University of Novi Sad
Faculty of Sciences
Department of Mathematics and Informatics

Tatjana Naumov

The capacity of multiple-access channels

Master thesis

Mentor:
Mladen Kovačević, PhD

Novi Sad, 2024

I would like to express my heartfelt gratitude to my mentor Dr Mladen Kovačević for his advices, help, support, suggestions, invaluable guidance and expertise during the research and writing of this thesis.

Additionally, I extend my thanks to the members of the board for their valuable insights, feedback throughout the process and for reviewing this thesis.

I could not have undertaken this journey without the support of my family, partner and friends for their unconditional love, understanding and encouragement. Their belief in me and unwavering support have kept my spirits, strength and motivation high during this process. I am forever thankful for that.

I am grateful to my colleagues for their encouragement, support and willingness to lend a helping hand whenever needed.

Finally, I would like to thank the faculty and all other professors for providing an enriching academic environment, as this master's program has been inspiring experience to me.

This achievement would not have been possible without the support and guidance of each every one of you. Thank you for believing in me and being a part of this journey.

Contents

Chapter 1

Introduction

When we say that A communicates with B, it implies that the actions performed by A have resulted in a desired physical state in B. This information transfer is a physical process and is susceptible to uncontrollable ambient noise and imperfections in the signaling process itself. The communication is considered successful when both the sender A and the receiver B agree on the transmitted content. We will explore the maximum number of distinguishable signals achievable through n uses of a communication channel. This number increases exponentially with n, and the exponent is referred to as the channel capacity. The central and most renowned achievement of information theory is the characterization of channel capacity (represented as the logarithm of the number of distinguishable signals) as the maximum mutual information.

A communication system that involves multiple senders and receivers introduces new elements such as interference, cooperation, and feedback. These elements fall within the domain of network information theory. The overarching problem is straightforward: given numerous senders and receivers, along with a channel transition matrix that describes interference and noise effects in the network, the task is to determine whether the sources can be successfully transmitted over the channel. This problem encompasses distributed source coding (data compression) and distributed communication (identifying the capacity region of the network). Notable examples of communication networks include computer networks, satellite networks, and the telephone system. Even within a single computer, different components engage in communication with one another. A comprehensive theory of network information would have significant implications for the design of communication and computer networks.

Chapter 2

Channel capacity

2.1 Communication system

Figure 2.1 illustrates a schematic representation of a physical signaling system. Source symbols, originating from a finite alphabet, are mapped into a sequence of channel symbols, which generates the output sequence of the channel. Although the output sequence is random, it follows a distribution that depends on the input sequence. Our goal is to recover the transmitted message based on the output sequence. Let us explain how communication system works. The message will be modeled as a random variable W because the receiver does not know in advance which message will be sent. It is some random message selected from the set of all possible messages. When we want to transmit messages, we do not want to transmit them in their original form for the following reason: the channel is going to introduce some noise, so the receiver on the other side will not receive the same message and, if we want to prevent errors from happening, first we will have to perform operations on that message and transmit X^n instead. It is another random object which is obtained by encoding the message. We are performing some function on the message and we are producing sequence of symbols, sequence of n random variables. We encode our message and we produce sequence of n random symbols. The sequence is transmitted over communication channel. Now at the output of this channel receiver obtains another sequence which is not necessarily equal to the transmitted sequence and is denoted by Y^n. The receiver will try to recover the original message by applying reverse operation called decoding. The receiver's estimate of the message is denoted by $\hat{W}$; it is in general no equal to W, because sometimes receiver will make a mistake. Our goal is to achieve the probability of correct decoding, $P(\hat{W} = W)$, as high as possible (ideally, it should be nearly 1).

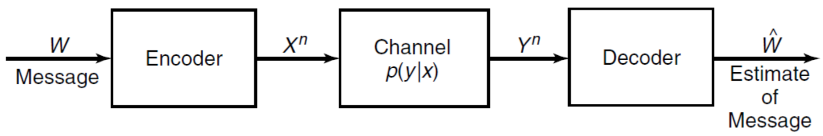

Figure 2.1: Communication system. (taken from [3])

2.2 Example of a communication system

If we simply send a sequence of bits of length n, approximately pn of them will be received incorrectly and receiver does not know which ones are wrong. Receiver has no way of recovering original message. What can we do? We can do some encoding operation. Instead of sending individual bits we will send code words. Let us give an example of sending code words, see Figure 2.2. Instead of 0 we will send through communication channel 000 and instead of 1 we will send 111, so each time we want to send a bit we will repeat 3 times. What happens now? If we transmit 000, the receiver could attain 010. The only two possible code words that can be transmitted are 000 which represents message 0 and 111 which represents message 1. If receiver gets 010 it knows that something is wrong, because the sequence 010 could not have been transmitted, so it is going to assume that 000 was sent. Channel is introducing some errors. The channel could also have produced 010 from 111, but probability of this happening is much smaller. We say that the 000 is more probable and we will decode it as 000. This is the whole point of coding. So we have list of all forbidden sequences, these cannot be at the channel input and therefore when receiver obtains them at the channel output he knows something is wrong.

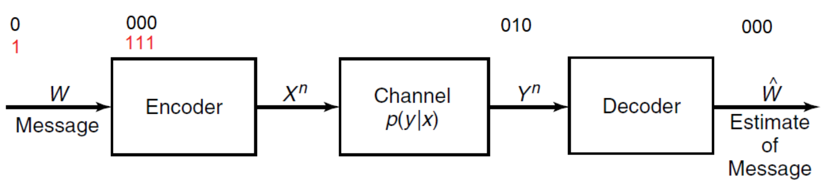

Figure 2.2: Example of communication system sending code words. (taken from [3])

This is the point of communication. We represent message W in some way by using sequence of symbols X^n and this operation is called encoding. This sequence X^n is always longer than the original message. We are adding symbols which are not carrying information, but whose purpose is just to protect our message. Then we transmit this over communication channel, something is going to happen in the channel and the receiver is going to get another sequence which is possibly different and then, based on what he received, it will produce estimate of message. It is going to produce best estimate which is not always going to be correct, but it will be correct most of the time (and this is called reliable communication).

2.3 Definition of channel capacity

Definition 2.3.1 (Channel capacity)**.** The *"information" capacity* of a discrete memoryless channel is defined by

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

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

The mutual information, denoted as $I(X;Y)$, is a fundamental measure in information theory that measures the amount of information shared between two random variables, X and Y. It provides insights into how much information X reveals about Y (or vice versa), and it is defined by

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

Explanation: Channel capacity is the maximum of mutual information between input and the output. The mutual information depends on the joint distribution. When we maximize the marginal distribution $p(x)$, the resulting quantity depends only on the conditional distribution $p(y|x)$ (the channel). It is a celebrated result of Shannon – the so-called *channel coding theorem* – that C represents the maximum number of bits per symbol that can be transmitted through a given channel with arbitrarily small error probability.

2.4 Examples of channel capacity

2.4.1 Binary symmetric channel

Let us examine the binary symmetric channel (BSC), showed in Figure 2.3. This channel operates with binary input symbols, which are flipped with a probability denoted as p. Although it represents the simplest model of a channel with errors, it captures most of the complexity of the general problem.

When an error arises, a received 0 is interpreted as 1, and vice versa. Notably, the received bits do not disclose the error locations. In a way, all the received bits lack reliability.

Suppose that there are only 2 possible messages. We are trying to send only two possible messages through channel, 0 or 1. If we transmit 0 or 1 receiver on the other side might not receive 0 or 1, so channel will introduce some distortion. In the channel input we have 0 or 1 and on the output we can receive also 0 or 1. If 0 was transmitted, 0 will be received with some probability $1-p$, but

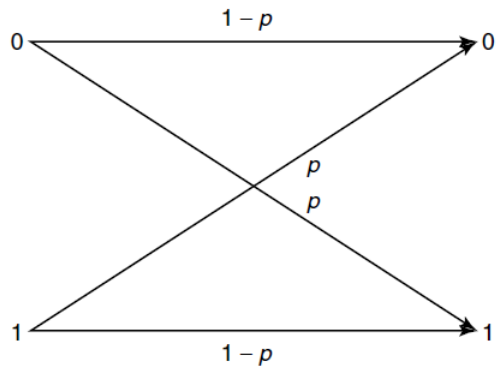

Figure 2.3: Binary symmetric channel. $C = 1 - H(p)$ bits. (taken from [3])

1 will be received with some small probability p. Even if 0 was transmitted that does not mean that 0 will be received all the time and vice versa. If we transmit a bit, we can sometimes receive a wrong bit and this is what we call noisy channel. This channel is one of the basic examples of noisy channel.

This diagram represents conditional distribution, how we describe communication channel, so we describe what is going to happen at the channel output given the channel input. Capacity is $1 - H(p)$, where $H(p) = -p \log p - (1 - p)\log(1-p)$ is binary entropy function and it is less then one, because receiver is confused sometimes and channel is noisy, we cannot send one bit per symbol.

2.4.2 Binary erasure channel

The analog to the binary symmetric channel, where certain bits are not corrupted but lost, is the binary erasure channel. In this channel, a fraction α of the bits are erased. The receiver possesses information about which bits have been erased. The binary erasure channel has two inputs and has three outputs.

As said, binary erasure channel has binary input alphabet, output alphabet has three symbols and it is called erasure for the following reason: If the symbol 0 was transmitted, 0 will be received with prob $1 - \alpha$ and with probability α symbol e will be received, so each input bit will be erased with probability α and input will be received with probability $1 - \alpha$. Capacity is $1 - \alpha < 1$. When receiver gets e it is confused.

2.5 Network information theory

Consider a scenario where m stations aim to communicate with a satellite through a common channel. This scenario is referred to as a multiple-access channel. How do the senders collaborate to transmit information to the receiver?

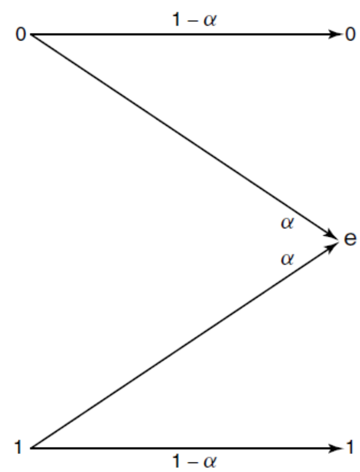

Figure 2.4: Binary erasure channel. (taken from [3])

What communication rates can be achieved simultaneously? How does interference among the senders limit the total communication rate? These questions are well-understood in the context of the multiuser channel and have satisfactory answers.

On the other hand, let us envision a scenario where one TV station transmits information to m TV receivers, as depicted in Figure 2.5. How does the sender encode information intended for different receivers in a unified signal? What rates of information transmission are achievable for the various receivers? Regarding this channel, answers are only known for specific cases. Several other

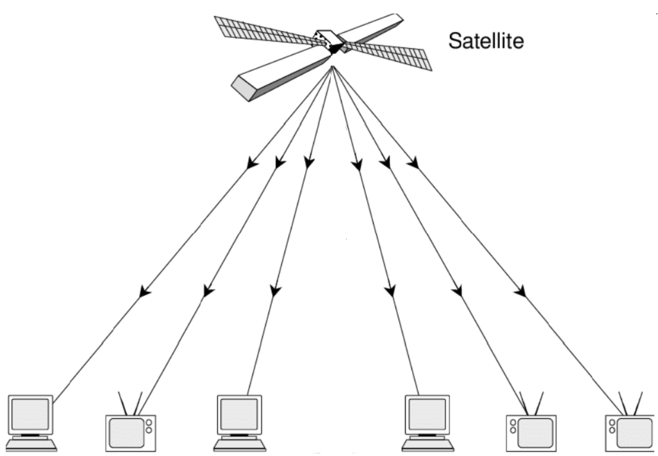

Figure 2.5: Broadcast channel. (taken from [8])

types of channels exist within network information theory. These include the relay channel, where there is a single source and destination, but intermediate sender-receiver pairs act as relays to facilitate communication; the interference channel, involving two senders and two receivers with crosstalk; and the two-way channel, where two sender-receiver pairs exchange information. While some answers regarding achievable communication rates and coding strategies exist for these channels, they are not yet fully understood. All these channels can be viewed as special cases of a communication network, as illustrated in Figure 2.6. In this network, m nodes attempt to communicate with each other. At any

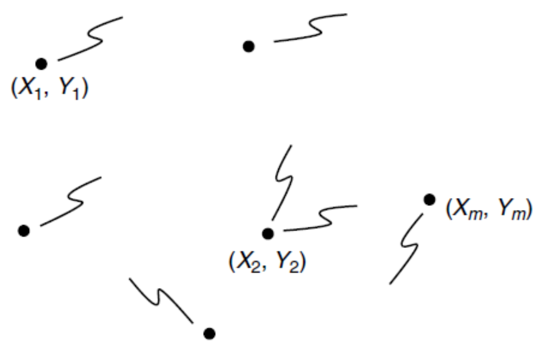

Figure 2.6: Communication network. (taken from [3])

given time, the i node transmits a symbol x_i based on the messages it intends to send and the symbols it has previously received. The simultaneous transmission of symbols (x_1, x_2, ..., x_m) results in random received symbols (Y_1, Y_2, ..., Y_m), which are drawn from a conditional probability distribution p($y^{(1)}$, $y^{(2)}$, ..., $y^{(m)}|x^{(1)}$, $y^{(1)}$, ..., $x^{(1)}$). This distribution accounts for the effects of noise and interference present in the network. If $p(\cdot|\cdot)$ can only take on the values 0 and 1, the network is deterministic.

Chapter 3

Multiple-access channel

In this chapter we will begin by closely analyzing the multiple-access channel, where two or more senders transmit information to a shared receiver. The channel is depicted in Figure 3.1. A typical instance of this channel is a satellite receiver with many independent ground stations, or a group of cell phones communicating with a central base station. It is important to note that the senders not only have to deal with receiver noise but also contend with interference caused by each other.

