UNIVERSITAT POLITÈCNICA DE CATALUNYA
Department of Signal Theory and Communications

Master's Thesis

Expected Distortion with Fading MIMO Channel and Side Information Quality

by

Iñaki Estella Aguerri

Supervisor:

Dr.Deniz Gündüz (CTTC)

Barcelona, January 2011

Abstract

English

We consider the joint source-channel coding problem of sending a Gaussian source over a multiple input-multiple output (MIMO) fading channel when the decoder has additional correlated side information whose quality is also time-varying. We assume a block fading model for both the channel and side information qualities, and assume perfect state information at the receiver, while the transmitter has only a statistical knowledge. We are interested in the expected squared-error distortion for this system.

We study separate source-channel coding, uncoded transmission and two joint source-channel transmission schemes based on joint decoding at the receiver: NBJD, that uses no explicit binning and joint decoding of the side information and the channel output at the decoder and HDA, that compresses the source and transmits the error. At the decoder, the quantized codeword is recovered by means of joint decoding of the error and the side information. We extend such techniques to hybrid digital-analog and multi-layer schemes. We study numerically the problem and give results in the finite SNR regime. We provide closed form expressions for the distortion exponent in the high SNR regime.

Keywords

Joint source-channel coding, Distortion exponent, fading channel, fading sideinformation, Gaussian sources.

Contents

Abstract **i**

Table of Contents **iii**

List of Acronyms & Abbreviations **v**

1 Introduction **1**

2 Background **5**
- 2.1 System model 5
- 2.2 The joint source-channel problem with CSIST 6
  - 2.2.1 MMSE estimation 7
- 2.3 Expected distortion function: when CSIST is not available. 8
- 2.4 Characterization at high SNR: distortion exponent 8

3 Transmission Schemes **11**
- 3.1 Expected distortion lower bound 12
- 3.2 Single layer schemes: separation, uncoded and NBJD 12
  - 3.2.1 Separate source and channel coding 12
  - 3.2.2 Separate source and channel coding without binning 14
  - 3.2.3 Uncoded transmission 14
  - 3.2.4 No binning with joint decoding (NBJD) 15
  - 3.2.5 NBJD-analog hybrid scheme 15
- 3.3 Single layer schemes: HDA 17
  - 3.3.1 Hybrid analog-HDA when CSIST is available 18
  - 3.3.2 MIMO HDA 21
- 3.4 Multi-layers schemes 25
  - 3.4.1 Multi-layer separate coding without binning 25
  - 3.4.2 Multi-layer NBJD 26
- 3.A Proof of the expected distortion lower bound 28
- 3.B Proof of distortion for uncoded transmission 28
- 3.C Proof of distortion for NBJD-analog hybrid scheme 28
  - 3.C.1 Case $M_* b \leq 1$ 28
  - 3.C.2 Case $M_* b > 1$ 29

4 Numerical Results **31**
- 4.1 Numerical optimization 31

5 High SNR Results: Distortion Exponent Analysis **35**
5.1 Distortion exponent upper bound 35
5.2 Single layer schemes: separation, uncoded and NBJD 35
5.3 Single layer schemes: HDA 37
5.4 Multi-layer schemes 37
5.5 Comments on the results 39
5.5.1 MIMO and MISO results 39
5.5.2 MIMO multi-layer results 40
5.6 SISO results 40
5.6.1 Comments on SISO results 43
5.A Proof for distortion exponent: lower bound 44
5.B Proof for distortion exponent: separate source and channel coding 46
5.C Proof for distortion exponent: uncoded 49
5.D Proof for distortion exponent : NBJD 51
5.E Proof for distortion exponent: hybrid digital-analog, case $M_* b \leq 1$ 54
5.F Proof for distortion exponent: hybrid digital-analog, case $M_* b > 1$ 56
5.G Proof for distortion exponent: HDA 58
5.H Proof for distortion exponent: multi-layer no binning 61
5.I Proof for distortion exponent: multi-layers NBJD 63
5.J Proof for distortion exponent: multi-layers NBJD with improved power allocation 66

6 Conclusions **71**

List of Figures **72**

Bibliography **73**

List of Acronyms & Abbreviations

SISO	Single Input-Single Output systems
MIMO	Multiple Input-Multiple Output systems
MISO	Multiple Input-Single Output systems
SIMO	Single Input-Multiple Output systems
NBJD	Non Binning Joint Decoding
HDA	Hybrid Digital-Analog
SNR	Signal to Noise ratio
CSI	Channel State Information available at the transmitter and the receiver
MMSE	Minimum Mean Squared Error
CSIST	Channel and Side Information State at the Transmitter

Chapter 1

Introduction

Many applications in wireless networks require the transmission of a source signal over a fading channel, to be reconstructed with the minimum distortion possible, i.e., multimedia signals over cellular networks or the accumulation of local measurements at a fusion center in sensor networks. In classical networks, transmission is based on Shannon's separation theorem, in which the source is first compressed at the minimum possible distortion and then transmitted over the channel. While source-channel separation is optimal in point-to-point systems (under certain theoretical assumptions), it does not extend to many multi-user scenarios or non-ergodic channels.

In many practical scenarios, the destination receives additional correlated side information either form other transmitters in the network or through its own sensing devices. For example, measurements from other sensors at a fusion center, signals from repeaters in digital TV broadcasting or relay signals in future mobile networks. However, similar to estimating the channel state information at the transmitter, it is costly to provide an estimate of the available side information to the transmitter. Hence, the transmitter needs to transmit in a way to adapt dynamically to the time-varying channel and side information qualities without knowing their realizations. We model this scenario by a Gaussian source X to be transmitted over a block-fading channel to a receiver that has correlated side information Y, modeled also by a fading gain.

When the knowledge of the channel state information is available at both the transmitter and the receiver (CSI), Shannon's separation theorem claims that digital transmission is optimal in point to point communication under certain conditions. Digital transmission implies compressing the source and sending the compressed bits through the channel at a rate not greater than the capacity, determined by the actual channel state. This channel transmission is designed independently of the source statistics, i.e., the source and the channel codes are independent. We call this separate source and channel coding. However, in some practical scenarios, the channel state is not fully known and the transmitter has to blindly set a rate. If the channel is 'good' the rate is likely to be lower than the capacity and the transmission will be successful, but if the channel is 'bad' the message will not be decoded. Note also that digital transmission suffers from a threshold effect, that is, the quality of the received signal does not improve as the channel quality improves once it is above the rate of the used channel code. If the transmitter only knows the statistics of the channel and the side information gains, but not the realizations, the transmission scheme needs to be designed based on these statistics, and only an average performance in terms of the expected distortion can be guaranteed. For this digital schemes, a tradeoff between fidelity of the source code and the reliability of the channel appears.

On the other hand, it is known that uncoded transmission, in which the transmitter simply transmits a scaled version of the source, provides both optimality [1] and robustness in terms of distortion in point to point links. Robustness refers to the gradual decrease in the average distortion as the channel or side information quality improves.Robustness in uncoded transmission is because the threshold effect disappears. However, the optimality of uncoded transmission is lost when multiple antennas are used, or when the bandwidth ratio between the channel and the source is not one in the transmission.

Here, we consider the problem of transmitting a Gaussian source over a MIMO channel when the receiver has access to a time varying correlated side information. This problem, in the absence of side information at the receiver has been studied extensively in recent years. Gündüz and Erkip propose in [2] a layered broadcast scheme consisting of successive refinement codes and superposition coding which is shown to achieve the optimal exponential high SNR behavior in MISO/SIMO systems and in some ranges for MIMO systems. The minimum expected distortion for the layered broadcast scheme is studied for finite SNR in [3]. In [4], a hybrid digital-analog scheme is proposed that achieves the optimal high SNR behavior for low bandwidth ratios. The related problem of an uncertain side information when the channel is an error-free bit-pipe at a given rate is studied in [5]. In general, analog transmission provides robustness that benefits the overall performance of the systems, even in MIMO scenarios, and hence, hybrid digital-analog schemes to improve the performance at least in the finite SNR regime.

In the presence of side information, separate source and channel coding achieves the optimal distortion when the channel and the side information quality is known at the transmitter, while, uncoded is no more optimal unlike the scenario without side information. However, despite the lack of optimality when the channel and side information state at the transmitter (CSIST) is available, uncoded transmission can still provide robustness when CSIST is not available. Hence, our goal is to design transmission schemes that can benefit from both the higher transmission capability of digital codes and the robustness provided by uncoded transmission.

Apart from the classical hyrid digital-analog transmission schemes, a technique that naturally combines the features of digital and analog transmission has been recently presented in [6]. This scheme benefits from its digital nature while providing robustness similar to uncoded transmission.

In this thesis, we consider various transmission schemes for this problem and analyze the expected distortion performance achieved by these schemes. While we study the expected distortion in the finite SNR regime numerically, we also consider its high SNR behavior, characterized by the distortion exponent, and obtain closed-form expressions for the achieved distortion exponent.

The problem of identifying the minimum achievable expected distortion is an open problem even in the absence of side information. Hence, our numerical results are non conclusive and we provide comparisons among various schemes as was as an informed transmitter lower bound. The optimal distortion expression can be characterized for most cases in the absence of side-information. Unfortunately this is not the case when there is time-varying side-information. We provide lower and upper bounds for the distortion exponent as a function of the bandwidth ratio.

The rest of the thesis is organized as follows. In Chapter 2, we introduce the system model and the basic ideas used along this work. We start by briefly presenting the Expected Distortion (ED) optimization problem and the nature of the tradeoff we deal with. We then develop the ideas behind the high SNR asymptotic characterization of the ED function, and define the so called Distortion Exponent, Δ. A short review on MMSE estimation for Gaussian sources is also included.

We start Chapter 3 by providing a lower bound on the ED function. Some achievable schemes are then proposed in three sections. Firstly, a separation scheme based on Wyner-Ziv coding [7] is presented. We also propose a simpler digital scheme that refines the digital transmission with the side information. Pure analog transmission, in which the source is scaled and directly transmitted over the channel, is also considered in this section. We propose a joint source-channel coding scheme, based on the joint decoding idea of [8], for which the success of decoding the message depends on the joint quality of the channel and the side information. This techniques is called NBJD. Following [4] we extend this scheme to hybrid digital-analog transmission. We then study another joint source-channel coding scheme: HDA, introduced in [6], that generates a codeword at the transmitter to be jointly decoded using the side information available. We extend this scheme to MIMO systems and the bandwidth regime expansion. Using HDA we also provide a continuum of schemes that reduces to the scheme proposed by Lapidoth in [9] when there is no side information. We show that this continuum does not hold in the presence of side information. Following [2], the last section extends NBJD to two different multilayer schemes.

In Chapter 4, we study the expected distortion at finite SNR by numerically optimizing over the rates. We prove analytically that the joint decoding scheme outperforms separation at any finite SNR.

In Chapter 5, the performance at high SNR is studied through distortion exponent [10], which characterizes the high SNR slope of the expected distortion. We provide the general results for the MIMO scenario as well as the particularization to SISO.

Finally, the conclusions are drawn in Chapter 6.

Chapter 2

Background

We start this chapter by providing a rigorous model of the problem. Then, we give some insights on the joint source-channel coding problem when CSISIT is available or not. Finally, we provide the rigorous definition of the distortion exponent, that characterizes the high SNR behavior of the expected distortion function.

2.1 System model

We wish to transmit a zero mean, unit variance complex Gaussian source sequence $X^m \in \mathbb{C}^m$ of independent and identically distributed (i.i.d.) random variables, i.e. $X_i \sim \mathcal{CN}(0,1)$, over a MIMO block Rayleigh-fading channel with M_t transmit antennas and M_r receiver antennas. In addition to the channel output, correlated source side information is also available at the decoder. Similar to the channel model, a block fading model is used for the side information. The channel and the side information states are assumed to be constant for the duration of one block and independent among different blocks. The transmitter is required to transmit a block of m source samples over a block of n channel uses. We define the *bandwidth ratio* of the system as $b \triangleq \frac{n}{m}$ channel uses per source sample. We assume that m is large enough to achieve the rate-distortion performance of the underlying source sequence in the presence of side information, and n is large enough to design codes that can achieve all rates below the instantaneous capacity of the block fading channel.

The encoder maps the source sequence X^m to a channel input sequence $\mathbf{U} \in \mathbb{C}^{M_t \times n}$ using an encoding function $f^{(m, M_t \times n)} : \mathbb{C}^m \to \mathbb{C}^{M_t \times n}$ such that the average power constraint is satisfied: $Tr\{E[\mathbf{U}^\dagger \mathbf{U}]\} \leq M_t n$. The memoryless slow fading channel is modeled as

$$\mathbf{V}_i = \mathbf{H}\sqrt{\rho}\mathbf{U}_i + \mathbf{N}_i, \qquad i = 1, ..., n, \tag{2.1}$$

where $\mathbf{H} \in \mathbb{C}^{M_r \times M_t} \sim p_h(\mathbf{H})$ is the channel matrix with i.i.d. entries $\sim \mathcal{CN}(0,1)$, $\rho \in \mathbb{R}^+$ is the transmit power and $\mathbf{N}_i$ models the additive noise with $\mathbf{N}_i \sim \mathcal{CN}(0, \mathbf{I})$. Let $M^* = \max\{M_t, M_r\}$ and $M_* = \min\{M_t, M_r\}$. The decoder, in addition to $\mathbf{V} = [\mathbf{V}_1, ..., \mathbf{V}_n] \in \mathbb{C}^{n \times M_r}$, observes $Y^m \in \mathbb{C}^m$ with a random degradation modeled as,

$$Y_j = \gamma\sqrt{\rho_s}X_j + Z_j, \qquad j = 1, ..., m, \tag{2.2}$$

where $\gamma \in \mathbb{C} \sim p_\gamma(\gamma)$, $Z_j \sim \mathcal{CN}(0,1)$ and $\rho_s \in \mathbb{R}^+$ models the power of the side information. The decoder reconstructs the source sequence $\hat{X}^m = g(\mathbf{V}, Y^m, \mathbf{H}, \gamma)$ with a mapping $g^{(n \times M_r, m)} : \mathbb{C}^{n \times M_r} \times \mathbb{C}^m \times \mathbb{C}^{M_t \times M_r} \times \mathbb{C} \to \mathbb{C}^m$.

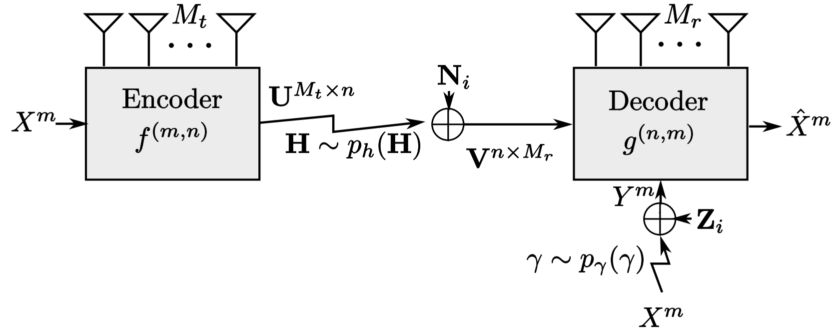

Figure 2.1: Block diagram of the joint source-channel coding problem with fading channel and side information qualities.

The distortion between the source sequence and the reconstruction is measured by

$$D \triangleq \frac{1}{m}\sum_{i=1}^{m} d(X_i,\hat{X}_i), \tag{2.3}$$

where $d(X_i,\hat{X}_i):\mathbb{C}\times\mathbb{C}\rightarrow\mathbb{R}^+$ is a single letter distortion measure. We focus on the quadratic distortion measure, i.e. $d(X_i,\hat{X}_i)=\|X_i-\hat{X}_i\|^2$.

We are interested in minimizing the *expected distortion* (ED), $E[D]$, where the expectation is taken with respect to the source and the side information realizations as well as the channel state and the channel noise.

In the considered scenario, the receiver knows the side information and the channel realization at each block. However, the encoder is only aware of the distribution of both parameters.

In this work, we are interested in characterizing the minimum expected distortion for each possible set of parameters $\Omega=(\rho,\rho_s,b)$,

$$ED^*(\Omega)\triangleq \min_{f,g} E[D]. \tag{2.4}$$

In next sections we try to give an intuitive vision of the problem using an example. We consider the scenario without side information for simplicity.

2.2 The joint source-channel problem with CSIST

In this section we consider the transmission of a source sequence when the channel state information, i.e. the actual realization of $\mathbf{H}$ is available at both the encoder and the decoder. Consider the classical digital transmission of a source sequence X^m over a channel with capacity $\mathcal{C}(\mathbf{H})$. First, the source sequence is quantized to a source codeword from a quantization codebook. The size of this codebook, that is given by 2^{mR_s} will depend on the channel realization. A channel codeword $\mathbf{U}^{M_t\times n}(i)$ is assigned to each of the quantization codewords, forming the 'channel codebook'. The size of the channel codebook is 2^{nR_c} i.e. arbitrarily close to the capacity so that the channel input can be recovered with high probability. Each $\mathbf{U}^{M_t\times n}(i)$ has to have a one to one correspondence with a $W(i)$, i.e. $R_s=bR_C$. For a given source sequence, a quantized codeword, indexed with i, is chosen and the corresponding channel codeword $\mathbf{U}(i)$ is used as the channel input. This process from the source sequence X^m to the channel input $\mathbf{U}$ is the encoding mapping, $f^{(m,M_t\times n)}:\mathbb{C}^m\rightarrow\mathbb{C}^{M_t\times n}$.

At the decoder, a channel output $\mathbf{V}^{M_r \times n}$ is available that depends on the channel distribution. In the example of digital transmission, the channel input is decoded, i.e. $\mathbf{U}(i)$ is uniquely recovered from the channel, as the channel codebook rate R_c has been design so that it is below the capacity of the channel, $\mathcal{C}$. Note that this design depends on the actual realization $\mathbf{H}$. Then, index i is known and the codeword $W(i)$ is available for the source reconstruction, that is directly $\hat{X}^m = W^m$.

In general, the correspondence of a channel output to a source sequence reconstruction is the mapping $g^{(n \times M_r, m)} : \mathbb{C}^{n \times M_r} \times \mathbb{C}^m \times \mathbb{C}^{M_t \times M_r} \times \mathbb{C} \to \mathbb{C}^m$. Along this work, the design of such mappings will consist on two basic stages. The first stage will consist on a decoding stage that will recover from the channel, either with or without the side information, as many information on the source sequence as possible, as the recovery of the quantized version of the source that is decoded from the digital channel output. This stage can, as we will see, include the recovery of source information in many ways. After obtaining all possible information, the source sequence is reconstructed. For our particular case of Gaussian source, the optimal way to reconstruct the source such that the mean square error between the source sequence and the reconstruction is minimized is by means of mmse estimation. It is well known, that the source sequence will be possible to be recovered with a distortion that is equal to $D = 2^{-R_s}$. In this case, is the minimum distortion is achieved by transmitting at the maximum possible channel rate R_c. In next subsection we give a brief overview of the techniques used to calculate the distortion achieved by mmse estimation depending on the available information.

2.2.1 MMSE estimation

The minimum mean square error (mmse) estimation is used to reconstruct the Gaussian source sequence at the receiver with available information $\mathbf{y}$. The mmse is defined as the the solution to the problem

$$\begin{aligned} mmse &= \min_{\hat{\mathbf{s}}=f(\mathbf{y})} E[(\mathbf{s} - \mathbf{f}(\mathbf{y}))^\dagger (\mathbf{s} - \mathbf{f}(\mathbf{y}))] \\ &= \min_{\hat{\mathbf{s}}=f(\mathbf{y})} Tr\{\mathbf{D}\} \end{aligned} \tag{2.5}$$

where the minimization is taken over all possible reconstruction functions $f(\mathbf{y})$, as $\hat{\mathbf{s}} = \mathbf{f}(\mathbf{y})$), and $\mathbf{D} = E_s[(\mathbf{s} - \mathbf{f}(\mathbf{y})))(\mathbf{s} - \mathbf{f}(\mathbf{y}))^\dagger]$ is the error covariance matric. The solution to this problem is known to be

$$\hat{\mathbf{s}} = E[\mathbf{s}|\mathbf{y}], \tag{2.6}$$

The optimal mmse estimator of a random vector $\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{R_s})$ under a linear model $\mathbf{y} = \mathbf{Hs} + \mathbf{N}$, where $\mathbf{H}$ is a known matrix and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \mathbf{R_N})$, is explicitly calculated as

$$\hat{\mathbf{s}} = (\mathbf{H^T R_N^{-1} H})^{-1} \mathbf{H^T R_N^{-1} y}, \tag{2.7}$$

where $\mathbf{R_N} = E[\mathbf{NN}^\dagger]$ and the corresponding error covariance matrix is given by

$$\mathbf{D} = (\mathbf{R_s^{-1}} + \mathbf{H^T R_N^{-1} H})^{-1}. \tag{2.8}$$

2.3 Expected distortion function: when CSIST is not available.

Now consider the system described in the previous system when the transmitter does not know the actual channel state $\mathbf{H}$. In this situation, the transmitter is not aware of the capacity of the channel and cannot send at a rate arbitrarily close to the capacity. Consider that the transmitter decides to transmit at a rate R_c. Now, the decoding of $\mathbf{U}$ is not granted. We have that if $R_c > \mathcal{C}(\mathbf{H})$ then $\mathbf{U}$ will not be recovered. We call this situation 'Outage event'. In this situation, the distortion of the source is maximum, i.e $D[\text{Outage event}] = 1$. On the contrary, if $R_c \leq \mathcal{C}(\mathbf{H})$ we will be able to reconstruct the source with a distortion $D[\text{No Outage event}] = 2^{-bR_s}$.

In order to measure the performance of the system, we are interested in the average performance of the system, i.e the average distortion. We call this performance the Expected Distortion function, $E[D]$, i.e.

$$\begin{aligned} E[D] &= (1 - P_{out})D[\text{No Outage event}] + P_{out}D[\text{Outage event}] \\ &= (1 - P_{out})2^{-bR_c} + P_{out} \cdot 1. \end{aligned} \tag{2.9}$$

where $P_{out} = \Pr\{R_c > \mathcal{C}(\mathbf{H})\}$ is the probability of outage of the system. Note that there is clear tradeoff in R_c. The higher R_c, the better the achieved distortion when there is no error. However, the higher R_c, the higher the probability of outage, and hence the probability of having the maximum distortion. As the encoder is only aware of the distribution of both parameters, we want to design R_c such that we minimize $E[D]$, with respect to the source and the channel state and the channel noise.

As commented, we are interested in characterizing the minimum expected distortion for each possible set of parameters $\Omega = (\rho, \rho_s, b)$,

$$ED^*(\Omega) \triangleq \min_{f,g} E[D]. \tag{2.10}$$

An example of expected distortion function can be seen in Figure 2.2 for different pairs of (f, g) in function or R_c for a particular scenario. Note that the distortion function becomes smooth on averaging, with a clear minimum.

2.4 Characterization at high SNR: distortion exponent

In general it is very complicated or impossible to obtain close form solution for the Expected Distortion function. A method that provides a good insight on the performance of different transmission schemes is the characterization of the slope of $E[D]$ in the high SNR regime, i.e. when $\rho \to \infty$. The performance measure is the *distortion exponent*, defined as

$$\Delta(b, x) \triangleq -\lim_{\rho\to\infty} \frac{\log ED}{\log \rho}, \tag{2.11}$$

In this section we will give a brief example of how to calculate the best Distortion Exponent achievable by the previous scheme. Consider the previous digital scheme for the SISO and Rayleigh fading. Outage events were proven in [10] to be the dominating in the probability of

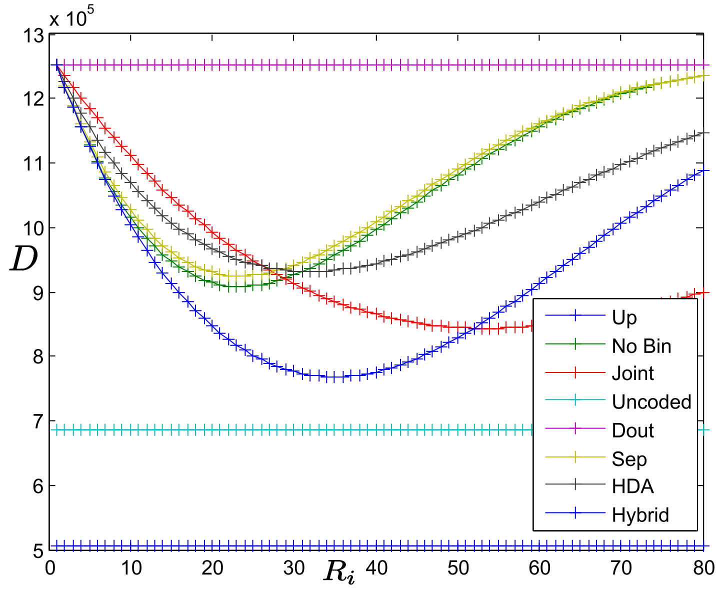

Figure 2.2: ED function for $\rho = 10dB$, with $b = 2$, $x = 1$ for many schemes presented in Chapter 3.

error in high SNR and hence, the probability of outage is equivalent to the error probability in high SNR. For this particular case, P_{out} can be explicitly calculated as

$$P_{out} = \Pr(R_c > \log(1 + \rho h^2)) = 1 - e^{\frac{(2^{R_c}-1)}{\rho}}. \tag{2.12}$$

Now, we will characterize the high SNR behavior of the probability of outage Consider the family of codes such that their dimensionality increases in ρ as $R_c = r \log \rho$, also known as *multiplexing gain* in literature. The diversity gain $d(r)$ is defined as the supremum of the diversity advantage over all possible code families with multiplexing gain r, and given by

$$d(r) \triangleq -\lim_{\rho\to\infty} \frac{\log P_{out}}{\log \rho}. \tag{2.13}$$

Then,

$$P_{out} = 1 - e^{\frac{(\rho^{r-1})}{\rho}} = 1 - \left(1 - (\rho^{r-1})\right) + \mathcal{O}(\rho) = \rho^{-(1-r)} = \rho^{-d(r)}. \tag{2.14}$$

where we have used the Taylor expansion series at $x = 0$ as $e^x = 1 - x + \mathcal{O}(x)$, which is true if $r \leq 1$. If $r > 1$, then $P_{out} = 1$. Then, $d(r)$ is the decaying slope for the probability of outage. The rigorous method to do those approximations was introduced by Tse in [10], and is extensively used in chapter 5. In this work the existence of a fundamental tradeoff between $d(r)$ and r is proven.

Now, the Expected Distortion function asymptotical behavior can be equivalently approximated as

$$\begin{aligned} E[D] &= (1 - P_{out})2^{-bR_c} + P_{out} \\ &\doteq \left(1 - \rho^{-d(r)}\right)\rho^{-br} + \rho^{-d(r)} \\ &\doteq \rho^{-br} + \rho^{-(1-r)}, \end{aligned} \tag{2.15}$$

where $(1-\rho^{-d(r)}) \doteq 1$ as $d(r) \geq 0$, as $r \leq 1$. As we are interested in characterizing the best performance of $E[D]$ in the high SNR regime for a given transmission scheme, we want to find r such that the decay of the expected distortion is maximized. Note that the first term decreases in r and the second one increases in r. The decay Δ can be maximized by equaling both terms, i.e. $br = (1-r)$ and then the optimal $r = \frac{1}{1+b}$. Then the the best distortion exponent achieved by digital transmission in a Rayleigh fading scenario is given by

$$E[D] \doteq \rho^{-\Delta} \doteq \rho^{-\frac{b}{1+b}}. \tag{2.16}$$

This simple example, outlines the ideas behind the optimization procedures used in Chapter 5.

Chapter 3

Transmission Schemes

In this chapter, we present the transmission techniques for MIMO channels when time varying side information is present at the receiver based on the state of the art and two novel joint source-channel coding transmission techniques. The chapter is divided in four sections. In the first section, we derive a lower bound for the expected distortion assuming that the transmitter is informed of the actual channel state and side information quality at the time of designing the transmission.

