Imperial College London

Department of Electrical and Electronic Engineering

Side Information Aware Source and Channel Coding in Wireless Networks

Iñaki Estella Aguerri

September 2014

Supervised by Dr. Deniz Gündüz

Submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy in Electrical and Electronic Engineering of Imperial College London and the Diploma of Imperial College London

Abstract

Signals in communication networks exhibit significant correlation, which can stem from the physical nature of the underlying sources, or can be created within the system. Current layered network architectures, in which, based on Shannon's separation theorem, data is compressed and transmitted over independent bit-pipes, are in general not able to exploit such correlation efficiently. Moreover, this strictly layered architecture was developed for wired networks and ignore the broadcast and highly dynamic nature of the wireless medium, creating a bottleneck in the wireless network design. Technologies that exploit correlated information and go beyond the layered network architecture can become a key feature of future wireless networks, as information theory promises significant gains. In this thesis, we study from an information theoretic perspective, three distinct, yet fundamental, problems involving the availability of correlated information in wireless networks and develop novel communication techniques to exploit it efficiently.

We first look at two joint source-channel coding problems involving the lossy transmission of Gaussian sources in a multi-terminal and a time-varying setting in which correlated side information is present in the network. In these two problems, the optimality of Shannon's separation breaks down and separate source and channel coding is shown to perform poorly compared to the proposed joint source-channel coding designs, which are shown to achieve the optimal performance in some setups. Then, we characterize the capacity of a class of orthogonal relay channels in the presence of channel side information at the destination, and show that joint decoding and compression of the received signal at the relay is required to optimally exploit the available side information.

Our results in these three different scenarios emphasize the benefits of exploiting correlated side information at the destination when designing a communication system, even though the nature of the side information and the performance measure in the three scenarios are quite different.

Declaration

I hereby certify that this thesis is the result of my own work under the guidance of my thesis advisor, Dr. Deniz Gündüz. Any ideas or quotations from the work of other people are appropriately referenced.

Iñaki Estella Aguerri
Imperial College London
15 September 2014

Acknowledgments

First of all, I would like to express my deepest gratitude to my supervisor, Dr. Deniz Gündüz, for believing in my abilities and guiding me throughout all these years together. From the very beginning, not only has he been a source of inspiration and encouragement, but a friend and supporter, in both my good and bad moments. Along with him, I have learnt that research is full of unexpected subtleties and that meaningful results are much harder to obtain than I thought before. But, above all, I have discovered the joy and excitement when, little by little, the final goal becomes, at least, one step closer. I am proud and happy to have gone over all this process with Deniz as my guide.

I would also like to thank my examiners Dr. Pier Luigi Dragotti and Dr. Pablo Piantanida for carefully reading this thesis, and for the challenging and encouraging discussion we had during the thesis defense. I really appreciate the comments and suggestions they provided, which improved the exposition and clarity of this work.

I am also very thankful to my friends from Imperial College London and outside the college, who made my life in London more joyful and enriching. They have offered me a healthy, productive and enjoyable environment, that has extended in many cases far beyond the research. Nothing but thank them for the coffee breaks, the motivating conversations, the inspirational Union nights we shared, the exhausting walks, the noisy BBQs, as well as all many unforgettable experiences in London.

I could not forget my friends and colleagues from Centre Tecnològic de Telecomunicomunicacions de Catalunya (CTTC) where I spent the beginning of this Ph.D. During all the time we spent together in Barcelona, they have been more than friends, but a family. From there, I also feel deeply in debt with Miguel Ángel Lagunas, for inspiring me during my undergraduate studies and motivating me to pursue the Ph.D. degree. His engaging classes made me passionate about research, always eager to learn more.

I would like to thank all my friends, some of them new, some of them very old, and flatmates I have lived with. During all this years, they have become many times my home and shelter, and together we have shared the most exciting moments.

Finally, I would like to thank my parents, Gerardo and Isabel, my brother, Albert, and Javi for their love, encouragement, never-ending patience, unconditional support and belief in me throughout all my life; without them this work could not have been possible. I dedicate this thesis to all of them, for everything they have done for me. They were always there in hard times and they were the first to share happy moments with.

Well, that is the beginning of many new things. Many many thanks to everyone!

Contents

List of Figures

Notation

Throughout this dissertation, we will use the following notation. We denote random variables with upper-case letters, e.g., X, their realizations with lower-case letters, e.g., x, and the sets with calligraphic letters $\mathcal{A}$, with cardinality $|\mathcal{A}|$. We denote by $p_X(x)$ the distribution of the random variable X taking realizations x over the set $\mathcal{X}$. We might use $p(x)$ to refer to $p_X(x)$ when there is no ambiguity.

For sequences, we denote sequences of n random variables as $X^n \triangleq (X_1, ..., X_n)$, and denote the i-th term as X_i. We denote partial sequences of X^n as $X_i^j \triangleq (X_i, X_{i+1}, ..., X_j)$ for $i < j$, $X_{n+1}^n \triangleq \emptyset$, and $X^{n\setminus i} \triangleq (X_1, ..., X_{i-1}, X_{i+1}, ..., X_n)$.

We denote random matrix column vectors by $\mathbf{X}$ with realizations $\mathbf{x}$. Similarly, for random matrix we use X and denote its realizations by $\mathbf{X}$. Yet using the same notation, the difference between random vector variable and random matrix realization will be clear from the context. Sequences of random variables and random matrix are defined as for the scalar case. We denote the transpose of a matrix $\mathbf{A}$ as $\mathbf{A}^T$, the conjugate transpose as $\mathbf{A}^H$, the trace as $\mathrm{Tr}\{\mathbf{A}\}$ and the determinant as $|\mathbf{A}|$ or $\det(\mathbf{A})$.

We denote $\mathrm{E}_X[\cdot]$ as the expectation with respect to X, and $\mathrm{E}_{\mathcal{A}}[\cdot]$ as the expectation over the set $\mathcal{A}$. We denote by $\mathbb{R}_n^+$ the set of positive real numbers, and by $\mathbb{R}_n^{++}$ the set of strictly positive real numbers in $\mathbb{R}^n$, respectively. We define $(x)^+ = \max\{0, x\}$ and $\log^+(x) = \max\{\log(x), 0\}$.

Given two functions $f(x)$ and $g(x)$, we use $f(x) \doteq g(x)$ to denote the exponential equality $\lim_{x\to\infty} \frac{\log f(x)}{\log g(x)} = 1$, while $\dot{\geq}$ and $\dot{\leq}$ are defined similarly.

We use $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ to denote a real multivariate Gaussian random variable with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ and use $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$ to denote circular symmetric complex multivariate Gaussian random variables with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$.

In general, in optimization problems, we will denote the optimal variable with a star, e.g., x^* if the optimization is over x.

Chapter 1

Introduction

Wireless communication has become the ubiquitous means of information transfer, ranging across machine-to-machine (M2M), Wi-Fi and cellular networks. An unprecedented mobile network traffic growth is foreseen within the next decade; whereby billions of connected devices will exchange massive amounts of data from applications, cloud services and multimedia content providers. In just a decade, the amout of data handled by wireless networks is expected to increase by more than a factor of 100: surging from under 3 exabites in 2010 to over 190 exabites in 2018, on pace to exceed 500 exabites by 2020 [1]. In addition to the traffic volume, in the next decade the number of devices and the data rates will continue to grow exponentially, reaching the tens of even hundreds or billions of devices and aggregate rates of 1000x with respect to the current 4G mobile networks. This traffic demand cannot be supported within the capacity of current wireless networks, and novel techniques and system architectures are required to accommodate the increasing traffic load.

Both industry and academia are intensively engaged in developing disruptive solutions to provide the network with sufficient capabilities to satisfy the expected demands. Although many challenges and requirements are still to be addressed, the key technologies envisioned to overcome the foreseen capacity crunch are the following: i) extreme network densification, that is, to significantly increase the number of access points per unit area; ii) increased bandwidth using new spectrum bands, such as millimeter wave bands or WiFi's unlicensed bands at 5GHz; and iii) increased spectral efficiency, by employing advanced communication schemes and nodes with multiple antennas [2]. Fully exploiting the potential of these technologies will require a complete understanding of the fundamental challenges and opportunities, and a revision of the current system design paradigm.

Information theory has been instructive in the design of communications networks since its origin in 1948, when Shannon settled the fundamental principles of reliable digital communication and data recording in his groundbreaking paper [3], and showed the optimality of decoupling the communication problem in point-to-point links into two separate simpler problems: data compression and channel communication.

Following the insights of information theory, network system design has traditionally followed the division of data transmission between nodes into independent layers: in the Application Layer, data is compressed into bits, which are transmitted over non-interfering bit-pipes by the Physical Layer, within a certain error probability. This separate operation framework presents significant advantages, such as simplifying the network design and providing a common transmission structure for all types of data and communication channels independently of their nature, which lead to the development of highly complex wired networks such as the Internet.

However, this approach has created a bottleneck in the design of wireless technologies. Communication over wireless networks differs significantly from wired communication: unlike wired channels, where the channel is time-invariant, wireless channels are *highly dynamic and unreliable* due to the particular propagation physics of the wireless medium and the potential user mobility. More importantly, the bit-pipe approach ignores the *broadcast nature* of the wireless medium, which generates interference in environments with many users competing for the limited network resources.

Additionally, both sources and channels in communication networks exhibit significant statistical correlations. Signal correlation can stem from the *physical nature* of the underlying sources. For example, in M2M and sensor networks signals from nearby devices, such as temperature measurements, or traffic logs at different routers in a network, show common statistical properties. Besides, signal correlation might also be *created* within the network. For instance, as densification increases in cellular networks, signals received at nearby nodes become highly correlated. In general, current layered system architectures ignore signal correlations and do not effectively exploit them. Nevertheless, communication technologies that exploit correlated information can become a key feature of future high performance networks, as information theory promises significant gains. In the same way information theory has been fundamental in the development of high performance wired networks, as well as many of the fundamental ideas of existing wireless networks, information theory will undoubtedly lay the foundations of the wireless networks of tomorrow, and this thesis is one step forward in identifying potential novel techniques for next generation wireless networks using information theoretic principles.

1.1 Motivation

From the perspective of information theory, the transmission of digital or analog data between terminals in the network is a joint source-channel coding problem, in which the encoding and decoding strategies have to be jointly designed based on the source and channel properties. While, in general, the design of the optimal transmission schemes is a challenging problem, Shannon's well-known separation theorem [3] reduces the communication problem, in point-to-point time-invariant settings, to the following two separate, simpler and independent problems without losing the end-to-end optimality:

- **Source coding problem**, which focuses on the design of compression and decompression schemes for the source, independent of the channel statistics, that allows to reconstruct the source at the destination within a certain quality.
- **Channel coding problem**, which studies the design of coding and decoding schemes that allow the transmission of data bits with vanishing error probability over noisy channels, independently of the source statistics.

This decoupled approach, in which source and channels statistical properties are dealt with independently as source and channel coding problems, was extended to multi-terminal scenarios soon after information theory was born. The source coding problem was extended to a wide range of setups such as source coding with side information [4, 5], distributed compression in multi-user scenarios [6, 7], or multiple descriptions coding [8, 9]. From the channel point of view, the effects of multi-user interference, as well as the benefits of coordination and availability of feedback, were considered in the basic network units, such as the broadcast channel (BC), the multiple access channel (MAC), the relay channel, as well as the interference channel (IC) [10].

Inspired by Shannnon's result, separation theorem became the cornerstone of today's communication system design and lead to the popular layered approach. However, the optimality of separation does not necessarily generalize to all networks. Information theory indicates that this approach can be strictly suboptimal in the wireless context. Indeed, the optimality of separation breaks down in:

1. **multi-terminal networks**. The separation theorem is valid for point-to-point communication channels and does not extend directly to networks, as first shown by Shannon in [11]. See [12], [13] and [14] for examples proving the suboptimality of separate source and channel coding in multi-user systems.
2. the presence of **channel fading** and applications with **delay limitations**. Separation theorem is limited to ergodic sources and channels. If slow fading or delay restrictions are present, ergodicity is lost; and separation fails. See [15], [16] and [17] for examples in which source-channel separation is not optimal in the wireless setting.
3. in **complexity-limited systems**. The separation theorem is proven assuming asymptotically long source and channel codes and it is based on the assumption of infinite complexity and delay. Therefore, it does not apply to practical systems. Recently the finite block-length regime has been studied in [18] and [19], while

zero-delay transmission schemes have been considered in [20–23].

While assumptions 1, 2 and 3 are reasonable for wired networks, they are, in general, unrealistic for wireless networks. Therefore, novel communication schemes that go beyond separate source and channel coding and jointly exploit the correlated information present in the network are required in multi-terminal networks, fading channels and delay constrained systems.

1.1.1 Beyond separate source and channel coding

Separation theorem plays an important role in this thesis. In this section, we present the formalization of Shannon's point-to-point communication setup and the optimal solution given by the separation theorem. While in general separate source and channel coding is seen as the unique optimal communication strategy, from a joint source-channel coding (JSCC) perspective, there are many alternative transmission schemes achieving the optimal performance, each with its benefits and drawbacks. Although all these schemes are just alternatives in the point-to-point setup, when the optimality of separation breaks down, these schemes may become more attractive, and sometimes significantly superior.

The point-to-point communication problem

The original point-to-point communication problem studied by Shannon in his seminal paper [3] consists of five fundamental parts: a stochastic *informative source* to be transmitted, a *transmitter*, a memoryless *channel*, a *receiver* and a *destination*. Informally, the communication problem can be stated as follows: given a certain channel and transmitter resources (e.g., power), which is the highest quality at which the source can be transmitted over the channel and reconstructed at the destination? The task is to characterize the optimal reconstruction distortion (under some given distortion measure), given the source and the channel properties, and design the transmitter and receiver strategies that achieve it.

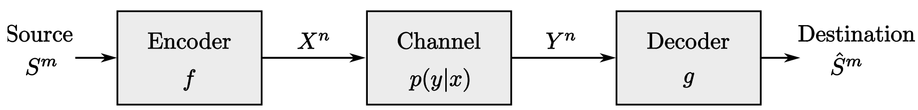


Figure 1.1: The point-to-point communication problem.

More formally, the communication problem is formulated as follows[1] (see Figure 1.1). A source sequence of m i.i.d. symbols $\{S_i\}_{i=1}^m$ from an alphabet $\mathcal{S}$ with distribution $p(s)$ has to be transmitted over a discrete memoryless channel (DMC), characterized by the

[1]While more general models can be considered, we restrict our attention to independent identically distributed (i.i.d.) sources, discrete memoryless channels and single letter distortion measures for ease of exposition.

conditional distribution $p(y|x)$, using n channel accesses. The channel input, given by X^n, and the channel output, given by Y^n, are from alphabets $|\mathcal{X}|$ and $|\mathcal{Y}|$, respectively. The transmitter maps the source block S^m to a channel input X^n using the encoding mapping $f^{(m,n)} : |\mathcal{S}|^m \to |\mathcal{X}|^n$. It is assumed that the channel input satisfies an input cost constraint given by $\mathrm{E}[\frac{1}{n}\sum_{i=1}^{n} c(X_i)] \leq P$, where $c(\cdot) : |\mathcal{X}| \to \mathbb{R}^+$ and P is the transmitter budget (for example, available power).

At the receiver, the channel output is used to reconstruct the source sequence as $\hat{S}^m$ with entries in $\hat{\mathcal{S}}$ using the receiver mapping $g^{(m,n)} : |\mathcal{Y}|^n \to |\hat{\mathcal{S}}|^m$. The distortion between the source sequence and the reconstruction is calculated using a distortion measure $d^m(S^m, \hat{S}^m) \triangleq \frac{1}{m}\sum_{i=1}^{m} d(S_i, \hat{S}_i)$. The *bandwidth ratio* of the system is defined as the average channel accesses per source sample as follows:

$$b \triangleq \frac{n}{m} \quad \text{channel dimensions per source sample.}$$

In the communication problem, we want to characterize the lowest achievable average distortion. We say that, given $p(s)$, $p(y|x)$, $c(\cdot)$, P and $d(\cdot,\cdot)$, average distortion D is *achievable* if there exists a sequence of encoding and decoding functions $\{f^{(m,n)}, g^{(m,n)}\}$ satisfying the channel constraint such that

$$\mathrm{E}\left[\sum_{i=1}^{m} d(S_i, \hat{S}_i)\right] \leq D, \tag{1.1}$$

for sufficiently large m, n satisfying the bandwidth ratio relation.

The characterization of this problem is in general very complicated. However, Shannon's separation theorem simplifies the problem by dividing it into two simpler parts: the source coding part and the channel coding part, which we describe next.

Source coding problem

In the source coding problem, the channel is substituted by an errorless bit-pipe of rate R. The minimum rate required to reconstruct the source sequence at average distortion D is characterized by the rate-distortion function, given by

$$R(D) \triangleq \min_{p(\hat{s}|s):\mathrm{E}[d(S;\hat{S})]\leq D} I(S;\hat{S}),$$

where $I(X;Y) \triangleq -\mathrm{E}\left[\log \frac{p(x)p(y)}{p(x,y)}\right]$ is defined as the *mutual information* between random variables X and Y.

Conversely, for rates below $R(D)$, distortion D is not achievable; that is, no matter which encoding and decoding method is used, the target distortion cannot be achieved.

The optimal distortion is achievable by the following random coding scheme. Generate a quantization codebook consisting of 2^{mR_s} length-m codewords $\hat{S}^m(i)$, $i = 1, ..., 2^{mR_s}$ with i.i.d. components following the optimal rate-distortion minimizing distribution $p(\hat{s}|s)$. Given a source sequence S^m, the source encoder looks for an index i such that $(S^m, \hat{S}^m(i))$ are jointly typical[2] and provides it to the destination through the errorless bit-pipe, where the source sequence is reconstructed as $\hat{S}^m(i)$.

In particular, for i.i.d. Gaussian[3] sources, $S \sim \mathcal{N}(0, \sigma^2)$ and a quadratic distortion measure $d(S, \hat{S}) = (S - \hat{S})^2$, the rate-distortion function is found as

$$R(D) = \frac{1}{2} \log^+ \left(\frac{\sigma^2}{D} \right), \tag{1.2}$$

and the quantization codewords with i.i.d. components following $p(\hat{s}|s)$ can be generated by a test channel as $\hat{S} = S + Q$, where $Q \sim \mathcal{N}(0, \sigma_Q^2)$ is independent of S and $\sigma_Q^2 \triangleq \sigma^2 D/(\sigma^2 - D)$.

Channel coding problem

In the channel coding part, the source sequence is substituted by a message set $w \in [1, ..., 2^{nR_c}]$ with uniform distribution. The objective is to characterize the largest rate $R_c \triangleq \frac{1}{n} \log M$ such that the probability of error of recovering m at the destination can be made arbitrarily small for sufficiently large n. The probability of error is defined as

$$P_e \triangleq \frac{1}{M} \sum_{w=1}^{M} \Pr\{g(Y^n) \neq w | W = w\}.$$

The maximum rate R_c for which P_e can be made arbitrarily small is characterized by the channel capacity for given cost function $c(\cdot)$, and budget P, defined as

$$C(P) \triangleq \max_{p(x): \mathrm{E}[c(X)] \leq P} I(X; Y).$$

Conversely, for any transmission rate above the channel capacity, the probability of error cannot be made arbitrarily small.

The capacity of a DMC can be achieved by using random coding as follows. We fix the input distribution $p(x)$ which maximizes $C(P)$ and satisfies the input cost constraint. We generate a channel codebook of 2^{nR_c} length-n codewords $X^n(i)$, $i = 1, .., 2^{nR_c}$, with distribution $p(x)$. Given a message w, the encoder transmits $X^n(w)$. At the destination, the receiver looks for the index $\hat{w}$ such that $(X^n(\hat{w}), Y^n)$ are jointly typical. For sufficiently large n, the correct w is successfully decoded as long as the code rate is

[2] We refer the reader to [24] for definitions and properties of typicality and joint typicality and details of the achievable scheme.

[3] Although the typicality arguments do not directly apply to Gaussian distributions due to its continuous alphabet, they can be extended using conventional discretization arguments [25].

below the channel capacity, i.e., $R_c \leq \mathcal{C}(P)$.

For channels with additive white Gaussian noise (AWGN), the channel output is given by

$$Y = X + N,$$

where $N \sim \mathcal{N}(0, \sigma_N^2)$. The input is constrained by an average power constraint, i.e., $c(X) = X^2$. The capacity is found as

$$\mathcal{C}(P) = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_N^2} \right), \tag{1.3}$$

and is achievable by i.i.d. channel codewords with Gaussian entries, $X \sim \mathcal{N}(0, P)$. The ratio between the average transmit power and the channel noise power, P/σ_N^2, is commonly referred to as the signal-to-noise (SNR) ratio.

Separation theorem

Shannon's separation theorem states that the optimal performance in the point-to-point communication system in Figure 1.1 is achievable by concatenating an optimal source code at rate that achieves the distortion-rate function, with an optimal channel code at a rate arbitrarily close to the channel capacity. Hence, reliable communication is feasible if $bR(D) < \mathcal{C}(P)$, and conversely, if for b, P given, D is achievable, then

$$bR(D) \leq \mathcal{C}(P). \tag{1.4}$$

In the Gaussian setup, with an i.i.d. source sequence $\{S_i\}_{i=1}^m \sim \mathcal{N}(0, \sigma^2)$, and an AWGN channel $Y^n = X^n + N^n$, where $N_j \sim \mathcal{N}(0, \sigma_N^2)$, condition (1.4) implies that the achievable distortion, with quadratic distortion measure, satisfies

$$D \geq \frac{\sigma^2}{(1 + P/\sigma_N^2)^b}, \tag{1.5}$$

where the equality is met by the concatenation of the optimal source code achieving (1.2) and the optimal channel code achieving (1.3).

Alternatives to separation

Separate source and channel coding transmission is the most common, and almost the only practically used, approach to achieve the optimal performance in a point-to-point setup. Nevertheless, there are alternative transmission schemes that achieve the optimal performance as well. While in separation the source encoding and the channel encoding are done independently, most of these alternative schemes are JSCC schemes, in the sense that, the channel encoding is not independent of the source statistics.

In the following, we consider some of these schemes. Each of them has benefits and limitations with respect to separation in performance features such as complexity, delay or robustness. Although these schemes are not universally optimal for all sources and channels, they are optimal in the Gaussian setup:

- **Uncoded transmission:** In this simple zero-delay scheme, the encoder transmits sample by sample a scaled version of the source sample, to satisfy the cost constraint. At the receiver, the source is reconstructed by applying an optimal estimator using the channel output. For example, in the Gaussian setup for matched bandwidth ratio, i.e., $b = 1$, the channel input is generated as $X_i = \sqrt{P/\sigma^2}S_i$, $i = 1, ..., n$. At the destination, the channel output is given by $Y_i = \sqrt{P/\sigma^2}S_i + N_i$ and the source sequence is reconstructed using a minimum mean square error (MMSE)[4] estimator, i.e., $\hat{S}_i = \mathrm{E}[S_i|Y_i]$, $i = 1, ..., n$. Surprisingly, this very simple scheme achieves the optimal distortion $D_u \triangleq \sigma^2/(1+P/\sigma_N^2)$ in (1.5) for $b = 1$ [26]. However, the optimality of uncoded transmission is very sensitive to the matching between the source/ channel distributions, input cost constraint, the distortion measure and the source and channel bandwidths [27]. In addition, due to the absence of coding, uncoded transmission is not capable of fully exploiting the degrees-of-freedom available in the system in general, and its optimality breaks down when multiple degrees-of-freedom are available, e.g., in the case of multiple-input multiple-output (MIMO) channels, bandwidth mismatch (i.e., $b \neq 1$) [15,16], or when correlated side information is available at the destination [28]. Despite these limitations, given the prevalence of Gaussian source and channel assumptions in the literature and its simplicity, it has received significant attention from the research community in recent years. In addition, while uncoded transmission in the point-to-point setting is just an alternative to optimal separate source and channel coding, and its main advantages are simplicity and zero delay; surprisingly, it is shown to achieve the optimal performance in various other scenarios, such as Gaussian MAC with correlated Gaussian sources [13, 29], or broadcasting a common source to multiple receivers over Gaussian channels, for which uncoded transmission is the only known optimal transmission scheme. Uncoded transmission has also been shown to outperform the best random JSCC in some setups in the finite blocklength regime [30]. More general versions of zero-delay transmission can be considered. On the one hand, zero-delay schemes that consider non-linear mappings have been shown to outperform linear transmission in certain

[4]Observing vector $\mathbf{A} \sim \mathcal{N}(0, \mathbf{C}_a)$, the MMSE in estimating the Gaussian vector $\mathbf{X} \sim \mathcal{N}(0, \mathbf{C}_x)$ is achieved with a conditional mean estimator, $\hat{\mathbf{X}} = \mathrm{E}[\mathbf{X}|\mathbf{A}]$, and is given by $D_{\mathrm{MMSE}} \triangleq (\mathbf{C}_x + \mathbf{C}_{xa}^H \mathbf{C}_a^{-1} \mathbf{C}_{xa})^{-1}$, where $\mathbf{C}_{xa} \triangleq \mathrm{E}[\mathbf{A}\mathbf{X}^H]$ [25].

scenarios [20–23]. On the other hand, linear schemes that are not zero-delay have been show to outperform zero-delay linear transmission in MAC setups [31].

- **Hybrid digital-analog (HDA) transmission:** Many limitations of uncoded transmission can be overcome by combining it with digital coding schemes in the form of *hybrid digital-analog transmission* (HDA). In HDA, the encoder generates a symbol-by-symbol mapping of the observed source (analog) and its digital compression codeword (digital). For example, for the point-to-point Gaussian setup, a continuum of optimal schemes, including uncoded transmission and separation as special cases, can be created by superposing an uncoded layer and a dirty paper coded digital layer, and optimally allocating the available power among the two. At the receiver, the digital layer is decoded and the source sequence is reconstructed using both the decoded layer and the channel output, which contains the uncoded layer [32]. This scheme combines the robustness of uncoded transmission with the flexibility of coded communication. In many scenarios, such as transmission with bandwidth mismatch [17], or broadcasting with correlated side information [33], HDA transmission is shown to improve the performance of both pure separation and uncoded transmission, and has been shown to outperform separation in cases of mismatched SNR [34] or delay constraint [15, 16]. HDA coding has been also considered for more general setups and multi-user scenarios [35–37].
- **Multi-layer transmission:** Schemes utilizing more layers than HDA schemes can be considered. In general, multi-layer schemes rely on the transmission of multiple layers that carry successive refinement layers of the source [38]. At the receiver, as many layers as possible are decoded depending on the channel quality. The better the channel quality, the more layers can be decoded and the smaller the distortion at the receiver. Multi-layer transmission schemes have been proposed to combat channel fading in the presence of time-varying channels, and have been shown to achieve the optimal performance in some high SNR scenarios [15, 39].

In this thesis, we exploit ideas from these JSCC schemes in order to efficiently exploit the correlated information available in the network, and develop high performance transmission schemes, particularly when the optimality of separation breaks down.

1.2 Objectives

Future networks are expected to be highly heterogeneous, combining several radio technologies and supporting a wide range of applications. This variety will place different performance requirements, which will have to be satisfied by tailoring the network configuration to the needs. This also places new challenges to exploit the correlation information available throughout the network.

In this thesis we study, from an information theoretic perspective, three distinct, yet fundamental, problems involving the availability of correlated information at the network terminals, and develop novel joint source-channel coding communication techniques to exploit it efficiently:

1. **The Helper Problem:** Consider a sensor network that gathers temperature measurements at different points in the network. At the moment of forwarding the collected data from one sensor to the fusion center, another nearby sensor sharing correlated measurements helps in the transmission of the sending sensor to improve the quality at which the data is recovered at the fusion center. We model this scenario as the *one-helper joint source-channel coding problem*, whereby two correlated sources are available at two separate terminals which transmit their observations to the destination over a Gaussian MAC. Of the two sources, the source of interest needs to be reconstructed at the destination with the minimum distortion possible. The second source is correlated with the source of interest, and acts as a helper source. From the source coding perspective, i.e., assuming finite bit-rate pipes from the transmitters to the destination, this problem reduces to the well-known one-helper source coding problem studied in [40]. However, in the presence of a noisy MAC, this is a multi-terminal JSCC problem for which the separation theorem fails. As opposed to separate source and channel coding transmission, schemes based on uncoded transmission and HDA are capable of generating *correlated channel inputs* at the transmitters in a distributed fashion, by exploiting the correlated source information. In this scenario we will study the potential gains through the generation of constructive interferences.
2. **Source-channel coding with time-varying channel and side information:** Consider streaming a high-definition video to a smartphone. Streaming transmission has stringent latency constraints compared to other high-rate applications such as file transfers, or non-critical data collection such as weather measurements. The video is compressed to minimize the amount of data to be downloaded, while the compressed bits are coded against channel uncertainty, all carried out under delay limitations. The more the data is compressed, the less the resolution, nonetheless, the stronger the data can be protected and the higher the probability of reception. Additionally, the decoder might have correlated side information about the video, either coming from previous transmissions or available through relay services. Due to the high variability of the wireless channel and network topology, the side-information available at the destination can vary significantly at different points of the network. We model this uncertainty as a

time-varying correlated side information available at the destination. This is a JSCC problem for which traditional source-channel coding schemes have been observed to suffer from severe outages in the absence of time-varying side information [15, 41]. In the absence of correlated side information, many schemes have been proposed in the literature for this problem, although the characterization of the optimal performance still remains an open problem. See for example [15, 16, 42, 43]. Contrary to single-layer coding schemes, multi-layer schemes that code the signal into different quality layers have been proposed to combat fading. Multi-layer transmission allows adaptation to the current quality of the channel without knowing its realizations. The available correlated side information at the destination provides additional diversity to the system as noisy uncoded versions of the source signal, which can be used to further combat fading. We will study new *coding strategies that jointly adapt to the variations* of the channel and the side information.

3. **A class of orthogonal relay channel with state:** Consider a cognitive network with a relay, in which the transmit signal of the secondary user interferes simultaneously with the received primary user signals at both the relay and the destination. After decoding the secondary user message, the destination obtains information about the interference affecting the source-relay channel, which can be exploited to decode the primary transmitter's message. This setup falls within a class of orthogonal relay channels in the presence of channel side information at the destination. We model the side information in this setting as follows: the source and the relay, and the source and the destination are connected through orthogonal channels that depend on a common state sequence, which is fully known at the destination, and unknown at the source and the relay. Note that this is essentially a channel coding problem in which source compression techniques are required to optimally exploit the side information: fully decoding the source message at the relay renders the side information at the destination useless, whereas compressing the signal received at the relay and forwarding it to the destination will allow the destination to exploit its side information and improve the rate of communications. From a joint source-channel coding perspective, the relay received signal acts a a source sequence which has to be partially transmitted to the destination. We will study the potential of *combining source and channel codes* in multi-user scenarios to exploit the channel state information available at different points of the network.

The nature of the side information in the three scenarios above is quite different. Individual chapters are dedicated to each of them, in which we characterize fundamental performance bounds and propose JSCC schemes that exploit the available correlated side information. Particular emphasis will be put into identifying optimal transmission strategies for the three problems.

1.3 Outline and Contributions

The technical content of this thesis is organized in four chapters. The second chapter is devoted to the Gaussian helper problem, the third and fourth to the transmission with delay constraints under time-varying channel and side-information, while the fifth chapter is dedicated to the orthogonal relay channel model with side information at the destination. In each of the chapters, a literature review and a formal problem statement are provided. Then, performance bounds and achievable schemes are considered, together with some optimality results and discussions. Each chapter finishes with a conclusion. In the following, we outline the content and results of each chapter and the publications related to each topic.

Chapter 2

In Chapter 2, we study the one-helper JSCC problem, in which a main source is to be reconstructed with minimum distortion with the help of a correlated helper source. Focusing on the case of Gaussian sources and a time-invariant Gaussian MAC, we derive a lower bound on the achievable distortion. Then, we consider separate source and channel coding as well as analog transmission, in which each transmitter sends a scaled version of the available source sequence. It is shown that in some regimes analog transmission outperforms separate source and channel coding, and that, in certain cases, it is sufficient to achieve the lowest distortion values among the considered schemes. We also present a hybrid digital-analog scheme, in which each user generates an analog signal in addition to the digital codewords using dirty paper coding, and transmits a superposition of the two. A second hybrid scheme is considered in which each transmitter quantizes the source sequence, and transmits a superposition of the quantized source codeword and an analog component. While different in nature, both schemes are numerically shown to achieve the same performance, and shown to achieve lower distortions than pure digital and analog transmission, in general.

The results in this chapter have been partially published in:

- I. Estella, D. Gündüz, "Hybrid Digital-Analog Transmission for the Gaussian One-helper Problem" , *in Proceedings of IEEE Global Communications Conference, (Globecom), 6-10 December 2010, Miami, Florida, USA.*

Chapter 3

In Chapter 3, the JSCC problem of transmitting a Gaussian source over a time-varying single input single output (SISO) Gaussian channel with minimum average end-to-end distortion is considered in the presence of time-varying correlated side information at the receiver. A block fading model is considered for both the channel and the side information, whose states are assumed to be known only at the receiver. As opposed to the previous chapter, the side information is provided through an orthogonal link to the destination. While separation is optimal with time-invariant channel and side information [28], the delay constraint breaks its optimality in the time-varying setup. However, we show the optimality of separate source and channel coding when the channel is static while the side information state follows a discrete or a continuous and quasiconcave distribution. When both the channel and the side information states are time-varying, separate source and channel coding is suboptimal in general.

A partially informed encoder lower bound is studied by providing the channel state information to the encoder. Several achievable transmission schemes are proposed based on uncoded transmission, separate source and channel coding, joint decoding as well as hybrid digital-analog transmission. Uncoded transmission is shown to be optimal for a class of continuous and quasiconcave side information state distributions, while the channel gain can have an arbitrary distribution. To the best of our knowledge, this is the first example in which the uncoded transmission achieves the optimal performance thanks to the time-varying nature of the states, while it is suboptimal in the static version of the same problem.

Then, we study this problem in the asymptotic SNR regime, in which the channel SNR and the side information quality increase asymptotically. This asymptotic notion is double: when the channel SNR increases in the network, in general the quality of the available side information, e.g., from previous transmissions and relay services, is assumed to increase accordingly. The study of the high SNR regime allows for deeper insights on this problem. In particular, we are interested in characterizing the optimal *distortion exponent*, which quantifies the exponential decay rate of the expected distortion in the high SNR regime. In this chapter, the optimal distortion exponent is characterized for Nakagami distributed channel and side information distributions and it is shown to be achieved by hybrid digital-analog transmission or joint decoding in certain cases, illustrating the suboptimality of pure digital or analog transmission in general. Although the analysis is asymptotic, our results are relevant for practical systems as well, since we observe through numerical simulations that they capture the behavior of the expected distortion at reasonable SNR levels as well.

The results in this chapter have been partially presented at:

- I. Estella, D.Gündüz, “Joint Source-Channel Coding with Time-Varying Channel and Side-Information", *submitted to Transactions on Information Theory.*

- I. Estella, D.Gündüz, "Distortion Exponent with Side-Information Diversity", *in Proceedings 2013 IEEE Global Conference on Signal and Information Processing (GlobalSIP), 3-5 December 2013, Austin, Texas.*
- I. Estella, D.Gündüz, "Systematic Lossy Source Transmission over Gaussian Time-Varying Channels", *in Proceedings IEEE International Symposium on Information Theory (ISIT), 7-12 July 2013, Istanbul, Turkey.*
- I. Estella, D. Gündüz, "Expected Distortion with Fading Channel and Side Information quality", *in Proceeding of IEEE International Conference on Communications (ICC), 5 June 2011, Kyoto, Japan.*

Chapter 4

In Chapter 4, the block-fading time-varying channel and side information setup of the previous chapter is generalized to include a MIMO fading channel and non-matched source and channel bandwidth, that is, the many source samples are transmitted per channel access (or conversely, more than one channel use is employed per source sample). The side information fading gain is assumed to have a Rayleigh distribution. In particular, the high SNR performance is studied by deriving the distortion exponent of various transmission schemes. Following similar techniques from the previous chapter, we derive upper bounds on the distortion exponent. Then we consider transmission schemes based on separate source and channel coding, uncoded transmission, joint decoding as well as hybrid digital-analog transmission. Multi-layer schemes, which transmit successive refinement layers of the source, are also proposed, based on progressive transmission or superposed transmission with joint decoding. While the optimal transmission strategy remains open for finite SNR values, we characterize the optimal distortion exponent for the single-input multiple-output (SIMO) and multiple-input single-output (MISO) by showing that the distortion exponent achieved by the multi-layer superpositon encoding with joint decoding meets the upper bound. In the MIMO scenario, the optimal distortion exponent is characterized in the low bandwidth expansion regime, and it is shown that the multi-layer superposition encoding performs very close to the upper bound in the high bandwidth expansion regime as well.

The results in this chapter have been partially published in:

- I. Estella, D.Gündüz, "Distortion Exponent in Fading MIMO Channels with Time-Varying Side Information", *submitted to Transactions on Information Theory.*
- I. Estella, D. Gündüz, "Distortion Exponent in Fading MIMO Channels with Time-varying Side Information", *in Proceedings IEEE International Symposium on Information Theory (ISIT), 31-5 August 2011, Saint Petersburg, Russia.*
- I. Estella, D. Gündüz, "Wireless Source Transmission with Time-varying Side

Information", *in Proceedings of 9th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 9-13 May 2011, Princeton, New Jersey, USA.*

Chapter 5

In Chapter 5, we consider a channel coding problem. We study the class of orthogonal relay channels in which the orthogonal channels connecting the source terminal to the relay and the destination depend on a state sequence. It is assumed that the state sequence is fully known at the destination while it is not known at the source or the relay. We study the performance of partial decode-and-forward (pDF), in which the relay decodes part of the source message from the received signal, reencodes it and forwards it to the destination; compress-and-forward (CF), in which the relay forwards a compressed version of the channel output to the destination, and partial decode-compress-and-forward (pDCF), that combines the two previous schemes. The capacity of this class of relay channels is characterized, and shown to be achieved by the pDCF scheme. Then the capacity of certain binary and Gaussian state-dependent orthogonal relay channels are studied in detail, and it is shown that the CF and pDF schemes are suboptimal in general. To the best of our knowledge, this is the first single relay channel model for which the capacity is achieved by pDCF, while pDF and CF schemes are both suboptimal. Furthermore, it is shown that the capacity of the considered orthogonal state-dependent relay channels is in general below the cut-set bound. The conditions under which pDF or CF suffices to meet the cut-set bound, and hence, achieve the capacity, are also derived.

The results in this chapter have been partially published in:

- I. Estella, D.Gündüz, "Capacity of a Class of Relay Channels with State", *submitted to Transactions on Information Theory.*
- I. Estella, D.Gündüz, "Capacity of a Class of Relay Channels with State", *in Proceedings 2012 IEEE Information Theory Workshop (ITW), 3-7 September 2012, Lausanne, Switzerland.*

Finally, in Chapter 6 we provide some discussions and conclusions arising from the research results in this thesis. We also indicate some future research directions in which the use of JSCC transmission can be beneficial.

Other Publications

In addition to the topics covered in this dissertation, other research areas have been addressed during the period of the Ph.D. studies. The resulting publications are listed below.

- I. Estella, A. Pascual-Iserte, M. Payaró, “OFDM and FBMC performance comparison for multistream MIMO systems”, *in Proceedings 2010 Future Network and Mobile Summit, 16 - 18 June 2010, Florence, Italy.*
- I. Estella, D.Gündüz, “Linear Transmission of Correlated Gaussian Sources over MIMO Channels”, *in Proceedings The Tenth International Symposium on Wireless Communication Systems (ISWCS), 27-30 August 2013, Ilmenau, Germany.*
- I. Estella, D.Gündüz, “Distortion exponent with side-information diversity”, *in Proceedings 2013 IEEE Global Conference on Signal and Information Processing (GlobalSIP), 3-5 December 2013, Austin, Texas, USA.*
- I. Estella, M. Varasteh, D.Gündüz, “Zero-delay joint source-channel coding”, *in Proceedings 2014 IEEE Iran Workshop on Communication and Information Theory (IWCIT), 7-8 May 2014, Tehran, Iran.*

Chapter 2

One-Helper Joint Source-Channel Coding

In this chapter, we consider the Gaussian one-helper problem introduced in Section 1.2, in which two correlated Gaussian sources, S_1 and S_2 are available at two separate terminals, which transmit their observations to the destination over a Gaussian MAC (see Fig. 2.1). Of the two sources, S_1 is the source of interest which needs to be decoded at the destination with minimum distortion. The second source S_2 is correlated with S_1 and acts as a helper source. From the source coding perspective, i.e., assuming finite rate bit pipes from the transmitters to the destination, this problem is a special case of the Berger-Tung problem [44], in which two sources are encoded separately, and decoded jointly, to satisfy two distortion criteria, and reduces to the well-known one-helper source coding problem studied in [40]. However, in the presence of a noisy MAC, this is a multi-terminal JSCC problem for which the separation theorem fails.

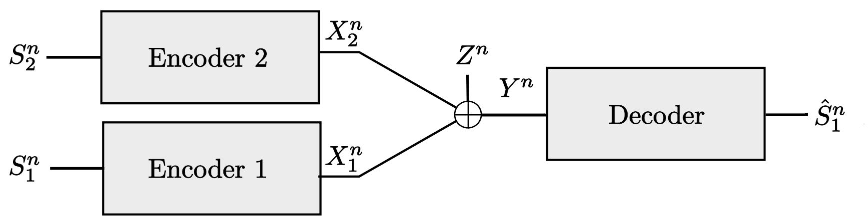


Figure 2.1: Gaussian one-helper problem.

A natural candidate for the JSCC of Gaussian sources over Gaussian channels is analog (uncoded) symbol-by-symbol lineal transmission, or simply SLU, as described in Section 1.1.1. Besides being optimal in the point-to-point Gaussian setup, even in the MAC setting when the destination is interested in both sources, it is shown in [13] that SLU achieves the optimal distortion pair when the SNR is below a certain threshold, which depends on the correlation among the sources. SLU for the one-helper problem is considered in [45] and [46], and it is shown to improve upon separation based digital transmission in some cases, depending on the available power at the users and the source correlation. This proves the suboptimality of SSCC in the Gaussian one-helper scenario.

In several multi-user scenarios, HDA transmission schemes as described in Section 1.1.1 improve the performance in terms of the achievable distortion or robustness to SNR variations (see [17] and [47] for examples). Here, we study an HDA scheme, which we call superposition vector-quantizer (S-VQ), also considered in [13] and [35] for a Gaussian MAC scenario. We consider this transmission scheme in the one-helper scenario. At each encoder, the available source is quantized, and each transmitter sends a superposition of the quantization codeword and an analog layer. The quantization codewords are then jointly decoded at the destination and the main source sequence is reconstructed.

Then, we consider an alternative HDA transmission scheme based on the HDA scheme proposed in [48] for the point-to-point Gaussian setup. In this scheme, the transmitter generates an analog signal and a digital signal using a dirty paper coding scheme [49] considering the interference caused by the analog layer, and, after allocating the available power, transmits a superposition of both signals. Here, we propose a similar HDA transmission scheme, which we call the interference-aware HDA (I-HDA), in the one-helper setting: both users divide their power among the analog and digital signals, and transmit a superposition of the two layers. The pure digital and the pure SLU schemes become two special cases of this HDA transmission strategy, obtained when both users allocate all their power to digital or to analog transmission, respectively, ignoring the interfering signal. We show in [50] that an HDA scheme that ignores the interference caused by the analog layer reduces to either pure SLU and digital transmission depending on the available power or the source correlation. The extra degrees-of-freedom obtained by the dirty paper encoding leads to a better performance. We numerically show that I-HDA achieves the same performance as the S-VQ scheme.

The main contributions of the chapter are the following:

- We derive a lower bound on the achievable distortion using cut-set arguments and by bounding the maximum correlation between the channel inputs of both encoders.
- We derive the minimum distortion achieved by separate source and channel coding.
- We show that when using the optimal symbol-by-symbol linear uncoded transmission, the helper does not use full power in general.
- We propose two different HDA schemes, S-VQ and I-HDA, and numerically show that they outperform the pure separate and analog transmission schemes, in general.

The rest of the chapter is organized as follows. In Section 2.1 we introduce the system model. In Section 2.2 a lower bound is presented. In Section 2.3 we introduce two hybrid digital-analog transmission strategies after reviewing pure digital and analog transmission schemes. Then, in Section 2.4 we provide numerical results, and finally we provide some conclusions in Section 2.5.

2.1 System Model

We consider the transmission of a length-n sequence of i.i.d. zero mean bivariate Gaussian source pair $\{S_{1j}, S_{2j}\}_{j=1}^{n}$ with a covariance matrix

$$\mathbf{K}_{S_1,S_2} = \begin{pmatrix} \sigma_{S_1}^2 & \rho\sqrt{\sigma_{S_1}^2\sigma_{S_2}^2} \\ \rho\sqrt{\sigma_{S_1}^2\sigma_{S_2}^2} & \sigma_{S_2}^2 \end{pmatrix}, \tag{2.1}$$

where $\rho \in [-1, 1]$ is the correlation coefficient, and $0 < \sigma_{S_i}^2 < \infty$ is the variance of the i-th source for $i = 1, 2$. Without loss of generality, we assume $\sigma_{S_1}^2 = \sigma_{S_2}^2 = 1$ and $\rho \in [0, 1]$ as one of the the transmitters can always multiply its source by -1 if $\rho < 0$.

Transmitter i observes the i-th source sequence and encodes it with function f_i^n : $\mathbb{R}^n \to \mathbb{R}^n$, such that $X_i^n = f_i^n(S_i^n)$ for $i = 1, 2$. The corresponding channel input vectors $X_i^n = [X_{i1}, ..., X_{in}]$ are subjected to individual average power constraints

$$\mathrm{E}[|X_i^n|^2] = \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}[|X_{ik}(S_i^n)|^2] \leq P_i, \qquad i = 1, 2. \tag{2.2}$$

The additive memoryless MAC is given by

$$Y_k = X_{1k} + X_{2k} + Z_k, \quad k = 1, ..., n, \tag{2.3}$$

where Z_k is the i.i.d. zero-mean Gaussian noise term with variance N, i.e., $Z_k \sim \mathcal{N}(0, N)$. The decoder consists of a decoding function $g^n : \mathbb{R}^n \to \mathbb{R}^n$, which reconstructs an estimate of the sequence of interest S_1^n, i.e., $\hat{S}_1^n = g^n(Y^n)$.

For a given system $\Omega \triangleq (\rho, P_1, P_2, N)$ we say that an average distortion D is achievable if there exists a sequence of encoding and decoding functions $\{f_1^n, f_2^n, g^n\}$ satisfying the power constraints in (2.2) and a mean square-error distortion of

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\mathrm{E}[(S_{1k} - \hat{S}_{1k})^2] \leq D. \tag{2.4}$$

2.2 Lower Bound

Before deriving the lower bound on the distortion, we provide some definitions that will be used in the derivation. The rate-distortion function of reconstructing $S_1^n \sim \mathcal{N}(0, \sigma^2)$ at distortion D is given, as in (1.2), by

$$R_{S_1}(D) = \min_{P_{W|S_1}:\mathrm{E}[(S_1-W)^2]\leq D} I(S_1;W) = \frac{1}{2}\log^+\left(\frac{\sigma^2}{D}\right). \tag{2.5}$$

For a bivariate Gaussian pair (S_1^n, S_2^n) with variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and correlation coefficient ρ, the conditional rate-distortion function $R_{S_1|S_2}(D_1)$ when S_2^n is available at both the encoder and the decoder, and S_1^n is reconstructed with distortion D_1 is given by

$$R_{S_1|S_2}(D) = \min_{P_{W|S_1S_2}:\mathrm{E}[(S_1-W)^2]\leq D} I(S_1;W|S_2) = \frac{1}{2}\log^+\left(\frac{\sigma^2(1-\rho^2)}{D}\right). \tag{2.6}$$

In the Gaussian helper problem, the encoders can generate physically correlated channel inputs by using the source sequences available at each transmitter. Naturally, the maximum correlation between the channel inputs is bounded by the correlation of the source sequence, as shown in [13] for the MAC setup where both sources have to be reconstructed at the destination. Using this fact together with cut-set arguments, we obtain the following necessary conditions on the achievable distortion.

Lemma 1. *A necessary condition for the achievability of distoriton D in the Gaussian helper problem is the existence of some $0 \leq \rho_x \leq \rho$ such that*

$$\begin{aligned} R_{S_1}(D) &\leq \frac{1}{2}\log\left(1 + \frac{P_1 + P_2 + 2\rho_x\sqrt{P_1P_2}}{N}\right), \\ R_{S_1|S_2}(D) &\leq \frac{1}{2}\log\left(1 + \frac{P_1(1-\rho_x^2)}{N}\right). \end{aligned}$$

Proof. The proof is given in Appendix B. □

Substituting the rate-distortion function expression for the bivariate Gaussian sources, the necessary condition in Lemma 1 can be expressed as the following lower bound.

$$D_l(\Omega) = \min_{0\leq\rho_x\leq\rho} \max\left\{\frac{N}{P_1 + P_2 + 2\rho_x\sqrt{P_1P_2} + N}, \frac{(1-\rho^2)N}{(1-\rho_x^2)P_1 + N}\right\}.$$

2.3 Achievable Schemes

In this section, we first propose "pure" schemes based on separate source and channel coding and uncoded transmission. Then, we consider two HDA schemes.

2.3.1 Pure Coding Schemes

Pure Digital Scheme

Pure digital transmission is based on separate source and channel coding. The sources are first compressed using one-helper source compression as in [40], then the compressed bits are transmitted over the MAC using independent channel inputs. The compression rates, and hence, the achieved distortion depends on the rates that are supported by the MAC. The distortion-rate function for the one-helper source coding problem is given as follows [40]

$$D(R_1, R_2) = (1 - \rho^2 + \rho^2 2^{-2R_2}) 2^{-2R_1}, \tag{2.7}$$

where R_i is the transmission rate of transmitter i.

It is easy to see that the distortion is minimized when the users operate on the corner point of the MAC capacity region that maximizes R_1, the rate from the main source to the decoder. Hence, the minimum achievable distortion with pure digital transmission is found to be

$$D_d^*(\Omega) = \frac{N(P_1 + P_2(1 - \rho^2) + N)}{(P_1 + P_2 + N)(P_1 + N)}. \tag{2.8}$$

Pure Analog Scheme

In pure symbol-by-symbol linear uncoded transmission (SLU), the encoders transmit scaled versions of the sources directly over the channel, i.e.,

$$X_i^n = \sqrt{\beta_i P_i} S_i^n, \quad i = 1, 2, \tag{2.9}$$

where $\beta_i \in [0, 1]$ is the scaling factor that allows to reduce the power assigned to each transmitter. Using an MMSE estimator at the receiver, the achievable distortion for this scheme is given by

$$D_u(\Omega, \beta_1, \beta_2) = \frac{\beta_2 P_2(1 - \rho^2) + N}{\beta_1 P_1 + \beta_2 P_2 + 2\rho\sqrt{\beta_1 \beta_2 P_1 P_2} + N}. \tag{2.10}$$

The next lemma characterizes the optimal distortion achievable by SLU by finding the optimal scaling factors at both transmitters.

Lemma 2. *For a given Ω, the optimal distortion achieved by SLU transmission is given by*

$$D_u^*(\Omega) = \begin{cases} \frac{P_2(1-\rho^2)+N}{P_1+P_2+2\rho\sqrt{P_1 P_2}+N}, & \textit{for } (P_1, P_2) \in \mathcal{P} \\ \frac{N(1-\rho^2)}{P_1(1-\rho^2)+N}, & \textit{for } (P_1, P_2) \notin \mathcal{P}, \end{cases}$$

where

$$\mathcal{P} \triangleq \left\{ (P_1, P_2) : P_1 \geq 0, P_2 \geq 0, \sqrt{P_1 P_2} \leq \frac{N\rho}{1-\rho^2} \right\}, \tag{2.11}$$

and the optimal scaling factors are

$$\beta_1^* = 1 \quad \textit{and} \quad \beta_2^* = \begin{cases} 1, & \textit{if } (P_1, P_2) \in \mathcal{P} \\ \frac{\rho^2 N^2}{P_1 P_2 (1-\rho^2)^2}, & \textit{if } (P_1, P_2) \notin \mathcal{P}. \end{cases} \tag{2.12}$$

Proof. Since $D_u(\Omega, \beta_1, \beta_2)$ in (2.10) is monotonically decreasing in β_1, it is minimized by $\beta_1^* = 1$. We also have that $D_u(\Omega, 1, \beta_2)$ is convex in $0 \leq \beta_2 \leq 1$, and therefore, the minimizing β_2^* is found as in (2.12) by using the standard Karush-Kuhn-Tucker conditions [51]. □

Lemma 2 indicates that the main source will always transmit at full power. On the other hand, all the helper power will be used if $(P_1, P_2) \in \mathcal{P}$, while if $(P_1, P_2) \notin \mathcal{P}$, the transmit power of the helper is reduced inversely proportional to the main source power. Increasing the helper power beyond the specified level in this regime increases the distortion.

