Author:
Chakravarty, Jennifer J

Title:
Physical Layer Security for Next Generation Wireless Systems

Physical Layer Security for Next Generation Wireless Systems

Jennifer Joya Chakravarty

School of Mathematics
University of Bristol

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Engineering.

August 2020

Word count: 20309 words

Abstract

Using information-theoretic constructions, it is possible to characterise the security of a communication system. This is called physical layer security. The intrinsic randomness of the wireless channel allows for provable security guarantees in the presence of an eavesdropper.

As telecommunications requirements and technologies evolve, questions about point to point systems are re-framed in ways which have not yet been explored. In this thesis we analyse the robustness of particular future wireless technologies against eavesdropping at the physical layer.

In the first of the original research chapters the secrecy capacity of a Gaussian multiple antenna system is considered. Despite the importance of the secrecy capacity metric, the general solution remains an open problem. This thesis resolves the secrecy capacity to be concave in a particular region in the single antenna eavesdropper regime. This allows for efficient computation of the secrecy capacity and gives communication rates which are secure.

In the second research chapter, we analyse a multiple antenna, multiple access scheme. We show that the system is inherently secure, since the eavesdroppers signal-to-noise ratio decreases with the number of users, amongst other results.

The third research chapter introduces a novel channel coding scheme, combining constant weight arithmetic coding with an existing combinatorial scheme. The codewords are designed to be low-power and robust against time dispersion. This has the advantage that several users may broadcast messages simultaneously. The codebook design uses characteristics of the legitimate channel, which the eavesdropper does not have access to. Simulation results show that the eavesdropper has a low probability of success.

We conclude with a discussion of future work.

Acknowledgements

I would like to thank my supervisors Olly Johnson and Rob Piechocki for their expertise, guidance and patience. You have been generous with your time, wisdom and encouragement and I am very grateful to have worked with you for the past 4 years.

This work was kindly funded and supervised by GCHQ, and particular thanks go to Dene Hedges for giving his time and valuable suggestions on research directions. Thank you to Paul Thomas for the support and opportunities he has provided. I would also like to thank my examiners, Ganesh and Michael Chai for their thorough and valuable comments.

I am thankful to my school, to have been surrounded by intelligent and kind people. Thank you for all of the conversations in the coffee room, the variety of seminars and an incredible team of administrative staff.

The support and distractions provided by my friends has kept me going and I am so lucky to have so many inspirational and interesting people in my life. Especially to Tom for being by my side through a PhD and a pandemic, it's not what we had planned but I'm glad to be with you.

Thanks to mum and Karen for always being there. My nephew Charlie - the work in this document is the same age as you which makes me feel old but I love watching you grow up. Finally, my dad who always has an anecdote for every occasion and who gently supports me with unwavering enthusiasm in everything I do.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .... DATE: ....

Contents

List of Tables

List of Figures

Notation

- Upper case letters denote **random variables**, such as X.
- **Vectors** will be denoted as bold lower case symbols, such as $\mathbf{x}$.
- Entries of vectors are denoted x_i.
- **Matrices** are denoted by upper case letters, such as H.
- $H_{i,j}$ denotes the (i, j)th entry of a matrix H.
- It will be clear from context whether an upper case letter denotes a matrix or a variable.
- **Conjugate transposes** of matrices are denoted with a *, such as H^*
- All **logarithms** are to the base 2 unless stated otherwise.
- **Matrix norms** are denoted $\|\cdot\|$, and the Frobenius norm is denoted $\|\cdot\|_F$.

Introduction

Communication security is traditionally provided by methods such as shared secret keys. Such techniques take place in the upper layers of the Open Systems Interconnection (OSI) reference model [76] (Table 1), a standardised model for telecommunications technologies, where each layer is independent of one another and deals with different types of data. The work in this thesis focuses on the physical layer, which is the layer concerning data at bit level.

Layer in OSI reference model	**Data type**
Application	Data
Presentation	Data
Session	Data
Transport	Segments
Network	Packets
Data Link	Frames
Physical	Bits

Table 1: The seven layers of the OSI reference model.

A layered architecture such as the OSI model means that it is possible to have security measures at each of the seven layers, which for something as crucial as data security, is surely desirable. The physical layer, the only layer which deals with data at bit level, is 'layer 1' and typically concerns matters of reliability and the physical medium for the transmission, such as the type of wire or frequency. In most modern day cases this medium will be wireless. The physical layer historically has not been used for security nearly as much as other, higher, layers where the security protocols assume that the physical layer is error free [7].

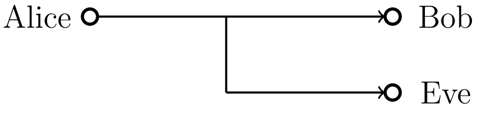
Classical security techniques have assumptions behind them which modern advances are surpassing. A cryptographic measure is considered secure if it would take an unfeasible amount of computational power for an adversary to break it [69]. Computational power available is increasing and thus this assumption may not hold in a modern day scenario. What is considered unfeasible depends on the current state of the art and is constantly changing.

All of this illustrates the desire to move away from a sole reliance on these classical techniques. The stack based model which most devices are based upon allows for security to be implemented at multiple layers. Therefore these technologies may be used in parallel and complement one another where feasible.

Fewer than 10% of the population used 1G [20] but developments in telecommunications has led to reduced costs for these technologies and a far greater uptake than could have been predicted in the days of Shannon. As 5G becomes a reality and 6G is being developed [74], much of the theory underlying physical layer security remains unknown. Although the fundamental ideas date back to Shannon in 1949 [68] and Wyner in 1975 [73], their work is based on classical point-to-point communication systems, and new versions of these results are required for multiple-input multiple-output (MIMO) and massive MIMO systems.

Physical layer security has an information-theoretic foundation and does not rely on the computing power available to the users, overcoming the concerns outlined above. It is instead based on the quality of the channel between the users and the blocklength of the messages. Rather than requiring users to generate random secret keys, physical layer security utilises the inherent randomness of the physical medium (in a wireless channel, this could be due to random electrical pulses in the environment) in order to improve secrecy. As long as the legitimate users maintain some advantage over the eavesdropper, their rate of perfectly secure communications may be positive.

This thesis studies security from the perspective of passive eavesdropping attacks. That is where the adversary does not actively jam, spoof or contaminate the legitimate signal, but simply overhears. Robustness against passive eavesdropping protects users against unwanted interception and traffic analysis. In this setup a legitimate user, Alice, is sending a message to a legitimate receiver, Bob. This message is intercepted by an eavesdropper, Eve, who observes the transmitted message through a different channel to Bob. The difference in their channel is utilised to provide secrecy.



Motivated by the evolution of telecommunications and security requirements, this thesis aims to address open problems and consider the inherent security of advancing technologies.

Chapter 1 provides the mathematical background required for this thesis. **Chapter 2** introduces the communications systems with the associated definitions and results required. **Chapter 3** surveys the literature in the relevant areas of physical layer security. **Chapter 4** states and proves results about the concavity of the MIMO secrecy capacity, the theoretical maximum rate for error free, perfectly secure communications for the Gaussian channel in the case of a single eavesdrop antenna, contributing to the literature for this open problem. **Chapter 5** considers a downlink MIMO NOMA setup and shows its robustness to eavesdropping. Using results from random matrix theory, it is shown that the secrecy is enhanced as the number of antennas increases. **Chapter 6** introduces a novel combinatorial coding scheme, which provides security against a passive eavesdropper while allowing several users to communicate in a time dispersive environment. The scheme uses properties of the legitimate channel to generate a sparse codebook, making the scheme robust to eavesdroppers who do not have access to these channels. **Chapter 7** summarises the key contributions of the preceding chapters and lays out future research directions.

Chapter 1

Background

In this chapter we outline some of the mathematical preliminaries required for analysing communication systems in later chapters. We begin by introducing the mathematical framework for communication systems and coding rates, information theory.

1.1 Information theory: an introduction

Information theory is the mathematical study of communications systems. The field was founded by Claude Shannon in the paper *'A Mathematical Theory of Communication'* [67], information theory concerns transmitting messages where noise is present.

1.1.1 Measuring uncertainty

A key metric in information theory is the *entropy*, which measures the uncertainty of a variable, or the surprise associated with an outcome of a random event. An event with high probability will have a low information content, as it has less of a 'surprise' factor, whereas an unlikely event occurring carries more information. For example if we see that it is raining in Bristol, a typically rainy city, this is not so surprising. The event has a high probability and thus low information value. If it is raining in the Atacama Desert, this is more surprising as it has a low probability and thus a high information value. Shannon quantified information of an event with probability p to be $-\log(p)$, and the entropy is the expected information content.

Definition 1.1.1: The *entropy* of a discrete random variable X, taking values in $\mathcal{X}$, is

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x), \tag{1.1}$$

where we let $0 \log 0 = 0$.

When the logarithm is base 2, entropy is measured in *bits*.

Example 1.1.2: Consider a Bernoulli distribution with probability of success p and probability of failure $1 - p$. The entropy may be written as

$$H(p) = -p \log p - (1 - p) \log(1 - p), \tag{1.2}$$

also known as the *binary entropy function*. If $p = 1$ or $p = 0$ then the entropy is zero. This is because there is no *surprise* as the variable is deterministic. The entropy of a Bernoulli random variable can be seen in Figure 1.1. It is maximised when $p = 0.5$, when the outcome is the least certain.

It is not only the binary entropy function which is maximised for a uniform probability distribution; this is the case for random variables taking values in larger sets as well. For a random variable X taking values in $\mathcal{X}$, the entropy is bounded above by $H(X) \leq \log|\mathcal{X}|$ [16, Theorem 2.6.4].

Definition 1.1.3: The *joint entropy* of discrete random variables X and Y, taking values in $\mathcal{X}$ and $\mathcal{Y}$ respectively, is

$$H(X, Y) = -\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x, y) \log p(x, y). \tag{1.3}$$

Definition 1.1.4: The *conditional entropy* of discrete random variables X and Y, taking values in $\mathcal{X}$ and $\mathcal{Y}$ respectively, is

$$H(X|Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x, y) \log p(x|y) \geq 0. \tag{1.4}$$

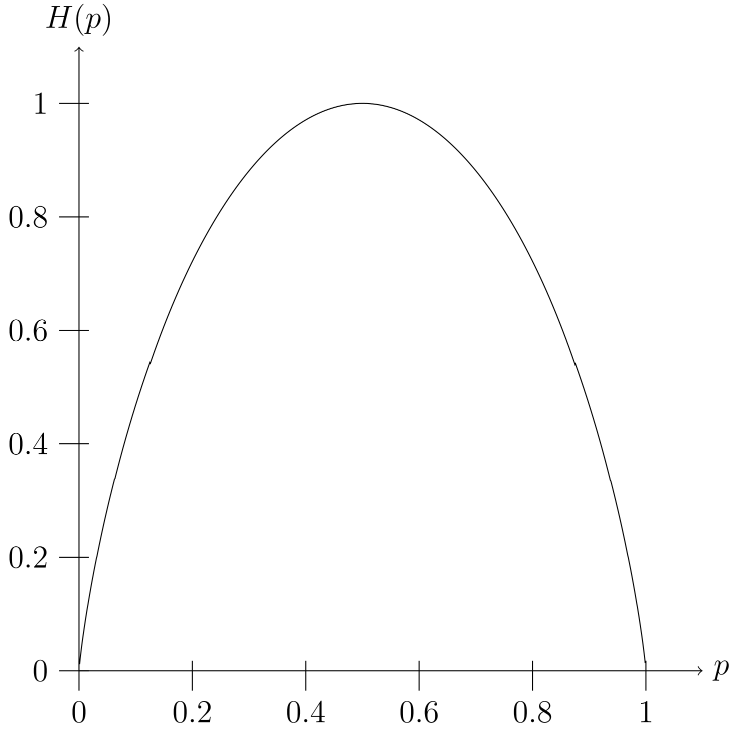


Figure 1.1: The entropy of a Bernoulli distribution with parameter p.

The conditional entropy is the residual uncertainty of X after observing $Y = y$ and indeed $H(X|Y)$ is related to $H(X,Y)$ as follows [52, §8.1]

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y). \tag{1.5}$$

From the above, we see that

$$H(X \mid Y) \leq H(X), \tag{1.6}$$

and thus conditioning may not increase entropy. More relationships between information theoretic quantities are shown in Figure 1.2.

Definition 1.1.5: The *mutual information* between random variables X and Y is defined as

$$I(X;Y) = H(Y) - H(Y|X) = H(X) - H(X|Y). \tag{1.7}$$

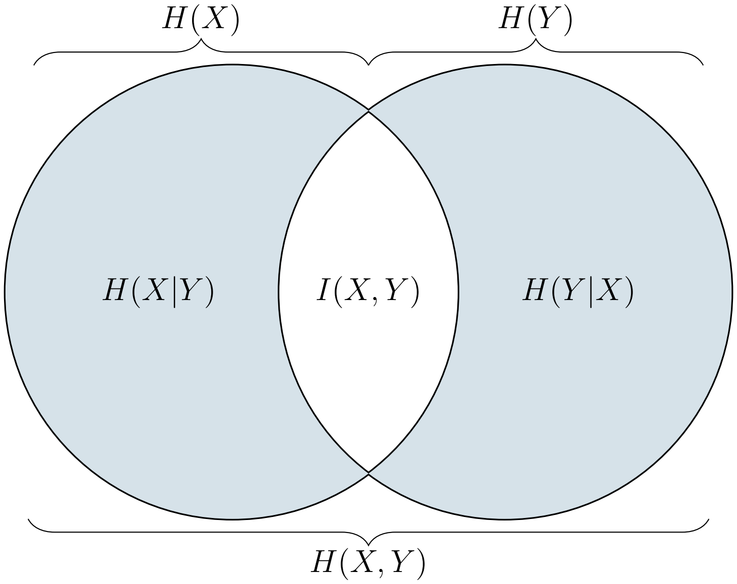


Figure 1.2: Relationships between entropy and mutual information [16, Theorem 2.4.1].

Theorem 1.1.6 ([16, Theorem 2.6.3])**:** The mutual information between random variables X and Y satisfies $I(X;Y) \geq 0$ with $I(X;Y) = 0$ if and only if X and Y are independent.

1.1.2 Channel coding

Information theory gives a the mathematical framework for communications in the presence of noise. In this section we outline the communication systems and the key information theoretic measures. A transmitter, *Alice*, sends a message to a receiver, *Bob*, over some channel. This channel may be a telephone line, piece of optical fibre or a wireless medium, and is formally defined later in Definition 2.1.2. If this channel is noisy, which almost all

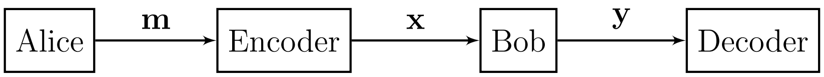


Figure 1.3: A general communication system from Alice to Bob

physical channels are, then how can Alice ensure Bob receives the correct message? This question is the basis for *channel coding*. Here, we use the term *code* to refer to the method used to transmit the message.

Definition 1.1.7: A symbol *code* of an ensemble X is a mapping from X into $\{s_1, \cdots, s_k\}^+$, or a *binary* code when $\{s_1 \cdots s_k\} = \{0, 1\}$. The representation of symbol x is called the *codeword* and the collection of codewords is the *codebook*.

Example 1.1.8: In *Morse code* the message is plaintext in the Latin alphabet. The *code* is the dots and dashes transmitted to represent the plaintext. The collection of all 36 *codewords* (representing a-z and 0-9) is the *codebook* for Morse code. The letters A-F are shown in Table 1.1. Note that a dot, ., has length one, whereas a dash, -, has length three meaning that the letter e is assigned the shortest possible codeword due to its frequency of use.

Plaintext symbol	Codeword
A	.-
B	-...
C	-.-.
D	-..
E	.
F	..-.

Table 1.1: Morse code plaintext and codewords.

In order for the code to be readable from left to right, one codeword must not be a prefix for any other. That is, the code design should be *prefix free.* Morse code is not prefix free, as the code for '`E`' is `.`, which occurs at the beginning of the code for '`A`'. In Example 1.1.8, symbols occurring with high probability (for example '`E`') are assigned short code lengths. Indeed, a 'good' code takes into account the underlying probability distribution in its design. The expected length of a code with probabilities p_i and corresponding codeword lengths l_i is

$$\mathbb{E}(L) = \sum_i p_i l_i \tag{1.8}$$

which we wish to minimise. The act of reducing the length of our codewords as much as possible is called *compression* and a lower bound for the expected lengths is $H(X)$, the Shannon entropy (Definition 1.1.1). Indeed, this is the *Source Coding Theorem* [67].

Theorem 1.1.9: For a random variable X, where x_i has probability p_i, there exists a prefix free code with an expected length $\mathbb{E}(L)$ satisfying

$$H(X) \leq \mathbb{E}L \leq H(X) + 1, \tag{1.9}$$

and no prefix free code has expected length less than the entropy.

A proof of Theorem 1.1.9 may be found in [52, §5.4]. This result shows that information may not be compressed below the entropy in an error free way. How well a code performs may be measured by its rate, defined as follows:

Definition 1.1.10: The *rate* of a code is the ratio of useful information bits to total information bits transmitted per second, measured in bits/s. A rate is said to be *achievable* if there exists a code which conveys information at that rate.

Codes so far have been designed per symbol, but often the underlying probability distribution will have implications for strings of symbols. In the

English language, for example, the pair ‘QU’ are far more likely to appear than ‘QJ’. We may think of our decoder reading a string of codewords from left to right, if they have seen a ‘Q’ they can expect to see a ‘U’ and so this string may be compressed further than the string ‘QJ’. The optimal coding scheme for this is *arithmetic coding*, introduced by [25]. In arithmetic coding, binary strings have a one to one mapping with an interval on the real line. These real intervals correspond to the probability that a sequence of symbols occurs.

The real interval corresponding to a generic string $x_1 \dots x_k$ has a width

$$p(x_k \mid x_1 \dots x_{k-1}), \tag{1.10}$$

which is mapped to a binary string.

Longer strings correspond to smaller intervals contained within the intervals of their prefixes. That is, the string `010` corresponds to a subinterval of `01`. The compression provided by arithmetic encoding is close to optimal [42].

Example 1.1.11: Consider random variable X with alphabet $\mathcal{X} = \{x_1, x_2, x_3\}$ with probabilities $(p_1, p_2, p_3) = (0.1, 0.3, 0.6)$ respectively. Consider strings where each symbol is independent and identically distributed (IID) from the previous choice. Initially, the interval $I_1 = [0, 1]$ is partitioned into the intervals $[0, p_1)$, $[p_1, p_1 + p_2)$ and $[p_1 + p_2, 1]$. For strings of length 1, the encoded string is the largest binary interval contained within these partitions. For longer strings, the intervals are updated and then the same rule applies. The following strings are depicted in Figure 1.4, with their probability intervals and binary intervals shown.

- The most likely string of length 1 to occur, x_3, corresponds to a partition width of 0.6 (which is just p_1) and is encoded as the binary string `1`.

- The length 2 string x_3x_3 (partition width of 0.36) is encoded as `11` while the string x_3x_2 (partition width of 0.18) is encoded as `100`.
- The length 3 string $x_3x_3x_3$ corresponds to a partition of width 0.216 and is encoded as `111`. The string $x_3x_3x_3$ corresponds to a partition width of 0.108 and is encoded as the longer message `1011`.

It can be seen that all strings exemplified above are encoded to messages beginning with `1`, this is because they all have x_3 as a prefix and are their intervals are contained within the x_3 interval.

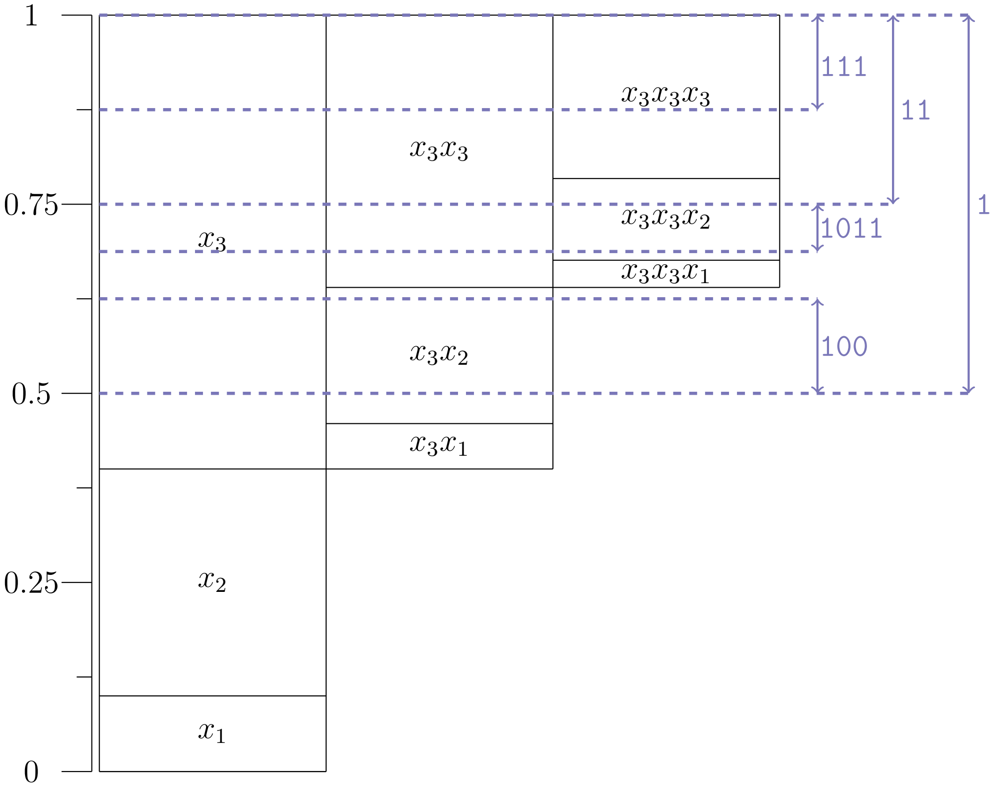


Figure 1.4: Arithmetic encoding partitions for Example 1.1.11.

In Example 1.1.11, the symbol strings were IID, but as was motivated earlier, arithmetic encoding is particularly useful when the probability distribution is dynamic as is the case in the English language.

Example 1.1.12: Consider encoding binary strings $x_1 \dots x_n$ of length n and fixed weight m. Initially the probability of observing a 1 is $\frac{m}{n}$. After observing x_1, the probabilities update

$$p(x_2 = 1 \mid x_1 = 1) = \frac{m-1}{n-1}, \tag{1.11}$$

$$p(x_2 = 1 \mid x_1 = 0) = \frac{m}{n-1}. \tag{1.12}$$

If we use the example length 5 strings with weight 1, the arithmetic encoding and probability intervals may be seen in Figure 1.5. Note that for the string 00100, the binary intervals of length 3 are not fully contained within the probability interval and thus the string is assigned a binary codeword of length 4.

1.2 Differential entropy

For continuous random variables, the summations in Definitions 1.1.1, 1.1.3 and 1.1.4 are replaced with an integral and the discrete probabilities are replaced with the probability density function. To justify why it is possible to do this, we first outline a quantisation argument from [16, §8.3].

Consider a random variable X with a continuous probability density function f. Split the real line into intervals of size δ: $(t\delta, (t+1)\delta)$ for $t \in \mathbb{Z}$. Then the probability that the quantised version of X, denoted X^δ, takes a certain value is given by

$$\mathbb{P}(X^\delta = t) = \int_{t\delta}^{(t+1)\delta} f(x)\, dx = \delta f(x_t) \tag{1.13}$$

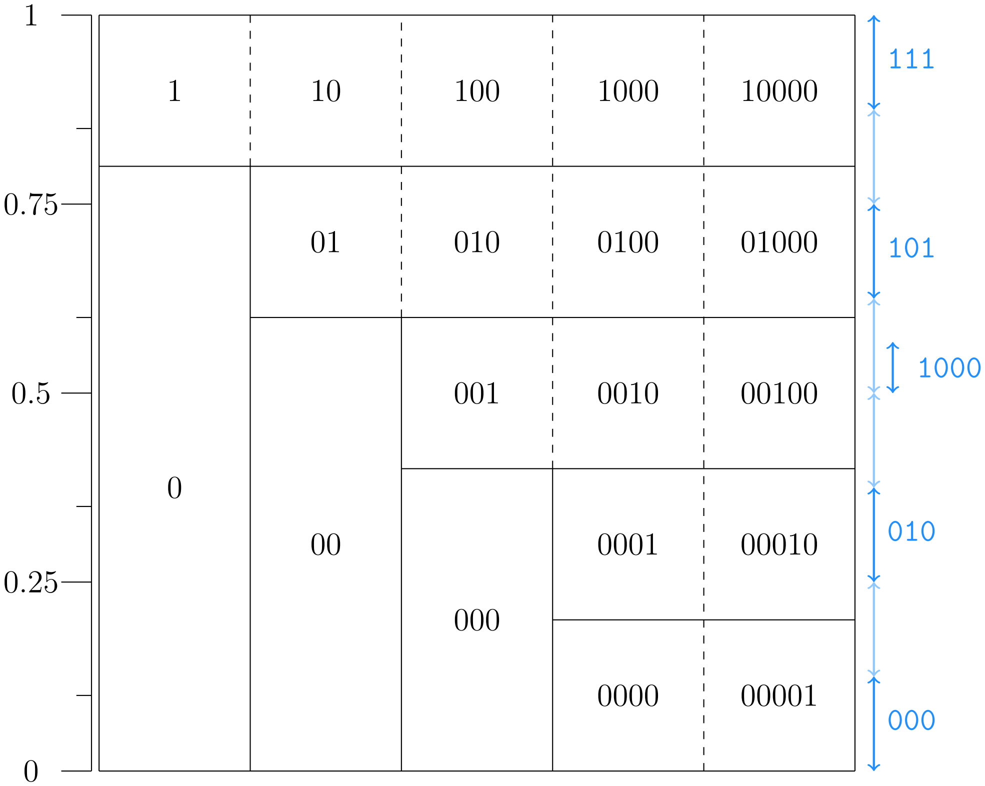


Figure 1.5: Arithmetic encoding intervals for length 5 strings of constant weight 1.

for some x_t in the interval. Then the entropy of the quantised variable is

$$\begin{aligned} H(X^\delta) &= -\sum_t \mathbb{P}(X^\delta = t)\log(\mathbb{P}(X^\delta = t)) \\ &= -\sum_t \delta f(x_t)\log f(x_t) - \log\delta. \end{aligned} \tag{1.14}$$

Then the differential entropy follows by the Riemann integrability of f.