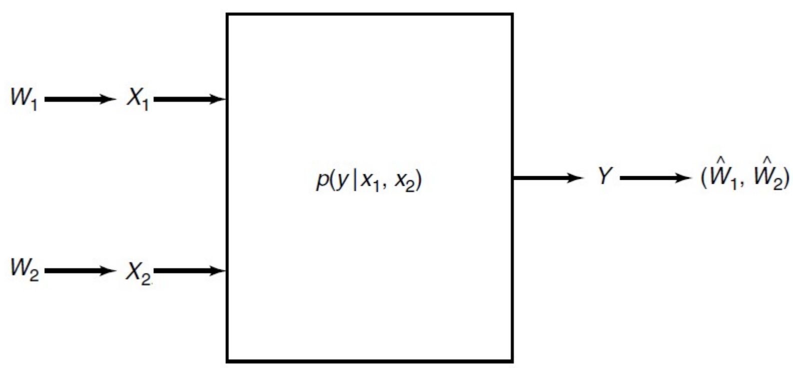

Figure 3.1: Multiple-access channel. (taken from [3])

3.1 Some properties and examples of multiple access channels

The code for the multiple-access channel, denoted as $((2^{nR_1}, 2^{nR_2}), n)$, involves two sets of integers: $\mathcal{W}_1 = \{1, 2, \ldots, 2^{nR_1}\}$ and $\mathcal{W}_2 = \{1, 2, \ldots, 2^{nR_2}\}$, which are referred to as the *message sets*. It consists of two *encoding functions*:

$$X_1 : \mathcal{W}_1 \to \mathcal{X}_1^n \tag{3.1}$$

$$X_1 : \mathcal{W}_2 \to \mathcal{X}_2^n \tag{3.2}$$

Additionally, there is a decoding function:

$$g : \mathcal{Y}^n \to \mathcal{W}_1 \times \mathcal{W}_2 \tag{3.3}$$

In this setup, there are two senders and one receiver. Sender 1 randomly selects an index from the set $\{1, 2, \ldots, 2^{nR_1}\}$, transmits the corresponding codeword through the channel, and Sender 2 does the same.
Assuming that the distribution of messages over the product set $\mathcal{W}_1 \times \mathcal{W}_2$ is uniform, meaning the messages are independent and equally likely, the *average probability of error* for the $((2^{nR_1}, 2^{nR_2}), n)$ code is defined as follows:

$$\mathcal{P}_e^{(n)} : \frac{1}{2^{n(R_1+R_2)}} \sum_{(\omega_1,\omega_2)\in\mathcal{W}_1\times\mathcal{W}_2} Pr\left\{g(Y^n) \neq (\omega_1, \omega_2)|(\omega_1, \omega_2) \text{ sent}\right\} \tag{3.4}$$

Definition 3.1.1. We consider a rate pair (R_1, R_2) to be *achievable* for the multiple-access channel when a sequence of $((2^{nR_1}, 2^{nR_2}), n)$ codes exists with

$$P_e^{(n)} \to 0. \tag{3.5}$$

Definition 3.1.2. The *capacity region* of the multiple-access channel refers to the closure of the set that includes all achievable rate pairs (R_1, R_2).

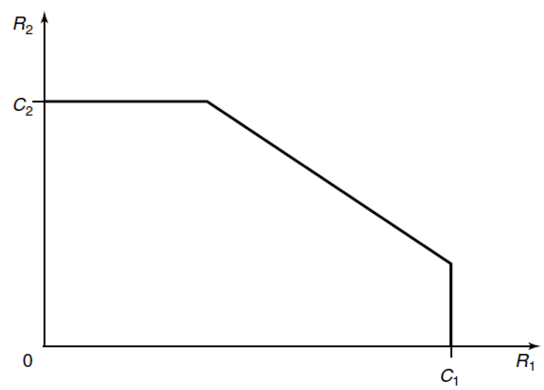

Figure 3.2: Capacity region for a multiple-access channel. (taken from [3])

An example of the capacity region for a multiple-access channel is shown in Figure 3.2. We will illustrate the capacity region with an example after we formally state theorem that presents capacity region.

Theorem 3.1.1. *The capacity of a multiple-access channel*

$$(\mathcal{X}_1 \times \mathcal{X}_2, \mathrm{p}(y|x_1, x_2), \mathcal{Y}) \tag{3.6}$$

that has two senders x_1, x_2 and a receiver y is given by the closure of the convex hull of all (R_1, R_2) pairs that satisfy the following conditions:

$$R_1 < I(X_1; Y|X_2) \tag{3.7}$$

$$R_1 < I(X_2; Y|X_1) \tag{3.8}$$

$$R_1 + R_2 < I(X_1, X_2; Y) \tag{3.9}$$

These conditions must hold for some product distribution $p_1(x_1)p_2(x_2)$ on $\mathcal{X}_1 \times \mathcal{X}_2$.

Before we present the proof that this forms the capacity region of the multiple-access channel, let us examine a few examples of multiple-access channels.

3.1.1 Independent binary symmetric channels

Let us consider a scenario where we possess two independent binary symmetric channels: one originating from sender 1 and the other from sender 2, as illustrated in Figure 3.3. It is clear that in this situation, we can transmit information at a rate of $1 - H(p_1)$ through the first channel and a rate of $1 - H(p_2)$ through the second channel. Due to the independence of the channels, there exists no mutual disruption between the senders. The capacity region in this scenario can be observed in Figure 3.4.

3.1.2 Binary multiplier channel

Let us consider a multiple access channel equipped with binary inputs and output

$$Y = X_1 X_2 \tag{3.10}$$

This channel is referred to as a binary multiplier channel. It is evident that by setting $X_2 = 1$, we can effectively transmit at a rate of 1 bit for each transmission from sender 1 to the receiver. Similarly, with $X_1 = 1$, we can attain a rate of $R_2 = 1$. Naturally, given that the output remains binary, the combined rates $R_1 + R_2$ for sender 1 and sender 2 must not exceed 1 bit. Through time division, we have the capability to realize any combination of rates such that the sum of $R_1 + R_2$ equals 1. Consequently, the capacity region is illustrated in Figure 3.5.

3.1.3 Binary erasure multiple-access channel

This particular multiple-access channel is characterized by binary inputs, where X_1 and X_2 can take values from the set $\{0, 1\}$ and it generates a ternary

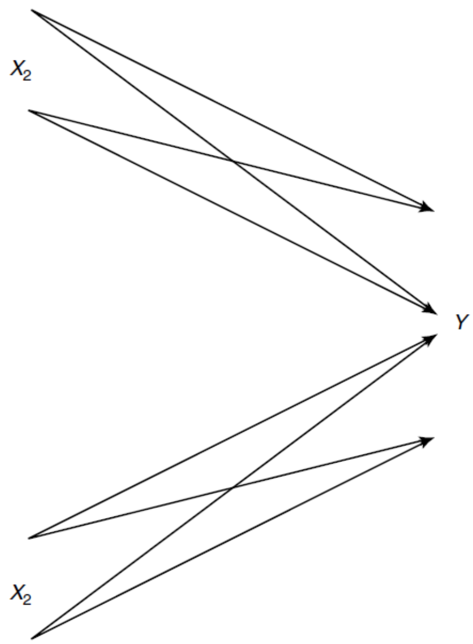

Figure 3.3: Independent binary symmetric channels. (taken from [3])

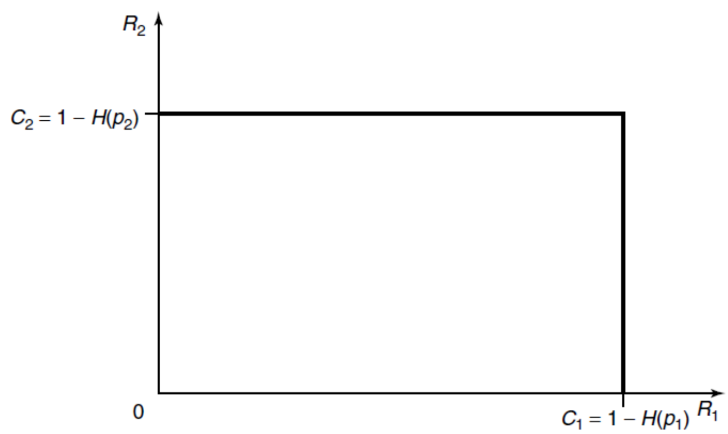

Figure 3.4: Capacity region for independent BSCs. (taken from [3])

output, Y, defined as the sum of X_1 and X_2,

$$Y = X_1 + X_2. \tag{3.11}$$

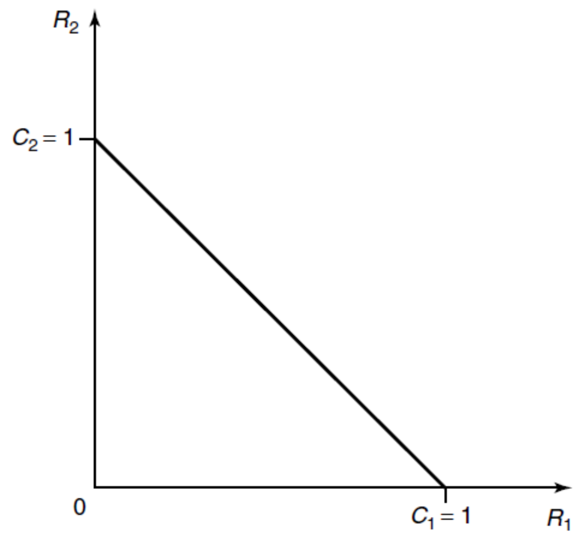

Figure 3.5: Capacity region for binary multiplier channel. (taken from [3])

There is no uncertainty in $(X_1 + X_2)$ if $Y = 0$ or $Y = 2$ is received, but $Y = 1$ can result from either (0,1) or (1,0).

Now, we will investigate the achievable rates along the axes. By setting $X_2 = 0$, we can transmit at a rate of 1 bit per transmission from sender 1. In a similar manner, with $X_1 = 0$, we can achieve a rate of $R_2 = 1$. This gives us two extreme points of the capacity region. Can improvements be made beyond these points? Let us assume $R_1 = 1$, leading to X_1 codewords must include all conceivable binary sequences, essentially resembling a Bernoulli(1/2) process. This behaves like noise during the transmission of X_2. For X_2, the channel looks like the channel in Figure 3.6. Drawing from the findings, the capacity of this channel amounts to 1/2 bit per transmission. Consequently, when transmitting at the maximum rate of 1 for sender 1, we can concurrently send an additional 1/2 bit from sender 2. Subsequently, once we derive the capacity region, we can substantiate that these rates represent the optimum achievable rates. The capacity region for a binary erasure channel is illustrated in Figure 3.7.

3.2 Achievability of the capacity region for the multiple-access channel

We are now going to establish the achievability of the rate region as stated in Theorem 3.1.1. The converse proof will be deferred until the subsequent section. The demonstration of achievability is very similar to the proof for the single-

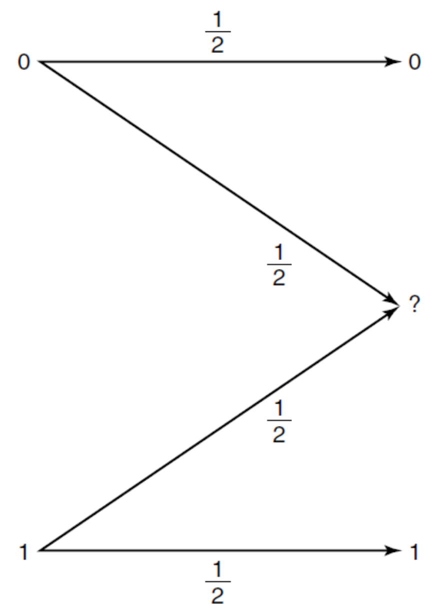

Figure 3.6: Equivalent single-user channel for user 2 of a binary erasure multiple-access channel. (taken from [3])

user channel, with a focus on highlighting the deviations from the single-user scenario. We start with proving the achievability that satisfy (3.9) for some fixed product distribution $p_1(x_1)p_2(x_2)$.

3.2.1 Achievability in theorem 3.1.1

Before we start with the proof, let us briefly explain what is asymptotic equipartition property (AEP). The concept of entropy in information theory is related to the concept of entropy in statistical. When we consider a sequence of n independent and identically distributed (i.i.d.) random variables, we can demonstrate that the likelihood of a "typical" sequence approximates $2^{-nH(X)}$ and there are approximately $2^{nH(X)}$ such typical sequences.

Let us state two theorems that are going to be useful.

Definition 3.2.1. Let $(X_1, X_2, \ldots, X_k)$ denote a finite collection of discrete random variables with some fixed joint distribution,

$$p(x^{(1)}, x^{(2)}, \ldots, x^{(k)}), \quad (x^{(1)}, x^{(2)}, \ldots, x^{(k)}) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_k.$$

Let S denote an ordered subset of these random variables and consider n copies of S.

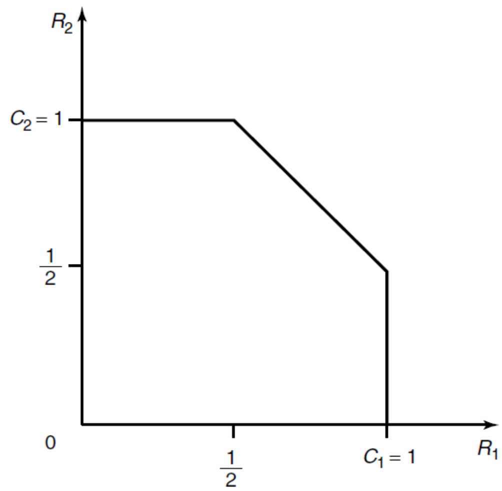

Figure 3.7: Capacity region for binary erasure multiple-access channel. (taken from [3])

Definition 3.2.2. The set $A_\epsilon^{(n)}$ of ϵ-typical n-sequences $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$ is defined by

$$A_\epsilon^{(n)}(X^{(1)}, X^{(2)}, \ldots, X^{(k)})$$
$$= A_\epsilon^{(n)}$$
$$= \left\{ (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k) : \left| -\frac{1}{n} \log p(\mathbf{s}) - H(S) \right| < \epsilon, \forall S \subseteq \{X^{(1)}, X^{(2)}, \ldots, X^{(k)}\} \right\}$$

Definition 3.2.3. Let $A_\epsilon^{(n)}(S)$ denote the restriction of $A_\epsilon^{(n)}$ to the coordinates of S. Thus, if $S = (X_1, X_2)$, we have:

$$\begin{aligned} A_\epsilon^{(n)}(X_1, X_2) = \{ & (\mathbf{x}_1, \mathbf{x}_2) : \\ & \left| -\frac{1}{n} \log p(\mathbf{x}_1, \mathbf{x}_2) - H(X_1, X_2) \right| < \epsilon, \\ & \left| -\frac{1}{n} \log p(\mathbf{x}_1) - H(X_1) \right| < \epsilon, \\ & \left| -\frac{1}{n} \log p(\mathbf{x}_2) - H(X_2) \right| < \epsilon \}. \end{aligned}$$

Theorem 3.2.1. *For any $\epsilon > 0$, for sufficiently large n,*

- $P(A_\epsilon^{(n)}(S)) \geq 1 - \epsilon, \quad \forall S \subseteq \left\{ X^{(1)}, X^{(2)}, \ldots, X^{(k)} \right\}$

- $\mathbf{s} \in A_\epsilon^{(n)}(S) \Rightarrow p(\mathbf{s}) = 2^{n(H(S)\pm\epsilon)}$.

