

Linear Coding for the Gaussian Multiple Access Channel that achieves Sum Capacity

Ioannis Themistoklis Papoutsidakis

Thesis submitted in partial fulfillment of the requirements for the
Masters' of Science degree in Computer Science and Engineering

University of Crete
School of Sciences and Engineering
Computer Science Department
Voutes University Campus, 700 13 Heraklion, Crete, Greece

Thesis Advisor: Assistant Prof. *Xenofontas Dimitropoulos*

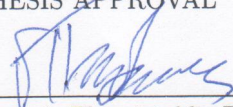
UNIVERSITY OF CRETE
COMPUTER SCIENCE DEPARTMENT

**Linear Coding for the Gaussian Multiple Access Channel that achieves
Sum Capacity**

Thesis submitted by
Ioannis Themistoklis Papoutsidakis
in partial fulfillment of the requirements for the
Masters' of Science degree in Computer Science

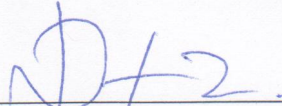
THESIS APPROVAL

Author:

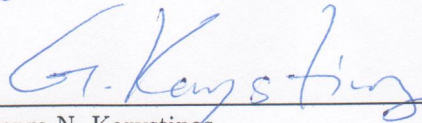


Ioannis Themistoklis Papoutsidakis

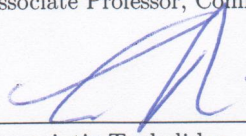
Committee approvals:



Xenofontas Dimitropoulos
Assistant Professor, Thesis Supervisor




George N. Karystinos
Associate Professor, Committee Member



Panagiotis Tsakalides
Professor, Committee Member

Departmental approval:



Antonios Argyros
Professor, Director of Graduate Studies

Heraklion, September 2018

Abstract

Multiple Access Channel (MAC) is a simple model for noisy many-to-one communication that is found in several applications, such as the uplink of a cellular system or medium access in a local area network. In this work we consider the practical aspects of the implementation of a Gaussian MAC. Specifically, we present several known results and techniques that are used on multiple access scenarios. The potential of uniform discrete inputs on a multiterminal environment with Gaussian noise is demonstrated. We give proof about the optimality of these inputs in terms of sum rate and highlight the duality between this problem and the optimal point-to-point transmission over a Gaussian channel using linear coding. The second part of this thesis is concerned with the construction of channel codes that achieve the rates that are previously produced. Encoding and decoding algorithms are analytically presented and experimental results are simulated.

Περίληψη

Το Κανάλι Πολλαπλών Προσβάσεων (ΚΠΠ) είναι ένα απλό μοντέλο για την θορυβώδη επικοινωνία πολλών χρηστών προς ένα που συναντάτε σε πολλές εφαρμογές, όπως το uplink ενός κυψελοειδούς συστήματος ή το medium access σε ένα τοπικό δίκτυο. Σε αυτή την εργασία εξετάζουμε τις πρακτικές πτυχές της εφαρμογής ενός Γκαουσιανού ΚΠΠ. Συγκεκριμένα, παρουσιάζουμε διάφορα γνωστά αποτελέσματα και τεχνικές που χρησιμοποιούνται σε σενάρια πολλαπλής πρόσβασης. Αναδεικνύεται η αξία των ομοιόμορφων διακριτών εισόδων σε ένα περιβάλλον με πολλούς χρήστες και Γκαουσιανό θόρυβο. Παρέχουμε απόδείξεις σχετικά με τη βελτιστότητα αυτών των εισόδων ως προς τον συνολικό ρυθμό μετάδοσης και τονίζουμε τη δυαδικότητα μεταξύ αυτού του προβλήματος και της βέλτιστης μετάδοσης σημείου - προς - σημείο σε ένα Γκαουσιανό κανάλι με την χρήση γραμμικής κωδικοποίησης. Το δεύτερο μέρος αυτής της εργασίας ασχολείται με την κατασκευή κωδίκων που επιτυγχάνουν τους ρυθμούς μετάδοσης που παρήχθησαν προηγουμένως. Οι αλγόριθμοι κωδικοποίησης και αποκωδικοποίησης παρουσιάζονται αναλυτικά και προσομοιώνονται πειραματικά αποτελέσματα.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τους επιβλεποντές μου, τον κ. Στέφανο Παπαδάκη και τον κ. Απόστολο Τραγανίτη, που με υποδέχθηκαν στο Πανεπιστήμιο Κρήτης και με καθοδήγησαν κατά την διάρκεια των μεταπτυχιακών σπουδών μου. Επίσης, θα ήθελα να ευχαριστήσω τους συμφοιτητές μου συνεργαστήκαμε ή με βοήθησαν με πολύτιμες συζητήσεις. Ιδιαίτερα τους κ.κ. Μ. Σουρλίγκα, Κ. Τριανταφυλλάκη, Γ. Βαρδάκη, Ν. Καραμολέγκο, Θ. Γκιόλια, Ν. Σδούκο, Α. Ηλία, Σ. Φαφαλιό και Κ. Μπουντρογιάννη.

Τέλος, αφιερώνω αυτή την εργασία στην οικογένεια μου, τους γονείς μου και την αδερφή μου, που με στήριξαν σε όλη την διάρκεια των σπουδών μου.

Contents

Table of Contents	i
List of Figures	v
1 Introduction	1
1.1 Network Information Theory	1
1.2 Coding Theory	2
1.3 Preliminaries	3
1.3.1 Random Variables	3
1.3.2 Entropy	3
1.3.3 Differential Entropy	4
1.3.4 Mutual Information	4
2 Gaussian Multiple Access Channel	7
2.1 Capacity of the Gaussian MAC	7
2.2 Common coding schemes	8
2.2.1 Treating other codeword as noise	9
2.2.2 Time-division multiple access	9
2.2.3 Time division with power control	9
2.2.4 Successive cancellation decoding	9
2.3 Uniform discrete input distributions	10
2.3.1 The case of the Gaussian point-to-point channel	10
2.3.2 A lower bound for the the Gaussian MAC	11
2.3.3 Input cardinality criterion	12
2.3.4 Sum-rate optimality for the k -sender Gaussian MAC	12
2.4 A dual problem	15
2.4.1 Shaping gain	15
2.4.2 Power distribution among senders	16
3 Polar Coding	21
3.1 Basic Notions	21
3.1.1 Symmetric Capacity	21
3.1.2 Binary Erasure Channel	22
3.2 Channel Polarization	22

3.3	Encoding	24
3.4	Decoding	24
3.5	Construction of Polar Codes	27
3.6	Performance Analysis	27
3.6.1	Hard-decision versus Soft-decision	27
3.6.2	Two-sender Gaussian MAC	28
3.6.3	Point-to-point Gaussian Channel	30
Appendix		33
1	Proof of theorem 4	33
Bibliography		35

List of Figures

1.1	The relationship between entropy and mutual information.	5
1.2	The additive white Gaussian noise channel.	5
2.1	Gaussian multiple access channel.	8
2.2	Capacity of the two-user Gaussian MAC.	8
2.3	Successive cancellation decoding of the two-user Gaussian MAC.	10
2.4	Rate regions of various methods for $S_1 = 10$ dB, $S_2 = 15$ dB and $r = 0.3$	13
2.5	Achievable rates with Gaussian and uniform inputs for the point-to-point Gaussian channel.	15
2.6	Shaping gain as a function of SNR.	16
2.7	Comparison of sum of three independent $\mathcal{U}(-1, 1)$ and four independent $\mathcal{U}(-\sqrt{3/4}, \sqrt{3/4})$ random variables with the normal distribution $\mathcal{N}(0, 1)$	18
2.8	Decay of absolute error $ F(z) - \Phi(z) $ as a function of n	18
2.9	Shaping gain as a function of SNR for various n 's.	19
3.1	The binary erasure channel with erasure probability ε	22
3.2	Basic polarization step.	22
3.3	Symmetric capacity for the BEC before and after the basic polarization step.	23
3.4	Generator matrix G_8 produced using (3.3).	24
3.5	Channel polarization for a BEC with $\varepsilon = 0.5$ and $N = 2^{11}$	25
3.6	Recursive construction of W_N from two copies of $W_{N/2}$	26
3.7	Comparison of soft-decision and hard-decision decoding on a binary transmission over the point-to-point Gaussian channel. The rate of the code is 0.5 bits/channel use and the block length is 2048.	28
3.8	Performance of the code of the first sender for several block lengths where $S_1 = 15$ dB.	29
3.9	Performance of the code of the second sender for block length $N = 4096$	30
3.10	Comparison between the virtual two-sender MAC as presented in section 2.4 and the straightforward use of uniform inputs. The block length is set to 2048 and the rate 1.5 bits/channel use.	31

Chapter 1

Introduction

In this chapter we introduce the basic theories and research fields behind this thesis. The fundamental preliminary notions that are used throughout this work are presented.

1.1 Network Information Theory

Network information theory studies the optimal flow of information through networks, that is the communication in multi-user scenarios in contrast to the original information theory that studies point-to-point schemes. It aims to establish the fundamental limits and the coding schemes that achieve these limits by extending the fundamental theorems of Shannon [1]. Even though the theory is far from complete, many elegant results and techniques have been developed over the past 45 years with a potential to be applied in real-world networks.

The first paper on network information theory was on the two-way channel by Shannon himself [2]. This was followed the next decade by several papers on multi-user schemes such as the work of Cover on the broadcast channel, the multiple access channel by Ahlswede and Liao, and the distributed lossless compression by Slepian and Wolf [3]. These results ignited the research in this field from the mid 1970s to the early 1980s with many new results and techniques developed. In spite of this early interest, many problems remained open and there was not any major research happening from the mid 1980s to the mid 1990s, with a lot of researchers altering their focus to other areas. The mid 1990s were a turning point for communication theorists and practitioners. The flourishing of wireless communications and the Internet was supported by advancements in semiconductor technology, channel and source coding, signal processing, and computer science. Since then the interest in the field is revived and in addition to the old open problems, new approaches, models, and concepts have been proposed. There are several techniques developed in network information theory that are used in real-world networks such as network coding and successive interference cancellation.

