SEMANTIC SECURITY FOR THE FAST FADING WIRETAP CHANNEL.

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Parker Lee Pavlicek

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Electrical and Computer Engineering

April 2018

Fargo, North Dakota

NORTH DAKOTA STATE UNIVERSITY

Graduate School

Title

SEMANTIC SECURITY FOR THE FAST FADING WIRETAP CHANNEL.

By

Parker Lee Pavlicek

The supervisory committee certifies that this thesis complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Sanjay Karmakar

Chair

Dr. Michael Cohen

Dr. Ivan Lima

Approved:

9 April 2018

Date

Dr. Benjamin Braaten

Department Chair

ABSTRACT

We provide a set of semantically secure achievable rates for the fast fading wiretap channel. In particular, we do so for the cases where there is channel state information at the transmitter (CSIT) for both the main and eavesdropper channels (full CSIT), for only the main channel (partial CSIT), and for neither channel (statistical CSIT).

In the case of partial CSIT and statistical CSIT fast-fading channels, we show that this coding scheme can achieve the best known achievable rates. In the case of full CSIT fast-fading wiretap channels, we show that this coding scheme can actually achieve the secrecy capacity. In particular, this implies that the semantic secrecy capacity for these channels is equivalent to the weak and strong secrecy capacities. Moreover, we achieve these rates in a way that is non-invasive to existing systems and also happens to be explicitly given as well as efficient in implementation.

ACKNOWLEDGEMENTS

I would like to acknowledge my research partner Eric Kubischta and advisor Dr. Sanjay Karmakar for their collaboration and guidance throughout the past two years. I would also like to thank Himanshu Tyagi of the Indian Institute of Science and Alexander Vardy of UC San Diego upon whose work this thesis is largely motivated.

DEDICATION

To Grandma Betty and Grandpa Izzy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CSIT . . . . . Channel State Information at the Transmitter

UHF . . . . . Universal Hash Family

AWGN . . . . . Additive White Gaussian Noise

i.i.d. . . . . Independent and Identically Distributed

1. INTRODUCTION

The goal of any security system is to provide a way for a transmitter and legitimate receiver to communicate while a third party, known as an adversary or eavesdropper, who also receives the transmitted signals, remains ignorant of their meaning. Most modern communication systems employ various security measures when transferring data from a transmitter to a legitimate receiver and these systems mostly obtain security via computational based cryptographic methods. This has been a suitable security solution for the past century but it relies on a large assumption which is being seriously challenged by the rise of quantum computers (and potentially other new computer architectures). The assumption made in the field of computational based cryptology is that the adversary (eavesdropper) is computationally bounded. This means that it is theoretically possible to break a cryptographic system, but it is not feasible to do so in any practical amount of time or with any practical amount of computing power.

Many cryptographic schemes are based on the assumption that integer factorization is a *hard* problem, that is, a problem whose solution is inefficient to find, and thus computationally bounded adversaries will not be able to break the scheme. However, this assumption is made using the further assumption that the factoring is being done on a classical computer. Integer factorization is not a hard problem for a quantum computer using quantum algorithms. It was shown in [22] that Shor's Algorithm can factor large integers in polynomial time on a quantum computer, rendering many cryptographic algorithms ineffective. Furthermore, to date, there has been no mathematical proof showing that integer factorization (and similar problems) is actually hard for classical computers as well. In other words, there is nothing precluding the discovery of an efficient algorithm for classical computers. Due to these risks, it is desirable to implement a new form of security on our communication systems. A leading candidate for this new form of security is information-theoretic security, also known as physical-layer security.

Information-theoretic security exploits the inherent randomness present in a communication channel due to ambient noise and interference. This form of security does not make the assumption that the adversary is computationally bounded; it is still provably unbreakable regardless of the adversary's computational capabilities. In an information-theoretic secure system, an adversary just

simply does not have enough information available to ever deduce the original message with a high probability. This idea of information-theoretic security was first formulated and put forth in 1949 by Claude Shannon in his seminal paper "Communication Theory of Secrecy Systems" [21]. After putting forth the idea, Shannon proved that it was indeed possible to realize information-theoretic security in real-world applications. However, information-theoretic security was soon abandoned by the security community as it was deemed impractical.

Although information-theoretic security offers many potential advantages over traditional cryptography, it has been largely ignored for decades and has been thought of mainly as a theoretical peculiarity. Shortly after Wyner revived the field of information-theoretic security in the seventies by formulating the mathematical model (the wiretap channel) for which we now rely, Diffie and Hellman introduced the world to public-key cryptography to which the attention of the security community soon shifted due to its readiness in applicability. Although Wyner's result was important in bringing back the field of information-theoretic secrecy, it was still highly theoretic - there were no practical ways to achieve it, further pushing the security community to tried and true (at the time) cryptographic methods. Focus on information-theoretic security has grown substantially since the turn of the century and continues to grow as researchers are realizing its potential. Further potential of the subject is being shown in a newly blossoming field known as quantum information theory which has a secrecy component to it as well.

As with most technologies, there are admittedly some down sides to information-theoretic security. To date, there are not many schemes in existence that can actually be implemented in a practical setting to provide this kind of security; most proposed schemes are too impractical. The impracticality of these proposed schemes comes in two forms. The first being the lack of a concrete method to implement the scheme, which we refer to as a scheme's lack of *explicitness*. Computational based cryptographic methods can be easily implemented in practice due to the fact that most of them can be coded in a programming language by a modest student of computer science and put to use immediately. In contrast, many existing techniques to ensure information-theoretic security rely on random coding arguments which have no way of being coded or implemented directly. The other form of impracticality is that of super-polynomial time complexity, which we will refer to as a scheme's lack of *efficiency*. Explicit techniques have been given that ensure information-theoretic security but the time or space complexity of the operations necessary to achieve it are

such that it is infeasible in practice. The simplest example of this is that of Shannon's one-time pad (although Shannon was not the original inventor of the one-time pad, that credit belongs to Frank Miller [18], he was the first to show its capabilities in an information-theoretic sense). This scheme ensures the best information-theoretic secrecy possible (this is given in more detail later in the thesis) and is explicitly given, but it has the downfall that it requires keys which are the same size as that of the message [21]. Furthermore, a new independent key must be generated for each message. Obviously, when working with gigabytes (or larger) of data such as are common in modern communication systems, generating and distributing keys of equal size is quite impractical.

Another deterrent to information-theoretic security being widely adopted in practice is that many proposed schemes (impractical or practical) do not provide a sufficient amount of security to be trusted with sensitive data. As will be shown in forthcoming chapters, there exist metrics to measure "how secure" a certain scheme is; many of the current schemes proposed in literature only provide security on the low end of this spectrum. So low, in fact, that in some cases an infinite amount of information can still be deduced by the adversary when using the proposed scheme. In order for information-theoretic security to become widely adopted, practical schemes which provide a large amount of security must be developed.

An additional assumption upon which many results in information-theoretic security rest is that of the adversary having a physically worse channel than the legitimate communicating parties. A precise definition of a channel and what it means to be "worse" are given in subsequent chapters but for now it suffices to interpret this assumption as meaning that the adversary is placed physically farther away from the legitimate receiver or suffers from more physical interference than that of the legitimate receiver. In many settings, this is quite a feasible assumption. One such suitable setting is that of wireless communications.

Wireless communication has become ubiquitous in modern society over the past century. A copious amount of data is being transmitted wirelessly every second; from social interactions and financial information, to military communications, data is traveling through our atmosphere like never before. With much of this data being highly sensitive, security measures must be devised to protect this data from malicious parties during transmission.

The goal of this thesis is to bring information-theoretic security for wireless communications from the realm of the theoretical to the realm of the practical. However, much more happens. We

do not just take what the current theoretical results are for wireless communications and make them practical; we first *improve* those theoretical results and then make *those* improved results practical. In other words, we prove that a higher level of information-theoretic security is possible in wireless communications than anything currently known and show *how* to attain it in a completely explicit and efficient way. What exactly is meant by all of this will become clear in the forthcoming chapters.

The remainder of the thesis is as follows. Chapter 2 introduces the reader to general information theory as well as information-theoretic secrecy. This chapter also lays some of the mathematical groundwork needed to progress through the thesis. The goal of Chapter 2 is to not only give the reader the tools needed for the remainder of the thesis but also attempts to give the reader intuition for many of the concepts used. With the needed background in hand, Chapter 3 gives the three-part problem statement in more specific detail as well an outline for the solution to it. Chapters 4, 5, and 6 are the actual solutions to the problem described in Chapter 3. Finally the thesis is concluded in Chapter 7 with a discussion.

2. BACKGROUND

2.1. Notation

We shall write a^n to denote an n-dimensional vector where a_i denotes the ith component; i.e., $a^n = (a_1, \ldots, a_n)$. We use the usual notation $\|a^n\|$ to denote the Euclidean norm. We shall denote the indicator (or characteristic) function by $\mathbb{1}(x \in \mathcal{A})$ and will take all logarithms in this paper to be base 2 unless specified otherwise. With a slight abuse of notation, we will write $\mathbb{R}_+$ to denote the set of *non-negative* reals.

Basic knowledge of probability theory is assumed throughout the thesis. We will denote random variables by capital letters, a realization of that random variable by lowercase letters, and will denote the spaces for which a random variable is defined by a respective scripted letter; e.g., A is a random variable with values in $\mathcal{A}$ and a is a realization of A. We write $A \perp B$ when random variable A is independent of B and write $\mathbb{E}[A]$ to denote the expected value of random variable A.

We denote all probability densities by $\omega(\cdot)$ unless otherwise noted and we define the conditional probability density in an analogous way and denote it by $\omega(\cdot|\cdot)$ where each is taken with respect to the random variable corresponding to the obvious choice; e.g., $\omega(b|a)$ will denote the conditional probability density of the random variable B given $A = a$.

2.2. Information Theory

Originally proposed and formulated by Claude Shannon in his seminal 1948 paper "A Mathematical Theory of Communication" [20], the field of information theory explores the fundamental limits of data storage and the communication of information. All of information theory can be thought of as answering two fundamental questions: What is the ultimate data compression and what is the ultimate transmission rate of communication [7]. In this paper, we do not concern ourselves with the question of data storage or compression and solely focus on the latter question. Although these two questions seem to be only related to communication theory, information theory plays an intricate role in many fields such as computer science, physics, and probability and statistics.

Abstractly, *information* can be thought of as a reduction in uncertainty of an event or random variable. More formally, information is typically measured by *entropy* which quantifies the

amount of uncertainty of a random variable. As an intuitive example, communicating the outcome of the flip of a fair coin toss gives less information and therefore has a lower entropy than that of a roll of a fair die since there was more uncertainty in the outcome of the latter case.

Definition. *The entropy H(X) of a discrete random variable X with alphabet $\mathcal{X}$ is defined by*

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

where $p(x)$ is the probability mass function of X and $H(X)$ is given in bits when the log is taken to be base two or given in nats when the log is taken to be the natural logarithm.

From this definition, we can immediately notice that deterministic events (X follows a deterministic distribution with $p(x) = 1$) have zero entropy, $H(X) = 0$, and thus communicating the realization of the random variable X carries no information. Consider now the example of the roll of a fair six sided die. Let X be the random variable representing the outcome of the roll. Then X takes any value from the set $\{1, 2, 3, 4, 5, 6\}$ with equal probability of $\frac{1}{6}$. The entropy is then calculated to be:

$$H(X) = -\sum_{x} \frac{1}{6} \log \frac{1}{6} = -\log \frac{1}{6} \approx 2.58 \text{ bits.}$$

Now suppose we wish to determine the output of the fair roll via a series of yes-no questions. A good first question to ask would be "Was the outcome an even number?" The next question could be "Was the outcome strictly smaller than four?" Depending on the actual outcome, after asking these questions we will have determined the actual value or we will have at most one more remaining question that needs to be asked. In other words, to determine the outcome we need to ask either two or three yes-no questions. By examining the value we calculated for $H(X)$, we can see that it roughly describes how many yes-no questions one needs to ask to determine the realization of X. This is not coincidental to this example; this example was meant to illustrate that another way to think of entropy (when measured in bits) is to think of it as giving the minimum expected number of binary questions required to determine the value of X. It turns out that the minimum expected number of binary questions needed to determine X is indeed between $H(X)$ and $H(X) + 1$ [7].

2.2.1. Mutual Information

An fundamental quantity necessary for the study of the rate of communication is that of *mutual information.*

Definition. *Let X and Y be random variables with joint probability mass function $p(x, y)$ and marginals $p(x)$ and $p(y)$ respectively. The mutual information $I(X \wedge Y)$ is given by:*

$$I(X \wedge Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

Mutual information can be thought of as answering the question of how much information one random variable carries about another. Phrasing it another way, it is a measure of how much the uncertainty of X is reduced by having knowledge of Y. It is easily seen that if X and Y are independent random variables then $I(X \wedge Y) = 0$ since knowing Y does not help to determine X in any way. In the other extreme case, if X is a deterministic function of Y, then knowledge of Y will surely determine X and therefore the mutual information is just the entropy of X, i.e., $I(X \wedge Y) = H(X)$.

The two previous definitions for discrete entropy and discrete mutual information also have analogs for the case of continuous random variables.

Definition. *Let X be a continuous random variable with probability density $\omega(x)$ with support $\mathcal{S}$. The differential entropy $h(X)$ is defined as*

$$h(X) = -\int_{\mathcal{S}} \omega(X) \log \omega(X) dx$$

Definition. *Let X and Y be two continuous random variables with joint probability density $\omega(x, y)$ and marginals $\omega(x)$ and $\omega(y)$ respectively. The mutual information $I(X \wedge Y)$ is defined as*

$$I(X \wedge Y) = \int_{\mathcal{S}} \omega(x, y) \log \frac{\omega(x, y)}{\omega(x)\omega(y)} dxdy$$

where $\mathcal{S}$ is the support of the random variables.

2.2.2. Capacity

As mentioned above, one of the goals of information theory is to find the optimal rate of communication from a transmitter A to a receiver B. Before addressing that goal, we need to first understand what it means for A to *communicate* with B. Communication can be defined as the process by which a desired physical state is induced at B due to some physical operation of A. Communication happens in our physical world and is therefore subject to external noise (in the form of electromagnetic interference, movement, physical objects blocking the line of sight from A to B, etc.) and flawed physical signaling by the parties themselves. We say a communication process was successful if the receiving party B and the transmitting party A agree on what physical state was meant to be induced at B. Informally, we can think of a communication channel as the physical path a signal takes from A to B and we say that A is using a communication channel to induce the desired physical state at B. Each time this is done constitutes one *use* of the channel.

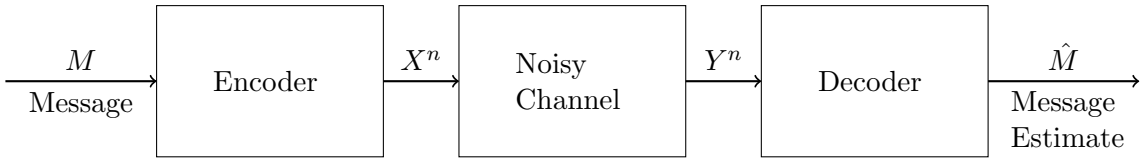

Figure 2.1. General communication system.

In more detail, as illustrated in Figure 2.1, a transmitter takes a message M from some finite set $\mathcal{M}$ and maps it to a sequence of n channel symbols X^n known as a *codeword*. These n channel symbols are then sent through the channel and are possibly altered according to the channel's noise and other characteristics to create an output sequence Y^n. The receiver then attempts to recover the transmitted message from that output. If one is not careful when choosing the sequence of channel symbols X^n, it is possible that due to the noise of the channel, two different input sequences, X_1^n and X_2^n (coming from different messages) may produce the same output sequence Y^n which would then only map to one message estimate. In this case, at least one of the messages was transmitted in error due to X_1^n and X_2^n being confusable. Information theory gives the tools to choose a subset of input sequences X^n for a channel such that with high probability, there is only one input sequence that could have produced a given output sequence. By doing this, even in the presence of noise,

we can recover the original input sequence and therefore the original message at the receiver with low probability of error.

It has been shown that the number of distinguishable messages able to be communicated successfully via n uses of the communication channel grows exponentially with n. In other words, suppose the size of the set of possible messages is α. Then α grows exponentially as n grows. The logarithm of the number of distinguishable messages divided by n is known as the *channel capacity*. Arguably the most prominent result of information theory is that this channel capacity has been characterized as the maximum amount of mutual information between the source and output.

Definition. *A discrete channel is a system consisting of an input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ and a conditional probability mass function known as a probability transition, $p(y|x)$, which gives the probability of observing the output y given that the input x was sent. In other words, $p(y|x)$ is characterizing the noise of a channel. The channel is said to be memoryless if the probability distribution of the output depends only on the current input and is independent of all others.*

Definition. *The channel capacity of a discrete memoryless channel is defined as:*

$$C = \max_{p(x)} I(X \wedge Y)$$

where the maximum is taken over all possible input distributions $p(x)$.

Intuitively, the mutual information between X and Y, $I(X \wedge Y)$, expresses how much information the output Y of a channel contains about the input X. Thus, for accurate communication, we desire that the output contains a large amount or all of the information about the input and thus would have a large value. The capacity as given above is in bits of information per channel use, and is known as the "information" channel capacity. There is also an operational definition of channel capacity which express the highest rate in bits per channel use where information can be transmitted with arbitrarily low probability of error. In his seminal paper, Shannon showed that these two definition are equivalent and thus we make no distinction between the two.

As one can see above, capacity is finding the upper limit of the mutual information between the input and output of a channel. Obviously, *how* we distribute X does not change the capacity

and since the only other factor affecting the output is the noise of the channel, we see that capacity is an intrinsic *physical* property of a communication channel.

Many of the above concepts were in terms of the discrete memoryless channel as a means to help garner intuition about the concepts. We now move into the continuous realm and begin to make rigorous many of the concepts described above.

2.3. Channel

Definition. *We define a channel as a stochastic mapping*

$$T : \mathcal{A} \to \mathcal{B}.$$

If A is a random variable on $\mathcal{A}$ then we let B be the random variable $B = T(A)$ on $\mathcal{B}$. We then associate the transition density $\omega(b|a)$ with T as a characterization of how the stochastic mapping is taking place. That is, given that $A = a$ was sent across the channel, the probability that B is in some $\mathcal{U} \subset \mathcal{B}$ is given by

$$\int_{\mathcal{U}} \omega(b|a)db$$

where the above integral is the Lebesgue integral. Note that the channel T is completely characterized by the tuple $(\mathcal{A}, \omega(b|a), \mathcal{B})$. When the channel is to be used n times it will be denoted as T^n.

Unless specified, for the remainder of the thesis we will be considering *continuous* channels where the input and output alphabets are both uncountable. Furthermore, we will be considering *subnormalized* channels; i.e., channels with transition densities such that

$$\int_{\mathcal{B}} \omega(b|a)db \leq 1.$$

For a subset $\mathcal{T} \subset \mathcal{A} \times \mathcal{B}$, define a *restricted* conditional density by

$$\omega_{\mathcal{T}}(b|a) = \begin{cases} \omega(b|a), & (a,b) \in \mathcal{T} \\ 0, & \text{Otherwise} \end{cases}.$$

This induces a new channel $T_{\mathcal{T}} : \mathcal{A} \to \mathcal{B}$ with transition density $\omega_{\mathcal{T}}(b|a)$. That is, given that $A = a$ was sent across channel $T_{\mathcal{T}}$, the probability that B is in some $\mathcal{U} \subset \mathcal{B}$ is given by

$$\int_{\mathcal{U}} \omega_{\mathcal{T}}(b|a)db = \int_{\mathcal{U}} \omega(b|a)\mathbb{1}\left((a,b) \in \mathcal{T}\right) db.$$

2.3.1. Gaussian Channel

The most common continuous input alphabet channel is the Additive White Gaussian Noise (AWGN) channel, or just the Gaussian channel given in Figure 2.2. This models real-world communication channels such as wired telephone, radio channels, and satellite links. For this thesis, we will only consider *discrete-time* channels where there is one input to the channel at each time instant. We let the input at time i be denoted by $X_i \in \mathbb{R}$ and the noise at time i be denoted by U_i and take the sum of the two to be the output Y_i:

$$Y_i = X_i + U_i.$$

We assume the noise component U_i is drawn i.i.d from a Gaussian distribution having zero mean and variance σ^2 and is independent of the input signal X_i.

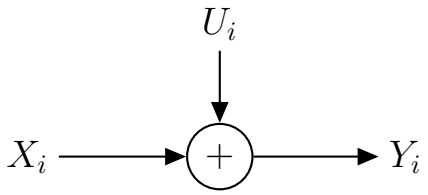

Figure 2.2. Gaussian channel.

Fact 1. *[7] The capacity of a Gaussian channel with input satisfying* $\mathbb{E}[X_i^2] \leq P$ *and noise variance* σ^2 *is*

$$C = \frac{1}{2}\log\left(1 + \frac{P}{\sigma^2}\right) \textit{ bits per transmission.}$$

2.3.2. Fading Channel

The general channel model used to model wireless communication environments is that of the *fading channel*, where the output signal is an attenuation of the input signal layered with additive white Gaussian noise. The attenuation, input, and noise are represented using the complex random variables H, X, and U respectively. The output of this channel at time i is then given as

$$Y_i = H_i X_i + U_i$$

where $X_i \in \mathbb{C}$, $H_i \in \mathbb{C}$, and $U_i \sim \mathcal{CN}(0, \sigma^2)$ and is illustrated in Figure 2.3. Here, $\mathcal{CN}(0, \sigma^2)$ is a *circularly-symmetric normal distribution* with 0 mean and variance σ^2. We shall refer to the random variable representing attenuation, H, as the *channel coefficient*.

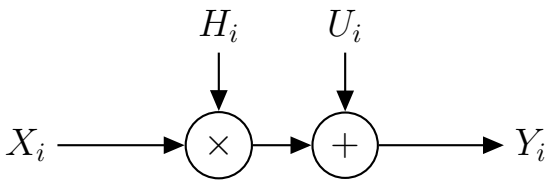

Figure 2.3. Fading channel.

For the purposes of this paper, we will be considering what is known as the *fast fading channel*. This is a channel model where the fading coefficient is sampled i.i.d. for each use of the channel and is commonly used to model wireless communications (cf. [24]).

2.4. Error Correcting Codes

In order to transmit information with a low probability of error, proper encoding of the source symbols is required so as not to confuse two inputs at the output. For the time being, we assume for ease that both the message alphabet and the channel symbol alphabet are binary alphabets and thus we are operating on a discrete channel for the moment. Error control coding is the process of encoding a k-bit message into an n-bit codeword to be sent over the channel where $k \leq n$. The set of all codewords constitutes a *codebook* and this codebook is assumed to be known at both the transmitter and receiver. Redundancy is purposely added to the message so that errors in the codeword during transmission do not hinder the receiver from deducing which message was

sent from the received codeword. For example, one of the simplest error correction codes (ECC) is repetition coding, where one bit is transformed into multiple copies of itself and that sequence is then sent across the channel. Suppose we wished to send the sequence

1011

across the channel as our message. Using (3) repetition coding we encode that sequence into

111000111111

where each bit in the message is repeated three times. Using this scheme, we can use majority decoding to deduce the original message. In other words, we observe three bits at a time and take which value occurs most often as the corresponding bit of the message. If the first three bits were received as 110 instead of 111, this would still be mapped to a 1 due to 1 being the majority. In this way, we can see that this code can always correct one error per message bit. If there happen to be two or three errors within a group of three bits, this scheme will fail and will decode the received codeword incorrectly.