- $\left|A_\epsilon^{(n)}(S)\right| = 2^{n(H(S)\pm 2\epsilon)}$.

- Let $S_1, S_2 \subseteq \left\{X^{(1)}, X^{(2)}, \ldots, X^{(k)}\right\}$. If $(s_1, s_2) \in A_\epsilon^{(n)}(S_1, S_2)$, then $p(s_1|s_2) = 2^{n(H(S_1|S_2)\pm 2\epsilon)}$.

Theorem 3.2.2. *Let $A_\epsilon^{(n)}$ denote the typical set for the probability mass function $p(s_1, s_2, s_3)$, and let*

$$P(S_1^{'} = s_1, S_2^{'} = s_2, S_3^{'} = s_3) = \prod_{i=1}^{n} p(s_{1i}|s_{3i})p(s_{2i}|s_{3i})p(s_{3i}). \tag{3.12}$$

Then

$$P\left\{(S_1^{'}, S_2^{'}, S_3^{'}) \in A_\epsilon^{(n)}\right\} = 2^{n(I(S_1;S_2|S_3)\pm 6\epsilon)}. \tag{3.13}$$

Proof. Fix $(p_1, p_2) = p_1(x_1)p_2(x_2)$.
Codebook Generation: Generate 2^{nR_1} independent codewords $\mathbf{X}_1(i)$, $i \in\{1,\ 2,\ \ldots,\ 2^{nR_1}\}$, of length n. Generate each element independently and identically distributed (i.i.d.) according to the distribution $\sim \prod_{i=1}^{n} p_1(x_{1i})$. Similarly, produce 2^{nR_2} independent codewords $\mathbf{X}_2(j)$, where $j \in \{1,\ 2,\ \ldots,\ 2^{nR_2}\}$, with each codeword element being i.i.d. $\sim \prod_{i=1}^{n} p_2(x_{2i})$. These codewords collectively form the codebook, which is disclosed to both senders and the receiver.

Encoding: For transmitting the index i, sender 1 sends the corresponding codeword $\mathbf{X}_1(i)$. Correspondingly, sender 2 transmits $\mathbf{X}_2(j)$ to convey the index j.

Decoding: Define $A_\epsilon^{(n)}$ as the set of typical $\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}$ sequences. The receiver Y^n selects the pair (i, j) such that

$$(\mathbf{x}_1(i), \mathbf{x}_2(j), \mathbf{y}) \in A_\epsilon^{(n)} \tag{3.14}$$

if such a unique pair (i, j) exists; otherwise, an error is indicated.

Analysis of the probability of error: Due to the symmetry inherent in the random code construction, the likelihood of error conditioned on specific index pairs transmitted remains unaffected. Consequently, the conditional probability of error coincides with the unconditional probability of error. Therefore, we can presume, without any loss of generality, that the pair $(i, j) = (1, 1)$ was transmitted.

An error is observed when either the accurate codewords fail to exhibit typicality with the received sequence, or when an erroneous pair of codewords displays typicality with the received sequence. Let us define events:

$$E_{i,j} = \left\{(\mathbf{X}_1(i), \mathbf{X}_2(j), \mathbf{Y}) \in A_\epsilon^{(n)}\right\} \tag{3.15}$$

Then by the union of events bound,

$$P_\epsilon^{(n)} = P\Big(E_{11}^c \bigcup \cup_{(i,j)\neq(1,1)} E_{ij}\Big) \tag{3.16}$$

$$\leq P(E_{11}^c) + \sum_{i\neq 1} P(E_{i1}) + \sum_{j\neq 1} P(E_{1j}) + \sum_{i\neq 1, j\neq 1} P(E_{ij}) \tag{3.17}$$

where P is the conditional probability given that $(1,1)$ was sent. From the AEP, $P(E_{11}^c) \to 0$. By Theorems 3.2.1 and 3.2.2, for $i \neq 1$, we have

$$P(E_{i1}) = P((\mathbf{X}_1(i), \mathbf{X}_2(j), \mathbf{Y}), \in A_\epsilon^{(n)}) \tag{3.18}$$

$$= \sum_{(x_1,x_2,y)\in A_\epsilon^{(n)}} p(\mathbf{x}_1)p(\mathbf{x}_2, \mathbf{y}) \tag{3.19}$$

$$\leq \left| A_\epsilon^{(n)} \right| 2^{-n(H(X_1)-\epsilon)} 2^{-n(H(X_2,Y)-\epsilon)} \tag{3.20}$$

$$\leq 2^{-n(H(X_1)+H(X_2,Y)-H(X_1,X_2,Y)-3\epsilon)} \tag{3.21}$$

$$= 2^{-n(I(X_1;X_2,Y)-3\epsilon)} \tag{3.22}$$

$$= 2^{-n(I(X_1;Y|X_2)-3\epsilon)} \tag{3.23}$$

where the equivalence of (3.22) and (3.23) follows from the independence of X_1 and X_2, and the consequent $I(X_1;X_2,Y) = I(X_1;X_2) + I(X_1;Y|X_2) = I(X_1;Y|X_2)$. Similarly, for $j \neq 1$,

$$P(E_{i1}) \leq 2^{-n(I(X_2;Y|X_1)-3\epsilon)} \tag{3.24}$$

and for $i \neq 1, j \neq 1$

$$P(E_{i1}) \leq 2^{-n(I(X_1,X_1;Y)-4\epsilon)} \tag{3.25}$$

It follows that

$$P_\epsilon^{(n)} \leq P(E_{11}^c) + 2^{nR_1}2^{-n(I(X_1;Y|X_2)-3\epsilon)} + 2^{nR_2}2^{-n(I(X_2;Y|X_1)-3\epsilon)} + 2^{n(R_1+R_2)}2^{-n(I(X_1,X_1;Y)-4\epsilon)} \tag{3.26}$$

As $\epsilon > 0$ is arbitary, the theorem's criteria lead to the conclusion that each term tends to zero as n approaches infinity, $n \to \infty$. Consequently, if the theorem's conditions are satisfied, the likelihood of error, contingent on a specific codeword being transmitted, tends to zero. The earlier stated bound indicates that the average probability of error, uniform due to symmetry and equal to the probability concerning a single codeword, averages across all potential codebook selections in the random code construction, is arbitrarily small. Therefore, there exists at least one code C^* with arbitrarily small probability of error. □

3.2.2 Comments on the capacity region for the multiple-access channel

We have proved the achievability of the capacity region for the multiple-access channel. This region represents the closure of the convex hull formed by the collection of points R_1, R_2 that adhere to the conditions:

$$R_1 < I(X_1; Y|X_2) \tag{3.27}$$

$$R_1 < I(X_2; Y|X_1) \tag{3.28}$$

$$R_1 + R_2 < I(X_1, X_2; Y) \tag{3.29}$$

for some product distribution $p_1(x_1)p_2(x_2)$ on $\mathcal{X}_1 \times \mathcal{X}_2$. Illustrated in Figure 3.8 is the region corresponding to a specific $p_1(x_1)p_2(x_2)$. Now, let us analyze the

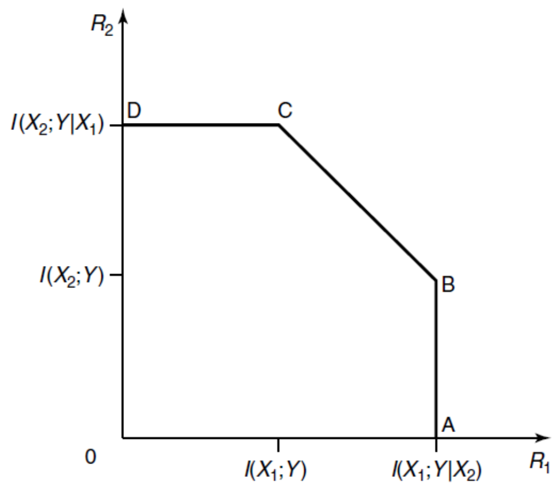

Figure 3.8: Achievable region of multiple-access channel for a fixed input distribution. (taken from [3])

significance of the corner points within the region. Point A corresponds to the maximum rate from sender 1 to the receiver under the condition that sender 2 is not transmitting any information. This corresponds to

$$\max R_1 = \max_{p_1(x_1)p_2(x_2)} I(X_1; Y|X_2). \tag{3.30}$$

Now, considering any given distribution $p_1(x_1)p_2(x_2)$,

$$I(X_1; Y|X_2) = \sum_{x_2} p_2(x_2) I(X_1; Y|X_2 = x_2) \tag{3.31}$$

$$\leq \max_{x_2} I(X_1; Y|X_2 = x_2), \tag{3.32}$$

since the average is less than the maximum. The maximum value in (3.30) is achieved by assigning $X_2 = x_2$, with x_2 being value that maximizes the conditional mutual information between X_1 and Y. The distribution of X_1 is chosen to maximize this mutual information. Consequently, X_2 needs to aid the transmission of X_1 by being set as $X_2 = x_2$.

Point B signifies the maximum rate at which sender 2 can transmit while ensuring sender 1 operates at their maximum rate. If X_1 is considered as noise for the channel from X_2 to Y, this is the rate that is obtained. In this scenario, when using the results from single-user channels, X_2 can transmit information at a rate of $I(X_2; Y)$. The receiver possesses knowledge about the specific X_2 codeword employed, allowing for the "subtraction" of its influence from the channel. This situation can be viewed as an assortment of single-user channels indexed by the X_2 symbol and in this case X_1 rate achieved is the average mutual information, where the average is conducted across these channels, with each channel repeating as frequently as the corresponding X_2 symbol emerges in the codewords. Therefore, the rate achieved is

$$\sum_{x_2} p(x_2) I(X_1; Y|X_2 = x_2) = I(X_1; Y|X_2) \tag{3.33}$$

Points C and D represent B and A, respectively, but with the positions of the senders switched. The intermediary points can be realized through time-sharing. Hence, we have provided an single-user interpretation and rationale for the capacity region of a multiple-access channel.

The concept of treating additional signals as components of noise, decoding one signal and "subtracting" it from the received signal is an exceptionally valuable approach.

3.3 Convexity of the capacity region of the multiple-access channel

We are currently reconfiguring the capacity region of the multiple-access channel to incorporate the process of forming the convex hull, and this involves the introduction of a new random variable. Our initial step involves demonstrating the convexity of the capacity region through a proof.

Theorem 3.3.1. *The capacity region C of a multiple-access channel is convex, meaning that if $R_1, R_2 \in C$ and $R_1^{'}, R_2^{'} \in C$, then for any value of λ between 0 and 1, $0 \leq \lambda \leq 1$, the point $(\lambda R_1 + (1-\lambda)R_1^{'}, \lambda R_2 + (1-\lambda)R_2^{'}) \in C$.*

Proof. The concept involves employing time-sharing. With two sets of code sequences operating at distinct rates, $\mathbf{R} = (R_1, R_2)$ and $\mathbf{R}^{'} = (R_1^{'}, R_2^{'})$, we have the ability to create a third codebook operating at a mixed rate of $\lambda\mathbf{R} + (1-\lambda)\mathbf{R}^{'}$, achieved by using the first codebook for the initial λn symbols and using to the second codebook for the remaining $(1-\lambda)n$ symbols. The count of X_1 codewords present in the updated code can be expressed as

$$2^{n\lambda R_1} 2^{n(1-\lambda R_1^{'})} = 2^{n(\lambda R_1 + (1-\lambda) R_1^{'})} \tag{3.34}$$

Consequently, the rate of the new code equals to $\lambda \mathbf{R} + (1-\lambda)\mathbf{R}^{'}$. Given that the collective likelihood of an error is lower than the total of the error probabilities for each individual segment, the error probability of the new code approaches zero. This signifies that the rate is achievable. □

We can proceed to reformulate the expression of the capacity region concerning the multiple-access channel by introducing a random variable Q that involves time-sharing. However, before we establish the proof for this outcome, it becomes necessary to prove a property of convex sets that are defined by linear inequalities similar to those of the capacity region of the multiple-access channel. Specifically, our aim is to establish that the convex hull formed by two such regions, defined by linear constraints corresponds to the region defined by the convex combination of these conditions. While initially, the equivalence between these two sets might appear apparent, upon closer analysis, a subtle complexity arises due to the possibility of certain constraints not being actively involved. This concept is most effectively demonstrated through the use of an example. Consider these following two sets that are defined by linear inequalities:

$$C_1 = \{(x, y) : x \geq 0, y \geq 0, x \leq 10, y \leq 10, x + y \leq 100\} \tag{3.35}$$

$$C_2 = \{(x, y) : x \geq 0, y \geq 0, x \leq 20, y \leq 20, x + y \leq 20\} \tag{3.36}$$

The $(\frac{1}{2}, \frac{1}{2})$ convex combination of the constraints, in this case, defines the region

$$C = \{(x, y) : x \geq 0, y \geq 0, x \leq 15, y \leq 15, x + y \leq 60\} \tag{3.37}$$

It is easy to observe that any point within C_1 or C_2 adheres to the condition $x + y < 20$. Therefore, any point situated within the convex combination of the combined regions C_1 and C_2 fulfills this criterion. Consequently, the point (15,15), which belongs to C, does not fall within the convex combination of $C_1 \cup C_2$. This example also hints at the cause of the problem and that is definition of C_1 where the constraint $x + y \leq 100$ remains inactive. If this constraint were substituted with a constraint $x + y \leq a$, where $a \leq 20$, the previously mentioned equivalence between the two regions would hold true. We will prove now that above result of the equality of the two regions would be true.