1.2 Coding Theory

Coding theory provides the mathematical tools used to construct codes for error correction, compression, and cryptography. Despite the fact that it is an older area compared to information theory it should not be considered as a discrete theory. The main reason is that before the seminal paper of Shannon the results were not produced in relevance to the fundamental limits and the codes were neither created nor benchmarked with the appropriate criteria in mind. The practice of designing codes without the use of Shannon’s information measurements exists even today and some of the state-of-the-art channel codes, such as Turbo codes, are meant to “approach” and not “achieve” the theoretical limits.

Research interest in coding followed a similar path to network information theory. The defining of information theory limits challenged the theorist for the first 20 years to invent codes that achieve the “capacity” of the a given channel. Two of the most prominent achievements of this period were the Convolutional codes and the Viterbi algorithm, which was used to optimally decode these codes. Most of the early space missions, such as the Voyager program, used these communication schemes. The years that followed were characterised by a decreasing interest from the research community until the mid 1990s, when Turbo codes were invented and performed incredibly better than the state-of-the-art at the time. A few years later, Low-density parity-check (LDPC) codes were rediscovered and studied again, since they were impractical back in the 1960s, when they were first invented by Gallager. The most recent milestone in the progress of the field is the invention of Polar codes by Arikan in the late 2000s, which were the first provably capacity-achieving codes.

This thesis focuses on a specific type channel codes, namely linear block codes. Linear block codes encode information in blocks and have the property of linearity, that is the encoding procedure can be executed by a linear transformation over a Galois Field $GF(q)$,

$$c = uG \tag{1.1}$$

where q is the cardinality of the used alphabet, c is the $1 \times N$ produced codeword, u is the $1 \times K$ information vector, and G is the $K \times N$ Generator matrix. The decoding is equivalent to the solution of the following linear system,

$$c = uG + z \tag{1.2}$$

where z is the $1 \times N$ noise vector. An important aspect of linear block codes is that they are appropriate for symmetric channels, that are the channels that the use of uniformly distributed inputs is essential for optimal transmission rates over them. For this work we are particularly attracted to the capacity-achieving Polar codes, since they achieve the theoretical transmission rates with absolute certainty.

1.3 Preliminaries

1.3.1 Random Variables

The transmitted signals are simulated with equivalent random variables. We are particularly interested in two classes of continuous probability distributions for our analysis. The first one the Normal distribution also known as Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, where μ is the mean value and σ^2 is the variance. The probability density function of this distribution is the following,

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (1.3)$$

Throughout this work we assume that the noise is normally distributed and in some cases the inputs of the given channel.

The second important continuous probability distributions is the Uniform distribution $\mathcal{U}(-a, a)$, where a defines the support $-a \geq x \geq a$. The probability density function is

$$f(x|a) = \begin{cases} \frac{1}{2a}, & -a \geq x \geq a \\ 0, & \text{otherwise.} \end{cases} \quad (1.4)$$

Clearly, the mean value of this distribution is 0. The variance is

$$\sigma^2 = \frac{a^2}{3}. \quad (1.5)$$

Variance is an important metric since in the case of zero-mean distributions this value is equivalent to the power of the random variable. In the following chapters we use the comparable notation that uses the variance instead of the support, $\mathcal{U}(-\sqrt{3\sigma^2}, \sqrt{3\sigma^2})$.

1.3.2 Entropy

Entropy is a measurement of uncertainty as described in [1]. Let a discrete random variable X with alphabet \mathcal{X} and probability mass function $p(x) = Pr\{X = x\}$. The entropy $H(X)$ is defined as,

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x). \quad (1.6)$$

The base of the logarithm determines the unit of measurement and it is usually 2, i.e. the entropy is measured in bits. Based on this definition it is derived that

$$H(X) \geq 0. \quad (1.7)$$

Another important formula which combines conditional and joint entropies is the chain rule. Specifically,

$$H(Y|X) = H(X, Y) - H(X). \quad (1.8)$$

1.3.3 Differential Entropy

A similar measurement to entropy but not exactly equivalent is the differential entropy. It is the straightforward conversion of (1.6), which is defined for discrete random variables, to the equivalent formula for the continuous distributions. Let X a continuous random variable, then

$$h(X) = - \int p(x) \log p(x) dx. \quad (1.9)$$

Chain rule remains relevant, however differential entropy does not retain, in general, many properties of entropy that make it a useful measurement of uncertainty. In particular, it is not invariant under a change of variables and can become negative. An adjustment to Shannon's definition for differential entropy was proposed by Jaynes in [4] called *limiting density of discrete points*. Despite these defects in the original definition of differential entropy, this measurement remains extremely important and works as it should in cases where differential entropies are compared as in the following section 1.3.4.

At this point we give the entropies of the two discussed continuous distributions in 1.3.1. The differential entropy of a random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is

$$h(X) = \frac{1}{2} \log(2\pi e\sigma^2). \quad (1.10)$$

The differential entropy of a random variable $X \sim \mathcal{U}(-\sqrt{3\sigma^2}, \sqrt{3\sigma^2})$ is

$$h(X) = \frac{1}{2} \log(12\sigma^2). \quad (1.11)$$

1.3.4 Mutual Information

Mutual information is a measurement of dependence between two random variables, or in other words, the information that holds one random variable about the other. The following results are exactly similar for discrete and continuous random variables alike. We present the discrete random variables case. Let two discrete random variables X and Y with joint probability mass function $p(x, y)$ and marginal probability functions $p(x)$ and $p(y)$, then

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \quad (1.12)$$

Some of the most useful mutual information identities are the following,

$$\begin{aligned} I(X; Y) &= I(Y; X) \\ &= H(Y) - H(Y|X) \\ &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X, Y). \end{aligned} \quad (1.13)$$

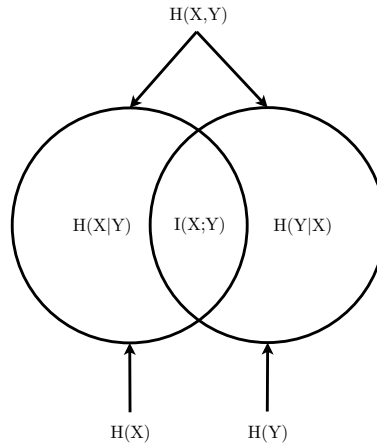


Figure 1.1: The relationship between entropy and mutual information.

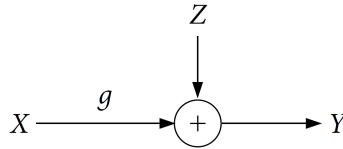


Figure 1.2: The additive white Gaussian noise channel.

In Figure 1.1 the relationship between entropy and mutual information is presented with a Venn diagram.

Based on the aforementioned results, Shannon gave the definition of *Channel Capacity* in [1], which is the maximum achievable transmission rate.

Theorem 1 (*Channel Coding Theorem*). *The capacity of the discrete memoryless channel $p(y|x)$ is given by the information capacity formula*

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

At this point we present the discrete-time additive white Gaussian noise channel as depicted in Figure 1.2, since it is one of the most important models that simulate real-world applications and its capacity is single-letter characterized and used throughout this thesis. The output of the channel which corresponds to input X is

$$Y = gX + Z, \tag{1.14}$$

where g is the channel gain and $Z \sim \mathcal{N}(0, \sigma_Z^2)$ is the noise. For the capacity characterization of this channel we use the equivalent to Theorem 1 for continuous

signals with input cost. For any $X \sim \mathcal{F}(x)$ with $E(X^2) = \sigma_X^2$,

$$\begin{aligned}
 I(X; Y) &= H(Y) - H(Y|X) \\
 &= H(X + Z) - H(X + Z|X) \\
 &= H(X + Z) - H(Z) \\
 &\stackrel{(a)}{=} \frac{1}{2} \log(2\pi e(\sigma_X^2 + \sigma_Z^2)) - \frac{1}{2} \log(2\pi e\sigma_Z^2) \\
 &= \frac{1}{2} \log \left(1 + \frac{\sigma_X^2}{\sigma_Z^2} \right) \\
 &= \mathsf{C} \left(\frac{\sigma_X^2}{\sigma_Z^2} \right) \\
 &= \mathsf{C}(S),
 \end{aligned} \tag{1.15}$$

where S is the signal-to-noise ratio and (a) follows from the fact that the maximum entropy distribution with support in $(-\infty, \infty)$, mean value $E(X) = a_1$, and $E(X^2) = a_2$ is $\mathcal{N}(a_1, a_2 - a_1^2)$ [5].

Chapter 2

Gaussian Multiple Access Channel

We consider the practical aspects of the implementation of the Gaussian Multiple Access Channel. We present the capacity region of this channel and results on the use of uniform inputs with linear codes. Additionally, the dual problem of optimal point-to-point Gaussian coding is revealed.