Then, we present two families of transmission schemes, each one constituting a section. The first family, that we will call 'Single layer schemes', consist of transmitting the source by using a unique layer of transmission. In the first section, 'Single layer schemes: Separation, Uncoded and NBJD', we study the classical separation in which the source is first compressed using binning and the compressed bits are sent though a digital transmission. We then present a simpler scheme that does not use binning in the compression. We also study the uncoded transmission.

Next we introduce a novel joint source-channel coding techniques: NBJD, that uses no explicit binning and applies joint decoding [8]. We study transmission schemes based on pure NBJD and construct transmission techniques that are known to have good, or even optimal, performance when time varying side information is not present at the receiver. In this section we also present an Hybrid analog-digital scheme using NBJD. This scheme superposes an analog transmission layer with a digital layer on top of it. This scheme is known to achieve the optimal performance for $M_*b < 1$ in MIMO systems when no side information is available [4].

In the second single layer section: 'Single layer schemes: HDA', we introduce a second novel joint source-channel coding technique, HDA, that quantizes the source and uses joint decoding with side information and the error in the quantization [6]. We then extend HDA to multiple-input multiple-output (MIMO) and bandwidth expansion regime, i.e. $b > 1$. Before constructing the scheme for time varying states, we first study the scenario when the channel and side information states are known at the transmitter and the receiver. We study a superposition scheme that combines quantization of the source and HDA and generalizes the result provided by Lapidoth in [9], in which a continuum of optimal transmission schemes when no side information is present is obtained. Contrary to the result when no side information is present, we prove, for the single-input single-output scheme, that such continuum of schemes does not extend when side information is present, and that the is only optimal configuration when side information is present coincides with pure HDA transmission. A continuum of optimal transmission schemes when side information is present has been recently given by [11] combining HDA with a digital layer superposed. We then construct the MIMO HDA based on pure HDA for an scenario in which the states are not available at the tranmitter.

In the last section, we introduce the family of 'Multi-layer Schemes'. This group consist of schemes that use multiple layers of coding to transmit the source and combat fading. This multilayer schemes where proven to be optimal in the high SNR regime [2] for SISO sytems and in MIMO systems for $b \geq M_t M_r$ in the continuum of infinit layers. We propose a separation based multilayer scheme and a NBJD scheme.

3.1 Expected distortion lower bound

We derive a lower bound on the expected distortion by providing the transmitter with the actual channel and side information states $(\gamma, \mathbf{H})$. A transmitter with the actual $\mathbf{H}$ and γ realizations at each block can use the optimal transmission scheme based on separate source-channel coding in which the source sequence is compressed using Wyner-Ziv coding and the compressed bits are transmitted at the channel capacity [12], such that no outage occurs and the minimum possible distortion is achieved at each block. The distortion at state $(\mathbf{H}, \gamma)$ is

$$D_{low}(\mathbf{H}, \gamma) = \frac{2^{-b\mathcal{C}(\mathbf{H})}}{\gamma^2 \rho_s + 1}, \tag{3.1}$$

where we drop the explicit dependance on b, ρ and ρ_s for the sake of brevity and the capacity of the MIMO channel is given by

$$C(\mathbf{H}) = \sup_{\mathbf{Q} \succcurlyeq 0, \mathrm{Tr}\{\mathbf{Q}\} \leq M_*.} \log\det\left(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\right), \tag{3.2}$$

The lower bound on the expected distortion is then found as

$$ED_{low} = \iint_{(\mathbf{H},\gamma)} \frac{\log(1 + \frac{\rho}{M_*}\sum_{i=1}^{M_*}\lambda_i q_i)^{-b}}{1 + \gamma^2\rho_s} p_h(h) p_\gamma(\gamma) d\mathbf{H} d\gamma. \tag{3.3}$$

where $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_{M_*}$ are the ordered eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$, with $\mathrm{rank}\{\mathbf{H}\} = \min\{M_t, M_r\} = M_*$ and

$$q_i = \left(\mu - \frac{1}{\lambda_i}\right)^+, \tag{3.4}$$

with μ such that the power constraint equation is satisfied

$$\sum_{i=1}^{M^*}\left(\mu - \frac{1}{\lambda_i}\right)^+ \leq 1. \tag{3.5}$$

Proof. See Appendix 3.A. □

3.2 Single layer schemes: separation, uncoded and NBJD

3.2.1 Separate source and channel coding

We first consider separate source and channel coding with a single layer, which is based on Wyner-Ziv coding followed by channel coding. However, unlike the optimization and binning rates are fixed at all channel states due to the lack of CSI at the transmitter.

The quantization codebook consists of $2^{m(bR_c+R_s)}$ length-m codewords, $W^m(i)$, $i = 1, ..., 2^{m(bR_c+R_s)}$, generated through a 'test channel' given by $W = X + Q$, where $Q \sim \mathcal{CN}(0, \sigma_Q^2)$ and independent of X. The quantization noise variance is chosen such that $I(X; W) = R_s + bR_c$, i.e. $\sigma_Q^2 = (2^{R_s+bR_c} - 1)^{-1}$. The generated quantization codewords are then uniformly distributed among 2^{nR_c} bins. On average, each bin contains 2^{mR_s} codewords. Additionally, a Gaussian channel codebook with 2^{nRc} length-n codewords $\mathbf{U}(s)$ is generated independently with i.i.d. components $U \sim \mathcal{CN}(0, 1)$ and each codeword $\mathbf{U}(s)$, $s \in [1, ..., 2^{nR_c}]$, is assigned to the bin index s. Note that the power at each antenna is uniformly allocated as $E[\mathbf{U}_i^\dagger \mathbf{U}_i] = \mathbf{I}$.

Given a source realization X^m, the encoder searches for a quantization codeword $W^m(i)$ that is jointly typical with X^m. Assuming one such codeword is found, the channel codeword $\mathbf{U}(s)$ is transmitted over the channel, where s is the bin index of $W^m(i)$. At reception, the bin index s is recovered with high probability using the channel output $\mathbf{V}$ if,

$$R_c \leq I(\mathbf{U}_i; \mathbf{V}_i). \tag{3.6}$$

The decoder then looks for a quantization codeword within the estimated bin that is jointly typical with the side information sequence Y^n. The correct codeword will be decoded with high probability if,

$$bR_c \geq I(X; W|Y). \tag{3.7}$$

If the quantization codeword is successfully decoded, then $\hat{X}^m$ is reconstructed with an optimal minimum mean square error (MMSE) estimator where $\hat{X}_i = E[X_i|Y_i, W_i]$ for $i = 1, ..., n$.

An outage is declared whenever, due to the randomness of the channel and the side information, the quantization codebook cannot be correctly decoded, i.e., when conditions (3.6) or (3.7) are not satisfied. In case of an outage, the side-information is used to estimate the source with $\hat{X}_i = E[X_i|Y_i]$. The distortion achieved when the quantization rate is R and the side information quality is fixed by γ is given by

$$D_d(R, \gamma) \triangleq (\gamma^2 \rho_s + 2^R)^{-1} \tag{3.8}$$

when the quantization codeword is decoded by the receiver. When there is an outage over the channel, the minimum distortion is given by $D(0, \gamma)$. Then, the expected distortion of the separation scheme is given by

$$\begin{aligned} ED_{sb}(R_s, R_c) = &\iint_{\mathcal{O}_{sb}^c} D_d(R_s + bR_c, \gamma) p_\gamma(\gamma) p_h(\mathbf{H}) d\gamma d\mathbf{H} \\ &+ \iint_{\mathcal{O}_{sb}} D_d(0, \gamma) p_\gamma(\gamma) p_h(\mathbf{H}) d\gamma d\mathbf{H}, \end{aligned} \tag{3.9}$$

where $\mathcal{O}_{sb}^c$ is the complement of the outage event defined as

$$\mathcal{O}_{sb} = \{(\mathbf{H}, \gamma) : I(\mathbf{U}_i; \mathbf{V}_i) < R_c \text{ or } bR_c < I(X; W|Y)\}.$$

We have $I(\mathbf{U}_i; \mathbf{V}_i) = \log\det(\mathbf{I} + \frac{\rho}{M_t} \mathbf{H}^\dagger \mathbf{H})$ and $I(X; W|Y) = \log\left(1 + (2^{Rs+bR_c} - 1)(\gamma^2 \rho_s + 1)^{-1}\right)$ for the model under consideration.

3.2.2 Separate source and channel coding without binning

We can reduce the complexity of the previous scheme by including a single codeword in each bin, i.e. by letting $R_s = 0$. This way, we get rid of the outage event corresponding to a poor side information realization. However, to achieve the same quantization noise, we need to transmit at a higher rate over the channel, which increases the channel outage probability. The expected distortion for this scheme is then found as

$$\begin{aligned} ED_{nb}(R_c) = &(1 - P_{out}(h))E_\gamma[D_d(bR_c, \gamma)] \\ &+ P_{out}(h)E_\gamma[D_d(0, \gamma)], \end{aligned} \tag{3.10}$$

where $P_{out}(h) \triangleq Pr\{R_c > \log\det(\mathbf{I} + \frac{\rho}{M_t}\mathbf{H}^\dagger\mathbf{H})\}$ is the probability of having a channel outage event.

3.2.3 Uncoded transmission

Uncoded transmission is a robust joint source-channel transmission technique that has a gradual degradation with worsening channel quality. Since rank$\mathbf{H} \leq M_*$, M_* samples are effectively transmitted at each channel use using M_* of the M_t tx antennas. For $b \leq \frac{1}{M_*}$, the channel codeword $\mathbf{U}^{M_* \times n}$ is generated scaling the first $M_* n$ samples of the source and mapping them into the channel codeword, i.e., $\mathbf{U}^{M_* \times n} = [X_1^{M_*}; X_{M_*+1}^{2M_*}; ...; X_{(n-1)M_*+1}^{nM_*}]$, so that the power constraint is satisfied[1]. At reception, the transmitted samples are estimated with an MMSE estimator using the channel output $\mathbf{V}$ and the available side information Y_1^n. The remaining $m - M_* n$ source samples that have not been transmitted are estimated using only the available side information $Y_{M_* n+1}^m$.

When $b > \frac{1}{M_*}$ the source sequence is transmitted in the first $\frac{m}{M_*}$ channel uses scaling the power by bM_*.

Let $\mathbf{H}_{M_*}$ be the submatrix of $\mathbf{H}$ obtained by taking the M_* columns corresponding to the antennas effectively used for the transmission. Then, the minimum achievable distortion when a source sample is transmitted uncoded at uniform power P at state $(\mathbf{H}, \gamma)$ is given by

$$D_p(P, \gamma, \mathbf{H}) \triangleq \sum_{i=i}^{M_*} \frac{1}{1 + P\mu_i\rho + \gamma^2\rho_s}. \tag{3.11}$$

where $\mu_1 \leq ... \leq \mu_{M_*}$ are the ordered eigenvalues of the matrix $\mathbf{H}_{M_*}\mathbf{H}_{M_*}^\dagger$.

Proof. See Appendix 3.B ☐

Then, the uncoded expected distortion can be expressed as

$$ED_u = \begin{cases} bM^* D_p(1, \mathbf{H}, \gamma) + (1 - bM_*)D_p(0, \mathbf{H}, \gamma) & \text{if } M_* b < 1, \\ D_p(bM^*, \mathbf{H}, \gamma) & \text{if } M_* b \geq 1. \end{cases}$$

Note that the uncoded scheme has no outage events.

[1] X_l^k is used to refer to the vector $[X_l, ..., X_k]$.

3.2.4 No binning with joint decoding (NBJD)

Next, we consider a source-channel coding scheme that does not involve any explicit binning at the encoder and uses joint decoding at the decoder. This coding scheme is introduced in [8] in the context of broadcasting a common source to receivers with different side information qualities. The success of the decoding process depends on the joint quality of the channel and the side information.

At the encoder, a codebook of 2^{nR_j} length-m quantization codewords is generated with $R_j = I(X;W)$ as in Section 3.2.1. Then, an independent Gaussian codebook of size 2^{nR_j} is generated with length-n codewords $\mathbf{U}(i) \in \mathbb{C}^{M_t \times n}$ with $U \sim \mathcal{CN}(0,1)$. Given a source outcome X^m, the transmitter finds the quantization codeword $W^m(i)$ jointly typical with the source outcome and transmits the corresponding channel codeword $\mathbf{U}(i)$ over the channel. At reception, joint typicality decoding is performed over the channel codewords such that the decoder looks for an index i for which both $(\mathbf{U}^n(i), V^n)$ and $(Y^m, W^m(i))$ are jointly typical. The outage event is given by

$$\mathcal{O}_j = \{(\mathbf{H}, \gamma) : I(X;W|Y) > bI(\mathbf{U}_i;\mathbf{V}_i)\}. \tag{3.12}$$

If the quantized codeword has been successfully decoded, the source X^m is estimated using an MMSE estimator. On the contrary, if an outage has occurred, the source X^m is reconstructed using only the side information.

The expected distortion for NBJD can be expressed as

$$\begin{aligned} ED_j(R_j) = &\iint_{\mathcal{O}_j^c} D_d(R_j,\gamma)p_\gamma(\gamma)p_h(\mathbf{H})d\gamma d\mathbf{H} \\ &+ \iint_{\mathcal{O}_j} D_d(0,\gamma)p_\gamma(\gamma)p_h(\mathbf{H})d\gamma d\mathbf{H}. \end{aligned} \tag{3.13}$$

3.2.5 NBJD-analog hybrid scheme

Case $M_* b \leq 1$

Mittal and Phamdo introduce in [13] the hybrid analog-digital schemes in the context of robust transmission in which the source sequence is transmitted using an uncoded signal and a digital codeword. Inspired by this work, we propose a hybrid scheme where the digital signal is transmitted using NBJD.

When $M_* b \leq 1$, we propose a hybrid scheme which consists of the superposition of an uncoded signal and a signal transmitted using NBJD. At the encoder, the source sequence is divided into two blocks: $X_u = X_1^{nM_*}$ is transmitted uncoded and $X_d = X_{nM_*+1}^m$ is transmitted with NBJD. The digitally transmitted part, X_d is quantized to a codeword W_d of a codebook of size $2^{(m-n)R_j}$ using the NBJD encoder and the corresponding channel codeword $\mathbf{U}_d \in \mathbb{C}^{n \times M_*}$ is picked from an independent codebook of the same size. Next, the analog channel input is generated by mapping X_u into $\mathbf{U}_u \in \mathbb{C}^{n \times M_*}$ and scaling to make it unit power as in Section 3.2.3. Then the channel codeword is generated as the superposition of these two signals with power allocation as

$$\mathbf{U} = \sqrt{1-\rho^{-\eta}}\mathbf{U}_d + \sqrt{\rho^{-\eta}}\mathbf{U}_u. \tag{3.14}$$

Again, let $\mathbf{H}_{M_*}$ be the submatrix of $\mathbf{H}$ obtained by taking the M_* columns corresponding to the antennas effectively used for the transmission. Then $\mathbf{U}$ is transmitted using M_* antennas thought $\mathbf{H}_{M_*}$. the selection of those antennas is based on some arbitrary prearranged strategy.

The decoder uses joint typicality decoding to recover W_d and $\mathbf{U}_d$ from the channel output. The outage is defined as

$$\mathcal{O}_h \triangleq \{(\mathbf{H}, \gamma) : (1 - bM_*)I(X_d; W_d|Y) > I(\mathbf{U}_d; \mathbf{V})\}.$$

with

$$I(\mathbf{U}_d; \mathbf{V}) = \log\det\left(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}_{M_*}\mathbf{H}_{M_*}^{\dagger}\right) - \log\det\left(\mathbf{I} + \frac{\rho^{1-\eta}}{M_*}\mathbf{H}_{M_*}\mathbf{H}_{M_*}^{\dagger}\right). \tag{3.15}$$

If W_d is successfully recovered, the receiver uses it and the available side information to reconstruct X_d with an MMSE estimator. Then, $\mathbf{U}_d$ is substracted from the channel output $\mathbf{V}$ to obtain a noisy version of $\mathbf{U}_u$ and reconstruct X_u with an MMSE estimator using also the side information. If there has been an outage and $\mathbf{U}_d$ is not recovered, X_d is reconstructed using only the side information and X_u is estimated using the channel output, now noisier than the previous case since $\mathbf{U}_d$ cannot be substracted and the side information. Finally, the source sequence is reconstructed as $\hat{X}^m = [\hat{X}_u, \hat{X}_d]$.

Then, the expected distortion for NBJD-analog hybrid scheme is given by

$$\begin{aligned}
ED_h(R_h) = (1 - bM_*) &\iint_{\mathcal{O}_h^c} D_p(R_h, \gamma)p_\gamma(\gamma)p_h(\mathbf{H}_{M_*})d\gamma d\mathbf{H}_{M_*} \\
&+ bM_* \iint_{\mathcal{O}_h^c} D_l(\rho^{-\eta}, \mathbf{H}_{M_*}, \gamma)p_\gamma(\gamma)p_h(\mathbf{H}_{M_*})d\gamma d\mathbf{H}_{M_*} \\
&+ (1 - bM_*) \iint_{\mathcal{O}_h} D_d(0, \gamma)p_\gamma(\gamma)p_h(\mathbf{H}_{M_*})d\gamma d\mathbf{H}_{M_*} \\
&+ bM_* \iint_{\mathcal{O}_h} D_i(\mathbf{H}_{M_*}, \eta)p_\gamma(\gamma)p_h(\mathbf{H}_{M_*})d\gamma d\mathbf{H}_{M_*}.
\end{aligned} \tag{3.16}$$

where $D_i(\mathbf{H}_{M_*}, \eta)$ is the distortion achieved by the mmse estimation of X_u treating $\mathbf{U}_d$ as noise when there has been an outage. It is given by

$$D_i(\mathbf{H}_{M_*}, \eta) = \frac{1}{M_*}\sum_{i=1}^{M_*}\left(1 + \gamma^2 + \frac{\rho^{1-\eta}\mu_i}{1 + \rho(1 - \rho^{1-\eta})\mu_i}\right)^{-1} \tag{3.17}$$

where $\mu_1 \leq ... \leq \mu_{M_*}$ are the ordered eigenvalues of $\mathbf{H}_{M_*}\mathbf{H}_{M_*}$.

Proof. See Appendix 3.C.1. □

Case $M_* b > 1$

When $M_* b > 1$, the channel codeword is generated at the encoder as a concatenation of an NBJD digital codeword and the scaled version of the error produced in the quantization. The NBJD codebook W_h is generated at a quantization rate R_h. The digital channel codebook is generated with $\mathbf{U}_d^{n-mM_*\times M_t}$. The analog signal $\mathbf{U}_a^{m\times M_*}$ is a mapping of the quantization error sequence, $X^m - W(i)^m$, and scaled by the error variance. Then, the channel codeword is constructed as $\mathbf{U} = [\mathbf{U}_a\ \mathbf{U}_d]$ and scaled to satisfy the power constraint to transmitted though M_* antennas.

At the receiver, $\mathbf{U}_d$ is decoded with joint typicality decoding using Y and the channel output of $\mathbf{U}_d$. The outage event is defined as

$$\mathcal{O}_h = \left\{(\mathbf{H},\gamma) \in \mathbb{C}^2 : I(W_h;X|Y) > (bM_* - 1)I(\mathbf{U}_d;\mathbf{V}_d)\right\},$$

where $\mathbf{V}_a$ is the channel output sequence $\mathbf{V}_1^m$ corresponding to the transmission of the quantization noise and $\mathbf{V}_b = \mathbf{V}_{m+1}^n$ corresponds to the NBJD part. We have, after some manipulation

$$\begin{aligned} I(W_h;X|Y,\mathbf{V_a}) &= \log\det\left(\sigma^2(\gamma^2+1)\mathbf{I} + \mathbf{I} + \frac{1}{M_*}\mathbf{H}\mathbf{H}^\dagger\right) - \log\det\left(\sigma^2(\gamma^2+1)\left(\mathbf{I} + \frac{1}{M_*}\mathbf{H}\mathbf{H}^\dagger\right)\right) \\ &= \log\det\left(\left(\mathbf{I} + \frac{1}{M_*}\mathbf{H}\mathbf{H}^\dagger\right)^{-1} + \frac{1}{\sigma^2(\gamma^2+1)}\mathbf{I}\right) \end{aligned} \quad (3.18)$$

and $I(\mathbf{U}_d;\mathbf{V}_d) = \log\det(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}_{M_*}\mathbf{H}_{M_*}^\dagger)$i.e.

$$\begin{aligned} \mathcal{O}_h = \Big\{(\mathbf{H},\gamma) \in \mathbb{C}^2 \ : \ &\log\det\left(\left(\mathbf{I} + \frac{1}{M_*}\mathbf{H}\mathbf{H}^\dagger\right)^{-1} + \frac{1}{\sigma^2(\gamma^2+1)}\mathbf{I}\right) \\ &> (bM_* - 1)\log\det(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}_{M_*}\mathbf{H}_{M_*}^\dagger)\Big\}. \end{aligned} \quad (3.19)$$

If $\mathbf{U}_d$ is decoded, the decoder reconstructs the source sequence with an MMSE estimator using the analog error transmission in V_1^m, Y, and $\mathbf{U}_d$. If $\mathbf{U}_d$ could not be recovered, the source is reconstructed with Y and $\mathbf{V}_a$. The distortion achieved when there is no outage is given by

$$D_{pd}(R,h,\gamma) \triangleq \sum_{i=1}^{M_*}\left(\frac{1}{\sigma^2}(1+\frac{\rho\lambda_i}{M_*}) + \gamma^2\rho^x + 1\right)^{-1}.$$

Proof. See Appendix 3.C.2. □

Hence, the expected distortion is found as

$$\begin{aligned} ED_h(R_h) = &\iint_{\mathcal{O}_h} D_d(0,\gamma)p_h(\mathbf{H}_{M_*})p_\gamma(\gamma)d\mathbf{H}_{M_*}d\gamma \\ &+ \iint_{\mathcal{O}_h^c} D_{pd}(R_h,\mathbf{H}_{M_*},\gamma)p_h(\mathbf{H}_{M_*})p_\gamma(\gamma)d\mathbf{H}_{M_*}d\gamma. \end{aligned} \quad (3.20)$$

3.3 Single layer schemes: HDA

In this section, we present an superposition scheme that makes use of the side information. Similarly to Lapidoth's scheme [14], we propose a scheme that superposes a correlated quantized version of the source with an uncorrelated codeword generated with an hybrid digital-analog coding scheme. We study the performance of such scheme under the scenario of known channel state and side information state and the more practical scenario in which the channel and the side information are Rayleigh block fading with states unknown at the transmitter. We find that the optimal configuration in presence of side information is given by a unique configuration that coincides with pure HDA. We then extend HDA scheme to MIMO systems.

3.3.1 Hybrid analog-HDA when CSIST is available

Similarly to the scheme introduced by Lapidoth we present an hybrid scheme for $b \geq 1$ that superposes an scaled version of the source sequence with an digital scheme. At the encoder, consider the auxiliary a random variable such that the first m samples are

$$W_1 = Q + \kappa X, \tag{3.21}$$

and the samples from $m+1$ to n

$$W_2 = Q. \tag{3.22}$$

with $Q \sim \mathcal{CN}(0,1)$ independent of X. Then generate a codebook of 2^{mR_C} length-n codewords W^n distributed according to P_W such that $R_C = I(W;X) = \log(1+\kappa^2)$. At a given source sequence X^m, the encoder looks for the i^*-th auxiliary codeword $W^n(i^*)$ such that $(W_1^m(i^*), X^m)$ are jointly typical. Then the channel input is generated as $U^n = [U_1^m, U_{m+1}^n]$, where U_1^n is a superposition of the auxiliary random variable W_1^n and $\kappa X^m - W_1^m$ as

$$\begin{aligned} U_1^n &= \alpha(W_1^m(i^*) - kX^n) + \beta W_1^m(i^*) \\ &= (\alpha+\beta)Q_1^m + \beta\kappa X^m. \end{aligned} \tag{3.23}$$

and $U_{m+1}^n = Q_{m+1}^n$.

In order to satisfy the power constraint, α and β have to satisfy

$$\kappa^2\beta^2 + (\alpha+\beta)^2 \leq 1. \tag{3.24}$$

At a given channel output V^n, the decoder looks for the auxiliary random variable $W^n(i)$ such that W_1^n is simultaneously jointly typical with the current side information block sequence Y^m and the channel output V_1^m, i.e. $(W_1^m(i); V_1^m, Y^m)$ are jointly typical and the channel output V_{m+1}^n is jointly typical with $W_{m+1}^n(i)$.

Theorem 1. *The decoding will be successful with high probability, using typicality arguments if*

$$I(X;W) \leq R_C \leq I(W_1; V_1, Y) + (b-1)I(W_2; V_2). \tag{3.25}$$

where V_1 and V_2 are a sample of the channel output in V_1^m and V_{m+1}^n, respectively.

Proof. Let $\tilde{Y}^n \doteq [Y^m, 0_{m+1}^n]$. Using typicality arguments, we have that to have W_1^n

$$\begin{aligned} mR_c &\leq I(W^n; V^n, \tilde{Y}^n) \\ &= I(W_1^m; V_1^m, \tilde{Y}_1^m) + I(W_{m+1}^n; V_{m+1}^n, Y_{m+1}^n) \\ &= mI(W_1; V_1, Y) + (n-m)I(W_2, V_2) \end{aligned} \tag{3.26}$$

where V_1 and V_2 is a sample that behaves statistically as a sample in V_1^m and V_{m+1}^n respectively. Equalities comes from the chain rule and the i.i.d generation of W. Dividing both sizes leads to the desired result. ☐

When we have Gaussian sources, condition (3.41) becomes

$$R_c \leq \log\left(\frac{\left(1+\kappa^2\right)\left(h^2\beta^2\kappa^2 + a\left(h^2(\alpha+\beta)^2+1\right)\right)}{a+\kappa^2\left(h^2\alpha^2+1\right)}\right) + \log\left(1+h^2\right)^{b-1}. \tag{3.27}$$

with $a = 1+\gamma^2$ and where to compute the mutual information, we use

$$C_{w;vy} = \begin{pmatrix} 1+\kappa^2 & \alpha+\beta+\beta\kappa^2 & \gamma\kappa \\ \alpha+\beta+\beta\kappa^2 & (\alpha+\beta)^2+\beta^2\kappa^2+\sigma & \beta\gamma\kappa \\ \gamma\kappa & \beta\gamma\kappa & 1+\gamma^2 \end{pmatrix}.$$

and $\sigma^2 = 1/h^2$.

Then, the source sequence is reconstructed with an mmse estimator $\hat{X} = E[X|V,Y,W]$. The achievable distortion is given by

$$D_C(\kappa, h^2, \gamma) \triangleq \left(1+\gamma^2+\kappa^2(h^2\alpha^2+1)\right)^{-1}. \tag{3.28}$$

Any power allocation has to satisfy the power (3.24) and decoding (3.27) conditions. Hence, for a given (h,γ) pair, the set of power allocations such that W is decoded, $\Lambda(h,\gamma)$, is the set of pairs $(\alpha,\beta) \in \mathbb{R}^2$ in

$$\Lambda(h,\gamma) = \Big\{(\alpha,\beta) \in \mathbb{R}^2 : (\alpha+\beta)^2+\kappa^2\beta^2 = 1, \quad 1 \leq \frac{h^2\beta^2\kappa^2 + a(h^2(\alpha+\beta)^2+1)}{a+\kappa^2(h^2\alpha^2+1)}\left(1+h^2\right)^{b-1}\Big\}. \tag{3.29}$$

Next theorem shows that, for a fixed (h,γ) there exists a single power allocation (α,β) in $\Lambda(h,\gamma)$ such that the optimal distortion is achieved. This particular choice reduces to the scheme presented by Wilson in [6] for $b=1$. When no side information is available, i.e. $\gamma^2 = 0$, a continuum of optimal schemes can be found. This result coincides with the schemes presented by Bross in [15] for $b = 1$.