We confine our focus to the pentagonal regions that emerge as constituents within the capacity region of a two-user multiple-access channel. In this specific instance, the capacity region for a fixed $p(x_1)p(x_2)$ is characterized by three distinct mutual information values: $I(X_1; Y|X_2)$, $I(X_2; Y|X_1)$ and $I(X_1, X_2; Y)$, which we shall denote as I_1, I_2 and I_3 respectively. For every $p(x_1)p(x_2)$ distribution, a corresponding vector $I = (I_1, I_2, I_3)$ is assigned, and a rate region defined by:

$$C_{\mathbf{I}} = \{(R_1, R_2) : R_1 \geq 0, R_2 \geq 0, R_1 \leq I_1, R_2 \leq I_2, R_1 + R_2 \leq I_3\} \tag{3.38}$$

Moreover, as applicable to any given distribution $p(x_1)p(x_2)$, the equation

$$\begin{aligned} I(X_2; Y|X_1) &= H(X_2|X_1) - H(X_2|Y, X_1) \\ &= H(X_2) - H(X_2|Y, X_1) \\ &= I(X_2; Y, X_1) \\ &= I(X_2; Y) + I(X_2; X_1|Y) \end{aligned}$$

holds. This implies that $I(X_2; Y|X_1)+I(X_1; Y|X_2) \geq I(X_1; Y|X_2)+I(X_2; Y) = I(X_1, X_2; Y)$. Consequently, for all vectors I, the relationship $I_1 + I_2 \geq I_3$ remains valid. This property will turn out to be critical for the theorem.

Lemma 3.3.1. *Consider two vectors of mutual information, denoted as* $\mathrm{I}_1, \mathrm{I}_2 \in R_3$*, defining rate regions* $C_{\mathrm{I}_1}, C_{\mathrm{I}_2} \in R_3$*, respectively, as given in equation (3.38).* $0 \leq \lambda \leq 1$*, define* $\mathrm{I}_\lambda = \lambda \mathrm{I}_1 + (1-\lambda)\mathrm{I}_2$*, and let* C_{I_λ} *represent the rate region determined by* I_λ*. Then*

$$C_{\mathrm{I}_\lambda} = \lambda C_{\mathrm{I}_1} + (1-\lambda)C_{\mathrm{I}_2} \tag{3.39}$$

Proof. Theorem will be proved in two parts. Firstly, let us show that any point in the $\lambda, (1-\lambda)$ mix of the sets C_{I_1} and C_{I_2} satisfies the constraints I_λ. It follows that

$$\lambda C_{I_1} + (1-\lambda)C_{I_2} \subseteq C_{I_\lambda} \tag{3.40}$$

In order to establish the opposite inclusion, we examine the extreme points of the pentagonal regions. It is evident that the rate regions, as defined in equation (3.38), consistently take the shape of a pentagon. In the exceptional scenario where $I_3 = I_1 + I_2$, are in the form of a rectangle. Consequently, the capacity region C_I can also be described as the convex hull formed by five points:

$$(0,0), (I_1, 0), (I_1, I_3 - I_1), (I_3 - I_2, I_2), (0, I_2). \tag{3.41}$$

Let us consider the region defined by I_λ, which also consists of five defining points. Select any one of these points, for instance, $(I_3^{(\lambda)} - I_2^{(\lambda)}, I_2^{(\lambda)})$. This specific point can be expressed as a combination of the points $(I_3^{(1)} - I_2^{(1)}, I_2^{(1)})$ and $(I_3^{(2)} - I_2^{(2)}, I_2^{(2)})$ with weights $(\lambda, 1-\lambda)$. Therefore, it falls within the convex mixture of C_{I_1} and C_{I_2}. Consequently, all extreme points of the pentagon C_{I_λ} lie in the convex hull of C_{I_1} and C_{I_2}, or

$$C_{I_\lambda} \subseteq \lambda C_{I_1} + (1-\lambda)C_{I_2} \tag{3.42}$$

Bringing together the two parts, we establish the theorem. □

In the theorem's proof, we have implicitly relied on the notion that every rate region is delineated by five extreme points (in some cases, some points could be identical). All five points outlined by the I vector fell within the rate region. If the condition $I_3 \leq I_1 + I_2$ is not satisfied, a few points in (3.41) could lie outside the rate region and the proof collapses.

As a direct result of the lemma mentioned above, we obtain the following theorem:

Theorem 3.3.2. *The convex hull of the union of the rate regions defined by individual* I *vectors is equal to the rate region defined by the convex hull of the* I *vectors.*

Theorem 3.3.3. *The achievable rates in a discrete memoryless multiple-access channel are defined by the closure of the set containing all pairs* (R_1, R_2) *that meet the following conditions:*

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2, Q), \\ R_2 &\leq I(X_2; Y|X_1, Q), \\ R_1 + R_2 &\leq I(X_1, X_2; Y|Q) \end{aligned} \tag{3.43}$$

for some choice of the joint distribution $p(q)p(x_1|q)p(x_2|q)p(y|x_1, x_2)$ *with* $|\mathcal{Q}| \leq 4$.

Proof. We will show that each rate pair lying in the region specified in equation (3.43) is achievable, meaning it falls within the convex closure of the rate pairs that fulfill Theorem 3.1.1. Additionally, we will prove that every point within the convex closure of the region described in Theorem 3.1.1 also belongs to the region defined by equation (3.43).

Let us examine a rate point R that satisfy the inequalities stated in equation (3.43) of the theorem. We can rephrase the right-hand side of the initial inequality as follows.

$$I(X_1; Y|X_2, Q) = \sum_{q=1}^{m} p(q) I(X_1; Y|X_2, Q = q) \tag{3.44}$$

$$= \sum_{q=1}^{m} p(q) I(X_1; Y|X_2)_{p_{1_q}, p_{2_q}}, \tag{3.45}$$

where m is the cardinality of the support set of Q.

To simplify our notation, we treat a rate pair as a vector and label a pair that fulfills the inequalities in equation (3.9) for a given input product distribution $p_{1_q}(x_1)p_{2_q}(x_2)$ as R_{p_1,p_2} as R_q. In particular, let $R_q = (R_{1_q}, R_{2_q})$ represent a rate pair that meets these conditions.

$$R_{1_q} < I(X_1; Y|X_2)_{p_{1_q}(x_1)p_{2_q}(x_2)}, \tag{3.46}$$

$$R_{2_q} < I(X_2; Y|X_1)_{p_{1_q}(x_1)p_{2_q}(x_2)}, \tag{3.47}$$

$$R_{1_q} + R_{2_q} < I(X_1, X_2; Y)_{p_{1_q}(x_1)p_{2_q}(x_2)}. \tag{3.48}$$

Subsequently, based on Theorem 3.1.1, $R_q = (R_{1_q}, R_{2_q})$ is achievable. Given that R satisfies the conditions in (3.43) and we can elaborate on the right-hand sides as shown in (3.45), there exists a collection of pairs R_q that adhere to (3.48), such that

$$R = \sum_{q=1}^{m} p(q) R_q. \tag{3.49}$$

As a convex combination of achievable rates is also achievable, this affirms the achievability of R. Consequently, we have proven the achievability of the region outlined in the theorem. A similar argument can be applied to show that any point within the convex closure of the region described in (3.9) can be expressed as a combination of points satisfying (3.48) and thus can be represented in the form (3.43).

The following section provides the proof for the converse. The converse shows that all achievable rate pairs are of the form (3.43), thus confirming it as the capacity region for the multiple-access channel. The limit on the cardinality of the time-sharing random variable Q is a result derived from Caratheodory's theorem on convex sets, as discussed further below. □

Demonstration of the capacity region's convexity in the proof illustrates that achieving a convex combination of rate pairs also results in achievability. We can extend this principle by taking convex combinations of additional points. Do we need to use an arbitrary number of points? Will it expand the capacity region? The subsequent theorem asserts otherwise.

Theorem 3.3.4 (Caratheodory's Theorem)**.** *Every point within the convex closure of a compact set A in a d-dimensional Euclidean space can be expressed as a convex combination of $d+1$ or fewer points from the original set A.*

This theorem permits us to focus on a specific finite convex combination when determining the capacity region. This is a crucial characteristic, as without it, calculating the capacity region as outlined in (3.43) would be infeasible, since we would never know whether using a larger alphabet $\mathcal{Q}$ would increase the region.

Within the domain of the multiple-access channel, the bounds define a connected compact set in three dimensions. As a result, any point within its closure can be expressed as a convex combination of no more than four points. Thus, we can restrict the cardinality of $\mathcal{Q}$ to a maximum of 4 in the provided definition of the capacity region.

3.4 Converse for the multiple-access channel

Until now, we have proved the achievability of the capacity region. In this section, we are going to prove the converse.

Proof. Converse to Theorems 3.1.1 and 3.3.3.
We will show that given any sequence of $((2^{nR_1}, 2^{nR_2}), n)$ codes with $P_e^{(n)} \to 0$, the rates must satisfy

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2, Q), \\ R_2 &\leq I(X_2; Y|X_1, Q), \\ R_1 + R_2 &\leq I(X_1, X_2; Y|Q) \end{aligned} \tag{3.50}$$

considering a random variable Q defined on $\{1, 2, 3, 4\}$ with a specific joint distribution $p(q)p(x_1|q)p(x_2|q)p(y|x_1, x_2)$. Let us fix n and consider the provided code with a block length of n. The joint distribution on $\mathcal{W}_1 \times \mathcal{W}_1 \times \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}^n$ is clearly defined. The randomness arises from the uniform selection of indices $\mathcal{W}_1$ and $\mathcal{W}_2$, along with the inherent randomness induced by the channel. The joint distribution is

$$p(\omega_1, \omega_2, x_1^n, x_2^n, y_n) = \frac{1}{2^{nR_1}} \frac{1}{2^{nR_2}} p(x_1^n|\omega_1)p(x_2^n|\omega_2) \prod_{i=1}^{n} p(y_i|x_{1i}, x_{2i}), \quad (3.51)$$

$p(x_1^n|\omega_1)$ takes a value of either 1 or 0, based on whether $x_1^n = \mathbf{x}_1(\omega_1)$, which is the codeword corresponding to ω_1 or not. Similarly, $p(x_2^n|\omega_2)$ is either 1 or 0, depending on whether $x_2^n = \mathbf{x}_2(\omega_2)$ or not. The subsequent calculations of mutual information are based on this specified distribution.

Due to the way the code is contruction, we can accurately estimate (W_1, W_2) from the received sequence Y^n with a minimal probability of error. Consequently, the conditional entropy of (W_1, W_2) given Y^n should be minimal, as indicated by Fano's inequality,

$$H(W_1, W_2|Y^n) \leq n(R_1 + R_2)P_e^{(n)} + H(P_e^{(n)}) \stackrel{\triangle}{=} n\epsilon_n \quad (3.52)$$

It is clear that $\epsilon_n \to 0$, because $P_e^{(n)} \to 0$. Then we have

$$H(W_1|Y^n) \leq H(W_1, W_2|Y^n) \leq n\epsilon_n, \quad (3.53)$$

$$H(W_2|Y^n) \leq H(W_1, W_2|Y^n) \leq n\epsilon_n. \quad (3.54)$$

We can now bound the rate R_1 as

$$
\begin{aligned}
nR_1 &= && (3.55)\\
&= I(W_1;Y^n) + H(W_1|Y^n) && (3.56)\\
&\overset{(a)}{\leq} I(W_1;Y^n) + n\epsilon_n && (3.57)\\
&\overset{(b)}{\leq} I(X_1^n(W_1);Y^n) + n\epsilon_n && (3.58)\\
&= H(X_1^n(W_1)) - H(X_1^n(W_1)|Y^n) + n\epsilon_n && (3.59)\\
&\overset{(c)}{\leq} H(X_1^n(W_1)|X_2^n(W_2)) - H(X_1^n(W_1)|Y^n, X_2^n(W_2)) + n\epsilon_n && (3.60)\\
&= I(X_1^n(W_1);Y^n|X_2^n(W_2)) + n\epsilon_n && (3.61)\\
&= H(Y^n|X_2^n(W_2)) - H(Y^n|X_1^n(W_1), X_2^n(W_2)) + n\epsilon_n && (3.62)\\
&\overset{(d)}{=} H(Y^n|X_2^n(W_2)) - \sum_{i=1}^{n} H(Y_i|Y^{i-1}, X_1^n(W_1), X_2^n(W_2)) + n\epsilon_n && (3.63)\\
&\overset{(e)}{=} H(Y^n|X_2^n(W_2)) - \sum_{i=1}^{n} H(Y_i|X_{1i}, X_{2i}) + n\epsilon_n && (3.64)\\
&\overset{(f)}{\leq} \sum_{i=1}^{n} H(Y_i|X_2^n(W_2)) - \sum_{i=1}^{n} H(Y_i|X_{1i}, X_{2i}) + n\epsilon_n) && (3.65)\\
&\overset{(g)}{\leq} \sum_{i=1}^{n} H(Y_i|X_{2i}) - \sum_{i=1}^{n} H(Y_i|X_{1i}, X_{2i}) + n\epsilon_n && (3.66)\\
&= \sum_{i=1}^{n} I(X_{1i};Y_i|X_{2i}) + n\epsilon_n, && (3.67)
\end{aligned}
$$

where
(a) can be derived using Fano's inequality.
(b) can be derived using the data-processing inequality.
(c) can be deduced from the independence of W_1 and W_2, resulting in the independence of $X_1^n(W_1)$ and $X_2^n(W_2)$. This leads to $H(X_1^n(W_1)|X_2^n(W_2)) = X_1^n(W_1)$, and $H(X_1^n(W_1)|Y^n, X_2^n(W_2)) \leq H(X_1^n(W_1)|Y^n)$ through conditioning.
(d) follows from the chain rule.
(e) can be deduced from the fact that Y_i depends only on X_{1i} and X_{2i}, by the memoryless property of the channel.
(f) can be derived using the chain rule and removing conditioning.
(g) follows from removing conditioning.