One can say that the goal of error control coding is that of "spacing" codewords "far enough" apart from one another in order to distinguish one from another after they have been disturbed by the channel. The metric used to determine the distance between two codewords and what constitutes "far enough" is specific to each devised ECC and also by what values are contained in the input alphabet. In the above example, distance would be defined by Hamming distance which is the number of bits in which the received vector differs from one of the possible input vectors. The codeword giving the smallest Hamming distance from the received vector is taken to be the input codeword which was sent. In the case where the input and output alphabets are continuous, Euclidean distance is often the metric used to determine distance between codewords.

The ratio $R = k/n$ is known as the code rate, or just rate, of the ECC and n is referred to as the *block length* of the code. The rate can be interpreted as the number of information bits being transferred over the channel per channel use. In the above example, the rate is $4/12 = 1/3$. In Shannon's original paper he showed that for any rate R which is smaller than or equal to the

capacity of the channel being operated on, there *will* exist an ECC that such that the probability of error at the output can be made arbitrarily small with sufficiently large block length. This result is known as Shannon's *noisy channel coding theorem.* Furthermore, if one uses an ECC that is operating at a rate *above* capacity, the probability of error at the receiver will be bounded away from 0 as the block length goes to infinity. We now make all these concepts rigorous.

Definition. *[7] Suppose we are given a channel $T : \mathcal{X} \to \mathcal{Y}$ for any alphabets $\mathcal{X}, \mathcal{Y}$. We call $\mathcal{C}_n$ an error correction code (ECC) for T of length n if it consists of:*

1. *A finite index set $\mathcal{M}$.*
2. *An encoder $e_n : \mathcal{M} \to \mathcal{X}^n$ (we assume e_n is injective).*
3. *A decoder $d_n : \mathcal{Y}^n \to \mathcal{M}$.*

The rate of an ECC $\mathcal{C}_n$ is defined to be

$$R_{\mathcal{C}_n} = \frac{\log(|\mathcal{M}|)}{n}.$$

We define a coding scheme for T as a set of codes

$$\mathcal{C} = \{\mathcal{C}_n\}_{n\in\mathbb{N}}$$

and the rate of the coding scheme $\mathcal{C}$ to be

$$R_{\mathcal{C}} = \lim_{n\to\infty} R_{\mathcal{C}_n}$$

when this exists. The image of the encoder e_n will be called the codebook and shall be denoted by

$$\mathscr{C}_n = \{x^n \in \mathcal{X}^n \,|\, x^n = e_n(m)\,,\, m \in \mathcal{M}\}.$$

We say the encoder e_n has power constraint P if:

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P \qquad \forall x^n \in \mathscr{C}_n.$$

For a coding scheme to be useful in transmitting error free data across communication channels, we need to guarantee that it is *reliable*.

Definition. *Let $M \in \mathcal{M}$ be the message input to the encoder, $e_n(M) = X^n$ and take Y^n to be the output of the channel T^n given that X^n was sent. We define the probability of error of an ECC as*

$$\mathbb{P}_e(\mathcal{C}_n) = \mathbb{P}[M \neq d_n(Y^n)].$$

We say a coding scheme $\mathcal{C} = \{\mathcal{C}_n\}$ is reliable if

$$\lim_{n\to\infty} \mathbb{P}_e(\mathcal{C}_n) = 0.$$

Finally, we need one more definition to characterize at what rates we can successfully communicate across the channel.

Definition. *We say that rate $R_{\mathcal{C}}$ is achievable across the channel T if there exists a reliable coding scheme $\mathcal{C} = \{\mathcal{C}_n\}$ such that $R_{\mathcal{C}_n} \to R_{\mathcal{C}}$ as $n \to \infty$.*

2.5. Wiretap Channel

The wiretap channel was originally posed and formulated by Wyner in 1975 [26]. Although he originally only considered what the information theory community now considers a special case of the general wiretap channel, he still put forth the concept we use today. A *wiretap channel* consists of a transmitter A, commonly referred to as Alice, a legitimate receiver B, commonly referred to as Bob, and a passive eavesdropper E, commonly referred to as Eve. Between Alice and Bob, there exists a channel commonly referred to as the main channel T; between Alice and Eve, there exists a channel commonly referred to as the eavesdropper channel A (A for *adversary*; whether A is referring to the channel or to the transmitter will be clear from context) as illustrated in Figure 2.4. For the remainder of this thesis, the output of the transmission through the channel at Bob and Eve will be denoted as Y^n and Z^n respectively. Finally, we will denote a wiretap channel by its pair of point-to-point channels $W = (T, A)$.

When operating on a wiretap channel, we employ the use of *wiretap codes* which are codes that provide two necessary utilities simultaneously - *reliability* and *security*. Reliability ensures error free communication between the transmitter and legitimate receiver in the same way that

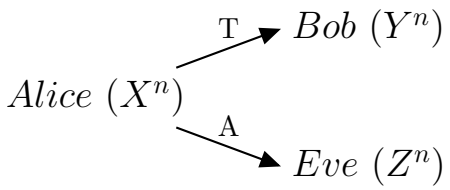

Figure 2.4. General wiretap channel.

an ECC provides reliability on a point-to-point channel. Security ensures that the message being transmitted is kept secret from the eavesdropper. The reliability of a wiretap code is measured in the same manner as that of an ECC, that being its probability of error $\mathbb{P}_e$. The security of a wiretap code is measured via various security metrics as described in the next subsection.

2.5.1. Security Metrics

To measure the amount of security provided by a wiretap code, we must first provide a metric by which that measurement is based. Shannon originally proposed what is known as *perfect secrecy* which states that given an encrypted message from a perfectly secure encryption scheme, precisely *nothing* will be revealed about the original message, i.e.,

$$I(M \wedge Z^n) = 0.$$

This is ideally what all information theorists and cryptographers would like to happen, but achieving this is often impractical in real-world scenarios as it requires large random keys such as in Shannon's one-time pad. Due to this restriction, other security metrics were proposed to offer potentially less security, but in a way that is more practical.

Table 2.1. Security metrics

Security Metric	
Weak	$\frac{1}{n} I(M \wedge Z^n), \quad M \sim \text{unif}(\mathcal{M})$
Strong	$I(M \wedge Z^n), \quad M \sim \text{unif}(\mathcal{M})$
Semantic	$\max_M I(M \wedge Z^n)$

In Wyner's original paper [26], he provided what is now known as the *weak secrecy* metric (also known as the Wyner metric). This metric measures the average information rate leaked to the eavesdropper for uniformly distributed messages and a scheme measured under this metric is said to be *weakly secure* if this rate goes to 0 as n goes to infinity, i.e.,

$$\frac{1}{n} I(M \wedge Z^n) \to 0 \text{ as } n \to \infty, \quad M \sim \text{unif}(\mathcal{M}).$$

This metric was soon shown to be too weak for practical purposes as it potentially allows an unbounded amount of information to leak to the eavesdropper over a long period of time. Another metric was soon developed known as the *strong secrecy* metric which measures the average amount of information leaked to the eavesdropper (note the difference with weak secrecy which measures the *rate*). Thus a scheme under this metric is deemed *strongly secure* if the average information leaked to the eavesdropper for uniformly distributed messages goes to 0 with n, i.e.,

$$I(M \wedge Z^n) \to 0 \text{ as } n \to \infty, \quad M \sim \text{unif}(\mathcal{M}).$$

For a long period of time, the strong security metric was held to be the gold standard for measuring information-theoretic secrecy. Recently, this was called into question by those in the cryptography community as again being too weak of a security metric, leading to the development of what is known as *semantic secrecy* (or mutual-information secrecy) [2]. Strong secrecy requires a uniform distribution of the message space, but this is not always a practical assumption for real-life applications. As such, might it be possible that a transmitter could leak more information to an eavesdropper if the message distribution were not uniform? To ensure that this does not happen, semantic security removes the assumption that the message is uniformly distributed and instead computes the largest amount of leakage over *any* message distribution. If this value tends to 0 as n goes to infinity, we say that the scheme is *semantically secure*, i.e.,

$$\max_M I(M \wedge Z^n) \to 0 \text{ as } n \to \infty.$$

Originally defined for computational based security [10], semantic security was later extended into an information-theoretic context as given above and is now held to be the gold standard

of information-theoretic security by information theorists and cryptographers alike. At its core, a semantically secure system is one where the message and the eavesdropper's output are *statistically independent* (asymptotically with block length). The careful reader may note that semantic security and mutual-information security are not technically the same and that the above definition is actually the definition for mutual-information security. However, they are equivalent as $n \to \infty$ and thus we refer to the metric given above as the semantic security metric as this matches up with a cryptographer's definition of semantic security [2, 17].

We say that under a certain metric, κ, a rate R is an *achievable secrecy rate* if there exists some coding scheme for the wiretap channel, known as a wiretap code, that satisfies both the reliability and secrecy constraints. We call the supremum of all achievable secrecy rates under security metric κ the κ *secrecy capacity* C_S. When the metric is clear from context, we will only refer to the secrecy capacity as C_S.

Fact 2. *[1] Let $(C_S)_{weak}$ denote the weak secrecy capacity of a channel and $(C_S)_{semantic}$ denote the semantic secrecy capacity of a channel. If all secure rates R_s achievable under the weak secrecy metric are also achievable under the semantic secrecy metric, then:*

$$(C_S)_{weak} = (C_S)_{semantic}.$$

2.5.2. Fast Fading Wiretap Channel

We wish to consider the case of the *fast fading wiretap channel* thus we take channels T and A to both be fast fading channels as described in Section 2.3.2 and refer to them as the main channel and eavesdropper channel respectively. More specifically, during the ith symbol of the codeword, the outputs at Bob from channel T and at Eve from channel A are given respectively as

$$Y_i = H_{m,i}X_i + U_{m,i}$$
$$Z_i = H_{e,i}X_i + U_{e,i},$$

where $U_{m,i}$ and $U_{e,i}$ are i.i.d. $\mathcal{CN}(0, \sigma_m^2)$ and $\mathcal{CN}(0, \sigma_e^2)$ additive noise, $X_i \in \mathbb{C}$ is subject to the power constraint $\mathbb{E}\left[|X|^2\right] \leq P'$, and the coefficients $H_{m,i}, H_{e,i} \in \mathbb{C}$ are also i.i.d. In the fading channel we use complex random variables to capture the fact that we have two degrees of freedom

when transmitting on a wireless channel. For technical reasons we assume that the second order moment of H_e exists; i.e., $\mathbb{E}[H_e^2] < \infty$. We note that this is not a very limiting constraint since it can be interpreted as the eavesdropper channel having an attenuation with finite energy. Apart from this, we do not assume *which* distribution the channel coefficients follow so as to remain as general as possible.

Achievability results for fading channels depend on which parties have instantaneous access to the realizations of $H_{m,i}$ and $H_{e,i}$, or rather, which parties have *full channel state information*. If a party only has access to the *statistics* of $H_{m,i}$ or $H_{e,i}$ we say that party has *statistical channel state information*.

Proposition 1. *Consider a complex fast fading channel. If the receiver has full channel state information (CSIR) then this complex channel can be decomposed into two real parallel channels.*

Proof. Without loss of generality, consider the intended receiver's channel given above and drop the index i for simplicity. Therefore, we are working with the complex fading channel $Y = H_m X + U_m$. Since $H_m \in \mathbb{C}$ we can write $H_m = |H_m|e^{i\theta}$ and thus, the receiver will receive the random variable $Y = |H_m|e^{i\theta}X + U_m$. However, since we are assuming channel state information is available at the receiver, the receiver actually knows the realization of H_m and hence knows the value $e^{i\theta}$. The receiver thus *adjusts* his output Y accordingly: $Ye^{-i\theta} = |H_m|X + U_m e^{-i\theta}$. Also, the additive white Gaussian noise is assumed to be circularly symmetric, so that $U_m e^{-i\theta}$ is actually distributed the same way as was U_m. Therefore, if we define $\tilde{Y} = Ye^{-i\theta}$ as the new output and $\tilde{U}_m = U_m e^{-i\theta}$ as the rotated noise, under the assumption of CSIR, the receiver can convert the original channel into the new channel: $\tilde{Y} = |H_m|X + \tilde{U}_m$. Now we can break up *this* channel into its real and imaginary parts:

$$\tilde{Y}_R + i\tilde{Y}_I = (|H_m|X_R + i|H_m|X_I) + \left((\tilde{U}_m)_R + i(\tilde{U}_m)_I\right).$$

Combining the real and imaginary parts respectively yields two parallel channels

$$\tilde{Y}_R = |H_m|X_R + (\tilde{U}_m)_R$$
$$\tilde{Y}_I = |H_m|X_I + (\tilde{U}_m)_I.$$

Here each output is identically given as

$$Y' = |H_m|X' + U'_m$$

where $|H_m| \in \mathbb{R}_+, X' \in \mathbb{R}$, $U'_m \sim \mathcal{N}(0, \sigma_m^2)$, and $\mathbb{E}\left[(X')^2\right] \leq P$.

□

Note that this proposition also holds for the eavesdropper's channel when the eavesdropper also has full channel state information. It is in fact very reasonable to assume that a receiver has channel state information as it can be accomplished by means of training the channel (cf. [24]).

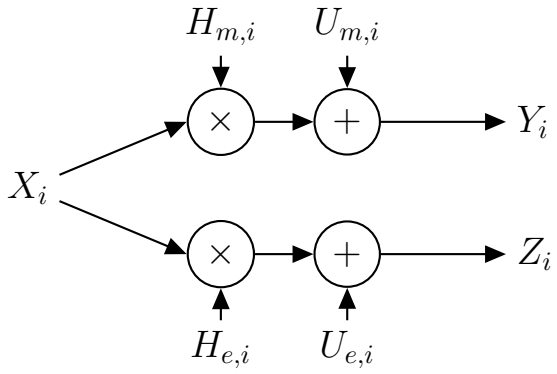

Figure 2.5. Fading wiretap channel model.

For the remainder of this thesis, we will focus on fast fading channels where both receivers have full channel state information (CSIR) about their respective channels. In particular, this means that we will only be considering the *real* fast fading channels given at time i as $Y_i = |H_{m,i}|X_i + U_{m,i}$ and $Z_i = |H_{e,i}|X_i + U_{e,i}$ due to Proposition 1. Since carrying around the modulus on the channel coefficients is cumbersome, we shall simply write H_m and H_e for the remainder of the paper where it will be clear that both are *non-negative real* random variables instead of complex as previously mentioned. Finally, we will denote the capacity of the main channel T as C_T, and that of the eavesdropper channel A as C_A. We will refer to both of these as the *point-to-point* capacity of their respective channels. An illustration of our setup is given in Figure 2.5.

We also assume that not only does the eavesdropper know the instantaneous realizations of her channel, but also the statistics of the main channel's fading coefficient. The main channel's fading statistics being public knowledge is a very plausible scenario and thus we take that extra knowledge given to eavesdropper into account. For the remainder of this thesis we also assume that the channel coefficients are not correlated, i.e., $H_{m,i} \perp H_{e,j}$ for all i, j.

Thus far, we have made no assumptions as to what information the transmitter has about the channel coefficients H_m and H_e (channel state information at the transmitter will be denoted as CSIT). As will become clear in the coming chapters, the amount of information present at the transmitter regarding the channel states plays a major role in determining an achievable rate for a fast fading wiretap channel. In particular, we will focus on three separate cases: that of *statistical CSIT* (S-CSIT) where the transmitter only has knowledge of the main channel and eavesdropper channel statistics. That which we will refer to as *partial CSIT* where the transmitter has knowledge of the main channel's instantaneous realizations of H_m at each time i but no knowledge of the eavesdropper's instantaneous channel coefficient - only its statistics. Finally, that of *full CSIT* where the transmitter has knowledge of both the main channel's and eavesdropper channel's instantaneous realizations of H_m and H_e respectively.

2.6. Further Mathematical Background

This section introduces some of the more complicated or important mathematical tools which will be used throughout the proofs in the remainder of the paper. These have no dependence on the above concepts and stand independently. They are presented here for the completeness and ease of the reader.

2.6.1. Hoeffding Bounds

Developed in 1963 by Wassily Hoeffding, Hoeffding's inequality gives an upper bound on the probability that the sum of independent random variables diverges from the expected value. In [12], Hoeffding first gives the bound for when the random variables are bounded between 0 and 1. He then generalizes this result for when the random variables are bounded by arbitrary finite bounds, which is the one we will focus on here. More clearly, take $X_1, \ldots, X_i, \ldots, X_n$ to be independent random variables bounded by $a_i \leq X_i \leq b_i$ where $a_i, b_i < \infty$. Define the empirical

mean of the random variables by

$$\bar{X} = \frac{1}{n}\left(X_1 + \cdots + X_n\right).$$

Then [12, Theorem 2] gives us that:

$$\mathbb{P}\left(\bar{X} - \mathbb{E}\left[\bar{X}\right] \geq t\right) \leq \exp\left(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

for $t > 0$, where $\mathbb{E}[\bar{X}]$ is the expected value of $\bar{X}$. More simply, this bound is showing the probability by which the empirical mean deviates from its expected value. It is generalized further to the case of *unbounded* random variables in Lemma 2 which is the bound that will be used in the proof of Lemma 3.

2.6.2. Fréchet Inequalities

Originally explicitly expressed by Maurice Fréchet [8], the Fréchet inequalities (also known as the Boole-Fréchet inequalities due to the fact that Boole's work inherently contained these bounds) give bounds for calculating probabilities of logical propositions with no need to assume independence or dependence of the propositions or events in question. Usually given as inequalities for logical conjunctions and disjunctions for logical propositions, the forms we will be using in this paper are that of set intersection and Cartesian products of sets. Let A and B be subsets of a set U and take $\mathbb{P}(A)$ to mean the probability that some element, $x \in U$ is in set A, i.e., $\mathbb{P}[x \in A]$. The Fréchet inequalities are then given as:

$$\text{Fréchet Inequality for intersections: } \max(0, \mathbb{P}(A) + \mathbb{P}(B) - 1) \leq \mathbb{P}(A \cap B)$$

$$\text{Fréchet Inequality for Cartesian products: } \max(0, \mathbb{P}(A) + \mathbb{P}(B) - 1) \leq \mathbb{P}(A \times B)$$

Notice that these can be used repeatedly in order to bound probabilities involving more than two sets (or events).

2.6.3. Typical Sets

Often times while working with random variables, we do not wish to consider the whole space upon which a random variable can take values; rather, we are more (or only) concerned with

the realizations of that random variable which have a sufficiently high chance of happening. Only focusing on the sets of realizations which have a sufficiently high probability of occurring greatly simplifies mathematical calculations without sacrificing much accuracy. This leads to our definition of a *typical set*.

Definition. *Let $\mathcal{A}$ be the support of random variable A. For $\epsilon \geq 0$, we call a subset $\mathcal{T} \subset \mathcal{A}$ a $(1-\epsilon)$-typical set if*

$$\mathbb{P}\left[A \in \mathcal{T}\right] \geq 1 - \epsilon.$$

Typical sets intuitively contain almost all that there is to know about our space up to some ϵ. We will sometimes refer to ϵ as the *threshold probability*.

3. PROBLEM STATEMENT

3.1. Literature Review

Information-theoretic security for fading channels is a relatively new field, arguably starting in 2006 when Gopala et al. [11] determined the secrecy capacity of the *slow fading* channel with full CSIT at the transmitter. Slow fading channels are fading channels where the channel coefficients vary at asymptotically long time intervals, rather than changing with each time instant as is the case in fast fading. In other words, on a slow fading channel, the channel is constant for long periods of time before changing and staying constant again. This result was found using the *weak* secrecy metric. In the same work, the weak secrecy capacity was also found for the case of slow fading channels with partial CSIT. This paper was the first to show that fading is actually *beneficial* to secrecy as it allows a transmitter to take advantage of additional randomness that is not present in the case of a Gaussian channel. A year later, Liang, Poor, and Shamai upgraded this result in the sense that they no longer considered slow fading channels but rather that of fast fading channels as described above. They found the weak secrecy capacity of the fast fading wiretap channel but only with the assumption of full CSIT [15]. This was again later improved by Bloch and Laneman in [5] where they determined the secrecy capacity of this channel under strong secrecy; however, they did not provide an explicit means of doing so.

The field was further extended again by Bloch and Laneman to the case of fast fading channels with partial CSIT, where an achievable secrecy rate was given under a secrecy constraint that is stronger than weak secrecy, yet weaker than strong secrecy known as *variational distance* [3]. Their solution relies on an optimization problem that has no closed form solution and thus it represents the *best known* secrecy rate on the fast fading channel with partial CSIT although further work needs to be done to determine if this actually represents the secrecy capacity. In the case of fast fading channels with statistical CSIT, it was only recently shown in [16, 19] that positive rates are actually achievable and an upper bound for the secrecy capacity is also derived. For a special class of fast fading statistical CSIT channels, [16] actually finds the secrecy capacity of these channels under the weak secrecy constraint. A summary of where the current state of the art stands for fast fading channels is given in Table 3.1.

Table 3.1. Current state of the art

Current State of the Fast Fading Wiretap Channel			
CSIT	Secrecy Capacity	Security Metric	Explicit
Full	Yes	Strong	No
Partial	No	Variational Distance	No
Statistical	No	Weak	No

3.2. Motivation

As seen in the previous subsection, the current results for the fast fading wiretap channel are insufficient for security in real world systems. Furthermore, few of the results include practical ways to actually achieve said security. For information-theoretic security to be used in wireless communications, it is necessary to improve the security metric in every case above to that of semantic, as well as give an explicit method to actually achieve that security in each case. This thesis intends to do just that.

3.3. Problem Statement

We wish to find a scheme that will provide *semantically secure* achievable rates on the fast fading wiretap channel in the cases of full CSIT, partial CSIT, and S-CSIT. Furthermore, we wish to do so in a way that is explicit and efficient so as to be implementable in practice.

In [13], a procedure was provided based on [25] which converts the problem of finding a semantically secure wiretap code into the problem of just finding an ECC for the main channel of the wiretap in question. We omit the details of why this procedure works as it is out of the scope of this thesis. The procedure given is general enough to be applied to any channel, but this thesis is only focused on fast fading channels and thus we will apply it in that setting. The first step in the procedure is calculating a parameter of the wiretap channel known as ξ.

3.4. Max-information

The parameter ξ mentioned above is directly related to a term known as "max-information." Intuitively, max-information measures the maximum amount of "information" that can be sent over the channel using a specific channel code. Thinking of it this way is *only* for intuitive purposes as it does not necessarily coincide with mutual information mathematically. Characterizing the relationship between these two quantities is an interesting future line of work.