2.3.2 Hybrid Digital-Analog (HDA) Schemes

In this section, we consider hybrid digital analog (HDA) schemes that transmit the superposition of a digital layer with an uncoded layer. First, we consider the most general HDA scheme known in the literature, proposed in [35] for the transmission of two source sequences (S_1^n, S_2^n) over a discrete memoryless MAC channel $p(y|x_1 x_2)$, in which the destination is interested in reconstructing both S_1^n and S_2^n at an average distortion (D_1, D_2).

In this scheme, each source sequence S_j^n is mapped to one of the 2^{nR_j} digital codewords $W_j^n(m_j)$. Then, each pair $(S_j^n, W_j^n(m_j))$ is mapped symbol-by-symbol to the channel input sequence X_j^n, that is transmitted over the interference channel. Upon receiving Y^n, the decoder jointly recovers the digital components $(W_1^n(m_1), W_2^n(m_2))$ by joint typicality, and reconstructs $\hat{S}_j^n$ by mapping symbol-by-symbol the analog channel output Y^n and the codewords corresponding to the two decoded digital messages. The general conditions for successful decoding of the messages and the achievable distortion pairs (D_1, D_2) for the transmission of (S_1^n, S_2^n) over a discrete memoryless MAC channel $p(y|x_1 x_2)$ are given in the next theorem.

Theorem 1. *[35] A distortion pair (D_1, D_2) is achievable for communication of (S_1, S_2) over a MAC channel $p(y|x_1, x_2)$ if*

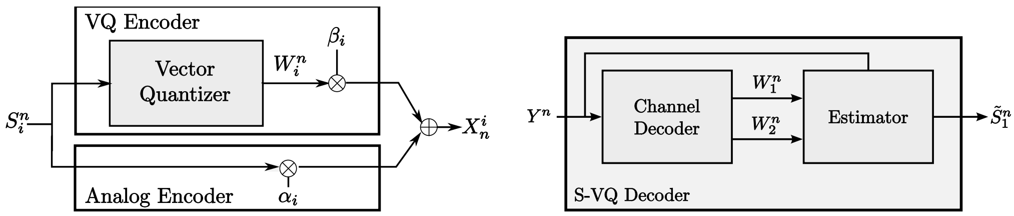


Figure 2.2: S-VQ encoder and decoder.

$$I(W_1; S_1|Q) < I(W_1; YW_2|Q)$$
$$I(W_2; S_2|Q) < I(W_2; YW_1|Q)$$
$$I(W_1; S_1|Q) + I(W_2; S_2|Q) < I(W_1W_2; Y|Q) + I(W_1; W_2|Q)$$

for some joint distribution $p(s_1s_2)p(q)p(w_1x_1|s_1q)p(w_2x_2|s_2q)$ *and reconstruction functions* $\hat{s}_i(w_1, w_2, y, q)$*, such that* $\mathrm{E}[d_i(S_i; \hat{S}_{1i})] \leq D_i$ *for* $i = 1, 2$.

In the one helper problem, the receiver is only interested in recovering S_1^n. Hence, Theorem 1 can be adapted to the one helper setup by dropping the distortion condition on D_2. Next, we study the achievable distortion for the Gaussian one-helper problem under different structures of the digital codewords W_1 and W_2.

Superposed Vector Quantizer (S-VQ)

Here, we consider a structure for the digital components of the HDA scheme following the scheme proposed in [13] for the MAC scenario. In this scheme, each transmitter quantizes the source with an optimal vector quantizer. Then, the quantized codeword is scaled and used directly as channel input superposed with an uncoded layer. The decoder jointly recovers the quantized codewords, and reconstructs the source using these codewords together with the channel output. See Fig. 2.2 for an illustration of the S-VQ encoder and decoder.

The distortion achievable by this scheme follows from Theorem 1. We let $Q = \emptyset$ for $i = 1, 2$. We consider rates $R_i > 0$, and, while potentially suboptimal, Gaussian random variables given by $W_i = S_i + Q_i$, where $Q_i \sim \mathcal{N}(0, 2^{-2R_i})$ independent of S_i and we generate the channel input with the symbol-by-symbol mapping $X_i = \alpha_i S_i + \beta_i W_i$. Parameters α_i, β_i are chosen such that the power constraint is satisfied, i.e.,

$$(\alpha_i + \beta_i)^2(1 - 2^{-2R_i}) + \alpha_i^2 2^{-2R_i} \leq P_i, \quad i = 1, 2. \tag{2.13}$$

The region given by (2.13) of feasible (α_i, β_i) pairs describes an ellipsoid. We consider the pairs which satisfy (2.13) with equality. It can be shown that it suffices to consider

$\alpha_i \geq 0$. We define the region of feasible pairs at each transmitter as

$$\Gamma_i \triangleq \left\{ (\alpha_i, \beta_i) \in \mathbb{R}^2 : \alpha_i \in \left[0, -\beta_i + \sqrt{P_i - \beta_i^2 2^{-2R_i}} \right], \quad \beta_i \in \left[-\sqrt{P_i 2^{2R_i}}, \sqrt{P_i 2^{2R_i}} \right] \right\}.$$

Let the covariance matrix of Y, W_1 and W_2 be given by

$$\mathbf{C}_{W_1 W_2 Y} = \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{12} & k_{22} & k_{23} \\ k_{13} & k_{23} & k_{33} \end{pmatrix} \text{ and } \mathbf{C}_{W_i Y} = \begin{pmatrix} k_{1i} & k_{i3} \\ k_{i3} & k_{33} \end{pmatrix}, \tag{2.14}$$

where

$$k_{ii} = (1 + 2^{-2R_i}), \quad j = 1, 2, \qquad k_{12} = \rho,$$
$$k_{13} = \alpha_1 + \beta_1 k_{11} + \rho(\alpha_2 + \beta_2), \qquad k_{23} = \alpha_2 + \beta_2 k_{22} + \rho(\alpha_1 + \beta_1),$$
$$k_{33} = P_1 + P_2 + 2\rho(\alpha_1 + \beta_1)(\alpha_2 + \beta_2) + N.$$

Using joint typicality decoding, W_1^n and W_2^n are decoded with high probability if $(R_1, R_2) \in \mathcal{R}$, where $\mathcal{R}$ is given by

$$\mathcal{R} \triangleq \left\{ (R_1, R_2) : R_1 + R_2 \leq \frac{1}{2} \log \frac{k_{11} k_{22} k_{33}}{|\mathbf{C}_{W_1 W_2 Y}|}, \ R_j \leq \frac{1}{2} \log \frac{k_{jj} |\mathbf{C}_{W_{j^c} Y}|}{|\mathbf{C}_{W_1 W_2 Y}|}, \ j = \{1, 2\}, \right\}.$$

Once W_1^n and W_2^n are recovered at the receiver, an MMSE estimator is used to reconstruct S_1^n with the available data $\tilde{S}_i^{vq} = \mathrm{E}[S_i | W_{1i}, W_{2i}, Y_i]$. We have $\tilde{S}_i^{vq} = \mathbf{h}_{vq} S_i + \mathbf{N}_i^{vq}$, where $\mathbf{h}_{vq} = [1, \rho, (\alpha_1 + \beta_1) + \rho(\alpha_2 + \beta_2)]$ and $\mathbf{N}_i^{vq} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{vq})$, where

$$\mathbf{C}_{vq} = \begin{pmatrix} 2^{-R_1} & 0 & \beta_1 2^{-2R_1} \\ 0 & 1 - \rho^2 + 2^{-2R_2} & (1-\rho^2)(\alpha_2 + \beta_2) + \beta_2 2^{-2R_2} \\ \beta_1 2^{-2R_1} & (1-\rho^2)(\alpha_2 + \beta_2) + \beta_2 2^{-2R_2} & (1-\rho^2)(\alpha_2 + \beta_2)^2 + \beta_1^2 2^{-2R_1} + \beta_2^2 2^{-2R_2} + N \end{pmatrix}.$$

Given $(\alpha_1, \alpha_2, R_1, R_2)$, the achievable distortion is found as

$$D_{vq}(\Omega, \alpha_1, \alpha_2, R_1, R_2) = (1 + \mathbf{h}_{vq}^H \mathbf{C}_{vq}^{-1} \mathbf{h}_{vq})^{-1}. \tag{2.15}$$

Minimized over the feasible parameters, the distortion achievable by S-VQ is given by

$$D_{vq}^*(\Omega) = \min_{\substack{(R_1, R_2) \in \mathcal{R} \\ (\alpha_i, \beta_i) \in \Gamma_i}} D_{vq}(\Omega, \alpha_1, \alpha_2, R_1, R_2) \tag{2.16}$$

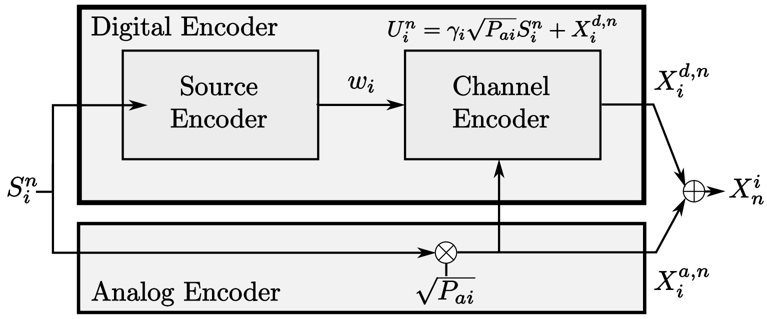


Figure 2.3: I-HDA encoder for transmitter i.

Interference Aware HDA Scheme (I-HDA)

Next, we propose an HDA scheme, based on the continuum of optimal HDA schemes for the point-to-point channel from [48]. This HDA scheme generalizes the pure digital transmission strategy, based on separate source and channel coding, as well as analog schemes, and takes into account the interference caused by the analog transmission in channel by using dirty paper coding (DPC) [49]. We denote this scheme by *interference aware HDA scheme* (I-HDA). As noted below, the achievable distortion of this scheme can be derived from Theorem 1. However, in order to reduce the complexity of the transmission scheme, here we propose the successive decoding of the digital components. In the next section, we numerically show that successive decoding is sufficient to achieve the same performance as S-VQ in the one-helper setup, which jointly decodes the digital components.

In I-HDA, each encoder transmits a superposition of digital and analog signals, i.e.,

$$X_i^n = X_i^{d,n} + X_i^{a,n}, \qquad i = 1, 2$$

where $X_i^{d,n}$ and $X_i^{a,n}$ are the length-n channel input vectors corresponding to digital and analog signals, respectively. The analog part of the transmitted signal, $X_i^{a,n}$ is a scaled uncoded version of the source sequence, S_i^n. We have $X_i^{a,n} = \sqrt{P_{ai}}S_i^n$, $i = 1, 2$, where factor $\alpha_i \in [0, 1]$, and $P_{ai} \triangleq \alpha_i P_i$ is the portion of the power at transmitter i dedicated to analog transmission. The digital portion $X_i^{d,n}$ transmits a quantized version of the source sequence, and is generated using a digital scheme based on dirty paper encoding, which considers the analog layers in the channel as interference known at the encoder. See Fig. 2.3 and Fig. 2.4 for an illustration of the encoder and decoder.

At the source encoder of user i, S_i^n is quantized with an optimal vector quantizer. The quantization codebook can be modeled with a "test channel", $\tilde{W}_i = S_i + Q_i$ where $Q_i \sim N(0, \sigma_{qi}^2)$ is independent of source S_i. The $2^{nI(S_i;\tilde{W}_i)}$ quantized codewords are randomly and uniformly assigned into $2^{nR_i^s}$ bins. For each source outcome, the source

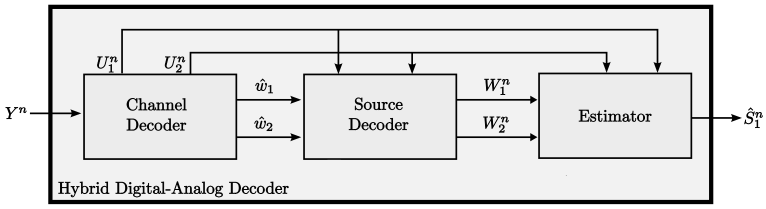


Figure 2.4: I-HDA decoder.

encoder determines the bin that the quantized source vector belongs to, and forwards the bin index $w_i \in [1, ..., 2^{nR_i^s}]$ to the digital channel encoder.

The channel codebook at each transmitter is generated using codeword as in the DPC for the Gaussian point-to-point setup as follows. We define the auxiliary random variable U_1 and U_2 as

$$U_i = \gamma_i \sqrt{P_{ai}} S_i + X_i^d, \quad i = 1, 2, \tag{2.17}$$

where $\gamma_i \in \mathbb{R}$, and X_i^d are zero mean Gaussian distributed variables with variance $P_{di} \triangleq (1 - \alpha_i)P_i$, i.e., $X_i^d \sim \mathcal{N}(0, P_{di})$. Note that relating (2.17) to the DPC scenario, S_i acts as an interference and $\gamma_i \sqrt{P_{ai}}$ has the role of the Costa parameter [49]. Then, at Transmitter 1 we generate $2^{nI(U_1;YU_2)}$, length-n i.i.d. codewords U_1^n, with $U_{1,k}$ $k = 1, ..., n$ following the distribution of (2.17), and at transmitter 2, we similarly generate $2^{nI(U_2;Y)}$ length-n i.i.d. codewords U_2^n.

Next, sequences U_i^n are uniformly distributed into $2^{nR_i^c}$ bins, $i = 1, 2$. For each sequence u_i^n, we let $j_i(u_i^n)$ be the index of the bin containing u_i^n. Then, at encoder i, given a source realization S_i^n, and a message from the source encoder w_i, in the bin w_i we search for a sequence U_i^n such that (U_i^n, S_i^n) are jointly typical, and declare an error if no or more than one such U_i^n can be found. Such a sequence is found with high probability, for large enough n, if the number of sequences in bin w_i is larger than $2^{nI(U_i;S_i)}$. Next, we transmit $X_i^{d,n} = U_i^n - \gamma_i \sqrt{P_{ai}} S_i^n$ as the channel input, and each encoder transmits a superposition of the analog signal and the digital codeword.

This scheme is a particular case of the general HDA scheme and the achievable distortion can be obtained by evaluating Theorem 1 with the digital codewords $W_i = (\tilde{W}_i, U_i)$ for $i = 1, 2$, where W_i and U_i are defined as above, and the symbol-by-symbol mappings $X_i = U_i + (-\gamma_i \sqrt{P_{ai}} + \sqrt{P_{ai}})S_i$. However, in the following we will consider successive decoding of the digital messages instead of the joint decoding of $(\tilde{W}_1, \tilde{W}_2, U_1, U_2)$, to reduce the complexity of the receiver[1]. We also note that unlike S-VQ, here the

[1]As noted in [13] and [35], messages ω_1 and ω_2 are correlated to the source sequences S_1^n and S_2^n, respectively, and common typical random coding techniques have to be modified accordingly. The same tools developed in [35] to prove Theorem 1 can be used to derive the sufficient conditions for successful encoding and successive decoding provided in this section.

quantization codewords, i.e., $\tilde{W}_i$, is not mapped to the channel input.

At the receiver, we first apply successive decoding of the auxiliary random variables U_1^n and U_2^n. Unlike in the usual superposition schemes, the channel codewords $X_1^{d,n}$ and $X_2^{d,n}$ are not recovered at the decoder; and hence, cannot be removed from the channel output. However, since U_1^n and U_2^n are correlated, the decoder first decodes U_2^n and uses it as side information to decode U_1^n. First, the decoder looks for the unique sequence $\hat{U}_2^n$ such that $(\hat{U}_2^n, Y^n)$ is jointly typical. We declare an error if more than one or no such sequence exist. The estimate $\hat{w}_2$ is equal to the index of the bin containing the sequence $\hat{U}_2^n$. It will be decoded correctly with high probability if,

$$R_2^c \leq I(U_2;Y) - I(U_2;S_2) = \frac{1}{2}\log\frac{P_{d2}(P_1+P_2+2\rho\sqrt{P_{a1}P_{a2}}+N)}{|\mathbf{C}_{U_2Y}|},$$

where

$$\mathbf{C}_{U_2Y} \triangleq \begin{pmatrix} P_1+P_2+2\rho\sqrt{P_{a1}P_{a2}}+N & P_{d2}+P_{a2}\gamma_2+\gamma_2\rho\sqrt{P_{a1}P_{a2}} \\ P_{d2}+P_{a2}\gamma_2+\gamma_2\rho\sqrt{P_{a1}P_{a2}} & P_{d2}+P_{a2}\gamma_2^2 \end{pmatrix}.$$

Once U_2^n is recovered, the decoder tries to decode U_1^n using U_2^n and Y^n. The decoder looks for $\hat{U}_1^n$ such that $(\hat{U}_1^n, U_2^n, Y^n)$ are jointly typical. For sufficiently large n, the decoding is successful if

$$\begin{aligned} R_1^c &\leq I(U_1;U_2Y) - I(U_1;S_1) \\ &= I(U_1;Y|U_2) - I(U_1;S_1|U_2) \\ &= \frac{1}{2}\log\frac{P_{d1}\,|\mathbf{C}_{U_2Y}|}{|\mathbf{C}_{U_1U_2Y}|}, \end{aligned}$$

where the first equality follows from the Markov Chain $U_1 - S_1 - U_2$, and the second one from

$$\mathbf{C}_{U_1U_2Y} \triangleq \begin{pmatrix} P_{d1}+P_{a1}\gamma_1^2 & \gamma_1\gamma_2\rho\sqrt{P_{a1}P_{a2}} & P_{d1}+P_{a1}\gamma_1+\gamma_1\rho\sqrt{P_{a1}P_{a2}} \\ \gamma_1\gamma_2\rho\sqrt{P_{a1}P_{a2}} & P_{d2}+P_{a2}\gamma_2^2 & P_{d2}+P_{a2}\gamma_2+\gamma_2\rho\sqrt{P_{a1}P_{a2}} \\ P_{d1}+P_{a1}\gamma_1+\gamma_1\rho\sqrt{P_{a1}P_{a2}} & P_{d2}+P_{a2}\gamma_2+\gamma_2\rho\sqrt{P_{a1}P_{a2}} & P_1+P_2+2\rho\sqrt{P_{a1}P_{a2}}+N \end{pmatrix}.$$

Factors γ_1 and γ_2 have to be chosen such that $R_1^c \geq 0$ and $R_2^c \geq 0$. This condition is satisfied by imposing $\gamma_{2L} \leq \gamma_2 \leq \gamma_{2H}$, where γ_{2L} and γ_{2H} are the two unique solutions to equation $R_2^c = 0$ given by

$$\gamma_{2H,2L} = P_{d2}\frac{P_{a2}+\rho\sqrt{P_{a1}P_{a2}} \pm \sqrt{P_{a2}\left(1+P_1+P_2+2\rho\sqrt{P_{a1}P_{a2}}\right)}}{P_{a2}\left(1+P_{d1}+P_{d2}+P_{a1}\left(1-\rho^2\right)\right)}.$$

For each feasible γ_2, any feasible γ_1 satisfies $\gamma_{1L} \leq \gamma_1 \leq \gamma_{1H}$, where γ_{1L} and γ_{1H} are the two unique solutions to equation $R_1^c = 0$ given by

$$\gamma_{1H,1L} = P_{d1} \frac{P_{d2}\rho\sqrt{P_{a1}P_{a2}}(1-\gamma_2) + P_{a1}\left(P_{d2} + P_{a2}\gamma_2^2(1-\rho^2)\right) \pm \sqrt{\Phi}}{P_{a1}\left(P_{a2}(1+P_{d1})\gamma_2^2\left(1-\rho^2\right) + P_{d2}\left(1+P_{d1}+P_{a2}(1-\gamma_2)^2\left(1-\rho^2\right)\right)\right)},$$

where

$$\begin{aligned}\Phi \triangleq\ & P_{a1}\left(P_{d2} + P_{a2}\gamma_2^2\left(1-\rho^2\right)\right)\left(P_{d2}\left(1+P_1+P_{a2}(1-\gamma_2)^2 + 2\sqrt{P_{a1}P_{a2}}\rho(1-\gamma_2)\right)\right.\\ & \left. + P_{a2}\gamma_2^2\left(1+P_{d1}+P_{a1}\left(1-\rho^2\right)\right)\right).\end{aligned}$$

In addition to the recovered bin indices w_1 and w_2, the decoded codewords U_1^n and U_2^n are correlated with the source sequence S_1^n and can be used as side information as in Wyner-Ziv source coding with correlated side information available at the decoder [4]. The source decoder at the receiver first decodes the quantized version of the helper source W_2^n jointly typical with (Y^n, U_1^n, U_2^n). Assuming it has access to the correct indices from the channel decoder, it can be decoded correctly with high probability, for large enough n, if,

$$R_2^s \geq I(W_2; S_2|YU_1U_2) = \frac{1}{2}\log\frac{|\mathbf{C}_{U_1U_2YW_2}|}{|\mathbf{C}_{U_1U_2Y}|\sigma_{q2}^2},$$

where

$$\mathbf{C}_{U_1U_2YW_2} = \begin{pmatrix} & & & \gamma_1\rho\sqrt{P_{a1}} \\ & \mathbf{C}_{U_1U_2Y} & & \gamma_2\sqrt{P_{a2}} \\ & & & \sqrt{P_{a2}}+\rho\sqrt{P_{a1}} \\ \gamma_1\rho\sqrt{P_{a1}} & \gamma_2\sqrt{P_{a2}} & \sqrt{P_{a2}}+\rho\sqrt{P_{a1}} & 1+\sigma_{q2}^2 \end{pmatrix}.$$

Once the side information from the helper is decoded, the source decoder tries to decode the quantized version of the main source W_1^n by looking for the sequence W_1^n such that $(W_1^n, W_2^n, Y^n, U_1^n, U_2^n)$ are jointly typical. A unique W_1^n is correctly decoded if,

$$R_1^s \geq I(W_1; S_1|YU_1U_2W_2) = \frac{1}{2}\log\frac{|\mathbf{C}_{U_1U_2YW_2W_1}|}{|\mathbf{C}_{U_1U_2YW_2}|\sigma_{q1}^2},$$

where

$$\mathbf{C}_{U_1U_2YW_2W_1} = \begin{pmatrix} & & & & \gamma_1\sqrt{P_{a1}} \\ & \mathbf{C}_{U_1U_2YW_2} & & & \gamma_2\rho\sqrt{P_{a2}} \\ & & & & \sqrt{P_{a1}}+\rho\sqrt{P_{a2}} \\ & & & & \rho \\ \gamma_1\sqrt{P_{a1}} & \gamma_2\rho\sqrt{P_{a2}} & \sqrt{P_{a1}}+\rho\sqrt{P_{a2}} & \rho & 1+\sigma_{q1}^2 \end{pmatrix}.$$

The channel coding rates R_1^c and R_2^c are determined for each Ω and $(\alpha_1, \alpha_2, \gamma_1, \gamma_2)$. To minimize the quantization error generated by the digital vector quantizers, their rates are chosen to be $R_i^s = R_i^c$ for $i = 1, 2$.

Finally, an MMSE estimator is used to reconstruct the source sequence S_1^n using all the available information at the decoder, $\tilde{S}_i^h = [U_{1i}, U_{2i}, Y_i, W_{1i}, W_{2i}]^H$, $i = 1, ..., n$, that is $\hat{S}_i = \mathrm{E}[S_i|\tilde{S}_i]$. The available information can be modeled as $\tilde{S}_i^h = \mathbf{h}_h S_{1i} + \mathbf{N}_i^h$ where $\mathbf{h}_h = [\gamma_1\sqrt{P_{a1}}, \gamma_2\sqrt{P_{a2}}\rho, \sqrt{P_{a1}} + \rho\sqrt{P_{a2}}, \rho, 1]^H$, and $\mathbf{N}_i^h$ is the signal components uncorrelated with S_i, distributed as $\mathbf{N}_i^h \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_h)$ where the $\mathbf{C}_h$ is given by

$$\mathbf{C}_h = \begin{pmatrix} P_{d1} & 0 & P_{d1} & 0 & 0 \\ 0 & P_{d2}+P_{a2}\gamma_2^2\left(1-\rho^2\right) & P_{d2}+P_{a2}\gamma_2\left(1-\rho^2\right) & \sqrt{P_{a2}}\gamma_2\left(1-\rho^2\right) & 0 \\ P_{d1} & P_{d2}+P_{a2}\gamma_2\left(1-\rho^2\right) & 1+P_{d1}+P_{d2}+P_{a2}\left(1-\rho^2\right) & \sqrt{P_{a2}}\left(1-\rho^2\right) & 0 \\ 0 & \sqrt{P_{a2}}\gamma_2\left(1-\rho^2\right) & \sqrt{P_{a2}}\left(1-\rho^2\right) & 1-\rho^2+\sigma_{q2}^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{q1}^2 \end{pmatrix}.$$

For a given Ω, fixed power allocation parameters (α_1, α_2), and fixed feasible (γ_1, γ_2) the distortion achievable by the I-HDA scheme is given by

$$D_h(\Omega, \alpha_1, \alpha_2, \gamma_1, \gamma_2) = (1 + \mathbf{h}_h^H \mathbf{C}_h^{-1} \mathbf{h}_h)^{-1}. \tag{2.18}$$

In general, the I-HDA scheme provides additional degrees-of-freedom through the two digital power allocation parameters α_1 and α_2 and the DPC parameters γ_1 and γ_2. For any given Ω we can optimize these parameters to minimize the distortion. We have

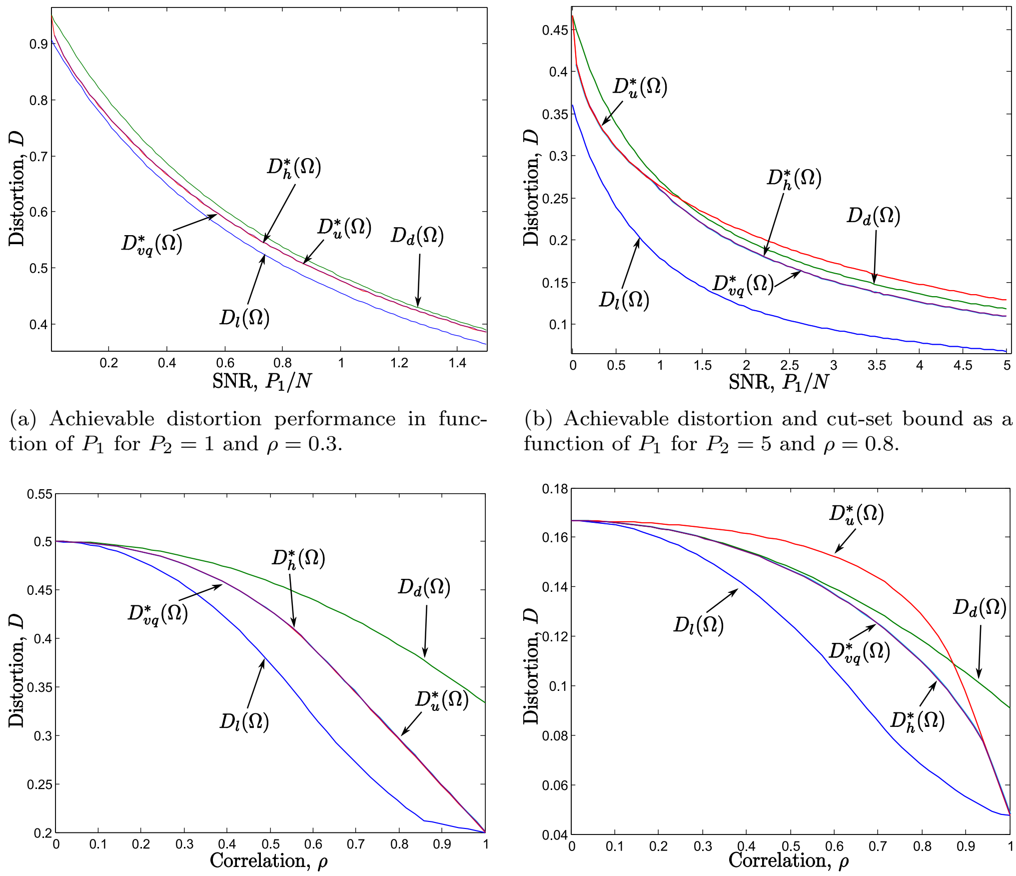
$$D_h^*(\Omega) = \min_{\substack{0\leq\alpha_1,\alpha_2\leq 1\\ \gamma_{1L}\leq\gamma_1\leq\gamma_{1H}\\ \gamma_{2L}\leq\gamma_2\leq\gamma_{2H}}} D_h(\Omega, \alpha_1, \alpha_2, \gamma_1, \gamma_2). \tag{2.19}$$

We note here that in general the optimal DPC parameters γ_1 and γ_2 do not coincide with the DPC Costa parameter from the point-to-point channel [49]. This was also noted in [48] for the point-to-point setup.

2.4 Numerical Results

In this section, we numerically evaluate the performance of the schemes proposed in Section 2.3, and we compare them with the lower bound from Section 2.2.

In Fig. 2.5(a) we let the power at the helper be fixed at $P_2 = 1$, and $N = 1$, and we plot the achievable distortion D with respect to the SNR, given by P_1/N. It can be seen that SLU achieves the lowest distortion among the considered schemes, and the performance of I-HDA and S-VQ, denoted by $D_h^*(\Omega)$ and $D_{vq}^*(\Omega)$, respectively,



(a) Achievable distortion performance in function of P_1 for $P_2 = 1$ and $\rho = 0.3$.

(b) Achievable distortion and cut-set bound as a function of P_1 for $P_2 = 5$ and $\rho = 0.8$.

(c) Achievable distortions and cut-set bound for $P = P_1 = P_2 = 1$ as a function of ρ.

(d) Achievable distortions and cut-set bound for $P = P_1 = P_2 = 5$ as a function of ρ.

Figure 2.5: Upper and lower bounds on the distortion with respect to SNR and correlation for the Gaussian one-helper problem.

reduce to $D_u^*(\Omega)$ in this regime of operation. On the other hand, separate source and channel coding achieves the worst performance among the considered schemes. In Fig. 2.5(b), we let $P_2 = 5$. Contrary to the previous case, while uncoded transmission still achieves the best distortion for low SNR values, its performance deteriorates as SNR increases, and the pure digital scheme has a better performance in this regime. I-HDA and S-VQ schemes reduce to the pure uncoded performance at low SNR values, while they outperform both digital and pure uncoded transmission schemes at higher SNR values. We note that both I-HDA and S-VQ achieve the same distortion in general, although I-HDA uses successive decoding, and are operationally different in their digital components. In S-VQ, the sources are quantized and are directly mapped to de channel input, and hence, the correlation between the quantization codewords is exploited. On the other hand, in I-HDA, the sources are quantized and mapped to DPC channel inputs. In this case, the correlation between the DPC codewords is exploited instead. While in this multi-terminal setup both structures achieve the same performance, we believe that in other communication scenarios their performance will be different.

Now, we consider a symmetric power scenario for which $P_1 = P_2 = P$, and plot the upper and lower bounds on the achievable distortion D with fixed SNR with respect to the source correlation ρ, which quantifies the quality of the helper's observation. We consider $P = 1$ in Fig. 2.5(c). Observe that all schemes achieve the lower bound, and are thus optimal, when $\rho = 0$, which corresponds to the case with independent, hence useless, helper observation. Since the helper is useless, the setup reduces to a Gaussian point-to-point channel, for which both separation and ULC are optimal. Uncoded transmission, I-HDA and S-VQ achieve the optimal performance at $\rho = 1$, i.e., when both users have access to the main source signal, while digital transmission is suboptimal. In this case, the helper and the main transmitter can fully cooperate by generating correlated channel inputs by exploiting the source correlation, although they still have individual power constraints. However, separation based schemes cannot generate correlated inputs distributedly, since source and channel coding are done independently. The suboptimality of digital transmission with respect to uncoded transmission for MMSE reconstruction in this setup was proven in [52]. Note that digital transmission is outperformed by the other schemes for any $\rho > 0$, while I-HDA, S-VQ and uncoded transmission achieve the same distortion. In Fig 2.5(c) we consider the upper and lower bounds for $P = 5$. In this case, pure digital transmission outperforms analog transmission for low ρ, while analog transmission achieves lower distortions for high correlation values. In general, the gains from separation based schemes are obtained only by the distributed compression of the source, while gains in SLU are obtained only by generating correlated channel inputs that result in beamforming gains. When the correlation is low, higher gains can be obtained from distributed compression, whereas when the correlation is high, distributed beamforming provides higher performance. Nevertheless, I-HDA and S-VQ schemes outperform both pure schemes and achieve lower distortions, since both schemes exploit both types of gains. We observe that for high correlation values, HDA schemes reduce to the performance of uncoded transmission. Note that, as expected, at $\rho = 0$ and $\rho = 1$ we have the same optimality results as before.

While we have not been able to prove it analytically, we believe that the uncoded transmission is optimal in those regions where the performance of HDA and uncoded transmission coincide. This is reminiscent of the optimality of uncoded transmission in the MAC setup considered in [13]. However, we note that the optimality conditions in [13] differ from the conditions under which uncoded transmission achieves the lowest distortion for intermediate correlation values.

In general, we observe that by using the correlated source sequences available at the transmitters, the encoders can generate correlated channel inputs that outperforms the pure digital transmission scheme based on separate source and channel coding, for which the channel inputs are independent. However, while the proposed HDA schemes achieve significant better performance than both the pure analog and pure digital schemes, in general the proposed transmission schemes are far from the derived distortion lower bound. We believe that this mainly stems from the looseness of the proposed lower bound, and tighter lower bounds in this setting is a challenging future research problem.

2.5 Conclusions

We have studied the JSCC one-helper problem in the Gaussian setting. We have proposed a lower bound on the achievable distortion using cut-set bound arguments and bounding the maximum correlation between the channel inputs. We have considered the achievable distortion, and have derived the optimal performance of pure digital and symbol-by-symbol linear analog transmission schemes. Then, we have proposed a generalized HDA transmission scheme based on power allocation among digital and analog signals, and studied two operational approaches. First, an HDA scheme has been considered, in which an analog component is superposed with a quantized version of the source sequence at each encoder, and joint decoding is employed at the destination. A second HDA scheme has been considered, which exploits the analog components through dirty paper coding and applies successive decoding. It is shown that in certain regimes, analog transmission outperforms pure digital transmission. It is numerically shown that both HDA schemes achieve the same distortion, which is significantly lower than pure digital and analog transmission in some regimes.

Chapter 3

Joint Source-Channel Coding with Time-Varying Channel and Side-Information: SISO

In this chapter, we consider the transmission of analog data, such as video or voice, under delay constraints in the presence of time-varying correlated information available at the receiver considered in Section 1.1.1. We model this important practical communication scenario as a JSCC problem of transmitting a Gaussian source over a time-varying Gaussian channel with the minimum average end-to-end distortion in the presence of time-varying correlated side information at the receiver. We consider a block fading model for the states of both the channel and the side information, and these states are assumed to be known perfectly at the receiver.

When both the channel and the side information are static, Shannon's separation theorem applies [28], and the optimal performance is achieved by separate source and channel coding; that is, the concatenation of an optimal Wyner-Ziv source code [4], which exploits the side information available at the decoder, with an optimal capacity achieving channel code. However, in delay-limited transmission, if the channel and the side information are time-varying, and the channel state information (CSI) is available only at the receiver, the transmitter cannot use the optimal source and channel codes without being prone to outages, and the separation theorem fails. In order to have a good performance on average, the transmitter has to adapt to the time-varying nature of both the channel and the side information without knowing their realizations.

Strategies based on separate source and channel coding suffer from the threshold effect and do not adapt well to the uncertainties of the channel [17]. On the other hand, simple uncoded transmission is robust to SNR mismatch, and does not suffer from the threshold effect. However, despite being optimal in the point-to-point Gaussian setup, it becomes suboptimal in the presence of correlated side information. In [34] an HDA scheme, which we denote by HDA-WZ, is proposed and shown to be robust to SNR mismatch and, unlike uncoded transmission, HDA-WZ is optimal even in the presence of side information at the receiver, or known interference in the channel. HDA-WZ is also shown to outperform separate source and channel coding and uncoded transmission in certain static setups, such as the transmission of a Gaussian source in the presence of correlated interference [53, 54], or to achieve the optimal distortion in the transmission of a bivariate Gaussian source over a broadcast channel [55]. In addition to HDA-WZ or various other HDA schemes, pure digital JSCC, based on joint decoding of the channel and source codewords, is also shown to exhibit improved robustness to the threshold effect, and to achieve the optimal performance in certain broadcasting scenarios [56–58].

The characterization of the optimal expected distortion for the proposed model in the absence of time-varying side information has received a lot of interest in recent years. Despite the ongoing efforts, the optimal performance remains an open problem. The expected distortion in this model is studied using multi-layer source codes concatenated with time-division [59] and superposition [42, 60] coding schemes. In general, more conclusive results on the performance can be obtained by studying the *distortion exponent*, which characterizes the exponential decay of the expected distortion in the asymptotically high SNR regime. The distortion exponent, which was introduced in [61], has been considered as a figure of merit in many scenarios: for parallel fading channels in [62], for the relay channel in [63], for channels with feedback in [64], for the two-way relay channel in [65], for the interference channel in [66], and in the presence of side information that might be absent in [67]. The distortion exponent, is characterized in the multi-antenna setup in certain regimes in [15],[16] and [39], and it is shown that multi-layer source and channel codes, or hybrid digital-analog coding schemes, are needed to achieve the optimal distortion exponent.

The pure source coding version of our problem, in which the channel is considered as an error-free constant-rate link, is studied in [43], and it is shown that, contrary to the channel coding problem, when the side information follows a continuous quasi-concave fading distribution, a single layer source code suffices to achieve the optimal performance. Recently, the JSCC problem has also been considered in [68] and [69]. In [68], the distortion exponent for separate source and channel coding is derived when the side information sequence has two states, the side information average gain does not increase with the SNR, and the channel follows a Rayleigh fading. In [69], HDA and joint decoding schemes are considered, and their performance is studied using the distortion loss, which quantifies the loss with respect to a fully informed encoder that perfectly knows the channels and the side information states.

In this chapter, we consider the JSCC problem both in the finite and asymtotically high SNR regimes for single antenna setups and defer the analysis of multi-antenna scenarios to Chapter 4. We first consider two lower bounds on the expected distortion by providing the encoder with different channel and side information state information. We then study achievable schemes based on uncoded transmission, SSCC, joint decoding, as well as HDA transmission and compare the performance of these schemes with the lower bound. The main contributions of this chapter are the following:

- We prove the optimality of separate source and channel coding when the channel is static and the side information state has a discrete or a continuous quasiconcave gain distribution. Remarkably, most common distributions used to model wireless communication channels, e.g., Rayleigh, Rician, Nakagami, have continuous and quasiconcave gains.
- When both the channel and the side information are time-varying, and the side information gain distribution is discrete or continuous quasiconcave, we derive a lower bound on the expected distortion called the partially informed encoder lower bound, by providing only the current channel state to the encoder while the side information state remains unknown.
- We show that uncoded transmission meets this lower bound when the side information fading state belongs to a certain class of continuous quasiconcave distributions, while separate source and channel coding is suboptimal. This class includes monotonically decreasing functions which occur, for example, under Rayleigh fading. To the best of our knowledge, this is the first result showing the optimality of uncoded transmission in a fading channel scenario while it would be suboptimal in the static case.
- We propose achievable schemes based on separate source and channel coding (SSCC), joint decoding (JDS) and hybrid digital analog transmission with a superposed analog layer (SHDA). We show that JDS always outperforms SSCC and numerically show that SHDA performs very close to the partially informed encoder lower bound, although in general no particular scheme outperforms the others.
- We obtain the distortion exponent corresponding to the proposed upper and lower bounds for Nakagami distributed channel and side information. We parameterize the uncertainty by the shape parameter, given by L_c for the channel and by L_s for the side information. For $L_c \geq 1$, we characterize the optimal distortion exponent and show that it is achieved by SHDA, in line with the numerical results. For $L_c < 1$, we show that JDS achieves the optimal distortion exponent in certain regimes, while SHDA is suboptimal. However, as L_s increases, the performance of JDS saturates and becomes worse than SHDA, whose distortion exponent converges to the upper bound.

The rest of the chapter is organized as follows: in Section 3.1 we introduce the

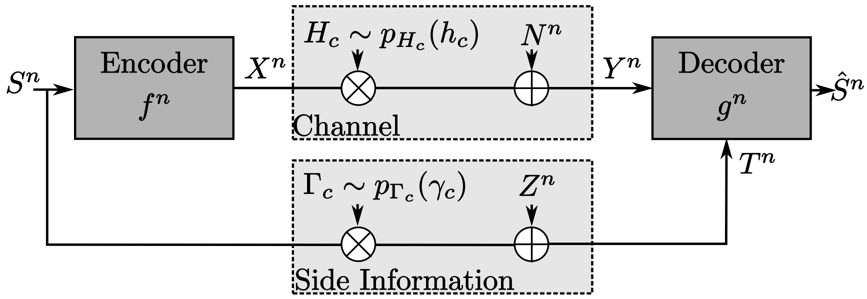


Figure 3.1: Block diagram of the joint source-channel coding problem with fading channel and side information.

system model. In Section 3.2 we provide some previous results and characterize the optimal performance for a static channel; while in Section 3.3, we propose upper and lower bounds on the performance. In Section 3.4 we prove the optimality of uncoded transmission under certain side information fading distributions. In Section 3.5 we provide numerical results for the finite SNR regime, while in Section 3.6 we consider a high SNR analysis and characterize the optimal distortion exponent. Finally, in Section 4.8 we provide the conclusions.

3.1 System Model

We consider the transmission of a random source sequence S^n of independent and identically distributed (i.i.d.) entries form a zero mean, unit variance real Gaussian distribution, i.e., $S_i \sim \mathcal{N}(0,1)$, over a time-varying channel (see Fig. 4.1). An encoder $f^n : \mathbb{R}^n \to \mathbb{R}^n$ maps the source sequence S^n to the input of this channel, $X^n \in \mathbb{R}^n$, i.e., $x^n = f^n(s^n)$, while satisfying an average power constraint: $\frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_i^2] \le 1$. The block-fading channel is given by

$$Y^n = H_c X^n + N^n, \tag{3.1}$$

where $H_c \in \mathbb{R}$ is the channel fading state with probability density function (pdf) $p_{H_c}(h_c)$, and N^n is the additive white Gaussian noise $N_i \sim \mathcal{N}(0,1)$.

In addition, there is an orthogonal block-fading side information channel connecting the source to the destination, which provides an uncoded noisy version of the source sequence to the destination. This second channel models the time-varying correlated side-information at the destination. Similarly to the communication channel, we model this side information channel as a memoryless block fading channel given by

$$T^n = \Gamma_c S^n + Z^n, \tag{3.2}$$

where $\Gamma_c \in \mathbb{R}$ is the side information fading state with pdf $p_{\Gamma_c}(\gamma_c)$, S^n is the uncoded channel input, and Z^n is the additive white Gaussian noise, i.e., $Z_i \sim \mathcal{N}(0,1)$, $i = 1, ..., n$.

We define $H \triangleq H_c^2 \in \mathbb{R}^+$ and $\Gamma \triangleq \Gamma_c^2 \in \mathbb{R}^+$ as the instantaneous *channel gain* and the instantaneous *side information gain*, with pdfs $p_H(h)$ and $p_\Gamma(\gamma)$, respectively.

We assume a stringent delay constraint that imposes each source block of n source samples to be transmitted over one block of the channel, consisting of n channel uses, i.e., we consider matched source and channel bandwidths. We study the more general case with an arbitrary bandwidth ratio in Chapter 4. Some additional results considering a mismatched bandwidth ratio in SISO channels have been reported in [70].

Both the channel and side information states, H_c and Γ_c, are assumed to be constant, with values h_c and γ_c, respectively, for the duration of one channel block, and independent among different blocks. The channel and side information state realizations h_c and γ_c are assumed to be known at the receiver, while the encoder is only aware of their distributions.

The decoder reconstructs the source sequence from the channel output Y^n, the side information sequence T^n, and the channel and side information states h_c and γ_c using a mapping $g^n : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^n$, where $\hat{S}^n = g^n(Y^n, T^n, h_c, \gamma_c)$.

For given channel and side information distributions, we are interested in characterizing the minimum *expected distortion*, $\mathrm{E}[D]$, where the quadratic distortion between the source sequence and the reconstruction is given by

$$D \triangleq \frac{1}{n} \sum_{i=1}^{n} (S_i - \hat{S}_i)^2. \tag{3.3}$$

The expectation is taken with respect to the source, channel and side information states, and the noise distributions. The minimum expected distortion can be expressed as

$$ED^* \triangleq \lim_{n\rightarrow\infty} \min_{f^n, g^n} \mathrm{E}[D]. \tag{3.4}$$

3.2 Preliminary Results

We first review some of the existing results in the literature for the source coding version of the problem under consideration, in which the fading channel is substituted by an error-free channel of finite capacity. We then focus on the scenario in which the channel is noisy but static, i.e., the channel gain is constant and known both at the encoder and the decoder. We show that separate source and channel coding is optimal in the case of a static channel.

3.2.1 Background: Lossy Source Coding with Fading Side Information

The source-coding version of this problem in which the fading channel is substituted by an error-free channel of rate R and a time-varying side information sequence T^n is available at the destination is considered in [43]. Here we briefly review the results of [43] which will be used later in the chapter.

Let the distribution $p_\Gamma(\gamma)$ be discrete with M states $\gamma_1 \leq \cdots \leq \gamma_M$ with probabilities $\Pr[\Gamma = \gamma_i] = p_i$. We define the side information sequence available at the decoder when the realization of the side information fading gain is γ_{si} as $T_{i,1}^n \triangleq \sqrt{\gamma_i} S^n + Z^n$ [1]. Note that the side information has a degraded structure, characterized by the Markov chain

$$T_{1,j} - \cdots - T_{M-1,j} - T_{M,j} - S_j, \quad j = 1, ..., n. \tag{3.5}$$

This is equivalent to the Heegard-Berger source coding problem with degraded side information [5], in which an encoder is connected by an error-free channel of rate R to M receivers, and receiver i has access to side information $T_{i,1}^n$. The minimum expected distortion is given by the solution to the following problem,

$$ED^*(R) = \min_{\mathbf{D}: R_{HB}(\mathbf{D}) \leq R} \mathbf{p}^T \mathbf{D}, \tag{3.6}$$

where $\mathbf{p} \triangleq [p_1, ..., p_M]$, $\mathbf{D} = [D_1, ..., D_M]$ with D_i defined as the achievable distortion at receiver i and $R_{HB}(\mathbf{D})$ is the Heegard-Berger rate-distortion function given by

$$R_{HB}(\mathbf{D}) = \min_{W_1^M \in \mathcal{P}(\mathbf{D})} \sum_{i=1}^{M} I(S; W_i | W_1^{i-1}, T_i), \tag{3.7}$$

where W_1^i denotes the auxiliary random variables $W_1, ..., W_i$, and $\mathcal{P}(\mathbf{D})$ is the set of random variables W_1^M satisfying the Markov chain condition

$$W_M - \cdots - W_1 - S - T_M - T_{M-1} - \cdots - T_1,$$

for which there exist source reconstructions $\hat{S}_i(T_i, W_1^i)$ satisfying $\mathrm{E}[d_i(S, \hat{S}_i)] \leq D_i$, $i = 1, ..., M$.

When the source X^n is Gaussian, it can be shown that the optimal auxiliary random variables W_1^M minimizing (3.6) are jointly Gaussian. Then, the minimum expected distortion for a Gaussian source with finite number of side information states can be

[1] To avoid confusion in the indexing, we use $T_{i,1}^n \triangleq [T_{i,1}, ..., T_{i,n}]$ to denote all the elements $T_{i,j}$, $j = 1, ..., n$ for the i-th side information state.

found by solving the following convex optimization problem [43, Eq. (59)-(62)]:

$$
\begin{aligned}
ED_F^*(R) = & \min_{D_1,\dots,D_M \in \mathbb{R}^{++}} \sum_{i=1}^{M} p_i D_i \\
& \text{s.t. } -\frac{1}{2}\sum_{i=0}^{M-1} \log(1+(\gamma_{i+1}-\gamma_i)D_i) - \frac{1}{2}\log D_M \le R, \\
& D_i \le (D_{i-1}^{-1} + \gamma_i - \gamma_{i-1})^{-1}, \ i = 1, ..., M,
\end{aligned} \tag{3.8}
$$

where $D_0 \triangleq \sigma_x^2 = 1$ and $\gamma_0 \triangleq 0$. The Heegard-Berger rate distortion function also extends to the set of infinitely many degraded fading states, $\gamma_1 \le \gamma_2 \le \cdots$ with $\sum_{i=1}^{\infty} p_i = 1$ [43]. For a countable number of states, the expected distortion is given in [43, Eq. (75)-(78)] as the solution to

$$
\begin{aligned}
ED_C^*(R) = & \min_{D_1, D_2 \dots \in \mathbb{R}^{++}} \sum_{i=1}^{\infty} p_i D_i \\
& \text{s.t. } -\frac{1}{2}\sum_{i=0}^{\infty} (\log(D_{i-1}^{-1} + \gamma_i - \gamma_{,i-1}) + \log D_i) \le R, \\
& D_i \le (D_{i-1}^{-1} + \gamma_i - \gamma_{,i-1})^{-1}, \ i = 1, 2, ...
\end{aligned} \tag{3.9}
$$

When the side information distribution $p_\Gamma(\gamma)$ is continuous and quasiconcave[2], the optimal expected distortion is achieved by single-layer rate allocation such that all the available rate R is targeted to a single side information state $\bar{\gamma}$ [43]. Then, the optimal expected distortion is given by

$$
ED_Q^*(R) = \int_0^{\bar{\gamma}} \frac{p_\Gamma(\gamma)}{1+\gamma} d\gamma + \int_{\bar{\gamma}}^{\infty} \frac{p_\Gamma(\gamma)}{(\bar{\gamma}+1)2^{2R} + \gamma - \bar{\gamma}} d\gamma, \tag{3.10}
$$

where $\bar{\gamma}$ minimizing (3.10) is determined as follows: Let a super-level set be defined as $[\gamma_l(\alpha), \gamma_r(\alpha)] \triangleq \{\gamma | p_\Gamma(\gamma) \ge \alpha\}$. Then, $\bar{\gamma}$ is defined as the left endpoint of the super-level set induced by α^*, i.e., $\bar{\gamma} = \gamma_l(\alpha^*)$, where $\alpha^* \in [0, \max p_\Gamma(\gamma)]$ is found by solving the equation

$$
\int_{\gamma_l(\alpha^*)}^{\infty} \frac{p_\Gamma(\gamma) - \alpha^*}{((1+\gamma_l(\alpha^*))2^{2R} + \gamma - \gamma_l(\alpha^*))^2} d\gamma = 0. \tag{3.11}
$$

When the side information state is Rayleigh distributed, the side information gain Γ is exponentially distributed. Then it can be seen that $\bar{\gamma} = 0$ and the optimal expected distortion becomes

$$
ED_{Ray}^*(R) = \frac{1}{\mathrm{E}[\Gamma]} e^{\frac{2^{2R}}{\mathrm{E}[\Gamma]}} E_1\left(\frac{2^{2R}}{\mathrm{E}[\Gamma]}\right), \tag{3.12}
$$

[2] A function $g(x)$ is quasiconcave if its supersets $\{x | g(x) \ge \alpha\}$ are convex for all α.

where $E_1(x) \triangleq \int_x^\infty t^{-1}e^{-t}dt$ is the exponential integral [43].

In the following sections we use $ED_F^*(R)$, $ED_Q^*(R)$ and $ED_{Ray}^*(R)$ to generate lower bounds on the expected distortion. To unify these results, we define the function $ED_s^*(R)$ as the minimum expected distortion in the source coding problem for these three setups. Therefore, while the achievability results are valid for any distribution, the optimality results in this chapter are valid for discrete, i.e., finite or countable number of states, as well as continuous quasiconcave distributions of the side information.

3.2.2 Static Channel and Fading Side Information

In this section we consider a static channel and prove the optimality of separate source and channel coding in this setting. We consider a channel from X^n to Y^n, not necessarily the fading Gaussian channel characterized in (3.1), of fixed capacity C. The side information is still block-fading as in (3.2) with the side information gain following a distribution $p_\Gamma(\gamma)$. Note that it is a JSCC generalization of the source coding problem reviewed in Section 3.2.1. We denote the minimum expected distortion in the case of a static channel by ED_{sta}^*.

Optimality of separate source and channel coding can be proven when Γ, the side information gain, has finite or countable number of states, or when it has a continuous quasiconcave distribution. This reduces the problem to the source coding problem of Section 3.2.1 with $R = \mathrm{C}$.

Theorem 2. *Assume that the channel is static with capacity* C. *When the side information gain* Γ *has a discrete number of states, or a continuous quasiconcave pdf* $p_\Gamma(\gamma)$*, the minimum expected distortion,* ED_{sta}^**, is achieved by separate source and channel coding, and is given by*

$$ED_{sta}^* = ED_s^*(\mathrm{C}). \tag{3.13}$$

Proof. The theorem is first proven when Γ has a discrete distribution. Then, to show the optimality of separation when $p_\Gamma(\gamma)$ is continuous and quasiconcave we construct a lower bound on the expected distortion ED_{sta}^* by discretizing the continuum of analog side information states, and show that this bound is achievable in the limit of finer discretizations. See Appendix C for details. □

3.3 Upper and Lower Bounds

In this section we return to the problem presented in Section 3.1 in which both the channel and the side information are block-fading. We construct two lower bounds on ED^*. The first one is obtained by informing the encoder with both of the channel and side information states H and Γ. Then, we construct a tighter lower bound by informing the encoder only with the channel state H. Next, we propose achievable schemes based on uncoded transmission, separate source and channel coding, joint decoding and hybrid digital-analog transmission. Comparison of the proposed upper and lower bounds in different regimes of operation is relegated to Sections 3.4, 3.5 and 3.6.