Theorem 2. *In presence of side information, i.e. $\gamma^2 \neq 0$, the unique power allocation achieving the optimal distortion, that for a SISO system is given by*

$$D^*(h^2,\gamma) = \frac{1}{(1+\gamma^2)\left(1+h^2\right)^b}. \tag{3.30}$$

is given by

$$\alpha = 1, \qquad \beta = 0, \qquad \kappa^2 = \frac{a}{h^2+1}\left[\left(1+h^2\right)^b - 1\right]. \tag{3.31}$$

If $\gamma^2 = 0$, i.e no side information is available, there exists a continuum of schemes achieving the optimal distortion with the following power allocation for a given $\kappa^2 \leq h^2$

$$\alpha^2 = \left(\frac{1}{\kappa^2} - \frac{1}{h^2}\right)^+,$$

$$\beta = -\frac{\alpha \pm \sqrt{1+\kappa^2-\alpha^2\kappa^2}}{1+\kappa^2}.$$

If $\kappa^2 > h^2$, the continuum reduces to a single power allocation achieving the optimal distortion satisfying the same equations given by $\alpha^2 = 0$, $\beta^2 = \frac{1}{1+\kappa^2}$ and $\kappa^2 = h^2$.

Proof. From equation (3.28) observe that any power allocation achieving the optimal distortion has to satisfy

$$a + \kappa^2(h^2\alpha^2 + 1) = a\left(1 + h^2\right)^b. \tag{3.32}$$

A power allocation achieving the optimum has to simultaneously satisfy (3.32) and belong to (3.29). Substituting (3.32) in (3.29), the optimal allowable power allocation have to satisfy

$$\beta^2\kappa^2 + (\alpha+\beta)^2 \leq 1, \tag{3.33}$$
$$\beta^2\kappa^2 + a(\alpha+\beta)^2 \geq a. \tag{3.34}$$

The allowable point are then in the intersection of two ellipsoids. We first prove that any point satisfying (3.33) is included in the complementary of (3.34), i.e.

$$\begin{aligned}
&\beta^2\kappa^2 + a(\alpha+\beta)^2 \\
&= \beta^2\kappa^2 + a(1-\kappa^2\beta^2) \\
&= a - \beta^2\kappa^2(1-a) \\
&= a + \beta^2\kappa^2(a-1) \\
&\geq a
\end{aligned} \tag{3.35}$$

Hence, the intersection can only happen in the frontier. Any point in the frontier satisfies $(\alpha+\beta)^2 = 1 - \kappa^2\beta^2$. The allowable (α, β), then have to satisfy

$$\begin{aligned}
&\beta^2\kappa^2 + a(\alpha+\beta)^2 \geq 1 \\
&\beta^2\kappa^2 + a(1-\kappa^2\beta^2) \geq 1. \\
&\beta^2\kappa^2(1-a) + a \geq 1.
\end{aligned} \tag{3.36}$$

The intersection is only no null for $a = 1$, i.e. $\gamma^2 = 0$ or $\beta^2 = 0$. Note that for $a = 1$, the frontiers of both ellipses are the same.

Now we study those particular cases. Observe that $\beta^2 = 0$ reduces to the scheme presented in [6] in the case of $b = 1$. From (3.24) we have that $\alpha^2 = 1$ and then, from (3.32), we have

$$\kappa^2 = \frac{a}{h^2+1}\left[\left(1+h^2\right)^b - 1\right]. \tag{3.37}$$

and achieves thus the optimal distortion.

If $\gamma^2 = 0$, the scheme becomes the optimal continuum of schemes proposed by Lapidoth in [15] in the case of $b = 1$. Any power allocation in (3.29) satisfies, for $a = 1$

$$\begin{aligned}
\beta^2\kappa^2 + (\alpha+\beta)^2 &= 1, \\
1 + \kappa^2(h^2\alpha^2 + 1) &= \left(1+h^2\right)^b.
\end{aligned} \tag{3.38}$$

Then, if $\kappa^2 + 1 \leq \left(1+h^2\right)^b$, we can solve problem (3.38) with equality with

$$\begin{aligned}
\alpha^2 &= \frac{1}{\kappa^2 h^2}\left[\left(1+h^2\right)^b - 1\right] - \frac{1}{h^2}, \\
\beta &= \frac{-\alpha \pm \sqrt{1+\kappa^2-\alpha^2\kappa^2}}{1+\kappa^2}.
\end{aligned}$$

which is a continuum of pairs (α, β) for the range of valid κ^2.

If $\kappa^2 + 1 > \left(1 + h^2\right)^b$, the continuum reduces to

$$\begin{aligned} \alpha^2 &= 0, \\ \beta^2 &= \frac{1}{1+\kappa^2}, \\ \kappa^2 &= \frac{1}{1+h^2}\left[\left(1+h^2\right)^b - 1\right]. \end{aligned} \tag{3.39}$$

which coincides with the scheme for $\beta^2 = 0$ for $a = 1$. □

3.3.2 MIMO HDA

In this section, we extend the HDA scheme to multiple input multiple output systems (MIMO) and $b > 1$. At the encoder, consider the auxiliary random variables $\mathbf{W} \in \mathbb{C}^{m \times M_t}$ and $\mathbf{T} \in \mathbb{C}^{n-m \times M_t}$ such that

$$\mathbf{W}_i = \boldsymbol{\kappa} X_i + \mathbf{Q}_i, \qquad i = 1...m$$

where $\boldsymbol{\kappa} \in \mathbb{R}^{M_t}$ and $\mathbf{Q}^{m \times M_t}$ such that $\mathbf{Q}_i$ is Gaussian distributed as $\mathbf{Q}_i \sim \mathcal{CN}(0, \mathbf{C_Q})$ independent of X and others $\mathbf{Q}_j$ for $j \neq i$, and $\mathbf{T}_i \sim CN(0, \mathbf{C_Q})$, $\qquad i = 1...n-m$ also independent of any $\mathbf{Q}_i$. Then generate two codebooks of 2^{mR_C} length-m codewords $\mathbf{W}(s)$ and $\mathbf{T}(s)$ length-$(n-m)$, $s = 1, ..., 2^{mR_C}$, distributed according to P_W and P_T such that $R_C = I(\mathbf{W}_i; X)$. At a given source sequence X^m, the encoder looks for the s^*-th auxiliary codeword $\mathbf{W}(s^*)$ such that $(\mathbf{W}(s^*), X^m)$ are jointly typical. Then, $\mathbf{T}(s^*)$ is used to generate the channel input as

$$\begin{aligned} \mathbf{U}_i &= \mathbf{B}[\boldsymbol{\kappa} X_i - \mathbf{W}_i(s^*)], \qquad i = 1, ..., m \\ \mathbf{U}_i &= \mathbf{B}\mathbf{T}_{i-m}(s^*), \qquad i = m+1, ..., n. \end{aligned} \tag{3.40}$$

where $\mathbf{B}$ is a channel matrix such that the power constraint is satisfied, i. e. $Tr\{\mathbf{U}_i \mathbf{U_i}^\dagger\} \leq 1$. For the sake of notation we consider the equivalent channel $\mathbf{H} = \tilde{\mathbf{H}}\mathbf{B}$ where $\tilde{\mathbf{H}}$ is the actual channel state.

At a given channel output $\mathbf{V}_i = \tilde{\mathbf{H}}\mathbf{U}_i + \mathbf{N}_i$ $i = 1, ..., n$, the decoder looks for the auxiliary random variable $\mathbf{W}(s)$ such that $\mathbf{W}(s)$, and $\mathbf{T}(s)$ are simultaneously jointly typical with the current side information block sequence $\mathbf{Y}^m$ and the channel output $\mathbf{V}$.

Theorem 3. *The decoding will be successful with high probability, using typicality arguments if*

$$H(\mathbf{W}|\mathbf{V}_1, Y) \leq H(\mathbf{Q}_1) + (b-1)I(\mathbf{T}, \mathbf{V}_2) \tag{3.41}$$

where $\mathbf{V}_1$ and $\mathbf{V}_2$ are a sample of the channel output in $\mathbf{V}_1^m$ and $\mathbf{V}_{m+1}^n$, respectively.

Proof. From the HDA scheme, we have that

$$I(\mathbf{W}; X^m) \leq I(\mathbf{W}, \mathbf{T}; \mathbf{V}, Y^m) \tag{3.42}$$

The right hand side is equivalent to

$$\begin{aligned} I(\mathbf{W}; X^m) &= \sum_{i=1}^{m} I(\mathbf{W}_i; X_i) = \\ &= m[H(\mathbf{W}_1) - H(\mathbf{W_1}|X)] = \\ &= m[H(\mathbf{W}_1) - H(\mathbf{Q}_1)]. \end{aligned} \tag{3.43}$$

The left hand side

$$
\begin{aligned}
I(\mathbf{W},\mathbf{T};\mathbf{V},Y^m) &= I(\mathbf{W};\mathbf{V}_1^m,Y_1^m)+I(\mathbf{T};\mathbf{V}_{m+1}^n)\\
&= mI(\mathbf{W}_1;\mathbf{V}_1,Y)+(n-m)I(\mathbf{T},\mathbf{V}_2)\\
&= m[H(\mathbf{W}_1)-H(\mathbf{W}_1|\mathbf{V}_1;Y)]\\
&+ (n-m)I(\mathbf{T},\mathbf{V}_2)
\end{aligned}
\tag{3.44}
$$

where $\mathbf{V}_1$ and $\mathbf{V}_2$ is a sample that behaves statistically as a sample in $\mathbf{V}_1^m$ and $\mathbf{V}_{m+1}^n$ respectively. Similarly, $\mathbf{W}_2$ is used for a sample in $\mathbf{W}_{m+1}^n$. Equalities comes from the chain rule and the i.i.d generation of $\mathbf{W}_i$. Dividing both sides leads to the desired result. □

For the Gaussian case, the outage region is then given by

$$
\begin{aligned}
&\mathcal{O} = \Big\{(\mathbf{H},\gamma^2):\\
&\log\det\left(\left(\frac{\mathbf{K}}{a}+\mathbf{C_Q}\right)+[\left(\frac{\mathbf{K}}{a}+\mathbf{C_Q}\right)-\mathbf{I}]\mathbf{C_Q}\mathbf{H}^\dagger\mathbf{H}\mathbf{C_Q}\right)\\
&> \log\left(\det(\mathbf{I}+\mathbf{H}\mathbf{C_G}\mathbf{H}^\dagger)\right)^b-\log\det(\mathbf{C_Q})^{b-1}\Big\}
\end{aligned}
\tag{3.45}
$$

where $\mathbf{K}=\kappa\kappa^\dagger$.

Proof. We have that $H(\mathbf{W}_1|\mathbf{V}_1,Y) = H(\mathbf{W}_1,\mathbf{V}_1,Y) - H(\mathbf{V}_1,Y)$. Let $\mathbf{G} = [\mathbf{W},\mathbf{V},Y]^\dagger$. The differential entropy of $\mathbf{G}$, a multivariate gaussian random variable, is given by $h(\mathbf{G}) = \log((2\pi e)^{M_t}\det(\mathbf{C_G}))$ where $\mathbf{C_G} = E[\mathbf{G}\mathbf{G}^\dagger]$ and given by

$$
\mathbf{C_G} = \begin{bmatrix} \mathbf{K}+\mathbf{C_Q} & \mathbf{C_Q}\mathbf{H}^\dagger & \gamma\boldsymbol{\kappa}\\ \mathbf{H}\mathbf{C_Q} & \mathbf{I}+\mathbf{H}\mathbf{C_Q}\mathbf{H}^\dagger & \mathbf{0}\\ \gamma\boldsymbol{\kappa}^\dagger & \mathbf{0} & a \end{bmatrix}
\tag{3.46}
$$

where $\mathbf{K}=\boldsymbol{\kappa}\boldsymbol{\kappa}^\dagger$ and $a = 1+\gamma^2$. Using the determinant property of a block matrix,

$$
\begin{aligned}
&\det(\mathbf{C_G}) =\\
&= a\det\left(\begin{bmatrix} \mathbf{K}+\mathbf{C_Q} & \mathbf{C_Q}\mathbf{H}^\dagger\\ \mathbf{H}\mathbf{C_Q} & \mathbf{I}+\mathbf{H}\mathbf{C_Q}\mathbf{H}^\dagger\end{bmatrix} - \frac{1}{a}\begin{bmatrix}\gamma\boldsymbol{\kappa}\\ 0\end{bmatrix}\begin{bmatrix}\gamma\boldsymbol{\kappa} & 0\end{bmatrix}\right)\\
&= a\det\left(\begin{bmatrix} \frac{1}{a}\mathbf{K}+\mathbf{C_Q} & \mathbf{C_Q}\mathbf{H}^\dagger\\ \mathbf{H}\mathbf{C_Q} & \mathbf{I}+\mathbf{H}^\dagger\end{bmatrix}\right)\\
&= a\det\left(\frac{\mathbf{K}}{a}+\mathbf{C_Q}\right)\det\Big(\mathbf{I}+\mathbf{H}\mathbf{C_Q}\mathbf{H}^\dagger\\
&- \mathbf{H}\mathbf{C_Q}\left(\frac{\mathbf{K}}{a}+\mathbf{C_Q}\right)^{-1}\mathbf{C_Q}\mathbf{H}^\dagger\Big)
\end{aligned}
\tag{3.47}
$$

By Silvester determinant theorem, $\det(I_m + AB) = \det(I_n + BA)$, for A $m \times n$ and B $n \times m$

$$\det\left(\mathbf{I}_{M_t} + \mathbf{H}\mathbf{C}_{\mathbf{Q}}[\mathbf{I} - \left(\frac{\mathbf{K}}{a} + \mathbf{C}_{\mathbf{Q}}\right)^{-1}]\mathbf{C}_{\mathbf{Q}}\mathbf{H}^{\dagger}\right)$$
$$= \det\left(\mathbf{I}_{M_t} + [\mathbf{I} - \left(\frac{\mathbf{K}}{a} + \mathbf{C}_{\mathbf{Q}}\right)^{-1}]\mathbf{C}_{\mathbf{Q}}\mathbf{H}^{\dagger}\mathbf{H}\mathbf{C}_{\mathbf{Q}}\right) \tag{3.48}$$

Now, we use that $\det(X + AB) = \det(X)\det(I + AX^{-1}B)$ for X invertible to have

$$\det\left(\frac{\mathbf{K}}{a} + \mathbf{C}_{\mathbf{Q}}\right)\det\left(\mathbf{I} + [\mathbf{I} - \left(\frac{\mathbf{K}}{a} + \mathbf{C}_{\mathbf{Q}}\right)^{-1}]\mathbf{C}_{\mathbf{Q}}\mathbf{H}^{\dagger}\mathbf{H}\mathbf{C}_{\mathbf{Q}}\right)$$
$$= \det\left(\left(\frac{\mathbf{K}}{a} + \mathbf{C}_{\mathbf{Q}}\right) + [\left(\frac{\mathbf{K}}{a} + \mathbf{C}_{\mathbf{Q}}\right) - \mathbf{I}]\mathbf{C}_{\mathbf{Q}}\mathbf{H}^{\dagger}\mathbf{H}\mathbf{C}_{\mathbf{Q}}\right)$$

Similarly, the differential entropy $H(\mathbf{V}_i, Y)$ is given by

$$H(\mathbf{V}, Y) = \log\left((2\pi e)^{M_t} a \det(\mathbf{I} + \mathbf{H}\mathbf{C}_{\mathbf{Q}}\mathbf{H}^{\dagger})\right). \tag{3.49}$$

and the mutual information

$$I(\mathbf{V}_2, \mathbf{Q}) = \log\left(\det(\mathbf{I} + \mathbf{H}\mathbf{C}_{\mathbf{Q}}\mathbf{H}^{\dagger})\right) - \log\det(\mathbf{C}_{\mathbf{Q}})$$
$$= \log\left(\det({\mathbf{C}_{\mathbf{Q}}}^{-1} + \mathbf{H}\mathbf{H}^{\dagger})\right). \tag{3.50}$$

Substituting we have

$$\mathcal{O} = \Big\{(\mathbf{H}, \gamma^2) :$$
$$v \log \frac{\det\left(\left(\frac{\mathbf{K}}{a} + \mathbf{C}_{\mathbf{Q}}\right) + [\left(\frac{\mathbf{K}}{a} + \mathbf{C}_{\mathbf{Q}}\right) - \mathbf{I}]\mathbf{C}_{\mathbf{Q}}\mathbf{H}^{\dagger}\mathbf{H}\mathbf{C}_{\mathbf{Q}}\right)}{\det(\mathbf{I} + \mathbf{H}\mathbf{C}_{\mathbf{G}}\mathbf{H}^{\dagger})}$$
$$> \log\left(\det(\mathbf{I} + \mathbf{H}\mathbf{C}_{\mathbf{G}}\mathbf{H}^{\dagger})\right)^{b-1} - \log\det(\mathbf{C}_{\mathbf{Q}})^{b-1}\Big\} \tag{3.51}$$

and operating leads to the result. □

If $\mathbf{W}$ could be decoded, the source sequence is reconstructed with an mmse estimator $\hat{X} = E[X|V, Y, W]$. The achievable distortion is given by

$$D(\boldsymbol{\kappa}, \mathbf{H}, \gamma^2) = (1 + \gamma^2 + \boldsymbol{\kappa}(\mathbf{I} + \mathbf{H}\mathbf{C}_{\mathbf{Q}}\mathbf{H}^{\dagger})\boldsymbol{\kappa}^{\dagger})^{-1} \tag{3.52}$$

Proof. The available information to reconstruct source sample X_i, $i = 1, ..., m$ with an mmse estimator can be modeled by the lineal model

$$\begin{bmatrix} \mathbf{W}_i \\ \mathbf{V}_i \\ Y_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{\kappa} \\ \mathbf{0} \\ \gamma \end{bmatrix} X + \begin{bmatrix} \mathbf{Q}_i \\ \mathbf{H}\mathbf{Q}_i + \mathbf{N}_i \\ Z_i \end{bmatrix} \tag{3.53}$$

Let $\mathbf{t}_i = \begin{bmatrix}\boldsymbol{\kappa} & \mathbf{0} & \gamma\end{bmatrix}^\dagger$ and $\mathbf{s}_i = \begin{bmatrix}\mathbf{Q}_i & \mathbf{H}\mathbf{Q}_i + \mathbf{N}_i & Z_i\end{bmatrix}^\dagger$. The distortion achieved with the optimal estimator is given by $D = [1 + \mathbf{t}\mathbf{C}_s\mathbf{t}^\dagger]^{-1}$ where $\mathbf{C_g}$

$$\mathbf{C_g} = \begin{bmatrix}\mathbf{C_G} & \mathbf{C_G}\mathbf{H}^\dagger & \mathbf{0} \\ \mathbf{H}\mathbf{C_G} & \mathbf{H}\mathbf{C_G}\mathbf{H}^\dagger + \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1\end{bmatrix}. \tag{3.54}$$

Using the block inverse properties, we have

$$\mathbf{C_g}^{-1} = \begin{bmatrix}\mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & 1\end{bmatrix} \tag{3.55}$$

where

$$\mathbf{A}^{-1} = \begin{bmatrix}\mathbf{C_G} & \mathbf{C_G}\mathbf{H}^\dagger \\ \mathbf{H}\mathbf{C_G} & \mathbf{H}\mathbf{C_G}\mathbf{H}^\dagger + \mathbf{I}\end{bmatrix}^{-1} = \begin{bmatrix}\mathbf{H}\mathbf{C_G}\mathbf{H}^\dagger + \mathbf{I} & -\mathbf{H}^\dagger \\ -\mathbf{H} & \mathbf{I}\end{bmatrix}$$

where we have used the block matrix inversion and the Woodbury matrix identity. Finally, by multiplying $\mathbf{t}\mathbf{C}_s\mathbf{t}^\dagger$, the expression is

$$D(\boldsymbol{\kappa}, \mathbf{H}, \gamma^2) = (1 + \gamma^2 + \boldsymbol{\kappa}(\mathbf{I} + \mathbf{H}\mathbf{C_Q}\mathbf{H}^\dagger)\boldsymbol{\kappa}^\dagger)^{-1} \tag{3.56}$$

Note that $\boldsymbol{\kappa}(\mathbf{I} + \mathbf{H}\mathbf{C_Q}\mathbf{H}^\dagger)\boldsymbol{\kappa}^\dagger$ is a quadratic form and then we can express it as

$$\boldsymbol{\kappa}(\mathbf{I} + \mathbf{H}\mathbf{C_Q}\mathbf{H}^\dagger)\boldsymbol{\kappa}^\dagger \tag{3.57}$$

□

The transmitter is not aware of the actual realization of the channel and hence, as in the digital case, it seems that there is no reason to benefit some dimensions over others in the channel transmission. Hence, we choose $\mathbf{B} = \frac{1}{\sqrt{M_*}}\mathbf{C_Q^{-1}}$. Similarly the influence of C_Q in the expressions should depend on $\mathbf{H}$. Similarly we pick $\mathbf{C_Q} = \mathbf{I}$. For this particular choice, the outage expression in (3.45) becomes, using Silvester determinant theorem

$$\begin{aligned}\mathcal{O} = \Big\{&(\mathbf{H}, \gamma^2) : \\ &\det\left(\mathbf{I} + \frac{\mathbf{K}}{a}\left(\mathbf{I} + \frac{1}{M_*}\mathbf{H}^\dagger\mathbf{H}\right)\right) > \det(\mathbf{I} + \frac{1}{M_*}\mathbf{H}\mathbf{H}^\dagger)^b\Big\} \\ = 1 + &\frac{1}{a}\boldsymbol{\kappa}\left(\mathbf{I} + \frac{1}{M_*}\mathbf{H}^\dagger\mathbf{H}\right)\boldsymbol{\kappa}^\dagger > \det(\mathbf{I} + \frac{1}{M_*}\mathbf{H}\mathbf{H}^\dagger)^b\end{aligned} \tag{3.58}$$

Now we use that any Wishart matrix $\mathbf{H}\mathbf{H}^\dagger$ can be decomposed as $\mathbf{U}\boldsymbol{\Lambda}(\rho)\mathbf{U}^\dagger$ where $\boldsymbol{\Lambda}(\rho)$ is a diagonal matrix that contains the eigenvalues of the matrix and $\mathbf{U}$ is a Haar distributed matrix. For any Haar distribute matrix $\mathbf{U}$ we have that for any unit vector $\mathbf{w}$, with unit norm $|\mathbf{w}| = 1$, $\mathbf{U}\mathbf{w} = \frac{\mathbf{s}}{|\mathbf{s}|}$ with $\mathbf{s} \sim \mathcal{CN}(0, \mathbf{I})$. Let $\boldsymbol{\kappa} = \kappa\hat{\boldsymbol{\kappa}}$ where $\hat{\boldsymbol{\kappa}}$ is a unit vector colineal with $\boldsymbol{\kappa}$ and $\kappa \geq 0$ and $\left(\mathbf{I} + \frac{1}{M_*}\mathbf{H}^\dagger\mathbf{H}\right) = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\dagger$ with $\Lambda = \text{diag}(1 + \frac{\lambda_i}{M_*})$ for $i = 1, ..., M_*$, where

$\lambda_1 \leq ... \leq \lambda_{M_*}$ are the eigenvalues of $\tilde{\mathbf{H}}\tilde{\mathbf{H}}^\dagger$. Note that the eigenvalues of $\mathbf{H}^\dagger\mathbf{H}$ are the same of $\mathbf{H}\mathbf{H}^\dagger$. Then,

$$\begin{aligned} &\boldsymbol{\kappa}\left(\mathbf{I}+\frac{1}{M_*}\mathbf{H}^\dagger\mathbf{H}\right)\boldsymbol{\kappa}^\dagger \\ &= \boldsymbol{\kappa}\left(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\dagger\right)\boldsymbol{\kappa}^\dagger \\ &= \frac{\mathbf{s}}{|\mathbf{s}|}\Lambda\frac{\mathbf{s}}{|\mathbf{s}|}^\dagger \\ &= \frac{\kappa^2}{|\mathbf{s}|^2}\sum_{i=1}^{M_*}(1+\lambda_i)s_i^2 \end{aligned} \tag{3.59}$$

where $s_i \sim \mathcal{CN}(\prime, \infty)$ and then

$$\mathcal{O} = \left\{(\mathbf{H},\gamma^2) : 1 + \frac{\kappa^2}{|\mathbf{s}|^2}\sum_{i=1}^{M_*}(1+\lambda_i)s_i^2 > \prod_{i=1}^{M_*}\left(1+\frac{\lambda_i}{M_*}\right)^b\right\},$$

and

$$D(\boldsymbol{\kappa},\mathbf{H},\gamma^2) = (1+\gamma^2+\frac{\kappa^2}{|\mathbf{s}|^2}\sum_{i=1}^{M_*}(1+\lambda_i)s_i^2)^{-1} \tag{3.60}$$

Then, the expected distortion can be expressed as

$$\begin{aligned} ED_h(\boldsymbol{\kappa}) = &\iint_{\mathcal{O}_h^c} D(\boldsymbol{\kappa},\mathbf{H},\gamma^2)p_\gamma(\gamma)p_h(\mathbf{H})d\gamma d\mathbf{H} \\ &+ \iint_{\mathcal{O}_h} D(\mathbf{0},\mathbf{H},\gamma^2)p_\gamma(\gamma)p_h(\mathbf{H})d\gamma d\mathbf{H}. \end{aligned} \tag{3.61}$$

3.4 Multi-layers schemes

3.4.1 Multi-layer separate coding without binning

A multi-layer transmission scheme is proposed in [2] in which the transmitter combats channel fading by superposing multiple layers, where each layer is the successive refinement for the previous layers. The decoder decodes as many layers as possible depending on the channel state and reconstructs the source. The better the channel state, the more layers can be decoded and the smaller is the achieved distortion.

At the encoder, we generate L Gaussian quantization codebooks, at rates $R_k = I(X; W_k|W_1^{k-1})$, $k = 1, ..., L$, such that each Gaussian codebook is a refinement codebook for the previous layers. The quantization codewords can be modeled, for $k = 1, ..., L$, as $W_k = X + \sum_{i=k}^L Q_i = W_{k+1} + Q_k = X + \bar{Q}_i$ with $Q_i \sim (0, \sigma_i^2)$ independent of each other, and $\overline{Q}_k \triangleq \sum_{i=k}^L Q_k$. The quantization noise is found to be $\sum_{i=k}^L \sigma_i^2 = (2^{\sum_{i=1}^l R_i} - 1)^{-1}$. For the channel codewords, generate L codebooks $\mathbf{U}_{l,i}$ $i = 1, ..., n$ generated i.i.d. with $\mathcal{CN}(0, 1)$ and let $\boldsymbol{\rho} = [\rho_1, ..., \rho_L]^T$ be the power allocation among each channel codeword such that $\rho = \sum_{i=1}^L \rho_i$. Then the channel codeword $\mathbf{U}_i$ is generated as the addition of the codewords $\mathbf{U}_{k,i}$ scaled with the corresponding power allocation $\sqrt{\rho_k}$. Define $\bar{\rho}_k = \sum_{i=k}^L \rho_i$. The output of the channel is given by $\mathbf{V}_i = \mathbf{H}\sum_{j=1}^L \sqrt{\rho_j}\mathbf{U}_{j,i} + \mathbf{N}_i$, $i = 1, ..., n$. At reception, the decoder applies successive refinement decoding starting from the first layer and at each layer it uses typical decoding as in

Section 3.2.1. If the receiver cannot decode layer k, it does not try to decode the subsequent $j > k$ layers. Provided $k-1$ layers are decoded, the outage event at layer k is given by

$$\mathcal{A}_k = \{\mathbf{H} : I(\mathbf{U}_{k,i}; \mathbf{V}_{k,i} | \mathbf{U}_{1,i}^{k-1}) < R_k\}, \tag{3.62}$$

where, for $k = 1, ..., L$,

$$\begin{aligned} I(\mathbf{U}_{k,i}, \bar{\mathbf{V}}_{k,i}) &= I(\bar{\mathbf{V}}_{k,i}; \mathbf{U}_{k,i}, \bar{\mathbf{U}}_{k+1,i}) - I(\bar{\mathbf{V}}_{k,i}; \bar{\mathbf{U}}_{k+1,i} | \mathbf{U}_{k,i}) \\ &= \log \frac{\det\left(\mathbf{I} + \frac{\bar{\rho}_k}{M_t}\mathbf{H}\mathbf{H}^\dagger\right)}{\det\left(\mathbf{I} + \frac{\bar{\rho}_{k+1}}{M_t}\mathbf{H}\mathbf{H}^\dagger\right)}, \end{aligned} \tag{3.63}$$

with $\bar{\rho}_{L+1} = 0$. We define $\mathcal{B}_k \triangleq \bigcup_{i=1}^{k} \mathcal{A}_i$. The receiver uses the side information available together with the decoded layers to reconstruct the source with an MMSE estimator. The expected distortion for a multiple-layer scheme with L layers is given by

$$ED_{ml}(\mathbf{R}) = \sum_{k=0}^{L} E_\gamma \left[D_d \left(b \sum_{j=0}^{k} R_j, \gamma \right) \right] (P_{out}^{i+1} - P_{out}^{i}),$$

where $\mathbf{R} \triangleq [R_1, ..., R_L]$, $P_{out}^i \triangleq Pr\{\mathcal{B}_i\}$, $R_0 \triangleq 0, P_{out}^0 = 0$ and $P_{out}^{L+1} = 1$.