Definition. *Let H_m and H_e be the random variables with alphabet $\mathcal{H}$ corresponding to the channel coefficients of the main channel and eavesdropper channel respectively. With respect to an n length ECC with codebook $\mathscr{C}_n$ used over channel T^n, we define max-information by*

$$\boldsymbol{\mathcal{I}}_n = \log\left(\mathbb{E}_{H_m^n H_e^n} \int_{\mathcal{Z}^n} \max_{x^n \in \mathscr{C}_n} \omega(z^n | x^n, H_m^n, H_e^n) dz^n\right).$$

In general, $\omega(z^n|x^n, h_m^n, h_e^n)$ finds the *relative likelihood* that an output z^n occurs given that $x^n \in \mathscr{C}_n$ and channel coefficients $h_m^n, h_e^n \in \mathcal{H}^n$ occurred. Therefore $\max_{x^n \in \mathscr{C}_n} \omega(z^n|x^n, h_m^n, h_e^n)$ measures the *highest likelihood* a particular eavesdropper output z^n could have over all codewords with channel coefficients h_m^n and h_e^n. Integrating with respect to z^n converts this likelihood into a conditional "*probability*"; however, it is not a true probability due to the fact that it is taking only the highest likelihood at each point. Taking the expected value with respect to both H_m and H_e normalizes the "probability" with respect to how likely these channel coefficients were to happen. Taking the log further normalizes the "probability" between 0 and $\log|\mathscr{C}_n|$.

Max-information is thus concerned with the space of events $\mathscr{C}_n \times \mathcal{H}^n \times \mathcal{H}^n \times \mathcal{Z}^n$; however, as mentioned in Section 2.6.3, it is intractable to work with this whole space and thus we wish to only consider the space of tuples that have a sufficiently high probability of occurring. In other words, we want to consider input-output pairs such that an output has a sufficiently high probability of occurring with a given input. With this motivation we restate our typical set definition as follows:

Definition. *For $\epsilon \geq 0$, we call a subset $\mathcal{T} \subset \mathscr{C}_n \times \mathcal{H}^n \times \mathcal{H}^n \times \mathcal{Z}^n$ a $(1-\epsilon)$-typical set if*

$$\mathbb{P}\left[(X^n, H_m^n, H_e^n, Z^n) \in \mathcal{T} \,|\, X^n = x^n\right] \geq 1 - \epsilon, \qquad \forall x^n \in \mathscr{C}_n.$$

Furthermore, we will denote the set of all $(1-\epsilon)$-typical sets by $\mathscr{T}_\epsilon$.

Using $(1-\epsilon)$-typical sets, we can define another max-information over this reduced space that will be crucial to our proofs later on.

Definition. *Given $\epsilon \geq 0$ and $\mathcal{T} \in \mathscr{T}_\epsilon$, consider*

$$\boldsymbol{\mathcal{I}}_n(A_{\mathcal{T}}^n) = \log\left(\mathbb{E}_{H_m^n H_e^n} \int_{\mathcal{Z}^n} \max_{x^n \in \mathscr{C}_n} \omega_{\mathcal{T}}(z^n | x^n, H_m^n, H_e^n) dz^n\right).$$

We define ϵ-smooth max-information by

$$\boldsymbol{\mathcal{I}}_n^{\epsilon} = \inf_{\mathcal{T} \in \mathscr{T}_{\epsilon}} \boldsymbol{\mathcal{I}}_n(A_{\mathcal{T}}^n).$$

That is, given some threshold ϵ, we find the smallest value that max-information could possibly be when defined on the subnormalized channels corresponding to those sets that contain *enough* probability with respect to our threshold. In this paper we will only be concerned with ϵ as a function of n and will mainly be concerned with the cases for which $\epsilon \to 0$ as $n \to \infty$. We now give a lemma due to [13] which gives the relationship between ξ and max-information.

Lemma 1. *[13] Suppose that an asymptotic upper bound ξ can be found:*

$$\lim_{\substack{n\to\infty \\ \epsilon\to 0}} \left(\frac{\boldsymbol{\mathcal{I}}_n^{\epsilon}}{n} \right) \leq \xi,$$

that holds for any sequence of codes $\{\mathcal{C}_n\}$ each with rate $R_{\mathcal{C}_n}$ having asymptotic rate $R_{\mathcal{C}} \leq C_T$. It is then possible to achieve an asymptotic overall transmission rate R_s under semantic security so long as

$$R_s < R_{\mathcal{C}} - \xi.$$

In particular, suppose $\xi = C_A$ and $R_{\mathcal{C}} = C_T$; then, it is possible to achieve an overall secrecy rate $R = C_T - C_A$ under semantic secrecy.

The preceding lemma gives a set of achievable rates under *semantic* security as a function of the error correcting code rate so long as we can find a proper asymptotic upper bound on the max-information per channel use. In order to achieve a positive secrecy rate, it is necessary to find an error correcting coding scheme such that $R_{\mathcal{C}} > \xi$. Step two of the procedure is exactly that, one must next find an error correcting coding scheme which operates successfully over the point-to-point main channel and satisfies the previous inequality. By Shannon's noisy channel coding theorem, we know a coding scheme does exist which will satisfy that inequality so long as $C_T > \xi$ and $C_T \geq R_{\mathcal{C}}$.

3.5. Transmission Scheme

With an ECC scheme in hand which satisfies the inequality of step 2, the last step of the procedure is to concatenate a *Universal Hash Family (UHF)* based "preprocessing" scheme with that ECC to produce the full coding scheme for the wiretap channel. Therefore, the wiretap coding scheme of [13] is a modular scheme that concatenates a preprocessing layer (to provide security) with *any* error correcting code (to provide reliability). Since no constraints have been imposed on the ECC apart from it needing to have a large enough rate, this preprocessing scheme can be added on to any existing scheme non-invasively. For completeness of the reader, we give a description of the preprocessing scheme next.

Let $\mathcal{M} = \{0,1\}^k$ and $\mathcal{M}' = \{0,1\}^l$ be the sets of all k-length and l-length bit strings respectively, where $k < l < \infty$. We will refer to these finite sets as the message set and the pseudo-message set respectively, and an element $M \in \mathcal{M}$ as the *actual message* and an element $M' \in \mathcal{M}'$ as the *pseudo-message*. The actual message is the information the transmitter wishes to transmit to the legitimate receiver, and the pseudo-message is a randomly chosen variation of the actual message. Also, let $\{0,1\}^l$ correspond to elements of the Galois Field $GF(2^l)$ and define $\mathcal{S}$ as $\mathcal{S} = \{0,1\}^l \setminus 0^l \times \{0,1\}^l$, where 0^l represents the all 0 bit string.

A UHF, first given in [6], is a tool generally found in computer science applications and is defined as a family of functions $\mathcal{F} = \{f \mid f : \mathcal{M}' \to \mathcal{M}\}$ such that for all $m' \neq m''$ it follows that

$$\frac{1}{|\mathcal{F}|}|\{f \in \mathcal{F} \mid f(m') = f(m'')\}| \leq \frac{1}{2^k}.$$

It is shown in [13] that any UHF satisfying certain additional properties will work as a preprocessing layer. These properties are omitted here for brevity and also they are not very enlightening for the purposes of this thesis as they are mostly mathematical technicalities. For ease, there is also given an explicit instance of a UHF with quadratic time complexity that meets the required properties:

$$\mathcal{F}' = \{f_{s,t}(m') = (s \odot m' + t)_k \mid (s,t) \in \mathcal{S}\}$$

where $a \odot b$ and $a + b$ are multiplication and addition defined on the finite field $GF(2^l)$ respectively and the operator $(\cdot)_k$ selects the k most significant bits. For the remainder of the paper, we will

employ this as our UHF, but note that the results of this paper hold for *any* UHF satisfying the requirements given in [13].

To start a transmission, the transmitter first uniformly chooses random seeds $(s,t) \in \mathcal{S}$ and shares them publicly. The seeds themselves are not transmitting any secret information, they are just required to set up the scheme. It is shown in [13] that sending these across the channel does not affect the rate of transmission. Next, the transmitter encodes a chosen message $m \in \mathcal{M}$ via the inverse universal hash function $f_{s,t}(m')^{-1}$. Using this specific UHF, this is given as $\phi : \mathcal{M} \to \mathcal{M}'$, defined by

$$\phi_{s,t}(m,R) = s^{-1} \odot ((m,R) - t) = m'$$

where R is a randomly chosen $l-k$ bit string, (m,R) is the concatenation of the bit strings m and R, s^{-1} is the inverse of s in $GF(2^l)$, and once again all multiplication and addition is done in $GF(2^l)$. This l-length pseudo-message m' is then passed to the encoder e_n of the error correcting code to be coded into an n-length channel code x^n which is then sent over the channels.

On the receiving end, the legitimate receiver will "undo" this process. He will receive an output vector y^n which is an altered version of x^n due to the channel attenuation and additive noise. y^n is passed to the decoding function d_n to produce an estimate $\hat{m}'$ of the pseudo-message. Recall that the error correcting code is chosen to operate successfully over the point-to-point channel from the transmitter to legitimate receiver and therefore the estimate $\hat{m}'$ does indeed equal m' with high probability. Finally, the receiver undoes the preprocessing layer by inputting the pseudo-message estimate into the UHF function $f_{s,t}$ to produce a message estimate $\hat{m}$. If, in fact, $\hat{m}' = m'$ then the UHF guarantees $\hat{m} = m$. This scheme is summarized in the following tables as well as illustrated in Figure 3.1.

Table 3.2. Transmission procedure

Transmission Procedure
Seeds $(s,t) \in \mathcal{S}$ are chosen uniformly and publicly shared.
Choose actual message $m \in \mathcal{M}$.
Generate pseudo-message $m' = s^{-1} \odot ((m,R) - t)$.
Generate channel codeword $x^n = e_n(m')$ and send over channel.

Table 3.3. Receiving procedure

Receiving Procedure
Receive channel codeword y^n.
Decode channel codeword into an estimate of the pseudo-message $\hat{m}' = d_n(m')$.
Obtain estimate of the actual message $\hat{m} = (s \odot m' + t)_k$.

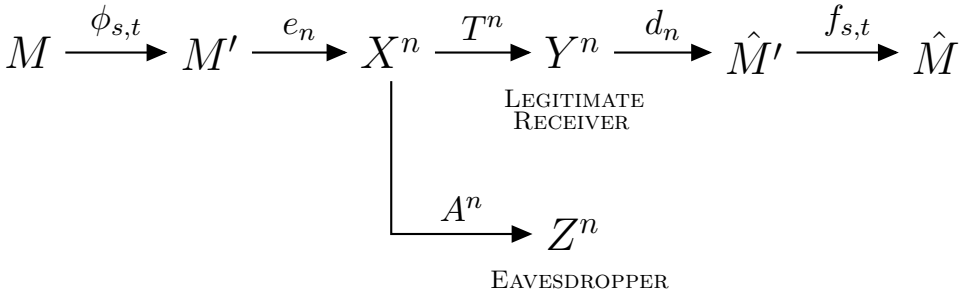

Figure 3.1. Transmission scheme.

Let's now consider what happens at the eavesdropper. Suppose we chose an ECC with a rate that was *above* the point-to-point capacity of the eavesdropper channel as the ECC of this scheme. Then we know that z^n will be decoded in error by the eavesdropper with probability bounded away from 0. In other words, let $\tilde{m}' = d_n(z^n)$ be the estimate of the pseudo-message at the eavesdropper. Then $\tilde{m}'$ will be in error with high probability and there is no possible way to rectify this as per Shannon's capacity results. The eavesdropper now must decide how to make an estimate of m from $\tilde{m}'$. The UHF has the property that passing an erroneous version of $\tilde{m}'$ into $f_{s,t}$ will yield the correct message m with a probability smaller than a uniform probability over the message space. Put another way, the eavesdropper has a *higher* probability of choosing the sent message by choosing a message uniformly from the message space *a priori* than to use the UHF. The eavesdropper is certainly not limited to only these two methods for decoding; [13] shows that no matter how she does her decoding the leakage information will still go to 0. Note that this argument relies on the fact that the transmitter was transmitting higher than the capacity of the eavesdropper channel. Thus when the main channel capacity is higher than that of the eavesdropper we can always ensure secrecy. If the eavesdropper channel capacity is *always* larger than that of the

main channel capacity, the secrecy rate will indeed be zero. However, if the eavesdropper channel capacity is just larger than that of the main channel *on the average*, a positive secrecy rate can still be achieved. What rates are actually possible given a fast fading wiretap channel are explicitly given in Chapters 4,5, and 6.

3.6. Discussion

As a summary, the procedure given in [13] to achieve positive semantically secure rates on the fast fading wiretap channel is as follows:

1. Find an asymptotic upper bound ξ to $\frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n}$ for the wiretap channel in question.
2. Find an error correcting scheme $\mathcal{C}$ of rate $R_\mathcal{C}$ such that $R_\mathcal{C} > \xi$.
3. Concatenate a UHF based preprocessing scheme as outlined above with $\mathcal{C}$.

In other words, given any wiretap channel, all one needs to do is find an asymptotic upper bound to $\frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n}$ for the wiretap channel in question to guarantee semantically secure achievable rates using the above transmission scheme so long as $R_s < R_\mathcal{C} - \xi$. The remainder of the paper is dedicated to doing just that. Chapter 4 finds an asymptotic upper bound to $\frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n}$ for the S-CSIT fast fading wiretap channel, Chapter 5 does so for the partial CSIT fast fading wiretap channel, and Chapter 6 does so for the case of full CSIT. With these bounds, we have successfully converted the problem of finding a wiretap code into that of just finding an ECC for the main channel in each of these three cases. Design of error correction codes is already a major subfield of communication theory and thus we have brought the problem of security for the fast fading wiretap channel into the world of coding theory.

4. S-CSIT

The case of S-CSIT, where the transmitter only knows the channel statistics of both the main and eavesdropper channels, is arguably the most realistic scenario of a modern wireless communication environment. It requires no special real-time feedback implementation for the main channel to give the instantaneous channel states and can assume that the eavesdropper is purely a malicious party (although still passive). Under this assumption, in this chapter we give a set of *semantically secure* achievable rates for the fast fading wiretap channel. To do so, we find an asymptotic upper bound, ξ, to $\frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n}$ for any choice of code so as to use Lemma 1 which will provide the achievable rates. In particular, we will be focused on $\xi = C_A$, where C_A denotes the point-to-point channel capacity of the eavesdropper's channel.

We start by first simplifying the expression for max-information in the case of S-CSIT.

Proposition 2. *On the S-CSIT real fast fading channel, max-information can be simplified as*

$$\boldsymbol{\mathcal{I}}_n(A_{\mathcal{T}}^n) = \log\left(\mathbb{E}_{H_e^n}\int\limits_{\mathbb{R}^n}\max_{x^n\in\mathscr{C}_n}\omega_{\mathcal{T}}(z^n|x^n,H_e^n)dz^n\right).$$

Proof. See Appendix. □

With codeword power constraint P and noise variance σ^2, we denote the signal to noise ratio by $\mathsf{SNR} = \frac{P}{\sigma^2}$.

Fact 3. *[24] The point-to-point capacity of a real fast fading channel with S-CSIT is given by*

$$C = \frac{1}{2}\mathbb{E}_H\left[\log(1+H^2\mathsf{SNR})\right].$$

where H is the random variable representing the channel coefficient.

To this end, our goal for the remainder of this chapter will be to show (for any code)

$$\lim_{\substack{n\to\infty\\ \epsilon\to 0}}\left(\frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n}\right) \le \frac{1}{2}\mathbb{E}_{H_e}\left[\log(1+H_e^2\mathsf{SNR})\right]$$

where SNR now represents the eavesdropper's SNR: $\frac{P}{\sigma_e^2}$.

4.1. Typical Set Motivation

In the next section, we will construct a typical set $\mathbb{R}_+ \times \mathcal{T}_n$ where $\mathcal{T}_n \subset \mathscr{C}_n \times \mathbb{R}_+^n \times \mathbb{R}^n$ which contains *probabilistically enough* content about our space. However, in this section, we will first provide motivation for choosing such a set.

The capacity expression for an additive white Gaussian noise channel (AWGN) is motivated by an intuitive argument called *sphere packing* (cf. [7, 24]). The argument asserts that due to properties of Gaussian random variables, a received output vector should be contained in some small n-dimensional ball around the transmitted codeword with high probability. In other words, the noise of the channel will only disturb the input vector by a certain amount (the radius of the small ball) with high probability. Furthermore, all received outputs should be contained in some larger ball with high probability since we are assuming that all the codewords are being transmitted while obeying the power constraint. If we use maximum likelihood decoding, given an output that resides in one of the small balls, the receiver assumes it came from the codeword that generated said ball. Therefore, the maximum number of small spheres we can pack into the larger ball roughly corresponds to how many codewords we can transmit reliably. This technique is called sphere packing since we are attempting to *pack* the larger ball with smaller spheres. Exact calculation is quite challenging; however, simply dividing the volume of the large ball by the volume in a small sphere gives an upper bound. What is perhaps surprising is that as the block length approaches infinity, this upper bound is actually achievable and is exactly the capacity of the AWGN channel.

We will provide a symmetric argument for the fast fading channel as justification for how and why we choose our typical sets the way we do in the next section. Given an input x^n and channel coefficient h_e^n (for the remainder of this subsection we will drop the subscript and refer to h_e as just h for ease), we know the output z^n will reside in some small ball about the point $h^n x^n$ with high probability since we assume the noise follows a Gaussian random variable. In fact, such a ball will have radius $\sqrt{n\sigma_e^2(1+\delta)}$ for every $\delta > 0$ sufficiently small.

In the case of the AWGN channel, the larger ball's dimensions were derived using the fact that we expect our channel to obey the law of conservation of energy; that is, the maximum output energy should be equal to the summation of the maximum input energy and noise energy. We expect a similar phenomenon to hold on the fast fading channel; however, the input energy will

also depend on the channel coefficient realization. During the ith symbol transmission, suppose h_i is the realized channel coefficient; then the effective maximum input power is given by $h_i^2 P$ so that the effective maximum average output power $\frac{1}{n} Z_i^2$ is given by $h_i^2 P + \sigma_e^2$. Therefore we expect the realization z_i^2 to be less than $n(h_i^2 P + \sigma_e^2)(1+\delta)$.

Since i is a *dimension* of the vector z^n, we should then expect z^n to be found in some volume where each component z_i is bounded by $\pm\sqrt{n(h_i^2 P + \sigma_e^2)(1+\delta)}$. Because h_i is *changing* for each use of the channel, each of these bounds will be different. Therefore, in contrast to the AWGN channel where each upper bound was constant, the volume we expect to contain most output vectors z^n is actually an n-dimensional ellipsoid with radii $\sqrt{n(h_i^2 P + \sigma_e^2)(1+\delta)}$. Thus, if we try to pack as many spheres into this ellipsoid as possible as illustrated in the (2-dimensional) Figure 4.1, we should come up with the maximum number of codewords we can transmit reliably, i.e., an expression for capacity.

Using the same technique as [7], we simply divide the volume of the ellipsoid by the volume of the small balls. That is, since the volume of an ellipsoid with radii r_i is given by $\eta_n \prod_{i=1}^n r_i$ where η_n is the same constant factor used to calculate the volume of an n-ball, it follows that an upper bound to the max number of codewords is given by:

$$\begin{aligned}\frac{\eta_n \prod_{i=1}^{n} \sqrt{n(h_i^2 P + \sigma_e^2)(1+\delta)}}{\eta_n \sqrt{n\sigma_e^2(1+\delta)}^{n}} &= \frac{\prod_{i=1}^{n} \sqrt{n\sigma_e^2(1+h_i^2\mathsf{SNR})(1+\delta)}}{\sqrt{n\sigma_e^2(1+\delta)}^{n}} \\ &= \prod_{i=1}^{n} \sqrt{1+h_i^2\mathsf{SNR}}.\end{aligned}$$

Since rate is usually defined as the logarithm of the number of codewords normalized by n, an upper bound to the max achievable rate is given by:

$$\begin{aligned}\frac{1}{n}\log\prod_{i=1}^{n}\sqrt{1+h_i^2\mathsf{SNR}} &= \frac{1}{2}\left(\frac{1}{n}\log\prod_{i=1}^{n}(1+h_i^2\mathsf{SNR})\right) \\ &= \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\log(1+h_i^2\mathsf{SNR})\right) \\ &\xrightarrow{n\to\infty} \frac{1}{2}\mathbb{E}\left[\log(1+H_e^2\mathsf{SNR})\right] \\ &= C_A,\end{aligned}$$

where the convergence in the penultimate line follows from the law of large numbers.

Since the above characterizations correctly estimated the asymptotic upper bound for the fading channel using the same sphere packing argument as in the AWGN case, we are confident moving forward that these bounds will produce sets that are typical in the proper sense. We will define these sets more rigorously in the forthcoming section.

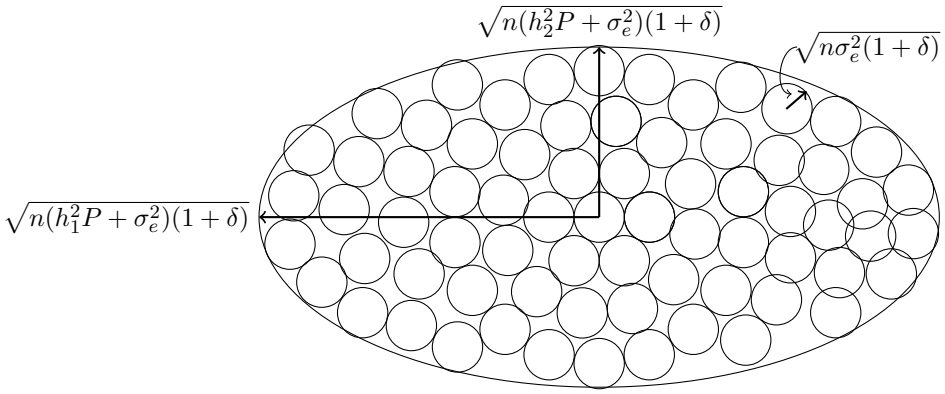

Figure 4.1. Sphere packing for the fading channel.

4.2. Constructing a Typical Set

In this section, we will be constructing a set and showing that it is typical by our definition. The set is made up of four constituent sets; one each concerning the output power, noise power, main channel coefficient power, and eavesdropper channel coefficient power. We begin by stating a lemma due to [23] where we have modified its form so as to be easily utilized in the following proofs. It can be considered a generalization of Hoeffding's inequality [12] to the case of *unbounded* random variables.

Lemma 2. *[23, Theorem 2.1] Let $\{W_i\}_{i=1}^n$ be a sequence of independent random variables. Suppose for all i there exists a $\gamma_i > 0$ such that $\mathbb{E}\left[e^{\gamma_i|W_i|}\right] < \infty$. Then for any sufficiently small $\alpha > 0$,*

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}(W_i - \mathbb{E}[W_i])\right| \leq \alpha\right] \geq 1 - 2e^{-\frac{n\alpha^2}{4K^*}}$$

where $K_i = 2(\mathbb{E}\left[W_i^4\right])^{\frac{1}{2}}\mathbb{E}\left[e^{\alpha|W_i|}\right]$ and $K^* = \max_i K_i$.

