TEL AVIV UNIVERSITY אוניברסיטת תל-אביב
FACULTY OF ENGINEERING הפאקולטה להנדסה
The Iby and Aladar Fleischman Faculty of Engineering
The Zandman-Slaner School of Graduate Studies
The Department of Electrical Engineering - Systems

# The Robustness of Dirty Paper Coding and The Binary Dirty Multiple Access Channel with Common Interference

Thesis submitted toward the degree of
Master of Science in Electrical and Electronic Engineering

by

**Anatoly Khina**

April, 2010

TEL AVIV UNIVERSITY אוניברסיטת תל-אביב
FACULTY OF ENGINEERING הפאקולטה להנדסה
The Iby and Aladar Fleischman Faculty of Engineering
The Zandman-Slaner School of Graduate Studies
The Department of Electrical Engineering - Systems

# The Robustness of Dirty Paper Coding and The Binary Dirty Multiple Access Channel with Common Interference

Thesis submitted toward the degree of
Master of Science in Electrical and Electronic Engineering

by

**Anatoly Khina**

This research was carried out at the
Department of Electrical Engineering - Systems,
Tel-Aviv University

Advisor:
**Dr. Uri Erez**
April, 2010

*"The least initial deviation from the truth*
*is multiplied later a thousandfold."*
Aristotle

**Acknowledgements**

I wish to express my utmost appreciation and gratitude to Dr. Uri Erez, who took me under his wing as a third year undergraduate student, and helped me making my first steps in the exciting, and to me, new world of information theory and communication. For his professional guidance and dedicated supervising, which shaped the way I approach and handle new theoretical problems. And for his patience and invaluable advice.

I thank Prof. Ram Zamir for serving as a true "academic grandfather" to me. I first met Rami in the undergraduate course "random signals and noise", which fascinated me and convinced me to pursue in this direction. Later on, his advanced information theory course gave me most of the basic tools I needed as a young researcher in the area of information theory, at the end of which I was able to conduct my research.

In the course of this work, I had the privilege of working together with Yuval Kochman and Tal Philosof on things within and outside the scope of this work. Yuval always sees things in a unique way, which frequently makes things look much simpler and their solutions - "natural". Tal has a broad vision and vast knowledge and understanding of both, theoretical and practical, aspects of communication systems. I learned a lot from both of them and for that I am grateful.

I would like to thank my other colleagues from 102 and 108 labs, as well, who made this period enjoyable and full of interesting interactions, both in academic and non-academic issues: Amir Alfandary, Ohad Barak, Ohad Ben-Cohen, Idan Goldenberg, Eli Haim, Amir Ingber, Roy Jevnisek, Oron Levy, Yuval Lomnitz, Eado Meron, Noam Presman, Ofer Shayevitz, Mikhal Shemer, Alba Sloin, Nir Weinberger, Yair Yona and Pia Zobel.

I would like to acknowledge the support of the Israeli ministry of trade and commerce as part of Nehusha/iSMART project and the Yitzhak and Chaya Weinstein Research Institute for Signal Processing at Tel Aviv University.

Finally warm thanks to my parents Tatyana and Alexander, for their endless support and caring.

**Abstract**

A dirty-paper channel is considered, where the transmitter knows the interference sequence up to a constant multiplicative factor, known only to the receiver. Lower bounds on the achievable rate of communication are derived by proposing a coding scheme that partially compensates for the imprecise channel knowledge. We focus on a communication scenario where the signal-to-noise ratio is high. Our approach is based on analyzing the performance achievable using lattice-based coding schemes. When the power of the interference is finite, we show that the achievable rate of this lattice-based coding scheme may be improved by a judicious choice of the scaling parameter at the receiver. We further show that the communication rate may be improved, for finite as well as infinite interference power, by allowing randomized scaling at the transmitter. This scheme and its analysis are used to compare the performance of linear and dirty paper coding transmission techniques over the MIMO broadcast channel, in the presence of channel uncertainty.

We also consider a binary dirty multiple-access channel with interference known at both encoders. We derive an achievable rate region for this channel which contains the sum-rate capacity and observe that the sum-rate capacity in this setup coincides with the capacity of the channel when full-cooperation is allowed between transmitters, contrary to the analogous Gaussian case.

# Nomenclature

| | |
|---|---|
| AWGN | Additive White Gaussian Noise |
| BC | Broadcast |
| BSC | Binary Symmetric Channel |
| DMAC | Dirty Multiple-Access Channel |
| DMC | Discrete Memoryless Channel |
| DP | Dirty Paper |
| DPC | Dirty Paper Coding |
| MAC | Multiple-Access Channel |
| MIMO | Multiple-Input Multiple-Output |
| MMSE | Minimum Mean-Square Error |
| MSE | Mean-Square Error |
| SI | Side Information |
| SIR | Signal-to-Interference Ratio |
| SNR | Signal-to-Noise Ratio |
| THP | Tomlinson-Harashima Precoding |
| ZF | Zero-Forcing |
| $\boldsymbol{a}$ | vector |
| $a_1^n$ | $a_1, ..., a_n$ |
| $\|\|\boldsymbol{a}\|\|$ | The $L2$-norm of $\boldsymbol{a}$ |
| $X$ | Random variable |
| $\mathcal{X}$ | The alphabet of the random variable $X$ |
| $\|\mathcal{X}\|$ | The Cardinality of the alphabet $\mathcal{X}$ |
| $\boldsymbol{X}$ | Random vector |
| $p(x)$ | Probability density function |
| $p(x, y)$ | Joint probability density function |
| $p(y\|x)$ | Conditional probability density function |

| | |
|---|---|
| $EX$ | The expectation of $X$ |
| Bernoulli$(p)$ | Bernoulli distribution with parameter $p$ |
| Unif $(\mathcal{R})$ | Uniform distribution over region $\mathcal{R}$ |
| $\mathcal{N}(\mu, \sigma^2)$ | Gaussian distribution with expectation $\mu$ and variance $\sigma^2$ |
| $H(X)$ | The entropy of a discrete random variable $X$ |
| $h(X)$ | The differential entropy of a continuous random variable $X$ |
| $H_b(p)$ | The entropy of a binary random variable $X \sim$ Bernoulli$(p)$ |
| $H_b^+(p)$ | $H_b\left(\min\left\{p, \frac{1}{2}\right\}\right)$ |
| $I(X;Y)$ | The mutual information of two random variables $X, Y$ |
| cl conv$\{\mathcal{R}\}$ | Closure and Convex hull of the region $\mathcal{R}$ |
| u.c.e{f(x)} | The upper convex envelope of $f(x)$ w.r.t $x$ |
| $\oplus$ | Modulo-two addition |
| $w_H$ | Hamming weight |
| $\mathbb{R}$ | The set of real numbers |
| $\mathbb{Z}$ | The set of integer numbers |
| $\mathbb{Z}_2$ | Galois field of size 2 |
| mod 2 | modulo-2 operation |
| mod $\Lambda$ | modulo lattice $\Lambda$ operation |
| $q_1 \circledast q_2$ | $(1-q_1)q_2 + q_1(1-q_2)$ |
| $\|\| \cdot \|\|$ | Euclidean norm |
| $\langle \cdot, \cdot \rangle$ | Euclidean inner-product |
| $\csc(x)$ | $1/\sin(x)$ |
| $\sec(x)$ | $1/\cos(x)$ |

# Contents

# List of Figures

# Chapter 1

# Introduction

## 1.1 Dirty Paper Coding Robustness

The dirty-paper (DP) channel, first introduced by Costa [11], provides an information theoretic framework for the study of interference cancellation techniques for interference known to the transmitter. The DP channel model has since been further studied and applied to different communication scenarios such as ISI channels (see, e.g., [30]), the MIMO Gaussian broadcast channel [6, 51, 40, 46] and information embedding [2]. The DP model, given by,

$$Y = X + S + N \,, \tag{1.1}$$

is composed of an input signal $X$, subject to a power constraint, corrupted by additive white Gaussian noise (AWGN) $N$ and additive interference $S$ which is known to the transmitter but not to the receiver, causally ("causal DP")

$$x_i = \phi(w, s_0, ..., s_i) \,,$$

or non-causally ("non-causal DP")

$$x_i = \phi(w, s_1, ..., s_n) \,,$$

where $w$ is the transmitted message, $\phi$ is a function satisfying the input constraint and $x_i$ and $s_i$ are the channel input and interference at time instance $i$ ($1 \leq i \leq n$), respectively.

Costa [11] showed that, for an i.i.d. Gaussian interference with arbitrary power, the capacity in the non-causal scenario is equal to that of the interference-free AWGN channel. This result was extended in [9] to the case of general ergodic interference and to arbitrary interference in [17].

The capacity of the DP channel with causal knowledge of the interference, first considered by Willems [47], is not known but upper and lower bounds for the case of arbitrary interference were found in [17], which coincide in the high SNR regime, thus establishing the capacity for this case to be the same as for the interference-free AWGN channel (or equivalently for the non-causal DP channel) up to a shaping loss. Thus, causality incurs a rate loss of $\frac{1}{2}\log(\frac{2\pi e}{12})$, relative to the capacity of the interference-free AWGN channel, in the high SNR regime. This result implies that in the limit of strong interference and high SNR, the well-known Tomlinson-Harashima precoding (THP) technique [45, 22] is optimal. For general SNRs, the lattice-based coding techniques of [7, 15, 17] are an extension of Tomlinson-Harashima precoding, sometimes referred to as MMSE (minimum mean-square error) Tomlinson-Harashima precoding, where a scaling parameter is introduced at the transmitter and receiver. In this thesis the term Tomlinson-Harashima precoding is used in this wider sense.

The causal and non-causal DP channels are special cases of the problem of a general state-dependent memoryless channel. This problem was first introduced by Shannon in 1958 [42], who found the capacity for the case of a causally known state. Kuznetzov and Tsybakov considered the non-causal scenario [29], the general capacity of which was found by Gel'fand and Pinsker in 1980 [19].

We further note that these channels model communication scenarios where the channel (i.e., all channel coefficients) is known *perfectly* to both, the transmitter and the receiver.

In this thesis we focus our attention on scalar precoding, both since it results in simpler coding schemes but also since the benefit of using a vector approach (at least using the methods we study) diminishes in the presence of imprecise channel knowledge, as will be shown in the sequel. Note that scalar precoding is applicable when the interference is known causally ("Shannon scenario"), whereas vector approaches require non-causal knowledge ("Ge'fand-Pinsker scenario"). See, e.g., [17].

In many cases of interest, the transmitter has imprecise channel knowledge.

For instance in a multi-user broadcast scenario, the interference sequence $S$ corresponds to the signal intended to another user *multiplied* by a channel gain. While the transmitter knows the transmitted interfering signal, only an estimate of the channel gain is known (for instance by quantized feedback; see, e.g., [24]). This leads to the question, studied in this work, of how sensitive dirty paper coding (DPC) is to imprecise channel knowledge. We address this question by adapting the extended Tomlinson-Harashima precoding, as presented in [17], to the case of imprecise channel knowledge. We consider the *real* channel case; for treatment of the case of imperfect phase knowledge, in the *complex* channel case, see [21, 3].

Caire and Shamai [6] and Weingarten, Steinberg and Shamai [46] showed that the private-message capacity of the Gaussian MIMO broadcast (BC) channel can be achieved using DPC. Nonetheless, it has been speculated in some works, e.g., [50, 8, 5], that DPC has a significant drawback in the presence of channel estimation errors, compared to linear approaches such as linear ZF. In this work, we analyze the performance of both linear ZF and DPC for the 2-user MIMO BC channel and observe that such claims are unqualified.

For the performance analysis of this scheme, we note that the DP channel with imprecise channel knowledge problem is a special case of the compound channel with side information at the transmitter problem, first introduced by Mitran, Devroye and Tarokh [33], which generalizes both the state-dependent memoryless channel problem and the compound channel problem, considered in several works [4, 14, 48]. Mitran, Devroye and Tarokh considered the non-causal scenario, for which they were able to derive upper and lower bounds, following the steps of Gel'fand-Pinsker [19] and adjusting their proof to the compound case. Nevertheless, the lower and upper bounds of [33] do not coincide in general, and the capacity for the non-causal case is yet to be determined. Since, we focus mainly on the causal DP scenario, we consider the problem of the compound channel with side information known *causally* at the transmitter, and derive its *capacity*, by adjusting the proof by Shannon [42] to the compound case.

## 1.2 Binary Dirty MAC with Common Interference

One possible scenario, which generalizes the point-to-point channel with side information (SI) at the transmitter [42, 19] (and the classical multiple-access channel (MAC) [1, 32]), is the state-dependent multiple-access channel (MAC).

An important special case of this problem, called the "dirty" MAC in [34] (after Costa's "Writing on Dirty Paper" [11]), is the MAC with additive messages, interference and noise, where different parts of the interference are known to different users causally or non-causally. Interestingly, the dirty MAC (DMAC) appears to be a bottleneck in many wireless networks, ad hoc networks and relay problems.

Different efforts towards determining the capacity region of the DMAC were made. In [43, 23], extensions are derived for the achievable of Gel'fand-Pinsker [19] to the case of state-dependent MAC with different SI availability scenarios, and some outer bounds are established. Nevertheless, trying to extend the capacity-achieving auxiliary selection of Costa for the Gaussian DP channel problem, falls through, as discussed in [36, 35]. Trying to shed light on this problem, the binary modulo-additive DMAC is discussed in [37], where capacity regions are found for the cases of two independent interferences, each known at different transmitter ("doubly-dirty" MAC), and for the case where the interference is known only to one of the transmitters (DMAC with "single informed user"). [1] Also note that some of the results of [37] are given also in [28, 44].

Unlike in the Gaussian DP channel problem, for which Costa [11] showed that the capacity is equal to the AWGN channel, i.e., as if the interference $S$ were not present, in the binary modulo-additive case, this does not carry on to the binary case: the capacity of the binary dirty channel is strictly smaller than that of the interference-free channel [2, 53]. [2] Hence, in the various binary DMAC scenarios, rate loss, relative to the interference-free ("clean") MAC is inevitable in general, due to the presence of the interference.

[1] For this case, both the common message and the private message capacities are determined, unlike for the doubly-dirty MAC, for which only the private message capacity was given.

[2] Unless the noise is absent or if the problem is not constrained by power.

In the second part of this work we focus on the binary dirty MAC with non-causal common interference, as this problem has not yet fully treated. To this end, we examine the capacity region and different coding strategies for the binary clean MAC.

## 1.3 Thesis Organization

The thesis is organized as follows.

Chapter 1: In Chapter 1.1 and Chapter 1.2 a short introduction is given of the two parts of this work, resp., followed by a more comprehensive theoretical background in Chapter 1.4.

Chapter 2: In Chapter 2.1 we discuss the compound causal dirty-paper channel model. We then turn, In Chapter 2.2, to the more general problem of the compound state-dependent discrete memoryless channel (DMC) and determine its capacity where the state is known causally. In Chapter 2.3 we consider the case where the interference $S$ is i.i.d. (of some distribution) with power $P_S$, and show how using a modified front-end can outperform the regular DP channel receiver, which ignores the inaccuracy in the channel knowledge. We then concentrate on the high SNR regime and show that using *random* scaling improves performance further, in Chapter 2.4. In Chapter 2.5, we discuss the extension of the scheme to the non-causal case, as well as presenting its implications to multiple-input multiple-output (MIMO) broadcast channels with imperfect channel knowledge at the transmitter in Chapter 2.6.

Chapter 3: In Chapter 3.1 we discuss the binary "dirty" MAC with common interference model. In Chapter 3.2 we discuss the clean binary MAC, followed by the treatment of the binary dirty MAC in Chapter 3.3.

Chapter 4: Summary of the main results.

## 1.4 Background

### 1.4.1 Channels with Side Information at the Transmitter

The problem of a state-dependent channel, where the state is known only to the transmitter ("SI"), depicted in Figure 1.1, was first introduced by Shannon [42], who considered a DMC whose transition matrix depends on the channel state $s$, where the latter is independent of the message $W$ that is sent, i.i.d. and known *causally* to the transmitter but not to the receiver. This channel is described by

$$p(\boldsymbol{y}|\boldsymbol{s},\boldsymbol{x}) = \prod_i p(y_i|s_i,x_i)$$
$$p(\boldsymbol{s}) = \prod_i p(s_i),$$

where $s \in \mathcal{S}$, $x \in \mathcal{X}$ is the channel input, $y \in \mathcal{Y}$ is the channel output; and $\mathcal{X}, \mathcal{Y}$ and $\mathcal{S}$ denote the channel input alphabet, channel output alphabet and state alphabet, respectively, all of which are finite sets. Shannon showed that the capacity of the above channel is equal to that of an equivalent derived DMC whose inputs are mappings $t \in \mathcal{T}$, which will be referred to hereafter as strategies, from $\mathcal{S}$ to $\mathcal{X}$, where $\mathcal{T}$ denotes the set of all mappings from $\mathcal{S}$ to $\mathcal{X}$, and therefore is of cardinality $|\mathcal{T}| = |\mathcal{X}|^{|\mathcal{S}|}$. The corresponding derived transition probabilities of this channel are

$$p(y|t) = \sum_s p(s)p(y|x=t(s),s).$$

Note that this result uses mappings of the *current* state only, even though the transmitter has access to all past states.

Thus, the capacity of this channel is

$$C^{\mathrm{Causal}} = \max_{p(t)\in\mathcal{P}(\mathcal{T})} I(T;Y), \tag{1.2}$$

where $\mathcal{P}(\mathcal{T})$ is the set of all probability vectors over $\mathcal{T}$.

Gel'fand and Pinsker [19] showed, using random binning for the direct part,

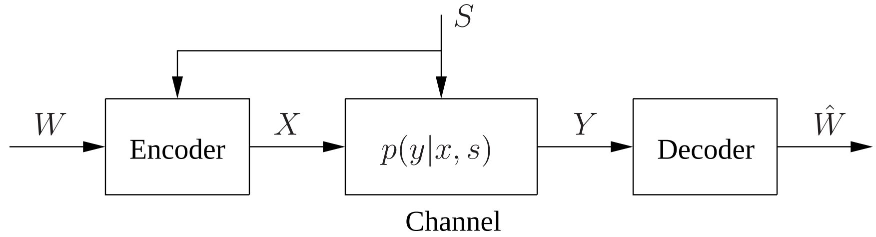


Figure 1.1: The discrete memoryless channel with SI at the transmitter.

that the capacity of the above channel, when the state $s$ is known *non-causally* to the transmitter, is given by

$$C^{\text{noncausal}} = \max_{p(u,x|s)} \{I(U;Y) - I(U;S)\}, \tag{1.3}$$

where the maximum is over all joint distributions of the form $p(s)p(u,x|s)p(y|x,s)$, and $U$ is an "auxiliary" random variable from a finite set, whose cardinality need not exceed $|\mathcal{U}| \leq |\mathcal{X}| + |\mathcal{S}|$.

Both of this results can be extended to continuous memoryless channels.

### 1.4.2 Writing on Dirty Paper

A well-known scenario of a channel with side information at the transmitter was introduced by Costa [11]. In this case, the channel, as shown in Figure 1.2, is given by

$$Y = X + S + N\,,$$

where $N \sim \mathcal{N}(0, P_N)$ is i.i.d Gaussian noise and $S \sim \mathcal{N}(0, P_S)$ is i.i.d Gaussian interference. The transmitter observes the interference *non-causally* and generates the transmitted codeword

$$\boldsymbol{x} = \phi(w, \boldsymbol{s})$$

where $\boldsymbol{x} = x_1^n$, $\boldsymbol{s} = s_1^n$ and $w \in \mathcal{W}$ is the transmitted message. The input is subject to a power constraint $\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P_X$. The receiver reconstructs the transmitted message using the following mapping

$$\hat{w} = \psi(\boldsymbol{y})\,,$$

where $\boldsymbol{y} = y_1^n$. The signal-to-noise ratio is defined as $\text{SNR} \triangleq P/N$.

