{P-\sqrt{P}}{p}\right)$ and $\mathbf{u} = [0, \ldots, 0, u_\nu, \ldots, u_s]^T$ to be a vector of independent Gausian random variables. Here ν is the smallest integer such that $\sigma_j > 1$ for all $j \geq \nu$ and $\sigma_j \leq 1$ otherwise. Each $u_j \sim \mathcal{CN}(0, \alpha\sqrt{P})$, where $\alpha = \frac{1}{n_t \sigma_{\max}(\mathbf{\Omega}_2)}$, is chosen to meet the power constraint.

An achievable rate for this choice of parameters is

$$R = I(\mathbf{u}, \mathbf{v}; \mathbf{y}_{\mathrm{r}}) - I(\mathbf{u}, \mathbf{v}; \mathbf{y}_{\mathrm{e}}) \tag{5.79}$$

$$= I(\mathbf{u}; \mathbf{y}_{\mathrm{r}}) - I(\mathbf{u}; \mathbf{y}_{\mathrm{e}}) + I(\mathbf{v}; \mathbf{y}_{\mathrm{r}} | \mathbf{u}), \tag{5.80}$$

where the last step follows from the fact that $\mathbf{v}$ is independent of $(\mathbf{y}_{\mathrm{e}}, \mathbf{u})$ (c.f. (5.78)). Following (5.69), we have that

$$I(\mathbf{u}; \mathbf{y}_{\mathrm{r}}) - I(\mathbf{u}; \mathbf{y}_{\mathrm{e}}) = \sum_{j:\sigma_j > 1} \log \sigma_j^2 - o_P(1) \tag{5.81}$$

and

$$I(\mathbf{v};\mathbf{y}_{\mathrm{r}}|\mathbf{u}) = \log\det\left(\mathbf{I} + \frac{P-\sqrt{P}}{p}\boldsymbol{\Omega}_3^{-1}\boldsymbol{\Omega}_3^{-\dagger}\right) \tag{5.82}$$

$$= \log\det\left(\mathbf{I} + \frac{P}{p}\boldsymbol{\Omega}_3^{-1}\boldsymbol{\Omega}_3^{-\dagger}\right) - o_P(1) \tag{5.83}$$

$$= \log\det\left(\mathbf{I} + \frac{P}{p}\mathbf{H}_{\mathrm{r}}\mathbf{H}_{\mathrm{e}}^{\perp}\mathbf{H}_{\mathrm{r}}^{\dagger}\right) - o_P(1), \tag{5.84}$$

where (5.83) follows from the fact that $\log(1+x)$ is a continuous function of x and $\log\det(\mathbf{I}+\mathbf{X}) = \sum\log(1+\lambda_i(\mathbf{X}))$ and the last step follows from (5.61).

To establish the upper bound, use the following choices for $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$ in (5.70).

$$\boldsymbol{\Theta} = \boldsymbol{\Psi}_{\mathrm{r}}\begin{Bmatrix} & \overset{k-s-p}{} & \overset{s}{} & \overset{n_e+p-k}{} \\ \overset{n_r-s-p}{} & \mathbf{0} & & \\ \overset{s}{} & & \mathbf{D}_{\mathrm{r}}\mathbf{D}_{\mathrm{e}}^{-1} & \\ \overset{p}{} & \mathbf{F}_{31} & \mathbf{F}_{32} & \mathbf{0} \end{Bmatrix}\boldsymbol{\Psi}_{\mathrm{e}}^{\dagger}, \tag{5.85}$$

and

$$\boldsymbol{\Phi} = \boldsymbol{\Psi}_{\mathrm{r}}\begin{Bmatrix} & \overset{k-s-p}{} & \overset{s}{} & \overset{n_e+p-k}{} \\ \overset{n_r-s-p}{} & \mathbf{0} & & \\ \overset{s}{} & & \boldsymbol{\Delta} & \\ \overset{p}{} & & & \mathbf{0} \end{Bmatrix}\boldsymbol{\Psi}_{\mathrm{e}}^{\dagger} \tag{5.86}$$

where $\boldsymbol{\Delta}$ is defined in (5.75), and the matrices

$$\begin{aligned} \mathbf{F}_{32} &= \mathbf{T}_{32}\boldsymbol{\Omega}_2\mathbf{D}_{\mathrm{e}}^{-1} \\ \mathbf{F}_{31} &= (\mathbf{T}_{31} - \mathbf{F}_{32}\mathbf{D}_{\mathrm{e}}\mathbf{T}_{21})\boldsymbol{\Omega}_1 \end{aligned} \tag{5.87}$$

are selected such that

$$\begin{aligned} &\mathbf{H}_{\mathrm{r}} - \boldsymbol{\Theta}\mathbf{H}_{\mathrm{e}} \\ &= \boldsymbol{\Psi}_{\mathrm{r}}([\boldsymbol{\Sigma}_{\mathrm{r}}\boldsymbol{\Omega}^{-1}, \mathbf{0}_{n_r\times n_t-k}] \\ &\quad - \boldsymbol{\Psi}_{\mathrm{r}}^{\dagger}\boldsymbol{\Theta}\boldsymbol{\Psi}_{\mathrm{e}}[\boldsymbol{\Sigma}_{\mathrm{e}}\boldsymbol{\Omega}^{-1}, \mathbf{0}_{n_e\times n_t-k}])\boldsymbol{\Psi}_{\mathrm{t}}^{\dagger} \end{aligned} \tag{5.88}$$

$$= \boldsymbol{\Psi}_{\mathrm{r}}\begin{Bmatrix} & \overset{k-p-s}{} & \overset{s}{} & \overset{p}{} & \overset{n_t-k}{} \\ \overset{n_r-s-p}{} & \mathbf{0} & & & \\ \overset{s}{} & & \mathbf{0} & & \\ \overset{p}{} & & & \boldsymbol{\Omega}_3^{-1} & \mathbf{0} \end{Bmatrix}\boldsymbol{\Psi}_{\mathrm{t}}^{\dagger}. \tag{5.89}$$

$$\mathbf{I} + \mathbf{H}_{\text{eff}}\boldsymbol{\Theta}\mathbf{H}_{\text{eff}}^{\dagger} + \boldsymbol{\Theta}\boldsymbol{\Theta}^{\dagger} - \boldsymbol{\Theta}\boldsymbol{\Phi}^{\dagger} - \boldsymbol{\Phi}\boldsymbol{\Theta}^{\dagger}$$

$$= \boldsymbol{\Psi}_{\text{r}} \left\{ \begin{array}{c} \\ n_r-s-p \\ s \\ p \end{array} \begin{array}{c} \begin{array}{ccc} \overset{n_r-s-p}{} & \overset{s}{} & \overset{p}{} \end{array} \\ \left[\begin{array}{ccc} \mathbf{I} & & \\ & \mathbf{I} + (\mathbf{D}_{\text{r}}\mathbf{D}_{\text{e}}^{-1})^2 - 2\mathbf{D}_{\text{r}}\mathbf{D}_{\text{e}}^{-1}\boldsymbol{\Delta} & (\mathbf{D}_{\text{r}}\mathbf{D}_{\text{e}}^{-1} - \boldsymbol{\Delta})\mathbf{F}_{32}^{\dagger} \\ & \mathbf{F}_{32}(\mathbf{D}_{\text{r}}\mathbf{D}_{\text{e}}^{-1} - \boldsymbol{\Delta}) & \mathbf{I} + \mathbf{F}_{31}\mathbf{F}_{31}^{\dagger} + \mathbf{F}_{32}\mathbf{F}_{32}^{\dagger} + \boldsymbol{\Omega}_3^{-1}\mathbf{Q}\boldsymbol{\Omega}_3^{-\dagger} \end{array} \right] \end{array} \right\} \boldsymbol{\Psi}_{\text{r}}^{\dagger} \tag{5.92}$$

The upper bound expression (5.70) can now be simplified as follows.

$$\mathbf{H}_{\text{eff}}\mathbf{K}_{\text{P}}\mathbf{H}_{\text{eff}}^{\dagger} = (\mathbf{H}_{\text{r}} - \boldsymbol{\Theta}\mathbf{H}_{\text{e}})\mathbf{K}_{\text{P}}(\mathbf{H}_{\text{r}} - \boldsymbol{\Theta}\mathbf{H}_{\text{e}})^{\dagger}$$

$$= \boldsymbol{\Psi}_{\text{r}} \left\{ \begin{array}{c} \\ n_r-p-s \\ s \\ p \end{array} \begin{array}{c} \begin{array}{ccc} \overset{n_r-p-s}{} & \overset{s}{} & \overset{p}{} \end{array} \\ \left[\begin{array}{ccc} \mathbf{0} & & \\ & \mathbf{0} & \\ & & \boldsymbol{\Omega}_3^{-1}\mathbf{Q}\boldsymbol{\Omega}_3^{-\dagger} \end{array} \right] \end{array} \right\} \boldsymbol{\Psi}_{\text{r}}^{\dagger}, \tag{5.90}$$

where $\mathbf{Q}$ is related to $\mathbf{K}_{\text{P}}$ by,

$$\boldsymbol{\Psi}_{\text{t}}^{\dagger}\mathbf{K}_{\text{P}}\boldsymbol{\Psi}_t = \begin{array}{c} \\ k-p-s \\ s \\ p \\ n_t-k \end{array} \begin{array}{c} \begin{array}{cccc} \overset{k-p-s}{} & \overset{s}{} & \overset{p}{} & \overset{n_t-k}{} \end{array} \\ \left[\begin{array}{cccc} \star & \star & \star & \star \\ \star & \star & \star & \star \\ \star & \star & \mathbf{Q} & \star \\ \star & \star & \star & \star \end{array} \right] \end{array} \tag{5.91}$$

and satisfies $\text{tr}(\mathbf{Q}) \leq P$. From (5.90), (5.86) and (5.85), we have that the numerator in the upper bound expression (5.70) simplifies as in (5.92).

Using (5.92) and the Hardamard inequality, we have

$$\begin{aligned} &\log\det(\mathbf{I} + \mathbf{H}_{\text{eff}}\boldsymbol{\Theta}\mathbf{H}_{\text{eff}}^{\dagger} + \boldsymbol{\Theta}\boldsymbol{\Theta}^{\dagger} - \boldsymbol{\Theta}\boldsymbol{\Phi}^{\dagger} - \boldsymbol{\Phi}\boldsymbol{\Theta}^{\dagger}) \\ &\leq \log\det(\mathbf{I} + (\mathbf{D}_{\text{r}}\mathbf{D}_{\text{e}}^{-1})^2 - 2\mathbf{D}_{\text{r}}\mathbf{D}_{\text{e}}^{-1}\boldsymbol{\Delta}) \\ &\quad + \log\det(\mathbf{I} + \mathbf{F}_{31}\mathbf{F}_{31}^{\dagger} + \mathbf{F}_{32}\mathbf{F}_{32}^{\dagger} + \boldsymbol{\Omega}_3^{-1}\mathbf{Q}\boldsymbol{\Omega}_3^{-\dagger}) \end{aligned} \tag{5.93}$$

Substituting this relation in (5.70), the upper bound reduces to,

$$\begin{aligned} R_{+}(P) \leq &\log\frac{\det(\mathbf{I} + (\mathbf{D}_{\text{r}}\mathbf{D}_{\text{e}}^{-1})^2 - 2\mathbf{D}_{\text{r}}\mathbf{D}_{\text{e}}^{-1}\boldsymbol{\Delta})}{\det(\mathbf{I} - \boldsymbol{\Delta}^2)} \\ &+ \max_{\substack{\mathbf{Q}\succeq 0: \\ \text{tr}(\mathbf{Q})\leq P}} \log\det(\mathbf{I} + \mathbf{F}_{31}\mathbf{F}_{31}^{\dagger} + \mathbf{F}_{32}\mathbf{F}_{32}^{\dagger} + \boldsymbol{\Omega}_3^{-1}\mathbf{Q}\boldsymbol{\Omega}_3^{-\dagger}) \end{aligned} \tag{5.94}$$

Substituting for $\mathbf{D}_{\mathrm{r}}$ and $\mathbf{D}_{\mathrm{e}}$ from (5.51) and for $\boldsymbol{\Delta}$ from (5.75), we have that

$$\log \frac{\det(\mathbf{I} + (\mathbf{D}_{\mathrm{r}}\mathbf{D}_{\mathrm{e}}^{-1})^2 - 2\mathbf{D}_{\mathrm{r}}\mathbf{D}_{\mathrm{e}}^{-1}\boldsymbol{\Delta})}{\det(\mathbf{I} - \boldsymbol{\Delta}^2)} = \sum_{j:\sigma_j>1} \log \sigma_j^2. \tag{5.95}$$

It remains to establish that

$$\begin{aligned} &\max_{\substack{\mathbf{Q}\succeq 0:\\ \mathrm{tr}(\mathbf{Q})\leq P}} \log\det(\mathbf{I} + \mathbf{F}_{31}\mathbf{F}_{31}^{\dagger} + \mathbf{F}_{32}\mathbf{F}_{32}^{\dagger} + \boldsymbol{\Omega}_3^{-1}\mathbf{Q}\boldsymbol{\Omega}_3^{-\dagger}) \\ &\qquad \leq \log\det\left(\mathbf{I} + \frac{P}{p}\mathbf{H}_{\mathrm{r}}\mathbf{H}_{\mathrm{e}}^{\perp}\mathbf{H}_{\mathrm{r}}^{\dagger}\right) + o_P(1), \end{aligned} \tag{5.96}$$

which we now do.

Let

$$\gamma = \sigma_{\max}(\mathbf{F}_{31}\mathbf{F}_{31}^{\dagger} + \mathbf{F}_{32}\mathbf{F}_{32}^{\dagger}), \tag{5.97}$$

denote the largest singular value of the matrix $\mathbf{F}_{31}\mathbf{F}_{31}^{\dagger} + \mathbf{F}_{32}\mathbf{F}_{32}^{\dagger}$. Since log-det is increasing on the cone of positive semidefinite matrices, we have that,

$$\begin{aligned} &\max_{\substack{\mathbf{Q}\succeq 0:\\ \mathrm{tr}(\mathbf{Q})\leq P}} \log\det(\mathbf{I} + \mathbf{F}_{31}\mathbf{F}_{31}^{\dagger} + \mathbf{F}_{32}\mathbf{F}_{32}^{\dagger} + \boldsymbol{\Omega}_3^{-1}\mathbf{Q}\boldsymbol{\Omega}_3^{-\dagger}) \\ &\leq \max_{\substack{\mathbf{Q}\succeq 0:\\ \mathrm{tr}(\mathbf{Q})\leq P}} \log\det((1+\gamma)\mathbf{I} + \boldsymbol{\Omega}_3^{-1}\mathbf{Q}\boldsymbol{\Omega}_3^{-\dagger}) && (5.98) \\ &= \log\det\left((1+\gamma)\mathbf{I} + \frac{P}{p}\boldsymbol{\Omega}_3^{-1}\boldsymbol{\Omega}_3^{-\dagger}\right) + o_P(1) && (5.99) \\ &= \log\det\left(\mathbf{I} + \frac{P}{p}\boldsymbol{\Omega}_3^{-1}\boldsymbol{\Omega}_3^{-\dagger}\right) + o_P(1) \\ &= \log\det\left(\mathbf{I} + \frac{P}{p}\mathbf{H}_{\mathrm{r}}\mathbf{H}_{\mathrm{e}}^{\perp}\mathbf{H}_{\mathrm{r}}^{\dagger}\right) + o_P(1) && (5.100) \end{aligned}$$

where (5.98) follows from the fact that $\mathbf{F}_{31}\mathbf{F}_{31}^{\dagger} + \mathbf{F}_{32}\mathbf{F}_{32}^{\dagger} \preceq \gamma\mathbf{I}$, and (5.99) follows from the fact that water-filling provides a vanishingly small gain over flat power allocation when the channel matrix has a full rank (see e.g., [36]) and (5.100) follows via (5.61).

5.4.2 Synthetic noise transmission strategy

The transmission scheme is based on a particular choice of $(\mathbf{x}, \mathbf{u})$ in the binning scheme (5.6). Let $b_1, \ldots, b_{n_t}$ be independent Gaussian random variables sampled according to $\mathcal{CN}(0, P_{\mathrm{t}})$, where $P_{\mathrm{t}} = \frac{P}{n_t}$. Let $\mathbf{H}_{\mathrm{r}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}_{\mathrm{r}}^{\dagger}$, be the compact SVD of $\mathbf{H}_{\mathrm{r}}$. Since $\mathrm{rank}(\mathbf{H}_{\mathrm{r}}) = n_r$, note that $\mathbf{U} \in \mathbb{C}^{n_r \times n_r}$ is a unitary matrix and $\boldsymbol{\Lambda} \in \mathbb{C}^{n_r \times n_r}$ is a diagonal matrix. Let $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_{n_r}] \in \mathbb{C}^{n_t \times n_r}$ and let $\{\mathbf{v}_j\}_{j=1}^{n_t}$ constitute an

orthogonal basis in $\mathbb{C}^{n_t}$. Our choice of parameters is,

$$\mathbf{x} = \sum_{j=1}^{n_t} b_j \mathbf{v}_j, \quad \mathbf{u} = (b_1, \ldots, b_{n_r}). \tag{5.101}$$

Here the symbols in $\mathbf{u}$ are the information bearing symbols from a corresponding codeword, while the symbols $(b_{n_r+1}, \ldots, b_{n_t})$ are synthetic noise symbols transmitted in the null space of the legitimate receiver's channel in order to confuse a potential eavesdropper. We first show, via straightforward computation, that this choice of parameters, results in a rate of

$$\begin{aligned} R_{\mathrm{SN}}(P) &= \log\det\left(\mathbf{I} + \varepsilon_t \mathbf{\Lambda}^{-2}\right) \\ &\quad + \log\det(\mathbf{H}_{\mathrm{r}}(\varepsilon_t \mathbf{I} + \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{H}_{\mathrm{r}}^{\dagger}). \end{aligned} \tag{5.102}$$

where $\varepsilon_t = \frac{1}{P_t}$. First note that

$$\begin{aligned} I(\mathbf{u}; \mathbf{y}_{\mathrm{e}}) &= \log\det(\mathbf{I} + P_t \mathbf{H}_{\mathrm{r}}\mathbf{H}_{\mathrm{r}}^{\dagger}) \\ &= \log\det(\mathbf{I} + P_t \mathbf{\Lambda}^2) \end{aligned} \tag{5.103}$$

In the following, let $\mathbf{V}_{\mathrm{n}} = [\mathbf{v}_{n_r+1}, \ldots, \mathbf{v}_{n_t}]$ denote the vectors in the null space of $\mathbf{H}_{\mathrm{r}}$.

$$\begin{aligned} &I(\mathbf{u}; \mathbf{y}_{\mathrm{e}}) = h(\mathbf{y}_{\mathrm{e}}) - h(\mathbf{y}_{\mathrm{e}}|\mathbf{u}) \\ &= \log\det(\mathbf{I} + P_t \mathbf{H}_{\mathrm{e}}\mathbf{H}_{\mathrm{e}}^{\dagger}) - \log\det(\mathbf{I} + P_t \mathbf{H}_{\mathrm{e}}\mathbf{V}_{\mathrm{n}}\mathbf{V}_{\mathrm{n}}^{\dagger}\mathbf{H}_{\mathrm{e}}^{\dagger}) \\ &= \log\det(\mathbf{I} + P_t \mathbf{H}_{\mathrm{e}}\mathbf{H}_{\mathrm{e}}^{\dagger}) - \log\det(\mathbf{I} + P_t \mathbf{H}_{\mathrm{e}}(\mathbf{I} - \mathbf{V}\mathbf{V}^{\dagger})\mathbf{H}_{\mathrm{e}}^{\dagger}) \\ &= \log\det(\mathbf{I} + P_t \mathbf{H}_{\mathrm{r}}^{\dagger}\mathbf{H}_{\mathrm{r}}) - \log\det(\mathbf{I} + P_t (\mathbf{I} - \mathbf{V}\mathbf{V}^{\dagger})\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}) \\ &= -\log\det(\mathbf{I} - P_t(\mathbf{I} + P_t \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}(\mathbf{V}\mathbf{V}^{\dagger}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})) \\ &= -\log\det(\mathbf{I} - P_t \mathbf{V}^{\dagger}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}(\mathbf{I} + P_t \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{V})) \\ &= -\log\det(\mathbf{V}^{\dagger}(\mathbf{I} + P_t \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{V}) \end{aligned}$$

Where we have repeatedly used the fact that $\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA})$ for any two matrices $\mathbf{A}$ and $\mathbf{B}$ of compatible dimensions.

$$R_{\mathrm{SN}}(P) = \log\det(\mathbf{I} + P_t \mathbf{\Lambda}^2) + \log\det(\mathbf{V}^{\dagger}(\mathbf{I} + P_t \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{V}).$$

Since $\mathbf{U}$ and $\mathbf{\Lambda}$ are square and invertible,

$$\begin{aligned} &R_{\mathrm{SN}}(P) \\ &= \log\det(\mathbf{I} + \varepsilon_t \mathbf{\Lambda}^{-2}) \\ &\qquad + \log\det(\mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\dagger}(\varepsilon_t \mathbf{I} + \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{V}\mathbf{\Lambda}\mathbf{U}^{\dagger}) \\ &= \log\det(\mathbf{I} + \varepsilon_t \mathbf{\Lambda}^{-2}) + \log\det(\mathbf{H}_{\mathrm{r}}(\varepsilon_t \mathbf{I} + \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{H}_{\mathrm{r}}^{\dagger}), \end{aligned}$$

as required.