Definition 1.2.1: The differential entropy $h(X)$ of a continuous random variable X with probability density function $f(x)$ is

$$h(X) = -\int_{\mathrm{Supp}(X)} f(x) \log f(x)\, dx, \tag{1.15}$$

where the integral is taken over the support of X.

One key difference from the discrete entropy is that differential entropy can take negative values. Consider the uniform distribution over the interval $[0, a]$ for some $a > 0$. The differential entropy is

$$h(X) = -\int_0^a \frac{1}{a} \log \frac{1}{a}\, dx = \log a, \tag{1.16}$$

this is negative when $a < 1$.

1.3 Mathematical preliminaries

The following section outlines some required definitions and theorems for studying convex functions and matrices. Since this thesis concerns multiple antenna systems, this will mean understanding their channel matrices (introduced in Section 2.4.1) and some convexity results for functions of matrices.

1.3.1 Complex random vectors

Complex random variables and vectors are defined similarly to continuous real random variables, but with entries drawn from $\mathbb{C}$ rather than $\mathbb{R}$. More information about complex random vectors and Gaussian random vectors can be found in [63, §7.9]. Throughout, we let i denote $\sqrt{-1}$.

Definition 1.3.1: A *complex random variable*, $Z \in \mathbb{C}$ is a variable of the form

$$Z = X + iY, \tag{1.17}$$

where both X and Y are real random variables.

In the real case, there is an inherent ordering of variables so the cumulative distribution function (CDF) makes sense to be $P(X \leq x)$ for some real x. This ordering does not exist in the complex plane, and thus the complex random variable is defined by the joint distribution of its real constituents.

Definition 1.3.2: A *complex Gaussian random vector*, is one which can be written as $\mathbf{z} = \mathbf{x} + i\mathbf{y}$ where both $\mathbf{x}$ and $\mathbf{y}$ are jointly Gaussian random vectors.

Definition 1.3.3: A *circularly symmetric Gaussian random variable* denoted $Z \sim \mathbb{C}N(0, \sigma^2)$, with variance $\mathbb{E}(Z)^2 = \sigma^2$ is one where Z and $Ze^{i\theta}$ have the same distribution for all $\theta \in [0, 2\pi)$. Z has probability distribution function

$$p(z) = \frac{1}{\pi\sigma^2} e^{-|z|^2/\sigma^2} \tag{1.18}$$

for $z \in \mathbb{C}$.

Definition 1.3.4 ([28, §7.7])**:** For matrices A and B, the *generalised eigenvalues* of $A - \lambda B$ are the values λ for which

$$\det(A - \lambda B) = 0. \tag{1.19}$$

The *generalised eigenvectors* are the non zero vectors $\mathbf{v}$ satisfying

$$A\mathbf{v} = \lambda B\mathbf{v}. \tag{1.20}$$

Useful definitions and further background on random vectors, matrices and their properties may be found in [58, §8].

1.3.2 Convexity

Definition 1.3.5: A function $f : \mathbb{R}^d \to \mathbb{R}$ is *convex* if its domain is a convex set and for any any pair x, y in the domain, and any $\lambda \in [0, 1]$

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \tag{1.21}$$

If the inequality is strict for all $\lambda \in (0,1)$, then the function is *strictly convex*. Likewise, a function is *concave* if the inequality in Equation (1.21) is reversed.

The binary entropy function seen previously in Figure 1.1 is a concave function. Convex functions are 'well behaved' in the sense that there are established methods for finding their minimum values. Boyd's book [10] on the optimisation of convex functions details many such methods. One family of optimisation algorithms are descent methods, which take a sequence

$$x_j = x_{j-1} + t\Delta x_j$$

such that

$$f(x_j) \leq f(x_{j-1})$$

until the minimum, or a value close to the minimum is reached.

Example 1.3.6: The *gradient descent method* involves searching in the direction of $-f'(x)$ as in Algorithm 1.

Algorithm 1: Gradient descent method

```
input  : $x \in \mathrm{dom}(f)$.
output: Minimum of $f(x)$ within a precision of $\eta > 0$.
while $\|f'(x)\| > \eta$ do
    Determine a descent direction $\Delta x = -f'(x)$.
    Choose a step size $t > 0$.
    Update $x = x + t\Delta x$.
end
```

Theorem 1.3.7: For positive semidefinite matrices X, $f(X) = \log\det(X)$ is concave.

To prove Theorem 1.3.7, we follow the approach of [10, p74] and consider taking an arbitrary line

$$X = Y + tZ$$

where X, Y, Z are positive symmetric matrices and t is some real number. We may now define

$$g(t) = \log\det(X) = \log\det(Y + tZ). \tag{1.22}$$

By restricting t to be such that $Y + tZ$ is positive semidefinite, we may assume without loss of generality that Y is positive semidefinite and $t = 0$ within this interval. Since X is positive semidefinite, there exists a matrix $X^{\frac{1}{2}}$ such that $X = X^{\frac{1}{2}}X^{\frac{1}{2}}$. Therefore

$$\begin{aligned} g(t) &= \log\det\left(X^{\frac{1}{2}}X^{\frac{1}{2}} + t(X^{\frac{1}{2}}X^{-\frac{1}{2}}ZX^{-\frac{1}{2}}X^{\frac{1}{2}})\right) && (1.23)\\ &= \log\det\left(X^{\frac{1}{2}}(I + tX^{-\frac{1}{2}}ZX^{-\frac{1}{2}})X^{\frac{1}{2}}\right). && (1.24)\end{aligned}$$

Since $\det(AB) = \det(A)\det(B)$, we may write

$$\begin{aligned} g(t) &= \log\left(det(X)det(I + tX^{-\frac{1}{2}}ZX^{-\frac{1}{2}}\right) && (1.25)\\ &= \log\det(X) + \sum_{j=1}^{n}\log(1 + t\lambda_j), && (1.26)\end{aligned}$$

where $\lambda_1, \ldots, \lambda_n$ denote the n eigenvalues of $X^{-\frac{1}{2}}ZX^{-\frac{1}{2}}$ (where $I + X^{-\frac{1}{2}}ZX^{-\frac{1}{2}}$ is a positive semidefinite matrix and so $1 + \lambda_j \geq 0$ for each j). Standard differentiation results give that $g''(t) \leq 0$ and thus $f(X)$ is concave.

Theorem 1.3.8 ([52, §2.7])**:** If f is a convex function then for any random variable X

$$\mathbb{E}f(X) \geq f(\mathbb{E}X). \tag{1.27}$$

This is known as *Jensen's inequality.*

1.3.3 Linear Algebra

Wireless channels are modelled as matrices, as we will see in Chapter 2, and consequently analysing them requires some results matrix algebra. This section provides the necessary definitions and theorems for matrices and vectors used throughout this thesis.

Definition 1.3.9: For a matrix A, the *Frobenius norm* is

$$A_F = \sqrt{\operatorname{Tr}(A^*A)}. \tag{1.28}$$

It can be seen by properties of the trace that

$$\|A^*A\|_F = \|AA^*\|_F \leq \|A\|_F^2. \tag{1.29}$$

Since the trace is the sum of eigenvalues, the Frobenius norm can be written as

$$\|A\|_F = \sqrt{\sum \lambda_j}, \tag{1.30}$$

where λ_j are the eigenvalues of A^*A.

Theorem 1.3.10 ([33, Example 5.6.0.2])**:** For square matrices, the Frobenius norm satisfies the *submultiplicative property.* That is, for square matrices A and B

$$\|AB\|_F \leq \|A\|_F\|B\|_F. \tag{1.31}$$

Definition 1.3.11 ([33, §4.2])**:** For a given Hermitian matrix, A, and a nonzero vector $\mathbf{x}$ the *Rayleigh Quotient* $R(A, x)$ is

$$R(A, \mathbf{x}) = \frac{\mathbf{x}^*A\mathbf{x}}{\mathbf{x}^*\mathbf{x}}. \tag{1.32}$$

Theorem 1.3.12: The *standard complex polarisation identity* states that

$$2\operatorname{Re}\langle \mathbf{u}, \mathbf{v}\rangle = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2, \tag{1.33}$$

meaning that $\|\mathbf{u}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2 \geq 0$ if and only if $2\operatorname{Re}\langle \mathbf{u}, \mathbf{v}\rangle \geq \|\mathbf{v}\|^2$.

Definition 1.3.13: For a real positive definite matrix the *Cholesky decomposition* is the factorisation

$$A = LL^T, \tag{1.34}$$

where L is a unique lower triangular matrix with positive diagonal entries. For proof that such a decomposition always exists, see [28, Theorem 4.2.7].

Definition 1.3.14: The *Kronecker product* of $m \times n$ matrix A and $p \times q$ matrix B is the $mp \times nq$ matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mm}B \end{pmatrix}. \tag{1.35}$$

Theorem 1.3.15: Let A be an $n \times n$ matrix, the derivative of the quadratic form (see [57, §IV] for more on quadratic forms) is

$$\frac{\partial}{\partial \mathbf{u}}[\mathbf{u}^* A \mathbf{u}] = (A + A^T)\mathbf{u}, \tag{1.36}$$

where $\mathbf{u}$ is some $n \times 1$ vector. When A is Hermitian, $A = A^*$ and thus the derivative becomes $2A\mathbf{u}$.

Chapter 2

Communication Theory

In this chapter, we give the required background knowledge to understand the communication models used in later chapters. We begin with the fundamental definitions and finish by introducing the 5G technologies which are studied in later chapters.

2.1 Channels and capacity

So far we have considered coding and compression of data in an error free sense. That is, what Alice sends is what Bob receives. In reality, there may be some corruption or noise which alters Bob's received message. At one end of the spectrum, Bob may receive nothing useful and entirely fail to understand what Alice sent. At the other end, Bob may receive the message perfectly. In reality, the channel will be noisy and a scenario somewhere in between will occur; Bob will make some errors. If Bob decodes a message which is believable, but incorrect, how will they know that they have made an error?

Example 2.1.1: To avoid errors, Alice may send each message T times. Bob can then take a majority vote on the most likely message based on the T received versions. This is called *repetition coding* and while it may work, it takes a factor of T times as long to send each message, and a rate of $1/T$ in the sense of Definition 1.1.10. As T increases, the error probability decreases, but the communication rate is sacrificed. This is not always practical, as information may be required quickly, or the cost of using the communication channel might be high.

The two key factors to consider in a communication system are the *error probability* and the *rate*. Evidently, there is a trade off to be made between the two. In the Example 2.1.1, we saw that by using repetition coding we may sacrifice rate to improve our accuracy. For a given system, there is an error threshold which is acceptable for its purpose. Teleconferencing or gaming demands a high rate with a low latency. In these cases, more errors are acceptable to the user. On the other hand for military communications, for example, accuracy may take precedence over the rate. Naively one might think the only way to achieve a zero error communication would mean the rate of communication tends to zero. However Shannon's notion of a system's *capacity* showed that it is possible to do far better than this and the rate of communication can be positive for arbitrarily small error.

The plaintext that Alice wishes to send will be called *the message* denoted $\mathbf{m}$. Alice then *encodes* the message to a *codeword*, $\mathbf{x}$, which they send over a *channel*. Bob receives a potentially corrupted version, $\mathbf{y}$, of this codeword and aims to *decode* it, hopefully recovering $\mathbf{m}$ correctly.

Definition 2.1.2: A *channel* W is a function

$$W : \mathcal{X} \times \mathcal{Y} \to [0, 1]$$

for input alphabets $\mathcal{X}$ and output alphabet $\mathcal{Y}$ which satisfies

$$\forall x \in \mathcal{X} : \sum_{y \in \mathcal{Y}} W(y|x) = 1. \tag{2.1}$$

The channel function can be thought of as a transition probability $p(y|x)$.

Example 2.1.3: The simplest example of a channel is the *binary symmetric channel* (BSC). This is the channel with a binary input and output alphabet $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and transition probabilities

$$p(0|1) = p(1|0) = p$$

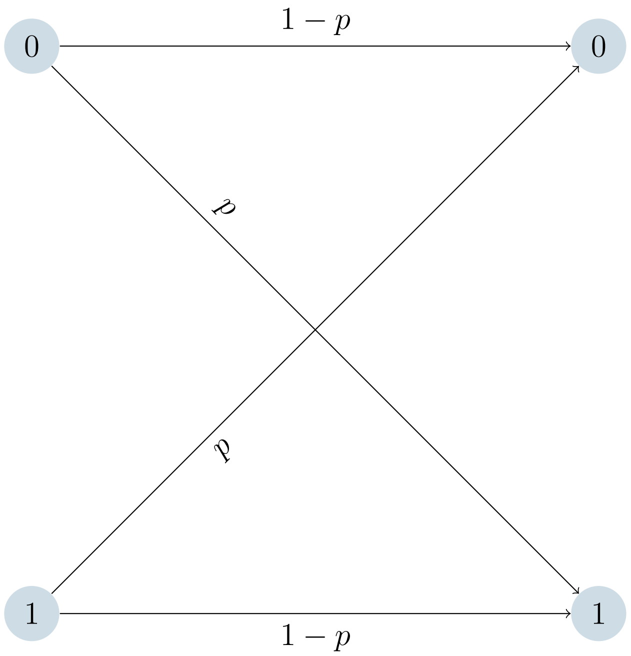


Figure 2.1: Binary symmetric channel with parameter p.

and

$$p(1|1) = p(0|0) = 1 - p.$$

In other words, this is the channel where the bits are flipped with probability p and correctly received with probability $1-p$. This can be seen in Figure 2.1.

We are now ready to define the capacity of a channel. With an arbitrarily small error, information can be transmitted across the channel at a rate less than C. If the rate of transmission exceeds C then the system will no longer be considered reliable and the probability of errors tends to 1. Shannon classified this in terms of the channel statistics, meaning that this capacity is prescribed from the fundamental properties of the channel.

Definition 2.1.4: The *capacity*, C, of a channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ is the supremum over all achievable rates of communication

$$C = \sup\{R : R \text{ is an achievable rate for a reliable code}\}.$$

Theorem 2.1.5: Shannon's Coding Theorem The *capacity*, C, of a communication system with discrete input X and output Y is

$$C = \max_{p(X)} I(X;Y), \tag{2.2}$$

where the maximum is taken over all input distributions and $I(\cdot,\cdot)$ denotes the mutual information.

2.1.1 Continuous signals

In reality, all signals are continuous but computers only have a finite amount of storage, we will first define how signals are reduced in order to store them. Firstly, the signal is *sampled* at a rate and then these samples are *quantised.*

Definition 2.1.6: The process of taking a continuous range of numbers and mapping these to a finite range of discrete values is called *quantisation.*

Example 2.1.7: Consider a continuous, real valued signal $f(t)$ which fluctuates above and below zero. We take taps of the channel at intervals of length T and quantise as follows for $n \in \mathbb{N}$

$$F(nT) = \begin{cases} 1, & \text{if } f(nT) > 0 \\ 0, & \text{otherwise.} \end{cases}$$

This process outputs a binary string.

2.2 SISO Channel

In a wireless communication, the conventional model is that of a single antenna at both the transmitter and receiver. This set up is also known as a single input single output (SISO) system, which has been the basis for many historical results.

Definition 2.2.1: The SISO channel has input $x \in \mathcal{X}$ and output $y \in \mathcal{Y}$ defined by the relationship

$$y = hx + n, \tag{2.3}$$

where y is the received message, x is the transmitted message, h is the channel coefficient and n is the noise added during the transmission.

2.2.1 AWGN Channel

A particularly useful channel model is that of a Gaussian channel. This closely resembles a real life wireless communication system [72, Section 5.1] and conveniently, is the most tractable, mathematically speaking.

Definition 2.2.2: The *SISO Gaussian channel* is the channel with input and output alphabets $\mathcal{X} = \mathcal{Y} = \mathbb{C}$. The transition $p(Y \mid x)$ is defined by

$$Y = x + Z$$

where the noise Z is drawn from a Gaussian distribution

$$Z \sim \mathbb{C}N(0, \sigma^2)$$

with σ^2 denoting the variance of the power, which since the mean is 0 is equivalent to the channels noise power.

Noise can be a result of random electrical processes in the atmosphere or agitation of electrons in the hardware. Thus the total noise is a summation of several small random processes. By the central limit theorem, it follows that this sum will be roughly Gaussian.

Theorem 2.2.3: The capacity of the SISO Gaussian channel with a transmit power of P and noise variance σ^2 is

$$C_{AWGN} = \log\left(1 + \frac{P}{\sigma^2}\right).$$

Note that P/σ^2 is the signal to noise ratio (SNR) of the channel.

This tells us, somewhat unsurprisingly, that the optimal transmit strategy to achieve capacity is to use all of the available power resource. It is also known that Gaussian signalling achieves the capacity (see [16, §9.1]) meaning that the input signal has the same shape of distribution as the noise, but a different power.

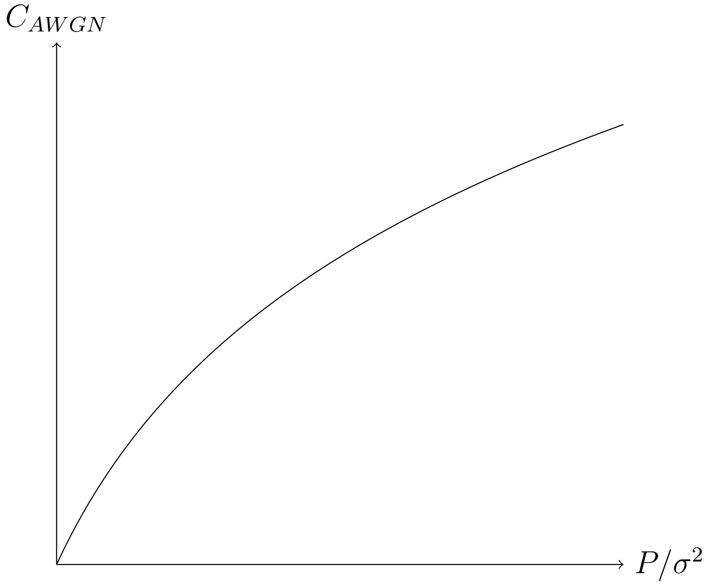


Figure 2.2: The capacity of a Gaussian channel vs SNR.

Heuristic proof of Theorem 2.2.3.

A heuristic proof of the AWGN capacity, found in [72, §5.1.2] is detailed below. For a more rigorous approach, see [16, §10.1], where the proof of both

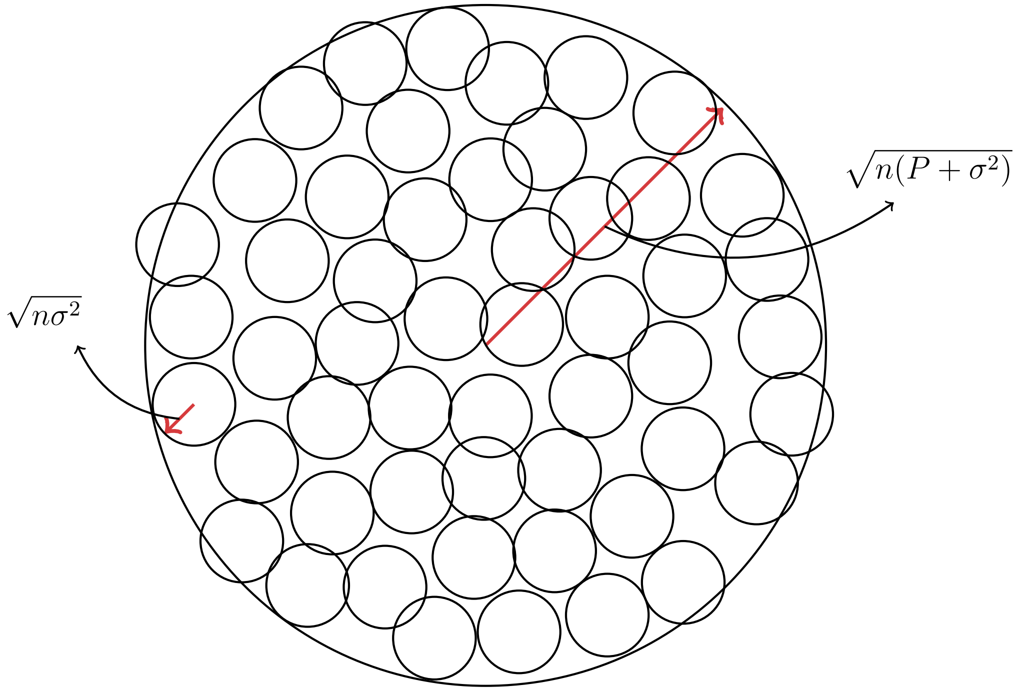


Figure 2.3: The capacity can be seen as the maximal number of non overlapping noise spheres inside the main sphere.

achievability and converse of the theorem can be found.

Let $\mathbf{x}$ be a blocklength n message with Gaussian entries and transmit power of P. The received message is $\mathbf{y} = \mathbf{x} + \mathbf{n}$ where $\mathbf{n}$ is a length n vector of Gaussian noise, with noise variance σ^2.

By the law of large numbers, $\mathbf{y}$ lies, with high probability, in a sphere of radius $\sqrt{n(P+\sigma^2)}$. As the blocklength n increases, the observed variance of the noise will approximate σ^2 and thus the observed signal $\mathbf{y}$ will, with high probability, lie near the surface of a noise sphere of radius $\sqrt{n\sigma^2}$ as seen in Figure 2.3.

To achieve a zero error probability, it is required that the noise spheres do not overlap, so that each $\mathbf{y}$ may be decoded uniquely to the corresponding $\mathbf{x}$. The volume of a general n dimensional sphere with radius r is proportional to r^n [14, §1.4]. Therefore the maximum number of messages we can send with zero errors is equivalent to the ratio

$$\frac{r_1^n}{r_2^n}, \tag{2.4}$$

where

$$r_1 = \sqrt{n(P + \sigma^2)} \tag{2.5}$$

$$r_2 = \sqrt{n\sigma^2} \tag{2.6}$$

which gives the maximum number of noise spheres that fit inside the larger sphere. Hence the maximum bits per symbol which may be communicated, or equivalently the capacity of the system, is given by

$$\frac{1}{n}\log\left(\frac{\sqrt{n(P+\sigma^2)}^n}{\sqrt{n\sigma^2}^n}.\right) = \frac{1}{2}\log\left(1 + \frac{P}{\sigma^2}\right), \tag{2.7}$$

which is the desired result.

2.3 Diversity and fading

Note that all channels defined in Chapter 1 were *static*. That is, the channel transition probabilities are fixed. In any physical channel there is fading and noise which varies over time and due to other factors such as the physical location or interference from other devices. Any wireless signal will have multipath components due to reflection, refraction etc. These multipaths will have differing arrival times at the receiver due to varying delays.

Diversity exploits the random fading of channels and is based on the idea that several statistically independent channels are unlikely to experience severe fading in the same places of the signal. A typical diversity system would sum at least two, but often many more, received versions of the same signal transmitted over multiple paths, each equipped with different fading statistics. A wireless communication between two devices happens over a fading channel. If we move one of those devices, the channel statistics will be different. Thus it is possible to exploit physical locations of transmitting devices in relation to the receiving device, known as spatial diversity.

Although the channel is modelled as a random variable, in this thesis we often refer to the channel coefficients as though they are fixed. This is because we assume we are working within the coherence time of the channel, defined as follows.

Definition 2.3.1: The *coherence time* of a channel is the duration of time in which the channel statistics are considered to be static.

2.3.1 Dispersive channels

The random fading described in Section 2.3 may also contribute to dispersive channels. An environment is described as highly dispersive if the characteristics change vastly over time, or it has a short coherence time (Definition 2.3.1). A simple multipath scenario is one where two multipath compononents arrive with similar power. These multipaths arrive with a time spread. Since each multipath component takes a unique path from transmitter to receiver, they experience a unique time of flight. This spread of timing leads to such a channel model being described as *time dispersive* and the ouput is based on taking taps of the channel.

Definition 2.3.2: A channel is said to be *time dispersive* if several multipaths arrive at different times. For L taps of the channel, the signal input is $\mathbf{x}$

$$\mathbf{y} = \mathbf{h} \star \mathbf{x} + \mathbf{n}, \tag{2.8}$$

where $\star$ represents convolution. The L-tap channel is represented by vector $\mathbf{h}$ of length L and the channel noise is the length L vector $\mathbf{n}$. The received signal in time slot t is

$$y_t = \sum_{i=0}^{L-1} h_i x_{t-i}, \tag{2.9}$$

where $x_j = 0$ for $j < 0$.

2.4 The Evolution of Telecommunications

When Shannon wrote *'A Mathematical Theory of Communications'* [67], the world was a different place and he based most of his work on wired telegraphs as the basic communication model [70, §16]. In the years that have passed since, there have been several generations of cellular communications. Despite this, the basic theories developed by Shannon remain the building blocks for studying these systems and are still of the utmost importance.

Generation	**Changes in services offered**	**Year**
1G	Voice calling	1979
2G	SMS capabilities, data rates up to 200kb/s	1991
3G	Data rates of 2Mb/s	1998
4G	Reduced cost of data, voice over IP	2008

Table 2.1: A high level overview of the new services offered in the evolution from 1G to 4G [20].