2.1 Capacity of the Gaussian MAC

The majority of classical network information theory problems remain open. Despite this fact, multiple access channels are well understood. We consider the channel depicted in Figure 2.1, where X_1 and X_2 are the inputs, g_1 , g_2 are the gains of each user and Z is the additive white Gaussian noise. We assume, without loss of generality, that the variance of the noise is equal to 1. Hence, the signal-to-noise ratios are defined as follows.

$$\begin{aligned} S_1 &= g_1 P \\ S_2 &= g_2 P \end{aligned} \tag{2.1}$$

Theorem 2 *The capacity region of the two-user Gaussian MAC is the set of rate pairs (R_1, R_2) such that*

$$\begin{aligned} R_1 &\leq C(S_1) \\ R_2 &\leq C(S_2) \\ R_1 + R_2 &\leq C(S_1 + S_2) \end{aligned}$$

where $C(x)$ is the Gaussian capacity function. [3]

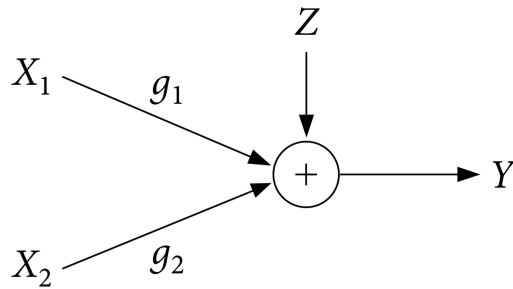


Figure 2.1: Gaussian multiple access channel.

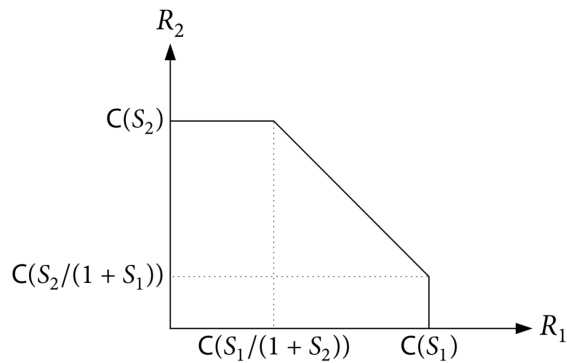


Figure 2.2: Capacity of the two-user Gaussian MAC.

It is notable that in the case of Gaussian MAC a simple inner and outer bound coincide and define the capacity of the channel as in Figure 2.2. This inner bound is obtained by using successive cancellation decoding and the outer bound is the capacity of the point-to-point Gaussian channel when SNR is $S_1 + S_2$.

In the case of k -sender Gaussian MAC the aforementioned results are generalized in a straightforward way [3]. Let the received SNRs S_j for $j \in [1 : k]$, then the capacity region is the set of rate tuples such that

$$\sum_{j \in \mathcal{J}} R_j \leq C \left(\sum_{j \in \mathcal{J}} S_j \right) \text{ for every } \mathcal{J} \subseteq [1 : k]. \quad (2.2)$$

2.2 Common coding schemes

We demonstrate practical schemes that use point-to-point Gaussian channel codes. We further show that such codes, when used with successive cancellation decoding and time sharing, can achieve the entire capacity region.

2.2.1 Treating other codeword as noise

In this scheme, each message is decoded by treating the other message as noise. This scheme achieves the set of rate pairs (R_1, R_2) such that,

$$\begin{aligned} R_1 &\leq C\left(\frac{S_1}{S_2 + 1}\right), \\ R_2 &\leq C\left(\frac{S_2}{S_1 + 1}\right). \end{aligned} \tag{2.3}$$

2.2.2 Time-division multiple access

A naive time-division scheme achieves the set of rate pairs (R_1, R_2) such that,

$$\begin{aligned} R_1 &\leq aC(S_1), \\ R_2 &\leq \bar{a}C(S_2). \end{aligned} \tag{2.4}$$

for some $a \in [0, 1]$ and $\bar{a} = 1 - a$.

Note that, despite the fact that the method of treating other codeword as noise is not close to capacity, in low SNRs it performs better than time-division multiple access.

2.2.3 Time division with power control

Note that the average power used by the senders in time-division multiple access is strictly lower than the average power constraint P for $a \in (0, 1)$. If the senders are allowed to use higher power in each transmission with the preservation of the average power P , the transmission rates are strictly greater than naive time-division multiple access scheme. This scheme achieves the set of rate pairs (R_1, R_2) such that,

$$\begin{aligned} R_1 &\leq aC\left(\frac{S_1}{a}\right), \\ R_2 &\leq \bar{a}C\left(\frac{S_2}{\bar{a}}\right). \end{aligned} \tag{2.5}$$

for some $a \in [0, 1]$ and $\bar{a} = 1 - a$.

Note that for $\alpha = S_1/(S_1 + S_2)$ the sum-rate $R_1 + R_2$ of this method is equal to the sum-capacity.

2.2.4 Successive cancellation decoding

The corner points of the Gaussian MAC capacity region can be achieved using successive cancellation decoding as depicted in Figure 2.3.

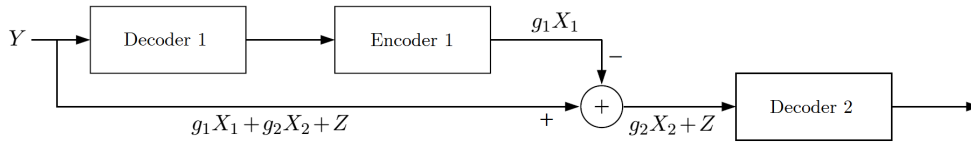


Figure 2.3: Successive cancellation decoding of the two-user Gaussian MAC.

- Upon receiving $Y = g_1X_1 + g_2X_2 + Z$, the receiver recovers X_2 while treating the received signal g_1X_1 from sender 1 as part of the noise. The probability of error for this step tends to zero as the block length of the code goes to infinity if $R_2 < C(S_2/(S_1 + 1))$.
- The receiver then subtracts g_2X_2 from Y and decodes $g_1X_1 + Z$ to recover X_1 . The probability of error for this step tends to zero as the block length of the code goes to infinity if $R_1 < C(S_1)$.

The other corner point can be achieved by changing the decoding order and any point on the $R_1 + R_2 = C(S_1 + S_2)$ line can be achieved by time sharing between the two corner points. Thus, any point inside the capacity region can be achieved using good point-to-point Gaussian channel codes.

2.3 Uniform discrete input distributions

As presented above, one can achieve the capacity of Gaussian MAC by using successive cancellation decoding and optimal Gaussian point-to-point random codes. A straightforward Gaussian signaling is not practical in every case. Most applications trade off a part of the capacity for the use of discrete uniform inputs. Also, state-of-the-art coding solutions such as LDPC and Polar codes are linear codes, hence, they need uniform inputs to work properly. In this section we derive a worst-case gap to capacity when uniform inputs are used for the Gaussian MAC.

2.3.1 The case of the Gaussian point-to-point channel

One of the most important and elegant results in information theory is the single-letter characterization of the point-to-point Gaussian channel capacity. This result shows that the use of normal distributed inputs is essential to achieve transmission with rate equal to capacity. Even so, most practical telecommunication schemes use sub-optimal uniform distributions. Ungerboeck in [6] was the first to observe that uniform discrete inputs perform quite close to the Gaussian ones. In [7], Ozarow and Wyner based on this observation derived a firm lower bound on the achievable rate for the case of N discrete uniform input levels.

$$I(X_N; Y_N) \geq C - \frac{1}{2} \log_2 \left(\frac{\pi e}{6} \right) - \frac{1}{2} \log_2 \left(\frac{1 + a^2}{a^2} \right) \quad (2.6)$$

Where C is the capacity of the channel and

$$a = N2^{-C} \quad (2.7)$$

We generalize this result to the multiple access channel and define a criterion for the density of the discrete input that is used.

2.3.2 A lower bound for the the Gaussian MAC

We derive the following result by employing successive decoding and (2.6). We assume, for a straightforward use of (2.6), that X_2 follows a Gaussian distribution. Since this assumption is not true and the Gaussian interference is the worst case scenario when it is treated as noise, we derive a strict inequality for the sum rate.

$$I(X_1; Y) > C \left(\frac{S_1}{S_2 + 1} \right) - \frac{1}{2} \log_2 \left(\frac{\pi e}{6} \right) - \frac{1}{2} \log_2 \left(\frac{1 + a_1^2}{a_1^2} \right) \quad (2.8)$$

$$I(X_2; Y|X_1) \geq C(S_2) - \frac{1}{2} \log_2 \left(\frac{\pi e}{6} \right) - \frac{1}{2} \log_2 \left(\frac{1 + a_2^2}{a_2^2} \right) \quad (2.9)$$

For the sum rate we have inequality (2.9).

$$\begin{aligned} I(X_1; Y) + I(X_2; Y|X_1) &> C(S_1 + S_2) - \log_2 \left(\frac{\pi e}{6} \right) - \frac{1}{2} \log_2 \left(\frac{1 + a_1^2}{a_1^2} \right) \\ &\quad - \frac{1}{2} \log_2 \left(\frac{1 + a_2^2}{a_2^2} \right) \end{aligned} \quad (2.10)$$

For our analysis we assume that $a = a_1 = a_2$, without loss of generality. Hence,

$$I(X_1; Y) + I(X_2; Y|X_1) > C(S_1 + S_2) - \log_2 \left(\frac{\pi e}{6} \right) - \log_2 \left(\frac{1 + a^2}{a^2} \right). \quad (2.11)$$

Assuming that the input distributions are continuous and uniform we derive inequality (2.12). This is, in practice, equivalent to a dense discrete input since the channel is noisy.

$$\begin{aligned} I(X_1; Y) + I(X_2; Y|X_1) &> C(S_1 + S_2) - \log_2 \left(\frac{\pi e}{6} \right) \\ &= C(S_1 + S_2) - 0.5092 \end{aligned} \quad (2.12)$$

This result proves that the use of uniform dense discrete distributions is asymptotically stable, i.e. sum capacity of the Gaussian MAC is achieved within 0.5092 bit for any signal-to-noise ratio. In practice, the sum-rate is closer to sum-capacity in comparison to the point-to-point scenario, even in the case of two senders. This can be observed in Figure 2.4.

2.3.3 Input cardinality criterion

In this part, we present a formula for the choice of the input cardinality, i.e. the number of symbols. We set a tolerance offset r in relation to the narrowest bound when input distribution is continuous and uniform. From (2.11) and (2.12) we have,

$$r = \log_2 \left(\frac{1 + a^2}{a^2} \right). \quad (2.13)$$

As the cardinality of the input becomes larger, r goes to zero. So, for our simulations we set a small target value for r , and by using equation (2.7), we derive the number of input levels for each user.