We now establish (5.13). First we use the following facts

Fact 12 (Taylor Series Expansion [44]) *Let* $\mathbf{M}$ *be an invertible matrix. Then*

$$(\varepsilon \mathbf{I} + \mathbf{M})^{-1} = \mathbf{M}^{-1} + O(\varepsilon), \tag{5.104}$$

where $O(\varepsilon)$ *represents a function that goes to zero as* $\varepsilon \to 0$.

Fact 13 *Suppose that* $\mathbf{H}_{\mathrm{r}}$ *and* $\mathbf{H}_{\mathrm{e}}$ *be the channel matrices as in* (5.1)*, and suppose that* $\mathrm{rank}(\mathbf{H}_{\mathrm{r}}) = n_r$ *and* $\mathrm{rank}(\mathbf{H}_{\mathrm{e}}) = n_t$ *and* $n_r \leq n_t \leq n_e$*. Let* $\sigma_1, \sigma_2 \ldots, \sigma_s$ *denote the generalized singular values of* $(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}})$ *(c.f.* (5.52)*). Then*

$$\det\left(\mathbf{H}_{\mathrm{r}}(\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{H}_{\mathrm{r}}^{\dagger}\right) = \prod_{j=1}^{s} \sigma_j^2 \tag{5.105}$$

The proof follows by direct substitution of the GSVD expansion (5.48) and will be omitted.

Finally, to establish (5.13), we take the limit $\varepsilon_t \to 0$ in (5.102)

$$\begin{aligned}
R_{\mathrm{SN}}(P) &= \log\det\left(\mathbf{I} + \varepsilon_t \mathbf{\Lambda}^{-2}\right) \\
&\quad + \log\det(\mathbf{H}_{\mathrm{r}}(\varepsilon_t \mathbf{I} + \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{H}_{\mathrm{r}}^{\dagger}) \\
&= \log\det(\mathbf{H}_{\mathrm{r}}((\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1} + O(\varepsilon_t))\mathbf{H}_{\mathrm{r}}^{\dagger}) + O(\varepsilon_t) \qquad (5.106) \\
&= \log\det(\mathbf{H}_{\mathrm{r}}(\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{H}_{\mathrm{r}}^{\dagger}) \\
&\quad + \log\det(\mathbf{I} + (\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1/2}O(\varepsilon)(\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-\dagger/2}) \\
&= \sum_{j=1}^{s} \log\sigma_j^2 + O(\varepsilon_t). \qquad (5.107)
\end{aligned}$$

where we use Facts 12 and 13 above in (5.106) and (5.107) above and the fact that $\log\det(\mathbf{I} + \mathbf{X}) = \sum_j \log(1 + \lambda_j(\mathbf{X}))$ is continuous in the entries of $\mathbf{X}$.

5.5 Zero-Capacity Condition and Scaling Laws

We first establish the zero capacity condition in Lemma 6

Proof. When $\mathrm{Null}(\mathbf{H}_{\mathrm{r}})^{\perp} \bigcap \mathrm{Null}(\mathbf{H}_{\mathrm{e}}) \neq \{\}$, clearly, $\sigma_{\max}(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}}) = \infty$. Otherwise, it is known (see e.g., [19]) that $\sigma_{\max}(\cdot)$ is the largest generalized singular value of $(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}})$ as defined in (5.52).

To establish that the capacity is zero, whenever $\sigma_{\max}(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}}) \leq 1$, it suffices to consider the high SNR secrecy capacity in (5.11) in Theorem 14, which is clearly zero whenever $\sigma_{\max}() \leq 1$.

If $\sigma_{\max} > 1$, let $\mathbf{v}$ there exists a vector $\mathbf{v}$ such that $||\mathbf{H}_{\mathrm{r}}\mathbf{v}|| > ||\mathbf{H}_{\mathrm{e}}\mathbf{v}||$. Select $\mathbf{x} = \mathbf{u} \sim \mathcal{CN}(0, P\mathbf{v}\mathbf{v}^{\dagger})$ in (5.6). Clearly $C(P) \geq R_{-}(P) > 0$ for all $P > 0$. ∎

For our scaling analysis, we use the following convergence property of the largest generalized singular value for Gaussian matrices.

Fact 14 ([47, 3]) *Suppose that* $\mathbf{H}_{\mathrm{r}}$ *and* $\mathbf{H}_{\mathrm{e}}$ *have i.i.d.* $\mathcal{CN}(0,1)$ *entries. Let* $n_{\mathrm{r}}, n_{\mathrm{e}}, n_{\mathrm{t}} \to \infty$, *while keeping* $n_{\mathrm{r}}/n_{\mathrm{e}} = \gamma$ *and* $n_{\mathrm{t}}/n_{\mathrm{e}} = \beta$ *fixed. If* $\beta < 1$, *then the largest generalized singular value of* $(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}})$ *converges almost surely to*

$$\sigma_{\max}(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}}) \xrightarrow{\text{a.s.}} \gamma \left[\frac{1 + \sqrt{1 - (1-\beta)\left(1 - \frac{\beta}{\gamma}\right)}}{1-\beta} \right]^2 . \tag{5.108}$$

By combining Lemma 6 and Fact 14, one can deduces the zero-capacity condition in Corollary 7.

5.6 Conclusion

We establish the secrecy capacity of the MIMOME channel as a saddle point solution to a minimax problem. Our capacity result establishes that a Gaussian input maximizes the secrecy capacity expression by Csiszár and Körner for the MIMOME channel. Our proof uses upper bounding ideas from the MIMO broadcast channel literature and the analysis of optimality conditions provides insight into the structure of the optimal solution. Next, we develop an explicit expression for the secrecy capacity in the high SNR regime in terms of the generalized singular value decomposition (GSVD) and show that in this case, an optimal scheme involves simultaneous diagonalization of the channel matrices to create a set of independent parallel channel and using independent codebooks across these channels. We also study a synthetic noise transmission scheme that is "semi-blind" as it selects the transmit directions based on the legitimate receiver's channel only and compare its performance with the capacity achieving scheme. Finally, we study the conditions under which the secrecy capacity is zero and study its scaling laws in the limit of many antennas.

Chapter 6

Secret-key generation with sources and channels

So far this thesis has focussed on variations of the wiretap channel model. As we discussed in the introductin, a related approach for generating secret keys between two terminals using correlated sources has been studied by Maurer [37] and Ahlswede and Csiszar [2]. As shows in Fig. 6-1, the two legitimate terminals, observe a pair of correlated sources (u^N, v^N) and through public discussion on the noiseless channel, distill a common secret key that must be concealed from the eavesdropper. In this chapter we extend their results to the case when the underlying channel is not a noiseless bit-pipe but rather a wiretap channel, see Fig. 6-2. Note that there are two types of uncertainties at the eavesdropper — correlated sources and wiretap channel. We develop insights into efficient code designs for secret-key generation in this joint source-channel setup.

6.1 Source-Channel Model

The channel from sender to receiver and wiretapper is a discrete-memoryless-channel (DMC), $p(y, z|x)$. The sender and intended receiver observe discrete-memoryless-

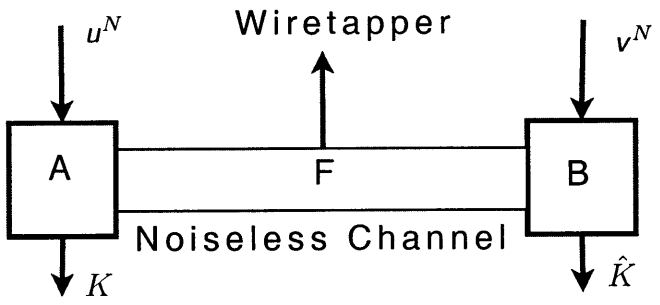


Figure 6-1: The problem of secret key generation using correlated sources. Terminals A and B observe a pair

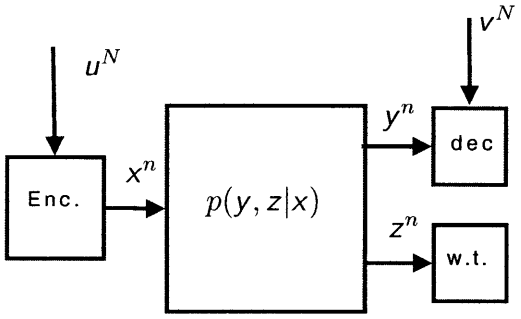


Figure 6-2: Wiretap channel model with Correlated sources

multiple-source (DMMS) $p(u,v)$ of length N and communicate over n uses of the DMC

A (n,N) secrecy code for this setup consists of a (possibly stochastic) function[1] $f_n : \mathcal{U}^N \to \mathcal{X}^n$, that maps the observed source sequence to the channel output, and two key generation functions $K_n = K_n(\mathcal{U}^N, \mathcal{X}^n)$ and $L_n = L_n(\mathcal{V}^N, \mathcal{Y}^n)$. A secret-key rate R is achievable with bandwidth expansion factor β if there exists a sequence of $(n, \beta n)$ codes, such that for a sequence ε_n that approaches zero as $n \to \infty$, we have (i) $\Pr(K_n \neq L_n) \leq \varepsilon_n$ (ii) $\frac{1}{n}H(K_n) \geq R - \varepsilon_n$ (iii)$\frac{1}{n}I(K_n; z^n) \leq \varepsilon_n$. The secret-key-capacity is the supremum of all achievable rates.

For some of our results, we will also consider the case when the wiretapper observes a side information sequence w^N sampled i.i.d. $p_w()$. In this case, the secrecy condition in (iii) above is replaced with

$$\frac{1}{n}I(K_n; z^n, w^N) \leq \varepsilon_n \tag{6.1}$$

6.2 Statement of Main Result

Lemma 11 *Suppose that t is a random variable such that $t \to u \to v$, and a and b are random variables such that $b \to a \to x \to (y,z)$ holds and $I(y;b) \leq I(z;b)$. Further define,*

$$R_{\text{ch}} = I(a;y), \tag{6.2a}$$
$$R_{\text{eq}}^- = I(a;y|b) - I(a;z|b) \tag{6.2b}$$
$$R_{\text{s}} = I(t;v), \tag{6.2c}$$
$$R_{\text{wz}} = I(t;u) - I(t;v). \tag{6.2d}$$

Suppose that the random variables t, a and b satisfy

$$\beta R_{\text{wz}} \leq R_{\text{ch}}, \tag{6.3}$$

[1]The alphabets associated with random variables will be denoted by calligraph letters. Random variables are denoted by sans-serif font, while their realizations are denoted by standard font. A length n sequence is denoted by x^n.

then

$$R^{-}_{\text{key}} = \beta R_{\text{s}} + R^{-}_{\text{eq}}, \tag{6.4}$$

is an achievable secret-key rate.

Lemma 12 *An upper bound on the secret-key rate is given by,*

$$R^{+}_{\text{key}} = \sup_{\{(x,t)\}} \left\{ \beta R_{\text{s}} + R^{+}_{\text{eq}} \right\}, \tag{6.5}$$

where the supremum is over all distributions over the random variables (x, t) that satisfy $t \to u \to w$, the cardinality of t is at-most the cardinality of u plus one, and

$$I(x; y) \geq \beta R_{\text{wz}}. \tag{6.6}$$

The quantities R_{s} and R_{wz} are defined in (6.2c) *and* (6.2d) *respectively and*

$$R^{+}_{\text{eq}} = I(x; y \mid z). \tag{6.7}$$

Furthermore, it suffices to consider only those distributions where (x, t) are independent.

6.2.1 Reversely degraded parallel independent channels

Our bounds coincide the the case of reversely degraded parallel independent channels. Consider M parallel independent channels, where channel i for $1 \leq i \leq M$ has transistion probability $p_{y_i,z_i|x_i}$ such that either $x_i \to y_i \to z_i$ or $x_i \to z_i \to y_i$ holds.

Corollary 8 *The secret-key-capacity for the reversely degraded parallel independent channels is given by*

$$C_{\text{key}} = \max_{\{(x_1,\ldots,x_M,t)\}} \left\{ \beta I(v; t) + \sum_{i=1}^{M} I(x_i; y_i | z_i) \right\}, \tag{6.8}$$

where the random variables $(x_1, \ldots, x_M, t)$ are mutually independent, $t \to u \to v$, and

$$\sum_{i=1}^{M} I(x_i; y_i) \geq \beta \{ I(u; t) - I(v; t) \} \tag{6.9}$$

A Gaussian reversely degraded parallel channel has $y_i = x_i + n_{r,i}$ and $z_i = x_i + n_{e,i}$ where $n_{r,i}$ and $n_{e,i}$ have variances equal to $\mathcal{N}(0, \sigma^2_{r,i})$ and $\mathcal{N}(0, \sigma^2_{e,i})$ respectively. Furthermore, if $x_i \to y_i \to z_i$ holds, then $y_i = x_i + n_{r,i}$ and $z_i = y_i + \Delta n_{e,i}$ else, if $x_i \to z_i \to y_i$ then $z_i = x_i + n_{e,i}$ and $y_i = z_i + \Delta n_{r,i}$ holds, where the random variables Δn are defined as the difference between the two noise variables. We assume that the input satisfies a sum power constraint i.e., $\sum_{i=1}^{n} E[x_i^2] \leq P$. Furthermore we assume

that u and v are jointly Gaussian (scalar valued) random variables, and without loss of generality we assume that $u \sim \mathcal{N}(0,1)$ and $v = u + s$, where $s \sim \mathcal{N}(0,S)$ is independent of u.

Corollary 9 *The secret-key capacity for the case of Gaussian parallel channels and Gaussian sources, as described above, is obtained by optimizing* (6.8) *and* (6.9) *over independent Gaussian distributions i.e., we can select* $x_i \sim \mathcal{N}(0,P_i)$ *and* $u = t + d$, *for some* $d \sim \mathcal{N}(0,D)$, *independent of* t *and* $\sum_{i=1}^{n} P_i \leq P$, $P_i \geq 0$, *and* $0 < D \leq 1$,

$$C_{\text{key}}^{G} = \max_{\{P_i\}_{i=1}^{M},D} \left\{ \frac{\beta}{2}\log\left(\frac{1+S}{D+S}\right) + \sum_{\substack{i:1\leq i\leq M \\ \sigma_{r,i}\leq\sigma_{e,i}}} \frac{1}{2}\log\left(\frac{1+P_i/\sigma_{r,i}^2}{1+P_i/\sigma_{e,i}^2}\right) \right\}, \tag{6.10}$$

where $D, P_1, \ldots, P_M$ *also satisfy the following relation:*

$$\sum_{i=1}^{M} \frac{1}{2}\log\left(1+\frac{P_i}{\sigma_{r,i}^2}\right) \geq \frac{1}{2}\log\left(\frac{1}{D}\right) - \frac{1}{2}\log\left(\frac{1+S}{D+S}\right) \tag{6.11}$$

A few remarks follow. Note that the secret-key capacity expression (6.8) exploits both the source and channel uncertainties at the wiretapper. By setting either uncertainty to zero, one can recover known results. $I(u;v) = 0$, i.e., there is no secrecy from the source, then the secret-key-rate equals the wiretap capacity [53]. If instead, $x \rightarrow z \rightarrow y$, i.e., there is no secrecy from the channel, then our result essentially reduces to the result by Narayan and Csiszar [10], that consider the case when the channel is a noiseless bit-pipe with finite rate.

In general, the setup of wiretap channel involves a tradeoff between information rate and equivocation. The secret-key generation setup provides an operational significance to this tradeoff. Note that the capacity expression (6.8) in Corollary 8 involves two terms. The first term $\beta I(t;v)$ is the contribution from the correlated sources. In general, this quantity increases by increasing the information rate $I(x;y)$ as seen from (6.9). The second term, $I(x;y|z)$ is the equivocation term and increasing this term, often comes at the expense of the information rate. Maximizing the secret-key rate, involves operating on a certain point on the rate-equivocation curve and thus provides an operational significance to the rate equivocation tradeoff.

Example

It is instructive to illustrate this tradeoff by a numerical example. Consider two parallel channels,

$$\begin{aligned} y_1 &= a_1 x + n_{r,1}, \quad & z_1 &= b_1 x + n_{e,2} \\ y_2 &= a_2 x + n_{r,2}, \quad & z_2 &= y_2 \end{aligned} \tag{6.12}$$

where $a_1 = 1$, $a_2 = 2$, and $b_1 = 0.5$. Furthermore, $u \sim \mathcal{N}(0,1)$ and $v = u + s$, where $s \sim \mathcal{N}(0,1)$ is independent of u. The noise rv's are all $\mathcal{CN}(0,1)$ and appropriately

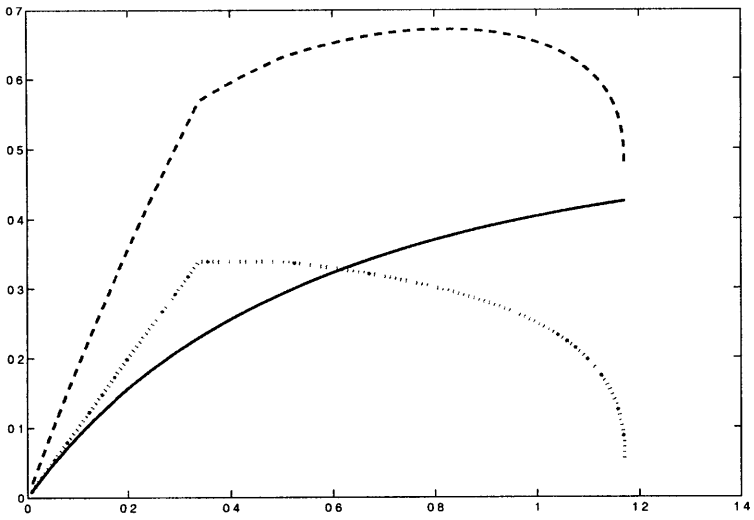


Figure 6-3: Tradeoff inherent in the secret-key-capacity formulation. The solid curve is the rate equivocation region for the parallel Gaussian channel (6.12). The dotted curve represents the quantity $I(t; v)$ as a function of the rate, while the dashed curve is the secret-key rate, which is the sum of the other two curves. The secret-key rate is maximized at a point between the maximum equivocation and maximum rate.

correlated so that the users are degraded on each channel. A total power constraint $P = 1$ is selected and the bandwidth expansion factor β equals unity.