3.3.1 Informed Encoder Lower Bound

A trivial lower bound on ED^* can be obtained by providing the encoder with the instantaneous states of the channel and the side information. We call this bound the *informed encoder lower bound.* At each realization, the problem reduces to the systematic model considered in [28] (see also [71]), for which the separation theorem holds. The encoder compresses the source sequence using Wyner-Ziv source coding considering the side information, and then transmits the compressed bits at the instantaneous capacity of the channel. For states (h, γ), the optimal distortion is given by $D_{\inf}(h, \gamma) \triangleq (1 + h)^{-1}(1 + \gamma)^{-1}$. Averaging over the channel and side information states, the informed encoder lower bound on the expected distortion is given by

$$ED^*_{\inf} = \mathrm{E}_{H,\Gamma}[D_{\inf}(H, \Gamma)]. \tag{3.14}$$

3.3.2 Partially Informed Encoder Lower Bound

We can obtain a tighter lower bound by providing the encoder only with the channel realization h. We call this the *partially informed encoder lower bound*, and denote it by ED^*_{pi}. For a given channel state realization h, the setup reduces to the one considered in Section 3.2.2, and for a discrete or continuous quasiconcave $p_\Gamma(\gamma)$, separation theorem applies for each channel realization. Averaging over the channel states, we have the following lower bound.

Lemma 3. *If $p_\Gamma(\gamma)$ is discrete or continuous quasiconcave, the minimum expected distortion is lower bounded by*

$$ED^*_{pi} \triangleq \mathrm{E}_H[ED^*_s(\mathcal{C}(H))], \tag{3.15}$$

where $\mathcal{C}(h) \triangleq \frac{1}{2}\log(1 + h)$ is the capacity of the channel for a given realization $h = h_c^2$.

Providing only the side information state to the encoder does not lead to a tight computable lower bound, since the optimality of separate source and channel coding does not hold in this case. Although the partially informed encoder lower bound is tighter, we will include the informed encoder bound in our analysis, as it provides a benchmark for the performance when both channel and side information states are available at the transmitter, which sheds light on the value of the CSI feedback for this JSCC problem.

Next, we study some achievable schemes (upper bounds) for the JSCC problem under consideration.

3.3.3 Uncoded Transmission

Uncoded transmission is a memoryless and zero-delay transmission scheme in which each channel input X_i is generated by scaling the source signal S_i while satisfying the power constraint. In our model both the source variance and power constraint of the encoder are 1, and hence, no scaling is needed, i.e., $X_i = S_i$. The received signal from the channel is then given by

$$Y_i = h_c S_i + N_i, \qquad i = 1, ..., n. \tag{3.16}$$

The receiver reconstructs each component with a MMSE estimator using both the channel output and the side information sequence, i.e., $\hat{S}_i = \mathrm{E}[S_i|Y_i, T_i]$, $i = 1, ..., n$. The distortion for each source component S_i for a given channel and side information realization h_c and γ_c is given by $D_u(h, \gamma) \triangleq (1 + h + \gamma)^{-1}$. Averaging over the channel and side information realizations, we have

$$ED_u = \mathrm{E}_{H,\Gamma}[D_u(H, \Gamma)]. \tag{3.17}$$

3.3.4 Separate Source and Channel Coding (SSCC)

Next, we consider separate source and channel coding with a single layer based on Wyner-Ziv source coding using the side information sequence followed by channel coding for the channel. Note that due to the lack of CSI at the transmitter the rates of the source and the channel codebooks are fixed at all channel and side information states. Since the number of bins and channel codewords are fixed without knowledge of the channel and the side information states, this scheme may suffer from outages both in the channel decoding and in the source decoding stages.

The quantization codebook consists of $2^{n(R_c+R_s)}$ length-n codewords, $W^n(i)$, $i = 1, ..., 2^{n(R_c+R_s)}$, generated through a 'test channel' given by $W = S + Q$, with $Q \sim \mathcal{N}(0, \sigma_Q^2)$ and independent of S. The quantization noise variance is chosen such that $R_s + R_c = I(S; W) + \epsilon$, for an arbitrarily small $\epsilon > 0$, i.e., $\sigma_Q^2 = (2^{2(R_s+R_c-\epsilon)} - 1)^{-1}$. The generated quantization codewords are then uniformly distributed into 2^{nR_c} bins. On average, each bin contains 2^{nR_s} codewords. Additionally, a Gaussian channel codebook with 2^{nRc} length-n codewords $X^n(s)$ is generated independently with $X \sim \mathcal{N}(0, 1)$, and the codeword $X^n(s)$, $s \in [1, ..., 2^{nR_c}]$, is assigned to the bin index s.

Given a source realization S^n, the encoder searches for a quantization codeword

$W^n(i)$ that is jointly typical with S^n. Assuming one such codeword is found, the channel codeword $X^n(s)$ is transmitted over the channel, where s is the bin index of $W^n(i)$. At reception, the bin index s is recovered with high probability using the channel output Y^n if,

$$R_c < I(X;Y). \tag{3.18}$$

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

$$R_c > I(S;W|T). \tag{3.19}$$

If the quantization codeword W^n is successfully decoded, then $\hat{S}^n$ is reconstructed with an optimal MMSE estimator as $\hat{S}_i = \mathrm{E}[S_i|T_i, W_i]$ for $i = 1, ..., n$.

An outage is declared whenever, due to the randomness of the channel or the side information, the quantization codebook cannot be correctly decoded, i.e., when condition (3.18) or (3.19) are not satisfied. In case of an outage, only the side information sequence is used to estimate the source, and we have $\hat{S}_i = \mathrm{E}[S_i|T_i]$. When the quantization rate is R and the side information state is γ, the distortion is

$$D_d(R,\gamma) \triangleq (\gamma + 2^{2R})^{-1}, \tag{3.20}$$

if the quantization codeword is decoded correctly. If an outage occurs, the achievable distortion is given by $D_d(0,\gamma)$. Then, the expected distortion of SSCC is given by

$$ED_{sb}(R_s, R_c) = \mathrm{E}_{\mathcal{O}_{sb}^c}[D_d(R_s + R_c, \Gamma)] + \mathrm{E}_{\mathcal{O}_{sb}}[D_d(0, \Gamma)],$$

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

$$\mathcal{O}_{sb} \triangleq \{(h,\gamma) : R_c \geq I(X;Y) \text{ or } R_c \leq I(S;W|T)\},$$

where $I(S;W|T) = \frac{1}{2}\log\left(1 + (2^{2(R_s+R_c+\epsilon)} - 1)/(\gamma+1)\right)$ and $I(X;Y) = \frac{1}{2}\log(1+h)$.

Since the source and channel rates R_s and R_c are fixed for all channel and side information states, we can chose those in order to minimize the expected distortion. Thus, we have

$$ED_{sb}^* \triangleq \min_{R_c, R_s} ED_{sb}(R_s, R_c). \tag{3.21}$$

When the side information has a continuous quasiconcave gain distribution, we can have a closed-form expression for the optimal source coding rate R_s, as given in the next lemma.

Lemma 4. *For a given R_c, if $p_\Gamma(\gamma)$ is continuous and quasiconcave, $ED_{sb}(R_s, R_c)$ is minimized by setting $R_s = \frac{1}{2}\log(1 + (1 + \bar{\gamma})(2^{2R_c} - 1)) - R_c + \epsilon$ where $\bar{\gamma}$ is the solution to (3.11).*

Proof. Once the channel rate has been fixed, i.e., once R_c is fixed, it follows from the results in Section 3.2.1 that $ED_{sb}(R_s, R_c)$ is minimized by compressing the source to a single layer targeted for side information state $\bar{\gamma}$, i.e., $R_c = I(S; W|T = \bar{\gamma}S + Z) = \frac{1}{2}\log\left(1 + \frac{2^{2(R_s+R_c-\epsilon)}-1}{1+\bar{\gamma}}\right)$, from where R_s is obtained. □

We can reduce the complexity of SSCC by having only a single codeword in each bin, that is, by letting $R_s = 0$. This way, we get rid of the outage event corresponding to a poor side information gain 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. Without binning, the minimum expected distortion is found as $ED^*_{nb} \triangleq \min_{R_c} ED_{sb}(0, R_c)$.

Note that when the side information fading distribution is such that $\bar{\gamma} = 0$, then, from Lemma 4, the optimal source coding rate is $R_s = 0$, i.e., the minimum expected distortion is achieved by ignoring the decoder side information in the encoding process.

Corollary 1. *If $\bar{\gamma} = 0$, the optimal SSCC does not utilize binning, that is, $R^*_s = 0$ and $ED^*_{sb} = ED^*_{nb}$.*

In this section, we have only considered a single layer source coding scheme since for continuous quasiconcave $p_\Gamma(\gamma)$, the optimal source code uses a single source code layer. However, in the case of discrete number of side information gain states, the optimal source code employs multiple source layers, one layer targeting each of the side information states [43]. For a channel code at rate R_c, the achievable expected distortion can be obtained similarly to the scheme described in this section, using $ED_F(R_c)$ and $ED_C(R_c)$ in (3.8), for finite and countable number of side information states, respectively.

3.3.5 Joint Decoding Scheme (JDS)

Here, we consider a source-channel coding scheme that does not involve any explicit binning at the encoder and uses joint decoding to reduce the outage probability. This coding scheme is introduced in [57] in the context of broadcasting a common source to multiple receivers with different side information qualities, and it is shown to be optimal in the case of lossless broadcasting over a static channel. The success of the decoding process depends on the joint quality of the channel and the side information states.

At the encoder, a codebook of $2^{nR_{jd}}$ length-n quantization codewords $W^n(i)$, $i = 1, ..., 2^{nR_{jd}}$, are generated through a 'test channel' $W = S + Q$, where $Q \sim \mathcal{N}(0, \sigma^2_Q)$

and is independent of S. The quantization noise variance is chosen such that $R_{jd} = I(S;W) + \epsilon$, for an arbitrarily small $\epsilon > 0$. Then, an independent Gaussian codebook of size $2^{nR_{jd}}$ is generated with length-n codewords $X^n(i)$ with $X \sim \mathcal{N}(0,1)$. Given a source outcome S^n, the transmitter finds the quantization codeword $W^n(i)$ jointly typical with the source outcome and transmits the corresponding channel codeword $X^n(i)$ over the channel. At reception, the decoder looks for an index i for which both $(x^n(i), Y^n)$ and $(T^n, w^n(i))$ are jointly typical. Then the outage event is given by

$$\mathcal{O}_{jd} \triangleq \{(h,\gamma) : I(S;W|T) \geq I(X;Y)\}, \tag{3.22}$$

where $I(S;W|T) = \frac{1}{2}\log\left(1 + (2^{2(R_{jd}-\epsilon)} - 1)/(\gamma+1)\right)$ and $I(X;Y) = \frac{1}{2}\log(1+h)$.

If decoding is successful, the source S^n is estimated using both the quantization codeword and the side information sequence, while if an outage occurs, the source S^n is reconstructed using only the side information sequence. Then, the expected distortion for the JDS scheme is found as

$$ED_{jd}(R_{jd}) = \mathrm{E}_{\mathcal{O}_{jd}^c}[D_d(R_{jd},\Gamma)] + \mathrm{E}_{\mathcal{O}_{jd}}[D_d(0,\Gamma)]. \tag{3.23}$$

Similarly to (3.21), the expected distortion can be optimized over R_{jd} to obtain the minimum expected distortion achieved by JDS, that is, $ED^*_{jd} \triangleq \min_{R_{jd}} ED_{jd}(R_{jd})$.

In SSCC, the quantization codeword is successfully decoded only if both the channel and the source codes are successfully decoded. On the other hand, JDS decodes the quantized codeword exploiting the joint quality of both the channel and the side information sequence. The joint decoding produces a binning-like decoding: only some Y^n are jointly typical with $X(s)$, generating a virtual bin of W^m codewords from which only one is jointly typical with T^m. The size of those bins depends on the particular realizations of H and Γ unlike in a Wyner-Ziv scheme, in which the bin sizes are designed in advance. Hence, a bad channel realization can be compensated with a sufficiently good side information realization, or viceversa, reducing the outage probability.

Indeed, the minimum expected distortion of JDS is always lower than that of SSCC, as stated in the next lemma.

Lemma 5. *For any given $p_H(h)$ and $p_\Gamma(\gamma)$, JDS outperforms SSCC at any SNR, i.e., we have $ED^*_{sb} \geq ED^*_{jd}$.*

Proof. Consider the SSCC scheme as in Section 3.3.4 with rates R_c and R_s. We will show that the JDS scheme with rate $R_{jd} = R_s + R_c$ achieves a lower expected distortion, i.e., $ED_{sb}(R_c, R_s) \geq ED_{jd}(R_c + R_s)$. If both schemes are in outage, or if the quantization codeword is decoded successfully in both, they achieve the same distortion. Thus, to prove our claim, it will suffice to show that $\mathcal{O}_{sb} \supseteq \mathcal{O}_{jd}$.

Let (h,γ) be such that $R_c \geq I(X;Y) = \frac{1}{2}\log(1+h)$, i.e., SSCC is in outage. Note that for given (h,γ), R_s and R_c, $I(X;Y)$ and $I(S;W|T)$ have the same values for both schemes. However, if $I(S;W|T) < I(X;Y)$, JDS is able to decode the quantization codeword successfully while SSCC would still be in outage. This condition is satisfied whenever $\frac{1}{2}\log\left(1+\frac{2^{2(R_{jd}-\epsilon)}-1}{\gamma+1}\right) < \frac{1}{2}\log(1+h)$, or equivalently $\gamma > \frac{2^{2(R_{jd}-\epsilon)}-1}{h} - 1$. If this condition does not hold, both schemes are in outage and have the same performance. Then, $\mathcal{O}_{sb} \supseteq \mathcal{O}_{jd}$. Conversely, if JDS is in outage, i.e., $I(S;W|T) \geq I(X;Y)$, then SSCC is also in outage since either $R_c \geq I(X;Y)$ or $R_c \leq I(X;Y) \leq I(S;W|T)$ holds. Therefore, we have $\mathcal{O}_{sb} \supseteq \mathcal{O}_{jd}$, which implies $ED_{sb}(R_c,R_b) \geq ED_{jd}(R_c+R_b)$ and $ED^*_{sb} \geq ED^*_{jd}$. This completes the proof. □

3.3.6 Superposed Hybrid Digital-Analog Transmission (SHDA)

In this section, we consider a general HDA scheme, and provide sufficient conditions for the achievable distortion for discrete memoryless channels with side information available at the decoder, similar to the general HDA scheme considered in [36] in the absence of side information. Then, using this result, we propose a particular HDA scheme for the time-varying setup that superposes a coded layer with an uncoded layer and allocates the power among the two layers.

Consider the transmission of a memoryless source sequence S^n over a discrete memoryless channel $p(y|x)$, in which the destination is interested in reconstructing the source sequence at an average distortion D. In addition, a memoryless sequence correlated with S^n, T^n, is available at the destination as side information. We consider a general HDA scheme in which the source sequence S^n is mapped to one of the 2^{nR} digital codewords $U^n(m)$. Then, each pair $(S^n, U^n(m))$ is mapped symbol-by-symbol to the channel input sequence X^n, which is transmitted over the channel. Upon receiving Y^n and together with the side information T^n, the decoder jointly recovers the digital components $U^n(m)$ by joint typicality, and reconstructs $\hat{S}^n$ by mapping symbol-by-symbol the analog channel output Y^n, the side-information T^n and the decoded digital message. The general conditions for successful decoding of the messages and the achievable distortion D are given in the next lemma, which follows from [36].

Lemma 6. *Let (S,T) be a pair of discrete memoryless sources and $d(s,\hat{s})$ be a distortion measure. A distortion D is achievable for communicating S over a memoryless channel $p(y|x)$ with side-information T available at the decoder if*

$$I(U;S) < I(U;YT) \tag{3.24}$$

for some conditional pdf $p(u|s)$, channel encoding $x(u,s)$ and a reconstruction function $\hat{s}(u,y,t)$, such that $\mathrm{E}[d(S;\hat{S})] \leq D$.

Proof. First, we note that the JSSC setup with side information can be converted into a point-to-point JSCC with two parallel channels: the original channel with channel input $x = (u, s)$ and channel output Y, characterized by $p(y|x)$, and an orthogonal channel corresponding to the side-information, with channel input $x' = s$ and channel output T, characterized by $p(t|x') = p(t|s)$. See Figure 3.1. The proof follows from the point-to-point version of Lemma 6 in the absence of side information, given in [36, Theorem 1]. Let the point-to-point channel $p(\bar{y}|x)$ be given by $\tilde{Y} = (Y, T)$, and $\tilde{X} = (X, S)$, and consider the transmission of S over the point-to-point memoryless channel $p(\tilde{y}|\tilde{x})$, with channel input $\tilde{X}$ and channel output $\tilde{Y}$. From [36, Theorem 1], it follows that an average distortion D is achievable if $I(\tilde{U}; S) \leq I(U; \tilde{Y})$ for some conditional pdf $p(\tilde{u}|s)$, channel encoding $\tilde{x}(\tilde{u}, s)$ and reconstruction function $\hat{s}(\tilde{u}, \tilde{y})$ such that $\mathrm{E}[d(S; \hat{S})] \leq D$. The proof is completed by substituting $\tilde{Y} = (Y, T)$ and $\tilde{X} = (X, S)$ in the sufficient conditions. □

Relying on this result, we propose a particular HDA scheme for the time-varying setup that superposes a coded layer with an uncoded layer and allocates the power among the two layers. The decoder uses joint decoding to recover the quantized codeword using the channel output and the side information sequence. The uncoded component in the channel causes an interference correlated with the source sequence, and thus, acts as side information in the decoding. On the contrary, if an outage occurs and the quantization codeword is not successfully decoded, the analog component provides additional robustness since the channel now contains a noisy uncoded version of the source sequence useful for the reconstruction. This scheme was presented without the uncoded layer in [34] for the static setting, i.e., static channel and static side information available at the receiver, and was shown to be robust against channel SNR mismatch.

The encoder transmits a superposition of digital and analog input signals as

$$X^n = X_d^n + X_a^n, \tag{3.25}$$

where X_d^n and X_a^n are the length-n channel input vectors corresponding to digital and analog input signals, respectively. The analog channel input X_a^n is a scaled version of the source sequence S^n with power P_a, given by $X_a^n = \sqrt{P_a} S^n$.

The digital portion of the transmitted signal X_d^n is generated as follows. We first define the auxiliary random variable $U \triangleq X_d + \eta S$, where X_d is independent of S and distributed as $X_d \sim \mathcal{N}(0, P_d)$, where P_d, is the power allocated to the digital channel input with $P_d = 1 - P_a$; and η and P_d satisfy, for an arbitrarily small $\epsilon > 0$, $R_h = I(U; S) + \epsilon = \frac{1}{2} \log \left(1 + \frac{\eta^2}{P_d}\right) + \epsilon$, i.e., $\eta^2 = P_d(2^{2(R_h - \epsilon)} - 1)$. Then, we generate a codebook of 2^{nR_h} length-n codewords U^n with i.i.d. components according to the auxiliary random variable U. For each source outcome, the encoder determines which of the $2^{n(I(U;S)+\epsilon)}$ codewords U^n in the codebook is jointly typical with S^n, and transmits $X_d^n = U^n - \eta S^n$. For sufficiently large n, a unique U^n satisfies the joint typicality condition with high probability since $R_h > I(U;S)$.

At the decoder, given the channel output Y^n and the side information sequence T^n, the receiver looks for an auxiliary codeword U^n which is simultaneously jointly typical with Y^n and T^n. From Lemma 6, the correct U^n codeword is decoded successfully if,

$$R_h < I(U;Y,T). \tag{3.26}$$

We define the matrix $\mathbf{C}_h \triangleq E[[U,Y,T][U,Y,T]^T]$. We have

$$\mathbf{C}_h = \begin{pmatrix} P_d+\eta^2 & \sqrt{h}(P_d+\eta\sqrt{P_a}) & \eta\sqrt{\gamma} \\ \sqrt{h}(P_d+\eta\sqrt{P_a}) & h(P_d+P_a)+1 & \sqrt{h\gamma P_a} \\ \eta\sqrt{\gamma} & \sqrt{h\gamma P_a} & \gamma+1 \end{pmatrix}.$$

Let $\mathbf{C}_h^{\{2,3\}}$ be defined as the submatrix of $\mathbf{C}_h$ with the first column and first row eliminated. Then, we have

$$\begin{aligned} I(U;Y,T) &= h(U) + h(Y,T) - h(U,Y,T) \\ &= \frac{1}{2}\log\left(\frac{(P_d+\eta^2)|\mathbf{C}_h^{\{2,3\}}|}{|\mathbf{C}_h|}\right) \\ &= \frac{1}{2}\log\left(\frac{\left(1+\gamma+h\left(1+P_d\gamma\right)\right)\left(P_d+\eta^2\right)}{P_d\left(1+\gamma+h\left(\sqrt{P_a}-\eta\right)^2\right)+\eta^2}\right). \end{aligned}$$

An outage will be declared whenever condition (3.26) does not hold due to the randomness of the channel and side information. Hence, the outage event is defined by

$$\mathcal{O}_h \triangleq \{(h,\gamma) : I(U;S) \geq I(U;Y,T)\}, \tag{3.27}$$

and is given by

$$\mathcal{O}_h \triangleq \{(h,\gamma) : P_d h(1+P_d\gamma) \leq P_d(h(\sqrt{P_a}-\eta)^2)+\eta^2\}. \tag{3.28}$$

If U^n is successfully decoded, each S_i is reconstructed using an MMSE estimator with all the information available at the decoder, $\hat{S}_i = \mathrm{E}[S_i|U_i,Y_i,T_i]$. The achievable distortion when U^n is successfully decoded is given by $D_h(P_d,\eta) \triangleq (1+\mathbf{c}^H\mathbf{C}_{\tilde{N}}^{-1}\mathbf{c})^{-1}$, where $\mathbf{c} \triangleq [\beta\sqrt{P_a} + \kappa, \sqrt{P_a}, \gamma]^T$ and

$$\mathbf{C}_{\tilde{N}} = \begin{pmatrix} P_d & P_d\sqrt{h} & 0 \\ P_d\sqrt{h} & hP_d + 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.29}$$

We have

$$D_h(P_d, \eta) = \frac{P_d}{\eta^2 + P_d\left(1 + \gamma + h\left(\sqrt{P_a} - \eta\right)^2\right)}. \tag{3.30}$$

If an outage occurs, the receiver estimates S^n from Y^n and T^n with an MMSE estimator, $\hat{S}_i = \mathrm{E}[S_i|Y_i, T_i]$. The achieved distortion is found to be

$$D_h^{out}(P_d, \eta) \triangleq \left(1 + \frac{hP_a}{1 + hP_d} + \gamma\right)^{-1}. \tag{3.31}$$

Finally, the expected distortion for SHDA is given by

$$ED_{shda}(P_d, \eta) \triangleq \mathrm{E}_{\mathcal{O}_h^c}[D_h(P_d, \eta)] + \mathrm{E}_{\mathcal{O}_h}[D_h^{out}(P_d, \eta)]. \tag{3.32}$$

Optimizing over P_d and η, we obtain $ED^*_{shda} \triangleq \min_{P_d,\eta} ED_{shda}(P_d, \eta)$. Note that uncoded transmission can be recovered from $ED_{shda}(P_d, \eta)$ with $P_d = 0$. The hybrid digital analog (HDA-WZ) scheme of [34] can be recovered by letting $P_a = 0$. We define the minimum expected distortion achievable with HDA-WZ as $ED^*_{hda} \triangleq \min_\eta ED_{shda}(1, \eta)$.

Alternatively to the derivation in Section 3.3.5, the performance of JDS can also be derived by using Lemma 6 with $U = (W, X)$, where $W = S + Q$ with $Q \sim \mathcal{N}(0, \sigma_Q^2)$ and $X \sim \mathcal{N}(0, P)$, and independent of each other, and using X as channel input. While in SHDA the quantization codeword is directly mapped to the channel input, in JDS the channel input is a codeword X independent of the quantization codeword W. Therefore, despite originating from the same general scheme, SHDA and JDS are operationally different. In Section 3.5 and Section 3.6, we observe that in general SHDA performs better than JDS, although the latter outperforms SHDA in certain regimes.

3.4 Optimality of Uncoded Transmission

In addition to separate source and channel coding, uncoded transmission is well known to achieve the minimum distortion in point-to-point static Gaussian channels [26], [27]. However, even in a point-to-point Gaussian channel, in the presence of static side information at the decoder, uncoded transmission becomes suboptimal. In this case, separate source and channel coding, concatenating a Wyner-Ziv source code with a capacity achieving channel code [28], or JSCC through the HDA-WZ scheme in [34] is required to achieve the optimal distortion. Surprisingly, in our setting, when Γ has a continuous and quasiconcave distribution for which $\bar{\gamma} = 0$ is the solution to equation (3.11), uncoded transmission achieves the lower bound ED^*_{pi} in (3.15) for any arbitrarily distributed channel, while both separate source and channel coding and HDA-WZ schemes are suboptimal. The optimality of uncoded transmission follows since, when $\bar{\gamma} = 0$, the side information renders useless for the partially informed lower bound. Then, this lower bound reduces to the fully informed transmitter lower bound of a point-to-point channel without side information, for which uncoded transmission is optimal. Similarly to the other results, the optimality of uncoded transmission in our setting is also sensitive to the source and channel distributions.

Theorem 3. *Let $p_H(h)$ be an arbitrary pdf while $p_\Gamma(\gamma)$ is a continuous and quasiconcave function satisfying equation (3.11) for $\bar{\gamma} = 0$. Then, the minimum expected distortion ED^* is achieved by uncoded transmission.*

Proof. For any pdf satisfying (3.11) with $\bar{\gamma} = 0$, the partially informed encoder lower bound is given by

$$\begin{aligned}
ED^*_{pi} &= \mathrm{E}_H \left[ED^*_Q \left(\frac{1}{2} \log(1+H) \right) \right] \Bigg|_{\bar{\gamma}=0} \\
&\overset{(a)}{=} \int_h \int_0^\infty \frac{p_H(h) p_\Gamma(\gamma)}{2^{\log(1+h)} + \gamma} d\gamma dh \\
&= \iint_{h,\gamma} \frac{p_H(h) p_\Gamma(\gamma)}{1+h+\gamma} d\gamma dh \\
&= ED_u,
\end{aligned}$$

where (a) is obtained by substituting $\bar{\gamma} = 0$ in (3.10). This completes the proof. □

The class of continuous quasiconcave functions for which any non-empty super-level set of $f_\Gamma(\gamma)$ begins at $\gamma = 0$, satisfies $\bar{\gamma} = 0$. It is not hard to see that the class of continuous monotonically decreasing functions in $\gamma \geq 0$ satisfy this condition.

Proposition 1. *Let $p_\Gamma(\gamma)$ be a continuous monotonically decreasing function for $\gamma > 0$. Then, (3.11) holds for $\bar{\gamma} = 0$; and hence, uncoded transmission achieves the optimal performance.*

Proof. By definition, $\bar{\gamma}$ is given by the left endpoint of the super-level set induced by α^*. For any monotonically decreasing function $p_\Gamma(\gamma)$, the left endpoint of the super-level set $\{\gamma : p_\Gamma(\gamma) \geq \alpha\}$ corresponds to $\gamma = 0$, and as a consequence, we have $\bar{\gamma} = 0$ for any value of α^*. □

3.5 Finite SNR Results

In the previous section we have seen the optimality of uncoded transmission when the side information fading state follows a continuous quasiconcave pdf for which $\bar{\gamma} = 0$. The exponential distribution, and the more general family of gamma distributions with shape parameter $L \leq 1$, are continuous monotonically decreasing distributions, and hence, the uncoded transmission is optimal when the side information gain Γ follows one of these distributions. Gamma distributed fading gains appear, for example, when the channel state follows a Nakagami distribution. The gamma distribution with shape parameter L and scale parameter θ, $\Gamma \sim \Upsilon(L, \theta)$, is given as

$$p_\Gamma(\gamma) = \frac{1}{\theta^L}\frac{1}{\Psi(L)}\gamma^{L-1}e^{-\frac{\gamma}{\theta}}, \text{ for } \gamma \geq 0, \text{and } L, \theta > 0, \tag{3.33}$$

where $\Psi(z) \triangleq \int_0^\infty t^{z-1}e^{-t}dt$ is the gamma function. The variance of Γ is $\sigma_\Gamma^2 = L\theta^2$ and its mean is $\mathrm{E}[\Gamma] = L\theta$. When $L \leq 1$, it is easy to check that $p_\Gamma(\gamma)$ is continuous monotonically decreasing, while it is continuous quasiconcave for $L > 1$. Note that when $L = 1$, the gamma distribution reduces to the exponential distribution.

Parameter L models the *side information diversity* since a time-varying side information sequence Y^m, with state distribution $p_\Gamma(\gamma)$, provides the equivalent information (in the sense of sufficient statistics) provided by L independent side information sequences each with i.i.d. Rayleigh block-fading gains. We note that despite the term "diversity", the side information diversity comes from uncoded noisy versions of the source sequence; hence, the gains it provides are limited compared to the channel diversity which can be better exploited through coding.

To illustrate the performance of the achievable schemes and compare them with the lower bounds, we consider Nakagami fading channel and side information distributions. We consider normalized channel and side information gains $H_c = \sqrt{\rho}H_{c0}$ and $\Gamma_c = \sqrt{\rho}\Gamma_{c0}$, such that

$$Y^n = \sqrt{\rho}H_{c0}X^n + N^n, \quad T^n = \sqrt{\rho}\Gamma_{c0}S^n + Z^n,$$

where H_{c0} and Γ_{c0} satisfy $\mathrm{E}[H_{c0}^2] = \mathrm{E}[\Gamma_{c0}^2] = 1$. Basically, H_{c0} and Γ_{c0} capture the randomness in the channels while ρ is the average SNR. We define the associated instantaneous gains $H_0 \triangleq H_{c0}^2$ and $\Gamma_0 \triangleq \Gamma_{c0}^2$.

We assume that the channel gain H_0 has a gamma distribution with scale parameter $L_c > 0$ and $\theta_c = L_c^{-1}$, i.e., $H_0 \sim \Upsilon(L_c, L_c^{-1})$, and similarly, the side information gain follows a gamma distribution with $L_s > 0$ and $\theta_s = L_s^{-1}$, i.e., $\Gamma_0 \sim \Upsilon(L_s, L_s^{-1})$. We have fixed the value of θ_c and θ_s such that $\mathrm{E}[H_{c0}^2] = \mathrm{E}[H_0] = 1$ and $\mathrm{E}[\Gamma_{c0}^2] = \mathrm{E}[\Gamma_0] = 1$, and both channels have the same average SNR ρ for any L_c and L_s. Note that the

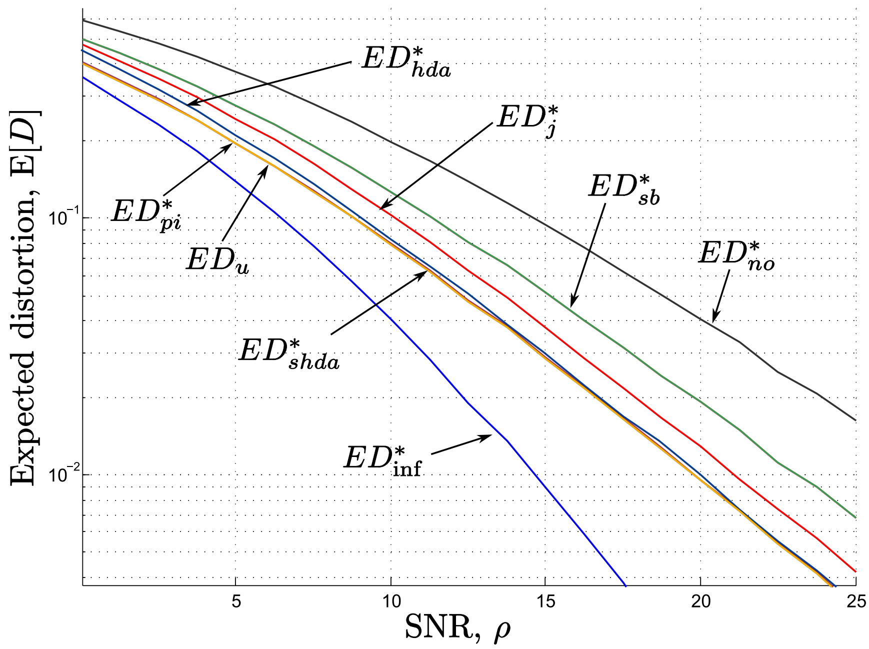


Figure 3.2: Upper and lower bounds on the expected distortion versus the channel SNR (ρ) for Rayleigh fading channel and side information gain distributions, i.e., $L_s = L_c = 1$, with $\rho = \mathrm{E}[H_c^2] = \mathrm{E}[\Gamma_c^2]$.

variance of Γ is $\sigma_\Gamma^2 = L_s\theta^2 = 1/L_s$. Thus, the side information gain Γ becomes more deterministic as L_s increases, and similarly, for L_c and H.

First we consider the case with $L_s = L_c = 1$, i.e., both the channel and the side information gains are Rayleigh distributed. In Fig. 3.2 we plot the expected distortion with respect to the channel SNR. As shown in Theorem 3, uncoded transmission achieves the partially informed encoder lower bound ED^*_{pi}. The minimum expected distortion is given by

$$ED^* = ED_u = \int_{h_0} \frac{1}{\rho} e^{\frac{1+\rho h_0}{\rho}} E_1\left(\frac{1+\rho h_0}{\rho}\right) p_{H_0}(h_0) dh_0. \tag{3.34}$$

We see from the figure that the informed encoder lower bound is significantly loose, especially at high SNR. This gap between the two lower bounds also illustrates the potential performance improvement that will be achieved by increasing the feedback resources. If both channel and side information states can be fed back to the encoder, instead of only CSI feedback, a significant improvement can be achieved. In relation to this observation, a problem that requires further research is the allocation of feedback resources between channel and side information states when a limited feedback channel is available from the decoder to the encoder.

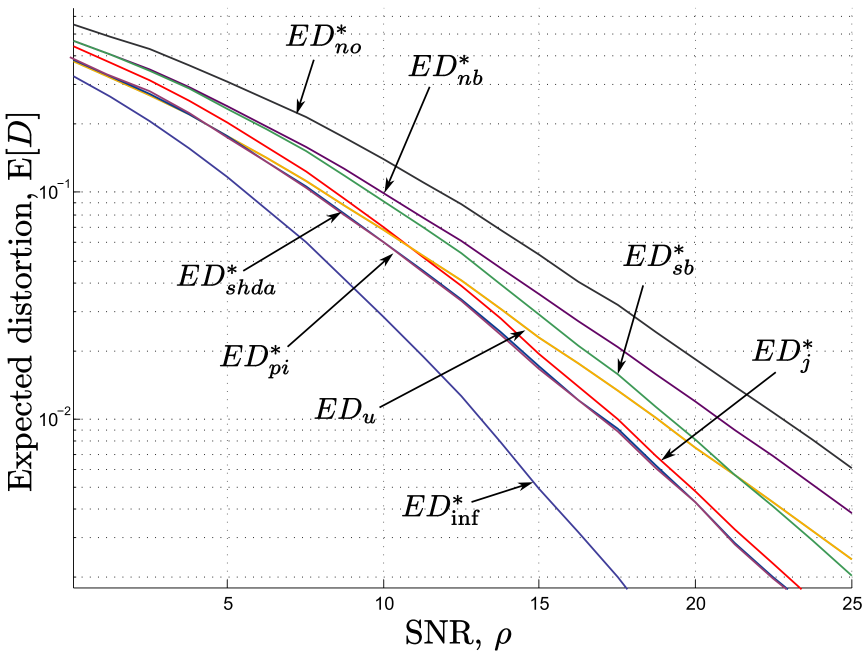


Figure 3.3: Lower and upper bounds on the expected distortion versus the channel SNR for $L_s = 2$ and $L_c = 1$ with $\rho = \mathrm{E}[H_c^2] = \mathrm{E}[\Gamma_c^2]$.

SHDA (ED^*_{shda}) also achieves the optimal performance by allocating all available power to the analog component, reducing it to uncoded transmission. Note that while the HDA-WZ scheme of [34] cannot reach ED^* in the low SNR regime, its performance gets very close to ED^* at high SNR values.

The expected distortion achievable by SSCC is minimized without any binning, since we have $\bar{\gamma} = 0$ for Rayleigh fading side information. Hence, $R_s^* = 0$ from Lemma 4, and therefore $ED^*_{sb} = ED^*_{nb}$. It is interesting to observe that for Rayleigh fading side information states, the uncertainty in the side information renders it useless in transmitting the quantized source codeword, and the side information is ignored to avoid outages in source decoding. The side information is used only in the estimation step. As will be seen next, this is not the case when the side information fading has a different distribution.

We also observe in Fig. 3.2 that JDS (ED^*_j) outperforms SSCC by exploiting the joint quality of the channel and the side information, as claimed by Lemma 5. We also see that JDS cannot achieve the optimal performance in this setting. Observe that the expected distortion achieved by MMSE estimation of the source using only the side information, which we denote by ED^*_{no}, has a constant gap with ED^* in this setup, as well as with the other schemes in the high SNR regime.

Observations above, including the optimality of uncoded transmission, hold for any

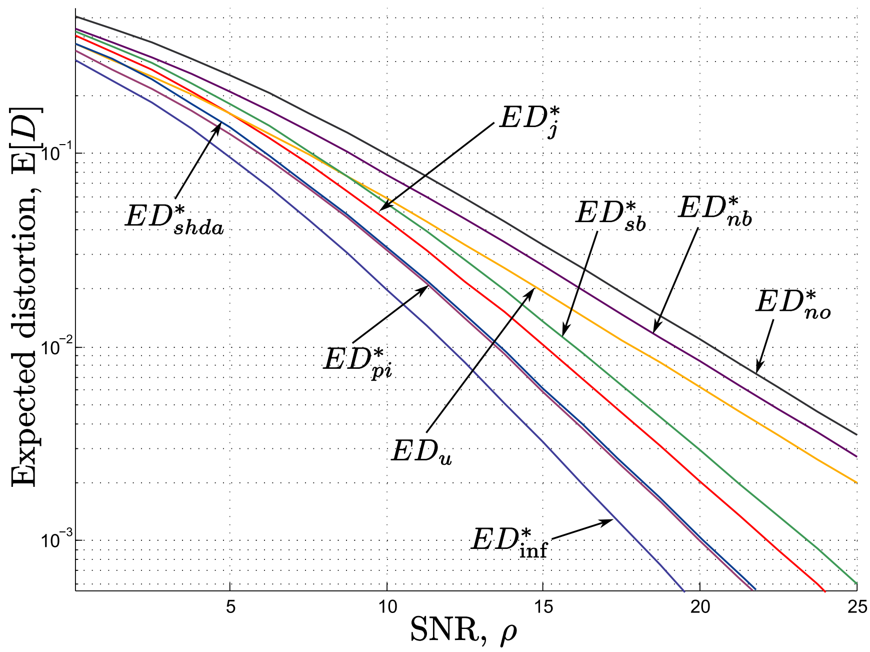


Figure 3.4: Lower and upper bounds on the expected distortion versus the channel SNR for $L_s = 10$ and $L_c = 1$ with $\rho = \mathrm{E}[H_c^2] = \mathrm{E}[\Gamma_c^2]$.

L_c value as long as $L_s \leq 1$. This follows from Proposition 1 since $p_\Gamma(\gamma)$ is monotonically decreasing if $L_s \leq 1$. However, while uncoded transmission is optimal when $L_s \leq 1$, this optimality does not hold in general. Next, it will be shown that for a wide variety of channel distributions, while uncoded transmission is suboptimal, SHDA performs very close to the partially informed encoder lower bound.

We consider the case with $L_s = 2$ and $L_c = 1$ in Fig. 3.3. We can see that SHDA achieves the lowest expected distortion among the proposed schemes and performs very close to the lower bound at all SNR values, while uncoded transmission is suboptimal. Although the performance of uncoded transmission is very close to ED^*_{pi} in the low SNR regime, as the SNR increases, the gap between uncoded transmission and the partially informed encoder bound increases. In addition, both SSCC and JDS surpass the performance of uncoded transmission as the SNR increases. In general, the robustness of uncoded transmission is helpful in the low SNR regime. However, in the high SNR regime uncoded transmission is not capable of exploiting the additional degrees-of-freedom in the system, given by the diversity in the side-information, i.e., when $L_s > 1$, and digital schemes exploit this additional degree-of-freedom better.

We see that SSCC with and without binning both have worse performance than JDS in all SNR regimes and, while at low SNR binning does not provide significant gains, as the SNR increases ED^*_{sb} starts to outperform ED^*_{nb}. On the other hand, ED^*_{nb} lies

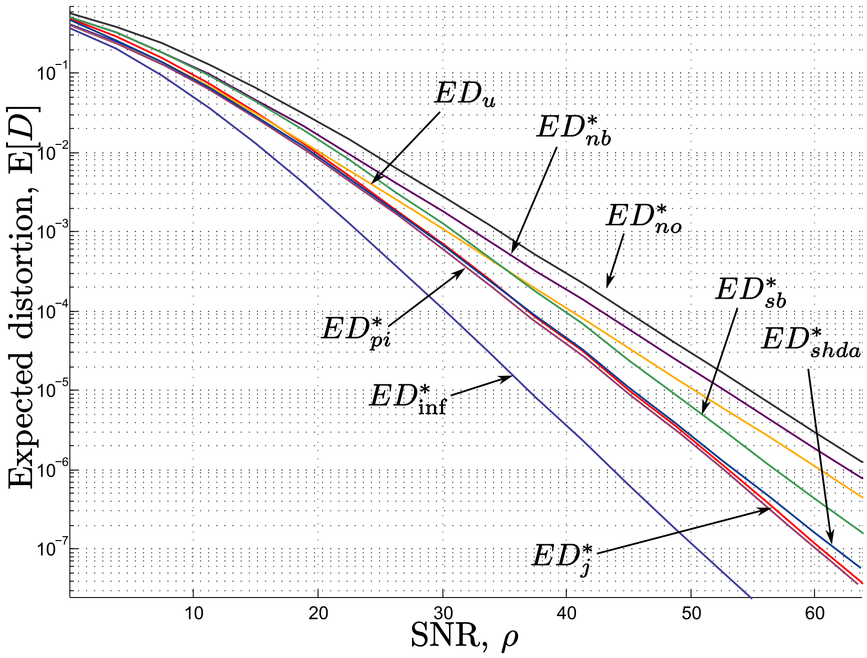


Figure 3.5: Lower and upper bounds on the expected distortion versus the channel SNR for $L_s = 1.5$ and $L_c = 0.5$ with $\rho = \mathrm{E}[H_c^2] = \mathrm{E}[\Gamma_c^2]$.

between ED_u and ED_{no}. These three schemes have the same decay rate and maintain a constant gap. The rate of decay in the high SNR regime is characterized in Section 3.6 for all the proposed schemes.

Similar behavior is observed in Fig. 3.4 for $L_s = 10$ and $L_c = 1$. The minimum distortion among the proposed transmission schemes is achieved by SHDA, which performs very close to the lower bound beyond SNR $\simeq$ 8dB. We can observe that as L_s increases, the performance of uncoded transmission is further away from the lower bound, and JDS outperforms it even at lower SNR values. However, the rate of decay of JDS is worse than the optimal decay in this setting. We also observe that when no binning is considered, the minimum expected distortion achieved by SSCC is still worse than that achieved by uncoded transmission, while the two have the same decay rate in the high SNR regime. However, the use of binning allows SSCC to outperform uncoded transmission, yet ED^*_{sb} is still far from the lower bound.

Finally, in Fig 3.5, we consider $L_c = 0.5$ and $L_s = 1.5$. Contrary to the previous scenarios, in this setup JDS outperforms SHDA for SNR values greater than SNR $\backsimeq$ 37dB. As the SNR increases, JDS performs close to the partially informed lower bound, while SHDA performance is further from the lower bound. Similarly to the previous scenarios, we observe that uncoded transmission performs close to the lower bound for low SNR values and that SSCC achieves lower distortion values if binning is considered.

Observe from Fig. 3.3 and Fig. 3.4 that, as the side information diversity, L_s, increases, the gap at any SNR between the informed encoder lower bound and the partially informed encoder lower bound reduces. The two bounds converge since for the studied setup $\sigma^2_{\Gamma_{c0}} = L_s^{-1}$, and as L_s increases, the variance decreases, and therefore, the level of uncertainty in the available side information gain state drops. In fact, the two bounds can be shown to converge at any SNR value and for any arbitrary side information gain whose variance decreases with some parameter, namely L_s, as given in the next lemma.

Lemma 7. *Let H be arbitrarily distributed and have a finite mean, i.e., $\mathrm{E}_H[H] < \infty$. Let $(\Gamma_L)_{L\geq 0}$ be a sequence of side information gain random variables such that, for every L, Γ_L follows an arbitrary distribution with variance σ^2_L. Assume that $\sigma^2_L \to 0$ for $L \to \infty$. Then, the partially informed encoder lower bound converges to the informed encoder lower bound, i.e., the following limit holds:*

$$\lim_{L\to\infty}(ED_{\mathrm{inf}} - ED^*_{\mathrm{pi}}) = 0. \tag{3.35}$$

Proof. See Appendix C. □

Although the side information available at the decoder becomes more deterministic with increasing L_s, the channel is still block-fading. Only SHDA performs close to the informed encoder lower bound, i.e., the optimal performance when the current channel and side information states are known. On the contrary, the rest of the studied schemes cannot fully exploit the determinism in the side information fading gain for $L_c \geq 1$, while it seems that for $L_c < 1$ JDS is the scheme achieving the lowest expected distortion. The performance of each scheme will be analyzed in the next section in terms of the exponential decay rate of the expected distortion in the high SNR regime.

3.6 High SNR Analysis

In the previous section we have seen the optimality of uncoded transmission in certain settings in which the proposed digital schemes are suboptimal. On the other hand, our numerical results have shown that the SHDA scheme has a good performance for a wide variety of channel distributions while the optimality of uncoded transmission is very sensitive to the distribution of the side information. We have also observed that JDS outperforms SHDA in certain regimes. Although we have characterized the optimal expected distortion in closed-form for the Rayleigh fading scenario in (3.34), a closed-form expression of the optimal expected distortion for general channel and side information distributions is elusive. Instead, we focus on the high SNR regime, and study the exponential decay rate of the expected distortion with increasing SNR, defined as the *distortion exponent*, and denoted by Δ [41]. We have,

$$\Delta \triangleq -\lim_{\rho\to\infty} \frac{\log \mathrm{E}[D]}{\log \rho}. \tag{3.36}$$

The study of the asymptotic behavior of the expected distortion in terms of the distortion exponent does not give exact results for the finite SNR regime. However, it provides relevant information on the average distortion at high SNR values, in terms of the exponential decay. We will see in the numerical evaluation in Section 3.5 that, the expected distortion converges to the asymptotic behavior for not so high SNR values, and therefore, the distortion exponent is a valuable metric to evaluate the performance of the transmission schemes.

In this section, we study the distortion exponent for the model considered in Section 3.5, i.e., a Nakagami fading channel and side information gains, i.e., $H_0 \sim \Upsilon(L_c, L_c^{-1})$ and $\Gamma_0 \sim \Upsilon(L_s, L_s^{-1})$. We are interested in characterizing the maximum distortion exponent over all encoder and decoder pairs, denoted by $\Delta^*(L_s, L_c)$.

We first provide an upper bound on the distortion exponent by studying the partially informed encoder lower bound on the expected distortion in (3.15). In determining the high SNR behavior of the partially informed encoder lower bound, it is challenging to characterize the optimal SNR exponent for the target side information state $\bar{\gamma}$ in (3.11) for different channel states. Hence, we further bound the expected distortion by considering the ergodic channel capacity as the channel rate.

Lemma 8. *The optimal distortion exponent is upper bounded by the exponent of the partially informed encoder lower bound at the ergodic channel capacity, given by*

$$\Delta_{pe}(L_s, L_c) = 1 + \left(1 - \frac{1}{L_s}\right)^+. \tag{3.37}$$

Proof. See Appendix C. □

We will see that $\Delta_{pe}(L_s, L_c)$ is tight only for $L_c \geq 1$, and the ergodic channel relaxation is loose for $L_c < 1$. In order to tighten the bound in these regimes, we consider the distortion exponent of the informed encoder upper lower proposed in Section 3.3.

Lemma 9. *The distortion exponent is upper bounded by the exponent of the informed encoder lower bound, given by*

$$\Delta_{\inf}(L_s, L_c) = \min\{L_c, 1\} + \min\{L_s, 1\}. \tag{3.38}$$

Proof. See Appendix C. □

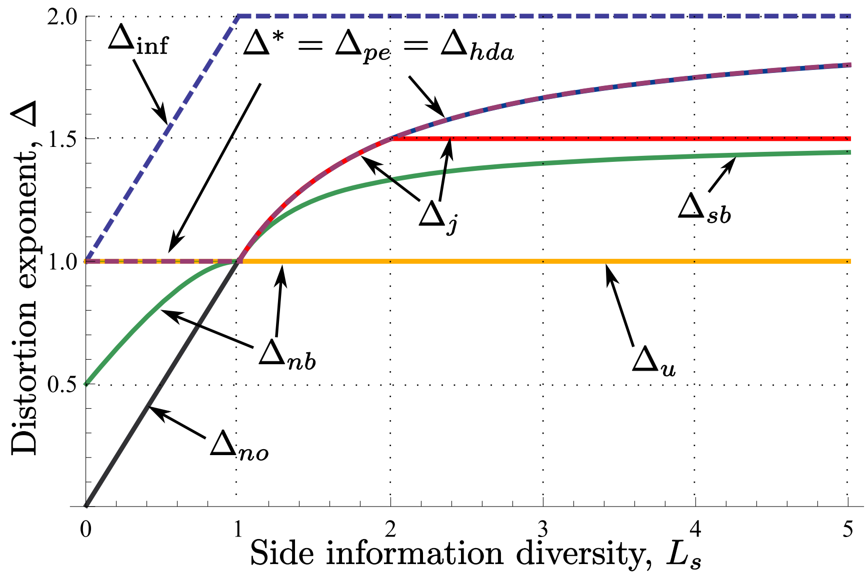


Figure 3.6: Distortion exponent upper and lower bounds for Nakagami fading channel and side information with $L_c = 1$, as a function of L_s.

While for $L_c \geq 1$, $\Delta_{pe}(L_s, L_c)$ is always tighter than $\Delta_{\text{inf}}(L_s, L_c)$, for $L_c < 1$ we have $\Delta_{pe}(L_s, L_c) \geq \Delta_{\text{inf}}(L_s, L_c)$ if $L_s \geq \frac{1}{1-L_c}$. In the next proposition, we combine the two upper bounds into a single upper bound on the distortion exponent.

Theorem 4. *For a Nakagami fading channel with $H_0 \sim \Upsilon(L_c, L_c^{-1})$, and a Nakagami fading side information with $\Gamma_0 \sim \Upsilon(L_s, L_s^{-1})$, the optimal distortion exponent is upper bounded by*

$$\min\{\Delta_{pe}(L_s, L_c), \Delta_{\text{inf}}(L_s, L_c)\} = \begin{cases} \min\{1, L_s + L_c\} & \text{if } L_s \leq 1, \\ 2 - \frac{1}{L_s} & \text{if } 1 < L_s \leq \frac{1}{(1-L_c)^+}, \\ 1 + L_c & \text{if } L_s > \frac{1}{(1-L_c)^+}. \end{cases} \quad (3.39)$$

In Fig. 3.6 and Fig. 3.7 we plot the distortion exponent upper and lower bounds with respect to the parameter L_s of the Nakagami distribution for $L_c = 1$ and $L_c = 0.5$, respectively.

Note that for $L_c \geq 1$, as L_s increases, the optimal distortion exponent $\Delta^*(L_s, L_c)$ converges to the informed encoder upper bound, which is obtained by assuming perfect knowledge of both channel and side information states at the encoder. This observation is parallel to the result in Lemma 7. However, this is not the case if $L_c < 1$. While Lemma 7 applies to any channel distribution, the partially informed bound with ergodic channel relaxation is loose in this regime.