With first generation communications (1G), voice calling was the main offering. Second generation (2G) continued to improve these offerings and was the first generation to introduce mobile data capabilities. Since 2G, demand for data has risen and continues to rise with modern needs superseding the offerings of fourth generation (4G) [20]. The key goals for 5G are detailed in Figure 2.4 and are driven by a number of industries and applications (see

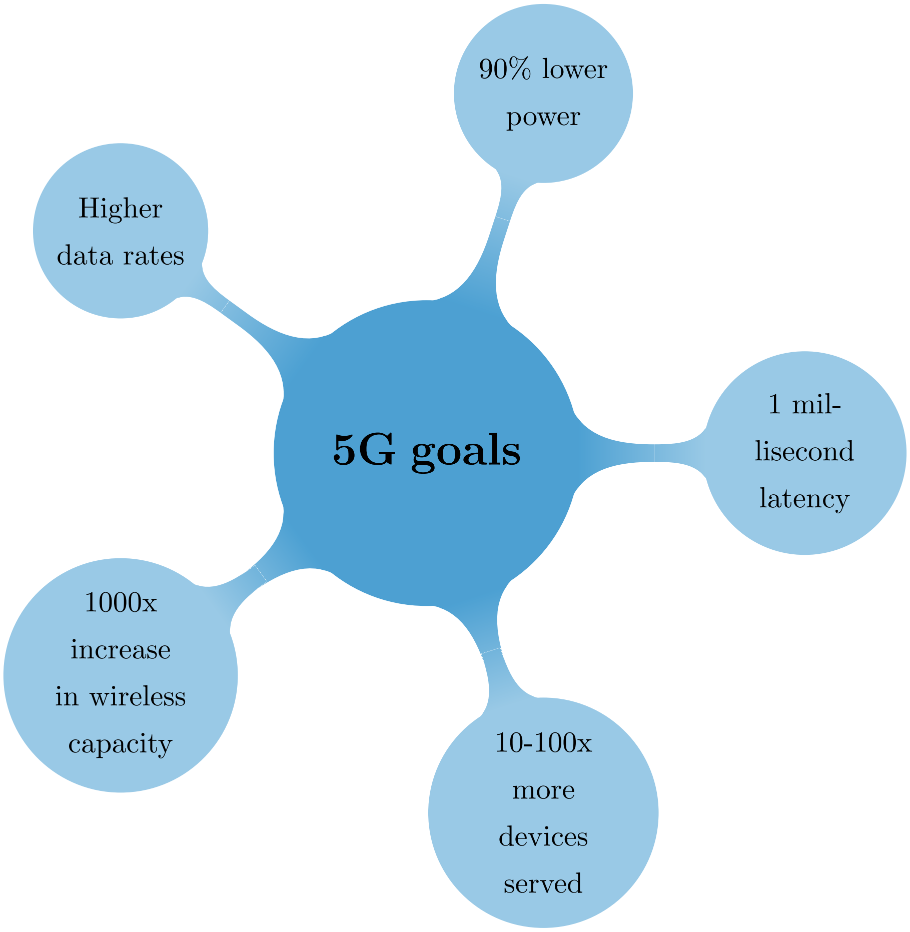

90% lower power
Higher data rates
5G goals
1 millisecond latency
1000x increase in wireless capacity
10-100x more devices served

Figure 2.4: Goals for 5G communications compared to 4G

Table 2.2). As the needs and requirements for communications grow and develop, so do the technologies to meet these demands [9].

An increase in data rates can come from several avenues. We can increase the power we transmit at, but this has an immediate limitation in a mobile device since the battery life is finite and is impractical beyond a certain level due to safety concerns for users and the interference caused for other devices. Increasing the frequency resource may enable a higher transmission rate theoretically but bandwidth is a limited and expensive resource with

Goal	**Example use case**
Low latency	Connected autonomous vehicles
Higher data rates	High definition video streaming
Higher wireless capacity	Dense areas of mobile users
Lower power	Sensor networks

Table 2.2: Driving factors for future wireless

access determined by policy. Utilising diversity, as explained in Section 2.3, is a fruitful way to achieve these higher data rates and indeed spatial and power diversity are present in the 5G specifications [55] as we will see in the following sections.

2.4.1 MIMO and massive MIMO

Using the same power and frequency resources, multiple antenna systems can achieve higher data rates than their single antenna counterparts by exploiting the spatial diversity of the antennas. Theoretical results by Telatar [71] and numerical results [27] in the late 1990s showed the potential gains for MIMO systems, even with small numbers of antennas. For multiple users, MIMO systems can improve data throughput by directing energy towards the required user [11, §1.2], reducing interference issues. Since the early results, MIMO has become well a well established technology, available in WiFi since 2006 [2, §C].

MIMO systems exploit spatial diversity by placing the multiple antennas far enough apart from one another that they can be assumed to be statistically independent. This distance is at least half of the wavelength. Each transmit antenna has a different channel between each receive antenna and the channel gains may now be represented as a matrix rather than a single number in the SISO case.

Definition 2.4.1: An $n \times m$ MIMO system is one with n transmit antennas and m receive antennas has a corresponding channel matrix, H where entry H_{ij} corresponds to the channel from antenna i at the transmitter to antenna j at the receiver.

Echoing Equation (2.3), the received vector in a MIMO system is

$$\mathbf{y} = H\mathbf{x} + \mathbf{n}. \tag{2.10}$$

Where $\mathbf{x}$ is the $n \times 1$ column vector containing the transmitted signal, $\mathbf{y}$ is the $m \times 1$ column vector containing the received signals at each antenna and $\mathbf{n}$ is the column vector of noise present in each channel.

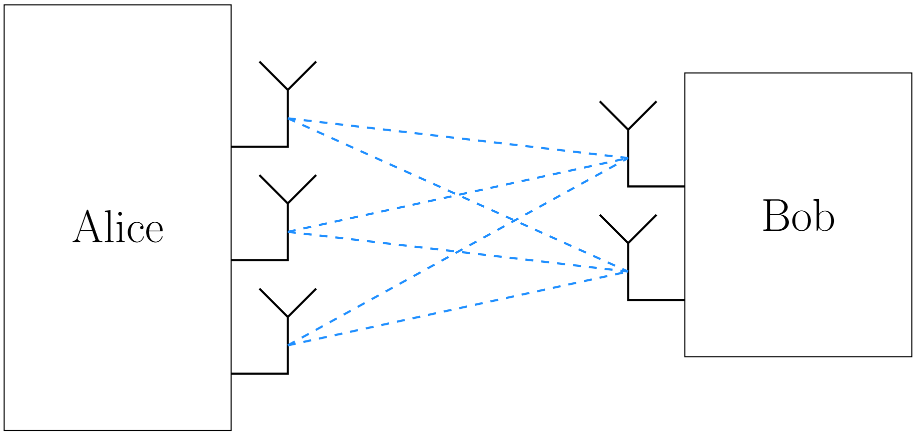


Figure 2.5: A 3×2 MIMO system.

Capacity of the Gaussian MIMO Channel

The Gaussian MIMO channel, which this thesis concentrates on, is a MIMO system with a Gaussian noise vector (Definition 2.2.2) and independent and identically distributed (IID) Gaussian entries for H.

Theorem 2.4.2: The capacity of a MIMO channel with Gaussian channel matrix H and input covariance matrix Q was found in [71] to be

$$\max_{Q} \log\det(I + HQH^{*}), \tag{2.11}$$

where the maximum is taken over input distributions and subject to a power constraint P.

Using Theorem 1.3.7 it can be seen that the capacity of a MIMO channel is a concave optimisation, and therefore mathematically tractable.

Massive MIMO

MIMO antenna systems are being scaled up in current research [64] as well as in practical applications for 5G to ‘massive’ MIMO. In a massive MIMO system the number of antennas (typically over one hundred) at the base station far exceeds the number of users. Massive MIMO is a technology for unlocking higher data rates and is considered to be a central technology for the development of 5G [1]. Practical results and trials such as those undertaken at the University of Bristol [31] have confirmed the theoretical promise.

2.4.2 Non-Orthogonal Multiple Access

Non-Orthogonal Multiple Access (NOMA) is a multiplexing technique in the code or power domain, which is particularly useful when users have very different channels and path loss characteristics since it exploits their channel diversity. In this thesis, NOMA in the power domain is considered. NOMA in this form was introduced by Saito et. al in [65] and is a part of the 5G specification [5, 19] due to the increased coverage and good spectral efficiency [18, 37].

It is typically considered for the *downlink* (base station to users) but may be implemented in the uplink also [32] however this section concerns a downlink NOMA system. Under this framework, the base station transmits a linear combination of messages which are allocated varying power resources depending on their channel quality. The receivers commonly use Successive Interference Cancellation (SIC) to retrieve their signal. Users share a frequency and time slot but the power allocated to each user differs depending on their channel quality. Simply, a user with a poor channel is allocated a higher power than a user with a better channel, as illustrated in Figure 2.6.

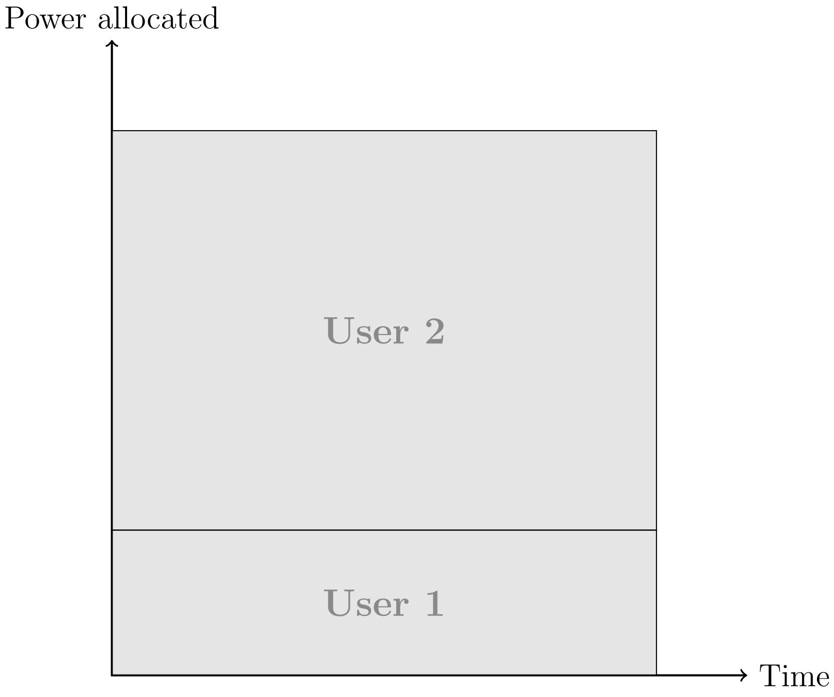


Figure 2.6: Power allocation in a NOMA system where User 2 has a worse channel than User 1.

In order to implement this, users with highly different channel characteristics are paired.

Example 2.4.3: Highly different channel characteristics may occur when users are physically located far apart. Consider a pair of one near and one far user. Suppose that User 1 is closer to the base station and User 2 is further away with channel coefficients h_1 and h_2 respectively, then the base station transmits the message $s = \alpha_1 s_1 + \alpha_2 s_2$ where s_i is the signal intended for user i and the α_i are power allocation coefficients with $\alpha_1^2 + \alpha_2^2 = 1$. In this example, $\alpha_1 \leq \alpha_2$.

For $i = 1, 2$, user i receives the message

$$y_i = h_i(\alpha_1 s_1 + \alpha_2 s_2) + n_i \tag{2.12}$$

where h_i is the channel coefficient and n_i is Gaussian noise and user interference with noise power N_i. User 2 treats the message for User 1 as noise as follows

$$y_2 = h_2 \alpha_2 s_2 + \underbrace{h_2 \alpha_1 s_1 + n_2}_{\text{noise}}. \tag{2.13}$$

User 1 uses SIC to retrieve their message; first they find s_2 (which is an easier problem than for User 2, because they are closer to the base station), then they subtract this and solve for s_1.

Assuming that User 1 can perfectly decode s_2, the two users have rates of transmission as follows:

$$R_1 = \log\left(1 + \frac{\alpha_1}{N_1}\right) \tag{2.14}$$

$$R_2 = \log\left(1 + \frac{\alpha_2}{\alpha_1 + N_2}\right). \tag{2.15}$$

It can be seen that the performance of the system depends heavily on the power allocations, α_1 and α_2. The further user, User 2, does not need a SIC receiver, which reduces the complexity requirement for their system.

2.4.3 MIMO NOMA

Non-orthogonal multiple access (NOMA), introduced in Section 2.4.2, is an enabling technology for 5G new radio [22, 35], due to the performance gains obtained when users have highly different channels. Performance of NOMA for single antennas at each of the users has been considered, both in the original NOMA paper by [65] and further in [24]. It was shown that, asymptotically, NOMA performs similarly to an opportunistic orthogonal multiple access schemes (OMA) [24] despite the potentially unfavourable conditions with many of the users having worse channel gains.

Since 5G uses MIMO and Massive MIMO technology [2], it is natural to ask whether MIMO and NOMA can be combined to deliver enhanced throughput relative to either scheme acting alone.

Indeed, these techniques have been successfully combined in a number of scenarios. It has been shown that, even with a loose interpretation of the need for differing channel conditions, gains from MIMO-NOMA schemes can be realised for internet of things (IoT) devices [21]. One multi-user MIMO-NOMA scheme of note was proposed by Ding, Schober and Poor [23], and has attracted considerable attention. Their scheme involves the base station transmitting a linear combination of messages, mixed using a precoding matrix P. This matrix P is carefully designed in terms of the row spaces of the downlink channel matrices, in order to achieve signal alignment. The key property is that, for each receiver, all but one of the interfering messages are aligned in the same vector subspace, and so can be removed by projection into an orthogonal space, effectively reducing the system to a standard two-user NOMA situation. We give more details in Section 5.2.

Chapter 3

Physical Layer Security

With modern day communications being used for a vast array of applications, from banking to healthcare, it is not surprising that security is of the utmost importance. 5G networks were required to provide 1000 times the data rates of 4G [30] which requires the emergence of new technologies. These technologies provide great promise for physical layer security but for many, this hasn't been investigated.

In this chapter, we review the current literature in the relevant areas of physical layer security. We begin with an example of a physical layer security scheme. We then outline the early evolution of physical layer security followed by the relevant literature, including the current state of the art, required for Chapters 4, 5 and 6.

3.1 Introduction

Physical layer security concerns any security measures and protocols occurring at layer 1 of the Open Systems Interconnection reference model (see Table 1). The core concept for security at this layer is to exploit the noise present in the communication channel to guarantee that a passive eavesdropper receives no useful information.

3.1.1 Shannon's Cryptosystem

The original model for studying physical layer security is known as *Shannon's cryptosystem* and comes from [68]. In this system, Alice and Bob share a secret key K which is used to encrypt Alice's message, M, into a codeword X. This system is noiseless, therefore Bob and Eve both receive X with no errors.

Definition 3.1.1: A system is said to have *perfect secrecy* when the mutual information

$$I(M; X) = 0,$$

where M is the original message and X is the encoded message. Equivalently $H(M \mid X) = H(M)$.

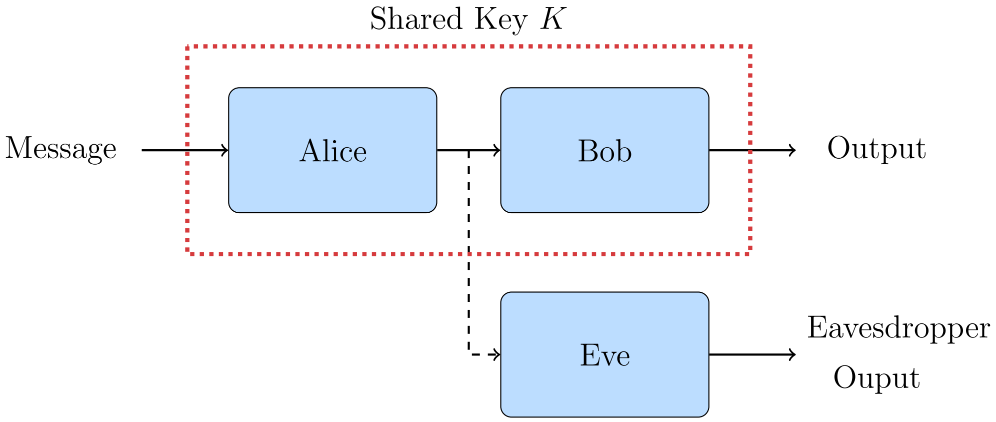


Figure 3.1: Shannon's cryptosystem

When a system has perfect secrecy, the best that an eavesdropper can do is randomly guess the transmitted message as they have gained no useful information from their observation, regardless of computational power, since M and X are statistically independent.

Theorem 3.1.2 (Shannon [68])**:** It is possible to achieve perfect secrecy if and only if $H(K) \geq H(M)$.

Proof. The proof uses a series of identities and inequalities. The following version follows that of [6, Proposition 3.1]. Since K, M and X are discrete, we have that $H\left(K|XM\right) \geq 0$, and so the first inequality follows:

$$\begin{aligned}
H(K) &\geq H\left(K\right) - H\left(K|XM\right) && (3.1)\\
&\geq H\left(K|X\right) - H\left(K|XM\right) && (3.2)\\
&= I\left(K;M|X\right) && (3.3)\\
&= H\left(M|X\right) - H\left(M|KX\right) && (3.4)\\
&= H\left(M|X\right) && (3.5)\\
&= H\left(M\right). && (3.6)
\end{aligned}$$

Here, Equation (3.2) follows since conditioning does not increase the entropy (see Definition 1.1.4) and Equations (3.3) and (3.4) follow by the definitions of mutual information (see Definition 1.1.5).

The quantity $H\left(M|KX\right) = 0$ by the definitions of M, K and X and so the equality in Equation (3.5) follows. Finally, if a coding scheme achieves perfect secrecy, then $H\left(M|X\right) = H\left(M\right)$ and so the result follows. □

Remark 3.1.3: The constraint introduced in Theorem 3.1.2 means that, in general, the secret key must be at least as long as the message Alice is transmitting to obtain perfect secrecy. In a realistic setup, this is highly restrictive as key management becomes difficult.

The assumed lack of noise and key management issues makes the results of Shannon's cryptosystem less applicable to a 'real world' noisy scenario however, the results show that it is possible to communicate securely without any requirement on computational power.

3.1.2 Wyner's Wiretap Channel

Shannon's results assume error free and noiseless channels, which means that Bob and Eve see the same message and Bob must gain their advantage only through the use of a secret key. Wyner proposed, in [73], a system without a shared key, but rather where the noise and channel properties are used to secure the communication. This approach solves both the problem of the length of the secret key required in Section 3.1.1 and the issue that the original cryptosystem was noiseless and therefore less applicable to real world systems. This setup is known as Wyner's *wiretap channel* and is the typical framework used for physical layer security. Alice is sending a message M, and encodes this to X. Bob receives Y and Eve receives a different signal, Z.

While maintaining a *reliable* communication with Bob, Alice now has the added requirement of ensuring their message is kept a secret from Eve.

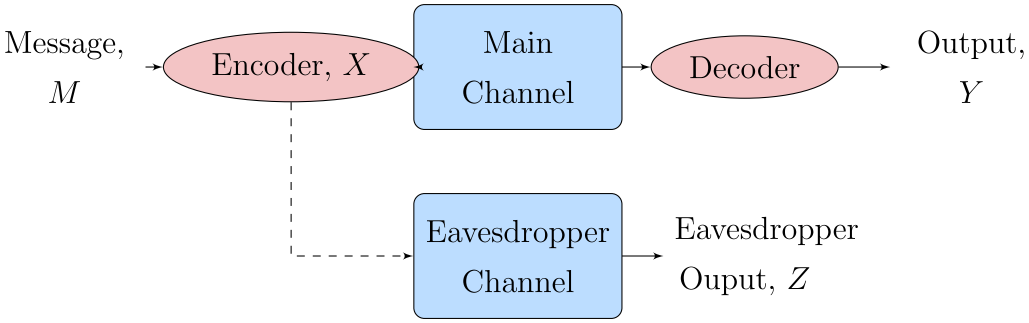


Figure 3.2: Wyner's wiretap channel [73].

3.1.3 Measuring secrecy

Intuitively, eavesdroppers fail if they make a mistake, that is, their error probability is 1 or very close. However Eve can always take a random guess, and they might get lucky, meaning their error probability is not quite 1. A better way of looking at secrecy is in terms of the encoding giving no useful information about a transmitted message, meaning that Eve's best method is to take a random guess. This is the definition of perfect secrecy (Definition 3.1.1) but as we have seen, this requires a shared secret key which is longer than the message and a noiseless environment. So how else may we measure secrecy when these ideal conditions no longer hold? Wyner's work requires that, asymptotically with the blocklength, n, of the code, the mutual information rate of the input message and Eve's observation is 0. This secrecy measure is known as *weak secrecy.*

Definition 3.1.4: The criterion for *weak secrecy* is met if for any $\epsilon > 0$ there exists some n such that

$$\frac{1}{n}H(M) - \frac{1}{n}H(M|Z) \leq \epsilon \tag{3.7}$$

or equivalently

$$\frac{1}{n}I(M;Z) \to 0 \tag{3.8}$$

where the limit is in the blocklength and taken symbol by symbol.

As the blocklength n increases, the expression above tends to 0 regardless of the scheme used. Thus it is possible to meet the criterion with a flawed scheme. The criterion was later strengthened in [54] to overcome these issues of aggregate information leakage - to obtain *strong secrecy.*

Definition 3.1.5: The criterion for *strong secrecy* is met if the limit

$$I(M;Z) \to 0, \tag{3.9}$$

tends to 0 with an increased blocklength.

Both M and Z are of length n, and the mutual information in Equation 3.9 is taken over the symbols and not the entire block. That is,

$$I(M_1, \ldots, M_n; Z_1, \ldots, Z_n) \to 0. \tag{3.10}$$

This criterion depends on the probability distribution of the message, and it has been argued (for example in [8]) that this is a drawback of using strong

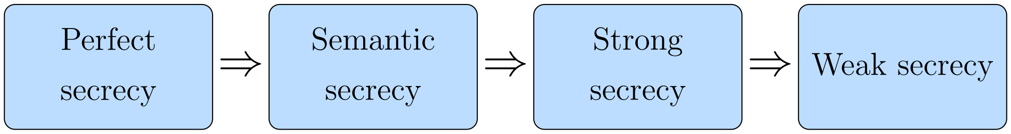


Figure 3.3: Implication chain for secrecy metrics

secrecy as a metric. This had led to metrics such as semantic secrecy, coined in [4], which is related to the cryptographic definition of semantic security. Semantic security removes this dependence on the probability distribution of the input message by measuring secrecy in terms of the *Advantage* of the eavesdropper. Having an advantage bounded above by some security threshold δ means that knowledge of Z may increase the probability of guessing some function of M by at most δ. Semantic secrecy is the 'strongest' metric, as semantic secrecy implies strong and weak secrecy. The full implication chain of the outlined metrics can be seen in Figure 3.3. We note that when a system has perfect secrecy, all outlined secrecy metrics are equal to zero.

For the purposes of this thesis, we concentrate on the idea of strong secrecy, as this is a fundamental property relying only on the channel characteristics. It has been shown that when a system meets the strong secrecy criterion, Eve's error rate approaches 1 exponentially fast [61], regardless of their decoding procedure.

3.1.4 Secrecy Capacity

Now that we have a measure of secrecy, we can return to the overarching question of how much information can Alice send securely and reliably to Bob in the presence of Eve in Wyner's model (Figure 3.2). Recall the notion of the channel capacity, given in Definition 2.1.4, which characterises the maximum rate at which Alice can transmit reliably. We now wish to extend this concept to account for an additional secrecy constraint. Firstly, we define the parameters of a code with reliability and secrecy constraints.

Definition 3.1.6: A code with parameters

$$(n, k_n, \epsilon_n, \delta_n)$$

is one where n denotes the blocklength, k denotes the number of distinct codewords in the codebook. The parameter ϵ_n denotes the *error threshold*, which is the maximum tolerable error rate for the system. The secrecy requirement is denoted by δ_n and is the maximum tolerable secrecy leakage subject to some measure of secrecy.

If strong secrecy is the metric of choice, then δ_n would be an upper bound on the mutual information in Equation (3.9).

Now the secrecy capacity is the maximum *achievable* rate for which codes above exist.

Definition 3.1.7: The *secrecy capacity*, C_s, is the supremum of all rates $R = k_n/n$ such that there exists sequences $(n, k_n, \epsilon_n, \delta_n)$ codes with the following properties

$$\lim_{n\to\infty} \frac{k_n}{n} \geq R$$
$$\lim_{n\to\infty} \epsilon_n = \lim_{n\to\infty} \delta_n = 0.$$

Maintaining reliability and secrecy seem to be conflicting goals. However, it is possible to achieve both simultaneously, with the rate of communication taking a penalty. This is perhaps the most important result, that the secrecy capacity can be non zero, as it gives traction to the field of physical layer security.

The central idea is to send useless information up to the capacity of the eavesdropper channel and then use the remaining rate to send secure communications across the main channel. This relies on the main channel having some sort of advantage over the eavesdropper.

Wyner was the first to establish the secrecy capacity for a discrete memoryless channel in the case of a degraded channel (a channel which is affected by noise) in [73]. This was then generalised by Csiszár and Körner for a non degraded case in [17] as follows.

Theorem 3.1.8 ([17])**:** For a discrete memoryless wiretap channel, with encoded message X, the secrecy capacity is characterised to be

$$C_s = \max\{I(V;Y) - I(V;Z)\}, \tag{3.11}$$

where Y is the random variable associated with the legitimate channel output and Z is the random variable associated with the eavesdrop channel output. The maximum is taken over all random variables V and X satisfying the Markov chain relationship $V - X - (Y, Z)$.

Remark 3.1.9: The V in Equation (3.11) can be thought of as the variable introducing noise in the channel, in Figure 3.2. Note that without the second term, this is the capacity of the main channel (see Theorem 2.1.5) and therefore the the secrecy capacity is similar to the difference between the main channel capacity and eavesdropper channel capacity.

3.2 The Gaussian Wiretap Channel

SISO Wiretap Channel

The most fundamental wiretap model is that of the Gaussian channel, described in Definition 2.2.2. For the single input single output (SISO) Gaussian channel, at time slot t, Bob receives

$$y = h_B x + n_B$$

and Eve receives

$$z = h_E x + n_E$$

where h_B and h_E denote the Gaussian channels and n_B and n_E are additive white Gaussian noise (AWGN) with zero mean and noise variance σ_B^2 and σ_E^2 respectively. Here, the secrecy capacity has been fully established.

Theorem 3.2.1 ([43, Theorem 1])**:** The secrecy capacity of the SISO Gaussian wiretap channel with a power constraint P is

$$C_s = \max\left\{\log\left(1+\frac{P|h_B|^2}{\sigma_B^2}\right) - \log\left(1+\frac{P|h_E|^2}{\sigma_E^2}\right), 0\right\}.$$

This result shows that the secrecy capacity in this case is equivalent to the difference of the capacity of the main channel and the capacity of the eavesdropper channel. Therefore it is possible to achieve a positive secrecy capacity if and only if the main channel capacity is higher than the eavesdrop channel capacity or equivalently, the SNR from Alice to Bob is higher than the SNR from Alice to Eve

$$\frac{|h_B|^2}{\sigma_B^2} > \frac{|h_E|^2}{\sigma_E^2}.$$

3.2.1 Gaussian MIMO Wiretap Channels

The conventional point to point results, also known as Single Input Single Output (SISO) systems are well understood in terms of physical layer security. We have already seen in Section 3.2 that the Gaussian wiretap channel's secrecy capacity is known, for example. However, many of these results do not generalise to the multiple antenna regime. Let N_A denote the number of antennas at the transmitter, N_B denote the number of antennas at the legitimate receiver and N_E the number of antennas at the eavesdropper.