In this example, the optimization in Corollary 9, takes the form:

$$C_{\text{key}} = \max_{P_1, P_2, D} R_{eq}(P_1, P_2) + R_{\Delta}(D), \tag{6.13}$$

$$\text{such that,} \tag{6.14}$$

$$R_{\Delta}(D) = \frac{1}{2} \log \frac{2}{1+D} \leq R(P_1, P_2), \tag{6.15}$$

$$P_1 + P_2 \leq P \tag{6.16}$$

where

$$R(P_1, P_2) = \frac{1}{2} \left(\log \left(1 + a_1^2 P_1\right) + \log(1 + a_2^2 P_2) \right), \tag{6.17}$$

and

$$R_{\text{eq}}(P_1, P_2) = \frac{1}{2} \left(\log(1 + a_1^2 P_1) - \log(1 + b_1^2 P_1) \right) \tag{6.18}$$

denote the rate and equivocation respectively.

First note that this example captures the inherent tradeoff between information rate and equivocation. To maximize the information rate, one maximizes $R(P_1, P_2)$ and this involves in general allocating power over both the sub-channels. To maximize equivocation, one must allocate all the power on the first sub-channel. The

second sub-channel is useless from the secrey point of view as $y_2 = z_2$. Figure 6-2 illustrates the (fundamental) tradeoff between rate and equivocation for this channel, which is obtained as we vary power allocation between the two sub-channels. The power allocation that maximizes the equivocation is $P_1 = 1$ and $P_2 = 0$ while the power allocation that maximizes the Shannn capacity is obtained by the water-filling equations (see e.g., [6]). On the other hand, the source-term $I(t; v)$ monotonically increases with the rate, as shown in the figure. The optimal operating point that maximizes the secret-key capacity (c.f. (6.14)) is also illustrated in the figure.

6.2.2 Side information at the wiretapper

So far, we have focussed on the case when there is no side information at the wiretapper. This assumption is valid for certain application such as biometrics, when the correlated sources constitute successive measurements of a person's biometric. In other applications, such as sensor networks, it is more realistic to assume that the wiretapper also has access to a side information sequence.

We consider the setup described in Fig. 6-2, but with a modification that the wiretapper observes a source sequence w^N, obtained by $N-$ independent samples of a random variable w. In this case the secrecy condition takes the form in (6.1). We only consider the case when the sources and channels satisfy a degradedness condition.

Lemma 13 *Suppose that the random variables (u, v, w) satisfy the degradedness condition $u \to v \to w$ and the broadcast channel is also degraded i.e., $x \to y \to z$. Then, the secret-key-capacity is given by*

$$C_{\text{key}} = \max_{(x,t)} \{\beta(I(t; v) - I(t; w)) + I(x; y|z)\}, \tag{6.19}$$

where the maximization is over all random variabes (t, x) that are mutually independent, $t \to u \to v \to w$ and

$$I(x; y) \geq \beta(I(v; t) - I(u; t)) \tag{6.20}$$

holds. Furthermore, it suffices to optimize over random variables t whose cardinality does not exceed that of u plus two.

6.3 Achievability: Coding Theorem

We demonstrate the coding theorem in the special case when $a = x$ and $b = 0$ in Lemma 11. Accordingly we have that (6.2a) and (6.2b) reduce to

$$R_{\text{ch}} = I(x; y) \tag{6.21a}$$

$$R_{\text{eq}}^{-} = I(x; y) - I(x; z) \tag{6.21b}$$

The more general case, can be incorporated by introducing an auxiliary channel $a \to x$ and superposition coding [9]. Furthermore, in our discussion below we will

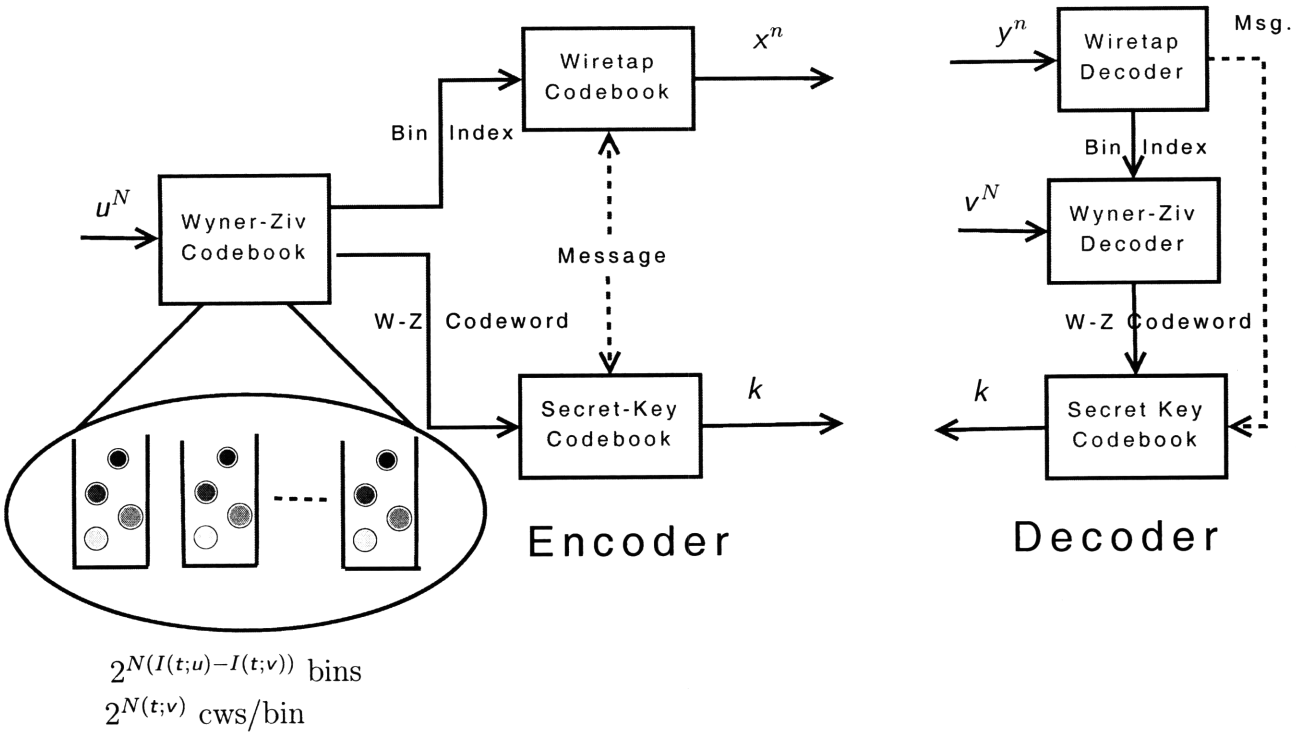


Figure 6-4: Source-Channel Code Design for secret-key distillation problem. The source sequence u^N is mapped to a codeword in a Wyner-Ziv codebook. This codeword determines the secret-key via the secret-key codebook. The bin index of the codeword constitutes a message in the wiretap codebook.

assume that the distributions $p_{t|u}$ and p_x are selected such that, for a sufficiently small but fixed $\delta > 0$, we have

$$\beta R_{\text{wz}} = R_{\text{ch}} - 3\delta. \tag{6.22}$$

We note that the optimization over the joint distributions in Lemma 11 is over the region $\beta R_{\text{wz}} \leq R_{\text{ch}}$. If the joint distributions satisfy that $\beta R_{\text{wz}} = \alpha(R_{\text{ch}} - 3\delta)$ for some $\alpha < 1$, one can use the code construction below for a bock-length αn and then transmit an independent message at rate R_{eq}^- using a perfect-secrecy wiretap-code. This provides a rate of

$$\alpha\left(\frac{\beta}{\alpha}R_{\text{wz}} + R_{\text{eq}}^-\right) + (1-\alpha)R_{\text{eq}}^- = R_{\text{eq}}^- + \beta R_{\text{wz}},$$

as required.

6.3.1 Codebook Construction

Our codebook construction is as shown in the Fig. 6-4.

It consists of three codebooks: Wyner-Ziv codebook, secret-key codebook and a wiretap codebook that are constructed via a random coding construction. In our discussion below we will be using the notion of strong typicality. Given a random

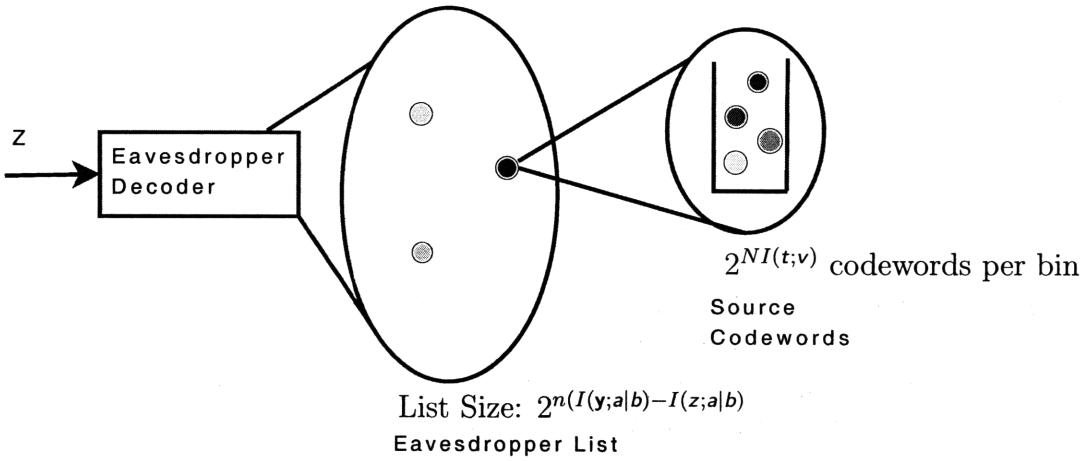


Figure 6-5: Equivocation at the eavesdropper through the source-channel codebook.

variable t, the set of all sequences of length N and type that coincides with the distribution p_t is denoted by T_t^N. The set of all sequences whose emperical type is in an ε-shell of p_t is denoted by $T_{t,\varepsilon}^N$. The set of jointly typical sequences are defined in an analogous manner. Given a sequence u^N of type T_u^N, the set of all sequences v^N that have a joint type of $p_{u,v}()$ is denoted by $T_{u,v}^N(u^N)$. We will be using the following properties of typical sequences

$$|T_{t,\varepsilon}^N| = \exp(N(H(t) + o_\varepsilon(1))) \tag{6.23a}$$

$$\Pr(t^N = t^N) = \exp(-N(H(t) + o_\varepsilon(1))), \quad \forall\, t^N \in T_{t,\varepsilon}^N \tag{6.23b}$$

$$\Pr(t^N \in T_{t,\varepsilon}^N) \geq 1 - o_\varepsilon(1), \tag{6.23c}$$

where $o_\varepsilon(1)$ is a term that approaches zero as $N \to \infty$ and $\varepsilon \to 0$.

For fixed, but sufficiently small constants $\delta > 0$ and $\eta = \delta/\beta > 0$, let,

$$M_{\text{WZ}} = \exp(N(R_{\text{s}} - \eta)) \tag{6.24a}$$

$$N_{\text{WZ}} = \exp(N(R_{\text{wz}} + 2\eta)) \tag{6.24b}$$

$$M_{\text{SK}} = \exp(n(I(x;z) - \delta)) \tag{6.24c}$$

$$N_{\text{SK}} = \exp(n(\beta R_{\text{s}} + R_{\text{eq}}^- - \delta)) \tag{6.24d}$$

Substituting (6.2a)-(6.2d) and (6.22) into (6.24a)-(6.24d) we have that

$$N_{\text{tot}} \triangleq M_{\text{SK}} \cdot N_{\text{SK}} = M_{\text{WZ}} \cdot N_{\text{WZ}} = \exp(N(I(t;u) + \eta)) \tag{6.25}$$

We construct the Wyner-Ziv and secret-key codebooks as follows. Randomly and independently select N_{tot} sequences from the set of $t-$typical sequences T_t^N. Denote this set $\mathcal{T}$. Randomly and independently partition this set into the following

codebooks[2]:

- *Wyner-Ziv* codebook with N_{WZ} bins consisting of M_{WZ} sequences. The j^{th} sequence in bin i is denoted by $t^N_{ij,\text{WZ}}$.
- *Secret-key* codebook with N_{SK} bins consisting of M_{SK} sequences. The j^{th} sequence in bin i is denoted by $t^N_{ij,\text{SK}}$.

We define two functions $\Phi_{\text{WZ}} : \mathcal{T} \to \{1, \ldots, N_{\text{WZ}}\}$ and $\Phi_{\text{SK}} : \mathcal{T} \to \{1, \ldots, N_{\text{SK}}\}$ as follows.

- $\Phi_{\text{WZ}}(t^N) = i$, if $\exists j \in [1, M_{\text{WZ}}]$, such that $t^N = t^N_{ij,\text{WZ}}$.
- $\Phi_{\text{SK}}(t^N) = i$, if $\exists j \in [1, M_{\text{SK}}]$ such that $t^N = t^N_{ij,\text{SK}}$.

The channel codebook consists of $N_{\text{WZ}} = \exp(n(R_{\text{ch}} - \delta))$ sequences x^n uniformly and independently selected from the set of $x-$typical sequences T^n_x. The channel encoding function maps message i into the sequence x^n_i, i.e., $\Phi_{\text{ch}} : \{1, \ldots, N_{\text{WZ}}\} \to \mathcal{X}^n$ is defined as $\Phi_{\text{ch}}(i) = x^n_i$.

6.3.2 Encoding

Given a source sequence u^N, the encoder produces a secret-key k and a transmit sequence x^N as shown in Fig. 6-4.

- Find a sequence $t^N \in \mathcal{T}$ such that $(u^N, t^N) \in T^N_{ut,\varepsilon}$. Let $\mathcal{E}_1$ be the even that no such t^N exists.
- Compute $\phi = \Phi_{\text{WZ}}(t^N)$ and $k = \Phi_{\text{SK}}(t^N)$. Declare k as the secret-key.
- Compute $x^n_i = \Phi_{\text{ch}}(\phi)$, and transmit this sequence over $n-$uses of the DMC.

6.3.3 Decoding

The main steps of decoding at the legitimate receiver are shown in Fig. 6-4.

- Given a received sequence y^n, the sender looks for a unique index i such that $(x^n_i, y^n) \in T^n_{xy,\varepsilon}$. An error event $\mathcal{E}_2$ happens if x^n_i is not the transmitted codeword.
- Given the observed source sequence v^N, the decoder then searches for a unique index $j \in [1, M_{\text{WZ}}]$ such that $(t^N_{ij,\text{WZ}}, v^N) \in T^N_{tv,\varepsilon}$. An error event $\mathcal{E}_3$ is declared if a unique index does not exist.
- The decoder finds indices $\hat{k}$ and $\hat{l}$ such that $t^N_{ij,\text{WZ}} = t^N_{\hat{k}\hat{l},\text{SK}}$. The secret-key is declared as $\hat{k}$.

[2]As will be apparent in the analysis, the only pairwise independence is required between the codebooks i.e., $\forall t^N, \hat{t}^N \in \mathcal{T}$, $\Pr\left(\Phi_{\text{WZ}}(t^N) = \Phi_{\text{WZ}}(\hat{t}^N) | \Phi_{\text{SK}}(t^N) = \Phi_{\text{SK}}(\hat{t}^N)\right) = \Pr\left(\Phi_{\text{WZ}}(t^N) = \Phi_{\text{WZ}}(\hat{t}^N)\right) = \frac{1}{N_{\text{WZ}}}$

6.3.4 Error Probability Analysis

The error event of interest is $\mathcal{E} = \{k \neq \hat{k}\}$. We argue that selecting $n \to \infty$ leads to $\Pr(\mathcal{E}) \to 0$.

In particular, note that $\Pr(\mathcal{E}) = \Pr(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3) \leq \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2) + \Pr(\mathcal{E}_3)$. We argue that each of the terms vanishes with $n \to \infty$.

Recall that $\mathcal{E}_1$ is the event that the encoder does not find a sequence in $\mathcal{T}$ typical with u^N. Since $\mathcal{T}$ has $\exp(NI(u;t) + \eta)$ sequences randomly and uniformly selected from the set T_t^N, we have that $\Pr(\mathcal{E}_1) \to 0$.

Since the number of channel codewords equals $N_{\text{WZ}} = \exp n(I(x;y) - \delta)$, and the codewords are selected uniformly at random from the set $T_{x,\varepsilon}^n$, the error event $\Pr(\mathcal{E}_2) \to 0$.

Finally, since the number of sequences in each bin satisfies $M_{\text{WZ}} = \exp(N(I(t;v) - \eta))$, joint typical decoding guarantees that $\Pr(\mathcal{E}_3) \to 0$.

6.3.5 Secrecy Analysis

In this section, that for the coding scheme discussed above, the equivocation at the eavesdropper is close (in an asymptotic sense) to R_{key}.

First we establish some uniformity properties which will be used in the subsequent analysis.

Uniformity Properties

In our code construction Φ_{WZ} satisfies some useful properties which will be used in the sequel.

Lemma 14 *The random variable Φ_{WZ} satisfies the following relations*

$$\frac{1}{n}H(\Phi_{\text{WZ}}) = \beta R_{\text{WZ}} + o_\eta(1) \tag{6.26a}$$

$$\frac{1}{n}H(t^N|\Phi_{\text{WZ}}) = \beta I(t;v) + o_\eta(1) \tag{6.26b}$$

$$\frac{1}{n}H(\Phi_{\text{WZ}}|z^n) = I(x;y) - I(x;z) + o_\eta(1) \tag{6.26c}$$

where $o_\eta(1)$ vanishes to zero as we take $\eta \to 0$ and $N \to \infty$ for each η.

Proof. Relations (6.26a) and (6.26b) can be established by using the properties of typical sequences (c.f. (6.23a)-(6.23c)).

Let us define the function $\Gamma_{\text{WZ}} : \mathcal{T} \to \{1, \ldots, M_{\text{WZ}}\}$ to identify the position of the sequence $t^N \in \mathcal{T}$ in a given bin i.e., $\Gamma_{\text{WZ}}(t_{ij,\text{WZ}}^N) = j$.