Hence, we have

$$
R_1 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i};Y_i|X_{2i}) + \epsilon_n \tag{3.68}
$$

Similarly, we have

$$R_2 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{2i};Y_i|X_{1i}) + \epsilon_n \tag{3.69}$$

To bound the sum of the rates, we have

$$\begin{aligned}
n(R_1+R_2) &= H(W_1,W_2) &(3.70)\\
&= I(W_1,W_2;Y^n) + I(W_1,W_2|Y^n) &(3.71)\\
&\overset{(a)}{\leq} I(W_1,W_2;Y^n) + n\epsilon_n &(3.72)\\
&\overset{(b)}{\leq} I(X_1^n(W_1),X_2^n(W_2);Y^n) + n\epsilon_n &(3.73)\\
&= H(Y^n) - H(Y^n|X_1^n(W_1),X_2^n(W_2)) + n\epsilon_n &(3.74)\\
&\overset{(c)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y^{i-1},X_1^n(W_1),X_2^n(W_2)) + n\epsilon_n &(3.75)\\
&\overset{(d)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_{1i},X_{2i}) + n\epsilon_n &(3.76)\\
&\overset{(e)}{\leq} \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_{1i},X_{2i}) + n\epsilon_n &(3.77)\\
&= \sum_{i=1}^{n} I(X_{1i},X_{2i};Y_i) + n\epsilon_n &(3.78)
\end{aligned}$$

where
(a) is a result of applying Fano's inequality.
(b) follows from the data-processing inequality.
(c) is a direct outcome of employing the chain rule.
(d) is result from the fact that Y_i depends only on X_{1i} and X_{2i} and is conditionally independent of everything else.
(e) is an outcome of the chain rule and of the removing of conditioning.

So, we have

$$R_1 + R_2 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i},X_{2i};Y_i) + \epsilon_n \tag{3.79}$$

The expressions found in (3.98), (3.99) and (3.79) are the averages of the mutual informations calculated at the empirical distributions in column i of the codebook. We can reformulate these equations using a new variable $\mathcal{Q}$, defined as $\mathcal{Q} = i \in \{1,2,\ldots,n\}$ with a probability of $\frac{1}{n}$.

$$R_1 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i};Y_i|X_{2i}) + \epsilon_n \tag{3.80}$$

$$= \frac{1}{n}\sum_{i=1}^{n} I(X_{1q};Y_q|X_{2q},\mathcal{Q}=i) + n\epsilon_n \tag{3.81}$$

$$= I(X_{1\mathcal{Q}};Y_{\mathcal{Q}}|X_{2\mathcal{Q}},\mathcal{Q}) + n\epsilon_n \tag{3.82}$$

$$= I(X_1;Y|X_2,\mathcal{Q}) + n\epsilon_n \tag{3.83}$$

where $X_1 \triangleq X_{2\mathcal{Q}}$, $X_2 \triangleq X_{1\mathcal{Q}}$ and $Y \triangleq Y_{\mathcal{Q}}$ represents new random variables and their distributions depend on $\mathcal{Q}$ in the same way as distributions of X_{1i}, X_{2i} and Y_i depend on i. We know that W_1 and W_2 are independent, that implies that $X_{1i}(W_1)$ and $X_{2i}(W_2)$ are as well and hence we know that

$$P(X_{1i}(W_1)=x_1, X_{2i}(W_2)=x_2)$$
$$\triangleq P(X_{1\mathcal{Q}}=x_1|\mathcal{Q}=i)P(X_{2\mathcal{Q}}=x_2|\mathcal{Q}=i) \tag{3.84}$$

When we take the limit, $n \to \infty$, $P_e^{(n)} \to 0$, we have the following converse:

$$R_1 \leq I(X_1;Y|X_2,Q),$$
$$R_2 \leq I(X_2;Y|X_1,Q),$$
$$R_1 + R_2 \leq I(X_1,X_2;Y|Q) \tag{3.85}$$

for some choice of joint distribution $p(q)p(x_1|q)p(x_2|q)p(y|x_1,x_2)$. As demonstrated in Section 3.4, the region is unchanged when we limit the size of $\mathcal{Q}$ to 4. This can conclude the proof of the converse. □

We proved the achievability of the region of the Theorem 3.1.1 in the Section 3.2. In Section 3.3 we proved that every point within the region specified by equation (3.50) is achievable. In the converse, we established that the region outlined in (3.50) represents the optimal achievable performance, that was the best we can do, confirming it as the capacity region of the channel. Hence, the region defined in (3.9) for the multiple-access channel cannot be larger than the region in (3.50), and this is the capacity region of the multiple-access channel.

3.5 m-User multiple-access channels

We will now extend the derived result for two senders to m senders, where $m \geq 2$. In this case, the multiple-access channel is shown in Figure 3.9. We transmit distinct indices $\omega_1, \omega_2, \ldots, \omega_m$ over the channel, each corresponding to the senders $1, 2, \ldots, m$, respectively. The codes, rates, and achievability follow the same definitions as in the case of two senders.

Let $S \subseteq \{1, 2, \ldots, m\}$ and let us denote S^c as complement of S. Let $R(S) = \sum_{i\in S} R_i$ and let $X(S) = \{X_i : i \in S\}$. Now we will prove the following theorem.

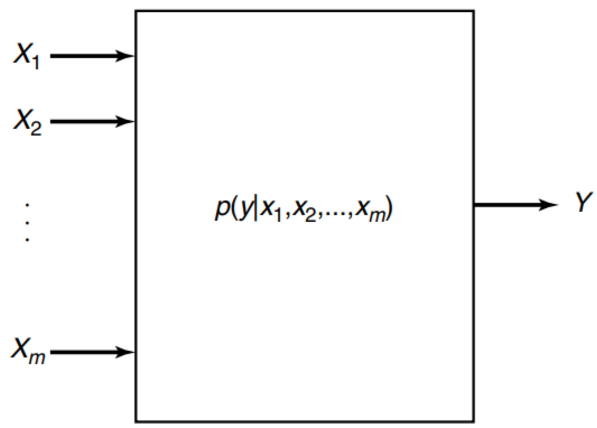

Figure 3.9: m-user multiple-access channel. (taken from [3])

Theorem 3.5.1. *The capacity region of the multiple-access channel with m users is represented by the closure of the convex hull of the rate vectors that meet the following conditions:*

$$R(S) \leq I(X(S); Y|X(S^c)), \quad \forall S \subseteq \{1, 2, \ldots, m\} \tag{3.86}$$

for some product distribution $p_1(x_1)p_2(x_2)\ldots p_m(x_m)$.

Proof. The probability of error in the achievability proof has now $2m-1$ terms and has an equal number of inequalities in the proof of the converse. □

3.6 Gaussian multiple-user channels

Gaussian multi-user channels illustrate crucial aspects of network information theory. Within this section, we will outline the fundamental concepts for defining the capacity regions of Gaussian channels, including multiple-access, broadcast, relay, and two-way channels, without presenting the proofs.

The basic discrete-time additive white Gaussian noise channel with input power P and noise variance N is modeled by:

$$Y_i = X_i + Z_i, \quad i = 1, 2, \ldots, \tag{3.87}$$

where the Z_i are i.i.d. Gaussian random variables with mean 0 and variance N.

The signal $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ has a power constraint:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \leq P. \tag{3.88}$$

The Shannon capacity C is derived by maximizing $I(X;Y)$ across all random variables X satisfying $EX^2 \leq P$ and it is given by:

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N}\right) \tag{3.89}$$

bits per transmission.

3.6.1 Gaussian multiple-access channel with m users

Let us consider m transmitters, each with a power P and let

$$Y = \sum_{i=1}^{n} X_i + Z. \tag{3.90}$$

Let

$$C\left(\frac{P}{N}\right) = \frac{1}{2} \log \left(1 + \frac{P}{N}\right) \tag{3.91}$$

denote as the capacity of a single-user Gaussian channel with signal-to-noise ratio P/N. The achievable rate region for the Gaussian channel is expressed in a simple form through the following equations:

$$R_i < C\left(\frac{P}{N}\right) \tag{3.92}$$

$$R_i + R_j < C\left(\frac{2P}{N}\right) \tag{3.93}$$

$$R_i + R_j + R_k < C\left(\frac{3P}{N}\right) \tag{3.94}$$

$$\vdots$$

$$\sum_{i=1}^{m} < C\left(\frac{mP}{N}\right). \tag{3.95}$$

We can see that all the rates are the same. Inequality (3.95) dominates the others.

In this scenario, we require m sets of codebooks, where the ith codebook has 2^{nR_i} codewords with power P. The transmission process is simple: each independent transmitter selects any codeword from its respective codebook. These vectors are then sent simultaneously by the users. At the receiver, the observed codewords are added together with the Gaussian noise $\mathbf{Z}$. For optimal decoding, the objective is to identify one codeword from each codebook (totaling m codewords) in a way that the vector sum is closest to Y in terms of Euclidean distance. And we can say that if $(R_1, R_2, \ldots, R_m)$ is in the capacity region given above, the probability of error goes to 0 as n tends to infinity.

Remarks: It is fascinating to observe in this scenario that the sum of user rates, denoted by $C(m\frac{P}{N})$, approaches infinity as m increases. Consequently, in a cocktail party, where we have m celebrants of power P and where the presence of ambient noise is N, we can say that intended listener receives an unbounded amount of information as the number of people grows to infinity at the party. It is also interesting to mention that each of the transmitters uses all of the bandwidth all of the time.

3.6.2 Gaussian multiple-access channels

We can now discuss the Gaussian multiple-access channels.

Two senders, denoted as X_1 and X_2 are communicating with a single receiver Y. Let us denote the received signal at time i:

$$Y_i = X_{1i} + X_{2i} + Z_i \tag{3.96}$$

where $\{Z_i\}$ represents a sequence of independent, identically distributed Gaussian random variables with zero mean and variance N. The channel is depicted in Figure 3.10.

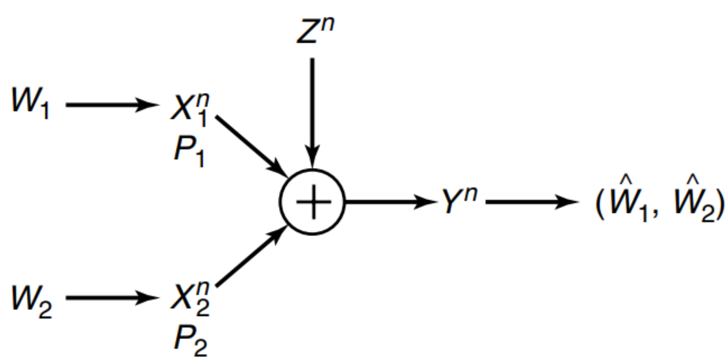

Figure 3.10: Gaussian multiple-access channel. (taken from [3])

We assume a power constraint P_j on sender j, implying that for each sender and all messages, we must satisfy:

$$\frac{1}{n}\sum_{i=1}^{n} x_{ji}^2(\omega_j) \leq P_j, \quad \omega_j \in \left\{1, 2, \ldots, 2^{nR_j}\right\}, j = 1, 2. \tag{3.97}$$

As the proof of achievability of channel capacity was extended from the discrete case to the Gaussian channel, we can further extend the proof from the discrete multiple-access channel to the Gaussian multiple-access channel. The converse can also be similarly extended. We expect the capacity region to be the convex hull of the set of rate pairs that satisfy the following conditions:

$$R_1 \leq I(X_1; Y|X_2), \tag{3.98}$$

$$R_2 \leq I(X_2; Y|X_1), \tag{3.99}$$

$$R_1 + R_2 \leq I(X_1, X_2; Y) \tag{3.100}$$

for some input distribution $f_1(x_1)f_2(x_2)$ that satisfies $EX_1^2 \leq P_1$ and $EX_2^2 \leq P_2$.

Definition 3.6.1. Let X be a random variable with a probability density function f whose support is a set $\mathcal{X}$. The *differential entropy* $h(X)$ or $h(f)$ is defined as

$$h(X) = \mathbf{E}[-\log(f(X))] = -\int_{\mathcal{X}} f(x) \log f(x) dx.$$

Now, we can express mutual information in terms of relative entropy, and as a result we have:

$$\begin{aligned}
I(X_1; Y|X_2) &= h(Y|X_2) - h(Y|X_1, X_2) && (3.101)\\
&= h(X_1 + X_2 + Z|X_2) - h(X_1 + X_2 + Z|X_1, X_2) && (3.102)\\
&= h(X_1 + Z|X_2) - h(Z|X_1, X_2) && (3.103)\\
&= h(X_1 + Z|X_2) - h(Z) && (3.104)\\
&= h(X_1 + Z) - h(Z) && (3.105)\\
&= h(X_1 + Z) - \frac{1}{2}\log(2\pi e N) && (3.106)\\
&\leq \frac{1}{2}\log(2\pi e(P_1 + N)) - \frac{1}{2}\log(2\pi e N) && (3.107)\\
&= \frac{1}{2}\log\left(1 + \frac{P_1}{N}\right), && (3.108)
\end{aligned}$$

equation (3.104), follows from the independence of Z from both X_1 and X_2. Equation (3.105) results from the independence of X_1 and X_2. Lastly, equation (3.108) follows from the fact that the normal distribution maximizes entropy for a given second moment. Consequently, the distribution that achieves the maximum is where $X_1 \sim \mathcal{N}(0, P_1)$, $X_2 \sim \mathcal{N}(0, P_2)$ and X_1 and X_2 are independent. This distribution simultaneously maximizes the mutual information bounds given in equations (3.98)–(3.100).