Definition. *Define the following sets where $\delta_n, \delta_n', \delta_n'' > 0$:*

- $\mathcal{P}_n^1$ *as the set of tuples* $(h_e^n, z^n) \in \mathbb{R}_+^n \times \mathbb{R}^n$ *such that*

$$\frac{1}{n}\sum_{i=1}^{n} \frac{z_i^2}{\sigma_e^2 + h_{e,i}^2 P} - 1 \leq \delta_n,$$

- $\mathcal{P}_n^2$ *as the set of* $z^n \in \mathbb{R}^n$ *that satisfy*

$$\|z^n - x^n h_e^n\|^2 \geq n\sigma_e^2(1 - \delta_n')$$

for a fixed $x^n \in \mathscr{C}_n$ *and* $h_e^n \in \mathbb{R}_+^n$,

- $\mathcal{P}_n^3$ *as the set of* $h_e^n \in \mathbb{R}_+^n$ *that satisfy*

$$\left| \frac{1}{n}\sum_{i=1}^{n} \log\left(1 + h_{e,i}^2 \mathsf{SNR}\right) - \mathbb{E}_{H_e}\left[1 + H_e^2 \mathsf{SNR}\right] \right| \leq \delta_n''.$$

These sets each correspond exactly to what we motivated in the previous section. $\mathcal{P}_n^1$ corresponds to the set of eavesdropper output powers and channel coefficients we wish to consider. $\mathcal{P}_n^2$ is essentially describing the *least* amount of noise added to $h_e^n x^n$ during transmission. We wish to only consider those channel coefficients that have a sufficiently high probability of occurring and not necessarily the entire space, which $\mathcal{P}_n^3$ is describing as will be proved in the following lemma.

Lemma 3.

1. *Let* $\epsilon_n^1 = 2e^{-\frac{n\delta_n^2}{4K^*}}$. *For any* $x^n \in \mathscr{C}_n$,

$$\mathbb{P}\left[(H_e^n, Z^n) \in \mathcal{P}_n^1 \,\middle|\, X^n = x^n\right] \geq 1 - \epsilon_n^1.$$

2. *Let* $\epsilon_n^2 = e^{-\frac{n{\delta_n'}^2}{4}}$. *For any* $x^n \in \mathscr{C}_n$ *and* $h_e^n \in \mathbb{R}_+^n$,

$$\mathbb{P}\left[Z^n \in \mathcal{P}_n^2 \,\middle|\, X^n = x^n, H_e^n = h_e^n\right] \geq 1 - \epsilon_n^2.$$

3. *Let* $\epsilon_n^3 = 2e^{-\frac{n{\delta_n''}^2}{4K}}$. *Then,*

$$\mathbb{P}\left[H_e^n \in \mathcal{P}_n^3\right] \geq 1 - \epsilon_n^3.$$

Proof. See Appendix. □

4.3. Typical Set

We now use the sets constructed above to create our typical set. Define each of the following sets:

$$\mathcal{T}_n^1 = \{(x^n, h_e^n, z^n) : x^n \in \mathscr{C}_n \text{ and } (h_e^n, z^n) \in \mathcal{P}_n^1\},$$
$$\mathcal{T}_n^2 = \{(x^n, h_e^n, z^n) : \text{ for each choice of } x^n \in \mathscr{C}_n \text{ and } h_e^n \in \mathbb{R}_+^n, z^n \in \mathcal{P}_n^2\},$$
$$\mathcal{T}_n^3 = \{(x^n, h_e^n, z^n) \in \mathscr{C}_n \times \mathcal{P}_n^3 \times \mathbb{R}^n\}.$$

We can think of each of these three sets as the *expansion* set that corresponds to each of the previous three sets $\mathcal{P}_n^1$, $\mathcal{P}_n^2$, and $\mathcal{P}_n^3$ but lives in the entire space $\mathscr{C}_n \times \mathbb{R}_+^n \times \mathbb{R}^n$.

We now take the intersection of these sets to construct one final set

$$\mathcal{T}_n = \mathcal{T}_n^1 \cap \mathcal{T}_n^2 \cap \mathcal{T}_n^3.$$

The following proposition shows that $\mathbb{R}_+^n \times \mathcal{T}_n$ is typical for any n (with a change of Cartesian ordering from the above definition of typical set).

Proposition 3. *Let* $\epsilon_n = \epsilon_n^1 + \epsilon_n^2 + \epsilon_n^3$ *then*

$$\mathbb{P}\left[(H_m^n, X^n, H_e^n, Z^n) \in \mathbb{R}_+^n \times \mathcal{T}_n | X^n = x^n\right] \geq 1 - \epsilon_n$$

for any $x^n \in \mathscr{C}_n$. *That is,* $\mathbb{R}_+^n \times \mathcal{T}_n$ *is a* $(1 - \epsilon_n)$*-typical set.*

Proof.

$$\begin{aligned}
&\mathbb{P}\left[(H_m^n, X^n, H_e^n, Z^n) \in \mathbb{R}_+^n \times \mathcal{T}_n | X^n = x^n\right] \\
&\overset{1}{\geq} \mathbb{P}\left[(H_m^n) \in \mathbb{R}_+^n | X^n = x^n\right] + \mathbb{P}\left[(X^n, H_e^n, Z^n) \in \mathcal{T}_n | X^n = x^n\right] - 1 \\
&= \mathbb{P}\left[(X^n, H_e^n, Z^n) \in \mathcal{T}_n | X^n = x^n\right] \\
&= \mathbb{P}\left[(X^n, H_e^n, Z^n) \in \mathcal{T}_n^1 \cap \mathcal{T}_n^2 \cap \mathcal{T}_n^3 | X^n = x^n\right] \\
&\overset{2}{\geq} \mathbb{P}\left[(X^n, H_e^n, Z^n) \in \mathcal{T}_n^1 | X^n = x^n\right] + \cdots
\end{aligned}$$

$$
\begin{aligned}
&\cdots + \mathbb{P}\left[(X^n, H_e^n, Z^n) \in \mathcal{T}_n^2 | X^n = x^n\right] + \mathbb{P}\left[(X^n, H_e^n, Z^n) \in \mathcal{T}_n^3 | X^n = x^n\right] - 2 \\
&\overset{3}{=} \mathbb{P}\left[(H_e^n, Z^n) \in \mathcal{P}_n^1 \,\middle|\, X^n = x^n\right] + \mathbb{E}_{H_e^n}\left(\mathbb{P}\left[Z^n \in \mathcal{P}_n^2 \middle| H_e^n = h^n, X^n = x^n\right]\right) + \mathbb{P}\left[H_e^n \in \mathcal{P}_n^3\right] - 2 \\
&\overset{4}{\geq} (1 - \epsilon_n^1) + (1 - \epsilon_n^2) + (1 - \epsilon_n^3) - 2 \\
&= 1 - (\epsilon_n^1 + \epsilon_n^2 + \epsilon_n^3) \\
&= 1 - \epsilon_n
\end{aligned}
$$

Justification.

1. *Fréchet inequality for Cartesian products.*

2. *Fréchet inequality for intersections.*

3. *The first, third, and fourth terms of the sum follow immediately. The second term is explained here:*

$$
\begin{aligned}
&\mathbb{P}\left[(X^n, H_e^n, Z^n) \in \mathcal{T}_n^2 | X^n = x^n\right] \\
&= \int_{\mathcal{H}^n} \int_{\mathcal{Z}^n} \omega(z^n, h_e^n | x^n) \mathbb{1}((x^n, h_e^n, z^n) \in \mathcal{T}_n^2) dz^n dh_e^n \\
&= \int_{\mathcal{H}^n} \int_{\mathcal{Z}^n} \frac{\omega(z^n | h_e^n, x^n) \omega(h_e^n x^n)}{\omega(x^n)} \mathbb{1}((x^n, h_e^n, z^n) \in \mathcal{T}_n^2) dz^n dh_e^n \\
&= \int_{\mathcal{H}^n} \omega(h_e^n) \int_{\mathcal{Z}^n} \omega(z^n | h_e^n, x^n) \mathbb{1}((x^n, h_e^n, z^n) \in \mathcal{T}_n^2) dz^n dh_e^n \\
&= \mathbb{E}_{H_e^n}\left(\mathbb{P}\left[Z^n \in \mathcal{P}_n^2 \middle| H_e^n = h^n, X^n = x^n\right]\right)
\end{aligned}
$$

4. *Lemma 3.*

□

With our typical set $\mathbb{R}_+ \times \mathcal{T}_n$ in hand, we are ready to prove the main result of this section and determine a characterization for semantically secure achievable rates for the fast fading wiretap channel with S-CSIT.

4.4. Set of Achievable Rates Under Semantic Security

Theorem 1. *Consider the fast fading wiretap channel with S-CSIT and let $\mathcal{T}_n$ and ϵ_n be as defined in Proposition 3. It follows that:*

$$\lim_{\substack{n\to\infty \\ \epsilon\to 0}} \left(\frac{\boldsymbol{\mathcal{I}}_n^{\epsilon}}{n}\right) \leq \frac{1}{2}\mathbb{E}_{H_e}\left[\log(1+H_e^2\mathsf{SNR})\right].$$

Proof.

$$\begin{aligned}
2^{\boldsymbol{\mathcal{I}}_n^{\epsilon}} &\overset{1}{\leq} 2^{\boldsymbol{\mathcal{I}}_n(A_{\mathcal{T}_n}^n)} \\
&\overset{2}{=} \mathbb{E}_{H_e^n} \int\limits_{\mathbb{R}^n} \max_{x^n\in\mathscr{C}_n} \omega_{\mathcal{T}_n}(z^n|x^n,H_e^n)dz^n \\
&\overset{3}{=} \mathbb{E}_{H_e^n} \int\limits_{\mathbb{R}^n} \max_{x^n\in\mathscr{C}_n} \left[\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_e^2}} e^{-\frac{(z_i-H_{e,i}x_i)^2}{2\sigma_e^2}}\right)\mathbb{1}((x^n,H_e^n,z^n)\in\mathcal{T}_n)\right]dz^n \\
&= \frac{1}{(2\pi\sigma_e^2)^{\frac{n}{2}}}\mathbb{E}_{H_e^n} \int\limits_{\mathbb{R}^n} \max_{x^n\in\mathscr{C}_n} \left(e^{-\frac{\|z^n-H_e^nx^n\|^2}{2\sigma_e^2}}\mathbb{1}((x^n,H_e^n,z^n)\in\mathcal{T}_n)\right)dz^n \\
&\overset{4}{\leq} \frac{e^{-\frac{n}{2}(1-\delta_n')}}{(2\pi\sigma_e^2)^{\frac{n}{2}}}\cdot\mathbb{E}_{H_e^n} \int\limits_{\mathbb{R}^n} \max_{x^n\in\mathscr{C}_n} \mathbb{1}((x^n,H_e^n,z^n)\in\mathcal{T}_n)dz^n \\
&\overset{5}{=} \frac{e^{-\frac{n}{2}(1-\delta_n')}}{(2\pi\sigma_e^2)^{\frac{n}{2}}}\cdot\mathbb{E}_{H_e^n} \int\limits_{\mathbb{R}^n} \max_{x^n\in\mathscr{C}_n} \mathbb{1}((x^n,H_e^n,z^n)\in\mathcal{T}_n^1\cap\mathcal{T}_n^2)\mathbb{1}((x^n,H_e^n,z^n)\in\mathcal{T}_n^3)dz^n \\
&= \frac{e^{-\frac{n}{2}(1-\delta_n')}}{(2\pi\sigma_e^2)^{\frac{n}{2}}}\cdot\mathbb{E}_{H_e^n}\left[\mathbb{1}(H_e^n\in\mathcal{P}_n^3)\left(\int\limits_{\mathbb{R}^n} \max_{x^n\in\mathscr{C}_n} \mathbb{1}((x^n,H_e^n,z^n)\in\mathcal{T}_n^1\cap\mathcal{T}_n^2)dz^n\right)\right]. \qquad \text{(E1)}
\end{aligned}$$

Justification.

1. *$\mathbb{R}_+^n\times\mathcal{T}_n$ is a $(1-\epsilon_n)$ typical set; however, it may not be the set corresponding to the "smallest" ϵ_n smooth max-information. Note that here we are labeling our typical set as just $\mathcal{T}_n$ for ease and dropping the subscript on ϵ_n.*

2. *Proposition 2. Since we no longer have any dependencies on H_m, we will henceforth write our typical set as just $\mathcal{T}_n$.*

3. *Each output, given $X_i = x_i$ and $H_{e,i} = h_{e,i}$, is $Z_i = h_{e,i}x_i + U_{e,i}$. This is simply a normal random variable that is shifted in mean by $h_{e,i}x_i$ with variance σ_e^2. Thus, the density for each*

transmission is given as

$$\omega(z_i|x_i, h_{e,i}) = \frac{1}{\sqrt{2\pi\sigma_e^2}} e^{-\frac{(z_i - h_{e,i}x_i)^2}{2\sigma_e^2}}.$$

Since we assume the channel is memoryless, we can split this density simply into a product.

4. *We are working on* $\mathcal{T}_n$ *and thus* $\mathcal{P}_n^2$*; thus,* $\|z^n - h_e^n x^n\|^2 \geq n\sigma_e^2(1 - \delta_n')$.

5. $\mathcal{T}_n^1, \mathcal{T}_n^2, \mathcal{T}_n^3$ *are defined in Section 4.3.*

Let us gain some intuition of what is happening at this point. In eq. (E1), suppose $\mathcal{T}_n^\star = \mathcal{T}_n^1 \cap \mathcal{T}_n^2$ and let us understand the term

$$\max_{x^n \in \mathscr{C}_n} \mathbb{1}((x^n, h_e^n, z^n) \in \mathcal{T}_n^\star).$$

If we temporarily *fix* z^n and h_e^n, then this maximization is simply asking if there exists some codeword $x^n \in \mathscr{C}_n$ that makes the sequence (x^n, h_e^n, z^n) an element of the set $\mathcal{T}_n^\star$. If there does exist such an x^n then this function returns 1; otherwise, it returns 0. If we now relax z^n and only fix h_e^n, $\mathcal{T}_n^\star$ can be thought of as a typical set as well: it is the set of typical input-output pairs. Thus the above function takes some output z^n and asks if there is possibly any codewords that could have generated such an output knowing the channel coefficient is h_e^n. It follows then, that the integral

$$\int_{\mathbb{R}^n} \max_{x^n \in \mathscr{C}_n} \mathbb{1}((x^n, h_e^n, z^n) \in \mathcal{T}_n^\star) dz^n,$$

roughly "counts" the number of valid input-output pairs given some h_e^n.

To calculate such an integral, we need to know the *shape* of $\mathcal{T}_n^\star$ and it is clear that $\mathcal{T}_n^\star = \mathcal{T}_n^1 \cap \mathcal{T}_n^2 \subset \mathcal{T}_n^1$ so that we can replace the $\mathcal{T}_n^\star$ with a $\mathcal{T}_n^1$ in the above integral at the expense of an inequality. However, this has *removed* the maximization since $\mathcal{T}_n^1$ has no dependence on codewords. Therefore the above integration is less than or equal to

$$\int_{\mathbb{R}^n} \mathbb{1}((h_e^n, z^n) \in \mathcal{P}_n^1) dz^n.$$

Now given some h_e^n, $\mathcal{P}_n^1$ is actually an ellipsoid with radii: $\sqrt{n\sigma_e^2(1+h_{e,i}^2\mathsf{SNR})(1+\delta_n)}$. Therefore, this integration is actually calculating the *volume* of such an ellipsoid, which is calculated to be

$$\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}\prod_{i=1}^{n}\sqrt{n\sigma_e^2(1+h_{e,i}^2\mathsf{SNR})(1+\delta_n)},$$

where Γ is the usual gamma function of analysis.

Let us return to Equation (E1); using the aforementioned reasoning above we have:

$$\begin{aligned}
(\text{E1}) &\leq \frac{e^{-\frac{n}{2}(1-\delta_n')}}{(2\pi\sigma_e^2)^{\frac{n}{2}}}\cdot\mathbb{E}_{H_e^n}\left[\left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}\mathbb{1}(H_e^n\in\mathcal{P}_n^3)\prod_{i=1}^{n}\sqrt{n\sigma_e^2(1+H_{e,i}^2\mathsf{SNR})(1+\delta_n)}\right)\right]\\
&= \frac{e^{-\frac{n}{2}(1-\delta_n')}}{(2\pi\sigma_e^2)^{\frac{n}{2}}}\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}(n\sigma_e^2(1+\delta_n))^{\frac{n}{2}}\mathbb{E}_{H_e^n}\left[\left(\mathbb{1}(H_e^n\in\mathcal{P}_n^3)\prod_{i=1}^{n}\sqrt{(1+H_{e,i}^2\mathsf{SNR})}\right)\right]\\
&= \left((1+\delta_n)e^{\delta_n'}\frac{n}{2e\cdot\Gamma(\frac{n}{2}+1)^{\frac{2}{n}}}\right)^{\frac{n}{2}}\int_{\mathcal{P}_n^3}\omega(h_e^n)\prod_{i=1}^{n}\sqrt{(1+h_{e,i}^2\mathsf{SNR})}dh_e^n\\
&\overset{6}{\leq} \left((1+\delta_n)e^{\delta_n'}\frac{n}{2e\cdot\Gamma(\frac{n}{2}+1)^{\frac{2}{n}}}\right)^{\frac{n}{2}}\int_{\mathcal{P}_n^3}\omega(h_e^n)2^{\frac{n}{2}\left(\delta_n''+\mathbb{E}_{H_e}[1+H_e^2\mathsf{SNR}]\right)}dh_e^n\\
&= \left((1+\delta_n)e^{\delta_n'}\frac{n}{2e\cdot\Gamma(\frac{n}{2}+1)^{\frac{2}{n}}}\right)^{\frac{n}{2}}2^{\frac{n}{2}\left(\delta_n''+\mathbb{E}_{H_e}[1+H_e^2\mathsf{SNR}]\right)}\int_{\mathcal{P}_n^3}\omega(h_e^n)dh_e^n\\
&= \left((1+\delta_n)e^{\delta_n'}\frac{n}{2e\cdot\Gamma(\frac{n}{2}+1)^{\frac{2}{n}}}\right)^{\frac{n}{2}}2^{\frac{n}{2}\left(\delta_n''+\mathbb{E}_{H_e}[1+H_e^2\mathsf{SNR}]\right)}
\end{aligned}$$

Justification.

6. *Due to the bounds of integration we know that every value of h_e will satisfy the definition of $\mathcal{P}_n^3$, thus it satisfies:*

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\log(1+h_{e,i}^2\mathsf{SNR})-\mathbb{E}_{H_e}[1+H_e^2\mathsf{SNR}]&\leq\delta_n''\\
\Rightarrow\frac{1}{n}\log\left(\prod_{i=1}^{n}(1+h_{e,i}^2\mathsf{SNR})\right)&\leq\delta_n''+\mathbb{E}_{H_e}[1+H_e^2\mathsf{SNR}]
\end{aligned}$$

Multiplying by n and exponentiating both sides:

$$\Rightarrow \prod_{i=1}^{n}(1+h_{e,i}^2\mathsf{SNR}) \leq 2^{n\left(\delta_n''+\mathbb{E}_{H_e}[1+H_e^2\mathsf{SNR}]\right)}$$
$$\Rightarrow \prod_{i=1}^{n}\sqrt{(1+h_{e,i}^2\mathsf{SNR})} \leq 2^{\frac{n}{2}\left(\delta_n''+\mathbb{E}_{H_e}[1+H_e^2\mathsf{SNR}]\right)}$$

Taking the logarithm of the beginning and end, and dividing by n, we continue as:

$$\begin{aligned}\frac{\boldsymbol{\mathcal{I}}_n^{\epsilon}}{n} &\leq \frac{1}{n}\log\left[\left((1+\delta_n)e^{\delta_n'}\frac{n}{2e\cdot\Gamma(\frac{n}{2}+1)^{\frac{2}{n}}}\right)^{\frac{n}{2}} 2^{\frac{n}{2}\left(\delta_n''+\mathbb{E}_{H_e}[1+H_e^2\mathsf{SNR}]\right)}\right] \\ &= \frac{1}{2}\log\left((1+\delta_n)e^{\delta_n'}\frac{n}{2e\cdot\Gamma(\frac{n}{2}+1)^{\frac{2}{n}}}\right)+\frac{1}{2}\left(\delta_n''+\mathbb{E}_{H_e}[1+H_e^2\mathsf{SNR}]\right) \\ &= \underbrace{\frac{1}{2}\log\left((1+\delta_n)e^{\delta_n'}\right)}_{\text{A1}}+\underbrace{\frac{1}{2}\log\left(\frac{n}{2e\cdot\Gamma(\frac{n}{2}+1)^{\frac{2}{n}}}\right)}_{\text{A2}}+\frac{1}{2}\left(\delta_n''+\mathbb{E}_{H_e}[1+H_e^2\mathsf{SNR}]\right).\end{aligned}$$

Let us see the asymptotic behavior of these first two terms.

A1. If we choose $\delta_n \to 0$ and $\delta_n' \to 0$ as $n \to \infty$ at rates sufficiently slow (so as to allow $1-\epsilon_n^1 \to 1$ and $1-\epsilon_n^2 \to 1$ resp.), then A1 $\to 0$ as $n \to \infty$.

A2. It can be shown that $\frac{n}{2e\cdot\Gamma(\frac{n}{2}+1)^{\frac{2}{n}}} \to 1$ as $n \to \infty$ so that A2 $\to 0$ as $n \to \infty$.

Since we can choose $\delta_n, \delta_n', \delta_n''$ in such a way so that $\delta_n'' \to 0$ and $\epsilon_n^1, \epsilon_n^2, \epsilon_n^3 \to 0$ as $n \to \infty$, it follows that $\epsilon_n \to 0$ as $n \to \infty$. Combing these previous steps yields our claim:

$$\lim_{\substack{n\to\infty\\ \epsilon\to 0}}\left(\frac{\boldsymbol{\mathcal{I}}_n^{\epsilon}}{n}\right) \leq \frac{1}{2}\mathbb{E}_{H_e}\left[\log(1+H_e^2\mathsf{SNR})\right].$$

□

We have now completed step one of the procedure given in Chapter 3. The following corollary then tells us what semantically secure rates we can achieve given this bound.

Corollary 1. *The transmission scheme of Section 3.5 can achieve an overall* ***semantic*** *secrecy rate of $C_T - C_A$ on the S-CSIT fast fading wiretap channel when $C_T > C_A$ and $R_{\mathcal{C}} = C_T$.*

Proof. We can combine the previous theorem with Lemma 1 and note that $\delta_n, \delta'_n, \delta''_n$ can be chosen in such a way that $\epsilon_n \to 0$ as $n \to \infty$ and the claim follows. □

Reiterating the significance of this result, we have found an achievable secrecy rate for the S-CSIT fast fading wiretap channel where we have not made *any* restrictive assumptions on how the channel fading is distributed, and we have used the semantic security metric. Thus, this result applies to *any* S-CSIT fast fading wiretap channel that exists. Furthermore, the achievability scheme is modular so as to be placed in existing systems not designed for security and when the error correcting code is efficient and explicit, the entire system is efficient and explicit. All one needs to do is concatenate the UHF preprocessing scheme with an ECC satisfying the rates in Lemma 1 and they are guaranteed that their data transmissions will be semantically secure!