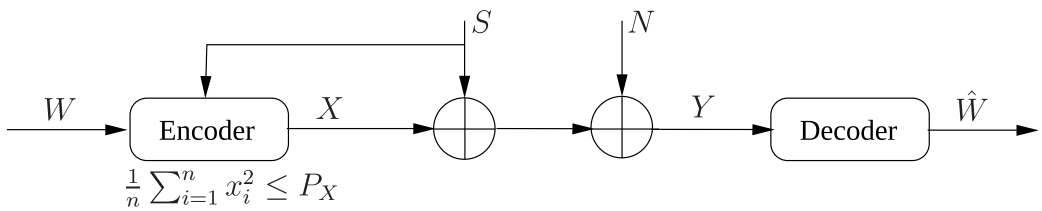


Figure 1.2: Dirty paper channel.

The proof of Costa [11] uses the general capacity formula derived by Gel'fand and Pinsker (1.3). Selecting the auxiliary random variable $U$ to be

$$U = X + \alpha S\,, \tag{1.4}$$

where $X \sim \mathcal{N}(0, P_X)$ and is independent of $S$, and taking $\alpha = \frac{\text{SNR}}{\text{SNR}+1}$ achieves the capacity of the interference-free AWGN channel which is given by

$$C = \frac{1}{2}\log\left(1 + \text{SNR}\right)\,.$$

Interestingly, the interference variance $P_S$ plays no role in the capacity expression. In later works it was shown that the capacity is the same for any ergodic interference distribution [9], and is in fact the same for arbitrary interference sequence, using lattice-based strategies [17].

The analogous causal side information scenario was first considered by Willems [47], who suggested to use a lattice-based scheme, which was later improved in [17], where also lower and upper bounds were given, that coincide in the limit of high SNR. As will be shown in the sequel in (1.10), in the limit of high SNR, the capacity of the causal DP channel is:

$$C^{\text{Causal}} = \frac{1}{2}\log\left(1 + \frac{P_X}{P_N}\right) - \frac{2\pi e}{12}\,,$$

for $P_X \to \infty$, i.e., in this limit there is a loss of $\frac{2\pi e}{12}$("shaping gain loss") due to causality.

The analogous binary setting is the "binary DP" channel, described by:

$$Y = X \oplus S \oplus N\,,$$

where $X, S, N \in \mathbb{Z}_2$ and $\oplus$ denotes addition mod 2 (XOR). The input constraint is $\frac{1}{n} w_H(\mathbf{x}) \leq q$, where $0 \leq q \leq 1/2$, $w_H(\cdot)$ denotes Hamming weight, and $n$ is the length of the codeword. The noise $N \sim \text{Bernoulli}(\varepsilon)$ is independent of $(S, X)$ (w.l.o.g. we assume $\varepsilon \leq \frac{1}{2}$) ; the state information ("interference") $S \sim \text{Bernoulli}\,(1/2)$ is known either causally or non-causally to the encoder.

The capacity of this binary DP channel with non-causal knowledge of the interference is equal to (see [2, 53]):

$$C_{\text{dirty}}^{\text{noncausal}} = \text{uch} \max \left\{ \mathrm{H_b(q)} - \mathrm{H_b}(\varepsilon), 0 \right\} , \tag{1.5}$$

where $H_b(\cdot)$ denotes the binary entropy [12] and uch is the upper convex hull operation with respect to $q$. Thus, unlike in the Gaussian setting, in the binary case the capacity of the dirty channel is *strictly lower* than that of the corresponding interference-free ("clean") channel,

$$C_{\text{clean}} = H_b(q \circledast \varepsilon) - H_b(\varepsilon) \,, \tag{1.6}$$

due to the binary convolution (denoted by $\circledast$) with $\varepsilon$ in the first element, which increases entropy, and is defined as:

$$q_1 \circledast q_2 \triangleq (1 - q_1) q_2 + q_1 (1 - q_2) \,.$$

The capacity of the causal binary DP channel can be easily derived from (1.2), and is equal to:

$$C_{\text{dirty}}^{\text{causal}} = 2q \left(1 - H_b(\varepsilon)\right) \,,$$

meaning that the best possible strategy, in the causal case, is to use all possible transmitting power to eliminate the interference $S$ from as much input symbols inside a block of size $n$ as possible ($2qn$ slots on average) and to send information over the binary symmetric channel (BSC) with error probability $\varepsilon$, that was obtained for these symbols.

### 1.4.3 Lattice-Strategies

#### Preliminary: Lattices

An $n$-dimensional lattice $\Lambda$ is a discrete group in the Euclidian space $\mathbb{R}^n$ which is closed with respect to the addition operation (over $\mathbb{R}$) [10]. The lattice is specified by

$$\Lambda = \{\lambda = G\boldsymbol{i} \ : \ \boldsymbol{i} \in \mathbb{Z}^n\},$$

where $G$ is an $n \times n$ real valued matrix, called the generator matrix of the lattice (whose choice is not unique). A coset of the lattice is any translation of the original lattice $\Lambda$, i.e., $\boldsymbol{a} + \Lambda$ where $\boldsymbol{a} \in \mathbb{R}^n$.

The nearest neighbor quantizer $Q_\Lambda(\cdot)$ associated with $\Lambda$ is defined by

$$Q_\Lambda(\boldsymbol{x}) = \lambda \in \Lambda \quad \text{if } ||\boldsymbol{x} - \lambda|| \leq ||\boldsymbol{x} - \lambda'||, \quad \forall\ \lambda' \in \Lambda,$$

where $||\cdot||$ denotes Euclidian norm. The Voronoi region associated with a lattice point $\lambda$ is the set of all points in $\mathbb{R}^n$ that are closer (in Euclidian distance) to $\lambda$ than to any other lattice point. Specifically, the fundamental Voronoi region is defined as the set of all points that are closest to the origin

$$\mathcal{V}_0 \triangleq \{\boldsymbol{x} \in \mathbb{R}^n : Q_\Lambda(\boldsymbol{x}) = \boldsymbol{0}\},$$

where ties are broken arbitrarily. The modulo lattice operation with respect to $\Lambda$ is defined as

$$\boldsymbol{x} \bmod \Lambda = \boldsymbol{x} - Q_\Lambda(\boldsymbol{x}).$$

This operation satisfies the following distributive property

$$[\boldsymbol{x} \bmod \Lambda + \boldsymbol{y}] \bmod \Lambda = [\boldsymbol{x} + \boldsymbol{y}] \bmod \Lambda.$$

The second moment of a lattice $\Lambda$ is given by

$$\sigma_\Lambda^2 \triangleq \frac{\frac{1}{n}\int_{\mathcal{V}_0} ||\boldsymbol{x}||^2 \mathbf{dx}}{V},$$

where $V$ is the volume of the fundamental Voronoi region, i.e., $V = \int_{\mathcal{V}_0} \mathbf{dx}$ (the same for all Voronoi regions of $\Lambda$). The normalized second moment is given by

$$G(\Lambda) \triangleq \frac{\sigma_\Lambda^2}{V^{2/n}}.$$

The normalized second moment is always greater than $1/2\pi e$ (see [52]). It is known [52] that for sufficiently large dimension there exist lattices that are good for quantization (these lattices are also known to be good for shaping [16]), in the sense that for any $\epsilon > 0$

$$\log(2\pi e G(\Lambda)) < \epsilon, \tag{1.7}$$

for large enough $n$. In addition, there exist lattices with second moment $P$ that are good for AWGN channel coding [16], satisfying

$$\Pr(\boldsymbol{X} \notin \mathcal{V}_0) < \epsilon, \text{ where } \boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, (P-\epsilon)I_n), \ \forall \epsilon > 0,$$

where $I_n$ is an $n \times n$ identity matrix.

The differential entropy of an $n$-dimensional random vector $\boldsymbol{U}$, which is distributed uniformly over the fundamental Voronoi cell, i.e., $\boldsymbol{U} \sim \mathrm{Unif}(\mathcal{V}_0)$, is given by [52]

$$\begin{aligned} h(\boldsymbol{U}) &= \log(V) \\ &= \log\left(\frac{\sigma_\Lambda^2}{G(\Lambda)}\right)^{n/2} \\ &= \frac{n}{2}\log\left(\frac{\sigma_\Lambda^2}{G(\Lambda)}\right) \\ &\approx \frac{n}{2}\log\left(2\pi e \sigma_\Lambda^2\right), \end{aligned}$$

where the last (approximate) equality holds for lattices that are good for quantization, for large $n$. For one-dimensional lattices, i.e., $c\mathbb{Z}$, the differential entropy of $\boldsymbol{U}$ is equal to

$$h(\boldsymbol{U}) = \log(12\sigma_\Lambda^2).$$

### Lattice-Strategies for Cancelling Known Interference

The capacity of the dirty-paper channel (1.1) can also be achieved using a coding scheme based on *lattice-strategies* (also known as *lattice precoding*) [17, 53]. Specifically, let $\Lambda$ be an $n$-dimensional lattice with second moment $P_X$ that is good for quantization (1.7). The information bearing signal $\boldsymbol{V}$ is distributed uniformly over the basic cell of $\Lambda$, i.e., $\boldsymbol{V} \sim \mathrm{Unif}(\mathcal{V}_0)$. The transmission scheme is shown in Figure 1.3.

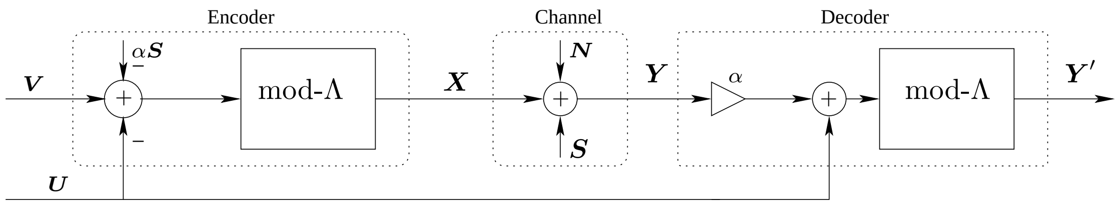


Figure 1.3: Lattice-strategies transmission scheme.

- *Transmitter*: The transmitter output is the error vector between $\boldsymbol{V}$ and $\alpha\boldsymbol{S} + \boldsymbol{U}$, i.e.,

$$\boldsymbol{X} = \Big[\boldsymbol{V} - \alpha\boldsymbol{S} - \boldsymbol{U}\Big] \bmod \Lambda,$$

  where $\boldsymbol{U} \sim \mathrm{Unif}(\mathcal{V}_0)$ is common randomness ("dither") which is known to both, the transmitter and the receiver. From the dither property [52], $\boldsymbol{X} \sim \mathrm{Unif}(\mathcal{V}_0)$ (and is independent of $\boldsymbol{V}$), and hence the power constraint is satisfied.

- *Receiver*: The channel output $\boldsymbol{Y}$ is multiplied by $\alpha$, followed by the dither addition modulo-$\Lambda$, i.e.,

$$\boldsymbol{Y}' = [\alpha\boldsymbol{Y} + \boldsymbol{U}] \bmod \Lambda.$$

Erez, Shamai and Zamir showed in [17] that the equivalent channel is an interference-free modulo-$\Lambda$ channel, i.e.,

$$\boldsymbol{Y}' = \Big[\boldsymbol{V} + \boldsymbol{N}_{\mathrm{eff}}\Big] \bmod \Lambda, \tag{1.8}$$

where $N_{\text{eff}}$ is the *effective noise* which is given by

$$\boldsymbol{N}_{\text{eff}} = -(1-\alpha)\boldsymbol{X} + \alpha\boldsymbol{N}\,, \tag{1.9}$$

and is independent of $\boldsymbol{V}$ since $\boldsymbol{X}$ is independent of $\boldsymbol{V}$ due to the dither and $(\boldsymbol{X}, \boldsymbol{V})$ are independent of $\boldsymbol{N}$. Moreover, $\boldsymbol{X}$ and $\boldsymbol{U}$ have the same distribution, and hence the effective noise of (1.9) is equivalent, in distribution, to

$$\boldsymbol{N}_{\text{eff}} = (1-\alpha)\boldsymbol{U} + \alpha\boldsymbol{N}\,.$$

For $\alpha = 1$, the *interference concentration* is reflected in the above modulo-$\Lambda$ equivalent channel (1.8) and (1.9). That is, the residual interference at the decoder is concentrated on discrete values (due to the modulo operation), which are the lattice points of $\Lambda$. Nevertheless, this is not the optimal selection of $\alpha$. A better selection is the one that minimize the power of $N_{\text{eff}}$, i.e., $\alpha = \frac{\text{SNR}}{\text{SNR}+1}$ (which is exactly the choice of $\alpha$ in the Costa scheme), which allows to achieve all rates satisfying:

$$R \leq \frac{1}{2}\log\left(1+\text{SNR}\right) - \frac{1}{2}\log\left(2\pi e G(\Lambda)\right)\,. \tag{1.10}$$

Taking a sequence of lattices $\Lambda$, with increasing dimension, that are good for quantization (1.7) ($G(\Lambda) \to \frac{1}{2\pi e}$), it follows that one can achieve rates approaching $\frac{1}{2}\log\left(1+\text{SNR}\right)$, i.e., the capacity of an interference-free AWGN channel [17, 53]. Nevertheless this is possible only when the interference is known non-causally. In case, the interference is known only causally, one cannot anticipate the interference of future symbols, and is limited to one-dimensional ("scalar") lattice strategies. For this channels only rates satisfying

$$R \leq \frac{1}{2}\log\left(1+\text{SNR}\right) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right)\,,$$

can be achieved, using this strategy, as $G(\Lambda) = 1/12$ for such lattices. Note that the one-dimensional lattice scheme can be seen as an extension of the intersymbol interference (ISI) cancellation scheme suggested independently by Tomlinson [45] and Harashima [22]. Hence, we shall refer to this scheme and its extensions as Tomlinson-Harashima precoding (THP).

### 1.4.4 Compound Channels

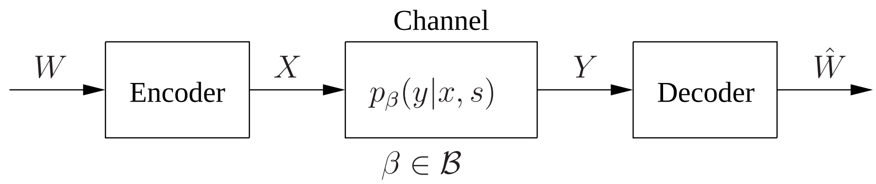


Figure 1.4: The discrete compound memoryless channel.

A discrete memoryless compound channel is a channel whose transition matrix depends on a parameter $\beta$, which is *constant* and *not known* to the transmitter but is known to the receiver[3] and takes values from $\mathcal{B}$, where the alphabet $\mathcal{B}$ is a finite set. See Figure 1.4.

The ("worst-case") capacity of this channel was found, by several different authors [4, 14, 48] (see also [49]), to be

$$C = \max_{p(x)\in\mathcal{P}(X)} \inf_{\beta\in\mathcal{B}} I_\beta(X;Y),$$

where $I_\beta(X;Y)$ denotes the mutual information of $X$ and $Y$ with respect to the transition matrix $p_\beta(y|x)$ and $\mathcal{P}(\mathcal{X})$ is the set of all probability vectors over $\mathcal{X}$.

This result can be extended to continuous memoryless channels, as well.

### 1.4.5 Compound Channels With SI at the Transmitter

The generalization of the two problems of Chapter 1.4.1 and Chapter 1.4.4 is that of a discrete-memoryless compound state-dependent channel, where the state is available (as SI) at the transmitter, depicted in Figure 1.5.

This problem was treated, for the non-causal case, by Mitran, Devroye and Tarokh in [33], where they extended the proof of Gel'fand and Pinsker [19] to the compound case, but due to the presence of the channel outputs $Y_1^i$ in the auxiliary variable $U$ in the converse part, their achievable rate and upper bound do not coincide in general, and thus they were only able to derive inner

[3]Sometimes a channel is said to be compound if $\beta$ is not known at *both* ends. The capacity however is the same in both scenarios (see, e.g., [49, chap. 4]), as the receiver may estimate $\beta$ to within any desired accuracy (with probability going to one), using a negligible portion of the block length.

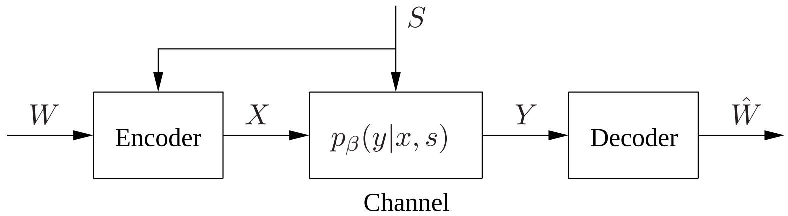


Figure 1.5: The compound discrete memoryless channel with SI at the transmitter.

and outer bounds on the capacity:

$$\begin{aligned}
C_l &\leq C \leq C_u \\
C_l &= \sup_{p(u|x,s,w)p(x|s,w)p(w)} \inf_{\beta\in\mathcal{B}} \left[I_\beta(U;Y|W) - I(U;S|W)\right] \\
C_u &= \sup_{p_\beta(u|x,s,w)p(x|s,w)p(w)} \inf_{\beta\in\mathcal{B}} \left[I_\beta(U;Y|W) - I(U;S|W)\right] ,
\end{aligned}$$

where the suprema are over all finite alphabet auxiliary random variables $U$ and finite alphabet time-sharing random variables $W$, and $\{p_\beta(u|x,s,w)\}$ denotes any family of distributions, where a distribution $p(u|x,s,w)$ is chosen for each value of $\beta$ before the infimum over $\beta$ is computed.

The authors of [33] extended these bounds to continuous alphabets and considered the following compound version of the DP channel:

$$Y = \beta_1 X + \beta_2 S + N \,, \tag{1.11}$$

where the interference sequence $S$ is known non-causally and the compound channel parameter is $\beta = (\beta_1, \beta_2)$. They suggest using the same auxiliary variable that was used by Costa for the non-compound case (given in (1.4)) and derive lower and upper bounds on the performance for this choice.

### 1.4.6 Gaussian MIMO Broadcast Channels

The general $K$-user real-valued [4] multiple-input multiple-output (MIMO) channel with $M$ antennas at the transmitter and $N$ antennas at each receiver, is

[4] The complex case is defined in a similar manner. See, e.g., [6, 51].

defined by

$$\boldsymbol{Y}_k = H_k \boldsymbol{X} + \boldsymbol{N}_k, \qquad k = 1, ..., K\,,$$

where $H_k \in \mathbb{R}^{N\times M}$ is the channel gain matrix of user $k$, $\boldsymbol{X}$ is the transmit signal vector, subject to some power constraint (depending on the scenario of interest and $\boldsymbol{N}_k$ is a Gaussian noise vector, which w.l.o.g. has zero-mean and identity covariance matrix $I_k$. [5]

Different scenarios for this channel were considered. We shall focus our interest on the private-message scenario, in which a different ("private") message needs to be conveyed to each of the users (in contrast to the common message scenario, in which the same message is transmitted to all users), and the power allocated to each of the messages is $P_i$.

For this scenario, different transmission schemes were proposed, the two most prominent being the linear transmission schemes and the ones that use DPC (see, e.g., [6]).

To further simplify the setting and give a geometrical view of this problem, we shall consider only the 2-user case with $K_t = 2$ transmit antennas and $K_r = 1$ receive antenna at each receiver:

$$Y_i = \boldsymbol{h}_i^T \boldsymbol{X} + N_i, \qquad i = 1, 2 \tag{1.12}$$

where $\boldsymbol{X}$ and $\boldsymbol{h}_i$ are $2 \times 1$ vectors.

Hence, for linear zero-forcing (ZF) or linear MMSE, as well as for DPC based schemes, the transmitted signal can be decomposed into a sum of the two message signals, meant for both users:

$$\begin{aligned}\boldsymbol{x} &= \boldsymbol{x}_1 + \boldsymbol{x}_2,\\ \boldsymbol{x}_i &= x_i \boldsymbol{t}_i\end{aligned}$$

where $x_i$ is the scalar information signal (can take, both, positive and negative values) intended for user $i$ of average power $P_i$, and $\boldsymbol{t}_i$ is a unit vector in the direction of the transmitted direction of this information signal. Without loss of generality, we shall assume that $P_2 \geq P_1$, and define $\mathrm{SNR}_i = P_i$ $(i = 1, 2)$.