Note that

$$\Pr(\Gamma_{\mathrm{WZ}} = j, \Phi_{\mathrm{WZ}} = i) \leq \sum_{u^N \in T_{u,t,\eta}(t^N_{ij,\mathrm{WZ}})} \Pr(u^N) \tag{6.27}$$

$$= \sum_{u^N \in T_{u,t,\eta}(t^N_{ij,\mathrm{WZ}})} \exp(-N(H(u) + o_\eta(1))) \tag{6.28}$$

$$= \exp(N(H(u|t) + o_\eta(1))) \exp(-N(H(u) + o_\eta(1))) \tag{6.29}$$

$$= \exp(-N(I(t;u) + o_\eta(1))) \tag{6.30}$$

where (6.27) follows from the construction of the joint-typicality encoder, (6.28) from (6.23b) and (6.29) from (6.23a). Marginalizing (6.27), we have that

$$\begin{aligned}\Pr(\Phi_{\mathrm{WZ}} = i) &= \sum_{j=1}^{M_{\mathrm{WZ}}} \Pr(\Gamma_{\mathrm{WZ}} = j, \Phi_{\mathrm{WZ}} = i) \\ &\leq M_{\mathrm{WZ}} \exp(-N(I(t;u) + o_\eta(1))) \\ &= \exp(-N(I(t;u) - I(t;v) + o_\eta(1))) \\ &= \exp(-N(R_{\mathrm{WZ}} + o_\eta(1))) \end{aligned} \tag{6.31}$$

Eq. (6.26a) follows from (6.31) and the continuity of the entropy function. Furthermore, we have from (6.30) that

$$\frac{1}{N} H(\Phi_{\mathrm{WZ}}, \Gamma_{\mathrm{WZ}}) = I(t;u) + o_\eta(1). \tag{6.32}$$

The relation (6.26b) follows by substituting (6.26a), since

$$\frac{1}{N} H(t^N|\Phi_{\mathrm{WZ}}) = \frac{1}{N} H(\Gamma_{\mathrm{WZ}}|\Phi_{\mathrm{WZ}}) = \frac{1}{N} H(\Gamma_{\mathrm{WZ}}, \Phi_{\mathrm{WZ}}) - \frac{1}{N} H(\Phi_{\mathrm{WZ}}) = I(t;v) + o_\eta(1). \tag{6.33}$$

Relation (6.26c) follows from the secrecy analysis of the channel codebook when the message is Φ_{WZ}. The details can be found in e.g., [53]. ∎

Furthermore the joint construction of the secret-key codebook and Wyner-Ziv codebook is such that the eavesdropper can decode the sequence t^N if it is revealed the secret-key $\Phi_{\mathrm{SK}} = k$ in addition to its observed sequence z^n. In particular

Lemma 15

$$\frac{1}{n} H(t^N|z^n, k) = o_\eta(1). \tag{6.34}$$

Proof. We show that there exists a decoding function $g : \mathcal{Z}^n \times \{1, 2, \ldots, N_{\mathrm{SK}}\} \to \mathcal{T}$ that such that $\Pr(t^N \neq g(z^n, k)) \to 0$ as $n \to \infty$. In particular, the decoding function $g(\cdot, \cdot)$ searches for the sequences in the bin associated with k in the secret-

key codebook, whose bin-index in the Wyner-Ziv codebook maps to a sequence x_i^n jointly typical with the received sequence z^n. More formally,

- Given z^n, the decoder constructs a the set of indices $\mathcal{I}_x = \{i : (x_i^n, z^n) \in T_{xz,\varepsilon}^n\}$.
- Given k, it constructs a set of sequences, $\mathcal{S} = \left\{ t_{kj\text{SK}}^N : \Phi_{\text{WZ}}(t_{kj,\text{SK}}^N) \in \mathcal{I}_x, 1 \le j \le M_{\text{SK}}, \right\}$.
- If S contains a unique sequence $\hat{t}^N$, it is declared to be the required sequence. An error event is defined as

$$\begin{aligned} \mathcal{J} &= \{\hat{t}^N \neq t^N\} \\ &= \left\{\exists j, 1 \le j \le M_{\text{SK}}, \Phi_{\text{WZ}}(t_{k,j,\text{SK}}^N) \in \mathcal{I}_x, j \neq j_0\right\}, \end{aligned} \tag{6.35}$$

  where j_0 is the index of the sequence t^N in bin k of the secret-key codebook, i.e., $t_{kj_0,\text{SK}}^N = t^N$.

We now use the properties of typical sequences (6.23a)-(6.23c) to show that $\Pr(\mathcal{J}) \to 0$ as $n \to \infty$. We begin by defining the event that the sequence $t^N \notin \mathcal{S}$, which is equivalent to

$$\mathcal{J}_0 = \left\{\Phi_{\text{WZ}}(t_{k,j_0,\text{SK}}^N) \notin \mathcal{I}_x\right\}.$$

From (6.23c) we have that $\Pr(\mathcal{J}_0) = o_\eta(1)$. Furthermore,

$$\begin{aligned} \Pr(\mathcal{J}) &\le \Pr(\mathcal{J}|\mathcal{J}_0^c) + \Pr(\mathcal{J}_0) \\ &\le \sum_{j=1}^{M_{\text{SK}}} \Pr(\mathcal{J}_j|\mathcal{J}_0^c) + o_\eta(1), \end{aligned} \tag{6.36}$$

where the event $\mathcal{J}_j$, defined as

$$\mathcal{J}_j = \left\{\Phi_{\text{WZ}}(t_{k,j,\text{SK}}^N) \in \mathcal{I}_x\right\}, \quad j = 1, 2, \ldots M_{\text{SK}}, j \neq j_0$$

is the event that the sequence $t_{kj\text{SK}}^N \in \mathcal{S}$.

To upper bound the event $\mathcal{J}_j$, we will consider the collision event that $t_{\text{kj,SK}}^N$ and $t_{\text{kj}_0,\text{SK}}^N$ belong to the same bins in the in the Wyner-Ziv codebook i.e.,

$$\mathcal{J}_{\text{col,j}} = \left\{\Phi_{\text{WZ}}(t_{kj,\text{SK}}^N) = \Phi_{\text{WZ}}(t_{kj_0,\text{SK}}^N)\right\}, \quad j = 1, 2, \ldots, M_{\text{sk}}, j \neq j_0.$$

By the union bound,

$$\Pr(\mathcal{J}_j|\mathcal{J}_0^c) \le \Pr(\mathcal{J}_j|\mathcal{J}_0^c \cap \mathcal{J}_{\text{col,j}}^c) + \Pr(\mathcal{J}_{\text{col,j}}|\mathcal{J}_0^c). \tag{6.37}$$

We bound each of the two terms in (6.37). The first term is conditioned on the event that the sequences $t_{kj,\text{SK}}^N$ and $t_{kj_0,\text{SK}}^N$ are assigned to independent bins in the Wyner-Ziv codebook. This event is equivalent to the event that a randomly selected sequence x^N belongs to the typical set $\mathcal{I}_x$. The error event is bounded as [6]

$$\Pr(\mathcal{J}_j|\mathcal{J}_0^c \cap \mathcal{J}_{\text{col,j}}^c) \leq \exp(-n(I(x;z) - 3\varepsilon)). \tag{6.38}$$

The second term in (6.37) is the collision event. Since the code construction partitions assigns the sequences $t_{kj,\text{SK}}^N$ and $t_{kj_0,\text{SK}}^N$ to independent bins, and channel error event is independent of this partitioning, we have

$$\begin{aligned}\Pr(\mathcal{J}_j|\mathcal{J}_0^c) &= \Pr(\mathcal{J}_j) \\ &= \exp(-n(\beta R_{\text{WZ}} + 2\delta)) \\ &= \exp(-n(\beta I(x;y) - \delta))\end{aligned} \tag{6.39}$$

Substituting (6.39) and (6.38) into (6.37), we have

$$\begin{aligned}\Pr(\mathcal{J}_j|\mathcal{J}_0^c) &\leq \exp(-n(I(x;z) - 3\varepsilon)) + \exp(-n(\beta I(x;y) - \delta)) \\ &\leq \exp(-n(I(x;z) - 4\varepsilon)), \quad n \geq n_0,\end{aligned} \tag{6.40}$$

where we use the fact that $I(x;y) > I(x;z)$ in the last step so that the required n_0 exists.

Finally substituting (6.40) into (6.36) and using relation (6.24c) for M_{SK}, we have that

$$\Pr(\mathcal{J}) \leq \exp(-n(\delta - 4\varepsilon)) + o_\eta(1), \tag{6.41}$$

which vanishes with n, whenever the decoding function selects $\varepsilon < \delta/4$. ∎

Equivocation Analysis

It remains to show that the equivocation rate at the eavesdropper approaches the secret-key rate as $n \to \infty$, which we do below.

$$\begin{aligned}
H(k|z^n) &= H(k, t^N|z^n) - H(t^N|z^n, k) \\
&= H(t^N|z^n) - H(t^N|z^n, k) && (6.42)\\
&= H(t^N, \Phi_{\mathrm{WZ}}|z^n) - H(t^N|z^n, k) && (6.43)\\
&= H(t^N|\Phi_{\mathrm{WZ}}, z^n) + H(\Phi_{\mathrm{WZ}}|z^n) - H(t^N|z^n, k) \\
&= H(t^N|\Phi_{\mathrm{WZ}}) + H(\Phi_{\mathrm{WZ}}|z^n) - H(t^N|z^n, k), && (6.44)\\
&= n\beta I(t; v) + n\{I(x; y) - I(x; z)\} + no_\eta(1) && (6.45)\\
&= n(R_{\mathrm{key}} + o_\eta(1)), && (6.46)
\end{aligned}$$

where (6.42) and (6.43) follow from the fact that Φ_{WZ} is a deterministic function of t^N and (6.44) follows from the fact that $t^N \to \Phi_{\mathrm{WZ}} \to z^n$ holds for our code construction. and (6.45) step follows from (6.26b) and (6.26c) in Lemma 14 and Lemma 15.

6.4 Proof of the Upper bound (Lemma 12)

Given a sequence of (n, N) codes that achieve a secret-key-rate R_{key}, there exists a sequence ε_n, such that $\varepsilon_n \to 0$ as $n \to \infty$, and

$$\frac{1}{n}H(k|y^n, v^N) \le \varepsilon_n \tag{6.47a}$$

$$\frac{1}{n}H(k|z^n) \ge \frac{1}{n}H(k) - \varepsilon_n. \tag{6.47b}$$

We can now upper bound the rate R_{key} as follows.

$$\begin{aligned}
nR_{\mathrm{key}} &= H(k) \\
&= H(k|y^n, v^N) + I(k; y^n, v^N) \\
&\le n\varepsilon_n + I(k; y^n, v^N) - I(k; z^n) + I(k; z^n) && (6.48)\\
&\le 2n\varepsilon_n + I(k; y^n, v^N) - I(k; z^n) && (6.49)\\
&= 2n\varepsilon_n + I(k; y^n) - I(k; z^n) + I(k; v^N|y^n) \\
&\le 2n\varepsilon_n + I(k; y^n) - I(k; z^n) + I(k, y^n; v^N) && (6.50)
\end{aligned}$$

where (6.48) and (6.49) follow from (6.47a) and (6.47b) respectively.

Now, let J be a random variable uniformly distributed over the set $\{1, 2, \ldots, N\}$ and independent of everything else. Let $t_i = (k, y^n, v_{i+1}^N, u_1^{i-1})$ and $t = (k, y^n, v_{J+1}^N, u_1^{J-1}, J)$, and v_J be a random variable that conditioned on $J = i$ has the distribution of p_{v_i}. Note that since $\mathbf{v}$ is memoryless, v_J is independent of J and has the same marginal distribution as v. Also note that $t \to u_J \to v_J$ holds.

$$
\begin{aligned}
I(k, y^n; v^N) &= \sum_{i=1}^{n} I(k, y^n; v_i | v_1^{i-1}) \\
&\leq \sum_{i=1}^{N} I(k, y^n, v_{i+1}^n; v_i) \\
&\leq \sum_{i=1}^{N} I(k, y^n, v_{i+1}^n, u_1^{i-1}; v_i) \\
&= NI(k, y^n, v_{J+1}^n, u_1^{J-1}; v_J | J) \\
&= NI(k, y^n, v_{J+1}^n, u_1^{J-1}, J; v_J) - I(J; v_J) \\
&= NI(t; v) \qquad (6.51)
\end{aligned}
$$

where (6.51) follows from the fact that v_J is independent of J and has the same marginal distribution as v.

Next, we upper bound $I(k; y^n) - I(k; z^n)$ as below. Let p_{x_i} denote the channel input distribution at time i and let p_{y_i,z_i} denote the corresponding output distribution. Let $p_x = \frac{1}{n}\sum_{i=1}^{n} p_{x_i}$ and let p_y and p_z be defined similarly.

$$
\begin{aligned}
I(k; y^n) - I(k; z^n) &\leq I(k; y^n | z^n) \\
&\leq I(x^n; y^n | z^n) \qquad (6.52) \\
&\leq \sum_{i=1}^{n} I(x_i; y_i | z_i) \qquad (6.53) \\
&\leq nI(x; y|z), \qquad (6.54)
\end{aligned}
$$

where (6.52) follows from the Markov condition $k \to x^n \to (y^n, z^n)$ and (6.53) follows from the fact that the channel is memoryless and (6.54) follows from Jensen's inequality since the term $I(x; y|z)$ is concave in the distribution p_x (see e.g., [23, Appendix-I]).

Combining (6.54) and (6.51) we have that

$$
R_{\text{key}} \leq I(x; y|z) + \beta I(v; t), \qquad (6.55)
$$

thus establishing the first half of the condition in Lemma 12. It remains to show that the condition

$$
I(t; u) - I(t; v) \leq I(x; y)
$$

is also satisfied. Since $u^N \to x^n \to y^n$ holds, we have that

$$
\begin{aligned}
nI(x; y) &\geq I(x^n; y^n) \qquad (6.56) \\
&\geq I(u^N; y^n) \qquad (6.57) \\
&\geq I(u^N; y^n, k) - I(v^N; y^n, k) - n\varepsilon_n, \qquad (6.58)
\end{aligned}
$$

where the last inequality holds, since

$$\begin{aligned} I(u^N; k|y^n) - I(v^N; y^n, k) &= -I(v^N; y^n) + I(u^N; k|y^n) - I(v^N; k|y^n) \\ &\leq I(u^N; k|y^n) - I(v^N; k|y^n) \\ &= H(k|y^n, v^N) - H(k|y^n, u^N) \\ &\leq n\varepsilon_n, \end{aligned}$$

where the last step holds via (6.47a) and the fact that $H(k|y^n, u^N) > 0$.

Continuing (6.58), we have

$$nI(x;y) \geq I(u^N; y^n, k) - I(v^N; y^n, k) - n\varepsilon_n \tag{6.59}$$

$$= \sum_{i=1}^{N} \{I(u_i; y^n, k, u_1^{i-1} v_{i+1}^n) - I(v_i; y^n, k, u_1^{i-1} v_{i+1}^n)\} + n\varepsilon_n \tag{6.60}$$

$$\begin{aligned} &= N\{I(u_J; y^n, k, u_1^{J-1} v_{J+1}^n | J) - I(v_J; y^n, k, u_1^{J-1} v_{J+1}^n | J) + \varepsilon_n\} \\ &= N\{I(u_J; y^n, k, u_1^{J-1} v_{J+1}^n | J) - I(v_J; y^n, k, u_1^{J-1} v_{J+1}^n | J) + \varepsilon_n\} \\ &= N\{I(u_J; t) - I(v_J; t) + I(v_J; J) - I(u_J; J) + \varepsilon_n\} \end{aligned}$$

$$= N\{I(u;t) - I(v;t) + \varepsilon_n\} \tag{6.61}$$

where (6.60) follows from the well known chain rule for difference between mutual information expressions and (6.61) follows from the fact that the random variables v_J and u_J are independent of J and have the same marginal distribution as v and u respectively.

The cardinality bound on t is obtained via Caratheordory's theorem and will not be presented here.

Finally, since the upper bound expression does not depend on the joint distribution of (t, x), it suffices to optimize over those distributions where (t, x) are independent.

6.5 Reversely Degraded Channels

6.5.1 Proof of Corollary 8

First we show that the expression is an upper bound on the capacity. From Lemma 12, we have that

$$C_{\text{key}} \leq \max_{(x,t)} I(x; y|z) + \beta I(t; v),$$

where we maximize over those distributions where (x, t) are mutually independent, $t \to u \to v$, and

$$I(x;y) \geq \beta(I(t;u) - I(t;v)).$$

For the reversely degraded parallel independent channels, note that

$$I(x;y) \leq \sum_{i=1}^{M} I(x_i;y_i)$$
$$I(x;y|z) \leq \sum_{i=1}^{M} I(x_i;y_i|z_i),$$

with equality when $(x_1,\ldots,x_M)$ are mutually independent. Thus it suffices to take $(x_1,\ldots,x_M)$ to be mutually independent, which establishes that the proposed expression is an upper bound on the capacity.

For achievability, we propose a choice of auxiliary random variables (a,b) in Lemma 11, such that the resulting expression reduces to the capacity. In particular, assume without loss in generality that for the first P channels we have that $x_i \to y_i \to z_i$ and for the remaining channels we have that $x_i \to z_i \to y_i$. Let $a = (x_1, x_2, \ldots, x_M)$ and $b = (x_{P+1}, \ldots, x_M)$ where the random variables $\{x_i\}$ are mutually independent. It follows from (6.2a) and (6.2b) that

$$R_{\text{ch}} = \sum_{i=1}^{M} I(x_i;y_i) \tag{6.62}$$
$$R_{\text{eq}}^{-} = \sum_{i=1}^{P} I(x_i;y_i|z_i) = \sum_{i=1}^{M} I(x_i;y_i|z_i), \tag{6.63}$$

where the last equality follows since for $x_i \to z_i \to y_i$, we have that $I(x_i;y_i|z_i) = 0$. Substituting in (6.4) and (6.3) we recover the capacity expression.

6.5.2 Gaussian Case (Corollary 9)

For the Gaussian case we show that Gaussian codebooks achieve the capacity as in Corollary 9.

Recall that the capacity expression involves maximizing over random variables $\mathbf{x} = (x_1,\ldots,x_M)$, and $t \to u \to v$,

$$C_{\text{key}} = \sum_i I(x_i;y_i|z_i) + I(t;v) \tag{6.64}$$

subjected to the constraint that $E[\sum_{i=1}^{M} x_i^2] \leq P$ and

$$\sum_i I(x_i;y_i) \geq I(t;u) - I(t;v). \tag{6.65}$$

Let us first fix the distribution $p_{\mathbf{x}}$ and upper bound the objective function (6.64). Let $R \triangleq \frac{1}{\beta}\sum_{i=1}^{M} I(x_i;y_i)$ and $v = u + s$, where $s \sim \mathcal{N}(0,S)$ is independent of u. We

will use the conditional entropy power inequality

$$\exp(h(u+s|t)) \geq \exp(h(u|t)) + \exp(h(s)) \tag{6.66}$$

for any pair of random variables (t, u) independent of s. The equality happens if (u, t) are jointly Gaussian.

Note that we can express (6.65) as

$$\begin{aligned} R + h(v) - h(u) &= h(v|t) - h(u|t) && (6.67)\\ &= h(u+s|t) - h(u|t) && (6.68)\\ &\geq \frac{1}{2}\log\left(\exp(h(u|t)) + 2\pi e S\right) - h(u|t) && (6.69) \end{aligned}$$

Letting

$$h(u|t)) = \frac{1}{2}\log 2\pi e D, \tag{6.70}$$

we have that

$$D \geq \frac{S}{\exp(2(R + h(v) - h(u))) - 1} \tag{6.71}$$

The term $I(t; v)$ in the objective function (6.64) can be upper bounded as

$$\begin{aligned} I(t; v) &= h(v) - h(v|t)\\ &= h(v) - h(u+s|t)\\ &\leq h(v) - \log(\exp(h(u|s)) + 2\pi e S) && (6.72)\\ &= \frac{1}{2}\log\frac{1+S}{D+S} && (6.73) \end{aligned}$$

where (6.72) follows by the application of the EPI (6.66) and (6.73) follows via (6.70). Thus the objective function (6.64) can be expressed as

$$C_{\text{key}} = \sum_i I(x_i; y_i|z_i) + \frac{1}{2}\log\frac{1+S}{D+S}, \tag{6.74}$$

where D satisfies (6.71).

It remains to show that the optimal $\mathbf{x}$ has a Gaussian distribution. Note that the set of feasible distributions for $\mathbf{x}$ is closed and bounded and hence an optimum exists. Also if $p_{\mathbf{x}}$ is any optimum distribution, we can increase both R and $I(x_i; y_i|z_i)$ by replacing $p_{\mathbf{x}}$ with a Gaussian distribution (see e.g., [24]) with the same second order moment. Since the objective function is increasing in both these terms, it follows that a Gaussian $p_{\mathbf{x}}$ also maximizes the objective function (6.64).