While Corollary 1 shows that we can achieve *a* positive semantically secure secrecy rate on any fast fading wiretap channel (and gives what that rate is), we actually have a stronger result for a certain class of fast fading channels. For the class of fast fading channels where the eavesdropper channel is *stochastically degraded* with respect to the main channel, which includes the case of when both channels are *Rayleigh faded*, we can actually achieve the *best* rate possible - the secrecy capacity!

4.4.1. Special Cases

Definition. *[4] We say that channel* $(\mathcal{X}, \omega(z|x), \mathcal{Z})$ *is stochastically degraded with respect to channel* $(\mathcal{X}, \omega(y|x), \mathcal{Y})$ *if*

$$w(z|x) = \sum_{y \in \mathcal{Y}} \omega(z|y)\omega(y|x) \qquad \forall (x, z) \in \mathcal{X} \times \mathcal{Z}.$$

Intuitively, a stochastically degraded channel is a channel where the eavesdropper's channel is a "noisy" version of the main channels output.

Fact 4. *[16] The weak secrecy capacity of a stochastically degraded fast fading wiretap channel with S-CSIT and i.i.d. channel coefficients is given by*

$$C_S = C_T - C_A.$$

Proposition 4. *Using the UHF based transmission scheme of Section 3.5 on any fast fading stochastically degraded wiretap channel,*

1. *It is possible to achieve the semantic secrecy capacity.*

2. $(C_S)_{weak} = (C_S)_{semantic}$.

Proof. Follows directly from combining Fact 4 with Corollary 1 and Fact 2. □

As a result of this, we have determined the semantic secrecy capacity of the popular Rayleigh fast faded wiretap channel (since it falls into the category of stochastically degraded [16]). This is a complex fast fading wiretap channel where each channel coefficient is distributed i.i.d. as $\mathcal{CN}(0, \sigma_{h,m}^2)$ and $\mathcal{CN}(0, \sigma_{h,e}^2)$ respectively.

To the extent of the author's knowledge, this is the first time *any* semantically secure achievable rate has been given for the S-CSIT fast fading channel. Furthermore, this is the first time the semantic secrecy capacity has been determined for stochastically degraded fast fading channels.

5. PARTIAL CSIT

We now turn to the case of partial CSIT, where the transmitter only has access to full CSI about the main channel but only has statistical CSI about the eavesdropper's channel. Our goal in this chapter is the same as in the previous chapter - we wish to characterize a set of semantically secure rates for the channel at hand. To do so, we find an asymptotic upper bound, ξ, to $\frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n}$ for any choice of code so as to use Lemma 1 which will provide the achievable rates. Note that the transmitter could choose to not use the extra information available to it (that being the knowledge of the main channel's instantaneous channel coefficients), in which case we are back to the case of S-CSIT and the results of Chapter 4 hold. The transmitter can still achieve those rates by ignoring the extra information. However, we do not want to restrict ourselves to that case but instead we wish to find out the extent to which that extra information can benefit the legitimate parties. Therefore, our goal for the remainder of this chapter is to take advantage of that extra information at the transmitter's disposal. To do this, we devise a power allocation scheme similar to that of [4, 9] and show its reliability, security, and the rates achievable by its use.

5.1. Intuition

In this section, we try to develop some intuition to assist in the reader's understanding of the power allocation scheme we are using to achieve security as well as that of the proof of Theorem 2. We do so by first considering a very simplified example of what will be generalized and made mathematically rigorous in later sections. Let $W^n = (T^n, A^n)$ be a fast fading wiretap channel with partial CSIT. We wish to find a set of semantically secure achievable rates for V^n. We accomplish this by decomposing W^n into multiple simpler channels and analyze those individually.

5.1.1. Channel Decomposition

Suppose for the sake of simplicity that the main channel coefficient H_m ever only takes two possible values: h_1 and h_2, while the eavesdropper channel is still left arbitrary. From the transmitter's point of view, since she knows instantaneously which of these two realizations is present on the main channel, she sees the main channel as a Gaussian channel with input weighted by a *constant*; in other words she knows the main channel, T, at that instant is represented by $h_\alpha X_i + U_{m,i}$ where $\alpha \in \{1, 2\}$ and X_i, $U_{m,i}$ are the input and noise as usual. From here, she can

redefine the input to be $\bar{X}_i = h_\alpha X_i$ making the main channel into the Gaussian channel $\bar{X}_i + U_{m,i}$. On the other hand, the transmitter is still completely oblivious to what the realizations of H_e are for the eavesdropper channel A. Thus, that channel will *always* be represented as $H_{e,i}X_i + U_{e,i}$ to her.

Since ECC's are individually tailored to the channel upon which they will be operating, given a channel model $h_1 X_i + U_{m,i}$ for all $i \leq n$, the transmitter will certainly use a different ECC than if she had been given $h_2 X_i + U_{m,i}$ for all $i \leq n$ since one of the channels may have a higher capacity and thus can support larger rates. However, we are assuming that the main channel coefficient is not being held constant for all n channel uses but rather varies back and forth randomly between h_1 and h_2. Suppose now that the transmitter knew before the transmission started that out of the n channel uses about to occur, n_1 of them would realize the channel gain h_1, while h_2 would be realized n_2 times such that $n_1 + n_2 = n$. Further suppose that the first n_1 channel uses all had the realization h_1 and the remaining n_2 channel uses all had the realization h_2. Knowing all of this a priori, the transmitter could design two ECC's, $\mathcal{C}^1_{n_1}$ (with encoder/decoder e_{n_1}, d_{n_1}) and $\mathcal{C}^2_{n_2}$ (with encoder/decoder e_{n_2}, d_{n_2}), designed to operate *reliably* over the point-to-point weighted Gaussian channels $h_1 X + U_m$ and $h_2 X + U_m$ respectively. If the transmitter has to satisfy a power constraint P (as is usually the case) over n channel uses, the ECCs must have their own power constraints such that their sum does not violate the original power constraint. In other words, we must impose a power constraint γ_1 on $\mathcal{C}^1_{n_1}$ and power constraint γ_2 on $\mathcal{C}^2_{n_2}$, such that $\gamma_1 + \gamma_2 \leq P$. Furthermore, these power constraints along with $\mathcal{C}^1_{n_1}$ and $\mathcal{C}^2_{n_2}$ are assumed to be public knowledge as usual to ensure successful decoding by the receiver. To successfully transmit n symbols, the transmitter would then use $\mathcal{C}^1_{n_1}$ for the first n_1 channel uses before switching to $\mathcal{C}^2_{n_2}$ for the remaining n_2 channel uses. Note that we have made no claims about security as of yet.

Let's take a step back and look at what we now have. We now essentially have *two* wiretap channels (separated in time) as shown in Figure 5.1: one which we are using n_1 times, whose main channel is given by $h_1 X_i + U_{m,i}$ and eavesdropper channel $H_{e,i}X_i + U_{e,i}$ for $1 \leq i \leq n_1$, call this wiretap channel $W_1^{n_1}$, and one which we are using n_2 times, whose main channel is given by $h_2 X_i + U_{m,i}$ for $n_1 < i \leq n_2$ and eavesdropper channel the same as before, call this $W_2^{n_2}$. This technique of breaking the one channel into multiple parallel channels across time is known as *demultiplexing*.

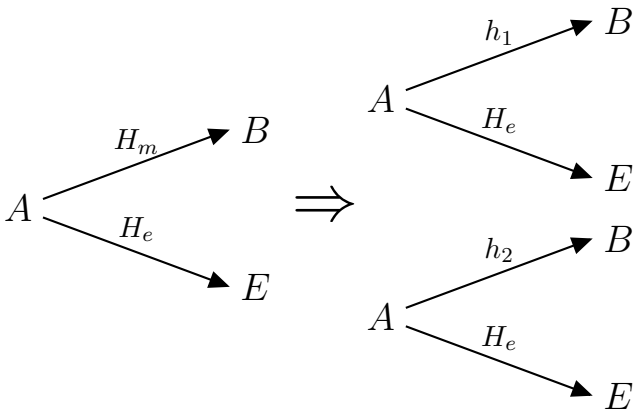

Figure 5.1. Decomposition of the fading wiretap channel with two channel coefficients.

5.1.2. Achievable Rates

Similarly to Chapter 4, we wish to bound $\frac{\mathcal{I}_n^\epsilon}{n}$ for the original channel W^n in order to invoke Lemma 1. Every wiretap channel in existence has its own max-information characterization and thus instead of trying to calculate a bound for the max-information of the original channel as a whole, we instead consider the max-information of each of the *individual* wiretap channels, $W_1^{n_1}$ and $W_2^{n_2}$, and find bounds on $\frac{\mathcal{I}_{n_1}^\epsilon}{n_1}$ and $\frac{\mathcal{I}_{n_2}^\epsilon}{n_2}$ respectively. Note that the bounds are being calculated over n_1 and n_2 channel uses respectively instead of the original n channel uses, hence the change in subscript. Combining the two yields a bound on $\frac{\mathcal{I}_n^\epsilon}{n}$ for the original channel. How they are combined exactly will come out in the calculations of Section 5.4. With the bound on $\frac{\mathcal{I}_n^\epsilon}{n}$ for W^n in hand, we can then invoke Lemma 1 to give us a set of achievable rates.

The results of Theorem 1 only required that the eavesdropper channel be a fast fading channel modeled by a random variable H_e and that the input X^n to the channel is independent of the channel coefficients (for Proposition 2 to hold). No restrictions were made as to what the main channel had to be distributed as. Indeed, in Theorem 1 and Proposition 2, the main channel coefficient H_m was kept arbitrary to produce the most general result possible. Letting H_m be a deterministic constant random variable (i.e. its realizations only take one value) allows the main channel to be a weighted Gaussian channel, exactly as we have here, one on each wiretap channel. Therefore, the results of Theorem 1 hold for both $W_1^{n_1}$ and $W_2^{n_2}$ giving us bounds on $\frac{\mathcal{I}_{n_1}^\epsilon}{n_1}$ and $\frac{\mathcal{I}_{n_2}^\epsilon}{n_2}$! Note that the input X^{n_1} for wiretap channel $W_1^{n_1}$ is indeed independent of the channel coefficients since although $\mathcal{C}_{n_1}^1$ was generated for a constant gain Gaussian channel with gain h_1, X^{n_1} is chosen

before transmission across the channel even begins. Thus current values of the channel coefficients play no part in the values of X^{n_1}. The analogous case holds for X^{n_2} on wiretap $W_2^{n_2}$. Since we now have bounds on $\frac{\boldsymbol{\mathcal{I}}_{n_1}^{\epsilon}}{n_1}$ and $\frac{\boldsymbol{\mathcal{I}}_{n_2}^{\epsilon}}{n_2}$ thanks to Theorem 1, we have a bound, ξ, on $\frac{\boldsymbol{\mathcal{I}}_{n}^{\epsilon}}{n}$ for the original wiretap channel W^n. Then Lemma 1 tells us that by using the transmission scheme of Section 3.5 we can achieve semantically secure rates on this fast fading wiretap channel when $R_{\mathcal{C}} > \xi$. However, we still have two technicalities to take care of.

In Lemma 1, $R_{\mathcal{C}}$ represented the asymptotic rate of *one* coding scheme. Here, we have two ECCs being used consecutively. Thus the overall rate of the ECCs in this scheme is just the sum of the rates of the ECCs weighted by the fraction of channel uses that ECC was being used. In this case it would be

$$R_{\mathcal{C}_n} = \frac{n_1}{n} R_1 + \frac{n_2}{n} R_2$$

where R_1 and R_2 are the rates of $\mathcal{C}_{n_1}^1$ and $\mathcal{C}_{n_2}^2$ respectively. $R_{\mathcal{C}}$ is then taken to be the asymptotic limit of this rate instead. The other technicality is that the preprocessing scheme cannot be applied directly due to our decomposition of the wiretap channel.

5.1.3. Transmission Scheme

Recall that the transmission scheme of Section 3.5 assumes that a k-bit message is preprocessed into an l-length pseudo-message which in turn is then encoded by an ECC into an n-symbol length codeword. In other words, the transmission scheme required an ECC of block length n be used to transmit data over the channel. The coding schemes above, $\mathcal{C}_{n_1}^1$ and $\mathcal{C}_{n_2}^2$, are obviously ECCs with block lengths n_1 and n_2 thus we cannot use the transmission scheme directly. The rectification of this is as follows: still preprocess the k-bit message into an l-length pseudo-message m'. Instead of encoding m' into an n-length codeword directly as before, we break apart m' into two bit strings, b_1 and b_2. b_1 is of length $n_1 R_1$ and b_2 is of length $n_2 R_2$, where R_1 and R_2 are the rates of $\mathcal{C}_{n_1}^1$ and $\mathcal{C}_{n_2}^2$ respectively, chosen in such a way that

$$l = n_1 R_1 + n_2 R_2$$

and R_1, R_2 are each less than or equal to the capacity of their respective channels. ECCs satisfying these rates can always be found due to Shannon's noisy channel coding theorem since we are

assuming they are each under capacity of their respective channels. We then pass as input b_1 and b_2 to their respective ECC encoders $e_{n_1}()$ and $e_{n_2}()$ to each be reliably transmitted. After the n_1-length output Y^{n_1} and n_2-length output Y^{n_2} are decoded at the receiver into n_1R_1-length message estimate $\hat{b_1}$ and n_2R_2-length message estimate $\hat{b_2}$ respectively, $\hat{b_1}$ and $\hat{b_2}$ are concatenated to form the pseudo-message estimate $\hat{m}'$. In this fashion, we can indeed use the preprocessor as before with a slight modification.

5.1.4. Removing Assumptions

Admittedly, the above paragraphs made many strong assumptions, most of which are completely unrealistic outside of a theoretical setting. As a recap, those unrealistic assumptions were:

1. The first n_1 channel uses all had channel coefficient h_1; the last n_2 channel uses all had channel coefficient h_2.
2. The transmitter knows a priori how many times a certain channel coefficient will occur.
3. H_m only takes two values: h_1 and h_2.

We now show how one can remove all of these assumptions. We will do so in the order they are presented above, but this is not necessary; we just remove them this way to keep the simplified example as long as possible.

5.1.4.1. Removing Assumption (1)

For now, we will assume that H_m still only takes values h_1 and h_2; however, we make no restrictions on when each occurs. Since the transmitter still knows how many times each coefficient will occur, she still uses $\mathcal{C}^1_{n_1}$ and $\mathcal{C}^2_{n_2}$ to generate codewords X^{n_1} and X^{n_2}. At each time i, the transmitter observes the channel coefficient h_α and transmits *one* symbol from the codeword corresponding to that realized channel coefficient. For example, at time 1 suppose the transmitter observed the main channel having coefficient h_2. Then she will send the first symbol of X^{n_2}. At time 2, suppose the coefficient was again h_2, the transmitter then transmits the second symbol of X^{n_2}. At time 3 suppose the coefficient was h_1, she then transmits the first symbol of X^{n_1} and so on. It is easy to see that proceeding in this fashion, the transmitter will have transmitted every symbol of both codewords, just in a mixed order. Since X^{n_1} is coming from a code with power constraint γ_1 and X^{n_2} is coming from a code with power constraint γ_2, we see that throughout the

transmission, we are sending inputs with (potentially) different powers dependent on the current channel coefficient. In essence, we are taking advantage of our extra knowledge of the channel coefficients by altering our power dynamically. This is what we will refer to as *power allocation*.

We have always assumed that the receiving parties know the instantaneous realizations of the channel coefficient belonging to their channel. Thus at time i the receiver knows which channel coefficient h_α occurred. With this knowledge the receiver can sort and store the received signals into those that belong to h_1 and those that belong to h_2 producing outputs Y^{n_1} and Y^{n_2} respectively. This technique is known as *code interleaving* and allows us to ignore in what order the channel coefficients occur thus removing assumption (1).

5.1.4.2. Removing Assumption (2)

The second assumption is obviously very unrealistic as it would imply that the transmitter has future knowledge of the channel. Dropping this assumption, the best a transmitter can do is make use of the channel statistics to try to predict how often the channel coefficients will occur. From the distribution of H_m, the transmitter can determine the probability each coefficient has of occurring, more specifically:

$$p_1 = \mathbb{P}[H_m = h_1]$$

$$p_2 = \mathbb{P}[H_m = h_2].$$

From this, the transmitter can estimate that np_1 channel uses will have channel coefficient h_1 and np_2 channel uses will have channel coefficient h_2. Suppose we redefined n_1 and n_2 from above as $n_1 = np_1$ and $n_2 = np_1$. Once again, before transmission across the channel, two codewords X^{n_1} and X^{n_2} are generated by $\mathcal{C}^1_{n_1}$ and $\mathcal{C}^2_{n_2}$ respectively. Since we now do not know exactly how many times a channel coefficient will occur, we must design our ECCs such that they probabilistically satisfy the overall power constraint P. In other words, $\mathcal{C}^1_{n_1}$ and $\mathcal{C}^2_{n_2}$ must have respective power constraints γ_1 and γ_2 such that $p_1\gamma_1 + p_2\gamma_2 \leq P$. As before, the transmitter sends symbols one at a time from the respective codeword depending on the current channel coefficient. However, since n_1 and n_2 are now just *estimates* of how many times the respective channel coefficients will appear on the channel, there are no guarantees that h_1 will appear n_1 times for example. Let N_1 be the number of times h_1 *actually* occurred and N_2 the number of times h_2 actually occurred. Note that we capitalize N_1 and N_2 since we do not know beforehand what values they will take and thus they

are random variables. One of three scenarios can then happen. The first possibility is the estimates being correct: $N_1 = n_1$ and $N_2 = n_2$. This is the case already considered above where we know that both codewords are sent over completely and thus successfully.

The second scenario being that of $N_1 < n_1$ and $N_2 > n_2$. In this case, we see that there were not enough occurrences of h_1 for the transmitter to send the entire n_1-length codeword X^{n_1}. By only sending *part* of the codeword, the transmitter is inherently sending a less or equal amount of information across the channel than if an entire codeword had been sent. A side effect of this is that similarly, only a smaller or equal amount of information can be leaked to the eavesdropper if only part of the codeword was sent rather than the entire codeword. At the same time, we see that there were more than enough occurrences of h_2 to send the entire n_2-length codeword X^{n_2}. Since in this case all of the information we wished to transmit via X^{n_2} has been transmitted in the first n_2 channel uses, we instruct the transmitter to not send any information through the channel for the remaining $N_2 - n_2$ time instances when channel coefficient h_2 appears. The third scenario is just the reverse of this one and thus follows similarly.

We see that by estimating via n_1 and n_2 as defined above will potentially lead to the transmitter not transmitting as much information as we would like, thus we must redefine n_1 and n_2 yet again to ensure that with high probability, both codewords will be fully sent across the channel. To this end, we redefine n_1 and n_2 as $n_1 = np_1 - \varepsilon_1$ and $n_1 = np_2 - \varepsilon_2$ respectively, where ε_1, ε_2 are chosen large enough such that the number of times h_1 and h_2 actually occur (N_1 and N_2 respectively) is larger than n_1 and n_2 with high probability. Furthermore, due to the law of large numbers, we know that as n grows, $N_1 \to np_1$ and $N_2 \to np_2$ thus we take ε_1, ε_2 to go to 0 as $n \to \infty$. By defining n_1 and n_2 in this fashion, we see that with high probability, X^{n_1} and X^{n_2} will be fully sent across the channel. We once again instruct the transmitter to not transmit any information on the remaining $N_1 - n_1$ time instances when h_1 is the channel coefficient after X^{n_1} has been fully sent. Similarly for the $N_2 - n_2$ time instances with channel coefficient h_2. Therefore, the transmitter does not know a priori how many times a certain channel coefficient occurs, but rather uses estimates.

5.1.4.3. Removing Assumption (3)

We now assume H_m is no longer restricted to two values, but can take any real number from $\mathbb{R}_+$. As is standard in literature [4], we assume that the channel coefficient H_m is bounded

and thus we take a value h_{max} to be a "probabilistic" upper bound for the realizations of H_m. In other words, we take h_{max} to be a large enough value such that all of the realizations of H_m will be smaller than h_{max} with high probability. From here, we partition the range of H_m, $[0, h_{max})$, into d smaller intervals $[h_{m,i}, h_{m,i+1})$ for $1 \leq i \leq d$ (i here is not the same as the time instant i). Note that both the transmitting party as well as the legitimate receiver and eavesdropper are assumed to know these partitions. With these partitions in hand, we extend all of the concepts above for d possible channel coefficients instead of just two. Thus we now define

$$p_i = \mathbb{P}\left[H_m \in [h_{m,i}, h_{m,i+1})\right].$$

and

$$n_i = np_i - \varepsilon_i.$$

Above, we designed ECCs $\mathcal{C}^1_{n_1}$ and $\mathcal{C}^2_{n_2}$ for Gaussian channels with *constant* channel gains meaning that we expected h_1 and h_2 to occur more than once. However, since $H_m \in \mathbb{R}_+$ we know that the main channel coefficient will take the same value twice with probability 0 and thus we cannot directly make ECC's based on the exact values of the channel coefficients otherwise they would all have length 1 (furthermore, since H_m is now coming from an uncountable set and we don't know a priori which n values will be realized, we would have to create an uncountable amount of ECCs - obviously not practical). To remedy this, we create d ECCs, $\{\mathcal{C}^i_{n_i}\}_d$, where each ECC is again designed for a Gaussian channel with constant gain, where the constant gain is given by the *lower* end of each of the d intervals described above, that being $h_{m,i}$ (we choose the first interval, where $h_{m,1} = 0$, to be arbitrarily small and choose to just not transmit when realizations fall into this interval). The block lengths of each of these ECCs is given by n_i respectively and each has power constraint γ_i such that the sum of the power constraints weighted by their probabilities of occurring is less than or equal to P.

The reason the lower end of each interval is chosen as the constant gain for the channel is to ensure the transmitter is not operating above the main channel's capacity. Consider all the realized channel coefficients that fall into the interval $[h_{m,i}, h_{m,i+1})$. The main channel will have an actual physical capacity, call it C_a, which is determined by the exact values of h_m. On the

other hand, before transmission, we can calculate a value for the capacity of a Gaussian channel with constant gain $h_{m,i}$, call it C_r. Since each of the actual realized values of h_m will be greater than or equal to the lower end of their respective interval, $h_{m,i}$, we know that $C_r \leq C_a$. Thus by always operating at a rate less than or equal to C_r, we are ensuring we are never operating above the channels true capacity C_a. By partitioning the main channel coefficients this way, we see that we have now demultiplexed the original channel W^n into d parallel wiretap channels as shown in Figure 5.2.