[5] Otherwise the receiver can subtract the noise mean vector from the channel output and multiply the result by a whitening-matrix.

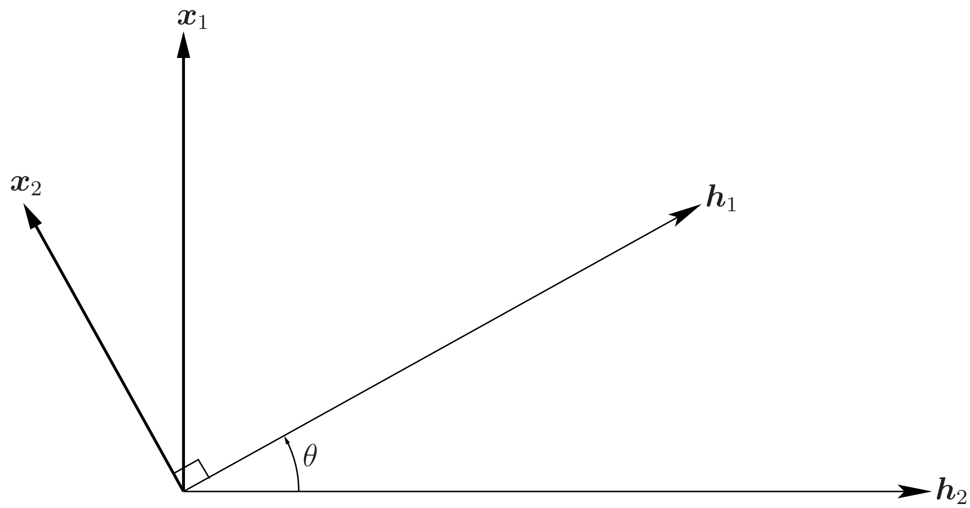


Figure 1.6: Pictorial representation of the zero-forcing technique for the MIMO broadcast channel.

In a similar way we shall rewrite the channel vectors $\boldsymbol{h}_i$ in the form

$$\boldsymbol{h}_i = h_i \boldsymbol{e}_i,$$

where $h_i$ is the signed-amplitude and $\boldsymbol{e}_i$ is a unit vector in the direction of $\boldsymbol{h}_i$. Let us denote the acute angle between $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ by $\theta$ (see Figure 1.6):

$$\theta \triangleq \min\left\{ \arccos\big(\langle \boldsymbol{e}_1, \boldsymbol{e}_2\rangle\big), \arccos\big( - \langle \boldsymbol{e}_1, \boldsymbol{e}_2\rangle \big) \right\},$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner-product in $\mathbb{R}^2$, and rewrite (1.12) as:

$$\begin{aligned} Y_1 &= \langle \boldsymbol{h}_1, \boldsymbol{X}_1 \rangle + \langle \boldsymbol{h}_1, \boldsymbol{x}_2 \rangle + N_1 \\ Y_2 &= \langle \boldsymbol{h}_2, \boldsymbol{X}_1 \rangle + \langle \boldsymbol{h}_2, \boldsymbol{x}_2 \rangle + N_2 \,. \end{aligned} \tag{1.13}$$

We focus our attention to the high SNR regime.

**Linear Zero-Forcing**

According to this strategy, the transmitter avoids interferences by transmitting $\boldsymbol{x}_1$ in an orthogonal direction to $\boldsymbol{h}_2$, and $\boldsymbol{x}_2$ - orthogonally to $\boldsymbol{h}_1$, as depicted in Figure 1.6 (see, e.g., [31]).

Hence, we may rewrite the channel outputs (1.13) as:[6]

$$
\begin{aligned}
Y_i &= \langle \boldsymbol{h}_i, \boldsymbol{X}_i \rangle + N_i \\
&= X_i h_i \cos\left(\frac{\pi}{2} - \theta\right) + N_i \\
&= X_i h_i \sin(\theta) + N_i, \qquad\qquad i = 1, 2.
\end{aligned}
$$

Note that this approach provides, effectively, two *parallel* channels. Finally, using codebooks generated in an i.i.d. Gaussian manner (with mean 0 and variance $P_i$), the following rates are achieved:

$$
R_i = I(X_i\,; Y_i) = \frac{1}{2}\log\left(1 + \mathrm{SNR}_i h_i^2 \sin^2(\theta)\right) \qquad i = 1, 2\,. \tag{1.14}
$$

**Zero-Forcing Dirty Paper Coding**

Instead of using linear precoding approaches, one may transmit the message to user 1 in an orthogonal direction to the channel vector of user 2, and apply dirty paper coding to eliminate the interference of user 2 on its own channel vector. This way, user 2 is free of interferences from the signal of user 1 and can transmit its information signal in the best possible direction, i.e., $\boldsymbol{e}_2$ (see Figure 1.7), and by this outperform the rates achievable via linear schemes. The expressions we provide below are for the non-causal case, i.e., correspond to using multi-dimensional THP where the dimension goes to infinity.[7]

Without loss of generality, we take the user that performs DPC to be user 1, i.e., $\langle \boldsymbol{h}_2, x_1 \rangle = 0$. Thus,

$$
\begin{aligned}
Y_2 &= \langle \boldsymbol{h}_2, X_2 \rangle + N_2 \\
&= h_2 X_2 + N_2 \\
Y_1 &= \langle \boldsymbol{h}_1, X_1 \rangle + \langle \boldsymbol{h}_1, x_2 \rangle + N_1\,, \\
&= h_1 X_1 \sin(\theta) + h_1 X_2 \cos(\theta) + N_1\,.
\end{aligned} \tag{1.15}
$$

[6] This is true up to a possible additional phase of $p$ inside the cosine, which has no effect on the effective channel, since the receiver knows the channel

[7] The results for the causal case are identical up to a subtraction of the shaping loss $\frac{1}{2}\log\left(\frac{2\pi e}{12}\right)$.

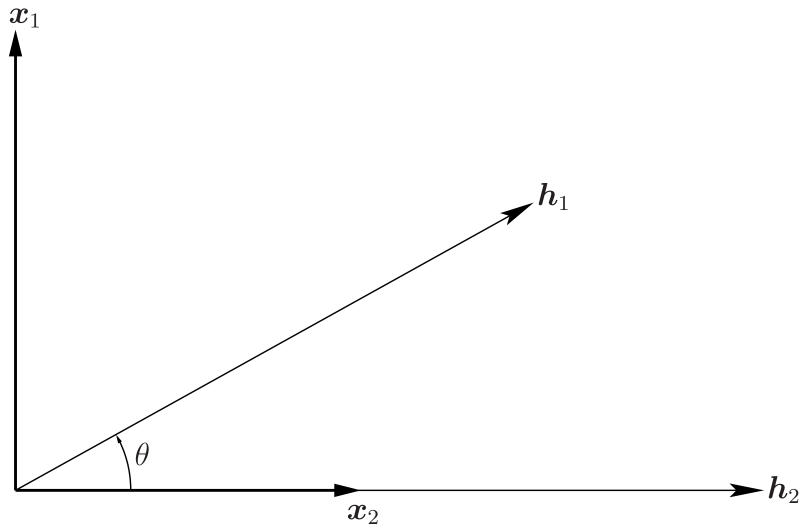


Figure 1.7: Pictorial representation of the ZF-DPC technique in the MIMO broadcast channel.

Dividing both sides of (1.15) by $h_1 \sin(\theta)$ gives rise to the equivalent channel

$$\tilde{Y}_1 = X_1 + X_2 \mathrm{ctg}(\theta) + \frac{1}{h_1 \sin(\theta)} N_1 \,.$$

Now, by using the dirty paper coding scheme of Chapter 1.4.3, user 1 can effectively eliminate the interference of user 2:

$$\begin{aligned} X_1 &= [v_1 - \alpha \cdot \mathrm{ctg}(\theta) X_2 - U] \bmod \Lambda \\ Y_1' &= [\alpha \tilde{Y}_1 + U] \bmod \Lambda \\ &= \left[ v_1 - (1-\alpha) X_1 + \frac{\alpha}{h_1 \sin(\theta)} N_2 \right] \bmod \Lambda \,, \end{aligned} \tag{1.16}$$

where $U$ is a dither distributed uniformly over the basic Voronoi cell $\mathcal{V}_0$ of the lattice $\Lambda$, whose second moment is set to be $P_1$. Finally, by setting the distributions of $V_1$ and $X_2$ to be uniform over $\mathcal{V}_0$ and Gaussian with power $P_2$, respectively, we obtain the following rates:

$$\begin{aligned} R_1 &= \frac{1}{2} \log \left( 1 + \mathrm{SNR}_1 h_1^2 \sin^2(\theta) \right) , \\ R_2 &= \frac{1}{2} \log \left( 1 + \mathrm{SNR}_2 h_2^2 \right) . \end{aligned}$$

Note that indeed, the rate of user 1 is the same as in (1.14), but the rate of user 2 has improved over the that of the linear ZF scheme.

**Remark 1.1.** Both the linear ZF and the ZF-DPC schemes can be improved by taking into account the noise power, rather than totaly eliminate the cross-interferences from the other user (see linear MMSE and MMSE-DPC in [6]). Nevertheless, when the SNRs are high the performance of the MMSE schemes coincide with those of the ZF ones.

### 1.4.7 Multiple-Access Channel

The multiple-access channel (MAC) problem was first considered by Shannon [41]. This problem consists of multiple distinct encoders who transmit private messages through a given channel (with the same number of inputs) to a single decoder, the aim of whom is to reconstruct the messages of all the encoders. The discrete memoryless multiple-access channel with $K$ users (encoders) is described by a channel matrix $p(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_K)$, with the memoryless property:

$$p\left(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_K\right) = \prod_i p\left(y_i|x_{1i}, x_{2i}, ..., x_{Ki}\right) ,$$

where $\{\boldsymbol{x}_k\}$ are the vectorial channel inputs and $\boldsymbol{y}$ is the channel output vector

Ahlswede [1] and Liao [32] found the capacity of this problem for two senders, to be:

$$\begin{aligned} \mathcal{C} \triangleq \text{cl conv} \Big\{ (\mathrm{R}_1, \mathrm{R}_2) :\ & \mathrm{R}_1 \leq I(X_1; Y|X_2) \\ & R_2 \leq I(X_2; Y|X_1) \\ R_1 + & R_2 \leq I(X_1, X_2; Y) \Big\}, \end{aligned} \tag{1.17}$$

where cl and conv are the closure and the convex hull operation, resp., over all admissible distributions of the form $p_1(x_1)p_2(x_2)$ on $\mathcal{X}_1 \times \mathcal{X}_2$.

This solution can be generalized to continuous alphabet channels. In the Gaussian additive MAC, any point within its capacity region can be achieved using *Gaussian stationary* inputs. Hence the convex hull operation is superfluous in the Gaussian case (see, e.g., [12]), and the capacity region is equal to:

$$\mathcal{C} \triangleq \Big\{(R_1, R_2): \; R_1 \leq \frac{1}{2}\log(1+\mathrm{SNR}_1)$$
$$R_2 \leq \frac{1}{2}\log(1+\mathrm{SNR}_2)$$
$$R_1 + R_2 \leq \frac{1}{2}\log(1+\mathrm{SNR}_1+\mathrm{SNR}_2)\Big\},$$

where $\mathrm{SNR}_1$ and $\mathrm{SNR}_2$ are the signal-to-noise ratios of users 1 and 2, respectively.

### 1.4.8 Dirty Multiple-Access Channel

Consider the two-user memoryless state-dependent multiple-access channel (MAC) with transition and state probability distributions

$$p(y|x_1, x_2, s) \;\text{ and }\; p(s),$$

where $s \in \mathcal{S}$ or parts of it are known causally or non-causally at one or both encoders. The channel inputs are $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, and the channel output is $y \in \mathcal{Y}$. The memoryless property of the channel implies that

$$p(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{s}) = \prod_{i=1}^{n} p(y_i|x_{1i}, x_{2i}, s_i).$$

Its capacity region is still not known in general, for the different SI scenarios, and remains an open problem. See, e.g., [36].

This model can be seen as a generalization of the point-to-point with SI at the transmitter, described in Chapter 1.4.1. Trying to generalize the random binning scheme of Gel'fand and Pinsker provides the achievable region (see, e.g., [36]):

$$\mathcal{R} \triangleq \mathrm{cl\ conv}\Big\{(\mathrm{R}_1, \mathrm{R}_2): R_1 \leq I(U;Y|V) - I(U;S|V)$$
$$R_2 \leq I(V;Y|U) - I(V;S|U)$$
$$R_1 + R_2 \leq I(U,V;Y) - I(U,V;S)\Big\}$$

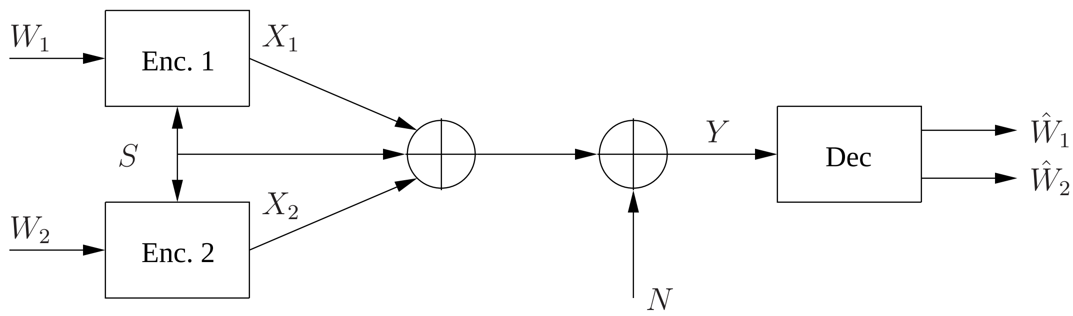


Figure 1.8: Dirty MAC with common state information.

where $(U, V)$ are auxiliary pairs satisfying:

$$(U, X_1) \leftrightarrow S \leftrightarrow (V, X_2)$$
$$(U, V) \leftrightarrow (X_1, X_2, S) \leftrightarrow Y.$$

However, this scheme was proved to be suboptimal by Philosof and Zamir [35], at least in certain cases, when the users have access to two distinct independent parts of the state $\boldsymbol{s}$.

Philosof et al. [34, 38, 35] considered a Gaussian additive MAC with additive interference and composed of a sum of two independent Gaussian interferences, where each interference is known non-causally only to one of the encoders. They called this channel the "doubly-dirty MAC". The capacity region of the Gaussian "dirty MAC", where the interference is known non-causally to both transmitters ("DMAC with common interference"), was found by Gel'fand and Pinsker [20] (and rediscovered by Kim, Sutivong and Sigurjónsson [27]), to be equal to the interference-free MAC channel, by applying DPC by both users.

Philosof, Zamir and Erez [37] considered a binary modulo-additive version of this channel ("binary DMAC"), depicted also in Figure 1.8:

$$Y = X_1 \oplus X_2 \oplus S \oplus N\,, \tag{1.18}$$

where $X_1, X_2, S, N \in \mathbb{Z}_2$. The input ("power") constraints are $\frac{1}{n} w_H(\mathbf{x}_i) \leq q_i$ for $i = 1, 2$, where $0 \leq q_1, q_2 \leq 1/2$. The noise $N \sim$ Bernoulli$(\varepsilon)$ and is independent of $S, X_1, X_2$; the state information $S \sim$ Bernoulli$(1/2)$ is known non-causally to both encoders.

They derived the capacities for two different scenarios:

- *The binary doubly-dirty MAC*: in this scenario $S = S_1 \oplus S_2$, where $S_1, S_2 \sim$ Bernoulli(1/2) are independent and known non-causally to encoders 1 and 2, respectively. The capacity region of this channel is given by the set of all rate pairs $(R_1, R_2)$ satisfying:

$$\mathcal{C}(q_1, q_2) \triangleq \Big\{(R_1, R_2) : R_1 + R_2 \leq \text{uch}\left[H_b(q_{min}) - H_b(\varepsilon)\right]\Big\},$$

  where $q_{min} \triangleq \min(q_1, q_2)$ and the upper convex hull operation is w.r.t. $q_1$ and $q_2$.

- *The single informed user*: in this scenario $S$ is known only to user 1. The capacity region of this channel is given by the set of all rate pairs $(R_1, R_2)$ satisfying:

$$\mathcal{C}(q_1, q_2) \triangleq \text{cl conv}\left\{(R_1, R_2) : \begin{array}{c} R_2 \leq H_b(q_2 \circledast \varepsilon) - H_b(\varepsilon) \\ R_1 + R_2 \leq H_b(q_1) - H_b(\varepsilon) \end{array}\right\}. \tag{1.19}$$

However, contrary to the Gaussian case, in which the common interference capacity region is the same as the interference-free region, and is achieved using *stationary* inputs, in the *binary* DP channel, there is a loss even in the point-to-point setting. Thus the capacity region of the binary DMAC with common interference is not known, and is yet to be determined.

# Chapter 2

# Robustness of Dirty Paper Coding

**In this chapter we consider a Gaussian DP channel, where the transmitter knows the interference sequence up to a constant multiplicative factor, known only to the receiver. we derive lower bounds on the achievable rate of communication by proposing a lattice-based coding scheme that partially compensates for the imprecise channel knowledge. We focus on a communication scenario where the SNR is high. When the power of the interference is finite, we show that the achievable rate of this coding scheme may be improved by a judicious choice of the scaling parameter at the receiver. We further show that the communication rate may be improved, for finite as well as infinite interference power, by allowing randomized scaling at the transmitter of the lattice-based scheme, as well as in Costa's random binning scheme. Finally we consider the implications of the results on the Gaussian MIMO BC channel with imprecise channel knowledge. We employ the derived technique on the DPC and linear transmission schemes, and compare their performance.**

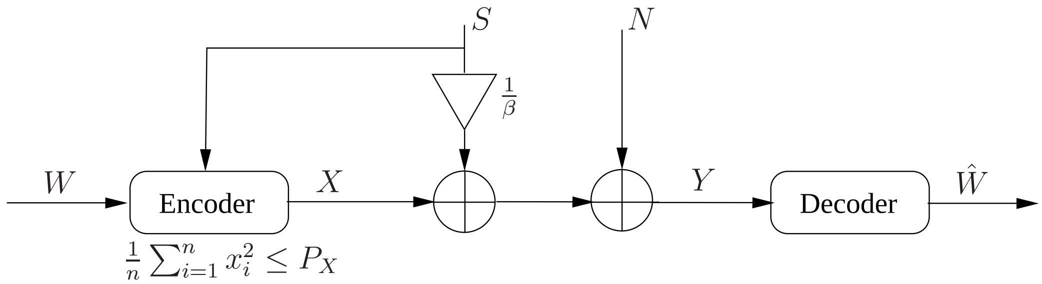


Figure 2.1: The compound dirty-paper channel.

## 2.1 Channel Model and Motivation

We consider the channel model, depicted in Figure 2.1:

$$Y = X + \frac{S}{\beta} + N, \tag{2.1}$$

where $\beta \in \mathcal{I}_\Delta = [1-\Delta, 1+\Delta]$ is a constant that is unknown to the transmitter ("compound"), but is known at the receiver. Thus, $\Delta$ is a measure of the degree of channel uncertainty. Note that we do not assume an additional compound parameter multiplying $X$, as opposed to the case treated by Mitran, Devroye and Tarokh in [33] (see (1.11)), since it does not conceal much added value over the case of (2.1), as will become evident in the sequel.

Consider first the limit of high SNR. At first glance, one might suspect that a reasonable approach could be to use standard THP since, as described in Chapter 1.4.3, it is optimal at high SNR in the perfect channel knowledge case. This would correspond to pre-subtracting the interference $S$ at the transmitter, applying a modulo operation and treating the residual interference as noise. However, the residual interference, namely $(1-\frac{1}{\beta})S$, left at the receiver, may be large if the power of the interference is large. In fact, in the limit $P_S \to \infty$, the achievable rate, for reliable communication using this approach, would vanish. Thus naïve implementation of THP is not robust to channel uncertainty.