The MIMO wiretap channel is the multiple antenna extension of the traditional point to point wiretap channel as depicted in Figure 3.4.

Definition 3.2.2: The (N_A, N_B, N_E) *MIMO wiretap channel* is one where Alice, Bob and Eve have N_A, N_B and N_E antennas, respectively, and is defined by the following relationships. Alice sends message vector $\mathbf{x}$ while Bob and Eve receive $\mathbf{y}$ and $\mathbf{z}$, respectively, defined as

$$\mathbf{y} = H_B\mathbf{x} + \mathbf{n_B}, \tag{3.12}$$

$$\mathbf{z} = H_E\mathbf{x} + \mathbf{n_E}, \tag{3.13}$$

where $\mathbf{n_B}$ and $\mathbf{n_E}$ are circularly symmetric Gaussian noise vectors, each with zero mean and identity covariance matrix. The system is subject to a power constraint P such that the covariance matrix of the input signal, Q, is bounded above by P. That is,

$$\operatorname{Tr} Q = \sum_{i=1}^{N_A} \mathbb{E}[\mathbf{x}_i\mathbf{x}_i^*] \leq P.$$

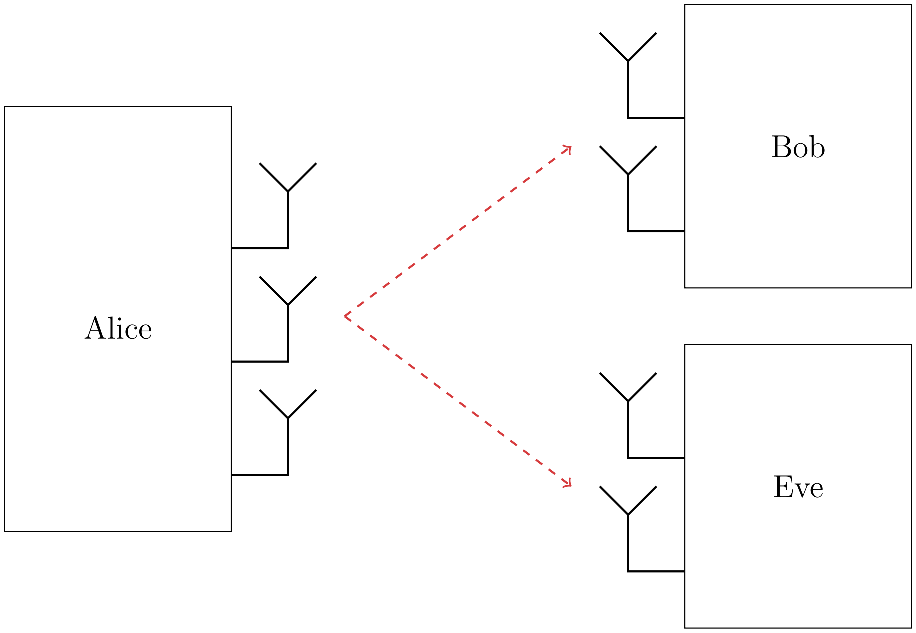


Figure 3.4: (3,2,2) MIMO wiretap channel.

The secrecy capacity for this type of wiretap channel was found in [40, 41, 56] and is stated in the following theorem.

Theorem 3.2.3: For the Gaussian MIMO wiretap channel, the secrecy capacity is, to take the form

$$\max_{Tr(Q)\leq P} \log\det(I + H_B Q H_B^T) - \log\det(I + H_E Q H_E^T), \tag{3.14}$$

$$\text{Such that} \quad Q \succeq 0$$

where P is the power constraint of the system.

The general solution to Equation (3.14) is unknown, since the optimisation is a non convex one and thus difficult to solve. It would be desirable to know the covariance matrix Q which maximises the secure transmission rate, as this would give an insight to the optimal secure signalling scheme. It is known that the solution to this is a low rank matrix [56] for the Gaussian wiretap channel however there is no known way of constructing this low rank matrix.

Table 3.1 gives an overview of the scenarios where Equation (3.14) is fully understood - that is, known in closed form and an optimal signalling scheme is known. Here, the single antenna is a subset of the multiple antenna case.

Number of Antennas at			
Alice	**Bob**	**Eve**	**Secrecy capacity fully understood?**
Single	Single	Single	Yes
Multiple	Multiple	Single	Only for (2,2,1) [66]
Multiple	Single	Multiple	Yes [40]
Multiple	Multiple	Multiple	No

Table 3.1: Overview of open cases for the secrecy capacity of the Gaussian MIMO wiretap channel. Single means one antenna where multiple means any positive integer, including one. Note that the work in Chapter 4 extends the knowledge in the highlighted row.

Fully Understood Scenarios

There are some families of MIMO systems where the secrecy capacity and optimal transmit strategy has been characterised. Some significant cases are outlined below.

MISOME Wiretap Channel

The secrecy capacity for a Multiple-Input Single-Output channel with multiple eavesdrop antennas (MISOME) with power constraint P is fully understood. This is the channel where there are multiple antennas at Alice, a single antenna at Bob and any number of antennas at Eve ($N_A \geq 1$, $N_B = 1$ and $N_E \geq 1$). The secrecy capacity from Equation (3.14) was derived explicitly [40] in closed form (without a maximisation) to be

$$C_s = \frac{1}{2} \log \left(\lambda_{\max}(I + P h_B h_B^T, I + P H_E^T H_E) \right) \tag{3.15}$$

where h_B is the main channel vector and H_E is the eavesdrop channel matrix. Here, $\lambda_{\max}$ denotes the largest generalised eigenvector (see Definition 1.3.4) of the two matrices $I + P h_B h_B^T$ and $I + P H_E^T H_E$. The authors showed that the scheme which is optimal for secrecy, achieving the secrecy capacity, is to transmit in the direction of the generalised eigenvector which corresponds to $\lambda_{\max}$.

This is the only ‘general’ multiple antenna case which is fully understood. That is, no additional requirements other than the Gaussian channel are necessary for these results to hold.

(2,2,1) channel

For the case where Alice and Bob have two antennas and Eve has one, known as the ‘(2,2,1) channel’, the form of Q is explicitly known [66]. Since the solution must be low rank, and the matrix in this case has dimensions 2×2, the rank of Q must be 1. The proof in this paper proposes Gaussian signalling as a scheme which achieves the optimal rate, and then provides a tight upper bound to meet the rate achieved.

It should be noted that [51] proposed an algorithm to solve the saddle point of a min-max problem to solve Equation (3.14). Their work gives an algorithm for solving the secrecy capacity of a general MIMOME wiretap.

Special cases

Table 3.1 gives the current state of knowledge for the general Gaussian MIMO wiretap channel. Although the general cases remain largely open, improvements have been made for cases with more constraints. A few are outlined below.

- **Constrained power:** When the input covariance matrix Q is bounded above by a general matrix power constraint $S \succeq 0$, the secrecy capacity is known in closed form and Q has been specified in [12].
- **Parallel channels:** For a number of *Gaussian parallel channels* a number of results have been established for the broadcast channel. The secrecy capacity regions were established in [45] and the secrecy capacity for transmitting a common message were found in [39].
- **Full rank channels:** A closed form expression for full rank Q has been found in [49]. This work was then extended to the rank deficient case in [50]. In these works, it is a necessary condition that the SNR is finite.

Also of note is the case of the isotropic eavesdropper, that is an eavesdropper with one parameter (the channel power gain). This differs to any cases outlined in this chapter, as full eavesdropper channel state information (CSI) is not considered. However it is proved in [50] that the case of an isotropic eavesdropper is the worst case for the MIMO wiretap channel. The optimal signalling strategy is known here in closed form.

3.3 Achieving secrecy

So far, we have encountered the secrecy capacity and measures of secrecy for communication systems. While these tell us that we *can* submit with perfect secrecy, they do not tell us *how* to do so. The secrecy capacity, C_s, being positive tells us that a code exists which can achieve the secrecy capacity but finding such codes is another problem space.

For the SISO Gaussian channel, the secrecy capacity is achieved using the full available power and Gaussian signalling [43]. Gaussian signalling is also the way to achieve capacity for a Gaussian channel, without a secrecy constraint, and this is why the secrecy capacity in this instance is exactly the difference of the capacity of the main channel and that of the eavesdropper channel (see Theorem 3.2.1). When considering semantic security as our secrecy metric, wiretap lattice coding may achieve the secrecy capacity [46], these are used since they maximise the error probability for the eavesdropper at their decoder [26]. For example, a particular coding scheme based on polar lattices [47] has been shown to achieve the secrecy capacity for the Gaussian case.

More generally, to confuse an eavesdropper, Alice wishes to exploit the difference in their channel when choosing a secure message. If the signal is based on the main (Alice to Bob) channel, then Eve's lack of knowledge will prevent them from decoding the message. This is the basis of secret key generation at physical layer, as Alice can use their channel with Bob to create a secret key. As seen previously, key storage and key generation is impractical for a number of reasons but this concept is used in the codebook design for secrecy at physical layer. For example, Alice may use *index modulation* where the information is transmitted in the *index of the codeword* rather than the codeword itself. An example of an index modulation scheme for secrecy is the work of [38] where Alice uses CSI of the legitimate channel to generate an integer value. This integer value is used as an antenna rotation index, which is easily undone by Bob since perfect CSI is assumed. They show that by doing this, their scheme had provable perfect secrecy. The work in Chapter 6 builds on such ideas to design a secure coding scheme and this specific scheme is explored in further detail in Section6.1.

Chapter 4

The Secrecy Capacity of a MIMOSE Wiretap Channel

In this chapter, the secrecy capacity for a Multiple-Input Multiple-Output (MIMO) wiretap channel is discussed. We consider a passive eavesdropper with a single antenna with Gaussian channels. The main result of this work (Theorem 4.2.1) provides a concavity result for an equivalent problem to find the secrecy capacity of such a system. This is done by reformulating the secrecy capacity (a non convex optimisation problem with no general solution) to a maximisation of a function with a scalar input. It is then shown that this equivalent function has a concave region, meaning that existing convex solvers (see Section 1.3.2) may be used to efficiently find the maximum and therefore the secrecy capacity. This work addresses the open problem of the secrecy capacity for a MIMO wiretap channel and contributes to the MIMO channel with a Single Eavesdropper (MIMOSE). The basis of this work has been published as joint work with Oliver Johnson and Robert Piechocki in [13], where all simulation and technical analysis was undertaken by myself as first author. Section 4.4 is additional to this publication.

4.1 Introduction

Multiple antenna systems play a large role in achieving higher capacities and thus are central in 5G technologies, with ‘massive’ MIMO being a central technology for 5G and future wireless [34]. Security for any modern day system is vital however there are several fundamental questions which remain open with regards to the physical layer security of a MIMO channel when compared to the equivalent model for point to point single antenna systems. Indeed, the secrecy capacity for a Gaussian MIMO wiretap channel, introduced in Section 3.2.1, is one of these open problems. The work in this chapter aims to addresses this, contributing a theorem which gives a region where a MIMOSE channel has a concave secrecy capacity equation. Knowing when the equation is concave allows for the problem to be efficiently solved, giving the secrecy capacity and thus the maximum rate for secure communications for the given channel.

4.1.1 Theoretical setup

We begin by laying out the notation and system setup. This work concerns a MIMO channel with N_A transmit antennas and N_B receive antennas at the legitimate receiver. The legitimate users, Alice and Bob, are communicating in the presence of a passive eavesdropper, Eve, with N_E antennas. For the results of this chapter to hold, Eve has a single eavesdrop antenna, that is $N_E = 1$ as depicted in Figure 4.1.

The channel between the transmitter and the legitimate receiver shall be referred to as the *main channel* while the channel between the transmitter and the eavesdropper shall be referred to as the *eavesdropper channel.* Their channel matrices are described by H_B, an $N_B \times N_A$ matrix for the main channel and H_E, an $N_E \times N_A$ matrix for the eavesdropper channel.

The input signal, $\mathbf{x}$, is drawn from a distribution with zero mean and covariance matrix $Q \succeq 0$, which is a positive semidefinite matrix. The received vectors at Bob and Eve, denoted $\mathbf{y}$ and $\mathbf{z}$ respectively, are:

$$\mathbf{y} = H_B\mathbf{x} + \mathbf{n}_B,$$
$$\mathbf{z} = H_E\mathbf{x} + \mathbf{n}_E.$$

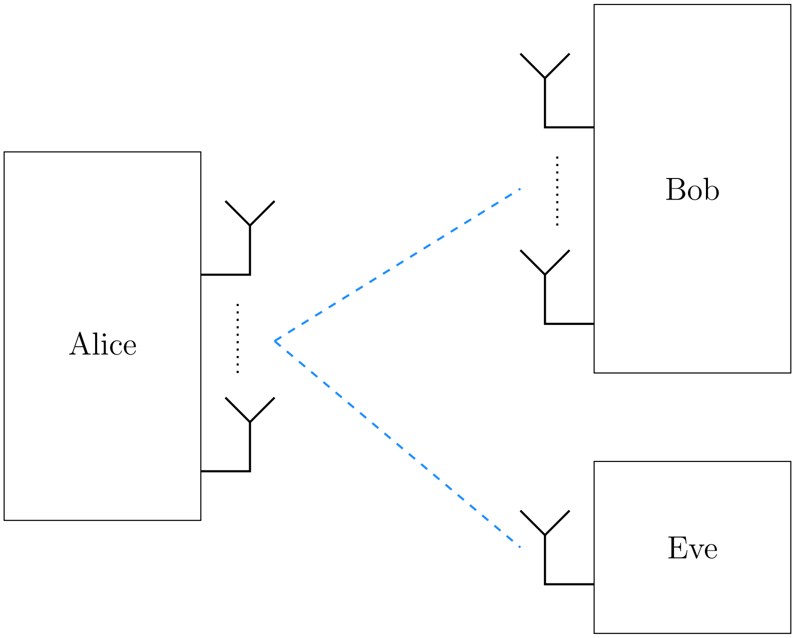


Figure 4.1: The MIMOSE wiretap channel

Definition	**Symbol**
Number of antennas at Alice	N_A
Number of antennas at Bob	N_B
Number of antennas at Eve	N_E
Main channel matrix	H_B
Eavesdropper channel matrix	H_E
Transmitted signal	$\mathbf{x}$
Received signal	$\mathbf{y}$
Eavesdropped signal	$\mathbf{z}$
Covariance matrix of input signal	Q
Power constraint of input	P

Table 4.1: Notation for Chapter 4.

where $\mathbf{n}_B$ and $\mathbf{n}_E$ are the Gaussian noise vectors for the two channels

$$\mathbf{n}_B \sim \mathbb{C}N(0, I_{N_B})$$

and

$$\mathbf{n}_E \sim \mathbb{C}N(0, I_{N_E})$$

where each element of the noise vector is statistically independent. Similarly, the channel matrices are modelled with IID entries assuming independence between each antenna element. The matrix I_k denotes the identity matrix of size $k \times k$. The input signal is subject to a power constraint P, meaning that the trace of the covariance matrix, Q, is bounded above by this quantity. That is,

$$\operatorname{Tr} Q = \sum_{i=1}^{N_A} \mathbb{E}[\mathbf{x}_i \mathbf{x}_i^*] \leq P.$$

Without the power constraint above, the capacity is theoretically infinite, which does not provide much insight in a practical setting.

4.1.2 Secrecy capacity

The open problem we are addressing in this chapter is the secrecy capacity for the outlined system setup. Recall from Equation (3.14) that the secrecy capacity, C_s, for the MIMO wiretap channel was established in [56], [40] and [41] to be of the form

$$C_s = \max_{Q:\operatorname{Tr}(Q) \leq P} \log\det(I_{N_B} + H_B Q H_B^*) - \log\det(I_{N_E} + H_E Q H_E^*) \tag{4.1}$$

where we note that, since the mean of the input signal is always zero, the maximum is being taken over all input distributions satisfying the power constraint.

The optimisation problem in Equation (4.1) is not easily solved for Q and the solution is only known for a subset of scenarios, which were outlined in

Section 3.2.1, it remains open in the general case. The difficulty lies in the fact that the optimisation is not convex and thus analytically challenging. Knowing the optimal Q is useful for a number of reasons, some of which are outlined below.

- The mean of the input signal is always zero, so the covariance matrix, Q, is the characterising variable for the input distribution.
- The input covariance gives details for the optimal input scheme for secrecy and rate requirements.
- Knowledge of the optimum covariance matrix gives the true secrecy capacity.
- Once the secrecy capacity is known, any rate of transmission below this is secure by definition, giving a secure region for reliable rates of transmission.

The key contribution of this chapter is for the Gaussian MIMO wiretap channel with a *single antenna eavesdropper*, a subset of the unknown MIMOSE family of wiretap channels. The secrecy capacity is examined for this open problem and a region is established where the problem is provably concave. The concavity of the problem gives an efficient method of determining the optimal input covariance matrix associated with the secrecy capacity of a system. The scheme given is valid for the MIMOSE channel where the receiver has at least as many antennas as the transmitter. That is, $N_B \geq N_A$ and $N_E = 1$.

This family of antenna configurations overlaps with only two known cases, the point to point single antenna case where Alice, Bob and Eve each have one antenna, and the so called '(2,2,1)' case. Both of these are detailed in Section 3.2.1. Our results are compared with their results in Section 4.2.2.

The theory of this chapter goes as follows: the secrecy capacity equation is reformulated into a problem which is convex, this allows existing convex optimisation tools and software to find the optimal solution to Equation (4.1).

The proof relies on properties of symmetric matrices and functions of the channel and thus for ease of notation we define the following positive semidefinite symmetric $N_A \times N_A$ matrices which are used in the statement of Theorem 4.2.1 and throughout the proof:

$$K_B = (H_B^* H_B)^{\frac{1}{2}}, \tag{4.2}$$

$$K_E = (H_E^* H_E)^{\frac{1}{2}}. \tag{4.3}$$

4.2 Concavity region for the secrecy capacity

The key limitation in solving Equation (4.1) is the fact that it is non-convex. In order to exploit existing convex solvers, we must first reformulate the secrecy capacity equation to an equivalent but tractable optimisation problem. Recall from Theorem 1.3.7 that $\log\det(\cdot)$ is known to be concave and twice differentiable for positive semidefinite arguments. It follows that each individual $\log\det(\cdot)$ term in Equation (4.1) is concave. This can be seen by considering their arguments. Since Q is a covariance matrix, it is restricted to positive semidefinite matrices by definition. The identity matrix is trivially a positive semidefinite matrix and thus the terms

$$I_{N_B} + H_B Q H_B^*$$

and

$$I_{N_E} + H_E Q H_E^*$$

will also be positive semidefinite. This means that both of the terms

$$\log\det(I_{N_B} + H_B Q H_B^*)$$

and

$$\log\det(I_{N_E} + H_E Q H_E^*)$$

are concave. However, in general, their difference is neither convex nor concave. We will reformulate the problem in order to restrict the problem space to a region where the difference is concave. Broadly speaking, this is done by fixing the second term and then varying its value. Hence we define the following problem:

$$\begin{aligned} &\max_{\mathrm{Tr}\,(Q)\leq P} \log\det(I_{N_B} + H_B Q H_B^*) - \log(s), \\ \text{such that} \quad & s = \det(I_{N_E} + H_E Q H_E^*) \\ \text{and} \quad & Q \succeq 0. \end{aligned} \tag{4.4}$$

The following work is constrained to a single eavesdrop antenna since, generally speaking, $\det(\cdot)$ is not a convex constraint. When the problem space is limited in this way, the matrix argument $I_{N_E} + H_E Q H_E^*$ is a scalar value. Since $\log\det(I_{N_B} + H_B Q H_B^*)$ is concave and the maximisation is taken over a convex set, it can be seen that by fixing the value of s, this becomes a concave problem.

With s fixed, Equation (4.4) is concave however it is no longer equivalent to Equation (4.1). In order to bridge this gap, we must vary our value of s and take an overall maximum. This is the overarching idea which is formally laid out in the following section.

For the optimal value of s, Equation (4.4) is an equivalent problem to Equation (4.1) and consequently will yield the same solution.

4.2.1 Statement of theorem

Each value of s gives a separate convex optimisation problem in Equation (4.4). For each optimisation, the output is a corresponding covariance matrix

Q and the maximum value of the argument. We aim to vary s and take the maximum over each of the aforementioned outputs.

We begin by defining functions of the input covariance matrix Q

$$s(Q) = \det(I_{N_E} + H_E Q H_E^*) \tag{4.5}$$

and

$$f(Q) = \log\det(I_{N_B} + H_B Q H_B^*) - \log s(Q). \tag{4.6}$$

We wish to fix values of s, where $s = s(Q)$ for some Q, and perform a convex optimisation for $f(Q)$ given this constraint. We then wish to take the maximum value of $f(Q)$ over all values of s. Therefore we define $\theta(\cdot)$ as:

$$\theta(s) = \max_{Q:s(Q)=s} f(Q). \tag{4.7}$$

A plot of $\theta(s)$ can be seen in Figure 4.2. Motivated by the apparent concavity of the simulation results, we aim to prove the concavity regions of these curves. The simulations and figures presented in this chapter runs the optimisation presented above for fixed values of s using convex optimisation software *CVX: Matlab Software for Disciplined Convex Programming* [29] but the theory holds for an arbitrary convex solver.

Finding the secrecy capacity is now a case of finding the maximum of $\theta(s)$. This is facilitated by the following Theorem, which gives a concavity result for θ which is the main result of our paper [13].

Let Q_i be a matrix achieving the maximum value in Equation (4.7) corresponding to s_i, that is $f(Q_i) = \theta(s_i)$, for $i \in \{1, 2\}$. By definition

$$s_i = I_{N_E} + H_E Q_i H_E^* \tag{4.8}$$

where the det is no longer required since $N_E = 1$. Without loss of generality, assume that $s_1 \geq s_2$. Let s_t be a convex combination of s_1 and s_2

$$s_t = ts_1 + (1-t)s_2 \tag{4.9}$$

for $t \in [0, 1]$.

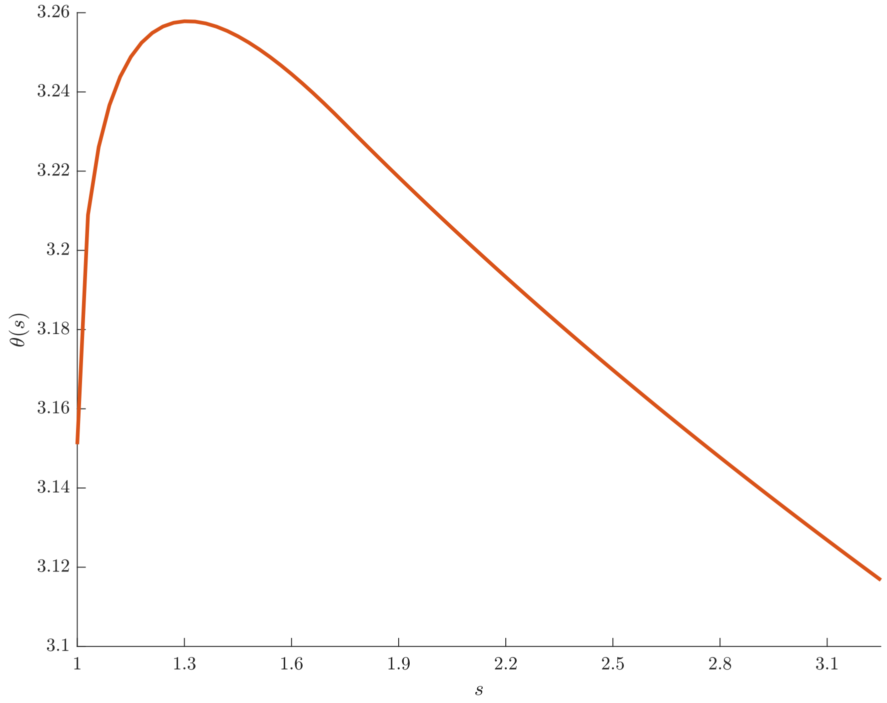


Figure 4.2: $\theta(s)$ vs s for $N_A = 2$, $N_B = 3$, $N_E = 1$ and $P = 10$ for a particular H_B and H_E.

Theorem 4.2.1: For $N_E = 1$ and any $N_B \geq N_A$, then

$$\theta(s_t) \geq t\theta(s_1) + (1-t)\theta(s_2), \tag{4.10}$$

if the matrices K_B and K_E from Equations (4.2) and (4.3) satisfy

$$\frac{s_1}{\|K_B^{-1}K_E^2K_B^{-1}\|_F} - 1 \geq \max\{\lambda_{\max}(H_BQ_1H_B^*), \lambda_{\max}(H_BQ_2H_B^*)\}. \tag{4.11}$$

4.2.2 Overlap with existing results

For the antenna configuration $N_E = 1$, $N_B \geq N_A$ required for Theorem 4.2.1 to hold there is only one fully understood case. This is the '(2,2,1)' case [66], where $N_A = 2$, $N_B = 2$ and $N_E = 1$.

Example 4.2.2 (2,2,1)**:** Figure 4.3 shows that the theoretical secrecy capacity found in [66] matches the maximum value of $\theta(s)$.

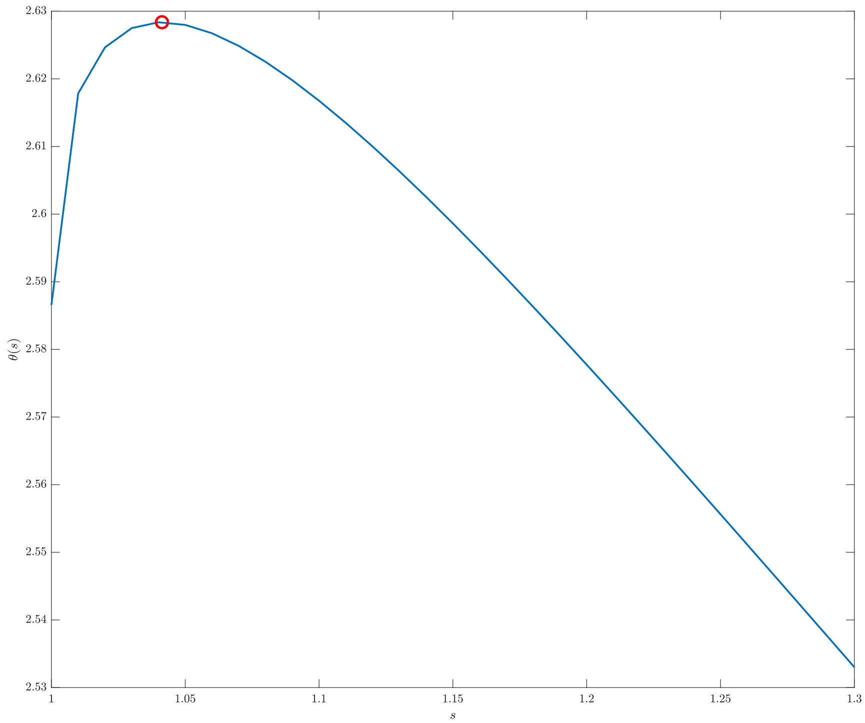


Figure 4.3: $\theta(s)$ vs s for the (2,2,1) case. The red mark indicates the theoretical secrecy capacity from the paper [66].