6.6 Side information at the Wiretapper

We now provide an achievability and a converse for the capacity stated in Lemma 13

6.6.1 Achievability

Our coding scheme is a natural extension of the case when $w = 0$. We only point out the main differences. Recall that for the degraded channel case, R_{ch} and R_{eq}^- are defined as

$$\begin{aligned} R_{\text{ch}} &= I(x;y) \\ R_{\text{eq}}^- &= I(x;y|z). \end{aligned}$$

Furthermore, we replace R_{s} in (6.2c) with

$$R_{\text{s}} = I(t;v) - I(t;w). \tag{6.75}$$

and the secret-key rate in (6.4) is

$$R_{\text{LB}} = \beta\{I(t;v) - I(t;w)\} + I(x;y|z). \tag{6.76}$$

The codebook construction, encoding and decoding are analogous to the descriptions in Sections 6.3.1, 6.3.2,and 6.3.3 respectively. The only difference is that the Secret-Key codebook rate is adjusted to reflect (6.76) i.e., the constant M_{SK} and N_{SK} in (6.24c) and (6.24d) are replaced with

$$M_{\text{SK}} = \exp\left(n(I(x;z) + \beta I(w;t) - \delta)\right) \tag{6.77}$$

$$N_{\text{SK}} = \exp\left(n(\beta R_{\text{s}} + R_{\text{eq}}^- - \delta)\right) \tag{6.78}$$

and R_{s} is defined in (6.75).

6.6.2 Secrecy Analysis

We show that the equivocation condition at the eavesdropper (6.1) in section ?? holds for the code construction. This is equivalent to showing that

$$\frac{1}{n}H(k|w^N, z^n) = \beta(I(t;v) - I(t;w)) + I(x;y|z) + o_\eta(n), \tag{6.79}$$

which we will now do.

We first provide an alternate expression for the left hand side in (6.79).

$$\begin{aligned} H(k|w^N, z^n) &= H(k, t^N|w^N, z^n) - H(t^N|k, w^N, z^n) && (6.80) \\ &= H(t^N|w^N, z^n) - H(t^N|k, w^N, z^n) \\ &= H(t^N, \Phi_{\text{WZ}}|w^N, z^n) - H(t^N|k, w^N, z^n) && (6.81) \\ &= H(\Phi_{\text{WZ}}|w^N, z^n) + H(t^N|\Phi_{\text{WZ}}, w^N) - H(t^N|k, w^N, z^n) && (6.82) \end{aligned}$$

where (6.81) follows from the fact that Φ_{WZ} is a deterministic function of t^N, while (6.82) follows from the fact that $t^N \to (w^N, \Phi_{\mathrm{WZ}}) \to z^n$ forms a Markov chain. The proof of (6.79) is completed by showing that

$$\frac{1}{n}H(\Phi_{\mathrm{WZ}}|w^N, z^n) \geq I(x; y|z) + o_\eta(1) \tag{6.83a}$$

$$\frac{1}{n}H(t^N|\Phi_{\mathrm{WZ}}, w^N) = \beta(I(t; v) - I(t; w)) + o_\eta(1) \tag{6.83b}$$

$$\frac{1}{n}H(t^N|k, w^N, z^n) = o_\eta(1). \tag{6.83c}$$

To interpret (6.83a), recall that Φ_{WZ} is the message to the wiretap codebook. The equivocation introduced by the wiretap codebook $\frac{1}{n}H(\Phi_{\mathrm{WZ}}|z^n)$ equals $I(x; y|z)$. Eq. (6.83a) shows that if in addition to z^n, the eavesdropper has access to w^N, a degraded source, the equivocation still remains the same. Eq. (6.83b) shows that the knowledge of w^N reduces the list of t^N sequences in any bin from $\exp(n(I(t; v)))$ to $\exp(n(I(t; v) - I(t; w)))$, while (6.83c) shows that for the code construction, the eavesdropper, if revealed the secret-key, can decode t^N with high probability.

To establish (6.83a),

$$\begin{aligned}
\frac{1}{n}H(\Phi_{\mathrm{WZ}}|w^N, z^n) &\geq \frac{1}{n}H(\Phi_{\mathrm{WZ}}|z^n, v^N) && (6.84)\\
&= \frac{1}{n}H(\Phi_{\mathrm{WZ}}|z^n) - \frac{1}{n}I(\Phi_{\mathrm{WZ}}; v^N|z^n)\\
&\geq I(x; y|z) + o_\eta(1) - \frac{1}{n}I(\Phi_{\mathrm{WZ}}; v^N|z^n), && (6.85)\\
&\geq I(x; y|z) + o_\eta(1) - \frac{1}{n}I(\Phi_{\mathrm{WZ}}; v^N), && (6.86)
\end{aligned}$$

where (6.84) follows from the fact that $w^N \to v^N \to u^N \to \Phi_{\mathrm{WZ}} \to z^n$, (6.85) from Lemma 14 and (6.86) from the fact that $v^N \to \Phi_{\mathrm{WZ}} \to z^n$ so that

$$\frac{1}{n}I(\Phi_{\mathrm{WZ}}; v^N|z^n) \leq \frac{1}{n}I(\Phi_{\mathrm{WZ}}; v^N). \tag{6.87}$$

Thus we need to show the following.

Lemma 16

$$\frac{1}{n}I(\Phi_{\mathrm{WZ}}; v^N) \leq o_\eta(1). \tag{6.88}$$

Proof. From Lemma 14 note that

$$\frac{1}{N}H(\Phi_{\mathrm{WZ}}) = I(t; u) - I(t; v) + o_\eta(1)$$

and hence we need to show that

$$\frac{1}{N}H(\Phi_{\mathrm{WZ}}|v^N) = I(t;u) - I(t;v) + o_\eta(1)$$

as we do below.

$$\begin{aligned}\frac{1}{N}H(\Phi_{\mathrm{WZ}}|v^N) &= \frac{1}{N}H(\Phi_{\mathrm{WZ}}, t^N|v^N) - \frac{1}{N}H(t^N|v^N, \Phi_{\mathrm{WZ}}) \\ &= \frac{1}{N}H(t^N|v^N) + o_\eta(1) \qquad (6.89)\end{aligned}$$

Where (6.89) follows since each bin has $M_{\mathrm{WZ}} = \exp\left(N(I(t;v) - \eta)\right)$ sequences, (from standard joint typicality arguments) we have that

$$\frac{1}{N}H(t^N|v^N, \Phi_{\mathrm{WZ}}) = o_\eta(1). \qquad (6.90)$$

Finally

$$\frac{1}{N}H(t^N|v^N) = I(t;u) - I(t;v) + o_\eta(1),$$

which follows by substituting $a = v$, $b = u$ and $c = t$ and $R = I(t;u) + \eta$, in Lemma 17 in Appendix D establishes (6.88).

■

To establish (6.83b), we again use Lemma 17 in Appendix D, with $a = w$, $b = u$ and $c = t$ and $R = I(t;v) - \eta$. Finally, to establish (6.83c), we construct a decoder as in section 6.3.5 that searches for a sequence t_{kj}^N such that $\Phi_{\mathrm{WZ}}(t_{kj}^N) \in \mathcal{I}_{\mathrm{x}}$ and which is also jointly typical with w^N. Since there are $\exp\{n(I(w;t) + I(x;z) - \eta)\}$ sequences in the set, we can show along the same lines as in the proof of Lemma 15 that t^N can be decoded with high probability given (k, z^n, w^N). The details will be omitted.

6.6.3 Converse

Suppose there is a sequences of (n, N) codes that achieves a secret key (k) rate of R, and $\beta = N/n$. Then from Fano's inequality,

$$H(k|y^n, v^N) \leq n\varepsilon_n, \quad H(k|x^n, u^N) \leq n\varepsilon_n$$

and from the secrecy constraint.

$$\frac{1}{n}I(k; z^n, w^N) \leq \varepsilon_n.$$

Combining these inequalities, we have that,

$$\begin{aligned}
nR_{\text{key}} &\leq I(k; y^n, v^N) - I(k; z^n, w^N) + 2n\varepsilon_n \\
&\leq I(k; y^n, v^N \mid z^n, w^N) + 2n\varepsilon_n \\
&\leq h(y^n \mid z^n) + h(v^N \mid w^N) - h(y^n \mid z^n, w^N, k) - h(v^N \mid y^n, z^n, w^N, k) + 2n\varepsilon_n \\
&\leq h(y^n \mid z^n) + h(v^N \mid w^N) - h(y^n \mid z^n, w^N, k, x^n) - h(v^N \mid y^n, z^n, w^N, k,) + 2n\varepsilon_n \\
&= h(y^n \mid z^n) + h(v^N \mid w^N) - h(y^n \mid z^n, x^n) - h(v^N \mid y^n, z^n, w^N, k,) + 2n\varepsilon_n \qquad (6.91) \\
&\leq \sum_{i=1}^{n} I(x_i; y_i \mid z_i) + h(v^N \mid w^N) - h(v^N | y^n, w^N, k) + 2n\varepsilon_n \qquad (6.92) \\
&\leq nI(x; y \mid z) + h(v^N \mid w^N) - h(v^N | y^n, w^N, k) + 2n\varepsilon_n \qquad (6.93)
\end{aligned}$$

where the (6.91) follows from the fact that $(w^N, k) \to (z^n, x^n) \to y^n$, and (6.92) follows from the Markov condition $z^n \to (y^n, w^n, k) \to v^N$ that holds for the degraded channel, while (6.93) follows from the fact that $I(x; y|z)$ is a concave function of p_{x_i} (see e.g., [23, Appendix-I]) and we select $p_x(\cdot) = \frac{1}{n}\sum_{i=1}^{n} p_{x_i}(\cdot)$. Now, let $t_i = (k, u_{i+1}^n v^{i-1}, y^n)$, J be a random variable uniformly distributed over the set $[1, 2, \ldots n]$ and $t = (J, k, u_{J+1}^n v^{J-1}, y^n)$ we have that

$$\begin{aligned}
h(v^N|y^n, w^N, k) &= \sum_{i=1}^{N} h(v_i|v^{i-1}, y^n, w^N, k) \\
&\geq \sum_{i=1}^{N} h(v_i|v^{i-1}, y^n, w^N, u_{i+1}^N, k) \\
&= \sum_{i=1}^{N} h(v_i|v^{i-1}, y^n, w_i, u_{i+1}^N, k) \qquad (6.94) \\
&= N \cdot h(v_J|t, w_J)
\end{aligned}$$

where we have used the fact that $(w^{i-1}, w_{i+1}^N) \to (v^{i-1}, y^n, w_i, u_{i+1}^N, k) \to v_i$ which can be verified as follows

$$\begin{aligned}
& p\left(v_i \mid w_i, w^{i-1}, w_{i+1}^N, v^{i-1}, u_{i+1}^N, y^n, k\right) \\
=& \sum_{u_i=u} p\left(v_i \mid w_i, u_i = u, w^{i-1}, w_{i+1}^N, v^{i-1}, u_{i+1}^N, y^n, k\right) p\left(u_i = u \mid w_i, w^{i-1}, w_{i+1}^N, v^{i-1}, u_{i+1}^N, y^n, k\right) \\
=& \sum_{u_i=u} p\left(v_i \mid w_i, u_i = u\right) p\left(u_i = u \mid w_i, v^{i-1}, u_{i+1}^N, y^n, k\right) \qquad (6.95) \\
=& p\left(v_i \mid w_i, v^{i-1}, u_{i+1}^N, y^n, k\right),
\end{aligned}$$

where (6.95) follows from the fact that since the sequence v^N is sampled i.i.d. , we have that

$$v_i \to (u_i, w_i) \to (w^{i-1}, w_{i+1}^N, v^{i-1}, u_{i+1}^N, y^n, k)$$

and since $u \to v \to w$, it follows that

$$u_i \to (v^{i-1}, u_{i+1}^N, y^n, w_i, k) \to (w^{i-1}, w_{i+1}^N).$$

Since, v_J and w_J are both independent of J, we from (6.93) that

$$R_{\text{key}} \leq I(x;y|z) + \beta I(t;v|w) + 2\varepsilon_n.$$

Finally, using the steps between (6.59)-(6.61) as in the converse for the case when $w = 0$, we have that

$$I(x;y) \geq \beta(I(t;u) - I(t;v)), \tag{6.96}$$

which completes the proof.

6.7 Conclusions

We study a joint-source channel setup for secret-key generation between two terminals. Lower and upper bounds on the secret-key capacity are presented and the capacity is established when the underlying channel constitutes parallel independent and reversely degraded channels. When the wiretapper also has access to a correlated source sequence, the secret-key-capacity is established when both the sources and the channels of the wiretapper are a degraded version of the legitimate receiver. This setup also provides an operational significance for the operating point on the rate-equivocation tradeoff for the wiretap channel. This is illustrated in detail with the example of Gaussian sources and Gaussian parallel channels.

In terms of future work, there can be many fruitful avenues to explore for secret-key distillation in a joint-source-channel setup. One can consider multi-user extensions of the secret-key generation problem along the lines of [11] and also consider more sophisticated channel models such as the compound wiretap channels, MIMO wiretap channels and wiretap channels with feedback and/or side information. Connections of this setup to wireless channels, biometric systems and other applications can also be interesting.

Chapter 7

Conclusion

This thesis explores the possibility of using ideas from information theory for providing data confidentiality. The focus of this thesis was on formulating new problems in information theory based on secrecy constraints. We studied the wiretap channel model and discussed its extensions to parallel channels, fading channels and multi-antenna channels. The role of source and channel coding techniques for the secret key generation was also studied. At the time of the writing of this thesis, there seems to be growing interest in formulating new multi-user information theory problems with secrecy constraints. We summarize a few directions of future work below.

7.1 Future Work

7.1.1 Practical code design

The design of practical code construction for the wiretap channel is not explored in this thesis. While Chapter 1 discusses a scalar code design for the uniform noise model, it is unclear if this design also extends to other noise models such as the, Gaussian noise model. A useful regime for practical code construction is the high signal-to-noise-ratio limit. Are there efficient scalar code constructions in this regime that achieve near optimal performance?

7.1.2 Equivocation criterion

Throughout this thesis the protocols that we consider measure equivocation level $\frac{1}{n}H(w|y_e^n)$ at the eavesdropper. The asymptotically perfect secrecy constraint requires that the equivocation rate equal the information rate as the block length goes to infinity.

At what equivocation level should one operate in practice? In general this depends on the application. If the goal is to use these protocols to transmit secret keys, a reasonable choice is the perfect secrecy condition. Perhaps in this case, it may be worth-while to consider even a stronger notions of secrecy as discussed in Chapter 1. On the other hand if these protocols are used for media content delivery, then clearly one can have a much smaller level of equivocation and operate close to the channel

capacity. In Chapter 6 we provided a framework of joint-source-channel coding for secret-key generation where the the equivocation point of operation lies in between the Shannon capacity and perfect secrecy. Developing further insights on the optimal operating point is an interesting area of further research.

7.1.3 Gains from Feedback

An interesting direction of further research is to study gains from feedback. In absence of feedback, this thesis considers the use of diversity techniques such as multiple antennas and fading to transmit secret information to legitimate receiver, even when the eavesdropper has on an average a stronger channel. Feedback provides yet another mechanism to establish secret keys when the legitimate receiver's channel is weaker, but statistically independent, of the eavesdropper.

7.1.4 Gaussian Model

In this thesis, our focus was on the case when both the channels of the legitimate receiver and the eavesdropper are subjected to Gaussian noise. In many systems, the noise may not be Gaussian, nevertheless the analysis with Gaussian noise is a worst case analysis. However it is unclear if the assumption of Gaussian noise for the eavesdropper's channel is a robust assumption. Formalizing the worst case model in a game theoretic setting could provide use insights.

Appendix A

Concavity of the conditional mutual information

We establish Fact 4 i.e., for any random variables x, y, and z the quantity $I(x;y|z)$ is concave in $p(x)$.

Proof. Let t be a binary valued random variable such that: if $t = 0$ the induced distribution on x is $p_0(x)$, i.e., $p(y,z,x|t=0) = p(y,z|x)p_0(x)$, and if $t = 1$ the induced distribution on $p(x)$ is $p_1(x)$ i.e. $p(y,z,x|t=1) = p(y,z|x)p_1(x)$. Note the Markov chain $t \rightarrow x \rightarrow (y,z)$. To establish the concavity of $I(x;y|z)$ in $p(x)$ it suffices to show that

$$I(x;y|z,t) \leq I(x;y|z). \tag{A.1}$$

The following chain of inequalities can be verified.

$$\begin{aligned} I(x;y|z,t) - I(x;y|z) &= \{I(x;y,z|t) - I(x;z|t)\} - \{I(x;y,z) - I(x;z)\} && \text{(A.2)} \\ &= \{I(x;y,z|t) - I(x;z|t)\} - \{I(t,x;y,z) - I(t,x;z)\} && \text{(A.3)} \\ &= \{I(x;y,z|t) - I(t,x;y,z)\} - \{I(x;z|t) - I(t,x;z)\} \\ &= I(t;z) - I(t;y,z) = -I(t;y|z) \leq 0. \end{aligned}$$

Equation (A.2) is a consequence of the chain rule for mutual information. Equation (A.3) follows from the fact that $t \rightarrow x \rightarrow (y,z)$ forms a Markov Chain, so that $I(t;z|x) = I(t;y,z|x) = 0$. ∎

Appendix B

Proof of Lemma 4

Suppose there exists a sequence of $(2^{nR}, n)$ codes such that for every $\varepsilon > 0$, and n sufficiently large we have that

$$\Pr(w \neq \hat{w}) \leq \varepsilon, \tag{B.1}$$

$$\frac{1}{n} I(w; \mathbf{y}_{\mathrm{e}}^n) \leq \varepsilon, \tag{B.2}$$

$$\frac{1}{n} \sum_{i=1}^{n} E[\|\mathbf{x}(i)\|^2] \leq P. \tag{B.3}$$

We first note that (B.1) implies, from Fano's inequality,

$$\frac{1}{n} I(w; y_{\mathrm{r}}^n) \geq R - \varepsilon_{\mathrm{F}}, \tag{B.4}$$

where $\varepsilon_{\mathrm{F}} \to 0$ as $\varepsilon \to 0$. Combining (B.2) and (B.4), we have for $\varepsilon' = \varepsilon + \varepsilon_{\mathrm{F}}$:

$$\begin{aligned}
nR - n\varepsilon' &\leq I(w; y_{\mathrm{r}}^n) - I(w; \mathbf{y}_{\mathrm{e}}^n) \\
&\leq I(w; y_{\mathrm{r}}^n, \mathbf{y}_{\mathrm{e}}^n) - I(w; \mathbf{y}_{\mathrm{e}}^n) && \text{(B.5)} \\
&= I(w; y_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n) && \text{(B.6)} \\
&= h(y_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n) - h(y_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n, w) \\
&\leq h(y_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n) - h(y_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n, w, \mathbf{x}^n) && \text{(B.7)} \\
&= h(y_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n) - h(y_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n, \mathbf{x}^n) && \text{(B.8)} \\
&= h(y_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n) - \sum_{t=1}^{n} h(y_{\mathrm{r}}(t) | \mathbf{y}_{\mathrm{e}}(t), \mathbf{x}(t)) && \text{(B.9)} \\
&\leq \sum_{t=1}^{n} h(y_{\mathrm{r}}(t) | \mathbf{y}_{\mathrm{e}}(t)) - \sum_{t=1}^{n} h(y_{\mathrm{r}}(t) | \mathbf{y}_{\mathrm{e}}(t), \mathbf{x}(t)) \\
&= nI(\mathbf{x}; y_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}, q) && \text{(B.10)} \\
&\leq nI(\mathbf{x}; y_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}), && \text{(B.11)}
\end{aligned}$$

where (B.5) and (B.6) each follow from the chain of mutual information, (B.7) follows from the fact that conditioning cannot increase differential entropy, (B.8) follows from the Markov relation $w \leftrightarrow (\mathbf{x}^n, \mathbf{y}_{\mathrm{e}}^n) \leftrightarrow y_{\mathrm{r}}^n$, and (B.9) follows from the fact the channel is memoryless. Moreover, (B.10) is obtained by defining a time-sharing random variable q that takes values uniformly over the index set $\{1, 2, \ldots, n\}$ and defining $(\mathbf{x}, y_{\mathrm{r}}, \mathbf{y}_{\mathrm{e}})$ to be the tuple of random variables that conditioned on $q = t$, have the same joint distribution as $(\mathbf{x}(t), y_{\mathrm{r}}(t), \mathbf{y}_{\mathrm{e}}(t))$. It then follows that for our choice of $\mathbf{x}$ and given (B.3), $E[\|\mathbf{x}\|^2] \leq P$. Finally, (B.11) follows from the fact that $I(\mathbf{x}; y_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}})$ is concave in $p_{\mathbf{x}}$ (see, e.g., [23, Appendix I] for a proof), so that Jensen's inequality can be applied.