We observe, in Chapter 2.3, that by using a carefully chosen scaling parameter at the receiver, reliable communication, at strictly positive rate, is possible, regardless of the interference power. The optimal scaling parameter does, however, depend on the power of the interference and should strike a balance between the residual interference, the "self-noise" component, and the

The material in this chapter was presented, in part, in [26, 25].

Gaussian noise.

We then show, in Chapter 2.4, that performance may further be improved by using randomized (time-varying) scaling at the transmitter. We begin by examining the more general problem of compound channel with side information, introduced in Chapter 1.4.5.

## 2.2 Compound Channels with Causal Side Information at the Transmitter

The compound DP channel of (2.1) is a compound memoryless state-dependent channel with SI at the transmitter, as argued in Chapter 1.4.5, where $S$ is the SI and $\beta$ plays the role of the compound component ($\mathcal{I}_\Delta$ plays the role of $\mathcal{B}$).

The (worst-case) capacity formula for the ("classical") compound channel, derived by Shannon [42], may be easily extended to the case of a compound channel with SI available *causally* to the transmitter, as implied by the following theorem, which is proved in Appendix A.1.

**Theorem 2.1.** *The worst-case capacity of a compound DMC with causal SI at the transmitter is given by*

$$C = \max_{p(t) \in \mathcal{P}(\mathcal{T})} \inf_{\beta \in \mathcal{B}} I_\beta(T;Y)\,,$$

*where $\mathcal{T}$ denotes the set of all strategy functions of the form $t : \mathcal{S} \longrightarrow \mathcal{X}$, and $\mathcal{P}(\mathcal{T})$ is the set of all probability vectors over $\mathcal{T}$.*

**Remark 2.1.**

- The result of Theorem 2.1 suggest that, like in the non-compound DMC with causal SI problem (see Chapter 1.4.1), only mappings of the *current* state needs to be considered.

- The case of non-causal SI is more difficult. The converse of Gel'fand-Pinsker [19] is not easily extended to the compound scenario, as briefly discussed in Chapter 1.4.5, and only upper and lower *single-letter* bounds on the capacity with non-causal SI, are known. Using Theorem 2.1, a *non single-letter* expression for the worst-case capacity in the non-causal

SI case, using $k$-dimensional vector strategies and taking $k$ to infinity, follows:

$$C^{\text{non-casual}} = \limsup_{k\to\infty} \max_{p(\boldsymbol{t})} \inf_{\beta} \frac{1}{k} I_\beta(\boldsymbol{T};\boldsymbol{Y})\,.$$

## 2.3 Compensation for Channel Uncertainty at the Transmitter

The compound DP channel was defined in (2.1). In this section, we consider the case of i.i.d. interference of finite power $P_S$. The results of Chapter 2.2 may readily be extended to continuous alphabet and to incorporate an input constraint (similarly to [33], Sec. IV). Thus, Theorem 2.1 holds for this setting as well.

Since the capacity of the dirty-paper channel with causal SI is unknown even in the standard (non compound) setting, we do not attempt to explicitly find the capacity in the compound setting. Rather, we shall examine the performance of THP-like precoding schemes and suggest methods by which the lack of perfect channel knowledge at the transmitter may be taken into account and partially compensated for.

### 2.3.1 THP With Imprecise Channel Knowledge

We shall concentrate on the performance of one-dimensional lattice based schemes, i.e., lattices of the form $\Lambda = L\mathbb{Z}$, whose fundamental Voronoi region is $\mathcal{V}_0 \triangleq \left[-\frac{L}{2}\frac{L}{2}\right)$, where $L$ is chosen such that the power constraint is satisfied: $P_X = \frac{L^2}{12}$. Denote by SIR $\triangleq \beta^2 \frac{P_X}{P_S}$ the signal-to-interference ratio. Let $U \sim \text{Unif}(\mathcal{V}_0)$ be a random variable (dither) known to both transmitter and receiver. We consider a variation of the THP scheme of Chapter 1.4.3, in which we distinguish between the inflation factors "$\alpha$", used at the transmitter and the receiver:

- Transmitter: for any $v \in \mathcal{V}_0$, the transmitted signal is

$$X = [v - \alpha_T S - U] \bmod \Lambda.$$

- Receiver: computes,

$$Y' = [\alpha_R Y + U] \bmod \Lambda.$$

The channel from $v$ to $Y'$ can be rewritten as:

$$\begin{aligned}
Y' &= [\alpha_R Y + U] \bmod \Lambda \\
&= \left[\alpha_R X + \alpha_R \frac{S}{\beta} + \alpha_R N + U\right] \bmod \Lambda \\
&= \left[v - (v - \alpha_T S - U) + \alpha_R X + (\alpha_R - \alpha_T \beta)\frac{S}{\beta} + \alpha_R N\right] \bmod \Lambda \\
&= \left[v - (1-\alpha_R) X + (\alpha_R - \alpha_T \beta)\frac{S}{\beta} + \alpha_R N\right] \bmod \Lambda\,.
\end{aligned}$$

Due to the dither $U$, $X$ is independent of $S$ and of the information signal $V$, and is uniform over $\Lambda$ (see, e.g., [17, 18]). Therefore, this channel is equivalent, in distribution, to the modulo-additive channel:

$$\begin{aligned}
Y' &= [v + N_{\text{eff}}^{\beta}] \bmod \Lambda \\
N_{\text{eff}}^{\beta} &\triangleq (1-\alpha_R) U + (\alpha_R - \alpha_T \beta)\frac{S}{\beta} + \alpha_R N,
\end{aligned} \tag{2.2}$$

where $N_{\text{eff}}^{\beta}$ is the "effective noise", composed of a "self noise" component $(1-\alpha_R)U$, a residual interference component $(\alpha_R - \alpha_T\beta)\frac{S}{\beta}$ and a Gaussian noise component $\alpha_R N$. The average power of the effective noise is

$$P_{N_{\text{eff}}^{\beta}} = (1-\alpha_R)^2 P_X + (\alpha_R - \alpha_T\beta)^2 \frac{P_S}{\beta^2} + {\alpha_R}^2 P_N.$$

and the corresponding signal-to-effective noise power is

$$\text{SNR}_{\text{eff}} \triangleq \frac{P_X}{P_{N_{\text{eff}}^{\beta}}} = \left[(1-\alpha_R)^2 + \frac{(\alpha_R - \alpha_T\beta)^2}{\text{SIR}} + \frac{{\alpha_R}^2}{\text{SNR}}\right]^{-1}.$$

We denote the maximal achievable rate under these settings by $R_{\text{THP}}^{d}$, where "d" stands for "deterministic" (choice of) $\alpha_T$ (in contrast to the random strategies treated later on in Chapter 2.4), and the achievable rate for a specific triplet $(\alpha_T, \alpha_R, \beta)$ - by $R_{\text{THP}}^{d}(\alpha_T, \alpha_R, \beta)$.

**Lemma 2.1.** *The maximal achievable rate using the scheme described above is lower-bounded by:*

$$R^d_{THP} \geq \max_{\alpha_T} \min_{\beta \in \mathcal{I}_\Delta} \max_{\alpha_R} \frac{1}{2}\log(SNR_{\mathit{eff}}) + \varepsilon(\beta, \alpha_T, \alpha_R) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right),$$

*where* $\varepsilon(\beta, \alpha_T, \alpha_R) \triangleq h\left(N^\beta_{\mathit{eff},G}\right) - h\left(N^\beta_{\mathit{eff}}\right)$, $h(\cdot)$ *denotes the differential entropy and* $N^\beta_{\mathit{eff},G}$ *is Gaussian with the same variance as* $N^\beta_{\mathit{eff}}$.

Thus, $\varepsilon(\beta, \alpha_T, \alpha_R) \geq 0$ is a measure of non-Gaussianity.

*Proof.* First note that for any triplet $(\alpha_T, \alpha_R, \beta)$, the mutual information is maximized by taking $V \sim \mathrm{Unif}(\mathcal{V}_0)$. Hence:

$$\begin{aligned} R^d_{\mathrm{THP}}(\alpha_T, \alpha_R, \beta) &= h(Y') - h(Y'|V) \\ &= \log(L) - h(\left[N^\beta_{\mathrm{eff}}\right] \bmod \Lambda). \end{aligned}$$

The maximal achievable rate $R^d_{\mathrm{THP}}$ is therefore lower bounded by

$$\begin{aligned} R^d_{\mathrm{THP}} &= \max_{\alpha_T} \min_{\beta \in \mathcal{I}_\Delta} \max_{\alpha_R} R^d_{\mathrm{THP}}(\alpha_T, \alpha_R, \beta) \\ &= \max_{\alpha_T} \min_{\beta \in \mathcal{I}_\Delta} \max_{\alpha_R} \left[\log(L) - h(\left[N^\beta_{\mathrm{eff}}\right] \bmod \Lambda)\right] \\ &\geq \max_{\alpha_T} \min_{\beta \in \mathcal{I}_\Delta} \max_{\alpha_R} \left[\frac{1}{2}\log(L^2) - h\left(N^\beta_{\mathrm{eff}}\right)\right] \\ &= \max_{\alpha_T} \min_{\beta \in \mathcal{I}_\Delta} \max_{\alpha_R} \left[\frac{1}{2}\log(L^2) - h\left(N^\beta_{\mathrm{eff,G}}\right) + \varepsilon(\beta, \alpha_T, \alpha_R)\right] \\ &= \max_{\alpha_T} \min_{\beta \in \mathcal{I}_\Delta} \max_{\alpha_R} \left[\frac{1}{2}\log(12P_X) - \frac{1}{2}\log\left(2\pi \mathrm{e} P_{N^\beta_{\mathrm{eff}}}\right) + \varepsilon(\beta, \alpha_T, \alpha_R)\right] \\ &= \max_{\alpha_T} \min_{\beta \in \mathcal{I}_\Delta} \max_{\alpha_R} \left[\frac{1}{2}\log(\mathrm{SNR}_{\mathrm{eff}}) + \varepsilon(\beta, \alpha_T, \alpha_R) - \frac{1}{2}\log\left(\frac{2\pi\mathrm{e}}{12}\right)\right], \end{aligned}$$

where $\varepsilon(\beta, \alpha_T, \alpha_R) \triangleq h\left(N^\beta_{\mathrm{eff,G}}\right) - h\left(N^\beta_{\mathrm{eff}}\right)$ and $N^\beta_{\mathrm{eff,G}}$ is Gaussian with the same power as $N^\beta_{\mathrm{eff}}$. $\square$

We are left with the task of choosing $\alpha_T, \alpha_R, \beta$.

### 2.3.2 Naïve Approach

One could ignore the presence of the inaccuracy factor $\beta$ and apply standard THP, using the parameters $\alpha_R = \alpha_T = \alpha_{\mathrm{MMSE}} \triangleq \frac{\mathrm{SNR}}{1+\mathrm{SNR}}$, which is the best selection of $\alpha_R$ and $\alpha_T$ in the perfect knowledge case, as discussed in Chapter 1.4.3. This gives rise to the following signal-to-effective noise ratio at the receiver:

$$\begin{aligned} \mathrm{SNR}_{\mathrm{eff}} &= \lambda_{\text{Naïve}}(\beta)(1+\mathrm{SNR}) \\ \lambda_{\text{Naïve}}(\beta) &\triangleq \frac{1}{1+\frac{1}{\mathrm{SIR}}+\frac{\mathrm{SNR}}{\mathrm{SIR}}(1-\beta)^2}\,. \end{aligned}$$

Note that since $(1 + \mathrm{SNR})$ is the output SNR in the perfect SI case, the loss due to the imprecision $(1 - \beta)$ is manifested in the multiplicative factor $0 < \lambda_{\text{Naïve}}(\beta) \leq 1$.

Moreover, when the interference is very strong, i.e., $\mathrm{SIR} \to 0$, even if the SNR is high, the effective SNR goes to zero along with the rate (as further explained in Chapter 2.4.1). Nonetheless, a strictly positive rate can be achieved in this scheme, using a smarter Rx-Tx pair, as is shown in the following sections.

### 2.3.3 Smart Receiver - Ignorant Transmitter

Using the fact that $\varepsilon(\beta, \alpha_T, \alpha_R) \geq 0$, we can further loosen the lower-bound of Lemma 2.1 to

$$R^d_{\mathrm{THP}} \geq \max_{\alpha_T} \min_{\beta \in \mathcal{I}_\Delta} \max_{\alpha_R} \frac{1}{2}\log(\mathrm{SNR}_{\mathrm{eff}}) - \frac{1}{2}\log\left(\frac{2\pi\mathrm{e}}{12}\right)\,. \tag{2.3}$$

Note that optimizing the r.h.s. of (2.3) is equivalent to maximizing $\mathrm{SNR}_{\mathrm{eff}}$ with respect to $\{\alpha_T, \alpha_R\}$. In this section we shall optimize with respect to $\alpha_R$ ("smart receiver") and use $\alpha_T = \alpha_{\mathrm{MMSE}} \triangleq \frac{\mathrm{SNR}}{1+\mathrm{SNR}}$ ("ignorant transmitter") as was done in Chapter 2.3.2, and leave the treatment of a smarter selection of $\alpha_T$ ("smart transmitter") to Chapter 2.4.

By solving the problem of maximizing the signal-to-effective noise ratio, the following $\alpha_R$ value and corresponding $\mathrm{SNR}_{\mathrm{eff}}$ are obtained:

$$\alpha_T^{\mathrm{MMSE}} = \alpha_{\mathrm{MMSE}} \triangleq \frac{\mathrm{SNR}}{1+\mathrm{SNR}}$$

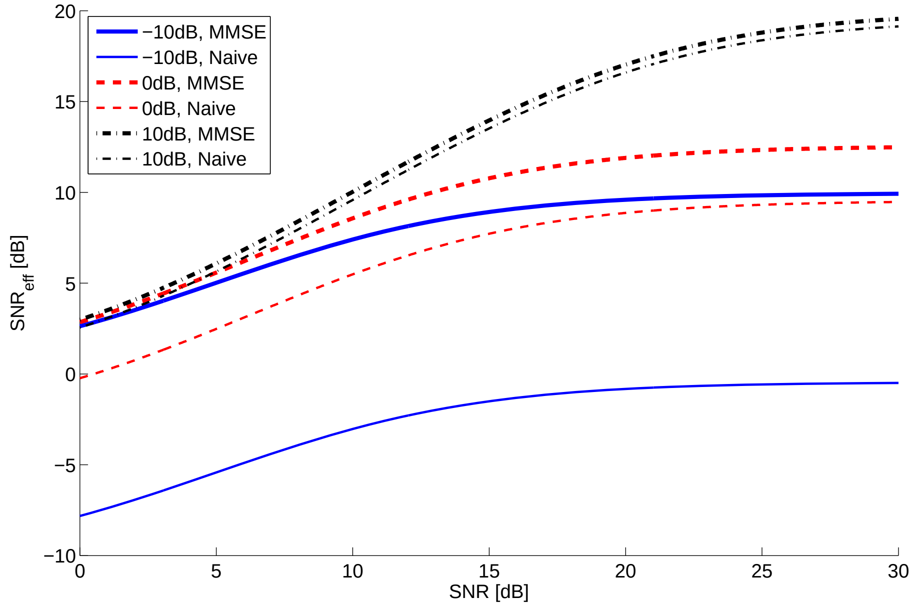


Figure 2.2: $\mathrm{SNR}_{\mathrm{eff}}$, for "naïve" and "smart receiver" approaches, as a function of SNR for different SIR values and $\Delta = 1/3$. Continuous line - $(-10)$dB; dashed line - $0dB$; dot-dashed - $10dB$. Within each pair: thick line - "Smart Rx" approach; thin line - "Naïve" approach.

$$
\begin{aligned}
\alpha_R^{\mathrm{MMSE}} &= \frac{1+\frac{\alpha_T^{\mathrm{MMSE}}\beta}{\mathrm{SIR}}}{1+\frac{1}{\mathrm{SIR}}+\frac{1}{\mathrm{SNR}}} \\
\mathrm{SNR}_{\mathrm{eff}} &= \lambda_{\mathrm{MMSE}}(\beta)(1+\mathrm{SNR}) \\
\lambda_{\mathrm{MMSE}}(\beta) &\triangleq \frac{1+\frac{1}{\mathrm{SIR}}+\frac{1}{\mathrm{SNR}}}{1+\frac{1}{\mathrm{SIR}}+\frac{1}{\mathrm{SNR}}+\frac{\mathrm{SNR}}{\mathrm{SIR}}(1-\beta)^2},
\end{aligned}
\tag{2.4}
$$

where again, the loss due to $\beta$ is manifested in $0 < \lambda_{\mathrm{MMSE}} \leq 1$. Note that the loss in $\mathrm{SNR}_{\mathrm{eff}}$ is smaller than that of the naïve approach since $\lambda_{\mathrm{Naïve}}(\beta) < \lambda_{\mathrm{MMSE}}(\beta)$, for every $\beta$.

Using $\alpha_R^{\mathrm{MMSE}}$, rather than the standard $\alpha_R = \frac{\mathrm{SNR}}{\mathrm{SNR}+1}$, improves $\mathrm{SNR}_{\mathrm{eff}}$ for all values of $\beta$. A lower-bound on the achievable rate is therefore given by,

$$
R_{\mathrm{THP}}^{d} \geq \frac{1}{2}\log(1+\mathrm{SNR}) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) - \frac{1}{2}\log\left(\frac{1}{\lambda_{\mathrm{MMSE}}(\beta = 1+\Delta)}\right).
$$

The gains of the this approach over the naïve one of Chapter 2.3.2, for different SNR values and $\Delta = 1/3$, are depicted in Figure 2.2.

**Remark 2.2.**

1. In the weak interference regime, SIR $\to \infty$, we have $\lambda_{\text{MMSE}}(\beta) \to 1$ (for all $\beta$) and hence $\alpha_R = \frac{\text{SNR}}{\text{SNR}+1}$ and $\text{SNR}_{\text{eff}} = 1 + \text{SNR}$. This is of course a non-interesting case as THP is unattractive in this regime.

2. In the strong interference regime, SIR $\to 0$, the residual interference component of $N_{\text{eff}}^{\beta}$ has to be completely cancelled. This is done by selecting $\alpha_R = \alpha_T \beta$ and results in an effective noise with finite power (dictated by the magnitude of $\Delta$). Thus reliable communication is possible at strictly positive rates, even when the interference is arbitrarily strong.

### 2.3.4 High SNR Regime

In the high SNR regime, i.e., SNR $\gg 1$, the choice $\alpha_T = 1$ becomes optimal. Using this choice of $\alpha_T$ in (2.4), we achieve the following effective SNR:

$$\text{SNR}_{\text{eff}} \geq \frac{1 + \text{SIR}}{(1-\beta)^2} \Big(1 - o(1)\Big) \,,$$

where $o(1) \to 0$ as SNR $\to \infty$. By substituting this effective SNR in the lower-bound of Lemma 2.1, we obtain the following achievable rate:

$$\begin{aligned} R_{\text{THP}}^{d} \geq & \frac{1}{2}\log(1+\text{SIR}) + \log\left(\frac{1}{\Delta}\right) \\ & - \frac{1}{2}\log\left(\frac{2\pi \mathrm{e}}{12}\right) + \min_{\beta \in \mathcal{I}_{\Delta}} \varepsilon(\beta, \alpha_T = 1, \alpha_R) - o(1) \,, \end{aligned} \tag{2.5}$$

where again, $o(1) \to 0$ as SNR $\to \infty$.

**Remark 2.3.**

1. In the case of strong interference and high SNR (SIR $\to 0$, SNR $\to \infty$), with the choice of $\alpha_T = 1$ and the corresponding optimal choice of $\alpha_R^{\text{MMSE}} = \frac{1}{\beta}$, the effective noise $N_{\text{eff}}^{\beta}$ has virtually only a self-noise component, i.e., $N_{\text{eff}}^{\beta} \approx (1-\alpha_R)U$. Hence, $\varepsilon(\beta, \alpha_T = 1, \alpha_R^{\text{MMSE}}) \to \frac{1}{2}\log\left(\frac{2\pi\mathrm{e}}{12}\right)$ as SNR $\to \infty$ (for $\forall \beta \in \mathcal{I}_{\Delta}$). Thus, there is no shaping loss compared to high-dimensional lattices in this case, as further explained in Chapter 2.5, and the corresponding achievable rate is $R_{\text{THP}}^{d} = \log\left(\frac{1}{\Delta}\right) - o(1)$.