3.4.2 Multi-layer NBJD

In the previous multi-layer scheme, the side information is not used in the coding and decoding of the multiple layers and is only used for the estimation at the receiver. We propose a multi-layer scheme that uses NBJD at each layer to use the joint quality of the channel and side information.

At the encoder, the transmission scheme is the same as in Section 3.4.1. At reception, the decoder uses a joint typicality decoder for each layer. The outage event at layer k, provided that $k-1$ layers have been decoded, is given by

$$\begin{aligned} \mathcal{L}_k &= \{(\mathbf{H}, \gamma) : bI(\mathbf{U}_{i,k}; \mathbf{V}_i | \mathbf{U}_{i,1}^{k-1}) < I(X; W_k | Y, W_1^{k-1}) | \mathcal{O}_{k-1}^c\} \\ &= \{(\mathbf{H}, \gamma) : bI(\mathbf{U}_{i,k}; \mathbf{V}_{i,k} | \mathbf{U}_{i,1}^{k-1}) < I(X; W_k | Y, W_1^{k-1})\}, \end{aligned} \tag{3.64}$$

where $\mathcal{O}_k$ is the event of having an outage at the previous $k-1$ layers, i.e. $\mathcal{O}_k \triangleq \bigcup_{i=1}^{k} \mathcal{L}_i$. Note that $\mathcal{L}_k$'s are mutually exclusive. Then, the event of being able to decode exactly k layers, $\bar{\mathcal{O}}_k$, is found as

$$Pr\{\bar{\mathcal{O}}_k\} = Pr\{\mathcal{O}_{k+1}\} - Pr\{\mathcal{O}_k\} = Pr\{\mathcal{L}_{k+1}\}. \tag{3.65}$$

The left hand side of the outage event in (3.64) is given in (3.63). Similarly, the right hand side of the outage event can be calculated as

$$\begin{aligned} I(X; W_k | W_1^{k-1}, Y) &= I(X; W_k | Y) - I(X; W_{k-1} | Y) \\ &\overset{(a)}{=} H(W_k | Y) - H(\overline{Q}_k) - H(W_{k-1} | Y) + H(\overline{Q}_{k-1}) \\ &= \log \left(\frac{\sum_{i=k-1}^{L} \sigma_i^2}{\sum_{i=k}^{L} \sigma_i^2} \frac{1 + (1 + \gamma^2 \rho^x) \sum_{j=k}^{L} \sigma_j^2}{1 + (1 + \gamma^2 \rho^x) \sum_{j=k-1}^{L} \sigma_j^2} \right), \end{aligned}$$

and

$$I(X;W_1|Y) = \left(1 + \frac{1}{(1+\gamma^2\rho^x)\sum_{i=1}^{L}\sigma_i^2}\right). \tag{3.66}$$

Equality (a) is due to the independence of $\bar{Q}_i$ with X and Y, and

$$H(W_k|Y) = \log\left(\sum_{i=k}^{L}\sigma_i^2 + \frac{1}{1+\gamma^2\rho^x}\right). \tag{3.67}$$

Including the quantization noise variances the expression is given, with $R_0 = 0$, by

$$I(X;W_k|W_1^{k-1},Y) = \log\left(\frac{2^{\sum_{i=0}^{k} R_i} + \gamma^2\rho^x}{2^{\sum_{i=0}^{k-1} R_i} + \gamma^2\rho^x}\right), \tag{3.68}$$

Then, the receiver reconstructs the source with an MMSE estimator with the side information and decode layers. The expected distortion can then be expressed as

$$ED_{mj}(\mathbf{R}) = \sum_{k=0}^{L}\iint_{\mathcal{L}_{k+1}} D_d\left(\sum_{i=0}^{k} R_i, \gamma\right) p_h(h)p_\gamma(\gamma)dhd\gamma$$

where $\mathcal{L}_{L+1}$ is the set of states in which all layers are decoded.

Appendix 3.A Proof of the expected distortion lower bound

We derive a lower bound on the expected distortion by providing the transmitter with the actual channel and side information states $(\gamma, \mathbf{H})$. A transmitter with the actual $\mathbf{H}$ and γ realizations at each block can use the optimal transmission scheme based on separate source-channel coding such that no outage occurs and the minimum possible distortion is achieved at each block. The following distortion is achieved at state $(\mathbf{H}, \gamma)$ [16]

$$D_{low}(\mathbf{H}, \gamma) = \frac{2^{-bC(\mathbf{H})}}{\gamma^2 \rho_s + 1}, \tag{3.69}$$

where $C(\mathbf{H})$ is the instantaneous channel capacity, i.e.

$$C(\mathbf{H}) = \sup_{\mathbf{Q} \succcurlyeq 0, \mathrm{Tr}\{\mathbf{Q}\} \leq M_*.} \log\det\left(\mathbf{I} + \frac{\rho_c}{M_*}\mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\right), \tag{3.70}$$

where we drop the explicit dependance on b, ρ and ρ_s for the sake of brevity as they are constants for a given scenario.

It is well known that for a channel state with an SVD decomposition $\mathbf{H}\mathbf{H}^\dagger = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^\dagger$, the instantaneous capacity is maximized by transmitting a channel input $\mathbf{U}$ with a covariance matrix $\mathbf{Q} = \mathbf{T}^\dagger\tilde{\mathbf{D}}\mathbf{T}$ where $\tilde{\mathbf{D}}$ is a diagonal matrix with $[D]_{ij}$ the solution to the water-filling problem [17].

Appendix 3.B Proof of distortion for uncoded transmission

The available information at the mmse estimation step can be modeled liniearly as

$$\begin{aligned}\begin{bmatrix}\mathbf{V}\\ \mathbf{Y}\end{bmatrix} &= \begin{bmatrix}\sqrt{\rho}\mathbf{H}_{M_*}\\ \sqrt{\rho_s}\gamma\mathbf{I}_{M_*}\end{bmatrix}\mathbf{x}_{M_*} + \begin{bmatrix}\mathbf{N}\\ \mathbf{Z}\end{bmatrix}\\ &= \tilde{\mathbf{H}}\mathbf{x} + \tilde{\mathbf{N}}.\end{aligned} \tag{3.71}$$

where $\tilde{\mathbf{H}} \in \mathbb{C}^{2M_* \times 2M_*}$ and $\tilde{\mathbf{N}} \sim \mathcal{CN}(0, \mathbf{I})$. Then the average achievable distortion is given by

$$\begin{aligned}\frac{1}{M_*}\mathrm{Tr}\{\mathbf{D}\} &= \frac{1}{M_*}\mathrm{Tr}\{(\mathbf{I}_{M_*} + \rho\tilde{\mathbf{H}}^\dagger\mathbf{I}\tilde{\mathbf{H}})^{-1}\}\\ &= \frac{1}{M_*}\mathrm{Tr}\{(\mathbf{I}_{M_*} + \rho\mathbf{H}^\dagger\mathbf{H} + \rho_s\gamma^2\mathbf{I})^{-1}\}\\ &= \frac{1}{M_*}\sum_{i=1}^{M_*}\frac{1}{1 + \rho\mu_i + \rho_s\gamma^2}.\end{aligned} \tag{3.72}$$

Appendix 3.C Proof of distortion for NBJD-analog hybrid scheme

3.C.1 Case $M_* b \leq 1$

The available information is given by

$$\begin{bmatrix}\mathbf{V}\\ \mathbf{Y}\end{bmatrix} = \begin{bmatrix}\frac{\sqrt{\rho^{1-\eta}}}{M_*}\mathbf{H}_{M_*}\\ \gamma\mathbf{I}\end{bmatrix}\mathbf{X}_u + \begin{bmatrix}\frac{\sqrt{\rho(1-\rho^{-\eta})}}{M_*}\mathbf{H}_m\mathbf{U}_\mathbf{a} + \mathbf{N}\\ \mathbf{Z}\end{bmatrix} \tag{3.73}$$

where we stack the source sequences and side information sequences as $\mathbf{X}_u = [X_{1u}, ..., X_{nM_*u}]$ and $\mathbf{Y} = [Y_1, ..., Y_n]$. Similarly for $\mathbf{Z}$.

Then, the average distortion for a given $\mathbf{H}_{M_*}$ can be calculated as

$$\begin{aligned}
D_i(\mathbf{H}_{M_*}, \eta) &= \frac{1}{M_*} tr \left\{ \left(\mathbf{I} + \begin{bmatrix} \frac{\sqrt{\rho^{1-\eta}}}{M_*} \mathbf{H}_{M_*} & \gamma \mathbf{I} \end{bmatrix} \begin{bmatrix} \frac{\rho(1-\rho^{-\eta})}{M_*} \mathbf{H}_{M_*} \mathbf{H}_{M_*}^\dagger + \mathbf{I} & \mathbf{0}_{M_*} \\ \mathbf{0}_{M_*} & \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\sqrt{\rho^{1-\eta}}}{M_*} \mathbf{H}_{M_*} \\ \gamma \mathbf{I} \end{bmatrix} \right)^{-1} \right\} \\
&= \frac{1}{M_*} tr \left\{ \left(\mathbf{I}(1+\gamma^2) + \rho^{1-\eta} \mathbf{H}_{M_*}^\dagger (\rho(1-\rho^{-\eta}) \mathbf{H}_{M_*} \mathbf{H}_{M_*}^\dagger + \mathbf{I})^{-1} \mathbf{H}_{M_*} \right)^{-1} \right\} \\
&= \frac{1}{M_*} \sum_{i=1}^{M_*} \left(1 + \gamma^2 + \frac{\rho^{1-\eta} \mu_i}{1 + \rho(1-\rho^{1-\eta})\mu_i} \right)^{-1}, \qquad (3.74)
\end{aligned}$$

where $\mathbf{0}_{M_*}$ is a $M_* \times M_*$ matrix with 0 entries.

3.C.2 Case $M_* b > 1$

The available information is given by

$$\begin{bmatrix} \mathbf{W} \\ \mathbf{V} \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{M_*} \\ \mathbf{0} \\ \gamma \mathbf{I}_{M_*} \end{bmatrix} \mathbf{X} + \begin{bmatrix} \frac{\sigma^2}{M_*} \mathbf{I} & \mathbf{H}\sigma & \mathbf{0} \\ \mathbf{H}^\dagger \sigma & \mathbf{I} + \frac{1}{M_t} \mathbf{H}\mathbf{H}^\dagger & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \qquad (3.75)$$

where $\mathbf{X}_i$ is a vector correspond to M_t source samples, and hence, using the Woodbury identity,

$$\begin{aligned}
D_{pd}(R, \mathbf{H}, \gamma) &= \mathrm{Tr} \left\{ \left(\frac{1}{\sigma^2} (\mathbf{I} + \frac{1}{M_*} \mathbf{H}\mathbf{H}^\dagger) + (\gamma^2 \rho^x + 1) \mathbf{I} \right)^{-1} \right\} \\
&= \sum_{i=1}^{M_*} \left(\frac{1}{\sigma^2} (1 + \frac{\lambda_i}{M_*}) + \gamma^2 \rho^x + 1 \right)^{-1} \qquad (3.76)
\end{aligned}$$

Chapter 4

Numerical Results

We have seen in previous chapters that closed from solutions for the expected distortion are complicated to obtain for the general finite snr regime.

In this chapter we obtain, through numerical optimization, the minimum expected distortion in function of the signal to noise ratio. Those numerical simulations are also interesting to check the range of validity of the high snr regime results for the Expected Distortion function with the Distortion Exponent that will be calculated in next chapter.

4.1 Numerical optimization

In this section, we compare numerically the optimal expected distortion, $ED^*(\Omega)$ for the schemes presented in the previous section by optimizing over the corresponding design parameters for a given discretized Ω.

For fixed b, x values, the numerical optimization to compute the best expected distortion over a set of design parameters $\Lambda(\mathcal{S})_n = [\lambda_{i_1,i_2,...,i_L}]$ where L is the number of design parameters in the system, and $i_l = 1, ..., R_l$,$l = 1, ..., L$ with outage event $\mathcal{O}_{\mathcal{S}}(\Lambda(\mathcal{S})_n)$, in N uniformly separated ρ_n, $n = 1, ..., N$ points for a particular scheme $\mathcal{S}$ is done by exhaustive numerical gridsearch

$$ED^*_{\mathcal{S}}(\rho_n) = \max_{\Lambda(\mathcal{S})_n} E[D](\rho_n, \Lambda(\mathcal{S})_n) \tag{4.1}$$

Construct a set D with the same dimensions of $\Lambda(\mathcal{S})_n$ and an indicator function I of the same dimensionality with a one to one correspondence between the 3 sets.

For simplicity in what follows we assume a single parameter to design, for example the rate R, and then $\Lambda(\mathcal{S})_n = [r_i]$ with $i = 1, ..., R$. If there are more parameters to design, for example power allocation and rate, the only difference is the increase in dimensionality of the gridsearch domain: a line for one parameter, a plane for two parameters, a cube for three parameters, etc. Hence, we have r_i, I_i and D_i for $i = 1, ...R$. Then, we can simplify

$$ED^*_{\mathcal{S}}(\rho_n) = \max_{r_i} E[D_i](\rho_n). \tag{4.2}$$

At each ρ_n, generate T channel states and side information state realizations pairs $(\mathcal{H}_t, \gamma_t)$ $t = 1, ..., T$. For each pair t, compute the outage event indicator I_{ti} for each i-th parameter in

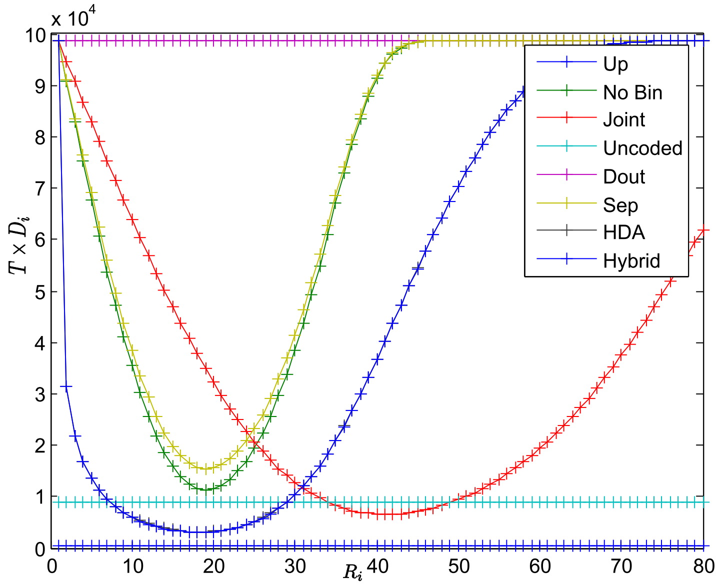

Figure 4.1: Numerical Expected Distortion function at $\rho_n = 10dB$.

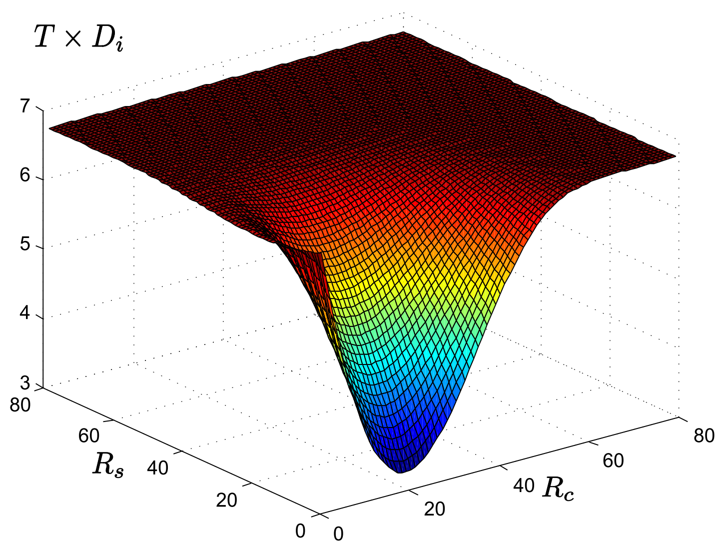

Figure 4.2: Numerical Expected Distortion function at $\rho_n = 10dB$ for Separation with binning scheme.

$\Lambda(\mathcal{S})_n$, the indicator $\mathbf{I}$ will be one if there has been an outge or 0 if it has not. Then, compute the distortion value D_{ti} for each parameter $\Lambda(\mathcal{S})_n$ according to I_{ti}, i.e. if there has been an outage or a correct transmission for parameter i at a particular realization t.

Once all D_{it} have been calculated for each realization pair, average over the realizations to obtain the expected distortion value at a particular r_i, $E[D_i](\rho_n)$. Then the optimal r_i is the one achieving the minimum $E[D_i](\rho_n)$. See Figure 4.1 for an example of gridsearch for a particular ρ_n in one dimension and Figure 4.2 for a gridsearch for two parameters.

We numerically obtain the lower bound for the expected distortion and the numerical optimized solution ED^* for separate source-channel coding scheme with and without binning, for NBJD, for uncoded transmission, HDA and digital-analog hybrid scheme in a SISO sce-

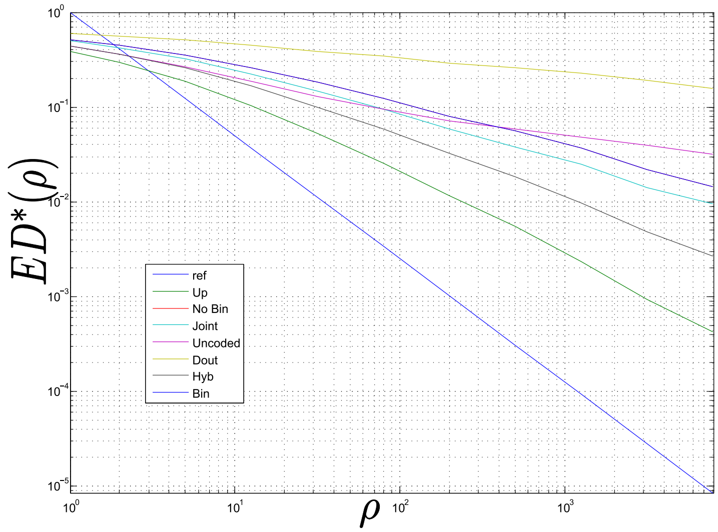

Figure 4.3: ED^* for $b = 0.8$ and $x = 0.3$.

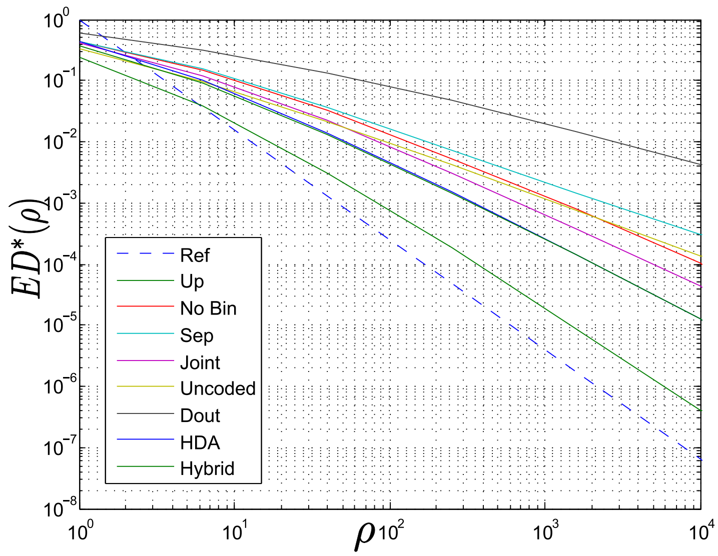

Figure 4.4: ED^* for $b = 2$ and $x = 0.8$

nario. We also show the achievable distortion D_{out} using only the available side information at the receive as a reference. Results are shown in Figure 4.3 for the bandwidth compression regime, i.e. $b < 1$, and for the bandwidth expansion regime, $b \geq 1$, in Figure 4.4.

For both the expansion and compression bandwidth regime, observe that ED^* coincide for the non binning scheme and the binning case. We can conclude that ED^* for separation is optimized by ignoring the side information when encoding, (i.e., no binning), as the scheme suffers from outage in both the channel and the side information. Observe this behavior in Figure 4.2. Unfortunately, he have been not able to give a proof for this observation.

Observe, in Figure 4.1 that for the same quantization rate, NBJD is able to decode higher rates and thus lower distortions are possible, i.e. at any given quantization rate, NBJD's probability of outage is always lower compared to the separation scheme. For both bandwidth regimes, the optimized NBJD performs better than separation with and without binning, i.e. $ED^*_{nb}(\Omega) \geq ED^*_{sb}(\Omega) \geq ED^*_j(\Omega)$, at moderate SNR. This can be proven analytically.

Lemma 1. *At any SNR, the expected distortion ED_j of NBJD outperforms the expected distortion ED_{sb} of the separation scheme, i.e.,*

$$ED_{sb}(R - bR_c, R) \geq ED_j(R) \tag{4.3}$$

where the equality only holds if $R = R_j = R_s + bR_c = 0$.

Observe that at high SNR, there seems to be a constant distance between NBJD and transmission based on separation.

The behavior of uncoded transmission at low SNR outperforms joint and separation schemes. On the contrary, at high SNR, its performance worsens. While in high SNR regimes it falls short from the others, in compression bandwidth regime, it tends to behave with a constant distance with D_{out}.

The behavior or hybrid analog-digital transmission gets the best of each of each scheme. At low SNR it benefits from the good performance of the analog part and as the SNR increases, it benefits from the digital component. Note that the hybrid of both performs better than separate NBJD and uncoded transmission in both bandwidth regimes.

In the expansion bandwidth regime, observe that HDA behaves very closely to hybrid analog-digital scheme. At low SNR, hybrid outperforms HDA, but they rapidly converge to the same performance as the SNR grows. In favor of HDA, note that for the special case of $b = 1$, hybrid reduces to pure uncoded transmission.

Chapter 5

High SNR Results: Distortion Exponent Analysis

In this chapter, we study the performance of the proposed schemes in the high SNR regime, i.e. when $\rho \to \infty$. The performance measure is the *distortion exponent*, defined as

$$\Delta(b,x) \triangleq -\lim_{\rho\to\infty} \frac{\log ED}{\log \rho}, \tag{5.1}$$

where we have the quality of the side information given by $x \triangleq \lim_{\rho\to\infty} \frac{\log \rho_s}{\log \rho}$.

5.1 Distortion exponent upper bound

We first provide an upper bound on the distortion exponent by analyzing the high SNR behavior of ED_{low}.

Theorem 4. *The distortion exponent $\Delta_{up}(b,x)$ is upper bounded by*

$$\Delta_{up}(b,x) = x + \sum_{i=1}^{M_*} \min\{b, 2i-1+M^*-M_*\} \tag{5.2}$$

Proof. See Appendix 5.A. □

5.2 Single layer schemes: separation, uncoded and NBJD

The following theorems characterize the best distortion exponents achievable by the presented schemes.