Using (2.13) and (2.7), we derive the following formulas for each sender, assuming that the message X_1 is decoded first,

$$\begin{aligned} N_1 &= \left\lceil \sqrt{\frac{S_1 + S_2 + 1}{(2^r - 1)(S_2 + 1)}} \right\rceil, \\ N_2 &= \left\lceil \sqrt{\frac{S_2 + 1}{2^r - 1}} \right\rceil. \end{aligned} \quad (2.14)$$

In Figure 2.4, we present the rate region of various well studied methods in comparison to the proposed scheme, for $S_1 = 10$ dB, $S_2 = 15$ dB and $r = 0.3$. For those parameters, we use 12 symbols for sender 1 and 3 for sender 2, based on the aforementioned equations (2.14). In the case where X_2 is decoded first, 5 symbols are used for the first sender and 7 symbols for the other one. We observe that uniform discrete inputs are quite competitive compared to capacity, despite the small number of used symbols. The trade-off of using this scheme with linear codes is very attractive compared to Gaussian codes.

2.3.4 Sum-rate optimality for the k -sender Gaussian MAC

Based on our observations, it is remarkable that in terms of sum-rate this multi-user arrangement seems to work better than a point-to-point Gaussian channel transmission scheme with the same aggregate power constraint.

We generalize this conjecture for the case of k -senders and prove it based on the central limit theorem (CLT). This theorem establishes that when independent random variables are added, their sum tends toward a normal distribution even if the original variables themselves are not normally distributed. There are several variants of the CLT specialized to different assumptions. We formally present the Lyapunov CLT [8].

Theorem 3 (*Lyapunov CLT*) Suppose $\{X_1, X_2, \dots\}$ is a sequence of independent

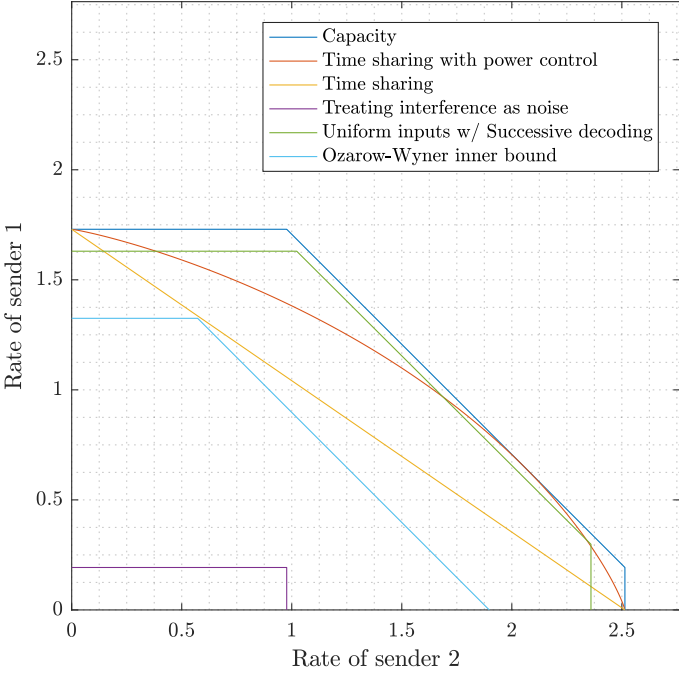


Figure 2.4: Rate regions of various methods for $S_1 = 10$ dB, $S_2 = 15$ dB and $r = 0.3$.

random variables, each with finite expected value μ_i and variance σ_i^2 . Define

$$s_n^2 = \sum_{i=1}^n \sigma_i^2.$$

If for some $\delta > 0$, Lyapunov's condition

$$\lim_{i \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n E[|X_i - \mu_i|^{2+\delta}] = 0$$

is satisfied, then a sum of $\frac{X_i - \mu_i}{s_n}$ converges in distribution to a standard normal random variable, as n goes to infinity,

$$\frac{1}{s_n} \sum_{i=1}^n (X_i - \mu_i) \xrightarrow{d} N(0, 1).$$

For our study, we are exclusively interested in zero-mean uniform distributions, since zero-mean distributions maximize the the variance of power constraint signals. The author in [8] proves that any zero-mean and uniformly bounded distribution satisfies the Lyapunov's condition for $\delta = 1$. Our distributions have finite support and are symmetric around the mean $\mu_i = 0$, therefore they are uniformly bounded and the Lyapunov's condition is satisfied.

Based on the aforementioned results we prove that uniform distributions are optimal in terms of sum-rate. Suppose X_1, X_2, \dots, X_k are independent continuous uniform random variables, i.e. the signals of each user. Each signal has expected value $\mu_i = 0$ and variance $\sigma_i^2 = S_i$. From Theorem 3 and [8], as $k \rightarrow \infty$,

$$\begin{aligned} I(X_1, X_2, \dots, X_k; Y) &= h(Y) - h(Y|X_1, X_2, \dots, X_k) \\ &= h\left(\sum_{i=1}^k X_i + Z\right) - h\left(\sum_{i=1}^k X_i + Z|X_1, X_2, \dots, X_k\right) \\ &= h\left(\sum_{i=1}^k X_i + Z\right) - h(Z) \\ &= \frac{1}{2} \log\left(2\pi e \sum_{i=1}^k S_i\right) - \frac{1}{2} \log(2\pi e) \\ &= \mathbb{C}\left(\sum_{i=1}^k S_i\right) \end{aligned} \tag{2.15}$$

Even though this proof stands asymptotically as k goes to infinity, the rate of convergence is very promising even for a small number of senders.

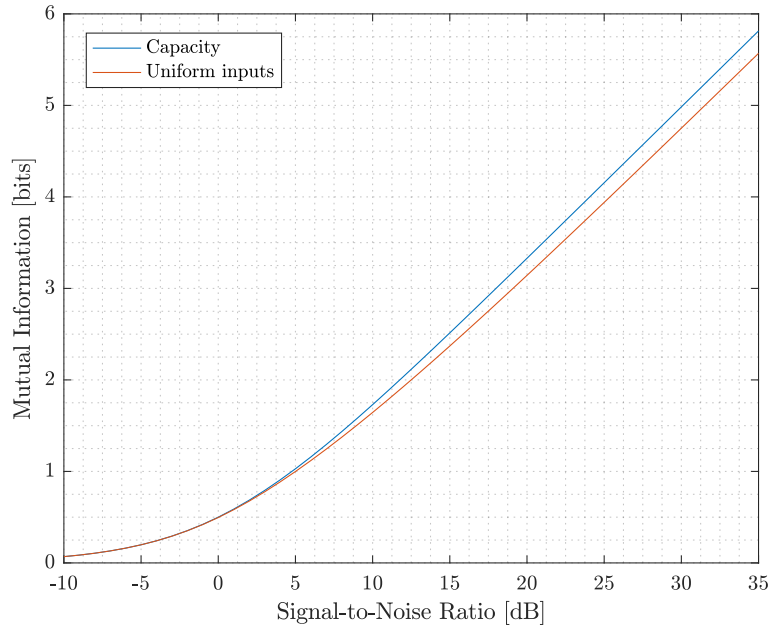


Figure 2.5: Achievable rates with Gaussian and uniform inputs for the point-to-point Gaussian channel.

2.4 A dual problem

There is duality between this problem and the optimal point-to-point transmission over a Gaussian channel using linear codes. The so-called shaping gain of 1.53 dB can be achieved [9]. There is some work done in this direction, such as [10], where the idea of using a multiple access scheme with binary convolutional codes to approach the capacity of the Gaussian noise channel is presented. In [11] the authors use the same scheme with binary Turbo codes to approach capacity without referring to the multiple access channel. The most recent work is [12], where a comparison is made between the method of [11] and a particular shaping technique using Polar codes. In this section we present an analysis of the shaping gain that can be achieved and we give a method for optimal split of the power among the k senders based on results on the Irwin-Hall distribution.

2.4.1 Shaping gain

Shaping gain is called the power gain that is achieved when particular methods are used to “correct” the sub-optimality of the usage of non-Gaussian, usually uniform, input distribution on a Gaussian channel. In Figure 2.5 this sub-optimality is depicted. Shaping therefore induces a Gaussian-like probability distribution on a constellation, rather than an equiprobable distribution.

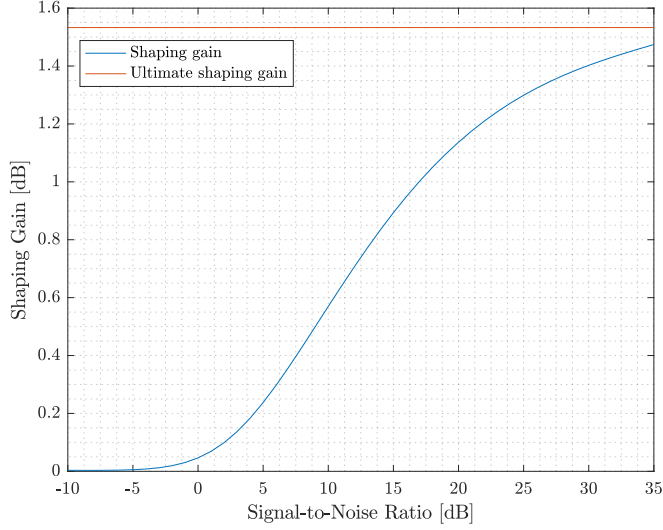


Figure 2.6: Shaping gain as a function of SNR.