4.3 Proof of the concave region

The main argument in the proof of Theorem 4.2.1 involves a Taylor expansion of a matrix term which is then bounded at the second order. The proof can be broken down into three key steps as follows.

1. Firstly, we consider the function $\theta(s)$, defined in Equation (4.7), for a convex combination of inputs, s_t (Equation (4.9)). Using Lemma 4.3.1, which is a second order concavity bound for the log det, we find a lower bound for $\theta(s_t)$.

2. We then minimise the difference between the bound from Step 1 with the lower bound required for concavity.

3. Finally, we rewrite these bounds in terms of symmetric matrices which allows us to exploit properties of the Frobenius norm resulting in the conditions stated in Theorem 4.2.1.

4.3.1 Step 1

In this step of the proof, concavity results from [15] are applied to the function $\theta(\cdot)$ defined in Equation (4.7). The use of these results allows us to find a tighter lower bound than the usual concavity lower bounds.

Lemma 4.3.1: Courtade et al. [15, Lemma 15] For positive definite matrices A and B and for any $t \in [0, 1]$

$$\begin{aligned}\log\det(tA + (1-t)B) \geq{}& t\log\det(A) + (1-t)\log\det(B)\\ &+ \frac{t(1-t)}{2\max\{\lambda^2_{\max}(A), \lambda^2_{\max}(B)\}}\|A - B\|_F^2, \end{aligned} \tag{4.12}$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue and $\|\cdot\|_F$ is the Frobenius norm.

For ease of notation, define

$$\mathcal{C}_{\max}(A, B) = \frac{\|A - B\|_F^2}{2\max\{\lambda^2_{\max}(A), \lambda^2_{\max}(B)\}}. \tag{4.13}$$

Considering the linear combination $Q_t = tQ_1 + (1-t)Q_2$, it can be seen that Q_t satisfies the constraint $s(Q_t) = s_t$ since $N_E = 1$ and

$$\begin{aligned}
s_t =& ts_1 + (1-t)s_2\\
=& t(I_{N_E}s_1 + H_E Q_2 H_E^*) + (1-t)(I_{N_E}s_2 + H_E Q_2 H_E^*)\\
=& I_{N_E} + H_E(tQ_1 + (1-t)Q_2)H_E^*\\
=& I_{N_E} + H_E Q_t H_E^*\\
:=& s(Q_t).
\end{aligned}$$

Hence

$$\theta(s_t) \geq f(Q_t). \tag{4.14}$$

By Lemma 4.3.1, taking $A = I_{N_B} + H_B Q_1 H_B^*$ and $B = I_{N_B} + H_B Q_2 H_B^*$ then $f(Q_t)$ is bounded below as follows.

$$\begin{aligned}
f(Q_t) =& \log\det(I_{N_B} + H_B Q_t H_B^*) - \log s_t \qquad (4.15)\\
\geq& t \log\det(I_{N_B} + H_B Q_1 H_B^*) + (1-t)\log\det(I_{N_B} + H_B Q_2 H_B^*)\\
&- \log s_t + t(1-t)\mathcal{C}_{\max} A, B).
\end{aligned}$$

Rewriting the lower bound in Equation (4.15) gives

$$\begin{aligned}
&t(\log\det(I_{N_B} + H_B Q_1 H_B^*) - \log s_1)\\
&\quad + (1-t)(\log\det(I_{N_B} + H_B Q_2 H_B^*) - \log s_2) + t(1-t)\mathcal{C}_{\max} A, B)\\
&\quad + t\log s_1 + (1-t)\log s_2 - \log(ts_1 + (1-t)s_2)\\
=& tf(Q_1) + (1-t)f(Q_2) + t(1-t)\mathcal{C}_{\max} A, B)\\
&\quad + t\log s_1 + (1-t)\log s_2 - \log(ts_1 + (1-t)s_2)\\
=& t\theta(s_1) + (1-t)\theta(s_2) + t(1-t)\mathcal{C}_{\max} A, B)\\
&\quad + t\log s_1 + (1-t)\log s_2 - \log(ts_1 + (1-t)s_2), \qquad (4.16)
\end{aligned}$$

since each of the Q_i are optimal by definition.

4.3.2 Step 2

In this step, we aim to minimise the difference between

$$tf(Q_1) + (1-t)f(Q_2)$$

in Equation (4.16) and the upper bound, $\theta(s_t)$ as defined in Equation (4.7). To do this, we introduce a constant $\kappa(s_1, s_2)$ and show that the following Lemma holds.

Lemma 4.3.2: For $t \in [0,1]$,

$$\begin{aligned} t\log(s_1) + &(1-t)\log(s_2) - \log(ts_1 + (1-t)s_2) \\ &\geq -t(1-t)\kappa(s_1, s_2), \end{aligned} \tag{4.17}$$

for

$$\kappa(s_1, s_2) = \frac{(s_1 - s_2)^2}{2s_1^2}. \tag{4.18}$$

Proof. Define a function g as:

$$g(t) := t\log(s_1) + (1-t)\log(s_2) - \log(ts_1 + (1-t)s_2) + t(1-t)\kappa(s_1, s_2) \tag{4.19}$$

where $\kappa(\cdot,\cdot)$ is a constant. We wish to show that $g(t) \geq 0$ for all $t \in [0,1]$.

By construction, $g(0) = g(1) = 0$ and therefore $g(t) \geq 0$ in the interval $t \in [0,1]$ is equivalent to $g(t)$ being concave in this interval or when $g''(t) \leq 0$.

The second derivative of g with respect to t is:

$$g''(t) = -2\kappa(s_1, s_2) + \frac{(s_1 - s_2)^2}{s_t^2}.$$

Since $s_2 \leq s_1$, $g(t)$ is concave for the value of $\kappa(s_1, s_2)$ in Equation (4.18) and thus $g(t) \geq 0$ on the interval. □

4.3.3 Step 3

Combining Lemma 4.3.2 with Equation (4.12), we see that Theorem 4.2.1 will follow from Equation (4.16) if

$$\begin{aligned}\frac{\|A-B\|_F^2}{2\max\{\lambda_{\max}^2(A),\lambda_{\max}^2(B)\}} &\geq \kappa(s_1,s_2)\\ &\geq \frac{(s_1-s_2)^2}{2s_1^2}\end{aligned} \tag{4.20}$$

where, as before,

$$A := I_{N_B} + H_B Q_1 H_B^* \tag{4.21}$$

and

$$B := I_{N_B} + H_B Q_2 H_B^*. \tag{4.22}$$

Writing $\overline{Q} := Q_1 - Q_2$ for simplicity, the Frobenius norm on the left of Equation (4.20) can be rewritten as

$$\begin{aligned}\|A-B\|_F^2 &= \operatorname{Tr}(H_B\overline{Q}H_B^*H_B\overline{Q}H_B^*)\\ &= \operatorname{Tr}(\overline{Q}K_B^2\overline{Q}K_B^2)\\ &= \operatorname{Tr}((K_B\overline{Q}K_B)(K_B\overline{Q}K_B))\\ &= \operatorname{Tr}(RR) = \operatorname{Tr}(RR^*)\\ &= \|R\|_F^2\end{aligned} \tag{4.23}$$

where R is the symmetric matrix

$$R := K_B\overline{Q}K_B. \tag{4.24}$$

In order to retrieve the value of $\overline{Q}$ from R requires that $H_B^*H_B$ is invertible. This implies that $N_B \geq N_A$.

Similarly, considering the numerator of the right hand side of Equation

(4.20) gives:

$$\begin{aligned}(s_1 - s_2)^2 &= (H_E \overline{Q} H_E^*)^2 \\ &= \mathrm{Tr}\,(\overline{Q} K_E^2 \overline{Q} K_E^2) \\ &= \mathrm{Tr}\,((K_E \overline{Q} K_E)(K_E \overline{Q} K_E)) \\ &= \mathrm{Tr}\,(RTRT) \\ &\leq \|RT\|_F^2 \qquad (4.25) \\ &\leq \|R\|_F^2 \|T\|_F^2. \qquad (4.26)\end{aligned}$$

where T is the symmetric matrix

$$T := K_B^{-1} K_E^2 K_B^{-1}. \qquad (4.27)$$

Here, Equation (4.25) follows by Cauchy-Schwarz, for any matrix C,

$$\mathrm{Tr}\,(C^2) \leq \mathrm{Tr}\,(C^* C) = \|C\|_F^2,$$

and Equation (4.26) follows by the submultiplicative property of the Frobenius norm (see Theorem 1.3.10). Since both R and T are symmetric, the following holds:

$$\mathrm{Tr}\,(RTRT) \leq \|RT\|_F^2. \qquad (4.28)$$

$$(s_1 - s_2)^2 \leq \|R\|_F^2 \|T\|_F^2. \qquad (4.29)$$

Therefore the inequality in Equation (4.20) is satisfied when

$$\frac{\|R\|_F^2}{2 \max\{\lambda_{\max}^2(A), \lambda_{\max}^2(B)\}} \geq \frac{\|R\|_F^2 \|T\|_F^2}{2 s_1^2}. \qquad (4.30)$$

Since each of $\lambda_{\max}^2(\cdot)$, $\|T\|_F^2$ and s_1^2 is positive, it is possible to present the conditions for satisfying Equation (4.30) as follows:

$$s_1 \geq \max\{\lambda_{\max}(A), \lambda_{\max}(B)\} \|T\|_F, \qquad (4.31)$$

and the proof of Theorem 4.2.1 is complete.

4.4 Outside the concave region

The function $\theta(\cdot)$ cannot be concave indefinitely, since the secrecy capacity must be non negative by definition. A negative secrecy capacity would be a worse regime than sending nothing, and thus a rate of 0 would be preferable. We wish to show that the function does not have another maximum, and therefore the maximum found in Equation 4.7 is the true secrecy capacity. If we can show that there exists a cutoff, a, such that $\theta(\cdot)$ is concave on $[0, a)$, is convex for (a, ∞) and tends to 0 then this is sufficient.

In the proof of Theorem 4.2.1, Lemma 4.3.1 was used to find a lower bound for $\log\det(\cdot)$. In this section, we prove a converse of Lemma 4.3.1 and then apply this to $\theta(\cdot)$.

Firstly, we define an $M(x, y)$-strongly concave function, and then apply this definition to $\log\det(\cdot)$ to find an upper bound on the log determinant of a convex combination of arguments (analogous to the lower bound in [15, Lemma 15]). This is then applied in a similar manner to the proof of Theorem 4.2.1 to give a result about $\theta(\cdot)$ outside of the concave region.

Definition 4.4.1: A twice differentiable function $f : \operatorname{dom} f \to \mathbb{R}$ is $M(x, y)$-strongly concave between $x, y \in \operatorname{dom} f$ if $\nabla^2 f(tx + (1-t)y) \leq M(x, y)I$ for all $t \in [0, 1]$.

Lemma 4.4.2: For all $t \in [0, 1]$, an $M(x, y)$-strongly concave function f satisfies

$$tf(x) + (1-t)f(y) \leq f(tx + (1-t)y) + t(1-t)\frac{M(x, y)}{2}|x - y|^2. \quad (4.32)$$

The proof of this lemma is largely the same as the proof of [Lemma 30] [15] but tackles the problem from the other side (that is, to give an upper bound rather than their lower bound).

Proof. The Taylor series expansion of f for any two points $x, y \in \mathrm{dom} f$ yields

$$\begin{aligned} f(x) =& f(y) + \langle \nabla f(y), y - x \rangle \\ &+ \frac{1}{2}\langle y - x, \nabla^2 f(t_0 a + (1-t_0)b)(y-x)\rangle \qquad (4.33) \\ \leq& f(y) + \langle \nabla f(y), y - x \rangle + \frac{M(x,y)}{2}|y-x|^2, \qquad (4.34) \end{aligned}$$

where Equation (4.33) holds for some $t_0 \in [0,1]$ and Equation (4.34) follows from Definition 4.4.1. Let $w = tx + (1-t)y$, for $t \in [0,1]$. Then applying the above inequality to $f(x)$ and $f(y)$ gives

$$f(x) \leq f(w) + \langle \nabla f(w), w - x \rangle + \frac{M(w,x)}{2}|w-x|^2 \qquad (4.35)$$

$$f(y) \leq f(w) + \langle \nabla f(w), w - y \rangle + \frac{M(w,y)}{2}|w-y|^2. \qquad (4.36)$$

Summing t(4.35) + $(1-t)$(4.36) yields

$$tf(x) + (1-t)f(y) \leq f(w) + \frac{t(1-t)^2 M(x,w) + t^2(1-t)M(y,w)}{2}|y-x|^2. \qquad (4.37)$$

By definition of w, $M(x,w) \leq M(x,y)$ and $M(y,w) \leq M(x,y)$ and therefore Equation (4.37) may be bounded above by

$$f(tx + (1-t)y) + t(1-t)\frac{M(x,y)}{2}|y-x|^2 \qquad (4.38)$$

which proves the lemma. □

We now give an upper bound for $\log\det(\cdot)$ for convex combinations, this is analagous to the lower bound of Lemma 4.3.1.

Lemma 4.4.3: For positive definite matrices A, B and $t \in [0,1]$,

$$\begin{aligned} \log\det(tA + (1-t)B) \leq& t\log\det(A) + (1-t)\log\det(B) \\ &+ \frac{t(1-t)}{2\min\{\lambda^2_{\min}(A), \lambda^2_{\min}(B)\}}\|A-B\|^2_F. \qquad (4.39) \end{aligned}$$

For ease of notation, we denote

$$\mathcal{C}_{\min}(A,B) = \frac{1}{2\min\{\lambda^2_{\min}(A), \lambda^2_{\min}(B)\}}. \tag{4.40}$$

Again, the proof closely follows that given in [15] for their equivalent Lemma, but uses the concavity of $\log\det(\cdot)$ rather than the convexity of $-\log\det(\cdot)$.

Proof. Since $f(\cdot) = \log\det(\cdot)$ is strictly concave and twice differentiable for positive semidefinite matrices, we may apply Lemma 4.4.2 to f. Therefore

$$\begin{aligned}\log\det(tA+(1-t)B) \leq t\log\det(A) + (1-t)\log\det(B)\\ + t(1-t)\frac{M(A,B)}{2}\|A-B\|_F^2.\end{aligned} \tag{4.41}$$

Since $\nabla^2 f(C) = C^{-1}\otimes C^{-1}$, where $\otimes$ denotes the Kronecker product (Definition 1.3.14). The maximum eigenvalue of this product is given by $1/\lambda_{\min(C)}$ (since eigenvalues of $X\otimes Y$ are the products of eigenvalues of X and eigenvalues of Y.). By the definition of $M(A,B)$ we have the following upper bounds

$$M(A,B) \leq \max_{t\in[0,1]} \frac{1}{\lambda^2_{\min}(tA+(1-t)B)} \tag{4.42}$$

$$\leq \frac{1}{\min\{\lambda^2_{\min}(A), \lambda^2_{\min}(B)\}}, \tag{4.43}$$

where Equation (4.43) follows by the concavity of the minimum eigenvalue. Combining Equations (4.41) and (4.43) gives

$$\begin{aligned}\log\det(tA+(1-t)B) \leq & t\log\det(A) + (1-t)\log\det(B) \\ & + t(1-t)\frac{1}{\min\{\lambda^2_{\min}(A), \lambda^2_{\min}(B)\}}\|A-B\|_F^2\end{aligned} \tag{4.44}$$

as desired. □

Using definitions and properties, we give a result describing the behaviour of $\theta(\cdot)$ outside the concave region.

Theorem 4.4.4: Given s_t, let Q_t be the corresponding optimal covariance matrix. Then

$$\theta(s_t) \leq t\theta(s(Q_1)) + (1-t)\theta(s(Q_2)) \tag{4.45}$$

for any positive semidefinite matrices Q_1, Q_2 such that $Q_t = tQ_1 + (1-t)Q_2$ if the matrices K_B and K_E from Equations (4.2) and (4.3) satisfy

$$\frac{s(Q_1)}{\|K_B^{-1}K_E^2K_B^{-1}\|_F} - 1 \leq \min\{\lambda_{\min}(H_BQ_1H_B^*), \lambda_{\min}(H_BQ_2H_B^*)\}. \tag{4.46}$$

Simulation results showing the cutoff points for the convex and concave regions can be seen as red markers in Figure 4.4. It can be seen that there is a gap between these two, and this is expected since in the proofs some conservative bounds are applied however, this is only a small region which can easily be searched across.

Analogously to the proof of the concave region, the main steps of this proof will be roughly the same.

1. Finding an upper bound for $f(Q_t)$ using Theorem 4.4.3.
2. Minimising the difference between the bound found in the first step with the desired convexity bound.
3. Rewriting these bounds in terms of symmetric matrices and applying properties of the Frobenius norm.

4.4.1 Step 1

Let Q_i denote the matrix which achieves the maximum value of θ for s_i. Recall the definition of $s(Q_i) = I_{N_E} + H_EQ_iH_E^*$, and for optimal Q_i, we have that $s(Q_i) = s_i$. Choose s_t and the corresponding Q_t. For some $t \in [0,1]$, write

$$s_t = ts(\overline{Q}_1) + (1-t)s(\overline{Q}_2) \tag{4.47}$$

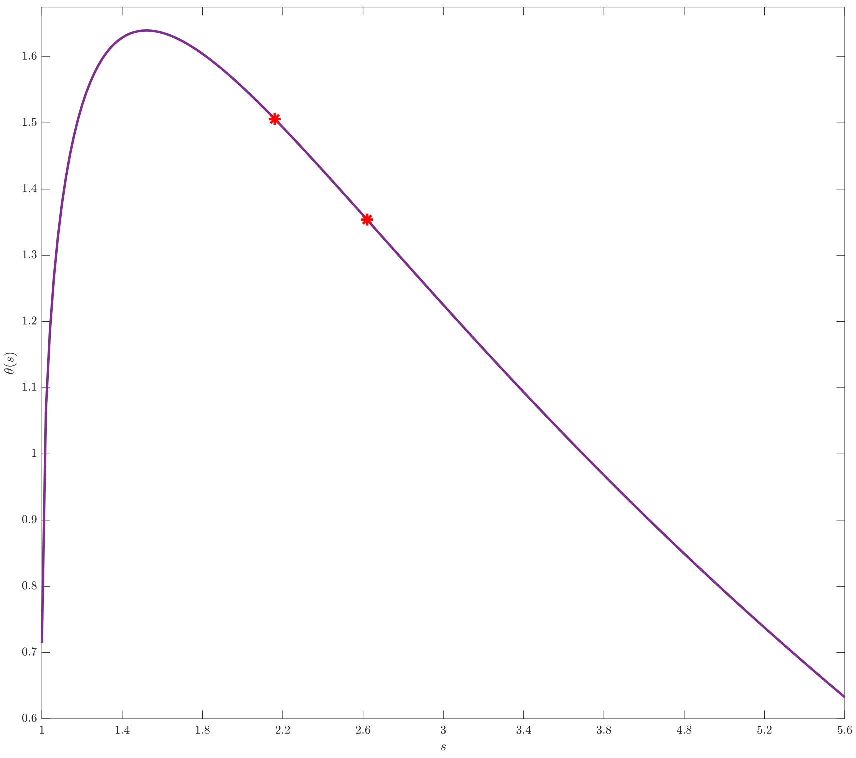


Figure 4.4: $\theta(s)$ vs s for a particular channel, showing the cut off points for the inequalities in Theorems 4.2.1 and 4.4.4.

$(s(\overline{Q}_1) \geq s(\overline{Q}_2))$ where $Q_t = t\overline{Q}_1 + (1-t)\overline{Q}_2$. Here we use the notation $\overline{Q}_i$ to distinguish this matrix from the optimal matrix Q_i.

By applying Lemma 4.4.3, and using the optimality of Q_t we may bound $\theta(s_t)$ as follows

$$\begin{aligned}\theta(s_t) =& f(Q_t) = \log\det(I + H_B Q_t H_B^*) - \log s_t \\ \leq& t \log\det(A) + (1-t)\log\det(B) + t(1-t)\mathcal{C}_{\min}(A,B) - \log s_t \quad (4.48)\end{aligned}$$

for $A = I + H_B\overline{Q}_1 H_B^*$ and $B = I + H_B\overline{Q}_2 H_B^*$. Equivalently, the upper bound of Equation (4.48) may be bounded by

$$\begin{aligned} tf(\overline{Q}_1) &+ (1-t)f(\overline{Q}_2) + t(1-t)\mathcal{C}_{\min}(A,B) \\ &+ t\log(s_1) + (1-t)\log(s_2) - \log(s_t) \qquad (4.49) \\ \leq t\theta(s_1) &+ (1-t)\theta(s_2) + t(1-t)\mathcal{C}_{\min}(A,B) \\ t\log(s_1) &+ (1-t)\log(s_2) - \log(s_t) \qquad (4.50) \end{aligned}$$

where Equation (4.50) follows by the definition of $Q_{1,2}$ and the fact that $\theta(\cdot)$ is a maximum.

4.4.2 Step 2

We require the following Lemma.

Lemma 4.4.5: For all $t \in [0,1]$,

$$\begin{aligned} t\log(s_1) + &(1-t)\log(s_2) - \log(ts_1 + (1-t)s_2) \\ &\leq -t(1-t)\kappa(s_1,s_2), \qquad (4.51) \end{aligned}$$

for

$$\kappa(s_1,s_2) = \frac{(s_1-s_2)^2}{2s_1^2}. \qquad (4.52)$$

Proof. Following the proof of Lemma 4.3.2, this is a matter of showing that the equivalent function g is convex in the interval for this value of κ. □

4.4.3 Step 3

The desired convexity constraints follows by combining Lemma 4.4.5 and Lemma 4.4.3. The desired bound is held if

$$\begin{aligned} \frac{t(1-t)}{2\min\{\lambda^2_{\min}(A), \lambda^2_{\min}(B)\}} &\leq \kappa(s_1,s_2) \qquad (4.53) \\ &\leq \frac{(s_1-s_2)^2}{2s_1^2}. \qquad (4.54) \end{aligned}$$

And so the result follows by the definitions of R and T given in the proof of Theorem 4.2.1.

4.5 Discussion

Although the expression for the secrecy capacity is known for the Gaussian wiretap channel, it is not generally known how to solve the optimisation problem for the covariance matrix, Q. The method presented in this chapter gives an efficient way to search for the secrecy capacity of a MIMO system and a corresponding covariance matrix for the transmission. The use of existing convex optimisation schemes makes the problem presented in Equation (4.1) manageable. We show that it is possible to efficiently search numerically for the maximum using linear combinations of variables.

For a fixed channel, the norm $\|T\|_F$ is simple to compute. To find the secrecy capacity, it is a case of picking a value of s_1 and s_2 and checking the constraint in Equation (4.31). If the criteria is satisfied, then these are in the concave region. It is therefore sufficient to use a standard concave optimisation technique, such as those outlined in Section 1.3.2. If Equation (4.31) is not satisfied, then an algorithm may be implemented to choose a different value until we are in the concave region.

The transmission scheme corresponding to this covariance matrix will be information theoretically secure since the user is guaranteed to be transmitting at or below the secrecy capacity.

This scheme is specific to the case with $N_E = 1$ and $N_B \geq N_A$. This is due to the requirements which arise in the derivation of the proof. Despite these restrictions, this work covers a family of MIMO systems which are not fully understood at the time of writing. For the situation with multiple antennas at the eavesdropper, the current state of the art is the algorithmic approach outlined by [51]. When the number of antennas at Eve is greater than 1, the

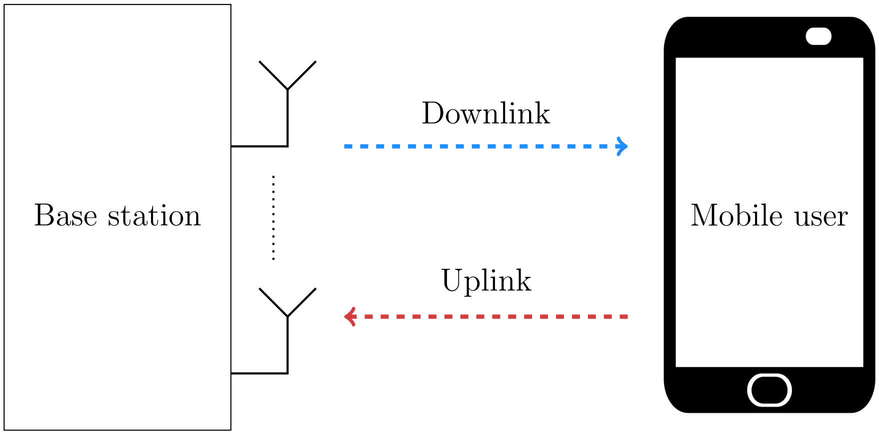


Figure 4.5: Massive MIMO basestations will have a far greater number of antennas than the mobile users

problem of the secrecy capacity cannot be written in the equivalent convex format as outlined in this chapter and the problem becomes far more difficult. In the Gaussian setup, multiple single antenna eavesdroppers behave in the same way as a multiple antenna eavesdropper. It is unclear whether this helps in this particular scenario, but is an avenue for future investigation.

In order to achieve the desired capacity gains for 5G, massive MIMO systems are a key technology [2]. This means that modern and future systems using massive MIMO will have a high number of antennas at the base station. Therefore the $N_B \geq N_A$ constraint in Theorem 4.2.1 would imply that these results are limited to the uplink for a massive MIMO system, as in Figure 4.5 since mobile users will have far fewer antennas. In future work, it would be interesting to generalise to the downlink of such channels.

It is important to note that this work assumes a static environment. Since the work considers the Gaussian wiretap channel with full channel state information (CSI), there is an inherent assumption that the channel statistics are fixed. Thus these results hold within the coherence time of the channel therefore the channel is fairly static, they are valid for a longer period of time. If we no longer assume a static channel, and instead suppose that the channel matrices are unknown, or fading, then the dimensions of the problem increase dramatically. For different types of fading, Equation (3.14) is no longer the agreed formula for the secrecy capacity, and there are far more degrees of freedom in the problem.