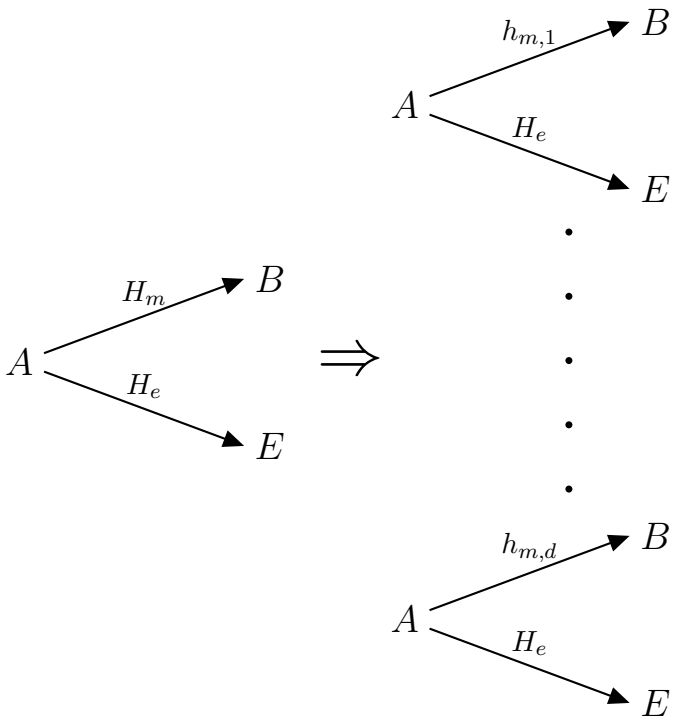

Figure 5.2. Decomposition of the fast fading wiretap channel with partial CSIT.

As before, at each time instant, the transmitter observes the channel coefficient h_m and transmits one symbol from the codeword corresponding to the interval h_m is contained in. For example, suppose at a certain time instant, $h_m \in [h_{m,7}, h_{m,8})$, the transmitter would then send one symbol from codeword X^{n_7} which was generated by $\mathcal{C}_{n_7}^7$ which in turn was designed for the Gaussian channel with constant gain $h_{m,7}$. In this way, code interleaving as illustrated in Figure 5.3 is still used to ensure that the order in which the channel coefficients appears does not matter. With this, we have now removed assumption (3) and hence all assumptions.

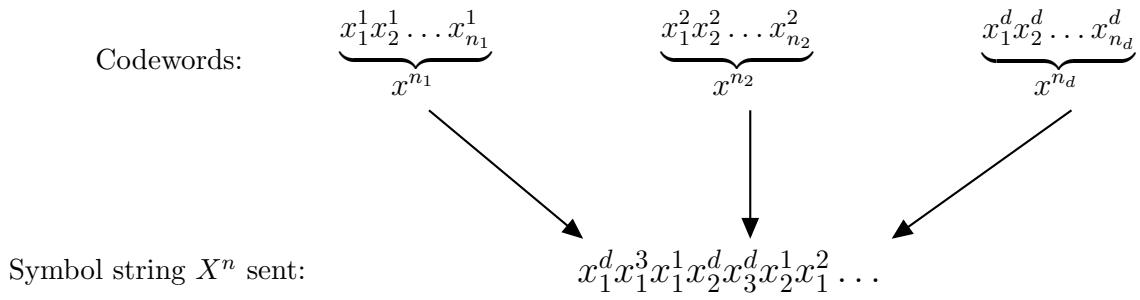

Figure 5.3. Code interleaving.

In Section 5.2 we restate the entire power allocation scheme in more succinct mathematical terms to make clear all of the assumptions being made. In Section 5.3 we then address this scheme's reliability, followed by its security in Section 5.4.

5.2. Power Allocation Scheme

Since the transmitter has access to CSI about the main channel, each party can demultiplex the fast fading wiretap channel into a set of d parallel channels by partitioning the channel coefficients of the main channel into d intervals. Each parallel wiretap channel is then composed of a time-invariant Gaussian main channel with a fast fading eavesdropper channel characterized by H_e as depicted in Figure 5.2. More specifically, we assume the fading gain of the main channel is bounded and divide the possible realizations of H_m into intervals $[h_{m,i}, h_{m,i+1})$ with $i \in [1, d]$. Let

$$p_i = \mathbb{P}\left[H_m \in [h_{m,i}, h_{m,i+1})\right].$$

Let N_i be the number of times channel i is actually used and let $n_i = p_i n - \varepsilon_i$, where ε_i is chosen sufficiently large such that the realization of N_i is greater than n_i with high probability and $\varepsilon_i \to 0$ as $n \to \infty$. For every index i, the transmitter and legitimate receiver will publicly agree on a transmit power $\gamma_i(H_m)$ where $\{\gamma_i\}_d$ is chosen such that

$$\sum_{i=1}^{d} p_i \gamma_i \leq P.$$

For $1 \leq i \leq d$, the transmitter and legitimate receiver also publicly agree upon an ECC $\mathcal{C}_{n_i}^i$ (with codebook $\mathscr{C}_{n_i}^i$) designed to operate on the Gaussian point-to-point channel with constant channel gain $h_{m,i}$. We denote by R_i the rate of $\mathcal{C}_{n_i}^i$ and the overall rate over the main channel to

be

$$R_{\mathcal{C}_n} = \sum_{i=1}^{d} p_i R_i.$$

The full coding scheme is then outlined as follows: a message $m \in \mathcal{M}$ is chosen which passes through the preprocessing layer to produce an l-length pseudo-message $m' \in \mathcal{M}'$. These l bits are then divided into sets of $n_i R_i$ bits such that

$$l = \sum_i n_i R_i.$$

A codeword is then generated for each of these sets by their respective $\mathcal{C}_{n_i}^i$ and the multiplexing strategy outlined above is then employed to transmit the ith codeword when the channel state is in the ith interval as illustrated in Figure 5.4. In more detail, at each time instant i the multiplexer will determine what the channel state is and send *one* symbol from the codeword associated with that channel gain.

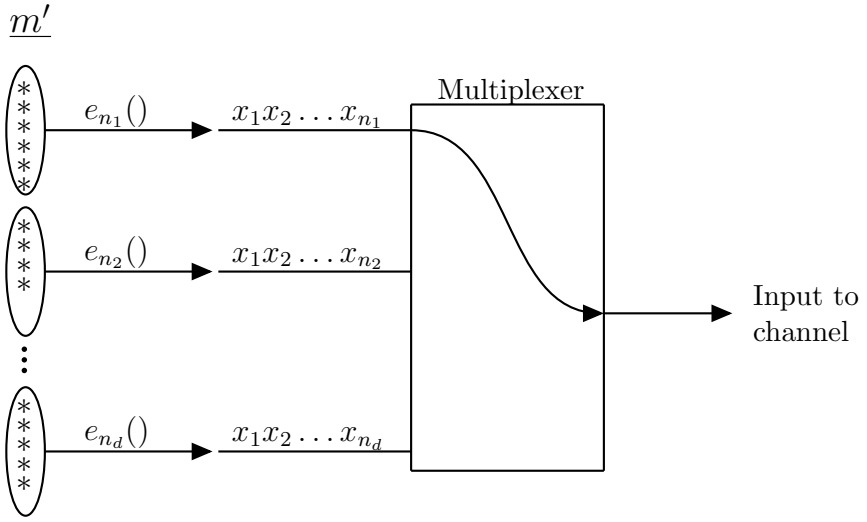

Figure 5.4. Multiplexing scheme.

5.3. Reliability

The reliability of this scheme comes from the aggregate reliability of all the ECC's being employed on the d parallel channels. Since we are assuming an ECC $\mathcal{C}_{n_i}^i$ is chosen to operate *successfully* over the ith point-to-point main channel, we know that $R_i \leq C_i$ where C_i here denotes

the point-to-point channel capacity between the transmitter and legitimate receiver. Since we are operating less than or equal to capacity, we know that the probability of error will be negligible, i.e., for any $\epsilon > 0$:

$$\mathbb{P}_e(\mathcal{C}_{n_i}^i) \leq \epsilon.$$

In other words, the receiver will be able to recover each n_i-length codeword with high probability. Thus the probability of error for the entire n-length transmission is just probability of error for each individual n_i-length codeword weighted by the probability that that code is used:

$$\mathbb{P}_e(\mathcal{C}_n) = \sum_i^d p_i \mathbb{P}_e(\mathcal{C}_{n_i}^i) \leq \sum_i^d p_i \epsilon = \epsilon$$

for any $\epsilon > 0$. Now that this scheme has been shown to be reliable, we now address its security.

5.4. Security

We wish to bound $\frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n}$ of this demultiplexed fast fading channel by considering the set of d parallel wiretap channels outlined above and each of *their* individual associated max-information terms for which we already know the bound found in Theorem 1. Again, the only way this differs from that of Chapter 4 is that in the case of S-CSIT, we are not allowed to vary the power we are transmitting at due to our lack of knowledge of instantaneous CSIT, whereas in the case of partial CSIT, we can vary our power to align with what the current channel gain is.

Definition. *Define the following sets:*

$$\mathcal{T}'_{n_i} = \{(x^{n_i}, h_m^{n_i}, h_e^{n_i}, z^{n_i}) : h_m^{n_i} \in \mathbb{R}_+^{n_i}, (x^{n_i}, h_e^{n_i}, z^{n_i}) \in \mathcal{T}_{n_i}\}$$
$$\mathcal{T}'_n = \bigotimes_i \mathcal{T}'_{n_i}$$

where $\mathcal{T}_{n_i}$ is defined in Section 4.3.

Similarly to the case of S-CSIT, we do not wish to consider the entire space from which inputs, outputs, and channel coefficients can take values; rather, we only wish to consider those tuples which have a high probability of occurring. Therefore, we create a typical set which will characterize the tuples we would "typically" expect to occur for the wiretap channel W^n. On each of the d parallel wiretap channels, we are using codewords from different codebooks, and thus

each parallel channel will have its own typical set characterizing the inputs, outputs, and channel coefficients that occur with high probability which is exactly given by $\mathcal{T}'_{n_i}$. The Cartesian product of $\{\mathcal{T}'_{n_i}\}$ defined above gives us a set with all high probability tuples from each channel. In other words, the typical set for the entire wiretap channel W^n is made up of the typical sets of its constituent d wiretap channels as proved next.

Lemma 4. *$\mathcal{T}'_n$ as defined above is a $(1-\epsilon_n)$ typical set.*

Proof. We first see that $\mathcal{T}'_{n_i}$ is $(1-\epsilon_{n_i})$ typical directly from Proposition 3. Then:

$$\begin{aligned}
&\mathbb{P}\left[(X^n, H_m^n, H_e^n, Z^n) \in \mathcal{T}'_n | X^n = x^n\right] \\
&= \mathbb{P}\left[(X^n, H_m^n, H_e^n, Z^n) \in \mathcal{T}'_{n_1} \times \cdots \times \mathcal{T}'_{n_d} | X^n = x^n\right] \\
&\overset{1}{\geq} \mathbb{P}\left[(X^{n_1}, H_m^{n_1}, H_e^{n_1}, Z^{n_1}) \in \mathcal{T}'_{n_1} | X^{n_1} = x^{n_1}\right] + \cdots \\
&\cdots + \mathbb{P}\left[(X^{n_d}, H_m^{n_d}, H_e^{n_d}, Z^{n_d}) \in \mathcal{T}'_{n_d} | X^{n_d} = x^{n_d}\right] - d + 1 \\
&\overset{2}{\geq} \sum_{i=1}^{d} (1-\epsilon_{n_i}) - d + 1 \\
&= d - d + 1 - \sum_{i=1}^{d} \epsilon_{n_i}
\end{aligned}$$

Let ϵ^* be the largest ϵ_{n_i} over all i

$$\geq 1 - d\epsilon^*$$

Since ϵ^* is going to 0 with $n \to \infty$ and we are free to choose d, we see that $\mathcal{T}'_n$ is a $(1-\epsilon_n)$ typical set.

Justification.

1. *Fréchet inequality for Cartesian products.*
2. *We know that $\mathcal{T}'_{n_i}$ is a $(1-\epsilon_{n_i})$ typical set for all i thus:*

$$\mathbb{P}\left[(X^{n_i}, H_m^{n_i}, H_e^{n_i}, Z^{n_i}) \in \mathcal{T}'_{n_i} | X^{n_i} = x^{n_i}\right] \geq 1 - \epsilon_{n_i}$$

and we sum over all i.

□

With the typical set $\mathcal{T}_n'$ in hand, we now aim to accomplish step 1 of the procedure outlined in Chapter 3 by finding the bound ξ for W^n. We do so by breaking up the wiretap channel into d parallel wiretap channels as described above. Finding an asymptotic bound for max-information on each of these channels using their respective typical sets yields us our bound ξ for the overall channel.

Theorem 2. *Consider a fast fading wiretap channel where the transmitter has partial CSIT with $\mathcal{T}_n'$ and ϵ_n as defined in Lemma 4. Using the power allocation scheme above, it follows that (dropping the subscript on ϵ):*

$$\lim_{\substack{n\to\infty \\ \epsilon\to 0}} \left(\frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n}\right) \leq \frac{1}{2}\mathbb{E}_{H_e H_m}\left[\log\left(1+\frac{\gamma(H_m)H_e^2}{\sigma_e^2}\right)\right].$$

Proof.

$$\begin{aligned}
2^{\boldsymbol{\mathcal{I}}_n^\epsilon} &\overset{1}{\leq} 2^{\boldsymbol{\mathcal{I}}_n(A^n_{\mathcal{T}_n'})} \\
&= \mathbb{E}_{H_m^n H_e^n} \int\limits_{\mathbb{R}^n} \max_{x^n\in\mathscr{C}_n} \omega_{\mathcal{T}_n'}(z^n|x^n,H_m^n,H_e^n)dz^n \\
&= \mathbb{E}_{H_m^n H_e^n} \int\limits_{\mathbb{R}^n} \max_{x^n\in\mathscr{C}_n} \omega(z^n|x^n,H_m^n,H_e^n)\mathbb{1}((x^n,H_m^n,z^n)\in\mathcal{T}_n')dz^n \\
&= \int \omega(h_m^n,h_e^n) \int\limits_{\mathbb{R}^n} \max_{x^n\in\mathscr{C}_n} \omega(z^n|x^n,h_m^n,h_e^n)\mathbb{1}((x^n,h_m^n,h_e^n,z^n)\in\mathcal{T}_n')dz^n dh_m^n dh_e^n \\
&\overset{2}{\leq} \prod_i \int \omega(h_m^{n_i},h_e^{n_i}) \int\limits_{\mathbb{R}^{n_i}} \max_{x^{n_i}\in\mathscr{C}_{n_i}^i} \omega(z^{n_i}|x^{n_i},h_m^{n_i},h_e^{n_i})\mathbb{1}((x^{n_i},h_m^{n_i},h_e^{n_i},z^{n_i})\in\mathcal{T}_{n_i}')dz^{n_i}dh_m^{n_i}dh_e^{n_i}
\end{aligned}$$

Let $\mathcal{J} = [h_{m,i}, h_{m,i+1})^{n_i} \times \mathbb{R}_+^{n_i}$

$$\overset{3}{=} \prod_i \int\limits_{\mathcal{J}} \omega(h_m^{n_i},h_e^{n_i}) \int\limits_{\mathbb{R}^{n_i}} \max_{x^{n_i}\in\mathscr{C}_{n_i}^i} \omega(z^{n_i}|x^{n_i},h_m^{n_i},h_e^{n_i})\mathbb{1}((x^{n_i},h_m^{n_i},h_e^{n_i},z^{n_i})\in\mathcal{T}_{n_i}')dz^{n_i}dh_m^{n_i}dh_e^{n_i}$$

Justification.

1. *$\mathcal{T}'_n$ is a $(1-\epsilon_n)$ typical set; however, it may not be the set corresponding to the "smallest" ϵ smooth max-information.*

2. *We wish to integrate over all n and to do so, we break up the integral into integrals over each n_i.*

   (a) *Suppose $N_i \geq n_i$. In this case, we have transmitted a full n_i length codeword over the ith channel and choose to not send information over the channel during the remaining $N_i - n_i$ channel uses. Then:*

$$\begin{aligned}
&\int_{\mathbb{R}^{N_i-n_i}} \max_{x^{N_i-n_i}} \omega(z^{N_i-n_i}|x^{N_i-n_i}, h_m^{N_i-n_i}, h_e^{N_i-n_i}) dz^{N_i-n_i} \\
&= \int_{\mathbb{R}^{N_i-n_i}} \omega(z^{N_i-n_i}|h_m^{N_i-n_i}, h_e^{N_i-n_i}) dz^{N_i-n_i} \\
&= 1
\end{aligned}$$

   Thus, if $N_i \geq n_i \, \forall i$ then we obtain equality at this line.

   (b) *Suppose $N_i < n_i$. In this case, the ith channel did not appear often enough for the transmitter to send an entire n_i length codeword. By not sending the full codeword, we are inherently limiting the amount of information sent across the channel and therefore the amount of information that can be leaked to the eavesdropper. Hence, sending the full n_i length codeword allows more information (or equal amount of information) to be leaked to the eavesdropper and therefore serves as an upper bound to the actual value. More clearly:*

$$\begin{aligned}
&\int_{\mathbb{R}^{N_i}} \omega(z^{N_i}|x^{N_i}, h_m^{N_i}, h_e^{N_i}) \mathbb{1}((x^{N_i}, h_m^{N_i}, h_e^{N_i}, z^{N_i}) \in \mathcal{T}'_{N_i}) dz^{N_i} \\
&\leq \int_{\mathbb{R}^{n_i}} \omega(z^{n_i}|x^{n_i}, h_m^{n_i}, h_e^{n_i}) \mathbb{1}((x^{n_i}, h_m^{n_i}, h_e^{n_i}, z^{n_i}) \in \mathcal{T}'_{n_i}) dz^{n_i}
\end{aligned}$$

3. *Due to the partitioning of the channel coefficients, we know that for each i, $H_m \in [h_{m,i}, h_{m,i+1})$.*

$$\stackrel{4}{=} \prod_i \mathbb{E}_{H_e^{n_i}} \int\limits_{\mathbb{R}^{n_i}} \max_{x^{n_i} \in \mathscr{C}_{n_i}^i} \omega_{\mathcal{T}_{n_i}}(z^{n_i}|x^{n_i}, H_e^{n_i}) dz^{n_i}$$
$$\stackrel{5}{=} \prod_i 2^{\boldsymbol{\mathcal{I}}_{n_i}(A_{\mathcal{T}_{n_i}}^{n_i})}$$

Taking the logarithm of each side and dividing by n:

$$\begin{aligned}
\frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n} &\leq \frac{1}{n} \log \left(\prod_i 2^{\boldsymbol{\mathcal{I}}_{n_i}(A_{\mathcal{T}_{n_i}}^{n_i})} \right) \\
&= \frac{1}{n} \sum_i \log \left(2^{\boldsymbol{\mathcal{I}}_{n_i}(A_{\mathcal{T}_{n_i}}^{n_i})} \right) \\
&\stackrel{6}{\leq} \frac{1}{n} \sum_i n_i \frac{1}{2} \mathbb{E}_{H_e} \left[\log \left(1 + H_e^2 \frac{\gamma_i(h_{m,i})}{\sigma_e^2} \right) \right] \\
&= \frac{1}{2n} \sum_i (p_i n - \varepsilon_i) \mathbb{E}_{H_e} \left[\log \left(1 + H_e^2 \frac{\gamma_i(h_{m,i})}{\sigma_e^2} \right) \right] \\
&= \frac{1}{2} \sum_i p_i \mathbb{E}_{H_e} \left[\log \left(1 + H_e^2 \frac{\gamma_i(h_{m,i})}{\sigma_e^2} \right) \right] - \frac{1}{2n} \sum_i \varepsilon_i \mathbb{E}_{H_e} \left[\log \left(1 + H_e^2 \frac{\gamma_i(h_{m,i})}{\sigma_e^2} \right) \right] \\
&\stackrel{7}{=} \frac{1}{2} \mathbb{E}_{H_e, H_m} \left[\log \left(1 + \frac{\gamma(H_m) H_e^2}{\sigma_e^2} \right) \right]
\end{aligned}$$

as $n \to \infty$ and $\epsilon \to 0$.

Justification.

4. We can split up the conditional density as

$$\begin{aligned}
\omega(z^{n_i}|x^{n_i}, h_m^{n_i}, h_e^{n_i}) &= \frac{\omega(z^{n_i}, x^{n_i}, h_e^{n_i}) \omega(h_m^{n_i})}{\omega(x^{n_i}, h_e^{n_i}) \omega(h_m^{n_i})} \\
&= \omega(z^{n_i}|x^{n_i}, h_e^{n_i})
\end{aligned}$$

where the first equality follows from the fact that $h_m^{n_i}$ is independent of z^{n_i}, x^{n_i}, and $h_e^{n_i}$. Note that $h_m^{n_i}$ was indeed used to determine which codebook to use on this channel, but at this point that has been determined and we have restricted the integration of $h_m^{n_i}$ to take this into

account, i.e. x^{n_i} is independent of $h_m^{n_i}$. Thus the multiplicand becomes:

$$\int_{\mathcal{J}} \omega(h_e^{n_i})\omega(h_m^{n_i}) \int_{\mathbb{R}^{n_i}} \max_{x^{n_i}\in\mathscr{C}_{n_i}^i} \omega_{\mathcal{T}^i}(z^{n_i}|x^{n_i}, h_e^{n_i}) dz^{n_i} dh_e^{n_i} dh_m^{n_i}$$
$$= \int_{\mathbb{R}_+^{n_i}} \omega(h_e^{n_i}) \int_{\mathbb{R}^{n_i}} \max_{x^{n_i}\in\mathscr{C}_{n_i}^i} \omega_{\mathcal{T}^i}(z^{n_i}|x^{n_i}, h_e^{n_i}) dz^{n_i} dh_e^{n_i}$$

Where the equality follows from the fact that we know $h_m \in [h_{m,i}, h_{m,i+1})$ for each component of the n_i length vector for every i. Therefore integrating $\omega(h_m^i)$ over the whole space where h_m is guaranteed to be will yield 1 for each of the $\sum_i n_i$ integrals. We then rewrite the integral over $h_e^{n_i}$ in the form of expected value.

5. *Definition of $\boldsymbol{\mathcal{I}}_{n_i}(A_{\mathcal{T}_{n_i}}^{n_i})$ and Proposition 2.*

6. *Upper bound as found in Theorem 1.*

7. *d can be made arbitrarily large and thus the channel coefficient intervals can be made arbitrarily small, hence the convergence of the first term to the expected value. For the second term:*

$$\lim_{n\to\infty} \frac{1}{2n} \sum_i \varepsilon_i \mathbb{E}_{H_e} \left[\log\left(1 + H_e^2 \frac{\gamma_i(h_{m,i})}{\sigma_e^2}\right)\right] = 0$$

Since $\mathbb{E}_{H_e}\left[\log\left(1 + H_e^2 \frac{\gamma_i(h_{m,i})}{\sigma_e^2}\right)\right]$ is constant with respect to n and $\varepsilon_i \to 0$. Also, $\delta_n, \delta_n', \delta_n''$ (from Theorem 1) can be chosen in such a way that $\epsilon \to 0$ as $n \to \infty$.