2. The lower bound of (2.5) can be evaluated for any specific distribution of $S$, by calculating $\min_\beta \varepsilon(\beta, \alpha_T = 1, \alpha_R^{\mathrm{MMSE}})$. For instance, if $S$ is uniform, that is the limit of an $M$-PAM constellation ($M \to \infty$), then $R^d_{\mathrm{THP}}$ can be lower-bounded by

$$R^d_{\mathrm{THP}} \geq \frac{1}{2}\log(1+\mathrm{SIR}) + \log(\frac{1}{\Delta}) - \frac{1}{2}\log\left(\frac{\mathrm{e}}{2}\right) - o(1)\,,$$

where $o(1) \to 0$ as SNR $\to \infty$. This can be done for a general SNR as well, viz., not only in the limit of high SNR.

3. Even in the limit of strong interference, i.e., SIR $\to$ 0, for the "smart-receiver" approach, $\mathrm{SNR}_{\mathrm{eff}} > 1$, due to the extra 1 in the nominator. Hence a strictly positive rate is achieved in this regime, contrary to the effective SNR of the naïve approach, $\frac{\mathrm{SIR}}{(1-\beta)^2}$ , which goes to zero along with the achievable rate.

4. In the case of equal interference and signal powers, SIR $= 1$, there is a gain of 3dB over the naïve approach, as is seen in Figure 2.2.

5. When the signal and interference have the same power, SIR $= 1$, $\alpha_R^{\mathrm{MMSE}}$ strikes a balance between the two effective noise components, the powers of which become both equal to $\frac{1}{4}(1-\beta)^2 P_X$ for $\alpha_R = \alpha_R^{\mathrm{MMSE}}$. Thus, $\alpha_R^{\mathrm{MMSE}}$ gives a total noise power of $P_{N^\beta_{\mathrm{eff}}} = \frac{1}{2}(1-\beta)^2 P_X$, which is half the noise power obtained by cancelling out the interference component completely ($\alpha_R = \beta$), or alternately, half the noise power obtained by cancelling out completely the self-noise component ($\alpha_R = 1$).

6. Due to the modulo operation at the receiver's side and since the effective noise is not Gaussian, the choice $\alpha_R = \alpha_R^{\mathrm{MMSE}}$ does not strictly maximize the mutual information $I(V; Y')$, but rather is a reasonable approximate solution. Moreover, in the compound case, in contrast to the perfect SI case, minimizing the mean-square error (MSE) is not equivalent to maximizing the effective SNR or the rate, as demonstrated in Example 2.1.

## 2.4 Randomized Scaling at Transmitter

For simplicity, we now restrict our attention to the case of strong interference and high SNR, i.e., $\mathrm{SIR} \to 0, \mathrm{SNR} \to \infty$. More specifically, we consider a noise-free channel model:

$$Y = X + \frac{S}{\beta}.$$

In this case, the receiver must completely cancel out the interference by choosing $\alpha_R = \beta \cdot \alpha_T$. Note that if $\beta$ were known at the transmitter, the capacity would be infinite.

We now investigate whether performance may be improved by introducing a random scaling factor $\alpha$ at the transmitter ($\alpha_T = \frac{1}{\alpha}$), which is chosen in an i.i.d. manner at each time instance and is assumed known to both transmitter and receiver. Thus, we consider the following transmission scheme:

- Transmitter: for any $v \in \mathcal{V}_0$, sends

$$X = [v - \frac{1}{\alpha} S - U] \bmod \Lambda.$$

- Receiver: applies the front end operation,

$$Y' = [\alpha_R Y + U] \bmod \Lambda,$$

  where $\alpha_R = \beta/\alpha$.

By substituting $\alpha_T = 1/\alpha$ and $\alpha_R = \beta/\alpha$ in (2.2), we arrive to the equivalent channel

$$Y' = \left[v + N_{\text{eff}}^{\beta}\right] \bmod \Lambda, \tag{2.6}$$

with $N_{\text{eff}}^{\beta} = \frac{\alpha-\beta}{\alpha} U$. Note that the average power of $N_{\text{eff}}^{\beta}$ now varies from symbol to symbol according to the value of $\alpha$.

The rationale for considering such scaling at the transmitter is that had the transmitter known $\beta$, it would choose $\alpha = \beta$ to match the actual interference as experienced at the receiver. By using randomization, this will occur some of the time. Since $\beta$ is unknown however (to the transmitter), one might suspect that using a deterministic selection of $\alpha = 1$ may be optimal, as was done in Chapter 2.3.1. However, due to convexity, it turns out that a better approach is to let $\alpha$ vary[1] from symbol to symbol (or block to block) within the interval of uncertainty $\mathcal{I}_\Delta$.

**Example 2.1.** To further motivate this we shall look at the simple case of a compound parameter with alphabet of size 2, $\beta \in \mathcal{B} = \{1 \pm \Delta\}$. In this case the best deterministic selection of $\alpha$ is $\alpha = 1$, which gives rise to a finite rate for every $\beta \in \mathcal{B}$. However, consider choosing $\alpha$ at random, in an i.i.d. manner for each symbol, according to

$$P(\alpha = 1 - \Delta) = P(\alpha = 1 + \Delta) = \frac{1}{2}\,.$$

When the transmitter uses this selection policy of $\alpha$, approximately for half of the transmitted symbols the chosen $\alpha$ will equal $\beta$, even though $\beta$ is unknown to the transmitter; while for the other half of the symbols, the mismatch between $\beta$ and the chosen $\alpha$ will be greater than that obtained by taking $\alpha = 1$. Since whenever the chosen $\alpha$ is (exactly) equal to $\beta$, the mutual information between the conveyed message signal $v$ and the channel output $Y$ is *infinite*, since the channel is noiseless, the total rate is *infinite* as well.

**Remark 2.4.** In the absence of noise, if $\beta$ takes only a finite number of values, i.e. $|\mathcal{B}| < \infty$, then the achievable rate is infinite. The achievability is shown by generalizing the idea of the binary case: by varying $\alpha$ in an i.i.d. manner from symbol to symbol according to the uniform distribution $\alpha \sim \mathrm{Unif}(\mathcal{B})$. However a straightforward extension to the case of an infinite countable cardinality (all the more to a continuous alphabet), is not possible.

We denote the maximal achievable rate of the “randomized ” scaling scheme by $R^r_{\mathrm{THP}}$, where “r” stands for “random”. It is given by:

$$R^r_{\mathrm{THP}} = \max_{f(\alpha)} R^r_{\mathrm{THP}}(f) = \max_{f(\alpha)} \min_{\beta \in \mathcal{I}_\Delta} I_\beta(V; Y'|\alpha), \tag{2.7}$$

where $f(\alpha)$ is the p.d.f. according to which $\alpha$ is drawn and $R^r_{\mathrm{THP}}(f)$ denotes the mutual information corresponding to the specific choice of $f(\alpha)$. Note that

[1] Note that by doing so, we in effect extend the class of strategies used in the transmission scheme.

in this case the distribution of $\alpha$ that minimizes the mean-square error (MSE) is not necessarily the one that maximizes $\mathrm{SNR}_{\mathrm{eff}}$ or the rate $R^r_{\mathrm{THP}}(f)$. The MMSE criterion provides the signal-to-effective noise ratio

$$\mathrm{SNR}_{\mathrm{eff}} = \max_{f(\alpha)} \min_{\beta} \frac{P_X}{E_\alpha \left(N^\beta_{\mathrm{eff}}\right)^2},$$

which differs from the optimal signal-to-effective noise ratio, that can be achieved by direct optimization:

$$\mathrm{SNR}_{\mathrm{eff}} = \max_{f(\alpha)} \min_{\beta} E_\alpha \left[ \frac{P_X}{\left(N^\beta_{\mathrm{eff}}\right)^2} \right].$$

Moreover, these optimizations are not equivalent in general to optimizing the achievable rate $R^r_{\mathrm{THP}}$. Hence a direct optimization of (2.7) needs to be done. Finally mind that in this case the effective noise will vary with time along with variations in the value of $\alpha$.

**Lemma 2.2.** *The maximal achievable rate, when $\Delta \leq \frac{1}{3}$, for the noiseless DP channel, using the "extended THP scheme", given in* (2.6)*, is*

$$R^r_{THP} = \max_{\substack{f(\alpha): \\ \mathrm{Supp}\{f(\alpha)\} \subseteq \mathcal{I}_\Delta}} \min_{\beta \in \mathcal{I}_\Delta} -E_\alpha \left[ \log \left| \frac{\alpha - \beta}{\alpha} \right| \right]. \tag{2.8}$$

The proof of this lemma is given in Appendix A.2 along with the treatment of the case of $\Delta > \frac{1}{3}$.

Finding the optimal distribution of $\alpha$ in (2.7) is cumbersome. Instead, we suggest several choices for the distribution $f(\alpha)$ which achieve better performance than that of any deterministic selection of $\alpha$ as well as derive an upper bound on $R^r_{\mathrm{THP}}$.

### 2.4.1 Quantifying the Achievable Rates

As indicated by Lemma 2.2, we restrict attention to the case of $\Delta \leq \frac{1}{3}$. We consider three different distributions for $\alpha$: deterministic selection, uniform distribution and V-like distribution.

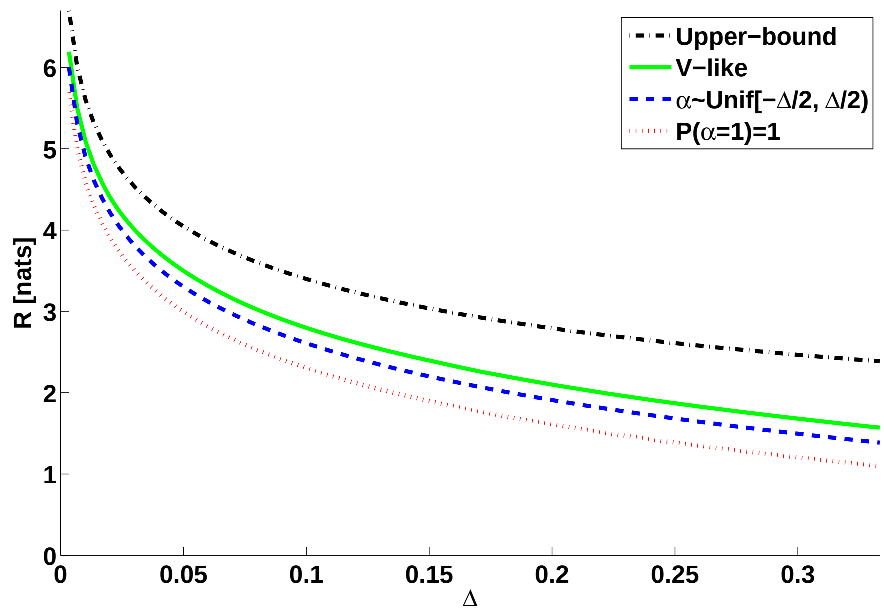


Figure 2.3: Achievable rates and upper bound on the THP scheme.

### Deterministic Selection

One easily verifies that the value of $\alpha$, which achieves the maximal rate, is $\alpha = 1$ and the corresponding rate is

$$R^r_{\mathrm{THP}}(f_{\mathrm{Deter}}) = -\log\Delta = \log\frac{1}{\Delta}\,.$$

Note that this result coincides with the result for $R^d_{\mathrm{THP}}$ of Section 2.3.1 ($\varepsilon(\beta, \alpha_R) - \frac{1}{2}\log(\frac{2\pi e}{12})$ is equal to zero in this case, as mentioned in Remark 2.3).

### Uniform Distribution

Taking $\alpha \sim \mathrm{Unif}(\mathcal{I}_\Delta)$ yields the following achievable rate:

$$\begin{aligned} R^r_{\mathrm{THP}}(f_{\mathrm{Unif}}) =& \frac{1}{2\Delta}\big[(1+\Delta)\log(1+\Delta) \\ &\quad - (1-\Delta)\log(1-\Delta) - 2\Delta\log(2\Delta)\big]\,. \end{aligned}$$

Hence, even this simple randomization improves on the deterministic selection, as may be seen in Figure 2.3.

**V-like Distribution**

A further improvement is obtained by taking a V-like distribution,

$$f_{\mathrm{V-like}}(\alpha) = \frac{|\alpha - 1|}{\Delta^2}, \qquad |\alpha - 1| \le \Delta\,.$$

The resulting rate is

$$R^r_{\mathrm{THP}}(f_{\mathrm{V-like}}) = -\frac{1}{2\Delta^2}\left[(1-\Delta^2)\log(1-\Delta^2) + \Delta^2\log(\Delta^2)\right].$$

We have not pursued numerical optimization of $f(\alpha)$. We note that none of the three distributions above are optimal since $I_\beta(V;Y')$ varies with $\beta$. Moreover, the optimal p.d.f. will not be totally symmetric around 1 due to the denominator in (2.8). This term becomes, however, less and less significant (and hence the optimal p.d.f. more and more symmetric) for small $\Delta$. We next derive an upper bound on the achievable rate which holds for any choice of $f(\alpha)$.

### 2.4.2 Upper Bound on Achievable Rates

**Lemma 2.3.** *The rate achievable using THP with randomized scaling is upper bounded by*

$$R^r_{THP} \le \log(1+\Delta) - \log(\Delta) + 1$$

*for any distribution $f(\alpha)$, when $\Delta \le \frac{1}{3}$.*

*Proof.* Using (2.8), for every distribution $f(\alpha)$, we have

$$\begin{aligned} I_\beta(V;Y') &= \min_\beta \left\{E_\alpha\left[\log\alpha\right] - E_\alpha\left[\log|\alpha-\beta|\right]\right\} \\ &\overset{(a)}{\le} \min_\beta \left\{\log(1+\Delta) - E_\alpha\left[\log(|\alpha-\beta| \bmod \Lambda)\right]\right\} \\ &\overset{(b)}{\le} \log(1+\Delta) - \frac{1}{2\Delta}\int_{-\Delta}^{\Delta} \log|x|dx \\ &= \log(1+\Delta) - \log(\Delta) + 1, \end{aligned}$$

where $(a)$ holds since $\mathrm{Supp}\,\{f(\alpha)\} \subseteq \mathcal{I}_\Delta$ and $(b)$ is true due to the monotonicity of the log function where equality is achieved for $\alpha \sim \mathrm{Unif}(\mathcal{I}_\Delta)$. $\square$

### 2.4.3 Noisy Case

The randomized approach taken may be extended to the noisy case:

$$Y' = \left[v + N_{\mathrm{eff}}^{\beta}\right] \bmod \Lambda\,,$$
$$N_{\mathrm{eff}}^{\beta} = (1-\alpha_R)U + \left(\alpha_R - \frac{\beta}{\alpha}\right)\frac{S}{\beta} + \alpha_R N.$$

This result is easily obtained by substituting $\alpha_T = 1/\alpha$ in (2.2).

Consider the case of SIR $\to 0$ (and finite SNR). In this case $\alpha_R$ has to be chosen to be equal to $\beta/\alpha$, in order to eliminate the residual interference component in the effective noise. The effective noise in this case is hence:

$$N_{\mathrm{eff}}^{\beta} = \frac{\alpha-\beta}{\alpha}U + \frac{\beta}{\alpha}N.$$

Note that, unlike in the noiseless case, in which the effective noise had only a finite support ("self-noise") component $\frac{\alpha-\beta}{\alpha}U$, here the noise has an additional Gaussian component $\frac{\beta}{\alpha}N$.

We only examine the deterministic and uniform distributions from Chapter 2.4 and minor variations on them, taking $\alpha_T = \frac{\alpha_{\mathrm{MMSE}}}{\alpha} \triangleq \frac{1}{\tilde{\alpha}}$, where $\alpha$ is selected according to the distributions of Chapter 2.4 and $\alpha^{\mathrm{MMSE}} \triangleq \frac{\mathrm{SNR}}{1+\mathrm{SNR}}$. The performances of the different choices of $\alpha_T$ are shown in Figure 2.4.

Note that in the high SNR regime, the non-deterministic distributions prove to be more effective than the best deterministic scheme, whereas in the low SNR regime the deterministic selection becomes superior. This threshold phenomenon can be explained by considering the two components of $N_{\mathrm{eff}}^{\beta}$: in the high SNR regime, the dominant noise component is the "self-noise" component $\frac{\tilde{\alpha}-\beta}{\tilde{\alpha}}U$, which is minimized by a "smart" selection of $f(\cdot)$; in the low SNR regime, on the other hand, the dominant noise component is the Gaussian part $\frac{\beta}{\tilde{\alpha}}N$, whose multiplicative factor $\frac{\beta}{\tilde{\alpha}}$ should be deterministic to minimize its average power. In general, there is a tradeoff between the best deterministic selection of $\alpha_T$ which minimizes the power of the Gaussian component and the self-noise component, which is to be minimized by a random $\alpha_T$ selection.

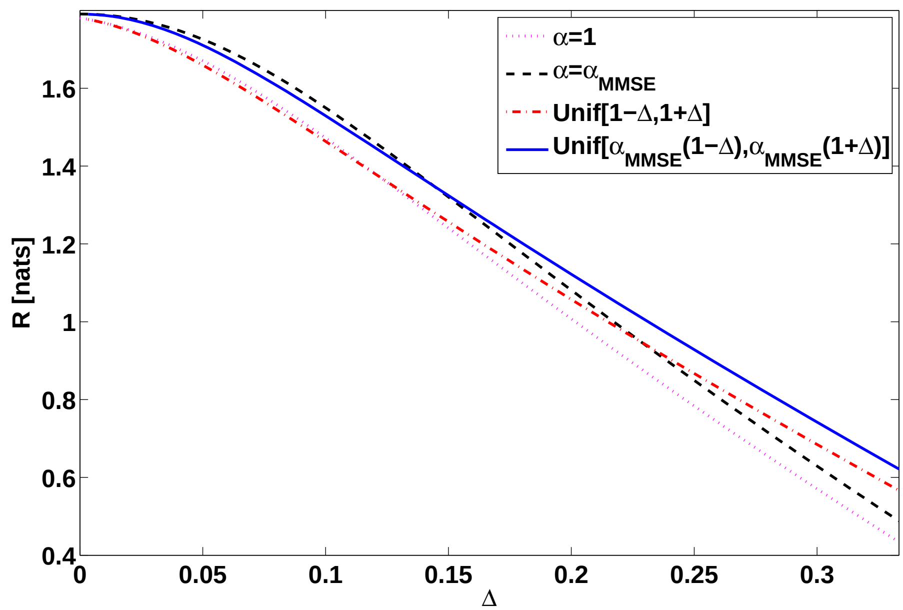


Figure 2.4: Achievable rates in the random THP scheme for SNR = 17dB.

## 2.5 Non-Causal Case and Multi-Dimensional Lattices

As discussed in Chapter 1.4.3, multi-dimensional extension of THP (i.e., lattice-based precoding), allows to approach the full capacity of the DP channel with *perfect channel knowledge* (and non-causal knowledge of the interference). Somewhat surprisingly, we observe that, when the channel knowledge is *imperfect*, multi-dimensional lattice precoding yields *identical* results to those obtained by scalar (one-dimensional lattice) precoding, in the limits of high SNR and low SIR. This is seen by simply repeating the proof of Lemma 2.2 for a multi-dimensional lattice $\Lambda$. It can be explained by the fact that, in this case, no shaping gain can be obtained using higher dimensional lattices, as the self-noise, being the only noise component, "gains shaping" just like the signal. Hence, using high-dimensional lattices does not increase the achievable rates of lattice-based precoding schemes in the absence of channel noise and when interference is strong. In the noisy case, as well as when the interference power is limited, however, multi-dimensional strategies allow gaining some of the shaping gain, due to the presence of a Gaussian noise component, as was discussed in Chapter 2.4.3.

Turning to the random binning scheme and the auxiliary used by Costa (see (1.4)), which was used for the compound case as well by Mitran, Devroye and Tarokh [33], we observe that by selecting the parameter $\alpha$, in the same manner as $\alpha_T$ of the THP schemes of Chapter 2.3 and Chapter 2.4, we arrive to the same performances when using multi-dimensional lattices of dimensions going to infinity, in all scenarios (finite/infinite SIR, finite/infinite SNR). Thus, the $\alpha$ parameter in this random binning scheme, takes the role of $\alpha_T$ in the THP scheme.