In this section we give a simple upper bound on the shaping gain, namely, *ultimate shaping gain*. The basic idea is that we compare the required power to represent the same amount of information with a zero-mean uniform distribution as with a zero-mean Gaussian distribution. Let $U \sim \mathcal{U}(-\sqrt{3\sigma_u^2}, \sqrt{3\sigma_u^2})$ and $G \sim \mathcal{N}(0, \sigma_g^2)$ where σ_u^2 and σ_g^2 are the variances, i.e. the powers of the signals. The amount of information that is represented by a probability distribution is calculated by the differential entropy,

$$\begin{aligned}
 h(U) = h(G) &\Rightarrow \\
 \frac{1}{2} \log(12\sigma_u^2) &= \frac{1}{2} \log(2\pi e\sigma_g^2) \Rightarrow \\
 12\sigma_u^2 &= 2\pi e\sigma_g^2 \Rightarrow \\
 \frac{\sigma_u^2}{\sigma_g^2} &= \frac{\pi e}{6}.
 \end{aligned} \tag{2.16}$$

Note that this gain of $\pi e/6$ (1.53 dB) is available as signal-to-noise ratio grows to infinity. In Figure 2.6 we present the actual shaping gains for finite SNRs. Despite the fact that the use uniform distributions is not detrimental, as presented in section 2.3, we observe shaping gains that correspond to a possible reduction in power up to roughly 30%.

2.4.2 Power distribution among senders

At this point, arises the question of how the power should be distributed among the senders with the objective of fast convergence to the capacity in mind. From

equation 2.15 we derive that this problem can be reduced to the fast convergence of the sum of input distributions to a Normal distribution. Through observation it seems that the optimal convergence is achieved when the inputs are identically distributed, i.e. each sender uses the same amount of power. Indeed, this conjecture is true.

Theorem 4 *The sum of n independent and identical uniform distributions converges faster to a normal distribution in comparison to the sum of n independent but not identical uniform distributions.*

In the first section of the appendix we provide the proof as presented in [13].

In this section, we demonstrate some interesting results on the equivalent Irwin-Hall distribution, which is the continuous probability distribution for the sum of n independent and identically distributed $\mathcal{U}(0, 1)$ random variables [14]. Despite the definition that is usually used in the literature, in this thesis we refer to the Irwin-Hall distribution as the distribution for the sum of n i.i.d. zero-mean $\mathcal{U}(-\sqrt{3\sigma^2}, \sqrt{3\sigma^2})$ random variables, where σ^2 is the variance. We are particularly interested on results about the rate of convergence of this distribution. Publication [14] provides an upper bound on the absolute difference of the cumulative distribution function $F(z)$ of the Irwin-Hall distribution and the cumulative distribution function $\Phi(z)$ of Normal distribution,

$$|F(z) - \Phi(z)| \leq \frac{\sqrt{3}}{20\sqrt{n}}. \quad (2.17)$$

This result stands true independently from the support of the i.i.d uniform random variables that are added. In Figure 2.7 we compare the probability density function of the Irwin-Hall distribution for $n = 3$, $n = 4$ with the probability density function of the Normal distribution $N(0, 1)$. All three resulting distributions have variances equal to 1 and they are indeed similar. In Figure 2.8 we plot the bound (2.17). In both cases the results are very promising for an actual implementation of this scheme, since with a small number of virtual senders we achieve distributions that are very close to the optimal Normal distribution.

In Figure 2.9 the power loss, i.e. the available shaping gain, for several n 's is illustrated. We observe that even with two virtual senders the majority of the available shaping gain is achieved.

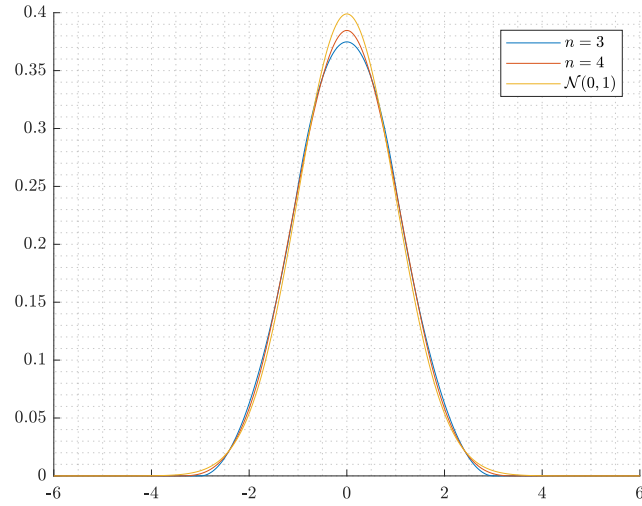


Figure 2.7: Comparison of sum of three independent $\mathcal{U}(-1, 1)$ and four independent $\mathcal{U}(-\sqrt{3}/4, \sqrt{3}/4)$ random variables with the normal distribution $\mathcal{N}(0, 1)$.

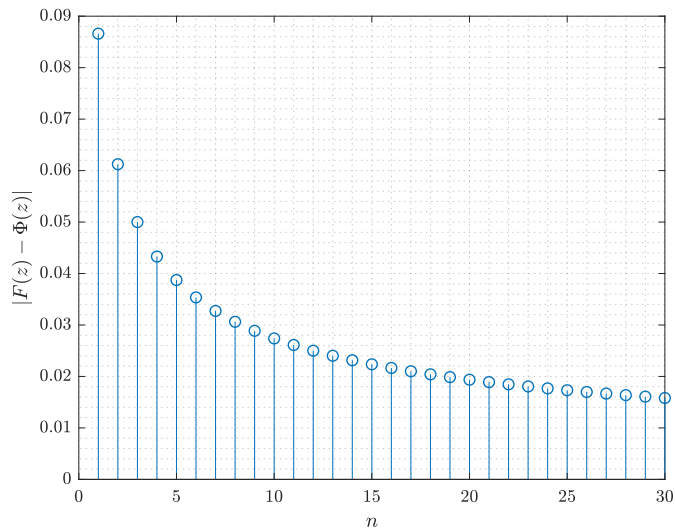


Figure 2.8: Decay of absolute error $|F(z) - \Phi(z)|$ as a function of n .

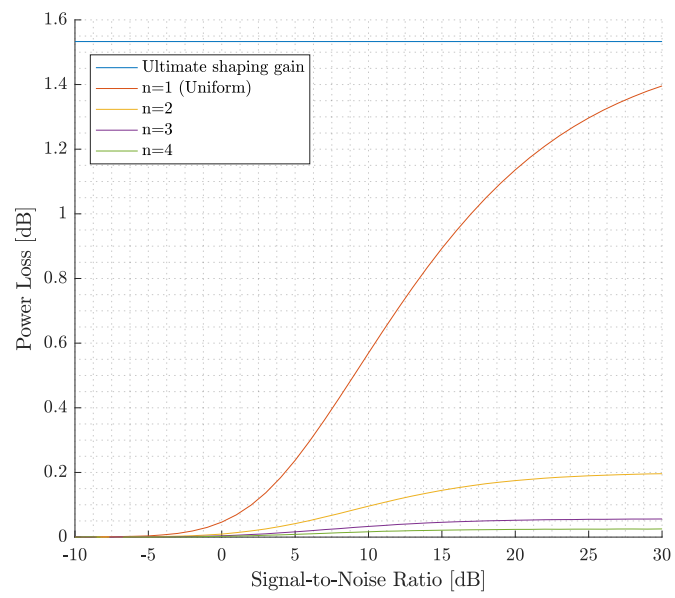


Figure 2.9: Shaping gain as a function of SNR for various n 's.

Chapter 3

Polar Coding

In this chapter we present the coding scheme that is used to implement the concepts of Chapter 2. Specifically we present the technique of channel polarization, which was originally proposed in [15] for binary-input discrete memoryless channels and is the basis of polar coding. These codes can achieve the “symmetric capacity,” i.e. the mutual information produced by uniformly distributed inputs, of any binary-input channel by employing low-complexity encoding and decoding algorithms. Later, it was proved that this code construction can achieve the capacity of any q -ary discrete memoryless channel [16]. This section rehearses the construction of Polar codes. Note that for the purposes of this work we use alphabets of prime cardinality, since the generalization from binary polar coding is straightforward.

3.1 Basic Notions

3.1.1 Symmetric Capacity

Given a q -ary input channel $W : \mathcal{X} \rightarrow \mathcal{Y}$ with $\mathcal{X} = \{0, 1, \dots, q-1\}$, its symmetric capacity is defined as

$$I(W) \doteq \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \frac{1}{q} W(y|x) \log_q \frac{W(y|x)}{\sum_{x' \in \mathcal{X}} \frac{1}{q} W(y|x')}. \quad (3.1)$$

Symmetric capacity is nothing but the mutual information between the input and the output of the channel when the input is uniformly distributed. Therefore, if the channel is symmetric, then its Shannon capacity is equal to its symmetric capacity. It is known that linear codes produce uniformly distributed codewords and, in the case of symmetric channels, uniform input distribution is needed to maximize mutual information between the transmitter and the receiver. Since we use base- q logarithm, we calculate capacity in q -ary symbols per channel use. Consequently,

$$0 \leq I(W) \leq 1.$$

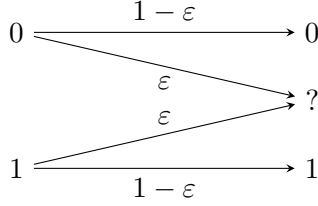
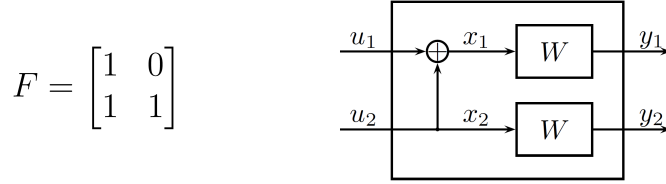
Figure 3.1: The binary erasure channel with erasure probability ε .

Figure 3.2: Basic polarization step.