B.1 Derivation of (4.49)

The argument of the logarithm on left hand side of (4.49) is convex in $\boldsymbol{\theta}$, so it is straightforward to verify that the minimizing $\boldsymbol{\theta}$ is

$$\boldsymbol{\theta} = (\mathbf{I} + P\mathbf{H}_e\mathbf{H}_e^\dagger)^{-1}(P\mathbf{H}_e\mathbf{h}_r + \boldsymbol{\phi}). \tag{B.12}$$

In the sequel, we exploit that by the definition of generalized eigenvalues via (4.1),

$$(\mathbf{I} + P\mathbf{h}_r\mathbf{h}_r^\dagger)\boldsymbol{\psi}_{\max} = \lambda_{\max}(\mathbf{I} + P\mathbf{H}_e^\dagger\mathbf{H}_e)\boldsymbol{\psi}_{\max}, \tag{B.13}$$

or, rearranging,

$$\left(\mathbf{h}_r\mathbf{h}_r^\dagger - \lambda_{\max}\mathbf{H}_e^\dagger\mathbf{H}_e\right)\boldsymbol{\psi}_{\max} = \frac{(\lambda_{\max} - 1)}{P} \cdot \boldsymbol{\psi}_{\max}. \tag{B.14}$$

First we obtain a more convenient expression for $\boldsymbol{\theta}$ as follows:

$$\boldsymbol{\theta} = (\mathbf{I} + P\mathbf{H}_e\mathbf{H}_e^\dagger)^{-1}\left(P\mathbf{H}_e\mathbf{h}_r + \frac{1}{\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}}\mathbf{H}_e\boldsymbol{\psi}_{\max}\right) \tag{B.15}$$

$$= (\mathbf{I} + P\mathbf{H}_e\mathbf{H}_e^\dagger)^{-1}\frac{\mathbf{H}_e(P\mathbf{h}_r\mathbf{h}_r^\dagger + \mathbf{I})\boldsymbol{\psi}_{\max}}{\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}}$$

$$= (\mathbf{I} + P\mathbf{H}_e\mathbf{H}_e^\dagger)^{-1}\frac{\lambda_{\max}\mathbf{H}_e(P\mathbf{H}_e^\dagger\mathbf{H}_e + \mathbf{I})\boldsymbol{\psi}_{\max}}{\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}} \tag{B.16}$$

$$= (\mathbf{I} + P\mathbf{H}_e\mathbf{H}_e^\dagger)^{-1}\frac{\lambda_{\max}\cdot(P\mathbf{H}_e\mathbf{H}_e^\dagger + \mathbf{I})\mathbf{H}_e\boldsymbol{\psi}_{\max}}{\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}} \tag{B.17}$$

$$= \lambda_{\max}\boldsymbol{\phi}, \tag{B.18}$$

where (B.15) follows from substituting (4.48) into (B.12), and (B.16) follows from substituting via (B.13).

Next we have that

$$\mathbf{h}_r - \mathbf{H}_e^\dagger\boldsymbol{\theta} = \mathbf{h}_r - \frac{\lambda_{\max}}{\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}}\mathbf{H}_e^\dagger\mathbf{H}_e\boldsymbol{\psi}_{\max} \tag{B.19}$$

$$= \frac{(\mathbf{h}_r\mathbf{h}_r^\dagger - \lambda_{\max}\mathbf{H}_e^\dagger\mathbf{H}_e)\boldsymbol{\psi}_{\max}}{\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}}$$

$$= \frac{(\lambda_{\max} - 1)\boldsymbol{\psi}_{\max}}{P\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}} \tag{B.20}$$

$$\tag{B.21}$$

where (B.19) follows from substituting from (B.18) with (4.48), and (B.20) follows by substituting (B.14). Thus,

$$P\|\mathbf{h}_r - \mathbf{H}_e^\dagger\boldsymbol{\theta}\|^2 = (\lambda_{\max} - 1)\left[\frac{(\lambda_{\max} - 1)}{P|\mathbf{h}_r^\dagger\boldsymbol{\psi}_{\max}|^2}\right]. \tag{B.22}$$

To simplify (B.22) further, we exploit that

$$1 - \lambda_{\max}\|\boldsymbol{\phi}\|^2 = 1 - \lambda_{\max}\frac{\boldsymbol{\psi}_{\max}^\dagger \mathbf{H}_{\mathrm{e}}^\dagger \mathbf{H}_{\mathrm{e}} \boldsymbol{\psi}_{\max}}{\boldsymbol{\psi}_{\max}^\dagger \mathbf{h}_{\mathrm{r}} \mathbf{h}_{\mathrm{r}}^\dagger \boldsymbol{\psi}_{\max}} \tag{B.23}$$

$$= \frac{\boldsymbol{\psi}_{\max}^\dagger (\mathbf{h}_{\mathrm{r}} \mathbf{h}_{\mathrm{r}}^\dagger - \lambda_{\max} \mathbf{H}_{\mathrm{e}}^\dagger \mathbf{H}_{\mathrm{e}}) \boldsymbol{\psi}_{\max}}{|\mathbf{h}_{\mathrm{r}}^\dagger \boldsymbol{\psi}_{\max}|^2}$$

$$= \frac{(\lambda_{\max} - 1)}{P|\mathbf{h}_{\mathrm{r}}^\dagger \boldsymbol{\psi}_{\max}|^2}, \tag{B.24}$$

where (B.23) follows by again substituting from (4.48), and (B.24) follows by again substituting from (B.14). In turn, replacing the term in brackets in (B.22) according to (B.24) then yields

$$P\|\mathbf{h}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}}^\dagger \boldsymbol{\theta}\|^2 = (\lambda_{\max} - 1)(1 - \lambda_{\max}\|\boldsymbol{\phi}\|^2). \tag{B.25}$$

Finally, substituting (B.25) then (B.18) into the left hand side of (4.49) yields, following some minor algebra, the right hand side as desired.

Appendix C

Appendix to the MIMOME Capacity derivation

In this appendix we derive several helper lemmas which were used in the derivation of the MIMOME secrecy capacity.

C.1 Optimality of Gaussian Inputs

We show that a Gaussian input maximizes the conditional mutual information term $I(\mathbf{x}; \mathbf{y}_\mathrm{r}|\mathbf{y}_\mathrm{e})$ when the noise distribution $[\mathbf{z}_\mathrm{r}^\dagger, \mathbf{z}_\mathrm{e}^\dagger]^\dagger \sim \mathcal{CN}(0, \mathbf{K}_\Phi)$.

Recall that $\mathbf{K}_\Phi$ has the form,

$$\mathbf{K}_\Phi = \begin{bmatrix} \mathbf{I}_{n_\mathrm{r}} & \mathbf{\Phi} \\ \mathbf{\Phi}^\dagger & \mathbf{I}_{n_\mathrm{e}} \end{bmatrix} \tag{C.1}$$

and $\mathbf{K}_\Phi \succ 0$ if and only if $||\mathbf{\Phi}||_2 < 1$. In this case we show that among all distributions $p_\mathbf{x}$ with a covariance of $\mathbf{K}_\mathrm{P}$, a Gaussian distribution maximizes $I(\mathbf{x}; \mathbf{y}_\mathrm{r}|\mathbf{y}_\mathrm{e})$. Note that

$$\begin{aligned} I(\mathbf{x}; \mathbf{y}_\mathrm{r}|\mathbf{y}_\mathrm{e}) &= h(\mathbf{y}_\mathrm{r}|\mathbf{y}_\mathrm{e}) - h(\mathbf{z}_\mathrm{r}|\mathbf{z}_\mathrm{e}) && \text{(C.2)} \\ &= h(\mathbf{y}_\mathrm{r}|\mathbf{y}_\mathrm{e}) - \log(2\pi e)^{n_r} \det(\mathbf{I} - \mathbf{\Phi}\mathbf{\Phi}^\dagger) \\ &\leq \log\det \Lambda(\mathbf{K}_\mathrm{P}) - \log\det(\mathbf{I} - \mathbf{\Phi}\mathbf{\Phi}^\dagger), && \text{(C.3)} \end{aligned}$$

where

$$\begin{aligned} \Lambda(\mathbf{K}_\mathrm{P}) \triangleq \mathbf{I} + \mathbf{H}_\mathrm{r}\mathbf{K}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger - \\ (\mathbf{\Phi} + \mathbf{H}_\mathrm{r}\mathbf{K}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger)(\mathbf{I} + \mathbf{H}_\mathrm{e}\mathbf{K}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger)^{-1}(\mathbf{\Phi}^\dagger + \mathbf{H}_\mathrm{e}\mathbf{K}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger) \end{aligned} \tag{C.4}$$

is the linear minimum mean squared error in estimating $\mathbf{y}_\mathrm{r}$ given $\mathbf{y}_\mathrm{e}$ and the last inequality is satisfied with equality if $p_\mathbf{x} = \mathcal{CN}(0, \mathbf{K}_\mathrm{P})$.

When $\bar{\mathbf{K}}_\Phi$ is singular, the expansion (C.2) is not well defined. Nevertheless, we can circumvent this step by defining an appropriately reduced channel. In particular,

let

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Delta} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^\dagger \\ \mathbf{V}_2^\dagger \end{bmatrix} \tag{C.5}$$

be the singular value decomposition of $\mathbf{\Phi}$, where $\sigma_{\max}(\mathbf{\Delta}) < 1$ then we have the following

Claim 5 *Suppose that the singular value decomposition of* $\mathbf{\Phi}$ *is given as in* (C.5) *and that for the input distribution* $p_{\mathbf{x}}$*, we have that* $I(\mathbf{x}; \mathbf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) < \infty$*, then,*

$$\mathbf{U}_1^\dagger \mathbf{z}_{\mathrm{r}} \overset{\text{a.s.}}{=} \mathbf{V}_1^\dagger \mathbf{z}_{\mathrm{e}} \tag{C.6a}$$

$$I(\mathbf{x}; \mathbf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) = I(\mathbf{x}; \mathbf{U}_2^\dagger \mathbf{y}_{\mathrm{r}} \mid \mathbf{y}_{\mathrm{e}}) \tag{C.6b}$$

The optimality of Gaussian inputs now follows from this claim, since the term $I(\mathbf{x}; \mathbf{U}_2^\dagger \mathbf{y}_{\mathrm{r}} \mid \mathbf{y}_{\mathrm{e}})$ can be expanded in the same manner as (C.2)-(C.3). The proof of Claim 5 is provided below.

Proof. To establish (C.6a), we simply note that

$$E[\mathbf{U}_1^\dagger \mathbf{z}_{\mathrm{r}} \mathbf{z}_{\mathrm{e}}^\dagger \mathbf{V}_1] = \mathbf{U}_1^\dagger \bar{\mathbf{\Phi}} \mathbf{V}_1 = \mathbf{I},$$

i.e., the Gaussian random variables $\mathbf{U}_1^\dagger \mathbf{z}_{\mathrm{r}}$ and $\mathbf{V}_1^\dagger \mathbf{z}_{\mathrm{e}}$ are perfectly correlated. Next note that

$$\begin{aligned} R_+(\mathbf{K}_{\mathrm{P}}, \bar{\mathbf{K}}_\Phi) &= I(\mathbf{x}; \mathbf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) \\ &= I(\mathbf{x}; \mathbf{U}_1^\dagger \mathbf{y}_{\mathrm{r}}, \mathbf{U}_2^\dagger \mathbf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) \qquad \text{(C.7)} \\ &= I(\mathbf{x}; \mathbf{U}_2^\dagger \mathbf{y}_{\mathrm{r}}, \mathbf{U}_1^\dagger \mathbf{y}_{\mathrm{r}} - \mathbf{V}_1^\dagger \mathbf{y}_{\mathrm{e}} | \mathbf{y}_{\mathrm{e}}) \\ &= I(\mathbf{x}; \mathbf{U}_2^\dagger \mathbf{y}_{\mathrm{r}}, \mathbf{U}_1^\dagger \mathbf{H}_{\mathrm{r}} \mathbf{x} - \mathbf{V}_1^\dagger \mathbf{H}_{\mathrm{e}} \mathbf{x} | \mathbf{y}_{\mathrm{e}}). \qquad \text{(C.8)} \end{aligned}$$

Since by hypothesis, $I(\mathbf{x}; \mathbf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) < \infty$, we have that $(\mathbf{U}_1^\dagger \mathbf{H}_{\mathrm{r}} - \mathbf{V}_1^\dagger \mathbf{H}_{\mathrm{e}})\mathbf{x} = 0$, and $I(\mathbf{x}; \mathbf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) = I(\mathbf{x}; \mathbf{U}_2^\dagger \mathbf{y}_{\mathrm{r}} \mid \mathbf{y}_{\mathrm{e}})$, establishing (C.6b). ∎

Finally if $p_{\mathbf{x}}$ is such that $I(\mathbf{x}; \mathbf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) = \infty$, then from (C.8), $(\mathbf{U}_1^\dagger \mathbf{H}_{\mathrm{r}} - \mathbf{V}_1^\dagger \mathbf{H}_{\mathrm{e}}) \mathbf{K}_{\mathrm{P}} (\mathbf{U}_1^\dagger \mathbf{H}_{\mathrm{r}} - \mathbf{V}_1^\dagger \mathbf{H}_{\mathrm{e}})^\dagger \neq \mathbf{0}$ and hence the choice of a Gaussian $p_{\mathbf{x}} = \mathcal{CN}(0, \mathbf{K}_{\mathrm{P}})$ also results in $I(\mathbf{x}; \mathbf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) = \infty$.

C.2 Matrix simplifications for establishing (5.24) from (5.34)

Substituting for $\bar{\mathbf{K}}_\Phi$ and $\mathbf{H}_t$ in (5.34) and carrying out the block matrix multiplication gives

$$\begin{aligned}
\mathbf{H}_r\bar{\mathbf{K}}_P\mathbf{H}_r^\dagger &= \boldsymbol{\Upsilon}_1(\mathbf{I}+\mathbf{H}_r\bar{\mathbf{K}}_P\mathbf{H}_r^\dagger)+\bar{\boldsymbol{\Phi}}\boldsymbol{\Upsilon}_2(\bar{\boldsymbol{\Phi}}^\dagger+\mathbf{H}_e\bar{\mathbf{K}}_P\mathbf{H}_r^\dagger)\\
\mathbf{H}_r\bar{\mathbf{K}}_P\mathbf{H}_e^\dagger &= \boldsymbol{\Upsilon}_1(\bar{\boldsymbol{\Phi}}+\mathbf{H}_r\bar{\mathbf{K}}_P\mathbf{H}_e^\dagger)+\bar{\boldsymbol{\Phi}}\boldsymbol{\Upsilon}_2(\mathbf{I}+\mathbf{H}_e\bar{\mathbf{K}}_P\mathbf{H}_e^\dagger)\\
\mathbf{H}_e\bar{\mathbf{K}}_P\mathbf{H}_r^\dagger &= \bar{\boldsymbol{\Phi}}^\dagger\boldsymbol{\Upsilon}_1(\mathbf{I}+\mathbf{H}_r\bar{\mathbf{K}}_P\mathbf{H}_r^\dagger)+\boldsymbol{\Upsilon}_2(\bar{\boldsymbol{\Phi}}^\dagger+\mathbf{H}_e\bar{\mathbf{K}}_P\mathbf{H}_r^\dagger)\\
\mathbf{H}_e\bar{\mathbf{K}}_P\mathbf{H}_e^\dagger &= \bar{\boldsymbol{\Phi}}^\dagger\boldsymbol{\Upsilon}_1(\bar{\boldsymbol{\Phi}}+\mathbf{H}_r\bar{\mathbf{K}}_P\mathbf{H}_e^\dagger)+\boldsymbol{\Upsilon}_2(\mathbf{I}+\mathbf{H}_e\bar{\mathbf{K}}_P\mathbf{H}_e^\dagger).
\end{aligned} \tag{C.9}$$

Eliminating $\boldsymbol{\Upsilon}_1$ from the first and third equation above, we have

$$(\bar{\boldsymbol{\Phi}}^\dagger\mathbf{H}_r-\mathbf{H}_e)\bar{\mathbf{K}}_P\mathbf{H}_r^\dagger=(\bar{\boldsymbol{\Phi}}^\dagger\bar{\boldsymbol{\Phi}}-\mathbf{I})\boldsymbol{\Upsilon}_2(\bar{\boldsymbol{\Phi}}^\dagger+\mathbf{H}_e\bar{\mathbf{K}}_P\mathbf{H}_r^\dagger). \tag{C.10}$$

Similarly eliminating $\boldsymbol{\Upsilon}_1$ from the second and fourth equations in (C.9) we have

$$(\bar{\boldsymbol{\Phi}}^\dagger\mathbf{H}_r-\mathbf{H}_e)\bar{\mathbf{K}}_P\mathbf{H}_e^\dagger=(\bar{\boldsymbol{\Phi}}^\dagger\bar{\boldsymbol{\Phi}}-\mathbf{I})\boldsymbol{\Upsilon}_2(\mathbf{I}+\mathbf{H}_e\bar{\mathbf{K}}_P\mathbf{H}_e^\dagger). \tag{C.11}$$

Finally, eliminating $\boldsymbol{\Upsilon}_2$ from (C.10) and (C.11) we obtain

$$\begin{aligned}
&(\bar{\boldsymbol{\Phi}}^\dagger\mathbf{H}_r-\mathbf{H}_e)\bar{\mathbf{K}}_P\mathbf{H}_r^\dagger\\
&=(\bar{\boldsymbol{\Phi}}^\dagger\mathbf{H}_r-\mathbf{H}_e)\bar{\mathbf{K}}_P\mathbf{H}_e^\dagger(\mathbf{I}+\mathbf{H}_e\bar{\mathbf{K}}_P\mathbf{H}_e^\dagger)^{-1}(\bar{\boldsymbol{\Phi}}^\dagger+\mathbf{H}_e\bar{\mathbf{K}}_P\mathbf{H}_r^\dagger)\\
&=(\bar{\boldsymbol{\Phi}}^\dagger\mathbf{H}_r-\mathbf{H}_e)\bar{\mathbf{K}}_P\mathbf{H}_e^\dagger\bar{\boldsymbol{\Theta}}^\dagger
\end{aligned} \tag{C.12}$$

which reduces to (5.24).