Definition 3.6.2. Let us define the channel capacity function:

$$C(x) \triangleq \frac{1}{2}\log(1 + x), \tag{3.109}$$

representing the channel capacity of a Gaussian white-noise channel with a signal-to-noise ratio x (as shown in Figure 3.11). We then express the constraint on R_1 as:

$$R_1 \leq C\left(\frac{P_1}{N}\right) \tag{3.110}$$

And similarly,

$$R_2 \leq C\left(\frac{P_2}{N}\right) \tag{3.111}$$

and

$$R_1 + R_2 \leq C\left(\frac{P_1 + P_2}{N}\right) \tag{3.112}$$

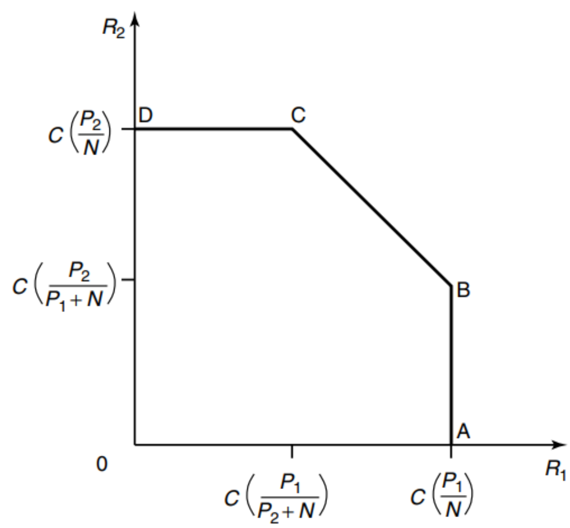

Figure 3.11: Gaussian multiple-access channel capacity region. (taken from [3])

These upper bounds are achieved when $X_1 \sim \mathcal{N}(0, P_1)$, $X_2 \sim \mathcal{N}(0, P_2)$, defining the capacity region. What is astonishing about these inequalities is that the sum of the rates can reach up to $C\left(\frac{P_1+P_2}{N}\right)$ matching the rate achieved by a single transmitter sending with a power equal to the sum of the powers.

The explanation of the corner points closely mirrors the interpretation of achievable rate pairs in a discrete multiple-access channel for a fixed input distribution. In the context of the Gaussian channel, decoding can be viewed as a two-step procedure. In the initial stage, the receiver decodes the second sender, considering the first sender as part of the noise. This decoding process is likely to have a low probability of error if $R_2 < C\left(\frac{P_2}{P_1+N}\right)$.

Once the second sender has been decoded successfully, it can be removed from consideration, allowing for successful decoding of the first sender if $R_1 <$

$C\left(\frac{P_1}{N}\right)$. Thus, this argument demonstrates that the rate pairs at the corner points of the capacity region can be realized through single−user sender operations. This method, known as *onion−peeling* can be applied and extended to any number of users. When we extend this to m senders with equal power, the total rate becomes $C\left(\frac{mP}{N}\right)$, which tends to ∞ as $m \to \infty$. On average, the rate per sender, $\frac{1}{m}C\left(\frac{mP}{N}\right)$, goes to 0. Consequently, with large number of senders, that are causing lot of interference, we can still transmit a total amount of information that is arbitrarily large despite the rate per individual sender goes to 0. The capacity region described above corresponds to *code-division multiple access*, or shortly CDMA. Here we say that separate codes are used for the different senders and the receiver decodes them one by one. However, in numerous real-world scenarios, simpler techniques like frequency-division multiplexing or time-division multiplexing are commonly used.

In frequency-division multiplexing, the achievable rates are contingent on the bandwidth assigned to each sender. Let us consider a scenario involving two senders, with powers P_1 and P_2 respectively, using nonintersecting frequency bands with bandwidths W_1 and W_2 where $W_1 + W_2 = W$ (representing the overall bandwidth). Utilizing the capacity formula for a single-user bandlimited channel, we can derive the following achievable rate pair:

$$R_1 = W_1 \log\left(1 + \frac{P_1}{NW_1}\right), \tag{3.113}$$

$$R_2 = W_2 \log\left(1 + \frac{P_2}{NW_2}\right). \tag{3.114}$$

By adjusting the values of W_1 and W_2, we generate the curve shown in Figure 3.12. This curve touches the boundary of the capacity region at a specific point, where the bandwidth is allocated to each channel in proportion to the power in that channel. In the context of *time-division multiple access*, shortly TDMA, time is divided into distinct slots. Within each slot, a designated user transmits while all other users remain inactive, quiet. If there are two users, both operating at a power level of P, the rate at which each transmits when the other is inactive is $C\left(\frac{P}{N}\right)$. Now if time is divided into equal-length slots, and every odd slot is assigned to user 1, while every even slot is assigned to user 2, the average rate achieved by each user is $\frac{1}{2}C\left(\frac{P}{N}\right)$. This system is called *naive time-division multiple access*, shortly TDMA.

Improved performance is achievable by recognizing that user 1 transmits only half the time. This allows user 1 to utilize twice the power during transmissions while adhering to the same average power constraint. With this modification, each user can transmit information at a rate of $\frac{1}{2}C\left(\frac{2P}{N}\right)$. By adjusting slot lengths for each sender and their respective instantaneous power during these slots, we can attain the same capacity region as FDMA with different bandwidth allocations.

As shown in Figure 3.12, the capacity region is typically larger compared to what can be achieved through time- or frequency-division multiplexing. It's important to observe that the multiple-access capacity region, as derived earlier,

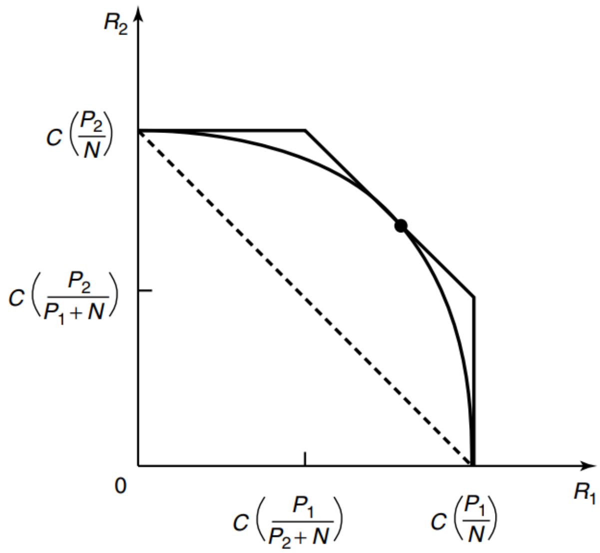

Figure 3.12: Gaussian multiple-access channel capacity with FDMA and TDMA. (taken from [3])

is achieved by use of a common decoder for all the senders. However, it is also feasible to achieve the capacity region through "onion-peeling" a technique that eliminates the need for a common decoder and instead, uses a sequence of single-user codes. CDMA fully achieves the capacity region and allows new users to be added easily without changing the codes of the current users. But on the other hand, when we speak about TDMA and FDMA systems, they are typically designed for a predefined number of users. It is possible that in such systems either some slots remain unoccupied (if the actual number of users is less than the number of slots) or some users might be excluded (if the number of users exceeds the available slots). In numerous real-world systems, design simplicity holds significant weight, and the improvement in capacity resulting from the multiple-access ideas may not justify the increased complexity.

In a Gaussian multiple-access system featuring m sources with respective powers $P_1, P_2, \ldots, P_m$ and ambient noise power N, we can express the equivalent of Gauss's law for any set S as follows:

$$\sum_{i \in S} R_i = \text{total rate of information flow from S} \tag{3.115}$$

$$\leq C\left(\frac{\sum_{i \in S} P_i}{N}\right). \tag{3.116}$$

Chapter 4

Gaussian vector channels

Gaussian vector channels represent models utilized in multiple-input multiple-output (MIMO) wireless communication setups, enabling both transmitters and receivers to employ more than a single antenna. The use of multiple antennas confers several advantages in a wireless multipath environment. Within this chapter, our focus is specifically on exploring and establishing the capacity of the Gaussian vector point-to-point channel. Subsequently, we show the capacity region of the Gaussian vector multiple access channel. Furthermore, we establish that the sum-capacity is achieved through iterative water-filling techniques. The rest of this chapter is dedicated to examining Gaussian vector broadcast channel.

4.1 Gaussian vector point-to-point channel

Consider the point-to-point communication system illustrated in Figure 4.1. The sender aims to reliably transmit a message M to the receiver over a MIMO

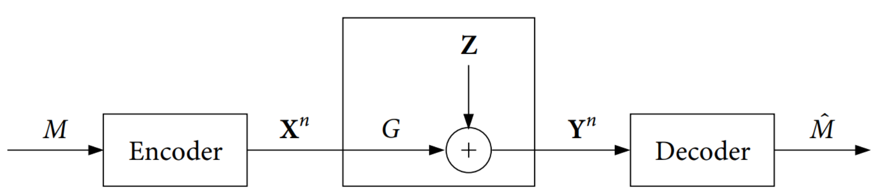

Figure 4.1: MIMO point-to-point communication system. (taken from [4])

communication channel.

We represent the MIMO communication channel as a Gaussian vector channel, where the output Y of the channel corresponding to the input X is

$$\mathbf{Y} = G\mathbf{X} + \mathbf{Z}.$$

In this context, $\mathbf{Y}$ is an r-dimensional vector, $\mathbf{X}$ is a t-dimensional vector, $\mathbf{Z} \sim N(0, K_{\mathbf{Z}})$, where $K_{\mathbf{Z}} \succ 0$ and G is an $r \times t$ constant channel gain matrix. The elements G_{jk} represent the gain of the channel from transmitter antenna k to receiver antenna j. The channel is discrete-time, and the noise vector process $\{\mathbf{Z}(i)\}$ is independent and identically distributed (i.i.d.) with $\mathbf{Z}(i) \sim N(0, K_{\mathbf{Z}})$ for every transmission $i \in [1:n]$. We assume an average transmission power constraint P on each codeword $\mathbf{x}^n(m) = (\mathbf{x}(m,1), \ldots, \mathbf{x}(m,n))$, i.e.,

$$\sum_{i=1}^{n} \mathbf{x}^T(m,i)\mathbf{x}(m,i) \leq nP, \quad m \in [1:2^{nR}].$$

Before we prove following theorem, we will state few points that will be useful later for proof.

Lemma 4.1.1 (Maximum Differential Entropy Lemma)**.** *Let $\boldsymbol{X} \sim f(x^n)$ be a random vector with covariance matrix $K_X = E[(\boldsymbol{X} - E(\boldsymbol{X}))(\boldsymbol{X} - E(\boldsymbol{X}))^T] \succ 0$. Then*

$$h(\boldsymbol{X}) \leq \frac{1}{2}\log((2\pi e)^n |K_{\boldsymbol{X}}|) \leq \frac{1}{2}\log((2\pi e)^n |\boldsymbol{E}(\boldsymbol{X}\boldsymbol{X}^T)|),$$

where $\boldsymbol{E}(\boldsymbol{X}\boldsymbol{X}^T)$ represents the correlation matrix of X^n. The first inequality holds with equality if and only if X is Gaussian and the second inequality holds with equality if and only if $\boldsymbol{E}(\boldsymbol{X}) = 0$. In a broader context, if $(\boldsymbol{X}, \boldsymbol{Y}) = (X^n, Y^k) \sim f(x^n, y^k)$ is a pair of random vectors $K_{\boldsymbol{X}|\boldsymbol{Y}} = E[(\boldsymbol{X} - E(\boldsymbol{X}|\boldsymbol{Y}))(\boldsymbol{X} - E(\boldsymbol{X}|\boldsymbol{Y}))^T]$ is the covariance matrix of the error vector of the minimum mean squared error (MMSE) estimate of X given Y, then

$$h(\boldsymbol{X}|\boldsymbol{Y}) \leq \frac{1}{2}\log((2\pi e)^n |K_{\boldsymbol{X}|\boldsymbol{Y}}|).$$

If $(\boldsymbol{X}, \boldsymbol{Y})$ is jointly Gaussian, then equality holds.

Theorem 4.1.1. *The capacity of the Gaussian vector channel is*

$$C = \max_{K_{\boldsymbol{X}} \succeq 0 : tr(K_X) \leq P} \frac{1}{2}\log|GK_{\boldsymbol{X}}G^T + I_r|.$$

Proof. Let us first note that the capacity with power constraint is upper bounded as

$$\begin{aligned} C &\leq \sup_{F(\mathbf{x}):E(\mathbf{X}^T\mathbf{X}) \leq P} I(\mathbf{X};\mathbf{Y}) \\ &= \sup_{F(\mathbf{x}):E(\mathbf{X}^T\mathbf{X}) \leq P} h(\mathbf{Y}) - h(\mathbf{Z}) \\ &= \max_{K_{\mathbf{X}} \succeq 0 : tr(K_X) \leq P} \frac{1}{2}\log|GK_{\mathbf{X}}G^T + I_r|, \end{aligned}$$

where the last step follows by the Lemma 4.1. The supremum is reached when $\mathbf{X}$ is a Gaussian variable with zero mean and covariance matrix $K_{\mathbf{X}}$. With this specific choice of $\mathbf{X}$, the output $\mathbf{Y}$ is also Gaussian and its covariance matrix is given by $GK_{\mathbf{X}}G^T + I_r$. □

The optimal covariance matrix $K_{\mathbf{X}}^*$ can be more explicitly defined. Assume that G has a rank d and is decomposed into singular values as $G = \Phi\Gamma\Psi^T$, where $\Gamma = \text{diag}(\gamma_1, \gamma_2, \dots, \gamma_d)$. Then

$$\begin{aligned}
C &= \max_{K_{\mathbf{X}} \succeq 0: tr(K_X) \leq P} \frac{1}{2} \log |GK_{\mathbf{X}}G^T + I_r| \\
&= \max_{K_{\mathbf{X}} \succeq 0: tr(K_X) \leq P} \frac{1}{2} \log |\Phi\Gamma\Psi^T K_{\mathbf{X}} \Psi\Gamma\Phi^T + I_r| \\
&\overset{(a)}{=} \max_{K_{\mathbf{X}} \succeq 0: tr(K_X) \leq P} \frac{1}{2} \log |\Phi^T\Phi\Gamma\Psi^T K_{\mathbf{X}} \Psi\Gamma + I_d| \\
&\overset{(b)}{=} \max_{K_{\mathbf{X}} \succeq 0: tr(K_X) \leq P} \frac{1}{2} \log |\Gamma\Psi^T K_{\mathbf{X}} \Psi\Gamma + I_d| \\
&\overset{(c)}{=} \max_{\tilde{K}_{\mathbf{X}} \succeq 0: tr(\tilde{K}_X) \leq P} \frac{1}{2} \log |\Gamma \tilde{K}_{\mathbf{X}} \Gamma + I_d|,
\end{aligned}$$

where *(a)* follows since $|AB+I| = |BA+I|$ with $A = \Phi\Gamma\Psi^T K_{\mathbf{X}} \Psi\Gamma$ and $B = \Phi^T$, *(b)* follows since $\Phi^T\Phi = I_d$ (using the definition of singular value decomposition in Notation) and *(c)* follows since the maximization problem is equivalent to that in *(b)* through the transformations $\tilde{K}_{\mathbf{X}} = \Psi^T K_{\mathbf{X}} \Psi$ and $K_{\mathbf{X}} = \Psi \tilde{K}_{\mathbf{X}} \Psi^T$. Utilizing Hadamard's inequality, the optimal $\tilde{K}_{\mathbf{X}}^*$ is represented by a diagonal matrix $\text{diag}(P_1, P_2, \dots, P_d)$, satisfying the water-filling condition, i.e.,

$$P_j = \left[\lambda - \frac{1}{\gamma_i^2}\right]^+,$$

where λ is chosen such that $\sum_{j=1}^{d} P_j = P$. Through the transformation between $K_{\mathbf{X}}$ and $\tilde{K}_{\mathbf{X}}$, the optimal $K_{\mathbf{X}}^*$ is determined as $K_{\mathbf{X}}^* = \Psi \tilde{K}_{\mathbf{X}}^* \Psi^T$. Therefore, the transmitter is advised to align its signal *direction* with the singular vectors of the effective channel and allocate an appropriate amount of in each direction to satisfy the *water-filling* principle over the singular values.