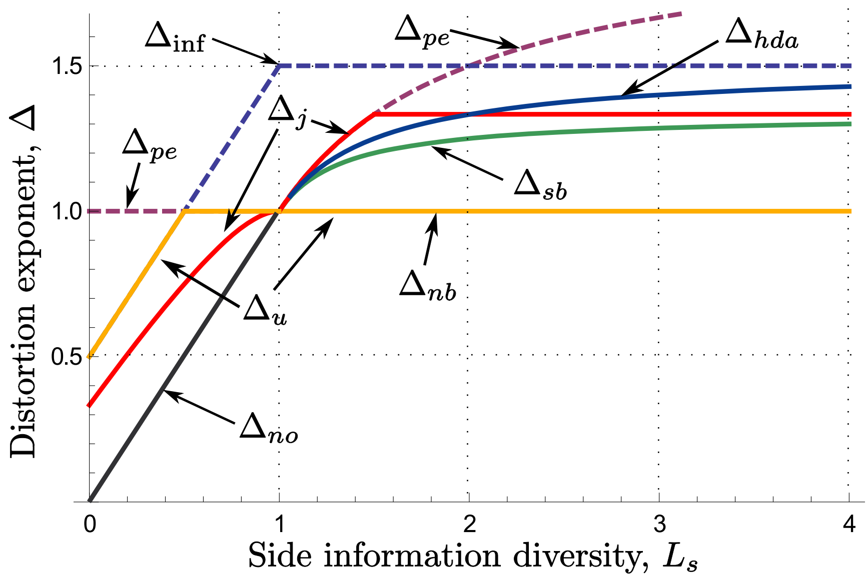


Figure 3.7: Distortion exponent upper and lower bounds for Nakagami fading channel and side information with $L_c = 0.5$, as a function of L_s.

Next, we consider the distortion exponent achievable by the transmission schemes proposed in Section 3.3. The proofs of the corresponding distortion exponent results can be found in Appendix C.

Lemma 10. *The distortion exponent achieved by uncoded transmission is given by*

$$\Delta_u(L_s, L_c) = \min\{L_s + L_c, 1\}. \tag{3.40}$$

As expected from Theorem 3, uncoded transmission achieves the optimal distortion exponent for $L_s \leq 1$. However, it is suboptimal for $L_s > 1$. We note that the distortion exponent of simple MMSE estimation using only the side information sequence, ED_{no}, is given by $\Delta_{no}(L_s, L_c) = \min\{L_s, 1\}$.

Lemma 11. *The distortion exponent achievable by SSCC with binning is given by*

$$\Delta_{sb}(L_s, L_c) = \begin{cases} 1 - \frac{(1-L_s)^2}{L_c+1-L_s} & \text{if } L_s \leq 1, \\ \frac{L_s(2L_c+1)-L_c-1}{L_s(L_c+1)-1} & \text{if } L_s > 1. \end{cases} \tag{3.41}$$

If binning is not used, the achievable distortion exponent is given by

$$\Delta_{nb}(L_s, L_c) = \begin{cases} 1 - \frac{(1-L_s)^2}{L_c+1-L_s} & \text{if } L_s \leq 1, \\ 1 & \text{if } L_s > 1. \end{cases} \tag{3.42}$$

From Lemma 4, we know that binning is suboptimal for $L_s \leq 1$ irrespective of the channel distribution, and both schemes achieve the same distortion exponent in this regime. Note also that when $L_s = 1$, SSCC achieves the optimal distortion exponent of 1. However, when $L_s > 1$, if binning is not used the scheme cannot exploit the side information state properly, and achieves the same distortion exponent as uncoded transmission. This proves that binning is required in this regime.

Lemma 12. *The distortion exponent achievable by JDS is given by*

$$\Delta_{jd}(L_s, L_c) = \begin{cases} 1 - \frac{(1-L_s)^2}{L_c+1-L_s} & \text{if } L_s \leq 1, \\ 2 - \frac{1}{L_s} & \text{if } 1 < L_s \leq 1 + L_c, \\ 1 + \frac{L_c}{L_c+1} & \text{if } L_s > L_c + 1. \end{cases} \tag{3.43}$$

JDS achieves the same distortion exponent as SSCC for $L_s \leq 1$. However, interestingly, for $1 \leq L_s \leq 1 + L_c$, JDS achieves the optimal distortion exponent and then saturates for $L_s > 1 + L_c$. Observe that, as L_s increases, the achievable distortion exponent with SSCC converges to the performance of JDS.

Lemma 13. *The distortion exponent achievable by SHDA and HDA-WZ is given by*

$$\Delta_{shda}(L_s, L_c) = \min\{1, L_s + L_c\} + \frac{\min\{1, L_c\}(L_s - 1)^+}{L_s - 1 + \min\{1, L_c\}}. \tag{3.44}$$

Lemma 13 reveals that the robustness provided by the uncoded layer in SHDA is not required in the high SNR regime to achieve the optimal distortion exponent, and allocating all the available power to the HDA-WZ layer of the SHDA scheme is sufficient. However, we remark that, in terms of the expected distortion in the low SNR regime pure HDA-WZ is not sufficient to achieve a performance close to the lower bound, and the uncoded layer improves the performance in general, as observed in the previous section.

HDA-WZ achieves the optimal distortion exponent for $L_c \geq 1$ while the rest of the proposed schemes are suboptimal. However, when $L_c < 1$, JDS outperforms HDA-WZ for $1 \leq L_s \leq 2$. Nevertheless, as L_s increases, HDA-WZ converges to the distortion exponent of the informed encoder lower bound, despite the uncertainty in the channel state.

We can see that in the limit $L_s \to \infty$, with $0 < L_c \leq 1$, we have

$$\Delta^*(\infty, L_c) = \Delta_{\text{inf}}(\infty, L_c) = \Delta_{hda}(\infty, L_c) = 1 + L_c,$$

whereas

$$\Delta_{sb}(\infty, L_c) = \Delta_j(\infty, L_c) = 1 + \frac{L_c}{L_c + 1} < 1 + L_c.$$

This result suggests that, as the side information fading state becomes more deterministic, the performance of HDA-WZ converges to the informed encoder lower bound, while the rest of the schemes perform significantly worse than HDA-WZ.

Combining the achievable distortion exponents of the JDS and HDA-WZ schemes, we can characterize the optimal distortion exponent $\Delta^*(L_s, L_c)$ in certain regimes, as given next.

Theorem 5. *Consider a Nakagami fading channel with $H_0 \sim \Upsilon(L_c, L_c^{-1})$ and a Nakagami fading side information with $\Gamma_0 \sim \Upsilon(L_s, L_s^{-1})$. If $L_c \geq 1$, the optimal distortion exponent is achieved by the HDA-WZ scheme, and is given by*

$$\Delta^*(L_s, L_c) = 1 + \left(1 - \frac{1}{L_s}\right)^+. \tag{3.45}$$

If $L_c < 1$, and $L_s \leq 1 + L_c$, the optimal distortion exponent is given by

$$\Delta^*(L_s, L_c) = \min\{1, L_s + L_c\} + \left(1 - \frac{1}{L_s}\right)^+, \tag{3.46}$$

and is achieved by uncoded transmission and HDA-WZ when $L_s \leq 1$, and by JDS when $1 \leq L_s \leq L_c + 1$.

These analytical results are in line with the numerical analysis carried out in Section 3.5. For $L_s = L_c = 1$, all the schemes achieve the optimal distortion exponent $\Delta^*(1, 1) = 1$, which is far from the informed encoder upper bound given by $\Delta_{\text{inf}}(1, 1) = 2$, as observed in Fig. 3.2. For $L_s = 2$ and $L_c = 1$, plotted in Fig. 3.3, the optimal distortion exponent is given by $\Delta^*(2, 1) = 3/2$, which is achieved by HDA-WZ, while uncoded transmission is suboptimal since $\Delta_u(2, 1) = 1$. In this case JDS also achieves the optimal distortion exponent, while SSCC with binning achieves a lower distortion exponent of $\Delta_{sb}(2, 1) = 4/3$. As observed in the numerical analysis, if no binning is used, SSCC achieves the same distortion exponent as the uncoded transmission, and the one achieved by using only the side information sequence, i.e., $\Delta_u(2, 1) = \Delta_{nb}(2, 1) = \Delta_{no}(2, 1) = 1$. Although a similar behavior is observed for higher values of L_s, JDS does not achieve the optimal distortion exponent in general. For the case of $L_s = 10$ and $L_c = 1$ plotted in Fig. 3.4, we have $\Delta^*(10, 1) = 19/10$, while $\Delta_{jd}(L_s, 1) = 3/2$ for $L_s \geq 2$. However, when $L_c = 0.5$ and $L_s = 1.5$, plotted in Fig. 3.5, JDS achieves the optimal distortion exponent of $\Delta^*(1.5, 0.5) = 4/3$, while HDA-WZ achieves a smaller distortion exponent given by $\Delta_{shda}(1.5, 0.5) = 5/4$. In this setup the performance of SSCC is improved if binning is used since $\Delta_{sb}(1.5, 0.5) = 6/5$, while if binning is not used we have $\Delta_{sb}(1.5, 0.5) = 1$, which coincides with the distortion exponent of uncoded transmission. In general, we observe that the decay of the expected distortion for finite SNR values converges to the distortion exponent in the asymptotic high SNR regime.

3.7 Conclusions

We have studied the JSCC problem of transmitting a Gaussian source over a delay-limited block-fading channel when block-fading side information is available at the decoder. We have assumed that only the receiver has full knowledge of the channel and side information states while the transmitter is aware only of their distributions. In the case of a static channel, we have shown the optimality of separate source and channel coding when the side information gain follows a discrete or a continuous quasiconcave distribution.

When both the channel and side information states are block-fading, the optimal performance is not known in general. We have proposed achievable schemes based on uncoded transmission, separate source and channel coding, joint decoding and hybrid digital-analog transmission. We have also derived a lower bound on the expected distortion by providing the encoder with the actual channel state. We call this the partially informed encoder lower bound, since the side information state remains unknown to the encoder. We have shown that this lower bound is tight for a certain class of continuous quasiconcave side information fading distributions, and the optimal performance is achieved by uncoded transmission. This, to the best of our knowledge, constitutes the first communication scenario in which the uncoded transmission would be suboptimal in the static setup and is optimal thanks to the existence of fading, while the known digital encoding schemes fall short of the optimal performance. We have also proved that joint decoding outperforms separate source and channel coding since the success of decoding at the receiver depends on the joint quality of the channel and side information states, rather than being limited by each of them separately. We have also shown numerically that hybrid digital-analog transmission performs very close to the lower bound for a wide range of channel and side-information distributions (in particular, we have considered Gamma distributed channel and side information gains with different shape parameters). However, it has also been observed that no unique transmission scheme outperforms others at all cases.

In the high SNR regime, we have obtained closed-form expressions for the distortion exponent, i.e., the optimal exponential decay rate of the expected distortion in the high SNR regime, of the proposed upper and lower bounds for Nakagami distributed channel and side information. Aligned with the numerical results in the finite SNR regime, we have shown that hybrid digital-analog transmission outperforms other schemes in most cases and achieves the optimal distortion exponent for certain values of channel and side information diversity, and joint decoding achieves the optimal distortion exponent for some values of side information diversity when the channel diversity is less than one, in which case hybrid digital-analog transmission is suboptimal.

Chapter 4

Joint Source-Channel Coding with Time-Varying Channel and Side-Information: MIMO

In this chapter, we study a generalization of the problem studied in Chapter 3. We consider the JSCC problem of transmitting a Gaussian source over a multiple-input multiple-output (MIMO) block-fading channel when the receiver has access to a time-varying correlated side information. As in the previous chapter, both the channel and the side-information quality states are assumed to follow block-fading models, whose states are unknown at the transmitter. Strict delay constraints apply, requiring the transmission of a block of source samples, for which the side-information quality state is constant, over a block of the channel, during which the channel state is constant. We assume that the two blocks do not necessarily have the same length, and their ratio is defined as the *bandwidth ratio* between the channel and the source bandwidths, similarly to Section 1.1.1. While in Chapter 3 we have studied the effects of the side information distribution on the expected distortion, here we assume that the side information fading follows a Rayleigh distribution.

The use of multiple antennas at the transmitter and the receiver (MIMO) has been proposed as a viable technology to significantly improve the performance over wireless channels and has been already adopted in many current standards. The use of MIMO provides additional degrees-of-freedom to the system which can be utilized in the form of spatial multiplexing gain and spatial diversity gain. How to translate this additional resources into performance improvements requires a careful design.

When the knowledge of the channel and side information states is available at both the transmitter and the receiver (CSI-TR), Shannon's separation theorem applies [28], that is, the optimal performance is achieved by first compressing the source with an optimal source code and transmitting the compressed bits with a capacity achieving channel code. However, as in Chapter 3, the optimality of source-channel separation does not extend to non-ergodic scenarios such as the model studied in this chapter.

This problem has been studied extensively in the literature for MIMO channels, mismatched bandwidth and in the absence of correlated side information at the receiver [42, 59, 72]. Despite the ongoing efforts, the minimum achievable average distortion remains an open problem; however, as observed in the previous chapter, more conclusive results on the performance can be obtained by studying the *distortion exponent*, which characterizes the exponential decay of the expected distortion in the high SNR regime [61]. In the absence of side information at the receiver, the optimal distortion exponent in MIMO channels is known in some regimes of operation, such as the large bandwidth regime [15] and the low bandwidth regime [16]. However, the general problem remains open. In [15] digital multi-layer superposition transmission schemes are shown to achieve the optimal distortion exponent for high bandwidth ratios in MIMO systems. The optimal distortion exponent in the low bandwidth regime is achieved through hybrid digital-analog transmission [15,16]. In [39], superposition multi-layer schemes are shown to achieve the optimal distortion exponent for some other bandwidth ratios. Overall, multi-layer transmission has been shown to achieve the largest distortion exponents among the existing schemes in the literature.

In this chapter, our goal is to find tight bounds on the distortion exponent when transmitting a Gaussian source over a time-varying MIMO channel in the presence of time-varying correlated side information at the receiver. We first derive upper bounds on the distortion exponent by providing the channel state information to the encoder. Then, we consider single layer encoding schemes based on separate source and channel coding (SSCC), joint decoding (JDS), uncoded transmission and hybrid digital-analog transmission. As shown in Chapter 3, in the SISO and matched bandwidth model, uncoded transmission achieves the minimum expected distortion for certain side information fading gain distributions. However, in the presence of additional degrees-of-freedom provided by the MIMO channel and the available bandwidth, SSCC, JDS and HDA schemes are expected to better exploit the additional resources and significantly outperform uncoded transmission, specially in terms of the distortion exponent. On the other hand, motivated by the improvements provided by multi-layer transmission in [15], we then consider two different multi-layer joint decoding schemes based on successive refinement of the source followed either by progressive transmission over the channel (LS-JDS), or by superposing JDS codes in a broadcast fashion (BS-JDS), and show that these schemes achieve the best distortion exponents.

The main results of this chapter can be summarized as follows:

- We first derive an upper bound on the distortion exponent by providing both the channel and the side information states to the encoder. Then, a tighter upper bound is obtained by providing only the channel state to the encoder.
- We characterize the distortion exponent achieved by JDS. While this scheme achieves a lower expected distortion than SSCC, we show that it does not improve the distortion exponent.
- We then consider a hybrid digital-analog scheme (HDA-WZ) that combines JDS with an analog layer. We show that HDA-WZ outperforms JDS not only in terms of the average distortion, but also the distortion exponent.
- We extend JDS by considering multi-layer transmission, where each layer carries successive refinement information for the source sequence. We consider both the progressive (LS-JDS) and superposition (BS-JDS) transmission of these layers, and derive the respective achievable distortion exponent expressions.
- We show that BS-JDS achieves the optimal distortion exponent for SISO/SIMO/MISO systems, thus characterizing the optimal distortion exponent in these scenarios. We also show that HDA-WZ achieves the optimal distortion exponent in SISO channels as well.
- In the general MIMO setup, we characterize the optimal distortion exponent in the low bandwidth ratio regime, and show that it is achievable by both HDA-WZ and BS-JDS. In addition, we show that in certain regimes of operation, LS-JDS outperforms all the other proposed schemes.

The rest of the chapter is organized as follows. The problem statement is given in Section 4.1. Then, known results on the diversity multiplexing tradeoff are provided in Section 4.2. Two upper bounds on the system performance are derived in Section 4.3 and the optimal distortion exponent in the low bandwidth regime is discusses in Section 4.4. Various single-layer achievable schemes are studied in Section 4.5, while multi-layer schemes are considered in Section 4.6. The characterization of the optimal distortion exponent for certain regimes is relegated to Section 4.7. Finally, the conclusions are presented in Section 4.8.

4.1 System Model

We wish to transmit a zero mean, unit variance real Gaussian source sequence $S^m \in \mathbb{R}^m$ of independent and identically distributed (i.i.d.) random variables, i.e., $S_i \sim \mathcal{N}(0,1)$, over a complex MIMO block Rayleigh-fading channel with M_t transmit and M_r receiver antennas, as shown in Figure 4.1. In addition to the channel output, time-varying correlated source side information is also available at the decoder. Time-variations in the source side information are assumed to follow a block fading model as well. The

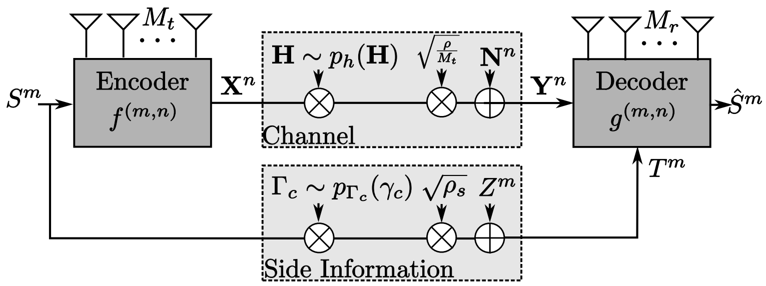


Figure 4.1: Block diagram of the joint source-channel coding problem with fading channel and side information qualities.

channel and the side information states are assumed to be constant for the duration of one block, and independent of each other, and among different blocks. We assume that each source block is composed of m source samples, which, due to the delay limitations of the underlying application, are supposed to be transmitted over one block of the channel, which consists of n channel uses. We define the *bandwidth ratio* of the system as[1]

$$b \triangleq \frac{2n}{m} \quad \textit{complex channel dimension per real source sample.}$$

The encoder maps each source sequence S^m to a channel input sequence $\mathbf{X}^n \in \mathbb{C}^{M_t \times n}$ using an encoding function $f^{(m,n)} : \mathbb{R}^m \to \mathbb{C}^{M_t \times n}$ such that the average power constraint is satisfied: $\sum_{i=1}^{n} \mathrm{Tr}\{\mathrm{E}[\mathbf{X}_i^H \mathbf{X}_i]\} \leq M_t n$. The memoryless slow fading channel is modeled as

$$\mathbf{Y}_i = \sqrt{\frac{\rho}{M_t}} \mathbf{H} \mathbf{X}_i + \mathbf{N}_i, \qquad i = 1, ..., n,$$

where $\mathsf{H} \in \mathbb{C}^{M_r \times M_t}$ is the channel matrix with i.i.d. zero mean complex Gaussian entries, i.e., $h_{ij} \sim \mathcal{CN}(0,1)$, whose realizations are denoted by $\mathbf{H}$, $\rho \in \mathbb{R}^+$ is the average signal to noise ratio (SNR) in the channel, and $\mathbf{N}_i$ models the additive noise with $\mathbf{N}_i \sim \mathcal{CN}(0, \mathbf{I})$. We define $M^* = \max\{M_t, M_r\}$ and $M_* = \min\{M_t, M_r\}$, and consider $\lambda_{M_*} \geq \cdots \geq \lambda_1 > 0$ to be the eigenvalues of $\mathbf{H}\mathbf{H}^H$.

In addition to the channel output $\mathbf{V}^n = [\mathbf{V}_1, ..., \mathbf{V}_n] \in \mathbb{C}^{M_r \times n}$, the decoder observes $T^m \in \mathbb{R}^m$, a randomly degraded version of the source sequence:

$$T^m = \sqrt{\rho_s} \Gamma_c S^m + Z^m,$$

where Γ_c models Rayleigh fading[2] in the quality of the side information satisfying

[1] This scaled definition is done for consistency of results with previous works in the distortion exponent literature, which use real/real or complex/complex sources and channels[15].

[2] The assumption of a real source sequence X^m and a real fading coefficient Γ_c is made in order to allow a degradation model possible. That is, the side-information qualities can be ordered among different channel states. Complex source and fading side information sequences would not allow an ordering in the quality of the side information sequences.

$\mathrm{E}[\Gamma_c^2] = 1$, $\rho_s \in \mathbb{R}^+$ models the average quality of the side information, and $Z_j \sim \mathcal{N}(0,1)$, $j = 1, ..., m$, models the noise. We define the *side information gain* as $\Gamma \triangleq \Gamma_c^2$, and its realization as γ. Then, Γ follows an exponential distribution with probability density function (pdf):

$$p_\Gamma(\gamma) = e^{-\gamma}, \qquad \gamma \geq 0.$$

In this work, we assume that the receiver knows the side information and the channel realizations, γ and $\mathbf{H}$, while the encoder is only aware of their distributions. The decoder reconstructs the source sequence $\hat{S}^m = g^{(m,n)}(\mathbf{Y}^n, T^m, \mathbf{H}, \gamma)$ with a mapping $g^{(m,n)}$: $\mathbb{C}^{n\times M_r} \times \mathbb{R}^m \times \mathbb{C}^{M_t \times M_r} \times \mathbb{R} \to \mathbb{R}^m$. The distortion between the source sequence and the reconstruction is measured by the quadratic average distortion $D \triangleq \frac{1}{m}\sum_{i=1}^m (S_i - \hat{S}_i)^2$.

We are interested in characterizing the minimum *expected distortion*, $\mathrm{E}[D]$, where the expectation is taken with respect to the source, the side information and channels state realizations, as well as the noise terms, and expressed as

$$ED^*(\rho, \rho_s, b) \triangleq \lim_{\substack{n,m\to\infty,\\ 2n\leq mb.}} \min_{f^{(m,n)}, g^{(m,n)}} \mathrm{E}[D].$$

In particular, we are interested in characterizing the optimal performance in the high SNR regime, i.e., when $\rho, \rho_s \to \infty$. We define x as a measure of the average *side information quality* in the high SNR regime, as follows:

$$x \triangleq \lim_{\rho\to\infty} \frac{\log \rho_s}{\log \rho}.$$

The performance measure we consider is the *distortion exponent*, defined as

$$\Delta(b, x) \triangleq -\lim_{\rho,\rho_s\to\infty} \frac{\log \mathrm{E}[D]}{\log \rho},$$

4.2 Diversity-Multiplexing tradeoff

Here we depart shortly from the distortion exponent problem introduces above, and briefly talk about another, more commonly used, performance measure in the high SNR regime, that will be instrumented in our analysis. The diversity-multiplexing tradeoff (DMT) measures the tradeoff between the rate and reliability in the transmission of a message over a MIMO fading channel in the asymptotic high SNR regime. Hence, the DMT is a performance measure for the channel coding problem over block-fading channels. In this section we briefly review some known results on the DMT, which will be useful in the distortion exponent analysis. We refer the reader to [73] for a more detailed exposition of the DMT.

For a family of channel codes with rate $R = r \log \rho$, where r is the *multiplexing gain*, the *diversity gain* is defined as

$$d(r) = -\lim_{\rho \to \infty} \frac{\log P_e(\rho)}{\log \rho},$$

where $P_e(\rho)$ is the probability of decoding error of the channel code. For each r, the supremum of the diversity gain $d(r)$ over all coding schemes is given by $d^*(r)$. The DMT for a MIMO channel is given as the solution to the following problem [73],

$$\begin{aligned} d^*(r) &= \inf_{\boldsymbol{\alpha}^+} \sum_{i=1}^{M_*} (2i - 1 + M^* - M_*)\alpha_i \\ &\text{s.t. } r \geq \sum_{i=1}^{M_*} (1 - \alpha_i), \end{aligned} \tag{4.1}$$

where $\boldsymbol{\alpha}^+ \triangleq \{(\alpha_1, ..., \alpha_{M_*}) \in \mathbb{R}^{M_*} : 1 \geq \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0\}$. The DMT obtained from (4.1) is a piecewise-linear function connecting the points $(k, d^*(k))$, $k = 0, ..., M_*$, where $d^*(k) = (M^* - k)(M_* - k)$. More specifically, for $r \geq M_*$, we have $d^*(r) = 0$, and for $0 \leq r \leq M_*$ satisfying $k \leq r \leq k + 1$ for some $k = 0, 1, ..., M_* - 1$, the DMT curve is characterized by

$$d^*(r) \triangleq \Phi_k - \Upsilon_k (r - k), \tag{4.2}$$

where we have defined

$$\Phi_k \triangleq (M^* - k)(M_* - k) \quad \text{and} \quad \Upsilon_k \triangleq (M^* + M_* - 2k - 1). \tag{4.3}$$

4.3 Distortion Exponent Upper Bound

In this section we derive two upper bounds on the distortion exponent by extending the two bounds on the expected distortion ED^* obtained in Chapter 3 to the MIMO setup with bandwidth mismatch, and analyzing their high SNR behavior.

4.3.1 Fully informed encoder upper bound

Following Chapter 3, the first upper bound, which we denote as the *fully informed encoder upper bound*, is obtained by providing the transmitter with both the channel state $\mathbf{H}$ and the side information state γ. At each realization, the problem reduces to the static setup studied in [28], and source-channel separation theorem applies; that is, the concatenation of a Wyner-Ziv source code with a capacity achieving channel code is optimal at each realization. Averaging the achieved distortion over the realizations of the channel and side information states, the expected distortion is found as

$$ED_{\text{inf}}(\rho, \rho_s, b) = \mathrm{E}_{\mathbf{H},\Gamma}\left[\frac{1}{1+\rho_s\gamma}2^{-b\mathcal{C}(\mathbf{H})}\right],$$

where $\mathcal{C}(\mathbf{H})$ is the capacity of the MIMO channel in bits/channel use.

Following similar derivations in [15] and the Appendix of Chapter 3, we find an upper bound on the distortion exponent, stated in the following lemma.

Lemma 14. *The distortion exponent is upper bounded by the informed encoder upper bound, given by*

$$\Delta_{\text{inf}}(x, b) = x + \Delta_{\text{MIMO}}(b), \tag{4.4}$$

where

$$\Delta_{\text{MIMO}}(b) \triangleq \sum_{i=1}^{M_*} \min\{b, 2i - 1 + M^* - M_*\}. \tag{4.5}$$

4.3.2 Partially informed encoder upper bound

As in Chapter 3, a tighter upper bound can be constructed by providing the transmitter with only the channel state realization $\mathbf{H}$ while the side information state γ remains unknown. We call this the *partially informed encoder upper bound*. The optimality of separate source and channel coding is shown in Chapter 3 when the side information fading gain distribution is discrete, or continuous and quasiconcave for $b = 1$. The proof easily extends to the non-matched bandwidth ratio setup and, since in our model $p_\Gamma(\gamma)$ is exponential, and hence, is continuous and quasiconcave, separation is optimal at each channel block.

As shown in Section 3.4, if $p_\Gamma(\gamma)$ is monotonically decreasing, the optimal source encoder ignores the side information completely, and the side-information is used only at the decoder for source reconstruction[3]. Concatenating this side-information-ignorant source code with a channel code at the instantaneous capacity, the minimum expected distortion at each channel state $\mathbf{H}$ is given by

$$D_{\text{op}}(\rho, \rho_s, b, \mathbf{H}) = \frac{1}{\rho_s} e^{\frac{2^{b\mathcal{C}(\mathbf{H})}}{\rho_s}} E_1\left(\frac{2^{b\mathcal{C}(\mathbf{H})}}{\rho_s}\right),$$

where $E_1(x)$ is the exponential integral given by $E_1(x) = \int_x^\infty t^{-1}e^t dt$. Averaging over

[3]We note that when the distribution of the side information is not Rayleigh, the optimal encoder follows a different strategy. For example, for quasiconcave continuous distributions the optimal source code compresses the source aiming at a single target side information state. See Chapter 3 for details.

the channel state realizations, the expected distortion is lower bounded as

$$ED^*_{\text{pi}}(\rho, \rho_s, b) = \mathrm{E}_{\mathbf{H}}[D_{\text{op}}(\rho, \rho_s, b, \mathbf{H})]. \tag{4.6}$$

An upper bound on the distortion exponent is found by analyzing the high SNR behavior of (4.6) as given in the next theorem.

Theorem 6. *For $M_*(2(l-1)-1+M^*-M_*) \leq x < M_*(2l-1+M^*-M_*)$ let $l \in [1, ..., M_*-1]$ be the integer satisfying the inequality, and let $l = 1$ for $x \leq M_*(M^*+M_*-1)$. Then, for $2k-1+M^*-M_* \leq b < 2k+1+M^*-M_*$, $k = l, ..., M_*-1$, the distortion exponent is upper bounded by*

$$\Delta_{up}(x, b) = \begin{cases} x & \text{if } 0 \leq b < \frac{x}{M_*}, \\ bM_* & \text{if } \frac{x}{M_*} \leq b < M^* - M_* + 1, \\ x + d^*\left(\frac{x}{b}\right) & \text{if } M^* - M_* + 1 \leq b < \frac{x}{M_*-k}, \\ \Delta_{\text{MIMO}}(b) & \text{if } \frac{x}{M_*-k} \leq b < M^* + M_* - 1, \\ x + d^*\left(\frac{x}{b}\right) & \text{if } b \geq M^* + M_* - 1. \end{cases}$$

If $x \geq M_*(M^*+M_*-1)$, then,

$$\Delta_{up}(b, x) = x + d^*\left(\frac{x}{b}\right),$$

where $d^*(r)$ is the DMT characterized in (4.2)-(4.3).

Proof. The proof is given in Appendix D. □

By comparing the two upper bounds in Lemma 14 and Theorem 6, we can see that the latter is always tighter. When $x > 0$, the two bounds meet only at the two extremes, when either $b = 0$ or $b \to \infty$. Note that these bounds provide the achievable distortion exponents when either both states or only the channel state is available at the transmitter, illustrating the gains from the channel state feedback in fading JSCC problems.

4.4 Optimal distortion exponent in the low bandwidth regime

In this section we use the upper bound derived in the previous section to characterize the optimal distortion exponent in the low bandwidth regime, i.e., $0 \leq bM_* \leq 1$. We show that, if the available bandwidth is small, the optimal distortion exponent is achieved by ignoring the channel and reconstructing the source sequence using only the side information. However, if more bandwidth is available, the optimal distortion exponent is achieved by ignoring the side information, and employing the optimal transmission scheme in the absence of side information.

First, we consider the MMSE reconstruction of S^m only from the side information sequence T^m available at the receiver, i.e., $\hat{S}_i = \mathrm{E}[S_i|T_i]$. The source sequence is reconstructed with distortion $D_{no}(\gamma) \triangleq (1+\rho_s\gamma)^{-1}$, and averaging over the side information realizations, the expected distortion is given by $ED_{no} = \mathrm{E}[D_{no}(\Gamma)]$. The achievable distortion exponent is found as $\Delta_{no}(x,b) = x$, which meets the upper bound $\Delta_{up}(x,b)$ for $0 \leq bM_* \leq x$, characterizing the optimal distortion exponent in this regime.

Lemma 15. *For $0 \leq bM_* \leq x$, the optimal distortion exponent $\Delta^*(b,x) = x$ is achievable by simple MMSE reconstruction of S^m only from the side information sequence T^m.*

Additionally, Theorem 6 reveals that in certain regimes, the distortion exponent is upper bounded by $\Delta_{\mathrm{MIMO}}(b)$, the distortion exponent upper bound in the absence of side information at the destination [15, Theorem 3.1]. In fact, for $x \leq bM_* \leq 1$, we have $\Delta_{up}(x,b) = bM_*$, which is achievable by ignoring the side information and using the hybrid digital-analog scheme proposed in [16]. In this scheme, which we denote by superposed HDA (HDA-S), the source sequence is divided and transmitted using two layers. The first layer transmits a part of the source sequence in an uncoded fashion, while the second layer digitally transmits the second sequence part. Both layers are superposed and the available power is allocated among them to maximize the achievable distortion exponent. At the destination, the digital layer is decoded treating the uncoded layer as noise. Then, the source sequence is reconstructed using both layers. The achievable distortion exponent is given by $\Delta_{sh}(x,b) = bM_*$ for $0 \leq bM_* \leq 1$.

Lemma 16. *For $x \leq bM_* \leq 1$, the optimal distortion exponent is given by $\Delta^*(b,x) = bM_*$, and is achievable by ignoring the side information sequence T^m and using HDA-S.*

In larger bandwidth ratio regimes, i.e., for $bM_* > 1$, transmission schemes using both the channel and the side information available are required.

4.5 Single layer transmission

In this section, we propose transmission schemes consisting of single layer code, and analyze their achievable distortion exponent performance. The illustration of the achievable distortion exponents and its comparison between transmission techniques and the proposed upper bound is deferred to Section 4.5.5.

4.5.1 Separate source and channel coding scheme (SSCC)

In this section we consider the generalization of the SSCC scheme considered in Chapter 3 to MIMO channels and general bandwidth ratios. As in the single antenna setup, the transmission suffers from two separate outage events: outage in channel decoding and outage in source decoding. It is shown in Corollary 1 in Section 3.3.4 that, for monotonically decreasing pdfs, such as $p_\Gamma(\gamma)$ considered here, the expected distortion is minimized by avoiding outage in source decoding, that is, by not using binning. Therefore, the optimal SSCC scheme compresses the source sequence at rate R_s ignoring the side information, and transmits the compressed bits over the channel with a channel code with rate R_c such that $\frac{b}{2}R_c = R_s$.

At the encoder, the quantization codebook consists of 2^{mR_s} length-m codewords, $W^m(i)$, $i = 1, ..., 2^{mR_s}$, generated through a 'test channel' given by $W = S + Q$, where $Q \sim \mathcal{N}(0, \sigma_Q^2)$, and is independent of S. The quantization noise variance is such that $R_s = I(S;W) + \epsilon$, for an arbitrarily small $\epsilon > 0$, i.e., $\sigma_Q^2 = (2^{2(R_s-\epsilon)} - 1)^{-1}$. For the channel code, a Gaussian channel codebook with 2^{nRc} length-n codewords $\mathbf{X}^n(s)$ is generated independently with $\mathbf{X} \sim \mathcal{CN}(0, \mathbf{I})$, and each codeword $\mathbf{X}^n(s)$, $s \in [1, ..., 2^{nR_c}]$, is assigned to a quantization codeword $W^m(i)$. Given a source sequence S^m, the encoder searches for a quantization codeword $W^m(i)$ jointly typical with S^m, and transmits the corresponding channel codeword $\mathbf{X}(i)$.

The decoder recovers the digital codeword with high probability if $R_c < I(\mathbf{X}, \mathbf{Y})$. An outage is declared whenever due to the channel randomness, the channel rate R_c is above the capacity and the codeword cannot be recovered. Then, the outage event is given by

$$\mathcal{O}_s = \{\mathbf{H} : R_c \geq I(\mathbf{X}; \mathbf{Y})\}, \tag{4.7}$$

where $I(\mathbf{X}; \mathbf{Y}) = \log\det(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}\mathbf{H}^H)$.

If W^m is successfully decoded, the source sequence is estimated with a MMSE estimator using the quantization codeword and the side information sequence, i.e., $\hat{S}_i = \mathrm{E}[S_i|W_i, T_i]$, and reconstructed with a distortion $D_d(bR_c/2, \gamma)$, where

$$D_d(R, \gamma) \triangleq (\rho_s\gamma + 2^{2R})^{-1}. \tag{4.8}$$

If there is an outage over the channel, only the side information is used in the source reconstruction and the corresponding distortion is given by $D_d(0, \gamma)$. The probability of outage depends only on the channel state $\mathbf{H}$. The expected distortion for SSCC can be written as

$$ED_s(bR_c) = \mathrm{E}_{\mathcal{O}_s^c}[D_d(bR_c/2,\Gamma)] + \mathrm{E}_{\mathcal{O}_s}[D_d(0,\Gamma)] \tag{4.9}$$
$$= (1-P_o(\mathbf{H}))\mathrm{E}_\Gamma[D_d(bR_c/2,\Gamma)] + P_o(\mathbf{H})\mathrm{E}_\Gamma[D_d(0,\Gamma)],$$

where $P_o(\mathbf{H}) \triangleq \Pr\{R_c \geq \log\det(\mathbf{I} + \frac{\rho}{M_*}\mathbf{HH}^H)\}$ is the channel probability of outage.

In the next theorem, the distortion exponent achievable by SSCC is provided.

Theorem 7. *The achievable distortion exponent for SSCC, $\Delta_s(b,x)$, is given by*

$$\Delta_s(b,x) = \max\left\{x, b\frac{\Phi_k + k\Upsilon_k + x}{\Upsilon_k + b}\right\}, \text{ for } b \in \left[\frac{\Phi_{k+1}+x}{k+1}, \frac{\Phi_k + x}{k}\right), k = 0,1,...,M_*-1,$$

where Φ_k and Υ_k are as defined in (4.3).

Proof. See Appendix D □

4.5.2 Joint Decoding Scheme (JDS)

In this section we consider the generalization of the JDS scheme considered in Chapter 3 to MIMO channels and general bandwidth ratios, which, by joint decoding of the channel and the source codewords, reduces the outage probability. It uses no explicit binning at the encoding, and the success of decoding depends on the joint quality of the channel and the side information. In the previous chapter, JDS is shown to outperform SSCC at any SNR and to achieve the optimal distortion exponent in certain regimes.

At the encoder, we generate a codebook of 2^{mR_j} length-m quantization codewords $W^m(i)$ and an independent Gaussian codebook of size $2^{n\frac{b}{2}R_j}$ with length-n codewords $\mathbf{X}(i) \in \mathbb{C}^{M_t \times n}$ with $\mathbf{X} \sim \mathcal{CN}(0,\mathbf{I})$, such that $\frac{b}{2}R_j = I(S;W) + \epsilon$, for an arbitrarily small $\epsilon > 0$. Given a source outcome S^m, the transmitter finds the quantization codeword $W^m(i)$ jointly typical with the source outcome and transmits the corresponding channel codeword $\mathbf{X}(i)$. Joint typicality decoding is performed such that the decoder looks for an index i for which both $(\mathbf{X}^n(i), \mathbf{Y}^n)$ and $(T^m, W^m(i))$ are jointly typical. Then the outage event is

$$\mathcal{O}_j = \left\{(\mathbf{H},\gamma) : I(S;W|T) \geq \frac{b}{2}I(\mathbf{X};\mathbf{Y})\right\}, \tag{4.10}$$

where $I(\mathbf{X};\mathbf{Y}) = \log\det(\mathbf{I} + \frac{\rho}{M_*}\mathbf{HH}^H)$ and $I(S;W|T) = \frac{1}{2}\log(1 + \frac{2^{R_j-\epsilon}-1}{\gamma\rho_s+1})$.

Similarly to SSCC, if there is no outage the source is reconstructed using the quantization codeword and the side information sequence with an MMSE estimator, while only the side information is used in case of an outage.

The joint decoding produces a binning-like decoding: only some $\mathbf{Y}^n$ are jointly typical with $\mathbf{X}(s)$, generating a virtual bin of W^m codewords from which only one is jointly typical with T^m. The size of those bins depends on the particular realizations of $\mathbf{H}$ and Γ unlike in a Wyner-Ziv scheme, in which the bin sizes are designed in advance. Since the outage event depends jointly on the channel and the side information states $(\mathbf{H}, \gamma)$, the expectation over the states is not separable as in (4.9). Then, the expected distortion for JDS is expressed as

$$ED_j(R_j) = \mathrm{E}_{\mathcal{O}_j^c}\left[D_d\left(\frac{b}{2}R_j, \Gamma\right)\right] + \mathrm{E}_{\mathcal{O}_j}[D_d(0, \Gamma)].$$

JDS reduces the probability of outage, and hence, the expected distortion compared to SSCC. However, both schemes achieve the same distortion exponent, as stated in the following theorem.

Theorem 8. *JDS achieves the same distortion exponent as SSCC characterized in Theorem 7, i.e.,* $\Delta_j(b,x) = \Delta_s(b,x)$.

Proof. See Appendix D. □

Although JDS and SSCC achieve the same distortion exponent in the current setting, JDS is shown to achieve larger distortion exponents than SSCC in general in Chapter 3. A comparison between the two schemes is deferred to Section 4.5.5.

4.5.3 Uncoded transmission

Uncoded transmission has been considered in Chapter 3, and shown to be exactly optimal in terms of the expected distortion when the side information gain follows a monotonically decreasing distribution function, such as $p_\Gamma(\gamma)$ in our model. However, for general MIMO channels and bandwidth ratios, it falls short of the optimal performance, since it cannot fully exploit the additional degrees-of-freedom in the system.

In uncoded transmission, the source samples are used directly as the channel inputs. Since the channel is complex, we reorder the source sequence as $\mathbf{S}_c^{\frac{m}{2}} \in \mathbb{C}^{\frac{m}{2}}$ given by

$$\mathbf{S}_c^{\frac{m}{2}} = \frac{1}{\sqrt{2}}\left([S_1, ..., S_{\frac{m}{2}}] + j[S_{\frac{m}{2}+1}, ..., S_m]\right)^T, \tag{4.11}$$

where $j = \sqrt{-1}$. In the transmission we consider M_* of the M_t transmit antennas since only M_* samples are effectively transmitted at each channel use, because $\mathrm{rank}\{\mathbf{H}\} \leq M_*$.

For $bM_* \leq 1$, the channel input $\mathbf{X}^n$ is generated scaling the first nM_* source samples of $X_c^{\frac{m}{2}}$ and mapping them into the channel input as $\mathbf{X}^n = [\mathbf{S}_{c,1}^{M_*}, \mathbf{S}_{c,M_*+1}^{2M_*}, .., \mathbf{S}_{c,(n-1)M_*+1}^{nM_*}]^T$. At reception, the transmitted nM_* source samples are reconstructed with an MMSE estimator using $\mathbf{Y}^n$ and T^{nM_*}, while the remaining $\frac{m}{2} - nM_*$ source samples that have not been transmitted, are estimated using only $T_{nM_*+1}^m$. For $bM_* \geq 1$, the whole source sequence is transmitted in the first $\frac{m}{2M_*}$ channel uses scaling the power by bM_*, and reconstructed at the decoder using an MMSE estimator. The minimum achievable distortion using uncoded transmission at uniform power P at state $(\mathbf{H}, \gamma)$ is given by

$$D_u(P, \gamma, \mathbf{H}) \triangleq \sum_{i=i}^{M_*} \frac{1}{1 + P\mu_i\rho + \gamma\rho_s}, \tag{4.12}$$

where $\mu_1 \geq \cdots \geq \mu_{M_*} \geq 0$ are the ordered eigenvalues of the matrix $\mathbf{H}_{M_*}\mathbf{H}_{M_*}^H$, where $\mathbf{H}_{M_*}$ is the submatrix of $\mathbf{H}$ obtained by taking the M_* columns corresponding to the antennas effectively used for transmission. Then, the expected distortion is found as

$$ED_u = \begin{cases} bM^*\mathrm{E}[D_u(1, \mathbf{H}, \Gamma)] + (1 - bM_*)\mathrm{E}[D_u(0, \mathbf{H}, \Gamma)] & \text{if } bM_* < 1, \\ \mathrm{E}[D_u(bM^*, \mathbf{H}, \Gamma)] & \text{if } bM_* \geq 1. \end{cases}$$

The distortion exponent for uncoded transmission is obtained similarly to $\Delta_d(b, x)$ and is given in the next theorem without proof.

Theorem 9. *The distortion exponent for uncoded transmission, $\Delta_u(b, x)$ is given by*

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

In Section 4.5.5, the performance of uncoded transmission will be compared to the proposed achievable schemes and upper bounds.

4.5.4 HDA Wyner-Ziv Coding (HDA-WZ)

In this section we consider a generalization of HDA-WZ in Chapter 3 to the MIMO channel and to bandwidth ratios satisfying $bM^* \geq 1$. This scheme quantizes the source, uses a scaled version of the quantization error as channel input, and applies joint decoding at the decoder. In the SISO fading setup with $b = 1$, HDA-WZ is shown to achieve the optimal distortion exponent for a wide family of side information distributions. We note that, HDA-S introduced in Section 4.4 for $bM_* < 1$, can be modified to include joint decoding to better exploit the available side information and reduce the expected distortion. However, the distortion exponent will not increase.

At the encoder, consider a quantization codebook of 2^{mR_h} length-m codewords $W^m(s)$, $s = 1, ..., 2^{mR_h}$, with a test channel $W = S + Q$, where $Q \sim \mathcal{N}(0, \sigma_Q^2)$ is independent of S, and quantization noise variance is chosen such that $\frac{R_h}{2} = I(W; S) + \epsilon$, for an arbitrarily small $\epsilon > 0$, i.e., $\sigma_Q^2 \triangleq (2^{R_h - \epsilon} - 1)^{-1}$. Then, each W^m is reordered into length-$\frac{m}{2M_*}$ complex codewords $\mathbf{W}(s) = [\mathbf{W}_1(s), ..., \mathbf{W}_{\frac{m}{2M_*}}(s)] \in \mathbb{C}^{\frac{m}{2M_*} \times M_*}$, where

$\mathbf{W}_i(s)$, $i = 1, ..., \frac{m}{2M_*}$, is given by

$$\mathbf{W}_i(s) = \frac{1}{\sqrt{2}}\left([W_{iM_*+1}(s); ...; W_{(i+1)M_*}(s)] + j[W_{(i+1)M_*+1}(s); ...; W_{2iM_*}(s)]\right)^T,$$

Similarly, we can reorder S^m and Q^m, and define $\mathbf{S}_i$ and $\mathbf{Q}_i$.

We then generate 2^{mR_h} independent auxiliary random vectors $\mathbf{U} \in \mathbb{C}^{\left(n-\frac{m}{2M_*}\right)\times M_*}$ distributed as $\mathbf{U}_i \sim \mathcal{CN}(0, \mathbf{I})$, for $i = 1, ..., n - \frac{m}{2M_*}$ and assign one to each $\mathbf{W}(s)$ to construct the codebook of size 2^{mR_h} consisting of the pairs of codewords $(\mathbf{W}(s), \mathbf{U}(s))$, $s = 1, ..., 2^{mR_h}$. For a given source sequence S^m, the encoder looks for the s^*-th codeword $\mathbf{W}(s^*)$ such that $(\mathbf{W}(s^*), S^m)$ are jointly typical. A unique s^* is found if $M_*R_h > I(\mathbf{W}; \mathbf{S})$. Then, the pair $(\mathbf{W}(s^*), \mathbf{U}(s^*))$ is used to generate the channel input, which is scaled to satisfy the power constraint:

$$\mathbf{X}_i = \begin{cases} \sqrt{\frac{1}{\sigma_Q^2}}[\mathbf{S}_i - \mathbf{W}_i(s^*)], & \text{for } i = 1, ..., \frac{m}{2M_*}, \\ \mathbf{U}_{i-\frac{m}{2M_*}}(s^*), & \text{for } i = \frac{m}{2M^*} + 1, ..., n. \end{cases}$$

Basically, in the first block of $\frac{m}{2M_*}$ channel accesses we transmit a scaled version of the error in the quantization $\mathbf{Q}_i$ in an uncoded fashion, while in the second block of $n - \frac{m}{2M_*}$ accesses we transmit a digital codeword.

The decoder looks for an index s such that $\mathbf{W}(s)$, T^m and the channel output corresponding to the uncoded input, $\mathbf{Y}_W^{\frac{m}{2M_*}} \triangleq [\mathbf{Y}_1, \ldots, \mathbf{Y}_{m/2M_*}]$, are jointly typical, while simultaneously $\mathbf{U}(s)$ is jointly typical with the channel output that corresponds to the coded input block, $\mathbf{Y}_U^{n-\frac{m}{2M_*}} \triangleq [\mathbf{Y}_{m/2M_*+1}, \ldots, \mathbf{Y}_n]$. Let $\mathbf{T}_i = [T_{(i-1)M_*+1}, ..., T_{iM_*}]^H$, for $i = 1, ..., \frac{m}{M_*}$, be blocks of T^m. At the receiver, it follows from Lemma 6 that decoding is successful with high probability if

$$I(\mathbf{W}; \mathbf{S}) < M_*R_h < I(\mathbf{WU}; \mathbf{YT}) \tag{4.13}$$

The outage event is obtained in Appendix D as

$$\mathcal{O}_h = \left\{(\mathbf{H}, \gamma) : I(\mathbf{W}, \mathbf{S}) \geq I(\mathbf{W}; \mathbf{Y}_W\mathbf{T}) + (bM_* - 1)I(\mathbf{U}; \mathbf{Y}_U)\right\}, \tag{4.14}$$

where $I(\mathbf{U}; \mathbf{Y}_U) = \log\det(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}\mathbf{H}^H)$ and,

$$I(\mathbf{W}; \mathbf{Y}_W\mathbf{T}) = \log\left(\frac{(\xi(1+\sigma_Q^2))^{M_*}\det(\mathbf{I} + \frac{\rho}{M_t}\mathbf{H}\mathbf{H}^H))}{\det(\mathbf{I} + \sigma_Q^2(\frac{\rho}{M_t}\mathbf{H}\mathbf{H}^H + \xi\mathbf{I}))}\right),$$

where $\xi \triangleq 1 + \rho_s\gamma$.

If $\mathbf{W}^{\frac{m}{2M_*}}$ is successfully decoded, each S^n is reconstructed with an MMSE estimator using $\mathbf{Y}$ and T^m with a distortion

$$D_h(\sigma_Q^2, \mathbf{H}, \gamma) = \frac{1}{M_*} \sum_{i=1}^{M_*} \left(1 + \rho_s \gamma + \frac{1}{\sigma_Q^2}\left(1 + \frac{\rho}{M_*}\lambda_i\right)\right)^{-1}. \tag{4.15}$$

The derivation of (4.15) is found in Appendix D.

If an outage occurs and $\mathbf{W}$ is not decoded, only T^m is used in the reconstruction, since $\mathbf{X}^n$ is uncorrelated with the source sequence by construction, and so is $\mathbf{Y}^n$. Using an MMSE estimator, the achievable distortion is given by $D_d(0, \gamma)$. Then, the expected distortion for HDA-WZ is found as

$$ED_h(R_h) = \mathrm{E}_{\mathcal{O}_h^c}[D_h(\sigma_Q^2, \mathbf{H}, \Gamma)] + \mathrm{E}_{\mathcal{O}_h}[D_d(0, \Gamma)].$$

The distortion exponent of HDA-WZ is given next.

Theorem 10. *The distortion exponent achieved by HDA-WZ, $\Delta_h(b, x)$, is given by $\Delta_h(b, x) = \max\{x, bM_*\}$ if $bM_* \leq 1$. If $bM_* > 1$, the distortion exponent is given by $\Delta_h(b, x) = x$ if $1 < bM_* < x$, and by*

$$\Delta_h(b, x) = 1 + \frac{(bM_* - 1)(\Phi_k + k\Upsilon_k - 1 + x)}{bM_* - 1 + M_*\Upsilon_k},$$

if

$$b \in \left[\frac{\Phi_{k+1} - 1 + x}{k+1} + \frac{1}{M_*}, \frac{\Phi_k - 1 + x}{k} + \frac{1}{M_*}\right), \quad \text{for } k = 0, ..., M_* - 1.$$

Proof. See Appendix D. □

4.5.5 Comparison of single layer tranmission schemes

Here, we compare the performance of the single layer schemes presented in this section. Figure 4.2 shows the expected distortion achievable by SSCC and JDS schemes in a SISO and a 3×3 MIMO setup for $b = 2$. It is observed that JDS outperforms SSCC in both SISO and MIMO scenarios. We also observe that both SSCC and JDS fall short of the expected distortion lower bound, ED^*_{pi}. Moreover the gap increases with the number of degrees-of-freedom in the system. We note that not only the gap between the achievable distortion exponent increase, but also the gap between the slopes of the curves, which means that the proposed transmission schemes perform especially poorly in the high SNR regime.

To illustrate this, we compare the distortion exponent achieved by SSCC, JDS, uncoded transmission, HDA-S and HDA-WZ in Figure 4.3 in a 2×2 MIMO channel. First, we note that, as discussed in Section 4.4, for $bM_* \leq 1$, the upper bound is achieved

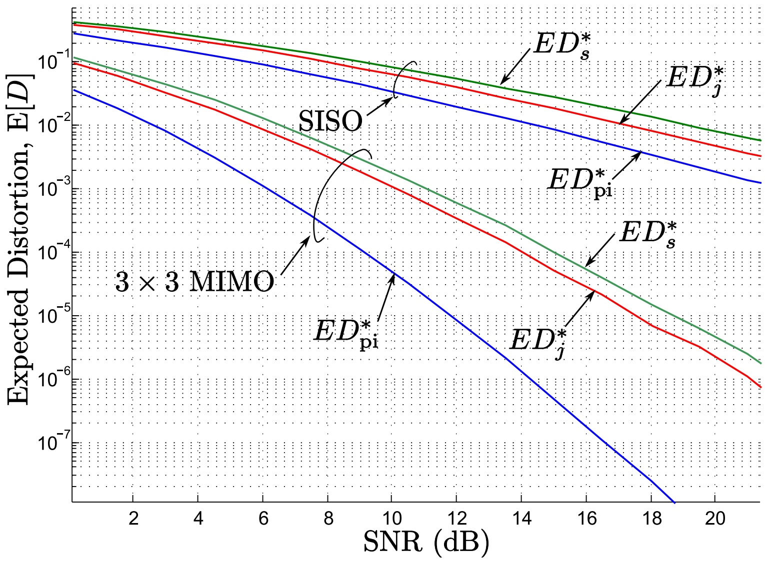


Figure 4.2: Minimum expected distortion achievable by SSCC and JDS for a SISO and a 3×3 MIMO channel for $b = 2$ and $x = 1$. The partially informed encoder bound is also included.