A practical limitation of any capacity result stemming from Shannon's work is the asymptotic nature of the results. While it is important to understand the fundamental measures of systems, there is evidence that the capacity of a system may be significantly lower for finite blocklength as shown in [60]. This means that the secrecy capacity could be an overestimate, particularly for low power devices with short blocklength such as internet of things devices.

Theorem 4.4.4 is a weaker statement than that in Theorem 4.2.1. This is because in finding the upper bound, and thus the concavity of $\theta(\cdot)$, firstly s_1 and s_2 are picked and then a convex combination $s_t = ts_1 + (1-t)s_2$ is taken. Since $\theta(\cdot)$ is a maximum of $f(Q)$ taken over all Q, we may upper bound our statement by $\theta(s_t)$. In the proof of Theorem 4.4.4, firstly s_t is picked. From here, it is not immediate that a value of s_1 and s_2 exist under the given constraints, and so the Theorem statement is looser.

Chapter 5

Eavesdropping a MIMO-NOMA Scheme

In this chapter, we consider a central base station communicating to many users by sharing their resource. As well as the typical security and low error constraints on the communication system, there is also a sense of 'fairness' which the base station must achieve. Our users may have multiple antennas, including the eavesdropper, and their Gaussian channels will vary in quality depending on their distance from the base station.

5.1 Introduction

Multiple access (MA) schemes enable many users to be served while sharing the same resource. Rather than the classic two user case considered in Chapter 4, a realistic scenario will consist of numerous users with a finite resource such as bandwidth or power. A base station in a city centre, for example, will be expected to serve all of the users on their network and must do so in a way which is fair, and provides a reasonable quality to all users. On the other hand, users aren't necessarily aware of the location of the base station, and will expect their phone signal to remain intact despite their physical location being at the cell edge.

The multiple access scheme in this chapter is Non-Orthogonal Multiple Access (NOMA), which is particularly useful for systems where users have highly different channel characteristics, for example near and far users.

NOMA is a multiplexing technique which is in the 5G specification, previously defined in Section 2.4.2. It has been discussed in Chapter 4 and Section 2.4.1 that Multiple-Input Multiple-Output (MIMO) systems unlock higher capacities and are being adopted in the communication systems of today. It is therefore natural to consider combining MIMO with NOMA. The NOMA system may utilise the channel differences to increase the throughput, while the MIMO exploits the additional degrees of freedom (DoF) to further enhance this.

Given the promise of both MIMO and NOMA, it is logical to ask about their security at physical layer. The work in this chapter demonstrates the robustness of a combined MIMO-NOMA scheme (outlined in Section 2.4.3) at physical layer, when in the presence of a passive eavesdropper. Bounds on the eavesdropper performance are presented and it is shown heuristically that, as the number of users and antennas increases, the eavesdropper's signal to interference and noise ratio (SINR) becomes small, regardless of how 'lucky' they may be with their channel.

We consider the scheme of [23] from the point of view of an eavesdropper. To implement NOMA, the base station applies a precoding scheme to the signal vector. This precoding gives a signal alignment to each user depending on their channel. Owing to the inherent randomness of the wireless medium we will assume that an eavesdropper has an independent randomly chosen channel and as a result, the eavesdropper is extremely unlikely to see the same signal alignment as the legitimate receivers. Hence, unlike the legitimate receivers, an eavesdropper cannot easily remove interfering messages meant for other receivers, and will see an inherently noisier signal. In other words, we argue that from the viewpoint of physical layer security, the MIMO-NOMA scheme [23] protects its messages from eavesdroppers by design. Further, from a Massive MIMO viewpoint, as the numbers of users and antennas grow, the eavesdropper's job becomes harder, and security is further enhanced.

5.2 Setup

We will look at a downlink (base station to user) NOMA setup, and use the same model and signal alignment scheme as [23, Section II.A]. Consider a base station equipped with M antennas and a collection of receivers each equipped with N antennas, where $N > M/2$ to allow for the use of signal alignment.

Definition	**Symbol**
Number of antennas each user	N
Number of basestation antennas	M
Number of user pairs	M
Distance of user m	d_m
Precoding matrix	P
Path loss funcion	$L(\cdot)$
Power allocation coefficient for user m	α_m

Table 5.1: Notation for Chapter 5.

Since we are considering near and far users, the channel gains are modelled to be worsened with distance from the base station. Assume the channel matrices from the base station to the particular users are of the form $G_m/\sqrt{L(d_m)}$ for a certain path loss function L which depends on the distance d_m. For brevity, we let L_m denote $L(d_m)$ for user m.

We select M 'near' users (within a radius r_1 of the base station) and M 'far' users (between r_1 and r_2 from the base station) and pair them up randomly. This setup can be seen in Figure 5.1. In particular, we consider pairing near users m and far users m' and creating an $M \times 1$ message vector

$\mathbf{s}$ of the form

$$\mathbf{s} = \begin{pmatrix} \alpha_1 s_1 + \alpha_{1'} s_{1'} \\ \vdots \\ \alpha_m s_m + \alpha_{m'} s_{m'} \\ \vdots \\ \alpha_M s_M + \alpha_{M'} s_{M'} \end{pmatrix}$$

where s_i is the signal intended for the ith user, and α_i are power allocation coefficients with $\alpha_m^2 + \alpha_{m'}^2 = 1$. Since user m' is further away, they require a greater power allocation and therefore $\alpha_{m'} > \alpha_m$.

The key to the scheme of [23] is the construction of an $M \times M$ precoding matrix P, which is designed to make it possible to remove interference at each pair of receivers, and to reduce the problem to standard 2-user NOMA by use of an appropriate detection vector $\mathbf{v}$. This is formed via constructing a matrix $G = [\mathbf{g}_1 \ \mathbf{g}_2 \ \ldots \mathbf{g}_M]^*$, with $\mathbf{g}_m$ being a particular vector in the intersection of the row spaces of G_m and $G_{m'}$ given by $\mathbf{g}_m^* = \mathbf{v}_m^* G_m$ for a certain $\mathbf{v}_m$. Then $P := G^{-1}F$, where F is a diagonal matrix chosen to ensure power control conditions are met at the base station[1].

The base station transmits the precoded signal, which is given by the product $P\mathbf{s}$ and user m receives (see [23, Eq. (2)]):

$$\mathbf{y}_m = \frac{G_m}{\sqrt{L_m}}(P\mathbf{s}) + \mathbf{n} \tag{5.1}$$

$$= \frac{G_m}{\sqrt{L_m}} \left(\sum_{i=1}^{M} (\alpha_m s_m + \alpha_{m'} s_{m'})\, \mathbf{p}_i \right) + \mathbf{n} \tag{5.2}$$

where $N \times 1$ vector $\mathbf{n}$ is circularly symmetric Gaussian noise with covariance proportional to $\sigma^2 \neq 0$. Note that the scheme in [23] has a factor ρ_I denoting

[1] Note this is different to [23, Eq. (10)] which defines $P = G^{-*}D$ for a different diagonal matrix. Since G has rows $\mathbf{g}_i^*$, and P has columns p_j, the necessary condition [23, Eq. (9)] that $\mathbf{g}_i^* p_j = 0$ for $i \neq j$ is achieved by taking GP diagonal. Here $F = \text{diag}(\mathbf{f})$ where $\mathbf{g}_i^* p_i = f_i$.

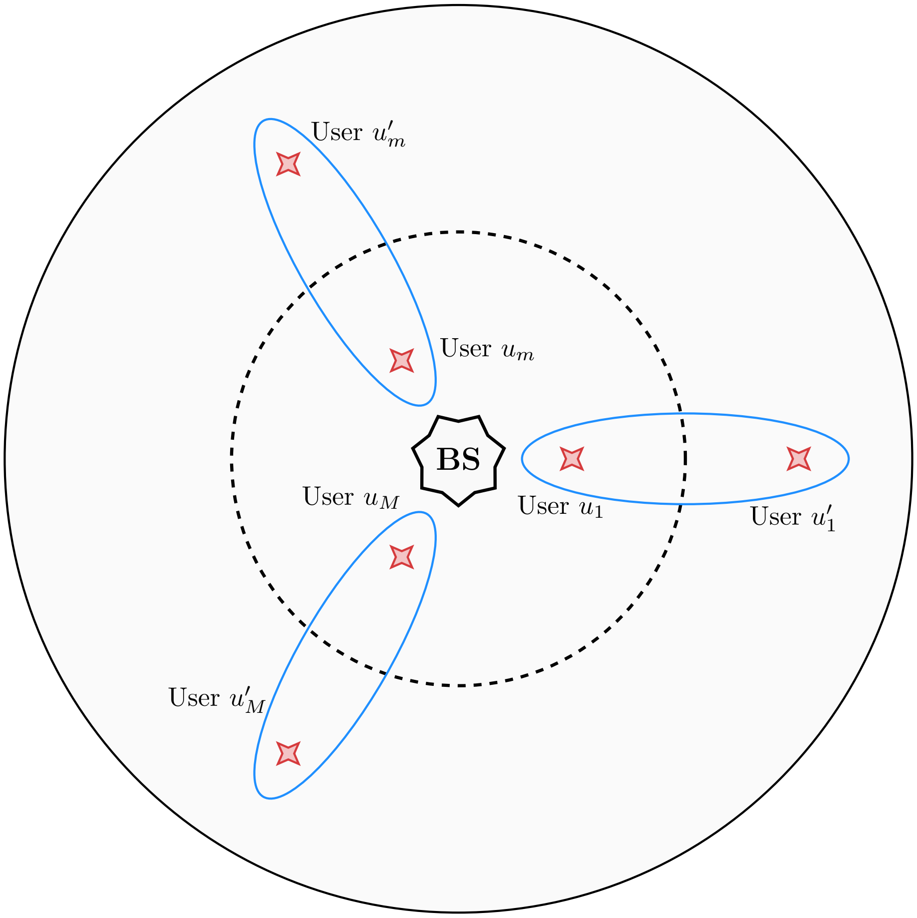


Figure 5.1: User pairings in the NOMA setup based on [23].

shot noise (noise from interference); for the purposes of this work we will assume there is no shot noise (ρ_I=0). If an eavesdropper cannot succeed without interference, then they cannot succeed with the additional noise.

An $N \times 1$ detection vector $\mathbf{u}$ is applied to $\mathbf{y}_m$. In [23], the choice $\mathbf{u} = \mathbf{v}_m$ is made, where the construction of the precoding matrix P ensures that $\mathbf{v}_m^* G_m \mathbf{p}_i = 0$ for $i \neq m$ and $\mathbf{v}_m^* G_m \mathbf{p}_m = \mathbf{g}_m^* \mathbf{p}_m = f_m$. This means that interference is removed and the problem is reduced to a one-dimensional

NOMA problem at each receiver, with

$$y_m := \mathbf{v}_m^* \mathbf{y}_m = \frac{f_m}{\sqrt{L_m}} (\alpha_m s_m + \alpha_{m'} s_{m'}) + n \tag{5.3}$$

where $n := \mathbf{v}_m^* \mathbf{n}$ is Gaussian noise.

5.3 Analysis of Eavesdropper Channel

Consider a passive eavesdropping receiver observing messages sent within the system. We will assume that the eavesdropper has the same number of antennas as the legitimate users. The eavesdropper has an $N \times M$ channel matrix of the form $K/\sqrt{L_E}$, where K has independent and identically distributed (IID) Rayleigh elements and $L_E = L(d_E)$ applies the same path loss function L to the eavesdropper distance from the base station. Without loss of generality, we will assume that the eavesdropper is listening into the message intended for Users 1 and $1'$. Since User $1'$ is further away, their signal receives a greater power allocation and thus will be easier to eavesdrop. If they are unsuccessful in obtaining the message for User $1'$, they will be unsuccessful in obtaining the message for User 1. We aim to show that, with high probability, the eavesdropper cannot gain useful information from the message for User $1'$.

The eavesdropper receives the $N \times 1$ vector

$$\begin{aligned} \mathbf{y}_e &= \frac{K}{\sqrt{L_E}} (P\mathbf{s}) + \mathbf{n} \\ &= \frac{1}{\sqrt{L_E}} \left(\sum_{i=1}^{M} (\alpha_m s_m + \alpha_{m'} s_{m'}) \mathbf{w}_i \right) + \mathbf{n} \end{aligned} \tag{5.4}$$

where $N \times 1$ vector $\mathbf{w}_i$ is the ith column of $W := KP$ and the other parameters and noise are as in Equation (5.1).

5.3.1 Optimal Detection Vector

We will consider the SINR for the eavesdropper, under the assumption that the signals s_i are independent with $\mathbb{E}|s_i|^2 = \rho\sigma^2$ for transmit SNR ρ. Consider trying to decode message $s_{1'}$ with detection vector $\mathbf{u}$. The eavesdropper will view all other signals as noise. The overall SINR for the communication in Equation (5.4) becomes

$$\mathrm{SINR}_E = \frac{\rho|\mathbf{u}^*\mathbf{w}_1|^2\alpha_{1'}^2}{\rho|\mathbf{u}^*\mathbf{w}_1|^2\alpha_1^2 + \rho\sum_{j=2}^{M}|\mathbf{u}^*\mathbf{w}_j|^2 + L_E\sum_{i=1}^{N}|u_i|^2}. \tag{5.5}$$

Given the assumption that the interference noise is 0, note that this is also the SNR.

Theorem 5.3.1: The optimal eavesdropper SINR is of the form

$$\mathrm{SINR}_E = \frac{\rho\alpha_{1'}^2}{\rho\alpha_1^2 + \left(\mathbf{w}_1^*\left(\rho(\overline{WW^*}) + L_E I_N\right)^{-1}\mathbf{w}_1\right)^{-1}}. \tag{5.6}$$

Proof. We can find the optimal detection vector by fixing

$$\mathbf{u}^*\mathbf{w}_1 = \mathbf{w}_1^*\mathbf{u} = |\mathbf{u}^*\mathbf{w}_1|^2 = 1 \tag{5.7}$$

and looking to minimise

$$\rho\sum_{j=2}^{M}|\mathbf{u}^*\mathbf{w}_j|^2 + L_E\sum_{i=1}^{N}|u_i|^2. \tag{5.8}$$

The first term may be rewritten as ρ multiplied by

$$\begin{aligned}
\sum_{j=2}^{M}\left(\sum_{r=1}^{M}u_r^*V_{rj}\right)\left(\sum_{s=1}^{M}u_sV_{sj}^*\right) &= \sum_{r,s=1}^{M}u_r^*u_s\sum_{j=2}^{M}V_{rj}V_{sj}^* &(5.9)\\
&= \sum_{r,s=1}^{M}u_r^*u_s(\overline{WW^*})_{rs} &(5.10)\\
&= \mathbf{u}^*(\overline{WW^*})\mathbf{u}, &(5.11)
\end{aligned}$$

where $\overline{W} = W - \mathbf{w}_1 \otimes (1, 0, \ldots, 0)$ is the matrix W with its first column set to zero.

Hence, a Lagrangian formulation gives

$$\mathcal{L}(\mathbf{u}, \lambda) = \mathbf{u}^* \left(\rho(\overline{W}\overline{W}^*) + L_E I_N\right) \mathbf{u} - \lambda \mathbf{u}^* \left(\mathbf{w}_1 \mathbf{w}_1^*\right) \mathbf{u}, \tag{5.12}$$

since this is a complex Hermitian quadratic form, we may apply Theorem 1.3.15 to give

$$\frac{\partial \mathcal{L}(\mathbf{u}, \lambda)}{\partial \mathbf{u}} = 2(\rho \overline{W}\overline{W}^* + L_E I_N)\mathbf{u} - 2\lambda \mathbf{w}_1 \mathbf{w}_1^* \mathbf{u} = 0. \tag{5.13}$$

And therefore the SINR (Equation (5.5)) may be written as

$$\begin{aligned} \text{SINR} =& \frac{\rho \alpha_{1'}^2}{\rho \alpha_1^2 + \mathbf{u}^* \left(\rho(\overline{W}\overline{W}^*) + L_E I_N\right) \mathbf{u}^*} && (5.14) \\ =& \frac{\rho \alpha_{1'}^2}{\rho \alpha_1^2 + \lambda \mathbf{u}^* \left(\mathbf{w}_1 \mathbf{w}_1^*\right) \mathbf{u}^*} = \frac{\rho \alpha_{1'}^2}{\rho \alpha_1^2 + \lambda}. && (5.15) \end{aligned}$$

Since $L_E \neq 0$, the matrix $\left(\rho(\overline{W}\overline{W}^*) + L_E I_N\right)$ is invertible. Hence, after some algebraic manipulation, the result follows. □

Remark 5.3.2: Note that the corresponding analysis will give the optimal detection vector and SINR for the legitimate user. In general this will not coincide with the choice $\mathbf{u} = \mathbf{v}_m$ made above in the analysis of Equation (5.1), since that choice removes interference potentially at the cost of increased noise, whereas our analysis considers interference and noise together.

From the point of view of Physical Layer Security, if the eavesdropper channel has smaller SINR than the legitimate channel, the true message can be protected by transmitting at the relevant rate. In order to compare the two channels, we will compare the optimal SINR in each case, though note that the expression [23, Eq. (15)] gives a tractable upper bound on the optimal legitimate SINR.

5.3.2 Bounding the Eavesdropper SINR

While Equation (5.6) gives a closed form expression for the optimal SINR, it is stated in terms of the random quantities $\mathbf{w}_1$ and $\overline{W}$. Hence, it is desirable to find an upper bound not dependant on these quantities.

Writing $Z = \rho(\overline{W}\overline{W}^*) + L_E I_N$, and $R(\cdot)$ for the Rayleigh quotient (Definition 1.3.11), note that

$$\begin{aligned}
\mathbf{w}_1^* \left(\rho(\overline{W}\overline{W}^*) + L_E I_N\right)^{-1} \mathbf{w}_1 &= \mathbf{w}_1^*\mathbf{w}_1 R(Z^{-1}; \mathbf{w}_1) \\
&\leq \frac{\mathbf{w}_1^*\mathbf{w}_1}{\lambda_{\min}(Z)} \\
&= \frac{\mathbf{w}_1^*\mathbf{w}_1}{\rho\lambda_{\min}(\overline{W}\overline{W}^*) + L_E} \qquad (5.16) \\
&\leq \frac{\mathbf{w}_1^*\mathbf{w}_1}{L_E}. \qquad (5.17)
\end{aligned}$$

This gives a conservative bound, since it considers the worst case and not the average case. However direct application of Equation (5.17) means that the SINR in Equation (5.6) is bounded above by

$$\text{SINR} \leq \frac{\rho\alpha_{1'}^2 \mathbf{w}_1^*\mathbf{w}_1}{\rho\alpha_1^2 \mathbf{w}_1^*\mathbf{w}_1 + L_E} \leq \frac{\rho\alpha_{1'}^2 EW}{\rho\alpha_1^2 EW + L_E}, \qquad (5.18)$$

where EW is the expectation of $\mathbf{w}_1^*\mathbf{w}_1$, and the second inequality follows by Jensen's inequality. We plot this result in Figure 5.2, which shows how eavesdropper SINR decays with distance as expected, and that (owing to lack of signal alignment) the eavesdropper performs worse than a legitimate receiver at the same distance.

5.4 Large Antenna Limits

To examine how the system fares in a massive MIMO setup, we consider the SINR of the eavesdropper as the number of antennas increases. We can also

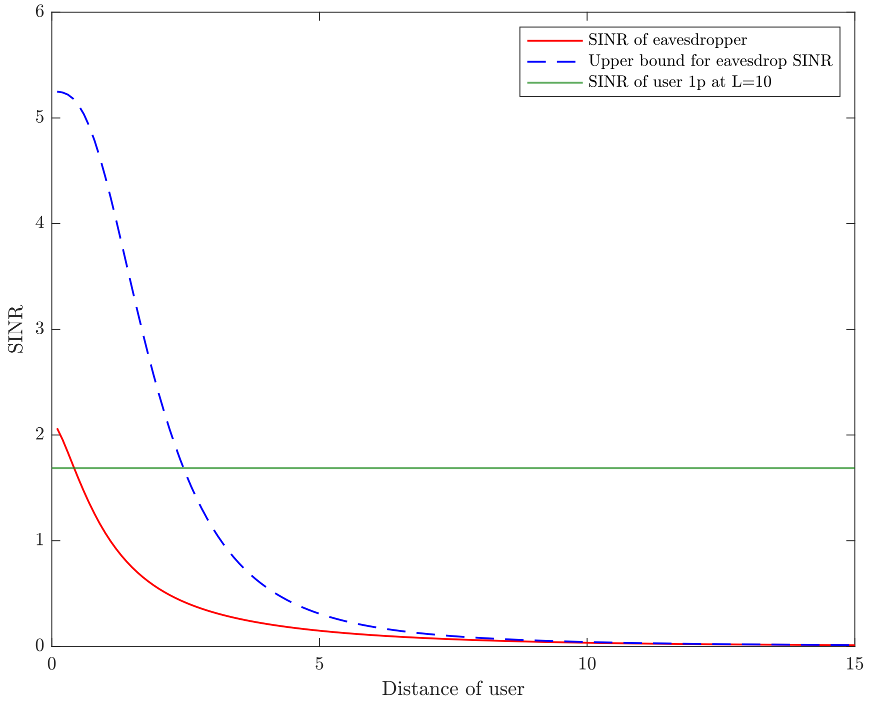


Figure 5.2: SINR vs User distance for $M = 7$, $N = 5$, $\rho = 5$ and legitimate users as in [23, Example 1]. We plot the upper bound on eavesdropper SINR from (5.18) in blue, the empirical eavesdropper SINR from simulation in red, and the legitimate SINR in green.

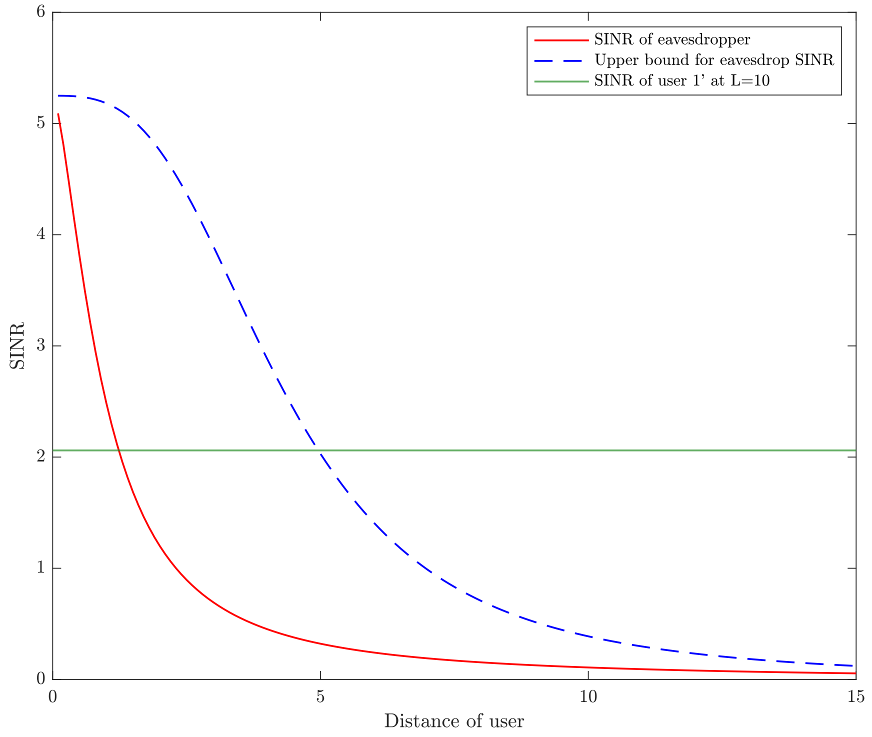


Figure 5.3: SINR vs User distance for $M = 50$ base station antennas.

argue heuristically in the large antenna limit, representing a Massive MIMO setup.

Proposition 5.4.1: In the limit of $N = \gamma M$ for $\frac{1}{2} < \gamma < 1$ then $\mathrm{SINR}_E \to 0$ at a rate of $\mathcal{O}\left(\frac{1}{M}\right)$.

Proof. Recall that $N > M/2$, so as M increases so does N. Thus we can apply the Marčenko–Pastur theory [53], in a regime where the number of antennas M is large and $N/M \to \gamma$ (for some $1/2 < \gamma < 1$), we have that

$$\lambda_{\min}(\overline{WW^*}) \simeq c(1 - \sqrt{\gamma})^2 M \tag{5.19}$$

for some positive constant c. Hence for any fixed distance L_E, for M sufficiently large the $\lambda_{\min}$ term will become the dominant one in Equation (5.16) which may be estimated as

$$\mathbf{w}_1^* \left(\rho(\overline{W}\overline{W}^*) + L_E I_N\right)^{-1} \mathbf{w}_1 \leq \frac{\mathbf{w}_1^*\mathbf{w}_1}{\rho\lambda_{\min}(\overline{W}\overline{W}^*) + L_E} \tag{5.20}$$

$$\simeq \frac{\mathbf{w}_1^*\mathbf{w}_1}{\rho c(1-\sqrt{\gamma})^2 M + L_E} \tag{5.21}$$

which is a scalar value. Consequently, the SINR of the eavesdropper will be bounded by

$$\mathrm{SINR}_E = \frac{\rho\alpha_{1'}^2}{\rho\alpha_1^2 + \left(\mathbf{w}_1^* \left(\rho(\overline{W}\overline{W}^*) + L_E I_N\right)^{-1} \mathbf{w}_1\right)^{-1}} \tag{5.22}$$

$$\leq \frac{\rho\alpha_{1'}^2}{\rho\alpha_1^2 + \frac{\rho c(1-\sqrt{\gamma})^2 M + L_E}{\mathbf{w}_1^*\mathbf{w}_1}}, \tag{5.23}$$

which becomes arbitrarily small for large M. That is, from any position, with enough antennas and user pairs, no eavesdropping is possible. □

5.5 Discussion

Schemes combining MIMO and NOMA provide great promise for the demands of 5G new radio and are likely to appear in real life systems imminently. Since security is a key factor in any communication system, it is vital to investigate their robustness to a passive eavesdropper. This work examined the combination of MIMO and NOMA in the system proposed by [23] where the message is precoded according to the legitimate user channels. This means that the message is easy to recover by a legitimate user, but difficult for users with a different channel.