□

5.5. Set of Secure Rates

Now that we have our bound, ξ, we have completed step 1 of the procedure. Lemma 1 immediately tells us that by using the UHF based preprocessing scheme we can achieve any positive rate, R_s, with semantic security satisfying $R_s < R_{\mathcal{C}} - \xi$. Let's see how this compares to previous results.

Fact 5. *[4] For the fast fading wiretap channel where the CSI of the main channel but not the CSI of the eavesdropper channel is known at the transmitter, all rates R_s such that*

$$R_s < \max_{\gamma} \left(\frac{1}{2}\mathbb{E}_{H_m}\left[\log\left(1+\frac{\gamma(H_m)H_m^2}{\sigma_m^2}\right)\right] - \frac{1}{2}\mathbb{E}_{H_m H_e}\left[\log\left(1+\frac{\gamma(H_m)H_e^2}{\sigma_m^2}\right)\right]\right)$$

where $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ obeys the constraint $\mathbb{E}[\gamma(H_m)] \leq P$ are achievable secrecy rates under the variational distance secrecy metric.

Corollary 2. *The transmission scheme of Section 3.5 can achieve all rates given in Fact 5 with* ***semantic*** *security on the partial CSIT fast fading wiretap channel when the main channel rate is taken as*

$$R_{\mathcal{C}} = \frac{1}{2}\mathbb{E}_{H_m}\left[\log\left(1+\frac{\gamma(H_m)H_m^2}{\sigma_m^2}\right)\right]$$

for any power $\gamma(H_m)$.

Proof. The result follows immediately after combining Lemma 1 with Theorem 2 and noting that there does exist some ECC which can achieve this rate due to the fact that the above expression is the point-to-point capacity of the fast fading channel with power $\gamma(H_m)$. □

In summary, we have completed step 1 of the procedure of Chapter 3 for all fast fading channels with partial CSIT. With this result, we have shown which rates can be achieved with semantic security using the UHF based preprocessing scheme. In particular, Corollary 2 shows that we can achieve the *best known* secrecy rate with semantic security. Thus all a user has to do is concatenate the UHF based preprocessing scheme with their existing ECC system and they will achieve semantic security with rates given in the above corollary. If they wish to achieve the optimal known secure rate, they just need to find an optimal ECC for the point-to-point fast fading channel. The problem of security on the fast fading wiretap channel with partial CSIT has now been converted into the problem of finding an ECC.

6. FULL CSIT

In this subsection, we shall assume full CSIT; that is, we assume the transmitter knows instantaneously the realizations at time instance i of both the main and eavesdropper coefficients.

The assumption of reliable feedback in a timely manner from the legitimate receiver is once again a valid assumption due to different channel testing and training techniques. However, it may not be very plausible that an eavesdropper will be feeding back reliable information about her channel to the transmitter. Real-world examples where this could be a valid assumption include situations where a superior wishes to speak to another superior while keeping a subordinate employee ignorant of the message. In this case, the eavesdropper is not malicious and thus the technology could be set up in such a way that the subordinate's communication equipment will feedback to his superior's equipment. Pathological examples aside, the case of full CSIT still carries theoretical importance. The results of this chapter carry further importance in the sense that for the first time known to the author, we have determined the *semantic secrecy capacity* of the fast fading wiretap channel with full CSIT. Not only have we determined it, but we also provide an explicit scheme by which one may achieve it!

To find an achievable rate with semantic security for this channel, we follow an almost identical power allocation scheme as presented in Chapter 5 and similar to that in [4, 9] and derive a bound for $\frac{\mathcal{I}_n^\epsilon}{n}$ of this channel. Due to the similarities with that of Chapter 5, we omit redundant explanations here. In short, the strategy is the same as that of Chapter 5 except instead of decomposing the wiretap channel into parallel channels according to the main channel coefficients, we decompose it according to *both* channel coefficients. This yields parallel *Gaussian* wiretap channels where both the main and eavesdropper channels are Gaussian wiretap channels instead of just the main being so as in Chapter 5.

Since full CSI is available at the transmitter, intended receiver, and eavesdropper, they each can demultiplex the fast-fading wiretap channel into d^2 parallel time-invariant *Gaussian* wiretap channels. This is accomplished by dividing the ranges of the channel coefficients into d intervals each. More specifically, we assume the fading gains are bounded and divide H_m into intervals $[h_{m,i}, h_{m,i+1})$ with $i \in [1, d]$, and H_e into intervals $[h_{e,j}, h_{e,j+1})$ with $j \in [1, d]$. Note that the ith

interval of the main channel coefficients does not have to contain the same values as the ith interval of the eavesdropper channel coefficients. Let

$$p_i = \mathbb{P}\left[H_m \in [h_{m,i}, h_{m,i+1})\right]$$

$$q_j = \mathbb{P}\left[H_e \in [h_{e,j}, h_{e,j+1})\right].$$

Each subchannel is composed of a main Gaussian channel with constant gain $h_{m,i}$ and eavesdropper channel with constant gain $h_{e,j+1}$. Again we take the lower bound of each interval to assume worst case for the main channel, and take the upper bound of each interval for the eavesdropper channel to assume best case scenario for the eavesdropper. In this way we are assuming worst case scenario for the legitimate parties which is the standard assumption in security.

Let $N_{i,j}$ be the number of times channel i, j is used. Let $n_{ij} = p_i q_j n - \varepsilon_{ij}$, where ε_{ij} is chosen sufficiently large such that the realization of $N_{i,j}$ is greater than n_{ij} with high probability and $\varepsilon_{ij} \to 0$ as $n \to \infty$. For every pair of indices (i, j), the transmitter and legitimate receiver will publicly agree on a transmit power $\gamma_{i,j}(H_m, H_e)$ where $\{\gamma_{i,j}\}_{d,d}$ is chosen such that

$$\sum_{i=1}^{d}\sum_{j=1}^{d} p_i q_j \gamma_{i,j} \leq P.$$

Furthermore, for each i, j, the transmitter and legitimate receiver publicly agree upon an ECC $\mathcal{C}_{n_{ij}}^{ij}$ (with codebook $\mathscr{C}_{n_{ij}}^{ij}$) designed to operate on the Gaussian point-to-point channel with constant channel gain $h_{m,i}$. We denote by R_{ij} the rate of $\mathcal{C}_{n_{ij}}^{ij}$. The full coding scheme is then outlined as follows: a message $m \in \mathcal{M}$ is chosen which passes through the preprocessing layer to produce an l-length pseudo-message $m' \in \mathcal{M}'$. These l bits are then divided into sets of $n_{ij}R_{ij}$ bits such that

$$l = \sum_{i,j} n_{ij} R_{ij}.$$

A codeword is then generated for each of these sets by their respective $\mathcal{C}_{n_{ij}}^{ij}$ and the multiplexing strategy outlined above is then employed to transmit the i, jth codeword when the channel state is in the i, jth interval.

We now wish to bound $\frac{\boldsymbol{\mathcal{I}}_n^\epsilon}{n}$ of this fast fading channel by considering the set of parallel Gaussian wiretap channels outlined above and each of *their* associated max-information terms for which we already know the bound as given in [25].

Let

$$\boldsymbol{\mathcal{I}}_{n_{ij}}^{\text{AWGN}}(A_{\mathcal{T}}^{n_{ij}}) = \log \int_{\mathcal{Z}^{n_{ij}}} \max_{x^{n_{ij}} \in \mathscr{C}_{n_{ij}}} \omega_{\mathcal{T}}(z^{n_{ij}}|x^{n_{ij}}) dz^{n_{ij}}$$

where $\mathcal{T}$ here is the typical set for the Gaussian channel given in [25]. In other words, $\boldsymbol{\mathcal{I}}_{n_{ij}}^{\text{AWGN}}$ is the max-information of the Gaussian channel over n_{ij} channel uses for the codebook $\mathscr{C}_{n_{ij}}$. This definition is equivalent to the definition of max-information given in Chapter 3 if the channel coefficients H_m and H_e are taken to be independent of the input X^n and output Z^n. Furthermore, if the eavesdropper channel A is a constant gain Gaussian channel then $\mathcal{T} = \mathcal{T}_{n_{ij}}^1 \cap \mathcal{T}_{n_{ij}}^2$ where $\mathcal{T}_{n_{ij}}^1$ and $\mathcal{T}_{n_{ij}}^2$ are given in Section 4.3.

Similar to Chapter 5, we wish to construct a typical set for the wiretap channel as a whole. To do so, we construct it out of the typical sets for each of the d^2 parallel wiretap channels.

Definition. *Define the following sets:*

$$\begin{aligned}
{}^\star\mathcal{T}_{n_{ij}}^1 &= \{(x^{n_{ij}}, h_m^{n_{ij}}, h_e^{n_{ij}}, z^{n_{ij}}) : h_m^{n_{ij}} \in \mathbb{R}_+^{n_{ij}}, (x^{n_{ij}}, h_e^{n_{ij}}, z^{n_{ij}}) \in \mathcal{T}_{n_{ij}}^1\} \\
{}^\star\mathcal{T}_{n_{ij}}^2 &= \{(x^{n_{ij}}, h_m^{n_{ij}}, h_e^{n_{ij}}, z^{n_{ij}}) : h_m^{n_{ij}} \in \mathbb{R}_+^{n_{ij}}, (x^{n_{ij}}, h_e^{n_{ij}}, z^{n_{ij}}) \in \mathcal{T}_{n_{ij}}^2\} \\
\mathcal{T}'_{n_{ij}} &= {}^\star\mathcal{T}_{n_{ij}}^1 \cap {}^\star\mathcal{T}_{n_{ij}}^2 \\
\mathcal{T}'_n &= \bigotimes_{i,j} \mathcal{T}'_{n_{ij}}
\end{aligned}$$

where $\mathcal{T}_{n_{ij}}^1$ and $\mathcal{T}_{n_{ij}}^2$ are defined in Section 4.3.

Lemma 5. *$\mathcal{T}'_n$ as defined above is a $(1-\epsilon_n)$ typical set.*

Proof. We see in the proof of Proposition 3 that ${}^\star\mathcal{T}_{n_{ij}}^1$ and ${}^\star\mathcal{T}_{n_{ij}}^2$ are directly $(1-\epsilon_n^1)$ and $(1-\epsilon_n^2)$ typical respectively, where ϵ_n^1 and ϵ_n^2 are given in Lemma 3. Now consider:

$$\begin{aligned}
&\mathbb{P}\left[(X^{n_{ij}}, H_m^{n_{ij}}, H_e^{n_{ij}}, Z^{n_{ij}}) \in \mathcal{T}'_{n_{ij}} | X^{n_{ij}} = x^{n_{ij}}\right] \\
&= \mathbb{P}\left[(X^{n_{ij}}, H_m^{n_{ij}}, H_e^{n_{ij}}, Z^{n_{ij}}) \in {}^\star\mathcal{T}_{n_{ij}}^1 \cap {}^\star\mathcal{T}_{n_{ij}}^2 | X^{n_{ij}} = x^{n_{ij}}\right]
\end{aligned}$$

$$
\begin{aligned}
&\overset{1}{\geq} \mathbb{P}\left[(X^{n_{ij}}, H_m^{n_{ij}}, H_e^{n_{ij}}, Z^{n_{ij}}) \in {}^{\star}\mathcal{T}_{n_{ij}}^1 | X^{n_{ij}} = x^{n_{ij}}\right] + \cdots \\
&\cdots + \mathbb{P}\left[(X^{n_{ij}}, H_m^{n_{ij}}, H_e^{n_{ij}}, Z^{n_{ij}}) \in {}^{\star}\mathcal{T}_{n_{ij}}^2 | X^{n_{ij}} = x^{n_{ij}}\right] - 1 \\
&\geq (1 - \epsilon_{n_{ij}}^1) + (1 - \epsilon_{n_{ij}}^2) - 1 \\
&= 1 - (\epsilon_{n_{ij}}^1 + \epsilon_{n_{ij}}^2)
\end{aligned}
$$

Therefore $\mathcal{T}'_{n_{ij}}$ is a $(1 - \epsilon_{n_{ij}})$ typical set. The proof of $\mathcal{T}'_n$ being typical then follows exactly as that of the proof of Lemma 4 and is thus omitted here.

Justification.

1. *Fréchet inequality for intersections.*

□

With the typical set $\mathcal{T}'_n$ now in hand, we proceed to accomplish step one of the procedure given in Chapter 3 by finding an asymptotic upper bound to max-information for this channel. We do so by splitting the channel up into the d^2 parallel wiretap channels as described above and find a bound for max-information on each of those individual channels.

Theorem 3. *Consider the fast fading wiretap channel with full CSIT at the transmitter with $\mathcal{T}'_n$ and ϵ_n be as defined in Lemma 5. Using power allocation scheme above, it follows that (dropping the subscript on ϵ):*

$$
\lim_{\substack{n\to\infty \\ \epsilon\to 0}} \left(\frac{\boldsymbol{\mathcal{I}}_n^{\epsilon}}{n}\right) \leq \frac{1}{2}\mathbb{E}_{H_e,H_m}\left[\log\left(1 + \frac{\gamma(H_m, H_e)H_e^2}{\sigma_e^2}\right)\right].
$$

Proof.

$$
\begin{aligned}
&2^{\boldsymbol{\mathcal{I}}_n^{\epsilon}} \overset{1}{\leq} 2^{\boldsymbol{\mathcal{I}}_n(A^n_{\mathcal{T}'_n})} \\
&= \mathbb{E}_{H_m^n H_e^n} \int_{\mathbb{R}^n} \max_{x^n \in \mathscr{C}_n} \omega_{\mathcal{T}'_n}(z^n | x^n, H_m^n, H_e^n) dz^n \\
&= \mathbb{E}_{H_m^n H_e^n} \left[\int_{\mathbb{R}^n} \max_{x^n \in \mathscr{C}_n} \omega(z^n | x^n, H_m^n, H_e^n) \mathbb{1}((x^n, H_m^n, H_e^n, z^n) \in \mathcal{T}'_n) dz^n\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{\mathbb{R}_+^n} \omega(h_m^n, h_e^n) \int_{\mathbb{R}^n} \max_{x^n \in \mathscr{C}_n} \omega(z^n | x^n, h_m^n, h_e^n) \mathbb{1}((x^n, h_m^n, h_e^n, z^n) \in \mathcal{T}_n') dz^n dh_m^n dh_e^n \\
&\overset{2}{\leq} \prod_{i,j} \int_{\mathbb{R}_+^{n_{ij}}} \omega(h_m^{n_{ij}}, h_e^{n_{ij}}) \int_{\mathbb{R}^{n_{ij}}} \max_{x^{n_{ij}} \in \mathscr{C}_{n_{ij}}^{i,j}} \omega(z^{n_{ij}} | x^{n_{ij}}, h_m^{n_{ij}}, h_e^{n_{ij}}) \cdots \\
&\qquad \cdots \mathbb{1}((x^{n_{ij}}, h_m^{n_{ij}}, h_e^{n_{ij}}, z^{n_{ij}}) \in \mathcal{T}_{n_{ij}}') dz^{n_{ij}} dh_m^{n_{ij}} dh_e^{n_{ij}}
\end{aligned}
$$

Let $\mathcal{J} = [h_{m,i}, h_{m,i+1}\ )^{n_{ij}} \times [h_{e,j}, h_{e,j+1}\ )^{n_{ij}}$

$$
\begin{aligned}
&\overset{3}{=} \prod_{i,j} \int_{\mathcal{J}} \omega(h_m^{n_{ij}}, h_e^{n_{ij}}) \int_{\mathbb{R}^{n_{ij}}} \max_{x^{n_{ij}} \in \mathscr{C}_{n_{ij}}^{i,j}} \omega_{\mathcal{T}_{n_{ij}}'}(z^{n_{ij}} | x^{n_{ij}}, h_m^{n_{ij}}, h_e^{n_{ij}}) dz^{n_{ij}} dh_m^{n_{ij}} dh_e^{n_{ij}} \\
&\overset{4}{=} \prod_{i,j} \int_{\mathcal{J}} \omega(h_m^{n_{ij}}, h_e^{n_{ij}}) \int_{\mathbb{R}^{n_{ij}}} \max_{x^{n_{ij}} \in \mathscr{C}_{n_{ij}}^{i,j}} \omega(z^{n_{ij}} | x^{n_{ij}}, h_e^{n_{ij}}) \cdots \\
&\qquad \cdots \mathbb{1}((x^{n_{ij}}, h_e^{n_{ij}}, z^{n_{ij}}) \in \mathcal{T}_{n_{ij}}^1 \cap \mathcal{T}_{n_{ij}}^2) dz^{n_{ij}} dh_m^{n_{ij}} dh_e^{n_{ij}} \\
&\overset{5}{=} \prod_{i,j} \int_{\mathcal{J}} \omega(h_m^{n_{ij}}, h_e^{n_{ij}}) \int_{\mathbb{R}^{n_{ij}}} \max_{\bar{x}^{n_{ij}} \in h_e^{n_{ij}} \mathscr{C}_{n_{ij}}^{i,j}} \omega(z^{n_{ij}} | \bar{x}^{n_{ij}}) \cdots \\
&\qquad \cdots \mathbb{1}((x^{n_{ij}}, h_e^{n_{ij}}, z^{n_{ij}}) \in \mathcal{T}_{n_{ij}}^1 \cap \mathcal{T}_{n_{ij}}^2) dz^{n_{ij}} dh_m^{n_{ij}} dh_e^{n_{ij}} \\
&\overset{6}{=} \prod_{i,j} \int_{\mathcal{J}} \omega(h_m^{n_{ij}}, h_m^{n_{ij}}) 2^{\boldsymbol{\mathcal{I}}_{n_{ij}}^{\text{AWGN}}\left(A_{\mathcal{T}_{n_{ij}}^{1,2}}^{n_{ij}}\right)} dh_m^{n_{ij}} dh_e^{n_{ij}}
\end{aligned} \tag{E2}
$$

Justification.

1. *$\mathcal{T}_n'$ is a $(1-\epsilon_n)$ typical set; however, it may not be the set corresponding to the "smallest" ϵ smooth max-information.*

2. *See justification (2) given in the proof of Theorem 2.*

3. *The n length transmission was split into n_{ij} length constituent parts based on the values of h_e and h_m at a given moment. In other words, during each entire n_{ij} length transmission, we know that $h_e \in [h_{e,j}, h_{e,j+1})$ and $h_m \in [h_{m,i}, h_{m,i+1})$ since the coding schemed determined n_{ij} that way, thus we can rewrite the bounds of integration for $h_m^{n_{ij}}$ and $h_e^{n_{ij}}$ to consider only this space and not the entire space as that is unneeded.*

4.

$$\begin{aligned}\omega(z^{n_{ij}}|x^{n_{ij}}, h_m^{n_{ij}}, h_e^{n_{ij}}) &= \frac{\omega(z^{n_{ij}}x^{n_{ij}}, h_e^{n_{ij}})\omega(h_m^{n_{ij}})}{\omega(x^{n_{ij}}, h_e^{n_{ij}})\omega(h_m^{n_{ij}})} \\ &= \omega(z^{n_{ij}}|x^{n_{ij}}, h_e^{n_{ij}})\end{aligned}$$

where the first equality follows from the fact that $h_m^{n_{ij}}$ is independent of $z^{n_{ij}}, x^{n_{ij}}$, and $h_e^{n_{ij}}$. Note that $h_m^{n_{ij}}$ was indeed used to determine which codebook to use on this channel, but at this point that has been determined and we have restricted the integration of $h_m^{n_{ij}}$ to take this into account, i.e. $x^{n_{ij}}$ is independent of $h_m^{n_{ij}}$. Finally, since the maximization no longer has any dependence on h_m, we then only need to consider tuples $(x^{n_{ij}}, h_e^{n_{ij}}, z^{n_{ij}}) \in \mathcal{T}_{n_{ij}}^1 \cap \mathcal{T}_{n_{ij}}^2$.

5. *Let $\bar{x}^{n_{ij}} = h_e^{n_{ij}} x^{n_{ij}}$. Then the codebook of $\bar{x}^{n_{ij}}$ becomes $h_e^{n_{ij}}\mathscr{C}_{n_{ij}}^{i,j}$, i.e. the original codebook of $x^{n_{ij}}$, $\mathscr{C}_{n_{ij}}^{i,j}$ where each codeword is multiplied by $h_e^{n_{ij}}$.*

6. *Definition of $\boldsymbol{\mathcal{I}}_{n_{ij}}^{AWGN}$ for a constant weight Gaussian channel.*

   We know that the power constraint of codebook $\mathscr{C}_{n_{ij}}^{i,j}$ is $\gamma_{i,j}$, thus:

$$\frac{1}{n_{ij}} \sum_{\iota=1}^{n_{ij}} x_\iota^2 \leq \gamma_{i,j}.$$

Therefore the power constraint of codebook $h_e^{n_{ij}}\mathscr{C}_{n_{ij}}^{i,j}$ is given by:

$$\frac{1}{n_{ij}} \sum_{\iota=1}^{n_{ij}} {h_{e\iota}}^2 x_\iota^2 \leq \gamma'_{i,j}.$$

for some finite $\gamma'_{i,j}$. From [25] we know that $\boldsymbol{\mathcal{I}}_{n_{ij}}^{\text{AWGN}}\left(A_{\mathcal{T}_{n_{ij}}^{1,2}}^{n_{ij}}\right)$ corresponding to the codebook $h_e^{n_{ij}}\mathscr{C}_{n_{ij}}^{i,j}$ has an upper bound given as:

$$\boldsymbol{\mathcal{I}}_{n_{ij}}^{\text{AWGN}}\left(A_{\mathcal{T}_{n_{ij}}^{1,2}}^{n_{ij}}\right) \leq \frac{n_{ij}}{2}\log\left(1 + \frac{\gamma'_{i,j}}{\sigma_e^2}\right) + n_{ij}\delta\log e + o(n_{ij})$$

for any sufficiently small $\delta > 0$. We can then upper bound each component of $h_e^{n_{ij}}$ by the largest value of the interval containing each component, that being the constant $h_{e,j+1}$. This produces a

new codebook $h_{e,j+1}^{n_{ij}} \mathscr{C}_{n_{ij}}^{i,j}$ having power constraint

$$\frac{1}{n_{ij}} \sum_{\iota=1}^{n_{ij}} {h_{e,j+1}}_{\iota}^{2} x_{\iota}^{2} = \frac{{h_{e,j+1}}^{2}}{n_{ij}} \sum_{\iota=1}^{n_{ij}} x_{\iota}^{2} \leq {h_{e,j+1}}^{2} \gamma_{i,j}.$$

Combining, we see that

$$\begin{aligned}
\boldsymbol{\mathcal{I}}_{n_{ij}}^{\text{AWGN}} \left(A_{\mathcal{T}_{n_{ij}}^{1,2}}^{n_{ij}} \right) &\leq \frac{n_{ij}}{2} \log \left(1 + \frac{\gamma'_{i,j}}{\sigma_e^2} \right) + n_{ij} \delta \log e + o(n_{ij}) \\
&\leq \frac{n_{ij}}{2} \log \left(1 + \frac{{h_{e,j+1}}^{2} \gamma_{i,j}}{\sigma_e^2} \right) + n_{ij} \delta \log e + o(n_{ij}) \\
&= n_{ij} \beta_{n_{ij}}
\end{aligned}$$

Continuing on from $(E2)$:

$$\begin{aligned}
(\text{E2}) &\leq \prod_{i,j} \int_{\mathcal{J}} \omega(h_m^{n_{ij}}, h_e^{n_{ij}}) 2^{n_{ij} \beta_{n_{ij}}} dh_m^{n_{ij}} dh_e^{n_{ij}} \\
&= \prod_{i,j} 2^{n_{ij} \beta_{n_{ij}}} \int_{\mathcal{J}} \omega(h_m^{n_{ij}}, h_e^{n_{ij}}) dh_m^{n_{ij}} dh_e^{n_{ij}} \\
&\overset{7}{=} \prod_{i,j} 2^{n_{ij} \beta_{n_{ij}}}
\end{aligned}$$

Taking the logarithm of each side and dividing by n we have:

$$\begin{aligned}
\frac{\boldsymbol{\mathcal{I}}_n^{\epsilon}}{n} &\leq \frac{1}{n} \log \left(\prod_{i,j} 2^{n_{ij} \beta_{n_{ij}}} \right) \\
&= \frac{1}{n} \sum_{i,j} \log \left(2^{n_{ij} \beta_{n_{ij}}} \right) \\
&= \frac{1}{n} \sum_{i,j} n_{ij} \beta_{n_{ij}} \\
&= \frac{1}{n} \sum_{i,j} (p_i q_j n - \varepsilon_n) \beta_{n_{ij}} \\
&= \sum_{i,j} p_i q_j \beta_{n_{ij}} - \frac{\varepsilon_n}{n} \sum_{i,j} \beta_{n_{ij}} \\
&\overset{8}{=} \frac{1}{2} \mathbb{E}_{H_e H_m} \left[\log \left(1 + \frac{\gamma(H_m, H_e) H_e^2}{\sigma_e^2} \right) \right]
\end{aligned}$$

as $n \to \infty$ and $\epsilon \to 0$.