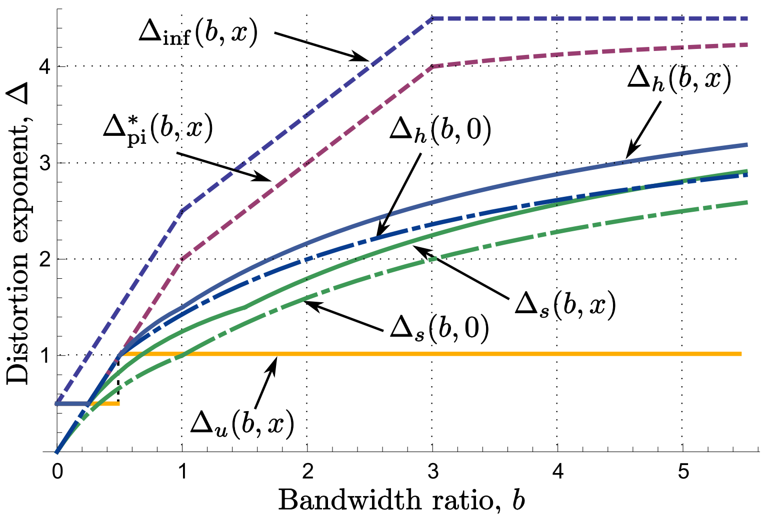


Figure 4.3: Distortion exponents upper bounds and lower bounds for single-layer schemes in function of b for $x = 0.5$ and 2×2 MIMO. The performance of these schemes is also shown in the absence of side information, i.e., $x = 0$.

by S-HDA and by only using the side information. SSCC, JDS and uncoded transmission also achieve the optimal performance for $0 \leq bM_* \leq x$, since these schemes also use the available side information. For larger bandwidth ratios, HDA-WZ improves upon SSCC and JDS, while uncoded transmission achieves the optimal distortion exponent for $bM_* = 1$ and then saturates becoming highly suboptimal for large b values. Note that uncoded transmission outperforms SSCC and JDS for the range $x \leq b \lesssim 0.7$. We also include the distortion exponent achievable when no side information is available, which can be modeled by letting $x = 0$. Significant gains can be obtained by exploiting the side information. However, this is not the case for uncoded transmission, for which $\Delta_u(b, x) = \Delta_u(b, 0) = 1$ for $M_* b \geq 1$. In general we observe that single layer schemes are not capable of fully exploiting the available degrees-of-freedom in the system, especially in the large bandwidth regime. Single layer schemes depend on a single parameter, the rate, and cannot adapt to the system, specially in those regimes in which the system has many degrees-of-freedom available. This motivates us to consider other achievability techniques, based on multi-layer transmission. In multi-layer transmission, each layer provides additional degrees-of-freedom to the system, and therefore can adapt better to the time variations and achieve higher distortion exponent values.

We also observe that the difference between the fully informed encoder upper bound and the partially informed encoder upper bound.

4.6 Multi-layer transmission

In the previous section, we have observed that the distortion exponent achievable with single layer schemes is far from the upper bound, especially in the high bandwidth regime. Here, we consider multi-layer schemes to improve the achievable distortion exponent in this regime. Multi-layer transmission is proposed in [15] to combat channel fading by transmitting multiple layers that carry successive refinements of the source [38]. At the receiver, as many layers as possible are decoded depending on the channel state. The better the channel state, the more layers can be decoded and the smaller is the distortion at the receiver. We propose the extension of the JDS schemes to *progressive multi-layer JDS transmission* and s*uperposed multi-layer JDS transmission*, and derive the corresponding distortion exponents.

4.6.1 Progressive multi-layer JDS transmission (LS-JDS)

In this section we consider the progressive transmission of JDS layers over the channel. The refinement codewords are transmitted one after the other over the channel using JDS transmission. Similarly to [15], we assume that each layer is allocated the same time resources (or number of channel accesses). In the limit of infinite layers, this assumption does not incur a loss in performance.

At the encoder, we generate L Gaussian quantization codebooks, at rates $bR_l/2L = I(S;W_l|W_1^{l-1})+\epsilon/2$, $l=1,...,L$, and an arbitrarily small $\epsilon>0$, such that each Gaussian codebook is a refinement for the previous layers [38]. The quantization codewords are generated as $W_l = S + \sum_{i=l}^{L} Q_i$, for $l=1,...,L$, where $Q_l \sim \mathcal{N}(0,\sigma_l^2)$ are independent of each other. As shown in Appendix D, for a given rate tuple $\mathbf{R} \triangleq [R_1,...,R_L]$ the quantization noises satisfy

$$\sum_{i=l}^{L} \sigma_i^2 = (2^{\sum_{i=1}^{l}(\frac{b}{L}R_i-\epsilon)}-1)^{-1}, \qquad l=1,...,L. \tag{4.16}$$

We generate L independent channel codebooks with $\frac{n}{L}$-lenght codewords $\mathbf{X}_l^n \in \mathbb{C}^{M_t\times n/L}$ with $\mathbf{X}_{l,i} \sim \mathcal{CN}(0,\mathbf{I})$. Each successive refinement codeword is transmitted using JDS as in Section 4.5.2. At the destination, the decoder successively decodes each refinement codeword using joint decoding from the first layer up to the L-th layer. Then, l layers will be successfully decoded if

$$I(S,W_l|T,W_1^{l-1}) < \frac{b}{2L}I(\mathbf{X};\mathbf{Y}) \le I(S,W_{l+1}|T,W_l^1),$$

that is, l layers are successfully decoded while there is an outage in decoding $l+1$ layers. Let us define the outage event, for $l=0,...,L$, as follows

$$\mathcal{O}_l^{ls} \triangleq \left\{(\mathbf{H},\gamma)\!:\! I(S,W_l|T,W_1^{l-1}) \ge \frac{b}{2L}I(\mathbf{X};\mathbf{Y})\right\}, \tag{4.17}$$

where $I(\mathbf{X},\mathbf{Y}) = \log\det\left(\mathbf{I}+\frac{\rho}{M_*}\mathbf{H}\mathbf{H}^H\right)$, and, for $R_0 \triangleq 0$,

$$I(S;W_l|W_1^{l-1},T) = \frac{1}{2}\log\left(\frac{2^{\sum_{i=1}^{l}\frac{b}{L}R_i}+\gamma\rho_s}{2^{\sum_{i=1}^{l-1}\frac{b}{L}R_i}+\gamma\rho_s}\right).$$

The details of the derivation are given in Appendix D. Due to the successive refinability of the Gaussian source, provided l layers have been successfully decoded, the receiver reconstructs the source with a MMSE estimator using the side information and the decoded layers with a distortion given by $D_d(\sum_{i=1}^{l} bR_l/2L,\gamma)$. The expected distortion can be expressed as follows.

$$ED_{ls}(\mathbf{R}) = \sum_{l=0}^{L} \mathrm{E}_{(\mathcal{O}_l^{ls})^c \bigcap \mathcal{O}_{l+1}^{ls}}\left[D_d\left(\sum_{i=1}^{l}\frac{bR_l}{2L},\gamma\right)\right]. \tag{4.18}$$

The distortion exponent achieved by LS-JDS is given next.

Theorem 11. *Let us define*

$$\phi_k \triangleq M^* - M_* + 2k - 1, \quad M_k \triangleq M_* - k + 1, \tag{4.19}$$

and the sequence $\{c_i\}$ *defined as*

$$c_0 = 0, \quad c_i = c_{i-1} + \phi_i \ln\left(\frac{M_* - i + 1}{M_* - i}\right),$$

for $i = 1, ..., M_* - 1$ *and* $c_{M_*} = \infty$.

The distortion exponent achieved by LS-JDS with infinite number of layers is given by $\Delta^*_{ls}(b, x) = x$ *for* $b \leq x/M_*$, *and for*

$$c_{k-1} + \frac{x}{M_* - k + 1} < b \leq c_k + \frac{x}{M_* - k},$$

$k = 1, ..., M_*$, *the achievable distortion exponent is given by*

$$\begin{aligned}\Delta^*_{ls}(b, x) = x + &\sum_{i=1}^{k-1}(M^* - M_* + 2i - 1) \\ &+ (M_* - k + 1)(M^* - M_* + 2k - 1) \times \left(1 - e^{-\frac{b(1-\kappa^*)-c_{k-1}}{M^*-M_*+2k-1}}\right),\end{aligned}$$

where

$$\kappa^* = \frac{\phi_k}{b}\mathcal{W}\left(\frac{e^{\frac{b-c_{k-1}}{\phi_k}}x}{M_k\phi_k}\right),$$

and $\mathcal{W}(z)$ *is the function W of Lambert, which gives the principal solution for* w *in* $z = we^w$.

Proof. See Appendix D □

The proof of Theorem 11 indicates that the distortion exponent for LS-JDS is achieved by allocating an equal rate among the first $\kappa^* L$ layers to guarantee that the distortion exponent is at least x. Then, the rest of layers, $(1 - \kappa^*)L$, are used to further increase the distortion exponent with the corresponding rate allocation. Note that for $x = 0$, we have $\kappa^* = 0$.

4.6.2 Superposed multi-layer JDS scheme (BS-JDS)

In this section, we consider that the successive refinements of the source are transmitted by a superposition of JDS layers. The receiver decodes as many layers a possible using successive joint decoding and reconstructs the source.

At the encoder, generate L Gaussian quantization codebooks, at rates given by $\frac{b}{2}R_l = I(S;W_l|W_1^{l-1}) + \epsilon/2$, $l = 1,...,L$, and an arbitrarily small $\epsilon > 0$, as in Section 4.6.1, and L channel codebooks $\mathbf{U}_l^n$, $l = 1,\dots,L$, i.i.d. with $\mathbf{U}_{l,i} \sim \mathcal{CN}(0,\mathbf{I})$. Let $\boldsymbol{\rho} = [\rho_1,...,\rho_L,\rho_{L+1}]^T$ be the power allocation among channel codebooks such that $\rho = \sum_{i=1}^{L+1}\rho_i$. We consider a power allocation $\rho_l = \rho^{\xi_{l-1}} - \rho^{\xi_l}$ with $1 = \xi_0 \geq \xi_1 \geq \dots \geq \xi_L \geq 0$ and define $\boldsymbol{\xi} \triangleq [\xi_1,...,\xi_L]$. In the last layer, the layer $L+1$, Gaussian i.i.d. noise $\tilde{\mathbf{N}}_i \sim \mathcal{CN}(0,\mathbf{I})$ is transmitted using the remaining power $\rho_{L+1} \triangleq \rho^{\xi_L}$ for mathematical convenience. Then, the channel input $\mathbf{U}^n$ is generated as the superposition of $\mathbf{U}_l^n$ with the corresponding power allocation $\sqrt{\rho_l}$ as

$$\mathbf{U}^n = \frac{1}{\sqrt{\rho}}\sum_{j=1}^{L}\sqrt{\rho_j}\mathbf{U}_j^n + \sqrt{\rho^{\xi_L}}\tilde{\mathbf{N}}^n.$$

At the receiver, the decoder uses successive joint decoding from layer 1 up to layer L considering the posterior layers as noise. Layer $L+1$, containing the noise, is ignored. The outage event at layer l, provided $l-1$ layers have been decoded, is given by

$$\mathcal{O}_l^{ml} = \left\{(\mathbf{H},\gamma) : \frac{b}{2}I(\mathbf{X}_l;\mathbf{Y}|\mathbf{X}_1^{l-1}) \leq I(S;W_l|T,W_1^{l-1})\right\}.$$

If l layers are decoded, the source is reconstructed at a distortion $D_d(\sum_{i=1}^{l} bR_i,\gamma)$ with a MMSE estimator, and the expected distortion is found as

$$ED_{ml}(\mathbf{R},\boldsymbol{\xi}) = \sum_{l=1}^{L}\mathrm{E}_{\mathcal{O}_{k+1}^{ml}}\left[D_d\left(\sum_{i=0}^{l}\frac{b}{2}R_i,\Gamma\right)\right],$$

where $\mathbf{R} \triangleq [R_1,...,R_L]$ and $\mathcal{O}_{L+1}^{ml}$ is the set of states in which the all L layers with information are decoded.

The distortion exponent for the transmission of L coded layers using BS-JDS is given in the next theorem.

Theorem 12. *Let us define*

$$\eta_k \triangleq \frac{b(k+1)-\Phi_{k+1}}{\Upsilon_k} \quad \text{and} \quad \Gamma_k \triangleq \frac{1-\eta_k^{L-1}}{1-\eta_k}. \tag{4.20}$$

The distortion exponent $\Delta_{ml}^L(b,x)$ achieved by BS-JDS with L layers with a power allocation $\rho_l = \rho^{\xi_{l-1}} - \rho^{\xi_l}$, $1 = \xi_0 \geq \xi_1 \geq \dots \geq \xi_L \geq 0$ and diversity multiplexing gain $\hat{r}_l = [(k+1)(\xi_{l-1}-\xi_l)-\epsilon_1]$, $\epsilon_1 \to 0$ is given by $\Delta_{ml}^L(b,x) = x$ for $bM_* \leq x$ and by

$$\Delta_{ml}^L(b,x) = x + \Phi_k - \frac{\Upsilon_k(\Upsilon_k(x+\Phi_k) + xb(k+1)\Gamma_k)}{(\Upsilon_k + b(1+k))(\Upsilon_k + b(1+k)\Gamma_k) - b(k+1)\Phi_k\Gamma_k}, \tag{4.21}$$

for

$$b \in \left[\frac{\Phi_{k+1}+x}{k+1}, \frac{\Phi_k + x}{k}\right), \quad k = 0, ..., M_* - 1.$$

The power allocation $\boldsymbol{\xi}$ *used given by*

$$\begin{aligned} \xi_1 &= \frac{(\Upsilon_k + \Phi_k\Gamma_k)(\Upsilon_k + b(k+1) - \Phi_k - x)}{(\Upsilon_k + b(1+k))(\Upsilon_k + b(1+k)\Gamma_k) - b(k+1)\Phi_k\Gamma_k}, \\ \xi_1 - \xi_2 &= \frac{\Phi_k(\Upsilon_k + b(k+1) - \Phi_k - x)}{(\Upsilon_k + b(1+k))(\Upsilon_k + b(1+k)\Gamma_k) - b(k+1)\Phi_k\Gamma_k}. \end{aligned} \tag{4.22}$$

and for $l = 2, ..., L$*,*

$$\xi_l = \xi_1 - (\xi_1 - \xi_2)\frac{1-\eta_k^{l-1}}{1-\eta_k}. \tag{4.23}$$

Proof. See Appendix D. □

An upper bound on the performance of BS-JDS is obtained for a continuum of infinite layers, i.e., $L \to \infty$.

Corollary 2. *The distortion exponent of BS-JDS in the limit of infinite layers,* $\Delta_{ml}^{\infty}(b,x)$*, is found, for* $k{=}0, ..., M_*{-}1$*, by*

$$\Delta_{ml}^{\infty}(b,x) = \max\{x, b(k+1)\} \qquad \textit{for } b{\in}\left[\frac{\Phi_{k+1}+x}{k}, \frac{\Phi_k}{k+1}\right),$$

and

$$\Delta_{ml}^{\infty}(b,x) = \Phi_k + x\left(\frac{b(1+k)-\Phi_k}{b(1+k)-\Phi_{k+1}}\right) \qquad \textit{for } b \in \left[\frac{\Phi_k}{k+1}, \frac{\Phi_k+x}{k}\right).$$

Proof. See Appendix D. □

The solution in Theorem 12 is obtained by fixing the diversity multiplexing gains of the code in each layer as $\hat{r}_l = b[(k+1)(\xi_{l-1} - \xi_l) - \epsilon_1]$. As discussed in Appendix D, this choice excludes single layer JDS from the set of feasible solutions. By choosing $r_2 = ... = r_L = 0$, BS-JDS scheme reduces to single layer JDS. Interestingly, for b in the regions

$$b \in \left[\frac{\Phi_k}{k}, \frac{\Phi_k + x}{k}\right), \qquad k = 1, ..., M_* - 1,$$

single layer JDS achieves a larger distortion exponent than $\Delta_{ml}^{\infty}(b,x)$ in Corollary 2, as shown in Figure 4.4. Note that this region is empty for $x = 0$, and thus, this phenomena does not appear in the absence of side information. Then, the achievable distortion exponent for BS-JDS can be given as follows.

Lemma 17. *BS-JDS achieves the distortion exponent*

$$\Delta^*_{ml}(b,x) \geq \max\{\Delta^{\infty}_{ml}(b,x), \Delta_j(b,x)\}.$$

The problem of optimizing the distortion exponent for BS-JDS can be formulated as a linear optimization program, as shown in (D.43) in Appendix D and efficiently solved numerically. In Figure 4.4 we show one instance of the numerical optimization for 2×2 MIMO, $x = 0.5$ and $L = 500$ layers. We also include the distortion exponent achievable by single layer JDS, i.e., when $L = 1$, and the exponent achievable by considering $\hat{\mathbf{r}}$ for $L = 2$ and in the limit of infinite layers, given by $\Delta^2_{ml}(b,x)$ and $\Delta^{\infty}_{ml}(b,x)$, respectively. We observe that the achievable distortion exponent for fixed multiplexing gains $\hat{\mathbf{r}}$ is not continuous, even in the limit of infinite layers. However, in general the distortion exponent is continuous when jointly optimized over the multiplexing gains and the power allocations. We also observe that there is a significant improvement in the distortion exponent just by using two layers. Also, we note that there is a tight match in the numerical and the achievable distortion given in Lemma 17. Many more numerical solutions suggest that, in fact, the optimal distortion exponent achievable by BS-JDS is given by the best of $\Delta^{\infty}_{ml}(b,x)$ and $\Delta_j(b,x)$.

Conjecture 1. *The optimal distortion exponent achievable by BS-JDS is given by*

$$\Delta^*_{ml}(b,x) = \max\{\Delta^{\infty}_{ml}(b,x), \Delta_j(b,x)\}.$$

In next section, we will see that fixing the diversity multiplexing gain to $\hat{\boldsymbol{r}}$ suffices for BS-JDS to meet the partially informed upper bound in the MISO/SIMO setup, and thus, this conjecture is resolved for the positive in these case.

4.7 Comparisons and Discussion

In this section, we discuss the performance of the proposed schemes with respect to the derived upper bounds and characterize the optimal distortion exponent for MISO/SIMO/SISO. In MISO/SIMO, i.e., $M_* = 1$, we show that BS-JDS achieves the partially informed encoder upper bound, thus characterizing the optimal distortion exponent. For SISO, i.e., $M^* = M_* = 1$, HDA-WZ also meets the optimal distortion exponent. For the general MIMO setup, the low bandwidth regime has been characterized in Section 4.4. However, the proposed schemes do not meet the upper bound for $bM_* > 1$. Nevertheless, multi-layer transmission schemes perform close to the upper bound, especially in the high bandwidth regime.

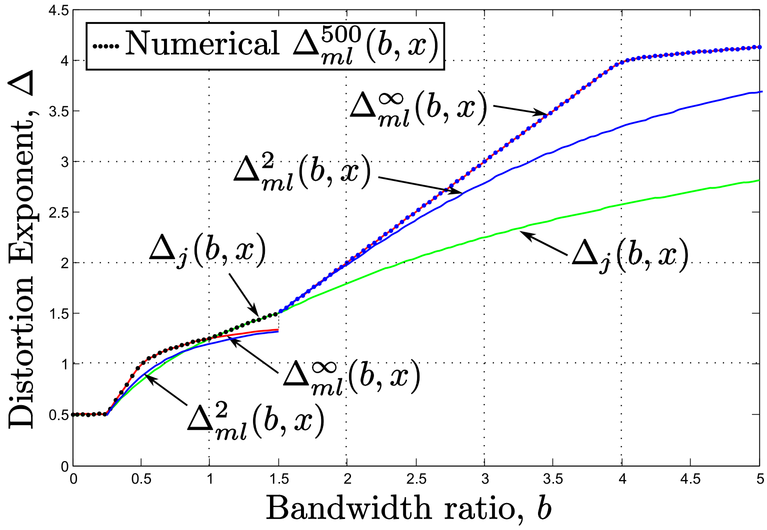


Figure 4.4: Distortion exponent achieved by BS-JDS with $L = 1, 2$ and in the limit of infinite layers with respect to the bandwidth ratio b for a 2×2 MIMO system and a side information quality given by $x = 0.5$. Numerical results on the achievable distortion exponent for $L = 500$ are also included.

4.7.1 Optimal Distortion Exponent for MISO/SIMO/SISO

We first particularize the upper bounds on the distortion exponent for $M_* = 1$. The informed encoder upper bound is found as

$$\Delta_{\text{inf}}(x, b) = x + \min\{b, M^*\},$$

and the partially informed encoder upper bound is given by

$$\Delta_{up}^*(b, x) = \begin{cases} \max\{x, b\} & \text{for } b \leq \max\{M^*, x\}, \\ M^* + x\left(1 - \frac{M^*}{b}\right) & \text{for } b > \max\{M^*, x\}. \end{cases}$$

Notice that as the bandwidth ratio increases, the partially informed encoder upper bound $\Delta_{up}^*(b, x)$ converges to the fully informed encoder upper bound $\Delta_{\text{inf}}(x, b)$.

Now we particularize the proposed lower bounds to $M_* = 1$. The distortion exponent for SSCC and JDS is given by

$$\Delta_j(b, x) = \max\left\{x, b\frac{x + M^*}{b + M^*}\right\},$$

while for uncoded transmission we have

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

Note that for $b = 1$, uncoded transmission meets $\Delta^*_{up}(b,x) = \max\{1,x\}$, while SSCC and JDS are both suboptimal. Similalry happens in the general MIMO channels.

The following distortion exponent is achievable by HDA-S, for $b \leq 1$ and HDA-WZ, for $b > 1$, in the MISO/SIMO setup.

$$\Delta_h(b,x) = \begin{cases} \max\{x,b\} & \text{for } b \leq 1, \\ \max\{x, \frac{M^*+(b-1)(M^*+x)}{M^*+b-1}\} & \text{for } b > 1. \end{cases}$$

As seen in Section 4.4, HDA-S meets the partially informed upper bound for $b \leq 1$. HDA-WZ is in general suboptimal.

For the multi-layer transmission schemes, the distortion exponent acheivable by LS-JDS is given by

$$\Delta^*_{ls}(b,x) = x+M^*\left(1-e^{-\frac{b(1-\kappa^*)}{M^*}}\right), \kappa^* = \frac{M^*}{b}\mathcal{W}\left(\frac{e^{\frac{b}{M^*}x}}{M^*}\right).$$

As for BS-JDS, considering the achievable rate in Corollary 2, this scheme meets the partially informed encoder lower bound in the limit of infinite layers, i.e., $\Delta^\infty_{ml}(b,x) = \Delta^*_{up}(b,x)$. This fully characterizes the optimal distortion exponent in the MISO/SIMO setup, as stated in the next theorem.

Theorem 13. *The optimal distortion exponent $\Delta^*(b,x)$ for MISO/SIMO systems is given by*

$$\Delta^*(b,x) = \begin{cases} \max\{x,b\} & \textit{for } b \leq \max\{M^*,x\}, \\ M^* + x\left(1-\frac{M^*}{b}\right) & \textit{for } b > \max\{M^*,x\}, \end{cases}$$

and is achieved by BS-JDS in the limit of infinite layers.

In Figure 4.5 we show the distortion exponent for a MISO/SIMO channel with $M^* = 4$ and $x = 0.5$, with respect to the bandwidth ratio b. We observe that, as given in Theorem 13, BS-JDS achieves the optimal distortion exponent. As discussed in Section 4.5.5, single layer schemes performs poorly as the bandwidth ration increases. We observe that HDA-WZ outperforms JDS in all regimes and that, although it outperforms the multi-layer LS-JDS for low b values, LS-JDS achieves larger distortion exponents than HDA-WZ for $b \geq 3$.

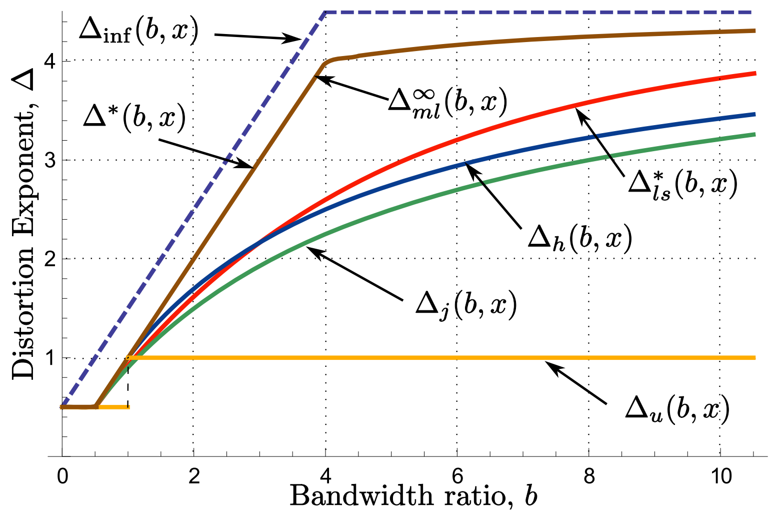


Figure 4.5: Distortion exponent Δ with respect to the bandwidth ratio b for a 4×1 MISO system and a side information quality given by $x = 0.5$.

In Figure 4.6, we show upper and lower bounds on the distortion exponent for the SISO case and $x = 0.4$. We observe that the performance of the schemes is similar to the MISO/SIMO case. However, LS-JDS achieves worse distortion exponents than HDA-WZ, which achieves the optimal distortion exponent for $b > 1$.

Lemma 18. *The optimal distortion exponent for SISO channels is achieved by BS-JDS, HDA-WZ and HDA-S.*

4.7.2 General MIMO

Here, we consider the general MIMO channel. Figure 4.7 shows the upper and lower bounds on the distortion exponents derived in the previous sections for a 2×2 MIMO channel with $x = 0.5$. First, it can be observed that, the optimal distortion exponent is achieved by HDA-S and BS-JDS for $b \leq 0.5$, as expected from Section 4.4. In addition, we note that BS-JDS with infinite layers also achieves the optimal distortion exponent in this regime, while the other schemes are suboptimal in general. In general, uncoded transmission achieves the optimal distortion exponent at $bM_* = 1$.

Lemma 19. *Uncoded transmission achieves the optimal distortion exponent for* $bM_*{=}1$.

For $0.5 < b \lesssim 2.4$, HDA-WZ is the scheme achieving the largest distortion exponent, and outperforms BS-JDS, and in particular, in all the regimes where the performance of

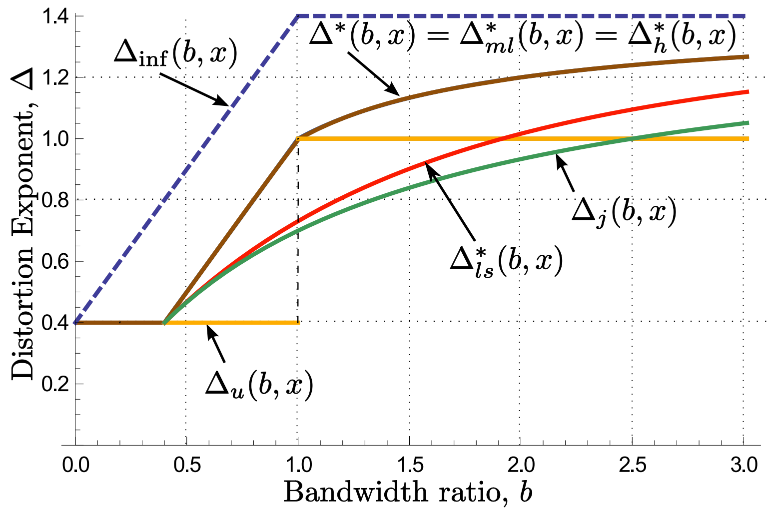


Figure 4.6: Distortion exponents in SISO channels in function of b for $x = 0.4$.

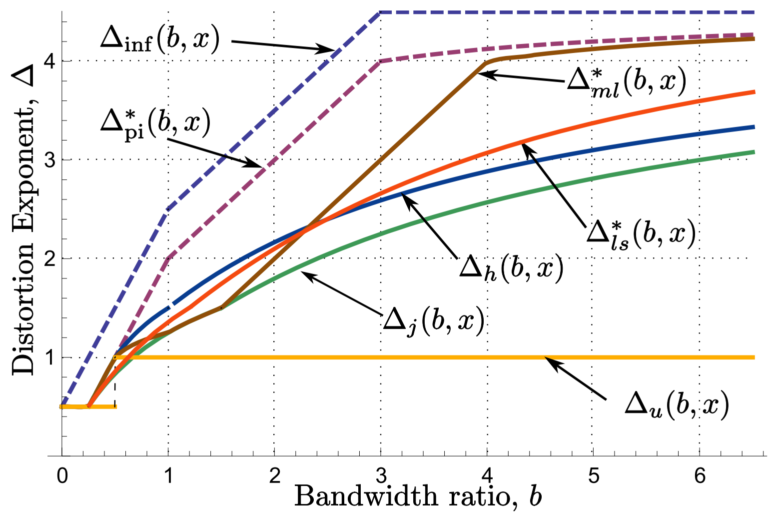


Figure 4.7: Distortion exponent Δ with respect to the bandwidth ratio b for a 2×2 MIMO system and a side information quality given by $x = 0.5$.

BS-JDS reduces to the performance of JDS, since HDA-WZ outperforms JDS in general. For larger b values, the largest distortion exponent is achieved by BS-JDS. Note that for $b \geq 4$, $\Delta^*_{ml}(b, 0.5)$ is very close to the partially informed encoder lower bound. We also observe that for $b \gtrsim 2.4$ LS-JDS outperforms HDA-WZ, but it is worse than BS-JDS. This is not the case in other regimes, as will be seen next.

In Figure 4.8, we show the upper and lower bounds proposed for a 4×4 MIMO channel with $x = 0.5$. We note that in this case, for $bM_* \leq \max\{1, x\}$, $\Delta^*(b, 3) = 3$, which is achievable by all schemes only using the side information sequence at the decoder. For this setup, LS-JDS achieves the best distortion exponent for intermediate

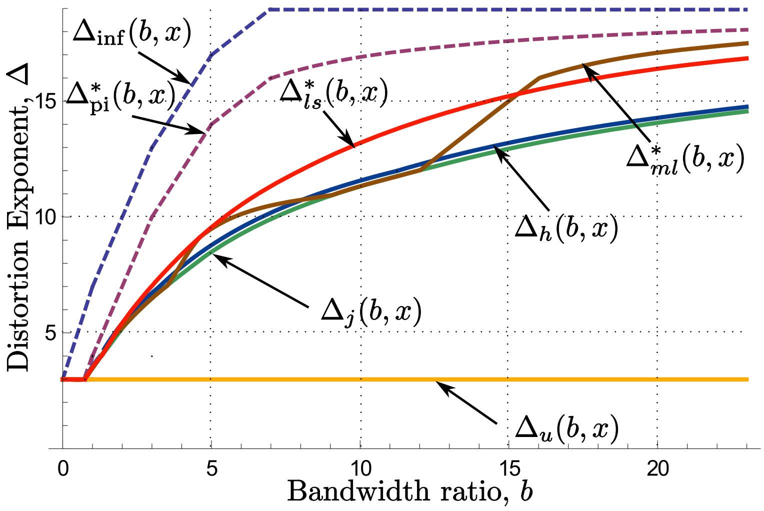


Figure 4.8: Distortion exponent Δ with respect to the bandwidth ratio b for a 4×4 MIMO system and a side information quality given by $x = 3$.

b values, outperforming both HDA-WZ and BS-JDS. Again, in the large bandwidth regime, BS-JDS achieves the best distortion exponent, and performs close to the upper bound. We note that as the number of antennas increases, the difference in performance between JDS and HDA-WZ decreases.

4.8 Conclusions

We have studied the distortion exponent when transmitting a Gaussian source over a time-varying fading MIMO channel in the presence of time-varying correlated side information at the decoder. We have assumed a block-fading model for both the channel and the side information states, and perfect state information of the channel and the side information at the receiver, while the transmitter has only a statistical knowledge. We have derived two upper bounds on the distortion exponent, as well as lower bounds based on separate source and channel coding, joint decoding, uncoded transmission and hybrid digital-analog transmission. We have proposed multi-layer transmission schemes based on progressive transmission of joint decoding codes or the superposition of them. We have considered the effects of the bandwidth ratio and the side information quality on the distortion exponent, and shown that the multi-layer transmission scheme based on superposition meets the upper bound in MISO/SIMO/SISO channels, solving the JSCC problem in the high SNR regime. For the general MIMO channel, we have characterized the optimal distortion in the low bandwidth regime and shown the multi-layer scheme based on superposition performs very close to the upper bound.

Chapter 5

A Class of Orthogonal Relay Channels with State

In this chapter, we consider a state-dependent orthogonal relay channel, in which the channels connecting the source to the relay and the destination are orthogonal, and are governed by a state sequence, which is assumed to be known only at the destination. We call this model the *state-dependent orthogonal relay channel with state information available at the destination*, and refer to it as the ORC-D model. See Fig. 5.1 for an illustration of the ORC-D channel model. While the setups considered in previous chapters are joint source-channel coding problems, this is a channel coding problem in which the use of source coding tools will be required to achieve the optimal performance.

As discussed in Section 1.2, many practical communication scenarios can be modelled by the ORC-D model. For example, consider a cognitive network with a relay, in which the transmit signal of the secondary user interferes simultaneously with the received primary user signals at both the relay and the destination. After decoding the secondary user message, the destination obtains information about the interference affecting the source-relay channel, which can be exploited to decode the primary transmitter's message, which may not be decoded at the relay. Similarly, consider a mobile network with a relay (e.g., a femtostation), in which the base station (BS) operates in the full-duplex mode, and transmits on the downlink channel to a user, in parallel to the uplink transmission of a femtocell user, causing interference for the first user's transmission at the femtostation. While the relay has no prior information about this interfering signal, the BS already knows it (if decoding of the secondary user's message is successful), which can be used to decode the primary user's message.

The best known transmission strategies for the three terminal relay channel are the decode-and-forward (DF), compress-and-forward (CF) and partial decode-compress-and-forward (pDCF) schemes, which were all introduced by Cover and El Gamal in [74]. In DF the relay decodes the source message and forwards it to the destination together with the source terminal. DF is generalized by the partial decode-and-forward (pDF) scheme in which the relay decodes and forwards only a part of the message. In the ORC-D model, pDF would be optimal when the channel state information is not available at the destination [75]; however, when the state information is known at the destination, fully decoding and re-encoding the message transmitted on the source-relay link renders the channel state information at the destination useless. Hence, we expect that pDF is suboptimal for ORC-D in general.

In CF, the relay does not decode any part of the message, and simply compresses the received signal and forwards the compressed bits to the destination using Wyner-Ziv coding followed by separate channel coding. Using CF in the ORC-D model allows the destination exploit its knowledge of the state sequence; and hence, it can decode messages that may not be decodable by the relay. However, CF also forwards some noise to the destination, and therefore, may be suboptimal in certain scenarios. For example, as the dependence of the source-relay channel on the state sequence weakens, i.e., when the state information becomes less informative, CF performance is expected to degrade.

pDCF combines both schemes: part of the source message is decoded by the relay, and forwarded, while the remaining signal is compressed and forwarded to the destination. Hence, pDCF can optimally adapt its transmission to the dependence of the orthogonal channels on the state sequence. Indeed, we show that pDFC achieves the capacity in the ORC-D channel model, while pure DF and CF are in general suboptimal. The main results of the chapter are summarized as follows:

- We derive an upper bound on the capacity of the ORC-D model, and show that it is achievable by the pDCF scheme. This characterizes the capacity of this class of relay channels.
- Focusing on the two-hop binary and Gaussian models, we show that applying either only the CF or only the DF scheme is in general suboptimal.
- We show that the capacity of the ORC-D model is in general below the cut-set bound. We identify the conditions under which pure DF or pure CF meets the cut-set bound. Under these conditions the cut-set bounds is tight, and either DF or CF scheme is sufficient to achieve the capacity.

While the capacity of the general relay channel is still an open problem, there have been significant achievements within the last decade in understanding the capabilities of various transmission schemes, and the capacity of some classes of relay channels has been characterized. For example, DF is shown to be optimal for physically degraded relay channels and inversely degraded relay channels in [74]. In [75], the capacity of the orthogonal relay channel is characterized, and shown to be achieved by the pDF scheme. It is shown in [76] that pDF achieves the capacity of semi-deterministic relay channels as well. CF is shown to achieve the capacity in deterministic primitive relay channels in [77]. While all of these capacity results are obtained by using the cut-set bound for the converse proof [24], the capacity of a class of modulo-sum relay channels is characterized in [78], and it is shown that the capacity, achievable by the CF scheme, can be below the cut-set bound. The pDCF scheme is shown to achieve the capacity of a class of diamond relay channels in [79].

The state-dependent relay channel has recently attracted considerable attention in the literature. Key to the investigation of the state-dependent relay channel model is whether the state sequence controlling the channel is known at the nodes of the network, the source, relay or the destination in a causal or non-causal manner. The relay channel in which the state information is non-causally available only at the source is considered in [80, 81], and both causally and non-causally available state information is considered in [82]. The model in which the state is non-causally known only at the relay is studied in [83] while causal and non-causal knowledge is considered in [84]. Similarly, the relay channel with state causally known at source and relay is considered in [85] and state non-causally known at source, relay and destination in [86]. The compound relay channel with informed relay and destination are discussed in [87] and [88]. The state-dependent relay channel with structured state has been considered in [89] and [90]. To the best of our knowledge, this is the first work that focuses on the state-dependent relay channel in which the state information is available only at the destination.

The rest of the chapter is organized as follows. In Section II we provide the system model and our main result. Section III is devoted to the proof of the achievability and converse of the main result. In section IV, we provide two examples showing the suboptimality of pDF and CF schemes, while in Section V we show that the capacity is in general below the cut-set bound, and we provide conditions under which pure DF and CF schemes meet the cut-set bound. Finally, Section VII concludes the chapter.

5.1 System Model and Main Result

We consider the class of orthogonal relay channels depicted in Fig. 5.1. The source and the relay are connected through a memoryless channel characterized by $p(y_R|x_1, z)$, while the source and the destination are connected through an orthogonal memoryless channel characterized by $p(y_2|x_2, z)$. Both memoryless channels depend on an independent and identically distributed (i.i.d.) state sequence $\{Z\}_{i=1}^n$, which is available at the destination. The relay and the destination are connected by a memoryless channel $p(y_1|x_R)$, which is independent of the state sequence z^n. The input and output alphabets are denoted by $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{X}_R$, $\mathcal{Y}_1$, $\mathcal{Y}_2$ and $\mathcal{Y}_R$, and the state alphabet by $\mathcal{Z}$.

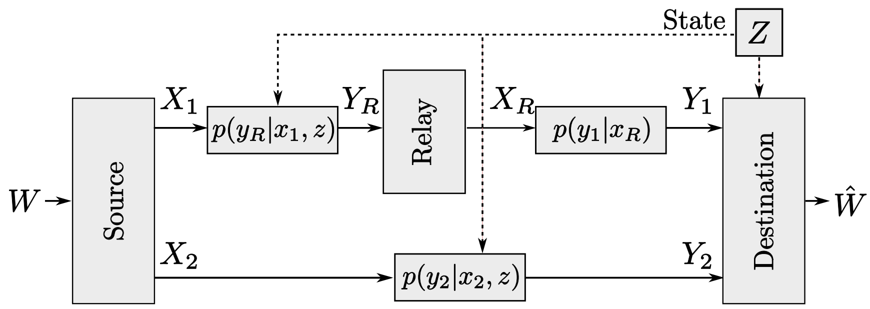


Figure 5.1: Orthogonal state-dependent relay channel with channel state information available at the destination, denoted by ORC-D model.

Let W be the message to be transmitted to the destination with the assistance of the relay. The message W is assumed to be uniformly distributed over the set $\mathcal{W} = \{1, ..., N\}$. An (M, n, ν_n) code for this channel consists of an encoding function at the source:

$$f : \{1, ..., M\} \to \mathcal{X}_1^n \times \mathcal{X}_2^n, \tag{5.1}$$

a set of encoding functions $\{f_{r,i}\}_{i=1}^n$ at the relay, whose output at time i depends on the symbols it has received up to time $i-1$:

$$X_{Ri} = f_{r,i}(Y_{R1}, ..., Y_{R(i-1)}), \quad i = 1, ..., n, \tag{5.2}$$

and a decoding function at the destination

$$g : \mathcal{Y}_1^n \times \mathcal{Y}_2^n \times \mathcal{Z}^n \to \{1, ..., M\}. \tag{5.3}$$

The probability of error, ν_n, is defined as

$$\nu_n \triangleq \frac{1}{M} \sum_{w=1}^{M} \Pr\{g(Y_1^n, Y_2^n, Z^n) \neq w | W = w\}. \tag{5.4}$$

The joint probability mass function (pmf) of the involved random variables over the set $\mathcal{W} \times \mathcal{Z}^n \times \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{X}_R^n \times \mathcal{Y}_R^n \times \mathcal{Y}_1^n \times \mathcal{Y}_2^n$ is given by

$$p(w, z^n, x_1^n, x_2^n, x_R^n, y_R^n, y_1^n, y_2^n) = p(w) \prod_{i=1}^{n} p(z_i) p(x_{1i}, x_{2i}|w) \cdot \\ p(y_{Ri}|z_i, x_{1i}) p(x_{Ri}|y_R^{i-1}) p(y_{1i}|x_{Ri}) p(y_{2i}|x_{2i}, z_i).$$

A rate R is said to be achievable if there exists a sequence of $(2^{nR}, n, \nu_n)$ codes such that $\lim_{n\to\infty}\nu_n = 0$. The capacity $\mathcal{C}$ of this class of state-dependent orthogonal relay channels, denoted as ORC-D, is defined as the supremum of the set of achievable rates.

We define R_0 as the capacity of the link connecting the relay to the destination, and R_1 as the capacity of the direct link connecting the source to the destination when the channel state sequence is available at the destination:

$$R_0 \triangleq \max_{p(x_R)} I(X_R;Y_1), \quad R_1 \triangleq \max_{p(x_2)} I(X_2;Y_2|Z), \tag{5.5}$$

and let $p^*(x_R)$ and $p^*(x_2)$ be the channel input distributions achieving R_0 and R_1, respectively.

Let us define $\mathcal{P}$ as the set of all joint pmf´s given by

$$\mathcal{P} \triangleq \{p(u,x_1,z,y_R,\hat{y}_R) : p(u,x_1,z,y_R,\hat{y}_R) = p(u,x_1)p(z)p(y_R|x_1,z)p(\hat{y}_R|y_R,u)\}, \tag{5.6}$$

where U and $\hat{Y}_R$ are auxiliary random variables defines over the alphabets $\mathcal{U}$ and $\hat{\mathcal{Y}}_R$, respectively.

The main result of this chapter, provided in the next theorem, is the capacity of the class of relay channels described above.

Theorem 14. *The capacity of the ORC-D relay channel is given by*

$$\begin{aligned}\mathcal{C} = \sup_{\mathcal{P}}\ & R_1 + I(U;Y_R) + I(X_1;\hat{Y}_R|UZ), \\ s.t.\ \ & R_0 \geq I(U;Y_R) + I(Y_R;\hat{Y}_R|UZ),\end{aligned} \tag{5.7}$$

where $|\mathcal{U}| \leq |\mathcal{X}_1| + 3$ *and* $|\hat{\mathcal{Y}}_R| \leq |\mathcal{U}||\mathcal{Y}_R| + 1$.

Proof. The achievability part of the theorem is proven in Section 5.2.1, while the converse proof is given in Section 5.2.2. □

In the next section, we show that the capacity of this class of state-dependent relay channels is achieved by the pDCF scheme. To the best of our knowledge, this is the first single relay channel model for which the capacity is achieved by pDCF, while the partial decode-and-forward (pDF) and compress-and-forward (CF) schemes are both suboptimal in general. In addition, the capacity of this relay channel is in general below the cut-set bound [24]. In certain cases, pDF or CF is sufficient to achieve the capacity, e.g., pDF is optimal when the channel state Z is absent or constant. These issues are discussed in more detail in Sections 5.3 and 5.4.

5.2 Proof of Theorem 14

We first show in Section 5.2.1 that the capacity region in Theorem 14 is achievable by pDCF. Then, we derive the converse for Theorem 14 in Section 5.2.2.

5.2.1 Achievability

We derive the rate achievable by pDCF scheme for ORC-D using the achievable rate expression for the pDCF scheme proposed in [74] for the general relay channel. The discrete memoryless relay channel consists of four finite sets $\mathcal{X}$, $\mathcal{X}_R$, $\mathcal{Y}$ and $\mathcal{Y}_R$ and a set of probability distribution $p(y, y_R|x, x_R)$. In this setup, X corresponds to the source input to the channel, Y to the channel output available at the destination, while Y_R is the channel output available at the relay, and X_R is the channel input symbol chosen by the relay. We note that the three terminal relay channel in [74] reduces to the ORC-D-channel by setting $X^n = (X_1^n, X_2^n)$ and $Y^n = (Y_1^n, Y_2^n, Z^n)$, and $p(y, r_R|x_1 x_R) = p(y_1, y_2, y_R, z|x_1, x_2, x_R) = p(z)p(y_R|x_1, z)p(y_1|x_R)p(y_2|x_2)$.

In pDCF for the general relay channel, the source applies message splitting, and the relay decodes only a part of the message. The part to be decoded by the relay is transmitted through the auxiliary random variable U^n, while the rest of the message is superposed onto this through channel input X^n. Block Markov encoding is used for transmission. The relay receives Y_R^n and decodes only the part of the message that is conveyed by U^n. The remaining signal Y_R^n is compressed into $\hat{Y}_R^n$. The decoded message is forwarded through V^n, which is correlated with U^n, and the compressed signal is superposed onto V^n through the relay channel input X_R^n. At the destination the received signal Y^n is used to recover the message. See [74] for details. The achievable rate of the pDCF scheme is given below.

Theorem 15. (Theorem 7,[74]) *The capacity of a relay channel $p(y, y_R|x, x_R)$ is lower bounded by the following rate:*

$$\begin{aligned} R_{pDCF} = \sup\min\{&I(X; Y, \hat{Y}_R|X_R, U) + I(U; Y_R|X_R, V), \\ &I(X, X_R; Y) - I(\hat{Y}_R; Y_R|X, X_R, U, Y)\}, \\ s.t.\ \ &I(\hat{Y}_R; Y_R|Y, X_R, U) \leq I(X_R; Y|V), \end{aligned} \tag{5.8}$$

where the supremum is taken over all joint pmf's of the form

$$p(v)p(u|v)p(x|u)p(x_1|v)p(y, y_R|x, x_R)p(\hat{y}_R|x_R, y_R, u).$$

Since ORC-D is a special case of the general relay channel model, the rate R_{pDCF} is achievable in an ORC-D as well. The capacity achieving pDCF scheme for the state-dependent channel from (5.8) is obtained by setting $V = \emptyset$ and generating X_R^n and X_1^n independent of the rest of variables with distribution $p^*(x_R)$ and $p^*(x_1)$, respectively, as given in the next lemma.

Lemma 20. *For the class of relay channels characterized by the ORC-D model, the capacity expression $\mathcal{C}$ defined in (5.7) is achievable by the pDCF scheme.*

Proof. See Appendix E. □

The optimal pDCF scheme for ORC-D applies independent coding over the source-destination and the source-relay-destination branches. The source applies message splitting. Part of the message is transmitted over the source-destination branch and decoded at the destination using Y_2^n and Z^n. In the relay branch, the part of the message to be decoded at the relay is transmitted through U^n, while the rest of the message is superposed onto this through the channel input X_1^n. At the relay the part conveyed by U^n is decoded from Y_R^n, and the remaining signal Y_R^n is compressed into $\hat{Y}_R^n$ using binning and assuming that Z^n is available at the decoder. Both U^n and the bin index corresponding to $\hat{Y}_R^n$ are transmitted over the relay-destination channel using X_R^n. At the destination, X_R^n is decoded from Y_1^n, and U^n and the bin index are recovered. Then, the decoder looks for the part of message transmitted over the relay branch jointly typical with $\hat{Y}_R^n$ within the corresponding bin and Z^n.

5.2.2 Converse

The proof of the converse consists of two parts. First we derive a single-letter upper bound on the capacity, and then, using the single-letter expression of the upper bound we provide an alternative expression for this bound, which coincides with the rate achievable by pDCF.

Lemma 21. *The capacity of the class of relay channels characterized by the ORC-D model is upper bounded by*

$$R_{up} = \sup_{\mathcal{P}} \min\{R_1 + I(U;Y_R) + I(X_1;\hat{Y}_R|UZ), R_1 + R_0 - I(\hat{Y}_R;Y_R|X_1UZ)\}. \quad (5.9)$$

Proof. See Appendix E. □

As stated in the next lemma, the upper bound R_{up}, given in Lemma 21, is equivalent to the capacity expression $\mathcal{C}$ given in Theorem 14. Since the achievable rate meets the upper bound, this concludes the proof of Theorem 14.

Lemma 22. *The upper bound on the achievable rate R_{up} given in Lemma 21 is equivalent to the capacity expression $\mathcal{C}$ in Theorem 14.*

Proof. See Appendix E. □

5.3 The Two-Hop Relay Channel with State: Suboptimality of Pure pDF and CF schemes

We have seen in Section 5.2 that the pDCF scheme is capacity-achieving for the class of relay channels characterized by the ORC-D model. In order to prove the suboptimality of the pure DF and CF schemes for this class of relay channels, we consider a simplified system model, called the *two-hop relay channel with state information available at the destination* (MRC-D), which is obtained by simply removing the direct channel from the source to the destination, i.e., $R_1 = 0$.

The capacity of this two-hop relay channel model and the optimality of pDCF follows directly from Theorem 14. However, the single-letter capacity expression depends on the joint pmf of X_1, Y_R, X_R and Y_1 together with the auxiliary random variables U and $\hat{Y}_R$. Unfortunately, the numerical characterization of the optimal joint pmf of these random variables is very complicated for most channels. A simple and computable upper bound on the capacity can be obtained from the cut-set bound [25]. For MRC-D, the cut-set bound is given by

$$R_{CS} = \min\{R_0, \max_{p(x_1)} I(X_1; Y_R|Z)\}. \tag{5.10}$$

Next, we characterize the rates achievable by the DF and CF schemes for MRC-D. Since they are special cases of the pDCF scheme, their achievable rates can be obtained by particularizing the achievable rate of pDCF for this setup.

DF Scheme

If we consider a pDCF scheme that does not perform any compression at the relay, i.e., $\hat{Y_R} = \emptyset$, we obtain the rate achievable by the pDF scheme. Note that the optimal distributions of X_R is given by $p^*(x_r)$. Then, we have

$$R_{pDF} = \min\{R_0, \sup_{p(x_1,u)} I(U; Y_R)\}. \tag{5.11}$$

From the Markov chain $U - X_1 - Y_R$, we have that $I(U; Y_R) \leq I(X_1; Y_R)$, where the equality is achieved by $U = X_1$. That is, the performance of pDF is maximized by letting the relay decode the whole message. Therefore, the maximum rate achievable by pDF and DF for MRC-D coincide, and is given by

$$R_{DF} = R_{pDF} = \min\{R_0, \max_{p(x_1)} I(X_1; Y_R)\}. \tag{5.12}$$

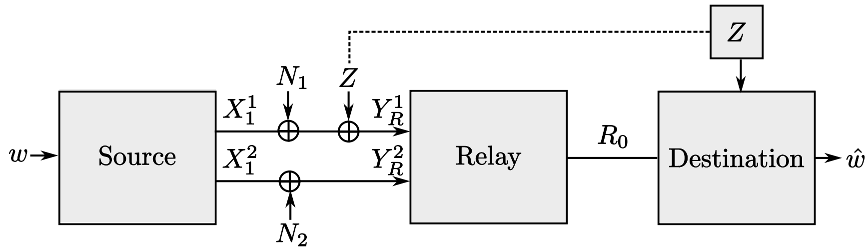


Figure 5.2: The parallel binary symmetric MRC-D with parallel source-relay links. The destination has side information about only one of the source-relay links.

CF Scheme

If the pDCF scheme does not perform any decoding at the relay, i.e., $U = \emptyset$, pDCF reduces to CF. Then, the achievable rate for the CF scheme in MRC-D is easily seen to be given by

$$\begin{aligned} R_{CF} = \sup & \, I(X_1; \hat{Y}_R | Z) \\ & \text{s.t. } R_0 \geq I(\hat{Y}_R; Y_R | Z), \\ & \text{over } p(x_1)p(z)p(y_R|x_1, z)p(\hat{y}_R|y_R). \end{aligned} \tag{5.13}$$

5.3.1 Two-Hop Parallel Binary Symmetric Channel

In this section we consider a special MRC-D as shown in Fig. 5.2, which we call the *parallel binary symmetric MRC-D*. For this setup, we characterize the optimal performance of the DF and CF schemes, and show that in general pDCF outperforms both, and that in some cases the cut-set bound is tight and coincides with the channel capacity. This example proves the suboptimality of both DF and CF on their own for the ORC-D.