C.3 Derivation of (5.24) when the noise covariance is singular

Consider the compact singular value decomposition of $\bar{\mathbf{K}}_\Phi$:

$$\bar{\mathbf{K}}_\Phi=\mathbf{W}\bar{\boldsymbol{\Omega}}\mathbf{W}^\dagger, \tag{C.13}$$

where $\mathbf{W}$ is a matrix with orthogonal columns, i.e., $\mathbf{W}^\dagger\mathbf{W}=\mathbf{I}$ and $\bar{\boldsymbol{\Omega}}$ is a non-singular matrix. We first note that it must also be the case that

$$\mathbf{H}_t=\mathbf{W}\mathbf{G}, \tag{C.14}$$

i.e., the column space of $\mathbf{H}_t$ is a subspace of the column space of $\mathbf{W}$. If this were not the case then clearly $I(\mathbf{x};\mathbf{y}_r,\mathbf{y}_e)=\infty$ whenever the covariance matrix $\mathbf{K}_P$ has a component in the null space of $\mathbf{W}$ which implies that,

$$\max_{\mathbf{K}_P\in\mathcal{K}_P} R_+(\mathbf{K}_P,\bar{\mathbf{K}}_\Phi)=\infty. \tag{C.15}$$

Since $(\bar{\mathbf{K}}_{\mathrm{P}}, \bar{\mathbf{K}}_{\Phi})$ is a saddle point, we must have that $R_+(\bar{\mathbf{K}}_{\mathrm{P}}, \bar{\mathbf{K}}_{\Phi}) \leq R_+(\bar{\mathbf{K}}_{\mathrm{P}}, \mathbf{I}) < \infty$, and hence (C.14) must hold. Also note that since

$$\begin{aligned} R_+(\bar{\mathbf{K}}_{\mathrm{P}}, \bar{\mathbf{K}}_{\Phi}) &= \log\det(\mathbf{I} + \mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_{\mathrm{P}}\mathbf{H}_{\mathrm{e}}^{\dagger}) \\ &\quad + \log\frac{\det(\mathbf{G}\bar{\mathbf{K}}_{\mathrm{P}}\mathbf{G}^{\dagger} + \boldsymbol{\Omega})}{\det(\boldsymbol{\Omega})} \end{aligned} \tag{C.16}$$

it follows that $\bar{\boldsymbol{\Omega}}$ in (C.13) is a solution to the following minimization problem,

$$\begin{aligned} &\min_{\boldsymbol{\Omega}\in\mathcal{K}_{\Omega}} R_{\Omega}(\boldsymbol{\Omega}), \\ &R_{\Omega}(\boldsymbol{\Omega}) = \log\frac{\det(\mathbf{G}\bar{\mathbf{K}}_{\mathrm{P}}\mathbf{G}^{\dagger} + \boldsymbol{\Omega})}{\det(\boldsymbol{\Omega})}, \\ &\mathcal{K}_{\Omega} = \left\{ \boldsymbol{\Omega} \;\middle|\; \mathbf{W}\boldsymbol{\Omega}\mathbf{W}^{\dagger} = \begin{bmatrix} \mathbf{I}_{n_r} & \boldsymbol{\Phi} \\ \boldsymbol{\Phi}^{\dagger} & \mathbf{I}_{n_e} \end{bmatrix} \succeq 0 \right\}. \end{aligned} \tag{C.17}$$

The Kuhn-Tucker conditions for (C.17) yield,

$$\begin{aligned} &\bar{\boldsymbol{\Omega}}^{-1} - (\mathbf{G}\bar{\mathbf{K}}_{\mathrm{P}}\mathbf{G}^{\dagger} + \bar{\boldsymbol{\Omega}})^{-1} = \mathbf{W}^{\dagger}\boldsymbol{\Upsilon}\mathbf{W}, \\ &\Rightarrow \mathbf{G}\bar{\mathbf{K}}_{\mathrm{P}}\mathbf{G}^{\dagger} = \bar{\boldsymbol{\Omega}}\mathbf{W}^{\dagger}\boldsymbol{\Upsilon}\mathbf{W}(\bar{\boldsymbol{\Omega}} + \mathbf{G}\bar{\mathbf{K}}_{\mathrm{P}}\mathbf{G}^{\dagger}) \end{aligned} \tag{C.18}$$

where $\boldsymbol{\Upsilon}$ has the block diagonal form in (5.29). Multiplying the left and right and side of (C.18) with $\mathbf{W}$ and $\mathbf{W}^{\dagger}$ respectively and using (C.13) and (C.14) we have that

$$\mathbf{H}_{\mathrm{t}}\bar{\mathbf{K}}_{\mathrm{P}}\mathbf{H}_{\mathrm{t}}^{\dagger} = \bar{\mathbf{K}}_{\Phi}\boldsymbol{\Upsilon}(\bar{\mathbf{K}}_{\Phi} + \mathbf{H}_{\mathrm{t}}\bar{\mathbf{K}}_{\mathrm{P}}\mathbf{H}_{\mathrm{t}}^{\dagger}), \tag{C.19}$$

establishing (5.34). Finally note that the derivation in Appendix C.2 does not require the non-singularity assumption on $\bar{\mathbf{K}}_{\Phi}$.

C.4 Proof of Claim 4

To establish (5.38) note that since $\mathcal{H}(\cdot)$ is a concave function in $\mathbf{K}_{\mathrm{P}} \in \mathcal{K}_{\mathrm{P}}$ and differentiable over $\mathcal{K}_{\mathrm{P}}$, the optimality conditions associated with the Lagrangian

$$\mathcal{L}_{\Theta}(\mathbf{K}_{\mathrm{P}}, \lambda, \boldsymbol{\Psi}) = \mathcal{H}(\mathbf{K}_{\mathrm{P}}) + \mathrm{tr}(\boldsymbol{\Psi}\mathbf{K}_{\mathrm{P}}) - \lambda(\mathrm{tr}(\mathbf{K}_{\mathrm{P}}) - P), \tag{C.20}$$

are both necessary and sufficient. Thus $\mathbf{K}_{\mathrm{P}}$ is an optimal solution to (5.38) if and only if there exists a $\lambda \geq 0$ and $\boldsymbol{\Psi} \succeq 0$ such that

$$\begin{aligned} &(\mathbf{H}_{\mathrm{r}} - \bar{\boldsymbol{\Theta}}\mathbf{H}_{\mathrm{e}})^{\dagger}[\boldsymbol{\Gamma}(\mathbf{K}_{\mathrm{P}})]^{-1}(\mathbf{H}_{\mathrm{r}} - \bar{\boldsymbol{\Theta}}\mathbf{H}_{\mathrm{e}}) + \boldsymbol{\Psi} = \lambda\mathbf{I}, \\ &\mathrm{tr}(\boldsymbol{\Psi}\mathbf{K}_{\mathrm{P}}) = 0, \quad \lambda(\mathrm{tr}(\mathbf{K}_{\mathrm{P}}) - P) = 0, \end{aligned} \tag{C.21}$$

where $\mathbf{\Gamma}(\cdot)$ is defined via

$$\mathbf{\Gamma}(\mathbf{K}_{\mathrm{P}}) \triangleq \mathbf{I} + \bar{\mathbf{\Theta}}\bar{\mathbf{\Theta}}^{\dagger} - \bar{\mathbf{\Theta}}\bar{\mathbf{\Phi}}^{\dagger} - \bar{\mathbf{\Phi}}\bar{\mathbf{\Theta}}^{\dagger} + (\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})\mathbf{K}_{\mathrm{P}}(\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})^{\dagger}. \quad \text{(C.22)}$$

To obtain these parameters note that since $(\bar{\mathbf{K}}_{\mathrm{P}}, \bar{\mathbf{K}}_{\Phi})$ constitutes a saddle point solution,

$$\bar{\mathbf{K}}_{\mathrm{P}} \in \arg\max_{\mathbf{K}_{\mathrm{P}} \in \mathcal{K}_{\mathrm{P}}} R_{+}(\mathbf{K}_{\mathrm{P}}, \bar{\mathbf{K}}_{\Phi}). \quad \text{(C.23)}$$

Since $R_{+}(\mathbf{K}_{\mathrm{P}}, \bar{\mathbf{K}}_{\Phi})$ is differentiable at each $\mathbf{K}_{P} \in \mathcal{K}_{\mathrm{P}}$ whenever $\bar{\mathbf{K}}_{\Phi} \succ \mathbf{0}$, $\bar{\mathbf{K}}_{\mathrm{P}}$ satisfies the associated KKT conditions — there exists a $\lambda_0 \geq 0$ and $\mathbf{\Psi}_0 \succeq 0$ such that

$$\begin{aligned} \nabla_{\mathbf{K}_{\mathrm{P}}} R(\mathbf{K}_{\mathrm{P}}, \bar{\mathbf{K}}_{\Phi})\Big|_{\bar{\mathbf{K}}_{\mathrm{P}}} + \mathbf{\Psi}_0 &= \lambda_0 \mathbf{I} \\ \lambda_0(\mathrm{tr}(\bar{\mathbf{K}}_{\mathrm{P}}) - P) = 0, \quad \mathrm{tr}(\mathbf{\Psi}_0 \bar{\mathbf{K}}_{\mathrm{P}}) &= 0. \end{aligned} \quad \text{(C.24)}$$

As we show below,

$$\nabla_{\mathbf{K}_{\mathrm{P}}} R(\mathbf{K}_{\mathrm{P}}, \bar{\mathbf{K}}_{\Phi})\Big|_{\bar{\mathbf{K}}_{\mathrm{P}}} = (\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})^{\dagger}[\mathbf{\Lambda}(\bar{\mathbf{K}}_{\mathrm{P}})]^{-1}(\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}}), \quad \text{(C.25)}$$

where

$$\mathbf{\Lambda}(\mathbf{K}_{\mathrm{P}}) \triangleq \mathbf{I} + \mathbf{H}_{\mathrm{r}}\mathbf{K}_{\mathrm{P}}\mathbf{H}_{\mathrm{r}}^{\dagger} - (\mathbf{\Phi} + \mathbf{H}_{\mathrm{r}}\mathbf{K}_{\mathrm{P}}\mathbf{H}_{\mathrm{e}}^{\dagger})(\mathbf{I} + \mathbf{H}_{\mathrm{e}}\mathbf{K}_{\mathrm{P}}\mathbf{H}_{\mathrm{e}}^{\dagger})^{-1}(\mathbf{\Phi}^{\dagger} + \mathbf{H}_{\mathrm{e}}\mathbf{K}_{\mathrm{P}}\mathbf{H}_{\mathrm{r}}^{\dagger}) \quad \text{(C.26)}$$

$\mathbf{\Lambda}(\cdot)$, satisfies[1] $\mathbf{\Lambda}(\bar{\mathbf{K}}_{\mathrm{P}}) = \mathbf{\Gamma}(\bar{\mathbf{K}}_{\mathrm{P}})$. Hence the first condition in (C.24) reduces to

$$(\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})^{\dagger}[\mathbf{\Gamma}(\bar{\mathbf{K}}_{\mathrm{P}})]^{-1}(\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}}) + \mathbf{\Psi}_0 = \lambda_0 \mathbf{I}. \quad \text{(C.27)}$$

Comparing (C.24) and (C.27) with (C.21), we note that $(\bar{\mathbf{K}}_{\mathrm{P}}, \lambda_0, \mathbf{\Psi}_0)$ satisfy the conditions in (C.21), thus establishing (5.38).

It thus remains to establish (C.25), which we do below.

$$\begin{aligned} &\nabla_{\mathbf{K}_{\mathrm{P}}} R_{+}(\mathbf{K}_{\mathrm{P}}, \bar{\mathbf{K}}_{\Phi}) \\ &= \mathbf{H}_{\mathrm{t}}^{\dagger}(\mathbf{H}_{\mathrm{t}}\mathbf{K}_{\mathrm{P}}\mathbf{H}_{\mathrm{t}}^{\dagger} + \bar{\mathbf{K}}_{\Phi})^{-1}\mathbf{H}_{\mathrm{t}} - \mathbf{H}_{\mathrm{e}}^{\dagger}(\mathbf{I} + \mathbf{H}_{\mathrm{e}}\mathbf{K}_{\mathrm{P}}\mathbf{H}_{\mathrm{e}}^{\dagger})^{-1}\mathbf{H}_{\mathrm{e}}. \end{aligned} \quad \text{(C.28)}$$

[1]To verify this relation, note that $\mathbf{\Gamma}(\mathbf{K}_{\mathrm{P}})$ is the variance of $\mathbf{y}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{y}_{\mathrm{e}}$. When $\mathbf{K}_{\mathrm{P}} = \bar{\mathbf{K}}_{\mathrm{P}}$, note that $\bar{\mathbf{\Theta}}\mathbf{y}_{\mathrm{e}}$ is the MMSE estimate of $\mathbf{y}_{\mathrm{r}}$ given $\mathbf{y}_{\mathrm{e}}$ and $\mathbf{\Gamma}(\mathbf{K}_{\mathrm{P}})$ is the associated MMSE estimation error.

Substituting for $\mathbf{H}_t$ and $\bar{\mathbf{K}}_\Phi$ from (5.33) and (5.22),

$$\begin{aligned}
&(\bar{\mathbf{K}}_\Phi + \mathbf{H}_t\bar{\mathbf{K}}_P\mathbf{H}_t^\dagger)^{-1} \\
&= \begin{bmatrix} \mathbf{I} + \mathbf{H}_r\bar{\mathbf{K}}_P\mathbf{H}_r^\dagger & \bar{\mathbf{\Phi}} + \mathbf{H}_r\bar{\mathbf{K}}_P\mathbf{H}_e^\dagger \\ \bar{\mathbf{\Phi}}^\dagger + \mathbf{H}_r\bar{\mathbf{K}}_P\mathbf{H}_e^\dagger & \mathbf{I} + \mathbf{H}_e\bar{\mathbf{K}}_P\mathbf{H}_e^\dagger \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \mathbf{\Lambda}(\mathbf{K}_P)^{-1} & -\mathbf{\Lambda}(\mathbf{K}_P)^{-1}\bar{\mathbf{\Theta}} \\ -\bar{\mathbf{\Theta}}^\dagger\mathbf{\Lambda}(\mathbf{K}_P)^{-1} & (\mathbf{I}+\mathbf{H}_e\bar{\mathbf{K}}_P\mathbf{H}_e)^{-1}+\bar{\mathbf{\Theta}}^\dagger\mathbf{\Lambda}(\mathbf{K}_P)^{-1}\bar{\mathbf{\Theta}} \end{bmatrix},
\end{aligned}$$

where we have used the matrix inversion lemma (e.g., [44]), and $\mathbf{\Lambda}(\bar{\mathbf{K}}_P)$ is defined in (C.4), and $\bar{\mathbf{\Theta}}$ is as defined in (5.23). Substituting into (C.28) and simplifying gives

$$\begin{aligned}
&\nabla_{\mathbf{K}_P} R_+(\mathbf{K}_P, \bar{\mathbf{K}}_\Phi)\Big|_{\bar{\mathbf{K}}_P} \\
=&\mathbf{H}_t^\dagger(\bar{\mathbf{K}}_\Phi + \mathbf{H}_t\bar{\mathbf{K}}_P\mathbf{H}_t^\dagger)^{-1}\mathbf{H}_t - \mathbf{H}_e^\dagger(\mathbf{I} + \mathbf{H}_e\bar{\mathbf{K}}_P\mathbf{H}_e^\dagger)^{-1}\mathbf{H}_e \\
&= (\mathbf{H}_r - \bar{\mathbf{\Theta}}\mathbf{H}_e)^\dagger[\mathbf{\Lambda}(\bar{\mathbf{K}}_P)]^{-1}(\mathbf{H}_r - \bar{\mathbf{\Theta}}\mathbf{H}_e)
\end{aligned}$$

as required.

C.5 Full Rank Condition for Optimal Solution

Claim 6 *Suppose that* $\bar{\mathbf{K}}_\Phi \succ \mathbf{0}$ *and* $\hat{\mathbf{K}}_P$ *be any optimal solution to*

$$\hat{\mathbf{K}}_P \in \arg\max_{\mathbf{K}_P \in \mathcal{K}_P} \log\det(\mathbf{I}+\mathbf{J}^{-\frac{1}{2}}(\mathbf{H}_r-\bar{\mathbf{\Theta}}\mathbf{H}_e)\mathbf{K}_P(\mathbf{H}_r-\bar{\mathbf{\Theta}}\mathbf{H}_e)^\dagger\mathbf{J}^{-\frac{1}{2}}) \tag{C.29}$$

for some $\mathbf{J} \succ 0$ *and* $\bar{\mathbf{\Theta}}$ *is defined in* (5.23). *Suppose that* $\mathbf{S}_P$ *is a matrix with a full column rank such that*

$$\hat{\mathbf{K}}_P = \mathbf{S}_P\mathbf{S}_P^\dagger \tag{C.30}$$

then $(\mathbf{H}_r - \bar{\mathbf{\Theta}}\mathbf{H}_e)\mathbf{S}_P$ *has a full column rank.*

Define

$$\mathbf{H}_{\text{eff}} \triangleq \mathbf{J}^{-\frac{1}{2}}(\mathbf{H}_r - \bar{\mathbf{\Theta}}\mathbf{H}_e).$$

It suffices to prove that $\mathbf{H}_{\text{eff}}\mathbf{S}_P$ has a full column rank, which we now do.

Let $\text{rank}(\mathbf{H}_{\text{eff}}) = \nu$ and let

$$\mathbf{H}_{\text{eff}} = \mathbf{A}\mathbf{\Sigma}\mathbf{B}^\dagger \tag{C.31}$$

be the singular value decomposition of $\mathbf{H}_{\text{eff}}$ where $\mathbf{A}$ and $\mathbf{B}$ are unitary matrices, and

$$\mathbf{\Sigma} = \begin{array}{c} \\ \nu \\ n_r-\nu \end{array} \begin{array}{c} \begin{array}{cc} \nu & n_t-\nu \end{array} \\ \begin{bmatrix} \mathbf{\Sigma}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \end{array}. \tag{C.32}$$

Note that it suffices to show that the matrix

$$\hat{\mathbf{F}} \triangleq \mathbf{B}^\dagger \hat{\mathbf{K}}_\mathrm{P} \mathbf{B} \tag{C.33}$$

has the form

$$\hat{\mathbf{F}} = \begin{array}{c} \\ \nu \\ n_\mathrm{t}-\nu \end{array} \begin{array}{c} \begin{array}{cc} \nu & n_\mathrm{t}-\nu \end{array} \\ \begin{bmatrix} \mathbf{F}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \end{array}. \tag{C.34}$$

Since,

$$\begin{aligned} \hat{\mathbf{K}}_\mathrm{P} &\in \arg\max_{\mathcal{K}_P} \log\det(\mathbf{I} + \mathbf{H}_\mathrm{eff} \mathbf{K}_\mathrm{P} \mathbf{H}_\mathrm{eff}^\dagger) \\ &= \arg\max_{\mathcal{K}_P} \log\det(\mathbf{I} + \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^\dagger \mathbf{K}_\mathrm{P} \mathbf{B}\boldsymbol{\Sigma}^\dagger \mathbf{A}^\dagger) \\ &= \arg\max_{\mathcal{K}_P} \log\det(\mathbf{I} + \boldsymbol{\Sigma}\mathbf{B}^\dagger \mathbf{K}_\mathrm{P} \mathbf{B}\boldsymbol{\Sigma}^\dagger), \end{aligned} \tag{C.35}$$

and $\mathbf{K}_\mathrm{P} \in \mathcal{K}_\mathrm{P}$ if and only if $\mathbf{B}^\dagger \mathbf{K}_\mathrm{P} \mathbf{B} \in \mathcal{K}_\mathrm{P}$, observe that,

$$\hat{\mathbf{F}} \in \arg\max_{\mathcal{K}_P} \log\det(\mathbf{I} + \boldsymbol{\Sigma}\mathbf{F}\boldsymbol{\Sigma}^\dagger) \tag{C.36}$$

$$= \arg\max_{\mathcal{K}_P} \log\det(\mathbf{I} + \boldsymbol{\Sigma}_0 \mathbf{F}_0 \boldsymbol{\Sigma}_0^\dagger), \tag{C.37}$$

where the $\mathbf{F}$ is of the form

$$\mathbf{F} = \begin{array}{c} \\ \nu \\ n_\mathrm{t}-\nu \end{array} \begin{array}{c} \begin{array}{cc} \nu & n_\mathrm{t}-\nu \end{array} \\ \begin{bmatrix} \mathbf{F}_0 & \mathbf{F}_1 \\ \mathbf{F}_1^\dagger & \mathbf{F}_2 \end{bmatrix} \end{array}. \tag{C.38}$$

We now note that $\hat{\mathbf{F}}_1 = 0$ and $\hat{\mathbf{F}}_2 = 0$. Indeed if $\hat{\mathbf{F}}_2 \neq 0$, then $\mathrm{tr}(\hat{\mathbf{F}}_2) > 0$. This contradicts the optimality claim in (C.37), since the objective function only depends on $\hat{\mathbf{F}}_0$ and one can strictly increase the objective function by increasing the trace of $\hat{\mathbf{F}}_0$. Finally since $\hat{\mathbf{F}} \succeq 0$ and $\hat{\mathbf{F}}_2 = 0$, it follows that $\hat{\mathbf{F}}_1 = 0$.