4.2 Gaussian vector multiple access channel

Consider the MIMO multiple access communication system illustrated in 4.2, where each sender aims to transmit a separate message to the receiver. This scenario operates under the assumption of a Gaussian vector multiple access channel (GV-MAC) model:

$$\mathbf{Y} = G_1\mathbf{X}_1 + G_2\mathbf{X}_2 + \mathbf{Z}$$

where $\mathbf{Y}$ is an r-dimensional output vector, $\mathbf{X}_1$ and $\mathbf{X}_2$ are t-dimensional input vectors, G_1 and G_2 are $r \times t$ channel gain matrices, $\mathbf{Z} \sim N(0, K_Z)$ is an r-dimensional noise vector and $K_Z \succ 0$ is an r-dimensional noise vector. We assume without loss of generality that $K_Z = I_r$. Let us further assume average power constraint P on each of X_1 and X_2, i.e.,

$$\sum_{i=1}^{n} \mathbf{x}_j^T(m_j,i)\mathbf{x}_j(m_j,i) \le nP, \quad m_j \in \left[1:2^{nR_j}\right], \; j=1,2.$$

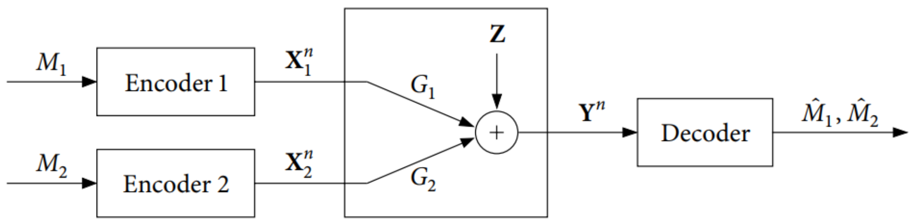

Figure 4.2: MIMO multiple access communication system. (taken from [4])

Theorem 4.2.1. *The capacity region of the GV-MAC is the set of rate pairs* (R_1, R_2) *such that*

$$\begin{aligned} R_1 &\le \frac{1}{2}\log\left|G_1K_1G_1^T + I_r\right|, \\ R_2 &\le \frac{1}{2}\log\left|G_2K_2G_2^T + I_r\right|, \\ R_1 + R_2 &\le \frac{1}{2}\log\left|G_1K_1G_1^T + G_2K_2G_2^T + I_r\right| \end{aligned}$$

for some $K_1, K_2 \succeq 0$ *with* $tr(K_j) \le P, \; j=1,2.$

4.3 GV-MAC with more than two senders

The capacity region of the GV-MAC can be extended to any number of senders. Let us take a look at the model of the GV-MAC involving k senders,

$$\mathbf{Y} = \sum_{j=1}^{k} G_j\mathbf{X}_j + \mathbf{Z},$$

where $\mathbf{Z} \sim N(0, K_Z)$ is the noise vector. Given an average power constraint P on each X_j, it can be shown that the capacity region is the set of rate tuples $(R_1, \ldots, R_k)$ where

$$\sum_{j\in\mathcal{J}} R_j \le \frac{1}{2}\log\left|\sum_{j\in\mathcal{J}} G_jK_jG_j^T + I_r\right|, \quad \mathcal{J} \subseteq [1:k],$$

for some $K_1, \ldots, K_k \succeq 0$ with $tr(K_j) \le P, \; j \in [1:k]$.

4.4 Gaussian vector broadcast channel

Examining the MIMO broadcast communication system illustrated in Figure 4.3, the sender aims to transmit a common message M_0 to the two receivers and a private message M_j to receiver $j = 1, 2$. The channel is modeled by a Gaussian vector broadcast channel (GV-BC)

$$\mathbf{Y}_1 = G_1\mathbf{X} + \mathbf{Z}_1,$$
$$\mathbf{Y}_2 = G_2\mathbf{X} + \mathbf{Z}_2,$$

where G_1, G_2 are $r \times t$ channel gain matrices and $\mathbf{Z}_1 \sim N(0, I_r)$ and $\mathbf{Z}_2 \sim N(0, I_r)$. Suppose the average transmission power constraint is given by:

$$\sum_{i=1}^{n} \mathbf{x}^T(m_0, m_1, m_2, i)\mathbf{x}(m_0, m_1, m_2, i) \leq nP$$

for $(m_0, m_1, m_2) \in [1 : 2^{nR_0}] \times [1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$.

Unlike the scalar Gaussian BC, it is important to note that the Gaussian vector BC is not generally degraded, and the capacity region is only known in special cases.

- If $t = r$ and G_1 and G_2 are diagonal, then the channel is a product of Gaussian BCs and the capacity region is known.
- If $M_0 = \emptyset$, then the (private-message) capacity region is known.
- If $M_1 = \emptyset$ (or $M_2 = \emptyset$), then the (degraded message sets) capacity region is known.

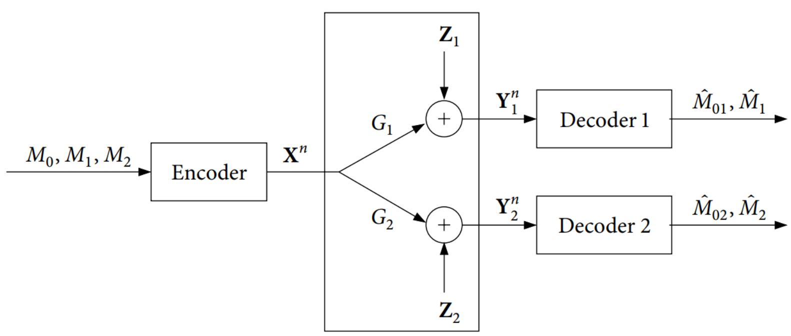

Figure 4.3: MIMO broadcast communication system. (taken from [4])

Chapter 5

Wireless fading channels

As the channel gain information is typically available at each receiver (obtained through training sequences) and potentially available at each sender (via feedback from receivers), wireless fading channels can be characterized as channels with a random state. In this context, the state refers to the channel gain information, which is available at the decoders and either fully or partially available at the encoders. The definition of capacity for wireless fading channels depends on factors such as the fading model and coding delay, and it may or may not be well-defined. Furthermore, even when capacity is well-defined, it may not serve as an accurate performance measure in practical scenarios. This can be attributed to its overly pessimistic nature or the necessity for extensive coding delays to achieve it.

We present various coding strategies considering both fast and slow fading assumptions when the channel gain is available only at the decoder and when the channel gain is available both at the encoder and the decoder, compound channel coding, outage capacity approach, broadcast channel approach, adaptive coding and adaptive coding with power control.

5.1 Gaussian fading model

Let us consider the Gaussian fading channel

$$Y_i = G_i X_i + Z_i, \quad i \in [1:n],$$

where G_i represents a channel gain process simulating fading in wireless communication, while Z_i denotes a WGN(1) process that is independent of G_i.

In practical scenarios, the channel gain typically exhibits variations over a significantly longer time scale than the symbol transmission time. This motivates the simplified *block fading model* illustrated in Figure 5.1. In this model, the gain G_i is presumed to remain constant within each coherence time interval $[(l-1)k+1 : lk]$ of length k, where $l = 1, 2, \ldots$. The block gain process

$\{\bar{G}_l\}_{l=1}^{\infty} = \{G_{lk}\}_{l=1}^{\infty}$ is considered stationary ergodic. Under this model, we explore two coding paradigms.

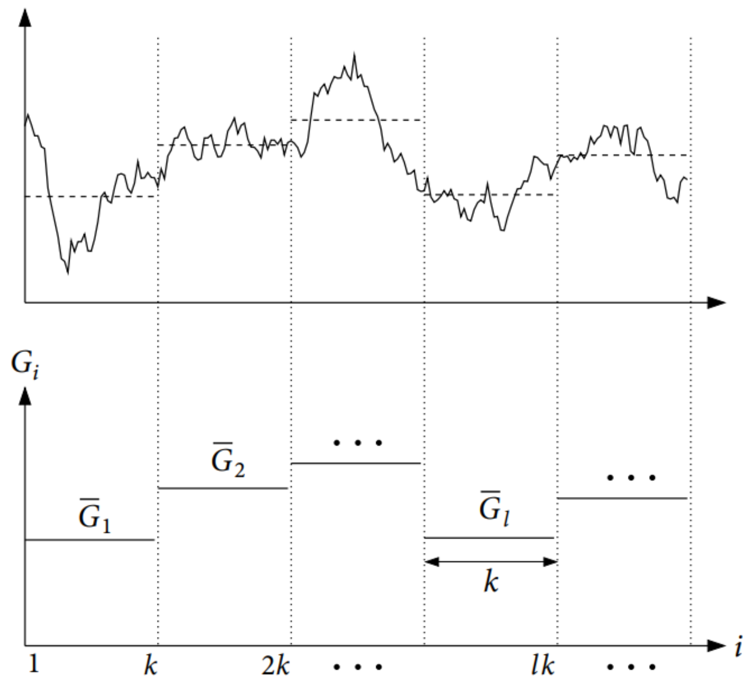

Figure 5.1: Wireless channel fading process and its block fading model. (taken from [4])

In the context of ***fast fading***, the code block length extends across numerous coherence time intervals, rendering the channel ergodic with a clearly defined Shannon capacity, often denoted as the *ergodic capacity*. Nevertheless, coding over a substantial number of coherence time intervals introduces significant delays.

In the context of ***slow fading***, the code block length aligns with the coherence time interval, resulting in a *non-ergodic channel* without Shannon capacity in general. We explore alternative coding approaches and associated performance metrics for this case.

In the upcoming two sections, we explore coding under fast and slow fading with channel gain availability only at the decoder or at both the encoder and the decoder. In cases where the channel gain is available only at the decoder, we assume an average power constraint P on X, denoted as $\sum_{i=1}^{n} x_i^2(m) \leq nP, m \in [1:2^{nR}]$. On the other hand, when the channel gain is available at both the encoder and the decoder, we assume an expected average power constraint P

on X, represented as

$$\sum_{i=1}^{n} E(x_i^2(m, G_i)) \leq nP, \quad m \in [1 : 2^{nR}].$$

5.2 Coding under fast fading

In fast fading scenarios, we code over many coherence time intervals, i.e. $n \gg k$. In this context, the block gain process $\{\bar{G}_l\}$ is stationary ergodic, for instance, it could be an i.i.d. process.
When the channel gain is available **only at the decoder**, we show that the ergodic capacity for this case is

$$C_{GI-D} = \mathbf{E}_G[\mathbf{C}(G^2 P)]$$

For a DMC with stationary ergodic state $p(y|x, s)$, the capacity when the state information is available at the decoder is

$$C_{SI-D} = \max_{p(x)} I(X; Y|S).$$

This outcome can be easily generalized to the Gaussian fading channel with a stationary ergodic block gain process $\{\bar{G}_l\}$, characterized by the marginal distribution $F_G(g_l)$ and adhering to a power constraint P, to obtain the channel capacity.

$$\begin{aligned} C_{SI-D}(P) &= \sup_{F(x):E(X^2)\leq P} I(X; Y|G) \\ &= \sup_{F(x):E(X^2)\leq P} (h(Y|G) - h(Y|G, X)) \\ &= \sup_{F(x):E(X^2)\leq P} h(GX + Z|G) - h(Z) \\ &= \mathbf{E}_G[\mathbf{C}(G^2 P)], \end{aligned}$$

where the supremum is attained by $X \sim N(0, P)$. In the context of fast fading, the Gaussian fading channel can be decomposed in time into k parallel Gaussian fading channels. The first channel corresponds to transmission times $1, k + 1, 2k + 1, \ldots$, while the second channel corresponds to transmission times $2, k + 2, \ldots$, and so on. All these channels share the same stationary ergodic channel gain process and average power constraint P. As a result, $C_{GI-D} \geq C_{SI-D}(P)$. The converse of this statement can be easily proven using standard arguments.
When the channel gain is **available both at the encoder and the decoder**,

the ergodic capacity under these conditions is given by:

$$\begin{aligned} C_{GI-ED} &= \max_{F(x|g):E(X^2)\le P} I(X;Y|G) \\ &= \max_{F(x|g):E(X^2)\le P} (h(Y|G) - h(Y|G,X)) \\ &= \max_{F(x|g):E(X^2)\le P} h(GX+Z|G) - h(Z) \\ &\overset{\text{(a)}}{=} \max_{\phi(g):E(\phi(G))\le P} \mathbf{E}_G[\mathbf{C}(G^2\phi(G))], \end{aligned}$$

where $F(x|g)$ represents the conditional cdf of X given $\{G = g\}$, and the expression (a) holds because the maximum is attained by $X|\{G = g\} \sim N(0, \phi(g))$.

5.3 Coding under slow fading

In the scenario of slow fading, we code over a single coherence time interval (i.e., $n = k$), and notion of channel capacity is not universally well-defined. As previously, we explore cases where the channel gain is available only at the decoder and scenarios where it is available at both the encoder and the decoder.