In this scenario, the source-relay channel consists of two parallel binary symmetric channels. We have $X_1 = (X_1^1, X_1^2)$, $Y_R = (Y_R^1, Y_R^2)$ and $p(y_R|x_R, z) = p(y_R^1|x_1^1, z)p(y_R^2|x_1^2)$ characterized by

$$Y_R^1 = X_1^1 \oplus N_1 \oplus Z, \quad \text{and} \quad Y_R^2 = X_1^2 \oplus N_2,$$

where N_1 and N_2 are i.i.d. Bernoulli random variables with $\Pr\{N_1 = 1\} = \Pr\{N_2 = 1\} = \delta$, i.e., $N_1 \sim \text{Ber}(\delta)$ and $N_2 \sim \text{Ber}(\delta)$. We consider a Bernoulli distributed state Z, $Z \sim \text{Ber}(p_z)$, which affects one of the two parallel channels, and is available at the destination. We have $\mathcal{X}_1^1 = \mathcal{X}_1^2 = \mathcal{Y}_R^1 = \mathcal{Y}_R^1 = \mathcal{N}_1 = \mathcal{N}_2 = \mathcal{Z} = \{0, 1\}$.

From (5.10), the cut-set bound is given by

$$\begin{aligned} R_{CS} &= \min\{R_0, \max_{p(x_1^1 x_1^2)} I(X_1^1 X_1^2; Y_R^1 Y_R^2 | Z)\} \\ &= \min\{R_0, 2(1 - h_2(\delta))\}, \end{aligned} \tag{5.14}$$

where $h_2(\cdot)$ is the binary entropy function defined as $h_2(p) \triangleq -p \log p - (1-p)\log(1-p)$.

The maximum DF rate is achieved by $X_1^1 \sim \text{Ber}(1/2)$ and $X_1^2 \sim \text{Ber}(1/2)$, and is found to be

$$\begin{aligned} R_{DF} &= \min\{R_0, \max_{p(x_1^1 x_1^2)} I(X_1^1 X_1^2; Y_R^1 Y_R^2)\} \\ &= \min\{R_0, 2 - h_2(\delta \star p_z) - h_2(\delta)\}, \end{aligned} \tag{5.15}$$

where $\alpha \star \beta \triangleq \alpha(1-\beta) + (1-\alpha)\beta$.

Following (5.13), the rate achievable by the CF scheme in the parallel binary symmetric MRC-D is given by

$$\begin{aligned} R_{CF} = &\max I(X_1^1 X_1^2, \hat{Y}_R | Z), \\ &\text{s.t. } R_0 \geq I(Y_R^1 Y_R^2; \hat{Y}_R | Z) \\ &\text{over } p(z)p(x_1^1 x_1^2)p(y_R^1|z, x_1^1)p(y_R^2|x_2)p(\hat{y}_R|y_R^1 y_R^2). \end{aligned} \tag{5.16}$$

Let us define $h_2^{-1}(q)$ as the inverse of the entropy function $h_2(p)$ for $q \geq 0$. For $q < 0$, we define $h_2^{-1}(q) = 0$.

As we show in the next lemma, the achievable CF rate in (5.16) is maximized by transmitting independent channel inputs over the two parallel links to the relay by setting $X_1^1 \sim \text{Ber}(1/2)$, $X_1^2 \sim \text{Ber}(1/2)$, and by independently compressing each of the channel outputs Y_R^1 and Y_R^2 as $\hat{Y}_R^1 = Y_R^1 \oplus Q_1$ and $\hat{Y}_R^2 = Y_R^2 \oplus Q_2$, respectively, where $Q_1 \sim \text{Ber}(h_2^{-1}(1 - R_0/2))$ and $Q_2 \sim \text{Ber}(h_2^{-1}(1 - R_0/2))$. Note that for $R_0 \geq 2$, the channel outputs can be compressed errorlessly. The maximum achievable CF rate is given in the following lemma.

Lemma 23. *The maximum rate achievable by CF in the parallel binary symmetric MRC-D is given by*

$$R_{CF} = 2\left(1 - h_2\left(\delta \star h_2^{-1}\left(1 - \frac{R_0}{2}\right)\right)\right). \tag{5.17}$$

Proof. See Appendix E. □

Now, we consider the pDCF scheme for the parallel binary symmetric MRC-D. Although we have not been able to characterize the optimal choice of $(U, \hat{Y}_R, X_1^1, X_1^2)$ in general, we provide an achievable scheme that outperforms both DF and CF schemes

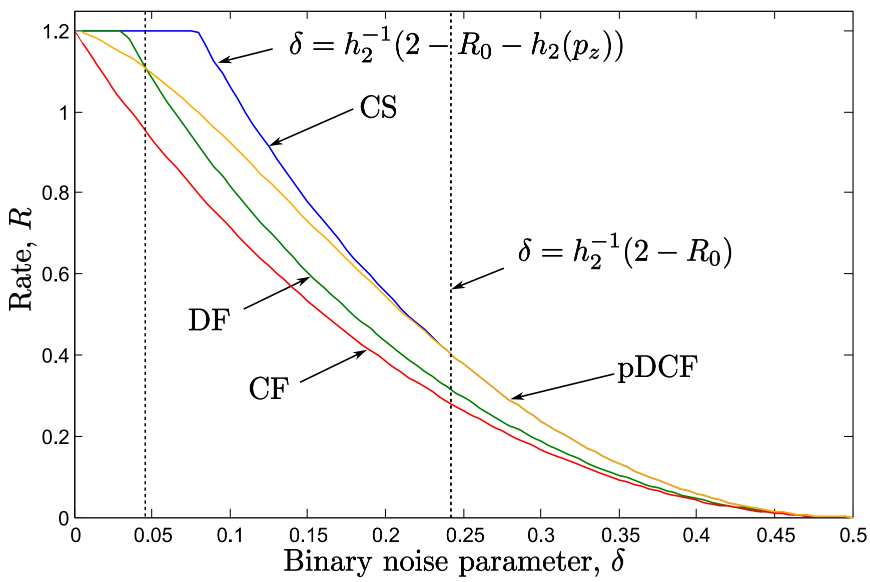


Figure 5.3: Achievable rates and the cut-set upper bound for the parallel binary symmetric MRC-D with respect to the binary noise parameter δ, for $R_0 = 1.2$ and $p_z = 0.15$.

and meets the cut-set bound in some regimes. Let $X_1^1 \sim \text{Ber}(1/2)$ and $X_1^2 \sim \text{Ber}(1/2)$ and $U = X_1^2$, i.e., the relay decodes the channel input X_1^2, while Y_R^1 is compressed using $\hat{Y}_R = Y_R^1 + Q$, where $Q \sim \text{Ber}(h_2^{-1}(2 - h_2(\delta) - R_0))$. The rate achievable by this scheme is given in the following lemma.

Lemma 24. *A lower bound on the achievable pDCF rate in the parallel binary symmetric MRC-D is given by*

$$R_{pDCF} \geq \min\{R_0, 2 - h_2(\delta) - h_2\left(\delta \star h_2^{-1}\left(2 - h_2(\delta) - R_0\right)\right)\}.$$

Proof. See Appendix E. □

We notice that for $p_z \leq h_2^{-1}(2 - h_2(\delta) - R_0)$, or equivalently, $\delta \leq h_2^{-1}(2 - h_2(p_z) - R_0)$, the proposed pDCF is outperformed by DF. In this regime, pDCF can achieve the same performance by decoding both channel inputs, reducing to DF.

Comparing the cut-set bound expression in (5.14) with R_{DF} in (5.15) and R_{CF} in (5.17), we observe that DF achieves the cut-set bound if $R_0 \leq 2 - h(\delta \star p_z) - h(\delta)$ while R_{CF} coincides with the cut-set bound if $R_0 \geq 2$. On the other hand, the proposed suboptimal pDCF scheme achieves the cut-set bound if $R_0 \geq 2 - h_2(\delta)$, i.e., for $\delta \geq h_2^{-1}(2 - R_0)$. Hence, the capacity of the parallel binary symmetric MRC-D in this regime is achieved by pDCF, while both DF and CF are suboptimal, as stated in the next lemma.

Lemma 25. *If $R_0 < 2$ and $\delta \geq h_2^{-1}(2 - R_0)$, pDCF achieves the capacity of the parallel binary symmetric MRC-D, while pure CF and DF are both suboptimal under these constraints. For $R_0 \geq 2$, both CF and pDCF achieve the capacity.*

The achievable rates of DF, CF and pDCF, together with the cut-set bound are shown in Fig. 5.3 with respect to δ for $R_0 = 1.2$ and $p_z = 0.15$. We observe that in this setup, DF outperforms CF in general, while for $\delta \leq h_2^{-1}(2 - R_0 - h_2(p_z)) = 0.0463$, DF outperforms pDCF as well. We also observe that pDCF meets the cut-set bound for $\delta \geq h_2^{-1}(2 - R_0) = 0.2430$, characterizing the capacity in this regime, and proving the suboptimality of both the DF and CF schemes when they are used on their own.

5.3.2 Two-Hop Binary Symmetric Channel

In order to gain further insights into the proposed pDCF scheme, we look into the *binary symmetric* MRC-D, in which, there is only a single channel connecting the source to the relay, given by

$$Y_R = X_1 \oplus N \oplus Z, \tag{5.18}$$

where $N \sim \text{Ber}(\delta)$ and $Z \sim \text{Ber}(p_z)$.

Similarly to Section 5.3.1, the cut-set bound and the maximum achievable rates for DF and CF are found as

$$R_{CS} = \min\{R_0, 1 - h_2(\delta)\}, \tag{5.19}$$

$$R_{DF} = \min\{R_0, 1 - h_2(\delta \star p_z)\}, \tag{5.20}$$

$$R_{CF} = 1 - h_2(\delta \star h_2^{-1}(1 - R_0))), \tag{5.21}$$

where R_{DF} is achieved by $X_1 \sim \text{Ber}(1/2)$, and R_{CF} can be shown to be maximized by $X_1 \sim \text{Ber}(1/2)$ and $\hat{Y}_R = Y_R \oplus Q$, where $Q \sim \text{Ber}(h_2^{-1}(1 - R_0))$ similarly to Lemma 23. Note that, for Y_R independent of Z, i.e., $p_z = 0$, DF achieves the cut-set bound while CF is suboptimal. However, CF outperforms DF whenever $p_z \geq h_2^{-1}(1 - R_0)$.

For the pDCF scheme, we consider binary $(U, X_1, \hat{Y}_R)$, with $U \sim \text{Ber}(p)$, a superposition codebook $X_1 = U \oplus W$, where $W \sim \text{Ber}(q)$, and $\hat{Y}_R = Y_R \oplus Q$ with $Q \sim \text{Ber}(\alpha)$. As stated in the next lemma, the maximum achievable rate of this pDCF scheme is obtained by reducing it to either DF or CF, depending on the values of p_z and R_0.

Lemma 26. *For the binary symmetric MRC-D, pDCF with binary $(U, X_1, \hat{Y}_R)$ achieves the following rate.*

$$R_{pDCF} = \max\{R_{DF}, R_{CF}\} = \begin{cases} \min\{R_0, 1 - h_2(\delta \star p_z)\} & \text{if } p_z < h_2^{-1}(1 - R_0), \\ 1 - h_2(\delta \star h_2^{-1}(1 - R_0)) & \text{if } p_z \geq h_2^{-1}(1 - R_0). \end{cases}$$

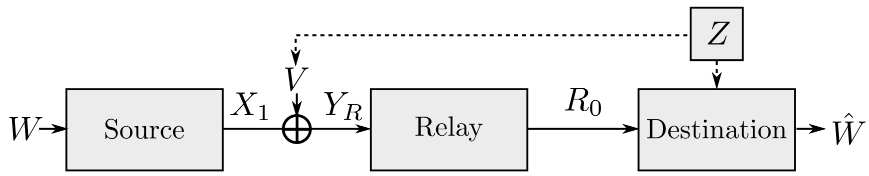


Figure 5.4: The two-hop Gaussian relay channel with source-relay channel state information available at the destination.

This result justifies the pDCF scheme proposed in Section 5.3.1 for the parallel binary symmetric MRC-D. Since the channel $p(y_1^2|x_2)$ is independent of the channel state Z, the largest rate is are achieved if the relay decodes X_1^2 from Y_R^2. However, for channel $p(y_1^1|x_1, z)$, which depends on Z, the relay either decodes X_1^1, or compress Y_R^1, depending on p_z.

5.3.3 Two-Hop Gaussian Channel with State

Next, we consider an AWGN two-hop channel, which we denote by *Gaussian* MRC-D, in which the source-relay link is characterized by $Y_R = X_1 + V$, while the destination has access to correlated state information Z. We assume that V and Z are zero mean jointly Gaussian random variables with a covariance matrix

$$\mathbf{C}_{ZV} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{5.22}$$

The channel input at the source has to satisfy the power constraint $E[|X_1^n|^2] \leq nP$. Finally, the relay and the destination are connected by a noiseless link of rate R_0 (see Fig. 5.4 for the channel model).

In this case, the cut-set bound is given by

$$R_{CS} = \min\left\{R_0, \frac{1}{2}\log\left(1 + \frac{P}{1-\rho^2}\right)\right\}. \tag{5.23}$$

It easy to characterize the optimal DF rate, achieved by a Gaussian input, as follows:

$$R_{DF} = \min\left\{R_0, \frac{1}{2}\log(1+P)\right\}. \tag{5.24}$$

For CF and pDCF, we consider the achievable rate when the random variables $(X_1, U, \hat{Y}_R)$ are constrained to be jointly Gaussian, which is a common assumption in evaluating achievable rates, yet potentially suboptimal. For CF, we generate the compression codebook using $\hat{Y}_R = Y_R + Q$, where $Q \sim \mathcal{N}(0, \sigma_q^2)$. Optimizing over σ_q^2,

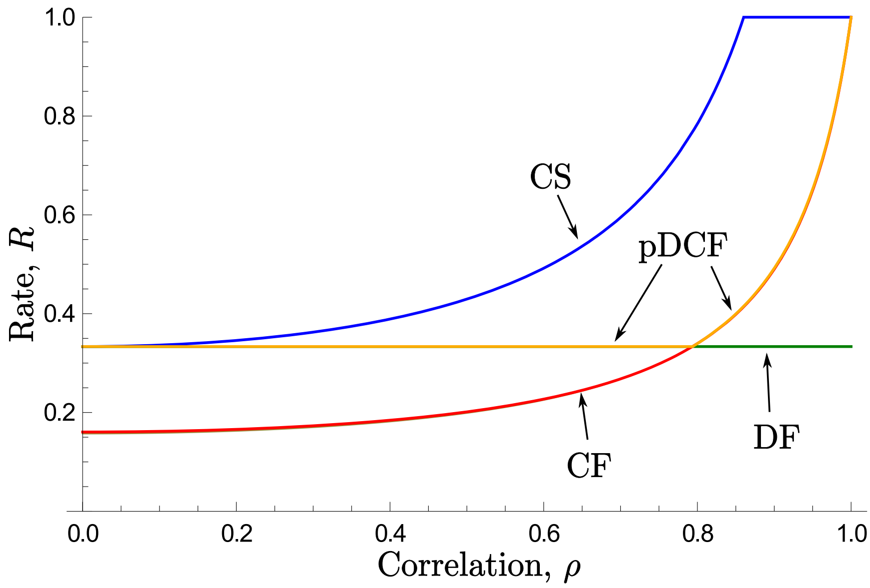


Figure 5.5: Achievable rates and the cut-set upper bound for the two-hop AWGN relay channel with source-relay channel state information at the destination for $R_0 = 1$ and $P = 0.3$.

the maximum achievable rate is given by

$$R_{CF} = R_0 - \frac{1}{2}\log\left(\frac{P + 2^{2R_0}(1-\rho^2)}{P+1-\rho^2}\right). \tag{5.25}$$

For pDCF, we let $U \sim \mathcal{N}(0, \alpha P_1)$, and $X_1 = U + T$ to be a superposition codebook where T is independent of U and distributed as $T \sim \mathcal{N}(0, \bar{\alpha} P_1)$, where $\bar{\alpha} \triangleq 1 - \alpha$. We generate a quantization codebook using the test channel $\hat{Y}_R = Y_R + Q$ as in CF. Next lemma shows that with this choice of random variables, pDCF reduces either to pure DF or pure CF, similarly to the two-hop binary model in Section 5.3.2.

Lemma 27. *The optimal achievable rate for pDCF with jointly Gaussian $(X_1, U, \hat{Y}_R)$ is given by*

$$R_{pDCF} = \max\{R_{DF}, R_{CF}\} = \begin{cases} \min\{R_0, 1/2\log(1+P)\} & \text{if } \rho^2 \leq 2^{-2R_0}(1+P), \\ R_0 - \frac{1}{2}\log\left(\frac{P+2^{2R_0}(1-\rho^2)}{P+1-\rho^2}\right) & \text{if } \rho^2 > 2^{-2R_0}(1+P). \end{cases}$$

Proof. See Appendix E. □

In Fig. 5.5 the achievable rates are compared with the cut-set bound. It is shown that DF achieves the best rate when the correlation coefficient ρ is low, i.e., when the destination has low quality channel state information, while CF achieves higher rates for higher values of ρ. It is seen that pDCF achieves the best of the two transmission schemes. Note also that for $\rho = 0$ DF achieves the cut-set bound, while for $\rho = 1$ CF achieves the cut-set bound.

Although this example proves the suboptimality of the DF scheme for the channel model under consideration, it does not necessarily lead to the suboptimality of the CF scheme as we have constrained the auxiliary random variables to Gaussian.

5.4 Comparison with the Cut-Set Bound

In the examples considered in Section 5.3, we have seen that for certain conditions, the choice of certain random variables allows us to show that the cut-set bound and the capacity coincide. For example, we have seen that for the parallel binary symmetric MRC-D the proposed pDCF scheme achieves the cut-set bound for $\delta \geq h_2^{-1}(2 - R_0)$, or Gaussian random variables meet the cut-set bound for $\rho = 0$ or $\rho = 1$ in the Gaussian MRC-D. An interesting question is whether the capacity expression in Theorem 14 always coincides with the cut-set bound or not; that is, whether the cut-set bound is tight for the relay channel model under consideration.

To address this question, we consider the two-hop binary channel in (5.18) for $Z \sim \text{Ber}(1/2)$. The capacity $\mathcal{C}$ of this channel is given in the following lemma.

Lemma 28. *The capacity of the binary symmetric MRC-D with $Y_R = X_1 \oplus N \oplus Z$, where $N \sim$ Ber(δ) and $Z \sim$ Ber$(1/2)$, is achieved by CF and pDCF, and is given by*

$$\mathcal{C} = 1 - h_2(\delta \star h_2^{-1}(1 - R_0)). \tag{5.26}$$

Proof. See Appendix E. □

From (5.19), the cut-set bound is given by $R_{CS} = 1 - h_2(\delta)$. It then follows that in general the capacity is below the cut-set bound. Note that for this setup, $R_{DF} = 0$ and pDCF reduces to CF, i.e., $R_{pDCF} = R_{CF}$. See Fig. 5.6 for comparison of the capacity with the cut-set bound for varying δ values.

CF suffices to achieve the capacity of the binary symmetric MRC-D for $Z \sim \text{Ber}(1/2)$. While in general pDCF outperforms DF and CF, in certain cases these two schemes are sufficient to achieve the cut-set bound, and hence, the capacity. For the ORC-D model introduced in Section 5.1, the cut-set bound is given by

$$R_{CS} = R_1 + \min\{R_0, \max_{p(x_1)} I(X_1; Y_R|Z)\}. \tag{5.27}$$

Next, we present four cases for which the cut-set bound is achievable, and hence, is the capacity:

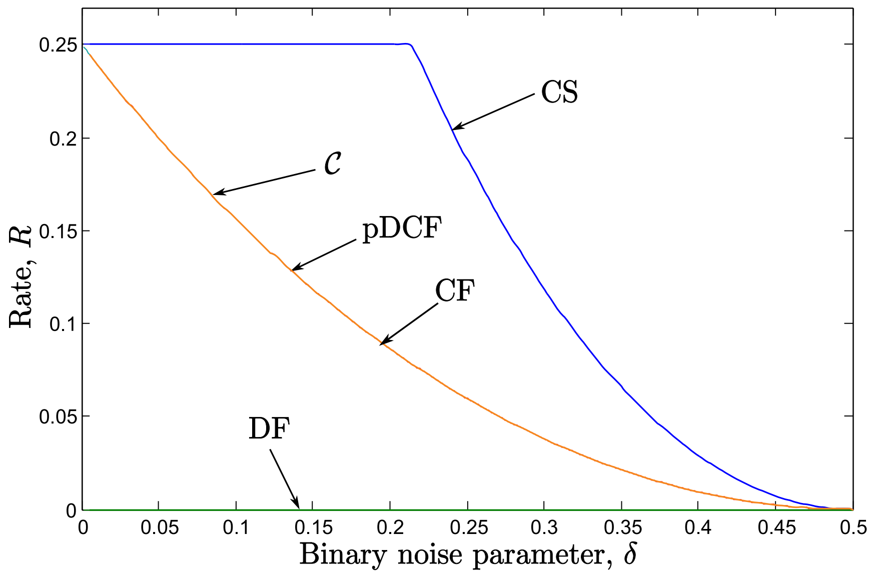


Figure 5.6: Achievable rates, capacity and cut-set upper bound for the two-hop binary relay channel with respect to δ for $R_0 = 0.25$ and $p_z = 0.5$.

1. If $I(Z;Y_R) = 0$, the setup reduces to the class of orthogonal relay channels studied in [91], for which the capacity is known to be achieved by pDF.
2. If $H(Y_R|X_1Z) = 0$, i.e., Y_R is a deterministic function of X_1 and Z, the capacity, given by
$$R_1 + \min\{R_0, \max_{p(x_1)} I(X_1;Y_R|Z)\},$$
is achievable by CF.
3. If $\max_{p(x_1)} I(X_1;Y_R) \geq R_0$, the capacity, given by $\mathcal{C} = R_1 + R_0$, is achievable by pDF.
4. Let $\arg\max_{p(x_1)} I(X_1;Y_R|Z) = \bar{p}(x_1)$. If $R_0 > H(\bar{Y}_R|Z)$ for $\bar{Y}_R$ induced by $\bar{p}(x_1)$, the capacity, given by $R_1 + I(\bar{X}_1;\bar{Y}_R|Z)$, is achievable by CF.

Proof. See Appendix E. □

These cases can be observed in the examples from Section 5.3. For example, in the Gaussian MRC-D, with $\rho = 0$, Y_R is independent of Z, and thus, DF meets the cut-set bound as stated in case 1. Similarly, for $\rho = 1$ CF meets the cut-set bound since Y_R is a deterministic function of X_R and Z, which corresponds to case 2.

For the parallel binary symmetric MRC-D in Section 5.3.1, pDCF achieves the cut-set bound if $\delta \geq h_2^{-1}(2-R_0)$ due to the following reasoning. Since Y_R^1 is independent of X_1^1, from case 1, DF should achieve the cut-set bound. Once X_1^1 is decoded, the available rate to compress Y_2 is given by $R_0 - I(X_1;Y_1) = R_0 - 1 + h_2(\delta)$, and the entropy of Y_2, given the channel state at the destination, is given by $H(Y_2|Z) = 1 - h_2(\delta)$. For $\delta \geq h_2^{-1}(2 - R_0)$ we have $R_0 - I(X_1; Y_1) \geq H(Y_2|Z)$. Therefore the relay can compress Y_2 losslessly, and transmit to the destination. This corresponds to case 4. Thus, the capacity characterization in the parallel binary symmetric MRC-D is due to a combination of cases 1 and case 4.

5.5 Conclusion

We have considered a class of orthogonal relay channels, in which the source and the relay are connected with a channel that depends on a state sequence, known at the destination. We have characterized the capacity of this class of relay channels, and shown that it is achieved by the partial decode-compress-and-forward (pDCF) scheme. This is the first three-terminal relay channel model for which the pDCF is shown to be capacity achieving while partial decode-and-forward (pDF) and compress-and-forward (CF) schemes are both suboptimal in general. We have also shown that, in general, the capacity of this channel is below the cut-set bound.

Chapter 6

Conclusions

In mid 20th century, Shannon settled the fundamental principles of information theory and reliable communication. In his groundbreaking paper, Shannon proved that, without incurring any performance loss, the communication problem can be divided into two separate simpler problems: data compression and data transmission. Since then, communication networks has systematically followed this division in their architecture design. This separate operation framework presents significant advantages in terms of simplicity and modularity, which led to the development of highly complex networks, such as the Internet. However, this approach has created a bottleneck in the design of wireless networks, since they significantly differ from wired networks, for which the layered framework was originally designed. Wireless channels are highly dynamic and present a broadcast nature, causing interferences in nearby devices, as opposed to wired channels, which are, essentially, time-invariant and non-interfering.

Moreover, sources and channels in wireless communication networks exhibit statistical correlation, which can stem from the physical nature of the underlying sources or can be created within the network. While current architectures ignore this correlation, communication technologies that exploit it, and go beyond the layered architecture approach, can become a key feature of future high performance networks, as information theory promises significant gains. However, translating the available correlation into performance improvements implies a careful system design and the use of appropriate communication strategies.

In this dissertation, we have studied potential novel technologies for next generation wireless networks from an information theoretic perspective. We have focused on three distinct problems involving the availability of correlated side information in wireless networks, and developed fundamental performance bounds and novel communication schemes that go beyond the classical separate source and channel coding approach.

We have identified operation regimes in which significant performance gains can be expected. In general, joint source-channel coding (JSCC) schemes, such as uncoded transmission, joint decoding, hybrid digital-analog (HDA) transmission and multi-layer transmission, have been shown to be able to provide significant performance gains over separate source and channel (SSCC) schemes. Under certain conditions, the proposed schemes have been shown to achieve the optimal performance, proving the necessity of JSCC in the presence of correlated information in wireless networks. Uncoded transmission plays a major role in the JSCC schemes considered in this thesis and, despite its simplicity, its optimality has been proven in many regimes of operation, in which SSCC is strictly suboptimal. On the other hand, we have also shown that the benefits of JSCC schemes are not restricted to JSCC problems. We have shown that a combination of channel coding and source coding techniques, are required in certain multi-terminal channel problems in order to achieve the channel capacity.

In Chapter 2, we have studied the joint source-channel Gaussian one-helper problem, in which two correlated Gaussian sources are available at two separate terminals and have to be transmitted over a time-invariant Gaussian MAC. Of the two sources, only one of the sources has to be reconstructed at the destination with minimum distortion, while the terminal with the second source acts as a helper.

We have characterized the optimal performance achievable by SSCC and uncoded transmission, and seen that, for low SNR and high correlation regimes, the latter outperforms SSCC. On the contrary, SSCC has been shown to perform better at high SNR and low correlation regimes. The optimality of separation breaks down in this scenario because by exploiting the correlated source sequences, each terminal can generate correlated channel inputs. This is unfeasible for SSCC, which can only generate independent inputs in this distributed setup. The correlation between the channel inputs brings an additional degree-of-freedom to the system, which potentially improves the performance. However, the amount of correlation that can be created is limited by the source correlation. We have used this fact, together with cut-set bound arguments to obtain a lower bound on the achievable distortion.

While we have seen that uncoded transmission outperforms SSCC, the good performance of uncoded transmission is dependent on certain source and channel matching conditions and, in general, uncoded transmission is not capable of exploiting the available degrees-of-freedom in the system. In order to benefit from both digital and analog transmission, we have considered a generalized HDA transmission scheme based on power allocation among digital and analog signals, denoted by I-HDA. The corresponding signals are transmitted by superposition, and the analog transmissions are treated as noise when decoding the digital codewords. This scheme includes pure SSCC and uncoded transmission schemes as particular cases. A second HDA transmission scheme, denoted by S-VQ, has also been considered. In the S-VQ scheme, at each terminal, the source is quantized and superposed onto an uncoded analog layer. These two schemes have been numerically shown to outperform pure SSCC and pure analog transmission by better exploiting the degrees-of-freedom in the system. We have observed that, both HDA schemes reduce to pure analog transmission in certain regimes, for which we conjecture that uncoded transmission is optimal. These results indicate that, even simple uncoded transmission is capable of outperforming SSCC, and in fact, achieve the best known performance, although more advanced schemes have been shown to provide higher gains in general.

In Chapter 3, we have looked at the JSCC problem of transmitting a Gaussian source over a time-varying fading channel with delay constraints and minimum expected distortion. We have studied the benefits of having correlated side information at the receiver whose quality, i.e., correlation with the source signal, varies over time, assuming that the states of the time-varying channel and side information are available only at the destination. In this case, contrary to the helper setup, the side information is provided to the destination through an orthogonal link. The optimality of Shannon's separation breaks down since, under delay constraints, SSCC cannot adapt to the channel and side information variations and suffers from outages in both the source and channel codes. Therefore, JSCC techniques that jointly adapt to the channel and the side information states are required.

We have derived a lower bound on the expected distortion by providing the encoder with the channel state, and shown that in this case, SSCC achieves the optimal performance, although the side information state is unknown at the transmitter. We have proved the optimality of uncoded transmission under discrete and continuous quasiconcave side information fading distributions, by showing that under these distributions uncoded transmission achieves the lower bound, rendering the available side information useless in transmission. This is the first known case, in which uncoded transmission becomes optimal due to fading while it would be suboptimal in the static case. We have also shown that, under this class of distributions, the optimal SSCC scheme ignores the available side information and uses the side information only for reconstruction at the destination. However, for other side information fading distributions, performance improvements can be achieved by exploiting the available side information.

We have shown that SSCC performs poorly compared to joint designs and considered a transmission schemes based on joint source and channel decoding (JDS), compensating bad quality channel states with good side information realizations (or the reverse), thus reducing the outage probability compared to SSCC. We have also considered an HDA scheme based on joint decoding that transmits an uncoded layer on top of a digital layer (SHDA). We have provided results in the finite SNR regime, and shown that, in general, JDS outperforms SSCC and, we have numerically observed that SHDA transmission performs very close to the lower bound. While the optimal transmission strategy remains open for finite SNR values, we have studied the high SNR performance, and characterized the distortion exponent in certain regimes of operation. We have shown that SHDA achieves the optimal distortion exponent for a family of side information distributions. However, in certain regimes of operation, JDS achieves the optimal distortion exponent while SHDA is suboptimal. Therefore, none of the schemes outperform the other in general.

The results in this chapter have been extended in Chapter 4 to MIMO channels and general bandwidth ratios, which provide the system with additional degrees-of-freedom. We have focused on the high SNR regime and assumed that the side information fading follows a Rayleigh distribution, and have considered the effects of its quality. By generalizing the bounds in the previous chapter, we have shown that in the very low bandwidth regime, the optimal distortion exponent is achieved by using only the side information, ignoring the channel output, and in the low bandwidth regime, by ignoring the side information and only using the optimal scheme in the absence of side information.

Then, we have considered larger bandwidth ratios. While in the SISO setup single layer schemes are sufficient, they fall short of the optimal performance in the MIMO setup. SSCC and JDS have been shown to achieve the same distortion exponent despite the latter suffering from fewer outages. Uncoded transmission has been shown to be highly suboptimal, specially for large bandwidth ratios, since it is not capable of exploiting the additional degrees-of-freedom in the system. We have also proposed an HDA scheme, which we call HDA-WZ, and shown that it outperforms the previous schemes. However, these schemes are not sufficient to exploit the degrees-of-freedom in the system, especially in the large bandwidth regime. We have considered multi-layer transmission schemes that transmit successive refinement layers of the source to combat the uncertainty, either in a progressive (LS-JDS) or with a broadcast (BS-JDS) approach. We have shown that these schemes achieve larger distortion exponents and that, in particular, BS-JDS achieves the optimal performance in MISO/SIMO/SISO setups, solving the JSCC problem in the high SNR regime. For the general MIMO channel, we have characterized the optimal distortion in the low bandwidth regime and shown that BS-JDS performs very close to the upper bound. However, we have also observed that LS-JDS outperforms BS-JDS for intermediate bandwidth ratio values, and thus both schemes are required to achieve the largest distortion exponent values, depending on the regime of operation.

In general, providing an estimation of the channel and side information state to the transmitter is costly. Although in Chapters 3 and 4 the transmitter is not aware of the current channel and side information states, the results in these chapters indicate that, in the high SNR regime, some schemes achieve the optimal performance of a transmitter that perfectly knows the channel state. We have also quantified the optimal performance of a transmitter that perfectly knows the channel and side information state, illustrating the gains from the channel state feedback in delay limited JSCC problems. The results in Chapters 3 and 4 indicate that, having correlated side information available at the receiver can be exploited to obtain significant performance gains, despite the transmitter not being aware of the current state of this information.

Finally, in Chapter 5, we have studied the capacity of a class of orthogonal relay channels in the presence of channel side information at the destination. We have modeled the side information in this setting as follows: the source and the relay, and the source and the destination are connected through orthogonal channels that depend on a common state sequence, which is fully known at the destination, and unknown at the source and the relay. While this is essentially a channel coding problem, source compression techniques are required to optimally exploit the side information. We have considered the partial-decode-and-forward (pDF) scheme, which decodes part of the message at the relay and forwards it to the destination, removing the channel noise but rendering the channel state available at the destination useless. We have also considered compress-and-forward (CF), which compresses the relay received signal and forwards to the destination, which decodes the message using the state information available. However, CF also forwards the channel noise. We have characterized the capacity of this class of relay channels, and proved the optimality of the partial decode-compress-and-forward scheme (pDCF), which combines both DF and CF. To the best of our knowledge, this is the first three terminal channel model for which partial-decode-compress-and-forward has been shown to be optimal. We have also proved that, in general, neither the pDF nor the CF scheme can achieve the capacity on its own. The results in this chapter show the importance of a fundamental understanding on how to exploit the side information in communication networks in order to translate the available side information to significant performance improvements.

Even though the nature of the side information and the performance measure in the three scenarios studied in this thesis are quite different, our results concerning these three different scenarios emphasize the significant benefits of exploiting correlated side information when designing a communication system.

Future Research Directions

In this thesis we have studied several open problems concerning the availability of correlated side information in the network and the use of JSCC schemes to efficiently exploit it. While in some cases we have fully solved these problems, there are still a plethora of questions that could be further explored to improve the understanding of fundamental benefits of side information. Here, we discuss some potential research direction that can be pursued.

In Chapter 2, we have provided achievable schemes based on separate source and channel coding, uncoded transmission and HDA schemes that combine the two schemes, however, their performance is far from the proposed lower bound. We believe that the large gap between the lower and upper bounds is caused by the looseness of the lower bound. Tightening this bound is a challenging problem that requires significant effort. Advances in this line of research could lead to improved performance bounds for the transmission of correlated sources in other multi-terminal setups, such as the JSCC interference channel, where little is known.

In Chapters 3 and 4, we have fully characterized the minimum expected distortion in some regimes, for which uncoded transmission has been shown to be optimal. However, most of the optimality results in these chapters have been obtained in terms of the distortion exponent, that is, in the high SNR regime. Further research for the expected distortion in the finite SNR regimes is required. Although this is a very challenging problem that remains open even in the absence of side information, it would be interesting to study the optimal performance of LS-JDS and BS-JDS in the finite SNR regime, and quantify the potential improvements with respect to SSCC based schemes. We expect this improvement to be quite significant. A noticeable performance improvement is observed in the simulations even when a single layer is used. It would also be interesting to further investigate why the achievable distortion exponent for BC-JDS reduces to the performance of JDS in certain regimes, while this does not occur in the absence of side-information. We have also quantified the optimal performance when perfect feedback of the channel and/or side information states is available to the transmitter. An interesting research direction would be to study this problem when the feedback is rate limited, and to characterize the optimal resource allocation among the channel and the side information feedback, as well as to identify the performance of the schemes proposed in this thesis in this model.

Extensions of the tools developed in Chapters 3 and 4 to multi-terminal problems, such as the MAC, the BC, the IC, the relay channel, or a time-varying version of the one-helper problem studied in Chapter 2, would be interesting. We believe that techniques that combine joint decoding, HDA and multi-layer schemes could provide significant gains in these JSCC scenarios, as well.

Finally, in Chapter 5, we have characterized the capacity of a class of relay channel with state information at the destination. More general relay models can be considered that go beyond the orthogonal setup studied in this chapter. However, we believe that this is an extremely challenging problem, since even simpler relay scenarios without channel state, have remained unsolved for many years.

There are still many challenges and open problems to be solved concerning the availability of side information. As wireless networks become more densified in future generations and the M2M paradigm becomes popular, we believe that the role of correlated side information in the network will become a key feature to improve the network performance. Information theory will be instrumental in developing novel transmission strategies that go beyond Shannon's separation paradigm, and settle the fundamental principles of the networks of tomorrow. We hope that the research work presented in this Ph.D. thesis has contributed to achieving this goal, as well as to raise awareness about the potential benefits of side information explotation in wireless networks.

Appendices

Appendix A

Some Notions and Results

In this appendix, we briefly review the notions of types and strong typicality that are used in this thesis, following [92] .

Let X and Y be a random variable over alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, jointly distributed following $p_{XY}(x, y)$, and with marginal distribution $p_X(x)$ and $p_Y(y)$.

Definition 1. *Given a distribution $p_X(x)$, the type P_{x^n} of an n-tuple x^n is the empirical distribution*

$$P_{x^n} = \frac{1}{n} N(a|x^n)$$

where $N(a|x^n)$ is the number of occurrences of the letter a in x^n.

Definition 2. *The set of all n-tuples x^n with type Q is called the type class Q and denoted by T_Q^n.*

Definition 3. *The set of δ-strongly typical n-tuples according to $p_X(x)$ is denoted by $T_{[X]_\delta}^n$ and is defined by*

$$T_{[X]_\delta}^n = \Big\{ x \in \mathcal{X}^n : \left| \frac{1}{n} N(a|x^n) - P_X(a) \right| \leq \delta \; \forall a \in \mathcal{X} \\ \text{and } N(a|x^n) = 0 \text{ whenever } p_X(x) = 0 \Big\}. \quad \text{(A.1)}$$

The definitions of type and strong typicality can be extended to joint and conditional distributions in a similar manner [92].

Next, we provide some results concerning typical sets used in the thesis.

Lemma 29. *Given a random variable X distributed following $p_X(x)$, for each $x^n \in T_{[X]_\delta}^n$ we have*

$$\left| \frac{1}{n} \log |T_{[X]_\delta}^n| - H(X) \right| \leq \frac{\delta}{|\mathcal{X}|} \quad \text{(A.2)}$$

for sufficiently large n.

Lemma 30. *Given a joint distribution $p_{XY}(x,y)$, if (x_i, y_i) is drawn independent and identically distributed (i.i.d.) with $p_X(x)p_Y(y)$ for $i = 1, \ldots, n$, then*

$$\Pr\{(x^n, y^n) \in T^n_{[XY]_\delta}\} \leq 2^{-n(I(X;Y)-3\delta)}. \tag{A.3}$$

Finally, we have the following lemma.

Lemma 31. *For a joint distribution $p_{XYZ}(x,y,z)$, if (x_i, y_i, z_i) is drawn i.i.d. with $p_X(x)p_Y(y)p_Z(z)$ for $i = 1, \ldots, n$, where $p_X(x)$, $p_Y(y)$ and $p_Z(z)$ are the marginals, then*

$$\Pr\{(x^n, y^n, z^n) \in T^n_{[XYZ]_\delta}\} \leq 2^{-n(I(X;Y,Z)+I(Y;X,Z)+I(Z;Y,X)-4\delta)}. \tag{A.4}$$

Appendix B

Proofs for Chapter 2

B.1 Proof of Lower Bound

Using the rate-distortion functions, we will lower bound some mutual information terms. We have,

$$
\begin{aligned}
I(S_1^n; Y^n | S_2^n) &\overset{(a)}{\geq} I(S_1^n; \hat{S}_1^n | S_2^n) \\
&= \sum_{i=1}^{n} I(S_{1i}; \hat{S}_1^n | S_2^n S_{11}^{i-1}) \\
&\overset{(b)}{\geq} \sum_{i=1}^{n} h(S_{1i} | S_{11}^{i-1} S_2^n) - h(S_{1i} | \hat{S}_{1i} S_{2i}) \\
&\overset{(c)}{=} \sum_{i=1}^{n} h(S_{1i} | S_{2i}) - h(S_{1i} | \hat{S}_{1i} S_{2i}) \\
&= \sum_{i=1}^{n} I(S_{1i}; \hat{S}_{1i} | S_{2i}) \\
&\overset{(d)}{\geq} n \sum_{i=1}^{n} \frac{1}{n} R_{S_1|S_2} \left(\mathrm{E}[(S_{1i} - \hat{S}_{1i})^2] \right) \\
&\overset{(e)}{\geq} n R_{S_1|S_2} \left(\frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[(S_{1i} - \hat{S}_{1i})^2] \right) \\
&\overset{(f)}{\geq} n R_{S_1|S_2}(D + \epsilon),
\end{aligned}
\tag{B.1}
$$

where (a) follows from the data processing inequality, (b) follows since conditioning reduces entropy, (c) follows due to the sources being i.i.d., (d) follows from the definition of $R_{S_1|S_2}(\cdot)$ in (2.6), (e) follows from the convexity of $R_{S_1|S_2}(\cdot)$, and (f) follows since D is achievable.

On the other hand, we have the following

$$
\begin{aligned}
I(S_1^n S_2^n; Y^n) &\overset{(a)}{\geq} I(S_1^n; \hat{S}_1^n) \\
&\overset{(b)}{\geq} \sum_{i=1}^{n} h(S_{1i}) - h(S_{1i}|\hat{S}_{1i}) \\
&\geq \sum_{i=1}^{n} I(S_{1i}; \hat{S}_{1i}) \\
&\overset{(c)}{\geq} n \sum_{i=1}^{n} \frac{1}{n} R_{S_1} \left(\mathrm{E}[(S_{1i} - \hat{S}_{1i})^2] \right) \\
&\overset{(d)}{\geq} n R_{S_1}(D + \epsilon),
\end{aligned}
\tag{B.2}
$$

where (a) follows from the data processing inequality; (b) follows since conditioning reduces entropy, (c) follows from the definition of $R_{S_1}(D_1)$ in (2.5), (d) follows from the convexity of $R_{S_1}(D_1)$ and from the achievability of D.

Next, we upper bound the mutual information terms in (B.2) and (B.1). For (B.2) we have

$$
\begin{aligned}
I(S_1^n; Y^n | S_2^n) &= \sum_{i=1}^{n} h(Y_i|S_2^n Y_1^{i-1}) - h(Y_i|S_1^n S_2^n Y_1^{i-1}) \\
&= \sum_{i=1}^{n} h(Y_i|S_2^n Y_1^{i-1} X_{2i}) - h(Y_i|S_1^n S_2^n Y_1^{i-1} X_{1i} X_{2i}) \\
&\overset{(a)}{\leq} \sum_{i=1}^{n} h(Y_i|X_{2i}) - h(Y_i|X_{1i} X_{2i}) \\
&= \sum_{i=1}^{n} I(X_{1i}; Y_i | X_{2i}),
\end{aligned}
\tag{B.3}
$$

where (a) follows from the Markov chain $Y_i - X_{1i}X_{2i} - S_1^n S_2^n Y_1^{i-1}$.

Then, for (B.1) we have

$$
\begin{aligned}
I(S_1^n S_2^n; Y^n) &= h(Y^n) - \sum_{i=1}^{n} h(Y_i|S_1^n S_2^n Y_1^{i-1}) \\
&\leq h(Y^n) - \sum_{i=1}^{n} h(Y_i|S_1^n S_2^n Y_1^{i-1} X_{1i} X_{2i}) \\
&\overset{(a)}{=} h(Y^n) - h(Y_i|X_{1i} X_{2i}) \\
&\leq \sum_{i=1}^{n} h(Y_i) - h(Y_i|X_{1i} X_{2i}) \\
&= \sum_{i=1}^{n} I(Y_i; X_{1i} X_{2i}),
\end{aligned}
\tag{B.4}
$$

where (a) is due to the Markov chain $Y_i - X_{1i}X_{2i} - S_1^n S_2^n Y_1^{i-1}$.

Expressions in (B.3) and (B.4) can be jointly upper bounded using the following lemma, derived in the context of a MAC channel with feedback in [93].

Lemma 32 (From [93])**.** *Let $\{X_{1i}\}$ and $\{X_{2i}\}$ be zero-mean satisfying $\frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_{ki}^2] \le P_k$ for $k = 1, 2$. Let $Y_i = X_{1i} + X_{2i} + Z_i$ where $Z_i \sim \mathcal{N}(0, N)$ and, for every i, is independent of (X_{1i}, X_{2i}). Let $\rho_n \in [0, 1]$ be defined as*

$$\rho_n \triangleq \frac{|\frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_{1i}X_{2i}]|}{\sqrt{(\frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_{1i}^2])(\frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_{2i}^2])}}. \tag{B.5}$$

Then,

$$\sum_{i=1}^n I(Y_i; X_{1i}X_{2i}) \le \frac{n}{2}\log\left(1 + \frac{P_1 + P_2 + 2\rho_n\sqrt{P_1P_2}}{N}\right),$$
$$\sum_{i=1}^n I(X_{ji}; Y_i|X_{ji}) \le \frac{n}{2}\log\left(1 + \frac{P_j(1-\rho_n^2)}{N}\right),\ j = 1, 2.$$

Next, we bound the correlation between the channel inputs (X_1^n, X_2^n) at each transmitter, denoted in Lemma 32 by ρ_n. While, in the presence of feedback, channel inputs can potentially be arbitrarily correlated, i.e., $0 \le \rho_n \le 1$, in the Gaussian helper setup, the correlation is limited by the correlation between the source sequences, therefore, we have $0 \le \rho_n \le \rho$.

Lemma 33. *The correlation between the channel inputs, denoted by ρ_n and defined in (B.5), is upper bounded by the correlation between the source sequences as follows*

$$0 \le \rho_n \le \rho. \tag{B.6}$$

Proof. The proof uses the following result from [13, Lemma B.2], given next.

Lemma 34. *For any coding scheme with encoding functions of the form $X_i^n = f_i^n(S_i^n)$ for $i = 1, 2$ that satisfy the power constraint (2.2) and reduction $\mathrm{E}[X_{i,k}] = 0$, for $i \in \{1, 2\}$ and $k = 1, ..., n$, and the encoder input sequences are jointly Gaussian as in (2.1) with $0 \le \rho \le 1$ and $\sigma_1^2 = \sigma_2^2 = \sigma^2$, any time-$k$ encoder output $X_{1,k}$ and $X_{2,k}$ satisfy*

$$\frac{\mathrm{E}[X_{1i}X_{2i}]}{\sqrt{\mathrm{E}[X_{1i}^2]}\sqrt{\mathrm{E}[X_{2i}^2]}} \le \rho. \tag{B.7}$$

Using Lemma 34, we can bound ρ_n in (B.5) as follows

$$\begin{aligned}
\rho_n &= \frac{|\frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_{1i}X_{2i}]|}{\sqrt{(\frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_{1i}^2])(\frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_{2i}^2])}} \\
&\overset{(a)}{\leq} \frac{|\frac{1}{n}\sum_{i=1}^n \rho\sqrt{\mathrm{E}[X_{1i}^2]}\sqrt{\mathrm{E}[X_{2i}^2]}|}{\sqrt{(\frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_{1i}^2])(\frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_{2i}^2])}} \\
&\overset{(b)}{\leq} \frac{|\frac{1}{n}\rho\sqrt{\sum_{i=1}^n \mathrm{E}[X_{1i}^2]}\sqrt{\sum_{i=1}^n \mathrm{E}[X_{2i}^2]}|}{\sqrt{(\frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_{1i}^2])(\frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_{2i}^2])}} \\
&= \rho. \qquad \text{(B.8)}
\end{aligned}$$

where (a) is due to (B.7) and (b) follows from the Cauchy-Schwarz inequality. □

Finally, in order to prove Lemma 1, we applying Lemma 32, to upper bound (B.3) and (B.4) and combining with the lower bound in (B.2) and (B.1) we obtain the inequalities in Lemma 1. Applying Lemma 34 we bound ρ_n, which completes the proof.

Appendix C

Proofs for Chapter 3

C.1 Proof of Theorem 2

C.1.1 Separation for Discrete Distributions

For Γ with two states optimality of separation can be obtained as a special case of the model studied in [71]. This result can be extended to M receivers (or states) by combining the converses in [71] and [5, Sec.VII] for M side information states, i.e., $T^n_{i,1}$, $i = 1, ..., M$. The direct part is shown by the concatenation of the optimal source code in [5] and an optimal channel code.

First, we consider the converse. We have,

$$\begin{aligned}
n\mathrm{C} &\overset{(a)}{\geq} I(X^n;Y^n)\\
&\overset{(b)}{\geq} I(S^n;Y^n)\\
&\overset{(c)}{=} I(S^n,T^n_{M,1};Y^n)\\
&\overset{(d)}{\geq} I(S^n;Y^n|T^n_{M,1})\\
&\overset{(e)}{=} I(S^n;Y^n,T^n_{1,1},T^n_{2,1},...,T^n_{M-1,1}|T^n_{M,1}) - I(S^n;T^n_{M-1,1}|Y^n,T^n_{M,1})\\
&\quad -I(X^n;Y^n_{M-2,1}|V^n,Y^n_{M-1,1},Y^n_{M,1}) - \cdots - I(S^n;T^n_{1,1}|Y^n,T^n_{2,1},...,T^n_{M,1})\\
&= \sum_{i=1}^{n}\Big[I(S_i;Y^n,T^n_{1,1},T^n_{2,1},...,T^n_{M-1,1}|T^n_{M,1},S_1^{i-1}) - I(S^n;T_{M-1,i}|Y^n,T^{i-1}_{M-1,1},T^n_{M,1})\\
&\quad -I(S^n;T_{M-2,i}|Y^n,T^i_{M-2,1},T^n_{M-1,1},T^n_{M,1})\\
&\quad -\cdots - I(X^n;Y_{1,i}|V^n,Y^i_{1,1},Y^n_{2,1},...,Y^n_{M,1})\Big],
\end{aligned}$$

where (a) is due to the definition of capacity, (b) is due to the data processing inequality, (c) is due to the Markov chain $T^n_{M,1} - S^n - Y^n$, (d) and (e) are due to the chain rule of the mutual information.

From this point, by applying the steps in [5, Sec.VII] with some slight modifications we obtain

$$\begin{aligned} n\mathrm{C} \geq \sum_{i=1}^{n} \Big[& I(S_i; W_{M,i}|T_{M,i}) + I(S_i; W_{M-1,i}|T_{M-1,i}, W_{M,i}) \\ & + I(S_i; W_{M-2,i}|T_{M-2,i}, W_{M-1,i}, W_{M,i}) + \cdots + I(S_i; W_{1,i}|T_{1,i}, W_{2,i}, ..., W_{M,i}) \Big] \\ = & \sum_{i=1}^{n} \sum_{l=1}^{M} I(S_i; W_{l,i}|T_{l,i}, W_{l+1,i}, ..., W_{M,i}), \end{aligned}$$

where we have defined the random variables

$$\begin{aligned} W_{M,i} &= (Y^n, T_{M-1,1}^{i-1}, T_M^{n/i}), \\ W_{M-1,i} &= (S_1^{i-1}, T_{M-2,1}^{i-1}, T_{M-1,i+1}^{n}, W_{M,i}), \\ W_{M-2,i} &= (T_{M-3,1}^{i-1}, T_{M-2,i+1}^{n}, W_{M-1,i}), \\ W_{M-3,i} &= (T_{M-4,1}^{i-1}, T_{M-3,i+1}^{n}, W_{M-2,i}), \\ &\vdots \\ W_{1,i} &= (T_{1,i+1}^{n}, W_{2,i}), \end{aligned}$$

for $i = 1, ..., n$. Note that the random variables satisfy the Markov chain condition

$$W_{M,i} - \cdots - W_{1,i} - S_i - T_{1,i} - \cdots - T_{M,i}.$$

Applying the usual techniques by defining the auxiliary random variable $Q \sim \mathrm{Unif}[1, n]$, S_{iQ}, W_{iQ}, and T_{iQ} for $i = 1, ..., M$, we obtain the single letter condition,

$$\mathrm{C} \geq R_{HB}(\mathbf{D}). \tag{C.1}$$

Finally, the right hand side of (C.1) is given by the Heegard-Berger rate-distortion function, $R_{HB}(\mathbf{D})$, and does not depend on the number of receivers but only on the sum of the mutual information terms, each one corresponding to a receiver with side information Y_i, as discussed in [43]. Hence, the converse applies for countably many receivers as well.

The achievability follows from Heegard-Berger source coding [5, Sec.VII] followed by channel coding at rate $R = \mathrm{C}$.

C.1.2 Separation for Continuous Quasiconcave Distributions

To prove the optimality of separation when $p_\Gamma(\gamma)$ is a continuous quasiconcave distribution, we construct a lower bound on the expected distortion ED^*_{sta} by discretizing the continuum of side information states, and show that this bound is achievable in the limit of finer discretizations.