Theorem 5. *The distortion exponents for the single layer separate source-channel coding, separation without binning, and single layer NBJD scheme, are equal to each other and given by,*

$$\Delta_{sb}(b,x) = \begin{cases} \max\left\{x, b\frac{M^*M_*+x}{b+M^*+M_*-1}\right\} & \text{if } b \geq x+(M^*-1)(M_*-1), \\ \max\left\{x, b\frac{M^*M_*+x-k(k+1)}{b+M^*+M_*-1-2k}\right\} & \text{if } b \in \left[\frac{(M^*-(k+1))(M_*-(k+1))}{k+1}, \frac{(M^*-k-1)(M_*-k-1)}{k}\right), \\ & k = 1, ..., M_*-1. \end{cases} \tag{5.3}$$

Proof. See Appendix 5.B. □

Remark 1. *The optimal distortion exponent for the separate scheme is obtained for $R_s = 0$, i.e. by including a single codeword in each bin. This result coincides with the numerical results for the finite SNR regime studied in Chapter 4, i.e. that binning does not improve the expected distortion performance.*

Theorem 6. *The distortion exponent of uncoded transmission is characterized by*

$$\Delta_u = \begin{cases} x & \text{if } M_* b < 1, \\ \max\{1, x\} & \text{if } M_* b \geq 1. \end{cases} \tag{5.4}$$

Proof. See Appendix 5.C. □

Theorem 7. *The distortion exponents for the single layer separate source-channel coding, separation without binning, and single layer NBJD scheme, are equal to each other and given by,*

$$\Delta_j(b,x) = \begin{cases} \max\left\{x, b\frac{M^* M_* + x}{b + M^* + M_* - 1}\right\} & \text{if } b \geq x + (M^* - 1)(M_* - 1), \\ \max\left\{x, b\frac{M^* M_* + x - k(k+1)}{b + M^* + M_* - 1 - 2k}\right\} & \text{if } b \in \left[\frac{(M^* - (k+1))(M_* - (k+1))}{k+1}, \frac{(M^* - k - 1)(M_* - k - 1)}{k}\right), \\ & k = 1, ..., M_* - 1. \end{cases} \tag{5.5}$$

Proof. See Appendix 5.D. □

Remark 2. *Note that it coincides with the distortion exponent achieved by separation with and without binning, given in 13.*

Theorem 8. *The distortion exponent $\Delta_h(b,x)$ of the NBJD-analog hybrid scheme is characterized by*

$$\Delta_h(b,x) = \begin{cases} \max\{x, M_* b + (1 - bM_*)x\} & \text{if } M_* b < x. \end{cases} \tag{5.6}$$

and for $b \geq M_* b$

$$\begin{aligned} \Delta_h(b,x) = 1 + & \frac{(b(M_* - k) - 1)((M^* - k)(M_* - k) - 1 + x)}{(M_* - k)(b + (M^* - k) + (M_* - k) - 1) - 1} \\ & \text{for } b \in \left[k - M^* + \frac{1}{M_*} - M_* + \frac{M^* M_* - 1 + x}{k}, k + 1 - M^* + \frac{1}{M_*} - M_* + \frac{M^* M_* - 1 + x}{k+1}\right) \\ & \text{for } k = 1, ..., M_* - 1. \end{aligned} \tag{5.7}$$

and

$$\Delta_h(b,x) = 1 + \frac{(bM_* - 1)(M^* M_* - 1 + x)}{M_*(b + M^* + M_* - 1) - 1} \qquad \text{for } \frac{1}{M_*} \geq b \geq \frac{M_*(1 + M^* - M_* + x)}{(M_* - 1)M_* - 1} \tag{5.8}$$

Proof. See Appendix 5.F for $M_* b > 1$ and 5.E for $M_* b \leq 1$. □

5.3 Single layer schemes: HDA

Theorem 9. *The distortion exponent of HDA source-channel coding scheme is given by*

$$\Delta_{hda}(b,x) = \begin{cases} x & \text{for } b \leq \frac{x}{M_*} \\ \frac{b(M^*M_* - (k(k+1)+x)}{b-1+M^*+M_*-2k} & \text{for } b \in \left[\frac{x+(M^*-(k+1))(M_*-(k+1))}{(k+1)}, \frac{x+(M^*-k)(M_*-k)}{k}\right) \\ & \text{for } k = 1, ..., M_* \\ \frac{M^*+M_*-1+(b-1)M^*M_*}{M^*+M_*-1+b-1+(b-1)x}, & \text{for } b \geq (M^*-1)(M_*-1)+x.. \end{cases} \tag{5.9}$$

Proof. See Appendix 5.G. □

5.4 Multi-layer schemes

Theorem 10. *The distortion exponent of L layer separate source-channel coding scheme without binning and for multi-layer NBJD schemes are equal to each other and given by*

$$\begin{aligned} \Delta_{ml}^{L}(b,x) = \max &\left\{ x, \frac{b(M_t-k)(M_r-k)(1-\eta_k^L)}{(M_t-k)(M_r-k)-b\eta_k^L} + x\frac{b(1-\eta_k)\eta_k^{L-1}}{(M_t-k)(M_r-k)-b\eta_k^L} \right\}, \\ \text{for} \quad & (M^*-k-1)(M_*-k-1) \leq b < (M^*-k)(M_*-k), \\ \text{for } & k = 0, ..., M_*-1. \end{aligned} \tag{5.10}$$

, where we define,

$$\eta_k \triangleq \frac{b-(M_t-k-1)(M_r-k-1)}{M_t-k+M_r-k-1}. \tag{5.11}$$

and

$$\begin{aligned} \Delta_{ml}^{L} = \max &\left\{ x, \frac{b(L-1)(b-M^*M_*)(x+M^*M_*)+b(x+M^*M_*+(L-1)M^*M_*)(M^*+M_*-1)}{b(L-1)(b-M^*M_*)+bL(M^*+M_*-1)+(M^*+M_*-1)^2} \right\}, \\ &\text{for } b \geq M^*M_*. \end{aligned} \tag{5.12}$$

Proof. See Appendix 5.H for multi-layer without binning and 5.J for multi-layer NBJD. □

Corollary 1. *The distortion exponent $\Delta^L(b,x)$ in the limit of infinite layers is given by*

$$\Delta_{ml}^{\infty}(b,x) = \begin{cases} x & \text{if } b \leq x, \\ b & \text{if } x < b < M_tM_r, \\ M^*M_* + x\left(\frac{b-M_*M^*}{b-(M^*-1)(M_*-1)}\right) & \text{if } x < M_tM_r \leq b. \end{cases} \tag{5.13}$$

Theorem 11. *By considering a more general power allocation, we achieve a better distortion exponent* $\Delta^L_{ml2}(b,x)$ *given by*

$$\Delta^L_{ml2}(b,x) = \max\left\{x, \frac{b(M_t-k)(M_r-k)(1-\eta_k^L)}{(M_t-k)(M_r-k)-b\eta_k^L} + x\frac{b(1-\eta_k)\eta_k^{L-1}}{(M_t-k)(M_r-k)-b\eta_k^L}\right\},$$
$$\text{for} \quad (M^*-k-1)(M_*-k-1) \leq b < (M^*-k)(M_*-k),$$
$$\text{for } k = 0, ..., M_*-1. \tag{5.14}$$

, *where we define,*

$$\eta_k \triangleq \frac{b-(M_t-k-1)(M_r-k-1)}{M_t-k+M_r-k-1}. \tag{5.15}$$

In the limit of infinite layers, the distortion exponent is given by

$$\Delta^L_{ml2} = \max\left\{x, \frac{b(L-1)(b-M^*M_*)(x+M^*M_*)+b(x+M^*M_*+(L-1)M^*M_*)(M^*+M_*-1)}{b(L-1)(b-M^*M_*)+bL(M^*+M_*-1)+(M^*+M_*-1)^2}\right\},$$
$$\text{for } b \geq M^*M_*. \tag{5.16}$$

Proof. See Appendix 5.H for multi-layer without binning and 5.J for multi-layer NBJD. □

Corollary 2. *The distortion exponent* $\Delta^L_{ml2}(b,x)$ *in the limit of infinite layers is given by*

$$\Delta^\infty_{ml2}(b,x) = \begin{cases} b(l+1) & \text{if } b \in [\frac{(M^*-l-1)(M^*-l-1)+x}{l+1}, \frac{(M^*-l)(M^*-l)+x}{l+1}), \\ (M^*-l)(M_*-l) & \text{if } b \in [\frac{(M^*-l-1)(M^*-l-1)+x}{l+1}, \frac{(M^*-l)(M^*-l)+x}{l}) \\ +x\left(\frac{b-(M_*-l)(M^*-l)}{b-((M^*-l)-1)((M_*-l)-1)}\right) & \text{for } k = 0, ..., M_*-1. \end{cases} \tag{5.17}$$

5.5 Comments on the results

In this section we provide a graphical representation of the distortion exponents to give a better understanding of the distortion exponent results for the described schemes in MIMO, MISO and SISO systems. We separate the section in three subsections, one concerning to all schemes for MIMO and MISO systems with different quality of side information x.

Then, next subsection is devoted to multilayer schemes and we show the behavior of the schemes for finite L and its convergence to the given continuum of schemes.

In the last subsection, we particularize to the SISO system.

5.5.1 MIMO and MISO results

Figure 5.1 and Figure 5.2 show the distortion exponent in function of b for a particular x. We can do the following observations:

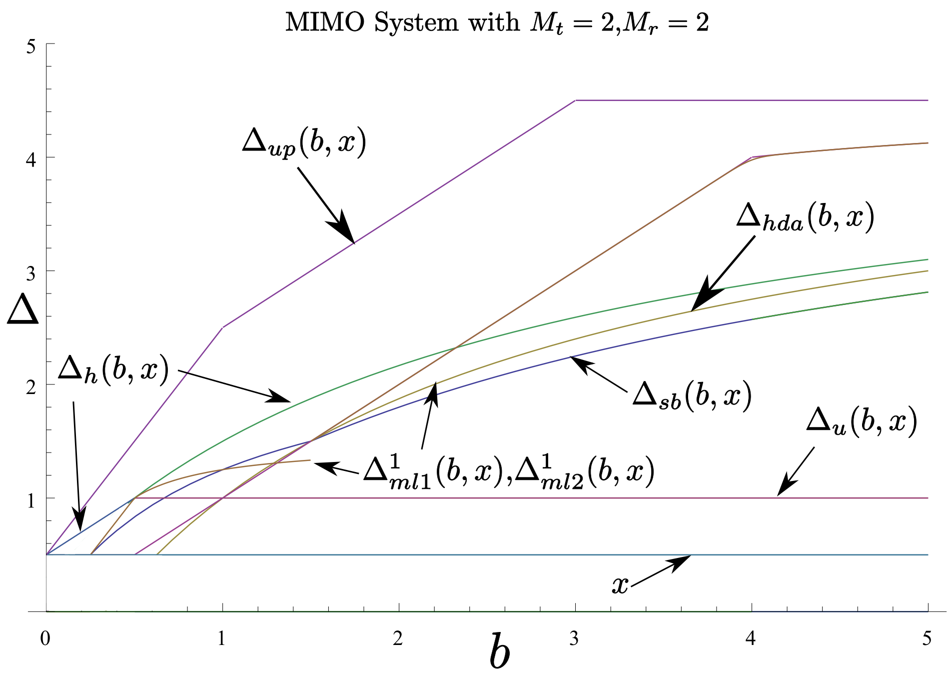

Figure 5.1: Distortion exponents in function of b for $x = 0.5$ for a MIMO 2×2 system.

- Uncoded distortion exponent $\Delta_u(b)$ performs poorly for large b as remains constant for $M_* b \geq 1$. It behaves as if there is no transmission for $M_* b \leq 1$.
- The distortion exponents $\Delta_{sb}(b)$ achieved by separate source-channel coding, separation without binning and single layer NBJD schemes are equal to each other.
- The distortion exponents of the layered NBJD in the limit of infinite layers $\Delta^{\infty}_{ml2}(b)$ is the best distortion exponent obtained in the large bandwidth region. Further details on multilayer schemes behavior is provided in next subsection.
- HDA scheme, $\Delta_{hda}(b)$, uniformly outperforms uncoded and single layer schemes $\Delta_{sb}(b)$ and is outperformed by NBJD-analog hybrid scheme for $b > 1$.

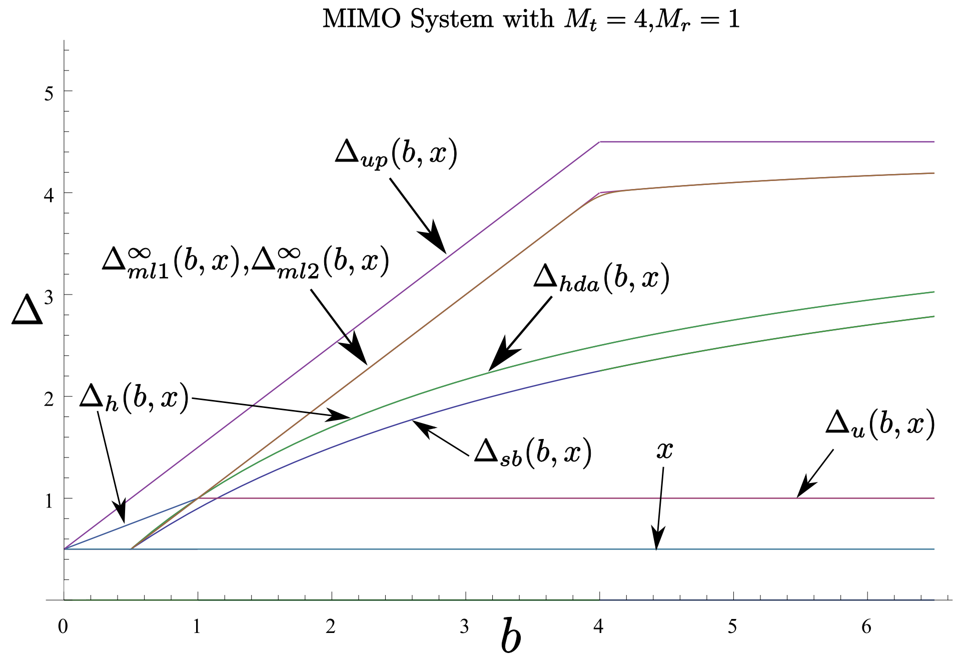

Figure 5.2: Distortion exponents in function of b for $x = 0.5$ for a MISO 4×1 system.

- The NBJD-analog hybrid scheme achieves the second best distortion exponent for $M_* b \geq 1$.

5.5.2 MIMO multi-layer results

Figure 5.3 shows the distortion exponent behavior of multilayer systems when there is no side information. Figure 5.4 shows the distortion exponent when time varying side information is available at the receiver. We can do the following observations:

- The distortion exponent of layered NBJD $\Delta^L_{ml2}(b)$ and $\Delta^L_{ml2}(b)$ with finite number of layers is not continuous for the proposed power allocations. On the contrary, $\Delta^{\infty}_{ml1}(b)$ and $\Delta^{\infty}_{ml2}(b)$ become continuous in the in the limit of infinite layers.
- Note that intuitively, $\Delta^{\infty}_{ml2}(b)$ behaves as shifted and scaled versions of $\Delta^{\infty}_{ml1}(b)$.
- When no side information is present, $\Delta^{\infty}_{ml1}(b)$ meets the upper bound $\Delta_{up}(b)$ for $b \geq M_t M_r$ and is hence optimal. $\Delta^{\infty}_{ml2}(b)$ is optimal for $b \geq M_t M_r$ and $b \leq \frac{(M^* - M_* + 1)}{M_*}$.
- The distortion exponents of the layered NBJD in the limit of infinite layers $\Delta^{\infty}_{ml2}(b)$ is the best distortion exponent obtained in the large bandwidth region.
- An important remark to recall is that for separation the distortion exponent is maximized by ignoring the side information when encoding, (i.e., no binning), as the scheme suffers from outage in both the channel and the side information.

5.6 SISO results

In this section we provide the results particularized to single antenna systems.

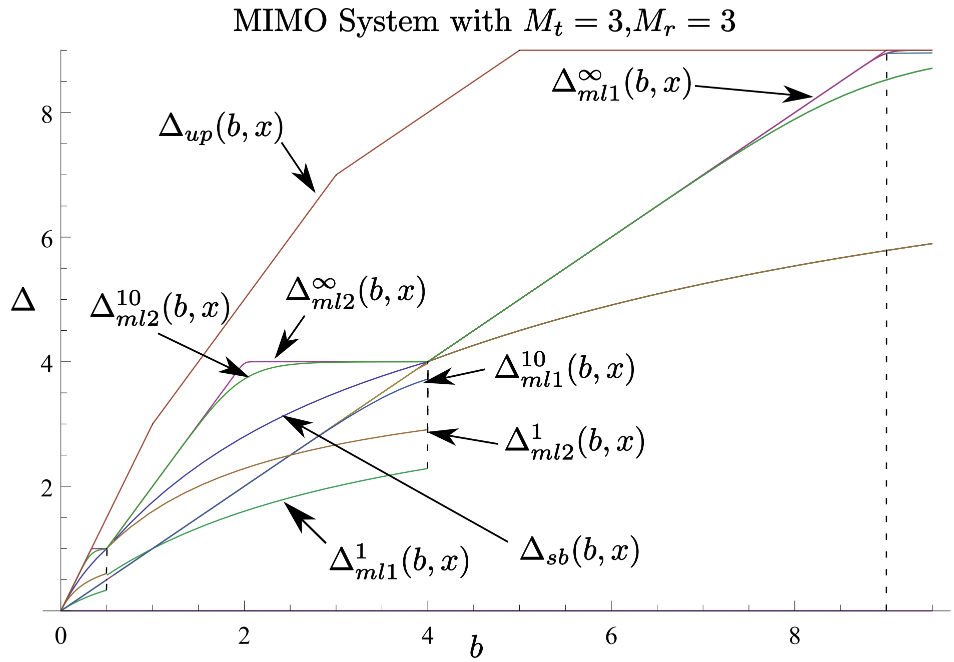

Figure 5.3: Distortion exponents of the multi-layer schemes in function of b for $x = 0$ in a MIMO 3×3 system.

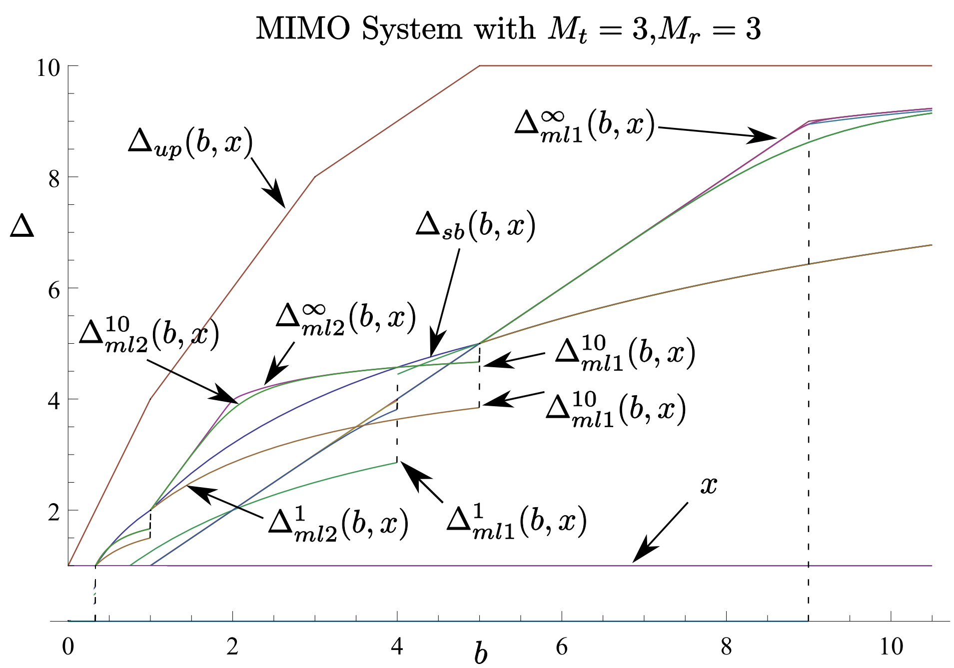

Figure 5.4: Distortion exponents of the multi-layer schemes in function of b for $x = 1$ in a MIMO 3×3 system.

Theorem 12. *The distortion exponent* $\Delta_{up}(b,x)$ *is upper bounded by*

$$\Delta_{up}(b,x) = \begin{cases} x + b & \text{if } b < 1, \\ 1 + x & \text{if } b \geq 1. \end{cases} \tag{5.18}$$

Theorem 13. *The distortion exponents for the single layer separate source-channel coding, separation without binning, and single layer NBJD scheme, for a SISO sytem are equal to each other and given by,*

$$\Delta^1(b,x) = \begin{cases} x & \text{if } b < x, \\ \frac{b(1+x)}{1+b} & \text{if } b \geq x. \end{cases}$$

Theorem 14. *The distortion exponent of uncoded transmission is a SISO system is characterized by*

$$\Delta_u(b,x) = \begin{cases} x & \text{if } b < 1, \\ \max\{1,x\} & \text{if } b \geq 1. \end{cases} \tag{5.19}$$

Theorem 15. *The distortion exponent $\Delta_h(b,x)$ of the NBJD-analog hybrid scheme is characterized, for a SISO sytem, by*

$$\Delta_h(b,x) = \begin{cases} b + x(1-b) & \text{if } b < 1, \\ 1 + x\left(1 - \frac{1}{b}\right) & \text{if } b \geq 1. \end{cases} \tag{5.20}$$

Theorem 16. *The distortion exponent $\Delta_{hda}(b,x)$ of the SISO HDA scheme is characterized by*

$$\Delta_{hda}(b,x) = 1 + x\left(1 - \frac{1}{b}\right) \qquad \text{if } b \geq 1. \tag{5.21}$$

Theorem 17. *The distortion exponent of L layer separate source-channel coding scheme without binning and for multi-layer NBJD schemes are equal to each other in SISO systems and given by*

$$\Delta^L(b,x) = \begin{cases} x & \text{if } b \leq x, \\ \frac{b+b^L x - b^{L+1}(1+x)}{1-b^{L+1}} & \text{if } x < b, b < 1, \\ \frac{b(1-(L-1)x + b(L-1)(1+x))}{1+b+b^2(L-1)} & \text{if } x < b, b \geq 1. \end{cases} \tag{5.22}$$

Corollary 3. *The distortion exponent $\Delta^L(b,x)$ in the limit of infinite layers for a SISO system is given by*

$$\Delta^\infty(b,x) = \begin{cases} x & \text{if } b \leq x, \\ b & \text{if } x < b, b < 1, \\ 1 + x\left(1 - \frac{1}{b}\right) & \text{if } x < b, b \geq 1. \end{cases} \tag{5.23}$$

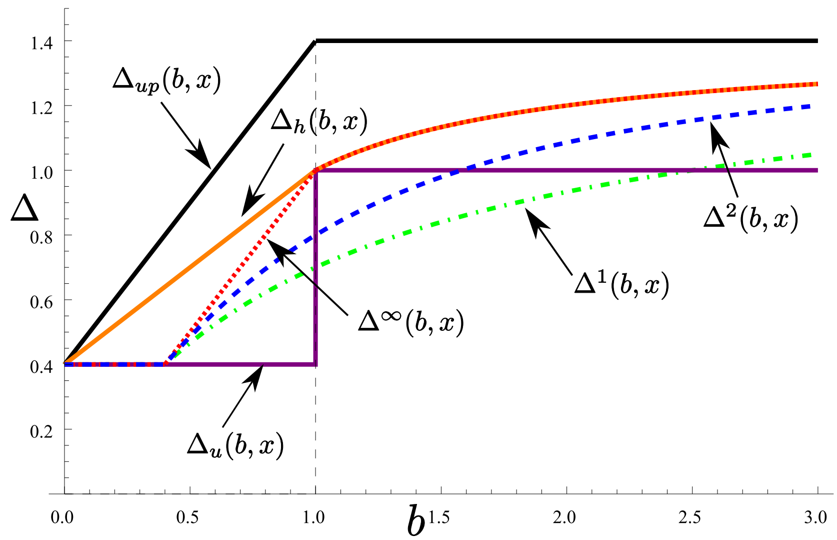

Figure 5.5: Distortion exponents in function of b for $x = 0.4$ for a SISO system.

5.6.1 Comments on SISO results

Results regarding the distortion exponent can be summarized as follows (See Figure 5.5 for illustration):

- Uncoded distortion exponent $\Delta_u(b)$ performs as no transmission for $b < 1$ and achieves $\Delta = 1$ for $b \geq 1$.
- The distortion exponents $\Delta^1(b)$ achieved by separate source-channel coding, separation without binning and single layer NBJD schemes are equal to each other.
- The distortion exponent of layered NBJD with two layers, $\Delta^2(b)$, notably improves upon single layer NBJD, $\Delta^1(b)$.
- Note that hybrid- NBJD reduces to pure uncoded at $b = 1$. On the contrary, HDA achieves the same distortion exponent by a technique that involves decoding.
- The distortion exponents of the layered NBJD in the limit of infinite layers $\Delta^\infty(b)$, the NBJD-analog hybrid scheme and HDA scheme $\Delta_h(b)$ are equal to each other for $b > 1$ and uniformly outperform uncoded and single layer schemes.

Appendix 5.A Proof for distortion exponent: lower bound

To obtain the distortion exponent lower bound, we notice that the capacity can be lower bounded by

$$C(\mathbf{H}) \leq \log\det\left(\mathbf{I} + \rho_c \mathbf{H}\mathbf{H}^\dagger\right), \tag{5.24}$$

as $\mathbf{I}M_* - \mathbf{Q} \succcurlyeq 0$ if $\text{Tr}\{\mathbf{Q}\} \leq M_*$ and $\log\det(\cdot)$ is an increasing function in the cone of positive-definite Hermitian matrices. Then the end-to-end distortion can be lower-bounded as

$$D_{low}(\mathbf{H}, \gamma^2) \geq \left\{ \frac{\det\left(\mathbf{I} + \rho\mathbf{H}\mathbf{H}^\dagger\right)^{-b}}{1 + \rho_s\gamma^2} \right\} \geq \left\{ \frac{\prod_{j=1}^{M_*}\left(1 + \rho_c\lambda_i\right)^{-b}}{1 + \rho_s\gamma^2} \right\}, \tag{5.25}$$

where $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_{M_*}$ are the ordered eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$.

Let $\lambda_i = \rho^{-\alpha_i}$, with $\alpha_1 \geq ... \geq \alpha_{M_*} \geq 0$ and $\gamma^2 = \rho_s^{-\beta} = \rho^{-x\beta}$. Then we have $(1 + \rho\lambda_i)^{-b} \doteq \rho^{-b(1-\alpha_i)^+}$ and $\left(1 + \rho_s\gamma^2\right)^{-1} \doteq \rho^{-x(1-\beta)^+}$. The joint probability density function (pdf) of $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_{M_*}]$, is given by

$$p(\boldsymbol{\alpha}) = K_{M_t,M_r}^{-1} (\log\rho)^{M_*} \prod_{i=1}^{M_*} \rho^{-(M^* - M_* + 1)\alpha_i} \cdot \left[\prod_{i<j} (\rho^{\alpha_i} - \rho^{\alpha_j})^2\right] \exp\left(-\sum_{i=1}^{M_*} \rho^{\alpha_i}\right), \tag{5.26}$$

where K_{M_t,M_r}^{-1} is a normalizing constant. For side information state, β is distributed as

$$p(\beta) = K_{1,1}^{-1} \log\rho^x \rho^{x\beta} \exp(-\rho^{x\beta}), \tag{5.27}$$

similarly. We can write using the independency of $\boldsymbol{\alpha}$ and β and following the arguments as in [18], as

$$
\begin{aligned}
&E\left[D_{low}(\mathbf{H},\gamma)\right] \\
&\doteq \int_{\mathbb{R}^{(M_*+1)+}} \frac{\prod_{i=1}^{M_*}(1+\rho\lambda_i)^{-b}}{1+\rho_s\gamma^2} p(\boldsymbol{\alpha})p(\beta)d\boldsymbol{\alpha}d\beta \\
&\doteq \int_{\mathbb{R}^{M_*+}} \prod_{i=1}^{M_*}(1+\rho\lambda_i)^{-b}\, p(\boldsymbol{\alpha})d\boldsymbol{\alpha} \\
&\quad \cdot \int_{\mathbb{R}^+} \left(1+\rho_s\gamma^2\right)^{-1} p(\beta)d\beta \\
&\doteq \int_{\mathbb{R}^{M_*+}} \prod_{i=1}^{M_*} \rho^{-b(1-\alpha_i)^+} \prod_{i=1}^{M_*} \rho^{-(2i-1+M^*-M_*)\alpha_i} d\boldsymbol{\alpha} \\
&\quad \cdot \int_{\mathbb{R}^+} \rho^{-x(1-\beta)^+} \rho^{x\beta} d\boldsymbol{\beta} \\
&\doteq \int_{\mathbb{R}^{M_*+}} \prod_{i=1}^{M_*} \rho^{-[b(1-\alpha_i)^+ + (2i-1+M^*-M_*)\alpha_i]} d\boldsymbol{\alpha} \\
&\quad \cdot \int_{\mathbb{R}^+} \rho^{-x(1-\beta)^+ + x\beta} d\beta \\
&\doteq \int_{\mathbb{R}^{M_*+}} \prod_{i=1}^{M_*} \rho^{-[S(\boldsymbol{\alpha}) + \sum_{i=1}^{M_*} b(1-\alpha_i)^+]} d\boldsymbol{\alpha} \\
&\quad \cdot \int_{\mathbb{R}^+} \rho^{-x[(1-\beta)^+ + \beta]} d\beta \\
&\doteq \rho^{-\Delta_1}\rho^{\Delta_2} \\
&\doteq \rho^{-\Delta_{up}}.
\end{aligned} \tag{5.28}
$$

where we define $p_\alpha(\boldsymbol{\alpha}) \doteq \rho^{-S(\boldsymbol{\alpha})}$

$$
S(\boldsymbol{\alpha}) \triangleq \sum_{i=1}^{M_*}(2i-1+M^*-M_*)\alpha_i \tag{5.29}
$$

The distortion exponent can be separated in two independent components corresponding to the channel state and the quality of the side information respectly. Δ_1 is a well known problem solved in proof of Theorem 4 in [18], and given by

$$
\Delta_1 = \sum_{i=1}^{M_*} \min\left\{b, 2i-1+M^*-M_*\right\}. \tag{5.30}
$$

Component corresponding to the side information quality is the solution to

$$
\begin{aligned}
\Delta_2 = \inf\, & x[(1-\beta)^+ + \beta] \\
\text{s.t. } & \beta \in \mathbb{R}^+.
\end{aligned} \tag{5.31}
$$

that solves for $\Delta_2 = x$ for any $\beta \in [0,1]$. That completes the proof.