## 2.6 Implications to MIMO BC Channels

Consider the Gaussian MISO model of (1.12):

$$Y_i = \boldsymbol{h}_i^T \boldsymbol{X} + N_i, \qquad i = 1, 2\,.$$

In practice, the channel vectors $\boldsymbol{h}_i$ are known up to some finite accuracy, due to estimation errors or limited feedback, at the transmitter. We assume that the transmitter knows the channel vectors $\boldsymbol{h}_i$ up to some *small* angular errors $\varepsilon_i \in [-\Delta, \Delta]$ ($\Delta \ll 1$),[2] that is:

$$\begin{aligned} h_i &\approx \tilde{h}_i \\ \langle \boldsymbol{e}_i, \tilde{\boldsymbol{e}}_i \rangle &= \cos(\varepsilon_i)\,, \end{aligned} \tag{2.9}$$

where $\boldsymbol{h}_i$ ($i = 1, 2$) are the estimations of the channel vectors available at the transmitter and $\tilde{\boldsymbol{h}}_i = \tilde{h}_i \tilde{\boldsymbol{e}}_i$ are the true channel realizations. See also Figure 2.5.

### 2.6.1 Linear Zero-Forcing

According to this strategy, the transmitter avoids interferences by transmitting $\boldsymbol{x}_1$ in an orthogonal direction to $\boldsymbol{h}_2$, and $\boldsymbol{x}_2$ - orthogonally to $\boldsymbol{h}_1$, as depicted in Figure 1.6 (see, e.g., [31]).

In the case of imperfect channel knowledge at the transmitter, described by (2.9), the presence of an additional residual noise component is inevitable.

[2] One may assume a presence of *small* magnitude errors as well. However, such errors would have no effect when performing first-order approximations.

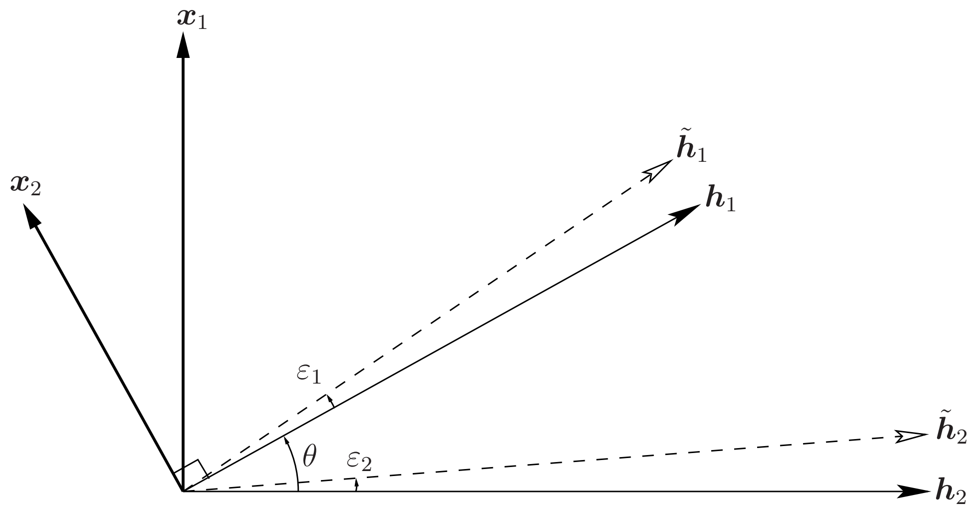


Figure 2.5: Pictorial representation of the zero-forcing technique in the MIMO broadcast channel.

The simplest approach to confront this problem is to ignore the estimation inaccuracy, that is, transmitting as if $\varepsilon_i$ were 0. This approach gives rise to the following equivalent channel:

$$\begin{aligned} Y_1 &= \left\langle \tilde{\boldsymbol{h}}_1, X_1 \right\rangle + \left\langle \tilde{\boldsymbol{h}}_1, X_2 \right\rangle + N_1 \\ &= X_1 h_1 \cos\left(\frac{\pi}{2} - \theta - \varepsilon_1\right) + X_2 h_1 \cos\left(\frac{\pi}{2} + \varepsilon_1\right) + N_1 \\ &= X_1 h_1 \sin(\theta + \varepsilon_1) - X_2 h_1 \sin\varepsilon_1 + N_1 \\ &\approx X_1 h_1 \sin(\theta) - X_2 h_1 \varepsilon_1 + N_1, \end{aligned} \tag{2.10}$$

where in the last equality we used a first-order approximation.

The average power of the effective noise is, therefore:

$$P_{N_{1,\text{eff}}} = P_2 h_1^2 \varepsilon_1^2 + 1,$$

(recall that we assumed, w.l.o.g, that the Gaussian noise power is 1).

The channel to user 2 can be derived in the same way:

$$\begin{aligned} Y_2 &\approx X_2 h_2 \sin(\theta) - X_1 h_2 \varepsilon_2 + N_2\,, \\ P_{N_{2,\text{eff}}} &= P_1 h_2^2 \varepsilon_2^2 + 1\,. \end{aligned}$$

Thus, by using codebooks that achieve capacity for the (interference-free)

AWGN channel, any rate pair $(R_1, R_2)$ satisfying:

$$R_1 \leq \frac{1}{2}\log\left(1 + \frac{\mathrm{SNR}_1 h_1^2 \sin^2(\theta)}{\mathrm{SNR}_2 h_1^2 \Delta^2 + 1}\right) - o(1)$$
$$R_2 \leq \frac{1}{2}\log\left(1 + \frac{\mathrm{SNR}_2 h_2^2 \sin^2(\theta)}{\mathrm{SNR}_1 h_2^2 \Delta^2 + 1}\right) - o(1) \tag{2.11}$$

is achievable, under first-order approximations ($\Delta \ll 1$), where $o(1) \to 0$ as $\Delta \to 0$.

We suggest improving the above scheme by working matched to the case of $\boldsymbol{h}_1$ tilted by an additional angle of our choice, when sending $\boldsymbol{x}_2$. That is, sending $\boldsymbol{x}_2$ in an orthogonal direction to $\boldsymbol{h}_1$ after subtracting a small angle $\alpha_1$. As the scheme is symmetric for both users, we apply the same strategy in the transmission direction of $\boldsymbol{x}_1$, by subtracting a small angle $\alpha_2$.

Repeating the steps of (2.10), we arrive to the channel

$$Y_1 \approx X_1 h_1 \sin(\theta) + X_2 h_1(\alpha_1 - \varepsilon_1) + N_1,$$

where again, first-order approximation is assumed. As in Chapter 2.4, we shall allow the use of a random selection of $\alpha_1$, according to some marginal distribution $f(\alpha_1)$. Hence, the (worst-case) achievable rates in this case are:

$$\begin{aligned} R_1 &= \max_{f(\alpha_1)} \min_{\varepsilon_1 \in [-\Delta,\Delta]} I(X_1\,; Y_1|\alpha_1) \\ &\approx \max_{f(\alpha_1)} \min_{\varepsilon_1} E_{\alpha_1} \frac{1}{2}\log\left(1 + \frac{\mathrm{SNR}_1 h_1^2 \sin^2(\theta)}{\mathrm{SNR}_2 h_1^2 (\alpha_1 - \varepsilon_1)^2 + 1}\right), \\ R_2 &\approx \max_{f(\alpha_2)} \min_{\varepsilon_2} E_{\alpha_2} \frac{1}{2}\log\left(1 + \frac{\mathrm{SNR}_2 h_2^2 \sin^2(\theta)}{\mathrm{SNR}_1 h_2^2 (\alpha_2 - \varepsilon_2)^2 + 1}\right). \end{aligned} \tag{2.12}$$

Similarly to the optimization problem of (2.8), the maximization problems in (2.12) are convex. Thus, the expressions in (2.12) are maximized for non-deterministic selections of $\alpha_1$ and $\alpha_2$.

### 2.6.2 Dirty Paper Coding

We now address the problem of working with DPC based schemes, when the channel knowledge is imperfect and given in the form of (2.9). We analyze

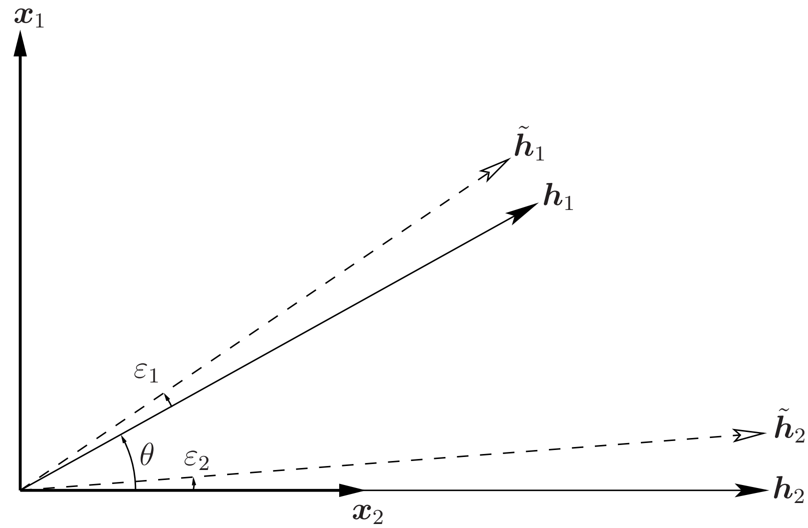


Figure 2.6: Pictorial representation of the DPC technique for the MIMO broadcast channel.

the performance of this scheme for the imperfect case, by repeating the steps in the treatment of the perfect knowledge case, described in Chapter 1.4.6. We distinguish between the inflation factors at the transmitter, $\alpha_T$, and the receiver, $\alpha_R$, as explained in Chapter 2.3.1, to facilitate the treatment for different approaches, similar to the ones discussed for the channel in (2.1) earlier in this section.

The channel to user 1 (the user that performs DPC), is (see also Figure 2.6):

$$\begin{aligned} Y_1 &= \left\langle \tilde{\boldsymbol{h}}_1, \boldsymbol{X}_1 \right\rangle + \left\langle \tilde{\boldsymbol{h}}_1, \boldsymbol{X}_2 \right\rangle + N_1 \\ &= X_1 h_1 \cos\left(\frac{\pi}{2} - \theta - \varepsilon_1\right) + X_2 h_1 \cos(\theta + \varepsilon_1) + N_1 \\ &= X_1 h_1 \sin(\theta + \varepsilon_1) + X_2 h_1 \cos(\theta + \varepsilon_1) + N_1 \,, \end{aligned}$$

or equivalently, after dividing both sides by $h_1 \sin(\theta + \varepsilon_1)$

$$\tilde{Y}_1 = X_1 + \mathrm{ctg}(\theta + \varepsilon_1) X_2 + \frac{1}{h_1 \sin(\theta + \varepsilon_1)} N_1 \,.$$

By constructing the coding scheme, similarly to the one in (1.16), we arrive to the (lattice) modulo-additive channel, of the form:

$$X_1 = [v_1 - \alpha_T \mathrm{ctg}(\theta) X_2 - U] \mod \Lambda \,,$$

$$\begin{aligned} Y_1' &= \left[\alpha_R \tilde{Y}_1 + U\right] \bmod \Lambda \\ &= \left[v_1 + N_{1,\text{eff}}^{\varepsilon_1}\right] \bmod \Lambda\,, \end{aligned}$$

where

$$\begin{aligned} N_{1,\text{eff}}^{\varepsilon_1} = -(1-\alpha_R)X_1 &+ \big(\alpha_R \text{ctg}(\theta+\varepsilon_1) - \alpha_T \text{ctg}(\theta)\big) X_2 \\ &+ \frac{\alpha_R}{h_1 \sin(\theta+\varepsilon_1)} N_1\,. \end{aligned} \tag{2.13}$$

Let us concentrate on the case in which $1 \ll P_1 \ll P_2$, that is, the case in which the interference (the message to user 2) is much stronger than the power of user 1 and both SNRs are high; we shall further assume that the angle between the two channel vectors satisfies $\theta < (\pi/2 - \Delta)$, since otherwise dirty paper coding is unattractive as explained in Remark 2.2.1., and also assume that $0 < \Delta \ll \theta$, since otherwise the directions of the two channel vectors are virtually indistinguishable. Note that this means that we continue to assume small "error intervals", viz., $\Delta \ll 1$, like was done for the linear ZF strategy.

Taking the naïve approach, viz., ignoring the imperfectness in the channel knowledge, suggests working with $\alpha_R = \alpha_T \approx 1$, as the SNRs are high. Thus, the effective noise in (2.13) is equal, under first order approximation, to:

$$N_{1,\text{eff}}^{\varepsilon_1} \approx -X_2 \csc^2(\theta)\varepsilon_1 + \frac{1}{h_1 \sin(\theta)} N_1\,,$$

where $\csc(x) \triangleq 1/\sin(x)$.

Note also, that the channel seen by user 2, in this case, is the same as in the linear ZF scheme. Hence, using the "naïve" approach, we achieve any rate pair $(R_1, R_2)$, satisfying:

$$\begin{aligned} R_1 &\le \frac{1}{2}\log\left(1 + \frac{\text{SNR}_1 h_1^2 \sin^2(\theta)}{\text{SNR}_2 h_1^2 \csc^2(\theta)\Delta^2 + 1}\right) - o(1)\,, \\ R_2 &\le \frac{1}{2}\log\left(1 + \frac{\text{SNR}_2 h_2^2}{\text{SNR}_1 h_2^2 \Delta^2 + 1}\right) - o(1)\,, \end{aligned} \tag{2.14}$$

where $o(1) \to 0$ as $\Delta \to 0$.

We now move to examining the performance of the "smart receiver" (see Chapter 2.3.3) system, which uses the same transmitter as in the "naïve" approach, and a receiver which performs MMSE estimation, to compensate for the additional $\varepsilon_1$ factor. Since we assumed $1 \ll P_1 \ll P_2$, the inflation factor at the transmitter is $\alpha_T \approx 1$ and the receiver ought to eliminate any residual interference by selecting

$$\alpha_R = \mathrm{tg}(\theta + \varepsilon_1)\mathrm{ctg}(\theta)\alpha_T \approx \mathrm{tg}(\theta + \varepsilon_1)\mathrm{ctg}(\theta)\,.$$

By using first-order approximations ($\Delta \ll 1$), we arrive to the following expression for the effective noise:

$$N_{1,\mathrm{eff}}^{\varepsilon_1} \approx 2X_2 \csc(2\theta)\varepsilon_1 + \frac{1}{h_1 \sin(\theta)} N_1\,,$$

Again, since user 2, sees the same channel as in the linear ZF scheme, the following rates are achievable:

$$\begin{aligned} R_1 &\le \frac{1}{2}\log\left(1 + \frac{\mathrm{SNR}_1 h_1^2 \sin^2(\theta)}{\mathrm{SNR}_1 h_1^2 \sec^2(\theta)\Delta^2 + 1}\right) - o(1) \\ R_2 &\le \frac{1}{2}\log\left(1 + \frac{\mathrm{SNR}_2 h_2^2}{\mathrm{SNR}_1 h_2^2 \Delta^2 + 1}\right) - o(1)\,, \end{aligned} \tag{2.15}$$

where again $o(1) \to 0$ as $\Delta \to 0$.

Finally, we introduce randomization at the transmission, in the form of a random inflation factor $\alpha_T$ for user 1 (the one that performs DPC), to assist in enlarging its own rate ($R_1$), and "the same" angular fluctuation to the transmission directions of both users. The perturbations in the transmission direction of user 1, improves the performance of user 2, just like in the linear ZF case. The fluctuation in the transmission direction of user 2, causes no additional (first order) improvement, nevertheless, tilting both vectors ($\boldsymbol{t}_1$ and $\boldsymbol{t}_2$) by the same angle, facilitates in the choice of the two fluctuations, as will become apparent in the sequel. We denote by $\alpha_2$ the angular perturbations of $t_1$ and $t_2$, and by $\alpha_1$ - the perturbation in the amplitude of user 1 (in the inflation factor of $x_1$).

The channel to user 1, in this case, is therefore

$$\begin{aligned} Y_1 &= \left\langle \tilde{\boldsymbol{h}}_1, \boldsymbol{X}_1 \right\rangle + \left\langle \tilde{\boldsymbol{h}}_1, \boldsymbol{X}_2 \right\rangle + N_1 \\ &= X_1 h_1 \cos\left(\frac{\pi}{2} - \theta - (\varepsilon_1 - \alpha_2)\right) + X_2 h_1 \cos(\theta + (\varepsilon_1 - \alpha_2)) + N_1 \end{aligned}$$

$$= X_1 h_1 \sin(\theta + (\varepsilon_1 - \alpha_2)) + X_2 h_1 \cos(\theta + (\varepsilon_1 - \alpha_2)) + N_1 ,$$

and after dividing by $h_1 \sin(\theta + \varepsilon_1 - \alpha_2)$:

$$\tilde{Y}_1 = X_1 + \mathrm{ctg}(\theta + \varepsilon_1 - \alpha_2) X_2 + \frac{1}{h_1 \sin(\theta + \varepsilon_1 - \alpha_2)} N_1 .$$

The (lattice) modulo-additive channel, in this case, is:

$$\begin{aligned} X_1 &= [v_1 - \mathrm{ctg}(\theta - \alpha_2 + \alpha_1) X_2 - U] \mod \Lambda , \\ Y_1' &= [\alpha_R \tilde{y}_1 + U] \mod \Lambda \\ &= \left[v_1 + N_{1,\text{eff}}^{\varepsilon_1}\right] \mod \Lambda , \end{aligned}$$

where

$$\begin{aligned} N_{1,\text{eff}}^{\varepsilon_1} &= -(1 - \alpha_R) X_1 + \big(\alpha_R \mathrm{ctg}(\theta + \varepsilon_1 - \alpha_2) X_1 - \mathrm{ctg}(\theta - \alpha_2 + \alpha_1)\big) X_2 \\ &\qquad + \frac{\alpha_R}{h_1 \sin(\theta + \varepsilon_1 - \alpha_2)} N_1 \\ &\approx 2 \csc(2\theta)(\varepsilon_1 - \alpha_1) X_1 + \frac{1}{h_1 \sin(\theta)} N_1 , \end{aligned} \tag{2.16}$$

and (2.16) holds under first order approximation and the choice $\alpha_R = \mathrm{ctg}(\theta - \alpha_2 \alpha_1) \mathrm{tg}(\theta + \varepsilon_1 - \alpha_2)$, which eliminates any residual interference.

This implicates the achievability of any rate pair $(R_1, R_2)$ satisfying:

$$\begin{aligned} R_1 &\le \max_{f(\alpha_1)} \min_{\varepsilon_1} E_{\alpha_1} \frac{1}{2} \log\left(1 + \frac{\mathrm{SNR}_1 h_1^2 \sin^2(\theta)}{\mathrm{SNR}_1 h_1^2 \sec^2(\theta)(\alpha_1 - \varepsilon_1)^2 + 1}\right) - o(1) \\ R_2 &\le \max_{f(\alpha_2)} \min_{\varepsilon_2} E_{\alpha_2} \frac{1}{2} \log\left(1 + \frac{\mathrm{SNR}_2 h_2^2}{\mathrm{SNR}_1 h_2^2 (\alpha_2 - \varepsilon_2)^2 + 1}\right) - o(1) , \end{aligned} \tag{2.17}$$

where $o(1) \to 0$ as $\Delta \to 0$.

Just like in (2.12), the maximization problems in (2.17) are convex, which means that non-deterministic selections of $\alpha_1$ and $\alpha_2$ need to be made to achieve optimum.

**Remark 2.5.**

- The intervals of uncertainty might be of different size, i.e., $\Delta_1 \neq \Delta_2$. Nonetheless, the treatment in this section is easily extended to this case.

- Note that the crosstalk of user 1 in the DPC based scheme (the "weak" user which performs DPC) depends on the power of this user rather than the power of user 2, which is the case in the linear ZF approach. Hence, in the regime $P_2 \ll P_1$, this suggest a great improvement over linear schemes.