3.1.2 Binary Erasure Channel

In this chapter, we use the binary erasure channel for our analysis, as in Figure 3.1. Once a bit is transmitted, the receiver either obtains the bit correctly with probability $1 - \varepsilon$ or receives a message that the bit was not received with probability ε . We choose this type of channel because it is relatively easy to construct algorithms for evaluating the polarized channels, due to the closed-form expression for the BEC capacity. The symmetric capacity of the BEC is

$$I(W) = 1 - \varepsilon. \quad (3.2)$$

The proof is provided in [5].

3.2 Channel Polarization

Channel polarization is an operation by which, out of N independent copies of a given discrete memoryless channel W , one manufactures a second set of N channels $\{W_N^{(i)} : 1 \leq i \leq N\}$ that show a polarization effect in the sense that, as N becomes large, the symmetric capacity terms $\{I(W_N^{(i)})\}$ tend towards 0 or 1 for all but a vanishing fraction of indices i [15].

To achieve this effect, we use a linear transformation over $\text{GF}(q)$, where q is a prime, for combining two identical channels to new synthetic channels W' and W'' ,

$$\{W, W\} \mapsto \{(y_1^2, u_1), (y_1^2, u_1; u_2)\} \iff \{W, W\} \mapsto \{W', W''\}.$$

The first step of this recursive transformation is shown in Figure 3.2. As this transformation occurs, the first channel degrades and the second upgrades in terms of symmetric capacity. In Figure 3.3, the latter is observed for the BEC.

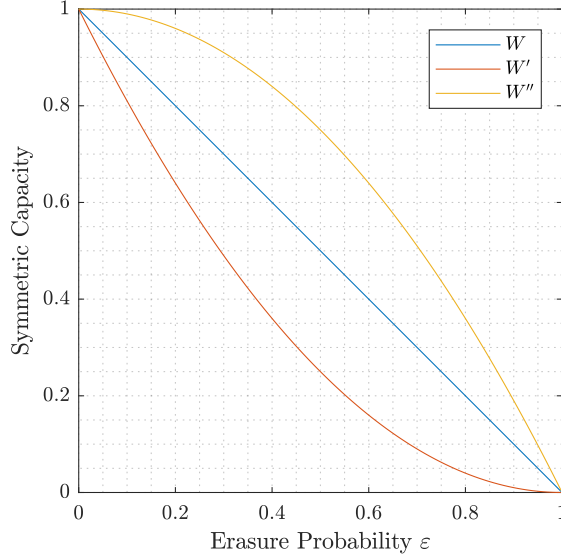


Figure 3.3: Symmetric capacity for the BEC before and after the basic polarization step.

Having shown how the basic step of original polarization works, we can define the recursion that constructs the generator matrix of Polar codes. As in [15], we define the Kronecker power $G^{\otimes n}$ as $G \otimes G^{\otimes n-1}$ for all $n \geq 1$, where \otimes denotes the Kronecker product and $G^{\otimes 0} = [1]$.

$$G_N = F^{\otimes n} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes n}. \quad (3.3)$$

With this method we can construct $N \times N$ generator matrices, where $N = 2^n$. In Figure 3.4 the generator matrix G_8 is presented. The next step is to define which rows of G correspond to the perfect channels and which to the useless. In the perfect rows, there will be put information bits, while in the useless rows, there will be put frozen (known to the decoder) bits. This task is easily managed for the BEC. We can recursively calculate the symmetric capacities of the manufactured channels using the following formulas [15].

$$I(W_N^{(2i-1)}) = I(W_{N/2}^{(i)})^2, \quad (3.4)$$

$$I(W_N^{(2i)}) = 2I(W_{N/2}^{(i)}) - I(W_{N/2}^{(i)})^2. \quad (3.5)$$

In the case of BEC(0.5), the symmetric capacity limit is at 0.5 bits per channel use. In Figure 3.5, using (3.4) and (3.5), we observe the effect of channel polarization. Indeed, almost half of the channels are perfect and the other half are useless.

$$G_8 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Figure 3.4: Generator matrix G_8 produced using (3.3).

3.3 Encoding

In this section, we consider the implementation of the encoder of Polar codes. Matrix multiplication of G with an information vector is easy for small block lengths but not convenient for bigger block lengths. We design a recursion based on the main channel combination W_2 that has been indicated in the Section 2.2.

The general form of the recursion is shown in Figure 3.6 where two independent copies of $W_{N/2}$ are combined to produce channel W_N . The operator R_N is a permutation, known as the reverse shuffle operation, and simply separates the odd-indexed from the even-indexed signals. Odd-indexed signals become input to the first copy of $W_{N/2}$ and even-indexed to the second.

In terms of complexity, if we take the complexity of a scalar mod- q addition as 1 unit and the complexity of the reverse shuffle operation R_N as N units of time we have,

$$\begin{aligned} T(N) &= \frac{N}{2} + O(N) + 2T\left(\frac{N}{2}\right) \Rightarrow \\ T(N) &= O(N \log_2 N). \end{aligned} \quad (3.6)$$

3.4 Decoding

The decoder introduced in [15] is called successive cancellation decoder. Its role is to decide with the rule of closest neighbour on the i^{th} symbol ($1 \leq i \leq N$) that is transmitted over $W_N^{(i)}$ by computing

$$\hat{u}_i = \begin{cases} u_i, & \text{when } u_i \text{ is a frozen symbol} \\ \arg \max_{x \in \{0, 1, \dots, q-1\}} W_N^{(i)}(y_1^N, u_1^{i-1} | x), & \text{otherwise,} \end{cases} \quad (3.7)$$

This decoding scheme estimates sequentially every information symbol. Each estimation is carried out by using the knowledge of frozen and previously estimated

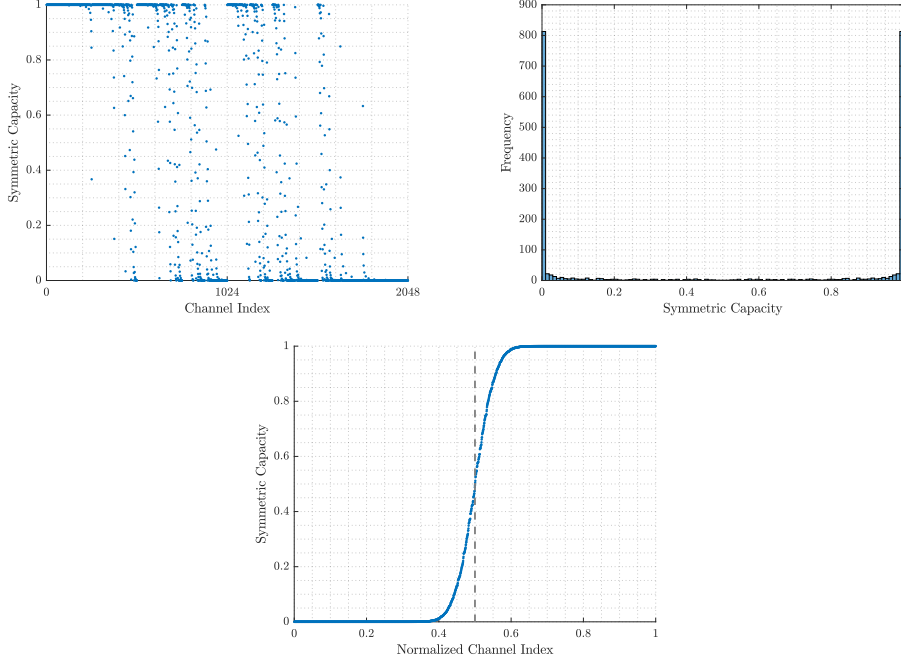


Figure 3.5: Channel polarization for a BEC with $\varepsilon = 0.5$ and $N = 2^{11}$.

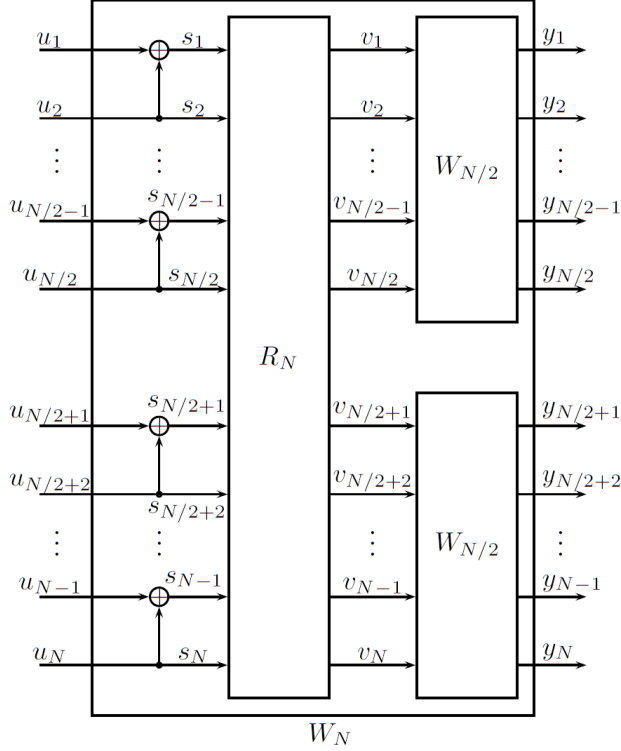
symbols. Note that the nature of the successive cancellation decoding of the multiple access channel has a distinct relationship with channel polarization and the successive cancellation decoding of Polar codes, as to the fact that in both cases the mutual information is conserved and achieved. We calculate the probabilities of (3.7) using the recursive formulas (3.8) and (3.9).

$$W_{2N}^{(2i-1)}(y_1^{2N}, u_1^{2i-2} | u_{2i-1}) = \sum_{u_{2i}} \frac{1}{q} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2} | u_{2i-1} \oplus u_{2i}) \cdot W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2} | u_{2i}), \quad (3.8)$$

$$W_{2N}^{(2i)}(y_1^{2N}, u_1^{2i-1} | u_{2i}) = \frac{1}{q} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2} | u_{2i-1} \oplus u_{2i}) \cdot W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2} | u_{2i}). \quad (3.9)$$

Every transition probability in this recursion is used over one time. For this reason, we implement a data structure to store these values in order not to calculate them again. We use q matrices of size $N \times (\log_2 N + 1)$. Each cell is filled after $\Theta(1)$ calculations, which implies that the complexity of the decoder is $O(N \log_2 N)$.