5.3.1 Channel gain available only at the decoder

In scenarios where the encoder does not know the gain, various coding options are available.

In the **compound channel approach**, we code against the worst channel to guarantee reliable communication. The (Shannon) capacity under this coding approach, following a straightforward extension of the capacity of the compound channel to the Gaussian case, can be expressed as:

$$C_{CC} = \inf_{g\in\mathcal{G}} \mathbf{C}(g^2P).$$

When fading results in extremely low channel gain, the compound channel approach, while effective, becomes less feasible. Therefore, we consider alternative coding approaches that are more practical in such situations.

Outage capacity approach: In this strategy, we transmit at a rate higher than the compound channel capacity (C_{CC}) and accept some information loss when the channel gain falls too low for message recovery. If the probability of such an *outage* event is low, we can achieve reliable communication most of the time. To be specific, if we can tolerate an outage probability p_{out}, representing an average loss of a fraction p_{out} of messages, then we can communicate at any rate below the outage capacity

$$C_{\text{out}} = \max_{g:P\{G\le g\}\le p_{\text{out}}} \mathbf{C}(g^2P).$$

Broadcast channel approach: For simplicity, let us consider two fading states g_1 and g_2 with $g_1 > g_2$. We view the channel as a Gaussian Broadcast Channel (BC) with gains g_1 and g_2, using superposition coding to transmit a common message to both receivers at a rate $\tilde{R}_0 < \mathbf{C}(g_2^2\bar{\alpha}P/(1+\alpha g_2^2 P))$, where $\alpha \in [0,1]$, and a private message to the stronger receiver at a rate $\tilde{R}_1 < \mathbf{C}(g_1^2\alpha P)$. If the gain is g_2, the receiver of the fading channel can recover the common message at a rate $R_2 = \tilde{R}_0$, and if the gain is g_1, it can recover both messages at a total rate $R_1 = \tilde{R}_0 + \tilde{R}_1$. Assuming $\mathbf{P}\{G = g_1\} = p$ and $\mathbf{P}\{G = g_2\} = \bar{p}$ we can calculate the *broadcast capacity* as:

$$C_{BC} = \max_{\alpha\in[0,1]} \left(p\mathbf{C}(g_1^2\alpha P) + \mathbf{C}\left(\frac{g_2^2\bar{\alpha}P}{1+\alpha g_2^2 P} \right) \right).$$

This strategy is most effective when transmitting multimedia content (such as video or music) over a fading channel, using successive refinement. When the channel gain is low, the receiver retrieves only the low-fidelity representation of the source. Conversely, when the gain is high, it can recover the refinement, obtaining the high-fidelity description as well.

5.3.2 Channel gain available at both the encoder and the decoder

Compound channel approach: In cases where the channel gain is available at the encoder, the compound channel capacity, denoted as C_{CC-E}, is determined by

$$C_{CC-E} = \inf_{g\in\mathcal{G}} \mathbf{C}(g^2 P) = C_{CC}.$$

Consequently, the capacity remains unchanged compared to scenarios where the encoder does not know of the state.

Adaptive coding: Rather than communicating at the capacity of the channel with the worst gain, we adjust the transmission rate according to the channel gain. We communicate at the maximum rate $C_g = \mathbf{C}(g^2P)$ when the gain is g. The *adaptive capacity* is defined as:

$$C_A = \mathbf{E}_G[\mathbf{C}(G^2P)].$$

It's important to note that this is identical to the ergodic capacity when the channel gain is available only at the decoder. However, the adaptive capacity is convenient performance metric and does not represent a capacity in the Shannon sense.

Adaptive coding with power control: As the encoder possesses knowledge of the channel gain, it can adapt both the power and transmission rate. In this scenario, we define the *power-control adaptive capacity* as

$$C_{PA} = \max_{\phi(g): E_G(\phi(G)))\leq P} \mathbf{E}_G[\mathbf{C}(G^2\phi(G))],$$

where the maximum is achieved through water-filling power allocation adhering to the constraint $\mathbf{E}_G(\phi(G)) \leq P$. It is worth noting that the power-control adaptive capacity is equivalent to the ergodic capacity when the channel gain is available at both the encoder and the decoder. Once again, C_{PA} does not represent a capacity in the Shannon sense.

Let us compare the performance of the above coding schemes in the following example. Consider two fading states, g_1 and g_2, with $g_1 > g_2$, and $P\{G = g_1\} = p$. In Figure 5.2, we evaluate the performance metrics C_{CC}, C_{out}, C_{BC}, C_A and C_{PA} for various values of $p \in [0, 1]$. The broadcast channel approach is effective when the better channel occurs more often ($p \approx 1$), while power control is particularly effective when the channel varies frequently ($p \approx \frac{1}{2}$).

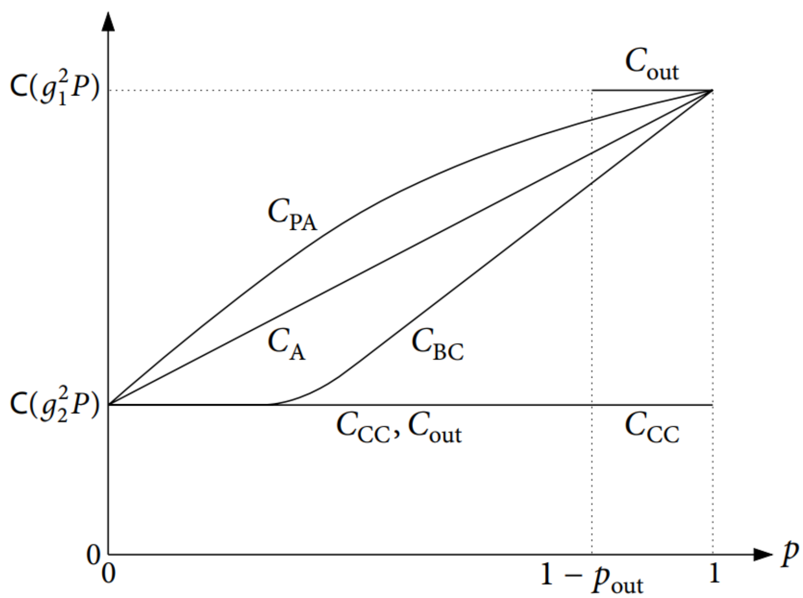

Figure 5.2: Comparison of performance metrics. (taken from [4])

Chapter 6

Conclusion

We analyzed multiple-access channel, properties and illustrations of several different examples. It has been established the achievability of the rate region and later on we reconfigured the capacity region of the multiple access channel so that we are able to incorporate the process of forming the convex hull. That brought us introducing new random variable Q. After the achievability of the capacity region has been proved, converse has been proved as well. It has been proved that every point within the region specified by equation $R_1 + R_2 \leq I(X_1, X_2; Y|Q)$ is achievable and in the converse and it has been established that the region outlined in the same equation represents the optimal achievable performance. We concluded that the region defined in the following way $R_1 + R_2 < I(X_1, X_2; Y)$ for the multiple-access channel cannot be larger than the region defined as $R_1 + R_2 \leq I(X_1, X_2; Y|Q)$, so this is the capacity region of the multiple-access channel. After that, it has been shown what Gaussian multiple-access channel is and explained what CDMA, TDMA and FDMA systems are and which role they have. We talked about Gaussian vector point-to-point channel and capacity of the Gaussian vector channel. Gaussian vector multiple access channel has been mentioned, where each sender aims to transmit a separate message to the receiver, model of the GV-MAC involving k senders and at the end we introduced Gaussian vector broadcast channel. At the end, in the last chapter, we analyzed Gaussian fading model in the context of fast fading and coding under slow fading, where it has been explained what is happening in the case when the channel gain available only at the decoder and the channel gain available both at the encoder and the decoder.

Bibliography

[1] R. Ahlswede, "Multi-way communication channels", In *Proc. 2nd Int. Symp. Inf. Theory* (Tsahkadsor, Armenian S.S.R.), pp. 23-52, Hungarian Academy of Sciences, Budapest, 1971.

[2] T. M. Cover and C. S. K. Leung, "An achievable rate region for the multiple access channel with feedback", *IEEE Trans. Inf. Theory*, vol. 27, no. 3, pp. 292-298, 1981.

[3] Thomas M. Cover, Joy A. Thomas, *Information Theory and Networks*, 2nd ed., John Wiley & Sons, Inc., 2006.

[4] Abbas El Gamal, Young-Han Kim, *Network Information Theory*, Cambridge University Press, 2011.

[5] T. Gaarder and J. K. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback", *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 100-102, 1975.

[6] L. Li, N. Jindal, and A. Goldsmith, "Outage capacities and optimal power allocation for fading multiple-access channels", *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1326-1347, 2005.

[7] H. Liao, *Multiple Access Channels*, Ph.D. thesis, Department of Electrical Engineering, University of Hawaii, Honolulu, 1972.

[8] P. H. Longstaff, *Networked Industries: Patterns in Development, Operation, and Regulation*, Syracuse University, 2000.

[9] L. H. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback", *IEEE Trans. Inf. Theory*, vol. 30, no. 4, pp. 623-629, 1984.

[10] Renato Renner, Ueli Maurer, "About the mutual (conditional) information", Department of Computer Science Swiss Federal Institute of Technology (ETH Zurich) CH-8092 Zurich, Switzerland, 2002.

[11] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels I: Polymatroid structure, optimal resource allocation and throughput capacities", *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796-2815, 1998.

[12] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels II: Delay limited capacities", *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2816- 2831, 1998.

[13] F. M. J. Willems, "The feedback capacity of a class of discrete memoryless multiple access channels", *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 93-95, 1982.

Biography

Tatjana Naumov is born on 12th of March 1996 in Sombor. She attended elementary school "Dositej Obradović" and grammar school "Veljko Petrović" in Sombor. She received her Bachelor degree in Applied Mathematics, module Financial Mathematics, in 2020 at Faculty of Sciences, University of Novi Sad. She continued her Master studies and enrolled in Data Science program at the same faculty. Tatjana began her career as a Financial Software Consultant at Synechron company in Novi Sad, where she applies her academic knowledge.

UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj:
RBR

Identifikacioni broj:
IBR

Tip dokumentacije: Monografska dokumentacija
BF

Tip zapisa: Tekstualni štampani materijal
TZ

Vrsta rada: Master rad
VR

Autor: Tatjana Naumov
AU

Mentor: dr Mladen Kovačević
MN

Naslov rada: Kapacitet kanala sa višestrukim pristupom
NR

Jezik publikacije: Engleski
JP

Jezik izvoda: Engleski
JI

Zemlja publikovanja: Republika Srbija
ZP

Uže geografsko područje: Vojvodina
UGP

Godina: 2024
GO

Izdavač: Autorski reprint

IZ

Mesto i adresa: Novi Sad, Prirodno-matematički fakultet, Departman za matematiku i informatiku, Trg Dositeja Obradovića 4
MA

Fizički opis rada: 6 poglavlja, 52 strane, 13 lit. citata, 23 slika, 0 tabele
FO

Naučna oblast: Matematika
NOS

Naučna disciplina: Primenjena matematika
ND

Ključne reči: Kapacitet kanala, komunikacioni sistem, kanal sa višestrukim pristupom, m-korisnik, Gausovi vektorski kanali, bežični kanali
KR

Univerzalna decimalna klasifikacija:
UDK:

Čuva se: Biblioteka Departmana za matematiku i informatiku Prirodno-matematičkog fakulteta, u Novom Sadu
ČU

Važna napomena:
VN

Izvod: U prvom delu ovog rada objašnjavamo šta je to komunikacioni sistem i kapacitet kanala kroz definicije i primere. U drugom delu rada počinjemo sa analizom kanala sa višestrukim pristupom kroz razne definicije, svojstva i teoreme koji nas kasnije dovode do poslednjeg dela rada, koji analiziraju složenije modele.
IZ

Datum prihvatanja teme od strane NN veća: 24.11.2023.
DP

Datum odbrane:
DO

Članovi komisije:
KO
Predsednik: dr Dušan Jakovetić, vanredni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu

Mentor: dr Mladen Kovačević, vanredni professor, Fakultet tehničkih nauka, Univerzitet u Novom Sadu
Član: dr Dejan Vukobratović, redovni profesor, Fakultet tehničkih nauka, Univerzitet u Novom Sadu

UNIVERSITY OF NOVI SAD
FACULTY OF SCIENCES
KEY WORDS DOCUMENTATION

Accession number:
ANO

Identification number:
INO

Document type: Monograph type
DT

Type of record: Printed text
TR

Contents code: Master's thesis
CC

Author: Tatjana Naumov
AU

Mentor: Dr Mladen Kovačević
MN

Title: The capacity of multiple-access channels
TI

Language of text: English
LT

Language of abstract: English
LA

Country of publication: Republic of Serbia
CP

Locality of publication: Vojvodina
LP

Publication year: 2024
PY

Publisher: Author's reprint

PU

Publ. place: Novi Sad, Department of Mathematics and Informatics, Faculty of Science and Mathematics, University of Novi Sad, Trg Dositeja Obradovića 4
PP

Physical description: 6 chapters, 52 pages, 13 references, 23 figures, 0 tables
PD

Scientific field: Mathematics
SF

Scientific discipline: Applied mathematics
SD

Key words: Channel capacity, communication system, multiple-access channel, m-User, Gaussian vector channels, wireless channels
KW

Universal decimal classification:
UDC

Holding data: The Library of the Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad
HD

Note:
N

Abstract: This paper explains and analyzes the capacity of multiple-access channels. In the first part of the paper, we explain what is a communication system and channel capacity through definitions and examples. In the second part of the paper, we begin with the analysis of multiple-access channels through various definitions, properties and theorems, that later lead us to the last part of the paper, which analyze more complex models.
AB

Accepted by the Scientific Board on: 24.11.2023.
ASB

Defended:
DE

Thesis defend board:
DB

Chairperson: Dr Dušan Jakovetić, associate professor, Faculty of Sciences, University of Novi Sad
Mentor: Dr Mladen Kovačević, associate professor, Faculty of Technical Sciences, University of Novi Sad
Member: dr Dejan Vukobratović, full professor, Faculty of Technical Sciences, University of Novi Sad