- Comparing the rates achievable by linear ZF, with the ones - via ZF DPC, that is, (2.11) with (2.15), and (2.12) with (2.17), we see that contrary to a common belief that the performance of robust DPC rapidly deteriorates with the growth of uncertainty in the channel coefficients, in many cases of interest, the opposite is true: for

$$\mathrm{SNR}_2 \cos^2(\theta) > \mathrm{SNR}_1 \,,$$

  both the deterministic and the "randomized" DPC schemes, (2.15) and (2.17), are in fact "more robust" then their linear counterparts, (2.11) and (2.12). Moreover, by comparing the rates of the "naïve" DPC scheme (2.14) with the ones of the parallel linear scheme (2.11), we see that even if the DPC user ignores the fact that there is a lack of channel knowledge, its performance does not collapse dramatically, and propose a trade-off between pairs that are achievable using linear schemes and those which are achievable using DPC.

# Chapter 3

# Binary Dirty Multiple-Access Channel

**The general two-user memoryless multiple-access channel with common channel state information among the encoders has no single-letter solution which explicitly characterizes its capacity region. In this chapter a binary "dirty" multiple-access channel with interference known at both encoders is considered. We derive an achievable rate region for this channel and determine its sum-rate capacity, which equals to the capacity when full-cooperation between transmitters is allowed, contrary to the Gaussian case.**

## 3.1 System Model and Motivation

We treat the binary DMAC with non-causal common interference, defined in Chapter 1.18, to be (see also Figure 1.8):

$$Y = X_1 \oplus X_2 \oplus S \oplus N \,,$$

with input ("power") constraints $\frac{1}{n} w_H(\mathbf{x}_i) \leq q_i$. We assume that the interference is "strong", i.e., $S \sim \text{Bernoulli}\,(1/2)$. This is the worst-case interference that can be assumed, as any other distribution of the interference can be transformed into this case by incorporating dithering at the receiver's end.

For simplicity, we concentrate on the "noiseless case" ($N = 0$).

In Chapter 1.4.2 a method to cancel known interference was discussed. In the sequel, we use a super-position of such codes along with successive decoding of the messages ("onion peeling") and derive an achievable rate region, which we conjecture to be the capacity region, and is equal to the binary clean MAC capacity region up to a loss which stems from the loss seen in the point-to-point case. Moreover, we show that these strategies achieve the sum-rate capacity of the binary dirty MAC with common interference, which is equal to the capacity of the clean MAC when full cooperation between the encoders is allowed.

## 3.2 Clean MAC

In this section we consider the "clean" binary modulo-additive channel:

$$Y = X_1 \oplus X_2 \oplus N$$

with input constraints $q_1, q_2$. Note that this channel is identical to our channel of interest (1.18) up to the interference which is equal to 0 in this case.

The capacity region of this channel contains the capacity region of the our channel of interest, and therefore serves as an outer bound on its capacity region. Again, we concentrate on the "noiseless case" ($N = 0$).

Contrary to the Gaussian case, in which only *stationary* inputs need to be considered to achieve capacity, in the binary case, the use of stationary inputs is not optimal and the convex hull is necessary to achieve the capacity region envelope. To see this we rewrite (1.17) explicitly for the binary case:

$$\begin{aligned} \mathcal{C} \triangleq \text{cl conv} \Big\{ (\mathrm{R}_1, \mathrm{R}_2) : R_1 &\leq H_b(X_1) \\ R_2 &\leq H_b(X_2) \\ R_1 + R_2 &\leq H_b(X_1 \oplus X_2) \Big\}, \end{aligned} \tag{3.1}$$

where the closure and the convex hull are taken over all admissible distributions of the form $p_1(x_1)p_2(x_2)$ on $\{0,1\} \times \{0,1\}$ and such that the input constraints are satisfied.

One easily verifies that, by allowing only stationary inputs in (3.1), .i.e., relinquishing the convex hull, the sum-rate $R_1 + R_2$ cannot exceed

$$R_1 + R_2 \leq H_b(q_1 \circledast q_2)\,. \tag{3.2}$$

However, as indicated by the following lemma, this is suboptimal.

**Lemma 3.1** (Sum-Rate Capacity of the Binary Clean MAC)**.** *The sum-rate capacity of the binary noiseless modulo-additive MAC with input constraints* $\frac{1}{n}w_H(\mathbf{x_i}) \leq q_i$, $i = 1, 2$, *is:*

$$C_{clean}^{sum} = H_b^+(q_1 + q_2)\,, \tag{3.3}$$

*where* $H_b^+(q) \triangleq H_b\left(\min\left\{q, \frac{1}{2}\right\}\right)$.

*Proof.*
*Direct:* Using time-sharing one can divide each block into two parts: in the first $\frac{q_1}{q_2+q_2}n$ block samples user 1 spends all of its power to convey his private message, while user 2 transmits 0, whereas in the remaining $\frac{q_2}{q_2+q_2}n$ block samples user 2 spends all of its transmission power to convey his message, while user 1 is silent. Thus in the first $\frac{q_1}{q_2+q_2}n$ samples, user 1 transmits over a binary (point-to-point) DP channel, whereas in the remaining samples, user 2 transmits over the same DP channel (with a different power constraint). This leads to the sum-rate

$$\begin{aligned} R_1 + R_2 &= \frac{q_1}{q_2 + q_2} H_b^+(q_1 + q_2) + \frac{q_2}{q_2 + q_2} H_b^+(q_1 + q_2) \\ &= H_b^+(q_1 + q_2)\,. \end{aligned}$$

*Converse:* Allow *full cooperation* between the transmitters. This can only increase the sum-rate capacity. Full cooperation transforms the problem into a point-to-point problem of transmitting over a binary clean channel with power constraint $\frac{1}{n}w_H(\mathbf{x}) \leq q_1 + q_2$, the capacity of which is $H_b^+(q_1 + q_2)$. □

Thus, the sum-rate capacity of the binary (clean) MAC (3.3) is strictly greater than the best rate achievable using only stationary inputs (3.2).

**Remark 3.1.**

- The sum-rate of the "noisy" clean MAC can be shown, using the same methods, to be:

$$C_{\text{clean}}^{\text{sum}} = H_b^+\Big((q_1+q_2) \circledast \varepsilon\Big) - H_b(\varepsilon)\,.$$

- If we allow full cooperation between the transmitters, the capacity of the channel does not outperform (3.3), as pointed out in the converse part of the proof. In the Gaussian case, on the other hand, the sum-rate capacity of the MAC channel is equal to $\frac{1}{2}\log(1+\text{SNR}_1+\text{SNR}_2)$, which is strictly smaller than the full-cooperation capacity, $\frac{1}{2}\log(1+\text{SNR}_1+\text{SNR}_2+2\sqrt{\text{SNR}_1\text{SNR}_2})$. This dissimilarity stems from the difference of the alphabets we work with in both problems and the nature of the addition: in the binary case, no "coherence" can be attained by transmitting the same message, and additional power can only assist in exploiting more time slots within a block. In the Gaussian case, on the other hand, cooperation allows additional coherence gain, which cannot be achieved otherwise.

To find the capacity region of (3.1) explicitly, we replace the convex hull with a time-sharing variable $Q$, with alphabet of size $|\mathcal{Q}| = 2$ (see, e.g., [13]).

$$\begin{aligned} \mathcal{C} \triangleq \bigcup \Big\{(R_1,R_2) : R_1 &\le H_b(X_1|Q) \\ R_2 &\le H_b(X_2|Q) \\ R_1+R_2 &\le H_b(X_1 \oplus X_2|Q)\Big\}\,, \end{aligned} \tag{3.4}$$

where the union is over all admissible Markov chains $X_1 \leftrightarrow Q \leftrightarrow X_2$, satisfying the input constraints $EX_i \le q_i$, $i = 1, 2$. The capacity region for $q_1 = 1/6$ and $q_2 = 1/10$ is depicted in Figure 3.1.

**Remark 3.2.** Note that $X_1$ and $X_2$ are not independent in (3.4), but rather independent *given* the time-sharing parameter $Q$.

Nonetheless, time-sharing of the form described in the proof of Lemma 3.1 is suboptimal in general, as can be seen in Figure 3.1.

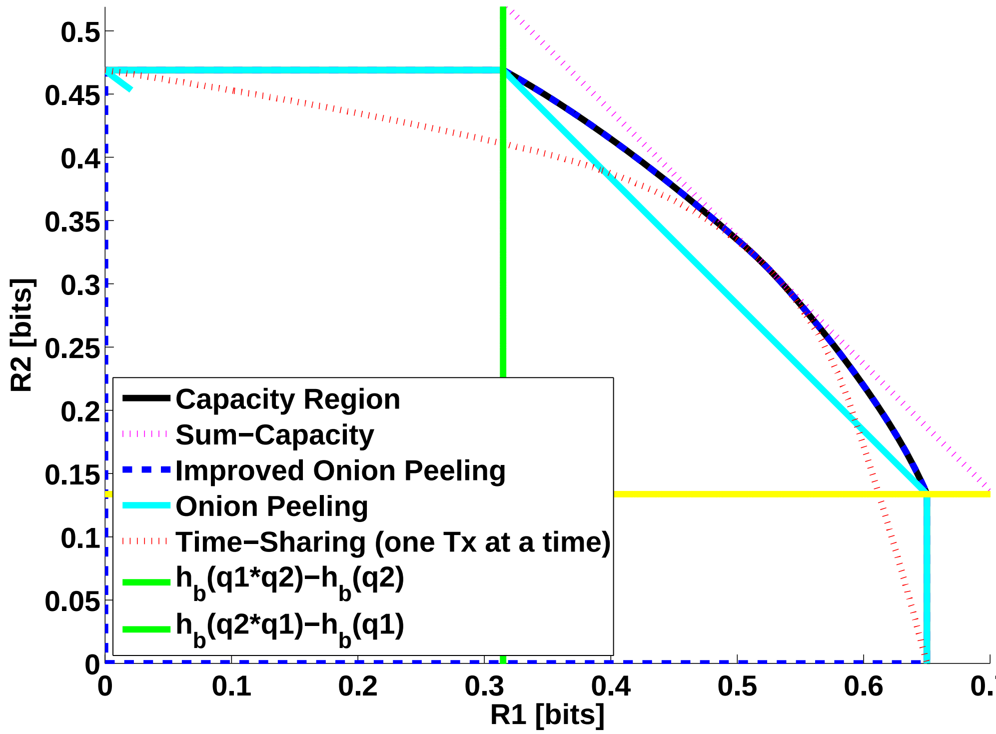


Figure 3.1: Rate Regions for the binary clean MAC and input constraints $q_1 = 1/6,\ q_2 = 1/10$.

### 3.2.1 Onion Peeling

Examining the capacity region in (3.4), we note that the corner points of the pentagons, which constitute the capacity region, i.e., points that satisfy one of the first two inequalities and the third one with equality, in (3.4), can be achieved by incorporating the "successive cancellation" method, a.k.a. "onion peeling", in which the decoder treats the message of one of the users as noise, recovers the message of the second user and subtracts it to recover the remaining message.

Due to the time-sharing variable $Q$ of cardinality 2, two such strategies need to be considered, to achieve a general point in the capacity region (3.4), such that the power constraints are satisfied on average. Nevertheless, we examine these rates for stationary points (viz. $P(Q = 0) = 1$) , to obtain better understanding. Thus, both users transmit simultaneously at all times, such that user 1 uses all of its available power $EX_1 = q_1$, whereas user 2 uses only some portion of its power $EX_2 = q_2'$ $(0 \le q_2' \le q_2)$. User 1 treats $q_2'$ as noise and can achieve a rate of $R_1 = H_b(q_1 \circledast q_2') - H_b(q_2')$. After recovering the message of user 1, it can be subtracted, such that user 2 sees a clean point-to-point channel and hence can achieve a rate of $R_2 = H_b(q_2')$. Note that even though the capacity region cannot be achieved using this strategy capacity, it does achieve certain rate pairs which cannot be achieved by simple time-sharing, like the one used in Lemma 3.1 and depicted in Figure 3.1.

**Remark 3.3.**

- When using this onion peeling strategy, user 2 does not exploit all of its power, but only a portion $0 \le q_2' \le q_2$. Hence a "residual" power of $q_2 - q_2'$ is left unexploited. This implies that this strategy is not optimal, as is, and a way to exploit this residual power needs to constructed.

- The roles of user 1 and user 2 are not symmetric: the achievable rate pairs, using onion peeling, when user 2 is peeled, differ from the rate pairs that are achieved when user 1 is peeled. Hence, by switching roles between the two users, one may achieve additional rate points.

### 3.2.2 Improving the Stationary Onion Peeling

The onion peeling considered in the previous section, can be easily improved, by dividing each block into two parts (i.e., time-sharing between two strategies): in the first sub-block, onion peeling is applied, whereas in the remaining time, the user that did not exploit all of its power during the first sub-block (viz. the onion peeling time period), uses all of its remaining "residual power", whereas the other user ought to transmit 0, as he already exploited all of its power. Let us denote by $\alpha$ the block portion allotted to onion peeling and by $q_2'$ $(0 \le q_2' \le q_2)$ the power of user 2, used during this period. Hence, the achievable rates, using this strategy, have the form:

$$\begin{aligned} R_1 &= \alpha H_b^+\left(\frac{q_1}{\alpha} \circledast \frac{q_2'}{\alpha}\right) - \alpha H_b^+\left(\frac{q_2'}{\alpha}\right), \\ R_2 &= \alpha H_b^+\left(\frac{q_2'}{\alpha}\right) + (1-\alpha) H_b^+\left(\frac{q_2 - q_2'}{1-\alpha}\right), \end{aligned} \tag{3.5}$$

**Remark 3.4.**

- In the noisy case, one may obtain rate pairs of the form

$$R_1 = \alpha H_b^+ \left( \frac{q_1}{\alpha} \circledast \frac{q_2'}{\alpha} \circledast \varepsilon \right) - \alpha H_b^+ \left( \frac{q_2'}{\alpha} \circledast \varepsilon \right) ,$$
$$R_2 = \alpha H_b^+ \left( \frac{q_2'}{\alpha} \circledast \varepsilon \right) + (1-\alpha) H_b^+ \left( \frac{q_2 - q_2'}{1-\alpha} \circledast \varepsilon \right) - H_b(\varepsilon)$$

  in a similar manner.

- Like in "stationary" onion peeling, the roles of the two users are not symmetrical. Hence by switching roles between the two users, one achieves points within the capacity region, that could not be achieved otherwise. See Remark 3.3.

In fact, by plotting the achievable rate region of this "improved" onion peeling scheme, for different cases (different $\{q_i\}$ values), [1] we see that it coincides with the capacity region (3.4). We conjecture that this is always true and that the capacity region of the binary modulo-additive MAC can be entirely described by (3.5) (and the symmetric expression in which user 1 and user 2 switch roles).

## 3.3 Dirty MAC with Common Interference

We adopt the strategies introduced in Chapter 3.2, to the dirty case, and derive an achievable rate region.

### 3.3.1 Sum-Rate Capacity

Similarly to the clean MAC case, the sum-rate capacity of the binary DMAC with common interference is equal to the capacity of this channel when both encoders can fully cooperate, as indicated by the following lemma.

**Lemma 3.2** (Sum-Rate Capacity of DMAC with Common SI)**.** *The sum-rate capacity of the binary noiseless modulo-additive dirty MAC with common*

[1] We do so for both when user 2 is peeled first and when user 2 is peeled first.

*interference and input constraints* $\frac{1}{n}w_H(\mathbf{x_i}) \leq q_i$, $i = 1, 2$, *is:*

$$C_{dirty}^{sum} = H_b^+(q_1 + q_2)\,. \tag{3.6}$$

*Proof.*

*Direct:* We repeat the proof of Lemma 3.1, only now the point-to-point BSC capacity (1.6) should be replaced by the binary DP channel capacity (1.5). Nevertheless, in the noiseless case ($N = 0$), there is no difference between the two expressions, and thus

$$R1 + R2 = H_b^+(q_1 + q_2)\,.$$

*Converse:* Again, like in the proof of Lemma 3.1, we allow *full cooperation* between the transmitters, which in turn transforms the problem into a point-to-point channel, the capacity of which is $H_b^+(q_1 + q_2)$. □

**Remark 3.5.**

- In the presence of noise, the sum-rate capacity of this channel is

$$C_{\mathrm{Dirty}}^{\mathrm{sum}} = \mathrm{uch}\max\left\{\mathrm{H_b^+(q_1 + q_2) - H_b(\varepsilon), 0}\right\}\,.$$

- In the noiseless case, the sum-rate capacities of the binary clean and dirty MACs are equal. However, in the presence of noise $N$, the sum-rate capacity of the dirty MAC channel is strictly smaller than that of the clean MAC channel (for $q_1 + q_2 < \frac{1}{2}$). This difference stems from the capacity loss, due to the presence of interference. in the point-to-point setting, as discussed in Chapter 1.4.2.

- As was mentioned in Remark 3.1, if we allow full cooperation between the transmitters, the capacity of the channel cannot exceed (3.6), in contrast to the Gaussian case, in which additional “coherence gain” can be achieved.

### 3.3.2 Onion Peeling

The capacity region of the “single informed user” (see Chapter 1.4.8) serves as an inner bound for the capacity region of the common interference dirty

MAC. To improve the achievable region of our channel of interest, we allow time-sharing between "single informed user" strategies, where the informed user is alternately user 1 and user 2.

The strategies used by Philosof, Zamir and Erez [37] to achieve the capacity region of the single informed user (1.19), can be viewed as onion peeling, where user 2 transmits in the same way, as described in Chapter 3.2.1, i.e., assumes a clean channel and input constraint $0 \le q_2' \le q_2$; and user 1 treats the signal of user 2, $X_2$, as noise and uses dirty paper coding of the form (1.5). The achievable rates, using this strategy, are of the form:

$$
\begin{aligned}
R_1 &= H_b(q_1) - H_b(q_2')\,, \\
R_2 &= H_b(q_2')\,,
\end{aligned}
$$

where since $q_2'$ can take any value in the interval $[0, q_2]$, the single informed user capacity (1.19) is achieved by time-sharing between such strategies.

**Remark 3.6.**

- Using such "stationary"strategy alone (with no time-sharing), one cannot hope to achieve the sum-rate capacity of Lemma 3.2 (or the whole capacity region of the single-informed user problem (1.19)).
- As in Remark 3.3, there is an average residual power of $q_2 - q_2'$, for each sample, left unexploited.
- This strategy is asymmetric in user 1 and user 2, as was explained in Remark 3.3.