In the special case of binary alphabets, it is convenient to define the likelihood ratio as

Figure 3.6: Recursive construction of W_N from two copies of $W_{N/2}$.

$$L_N^{(i)}(y_1^N, \hat{u}_1^{i-1}) = \frac{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|0)}{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|1)}. \quad (3.10)$$

This way, the SC decoder is defined as

$$\hat{u}_i = \begin{cases} u_i, & \text{if } u_i \text{ is a frozen bit} \\ 0, & \text{if } L_N^{(i)}(y_1^N, \hat{u}_1^{i-1}) \geq 1 \\ 1, & \text{otherwise.} \end{cases} \quad (3.11)$$

For computing $L_N^{(i)}(y_1^N, \hat{u}_1^{i-1})$, a straightforward calculation using the recursive formulas (3.8) and (3.9) gives

$$L_N^{(2i-1)}(y_1^N, \hat{u}_1^{2i-2}) = \frac{L_{N/2}^{(i)}(y_1^{N/2}, \hat{u}_{1,o}^{2i-2} \oplus \hat{u}_{1,e}^{2i-2}) L_{N/2}^{(i)}(y_{N/2+1}^N, \hat{u}_{1,e}^{2i-2}) + 1}{L_{N/2}^{(i)}(y_1^{N/2}, \hat{u}_{1,o}^{2i-2} \oplus \hat{u}_{1,e}^{2i-2}) + L_{N/2}^{(i)}(y_{N/2+1}^N, \hat{u}_{1,e}^{2i-2})}, \quad (3.12)$$

$$L_N^{(2i)}(y_1^N, \hat{u}_1^{2i-1}) = \left[L_{N/2}^{(i)}(y_1^{N/2}, \hat{u}_{1,o}^{2i-2} \oplus \hat{u}_{1,e}^{2i-2}) \right]^{1-2\hat{u}_{1,e}^{2i-1}} L_{N/2}^{(i)}(y_{N/2+1}^N, \hat{u}_{1,e}^{2i-2}). \quad (3.13)$$

3.5 Construction of Polar Codes

Having in our hands a low complexity encoding and decoding channel code, it is of great interest the efficient construction of such a code. That is, the determination of the frozen symbols that are constant and known to the receiver. This is a research subject by itself as only the q -ary erasure channel has closed-form formulas (3.4) and (3.5) for channel evaluation.

There are several results on this problem in the literature. The most acclaimed is [18], where upper and lower bounds of the capacity are calculated recursively, until a proper sorting of the channels is possible. Unfortunately, this scheme is efficient in terms of complexity only for binary-input channels. The challenge of this work is to construct polar codes for q -ary discrete-input continuous-output channels. With this object in mind, we make a trade-off between the theoretical optimal results and the complexity of our scheme. Particularly, we convert the given discrete-input continuous-output channel to a discrete-input discrete-output channel with very large output alphabet. This is not a trade-off in practice, since conventional systems cannot operate with analog inputs without quantization and the lost rate from this quantization is infinitesimal.

Considering that there is not any exact algorithm for a evaluating q -ary channels we use an approximate method. Specifically, we use a Monte Carlo approach to solve the problem of channel evaluation as it was proposed in the introductory paper of polar codes [15]. Using a transmission arrangement with successive cancellation decoding we compute the symbol error rate of the i -th W_i constructed channel by transmitting frozen symbols to all previously decoded channels W_1, \dots, W_{i-1} .

3.6 Performance Analysis

In this section we evaluate the performance of the proposed transmission schemes. We examine two particular cases. The first one is about the 2-sender Gaussian MAC and the second is about the optimal transmission over a point-to-point Gaussian channel as presented in section 2.4. The purpose of this analysis is to observe the polarization effect as block length becomes larger in the case of the Gaussian MAC and to compare the practical shaping gain that is achieved using the a virtual MAC in comparison to the straightforward use of uniformly distributed inputs.

3.6.1 Hard-decision versus Soft-decision

There are two considerable approaches for treating the output of a Gaussian channel. The most commonly used method is the hard-decision decoding, which is the process of assigning the real-valued output of the channel to a discrete symbol (demodulation) and then decoding the resulting codeword. The second and least used approach is the soft-decision decoding, in which the output of the channel is given directly to the decoder. In general, the superiority of soft-decision decoding is well known. However, due to the increased complexity in comparison to the

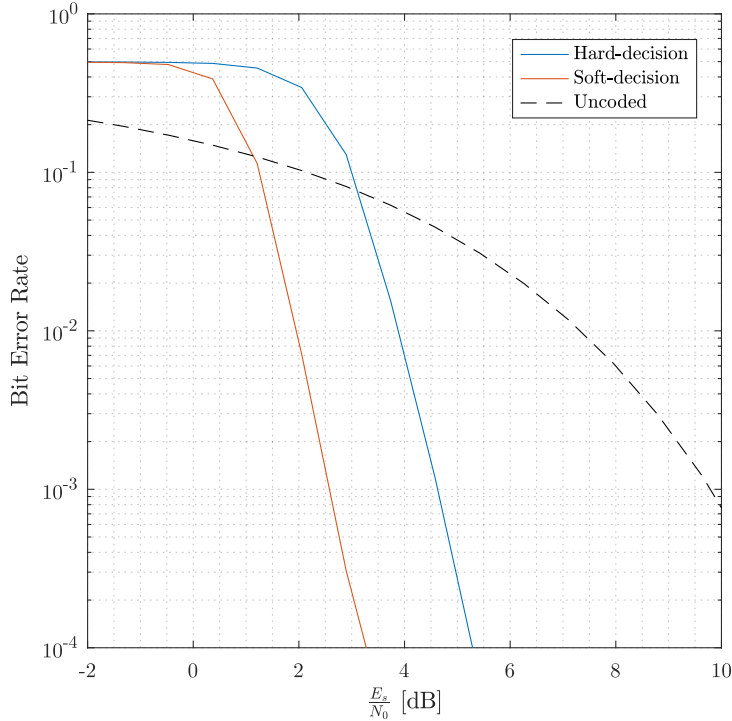


Figure 3.7: Comparison of soft-decision and hard-decision decoding on a binary transmission over the point-to-point Gaussian channel. The rate of the code is 0.5 bits/channel use and the block length is 2048.

hard-decision decoding this method is not extensively used. In our case, both decoding schemes are performed with the same complexity. This has to do with the nature of the successive cancellation decoder, since the only difference between the two methods is the initialization of the recursion of the decoder. In Figure 3.7, the performance of soft-decision and hard-decision decoding is depicted in a specific point-to-point transmission scenario. Every simulation that follows is performed using soft-decision decoding.

3.6.2 Two-sender Gaussian MAC

The first simulation that we perform is for the case presented in Figure 2.4, where the signal-to-noise ratio of the first sender is $S_1 = 15$ dB and the second is $S_2 = 10$ dB. Using the formulas of (2.14) and setting $r = 0.3$ we set the first sender to use a quinary alphabet and the second a septenary.

In general, the capacity is an asymptotic identity of the channel, that is it is achievable as block length grows to infinity. This becomes apparent in Figure 3.8 where the performance of the code of sender 1 is presented in terms of symbol error

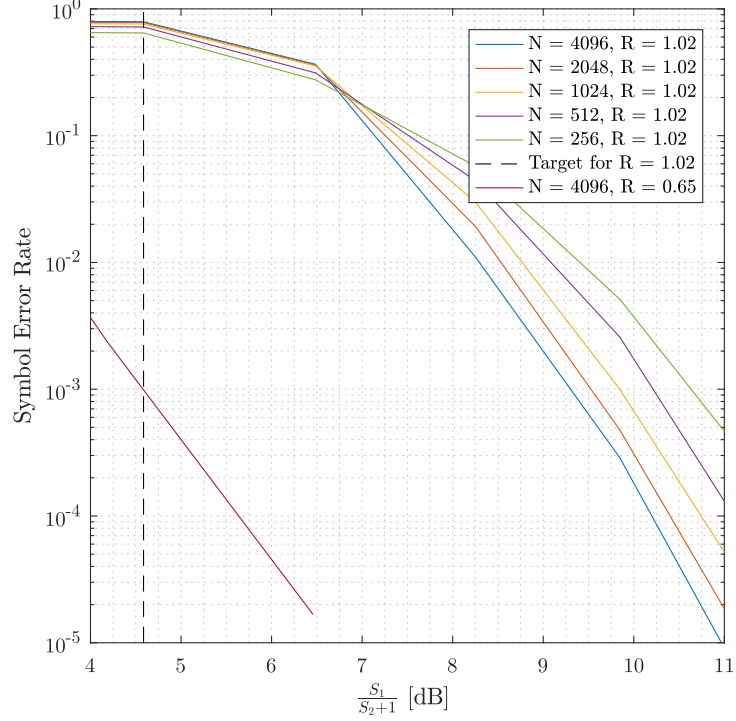


Figure 3.8: Performance of the code of the first sender for several block lengths where $S_1 = 15$ dB.

rate for several block lengths. In our scenario the “absolutely” correct reception of the first sender is extremely important, since only then the first codeword can be subtracted from the received signal. We set the tolerance limit to symbol error rate to 10^{-3} . To achieve such error rate for the first sender in moderate block lengths we have to choose one out of two options. The first one is to obligate the second sender to use less power, which means that $\frac{S_1}{S_2+1}$ becomes better for the first sender but at the same time the second one limits its rate. The second option is to reduce the transmission rate of the first sender and let the second sender unaffected. We choose the latter since in terms of sum-rate the second codeword usually carries the majority of the information and in the case discussed in section 2.4 it invariably does. In Figure 3.8, we also display the performance of the code when the rate is reduced. Using this rate, the target error rate is achieved and the first codeword can be subtracted without complications. In Figure 3.9, the performance of the code of the second sender is depicted for block length equal to 4096 symbols. For this relatively small block length, a sum-rate of 1.45 bits/channel use is achieved with probability of error equal to 10^{-3} .