Appendix 5.B Proof for distortion exponent: separate source and channel coding

The outage set $\mathcal{O}_{sb}$ from equation (3.10) can be decomposed in two non intersecting sets,

$$\mathcal{O}_{sb} = \mathcal{E}_1 \oplus \mathcal{E}_2, \tag{5.32}$$

where $\mathcal{E}_1$ is the set of $\mathbf{H}$ such that there is an outage in the channel transmission and $\mathcal{E}_2$ is the set for which there is an outage in the source decoder when the channel codeword has been correctly decoded, i.e

$$\begin{aligned} &\mathcal{E}_1 = \{\mathbf{H} : R_c > I(\mathbf{U}_i; \mathbf{V}_i)\}, \\ &\mathcal{E}_2 = \{(\mathbf{H}, \gamma) : bR_c < I(X; W|Y), R_c \leq I(\mathbf{U}_i; \mathbf{V}_i)\}. \end{aligned} \tag{5.33}$$

The exponential behavior for the outage set event $\mathcal{E}_1$, $\tilde{\mathcal{E}}_1$, is obtained by finding the equivalent outage probability set

$$\begin{aligned} &Pr\{\mathcal{E}_1\} \\ &= Pr\{R_c > I(\mathcal{U}_i; \mathcal{V}_i)\} \\ &= Pr\{r_c \log\rho > \log\det(\mathbf{I} + \rho\mathbf{H}^\dagger\mathbf{H})\} \\ &= Pr\{\log\rho^{r_c} > \log\prod_{i=1}^{M_*}(1+\rho^{1-\alpha_i}), \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0\} \\ &\doteq Pr\{r_c > \sum_{i=1}^{M_*}(1-\alpha_i)^+, \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0\}, \end{aligned} \tag{5.34}$$

where we have used the change of variables similarly to (5.26).

The outage event $\mathcal{E}_2$ is exponentially equivalent to $\tilde{\mathcal{E}}_2$

$$\begin{aligned} &Pr\{\mathcal{E}_2\} \\ &= Pr\{I(X; W|Y) > bR_c|\mathcal{E}_1^c\} \\ &= Pr\left\{\log\left(1 + \frac{1}{\sigma_Q^2(\gamma^2\rho+1)}\right) > bR_c|\mathcal{E}_1^c\right\} \\ &= Pr\left\{1 + \frac{\rho^{r_s+br_c}-1}{\rho^{x(1-\beta)}+1} > \rho^{br_c}|\mathcal{E}_1^c\right\} \\ &\doteq Pr\left\{1 + (\rho^{r_s+br_c})\rho^{-x(1-\beta)^+} > \rho^{br_c}|\mathcal{E}_1^c\right\} \\ &\doteq Pr\left\{\rho^{(r_s+br_c-x(1-\beta)^+)^+} > \rho^{br_c}|\mathcal{E}_1^c\right\} \\ &\doteq Pr\left\{r_s > x(1-\beta)^+|\mathcal{E}_1^c\right\}, \end{aligned} \tag{5.35}$$

where $\gamma^2 = \rho^{-\beta}$.

The exponential expected distortion can be computed as

$$
\begin{aligned}
&ED_{sb}(R_s, R_c) \\
&= \iint_{\mathcal{E}^c} D_d(R_s + bR_c, \gamma) p_h(\mathbf{H}) p_\gamma(\gamma) d\mathbf{H} d\gamma \\
&+ \iint_{\mathcal{E}_1} D_d(0, \gamma) p_h(\mathbf{H}) p_\gamma(\gamma) d\mathbf{H} d\gamma \\
&+ \iint_{\mathcal{E}_2} D_d(0, \gamma) p_h(\mathbf{H}) p_\gamma(\gamma) d\mathbf{H} d\gamma. \\
&\doteq \int_{\mathcal{E}_\alpha^c} p_\alpha(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \int_{\mathcal{E}_\beta^c} D_d(r_s + br_c, \beta) p_\beta(\beta) d\beta \\
&+ \int_{\mathcal{E}_\alpha^c} p_\alpha(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \int_{\mathcal{E}_\beta} D_d(0, \beta) p_\beta(\beta) d\beta \\
&+ \int_{\mathcal{E}_\alpha} p_\alpha(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \int_{\mathbb{R}^+} D_d(0, \beta) p_\beta(\beta) d\beta
\end{aligned}
$$

Where we have used that the equivalent outage sets can be separated for α and β as:

$$
\begin{aligned}
\tilde{\mathcal{E}}_1 &= \tilde{\mathcal{E}}_\alpha \oplus \mathbb{R}^+ = \left\{ \boldsymbol{\alpha} : \, r_c > \sum_{i=1}^{M_*} (1-\alpha_i)^+, \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0 \right\} \oplus \mathbb{R}^+, \\
\tilde{\mathcal{E}}_2 &= \tilde{\mathcal{E}}_\alpha^c \oplus \tilde{\mathcal{E}}_\beta \\
&= \left\{ \boldsymbol{\alpha} : r_c \leq \sum_{i=1}^{M_*} (1-\alpha_i)^+, \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0 \right\} \oplus \left\{ \beta : \, r_s > x(1-\beta)^+ \right\}, \quad (5.36)
\end{aligned}
$$

Note that the integrals dependant on the channel realization correspond to the outage probability of a MIMO channel, P_{out}, that is known to be exponential characterized by the dmt curve $d^*(r)$ of the system:

$$
P_{out} \doteq \int_{\mathcal{E}_\alpha} p_\alpha(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \doteq \int_{\mathcal{E}_\alpha} \rho^{-S(\boldsymbol{\alpha})} d\boldsymbol{\alpha} \doteq \rho^{-d^*(r_c)} \quad (5.37)
$$

where for an $M_t \times M_r$ MIMO block fading channel, the optimal tradeoff curve $d^*(r_c)$ is given by the piecewise-linear function connecting the points $(k, d^*(k))$,$k = 0, 1, ..., M_*$, where

$$
d^*(k) = (M_t - k)(M_r - k). \quad (5.38)
$$

Following the reasonings in [10], the remaining integral in the first term is exponentially equivalent

$$
\int_{\tilde{\mathcal{E}}_\beta^c} \rho^{-\max\{x(1-\beta)^+, r_s + br_c\}} \rho^{-\beta} d\beta \doteq \rho^{-\Gamma_{s1}(r_s, r_c)} \quad (5.39)
$$

where using the Laplace's method [18], we have

$$
\begin{aligned}
\Gamma_{s1}(r_s, r_c) = &\inf \max\{x(1-\beta)^+, r_s + br_c\} + x\beta \\
&\text{s.t. } \, r_s \leq x(1-\beta)^+. \quad (5.40)
\end{aligned}
$$

Similarly, the second integral can be expressed as

$$\int_{\mathcal{E}_\beta} \rho^{-(x(1-\beta)^+ + x\beta)} d\beta \doteq \rho^{-\Gamma_{s2}(r_s)}. \tag{5.41}$$

where

$$\begin{aligned} \Gamma_{s2}(r_s) = \inf\ & x[(1-\beta)^+ + \beta] \\ \text{s.t.}\ & r_s > x(1-\beta)^+. \end{aligned} \tag{5.42}$$

The third integral can be simplified likewise and

$$\begin{aligned} \Gamma_{s3} = \inf\ & x[(1-\beta)^+ + \beta] \\ \text{s.t.}\ & \beta \geq 0. \end{aligned} \tag{5.43}$$

Thus, the distortion exponent can be calculated as

$$\begin{aligned} & ED_{sb}(r_s \log\rho, r_c \log\rho) \\ & \doteq (1-P_{out})\rho^{-\Gamma_{s1}(r_s,r_c)} + (1-P_{out})\rho^{-\Gamma_{s2}(r_s)} + P_{out}\rho^{-\Gamma_{s3}} \\ & \doteq \rho^{-\Gamma_{s1}(r_s,r_c)} + \rho^{-\Gamma_{s2}(r_s)} + \rho^{-(d^*(r_c)+\Gamma_{s3})} \\ & \doteq \rho^{-\min\{\Gamma_{s1}(r_s,r_c), \Gamma_{s2}(r_s), d^*(r_c)+\Gamma_{s3}\}} \\ & \doteq \rho^{-\Delta_{sb}(r_s,r_c,b)}, \end{aligned} \tag{5.44}$$

The dominant exponent for the separate source and channel coding scheme can then be obtained as the solution of the following optimization problem

$$\Delta_{sb}(r_s, r_c, b) = \min\{\Gamma_{s1}(r_s, r_c), \Gamma_{s2}(r_s), d^*(r_c) + \Gamma_{s3}\} \tag{5.45}$$

Note that from (5.40), (5.42) and (5.43), one can conclude that considering $\beta > 1$ can only increase the expressions to minimize. Consequently, we restrict to $0 \leq \beta \leq 1$.

First we solve (5.40). From the constraint we see that β has to satisfy $\beta \leq 1 - \frac{r_c}{x}$. Hence, the infimum is obtained for $\beta = 0$ if $r_s \leq x$ and given by

$$\Gamma_{s1}(r_c, r_s) = \max\{x, r_s + br_c\} \text{ if } r_s \leq x. \tag{5.46}$$

Next we solve (5.42). Considering the restrictions, $r_s > x(1-\beta)^+$, the infimum is obtained for $\beta = (1 - \frac{r_s}{x} + epsilon)^+$ with $\epsilon > 0$. Note that if $r_s = 0$, there exists no β satisfying the constraint and the integral does not exist. Hence, letting $\epsilon \to \infty$, the infimum is given by

$$\Gamma_{s2}(r_s) = x \text{ for } r_s > 0. \tag{5.47}$$

Next, problem (5.43) was solved in proof 5.A and is given, for any $\beta \in [0, 1]$ as

$$\Gamma_{s3} = x. \tag{5.48}$$

Bringing all these together, we obtain from (5.45):

$$\Delta_{sb}(r_s, r_c, b) = \begin{cases} x & \text{if } r_s > 0, \\ \min\{x + d^*(r_c), br_c\} & r_c \leq 1,\ r_s = 0. \end{cases} \tag{5.49}$$

We now design the rate such that the minimum distortion exponent is maximized, i.e.

$$\Delta_{sb}(b) \triangleq \max_{r_s, r_c} \Delta_{sb}(r_s, r_c, b). \tag{5.50}$$

It can be seen from (5.49) that by letting $r_s = 0$, i.e. by transmitting without binning, we can improve the distortion exponent. In this case, the infimum is achieved equating the two terms inside $\min\{\cdot\}$: $x + d^*(r_c) = br_c$. In order to find the explicit expression. For $r_c = j$, where $j = 1, ..., M_* - 1$ are integers, the distortion exponent has solutions at $b = \frac{d_{ds}(j)-x}{j}$. Note that if $x \geq bd(M_*)$, the intersection is always at $br_c = x$. Hence, Δ_{sb} interpolates$(\frac{d_{ds}(j)-x}{j}, d(j) + x)$ points if $x < M_* b$ and is constant x if $x \geq M_* b$. For $b \in [(M^* - 1)(M_* - 1), M^* M_*)$, the dmt curve intersects as

$$x + M^* M_* - (M^* + M_* - 1) r_{sb} = br_{sb}, \tag{5.51}$$

and

$$\Delta_{sb}(b, x) = b \frac{M^* M_* + x}{M^* + M_* - 1 + b}, \tag{5.52}$$

for

$$r_{sb} = \frac{M^* M_* + x}{M^* + M_* - 1 + b}. \tag{5.53}$$

For $b \in [\frac{(M^*-k-1)(M_*-k-1)+x}{k}, \frac{(M^*-k)(M_*-k)+x}{k})$ we can consider the intersection of

$$(M^* - (k+1))(M_* - (k+1)) - (r_{sb} - (k+1))(1 - 2(k+1) + M^* + M_*) = br_{sb} \tag{5.54}$$

for $k = 1, ..., M_* - 1$ and hence,

$$r_{sb} = \frac{M^* M_* + x - k(k+1)}{b + M^* + M_* - 1 - 2k}. \tag{5.55}$$

Then, the distortion exponent is given by

$$\Delta_{sb}(b, x) = b \frac{M^* M_* + x - k(k+1)}{b + M^* + M_* - 1 - 2k}. \tag{5.56}$$

Finally, note that we have that $\Delta_{sb}(b, x) \leq x$ if $b \leq \frac{x(M^*+M_*-1)}{M^* M_*}$ and hence the distortion exponent $\Delta_{sb}(b, x) = x$ for this range. Notice that this range includes $b \leq \frac{x}{M_*}$, which completes the proof.

Remark 3. *The optimal exponent is obtained for $r_s = 0$ which indicates that the number of auxiliary codewords in each bin has to be one per bin. However, note that at finite SNR regime, separation with binning outperforms the scheme with no binning.*

Appendix 5.C Proof for distortion exponent: uncoded

Using the same reasonings as in the proof for Theorem 12, with $\mu_i \doteq \rho^{-\alpha_i}$ we have

$$\begin{aligned} \frac{1}{M_*} \sum_{i=1}^{M_*} \frac{1}{1 + \rho\mu_i + \rho_s \gamma^2} &\doteq \frac{1}{M_*} \sum_{i=1}^{M_*} \frac{1}{1 + \rho^{1-\alpha_i} + \rho^{x(1-\beta}} \\ &\doteq \frac{1}{M_*} \sum_{i=1}^{M_*} \rho^{-\max\{(1-\alpha_i)^+, x(1-\beta)^+\}} \\ &\rho^{-\min_{i=1,...,M_*} \max\{(1-\alpha_i)^+, x(1-\beta)^+\}} \end{aligned}$$

and hence, the exponential equivalent can be found for $M_*b < 1$ as

$$
\begin{aligned}
ED_u &\doteq (1 - M^*b) \int_{\mathbb{R}} \rho^{-x(1-\beta)^+} \rho^{-x\beta} d\beta \\
&+ M_*b \iint_{\mathbb{R}^{M_*+1}} \rho^{-\min_{i=1,\dots,M_*} \max\{(1-\alpha_i)^+, x(1-\beta)^+\}} \\
&\cdot \ \rho^{-(S(\boldsymbol{\alpha})+\beta)} d\boldsymbol{\alpha} d\beta \\
&\doteq (1 - M^*b)\rho^{-\Delta_{u\mathcal{N}}} + M_*b\rho^{-\Delta_{u\mathcal{R}}} \\
&\doteq \rho^{-\Delta_u},
\end{aligned} \tag{5.57}
$$

Note that the distortion for the source samples with uncoded transmission, i.e. $X_1^{nM_*}$, behave exponentially as $\rho^{-\Delta_{u\mathcal{R}}}$ and source samples without uncoded transmission, $X_{M_*n+1}^m$, behave as $\rho^{-\Delta_{u\mathcal{N}}}$. If $M_*b \geq 1$, all samples can be transmitted though the channel and equation (5.57) simplifies to

$$
\begin{aligned}
ED_u &\doteq \iint_{\mathbb{R}^{M_*+1}} \rho^{-\min_{i=1,\dots,M_*} \max\{(1-\alpha_i)^+, x(1-\beta)^+\}} \\
&\cdot \ \rho^{-(S(\boldsymbol{\alpha})+\beta)} d\boldsymbol{\alpha} d\beta \\
&\doteq \rho^{-\Delta_{u\mathcal{R}}},
\end{aligned} \tag{5.58}
$$

Note that increasing the channel input power does not modify the diversity of the system.

Then, the dominant exponent is the solution to the optimization problem

$$
\Delta_u = \begin{cases} \min\{\Delta_{u\mathcal{R}}, \Delta_{u\mathcal{N}}\} & \text{if } M_*b < 1, \\ \Delta_{u\mathcal{R}} & \text{if } M_*b \geq 1, \end{cases} \tag{5.59}
$$

where

$$
\begin{aligned}
\Delta_{u\mathcal{R}} \ &= \ \inf \min_{i=1,\dots,M_*} \max\{(1-\alpha_i)^+, x(1-\beta)^+\} + S(\boldsymbol{\alpha}) + x\beta \\
&\text{s.t. } \alpha_1 \geq \dots \geq \alpha_{M_*} \geq 0, \qquad \beta \geq 0,
\end{aligned} \tag{5.60}
$$

and

$$
\begin{aligned}
\Delta_{u\mathcal{N}} \ &= \ \inf x[(1-\beta)^+ + \beta] + S(\boldsymbol{\alpha}) \\
&\text{s.t. } \alpha_1 \geq \dots \geq \alpha_{M_*} \geq 0, \qquad \beta \geq 0,
\end{aligned} \tag{5.61}
$$

Next we solve (5.60). Picking $\beta = 0$ minimizes the expression for any $\boldsymbol{\alpha}$. We observe that $S(\boldsymbol{\alpha}) \geq 1$ and hence by picking $\alpha_i = 0$ the expression is minimized. The infimum is then given by $\Delta_{u\mathcal{R}} = \max\{1, x\}$. Problem (5.61) achieves the minimum $\Delta_{u\mathcal{N}} = x$ for $\boldsymbol{\alpha} = 0$ and $\beta = 0$.

Then (5.59), the distortion exponent for uncoded transmission is given by

$$
\Delta_u = \begin{cases} x & \text{if } M_*b < 1, \\ \max\{1, x\} & \text{if } M_*b \geq 1. \end{cases} \tag{5.62}
$$

Appendix 5.D Proof for distortion exponent : NBJD

The outage probability of the joint source-channel scheme is characterized by

$$Pr\{\mathcal{O}_j\} = Pr\{\ I(X;W|Y) > bI(\mathbf{U};\mathbf{V})\ \}. \tag{5.63}$$

By operating as in [10] we obtain

$$Pr\{\mathcal{O}_j\} = Pr\left\{1 + \frac{\rho^{r_j}-1}{\gamma^2\rho^x+1} > \prod_{i=1}^{M_*}(1+\rho\lambda_i)^b\right\}. \tag{5.64}$$

Now we apply the change of variables $\lambda_i = \rho^{-\alpha_i}, \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0$ and $\gamma^2 = \rho^{-x\beta}$ distributed as in (5.26). Then,

$$\begin{aligned}
&Pr\{1 + \frac{\rho^{r_j}-1}{\rho^{x(1-\beta)}+1} > \prod_{i=1}^{M_*}(1+\rho^{1-\alpha_i})^b\} \\
&\doteq Pr\{1 + (\rho^{r_j}-1)\rho^{-x(1-\beta)^+} > \rho^{b\sum_{i=1}^{M_*}(1-\alpha_i)^+}\} \\
&\doteq Pr\{1 + \rho^{r_j - x(1-\beta)^+} > \rho^{b\sum_{i=1}^{M_*}(1-\alpha_i)^+}\} \\
&\doteq Pr\{\rho^{(r_j - x(1-\beta)^+)^+} > \rho^{b\sum_{i=1}^{M_*}(1-\alpha_i)^+}\} \\
&\doteq Pr\left\{(r_j - x(1-\beta)^+)^+ > b\sum_{i=1}^{M_*}(1-\alpha_i)^+\right\}.
\end{aligned} \tag{5.65}$$

We define the exponentially equivalent outage event set $\tilde{\mathcal{O}}_j$ as

$$\tilde{\mathcal{O}}_j = \left\{(\boldsymbol{\alpha},\beta)\ :\ (r - x(1-\beta)^+)^+ > b\sum_{i=1}^{M_*}(1-\alpha_i)^+, \beta \geq 0\, \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0\right\}. \tag{5.66}$$

where we can restrict to $\alpha \geq 0$ and $\beta \geq 0$ with the reasoning used to obtain the sets in (5.36). The expected end-to-end distortion can be expressed as

$$\begin{aligned}
ED_j(r_j \log\rho) &\doteq \iint_{\tilde{\mathcal{O}}_j^c} D_d(r_j,\beta)p_\alpha(\boldsymbol{\alpha})p_\beta(\beta)d\boldsymbol{\alpha}d\beta \\
&+ \iint_{\tilde{\mathcal{O}}_j} \tilde{D}_d(0,\beta)p_\alpha(\boldsymbol{\alpha})p_\beta(\beta)d\boldsymbol{\alpha}d\beta \\
&\doteq \iint_{\tilde{\mathcal{O}}_j^c} \rho^{-\max\{x(1-\beta)^+, r_j\}}\rho^{-(S(\boldsymbol{\alpha})+\beta)}d\boldsymbol{\alpha}d\beta \\
&+ \iint_{\tilde{\mathcal{O}}_j} \rho^{-x(1-\beta)^+}\rho^{-(S(\boldsymbol{\alpha})+\beta)}d\boldsymbol{\alpha}d\beta. \\
&\doteq \rho^{-\Delta_{j1}(r_j)} + \rho^{-\Delta_{j2}(r_j)} \\
&\doteq \rho^{-\min\{\Delta_{j1}(r_j,b),\Delta_{j2}(r_j,b)\}} \\
&\doteq \rho^{-\Delta_j(r_j,b)}.
\end{aligned} \tag{5.67}$$

The dominant exponent for the joint scheme can then be then obtained as the solution of the following optimization problem

$$\Delta_j(r_j,b) = \min\{\Delta_{j1}(r_j,b), \Delta_{j2}(r_j,b)\}, \tag{5.68}$$

where

$$\Delta_{j1}(r_j, b) = \inf \max\{x(1-\beta)^+, r_j\} + x\beta + S(\boldsymbol{\alpha})$$
$$\text{s.t.}(r_j - x(1-\beta)^+)^+ \leq b\sum_{i=1}^{M_*}(1-\alpha_i)^+. \quad (5.69)$$

and

$$\Delta_{j2}(r_j, b) = \inf x[(1-\beta)^+ + \beta] + S(\boldsymbol{\alpha})$$
$$\text{s.t.}(r_j - x(1-\beta)^+)^+ > b\sum_{i=1}^{M_*}(1-\alpha_i)^+. \quad (5.70)$$

As in previous proofs, we can constrain to $0 \leq \alpha_i \leq 1$ and $0 \leq \beta \leq 1$. We first solve problem (5.69). If $r_j \leq x(1-\beta)$, we have

$$\begin{aligned}\Delta_{j1}(r_j, b) &= \inf x + S(\boldsymbol{\alpha}) \\ \text{s.t.}\, &0 \leq \alpha_i \leq 1 \\ &0 \leq \beta \leq 1 - \frac{r_j}{x}.\end{aligned} \quad (5.71)$$

The infimum corresponds to $\Delta_{j1}(r_j, b) = x$ for $\alpha_i = 0$ and $\beta \in [0, 1 - \frac{r_j}{x})$. Note that the solution is only valid for $r_j \leq x$. If $r_j > x(1-\beta)$, we have

$$\begin{aligned}\Delta_{j1}(r_j, b) &= \inf r_j + S(\boldsymbol{\alpha}) + x\beta \\ \text{s.t.}\, &0 \leq \alpha_i \leq 1 \\ &0 \leq \beta \leq 1 \\ &\beta \leq b\sum_{i=1}^{M_*}(1-\alpha_i)^+ + x - r_j \\ &\beta > 1 - \frac{r_j}{x}.\end{aligned} \quad (5.72)$$

By picking $\alpha_i = 0$, we minimize $S(\boldsymbol{\alpha})$ and increase the domain of β in the problem. Then

$$\begin{aligned}\Delta_{j1}(r_j, b) &= \inf r_j + x\beta \\ \text{s.t.}\, &0 \leq \beta \leq 1 \\ &\beta \leq bM_* + x - r_j \\ &\beta > 1 - \frac{r_j}{x}.\end{aligned} \quad (5.73)$$

The minimizing β is given by $\beta = (x - r_j)^+ + \epsilon$ if $r_j \leq bM_* + x$. Hence, letting $\epsilon \to 0$ the infimum distortion exponent is

$$\Delta_{j1}(r_j, b) = \begin{cases} r_j & \text{if } x \leq r_j \leq x + M_* b, \\ x & \text{if } r_j < x. \end{cases} \quad (5.74)$$

Next we solve the second optimization problem (5.70). If $r_j \geq x + bM_t$ there exist no elements $(\boldsymbol{\alpha}, \beta) \in \mathcal{O}_j^c$, i.e. there is always outage. This can be seen as

$$
\begin{aligned}
(r_j - (x-\beta)^+)^+ &\leq b\sum_{i=1}^{M_*}(1-\alpha_i)^+ \\
(bM_t + x + \epsilon - (x-\beta)^+)^+ &\leq b\sum_{i=1}^{M_*}(1-\alpha_i)^+
\end{aligned}
\tag{5.75}
$$

which can never be satisfied for a rate $r_j = x + bM_t + \epsilon$. If $\beta = 1$, $\Delta_{j2}(r_j, b)$ is minimized and the range of $\boldsymbol{\alpha}$ increases. Then, the problem becomes

$$
\begin{aligned}
\Delta_{j2}(r_j, b) = \inf\, & x + S(\boldsymbol{\alpha}) \\
\text{s.t.}\, & 0 \leq \alpha_i \leq 1 \\
& r_j > b\sum_{i=1}^{M_*}(1-\alpha_i)^+ .
\end{aligned}
\tag{5.76}
$$

that is an scaled version of the dmt problem in 5.37. Hence, we have

$$
\Delta_{2j}(r_j, b) = \begin{cases} x + d^*(\frac{r_j}{b}) & \text{if } r_j \leq b, \\ x & \text{if } r_j > b. \end{cases} \tag{5.77}
$$

Bringing all together, if $x \leq b$ the expected distortion exponent for the joint source-channel scheme in (5.68) is

$$
\Delta_j(r_j, b) = \begin{cases} x & \text{if } r_j \leq x, r_j > b, \\ \min\{x + d^*(\frac{r_j}{b})\} & \text{if } x \leq r_j \leq b. \end{cases} \tag{5.78}
$$

If $b < x$, the expected distortion exponent for the joint source-channel scheme is $\Delta_j(r_j, b) = x$.