C.6 Full rank condition when $\bar{\mathbf{K}}_\Phi$ is singular

In this section we establish 2) in Lemma 9 when $\bar{\mathbf{K}}_\Phi$ is singular. We map this case to another channel when the saddle point noise covariance is non-singular and apply the results for non-singular noise covariance.

When $\bar{\mathbf{K}}_\Phi$ is singular, we have that $\bar{\boldsymbol{\Phi}}$ has $d \geq 1$ singular values equal to unity and hence we express its SVD in (C.5), where $\sigma_{\max}(\boldsymbol{\Delta}) < 1$.

Following Claim 5 in Appendix C.1 we have that

$$\mathbf{U}_1^\dagger \mathbf{z}_\mathrm{r} \overset{\text{a.s.}}{=} \mathbf{V}_1^\dagger \mathbf{z}_\mathrm{e} \tag{C.39a}$$

$$\mathbf{U}_1^\dagger \mathbf{H}_\mathrm{r} = \mathbf{V}_1^\dagger \mathbf{H}_\mathrm{e}, \tag{C.39b}$$

$$R_+(\mathbf{K}_\mathrm{P}, \bar{\mathbf{K}}_\Phi) = I(\mathbf{x}; \mathbf{U}_2^\dagger \mathbf{y}_\mathrm{r} \mid \mathbf{y}_\mathrm{e}), \quad \forall\, \mathbf{K}_\mathrm{P} \in \mathcal{K}_\mathrm{P}. \tag{C.39c}$$

Thus with $\hat{\mathbf{H}}_\mathrm{r} = \mathbf{U}_2^\dagger \mathbf{H}_\mathrm{r}$, and $\hat{\mathbf{z}}_\mathrm{r} = \mathbf{U}_2^\dagger \mathbf{z}_\mathrm{r}$ and

$$\hat{\mathbf{y}}_\mathrm{r} = \mathbf{U}_2^\dagger \mathbf{y}_\mathrm{r} = \hat{\mathbf{H}}_\mathrm{r} \mathbf{x} + \hat{\mathbf{z}}_\mathrm{r}, \tag{C.40}$$

we have from (C.39c), that

$$\bar{\mathbf{K}}_\mathrm{P} \in \arg\max_{\mathcal{K}_\mathrm{P}} I(\mathbf{x}; \hat{\mathbf{y}}_\mathrm{r} \mid \mathbf{y}_\mathrm{e}). \tag{C.41}$$

Since the associated cross-covariance matrix $\hat{\mathbf{\Phi}} = E[\hat{\mathbf{z}}_\mathrm{r} \mathbf{z}_\mathrm{e}^\dagger]$ has all its singular values strictly less than unity, it follows from Claim 4 that

$$\bar{\mathbf{K}}_\mathrm{P} \in \arg\max_{\mathcal{K}_P} \hat{\mathcal{H}}(\mathbf{K}_\mathrm{P}) \tag{C.42}$$

where

$$\begin{aligned} \hat{\mathcal{H}}(\mathbf{K}_\mathrm{P}) &= h(\hat{\mathbf{y}}_\mathrm{r} - \hat{\mathbf{\Theta}} \mathbf{y}_\mathrm{e}), \\ \hat{\mathbf{\Theta}} &= \mathbf{U}_2^\dagger (\mathbf{H}_\mathrm{r} \bar{\mathbf{K}}_\mathrm{P} \mathbf{H}_\mathrm{e}^\dagger + \bar{\mathbf{\Phi}})(\mathbf{I} + \mathbf{H}_\mathrm{e} \bar{\mathbf{K}}_\mathrm{P} \mathbf{H}_\mathrm{e}^\dagger)^{-1}. \end{aligned}$$

Following the proof of Claim 6 in Appendix C.5 we then have that

$$(\hat{\mathbf{H}}_\mathrm{r} - \hat{\mathbf{\Theta}} \mathbf{H}_\mathrm{e}) \mathbf{S} = \mathbf{U}_2^\dagger (\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}} \mathbf{H}_\mathrm{e}) \mathbf{S}$$

has a full column rank. This in turn implies that $(\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}} \mathbf{H}_\mathrm{e}) \mathbf{S}$ has a full column rank.

C.7 Proof of Lemma 10 when $\bar{\mathbf{K}}_\Phi$ is singular

When $\bar{\mathbf{K}}_\Phi$ is singular, we assume that the singular value decomposition of $\bar{\mathbf{\Phi}}$ is given in (C.5). First let us consider the case that $\mathbf{H}_\mathrm{r} = \bar{\mathbf{\Theta}} \mathbf{H}_\mathrm{e}$ and show that $R_+(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi) = 0$. Indeed following claim 5 in Appendix C.1 we have that $R_+(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi) = I(\mathbf{x}; \mathbf{U}_2^\dagger \mathbf{y}_\mathrm{r} | \mathbf{y}_\mathrm{e})$ and expanding this expression in the same manner as (5.43)-(5.45), we establish the desired result.

When $\mathbf{H}_\mathrm{r} - \bar{\Theta} \mathbf{H}_\mathrm{e} \neq \mathbf{0}$, we show that the difference between the upper and lower bounds is zero.

$$\begin{aligned} \Delta R &= R_+(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi) - R_-(\bar{\mathbf{K}}_\mathrm{P}) \\ &= I(\mathbf{x}; \mathbf{y}_\mathrm{e} \mid \mathbf{y}_\mathrm{r}) \\ &= I(\mathbf{x}; \mathbf{V}_2^\dagger \mathbf{y}_\mathrm{e} \mid \mathbf{y}_\mathrm{r}), \end{aligned} \tag{C.43}$$

where the last step follows from the fact that $\mathbf{U}_1^\dagger \mathbf{z}_\mathrm{r} \overset{\text{a.s.}}{=} \mathbf{V}_1^\dagger \mathbf{z}_\mathrm{e}$ and $\mathbf{U}_1^\dagger \mathbf{H}_\mathrm{r} = \mathbf{V}_1^\dagger \mathbf{H}_\mathrm{e}$

(c.f. (C.39a), (C.39b)). Next, note that,

$$
\begin{aligned}
& h(\mathbf{V}_2^\dagger \mathbf{y}_e \mid \mathbf{y}_r) \\
& = \log\det(\mathbf{I} + \mathbf{V}_2^\dagger \mathbf{H}_e \bar{\mathbf{K}}_P \mathbf{H}_e^\dagger \mathbf{V}_2 - (\mathbf{V}_2^\dagger \mathbf{H}_e \bar{\mathbf{K}}_P \mathbf{H}_r^\dagger + \boldsymbol{\Delta}^\dagger \mathbf{U}_2^\dagger) \\
& \qquad (\mathbf{I} + \mathbf{H}_r \bar{\mathbf{K}}_P \mathbf{H}_r^\dagger)^{-1} (\mathbf{H}_r \bar{\mathbf{K}}_P \mathbf{H}_e^\dagger \mathbf{V}_2 + \mathbf{U}_2 \boldsymbol{\Delta})) \qquad \text{(C.44)} \\
& = \log\det(\mathbf{I} + \boldsymbol{\Delta}^\dagger \mathbf{U}_2^\dagger \mathbf{H}_r \bar{\mathbf{K}}_P \mathbf{H}_r^\dagger \mathbf{U}_2 \boldsymbol{\Delta} \\
& \qquad - \boldsymbol{\Delta}^\dagger \mathbf{U}_2^\dagger (\mathbf{I} + \mathbf{H}_r \bar{\mathbf{K}}_P \mathbf{H}_r^\dagger) \mathbf{U}_2 \boldsymbol{\Delta}) \\
& = \log\det(\mathbf{I} - \boldsymbol{\Delta}^\dagger \boldsymbol{\Delta}) \\
& = h(\mathbf{V}_2^\dagger \mathbf{z}_e \mid \mathbf{U}_2^\dagger \mathbf{z}_r) = h(\mathbf{V}_2^\dagger \mathbf{z}_e \mid \mathbf{z}_r), \qquad \text{(C.45)}
\end{aligned}
$$

where we have used (c.f. (5.46)) that

$$
\mathbf{V}_2^\dagger \bar{\boldsymbol{\Phi}}^\dagger \mathbf{H}_r \mathbf{S} = \mathbf{V}_2^\dagger \mathbf{H}_e \mathbf{S} \Rightarrow \boldsymbol{\Delta}^\dagger \mathbf{U}_2^\dagger \mathbf{H}_r \mathbf{S} = \mathbf{V}_2^\dagger \mathbf{H}_e \mathbf{S},
$$

in simplifying (C.44) and the equality in (C.45) follows from the fact that $\mathbf{U}_1^\dagger \mathbf{z}_r$ is independent of $(\mathbf{U}_2^\dagger \mathbf{z}_r, \mathbf{V}_2^\dagger \mathbf{z}_e)$.

Appendix D

Conditional Entropy Lemma

Lemma 17 *Suppose that the random variables a,b, and c are finite valued with a joint distribution $p_{a,b,c}(\cdot)$ that satisfies $a \rightarrow b \rightarrow c$ For some $N \geq 0$ and $R > I(c;a)$ suppose that a set $\mathcal{C}_c$ is selected by drawing $\exp(NR)$ sequences $\{c_i^N\}$ uniformly and at random from the set of p_c typical sequences T_c^N.*

Suppose that the pair of length-N sequences (a^N, b^N) are drawn i.i.d. from the distribution $p_{a,b}$ and a sequence $c_i^N \in \mathcal{C}_c$ is selected such that $(c_i^N, b^N) \in T_{cb,\eta}^N$. Then

$$\frac{1}{N} H(c_i^N | a^N) = R - I(c;a) + o_\eta(1), \tag{D.1}$$

where the term $o_\eta(1)$ vanishes to zero as $N \rightarrow \infty$ and $\eta \rightarrow 0$.

Proof. From (6.23c), for all pair of sequences (a^N, b^N), except a set of probability $o_\eta(1)$, we have that $(a^N, b^N) \in T_{ab,\eta}^N$. Furthermore, for each such typical pair, since $a \rightarrow b \rightarrow c$ and $(b^N, c_i^N) \in T_{bc,\eta}^N$ from the Markov Lemma it follows that $(a^N, c_i^N) \in T_{ac,\eta}^N$.

To establish (D.1) it suffices to show that for all sequences $a^N \in T_{a,\eta}^N$, except a set of size at most $o_\eta(1)$

$$\Pr(c^N = c_i^N | a^N = a^N) = \exp(-N(R - I(c;a) + o_\eta(1))). \tag{D.2}$$

The expression in (D.1) follows by due to the continuity of the log() function. To establish (D.2), we use the fact that

$$\Pr(c^N = c_i^N | a^N = a^N) = \frac{p(a^N | c_i^N) \Pr(c^N = c_i^n)}{p(a^n)(a^N)}. \tag{D.3}$$

From property (6.23b) of typical sequences $p(a^N) = \exp(-N(H(a)+o_\eta(1)))$, $p(a^N|c_i^N) = \exp(-N(H(a|a) + o_\eta(1)))$ and from symmetriy $\Pr(c^N = c_i^N) = \exp(-NR)$. Substituting these quantities in (D.3) establishes (D.2). ∎

Bibliography

[1] LAPACK users' guide, Third Edition. http://www.netlib.org/lapack/lug/lapack_lug.html, August 1999.

[2] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography – Part I: Secret sharing. *IEEE Trans. Inform. Theory*, 39:1121–1132, July 1993.

[3] Z. D. Bai and J. W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large dimensional random matrices. *Annals of Probability*, 26:316–345, 1998.

[4] G. Caire and S. Shamai. On the capacity of some channels with channel state information. *IEEE Trans. Inform. Theory*, 45:2007–2019, 1999.

[5] B. Chor, A. Fiat, M. Naor, and B. Pinkas. Tracing traitors. *IEEE Trans. Inform. Theory*, pages 893–910, May 2000.

[6] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley and Sons, 1991.

[7] I. Csiszár. Almost independence and secrecy capacity (in russian). *Probl. Inform. Transmission*, 32:48–57, 1996.

[8] I. Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Trans. Inform. Theory*, 24:339–348, 1978.

[9] I. Csiszár and J. Körner. *Information Theory, Coding Theorems for Discrete Memoryless Systems.* Akadémiai Kiadó, 1981.

[10] I. Csiszár and P. Narayan. Common randomness and secret key generation with a helper. *IEEE Trans. Inform. Theory*, 46, March 2000.

[11] I. Csiszár and P. Narayan. Secrecy capacities for multiple terminals. *IEEE Trans. Inform. Theory*, 50:3047–3061, 2004.

[12] H. A. David. *Order Statistics.* New York: Wiley, 1981.

[13] S. N. Diggavi and T. M. Cover. The worst additive noise under a covariance constraint. *IEEE Trans. Inform. Theory*, IT-47(7):3072–3081, 2001.

[14] S. C. Draper, A. Khisti, E. Martinian, J. Yedidia, and A. Vetro. Using distributed source coding to secure fingerprint biometrics. In *Proc. Int. Conf. Acoust. Speech, Signal Processing*, 2007.

[15] A. A. El Gamal. Capacity of the product and sum of two un-matched broadcast channels. *Probl. Information Transmission*, pages 3–23, 1980.

[16] A. Fiat and M. Naor. Broadcast encryption. In *Proceedings of the 13th annual international cryptology conference on Advances in cryptology*, pages 480–491, Santa Barbara, CA, 1994.

[17] S. I. Gel'fand and M. S. Pinsker. Coding for channels with random parameters. *Problems of Control and Information Theory*, 9:19–31, 1980.

[18] S. Goel and R. Negi. Secret communication in presence of colluding eavesdroppers. In *Proc. IEEE Military Commun. Conf.*, 2005.

[19] G. Golub and C. F. Van Loan. *Matrix Computations (3rd ed).* Johns Hopkins University Press, 1996.

[20] P. Gopala, L. Lai, and H. El Gamal. On the secrecy capacity of fading channels. *IEEE Trans. Inform. Theory*, submitted, 2006.

[21] M. Kang and M. S. Alouini. Hotelling's generalized distribution and performance of 2d-rake receivers. *IEEE Trans. Inform. Theory*, 49:317–23, January 2003.

[22] A. Khisti, A. Tchamkerten, and G. W. Wornell. Secure broadcasting with multiuser diversity. In *Proc. Allerton Conf. Commun., Contr., Computing*, 2006.

[23] A. Khisti, A. Tchamkerten, and G. W. Wornell. Secure Broadcasting over Fading Channels. *IEEE Trans. Inform. Theory, Special Issue on Information Theoretic Security*, pages 2453–2469, 2008.

[24] A. Khisti and G. W. Wornell. Secure transmission with multiple antennas: The MISOME wiretap channel. *Submitted Aug. 2007, IEEE Trans. Inform. Theory, available online, http://arxiv.org/abs/0708.4219.*

[25] A. Khisti, G. W. Wornell, A. Wiesel, and Y. Eldar. On the Gaussian MIMO wiretap channel. In *Proc. Int. Symp. Inform. Theory*, Nice, 2007.

[26] J. Korner and K. Marton. General broadcast channel with degraded message sets. *IEEE Trans. Inform. Theory*, 23:60–64, 1977.

[27] S. K. Leung-Yan-Cheong and M. E. Hellman. The Gaussian wiretap channel. *IEEE Trans. Inform. Theory*, 24:451–56, 1978.

[28] C. Li and R. Mathias. Extremal characterizations of the Schur complement and resulting inequalities. *SIAM Review*, 42:233–46, 2000.

[29] L. Li and A. J. Goldsmith. Optimal resource allocation for fading broadcast channels- part I: Ergodic capacity. *IEEE Trans. Inform. Theory*, 47:1083–1102, March 2001.

[30] Z. Li, W. Trappe, and R. Yates. Secret communication via multi-antenna transmission. In *Forty-First Annual Conference on Information Sciences and Systems (CISS)*, Baltimore, MD, March 2007.

[31] Z. Li, W. Trappe, and R. Yates. Secure communicatoin with a fading eavesdropper channel. In *Proc. Int. Symp. Inform. Theory*, Nice, France, 2007.

[32] Y. Liang and H. V. Poor. Secure communication over fading channels. In *Proc. Allerton Conf. Commun., Contr., Computing*, 2006.

[33] Y. Liang, H. V. Poor, and S. Shamai. Secure communication over fading channels. *IEEE Trans. Inform. Theory*, submitted.

[34] C. F. V. Loan. Generalizing the singular value decomposition. *SIAM Journal on Numerical Analysis*, 13:76–83, 1976.

[35] A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Applications.* Academic Press, 1979.

[36] E. Martinian. Waterfilling gains O(1/SNR) at high SNR. unpublished, http://www.csua.berkeley.edu/ emin/research/wfill.pdf, February 2004.

[37] U. M. Maurer. Secret key agreement by public discussion from common information. *IEEE Trans. Inform. Theory*, 39:733–742, March 1993.

[38] U. M. Maurer and S. Wolf. Information-theoretic key agreement: from weak to strong secrecy for free. In *EUROCRYPT*, 2000.

[39] Neri Merhav and Erdal Arikan. The shannon cipher system with a guessing wiretapper. *IEEE Transactions on Information Theory*, 45(6):1860–1866, 1999.

[40] R. J. Muirhead. *Aspects of Multivariate Statistical Theory.* Wiley, 1982.

[41] R. Negi and S. Goel. Secret communication using artificial noise. In *Proc. Vehic. Tech. Conf.*, 2005.

[42] B. Obama. *The Audacity of Hope: Thoughts on Reclaiming the American Dream.* Crown/Three Rivers Press, 2006.

[43] C. Paige and M. A. Saunders. Towards a generalized singular value decomposition. *SIAM J. Numer. Anal*, 18:398–405, 1981.

[44] K. Petersen and M. Pedersen. The Matrix Cookbook, September, 2007.

[45] S. Shaifee and S. Ulukus. Achievable rates in Gaussian MISO channels with secrecy constraints. In *Proc. Int. Symp. Inform. Theory*, June 2007.

[46] C. E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28:656–715, 1949.

[47] J. W. Silverstein. The limiting eigenvalue distribution of a multivariate F- matrix. *SIAM Journal on Mathematical Analysis*, 16:641–646, 1985.

[48] D. Tse. Optimal power allocation over parallel Gaussian broadcast channels. unpublished, 1999.

[49] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication.* Cambridge University Press, 2005.

[50] A. M. Tulino and S. Verdu. Random matrix theory and wireless communications. *Foundations and Trends in Communications and Information Theory, Now Publishers*, 2004.

[51] S. Wilks. *Mathematical Statistics.* John Wiley, 1962.

[52] R. Wilson, D. Tse, and R. Scholtz. Channel Identification: Secret Sharing using Reciprocity in UWB Channels. *submitted to IEEE Transactions on Information Forensics and Security*, March 2006.

[53] A. D. Wyner. The wiretap channel. *Bell Syst. Tech. J.*, 54:1355–87, 1975.