It may seem that the eavesdropper could become lucky and, if well aligned with the legitimate user, they could obtain the message. Proposition 5.4.1

shows that as the number of user pairs increases, this is untrue and, regardless of position, the SINR of the eavesdropper tends to zero and thus they can obtain no useful information from their eavesdropping. The results in this chapter assume that an eavesdropper treats messages other than the specific one they are eavesdropping, as noise. This means the results in this chapter look at the SINR rather than the secrecy capacity. It remains to consider the problem of a more sophisticated eavsdropper who may employ successive decoding, for example, based on their strongest channel.

A significant drawback of a standard NOMA system is that each user must decode the messages intended for all other users [36]. The scheme presented in [23] overcomes this through their choice of precoding. However, this drawback remains for the eavesdropper, therefore the scheme exploits a weakness of NOMA to the advantage for secrecy.

These results are promising for the inherent security of MIMO NOMA systems. Since 5G networks are expecting to be dense [2] and the number of devices continuing to grow, it can be expected that base stations will be required to serve a high number of user pairs. The results in Section 5.3.2 are particularly relevant for such systems, and show promise for the inherent security of MIMO NOMA. In the downlink, as the number of user pairs increases, so does their number of antennas due to the setup of the system based on [23]. A mobile user is unlikely to be equipped with a large number of antennas so the heuristic bounds in Section 5.4 is not immediately applicable to dense systems, although hints at their robustness. These results are, however, applicable in the uplink which the scheme of [23] is applicable to.

Since it is assumed that the eavesdropper views all signals as noise except for the one they are trying to decode, the scope of this work is limited to studying the SINR. If a more sophisticated eavesdropper is considered, perhaps one using successive decoding based on their signal strengths, the overall system security could be reduced.

Chapter 6

Secure channel coding scheme

In this chapter, we introduce a novel coding scheme which is robust to time dispersion and eavesdropping. Multiple users share the channel and time resource giving an almost duplex scenario. So far in this thesis, we have discussed achievable secrecy but have not specified a coding scheme. Here, we present a coding scheme with secrecy at the forefront of the design.

Definition	**Symbol**
Number of legitimate users	K
Channel length	L
Size of codebook per user	N
Number of active codes per user	n
Transmission length	M
Weight of codes	m

Table 6.1: Notation for Chapter 6

6.1 Introduction

Time dispersive environments, as introduced in Section 2.3.1, are common in urban environments, in this chapter we propose a novel and secure channel coding scheme to thrive in such an environment. Combining elements from the papers of [59], [38] and [62] we present a scheme which exploits characteristics of time dispersive environments to efficiently encode and transmit information using elements of the legitimate channel, making it robust to eavesdropping.

The key papers are outlined below:

- The scheme of [59], which is called Combinatorial Channel Signature Modulation (**CCSM**), portrays the information content in the choice of the codeword combination, rather than the codewords themselves. The signal is constructed by selecting a subset of n codewords from a codebook of size N to be 'active' (where $n \ll N$). Constructing the signal in such a way, and using sparse codewords, gives a scheme which is robust to time dispersion, and in fact excels in dispersive environments. Their work builds on the similar work of [75] but reduces the complexity by their choice of signal structure.

- The secrecy scheme in [38] is an **antenna rotation scheme** where the Multiple-Input Multiple-Output (MIMO) users generate a pair of indices and secure their message using the indices. Specifically the legitimate channel calculates an antenna and constellation rotation value based on the legitimate channel characteristics. Since the eavesdroppers' channel is modelled as statistically independent from the legitimate channel, the eavesdropper does not have access to the rotation values. They therefore cannot undo the rotation and therefore may not recover the original message. They prove that the eavesdroppers best strategy is to guess, and thus perfect secrecy is achieved.

- The paper of [62] designs a codebook of **constant weight codewords**. Their scheme gives an encoding procedure for constant weight code constructions using arithmetic encoding techniques. The encoding and decoding procedures outlined use simple logical and arithmetic operations and thus they can construct codes with long codelength.

In our scheme, each user creates a codebook based on their channel, where each codeword is sparse - allowing users to transmit simultaneously and overcoming the need for complex timetabling protocols which are usually required for a full duplex system. It is a hybrid of the above outlined schemes, combining several of their advantages and thus our scheme is:

- Resilient to time dispersion,
- Does not need complex time scheduling,
- Efficient error detection due to constant weight coding and
- No requirement for a collision avoidance technique as users have access to a shared channel.

We maintain the combinatorial aspect of the CCSM signal structure from [59] in our scheme, as this was shown this to be effective in dispersive environments presented by scattering effects. We also adopt their notion of turning channel state information into codewords.

Each user generates a codebook where each codeword has a fixed weight m and length M where $m \ll M$. Here, the constant weight aspect is based on [62], but used in reverse meaning that a short channel realisation is encoded to a longer constant weight codeword. Typically, one would use arithmetic coding to compress data and not expand however we wish to generate long, sparse codewords and therefore employ this in reverse.

6.2 System model

Consider a system with K active, legitimate users, all transmitting broadcast messages in the presence of a passive eavesdropper, Eve, who does not transmit.

User i transmits a message $\mathbf{x}^{(i)}$ which is passed through the dispersive channel and convolved with the channel impulse response. The channel between user i and user j is denoted by a vector of complex entries

$$\mathbf{h}^{(i,j)} = \left(h_1^{(i,j)}, \ldots, h_L^{(i,j)}\right)$$

of length L where L is the channel length.

Each user has a codebook of size N, and their contents is referred to as their codeword span. Each codeword is constructed to be of a fixed weight, m, and length M. The codebook is designed based on the channel state information (CSI) between users. The sparsity provided by the requirement that the weights $m \ll M$ allows for an efficient decoding scheme and for the system to behave as though it is full duplex, despite users only being equipped with half duplex transmitters/receivers.

The combinatorial method of [59] is adopted, meaning that the messages are encoded in the choice of *combination* of n out of N of the codewords span rather than in the codewords themselves. Here $n \ll N$ and thus there are $\binom{N}{n}$ choices for the signal. This means that the information rate is agnostic to the type of modulation, as the useful information is in the choice of the codeword combination.

To visualise this, we give a toy example where $N = 6$ and $n = 2$, the signal construction may be seen in Figure 6.1. Here, the active codewords (indices 1 and 5) are highlighted and summed to give the signal which is transmitted. There are $\binom{6}{2} = 15$ choices for this signal.

The receiver then uses its knowledge of the channel and performs a maximum likelihood (ML) estimate to find the most likely combination of codewords that made up $\mathbf{x}^{(i)}$. This is discussed in more detail in Section 6.3.2.

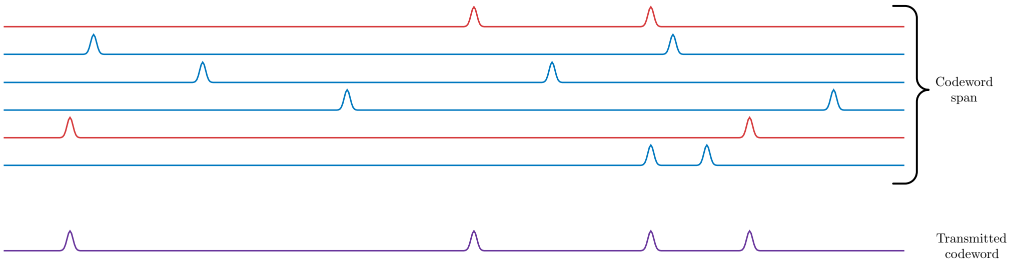


Figure 6.1: Example codebook and signal construction, the codewords in red are active. Here the codewords are of length $M = 100$ and weight $m = 2$. The codebook size, N, is 6 and the number of active codewords, n, is 2.

6.3 Constructing the codebook

We wish for the codebook to be some function of the legitimate user channel to reduce Eve's chance of success. Using channel quantisation, the legitimate users derive a set of indices from their channel. This set of indices is then used to permute the agreed codebook. This idea is related to, and generalises, the successful secrecy enhancing scheme of [38], where they rotate the antenna indices and prove that this obtains perfect secrecy. Eve has access to the agreed codebook and the scheme used to permute the codebook however has no knowledge of the channels used to perform the permutation and therefore cannot succeed at decoding.

6.3.1 Code construction

Arithmetic coding, introduced in Section 1.1.2, maps a string of symbols with an underlying probability distribution to a unique interval and corresponding binary codeword. This method of compression means that high probability variables are mapped to shorter codewords and low probabilities are assigned longer words.

We wish to create a codebook where each of the codewords is of a fixed constant weight, following the constructions of [62]. An advantage of constant weight coding is the simplicity of the error check. If the codeword is not of a specific weight, an error has certainly been made. This can be tested at virtually no cost to a decoder. Here, constant weight refers to the Hamming weight. That is, the number of symbols not zero is constant.

Definition 6.3.1: A codeword $c = c_1 \cdots c_M$ is a *constant weight codeword* of weight m if the Hamming weight of c is m we will denote this as $w(c) = m$.

In Example 1.1.12, we saw constant weight strings compressed into shorter binary strings, for our scheme we wish to implement this in reverse. That is, a binary string is elongated to a binary codeword of a fixed weight. In Example 1.1.12 the constant weight strings 10000, 01000 and 00100 are encoded to the shorter strings 111, 101 and 1000 respectively. In our scheme, we would take, for example, the string 111 and encode this to the constant weight codeword 10000 corresponding to the decoding the above scheme.

The codebook is entirely determined by the channel and the process for doing this is outlined in Algorithm 2. Algorithm 3 details the reverse arithmetic coding scheme to obtain a single codeword.

Example 6.3.2: Suppose we have a constant weight codeword $\mathbf{c}$ obtained from a channel vector $\mathbf{h}$ by Steps 1-3 above. Examples of how to carry out step 4 include:

- Let $r = \arg\max_i \|h_i\|$, and apply a cyclic shift of order r to $\mathbf{c}$.
- For $i = 1, \ldots, L$ let r_i be the order index of the entry h_i. The vector $(r_1, \ldots, r_L)$ defines a permutation, which may be applied to $\mathbf{c}$. This is the method used in the simulations for this chapter, unless explicitly stated otherwise.

Remark 6.3.3: The codebook generation outlined above is dependent on

Algorithm 2: Generating the codebook.

```
Input  : Channel realisation $\mathbf{h}$ of length $L$, desired codebook size $N$,
         codeword lengths $M$ and weights $m$.
Result: $\{\mathbf{c}_1, \dots, \mathbf{c}_N\}$ of length $M$ and weight $m$.
Initialise;
1. Quantise $\mathbf{h}$ to a binary string $\mathbf{v}$;
2. Apply Algorithm 3 to $\mathbf{v}$ to output constant weight codeword $\mathbf{c}$;
3. Modulate $\mathbf{c}$ according to the modulation scheme of choice to give $\mathbf{c}_1$;
4. Derive a permutation or rotation index from $\mathbf{h}$;
for $i = 2 : N$ do
    5. Apply the permutation derived in line 4 to $\mathbf{c}_{i-1}$ to give $\mathbf{c}_i$.
end
```

the legitimate channel state information (CSI). Without access to the correct CSI, the codebook is difficult to recover. This is exemplified in Figure 6.2 where the example codebook seen earlier (Figure 6.1) is generated once with correct CSI, and once with noisy CSI. Here, the correct channel is $\mathbf{h}$ and the noisy channel is $\mathbf{h} + \boldsymbol{\delta}$ where $\boldsymbol{\delta}$ is Gaussian noise with noise power of one tenth of the transmit power, showing that even small changes in the channel lead to an entirely different codebook. This is a property we will revisit when considering the security of this scheme.

6.3.2 Decoding process

Each legitimate user has access to all codebooks and therefore any receiver also has an effective codebook. That is, a codebook where the codewords have been convolved with the channel. Users are not fully duplex, therefore they cannot transmit and receive simultaneously.

Definition 6.3.4: To capture this, we define an *erasure pattern* for user j

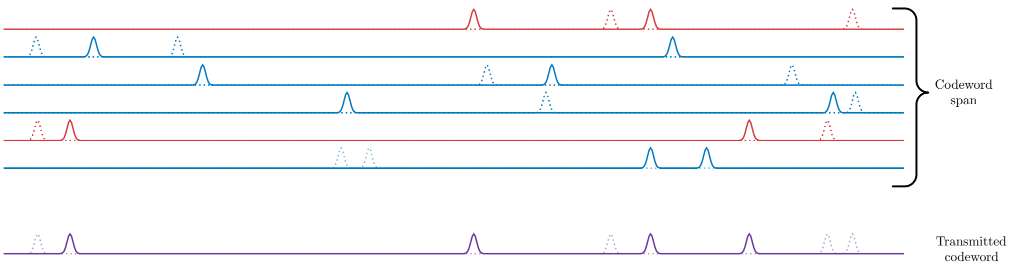


Figure 6.2: A correct codebook overlayed with a codebook generated from an incorrect channel (dotted lines).

to be the vector $\mathbf{e}^{(j)}$ with entries

$$e_i^{(j)} = \begin{cases} 0 & \text{if } x_i^{(j)} \neq 0 \\ 1 & \text{if } x_i^{(j)} = 0 \end{cases} \tag{6.1}$$

for $i = 1, \ldots, M$.

The effect of the erasure pattern is a 'puncturing' of the signal, where the received signal vector has zeros in the timeslots where they have transmitted. User j receives

$$\mathbf{y}^{(j)} = \mathbf{e}^{(j)} \left(\sum_{i=1}^{K} \mathbf{h}^{(i,j)} \star \mathbf{x}^{(i)} + \mathbf{z}^{(j)} \right), \tag{6.2}$$

where $\mathbf{z}^{(j)}$ is the additive Gaussian noise vector for their channel and the multiplication of $\mathbf{e}^{(j)}$ is elementwise (and thus $\mathbf{y}^{(j)}$ is a vector of length M). Due to the dispersive nature of the channel, there will be a self interference factor spread across multiple time slots, not just during their 'on' slots and thus the user may subtract any self interference (the $i = j$ term in the summation) to obtain

$$\tilde{\mathbf{y}}^{(j)} = \mathbf{e}^{(j)} \left(\sum_{i=1, i \neq j}^{K} \mathbf{h}^{(i,j)} \star \mathbf{x}^{(i)} + \mathbf{z}^{(j)} \right). \tag{6.3}$$

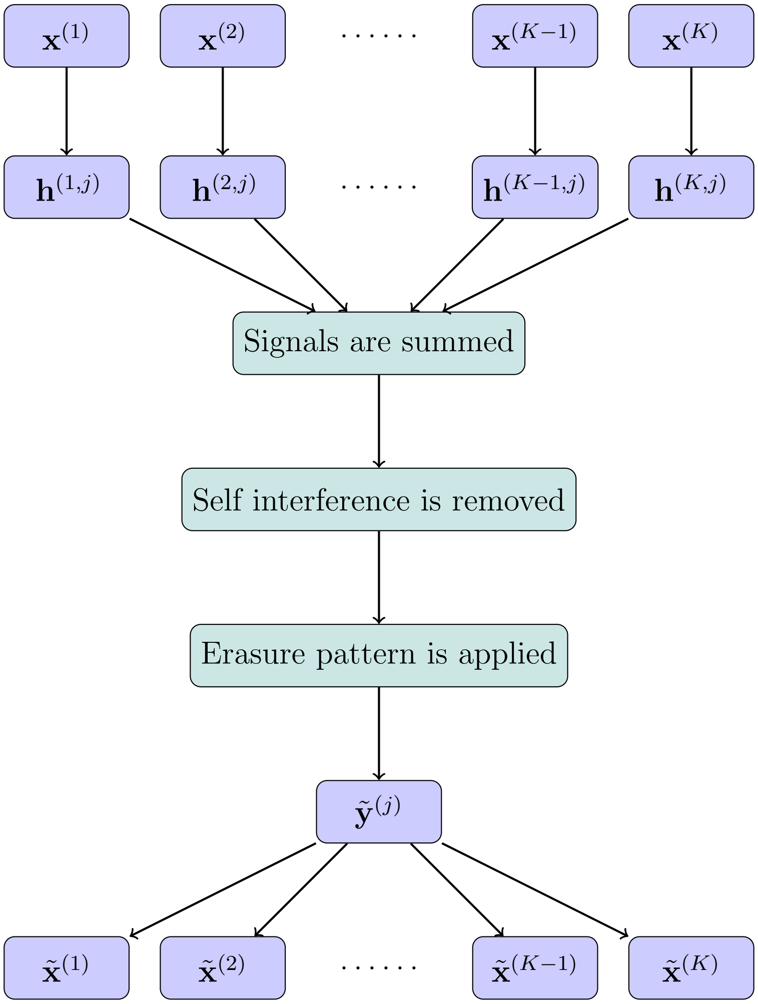


Figure 6.3: Decoding system for user j

In this setting, a maximum likelihood decoder would be too complicated, and given the sparsity of the messages we instead follow the CCSM paper [59] and implement a sparse recovery solver. User j has to solve the problem

$$\begin{aligned}\tilde{X} = \arg\min &\left\| \tilde{\mathbf{y}}^{(j)} - \sum_{i=1, i\neq j}^{K} \mathbf{h}^{(i,j)} \star \mathbf{x}^{(i)} \right\|^2, \\ &\text{Such that } w(\mathbf{x}^{(i)}) = mn \text{ for all } i\end{aligned} \tag{6.4}$$

where $\tilde{X}$ represents an $M \times K - 1$ matrix where column i is $\mathbf{x_i}$ and the minimum is taken over all possible $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(j-1)}, \mathbf{x}^{(j+1)}, \ldots, \mathbf{x}^{(K)}$.

Since the codewords are sparse and each signal $\mathbf{x}^{(i)}$ consists of exactly n codewords, the above decoding problem becomes a sparse recovery problem. In [59, Algorithm 1] they reduce the complexity of the problem by removing the requirement that *exactly* nm entries of each $\mathbf{x}^{(i)}$ are non-zero. Rather they solve using a Lasso algorithm and take the indices of the nm strongest values, setting the remaining entries to 0 afterwards. That is, we solve the problem

$$\begin{aligned}X = \arg\min &\left\| \tilde{\mathbf{y}}^{(j)} - \sum_{i=1, i\neq j}^{K} \mathbf{h}^{(i,j)} \star \mathbf{x}^{(i)} \right\|^2, \\ &\text{Such that } \left\|\mathbf{x}\right\|^2 \leq mn \text{ for all } i\end{aligned} \tag{6.5}$$

It is important to understand the effect of the codeword length, M, on the performance of the scheme. As the codeword length increases the effect of the signal puncturing occuring from the erasure pattern (Definition 6.3.4) lessens.

Figure 6.4 shows an increased performance for a longer transmission length, due to fewer clashes in transmitting and receiving.

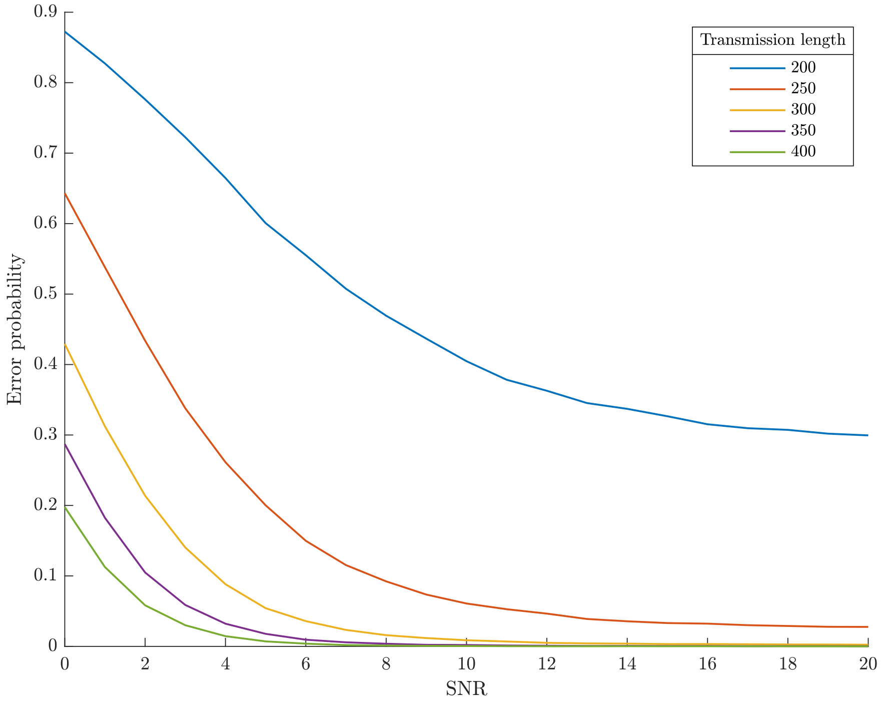


Figure 6.4: Error rate vs SNR for differing transmission lengths, where codewords have a fixed weight of 25 with 10 users. Here, $N = 32$ and $n = 4$.

6.4 Secrecy analysis

The design of the codebook in Section 6.3, is such that knowledge of the legitimate channels are required to find the codebook. As demonstrated in Figure 6.2, a small error in the channel can lead to a large error in the codebook. Since the performance of an eavesdropper is inherently linked to their knowledge of the codebook, this scheme promises security for the legitimate users. However, the eavesdropper is passive and does not have to content with self interference or the erasure patterns. In this section we show that the eavesdropper is unlikely to do well and through a series of simulations, that their error probability is close to 1 unless they are given a large (and unrealistic) advantage.

The received message at the eavesdropper is

$$\mathbf{y}^{(E)} = \sum_{i=1}^{K} \mathbf{h}^{(i,E)} \star \mathbf{x}^{(i)} + \mathbf{z}^{(E)}, \tag{6.6}$$

where $\star$ represents convolution. In other words, the signal received by the eavesdropper in time slot t is

$$y_t^{(E)} = \sum_{i=1}^{K} \sum_{j=0}^{L-1} h_j^{(i,E)} x_{t-j}^{(i)} + z_t, \tag{6.7}$$

where z_t are independent complex Gaussians with mean 0 and variance σ_E^2.

A maximum likelihood decoder looks to solve the problem

$$\min_{\mathbf{u}^{(1)},\cdots,\mathbf{u}^{(K)}} \left\| \mathbf{y}^{(E)} - \sum_{i=1}^{K} \mathbf{h}^{(i,E)} \star \mathbf{u}^{(i)} \right\|^2, \tag{6.8}$$

where the minimum is taken over sets of possible messages.

Following the paper of [48], we consider the case where a particular received message set is decoded to some other message. We suppose the true transmitted message set was $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}$ then the eavesdropper makes a mistake when the decoder incorrectly selects at least one of the $\mathbf{u}^{(j)}$. In general this is difficult to deal with, due to the size of the problem space so we instead consider the case where exactly one message is incorrectly decoded.

Proposition 6.4.1: Consider a system with K legitimate users, who have transmitted the message set $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}\}$ and suppose that an eavesdropper decodes these to $\{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(K)}\}$ where

$$\mathbf{u}^{(j)} = \mathbf{x}^{(j)} \text{if } j \neq l, \tag{6.9}$$

and $\mathbf{\Delta} = \mathbf{x}^{(j)} - \mathbf{u}^{(j)}$.

Then using an ML decoder, the eavesdropper will make a mistake when

$$\left\|\mathbf{h}^{(l,E)} \star \boldsymbol{\Delta}\right\|^2 \leq 2\left\|\mathbf{h}^{(l,E)} \star \boldsymbol{\Delta}\right\| \mathbf{z}^{(E)}, \tag{6.10}$$

where $\mathbf{h}^{(l,E)}$ denotes the channel between User l and the eavesdropper and $\mathbf{z}^{(E)}$ denotes the additive noise for the eavesdropper channel.

Proof. Equation (6.8) means that the ML decoder will make a mistake when the following inequality holds

$$\left\|\mathbf{y}^{(E)} - \sum_{i=1}^{K} \mathbf{h}^{(i,E)} \star \mathbf{u}^{(i)}\right\|^2 \leq \left\|\mathbf{y}^{(E)} - \sum_{i=1}^{K} \mathbf{h}^{(i,E)} \star \mathbf{x}^{(i)}\right\|^2. \tag{6.11}$$

The argument of the lower bound may be rewritten to give

$$\begin{aligned} \mathbf{y}^{(E)} - \sum_{i=1}^{K} \mathbf{h}^{(i,E)} \star \mathbf{u}^{(i)} &= \mathbf{y}^{(E)} - \sum_{i=1}^{K} \mathbf{h}^{(i,E)} \star \mathbf{x}^{(i)} + \mathbf{h}^{(l,E)} \star \boldsymbol{\Delta} && (6.12) \\ &= \mathbf{z}^{(E)} - \mathbf{h}^{(l,E)} \star \boldsymbol{\Delta} && (6.13) \end{aligned}$$

by Equation (6.6) and assumptions made above. And similarly, the argument of the upper bound of (6.11) is $\mathbf{z}^{(E)}$.

Inserting Equation (6.13) into Equation (6.11) and rearranging gives that the decoder will make a mistake when

$$\left\|\mathbf{z}^{(E)} - \mathbf{h}^{(l,E)} \star \boldsymbol{\Delta}\right\|^2 \leq \left\|\mathbf{z}^{(E)}\right\|^2. \tag{6.14}$$

By the Complex Polarisation Identity (Theorem 1.3.12), Equation (6.14) holds if and only if Equation (6.10) holds and so the proof is complete. □

Remark 6.4.2: If strict inequality holds in Equation (6.10), the decoder will definitely make a mistake, if equality holds, the decoder can't do better than guess so will make a mistake with probability at least 1/2.

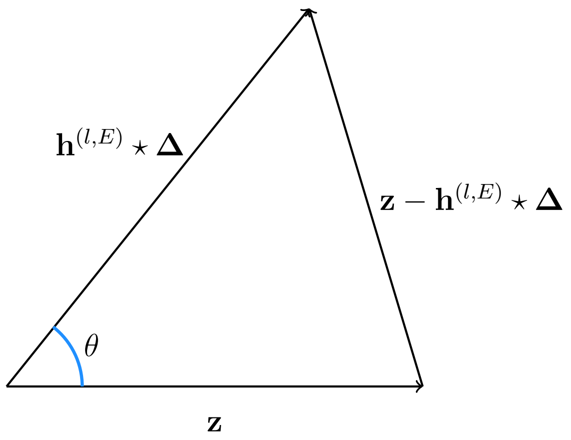

$\mathbf{h}^{(l,E)} \star \boldsymbol{\Delta}$
$\mathbf{z} - \mathbf{h}^{(l,E)} \star \boldsymbol{\Delta}$
θ
$\mathbf{z}$

Figure 6.5: Equation (6.10) holds when the angle, θ, is acute.