We now design the code such that the minimum distortion exponent is maximized, i.e.

$$
\Delta_j(b) \triangleq \max_{r_j} \Delta_j(r_j, b). \tag{5.79}
$$

In order to maximize the minimum distortion exponent, we chose a rate such that $x + d^*(\frac{r_j}{b}) = r_j$, that is an scaling of 5.45, and hence

$$
r_j = br_c. \tag{5.80}
$$

Appendix 5.E Proof for distortion exponent: hybrid digital-analog, case $M_*b \leq 1$

Let $b' \triangleq \frac{1}{1-bM_*}$. By applying the usual change of variables $R_h = r_h \log\rho$, $\lambda_i^2 = \rho^{-\alpha_i}$ and $\|\gamma\|^2 = \rho^{-\beta}$, in (3.15) we have

$$
\begin{aligned}
\tilde{\mathcal{O}}_h &= \left\{ b' \log \frac{\det\left(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}_{M_*}\mathbf{H}_{M_*}^\dagger\right)}{\det\left(\mathbf{I} + \frac{\rho^{1-\eta}}{M_*}\mathbf{H}_{M_*}\mathbf{H}_{M_*}^\dagger\right)} < \log\left(1 + \frac{2^{R_h}}{1+\rho^x\gamma^2}\right)\right\} \\
&= \left\{ b' \log \frac{\prod_{i=1}^{M_*}\left(1 + \frac{\rho}{M_*}\lambda_i\right)}{\prod_{i=1}^{M_*}\left(1 + \frac{\rho^{1-\eta}}{M_*}\lambda_i\right)} < \log\left(1 + \frac{2^{R_h}}{1+\rho^x\gamma^2}\right)\right\} \\
&\doteq \left\{ b'(\sum_{i=1}^{M_*}(1-\alpha_i)^+ - (1-\alpha_i-\eta)^+) < (r_h - (x-\beta)^+)^+, \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0 \right\}. \quad (5.81)
\end{aligned}
$$

We now examine each term in (3.16). The exponent for the first term is found to be, similarly to the other proofs as

$$
\begin{aligned}
\Delta_{h1}(r_h, \eta) \;&=\; \inf \max\{r_h, (x-\beta)^+\} + S(\boldsymbol{\alpha}) + \beta \\
&\text{s.t. } (\boldsymbol{\alpha}, \beta) \in \tilde{\mathcal{O}}_h^c. \quad (5.82)
\end{aligned}
$$

The second term is found to be

$$
\begin{aligned}
\Delta_{h2}(r_h, \eta) \;&=\; \inf \max\{(1-\alpha_1-\eta)^+, (x-\beta)^+\} + S(\boldsymbol{\alpha}) + \beta \\
&\text{s.t. } (\boldsymbol{\alpha}, \beta) \in \tilde{\mathcal{O}}_h^c. \quad (5.83)
\end{aligned}
$$

Finally, it is easy to see that the third and fourth term have the same exponential behaviour, given by

$$
\begin{aligned}
\Delta_{h34}(r_h, \eta) \;&=\; \inf (x-\beta)^+ + S(\boldsymbol{\alpha}) + \beta \\
&\text{s.t. } (\boldsymbol{\alpha}, \beta) \in \tilde{\mathcal{O}}_h. \quad (5.84)
\end{aligned}
$$

Then, similarly to the other proofs, the expected distortion is exponentially equivalent to

$$
\begin{aligned}
ED_h(R_h) &\doteq (1-bM_*)\rho^{-\Delta_{h1}(r_h,\eta)} + bM_*\rho^{-\Delta_{h2}(r_h,\eta)} + \rho^{-\Delta_{h34}(r_h,\eta)} \\
&\doteq \rho^{-\Delta_h(r_h,\eta)} \quad (5.85)
\end{aligned}
$$

where

$$\Delta_h(r_h, \eta) = \inf\{\Delta_{h1}(r_h,\eta), \Delta_{h2}(r_h,\eta), \Delta_{h34}(r_h,\eta)\}. \quad (5.86)$$

We can restrict $0 \leq \alpha_i \leq 1$ and $0 \leq \beta \leq x$ without loss of generlality. We first start by solving problem (5.84). It can be seen that $\beta = x$ as the distortion exponent is minimized and the domain of α_i is increased. Then,

$$
\begin{aligned}
\Delta_{h34}(r_h, \eta) \;&=\; \inf x + S(\boldsymbol{\alpha}) \\
&\text{s.t. } b' \sum_{i=1}^{M_*} [(1-\alpha_i)^+ - (1-\alpha_i-\eta)^+] < r_h, \\
&\alpha_1 \geq ... \geq \alpha_{M*} \geq 0. \quad (5.87)
\end{aligned}
$$

Let $\frac{r_h}{b'} = k\eta + \delta$ with $k \in \{0, ..., M_* - 1\}$ and $0 \le \delta \le \eta$. Then, the infimum is given by

$$\alpha_i = \begin{cases} 1, & 1 \le i < M_* - k, \\ 1 - \delta, & i = M_* - k, \\ 0, & M_* - k < j \le M_* \end{cases} \tag{5.88}$$

. Then we have that

$$\Delta_{h34} = x + d_k(r_h) \tag{5.89}$$

where

$$d_k(r_h) = (M^* - k)(M_* - k) - (M^* + M_* - 1 - 2k)\delta. \tag{5.90}$$

for $\frac{r_h}{b'} = k\eta + \delta$ with $k \in \{0, ..., M_* - 1\}$ and $0 \le \delta \le \eta$

Next, we solve (5.82). We have that $\boldsymbol{\alpha} = 0$ independently of the rest of parameters. If $r_h \le x - \beta$, then the problem simplifies to

$$\begin{aligned} \Delta_{h1}(r_h, \eta) \ &= \ \inf (x - \beta)^+ + \beta \\ &\text{s.t. } \ 0 \le \beta < x - r_h. \end{aligned} \tag{5.91}$$

The infimum is achieved for $\Delta_{h1} = x$ for any $\beta \in [0, x - r_h)$ if $x > r_h$. If $r_h > x - \beta$, we have

$$\begin{aligned} \Delta_{h1}(r_h, \eta) \ &= \ \inf r_h + \beta \\ &\text{s.t. } \ M_* b'(1 - (1 - \eta)^+) \ge r_h - (x - \beta) \\ &\qquad 0 \le \beta \le x \\ &\qquad \beta \ge x - r_h. \end{aligned} \tag{5.92}$$

The infimum is achieved for $\beta = (x - r_h)^+$. Finally,

$$\Delta_{h1}(r_h, \eta) = \begin{cases} x + \epsilon_1 & \text{if } r_h < x, \\ r_h & \text{if } x \le r_h \le x + \eta b' M_*. \end{cases} \tag{5.93}$$

Next we solve (5.83). We have that $\boldsymbol{\alpha} = 0$ is optimal indendently of β, r_h, x. If $r_h < x - \beta$, we have

$$\begin{aligned} \Delta_{h2}(r_h, \eta) \ &= \ \inf \max\{(1 - \eta)^+, x - \beta\} + \beta \\ &\text{s.t. } \ 0 \le \beta \le M^* b' \eta + x - r_h. \end{aligned} \tag{5.94}$$

The infimum is achieved for $\beta = 0$ if $x <\ge r_h$, and

$$\Delta_{h2}(r_h, \eta) = \max\{(1 - \eta)^+, x\}, \qquad r_h \le M^* b' \eta + x. \tag{5.95}$$

The distortion exponent is then given by the maximum infimum in (5.86), i.e,

$$\Delta_h(b, x) = \max_{r_h, \eta} \inf\{\Delta_{h1}(r_h, \eta), \Delta_{h2}(r_h, \eta), \Delta_{h34}(r_h, \eta)\}. \tag{5.96}$$

By letting $k = M_* - 1$, we have

$$d_k = (M^* - M_* + 1)(1 - \delta) \tag{5.97}$$

Notice that $0 \geq \delta \geq \eta$ and hence, $\Delta_3 4(b, x) \geq \Delta_2(b, x)$. Consequently, by picking $r_h = 1 - \eta$, we have $r_h \leq b' M_*(1-r)+x$, and the maximum exponent is found for $r_h = M_* b + (1 - bM_*)x - \epsilon$ and $\eta = 1 - r_h$. Then, the maximum distortion exponent is given, by letting $\epsilon \to 0$ by

$$\Delta_h(b, x) = M_* b + (1 - bM_*)x. \tag{5.98}$$

Appendix 5.F Proof for distortion exponent: hybrid digital-analog, case $M_* b > 1$

The expected distortion function behavior, at high SNR can be calculated, similarly to the the proofs in previous sections as

$$\begin{aligned} ED_h(R_h) = & \iint_{\mathcal{O}_h} D_d(0, \gamma) p_h(\mathbf{H}_{M_*}) p_\gamma(\gamma) d\mathbf{H}_{M_*} d\gamma + \iint_{\mathcal{O}_h^c} D_{pd}(R_h, \mathbf{H}_{M_*}, \gamma) p_h(\mathbf{H}_{M_*}) p_\gamma(\gamma) d\mathbf{H}_{M_*} d\gamma. \\ & \doteq \rho^{-\Delta_{h1}} + \rho^{-\Delta_{h2}}, \\ & \doteq \rho^{-\max\{\Delta_{h1}, \Delta_{h2}\}}, \end{aligned} \tag{5.99}$$

where Δ_{h1},Δ_{h2} are obtained similarly to other proofs and given next.

By applying the usual change of variables $R_h = r_h \log \rho$, $\sigma^2 = (2^{R_h} - 1) \doteq \rho^{-r_h}$ $\lambda_i^2 = \rho^{-\alpha_i}$ and $\|\gamma\|^2 = \rho^{-\beta}$ the outage event is given at high SNR by

$$\begin{aligned} \mathcal{O}_h &= \left\{ (\mathbf{H}, \gamma) : \log\det\left(\left(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}\mathbf{H}^\dagger \right)^{-1} + \frac{1}{\sigma^2(\gamma^2+1)}\mathbf{I} \right) > b' \log\det(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}_{M_*}\mathbf{H}_{M_*}^\dagger) \right\} \\ &= \left\{ (\mathbf{H}, \gamma) : \det\left(\left(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}\mathbf{H}^\dagger \right)^{-1} \right) \det\left(\mathbf{I} + \frac{\left(\mathbf{I} + \frac{1}{M_*}\mathbf{H}\mathbf{H}^\dagger\right)}{\sigma^2(\gamma^2+1)} \right) > \det(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}_{M_*}\mathbf{H}_{M_*}^\dagger)^{b'} \right\} \\ &= \left\{ (\mathbf{H}, \gamma) : \det\left(\mathbf{I} + \frac{\left(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}\mathbf{H}^\dagger\right)}{\sigma^2(\gamma^2+1)} \right) > \det(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}_{M_*}\mathbf{H}_{M_*}^\dagger)^{1+b'} \right\} \\ &= \left\{ (\mathbf{H}, \gamma) : \prod_{i=1}^{M_*} \left(1 + \frac{1 + \rho\lambda_i}{\sigma^2(\gamma^2+1)} \right) > \prod_{i=1}^{M_*} (1 + \rho\lambda_i)^{1+b'} \right\} \\ &\doteq \left\{ (\boldsymbol{\alpha}, \beta) : \prod_{i=1}^{M_*} \left(1 + \frac{\rho^{(1-\alpha_i)^+}}{\rho^{-r_h}(\rho^{(x-\beta)^+})} \right) > \prod_{i=1}^{M_*} \left(\rho^{(1-\alpha_i)^+} \right)^{1+b'}, \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0 \right\} \\ &\doteq \left\{ (\boldsymbol{\alpha}, \beta) : \sum_{i=1}^{M_*} [(1-\alpha_i)^+ + r_h - (x-\beta)^+]^+ > (1+b') \sum_{i=1}^{M_*} (1-\alpha_i)^+, \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0 \right\} \end{aligned} \tag{5.100}$$

where $b' = (M_* b - 1)$.

On the other hand, the exponential behavior equivalent of the distortion function $D_{pd}(R,h,\gamma)$ is given by

$$\begin{aligned}D_{pd}(R,h,\gamma) &= \sum_{i=1}^{M_*}\left(\frac{1}{\sigma^2}(1+\frac{\rho\lambda_i}{M_*})+\gamma^2\rho^x+1\right)^{-1}. \\ &\doteq \sum_{i=1}^{M_*}\left(\rho^{(x-\beta)^+}+\rho^{r_h}\rho(1-\alpha_i)^+\right)^{-1} \\ &\doteq \rho^{-\max\{(x-\beta)^+, r_h+(1-\alpha_{M_*})^+\}} \qquad (5.101)\end{aligned}$$

Similarly to other proofs, without loss of generality we can constrain $0 \leq \alpha_i \leq 1$ and $0 \leq \beta \leq x$. Hence,

$$\begin{aligned}\Delta_{h1}(r_h) &= \max\{(x-\beta)^+, r_h+(1-\alpha_{M_*})^+\}+S(\boldsymbol{\alpha})+\beta \\ &\text{s.t. } \sum_{i=1}^{M_*}[(1-\alpha_i)^+ + r_h-(x-\beta)^+]^+ \leq (1+b')\sum_{i=1}^{M_*}(1-\alpha_i)^+, \\ &\alpha_1 \geq ... \geq \alpha_{M_*} \geq 0. \qquad (5.102)\end{aligned}$$

and

$$\begin{aligned}\Delta_{h2}(r_h) &= \inf (x-\beta)^+ + S(\boldsymbol{\alpha})+\beta \\ &\text{s.t. } \sum_{i=1}^{M_*}[(1-\alpha_i)^+ + r_h-(x-\beta)^+]^+ > (1+b')\sum_{i=1}^{M_*}(1-\alpha_i)^+, \\ &\alpha_1 \geq ... \geq \alpha_{M_*} \geq 0. \qquad (5.103)\end{aligned}$$

First we solve (5.102). The infimum for this problem is obtained for $\boldsymbol{\alpha} = 0$ and $\beta = 0$, and is given by

$$\Delta_{1h}(b,x) = \max\{x, r_h+1\} \qquad \text{for } r_h \leq M_* b - 1 + x. \qquad (5.104)$$

Now we solve (5.105). Notice that by letting $\beta = x$, the range α_i increases while $\Delta_{2h}(b,x)$ is minimized. Hence, the problem to solve is

$$\begin{aligned}\Delta_{h2}(r_h) &= \inf x + S(\boldsymbol{\alpha}) \\ &\text{s.t. } M_* r_h^+ > b'\sum_{i=1}^{M_*}(1-\alpha_i)^+, \\ &\alpha_1 \geq ... \geq \alpha_{M_*} \geq 0. \qquad (5.105)\end{aligned}$$

This problem has been already solved in Section 5.37, and is given, changing $b' = M_* b - 1$, by

$$\Delta_{2h}(b,x) = x + d\left(\left(b - \frac{1}{M_*}\right) r_h\right). \qquad (5.106)$$

Gathering all results, the distortion exponent is maximized by letting $\Delta_1(r_h,x) = \Delta_2(r_h,x)$, i.e. $r_h + 1 = x + d((bM_* - 1)r_h)$ for $r_h > x$. First, let $r'_h = r_h(b - \frac{1}{M_*})$. Then, the problem can be expressed as $r'_h\left(b - \frac{1}{M_*}\right) + 1 = x + d(r'_h)$. From the points at which $r_h = k$ is an integer, $k = 1, ..., M_* - 1$, we have that the ranges of b for which $d(r_h)$ is the interpolating points between $d(k-1)$ and $d(k)$, by equaling $1 + i(b - \frac{1}{M_*}) = (M^* - 1)(M_* - 1) + x$, are given by

$$b \in [k - M^* + \frac{1}{M_*} - M_* + \frac{M^* M_* - 1 + x}{k}, k - 1 - M^* + \frac{1}{M_*} - M_* + \frac{M^* M_* - 1 + x}{k-1}) \quad \text{for } k = 1, ..., M_* - 1. \tag{5.107}$$

and the solution is given by

$$\Delta_h(b, x) = 1 + \frac{(b(M_* - k) - 1)((M^* - k)(M_* - k) - 1 + x)}{(M_* - k)(b + (M^* - k) + (M_* - k) - 1) - 1} \tag{5.108}$$

for

$$r'_h = \frac{(M_* - k)((M^* - k)(M_* - k) - 1 + x)}{(M_* - k)(b + (M^* - k) + (M_* - k) - 1) - 1}, \tag{5.109}$$

If$\frac{-1+M_*(1+M^*-M_*+x))}{(M_*-1)M_*} \leq b \leq \frac{1}{M_*}$, we have that the intersection is given by

$$1 + r'_h \left(b - \frac{1}{M_*} \right) = x + M^* M_* - (M^* + M_* - 1) r'_h \tag{5.110}$$

that solves as

$$r'_h = \frac{M_*(M^* M_* - 1 + x)}{M_*(b + M^* + M_* - 1) - 1}, \tag{5.111}$$

and the distortion exponent is

$$\Delta_h(b, x) = 1 + \frac{(bM_* - 1)(M^* M_* - 1 + x)}{M_*(b + M^* + M_* - 1) - 1} \tag{5.112}$$

Notice that if $x \geq bM_*$, the distortion exponent is always $\Delta_h(b, x) = x$, due to the intersection with $d(r) = x$, constant. Taking the maximum of of this expression with x at each interval completes the proof.

Appendix 5.G Proof for distortion exponent: HDA

We have that for $\mathbf{H}\mathbf{H}^\dagger = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$ the distortion expression can be expressed as

$$\begin{aligned} ED_h(\boldsymbol{\kappa}) = & \iint_{\mathcal{O}_h^c} D(\boldsymbol{\kappa}, \mathbf{H}, \gamma^2) p_\gamma(\gamma) p_\lambda(\mathbf{\Lambda}) p_u(\mathbf{U}) d\gamma d\mathbf{\Lambda} d\mathbf{U} \\ & + \iint_{\mathcal{O}_h} D(\mathbf{0}, \mathbf{H}, \gamma^2) p_\gamma(\gamma) p_\lambda(\mathbf{\Lambda}) p_u(\mathbf{U}) d\gamma d\mathbf{\Lambda} d\mathbf{U}. \end{aligned}$$

First note that $p(\mathbf{U})$ is a Haar distribution and does not depend on ρ and all dependency on ρ of $\mathbf{H}$ is included in $\mathbf{\Lambda}$.

Let $\lambda^i = \rho^{(1-\alpha_i)}$, $a = \rho^{(x-\beta)^+}$ and $\kappa^2 = \rho^{1-\nu}$ with $\nu \in [0, \infty)$, then,

$$1 + \frac{\rho^{1-\nu}}{\rho^{(x-\beta)^+}} \sum_{i=1}^{M_*} \frac{s_i^2}{|\mathbf{s}|^2} \left(1 + \frac{1}{M_t} \rho^{1-\alpha_i} \right) > \prod_{i=1}^{M_*} \left(1 + \frac{\rho^{1-\alpha_i}}{M_t} \right)^b$$

The exponential behavior of

$$\sum_{i=1}^{M_*} \frac{s_i^2}{|\mathbf{s}|^2}\left(1+\frac{1}{M_t}\rho^{1-\alpha_i}\right) \doteq \rho^{(1-\alpha_{M_*})^+} \tag{5.113}$$

this is true with probability one, as the multiplying component to the maximum eigenvalue is $s_{M_*}=0$ with probability 0. On the other hand, by definition of the Haar distribution, the Haar distributed vectors are unit norm and hence $\sum_{i=1}^{M_*} s_i^2 = 1$ and $\frac{s_i}{|\mathbf{s}|^2}$ is bounded. Finally, we can conclude that the maximum eigenvalue dominates the sum with probability one. Then,

$$1+\rho^{(1-\nu)-(x-\beta)^+ +(1-\alpha_{M_*})^+} > \rho^{b\sum_{i=1}^{M_*}(1-\alpha_i)^+} \tag{5.114}$$

that is equivalent to

$$\begin{aligned}\{(\boldsymbol{\alpha},\beta) : [(1-\nu)-(x-\beta)+(1-\alpha_{M_*})]^+ > b\sum_{i=1}^{M_*}(1-\alpha_i),\\ \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0\}\end{aligned} \tag{5.115}$$

The same reasoning can be done for the distortion exponential behavior, and

$$\begin{aligned}D(\boldsymbol{\kappa},\mathbf{H},\gamma^2) &= (1+\gamma^2+\frac{\kappa^2}{|\mathbf{s}|^2}\sum_{i=1}^{M_*}(1+\lambda_i)s_i^2)^{-1}\\ &\doteq (1+\rho^{(x-\beta)^+}+\frac{1-\nu}{|\mathbf{s}|^2}\sum_{i=1}^{M_*}(1+\lambda_i)s_i^2)^{-1}\end{aligned} \tag{5.116}$$

Using Varadhan integral lemma, we have that

$$\begin{aligned}&\Delta_{1hda}(b,x) =\\ &\inf \max\{x-\beta, 1-\nu+(1-\alpha_{M_*})\} + S(\boldsymbol{\alpha})+\beta\\ &\text{s.t. } [(1-\nu)-(x-\beta)+(1-\alpha_{M_*})]^+ \leq b\sum_{i=1}^{M_*}(1-\alpha_i),\\ &\alpha_1 \geq ... \geq \alpha_{M_*} \geq 0\}\end{aligned} \tag{5.117}$$

and

$$\begin{aligned}\Delta_{2hda}(b,x) = &\inf (x-\beta)+S(\boldsymbol{\alpha})+\beta\\ &\text{s.t. } [(1-\nu)-(x-\beta)+(1-\alpha_{M_*})]^+ > b\sum_{i=1}^{M_*}(1-\alpha_i),\\ &\alpha_1 \geq ... \geq \alpha_{M_*} \geq 0\}\end{aligned} \tag{5.118}$$

First we solve problem (5.117). If $1-\nu \geq x$, the minimum for this problem is given by $\beta=0$ and $\boldsymbol{\alpha}=0$ and is given by

$$\begin{aligned}\Delta_{1hda}(b,x) &= \inf \max\{x, 1-\nu+1\}\\ &\text{for } (1-\nu)-x+1 \leq bM_*.\end{aligned} \tag{5.119}$$

Now we solve problem (5.121). We can set $\beta = x$ as the optimal β as the domain of $\boldsymbol{\alpha}$ is increased while $\Delta_{2hda}(b,x)$ is minimized.

If $1-\nu \leq (1-b)$, we have that

$$\begin{aligned}\Delta_{2hda}(b,x) = &\inf(x-\beta) + S(\boldsymbol{\alpha}) + \beta \\ &\text{s.t. } [(1-\nu) - (x-\beta) + (1-\alpha_{M_*})]^+ > b\sum_{i=1}^{M_*}(1-\alpha_i), \\ &\alpha_1 \geq ... \geq \alpha_{M_*} \geq 0\} \end{aligned} \tag{5.120}$$

is minimized by $\alpha_{M_*} = 1 - \frac{1-\nu}{b-1} + \epsilon$ and $\alpha_i = 1$ for $i = 1, ..., M_* - 1$. And

$$\begin{aligned}\Delta_{2hda}(b,x) &= x + (M^* - 1)(M_* - 1) \\ &\quad + (M^* + M_* - 1)(1 - \frac{1-\nu}{b-1} - \epsilon) \end{aligned} \tag{5.121}$$

By equating $\Delta_{1hda}(b,x) = \Delta_{1hda}(b,x)$ we obtain

$$\begin{aligned}\Delta_{hda}(b,x) &= \frac{M^* + M_* - 1 + (b-1)M^*M_*}{M^* + M_* - 1 + b - 1 + (b-1)x}, \\ &\text{for } b \geq M^*M_* + x - (M^* + M_* - 1). \end{aligned} \tag{5.122}$$

for

$$1 - \nu = \frac{(b-1)(M^* + M_* - 1 + x)}{M1 + M2 - 2 + b}, \tag{5.123}$$

that corresponds to

$$b \geq (M^* - 1)(M_* - 1) + x. \tag{5.124}$$

If $1-\nu \geq (1-b)$, consider $b(k-1) + (b-1) \leq r_h \leq bk + (b-1)$ with $k = 1, ..., M_* - 1$, or equivalently, $r_h = b(k-1) + b - 1 + \delta$ with $\delta \in [0, b)$. We have that the problem is minimized by

$$\alpha_i = \begin{cases} 1 & 1 \leq i < M_* - k \\ 1 - \frac{\delta}{b} + \epsilon & i = M_* - k \\ 0, & M_* - k < i \leq M_* - 1. \end{cases} \tag{5.125}$$

with $\epsilon > 0$ and the infimum is given by

$$\begin{aligned}\Delta_{2hda}(b,x) &= x + (M^* - 1 - k)(M_* - 1 - k) \\ &\quad + (M^* + M_* - 1 - 2k)\left(1 - \frac{\delta}{b}\right) \\ &\quad \text{for } b(k-1) + (b-1) \leq r_h \leq bk + (b-1), \\ &\quad k = 1, ..., M_* - 1. \end{aligned} \tag{5.126}$$

For $b \in \left[\frac{x + (M^* - (k+1))(M_* - (k+1))}{(k+1)}, \frac{x + (M^* - k)(M_* - k)}{k}\right)$ for $k = 1, ..., M_*$, the solution is given by

$$\Delta_{2hda}(b,x) = \frac{b\,(M^*M_* - (k(k+1) + x)}{b - 1 + M^* + M_* - 2k}. \tag{5.127}$$

and $\Delta_{hda}(b,x) = x$ for $b \leq \frac{x}{M_*}$.

Appendix 5.H Proof for distortion exponent: multi-layer no binning

Following [2], we consider a power allocation satisfying

$$\bar{\rho}_k = \rho^{1-\sum_{i=0}^{k-1} r_i + \epsilon_{k-1}}. \tag{5.128}$$

with $0 < \epsilon_{k-1} < \epsilon_k$. It is straightforward to obtain that the exponential behavior of the expected distortion ED_{ml} is given by

$$ED_{ml} \doteq \sum_{k=0}^{L} \rho^{-d_{sd}(r_{k+1})} \rho^{-\Delta_k^d} \doteq \sum_{k=0}^{L} \rho^{-\Delta_k}. \tag{5.129}$$

where $\Delta_k^d \triangleq \max\{x, \sum_{i=0}^{k} r_i\}$ is the equivalent distortion exponent for the averaged distortion. The successive decoding diversity gain is found in [2] as the solution to the successive probability of outage at each layer, i.e.

$$\begin{aligned} d_{ds}(r_k) &= \inf S(\boldsymbol{\alpha}) \\ \text{s.t.} & \sum_{i=1}^{M_*}(1 - \sum_{i=1}^{k-1} r_i - \epsilon_{k-1} - \alpha_i)^+ - \sum_{i=1}^{M_*}(1 - \sum_{i=1}^{k} r_i - \epsilon_k - \alpha_i)^+ < r_{k+1}, \\ & \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0, \end{aligned} \tag{5.130}$$

that is explicitly given by

$$d_{ds}(r_k) = M^* M_* (1 - \sum_{i=1}^{k-1} r_i) - (M^* - M_* - 1) r_k. \tag{5.131}$$

The exponent distortion is given by the minimum of each layer exponent Δ_k, i.e.

$$\Delta_{ml}^L = \min_{0 \leq k \leq L} \left\{ d_{sd}(r_{k+1}) + \Delta_k^d \right\}. \tag{5.132}$$

If $x \geq br_1$, we have $\Delta_k^d = x$ for all k. The minimum exponent is then given by $\Delta_L = x$. If $x \leq br_1 x$, we have that $\Delta_k^d = b\sum_{i=1}^{k} r_i$ for all i. The system becomes

$$\begin{aligned} \Delta_0 &= x + M^* M_* - (M^* + M_* - 1) r_1, \\ \Delta_k &= M^* M_* - (M^* M_* - b) \sum_{i=1}^{k} r_i - (M^* - M_* - 1) r_{k+1}, \\ &\quad \text{for } \; k = 1 \dots L-1, \\ \Delta_L &= b \sum_{i=1}^{L} r_i. \end{aligned} \tag{5.133}$$

Now we design the rates r_k such that the minimum distortion exponent is maximized. Note that Δ_0 and Δ_k are decreasing in r_k while Δ_L is increasing. We first consider the case $b \geq (M_t - 1)(M_r - 1)$. Let

$$\eta_0 = \frac{b - (M_t - 1)(M_r - 1)}{M_t + M_r - 1} \geq 0. \tag{5.134}$$

For $0 \leq \eta_0 < 1$, i.e. $M^* M_* \geq b \geq (M^*-1)(M_*-1)$ we minimize by equaling all the exponents $\Delta_0 = \Delta_1 = \cdots = \Delta_L$. Then, we have that all distortion exponents are equal for

$$r_k = \left(\frac{1-\eta_0^{L-1}}{1-\eta_0}\right)^{k-2} r_2, \text{ for } k = 3, ..., L, \tag{5.135}$$

where

$$r_2 = \eta_0 r_1 - \frac{x}{M_t + M_r - 1}, \tag{5.136}$$

and

$$r_1 = \frac{1-\eta_0}{M_t M_r - b\eta_0^L}\left(M_t M_r + x\left(1 + \frac{b}{M_t + M_r - 1}\left(\frac{1-\eta_0^{L-1}}{1-\eta_0}\right)\right)\right).$$

the system solves for

$$\Delta_{ml}^L = \frac{bM_t M_r(1-\eta_0^L)}{M_t M_r - b\eta_0^L} + x\frac{b(1-\eta_0)\eta_0^{L-1}}{M_t M_r - b\eta_0^L}, \tag{5.137}$$

If $\eta_0 \geq 1$, i.e. $b \geq M^* M_*$, rates r_k cannot be chosen to be decreasing. consequently, we pick equal rates $r = r_2 = r_3 = \cdots = r_L$. Then we have that $\Delta_k \geq \Delta_2$. Hence we can only maximize the lower Δ_k, i.e Δ_2. With this choice, the system becomes:

$$\begin{aligned} \Delta_0 &= x + M^* M_* - (M^* + M_* - 1)r_1 \\ \Delta_2 &= M^* M_* - (M^* M_* - b)r_1 - (M^* + M_* - 1)r \\ \Delta_L &= b(r_1 + (L-1)r). \end{aligned} \tag{5.138}$$

By equating, the system is solved for

$$\Delta_{ml}^L = \frac{b(L-1)(b-\delta_1)(x+\delta_1) + b(x + \delta_1 + (L-1)\delta_1)\delta_2}{b(L-1)(b-\delta_1) + bL\delta_2 + \delta_2^2}, \tag{5.139}$$

with the rates

$$r = \frac{\delta_1(b - x - \delta_1 + \delta_2)}{b(L-1)(b-\delta_1) + bL\delta_2 + \delta_2^2}, \tag{5.140}$$

and

$$r_1 = \frac{b(L-1)x + (x+\delta_1)\delta_2}{b(L-1)(b-\delta_1) + bL\delta_2 + \delta_2^2}, \tag{5.141}$$

where

$$\delta_1 = M^* M_* \text{ and } \delta_2 = M^* + M_* - 1. \tag{5.142}$$

For $(M^* - k - 1)(M^* - k - 1) \leq b < (M^* - k)(M_* - k)$, $k = 1, ..., M_* - 1$, we can consider a $(M^* - k) \times (M_* - k)$ antenna system and following the same steps as above, we obtain,

$$\Delta_{ml}^L = \frac{b(M_t - k)(M_r - k)(1-\eta_k^L)}{M_t M_r - b\eta_k^L} + x\frac{b(1-\eta_k)\eta_k^{L-1}}{(M_t - k)(M_r - k) - b\eta_k^L}, \tag{5.143}$$

Note that if you consider $x > b\sum_{i=1}^{k} r_i$, i.e the first k rates are under the quality provided by the side information, by equaling the distortion exponents $\Delta_1 = \Delta_i$ for $i = 2...k$, we have that rates of such layers are forced to be 0, i.e. $r_i = 0$ for $i = 2...k$. By relabeling the rates as $r_{i+1} = r_{i-k+2}$, the distortion exponents system becomes equivalent to one with $L - k$ layers. Thus there is no improvement by pursuing this strategy.