We divide the side information state γ into some partition $\mathbf{s}$ given by $[s_0, s_1), [s_1, s_2), ...,$ such that $s_0 = 0 < s_1 < ... < s_i < \cdots$ and $\gamma \in [s_{i-1}, s_i)$ if $s_{i-1} \leq \gamma < s_i$ for some $i = 1, 2, ...$ The length of the partition $[s_{i-1}, s_i)$ is defined by Δs_i, i.e., $\Delta s_i \triangleq s_i - s_{i-1}$. Let us define $\bar{\gamma} > 0$ as the super-level set $\bar{\gamma}$ satisfying (3.11). The partition is chosen such that for some index j, we have $s_j = \bar{\gamma}$. A fading realization belongs to the interval $[s_{i-1}, s_i)$ with probability $p_i = \int_{s_{i-1}}^{s_i} p_\Gamma(\gamma) d\gamma$.

We assume that when γ belongs to the interval $[s_{i-1}, s_i)$, a genie substitutes the current side information sequence $T = \gamma_c S + N$ with a sequence with gain s_i, i.e., $\tilde{T} \triangleq \sqrt{s_i} S + N$. Note that this receiver has a better performance as noise can be added to $\tilde{Y}$ to recover the original side information sequence if required. Hence, the expected distortion for a given partition $\mathbf{s}$, denoted by $ED^*_{gen}(\mathbf{s})$, is a lower bound on the expected distortion of the continuous fading setup. The genie aided system now consists of a countable number of receivers and, due to the optimality of separation under countable number of side information states, $ED^*_{gen}(\mathbf{s})$ is achieved by the concatenation of a Heegard-Berger source encoder with side information states $s_1, s_2, ...$ and a capacity achieving channel code. Then, for a given partition $\mathbf{s}$, we have,

$$ED^*_{gen}(\mathbf{s}) = ED^*_C(\mathrm{C}), \tag{C.2}$$

where $ED^*_C(\cdot)$ is defined in (3.9).

With the channel state h_c known, expected distortion $ED^*_Q(\mathrm{C})$ is achievable with separate source and channel coding by concatenating a single layer source encoder for side information state $\bar{\gamma}$, and a channel code at a rate arbitrarily close to C. Then,

$$ED^*_{gen}(\mathbf{s}) \leq ED^*_{sta} \leq ED^*_Q(\mathrm{C}). \tag{C.3}$$

As the partition gets finer in the sense that $\max_i \Delta s_i \to 0$, the limiting behavior of $ED^*_{gen}(\mathbf{s})$ can be obtained by noting that once the optimality of separation is proved for each $ED^*_{gen}(\mathbf{s})$, the problem reduces to the problem studied in [43]. Hence, by [43, Proposition 4] and [43, Proposition 5], $ED^*_{gen}(\mathbf{s})$ converges to $ED^*_Q(\mathrm{C})$, i.e.,

$$\lim_{\max_i \Delta s_i \to 0} ED^*_{gen}(\mathbf{s}) = ED^*_Q(\mathrm{C}). \tag{C.4}$$

Then from inequality (C.3) we have, in the limit of finer partitions, $ED^*_{sta} = ED^*_Q(\mathrm{C})$. This completes the proof.

C.2 Proof of Lemma 7

In order to show the convergence of ED_{pi}^* to ED_{inf}, first, we construct an upper bound on ED_{pi}^* and we show that this bound converges to ED_{inf} for large enough L.

The lower bound ED_{pi}^* is achieved by the concatenation of a capacity achieving channel code with a single-layer source code targeting the side information state $\bar{\gamma}$, the solution to (3.11), for each realization of H. Instead, we consider that, for a given L the source coding is done targeting the state

$$\bar{\gamma}_L \triangleq \mu - \delta, \tag{C.5}$$

where $\mu \triangleq \mathrm{E}[\Gamma_L]$ is the mean of Γ_L and $\delta \triangleq \sqrt{\sigma_L^2}$. The expected distortion achieved by this scheme is an upper bound on ED_{pi}^* and is found, similarly to ED_{pi}^*, to be given by

$$\begin{aligned} ED_{lay} &\triangleq \mathrm{E}_H\left[ED_Q\left(\frac{1}{2}\log(1+H)\right)\right] \\ &= \int_0^{\bar{\gamma}_L} \frac{p_L(\gamma)}{1+\gamma}d\gamma + \int_h \int_{\bar{\gamma}_L}^{\infty} \frac{p_L(\gamma)p_H(h)}{(1+h)(1+\bar{\gamma}_L)+\gamma-\bar{\gamma}_L}d\gamma dh, \end{aligned}$$

where $ED_Q(R)$ is given as in (3.10) for $\bar{\gamma}$ substituted by $\bar{\gamma}_L$ and $p_L(\gamma)$ is the pdf of Γ_L.

Then, we have the following bound

$$\begin{aligned} ED_{pi^*} - ED_{\text{inf}} &\leq ED_{lay} - ED_{\text{inf}} \\ &= \int_0^{\bar{\gamma}_L} \frac{p_L(\gamma)}{1+\gamma}d\gamma + \int_h \int_{\bar{\gamma}_L}^{\infty} \frac{p_L(\gamma)p_H(h)}{(1+h)(1+\bar{\gamma}_L)+\gamma-\bar{\gamma}_L}d\gamma dh - \int_h \int_\gamma \frac{p_H(h)p_L(\gamma)}{(1+h)(1+\gamma)}d\gamma dh \\ &\overset{(a)}{\leq} \int_0^{\bar{\gamma}_L} p_L(\gamma)dh + \int_h \int_{\bar{\gamma}_L}^{\infty} \frac{p_L(\gamma)p_H(h)}{(1+h)(1+\bar{\gamma}_L)+\gamma-\bar{\gamma}_L}d\gamma dh - \int_h \int_{\mu-\delta}^{\mu+\delta} \frac{p_H(h)p_L(\gamma)}{(1+h)(1+\gamma)}d\gamma dh \\ &= \Pr[\Gamma_L < \bar{\gamma}_L] + \int_h \int_{\bar{\gamma}_L}^{\mu+\delta} \frac{p_L(\gamma)p_H(h)}{(1+h)(1+\bar{\gamma}_L)+\gamma-\bar{\gamma}_L}d\gamma dh \\ &\quad + \int_h \int_{\mu+\delta}^{\infty} \frac{p_L(\gamma)p_H(h)}{(1+h)(1+\bar{\gamma}_L)+\gamma-\bar{\gamma}_L}d\gamma dh - \int_h \int_{\mu-\delta}^{\mu+\delta} \frac{p_H(h)p_L(\gamma)}{(1+h)(1+\gamma)}d\gamma dh \\ &\overset{(b)}{\leq} \Pr[\Gamma_L < \bar{\gamma}_L] + \int_h \int_{\mu-\delta}^{\mu+\delta} \frac{p_L(\gamma)p_H(h)}{(1+h)(1+\bar{\gamma}_L)+\gamma-\bar{\gamma}_L}d\gamma dh \\ &\quad + \Pr[\Gamma_L \geq \mu+\delta] - \int_h \int_{\mu-\delta}^{\mu+\delta} \frac{p_H(h)p_L(\gamma)}{(1+h)(1+\gamma)}d\gamma dh \\ &\overset{(c)}{=} \Pr[|\Gamma_L - \mu| \leq \delta] + \int_h \int_{\mu-\delta}^{\mu+\delta} \frac{h(\gamma-\bar{\gamma}_L)p_L(\gamma)p_H(h)}{((1+h)(1+\bar{\gamma}_L)+\gamma-\bar{\gamma}_L)(1+h)(1+\gamma)}d\gamma dh \\ &\overset{(d)}{\leq} \Pr[|\Gamma_L - \mu| \leq \delta] + \mathrm{E}_H[H]\cdot 2\delta \\ &\overset{(e)}{\leq} \frac{\sigma_L^2}{\delta} + \mathrm{E}_H[H]\cdot 2\delta \end{aligned}$$

where (a) follows since $\frac{1}{(1+\gamma)} \leq 1$ for the first integral, and because we are reducing the integration region in the third one, (b) follows due to

$$\int_h \int_{\mu+\delta}^{\infty} \frac{p_L(\gamma)p_H(h)}{(1+h)(1+\bar{\gamma}_L)+\gamma-\bar{\gamma}_L} d\gamma dh \leq \int_h \int_{\mu+\delta}^{\infty} p_L(\gamma)p_H(h) d\gamma dh$$
$$= \Pr[\Gamma_L \geq \mu+\delta].$$

Then (c) follows since $\bar{\gamma}_L = \mu - \delta$, and subtracting the two integrals, (d) follows from the following bound,

$$\begin{aligned}
&\int_h \int_{\mu-\delta}^{\mu+\delta} \frac{h(\gamma-\bar{\gamma}_L)p_L(\gamma)p_H(h)}{((1+h)(1+\bar{\gamma}_L)+\gamma-\bar{\gamma}_L)(1+h)(1+\gamma)} d\gamma dh \\
&\leq \int_h \int_{\mu-\delta}^{\mu+\delta} h(\gamma-\bar{\gamma}_L)p_L(\gamma)p_H(h) d\gamma dh \\
&\stackrel{(f)}{\leq} \mathrm{E}[H] \cdot (\mu+\delta-\bar{\gamma}_L) \int_{\mu-\delta}^{\mu+\delta} p_L(\gamma) d\gamma \\
&\stackrel{(g)}{\leq} \mathrm{E}[H] \cdot 2\delta
\end{aligned}$$

where (f) follows since $\gamma \leq \mu+\delta$ in the integration region; (g) follows since $\bar{\gamma}_L = \mu - \delta$ and $\int_{\mu-\delta}^{\mu+\delta} p_L(\gamma)d\gamma \leq 1$. Finally, (e) follows from Chebyshev's inequality.

By the choice of $\delta = \sqrt{\sigma_L^2}$, we have

$$ED_{pi}^* - ED_{\inf} \leq \frac{\sigma_L^2}{\delta} + \mathrm{E}[H] \cdot 2\delta = \sqrt{\sigma_L^2} + \mathrm{E}[H] \cdot 2\sqrt{\sigma_L^2},$$

and the difference converges to 0 from the assumption $\sigma_L^2 \to 0$ for $L \to \infty$. This completes the proof.

C.3 Converse

C.3.1 Partially Informed Encoder Upper Bound

In Section 3.3.2 we have seen that for continuous quasiconcave pdfs, ED_{pi}^* is obtained by averaging the expected distortion achievable by the concatenation of a single layer source code designed for the side information state $\bar{\gamma}(h)$ and an optimal channel code for the current channel state h. For each h, the optimal $\bar{\gamma}(h)$ is determined by solving (3.11) with $R = \mathcal{C}(h) = \frac{1}{2}\log(1+h)$. Note that $\bar{\gamma}(h)$ is a random variable dependant on the realization of the channel fading H.

An upper bound on the distortion exponent can be found by lower bounding ED_{pi}^*. First, we note that $ED_Q^*(R)$ in (3.10) is a convex function of R. This follows from the time-sharing arguments and convexity of the Heegard-Berger rate-distortion function

[5]. Then, by Jensen's inequality, we have

$$ED^*_{pi} = \mathrm{E}_H[ED^*_Q(\mathcal{C}(H))] \geq ED^*_Q(\mathrm{E}_H[\mathcal{C}(H)]), \tag{C.6}$$

where

$$ED^*_Q(\mathrm{E}_H[\mathcal{C}(H)]) = \int_0^{\tilde{\gamma}} \frac{p_\Gamma(\gamma)}{1+\gamma} d\gamma + \int_{\tilde{\gamma}}^{\infty} \frac{p_\Gamma(\gamma)}{(\tilde{\gamma}+1)2^{2\mathrm{E}_H[\mathcal{C}(H)]} + \gamma - \tilde{\gamma}} d\gamma, \tag{C.7}$$

and $\tilde{\gamma}$ is the solution to (3.11) with $R = \mathrm{E}_H[\mathcal{C}(H)]$, that is, the ergodic capacity of the channel. Note that $\tilde{\gamma}$ depends only on the ergodic capacity of the channel and not on the current channel state realization, and therefore, is not a random variable, as opposed to $\bar{\gamma}(h)$.

Now, since $\mathcal{C}(h)$ is a concave function of h, applying Jensen's inequality again, we have

$$\mathrm{E}_H[\mathcal{C}(H)] = \mathrm{E}_H\left[\frac{1}{2}\log(1+H)\right] \leq \frac{1}{2}\log(1+\mathrm{E}[H]) = \frac{1}{2}\log(1+\rho), \tag{C.8}$$

that is, the ergodic capacity of the channel is lower than the capacity of a static channel with the same average SNR.

We define, for $\hat{\gamma} \geq 0$,

$$ED_{pe}(\hat{\gamma}) \triangleq \int_0^{\hat{\gamma}} \frac{p_\Gamma(\gamma)}{1+\gamma} d\gamma + \int_{\hat{\gamma}}^{\infty} \frac{p_\Gamma(\gamma)}{(\hat{\gamma}+1)(1+\rho) + \gamma - \hat{\gamma}} d\gamma. \tag{C.9}$$

Then, we have

$$\begin{aligned} ED^*_{pi} &\overset{(a)}{\geq} \int_0^{\tilde{\gamma}} \frac{p_\Gamma(\gamma)}{1+\gamma} d\gamma + \int_{\tilde{\gamma}}^{\infty} \frac{p_\Gamma(\gamma)}{(\tilde{\gamma}+1)(1+\rho) + \gamma - \tilde{\gamma}} d\gamma \\ &\overset{(b)}{\geq} \min_{\hat{\gamma} \geq 0.} ED_{pe}(\hat{\gamma}) \triangleq ED^*_{pe}, \end{aligned} \tag{C.10}$$

where (a) follows from inequality (C.8), and (b) follows from the definition in (C.9).

Now, we obtain the exponential behavior of ED^*_{pe}. Consider a sequence of normalized gamma distributed random variables $H_0 \sim \Upsilon(L, \theta)$ under the change of variables $A = -\frac{\log H_0}{\log \rho}$. The pdf for A is found as

$$p_A(\alpha) = \left|\frac{\partial H_0}{\partial \alpha}\right| p_{H_0}(h_0) = \rho^{-\alpha} p_{H_0}(\rho^{-\alpha}) \log \rho. \tag{C.11}$$

Then, $p_A(\alpha)$ is given by

$$p_A(\alpha) = \rho^{-\alpha} \frac{1}{\theta^L} \frac{1}{\Psi(L)} \rho^{-\alpha(L-1)} e^{-\frac{\rho^{-\alpha}}{\theta}} \log \rho = \frac{1}{\theta^L} \frac{1}{\Psi(L)} \rho^{-L\alpha} e^{-\frac{\rho^{-\alpha}}{\theta}} \log \rho,$$

and the exponential behavior is found as

$$S_A(\alpha) = -\lim_{\rho\to\infty} \frac{\log p_A(a)}{\log \rho} = \begin{cases} L\alpha & \text{if } \alpha \geq 0, \\ +\infty & \text{if } \alpha < 0. \end{cases} \tag{C.12}$$

For the model considered in Section 3.5, the SNR exponent for the Nakagami fading channel, $H_0 \sim \Upsilon(L_c, L_c^{-1})$, is given by $S_A(\alpha) = L_c\alpha$ for $\alpha \geq 0$, and for the Nakagami fading side information, $\Gamma_0 \sim \Upsilon(L_s, L_s^{-1})$, we have $S_B(\beta) = L_s\beta$ for $\beta \geq 0$.

Define $\kappa \triangleq \frac{\log \hat{\gamma}}{\log \rho}$, such that $\hat{\gamma} = \rho^\kappa$. Applying the change of variables to (C.9), in the high SNR regime, we have

$$\begin{aligned} ED_{pe}(\rho^\kappa) &= \int_{\mathcal{A}_{pe}} \frac{p_B(\beta)}{1+\rho^{1-\beta}} d\beta + \int_{\mathcal{A}_{pe}^c} \frac{p_B(\beta)}{(\rho^\kappa+1)(1+\rho)+\rho^{1-\beta}-\rho^\kappa} d\beta \\ &\doteq \int_{\mathcal{A}_{pe}} \rho^{-(1-\beta)^+} p_B(\beta) d\beta + \int_{\mathcal{A}_{pe}^c} \rho^{-(\kappa^+ +1)} p_B(\beta) d\beta \\ &\doteq \int_{\mathcal{A}_{pe}} \rho^{-[(1-\beta)^+ + S_B(\beta)]} d\beta + \int_{\mathcal{A}_{pe}^c} \rho^{-[\kappa^+ + 1 + S_B(\beta)]} d\beta, \end{aligned} \tag{C.13}$$

where we have defined

$$\mathcal{A}_{pe} \triangleq \{\beta : \hat{\gamma} \geq \rho^{1-\beta}\} = \{\beta : \kappa \geq 1-\beta\},$$

and we have used the fact that, in the high SNR asymptotic, and for $\beta \in \mathcal{A}_{pe}^c$, we have

$$\begin{aligned} [(\rho^\kappa+1)(1+\rho)+\rho^{1-\beta}-\rho^\kappa]^{-1} &\doteq [\rho^{\kappa^+}\rho^1 + \rho^{\max\{1-\beta,\kappa\}}]^{-1} \\ &\doteq \rho^{-\max\{\kappa^+ +1, (1-\beta)^+\}} \\ &= \rho^{-(\kappa^+ +1)}, \end{aligned}$$

which follows since $\rho^x + \rho^y \doteq \rho^{\max\{x,y\}}$ for $x, y \geq 0$, and we have $1-\beta > \kappa$ for $\beta \in \mathcal{A}_{pe}^c$. Similarly, in the high SNR limit we have $(1+\rho^{1-\beta})^{-1} \doteq \rho^{-(1-\beta)^+}$.

Since the exponents in the integral do not depend on ρ, the distortion exponent for each integral can be found by applying Varadhan's Lemma [94] separately for each integral term, similar to the proof of Theorem 4 in [73]. We define

$$\Delta_{p1}(\kappa) \triangleq \inf_{\mathcal{A}_{pe}} (1-\beta)^+ + S_B(\beta), \tag{C.14}$$

and

$$\Delta_{p2}(\kappa) \triangleq \inf_{\mathcal{A}_{pe}^c} \kappa^+ + 1 + S_B(\beta), \tag{C.15}$$

and write (C.10) as follows

$$ED^*_{pi} \geq \min_{\kappa\in\mathbb{R}}\{ED_{pe}(\rho^\kappa)\} \dot{\geq} \min_{\kappa\in\mathbb{R}}\{\rho^{-\Delta_{p1}(\kappa)} + \rho^{-\Delta_{p2}(\kappa)}\} \doteq \rho^{-\max_{\kappa\in\mathbb{R}}\min\{\Delta_{p1}(\kappa),\Delta_{p2}(\kappa)\}}.$$

Then, the distortion exponent is upper bounded by

$$\max_{\kappa\in\mathbb{R}}\min\{\Delta_{p1}(\kappa),\Delta_{p2}(\kappa)\}. \tag{C.16}$$

We solve the optimization problem in (C.16) with $S_B(\beta) = L_s\beta$, and denote the optimal value by $\Delta_{pe}(L_s, L_c)$. We note that we can restrict the domain of β in (C.14) and (C.15) to $\beta \geq 0$ without loss of optimality since $S_B(\beta) = +\infty$ for $\beta < 0$.

First, we consider the case $\kappa < 0$. In that case, $\Delta_{p1}(\kappa)$ is minimized by $\beta^* = 1-\kappa$ and we have $\Delta_{p1}(\kappa) = L_s(1-\kappa)$. On the other hand, we have

$$\begin{aligned}\Delta_{p2}(\kappa) = \inf_{\beta\geq 0}\ & 1 + L_s\beta \\ \text{s.t.}\ & \beta < 1-\kappa,\end{aligned} \tag{C.17}$$

which is minimized by $\beta^* = 0$, and $\Delta_{p2}(\kappa) = 1$. Then, from (C.16), we have $\Delta_{pe}(L_s, L_c) = \max_{\kappa<0}\min\{L_s(1-\kappa), 1\}$, which is maximized by $\kappa = -\infty$, and we have $\Delta_{pe}(L_s, L_c) = 1$.

Next, we consider the case $\kappa \geq 0$. Substituting $S_B(\beta) = L_s\beta$ in $\Delta_{p1}(\kappa)$ in (C.14), we note that we can constrain our search to $0 \leq \beta \leq 1$, since any $\beta > 1$ can only increase the objective function. We have,

$$\begin{aligned}\Delta_{p1}(\kappa) = \inf_{0\leq\beta\leq 1}\ & 1 + (L_s-1)\beta \\ \text{s.t.}\ & \beta \geq 1-\kappa.\end{aligned} \tag{C.18}$$

Since for $L_s > 1$, $1 + (L_s - 1)\beta$ is increasing in β, the minimum is achieved by $\beta^* = (1-\kappa)^+$ and $\Delta_{p1}(\kappa) = 1 + (L_s-1)(1-\kappa)^+$. On the contrary, for $L_s \leq 1$, the objective function is decreasing in β, and is minimized at $\beta^* = 1$, which yields $\Delta_{p1}(\kappa) = L_s$.

Similarly, for $\Delta_{p2}(\kappa)$ in (C.15), we have

$$\begin{aligned}\Delta_{p2}(\kappa) = \inf_{\beta\geq 0}\ & \kappa + 1 + L_s\beta \\ \text{s.t.}\ & \beta < 1-\kappa.\end{aligned} \tag{C.19}$$

This problem is minimized by $\beta^* = 0$, for which $\Delta_{p2}(\kappa) = 1+\kappa$, for $0 \leq \kappa < 1$, and has no solution for $\kappa \geq 1$, since there are no feasible β in the optimization set.

Then, substituting in (C.16), for $L_s \leq 1$, we have $\Delta_{pe}(L_s, L_c) = \max_{\kappa\geq 0}\min\{L_s, 1+\kappa\} = L_s$, and $\Delta_{pe}(L_s, L_c) = 1$. For $L_s > 1$, since $\Delta_{p1}(\kappa)$ is decreasing in κ while

$\Delta_{p2}(\kappa)$ is increasing in κ, the maximum $\Delta_{pe}(L_s, L_c)$ in (C.16) is achieved when the two exponents are equal, i.e., $1 + \kappa = 1 + (L_s - 1)(1 - \kappa)$, from which we find

$$\Delta_{pe}(L_s, L_c) = 2 - \frac{1}{L_s}, \quad \text{for } \kappa^* = \frac{L_s - 1}{L_s} \in (0, 1). \tag{C.20}$$

Now, we find the maximizing κ for each L_s regime to obtain $\Delta^*_{pe}(L_s, L_c)$. For $L_s \le 1$, the distortion exponent is maximized by $\kappa = -\infty$ and $\Delta_{pe}(L_s, L_c) = 1$, since $\Delta_{pe}(L_s, L_c) = L_s$ for any $\kappa \ge 0$. On the contrary, for $L_s \ge 1$, the distortion exponent is maximized as (C.20), while $\Delta_{pe}(L_s, L_c) = 1$ if we consider $\kappa < 0$.

Note that when $L_s \le 1$, the side information gain distribution is monotonically decreasing. Then $\bar{\gamma}(h) = 0$ for any h from Proposition 1, and therefore, from Theorem 3, uncoded transmission achieves the minimum expected distortion, i.e., $ED^*_{pi} = ED_u$. The distortion exponent for uncoded transmission $\Delta_u(L_s, L_c)$ is calculated in Appendix C as $\Delta_u(L_s, L_c) = \min\{1, L_s + L_c\}$. Comparing $\Delta_u(L_s, L_c)$ with $\Delta_{pe}(L_s, L_c)$, we observe that the proposed lower bound on ED^*_{pi} is in general not tight due to inequality (C.8).

C.3.2 Informed Encoder Upper Bound

Expressing the informed encoder lower bound ED_{inf} in (3.14) in terms of α and β, the distortion exponent is found by using Varadhan's Lemma as follows,

$$\begin{aligned}
ED_{\text{inf}} &= \iint_{h,\gamma} \frac{p_H(h)p_\Gamma(\gamma)}{(1+h)(1+\gamma)} dh d\gamma \\
&= \iint_{\alpha,\beta} \frac{p_A(\alpha)p_B(\beta)}{(1+\rho^{1-\alpha})(1+\rho^{1-\beta})} d\alpha d\beta \\
&\doteq \int_{\mathbb{R}^2} \frac{p_A(\alpha)p_B(\beta)}{\rho^{(1-\alpha)^+ + (1-\beta)^+}} d\alpha d\beta \\
&\doteq \rho^{-\Delta_{\text{inf}}(L_s)},
\end{aligned}$$

where the distortion exponent is found as the solution to the following optimization problem,

$$\Delta_{\text{inf}}(L_s, L_c) \triangleq \inf_{\mathbb{R}^2} (1-\alpha)^+ + (1-\beta)^+ + S_A(\alpha) + S_B(\beta). \tag{C.21}$$

We note that we can reduce the optimization domain to $\alpha, \beta \ge 0$ since $S_A(\alpha) = S_B(\beta) = +\infty$ for $\alpha, \beta < 0$. Evaluating for $S_A(\alpha) = L_c\alpha$ and $S_B(\beta) = L_s\beta$, the minimum in (C.21) is achieved by $\alpha^* = 1$ if $L_c < 1$ and $\alpha^* = 0$ if $L_c \ge 1$, and by $\beta^* = 1$ if $L_s < 1$, and $\beta^* = 0$ for $L_s \ge 1$. Then, the minimum is found to be given by $\Delta_{\text{inf}}(L_s, L_c) = \min\{1, L_c\} + \min\{1, L_s\}$.

C.4 Distortion Exponent Derivations

C.4.1 Uncoded transmission

Similarly to the proof in Appendix C.3.1 , applying the change of variables $H_0 = \rho^{-A}$ and $\Gamma_0 = \rho^{-B}$, and Varadhan's lemma, we have

$$ED_u = \iint_{\alpha,\beta} \frac{p_A(\alpha)p_B(\beta)}{1+\rho^{1-\alpha}+\rho^{1-\beta}} d\alpha d\beta \doteq \rho^{-\Delta_u(L_s)},$$

where the distortion exponent is found by substituting $S_A(\alpha) = L_c\alpha$ and $S_B(\beta) = L_s\beta$ as

$$\Delta_u(L, L_c) = \min_{0\leq\alpha,\beta\leq 1} \max\{1-\alpha, 1-\beta\} + L_c\alpha + L_s\beta.$$

Note that we can constraint to $0 \leq \alpha, \beta \leq 1$ without loss in optimality since any $\alpha, \beta > 1$ achieve a larger solution. Then, if $1-\alpha \geq 1-\beta$, the minimum is achieved by $\beta^* = \alpha^*$ and $\alpha^* = 0$ if $L_s + L_c \geq 1$, while $\alpha^* = 1$ if $L_s + L_c < 1$. Similarly occurs by symmetry when $1-\alpha < 1-\beta$. Then, the minimum is found to be given by $\Delta_u(L_s, L_c) = \min\{L_s+L_c, 1\}$.

C.4.2 Separate Source and Channel Coding (SSCC)

Here we find the distortion exponent of SSCC. Let us define the events

$$\mathcal{O}_1 \triangleq \{(h,\gamma) : R_c \geq I(U;V)\},$$
$$\mathcal{O}_2 \triangleq \{(h,\gamma) : R_c < I(U;V), R_c \leq I(X;W|Y)\}.$$

Event $\mathcal{O}_1$ corresponds to an outage due to bad quality of the channel, and $\mathcal{O}_2$ corresponds to a correct decoding of the channel codeword while an outage occurs due to the bad quality of the side information. It is readily seen that $\mathcal{O}_{sb} = \mathcal{O}_1 \bigcup \mathcal{O}_2$. Consider the change of variables $H_0 = \rho^{-A}$, $\Gamma_0 = \rho^{-B}$, $R_s = \frac{r_s}{2}\log\rho$ and $R_c = \frac{r_c}{2}\log\rho$, for $r_s \geq 0$ and $r_c > 0$. We consider $r_s = 0$ to allow SSCC to transmit without binning. We have

$$\begin{aligned}ED_{sb}(R_c, R_s) &= \int_{\mathcal{O}_{sb}^c} \frac{p_H(h)p_\Gamma(\gamma)}{2^{2(R_c+R_s-\epsilon)}+\gamma} dh d\gamma + \int_{\mathcal{O}_{sb}} \frac{p_H(h)p_\Gamma(\gamma)}{1+\gamma} dh d\gamma \\ &= \int_{\mathcal{A}_{sb}^c(\rho)} \frac{p_A(\alpha)p_B(\beta)}{\rho^{r_c+r_s}+\rho^{1-\beta}} d\alpha d\beta + \int_{\mathcal{A}_{sb}(\rho)} \frac{p_A(\alpha)p_B(\beta)}{1+\rho^{1-\beta}} d\alpha d\beta,\end{aligned}$$

where we have defined $\mathcal{A}_{sb}(\rho) \triangleq \mathcal{A}_1(\rho) \bigcup \mathcal{A}_2(\rho)$, and $\mathcal{A}_1(\rho)$ characterizes $\mathcal{O}_1$ in terms of α and β, and is given by

$$\mathcal{A}_1(\rho) \triangleq \left\{(h,\gamma) : R_c \geq \frac{1}{2}\log(1+h)\right\} = \{(\alpha,\beta) : \rho^{r_c} \geq 1+\rho^{1-\alpha}\},$$

and similarly for $\mathcal{O}_2$ we have

$$\mathcal{A}_2(\rho) \triangleq \left\{(h,\gamma) : R_c < \frac{1}{2}\log(1+h),\ R_c \leq \frac{1}{2}\log\left(1+\frac{2^{2(R_s+R_c-\epsilon)}-1}{1+\gamma}\right)\right\}$$
$$= \left\{(\alpha,\beta) : \rho^{r_c} < 1+\rho^{1-\alpha},\ \rho^{r_c} \leq 1+\frac{2^{-2\epsilon}\rho^{r_s+r_c}}{1+\rho^{(1-\beta)}}\right\}.$$

Using similar bounding techniques to the ones used in Appendix C, it is not hard to show that in the high SNR regime, we have

$$ED_{sb}(R_c,R_s) \doteq \int_{\mathcal{A}_1^c\cap\mathcal{A}_2^c} \frac{p_A(\alpha)p_B(\beta)}{\rho^{\max\{r_c+r_s,1-\beta\}}}d\alpha d\beta + \int_{\mathcal{A}_1\cup\mathcal{A}_2} \frac{p_A(\alpha)p_B(\beta)}{\rho^{(1-\beta)^+}}d\alpha d\beta,$$

where the equivalent outage sets in the high SNR are

$$\mathcal{A}_1 \triangleq \{(\alpha,\beta) : r_c \geq (1-\alpha)^+\},$$
$$\mathcal{A}_2 \triangleq \{(\alpha,\beta) : r_c < (1-\alpha)^+, r_c \leq (r_s+r_c-(1-\beta)^+)^+\}.$$

Let $\mathbf{r} \triangleq [r_c, r_s]$. Applying Varadhan's lemma, the distortion exponent of each integral term are found as

$$\Delta_{s1}(\mathbf{r}) = \inf_{\mathbb{R}^2} \max\{r_c+r_s, 1-\beta\} + S_A(\alpha) + S_B(\beta)$$
$$\text{s.t. } r_c < (1-\alpha)^+, \quad r_c > (r_s+r_c-(1-\beta)^+)^+,$$

and

$$\Delta_{s2}(\mathbf{r}) = \inf_{\mathbb{R}^2}(1-\beta)^+ + S_A(\alpha) + S_B(\beta) \tag{C.22}$$
$$\text{s.t. } r_c \geq (1-\alpha)^+,$$
$$\text{or } r_c < (1-\alpha)^+, \quad r_c \leq (r_s+r_c-(1-\beta)^+)^+.$$

We can limit the optimization to $0 \leq \alpha,\beta \leq 1$ without loss of optimality. First, we find the distortion exponent for $L_s \geq 1$. We start with $\Delta_{s1}(\mathbf{r})$. If $r_c + r_s \geq 1-\beta$, we have

$$\Delta_{s1}(\mathbf{r}) = \inf_{\alpha,\beta\geq 0} r_s + r_c + L_c\alpha + L_s\beta \tag{C.23}$$
$$\text{s.t. } \alpha < 1-r_c, \quad 1-(r_s+r_c) \leq \beta < 1-r_s.$$

The minimum is achieved by $\beta^* = (1-(r_s+r_c))^+$ and $\alpha^* = 0$ and we have $\Delta_{s1}(\mathbf{r}) = r_s + r_c + L_s(1-(r_s+r_c))^+$ for $r_c < 1$, $r_s < 1$. If $1-\beta > r_c + r_s$,

$$\Delta_{s1}(\mathbf{r}) = \inf_{\alpha,\beta\geq 0} 1 + L_c\alpha + (L_s-1)\beta \tag{C.24}$$
$$\text{s.t. } \alpha < 1-r_c, \quad \beta < 1-(r_s+r_c).$$

The minimum is achieved by $\alpha^* = \beta^* = 0$, and is found to be $\Delta_1(\mathbf{r}) = 1$ for $r_c < 1$ and $r_c + r_s < 1$. Then, putting all together, the infimum is given by $\Delta_{s1}(\mathbf{r}) = \max\{1, r_s + r_c\}$, for $r_s < 1$ and $r_c < 1$.

For $\Delta_{s2}(\mathbf{r})$, we first consider the case with constraint $r_c \geq (1-\alpha)^+$. The minimum is easily seen to be given by $\alpha^* = (1-r_c)^+$ and $\beta^* = 0$. Then $\Delta_{s2}(\mathbf{r}) = 1 + L_c(1-r_c)^+$. If $r_c \leq (1-\alpha)^+$, the second constraint is active. If $r_s + r_c < (1-\beta)^+$, $\Delta_{s2}(\mathbf{r})$ has no solution since this would require $r_c \leq 0$. If $r_s + r_c \geq (1-\beta)^+$, the minimum is achieved for $\alpha^* = 0$ and $\beta^* = (1-r_s)^+$, and is given by $\Delta_{s2}(\mathbf{r}) = 1 + (L_s - 1)(1-r_s)^+$ for $r_s > 0$ and $r_c < 1$.

The optimal distortion exponent of SSCC is found by maximizing over the rates as

$$\Delta_{sb}(L_s, L_c) = \max_{r_c, r_s \geq 0} \min\{\Delta_{s1}(\mathbf{r}), \Delta_{s2}(\mathbf{r})\}.$$

The distortion exponent is maximized when $r_s + r_c > 1$, $r_c < 1$ and $r_s < 1$. Then, we have $\Delta_{s1}(\mathbf{r}) = r_s + r_c$, $\Delta_{s2}(\mathbf{r}) = \min\{1 + L_c(1-r_c)^+, 1 + (L_s-1)(1-r_s)^+\}$. The maximum is achieved by r_c and r_s for which the left and right terms in the minimization in $\Delta_{s2}(\mathbf{r})$ are equal, i.e., $1 + L_c(1-r_c) = 1 + (L_s-1)(1-r_s)$, and $\Delta_{s1}(\mathbf{r}) = \Delta_{s2}(\mathbf{r})$. Solving this, we have

$$r_s^* = \frac{(L_c+1)(L_s-1)}{L_s(L_c+1)-1}, \qquad r_c^* = \frac{L_c L_s}{L_s(L_c+1)-1},$$

which satisfy $r_s < 1$, $r_c < 1$ and $r_s + r_c > 1$. Note that for $L_s = 1$, we have $r_s = 0$, i.e., no binning is optimal, as expected from Lemma 4.

Now we consider the case $L_s \leq 1$. In this regime, the gamma function is monotonically decreasing, and hence, $\bar{\gamma} = 0$ and from Lemma 4 we have $R_s^* = 0$, i.e., no binning achieves the minimum distortion for SSCC. Next, we derive the distortion exponent when no binning is considered, for general L_s to account for ED_{nb}^*.

Letting $R_s = 0$, the outage event $\mathcal{A}_2$ is empty. Then, we find the distortion exponent of $ED_{nb}(R_c)$ as

$$\begin{aligned}\Delta_{nb1}(r_c) = \inf_{\alpha,\beta \geq 0} & \max\{r_c, (1-\beta)^+\} + L_c\alpha + L_s\beta \\ \text{s.t. } & r_c < (1-\alpha)^+,\end{aligned}$$

and

$$\begin{aligned}\Delta_{nb2}(r_c) = \inf_{\alpha,\beta \geq 0} & (1-\beta)^+ + \alpha + L_s\beta \\ \text{s.t. } & r_c \geq (1-\alpha)^+.\end{aligned}$$

By solving the cases for $r_c < 1 - \beta$ and $r_c \geq 1 - \beta$, we find that $\Delta_{nb1}(r_c)$ attains its infimum at $\alpha^* = 0$ and $\beta^* = (1-r_c)^+$ as $\Delta_{nb1}(r_c) = 1 + (L_s - 1)(1-r_c)^+$ if $L_s < 1$. If $L_s \geq 1$, then the minimum is achieved by $\alpha^* = 0$ and $\beta^* = 0$, and is given by $\Delta_{nb1}(r_c) = 1$. On the other hand, $\Delta_{nb2}(r_c)$ is minimized by $\alpha^* = (1-r_c)^+$ and $\beta^* = 1$ when $L_s \leq 1$, and $\beta^* = 0$ when $L_s > 1$. Then we have $\Delta_{nb2}(r_c) = \min\{L_s, 1\} + L_c(1-r_c)^+$. The distortion exponent is found as $\Delta_{nb}(L_s, L_c) = \min_{r_c}\{\Delta_{nb1}(r_c), \Delta_{nb2}(r_c)\}$. The optimal distortion exponent without binning is achieved by setting

$$r_c = \frac{L_c}{1 - L_s + L_c} \text{ for } L_s \leq 1, \quad \text{and} \quad r_c = 1 \text{ for } L_s > 1.$$

C.4.3 Joint Decoding Scheme (JDS)

Here, we consider the distortion exponent for JDS. Applying the change of variables, $H_0 = \rho^{-A}$, $\Gamma_0 = \rho^{-B}$ and $R_{jd} = \frac{r_{jd}}{2}\log\rho$ for $r_h > 0$, form (4.11) we have

$$\begin{aligned} ED_j(R_{jd}) &= \int_{\mathcal{O}_j^c} \frac{p_H(h)p_\Gamma(\gamma)}{2^{2(R_{jd}-\epsilon)} + \gamma} dh d\gamma + \int_{\mathcal{O}_j} \frac{p_H(h)p_\Gamma(\gamma)}{1+\gamma} dh d\gamma \\ &\doteq \int_{\mathcal{A}_j^c} \frac{p_A(\alpha)p_B(\beta)}{\rho^{\max\{r_{jd},(1-\beta)^+\}}} d\alpha d\beta + \int_{\mathcal{A}_j} \frac{p_A(\alpha)p_B(\beta)}{\rho^{(1-\beta)^+}} d\alpha d\beta, \end{aligned}$$

where we define the outage event in the high SNR regime as

$$\mathcal{A}_j \triangleq \left\{(\alpha,\beta) : (r_{jd} - (1-\beta)^+)^+ \geq (1-\alpha)^+\right\}.$$

The distortion exponent for each term is found applying Varadhan's Lemma as

$$\Delta_{j1}(r_{jd}) = \inf_{\mathcal{A}_j^c} \max\{r_{jd}, (1-\beta)^+\} + S_A(\alpha) + S_B(\beta),$$

and

$$\Delta_{j2}(r_{jd}) = \inf_{\mathcal{A}_j} (1-\beta)^+ + S_A(\alpha) + S_B(\beta).$$

First we note that in both $\Delta_{j1}(r_{jd})$ and $\Delta_{j2}(r_{jd})$ we can restrict to $0 \leq \alpha, \beta \leq 1$ without loss of optimality since $S_A(\alpha) = L_c\alpha$ and $S_B(\beta) = L_s\beta$. Now we solve $\Delta_{j1}(r_{jd})$. If $r_{jd} < 1-\beta$, we have $\mathcal{A}_j = \{(\alpha,\beta) : (1-\alpha)^+ \geq 0, r_{jd} < 1-\beta\}$ and it is easily seen that $\alpha^* = 0$. Then if $L_s \geq 1$, we have $\beta^* = 0$ and $\Delta_{j1}(r_{jd}) = 1$ for $r_{jd} \leq 1$. If $L_s < 1$, then $\beta^* = (1-r_{jd})^+$ and $\Delta_{j1}(r_{jd}) = 1 + (L_s - 1)(1-r_{jd})^+$ for $r_{jd} \leq 1$. If $r_{jd} \geq 1-\beta$, we have

$$\begin{aligned} \Delta_{j1}(r_{jd}) = &\inf_{0\leq\alpha,\beta\leq 1} r_{jd} + L_c\alpha + L_s\beta \\ &\text{s.t. } \alpha + \beta < 2 - r_{jd}, \quad \beta \geq 1 - r_{jd}. \end{aligned} \tag{C.25}$$

The minimum is achieved by $\alpha^* = 0$ and $\beta^* = (1 - r_{jd})^+$ if $r_{jd} \leq 2$ and is given by $\Delta_{j1}(r_{jd}) = r_{jd} + L_s(1 - r_{jd})^+$ and has no feasible solutions if $r_{jd} \geq 2$. Then, the exponent $\Delta_{j1}(r_{jd})$ is given by the minimum of these solutions, given by

$$\Delta_{j1}(r_{jd}) = \begin{cases} 1 + (L_s - 1)^+(1 - r_{jd}) & \text{if } 0 \leq r_{jd} < 1, \\ r_{jd} & \text{if } 1 \leq r_{jd} < 2, \end{cases} \tag{C.26}$$

where we have used that for $L_s \leq 1$ and $0 \leq r_{jd} \leq 1$, we have $r_{jd} + L_s(1 - r_{jd})^+ = 1 + (1 - L_s)^+(1 - r_{jd})^+$, and for $L_s \geq 1$ and $0 \leq r_{jd} \leq 1$, we have $\min\{r_{jd} + L_s(1 - r_{jd})^+, 1\} = 1$.

Now, we solve $\Delta_{j2}(r_{jd})$. If $r_j < 1 - \beta$, the problem has no feasible solution due to the constraints. If $r_j \geq 1 - \beta$, we have

$$\begin{aligned} \Delta_{j2}(r_{jd}) = & \inf_{0 \leq \alpha, \beta \leq 1} 1 + (L_s - 1)\beta + L_c\alpha \\ & \text{s.t. } \alpha + \beta \geq 2 - r_{jd}, \quad \beta \geq 1 - r_{jd}. \end{aligned} \tag{C.27}$$

The minimum is achieved by $\alpha^* = (2 - r_{jd} - \beta)^+$, which satisfies $\alpha^* \leq 1$ due to $\beta \geq 1 - r_{jd}$. Then, if $\beta \geq 2 - r_{jd}$ and $L_s \geq 1$, we have $\beta^* = (2 - r_{jd})^+$ for $r_{jd} \geq 1$ and the minimum is given by $\Delta_{j2}(r_{jd}) = 1 + (L_s - 1)(2 - r_{jd})^+$. If $\beta \geq 2 - r_{jd}$ and $L_s < 1$ we have $\beta^* = 1$ and $\Delta_{j2}(r_{jd}) = L_s$ for $r_{jd} \geq 1$. If $\beta < 2 - r_{jd}$ and $L_s \geq 1 + L_c$, the minimum is achieved by $\beta^* = (1 - r_{jd})^+$ if $r_{jd} \leq 2$ and $\Delta_{j2}(r_{jd}) = 1 + (L_s - 1 - L_c)(1 - r_{jd})^+ + L_c(2 - r_{jd})$. If $L_s < 1 + L_c$, the solution is found as $\Delta_{j2}(r_{jd}) = L_s + L_c(1 - r_{jd})$ if $r_{jd} \leq 1$ for $\beta^* = 1$ and by $\Delta_{j2}(r_{jd}) = 1 + (L_s - 1)(2 - r_{jd})$ if $r_{jd} \geq 1$ for $\beta = (2 - r_{jd})^+ - \delta$, for arbitrarily small $\delta > 0$.

Finally, $\Delta_{j2}(r_{jd})$ is found by the minimum of these solutions in each regime. If $0 \leq r_{jd} \leq 1$, we have

$$\Delta_{j2}(r_{jd}) = \begin{cases} L_s + L_c(1 - r_{jd}) & \text{if } L_s < L_c + 1, \\ 1 + L_c + (L_s - 1)(1 - r_{jd}) & \text{if } L_s \geq L_c + 1. \end{cases} \tag{C.28}$$

If $1 \leq r_{jd} \leq 2$, we have

$$\Delta_{j2}(r_{jd}) = \begin{cases} L_s & \text{if } L_s < 1, \\ 1 + \min\{L_c, L_s - 1\}(2 - r_{jd})^+ & \text{if } L_s \geq 1, \end{cases} \tag{C.29}$$

where for the case $L_s < 1$ we have that $L_s \leq L_s + L_c(1 - r_{jd})$, and in the case $L_s \geq 1$, we have that $1 + L_c(2 - r_{jd}) \leq 1 + (L_s - 1)(2 - r_{jd})$ for $L_s \geq 1 + L_c$. Finally, for $r_{jd} \geq 2$ we have $\Delta_{j2}(r_{jd}) = \min\{1, L_s\}$.

The distortion exponent can be maximized over r_{jd}. If $L_s \leq 1$, the maximum is found by using a rate $0 \leq r_{jd} \leq 1$ and equating $\Delta_{j1}(r_{jd}) = 1 + (L_s - 1)(1 - r_{jd})$ and

$\Delta_{j2}(r_{jd}) = L_s + L_c(1 - r_{jd})$. The optimal rate is found as $r^*_{jd} = \frac{L_c}{1+L_c-L_s} \leq 1$. If $1 < L_s \leq L_c + 1$, the maximum distortion exponent is found with a rate $1 \leq r_{jd} \leq 2$ such that $\Delta_{j1}(r_{jd}) = r_{jd}$ and $\Delta_{j2}(r_{jd}) = 1 + (L_s - 1)(2 - r_{jd})$ are equal, given by $r^*_{jd} = 2 - \frac{1}{L_s}$. Finally, if $L_s > L_c + 1$, the distortion exponent is maximized when $1 \leq r_{jd} \leq 2$. By equaling $\Delta_{j1}(r_{jd}) = r_{jd}$ and $\Delta_{j2}(r_{jd}) = 1 + L_c(2 - r_{jd})$, the distortion exponent is maximized by $r^*_{jd} = 1 + \frac{L_c}{L_c+1}$.

C.4.4 Superposed Hybrid Digital-Analog Transmission (SHDA)

The performance of the SHDA scheme in Section 3.3.6 can be optimized over P_d, P_a and η^2. From the distortion exponent perspective, we have observed that it suffices to allocate all the power to the digital component, which reduces SHDA to HDA. Therefore, we let $P_d = 1$, $P_a = 0$. Applying the change of variables, we have from (3.30)-(3.32),

$$\begin{aligned}
ED_{shda}(1,\eta) &= E_{\mathcal{O}_h}[D_h^{out}(\eta,1)] + E_{\mathcal{O}_h^c}[D_h(\eta,1)] \\
&= \int_{\mathcal{O}_h} \frac{p_H(h)p_\Gamma(\gamma)}{1+\gamma} dh d\gamma + \int_{\mathcal{O}_h^c} \frac{p_H(h)p_\Gamma(\gamma)}{1+\gamma+\eta^2(1+h)} dh d\gamma \\
&= \int_{\mathcal{A}_h(\rho)} \frac{p_A(\alpha)p_B(\beta)}{1+\rho^{1-\beta}} d\alpha d\beta + \int_{\mathcal{A}_h^c(\rho)} \frac{p_A(\alpha)p_B(\beta)}{1+\rho^{1-\beta}+\eta^2(1+\rho^{1-\alpha})} d\alpha d\beta,
\end{aligned}$$

where $\mathcal{O}_h$ in (3.28) is found, in terms of α and β as

$$\mathcal{A}_h(\rho) \triangleq \left\{(\alpha,\beta) : \frac{\rho^{1-\alpha}}{1+\rho^{1-\alpha}}(1+\rho^{1-\beta}) \leq \eta^2\right\}.$$

In the high SNR regime, we let $\eta^2 = \rho^{r_h}$, for $r_h \in \mathbb{R}$,and the outage event $\mathcal{A}_h(\rho)$ is equivalent to

$$\mathcal{A}_h \triangleq \left\{(\alpha,\beta) : (1-\beta)^+ - (\alpha-1)^+ \leq r_h\right\}. \tag{C.30}$$

Then, we have

$$\begin{aligned}
&ED_{shda}(1,\rho^{r_h}) \qquad\qquad\qquad\qquad\qquad\qquad (C.31)\\
&\doteq \int_{\mathcal{A}_h} \rho^{-(1-\beta)^+} p_A(\alpha)p_B(\beta) d\alpha d\beta + \int_{\mathcal{A}_h^c} \rho^{-\max\{(1-\beta)^+,(1-\alpha)^+ + r_h\}} p_A(\alpha)p_B(\beta) d\alpha d\beta.
\end{aligned}$$

Using Varadhan's Lemma, the distortion exponent for the first integral in (C.31) is found as

$$\Delta_{h1}(r_h) \triangleq \inf_{\mathcal{A}_h} (1-\beta)^+ + S_A(\alpha) + S_B(\beta),$$

and for the second integral as

$$\Delta_{h2}(r_h) \triangleq \inf_{\mathcal{A}_h^c} \max\{(1-\beta)^+, (1-\alpha)^+ + r_h\} + S_A(\alpha) + S_B(\beta).$$

The distortion exponent for HDA can be optimized over the parameter r_h as

$$\Delta_{hda}(L_s, L_c) = \max_{r_h \in \mathbb{R}} \min\{\Delta_{h1}(r_h), \Delta_{h2}(r_h)\}. \tag{C.32}$$

First, we obtain the achievable distortion exponent when $r_h < 0$. To solve $\Delta_{h1}(r_h)$, note that if $0 \leq \alpha \leq 1$, there are no feasible solutions. Then, for $\alpha > 1$, we have

$$\begin{aligned} \Delta_{h1}(r_h) \triangleq & \inf_{\alpha > 1, \beta \geq 0} (1-\beta)^+ + L_c\alpha + L_s\beta \\ & \text{s.t. } \alpha \geq (1-\beta)^+ + 1 - r_h. \end{aligned} \tag{C.33}$$

We can constrain the optimization to $0 \leq \beta \leq 1$ without loss of optimality, and the minimum is achieved by $\alpha^* = 2 - \beta - r_h$. If $L_s \geq 1 + L_c$, the minimum is achieved by $\beta^* = 0$, and is given by $\Delta_{h1}(r_h) = 1 + L_c(2 - r_h)$. On the other hand, if $L_s < 1 + L_c$, $\beta^* = 1$, and $\Delta_{h1}(r_h) = L_s + L_c(1 - r_h)$. Putting all together, we have $\Delta_{h1}(r_h) = \min\{L_s, 1 + L_c\} + L_c(1 - r_h)$.

Now, we solve $\Delta_{h2}(r_h)$. Without loss of optimality, we can assume $0 \leq \alpha, \beta \leq 1$, as otherwise the feasible grows and $\alpha > 1$ or $\beta > 1$ can only increase the objective function. Then, the constraint is always satisfied, since $1 - \beta \geq r_h$ for any $0 \leq \beta \leq 1$. We have

$$\Delta_{h2}(r_h) = \max_{0 \leq \alpha, \beta \leq 1} \{1 - \beta, 1 - \alpha + r_h\} + L_s\beta + L_c\alpha. \tag{C.34}$$

If $1 - \beta \geq 1 - \alpha + r_h$, the minimum is achieved by $\alpha^* = \beta^* = 0$ when $L_s \geq 1$ and $\Delta_{h2}(r_h) = 1$. If $L_s < 1$, $\beta^* = \alpha - r_h$ if $\alpha - r_h \leq 1$, and $\alpha^* = 0$ when $L_s + L_c \geq 1$ and we have $\Delta_{h2}(r_h) = 1 - (L_s - 1)r_h$. When $L_s + L_c < 1$, we have $\alpha^* = 1 + r_h$ and $\Delta_{h2}(r_h) = L_s + L_c(1 + r_h)$, $-1 \leq r_h < 0$ and, when $\alpha > 1 + r_h$, we have $\beta^* = 1$ and $\Delta_{h2}(r_h) = L_s + L_c(1 + r_h)^+$. If $1 - \beta < 1 - \alpha + r_h$, we have $\beta^* = \alpha + \delta$, which has to satisfy $\beta^* \leq 1$, i.e., it is feasible whenever $\alpha \leq 1 + r_h$. Then, $\alpha^* = 0$ if $L_s + L_c \geq 1$ and the minimum is given by $\Delta_{h2}(r_h) = 1 - r_h(L_s - 1)$. If $L_s + L_c < 1$, we have $\alpha^* = 1 + r_h$ and $\Delta_{h2}(r_h) = L_s + L_c(1 + r_h)$, for $r_h \geq -1$. Putting all together, we have $\Delta_{h2}(r_h) = 1$ when $L_s \geq 1$ and $\Delta_{h2}(r_h) = \min\{1 - (L_s - 1)r_h, L_s + L_c(1 + r_h)\}$ for $L_s < 1$.

If $L_s \leq 1$, we have $\Delta_{h1}(r_h) \geq \Delta_{h2}(r_h)$, and the distortion exponent is maximized by letting $r_h \to 0$ and we get $\Delta_{hda}(L_s, L_c) = \min\{L_s + L_c, 1\}$. If $L_s \geq 1$, we have $\Delta_{hda}(L_s, L_c) = 1$ for any $r_h < 0$.