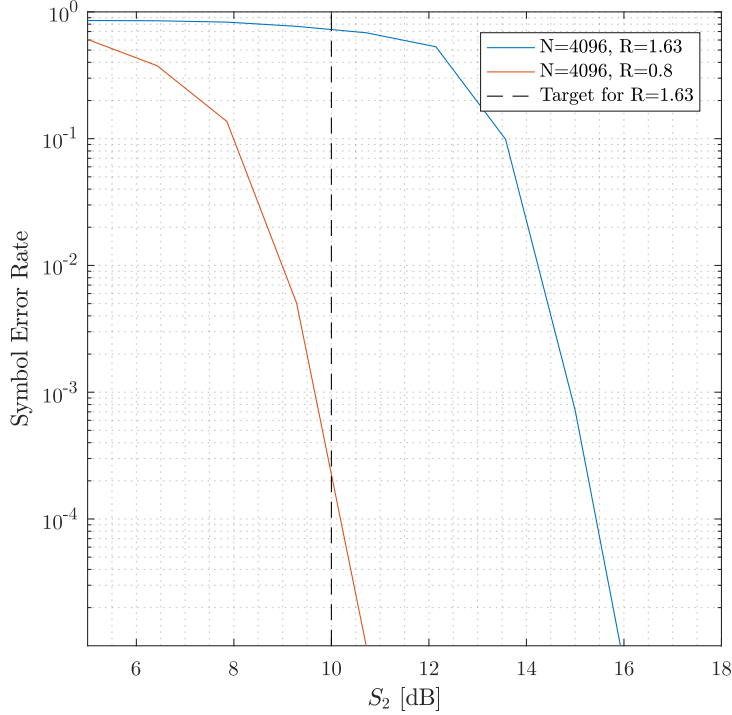


Figure 3.9: Performance of the code of the second sender for block length $N = 4096$.

3.6.3 Point-to-point Gaussian Channel

A performance comparison between a virtual two-sender MAC scheme and the straightforward use of uniform inputs is simulated. The comparison is made in terms of frame error rate since several different alphabets are used and symbol error rate is not appropriate. In the case of uniform inputs a quinary alphabet is used and the rate of the code is set to 1.5 bits/channel use. For the case of virtual senders the first sender uses a transmission rate that allows correct decoding almost always. We set this tolerance limit to the frame error rate of the first sender to 10^{-3} . The first virtual sender transmits with a rate of 0.45 bits/channel use and uses a binary alphabet. The second one uses a ternary alphabet and transmits with a rate of 1.05 bits/channel use. Note that in both cases the aggregate rate is 1.5 bits/channel use. In Figure 3.10 the dominance of our proposed method is depicted over the conventional one.

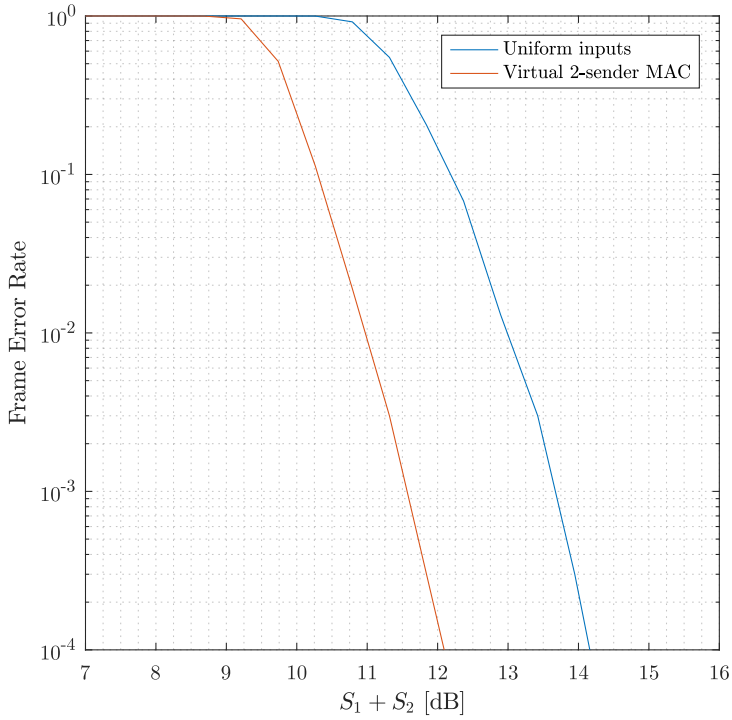


Figure 3.10: Comparison between the virtual two-sender MAC as presented in section 2.4 and the straightforward use of uniform inputs. The block length is set to 2048 and the rate 1.5 bits/channel use.

Appendix

1 Proof of theorem 4

Let U_1, U_2, \dots be i.i.d. random variables uniformly distributed on $[-1, 1]$. If a natural number n and a real a_1, \dots, a_n vary so that

$$\sum_{i=1}^n a_i^2 = 3 \text{ and } \max_{i=1}^n |a_i| \rightarrow 0 \quad (1)$$

(whence $n \rightarrow \infty$), then (say) by the Berry–Esseen inequality,

$$S_n := \sum_{i=1}^n a_i U_i$$

converges to a standard normal rv Z in distribution.

The closeness of the distribution of S_n to normality can be reasonably measured in an infinite variety of ways. One of them, in view of the Esseen smoothing inequality, is to consider the closeness of the characteristic function f_n of S_n to the characteristic function f of Z in a neighbourhood of 0. Given (1), we have

$$\ln f_n(t) - \ln f(t) = \sum_{i=1}^n \ln \frac{\sin(a_i t)}{a_i t} - \frac{t^2}{2} \sim \frac{t^4}{180} \sum_{i=1}^n a_i^4$$

uniformly over all t in any given neighbourhood of 0; here we use the asymptotic expansion

$$\ln \frac{\sin(at)}{at} - \left(\frac{-a^2 t^2}{6} \right) \sim \frac{a^4 t^4}{180}$$

for $a \rightarrow 0$ and t in any given neighbourhood of 0.

So, the closeness of the distribution of S_n to normality can be measured by

$$\sum_{i=1}^n a_i^4$$

which attains its minimum given the first condition in (1) when the a_i^2 's are the same for all $i = 1, \dots, n$, that is, when the random variables $a_i U_i$ are identically distributed [13].

Bibliography

- [1] C. E. Shannon, “A mathematical theory of communication,” *Bell System Tech. J.*, vol. 27, pp. 379-423, 623-656, July-Oct. 1948.
- [2] C. E. Shannon, “Two-way Communication Channels,” *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, vol. I, pp. 611-644, University of California Press, Berkeley, 1961.
- [3] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
- [4] E. T. Jaynes, “Information Theory and Statistical Mechanics”, *In K. Ford. Statistical Physics*, Benjamin, New York, p. 181, 1963.
- [5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd Edition, ISBN: 978-0-471-24195-9.
- [6] G. Ungerboeck, “Channel coding with multilevel/phase signals,” in *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 55-67, Jan 1982.
- [7] L. H. Ozarow and A. D. Wyner, “On the capacity of the Gaussian channel with a finite number of input levels,” in *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1426-1428, Nov 1990.
- [8] P. Billingsley, *Probability and Measure*, 3rd Edition, John Wiley & Sons, 1995.
- [9] G. D. Forney and G. Ungerboeck, “Modulation and coding for linear Gaussian channels,” in *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2384-2415, Oct 1998.
- [10] A. J. Viterbi, “Very low rate convolution codes for maximum theoretical performance of spread-spectrum multiple-access channels,” in *IEEE Journal on Selected Areas in Communications*, vol. 8, no. 4, pp. 641-649, May 1990.
- [11] Long Duan, B. Rimoldi and R. Urbanke, “Approaching the AWGN channel capacity without active shaping,” *Proceedings of IEEE International Symposium on Information Theory*, Ulm, 1997, pp. 374.

- [12] E. Abbe and A. Barron, “Polar coding schemes for the AWGN channel,” *2011 IEEE International Symposium on Information Theory Proceedings*, St. Petersburg, 2011, pp. 194-198.
- [13] Iosif Pinelis, Fastest convergence of sum of uniform independent distributions to a Gaussian, <https://mathoverflow.net/q/302174>.
- [14] James E. Marengo, David L. Farnsworth, and Lucas Stefanic, “A Geometric Derivation of the Irwin-Hall Distribution,” *International Journal of Mathematics and Mathematical Sciences*, vol. 2017, Article ID 3571419, 6 pages, 2017.
- [15] E. Arıkan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” *IEEE Trans. Inform. Theory*, vol. IT-55, pp. 3051-3073, July 2009.
- [16] E. Şaşıođlu, E. Telatar, and E. Arıkan, “Polarization for arbitrary discrete memoryless channels,” in *Proc. 2009 IEEE Inform. Theory Workshop*, Taormina, Italy, Oct. 2009, pp. 144-148.
- [17] N.L. Johnson, S. Kotz, N. Balakrishnan, *Continuous Univariate Distributions*, Volume 2, 2nd Edition, Wiley, ISBN 0-471-58494-0, 1995.
- [18] I. Tal and A. Vardy, “How to Construct Polar Codes,” *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6562-6582, Oct. 2013.