Finally, putting all the results together, the distortion exponent Δ_{ml}^{L} is proved to be

Theorem 18. *The distortion exponent $\Delta_{mle}^{L}(b)$ for n multiple layer and estimate coding scheme is characterized by*

$$\Delta_{ml}^{L}(b,x) = \begin{cases} x & \text{if } b \le x, \\ \frac{b+b^L x - b^{L+1}(1+x)}{1-b^{L+1}} & \text{if } x < b < 1, \\ \frac{b(1-(L-1)x + b(L-1)(1+x))}{1+b+b^2(L-1)} & \text{if } x < 1 \le b. \end{cases} \tag{5.144}$$

The distortion exponent $\Delta_{ml}^{\infty}(b,x)$ for an infinite continuum of multiple layer and estimate coding scheme can be proven to converge to

$$\Delta_{ml}^{\infty}(b,x) = \begin{cases} x & \text{if } b \le x, \\ b & \text{if } x < b < M_t M_r, \\ M^* M_* + x\left(\frac{b - M_* M^*}{b-(M^*-1)(M_*-1)}\right) & \text{if } x < M_t M_r \le b. \end{cases} \tag{5.145}$$

Appendix 5.I Proof for distortion exponent: multi-layers NBJD

The expected distortion can be expressed as

$$ED_{mj}(\mathbf{R}) = \sum_{k=0}^{L} \iint_{\mathcal{L}_{k+1}} D_d\left(\sum_{i=0}^{k} R_i, \mathbf{H}, \gamma\right) p_h(\mathbf{H}) p_\gamma(\gamma) d\mathbf{H} d\gamma \tag{5.146}$$

where $\mathcal{L}_{L+1}$ is the set such that all layers are decoded and is given by $\mathcal{L}_L^c$. We have used the fact that $\mathcal{L}_k$ are mutually exclusive to decompose the integral.

We consider a power allocation satisfying

$$\rho_k = \rho^{1-\sum_{i=0}^{k-1}\frac{r_i}{b}+\epsilon_{k-1}}, \tag{5.147}$$

such that$\sum_{i=0}^{L} r_i \le b$ for $k = 2, ..., L$ and $0 < \epsilon_{k-1} < \epsilon_k$. By applying the usual change of variables, $\lambda_i \doteq \rho^{-\alpha_i}$, $\alpha_1 \ge ... \ge \alpha_{M_*} \ge 0$, we have,

$$\begin{aligned} I(\mathbf{U}_{i,k}; \mathbf{V}_i | \mathbf{U}_{i,1}^{k-1}) \doteq & \sum_{i=1}^{M_*}(1 - \frac{r_1}{b} - ... - \frac{r_{k-1}}{b} - \epsilon_{k-1} - \alpha_i)^+ \\ & - \sum_{i=1}^{M_*}(1 - \frac{r_1}{b} - ... - \frac{r_k}{b} - \epsilon_k - \alpha_i)^+; \end{aligned} \tag{5.148}$$

By applying the change of variables, $R_i = r_i \log \rho$ and $\gamma^2 = \rho^{-x\beta}$, (3.68) is equivalent to

$$I(X; W_k | W_1^{k-1}, Y) = \log \left(\frac{\rho^{\sum_{i=0}^{k} r_i - x(1-\beta)^+} + 1}{\rho^{\sum_{i=0}^{k-1} r_i - x(1-\beta)^+} + 1} \right), \tag{5.149}$$

for $k = 1, ..., L$.

Then, at high SNR we have

$$Pr\{\mathcal{L}_k\} \doteq \left\{ b\left(1 - \sum_{i=1}^{k-1} \frac{r_i}{b} - \epsilon_{k-1} - \alpha\right)^+ - b\left(1 - \sum_{i=1}^{k} \frac{r_i}{b} - \epsilon_k - \alpha\right)^+ \right.$$
$$\left. < \left(\sum_{i=1}^{k} r_i - x(1-\beta)^+\right)^+ - \left(\sum_{i=1}^{k-1} r_i - x(1-\beta)^+\right)^+, \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0 \right\} \tag{5.150}$$

Then, similarly to the other proofs, the expected distortion function is exponentially equivalent to

$$\begin{aligned} &ED_{mj}(\mathbf{R}) \\ &= \sum_{k=0}^{L} \iint_{\mathcal{L}_{k+1}} D_d\left(\sum_{i=0}^{k} R_i, \mathbf{H}, \gamma\right) p_h(\mathbf{H}) p_\gamma(\gamma) d\mathbf{H} d\gamma \\ &\doteq \sum_{k=0}^{L} \iint_{\tilde{\mathcal{L}}_{k+1}} \rho^{-(\max\{\sum_{i=0}^{k} r_i, x(1-\beta)^+\} + x\beta + S(\boldsymbol{\alpha}))} d\boldsymbol{\alpha} d\beta \\ &\doteq \sum_{k=0}^{L} \rho^{-\Delta_k(\mathbf{r})} \\ &\doteq \rho^{-\Delta_{mj}(\mathbf{r})}. \end{aligned} \tag{5.151}$$

where

$$\Delta_{mj}(\mathbf{r}) = \min\left\{\Delta_k(\mathbf{r})\right\}. \tag{5.152}$$

and

$$\begin{aligned} \Delta_k(\mathbf{r}) = \inf \max &\left\{ \sum_{i=0}^{k} r_i, x(1-\beta)^+ \right\} + x\beta + S(\boldsymbol{\alpha}) \\ &\text{s.t.} (\boldsymbol{\alpha}, \beta) \in \tilde{\mathcal{L}}_{k+1}. \end{aligned} \tag{5.153}$$

For the case that any layer is decoded, Δ_0,

$$\begin{aligned} \Delta_0(\mathbf{r}) = \inf\, & x[(1-\beta)^+ + \beta] + S(\alpha) \\ \text{s.t.}\, & b\sum_{i=1}^{M_*}((1-\alpha_i)^+ - (1 - \frac{r_1 - \epsilon_1}{b} - \alpha_i)^+) < (r_1 - x(1-\beta))^+ \\ & \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0 \\ & 0 \leq \beta \leq 1. \end{aligned} \tag{5.154}$$

The infimum is achieved by and $\beta = 1$ and using, (5.166),

$$\Delta_0(\mathbf{r}) = x + d_{ds}\left(\frac{r_1}{b}\right). \tag{5.155}$$

At layer k, the problem is given by

$$\begin{aligned}
&\Delta_k = \max\{r_1^k, x(1-\beta)\} + x\beta + S(\boldsymbol{\alpha}) \\
&\text{s.t.} b\sum_{i^1}^{M_*}\left(\left(1-\alpha_i-\frac{r_1^k-\epsilon_k}{b}\right)^+ - \left(1-\frac{r_1^{k+1}-\epsilon_{k+1}}{b}-\alpha_i\right)^+\right) \\
&\quad < (r_1^{k+1}-x+\beta)^+ - (r_1^k - x+\beta)^+, \\
&\quad \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0.
\end{aligned} \tag{5.156}$$

If $r_1^k \geq x$, then the infimum is obtained for $\beta = 0$ and

$$\begin{aligned}
&\Delta_k = \max\{r_1^k, x\} + S(\boldsymbol{\alpha}) \\
&\text{s.t.} b\sum_{i^1}^{M_*}\left(\left(1-\alpha_i-\frac{r_1^k-\epsilon_k}{b}\right)^+ - \left(1-\frac{r_1^{k+1}-\epsilon_{k+1}}{b}-\alpha_i\right)^+\right) < r_{k+1}, \quad (5.157) \\
&\quad \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0. \quad (5.158)
\end{aligned}$$

that is a version of (5.166) and solved for

$$\Delta_k = \max\{x, \sum_{i=1}^{k} r_i\} + d(r_{k+1}). \tag{5.159}$$

If $r_1^k \leq x$, the infimum is given by $\beta = \frac{x-r_1^k}{x}$ and again, we have a version of (5.166) with the distortion exponent is

$$\Delta_k = x + d(r_{k+1}). \tag{5.160}$$

At layer L, the distortion exponent is the solution to the problem

$$\begin{aligned}
\Delta_L(\mathbf{r}) = \inf \max &\left\{\sum_{i=1}^{L} r_i, x(1-\beta)^+\right\} + x\beta + S(\boldsymbol{\alpha}) \\
\text{s.t.} &\sum_{i=1}^{M_*} b\left(1-\frac{\sum_{i=1}^{L-1} r_i - \epsilon_L}{b} - \alpha_i\right)^+ \\
\geq &\left(\sum_{i=1}^{L} r_i - (x-\beta)\right)^+ - \left(\sum_{i=1}^{L-1} r_i - (x-\beta)\right)^+, \\
&\alpha_1 \geq ... \geq \alpha_{M_*} \geq 0.
\end{aligned} \tag{5.161}$$

The infimum is achieved for $\alpha_i = 0$ and $\beta = 0$ and is given by

$$\Delta_L = \max\left\{\sum_{i=1}^{L} r_i, x\right\}. \tag{5.162}$$

Finally, gathering all results, and letting all $\epsilon_k \to 0$

$$\Delta_0 = x + d_{ds}(\frac{r_1}{b}),$$
$$\Delta_k = \max\left\{x, \sum_{i=1}^{k+1} r_i\right\} + d_{ds}\left(\frac{r_{k+1}}{b}\right),$$
$$\Delta_L = \max\left\{\sum_{i=1}^{L} r_i, x\right\}. \quad (5.163)$$

It is easy to see that by the scaling $r_i = br'_i$, the problem is equivalent to the multilayer scheme without binning and thus have the same distortion exponent, i.e. $\Delta^L_{mj} = \Delta^L_{ml}$.

Appendix 5.J Proof for distortion exponent: multi-layers NBJD with improved power allocation

We consider a power allocation satisfying

$$\bar{\rho}_k = \rho^{\gamma_{k-1}} - \rho^{\gamma_k}. \quad (5.164)$$

with $\gamma_k \geq 0$ a decreasing sequence with $\gamma_0 = 1$. Following the usual arguments, we have that the exponential behavior of the expected distortion ED_{ml} is given by

$$ED_{ml} \doteq \sum_{k=0}^{L} \rho^{-d_{sd}(r_{k+1},\gamma_k,\gamma_{k+1})}\rho^{-\Delta^d_k} \doteq \sum_{k=0}^{L} \rho^{-\Delta_k}. \quad (5.165)$$

where $\Delta^d_k \triangleq \max\{x, \sum_{i=0}^{k} r_i\}$ is the equivalent distortion exponent for the averaged distortion and the successive decoding diversity gain, defined in [2], is the solution to the successive probability of outage at each layer, i.e.

$$\begin{aligned} d_{ds}(r_k,\gamma_{k-1},\gamma_k) \;=\; & \inf S(\boldsymbol{\alpha}) \\ & \text{s.t. } \sum_{i=1}^{M_*}(\gamma_k - \alpha_i)^+ - \sum_{i=1}^{M_*}(\gamma_{k-1} - \alpha_i)^+ < r_k, \\ & \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0. \end{aligned} \quad (5.166)$$

Consider that rate at each layer k, $0 \leq r_k \leq M_*$ given by $r_k = l(\gamma_{k-1} - \gamma_{\gamma_k}) + \delta$ where $l \in [0, 1, ..., M_* - 1]$ and $0 \leq \delta < \gamma_{k-1} - \gamma_k$. Then, the infimum for 5.166 is explicitly given by

$$\alpha_i = \begin{cases} \gamma_{k-1}, & 1 \leq i < M_* - l, \\ \gamma_{k-1} - \delta, & i = M_* - l, \\ 0, & M_* - l < i \leq M_*. \end{cases} \quad (5.167)$$

and is given by

$$d_{ds}(r_k,\gamma_{k-1},\gamma_k) = (M^* - l)(M_* - l)\gamma_{k-1} - (M^* + M_* - 1 - 2k)\delta. \quad (5.168)$$

The exponent distortion is given by the minimum of each layer exponent Δ_k, i.e.

$$\Delta_{ml}^{L} = \min_{0\le k\le L} \left\{ d_{sd}(r_{k+1},\gamma_k,\gamma_{k+1}) + \Delta_k^d \right\}. \tag{5.169}$$

If $x \ge br_1$, we have $\Delta_k^d = x$ for all k. The minimum exponent is then given by $\Delta_L = x$. If $x \le br_1 x$, we have that $\Delta_k^d = b\sum_{i=1}^{k} r_i$ for all i. The system becomes

$$\begin{aligned}\Delta_0 &= x + d_{sd}(r_1,\gamma_0,\gamma_1),\\ \Delta_k &= b\sum_{i=1}^{k} r_i + d_{sd}(r_{k+1},\gamma_k,\gamma_{k+1}),\\ &\quad \text{for } \ k = 1\dots L-1,\\ \Delta_L &= b\sum_{i=1}^{L} r_i.\end{aligned} \tag{5.170}$$

Now we design the rates r_k such that the minimum distortion exponent is maximized. Note that Δ_0 and Δ_k are decreasing in r_k while Δ_L is increasing.

In the following, we fix $r_k = (l+1)(\gamma_{k-1}-\gamma_k) - \epsilon$, $\epsilon > 0$ and γ_k arbitrary. By equaling all distortion exponents Δ_k for $k = 1, ..., L-1$ equal, $\Delta_k = \Delta_{k+1}$, we have

$$d_{sd}(r_{k'},\gamma_{k'-1},\gamma_{k'}) = br_{k'} + d_{sd}(r_{k'+1},\gamma_{k'},\gamma_{k'+1}) \tag{5.171}$$

where $k' = k+1$ for notation. Since $r_k' = (l+1)(\gamma_{k'-1}-\gamma_{k'}) - \epsilon$, we have

$$\begin{aligned}d_{sd}(r_{k'},\gamma_{k'-1},\gamma_{k'}) = {}&(M^*-l)(M_*-l)\gamma_{k'-1}\\ &-(M^*+M_*-1-2k')(\gamma_{k'-1}-\gamma_k'-\epsilon).\end{aligned} \tag{5.172}$$

Substituting in (5.171) we obtain that the power allocations have to satisfy,

$$(\gamma_{k'}-\gamma_{k'+1}) = \eta_l(\gamma_{k'-1}-\gamma_{k'}) + \epsilon' \tag{5.173}$$

where $epsilon' \to 0$ when $\epsilon \to 0$, and we define

$$\eta_l \triangleq \frac{b(l+1)-(M^*-l-1)(M_*-l-1)}{M^*+M_*-1-2k'}. \tag{5.174}$$

Then, we have have the following recursion

$$\gamma_k' - \gamma_{k'+1} = \eta_l^{k'-1}(\gamma_1-\gamma_2) + \epsilon'. \tag{5.175}$$

Finally, we have that for $k = 2, ..., L-1$ the power allocation has to satisfy

$$\gamma_k - \gamma_{k+1} = \eta_l^{k-1}(\gamma_1-\gamma_2). \tag{5.176}$$

Now, from $\Delta_0 = \Delta_1$, we have that

$$\begin{aligned}&(\gamma_1-\gamma_2) =\\ &\frac{(b(1+k)-(M^*-1-k)(M_*-1-k))(\gamma_0-\gamma_1)-x}{M^*+M_*-1-2k}\end{aligned} \tag{5.177}$$

Form $\Delta_L = b\sum_{i=1}^{L} r_i = b\sum_{i=1}^{L}(l+1)(\gamma_{i-1}-\gamma_i))$ we have

$$\begin{aligned} & b(\gamma_0-\gamma_1)+b(\gamma_2-\gamma_1)\sum_{i=1}^{L}\eta_l^{i-1} \\ &= b(\gamma_0-\gamma_1)+b(\gamma_2-\gamma_1)\frac{1-\eta_l^L}{1-\eta_l}, \end{aligned} \tag{5.178}$$

if $\eta_l \leq 1$. Let

$$\Gamma \triangleq \frac{1-\eta^{L-1}}{1-\eta}. \tag{5.179}$$

From $\Delta_0 = \Delta_L$

$$(\gamma_0-\gamma_1) = \frac{(M^*-k)(M_*-k)+x-b(\gamma_2-\gamma_1)\Gamma}{-1+b-2k+M^*+M_*}. \tag{5.180}$$

From $\Delta_1 = \Delta_L$

$$(\gamma_2-\gamma_1) = \frac{(bk+(M^*-k)(M_*-k))(\gamma_0-\gamma_1)}{-1-2k+M^*+M_*+b\Gamma}. \tag{5.181}$$

Finally, putting all the results together, the distortion exponent Δ_{ml}^L is given by

$$\Delta_{ml2}^L = b(1+k)\frac{b(1+k)\Gamma(x+\Phi)+x(\Upsilon-\Gamma\Phi)+\Phi(\ \Upsilon+\Gamma\Upsilon-\Gamma\Upsilon)}{(b+bk+\Upsilon)(b(1+k)\Gamma+\Upsilon)-b(1+k)\Gamma\Phi} \tag{5.182}$$

where

$$\begin{aligned} \Phi &= (M1-k)(M2-k), \\ \Upsilon &= (M1+M2-1-2k). \end{aligned} \tag{5.183}$$

for the power allocation satisfying

$$(\gamma_2-\gamma 1) = \frac{(b+bk-x+\Upsilon-\Phi)\Phi}{(b+bk+\Upsilon)(b(1+k)\Gamma+\Upsilon)-b(1+k)\Gamma\Phi}, \tag{5.184}$$

and

$$\gamma_1 = \frac{(b(1+k)\Gamma+\Upsilon)(b+bk-x+\Upsilon-\Phi)}{(b+bk+\Upsilon)(b(1+k)\Gamma+\Upsilon)-b(1+k)\Gamma\Phi}. \tag{5.185}$$

and

$$\gamma_k-\gamma_{k+1} = \eta_l^{k-1}(\gamma_1-\gamma_2) fork = 2,...,L. \tag{5.186}$$

We omit the system resolution as it is a classical linear system. For $0 \leq \eta_k \leq 1$, i. e.

$$b \in [\frac{(M^*-l-1)(M^*-l-1)+x}{l+1}, \frac{(M^*-l)(M^*-l)+x}{l+1}). \tag{5.187}$$

The power allocation is valid. When

$$b \in [\frac{(M^*-l)(M^*-l)+x}{l+1}, \frac{(M^*-l)(M^*-l)+x}{l}). \tag{5.188}$$

we have that $\eta_k > 1$. In this region, a power allocation similar to the done previously by equating all power allocations in $k = 1, ..., L$ can be done.

In the continuum infinity of layers, this scheme can be proven to converge in

$$b \in [\frac{(M^* - l - 1)(M^* - l - 1) + x}{l + 1}, \frac{(M^* - l)(M^* - l) + x}{l + 1}). \tag{5.189}$$

to $\Delta_{ml2}(b, x)^\infty = b(l + 1)$.

When

$$b \in [\frac{(M^* - l - 1)(M^* - l - 1) + x}{l + 1}, \frac{(M^* - l)(M^* - l) + x}{l}). \tag{5.190}$$

It can be proven that it converges to

$$\Delta_{ml2}(b, x)^\infty = (M^* - l)(M_* - l) + x \left(\frac{b - (M_* - l)(M^* - l)}{b - ((M^* - l) - 1)((M_* - l) - 1)} \right) \tag{5.191}$$

Chapter 6

Conclusions

We have studied the transmission of a Gaussian source over a slow fading channel in the presence of fading side information at the receiver for MIMO systems. The performance has been studied by means of the achievable expected distortion under the assumption that the channel and side information states are available only at the receiver.

We have derived the expected distortion function for two separation based schemes, with and without binning, as well as uncoded transmission. Then, we have presented a joint source-channel coding technique, named NBJD, that does no explicit binning at the transmitter and applies joint decoding at the receiver. We have proposed a hybrid NBJD-analog scheme and a multiple-layer scheme with two different power allocations. We have also studied the joint source-channel coding scheme of HDA with single layer. In addition to these achievable schemes, we have provided a lower bound on the expected distortion derived by assuming the availability of channel and side information states at the transmitter. We have studied numerically the SISO scenario by exhaustive search over the design parameters in the finite SNR regime. In the high SNR regime, closed-form expressions for the distortion exponent are obtained for the proposed schemes as well as the upper bound.

Interestingly, the presented schemes do not achieve the distortion exponent upper bound as opposed to the scenario without side information. This might be due to the looseness of the informed transmitter upper bound in the scenario, or it might point to the need of better achievable techniques.

The following conclusions can be derived from the numerical results and the distortion exponent analysis in this thesis:

- We have seen that separate source-channel coding with binning is optimized by ignoring the side information when encoding, (i.e., no binning), and hence coinciding with the non binning case, in the numerical simulation for the SISO case. While we have been not able to give an analytical proof for this result in the finite SNR regime, we have proven that distortion exponent is maximized without binning.
- We have proved that the NBJD scheme outperforms the separation approach analytically in all bandwidth regimes. The numerical results reveal a constant performance improvement in the high SNR regime. We have proven that NBJD has the same distortion exponent performance as separation despite its improvement in terms of the expected distortion in the finite SNR regime, revealing that the improvement is not exponential, matching with the numerical result.
- Using the high SNR results, we have proved that hybrid-NBJD, HDA and multi-layer

NBJD scheme in the limit of a continuum of layers achieve the same distortion exponent performance in a SISO scenario.On the other hand, the numerical results in the finite SNR regime show that hybrid analog-NBJD and HDA schemes have very similar performances and they both perform better than NBJD as they benefit from the robustness provided by the analog transmission in the finite SNR regime.

- The equality of the distortion exponents of theses techniques does not extend to MIMO scenarios. In the case of MIMO, NBJD multi-layer scheme achieves the higher distortion exponents for large b regimes while the NBJD-analog hybrid scheme achieves higher distortion exponents in the low bandwidth regime. This is due to the limitations of analog transmission. The proposed HDA MIMO technique achieves a distortion exponent that is very close to the hybrid analog-NBJD performance. A better construction of the MIMO HDA should provide the same performance as the hybrid scheme.

Note that HDA scheme naturally combines the robustness of analog transmission with the higher transmission capability of digital transmission just with a single layer. Based on this property of HDA and the fact that it achieves the best distortion exponent among all single layer schemes, future research line aims at creating a multi-layer extension of the HDA scheme that can potentially improve the pure NBJD multi-layer scheme, which achieves the best known distortion exponent at the moment.

List of Figures

2.1 Block diagram of the joint source-channel coding problem with fading channel and side information qualities. 6
2.2 ED function for $\rho = 10dB$, with $b = 2$, $x = 1$ for many schemes presented in Chapter 3. 9

4.1 Numerical Expected Distortion function at $\rho_n = 10dB$. 32
4.2 Numerical Expected Distortion function at $\rho_n = 10dB$ for Separation with binning scheme. 32
4.3 ED^* for $b = 0.8$ and $x = 0.3$. 33
4.4 ED^* for $b = 2$ and $x = 0.8$ 33

5.1 Distortion exponents in function of b for $x = 0.5$ for a MIMO 2×2 system. . 39
5.2 Distortion exponents in function of b for $x = 0.5$ for a MISO 4×1 system. . . 40
5.3 Distortion exponents of the multi-layer schemes in function of b for $x = 0$ in a MIMO 3×3 system. 41
5.4 Distortion exponents of the multi-layer schemes in function of b for $x = 1$ in a MIMO 3×3 system. 41
5.5 Distortion exponents in function of b for $x = 0.4$ for a SISO system. 43

Bibliography

[1] T. J. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Information Theory*, vol. 11, no. 11, pp. 558–567, November 1965.

[2] D. Gündüz and E. Erkip, "Joint source–channel codes for MIMO block-fading channels," *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 116–134, 2008.

[3] C. Ng, D. Gündüz, A. Goldsmith, and E. Erkip, "Distortion minimization in Gaussian layered broadcast coding with successive refinement," vol. 55, no. 11, Nov., pp. 5074–5086.

[4] G. Caire and K. Narayanan, "On the distortion SNR exponent of hybrid digital–analog space–time coding," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2867–2878, Aug. 2007.

[5] C. Ng, C. Tian, A. Goldsmith, and S. Shamai, "Minimum expected distortion in Gaussian source coding with uncertain side information," in *IEEE Inf. Theory Workshop, 2007. ITW'07*, Sep. 2007, pp. 454–459.

[6] M. Wilson, K. Narayanan, and G. Caire, "Joint source channel coding with side information using hybrid digital analog codes," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 4922–4940, 2010.

[7] A. Wyner, "The rate-distortion function for source coding with side information at the decoder," *Information and Control*, vol. 38, no. 1, pp. 60–80, Jan. 1978.

[8] E. Tuncel, "Slepian-Wolf coding over broadcast channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1469–1482, Apr. 2006.

[9] A. Lapidoth and S. Tinguely, "Sending a bivariate Gaussian source over a Gaussian mac with feedback," in *Proc. of the IEEE International Symposium on Information Theory*, Nice, France, June 2007.

[10] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May. 2003.

[11] Y. Gao and E. Tuncel, "Wyner-ziv coding over broadcast channels: Hybrid digital/analog schemes," *submitted to IEEE Transactions on Information Theory.*

[12] N. Merhav and S. Shamai, "On joint source–channel coding for the Wyner–Ziv source and the Gelfand–Pinsker channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, Nov. 2003.

[13] U. Mittal and N. Phamdo, "Hybrid digital–analog (HDA) joint source–channel codes for broadcasting and robust communications," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, May. 2002.

[14] A. Lapidoth and S. Tinguely, "Sending a bi-variate Gaussian source over a Gaussian mac," in *Proc. of the IEEE International Symposium on Information Theory*, Seattle, WA, July 2006.

[15] ——, "Sending a bi-variate Gaussian source over a Gaussian MAC," in *2006 IEEE International Symposium on Information Theory*, 2006, pp. 2124–2128.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley-Interscience, 1991.

[17] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[18] D. Tse and P. Viswanath, *Multiantenna Channels: Capacity, Coding and Signal Processing*. DIMACS, American Mathematical Society, 2003, ch. On the capacity of the multiple antenna broadcast channel, pp. 87–105.