### 3.3.3 Improved Onion Peeling

As was done for the clean MAC in Chapter 3.2.2, we improve the achievable rate region of "stationary" onion peeling, by employing time-sharing between an onion peeling strategy (discussed in Chapter 3.3.2) and transmission of the residual power (where the other user is silent). This allows to achieve rate pairs of the form:

$$
R_1 = \alpha H_b^+\left(\frac{q_1}{\alpha}\right) - \alpha H_b^+\left(\frac{q_2'}{\alpha}\right)\,,
$$

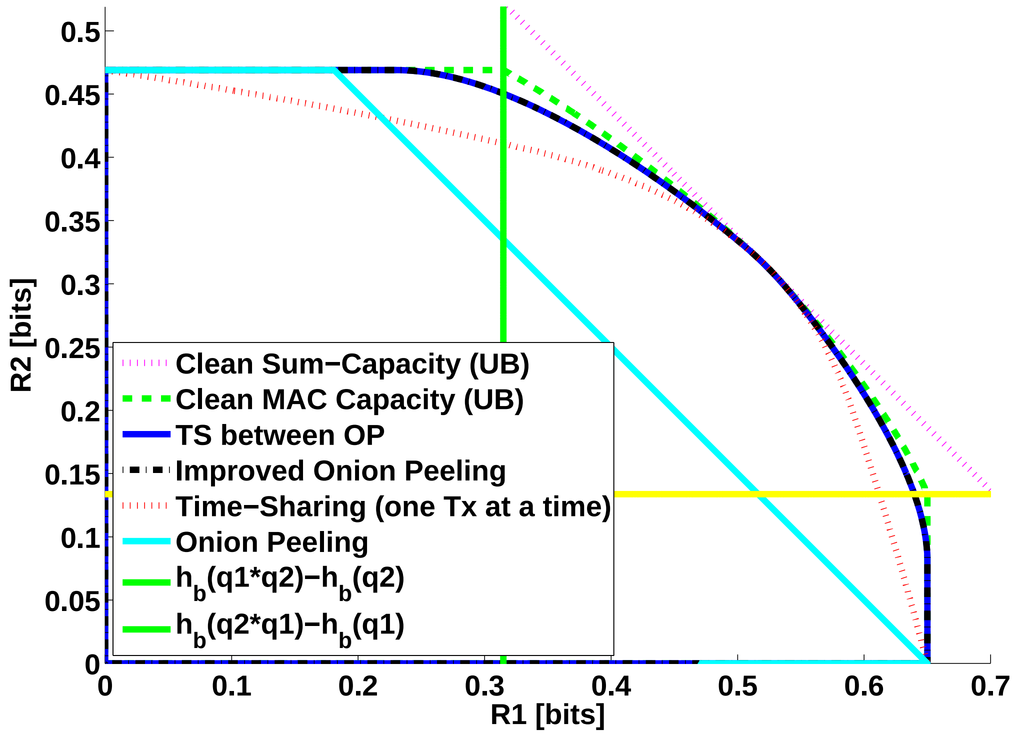


Figure 3.2: Rate regions for the binary DMAC with common interference and input constraints $q_1 = 1/6,\ q_2 = 1/10$.

$$R_2 = \alpha H_b^+\left(\frac{q_2'}{\alpha}\right) + (1-\alpha)H_b^+\left(\frac{q_2 - q_2'}{1-\alpha}\right)\ , \tag{3.7}$$

where $q_2' \in [0, q_2]$. See Figure 3.2.

**Remark 3.7.**

- Even in the noiseless case ($N = 0$) the achievable rate region of the dirty channel (3.7) is properly contained in its corresponding clean counterpart (3.5), as depicted in Figure 3.2. This difference stems from the fact that in the first sub-block, the user being peeled first, treats the signal of the other user as *noise*, in the presence of *interference*. Hence the achievable rate during this stage is strictly smaller, due to the point-to-point loss of binary DP, discussed in Chapter 1.4.2.

- In the noisy case, one may obtain rate pairs of the form

$$R_1 = \alpha H_b^+\left(\frac{q_1}{\alpha}\right) - \alpha H_b^+\left(\frac{q_2'}{\alpha} \circledast \varepsilon\right),$$
$$R_2 = \alpha H_b^+\left(\frac{q_2'}{\alpha} \circledast \varepsilon\right) + (1-\alpha) H_b^+\left(\frac{q_2 - q_2'}{1-\alpha} \circledast \varepsilon\right) - H_b(\varepsilon)$$

  in a similar manner.

- The roles of the two users are not symmetrical. Hence by switching roles between the two users, one achieves points within the capacity region, that could not be achieved otherwise. See Remark 3.3.

One could suspect that allowing time-sharing between a few onion peeling strategies can further improve this region. Nevertheless, by plotting the corresponding rate region (see Figure 3.2), one sees that no further improvement is gained this way. Thus, we conjecture, like for the clean MAC, that strategies of the form (3.7), are in fact sufficient to achieve any point in the capacity region of the binary DMAC with common interference. Nevertheless, the problem of determining the capacity region of this problem remains open.

# Chapter 4

# Summary

In this work, the compound dirty-paper channel was considered. The capacity of general compound channels with causal side information was determined. We further studied the performance that may be achieved for the specific case of the DP channel model by an extended Tomlinson-Harashima precoding scheme and derived lower bounds on the capacity of the channel. We derived the MMSE scaling that can be applied at the receiver to compensate for imprecise channel knowledge at the transmitter. We further showed that randomized inflation scaling at the transmitter may further improve the achievable rate. We also showed that the potential shaping gain of higher lattice dimensions diminishes with the increase of the channel estimation inaccuracy and that well designed dirty paper coding techniques for the MIMO broadcast channel can be even more robust than their linear counterparts.

This work focused exclusively on the performance achievable using THP-like schemes. It would be interesting to obtain an upper bound on the capacity (without any restriction on the coding technique) for the noiseless DP channel under channel uncertainty.

In the second part of this work, we studied the binary dirty multiple-access channel with common interference. We first studied the clean binary multiple-access channel. Using our understanding of the clean multiple-access channel, we were able to derive the sum-rate capacities of both the clean and the dirty MAC, which were shown to be equal to the capacities of these channels when full cooperation between the encoders is allowed. We also derived an achievable rate region for the clean and the dirty multiple-access channels and conjectured that the capacities of both of these channels can be described in a simple manner, which corresponds to a simple successive decoding technique.

# Appendix A

## A.1 Proof of Theorem 2.1

*Direct.* Denote by $\mathcal{T}$ the family of all mappings from $\mathcal{S}$ to $\mathcal{X}$. Use a transmitter that sends $x = t(s)$, where $t$ is chosen in an i.i.d. manner, according to some predefined probability distribution $p(t)$. In this case, the problem reduces to that of a compound channel with *no side information*, with an input alphabet $\mathcal{T}$ (see [42]), the same output alphabet $\mathcal{Y}$ and the corresponding transition probabilities

$$p(y|t) = \sum_{s \in \mathcal{S}} p(y|x = t(s), s)\,.$$

Hence, by maximizing over all possible input probabilities, $p(t)$, of the equivalent channel, we have an inner bound on the (worst-case) capacity (see, e.g., [4]):

$$C \geq \_p(t) \in \mathcal{P}(\mathcal{T}) \inf_{\beta \in \mathcal{B}} I_\beta(T;Y)\,.$$

*Converse.* For each $n$, let the information message $W$ be drawn according to a uniform distribution over $\{1, ..., 2^{nR}\}$. Denote the error probability corresponding to $\beta \in \mathcal{B}$ by $P_{e,\beta}^{(n)}$ and the error probability of the scheme as the supremum of these probabilities, $P_e^{(n)} \triangleq \sum_{\beta \in \mathcal{B}} P_{e,\beta}^{(n)}$. Then we have:

$$\begin{aligned} nR = H(W) &\leq 1 + P_{e,\beta}^{(n)} nR + I_\beta(W;Y_1^n) \\ &\leq 1 + P_e^{(n)} nR + \sum_{i=1}^{n} I_\beta(W;Y_i|Y_1^{i-1})\,, \end{aligned}$$

where the first inequality is due to Fano's inequality (see, e.g., [12]) and the second inequality follows from the chain-rule for mutual information. By retracing the steps of Shannon in [42], we have $I_\beta(W;Y_i|Y_1^{i-1}) \le I_\beta(W,S_1^{i-1};Y_i)$, for every $\beta \in \mathcal{B}$. Since $\{W,S_1^{i-1}\}$ does not depend on the value of $\beta$, the following inequality holds true as explained in detail in [42]:

$$nR \le P_e^{(n)} nR + \sum_{i=1}^{n} I_\beta(T_i;Y_i), \qquad \forall \beta \in \mathcal{B}\,.$$

The inequality above needs to be held for all $\beta \in \mathcal{B}$ simultaneously, and hence can be rewritten as

$$nR \le \max_{p(t)\in\mathcal{P}(\mathcal{T})} \inf_{\beta\in\mathcal{B}} P_e^{(n)} nR + nI_\beta(T;Y)\,.$$

Finally, dividing by $n$, taking $P_e \to 0$ and letting $n \to \infty$, we obtain

$$R \le \max_{p(t)\in\mathcal{P}(\mathcal{T})} \inf_{\beta\in\mathcal{B}} I_\beta(T;Y_i)\,.$$

□

## A.2 Proof of Lemma 2.2 and treatment for $\Delta > 1/3$

*Proof of Lemma 2.2.* The term

$$I_\beta(V;Y'|\alpha) = h_\beta(Y'|\alpha) - h_\beta(Y'|V,\alpha)$$

is maximized by taking $V \sim \mathrm{Unif}(\Lambda)$. Moreover, it is easily seen that the support of $f(\alpha)$ should be restricted to $\mathcal{I}_\Delta$. It follows that,

$$\begin{aligned} I_\beta(V;Y'|\alpha) &= h_\beta(Y'|\alpha) - h_\beta(Y'|V,\alpha) \\ &= \log(L) - h_\beta(Y'|V,\alpha) \\ &= \log(L) - h([N_{\text{eff}}^\beta] \bmod \Lambda) \\ &= \log(L) - E_\alpha\left[h\left(\left[\frac{\alpha-\beta}{\alpha}U\right] \bmod \Lambda\right)\right]. \end{aligned}$$

The term $\frac{\alpha-\beta}{\alpha}$ is maximized when $\alpha = 1-\Delta$ and $\beta = 1+\Delta$, and is equal to $\frac{2\Delta}{1-\Delta}$. Hence, for $\Delta \leq \frac{1}{3}$, we have $\frac{\alpha-\beta}{\alpha} \leq 1$. Therefore,

$$\begin{aligned} I_\beta(V;Y'|\alpha) &= \log(L) - E_\alpha\left[h\left(\frac{\alpha-\beta}{\alpha}U\right)\right] \\ &= \log(L) + E_\alpha\left[-\log(\Delta) - \log\left|\frac{\alpha-\beta}{\alpha}\right|\right] \\ &= -E_\alpha \log\left|\frac{\alpha-\beta}{\alpha}\right| . \end{aligned}$$

□

The case of $\Delta > 1/3$ can be treated in a similar manner by employing the following lemma.

**Lemma A.1.** *Suppose* $U \sim \mathrm{Unif}(\mathcal{V}_0)$. *Then for every* $a > 1$, *the entropy of* $\big([aU] \bmod \Lambda\big)$ *is bounded by*

$$\log(L) - \log\left(\frac{\lceil a\rceil}{a}\right) \leq h([aU] \bmod \Lambda) \leq \log(L)\,.$$

*Proof.* The upper-bound follows easily from the fact that differential entropy is maximized by a uniform distribution, when subject to an amplitude constraint, see, e.g. [12]. To prove the lower-bound, note that there is a unique index $k \in \mathbb{Z}$ which satisfies

$$aU = [aU] \bmod \Lambda + kL,$$

with alphabet cardinality $|\mathcal{K}| = \lceil a\rceil$. One may easily verify [39], that the following relation holds

$$h(aU) = h([aU] \bmod \Lambda) + H(k|[aU])\,,$$

which leads to the desired bound:

$$\begin{aligned} h([aU] \bmod \Lambda) &= \log(aL) - H(k|[aU]) \\ &\geq \log(aL) - H(k) \\ &= \log(aL) - \log(a+1) \end{aligned}$$

$$= \log(L) - \log\left(\frac{\lceil a \rceil}{a}\right).$$

$\square$

# Bibliography


[1] R. Ahlswede. Multi-way communication channels. In *Proceedings of 2nd International Symposium on Information Theory, Thakadsor, Armenian S.S.R., Sept. 1971. Akadémiai Kiadó, Budapest, 1973*, pages 23–52.

[2] R. J. Barron, B. Chen, and G. W. Wornell. The duality between information embedding and source coding with side information and some applications. *IEEE Trans. Information Theory*, 49:1159–1180, May 2003.

[3] A. Bennatan, V. Aggarwal, Y. Wu, A. R. Calderbank, J. Hoydis, and A. Chindapol. Bounds and lattice-based transmission strategies for the phase-faded dirty-paper channel. *IEEE Trans. Wireless Communications*.

[4] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacity of a class of channels. *The Annals of Mathematical Statistics*, 30:1229–1241, Dec. 1959.

[5] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran. Quantized vs. analog feedback for the MIMO broadcast channel: A comparison between zero-forcing based achievable rates. In *Proc. Int. Symp. Inform. Theory (ISIT), Nice, France*, pages 2046–2050, June 2007.

[6] G. Caire and S. Shamai (Shitz). On the achievable throughput of a multi-antenna Gaussian broadcast channel. *IEEE Trans. Information Theory*, IT–49:1649–1706, July 2003.

[7] B. Chen and G. W. Wornell. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Trans. Information Theory*, IT–47:1423–1443, May 2001.

[8] R. Chen, R.W. Heath Jr., and J.G. Andrews. Transmit selection diversity for unitary precoded multiuser spatial multiplexing systems with linear receivers. *IEEE Trans. Signal Processing*, vol. 55, no. 3.

[9] A. S. Cohen and A. Lapidoth. The Gaussian watermarking game. *IEEE Trans. Information Theory*, June, 2002, 1639–1667. See also, *On the Gaussian Watermarking Game*, Proc. Int. Symp. Inform. Theory (ISIT), page 48, Sorrento, Italy, June 2000.

[10] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups.* Springer-Verlag, New York, N.Y., 1988.

[11] M. H. M. Costa. Writing on dirty paper. *IEEE Trans. Information Theory*, IT-29:439–441, May 1983.

[12] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, New York, 1991.

[13] I. Csiszar and J. Korner. *Information Theory - Coding Theorems for Discrete Memoryless Systems.* Academic Press, New York, 1981.

[14] R. L. Dobrushin. Optimal information transmission over a channel with unknown parameters (in Russian). *Radiotekh. i Elektron.*, Vol. 4, No. 12:1961–1956, Dec. 1959.

[15] J. J. Eggers, R. Bäumel, R. Tzschoppe, and B. Girod. Scalar Costa scheme for information embedding. *IEEE Trans. Signal Processing*, pages 1003–1019, Apr., 2003.

[16] U. Erez, S. Litsyn, and R. Zamir. Lattices which are good for (almost) everything. *IEEE Trans. Information Theory*, IT-51:3401–3416, Oct. 2005.

[17] U. Erez, S. Shamai (Shitz), and R. Zamir. Capacity and lattice strategies for cancelling known interference. *IEEE Trans. Information Theory*, pages 3820–3833, Nov. 2005.

[18] G. D. Forney, Jr. On the role of MMSE estimation in approaching the information- theoretic limits of linear Gaussian channels: Shannon meets Wiener. In *41st Annual Allerton Conference on Communication, Control,*

*and Computing, Allerton House, Monticello, Illinois*, pages 430–439, Oct. 2003.

[19] S. I. Gel'fand and M. S. Pinsker. Coding for channel with random parameters. *Problemy Pered. Inform. (Problems of Inform. Trans.)*, 9, No. 1:19–31, 1980.

[20] S. I. Gel'fand and M. S. Pinsker. On Gaussian channels with random parameters. In *Abstracts of Sixth International Symposium on Information Theory, Tashkent, U.S.S.R*, pages 247–250, Sep. 1984.

[21] P. Grover and A. Sahai. What is needed to exploit knowledge of primary transmissions? *available at arXiv:cs/0702071.*

[22] H. Harashima and H. Miyakawa. Matched-transmission technique for channels with intersymbol interference. *IEEE Trans. Communications*, COM-20:774–780, Aug. 1972.

[23] S. A. Jafar. Capacity with causal and non-causal side information - a unified view. *IEEE Trans. Information Theory*, IT-52:5468–5475, Dec. 2006.

[24] N. Jindal. MIMO broadcast channels with finite-rate feedback. *IEEE Trans. Information Theory*, Vol. 52, No. 11:5045–5060, Nov. 2003.

[25] A. Khina and U. Erez. On the robustness of dirty paper coding. *IEEE Trans. Communications*, IT-58, accepted for publication, May 2010.

[26] A. Khina and U. Erez. On robust dirty paper coding. In *Proceedings of IEEE Information Theory Workshop, ITW 2008, Porto, Portugal*, pages 204–208, May 2008.

[27] Y. H. Kim, A. Sutivong, and S. Sigurjónsson. Multiple user writing on dirty paper. In *Proceedings of IEEE International Symposium on Information Theory, Chicago, USA*, June 2004.

[28] S. Kotagiri and J. N. Laneman. Multiple access channels with state information known at some encoders. *IEEE Trans. Information Theory*, July 2006, submitted for publication.

[29] A. V. Kuznetsov and B. S. Tsybakov. Coding in a memory with defective cells. *translated from Prob. Peredach. Inform.*, 10:52–60, April-June, 1974.

[30] E. A. Lee and D. G. Messerschmitt. *Digital Communication.* Kluwer, 3rd ed. Newell, MA, 2004.

[31] J. Lee and N. Jindal. High SNR analysis for MIMO broadcast channels: Dirty paper coding versus linear precoding. *IEEE Trans. Information Theory*, Vol. 53, No. 12:4787–4792, Dec. 2007.

[32] H. Liao. *Multiple Access Channels.* PhD thesis, Department of Electrical Engineering, University of Hawaii, Honolulu, 1972.

[33] P. Mitran, N. Devroye, and V. Tarokh. On compound channels with side information at the transmitter. *IEEE Trans. Information Theory*, 52:1745–1755, April 2006.

[34] T. Philosof, A. Khisti, U. Erez, and R. Zamir. Lattice strategies for the dirty multiple access channel. In *Proceedings of IEEE International Symposium on Information Theory, Nice, France*, June 2007.

[35] T. Philosof and R. Zamir. On the loss of single-letter characterization: The dirty multiple access channel. *IEEE Trans. Information Theory*, IT-55:2442–2454, June 2009.

[36] T. Philosof, R. Zamir, and U. Erez. Technical report: Achievable rates for the MAC with correlated channel-state information. In *arXiv:0812.4803*, 2008.

[37] T. Philosof, R. Zamir, and U. Erez. The capacity region of the binary dirty MAC. In *Proc. of Info. Th. Workshop, Sicily, Italy*, pages 273–277, Oct. 2009.

[38] T. Philosof, R. Zamir, U. Erez, and A. Khisti. Lattice strategies for the dirty multiple access channel. *arXiv:0904.1892v1, also submitted to Transaction on Information Theory*, 2009.

[39] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes.* Holden Day, San Francis. CA., 1964.

[40] N. Jindal S. Vishwanath and A. Goldsmith. Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels. *IEEE Trans. Information Theory*, 49:2658–2668, Oct. 2003.

[41] C. E. Shannon. Two-way communication channels. In *Proceedings of 4th Berkley Symposium on math. stat. and prob.*, volume 1, pages 611–644, 1962.

[42] C. E. Shannon. Channels with side information at the transmitter. *IBM Journal of Research and Development*, 2:289–293, Oct. 1958.

[43] S. Sigurjónsson and Y. Kim. On multiple user channels with causal state information at the transmitters. *cs.IT/0508096*, 2005. Available: http://arxiv.org.

[44] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdu. Cooperative multiple-access encoding with states available at one transmitter. *IEEE Trans. Information Theory*, IT-54:4448–4469, Oct. 2008.

[45] M. Tomlinson. New automatic equalizer employing modulo arithmetic. *Electronic Lett.*, vol. 7:138–139, Mar. 1971.

[46] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz). The capacity region of the Gaussian multiple-input multiple-output broadcast channel. *IEEE Trans. Information Theory*, 52:3936–3964, Sept. 2006.

[47] F. M. J. Willems. On Gaussian channels with side information at the transmitter. In *Proc. of the Ninth Symposium on Information Theory in the Benelux*, Enschede, The Netherlands, 1988.

[48] J. Wolfowitz. Simultaneous channels. *Arch. Rational Mech. Anal.*, 4:371–386, 1960.

[49] J. Wolfowitz. *Coding Theorems of Information Theory.* Springer-Verlag, Berlin Heidelberg, New York, third edition, 1978.

[50] S. Yang and J. Belfiore. The impact of channel estimation error on the DPC region of the two-user Gaussian broadcast channel. In *43th Annual Allerton Conference on Communication, Control, and Computing, Allerton House, Monticello, Illinois*, Oct. 2002.

[51] W. Yu and J. Cioffi. Sum capacity of a Gaussian vector broadcast channel. *IEEE Trans. Information Theory*, IT-50:1875–1892, Sept. 2004.

[52] R. Zamir and M. Feder. On lattice quantization noise. *IEEE Trans. Information Theory*, IT-42:1152–1159, July 1996.

[53] R. Zamir, S. Shamai (Shitz), and U. Erez. Nested linear/lattice codes for structured multiterminal binning. *IEEE Trans. Information Theory*, 48:1250–1276, June 2002.