Proposition 6.4.1 can be interpreted as how well aligned the noise is with the mixture $\mathbf{h}^{(l,E)} \star \boldsymbol{\Delta}$, which is represented geometrically in Figure 6.5.

In reality, using a maximum likelihood decoder is not practical for this scheme due to the size of the problem space. However if an eavesdropper cannot succeed with a maximum likelihood decoder, they certainly cannot succeed with another decoder.

6.4.1 Eavesdropper channel model

When the eavesdropper is modelled independently to the legitimate users they do not have access to the channel in order to generate the correct codebook. This assumption relies on the fact that they are over one half wavelength away from the legitimate user, which is likely.

We use the Cholesky decomposition of a correlation matrix (recall Definition 1.3.13) to generate an eavesdropper channel which is correlated to the main channel. The legitimate channel is generated, as before, with IID Gaussian entries. The eavesdropper channel is then designed according to the correlation model proposed by [38] as follows:

$$\mathbf{h}^{(i,E)} = \rho\mathbf{h}^{(i,j)} + \sqrt{(1-\rho^2)}\mathbf{g}^{(i,E)} \tag{6.15}$$

where $\rho \in [0,1]$ and $\mathbf{g}^{(i)}$ denotes an independent channel vector. Note that this model is similar to a channel with an estimation error (as shown in [3]). Under this model, the error probability for differing values of ρ is shown in Figure 6.6. Here, it can be seen that the eavesdroppers probability of error is at, or very close to, 1 each time. We infer that under this channel model, without the legitimate system being compromised, the eavesdropper may not succeed.

Results for a system with 10 users are shown in Figure 6.7. Here, the eavesdropper has access to different numbers of the legitimate codebooks meaning that their decoding problem has varying levels of difficulty. The error probability is taken across all 10 users, so if they were correctly decoding 5 out of 10 users, we would expect ot see an error probability of 0.5. These simulation results show that even in the unlikely case where the eavesdropper has perfect access to 9 out of 10 of the users codebooks, they still make errors at a rate greater than 0.2. In other words, they are incorrectly decoding more than one user, despite only missing one users codebook. From a legitimate users viewpoint, this is a pessimistic scenario, which would require perfect CSI for each compromised user, and full knowledge of the coding scheme. This is compounded as user numbers increase, and the broadcast nature of the system enhances security in this sense.

6.5 Discussion

This chapter introduces a novel channel coding and multiplexing method for time dispersive channels. This is particularly applicable to wireless ad-hoc networks since there is no one user ‘in charge’ in such systems. The broadcast

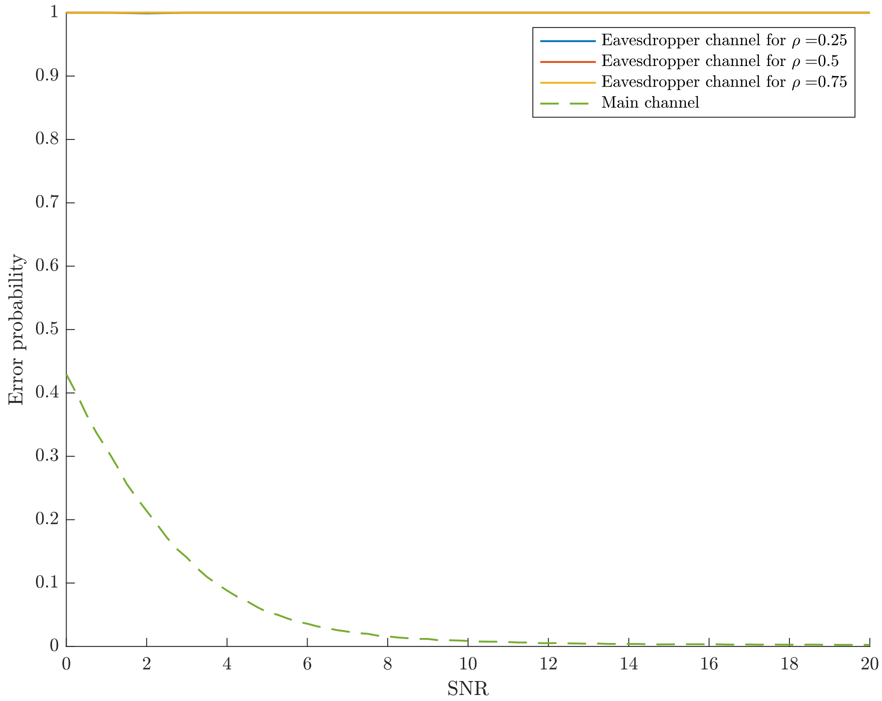


Figure 6.6: Error rates for a system with 10 legitimate users and one eavesdropper, with their channel generated as in Equation (6.15) for varied ρ. Here, $M = 300$, $m = 25$, $N = 32$ and $n = 4$.

nature of the system reduces the complexity requirements for access control, while the use of the channels in the encoding naturally enhances the security. As the length of the codewords used increases, so does the performance of the system since users are less likely to interrupt one another. This allows the non duplex transmitters and receivers to behave in a duplex manner, which increases the efficiency.

In order to compromise the system, Eve would need access to all users codebooks, and not just one. In this way, the broadcast nature of this scheme not only increases efficiency but also provides an additional layer of resilience

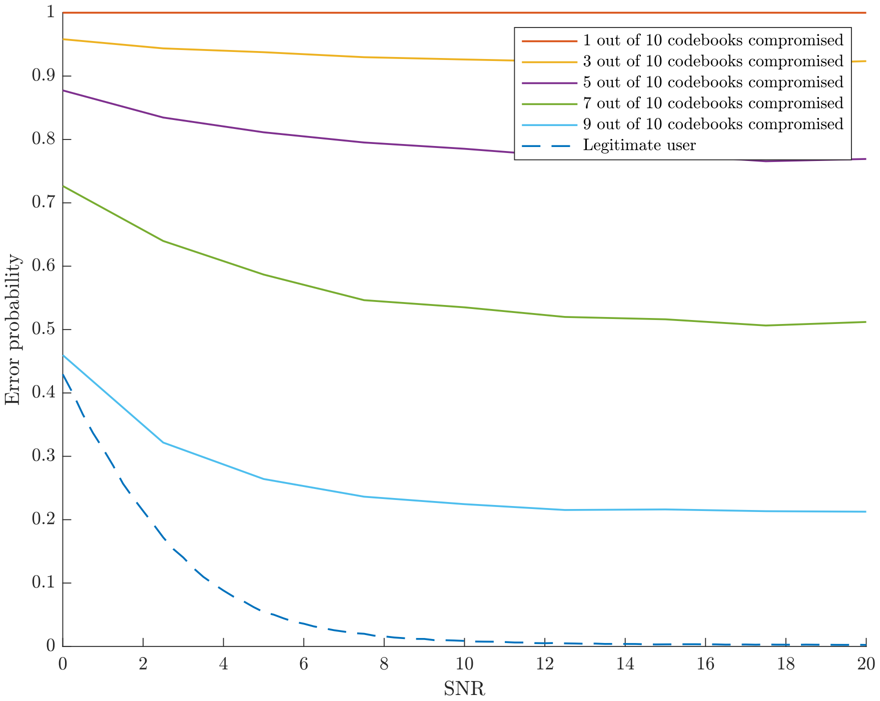


Figure 6.7: Error rates for a system with 10 legitimate users with different numbers of codebooks compromised to an eavesdropper. Here, $M = 300$, $m = 25$, $N = 32$, $n = 4$ and $\rho = 0.5$.

against eavesdropping. If just one codebook is compromised, the eavesdropper still has to contend with a large amount of uncertainty and fails with high probability, as is exemplified in Figure 6.7.

It remains to fully quantify the secrecy of the scheme. In Proposition 6.4.1 we presented results for the case where the eavsedropper correctly decodes all but one signal. To further generalise this is a non trivial problem and this is discussed in Chapter 7.

Algorithm 3: Generating a constant weight codeword using a reverse arithmetic coding scheme.

```
Input  : $\mathbf{v}$ of length $L$, length of desired codeword $M$ and weight $m$.
Result: $\mathbf{c}$ of length $M$ and weight $m$.
Initialise;
$a = 0$;
$b = 1$;
$p_1 = (M - m)/M$;
$\mathrm{cdf} = [0, \mathrm{p}_1, 1]$;
for $i = 1 : M$ do
    if $v_i = 0$ then
        Let $r = 0$;
    end
    else
        $r = 1$.
    end
    Update $a = a + r2^{-(i+1)}$.
end
Update $b = a + 2^{-L}$;
if $a \geq p_1$ then
    $c_1 = 1$;
    Update $p_1 = p_1 + (M - m + 1)/(M - 1)$.
end
else
    $c_1 = 0$;
end
for $j = 2 : M$ do
    if $a \geq p_1$ and $\sum_{i=1}^{j-1} \leq m$ then
        $c_j = 1$;
        Update $p_1 = p_1 + (M - m + \sum_{i=1}^{j} c_j)/(M - j)$.
    end
    else
        $c_j = 0$.
    end
end
```

Chapter 7

Conclusions and open problems

The work in this thesis considers topics in physical layer security for future telecommunications technologies. The closed form of the secrecy capacity of the Gaussian Multiple-Input Multiple-Output (MIMO) wiretap channel has remained an open problem in the general case. This thesis contributes to the $(N_A, N_B, 1)$ configuration, showing that the secrecy capacity is equivalent to the maximum of a provably concave region of a function. We examined the robustness of an innovative MIMO-NOMA system, showing that the eavesdropper SINR diminishes with a large number of users and with eavesdropper distance. Finally, a channel coding scheme which performs well in a time dispersive regime. It is shown that this scheme is robust to eavesdropping and interference. With the increasing use of multiple antenna systems, and considering power limitations, physical layer security provides an important way to improve and compound network security, and this thesis has presented novel ways to analyse the performance of such methods. A detailed conclusion and outline of the open problems are given for each topic below.

In Chapter 4 we presented results on the concavity for the $(N_A, N_B, 1)$ MIMO wiretap channel, giving a provably concave and convex region (Theorems 4.2.1 and 4.4.4 respectively). These results are notable since the current literature does not address a general case (that is, a case where the transmit regime is not constrained) other than with algorithmic and computational results. Our results are validated in Section 4.2.2 by considering the $(2, 2, 1)$ configuration. Simulation results agree with the theoretical secrecy capacity found by Shafiee et al. in [66].

Our results hold for the case where $N_A \leq N_B$, providing a baseline for future work without this restriction. In the process of calculating the cutoff points, there is a matrix inversion which introduces this requirement. To find an analogous result for the general antenna configuration, results on the channel matrices may be imposed.

Open problems: For a multiple antenna eavesdropper, the derivations in our work are no longer concave and the proof does not follow. In the Gaussian case, the multiple antenna eavesdropper is equivalent to multiple single antenna eavesdroppers colluding. Hence the MIMOME wiretap channel could be seen as a compound MIMOSE wiretap channel (see [44]).

The secrecy capacity, as with any capacity result, is an asymptotic result. These results are derived from blocklengths tending to infinity. Results by [60] show that for a finite blocklength, the realistic system capacity can be far lower than the asymptotic result.

In Chapter 5 a scheme combining NOMA, a multiple access scheme, with MIMO is presented. Both are enabling technologies for 5G and future wireless, and thus these results are particularly relevant to current and upcoming architectures. We showed that the system is inherently secure in the sense that the eavesdropper has a low probability of obtaining the message sent by the legitimate user. Further, the eavesdropper SINR diminishes with an increase in the number of users - representing a dense network, shown in Proposition 5.4.1.

As the number of user antennas increases, representative of a massive MIMO system, the eavesdropper SINR tends to zero. This shows that the shift towards massive MIMO for future wireless adds further levels of security to systems.

Open problems: We presented results for one eavesdropper, where multiple colluding eavesdroppers would be a valuable extension to the work. As discussed previously, one multiple antenna eavesdropper is equivalent to multiple single antenna eavsdroppers from a theoretical viewpoint. However, the system in Chapter 5 is affected by the positioning of the users. Therefore, multiple single antenna eavesdroppers may be able to optimise their location to detect the strongest signals.

In Chapter 6 we presented a novel and secure channel coding scheme. The scheme works particularly well in dispersive environments, where other schemes may fail. The information is transmitted in sparse codewords of constant weight where the information is encapsulated in the choice of the codewords rather than the codewords themselves. This allows for simple error detection at the receivers and means the scheme is independent of the modulation scheme used, making it applicable to a multitude of scenarios. Simulation showed that the systems performance is enhanced as the codewords become increasingly sparse (see Figure 6.4), due to the effect of erasures caused by a user transmitting (and therefore being unable to receive) being largely mitigated.

Open problems: It remains to quantify the secrecy in terms of the entropy. Proposition 6.4.1 considered the case where the eavesdropper only has to decode one message, and gave criteria for the success of the eavesdropper. That is, we considered the event where any codeword $\mathbf{u}_i$ has a higher probability than the real codeword $\mathbf{x}$ given the received signal, $\mathbf{y}$. For a specific codeword $\mathbf{u}_i$, the probability of this type of error is bounded by

$$p(\mathbf{u}_i \mid \mathbf{x}, \mathbf{y}) \leq \frac{p(\mathbf{y})}{p(\mathbf{y} \mid \mathbf{x})}, \tag{7.1}$$

where this corresponds to [48, Equation (2)] and is found by applying the Markov inequality. To bound the general case, we consider Equation (7.1) for all pairwise errors. By applying the union bound, Lomnitz and Feder bound this as follows

$$p\left(\bigcup \mathbf{u}_i \mid \mathbf{x}, \mathbf{y}\right) \leq 2^{MN} \frac{p(\mathbf{y})}{p(\mathbf{y} \mid \mathbf{x})} \tag{7.2}$$

where N is the size of the codebooks and M is the length of the codewords.

Given that the received codewords are not IID, and have a block dependency by their construction, we can't apply the law of large numbers to Equation (7.2) as is the next step in [48]. The quantity which they obtain ([48, Equation 4]) is the same as the metric of interest in [60], the information density. This work was previously outlined as a future research avenue for Chapter 4. The information density is

$$-\log \frac{p(\mathbf{y})}{p(\mathbf{y} \mid \mathbf{x})}, \tag{7.3}$$

and measures the amount of independence between the variables. The expectation of Equation (7.3) gives the mutual information. The problem of bounding the eavesdroppers failure rate is now the case of finding the information density. If the requirements for a secrecy metric are relaxed, to consider IID codewords, the problem of finding a bound is more tractable. In the interim, results for non-optimal decoders could be compared to provide insight to the general case.

Bibliography

[1] M. Agiwal, A. Roy, and N. Saxena, *Next generation 5G wireless networks: A comprehensive survey*, IEEE Communications Surveys Tutorials, 18 (2016), pp. 1617–1655.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, *What will 5G be?*, IEEE Journal on Selected Areas in Communications, 32 (2014), pp. 1065–1082.

[3] R. Annavajjala, P. C. Cosman, and L. B. Milstein, *Performance analysis of linear modulation schemes with generalized diversity combining on Rayleigh fading channels with noisy channel estimates*, IEEE Transactions on Information Theory, 53 (2007), pp. 4701–4727.

[4] M. Bellare, S. Tessaro, and A. Vardy, *A cryptographic treatment of the wiretap channel*, arXiv preprint arXiv:1201.2205, (2012).

[5] A. Benjebbour, A. Li, K. Saito, Y. Saito, Y. Kishiyama, and T. Nakamura, *NOMA: From concept to standardization*, in 2015 IEEE Conference on Standards for Communications and Networking (CSCN), Oct 2015, pp. 18–23.

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

[7] M. Bloch, J. Barros, M. R. D. Rodrigues, and S. W. McLaughlin, *Wireless information-theoretic security*, IEEE Transactions on Information Theory, 54 (2008), pp. 2515–2534.

[8] M. Bloch, M. Hayashi, and A. Thangaraj, *Error-control coding for physical-layer secrecy*, Proceedings of the IEEE, 103 (2015), pp. 1725–1746.

[9] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, *Five disruptive technology directions for 5G*, IEEE Communications Magazine, 52 (2014), pp. 74–80.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.

[11] T. Brown, P. Kyritsi, and E. De Carvalho, *Practical guide to MIMO radio channel: With MATLAB examples*, John Wiley & Sons, 2012.

[12] R. Bustin, R. Liu, H. V. Poor, and S. Shamai, *An MMSE approach to the secrecy capacity of the MIMO Gaussian wiretap channel*, in 2009 IEEE International Symposium on Information Theory, June 2009, pp. 2602–2606.

[13] J. Chakravarty, O. Johnson, and R. Piechocki, *A convex scheme for the secrecy capacity of a MIMO wiretap channel with a single antenna eavesdropper*, in ICC IEEE International Conference on Communications (ICC), May 2019, pp. 1–5.

[14] J. H. Conway and N. J. A. Sloane, *Sphere packings, lattices and groups*, vol. 290, Springer Science & Business Media, 2013.

[15] T. A. Courtade, M. Fathi, and A. Pananjady, *Quantitative stability of the entropy power inequality*, IEEE Transactions on Information Theory, 64 (2018), pp. 5691–5703.

[16] T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley & Sons, 2012.

[17] I. Csiszár and J. Körner, *Broadcast channels with confidential messages*, IEEE Transactions on Information Theory, 24 (1978), pp. 339–348.

[18] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, *A survey of non-orthogonal multiple access for 5G*, IEEE Communications Surveys Tutorials, 20 (2018), pp. 2294–2323.

[19] L. Dai, B. Wang, Y. Yuan, S. Han, C. I, and Z. Wang, *Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends*, IEEE Communications Magazine, 53 (2015), pp. 74–81.

[20] K. David and H. Berndt, *6G vision and requirements: Is there any need for beyond 5G?*, IEEE Vehicular Technology Magazine, 13 (2018), pp. 72–80.

[21] Z. Ding, L. Dai, and H. V. Poor, *MIMO-NOMA Design for Small Packet Transmission in the Internet of Things*, IEEE Access, 4 (2016), pp. 1393–1405.

[22] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, *A survey on Non-Orthogonal Multiple Access for 5G networks: Research challenges and future trends*, IEEE Journal on Selected Areas in Communications, 35 (2017), pp. 2181–2195.

[23] Z. Ding, R. Schober, and H. V. Poor, *A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment*, IEEE Transactions on Wireless Communications, 15 (2016), pp. 4438–4454.

[24] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, *On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users*, IEEE Signal Processing Letters, 21 (2014), pp. 1501–1505.

[25] P. Elias, *Universal codeword sets and representations of the integers*, IEEE transactions on information theory, 21 (1975), pp. 194–203.

[26] A. Ernvall-Hytönen and C. Hollanti, *On the eavesdropper's correct decision in Gaussian and fading wiretap channels using lattice codes*, in 2011 IEEE Information Theory Workshop, Oct 2011, pp. 210–214.

[27] G. J. Foschini and M. J. Gans, *On limits of wireless communications in a fading environment when using multiple antennas*, Wireless personal communications, 6 (1998), pp. 311–335.

[28] G. H. Golub and C. F. Van Loan, *Matrix computations*, vol. 3, JHU press, 2012.

[29] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming, version 2.1*, Mar. 2014.

[30] A. Gupta and R. K. Jha, *A survey of 5G network: Architecture and emerging technologies*, IEEE Access, 3 (2015), pp. 1206–1232.

[31] P. Harris, W. B. Hasan, S. Malkowsky, J. Vieira, S. Zhang, M. Beach, L. Liu, E. Mellios, A. Nix, S. Armour, A. Doufexi, K. Nieman, and N. Kundargi, *Serving 22 users in real-time with a 128-antenna massive MIMO testbed*, in 2016 IEEE International Workshop on Signal Processing Systems (SiPS), Oct 2016, pp. 266–272.

[32] K. Higuchi and Y. Kishiyama, *Non-orthogonal access with successive interference cancellation for future radio access*, APWCS2012, 8 (2012).

[33] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge university press, 1990.

[34] J. Hoydis, S. Ten Brink, and M. Debbah, *Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?*, IEEE Journal on selected Areas in Communications, 31 (2013), pp. 160–171.

[35] Huawei, *5G: A technology vision, White Paper*, 2013.

[36] S. R. Islam, N. Avazov, O. A. Dobre, and K. Kwak, *Power-domain Non-Orthogonal Multiple Access (NOMA) in 5G systems: Potentials and challenges*, IEEE Communications Surveys Tutorials, 19 (2017), pp. 721–742.

[37] S. R. Islam, M. Zeng, and O. A. Dobre, *NOMA in 5G systems: Exciting possibilities for enhancing spectral efficiency*, IEEE 5G Tech Focus, (2017).

[38] X. Jiang, M. Wen, H. Hai, J. Li, and S. Kim, *Secrecy-enhancing scheme for spatial modulation*, IEEE Communications Letters, 22 (2018), pp. 550–553.

[39] A. Khisti, A. Tchamkerten, and G. W. Wornell, *Secure broadcasting over fading channels*, IEEE Transactions on Information Theory, 54 (2008), pp. 2453–2469.

[40] A. Khisti and G. W. Wornell, *Secure transmission with multiple antennas – Part I: The MISOME wiretap channel*, IEEE Transactions on Information Theory, 56 (2010), pp. 3088–3104.

[41] ——, *Secure transmission with multiple antennas – Part II: The MIMOME wiretap channel*, IEEE Transactions on Information Theory, 56 (2010), pp. 5515–5532.

[42] G. G. Langdon, *An Introduction to Arithmetic Coding*, IBM Journal of Research and Development, 28 (1984), pp. 135–149.

[43] S. Leung-Yan-Cheong and M. Hellman, *The Gaussian wire-tap channel*, IEEE Transactions on Information Theory, 24 (1978), pp. 451–456.

[44] Y. Liang, G. Kramer, H. V. Poor, and S. Shamai, *The compound wire-tap channels,*, in Proc. of the 45th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, September 2007.

[45] Y. Liang, H. V. Poor, and S. Shamai, *Secrecy capacity region of parallel broadcast channels*, in 2007 Information Theory and Applications Workshop, Jan 2007, pp. 245–250.

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

[47] L. Liu, Y. Yan, and C. Ling, *Achieving secrecy capacity of the Gaussian wiretap channel with polar lattices*, IEEE Transactions on Information Theory, 64 (2018), pp. 1647–1665.

[48] Y. Lomnitz and M. Feder, *A simpler derivation of the coding theorem*, CoRR, abs/1205.1389 (2012).

[49] S. Loyka and C. D. Charalambous, *On optimal signaling over secure MIMO channels*, in 2012 IEEE International Symposium on Information Theory Proceedings, July 2012, pp. 443–447.

[50] ——, *Further results on optimal signaling over secure MIMO channels*, in 2013 IEEE International Symposium on Information Theory, July 2013, pp. 2019–2023.

[51] ——, *An algorithm for global maximization of secrecy rates in Gaussian MIMO wiretap channels.*, IEEE Trans. Communications, 63 (2015), pp. 2288–2299.

[52] D. J. MacKay, *Information theory, inference and learning algorithms*, Cambridge university press, 2003.

[53] V. A. Marčenko and L. A. Pastur, *Distribution of eigenvalues for some sets of random matrices*, Mathematics of the USSR-Sbornik, 1 (1967), pp. 507–536.

[54] U. M. Maurer, *The Strong Secret Key Rate of Discrete Random Triples*, Springer US, Boston, MA, 1994, pp. 271–285.

[55] i. NTT DOCOMO, *5G evolution and 6G, White Paper*, January 2020.

[56] F. Oggier and B. Hassibi, *The secrecy capacity of the MIMO wiretap channel*, IEEE Transactions on Information Theory, 57 (2011), pp. 4961–4972.

[57] O. T. O'Meara, *Introduction to quadratic forms*, vol. 117, Springer, 2013.

[58] K. Petersen, M. Pedersen, et al., *The matrix cookbook, vol. 7*, Technical University of Denmark, 15 (2008).

[59] R. J. Piechocki and D. Sejdinovic, *Combinatorial channel signature modulation for wireless ad-hoc networks*, in 2012 IEEE International Conference on Communications (ICC), June 2012, pp. 4684–4689.

[60] Y. Polyanskiy, H. V. Poor, and S. Verdu, *Channel coding rate in the finite blocklength regime*, IEEE Trans. Inform. Theory, 56 (2010), pp. 2307–2359.

[61] H. V. Poor and R. F. Schaefer, *Wireless physical layer security*, Proceedings of the National Academy of Sciences, 114 (2017), pp. 19–26.

[62] T. V. Ramabadran, *A coding scheme for m-out-of-n codes*, IEEE Transactions on Communications, 38 (1990), pp. 1156–1163.

[63] B. Rimoldi, *Passband communication via up/down conversion: Third layer*, Cambridge University Press, 2016, p. 232–283.

[64] F. Rusek, D. Persson, B. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, *Scaling up MIMO: opportunities and challenges with very large arrays*, IEEE Signal Processing Magazine, 30 (2013), pp. 40–60.

[65] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, *Non-orthogonal multiple access (NOMA) for cellular future radio access*, in 2013 IEEE 77th Vehicular Technology Conference (VTC Spring), IEEE, 2013, pp. 1–5.

[66] S. Shafiee, N. Liu, and S. Ulukus, *Towards the secrecy capacity of the Gaussian MIMO wire-tap channel: The 2-2-1 channel*, IEEE Transactions on Information Theory, 55 (2009), pp. 4033–4039.

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

[68] ——, *Communication theory of secrecy systems*, The Bell System Technical Journal, 28 (1949), pp. 656–715.

[69] N. P. Smart, *Cryptography Made Simple*, Springer, 2016.

[70] J. Soni and R. Goodman, *A mind at play: how Claude Shannon invented the information age*, Simon and Schuster, 2017.

[71] E. Telatar, *Capacity of multi-antenna Gaussian channels*, European transactions on telecommunications, 10 (1999), pp. 585–595.

[72] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.

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

[74] P. Yang, Y. Xiao, M. Xiao, and S. Li, *6G wireless communications: Vision and potential techniques*, IEEE Network, 33 (2019), pp. 70–75.

[75] L. Zhang and D. Guo, *Wireless peer-to-peer mutual broadcast via sparse recovery*, in 2011 IEEE International Symposium on Information Theory Proceedings, July 2011, pp. 1901–1905.

[76] H. Zimmermann, *OSI Reference Model - The ISO Model of Architecture for Open Systems Interconnection*, IEEE Transactions on Communications, 28 (1980), pp. 425–432.