Justification.

7. *For each of the n_{ij} components, we know that h_m and h_e will surely be found in $\mathcal{J}$ due to our choice of coding scheme.*

8. *d can be chosen arbitrarily large thus making each channel coefficient interval arbitrarily small, hence the convergence of the first term to the expected value. The second term goes to 0 with n as the summand only grows with $o(n_{ij})$ while $n \to \infty$ and $\varepsilon_n \to 0$. Also, δ (from the bound taken from [25]) can be chosen in such a way that $\epsilon \to 0$ as $n \to \infty$.*

□

Now that we have completed step one of the procedure given in Chapter 3, let's see what semantically secure rates we can achieve with this UHF scheme and compare this with previous work.

Fact 6. *[15] With full CSI for both the main channel and the eavesdropper channels available at the transmitter, the weak secrecy capacity of the ergodic fading channel is:*

$$C_s = \max_{\gamma} \left(\frac{1}{2}\mathbb{E}_{H_m H_e}\left[\log\left(1 + \frac{\gamma(H_m, H_e)H_m^2}{\sigma_m^2}\right)\right] - \frac{1}{2}\mathbb{E}_{H_m H_e}\left[\log\left(1 + \frac{\gamma(H_m, H_e)H_e^2}{\sigma_e^2}\right)\right]\right)$$

where $\gamma : \mathbb{R}_+^2 \to \mathbb{R}_+$ obeys the power constraint $\mathbb{E}\left[\gamma(H_m, H_e)\right] \leq P$.

Corollary 3. *The **semantic** secrecy capacity of the fading channel with Full CSIT is given by:*

$$C_s = \max_{\gamma} \left(\frac{1}{2}\mathbb{E}_{H_m H_e}\left[\log\left(1 + \frac{\gamma(H_m, H_e)H_m^2}{\sigma_m^2}\right)\right] - \frac{1}{2}\mathbb{E}_{H_m H_e}\left[\log\left(1 + \frac{\gamma(H_m, H_e)H_e^2}{\sigma_e^2}\right)\right]\right)$$

Furthermore, the transmission scheme of Section 3.5 can achieve the semantic secrecy capacity of the fast fading wiretap channel with full CSIT.

Proof. Let γ^* be the power allocation function that maximizes the expression in Fact 6 as found in [15]. Let the main channel rate be

$$R_{\mathcal{C}} = \frac{1}{2}\mathbb{E}_{H_m H_e}\left[\log\left(1 + \frac{\gamma^*(H_m, H_e)H_m^2}{\sigma_m^2}\right)\right]$$

We know by Shannon's noisy channel coding theorem that *some* ECC will exist which satisfies this rate due to the above expression being the point-to-point capacity of the fast fading channel under power allocation function γ^*. Since the bound found in Theorem 3 holds for any power allocation function γ, it holds for γ^* in particular. In Theorem 3 we found an upper bound to the right hand term of the difference in Fact 6, thus invoking Lemma 1 we know we can achieve any rate arbitrarily close to the secrecy capacity given in Fact 6. Therefore the semantic secrecy capacity is equal to the weak secrecy capacity by Fact 2 in the case of full CSIT and the given scheme achieves it. $\square$

In summary, we have shown that with the UHF preprocessing scheme provided in [13], we can achieve all rates up to and including the weak secrecy capacity which implies that the semantic secrecy capacity is also given by the expression for weak secrecy. To the extent of the author's knowledge, this is the first time the semantic secrecy capacity has been characterized for this fading channel. Doing so in an explicit manner gives communication system designers all the tools they may ever need to ensure the best semantically secure rate possible in their system.

7. CONCLUSION

In the preceding chapters, we have determined rates which can be achieved on a fast fading wiretap channel with semantic security. In particular, these can be achieved by an explicit and efficient UHF based preprocessing scheme which can be concatenated with any error correcting code. This differs substantially from the majority of current literature which has primarily aimed at designing "all-in-one" wiretap codes. In other words, they have focused on coming up with a scheme that provides reliability and security all at once. The scheme presented in [13] which we use here, separates the wiretap code into two distinct components - one to provide security and one to provide reliability. We have found the security component for the fast fading wiretap channels given in this thesis; therefore, all that remains is to find an explicit and efficient error correcting code for the main channel in question. Put another way, we have converted the problem of finding a good wiretap code into one of just finding a good error correcting code which is already an entire field of study in its own right and has considerable attention and resources devoted to it.

To date, in the case of S-CSIT, a general characterization of the secrecy capacity has not been determined under any secrecy metric. The weak secrecy capacity has indeed been characterized for certain subsets of the S-CSIT case which this thesis has now determined to be equivalent to the semantic secrecy capacity. However, a general expression for any fast fading channel with S-CSIT has eluded us. It could be the case that there exist fast fading channels with S-CSIT which do not fall into one of those special subsets and thus may have a different secrecy capacity. All we can say due to the work in this thesis is that on those channels, we can still achieve rates up to $C_T - C_A$ with semantic security and thus the secrecy capacity for those channels must be lower bounded by that value. It is an interesting and important line of future work to characterize the secrecy capacity of the fast fading channel with S-CSIT in general.

Although in the case of partial CSIT, we have not determined the secrecy capacity, we have shown that the best known rates achievable under the variational distance metric can also be achieved under the semantic security metric as well using the power allocation scheme described in Chapter 5 with the UHF based preprocessing layer. It is again of considerable future interest to characterize the actual secrecy capacity in the case of partial CSIT under any metric. In [4] a very

similar power allocation scheme is presented to achieve that previously best known rate (under a weaker secrecy metric); however, it differs from the scheme presented here at a crucial point. In [4], they demultiplex the channel as we do, after which they proceed to design *d wiretap codes* to operate on each of the parallel wiretap channels. This is in contrast to our scheme which requires the design of *d error correcting codes* which are connected together to only 1 preprocessor. Thus once again, we have converted the problem of designing wiretap codes to that of designing error correcting codes which is significantly easier.

Similarly, in the case of full CSIT, we have once again converted the problem of designing wiretap codes into that of designing good error correcting codes. However, in this case, the expression for the weak secrecy capacity of any fast fading channel with full CSIT had already been determined and we show that we can also achieve that rate with semantic security, therefore determining the semantic secrecy capacity for the first time.

The UHF based preprocessing scheme used here has also been shown to achieve the semantic secrecy capacity of both the Gaussian wiretap channel and any discrete memoryless channel [13]. By further providing an explicit means of achieving semantically secure rates for the first time on *any* fast fading channel, the author is confident that UHF based preprocessing schemes are one of the best candidates for implementing information-theoretic security in our modern communication systems.

REFERENCES

[1] J. Barros and M. R. D. Rodrigues. Secrecy capacity of wireless channels. In *2006 IEEE International Symposium on Information Theory*, pages 356–360, July 2006.

[2] Mihir Bellare, Stefano Tessaro, and Alexander Vardy. A cryptographic treatment of the wiretap channel. *CoRR*, abs/1201.2205, 2012.

[3] M. Bloch and J. Laneman. Information-spectrum methods for information-theoretic security. *2009 Information Theory and Applications Workshop*, 2009.

[4] Matthieu Bloch and Joao Barros. *Physical-layer security: from information theory to security engineering*. Cambridge University Press, 2011.

[5] Matthieu R. Bloch and J. Nicholas Laneman. Strong secrecy from channel resolvability. *IEEE Transactions on Information Theory*, 59(12):8077–8098, 2013.

[6] J.Lawrence Carter and Mark N. Wegman. Universal classes of hash functions. *Journal of Computer and System Sciences*, 18(2):143 – 154, 1979.

[7] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience, 2006.

[8] Maurice Fréchet. Généralisation du théorème des probabilités totales. *Fundamenta Mathematicae*, 25(1):379–387, 1935.

[9] A.j. Goldsmith and P.p. Varaiya. Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory*, 43(6):1986–1992, 1997.

[10] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *Journal of Computer and System Sciences*, 28(2):270–299, 1984.

[11] P. K. Gopala, L. Lai, and H. El Gamal. On the secrecy capacity of fading channels. In *2007 IEEE International Symposium on Information Theory*, pages 1306–1310, June 2007.

[12] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[13] Eric Kubischta, Parker Pavlicek, and Sanjay Karmakar. Universal hashing for semantically secure wiretap communication with application to fading channels (preprint). *IEEE Transactions on Information Theory*, 2018.

[14] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, 28(5):1302–1338, 10 2000.

[15] Yingbin Liang and H. Vincent Poor. Secure communications over fading channels. *CoRR*, abs/0708.2733, 2007.

[16] P. H. Lin and E. Jorswieck. On the fast fading gaussian wiretap channel with statistical channel state information at the transmitter. *IEEE Transactions on Information Forensics and Security*, 11(1):46–58, Jan 2016.

[17] C. Ling, L. Luzzi, J. C. Belfiore, and D. Stehlé. Semantically secure lattice codes for the gaussian wiretap channel. *IEEE Transactions on Information Theory*, 60(10):6399–6416, Oct 2014.

[18] F. Miller. *Telegraphic Code to Insure Privacy and Secrecy in the Transmission of Telegrams.* C.M. Cornwell, 1882.

[19] Pritam Mukherjee and Sennur Ulukus. Fading wiretap channel with no csi anywhere. *2013 IEEE International Symposium on Information Theory*, 2013.

[20] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[21] C. E. Shannon. Communication theory of secrecy systems*. *Bell System Technical Journal*, 28(4):656–715, 1949.

[22] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Review*, 41(2):303–332, 1999.

[23] Soo Hak Sung, Patchanok Srisuradetchai, and Andrei Volodin. A note on the exponential inequality for a class of dependent random variables. *Journal of the Korean Statistical Society*, 40(1):109–114, 2011.

[24] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge, 2013.

[25] H. Tyagi and A. Vardy. Universal hashing for information-theoretic security. *Proceedings of the IEEE*, 103(10):1781–1795, Oct 2015.

[26] A. D. Wyner. The wire-tap channel. *Bell System Technical Journal*, 54(8):1355–1387, 1975.

APPENDIX

A.1 Proof of Proposition 2

Proof. From the definition of $\boldsymbol{\mathcal{I}}_n(A_{\mathcal{T}}^n)$ we have:

$$
\begin{aligned}
2^{\boldsymbol{\mathcal{I}}_n(A_{\mathcal{T}}^n)} &= \mathbb{E}_{H_m^n H_e^n} \int\limits_{\mathbb{R}^n} \max_{x^n \in \mathscr{C}_n} \omega_{\mathcal{T}}(z^n | x^n, H_m^n H_e^n) dz^n \\
&= \int\limits_{\mathbb{R}_+^n} \int\limits_{\mathbb{R}_+^n} \omega(h_e^n, h_m^n) \int\limits_{\mathbb{R}^n} \max_{x^n \in \mathscr{C}_n} \omega_{\mathcal{T}}(z^n | x^n, h_e^n, h_m^n) dz^n dh_m^n dh_e^n \\
&\overset{1}{=} \int\limits_{\mathbb{R}_+^n} \omega(h_e^n) \int\limits_{\mathbb{R}_+^n} \omega(h_m^n) \int\limits_{\mathbb{R}^n} \max_{x^n \in \mathscr{C}_n} \omega_{\mathcal{T}}(z^n | x^n, h_e^n) dz^n dh_m^n dh_e^n \\
&= \int\limits_{\mathbb{R}_+^n} \omega(h_e^n) \int\limits_{\mathbb{R}^n} \max_{x^n \in \mathscr{C}_n} \omega_{\mathcal{T}}(z^n | x^n, h_e^n) dz^n dh_e^n \\
&= \mathbb{E}_{H_e^n} \int\limits_{\mathbb{R}^n} \max_{x^n \in \mathscr{C}_n} \omega_{\mathcal{T}}(z^n | m', H_e^n) dz^n.
\end{aligned}
$$

Justification.

1. *Independence of H_e^n and H_m^n. Also, $Z_i = H_{e,i} X_i + U_i$ and X_i is not a function of the channel coefficients since we have S-CSIT; therefore, Z^n is independent of H_m^n.*

□

A.2 Proof of Lemma 3

We need the following lemma for technical reasons.

Lemma 6.

$$
\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \frac{Z_i^2}{\sigma_e^2 + H_{e,i}^2 P} \,\middle|\, X^n = x^n \right] \leq 1.
$$

Proof.

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \frac{Z_i^2}{\sigma_e^2 + H_{e,i}^2 P} \,\middle|\, X^n = x^n \right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\frac{H_{e,i}^2 x_i^2 + U_{e,i}^2 + 2 H_{e,i} x_i U_{e,i}}{\sigma_e^2 + H_{e,i}^2 P} \right]
\end{aligned}
$$

$$
\begin{aligned}
&\overset{1}{=} \frac{1}{n}\sum_{i=1}^{n} x_i^2 \cdot \mathbb{E}\left[\frac{H_{e,i}^2}{\sigma_e^2 + H_{e,i}^2 P}\right] + \mathbb{E}U_{e,i}^2 \cdot \mathbb{E}\left[\frac{1}{\sigma_e^2 + H_{e,i}^2 P}\right] + \mathbb{E}U_{e,i} \cdot \mathbb{E}\left[\frac{2x_i^2 H_{e,i}}{\sigma_e^2 + H_{e,i}^2 P}\right] \\
&\overset{2}{=} \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) \cdot \mathbb{E}\left[\frac{H_e^2}{\sigma_e^2 + H_e^2 P}\right] + \mathbb{E}\left[\frac{\sigma_e^2}{\sigma_e^2 + H_e^2 P}\right] \\
&\overset{3}{\leq} \mathbb{E}\left[\frac{H_e^2 P}{\sigma_e^2 + H_e^2 P}\right] + \mathbb{E}\left[\frac{\sigma_e^2}{\sigma_e^2 + H_e^2 P}\right] \\
&= 1
\end{aligned}
$$

Justification.

1) Follows from independence of H_e, U_e.

2) U_e is i.i.d. and $\sim \mathcal{N}(0, \sigma_e^2)$.

3) Follows from the power constraint on all codewords.

□

Proof of Lemma 3

1. *Proof.* Let $\mu = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{Z_i^2}{\sigma_e^2 + H_{e,i}^2 P} \,\middle|\, X^n = x^n\right]$. Then,

$$
\begin{aligned}
&\mathbb{P}\left[(H_e^n, Z^n) \in \mathcal{P}_n^1 \,\middle|\, X^n = x^n\right] \\
&= \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{Z_i^2}{\sigma_e^2 + H_i^2 P} - 1 \leq \delta_n \,\middle|\, X^n = x^n\right] \\
&\geq \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{Z_i^2}{\sigma_e^2 + H_{e,i}^2 P} - \mu \leq \delta_n \,\middle|\, X^n = x^n\right] \\
&= \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{x_i^2 H_{e,i}^2 + U_i^2 + 2x_i H_{e,i} U_i}{\sigma_e^2 + H_{e,i}^2 P} - \mu \leq \delta_n\right], \qquad \text{(E3)}
\end{aligned}
$$

where the inequality follows from Lemma 6.

Since x^n is a constant and $\{H_{e,i}\}$ and $\{U_{e,i}\}$ are each mutually independent, the term $\frac{x_i^2 H_{e,i}^2 + U_i^2 + 2x_i H_{e,i} U_i}{\sigma_e^2 + H_{e,i}^2 P}$ is an independent random variable. Let us show that it also satisfies the main condition of Lemma 2 (dropping the subscript e of $H_{e,i}$ and σ_e^2 to remove clutter).

$$
\begin{aligned}
&\mathbb{E}\left[e^{\gamma\frac{x_i^2H_i^2+U_i^2+2x_iH_iU_i}{\sigma^2+H_i^2P}}\right]\\
&=\mathbb{E}\left[e^{\gamma\left(\frac{x_i^2H_i^2}{\sigma^2+H_i^2P}+\frac{U_i^2}{\sigma^2+H_i^2P}+\frac{2x_iH_iU_i}{\sigma^2+H_i^2P}\right)}\mathbb{1}\left(H_i>1\right)\right]+\mathbb{E}\left[e^{\gamma\left(\frac{x_i^2H_i^2}{\sigma^2+H_i^2P}+\frac{U_i^2}{\sigma^2+H_i^2P}+\frac{2x_iH_iU_i}{\sigma^2+H_i^2P}\right)}\mathbb{1}\left(H_i\leq 1\right)\right]\\
&\overset{1}{\leq}\mathbb{E}\left[e^{\gamma\left(\frac{x_i^2H_i^2}{H_i^2P}+\frac{U_i^2}{\sigma^2}+\frac{2x_iH_i^2U_i}{H_i^2P}\right)}\mathbb{1}\left(H_i>1\right)\right]+\mathbb{E}\left[e^{\gamma\left(\frac{x_i^2H_i^2}{H_i^2P}+\frac{U_i^2}{\sigma^2}+\frac{2x_iU_i}{\sigma^2}\right)}\mathbb{1}\left(H_i\leq 1\right)\right]\\
&=\mathbb{E}\left[e^{\gamma\left(\frac{x_i^2}{P}+\frac{U_i^2}{\sigma^2}+\frac{2x_iU_i}{P}\right)}\right]+\mathbb{E}\left[e^{\gamma\left(\frac{x_i^2}{P}+\frac{U_i^2}{\sigma^2}+\frac{2x_iU_i}{\sigma^2}\right)}\right]\\
&\leq\mathbb{E}\left[e^{2\gamma\left(\frac{U_i}{\sigma}+x_i\frac{P+\sigma^2}{2P\sigma}\right)^2}\right]\\
&\overset{2}{\leq}\mathbb{E}\left[e^{2\gamma(G_i)^2}\right]\\
&\overset{3}{<}\infty.
\end{aligned}
$$

Justification.

1) $H_i > 1$ implies $H_i \leq H_i^2$.

2) $G_i \sim \mathcal{N}(x_i\frac{P+\sigma^2}{2P\sigma}$, 1) implies that G_i^2 is a non-central χ^2 random variable.

3) Choosing γ appropriately ensures the moment generating function is finite.

Since a finite moment generating function implies *every* moment is finite, K_i exists for all i so that K^* is well defined. Therefore, using Lemma 2, it follows immediately that

$$(E3) \geq 1 - 2e^{-\frac{n\delta_n^2}{4K^*}}.$$

□

2. *Proof.*

$$
\begin{aligned}
&\mathbb{P}\left[Z^n\in\mathcal{P}_n^2\,\middle|\,X^n=x^n,H_e^n=h^n\right]\\
&=\mathbb{P}\left[\|Z^n-x^nh^n\|^2\geq n\sigma_e^2(1-\delta_n')\middle|X^n=x^n,H_e^n=h^n\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{P}\left[\sum_{i=1}^{n}(Z_i - x_i h_i)^2 \geq n\sigma_e^2(1-\delta_n') \middle| X^n = x^n, H_e^n = h^n\right] \\
&= \mathbb{P}\left[\sum_{i=1}^{n}(U_i + x_i h_i - x_i h_i)^2 \geq n\sigma_e^2(1-\delta_n')\right] \\
&= \mathbb{P}\left[\frac{1}{\sigma_e^2}\sum_{i=1}^{n} U_i^2 \geq n(1-\delta_n')\right] \\
&\overset{1}{\geq} 1 - e^{-\frac{n{\delta_n'}^2}{4}}.
\end{aligned}
$$

Justification.

1) Chi-squared tail bounds [14, Lemma 1].

□

3. *Proof.* To prove this, we will use Lemma 2 reduced to the i.i.d. case. We have that $\{\log(1 + H_i^2\mathsf{SNR})\}$ is a sequence of i.i.d. random variables; to employ Lemma 2 it remains to prove that $\mathbb{E}\left[e^{\gamma|\log(1+H_e^2 SNR)|}\right] < \infty$ for some $\gamma > 0$.

$$
\begin{aligned}
\mathbb{E}\left[e^{\gamma|\log(1+H_e^2\mathsf{SNR})|}\right] &= \mathbb{E}\left[e^{\gamma\frac{\ln(1+H_e^2\mathsf{SNR})}{\ln(2)}}\right] \\
&= \mathbb{E}\left[(1+H_e^2\mathsf{SNR})^{\frac{\gamma}{\ln(2)}}\right] \\
&= \mathbb{E}\left[(1+H_e^2\mathsf{SNR})\right] \text{ (letting } \gamma = \ln 2) \\
&= 1 + \mathbb{E}[H_e^2]\mathsf{SNR} \\
&< \infty.
\end{aligned}
$$

Then Lemma 2 gives us:

$$
\begin{aligned}
&\mathbb{P}\left[H_e^n \in \mathcal{P}_n^3\right] \\
&= \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\log\left(1+H_i^2\mathsf{SNR}\right) - \mathbb{E}\left[1+H_e^2\mathsf{SNR}\right]\right| \leq \delta_n''\right] \\
&\geq 1 - 2e^{-\frac{n{\delta_n''}^2}{4K}},
\end{aligned}
$$

where

$$K = 2\left(\mathbb{E}\left[\log(1+H_e^2\mathsf{SNR})^4\right]\right)^{\frac{1}{2}}\mathbb{E}\left[e^{\gamma\log(1+H_e^2\mathsf{SNR})}\right].$$

□