In the following, we derive the distortion exponent achievable by SHDA when $r_h \geq 0$. First, we solve $\Delta_{h1}(r_h)$. We can limit the optimization to $0 \leq \beta \leq 1$ without loss of optimality. Then, for $0 \leq \alpha \leq 1$ the minimum is achieved by $\alpha^* = 0$, and if $L_s \geq 1$, the minimum is achieved by $\beta^* = (1 - r_h)^+$ and $\Delta_{h1}(r_h) = 1 + (L_s - 1)(1 - r_h)^+$, and if $L_s < 1$, $\beta^* = 1$ and $\Delta_{h1}(r_h) = L_s$. If $\alpha > 1$, the constraint becomes $\alpha \geq 2 - \beta - r_h$, and the minimizing α is given by $\alpha^* = 2 - \beta - r_h$, which is feasible provided that $\beta < 1 - r_h$. Then, we have

$$\begin{aligned} \Delta_{h1}(r_h) = \inf_{0 \leq \beta \leq 1} \; & 1 + (L_s - 1 - L_c)\beta + L_c(2 - r_h) \\ \text{s.t.} \; & \beta < 1 - r_h. \end{aligned} \tag{C.35}$$

If $L_s \geq 1 + L_c$, we have $\beta^* = 0$ and $\Delta_{h1}(r_h) = 1 + L_c(2 - r_h)$ for $r_h \leq 1$, and if $L_s < 1 + L_c$, we have $\beta^* = 1 - r_h$ and $\Delta_{h1}(r_h) = 1 + L_c + (L_s - 1)(1 - r_h)$. Putting all together, $\Delta_{h1}(r_h)$ is found as

$$\Delta_{h1}(r_h) = \begin{cases} L_s & \text{if } L_s < 1, \\ 1 + (L_s - 1)(1 - r_h)^+ & \text{if } L_s \geq 1. \end{cases} \tag{C.36}$$

Next, we solve $\Delta_{h2}(r_h)$. First, we note that we can constrain to $0 \leq \beta \leq 1$, since the optimization set is empty if $\beta > 1$. Similarly, we assume $0 \leq \alpha \leq 1$, since any $\alpha > 1$ achieves a larger exponent. Then,

$$\begin{aligned} \Delta_{h2}(r_h) = \inf_{0 \leq \alpha, \beta \leq 1} \; & \max\{1 - \beta, 1 - \alpha + r_h\} + L_s\beta + L_c\alpha \\ \text{s.t.} \; & \beta < 1 - r_h. \end{aligned} \tag{C.37}$$

If $1 - \beta > 1 - \alpha + r_h$, we have $\alpha^* = \beta + r_h$, which satisfies $\alpha^* \leq 1$ since $\beta < 1 - r_h$. Then, $\beta^* = 0$ if $L_s + L_c \geq 1$ and $\Delta_{h2}(r_h) = 1 + L_c r_h$, and if $L_s + L_c < 1$, $\beta^* = 1 - r_h - \epsilon$ for an arbitrarily $\epsilon > 0$ and the infimum is found as $\Delta_{h2}(r_h) = 1 + L_c + (L_s - 1)(1 - r_h)$ for $r_h < 1$. If $1 - \beta \leq 1 - \alpha + r_h$, the infimum is given by $\beta^* = (\alpha - r_h)^+$. If $\alpha \geq r$ and $L_s + L_c \geq 1$, the minimum is found as $\alpha^* = r_h$ and $\Delta_{h2}(r_h) = 1 + r_h L_c$, while $\alpha^* = 1$ if $L_s + L_c < 1$, and $\Delta_{h2}(r_h) = 1 + L_c + (L_s - 1)(1 - r_h)$. If $\alpha < r_h$, we have $\alpha^* = 0$ if $L_c \geq 1$ and $\Delta_{h2}(r_h) = 1 + r_h$ and if $L_c < 1$, we have $\alpha^* = r_h + \epsilon$ for an arbitrarily small $\epsilon > 0$ and $\Delta_{h2}(r_h) = 1 + r_h L_c$. Putting all together, we have $\Delta_{h2}(r_h) = 1 + \min\{1, L_c\} r_h$ for $r_h \leq 1$.

We optimize over r_h to solve (C.32). For $L_s \leq 1$, we have $\Delta_{h1}(r_h) < \Delta_{h2}(r_h)$ for any $r_h \geq 0$ and $\Delta_{hda}(L_s, L_c) = L$. Then, the achievable distortion exponent is maximized, by using $r_h < 0$ and $r_h \to 0$, for which we obtain $\Delta_{hda}(L_s, L_c) = \min\{L_s + L_c, 1\}$. On the contrary, when $L_s \geq 1$, the distortion exponent is maximized for an $r_h > 0$ such that $\Delta_{h1}(r_h) = \Delta_{h2}(r_h)$, i.e.,

$$r_h^* = \frac{(L_s - 1)}{L_s - 1 + \min\{1, L_c\}}. \tag{C.38}$$

Putting all together we obtain the achievable distortion exponent in (3.44).

Appendix D

Proofs for Chapter 4

D.1 Proof of Theorem 6

The exponential integral can be bounded as follows [95, p.229, 5.1.20]:

$$\frac{1}{2}\ln\left(1+\frac{2}{t}\right) < e^t E_1(t) < \ln\left(1+\frac{1}{t}\right), \quad t>0. \tag{D.1}$$

Then, ED^*_{pi} in (4.6) is lower bounded by

$$ED^*_{\mathrm{pi}} \geq \int_{\mathbf{H}} \frac{1}{2\rho_s}\ln\left(1+\frac{2\rho_s}{2^{\mathcal{C}(\mathbf{H})}}\right)p_h(\mathbf{H})d\mathbf{H}. \tag{D.2}$$

Consider the change of variables $\lambda_i = \rho^{-\alpha_i}$, with $\alpha_1 \geq \ldots \geq \alpha_{M_*} \geq 0$. The joint probability density function (pdf) of $\boldsymbol{\alpha} \triangleq [\alpha_1, \ldots, \alpha_{M_*}]$ is given by [73]:

$$p_A(\boldsymbol{\alpha}) = K^{-1}_{M_t,M_r}(\log\rho)^{M_*}\prod_{i=1}^{M_*}\rho^{-(M^*-M_*+1)\alpha_i}\left[\prod_{i<j}(\rho^{\alpha_i}-\rho^{\alpha_j})^2\right]\exp\left(-\sum_{i=1}^{M_*}\rho^{\alpha_i}\right) \tag{D.3}$$

where $K^{-1}_{M_t,M_r}$ is a normalizing constant.

We define the high SNR exponent of $p_A(\boldsymbol{\alpha})$ as $S_A(\boldsymbol{\alpha})$, that is, we have $p_A(\boldsymbol{\alpha}) \doteq \rho^{-S_A(\boldsymbol{\alpha})}$, where

$$S_A(\boldsymbol{\alpha}) \triangleq \begin{cases} \sum_{i=1}^{M_*}(2i-1+M^*-M_*)\alpha_i & \text{if } \alpha_{M_*} \geq 0, \\ \infty & \text{otherwise.} \end{cases} \tag{D.4}$$

Following [73], the capacity of the MIMO channel is upper bounded as

$$\mathcal{C}(\mathbf{H}) = \sup_{\mathbf{C}_x:\mathrm{Tr}\{\mathbf{C}_x\}\leq M_t} \log\det\left(\mathbf{I}+\frac{\rho}{M_t}\mathbf{H}\mathbf{C}_x\mathbf{H}^H\right) \leq \log\det\left(\mathbf{I}+\rho\mathbf{H}\mathbf{H}^H\right),$$

where the inequality follows from the fact that $M_t\mathbf{I} - \mathbf{C}_x \succeq 0$ subject to the power constraint $\mathrm{Tr}\{\mathbf{C}_x\} \leq M_t$, and the function $\log\det(\cdot)$ is nondecreasing on the cone of positive semidefinite Hermitian matrices. Then, from (D.2) we have

$$\begin{aligned} ED^*_{\text{pi}} &\geq \int_{\mathbf{H}} \frac{1}{2\rho_s} \ln\left(1 + \frac{2\rho_s}{\prod_{i=1}^{M_*}(1+\rho\lambda_i)^b}\right) p_h(\mathbf{H})d\mathbf{H} \\ &= \int_{\boldsymbol{\alpha}} \frac{1}{2\rho_s} \ln\left(1+f(\boldsymbol{\alpha})\right) p_A(\boldsymbol{\alpha})d\boldsymbol{\alpha} \\ &\geq \int_{\boldsymbol{\alpha}^+} \frac{1}{2\rho_s} \ln\left(1+f(\boldsymbol{\alpha})\right) p_A(\boldsymbol{\alpha})d\boldsymbol{\alpha} \qquad \text{(D.5)} \\ &\geq \int_{\boldsymbol{\alpha}^+} \frac{1}{2\rho_s} \frac{f(\boldsymbol{\alpha})}{1+f(\boldsymbol{\alpha})} p_A(\boldsymbol{\alpha})d\boldsymbol{\alpha}, \qquad \text{(D.6)} \end{aligned}$$

where we define $f(\boldsymbol{\alpha}) \triangleq 2\rho_s \prod_{i=1}^{M_*}\left(1+\rho^{1-\alpha_i}\right)^{-b}$ and

$$G_\rho(\boldsymbol{\alpha}) \triangleq \frac{1}{2\rho_s}\frac{f(\boldsymbol{\alpha})}{1+f(\boldsymbol{\alpha})},$$

and the set $\boldsymbol{\alpha}^+ \triangleq \{\boldsymbol{\alpha} \in \mathbb{R}^{M_*} : 1 \geq \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0\}$ in (D.5). Inequality (D.6) follows from the lower bound $\ln(1+t) \geq \frac{t}{1+t}$, for $t > -1$.

Then, in the high SNR regime we have,

$$\begin{aligned} G(\boldsymbol{\alpha}) &\triangleq \lim_{\rho\to\infty} \frac{\log G_\rho(\boldsymbol{\alpha})}{\log\rho} = \lim_{\rho\to\infty} \frac{\log \rho^{-x}\frac{\rho^{x-b\sum_{i=1}^{M_*}(1-\alpha_i)^+}}{1+\rho^{x-b\sum_{i=1}^{M_*}(1-\alpha_i)^+}}}{\log\rho} \\ &= \begin{cases} -x & \text{if } x > b\sum_{i=1}^{M_*}(1-\alpha_i)^+, \\ -b\sum_{i=1}^{M_*}(1-\alpha_i)^+ & \text{if } x \leq b\sum_{i=1}^{M_*}(1-\alpha_i)^+, \end{cases} \end{aligned}$$

where we have used the exponential equalities $1+\rho^{1-\alpha_i} \doteq \rho^{(1-\alpha_i)^+}$, and $\rho_s \doteq \rho^x$.

Therefore, for sufficiently large ρ, we have

$$\begin{aligned} ED^*_{\text{pi}} &\geq \int_{\boldsymbol{\alpha}^+} \exp\left(\frac{\log G_\rho(\boldsymbol{\alpha})}{\log\rho}\log\rho\right) p_A(\boldsymbol{\alpha})d\boldsymbol{\alpha} \\ &\doteq \int_{\boldsymbol{\alpha}^+} \exp\left(G(\boldsymbol{\alpha})\log\rho\right) p_A(\boldsymbol{\alpha})d\boldsymbol{\alpha}. \end{aligned}$$

Defining $\Delta^*_{pi}(b,x) = -\lim_{\rho\to\infty}\frac{\log ED^*_{\text{pi}}}{\log\rho}$, the distortion exponent of the partially informed encoder is upper bounded by

$$\Delta^*_{pi}(b,x) \leq \lim_{\rho\to\infty}\frac{1}{\log\rho}\log\int_{\boldsymbol{\alpha}^+}\exp\left(G(\boldsymbol{\alpha})\log\rho\right)p_A(\boldsymbol{\alpha})d\boldsymbol{\alpha}.$$

From Varadhan's lemma [94], it follows that the distortion exponent of ED^*_{pi} is upper bounded by the solution to the following optimization problem,

$$\Delta_{up}(b,x) \triangleq \inf_{\boldsymbol{\alpha}^+}[-G(\boldsymbol{\alpha}) + S_A(\boldsymbol{\alpha})]. \tag{D.7}$$

In order to solve (D.7) we divide the optimization into two subproblems: the case when $x \leq b\sum_{i=1}^{M_*}(1-\alpha_i)$, and the case when $x > b\sum_{i=1}^{M_*}(1-\alpha_i)$. The solution is then given by the minimum of the solutions of these subproblems.

If $x \geq b\sum_{i=1}^{M_*}(1-\alpha_i)$, the problem in (D.7) reduces to

$$\Delta_{up}^1(b,x) = x + \inf_{\boldsymbol{\alpha}^+}\sum_{i=1}^{M_*}(2i-1+M^*-M_*)\alpha_i \qquad \text{s.t. } \sum_{i=1}^{M_*}(1-\alpha_i) \leq \frac{x}{b}. \tag{D.8}$$

The part inside the optimization in (D.8) can be identified with the DMT problem in (4.1) for a multiplexing gain of $r = \frac{x}{b}$. Next, we give an explicit solution for completeness.

First, if $bM_* \leq x$, the infimum is given by $\Delta_{up}^1(b,x) = x$ for $\boldsymbol{\alpha}^* = 0$. Then, for $k \leq \frac{x}{b} \leq k+1$, for $k = 0, ..., M_*-1$, i.e., $\frac{x}{k+1} \leq b \leq \frac{x}{k}$, the infimum is achieved by

$$\alpha_i^* = \begin{cases} 1 & \text{for } i = 1, ..., M_*-k-1, \\ k+1-\frac{x}{b} & i = M_*-k, \\ 0 & \text{for } i = M_*-k+1, ..., M_*. \end{cases}$$

Substituting, we have, for $k = 0, ..., M_*-1$,

$$\Delta_{up}^1(b,x) = x + \Phi_k - \Upsilon_k\left(\frac{x}{b} - k\right) = x + d^*\left(\frac{x}{b}\right),$$

where Φ_k and Υ_k are defined as in (4.3).

Now we solve the second subproblem with $x < b\sum_{i=1}^{M_*}(1-\alpha_i)$. Since $1 \geq \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0$ we can write (D.7) as

$$\Delta_{up}^2(b,x) = \inf_{\boldsymbol{\alpha}^+} bM_* - \sum_{i=1}^{M_*}\alpha_i\phi(i) \qquad \text{s.t. } \sum_{i=1}^{M_*}\alpha_i < M_* - \frac{x}{b}, \tag{D.9}$$

where we define $\phi(i) \triangleq [b - (2i-1+M^*-M_*)]$. Note that $\phi(1) > \cdots > \phi(M_*)$.

First, we note that for $bM_* < x$ there are no feasible solutions due to the constraint in (D.9).

Now, we consider the case $x \leq M_*(1+M^*-M_*)$. If $\frac{x}{M_*} \leq b < 1+M^*-M_*$, all the terms $\phi(i)$ multiplying α_i's are negative, and, thus, the infimum is achieved by $\boldsymbol{\alpha}^* = 0$, and is given by $\Delta_{up}^2(b,x) = bM_*$. If $1+M^*-M_* \leq b < 3+M^*-M_*$, then $\phi(1)$ multiplying α_1 is positive, while the other $\phi(i)$ terms are negative. Then $\alpha_i^* = 0$

for $i = 2, ..., M_*$. From (D.9) we have $\alpha_1 \leq M_* - \frac{x}{b}$. If $b \geq \frac{x}{M_*-1}$, the right hand side (r.h.s.) of (D.9) is greater than one, and smaller otherwise. Then, we have

$$\alpha_1^* = \begin{cases} 1 & \text{if } b \geq \frac{x}{M_*-1}, \\ M_* - \frac{x}{b} & \text{if } b < \frac{x}{M_*-1}. \end{cases}$$

Note that $\alpha_1^* \geq 0$ since $b > \frac{x}{M_*}$.

When $2k - 1 + M^* - M_* \leq b < 2k + 1 + M^* - M_*$ for $k = 2, ..., M_* - 1$, the coefficients $\phi(i)$, $i = 1, ..., k$, associated with the first k α_i terms are positive, while the others remain negative. Then,

$$\alpha_i^* = 0, \quad \text{for } i = k+1, ..., M_*. \tag{D.10}$$

Since $\phi(i)$, $i = 1, ..., k$, are positive and $\phi(1) > \cdots > \phi(k)$, we have $\alpha_i^* = 1$ for $i = 1, ..., k-1$, and the constraint becomes $\alpha_k < M_* - (k-1) - \frac{x}{b}$. If $b \geq \frac{x}{M_*-k}$, then the r.h.s. is greater than one, and smaller otherwise. In order for the solution to be feasible, we need $M_* - (k-1) - \frac{x}{b} \geq 0$. Then we have

$$\alpha_k^* = \begin{cases} 1 & \text{if } b \geq \frac{x}{M_*-k}, \\ M_* - (k-1) - \frac{x}{b} & \text{if } \frac{x}{M_*-(k-1)} \leq b < \frac{x}{M_*-k}. \end{cases} \tag{D.11}$$

If $b < \frac{x}{M_*-(k-1)}$, the solution in (D.11) is not feasible. Instead, we have $\alpha_k^* = 0$, since $\phi(k) < \phi(k-1)$, $\alpha_i^* = 0$ for $i = k+1, ..., M_*$, and $\alpha_i^* = 1$, for $i = 1, ..., k-2$. Then, the constraint in (D.9) is given by $\alpha_{k-1} \leq M_* - (k-2) - \frac{x}{b}$. Since $b < \frac{x}{M_*-(k-1)}$, the r.h.s. is always smaller than one, and for the existence of a feasible solution, it is required to be greater than zero. Then, we have

$$\alpha_{k-1}^* = M_* - (k-2) - \frac{x}{b}, \qquad \text{if } \frac{x}{M_* - (k-2)} \leq b < \frac{x}{M_* - (k-1)}.$$

In general, iterating this procedure, for

$$\frac{x}{M_* - (j-1)} \leq b < \frac{x}{M_* - j}, \quad j = 1, ..., k,$$

we have

$$\alpha_i = \begin{cases} 1 & \text{for } i = 1, ..., j-1, \\ M_* - (j-1) - \frac{x}{b} & \text{for } i = j, \\ 0 & \text{for } i = j+1, ..., M_*. \end{cases} \tag{D.12}$$

Note that for the case $j = 1$, we have $\alpha_1 = M_* - \frac{x}{b}$, which is always feasible.

We now evaluate (D.9) with the optimal $\boldsymbol{\alpha}^*$ for $2k - 1 + M^* - M_* \leq b < 2k + 1 + M^* - M_*$, $k = 2, ..., M_* - 1$. For $b \geq \frac{x}{M_* - k}$, we have $\alpha_1 = \cdots = \alpha_k = 1$ and $\alpha_{k+1} = \cdots = \alpha_{M_*} = 0$ and then

$$\Delta^2_{up}(b, x) = \sum_{i=1}^{M_*} \min\{b, 2i - 1 + M^* - M_*\} = \Delta_{\text{MIMO}}(b).$$

For $\frac{x}{M_*} \leq b \leq \frac{x}{M_* - k}$, substituting (D.12) into (D.9) we have

$$\Delta^2_{up}(b, x) = x + (M^* - M_* - 1 + j)(j - 1) + \left(M_* - (j - 1) - \frac{x}{b}\right)(2j - 1 + M^* - M_*),$$

for

$$\frac{x}{M_* - (j - 1)} \leq b \leq \frac{x}{M_* - j}, \quad j = 1, ..., k.$$

Note that with the change of index $j = M_* - j'$, we have, after some manipulation,

$$\Delta^2_{up}(b, x) = x + (M^* - j')(M_* - j') - \left(\frac{x}{b} - j'\right)(M^* + M_* - 2j' - 1),$$

in the regime

$$\frac{x}{j' + 1} \leq b < \frac{x}{j'}, \quad j' = M_* - k, ..., M_* - 1.$$

This is equivalent to the value of the DMT curve in (4.2) at multiplexing gain $r = \frac{x}{b}$. Then, for $\frac{x}{M_*} \leq b < \frac{x}{M_* - k}$ we have

$$\Delta^2_{up}(b, x) = x + d^*\left(\frac{x}{b}\right).$$

If $b \geq M^* + M_* - 1$, the infimum is achieved by $\alpha_i^* = 1$, for $i = 1, ..., M_* - 1$, and $\alpha^*_{M_*} = 1 - \frac{x}{b}$ if $b \geq x$. If $b < x$, this solution is not feasible, and the solution is given by (D.12). Therefore, in this regime we also have

$$\Delta^2_{up}(b, x) = x + d^*\left(\frac{x}{b}\right).$$

Putting all these results together, for $x \leq M_*(M^* - M_* + 1)$ and $2k - 1 + M^* - M_* \leq b < 2k + 1 + M^* - M_*$, for $k = 1, ..., M_* - 1$, we have

$$\Delta^2_{up}(x, b) = \begin{cases} bM_* & \text{for } \frac{x}{M_*} \leq b < M^* - M_* + 1, \\ x + d^*\left(\frac{x}{b}\right) & \text{for } M^* - M_* + 1 \leq b < \frac{x}{M_* - k}, \\ \Delta_{\text{MIMO}}(b) & \text{for } \frac{x}{M_* - k} \leq b < M^* + M_* - 1, \\ x + d^*\left(\frac{x}{b}\right) & \text{for } b \geq M^* + M_* - 1. \end{cases}$$

Now, we solve (D.9) for $M_*(M^* - M_* + 1) \leq x < M_*(M^* + M_* - 1)$. Let $2(l-1) - 1 + M^* - M_* \leq \frac{x}{M_*} < 2l - 1 + M^* - M_*$, for some $l = 2, ..., M_*$. The first interval of b in which a feasible solution exists is given by $\frac{x}{M_*} \leq b < 2l - 1 + M^* - M_* + 1$. From the sign of the coefficients $\phi(i)$ in this interval we have $\alpha_i^* = 0$ for $i = (l+1), ..., M_*$, and $\alpha_i^* = 1$ for $i = 1, ..., l-1$. Substituting, the constraint becomes $\alpha_l < M_* - (l-1) - \frac{x}{b}$. If $b > \frac{x}{M_* - l}$ the r.h.s. is larger than one, and $\alpha_l^* = 1$. On the contrary, if $b \leq \frac{x}{M_* - l}$, it is given by $\alpha_l^* = M_* - (l-1) - \frac{x}{b}$ if $b > \frac{x}{M_* - (l-1)}$, so that the r.h.s. of the constraint is larger than zero, and the solution is found following the techniques that lead to (D.12).

The problem is now solved as for the case $x \leq M_*(M^* - M_* + 1)$ in each interval $2k - 1 + M^* - M_* \leq b < 2(k+1) - 1 + M^* - M_*$ with $k = l, ..., M_* - 1$ instead of $k = 1, ..., M_* - 1$ and, thus, we omit the explicit resolution.

Putting all together, if x satisfies $2(l-1) - 1 + M^* - M_* \leq \frac{x}{M_*} < 2l - 1 + M^* - M_*$, $l = 2, ..., M_*$, for $2k - 1 + M^* - M_* \leq b < 2k + 1 + M^* - M_*$, for $k = l, ..., M_* - 1$, we have

$$\Delta_{up}^2(x, b) = \begin{cases} x + d^* \left(\frac{x}{b}\right) & \text{for } \frac{x}{M_*} \leq b < \frac{x}{M_* - k}, \\ \Delta_{\text{MIMO}}(b) & \text{for } \frac{x}{M_* - k} \leq b < M^* + M_* - 1, \\ x + d^* \left(\frac{x}{b}\right) & \text{for } b \geq M^* + M_* - 1. \end{cases}$$

Finally, the case $x \geq M_*(M^* + M_* - 1)$ can be solved as before. Notice that if $\alpha_i^* = 1$, $i = 1, ..., M_* - 1$ we have the constraint $\alpha_{M_*} \leq 1 - \frac{x}{b}$, that is, we never have the case $\alpha_{M_*}^* = 1$. Then, the optimal α_i^* are given as in (D.12), and we have

$$\Delta_{up}^2(x, b) = x + d^* \left(\frac{x}{b}\right) \quad \text{for } \frac{x}{M_*} \leq b.$$

Now, $\Delta_{up}(b, x)$ is given by the minimum of $\Delta_{up}^1(b, x)$ and $\Delta_{up}^2(b, x)$. First, we note that $\Delta_{2up}(b, x)$ has no feasible solution for $bM_* \leq x$, and we have $\Delta_{up}(b, x) = \Delta_{up}^1(b, x) = x$ in this region. For $bM_* > x$, both solutions $\Delta_{up}^1(b, x)$ and $\Delta_{up}^2(b, x)$ coincide except in the range $\frac{x}{M_* - l} \leq b \leq M^* + M_* - 1$. We note that $\Delta_{up}^1(b, x)$ in (D.8) is linear and increasing in $\boldsymbol{\alpha}$, and hence, the solution is such that the constraint is satisfied with equality, i.e., $x = \sum_{i=1}^{M_*} b(1 - \alpha_i)$. That is, $\Delta_{up}^2(b, x) \leq \Delta_{up}^1(b, x)$ whenever both solutions exist in the same $\boldsymbol{\alpha}$ region. Then, the minimizing $\boldsymbol{\alpha}$ will be one such that either $\Delta_{up}^1(b, x) < \Delta_{up}^2(b, x)$, or the one arbitrarily close to the boundary $x = b \sum_{i=1}^{M_*} (1 - \alpha_i)^+$, where $\Delta_{up}^1(b, x) = \Delta_{up}^2(b, x)$. Consequently, $\min\{\Delta_{up}^1(b, x), \Delta_{up}^2(b, x)\} = \Delta_{up}^1(b, x)$, whenever they are defined in the same region. Putting all the results together we complete the proof.

D.2 Proof of Theorem 7

To derive the distortion exponent of SSCC we first study the exponential behavior of $E_\Gamma[D_d(R,\Gamma)]$ in (4.9). We consider the change of variables $\gamma = \rho^{-\beta}$, with pdf $p_B(\beta)$ given as in (D.3) and $S_B(\beta) = \beta$, for $\beta \geq 0$, and $R = r\log\rho$. Then,

$$E_\Gamma[D_d(R,\Gamma)] = \int \frac{1}{\gamma\rho_s + 2^{2R}} p_\Gamma(\gamma)d\gamma = \int \exp(\log(\rho^{x-\beta} + \rho^{2r})^{-1})p_B(\beta)d\beta.$$

In the high SNR regime, we have

$$E_\Gamma[D_d(r\log\rho,\Gamma)] \doteq \int_{\mathbb{R}} \rho^{-\max\{(x-\beta)^+,2r\}} p_B(\beta)d\beta,$$

where we have used $(1+\rho^{x-\beta}+\rho^{2r})^{-1} \doteq \rho^{-\max\{(x-\beta)^+,2r\}}$. Applying Varadhan's lemma we have

$$E_\Gamma[D_d(R,\Gamma)] \doteq \inf_{\beta\in\mathbb{R}^+} \max\{(x-\beta)^+, 2r\}+\beta = \max\{x, 2r\}.$$

For a family of codes with rate $\frac{b}{2}R_c = \frac{b}{2}r_c\log\rho$, (4.9) is exponentially equivalent to

$$\begin{aligned}
ED_s(br_c\log\rho) &= (1-P_o(\mathbf{H}))\mathrm{E}_\Gamma[D_d(br_c/2\log\rho,\Gamma)] + P_o(\mathbf{H})\mathrm{E}_\Gamma[D_d(0,\Gamma)] \\
&\doteq (1-\rho^{-d^*(r_c)})\rho^{-\max\{x,br_c\}} + \rho^{-d^*(r_c)}\rho^{-x} \\
&\doteq \rho^{-\max\{x,br_c\}} + \rho^{-(d^*(r_c)+x)} \\
&\doteq \rho^{-\min\{\max\{x,br_c\},d^*(r_c)+x\}},
\end{aligned}$$

where we have used that the outage probability is exponentially equivalent to the probability of error [73], i.e., $P_o(\mathbf{H}) \doteq \rho^{-d^*(r_c)}$, and $d^*(r_c)$ is the DMT curve in (4.2).

The best distortion exponent achievable by SSCC, $\Delta_s(b,x)$, is found by maximizing over r_c as follows

$$\Delta_s(b,x) \triangleq \max_{r_c\geq 0}\{\min\{\max\{x, br_c\}, x + d^*(r_c)\}\}. \tag{D.13}$$

The maximum achieved when the two terms inside $\min\{\cdot\}$ are equal, i.e., $\max\{br_c, x\} = x+d^*(r_c)$. We chose a rate r_c such that $br_c > x$ and $r_c < M_*$, as otherwise, the solution is readily given by $\Delta_s(b,x) = x$. Note that for $bM_* \leq x$ this is never feasible, and thus, $\Delta_s(b,x) = x$, and if $x \geq bd^*(M_*)$, the intersection is always at $br_c = x$. Assuming $k \leq r_c \leq k+1$, $k = 0, ..., M_* - 1$, the optimal r_c satisfies at $x + br_c = d^*(r_c)$, or, equivalently, $br_c = x + \Phi_k - (r_c - k)\Upsilon_k$, and we have

$$r_c^* = \frac{\Phi_k + k\Upsilon_k + x}{\Upsilon_k + b}, \quad \Delta_s(b,x) = br_c^* = b\frac{\Phi_k + k\Upsilon_k + x}{\Upsilon_k + b}.$$

Since solution r_c^* is feasible whenever $k < r_c^* \leq k+1$, this solution is defined in

$$b \in \left[\frac{\Phi_{k+1}+x}{k+1}, \frac{\Phi_k + x}{k}\right), \quad \text{for } k = 0, ..., M_* - 1, \tag{D.14}$$

where we have used $\Phi_{k+1} = \Phi_k - \Upsilon_k$. Notice also that, we also need $br_c^* \leq x$, which holds whenever $\Delta_s(b,x) \leq x$ in (D.14). Under these conditions, we have $\Delta_s(b,x) = x$. Remember that for $bM_* \leq x$ we also have $\Delta_s(b,x) = x$. Gathering all results completes the proof of Theorem 7.

D.3 Proof for distortion exponent for JDS

In this section we derive the distortion exponent of JDS, and show that it coincides with the distortion exponent of SSCC. Applying the change of variables $\lambda_i = \rho^{-\alpha_i}$ and $\gamma = \rho^{-\beta}$, and considering a rate $R_j = r_j \log\rho$, $r_j > 0$, the outage event in (4.10) can be written as

$$\begin{aligned}\mathcal{O}_j &= \left\{(\mathbf{H},\gamma) : 1 + \frac{2^{-\epsilon}\rho^{br_j} - 1}{\gamma\rho^x + 1} \geq \prod_{i=1}^{M_*}(1+\rho\lambda_i)^b\right\}\\ &= \left\{(\boldsymbol{\alpha},\beta) : 1 + \frac{2^{-\epsilon}\rho^{br_j} - 1}{\rho^{(x-\beta)} + 1} \geq \prod_{i=1}^{M_*}(1+\rho^{1-\alpha_i})^b\right\}.\end{aligned}$$

For large ρ, we have

$$\frac{1+\frac{2^{-\epsilon}\rho^{br_j}-1}{\rho^{(x-\beta)}+1}}{\prod_{i=1}^{M_*}(1+\rho^{1-\alpha_i})^b} \doteq \frac{1+\rho^{br_j}\rho^{-(x-\beta)^+}}{\rho^{b\sum_{i=1}^{M_*}(1-\alpha_i)^+}} \doteq \rho^{(br_j-(x-\beta)^+)^+ - b\sum_{i=1}^{M_*}(1-\alpha_i)^+}.$$

Therefore, at high SNR, the achievable expected end-to-end distortion for JDS is found as,

$$\begin{aligned}ED_j(br_j\log\rho) &= \int_{\mathcal{O}_j^c} D_d(br_j/2\log\rho, \rho^{-\beta})p_A(\boldsymbol{\alpha})p_B(\beta)d\boldsymbol{\alpha}d\beta\\ &\quad + \int_{\mathcal{O}_j} D_d(0,\rho^{-\beta})p_A(\boldsymbol{\alpha})p_B(\beta)d\boldsymbol{\alpha}d\beta\\ &\doteq \int_{\mathcal{A}_j^c} \rho^{-\max\{(x-\beta)^+, br_j\}}\rho^{-(S(\boldsymbol{\alpha})+\beta)}d\boldsymbol{\alpha}d\beta\\ &\quad + \int_{\mathcal{A}_j}\rho^{-(x-\beta)^+}\rho^{-(S(\boldsymbol{\alpha})+\beta)}d\boldsymbol{\alpha}d\beta.\\ &\doteq \rho^{-\Delta_j^1(r_j)} + \rho^{-\Delta_j^2(r_j)}\\ &\doteq \rho^{-\min\{\Delta_j^1(r_j),\Delta_j^2(r_j)\}}\\ &\doteq \rho^{-\Delta_j(r_j)},\end{aligned} \tag{D.15}$$

where $D_d(R,\gamma)$ is as defined in (4.8), and we have used $D_d(r\log\rho,\beta) \doteq \rho^{-\max\{(x-\beta)^+,2r\}}$. We have also defined the high SNR equivalent of the outage event as

$$\mathcal{A}_j \triangleq \left\{ (\boldsymbol{\alpha},\beta) : (br_j - (x-\beta)^+)^+ \geq b\sum_{i=1}^{M_*}(1-\alpha_i)^+ \right\}.$$

We have applied Varadhan's lemma to each integral to obtain

$$\Delta_j^1(r_j) \triangleq \inf_{\mathcal{A}_j^c} \max\{(x-\beta)^+, br_j\} + \beta + S_A(\boldsymbol{\alpha}), \tag{D.16}$$

and

$$\Delta_j^2(r_j) \triangleq \inf_{\mathcal{A}_j} (x-\beta)^+ + \beta + S_A(\boldsymbol{\alpha}). \tag{D.17}$$

Then, the distortion exponent of JDS is found as

$$\Delta_j(r_j) = \min\{\Delta_j^1(r_j), \Delta_j^2(r_j)\}. \tag{D.18}$$

We first solve (D.16). We can constrain the optimization to $\boldsymbol{\alpha} \geq 0$ and $\beta \geq 0$ without loss of optimality, since for $\boldsymbol{\alpha},\beta < 0$ we have $S_A(\boldsymbol{\alpha}) = S_B(\beta) = +\infty$. Then, $\Delta_j^1(r_j)$ is minimized by $\boldsymbol{\alpha}^* = 0$ since this minimizes $S_A(\boldsymbol{\alpha})$ and enlarges $\mathcal{A}_j^c$. We can rewrite (D.16)

$$\Delta_j^1(r_j) = \inf_{\beta\geq 0} \max\{(x-\beta)^+, br_j\} + \beta \qquad \text{s.t. } (br_j - (x-\beta)^+)^+ < bM_*.$$

If $br_j < (x-\beta)^+$, the minimum is achieved by any $0 \leq \beta < x - r_j b$, and thus $\Delta_j^1(r_j) = x$ for $x > br_j$. If $br_j \geq (x-\beta)^+$, then

$$\Delta_j^1(r_j) = \inf_{\beta\geq 0} br_j + \beta \qquad \text{s.t. } br_j - bM_* < (x-\beta)^+ \leq br_j. \tag{D.19}$$

If $\beta > x$, the problem is minimized by $\beta^* = x + \epsilon$, $\epsilon > 0$, and $\Delta_j(r_j) = br_j + x + \epsilon$, for $r_j \leq M_*$. For $0 \leq \beta \leq x$, we have $\beta^* = (x - r_j b)^+$, and $\Delta_j^1(r_j) = \max\{br_j, x\}$ if $br_j \leq bM_* + x$. Putting all these together, we obtain

$$\Delta_j^1(r_j) = \max\{br_j, x\} \quad \text{if } br_j \leq x + bM_*. \tag{D.20}$$

If $br_j > x + bM_*$, $\mathcal{A}_j^c$ is empty, and there is always outage.

Next we solve the second optimization problem in (D.17). With $\beta = x$, $\Delta_j^2(r_j)$ is minimized and the range of $\boldsymbol{\alpha}$ is enlarged. Then, the problem to solve reduces to

$$\Delta_j^2(r_j) = \inf x + S(\boldsymbol{\alpha}) \qquad \text{s.t. } r_j \geq \sum_{i=1}^{M_*}(1-\alpha_i)^+,$$

which is the DMT problem in (D.8), and $\Delta_{2j}(r_j, b) = x + d^*(r_j)$. Bringing all together,

$$\Delta_j(b, x) = \max_{r_j \geq 0} \{\min\{\max\{x, br_j\}, x + d^*(r_j)\}\}. \tag{D.21}$$

Since $d^*(r_j) = 0$ for $r_j > M_*$, the constraint in (D.21) can be reduced to $0 \leq r_j \leq M_*$ without loss of optimality since $\Delta_j(b, x) = x$ for any $r_j \geq M_*$. Then, (D.21) coincides with (D.13), and thus, SSCC and JDS achieve the same distortion exponent.

D.4 Proof Of Distortion Exponent for HDA-WZ

In this Appendix we derive the outage region $\mathcal{O}_h$ in (4.14) and the distortion in (4.15). Then, using these we obtain the distortion exponent achievable by HDA-WZ.

D.4.1 Outage region for HDA-WZ

From joint typicality arguments similarly to [34], the decoding of $\mathbf{W}^{\frac{m}{2M_*}}$ is successful with high probability if

$$I(\mathbf{W}^{\frac{m}{2M_*}}; S^m) < \frac{m}{2} R_h < I(\mathbf{W}^{\frac{m}{2M_*}} \mathbf{U}^{n-\frac{m}{2M_*}}; \mathbf{Y}^n T^m). \tag{D.22}$$

For the r.h.s. of (D.22) we have

$$I(\mathbf{W}^{\frac{m}{2M_*}}; S^m) = \sum_{i=1}^{\frac{m}{2M_*}} I(\mathbf{W}_i; \mathbf{S}_i) = \frac{m}{2M_*} I(\mathbf{W}; \mathbf{S}) = mI(W; S) < \frac{m}{2} R_h, \tag{D.23}$$

due to the i.i.d. distribution of the source, $\mathbf{Q}_i$ and $\mathbf{X}_i$. Note that the l.h.s. of (D.22) always holds since R_h is chosen such that $\frac{R_h}{2} = I(W; S) + \epsilon$.

The r.h.s. of the decoding condition (D.22) is given by

$$\begin{aligned} I(\mathbf{W}^{\frac{m}{2M_*}} \mathbf{U}^{n-\frac{m}{2M_*}}; \mathbf{Y}^n T^m) &\overset{(a)}{=} \sum_{i=1}^{\frac{m}{2M_*}} I(\mathbf{W}_i; \mathbf{Y}_{W,i} \mathbf{T}_i) + \sum_{i=\frac{m}{2M_*}+1}^{n} I(\mathbf{U}_i; \mathbf{Y}_{U,i}) \\ &= \frac{m}{2M_*} I(\mathbf{W}; \mathbf{Y}_W \mathbf{T}) + \left(n - \frac{m}{2M_*}\right) I(\mathbf{U}; \mathbf{Y}_U), \end{aligned} \tag{D.24}$$

where (a) follows from the i.i.d. distribution of the implied variables.

Substituting (D.23) and (D.24) into (D.22) and dividing both sides by $m/2M_*$, the outage condition in (4.14) follows.

Next, we evaluate the outage region in (4.14) for Gaussian codewords. We have $I(\mathbf{W}; \mathbf{Y}_W \mathbf{T}) = H(\mathbf{Y}_W \mathbf{T}) - H(\mathbf{Y}_W \mathbf{W} \mathbf{T}) + H(\mathbf{W})$. Let $\mathbf{G} \triangleq [\mathbf{W}, \mathbf{Y}_W, \mathbf{T}]^H$. Since $\mathbf{G}$ is a complex multivariate Gaussian random vector, its differential entropy is given by

$H(\mathbf{G}) = \log((2\pi e)^{3M_*} \det(\mathbf{C}_\mathbf{G}))$, where $\mathbf{C}_\mathbf{G} = \mathrm{E}[\mathbf{G}\mathbf{G}^H]$ is given by

$$\mathbf{C}_\mathbf{G} = \begin{bmatrix} \mathbf{I} + \sigma_Q^2\mathbf{I} & \sqrt{\alpha}\sigma_Q^2\mathbf{H}^H & \gamma\sqrt{\rho_s}\mathbf{I} \\ \sqrt{\alpha}\sigma_Q^2\mathbf{H} & \mathbf{I} + \alpha\sigma_Q^2\mathbf{H}\mathbf{H}^H & \mathbf{0} \\ \gamma\sqrt{\rho_s}\mathbf{I} & \mathbf{0} & \xi\mathbf{I} \end{bmatrix},$$

with $\alpha \triangleq \rho/M_t$ and $\xi \triangleq 1 + \rho_s\gamma$. By using properties of the determinant of a block matrix and some algebra, we have

$$\det(\mathbf{C}_\mathbf{G}) = \det\left(\mathbf{I} + \alpha\sigma_Q^2\mathbf{H}\mathbf{H}^H + \xi\sigma_Q^2\mathbf{I}\right) = \prod_{i=1}^{M_*}\left(1 + \xi\sigma_Q^2 + \frac{\rho}{M_*}\lambda_i\right).$$

Similarly, we have

$$\begin{aligned} H(\mathbf{Y}_W\mathbf{T}) &= \log\left((2\pi e)^{2M_*}\xi^{M_*}\det\left(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}\mathbf{H}^H\right)\right), \\ H(\mathbf{W}) &= \log((2\pi e)^{M_*}(1+\sigma_Q^2)^{M_*}), \\ I(\mathbf{Y}_U;\mathbf{U}) &= \log\left(\det\left(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}\mathbf{H}^\dagger\right)\right). \end{aligned}$$

Substituting in (4.13) we have the outage region (4.14) .

D.4.2 Expected distortion achieved by HDA-WZ

We use a MMSE estimator to reconstruct each source block $\mathbf{S}_i$, $i = 1, ..., \frac{m}{2M_*}$, with the available information, which can be modeled by the linear model as follows:

$$\begin{bmatrix} \mathbf{W}_i \\ \mathbf{Y}_i \\ \mathbf{T}_i \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \\ \gamma\mathbf{I} \end{bmatrix}\mathbf{X}_i + \begin{bmatrix} \mathbf{Q}_i \\ \sqrt{\alpha}\mathbf{H}\mathbf{Q}_i + \mathbf{N}_i \\ \mathbf{Z}_i \end{bmatrix}.$$

Let $\mathbf{B} \triangleq \begin{bmatrix}\mathbf{I} & \mathbf{0} & \gamma\mathbf{I}\end{bmatrix}^H$ and $\mathbf{S}_i \triangleq \begin{bmatrix}\mathbf{Q}_i & \alpha\mathbf{H}\mathbf{Q}_i + \mathbf{N}_i & \mathbf{Z}_i\end{bmatrix}^H$. Then, the distortion for each source block is found to be given by $\mathrm{Tr}\{\mathbf{D}\} = \frac{1}{M_*}\sum_{i=1}^{M_*}\mathrm{Tr}[\mathbf{I} + \mathbf{B}\mathbf{C}_S\mathbf{B}^H]^{-1}$, where $\mathbf{D}$ is the distortion matrix in the reconstruction of each block, and

$$\mathbf{C}_\mathbf{S} \triangleq \mathrm{E}[\mathbf{S}_i\mathbf{S}_i^H] = \begin{bmatrix} \mathbf{I} & \sqrt{\alpha}\mathbf{H}^H & \mathbf{0} \\ \sqrt{\alpha}\mathbf{H} & \alpha\sigma_Q^2\mathbf{H}\mathbf{H}^H + \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

Using the block inverse properties,the singular value decomposition of $\mathbf{H}$ we obtain the expected distortion expression in (4.15).

D.4.3 Distortion exponent achieved by HDA-WZ

In this section we derive the distortion exponent for HDA-WZ. The outage region in

(4.14) is given by

$$\mathcal{O}_h = \left\{ (\mathbf{H},\gamma) : \left(1+\frac{1}{\sigma_Q^2}\right)^{M_*} \geq \left(((1+\rho_s\gamma)(1+\sigma_Q^2))^{M_*} \frac{\prod_{i=1}^{M_*}(1+\frac{\rho}{M_*}\lambda_i)^{bM_*}}{\prod_{i=1}^{M_*}(1+\frac{\rho}{M_*}\lambda_i + (1+\rho_s\gamma)\sigma_Q^2)} \right) \right\}.$$

Similar to the analysis for the other schemes, we consider the change of variables $\lambda_i = \rho^{-\alpha_i}$, and $\gamma = \rho^{-\beta}$, and a rate $R_h = r_h \log\rho$, for $r_h \geq 0$. Then, we start by finding the equivalent outage set in the high SNR regime. We have,

$$\prod_{i=1}^{M_*}\left(1+\frac{\rho}{M_*}\lambda_i\right)^{bM_*} \doteq \rho^{bM_*\sum_{i=1}^{M_*}(1-\alpha_i)^+},$$

and

$$\begin{aligned}\prod_{i=1}^{M_*}\left(1+\frac{\rho}{M_*}\lambda_i+(1+\rho_s\gamma)\sigma_Q^2\right) &\doteq \prod_{i=1}^{M_*}\left(1+\rho^{1-\alpha_i}+(1+\rho^{x-\beta})\rho^{-r_h}\right)\\ &\doteq \rho^{\sum_{i=1}^{M_*}\max\{(1-\alpha_i)^+,(x-\beta)^+-r_h\}},\end{aligned}$$

where we use $\sigma_Q^2 = (2^{R_h-\epsilon}-1)^{-1} = (2^{-\epsilon}\rho^{r_h}-1)^{-1} \doteq \rho^{-r_h}$. From the outage condition in (D.25), we have

$$\begin{aligned}\frac{\left(1+\frac{1}{\sigma_Q^2}\right)^{M_*}\prod_{i=1}^{M_*}(1+\frac{\rho}{M_*}\lambda_i+(1+\rho_s\gamma)\sigma_Q^2)}{((1+\rho_s\gamma)(1+\sigma_Q^2))^{M_*}\prod_{i=1}^{M_*}(1+\frac{\rho}{M_*}\lambda_i)^{bM_*}} &\doteq \frac{\rho^{M_*r_h}\rho^{\sum_{i=1}^{M_*}\max\{(1-\alpha)^+,(x-\beta)^+-r_h\}}}{\rho^{M_*(x-\beta)^+}\rho^{bM_*\sum_{i=1}^{M_*}(1-\alpha)^+}}\\ &\doteq \rho^{\sum_{i=1}^{M_*}(r_h-(x-\beta)^++(1-\alpha_i))^+-bM_*\sum_1^{M_*}(1-\alpha_i)^+}.\end{aligned}$$

Therefore, in the high SNR regime, the set $\mathcal{O}_h$ is equivalent to the set given by

$$\mathcal{A}_h \triangleq \left\{(\boldsymbol{\alpha},\beta)^+ : \sum_{i=1}^{M_*}(r_h-(x-\beta)^++(1-\alpha_i))^+ > bM_*\sum_{i=1}^{M_*}(1-\alpha_i)\right\}.$$

On the other hand, in the high SNR regime, the distortion achieved by HDA-WZ is equivalent to

$$\begin{aligned}D_h(\sigma_Q^2,\mathbf{H},\gamma) &= \frac{1}{M_*}\sum_{i=1}^{M_*}\left(1+\rho_s\gamma+\frac{1}{\sigma_Q^2}\left(1+\frac{\rho}{M_*}\lambda_i\right)\right)^{-1}\\ &\doteq \sum_{i=1}^{M_*}\left(1+\rho^{x-\beta}+\rho^{r_h+(1-\alpha_i)}\right)^{-1}\\ &\doteq \rho^{-\min_{i=1,\dots,M_*}\{\max\{(x-\beta)^+,r_h+1-\alpha_i\}\}}\\ &\doteq \rho^{-\max\{(x-\beta)^+,r_h+1-\alpha_1\}},\end{aligned}$$

where the last equality follows since $\alpha_1 \geq \ldots \geq \alpha_{M_*} \geq 0$. Then, in the high SNR regime, the expected distortion for HDA-WZ is given as

$$\begin{aligned}
ED_h(r_h \log\rho) &= \int_{\mathcal{O}_h^c} D_h(\sigma_Q^2, \mathbf{H}, \gamma) p_h(\mathbf{H}) p_\Gamma(\gamma) d\mathbf{H} d\gamma \\
&\quad + \int_{\mathcal{O}_h} D_d(0,\gamma) p_h(\mathbf{H}) p_\Gamma(\gamma) d\mathbf{H} d\gamma \\
&\doteq \int_{\mathcal{A}_j^c} \rho^{-\max\{(x-\beta)^+, r_h+(1-\alpha_1)\}} p_A(\boldsymbol{\alpha}) p_B(\beta) d\boldsymbol{\alpha} d\beta \\
&\quad + \int_{\mathcal{A}_j} \rho^{-(x-\beta)^+} p_A(\boldsymbol{\alpha}) p_B(\beta) d\boldsymbol{\alpha} d\beta.
\end{aligned}$$

Similarly to the proof of Theorem 9, applying Varadhan's lemma, the exponent of each integral is found as

$$\Delta_h^1(r_h) = \inf_{\mathcal{A}_h^c} \max\{(x-\beta)^+, r_h + 1 - \alpha_1\} + S_A(\boldsymbol{\alpha}) + \beta,$$

and

$$\Delta_h^2(r_h) = \inf_{\mathcal{A}_h} (x-\beta)^+ + S_A(\boldsymbol{\alpha}) + \beta, \tag{D.25}$$

First we solve $\Delta_h^1(r_h)$. The infimum for this problem is achieved by $\boldsymbol{\alpha}^* = 0$ and $\beta^* = 0$, and is given by

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

Now we solve $\Delta_h^2(r_h)$ in (D.25). By letting $\beta^* = x$, the range of $\boldsymbol{\alpha}$ is enlarged while the objective function is minimized. Thus, the problem to solve reduces to

$$\Delta_h^2(r_h) = \inf x + S(\boldsymbol{\alpha}) \qquad \text{s.t. } r_h > \frac{bM_* - 1}{M^*} \sum_{i=1}^{M_*} (1-\alpha_i)^+. \tag{D.26}$$

Again, this problem is a scaled version of the DMT curve in (D.8). Therefore, we have

$$\Delta_h^2(r_h) = x + d^* \left(\left(\frac{bM_* - 1}{M_*} \right)^{-1} r_h \right).$$

The distortion exponent is given by optimizing over r_h as

$$\Delta_h(b, x) = \max_{r_h} \min\{\Delta_h^1(r_h), \Delta_h^2(r_h)\}.$$

The maximum distortion exponent is obtained by letting $\Delta_h^1(r_h) = \Delta_h^2(r_h)$. We assume $r_h + 1 > x$ since otherwise $\Delta_h(b,x) = x$, and then, we have $r_h + 1 = x +$

$d^*\left((b-\frac{1}{M_*})^{-1}r_h\right)$. Let $r'_h = r_h(b-\frac{1}{M_*})^{-1}$. Using (4.2), for $k < r'_h \leq k+1$, $k = 0, ..., M_*-1$, the problem is equivalent to $r'_h\left(b-\frac{1}{M_*}\right)+1 = x+\Phi_k-(r'_h-k)\Upsilon_k$, where Φ_k and Υ_k are given as in (4.3). The r'_h satisfying the equality if given by

$$r'^*_h = \frac{\Phi_k + k\Phi_k - 1 + x}{b - \frac{1}{M_*} + \Phi_k},$$

and the corresponding distortion exponent is found as

$$\Delta_h(b, x) = 1 + \frac{(bM_* - 1)(\Phi_k + k\Upsilon_k - 1 + x)}{bM_* - 1 + M_*\Upsilon_k},$$

for

$$b \in \left[\frac{\Phi_{k+1} - 1 + x}{k+1} + \frac{1}{M_*}, \frac{\Phi_k - 1 + x}{k} + \frac{1}{M_*}\right), \qquad \text{for } k = 0, ..., M_* - 1.$$

Note that we have $r^*_h + 1 > x$ whenever $\Delta_h(b, x) > x$. Otherwise, r^*_h is not feasible and $\Delta_h(b, x) = x$. Note also that if $x \geq bM_*$, the distortion exponent is given by $\Delta_h(b, x) = x$.

D.5 Distortion exponent achieved by LS-JDS

D.5.1 Successively refinable codebooks

Consider a successively refinable codebook [38] at rate $tR_k = I(S; W_l|W_1^{l-1}) + \epsilon$ for each codebook layer, where $t > 0$. Then, we have

$$\begin{aligned} I(S; W_l|W_1^{l-1}) &\overset{(a)}{=} I(S; W_1^l) - I(S; W_1^{l-1}) \\ &\overset{(b)}{=} I(S; W_k) - I(S; W_{k-1}), \end{aligned} \tag{D.27}$$

where (a) is due to the chain rule, and (b) holds form the Markov chain $S - W_l - W_1^{l-1}$. We have

$$\begin{aligned} \sum_{i=1}^{l}(tR_i - \epsilon) &= \sum_{i=1}^{l} I(S; W_i|W_1^{i-1}) \\ &\overset{(a)}{=} \sum_{i=1}^{l} I(S; W_i) - I(S; W_{i-1}) \\ &= I(S; W_l) \\ &= \frac{1}{2}\log\left(1 + \frac{1}{\sum_{i=l}^{L}\sigma_i^2}\right), \ l = 1, ..., L, \end{aligned}$$

where (a) follows from (D.27) and $W_0 = \emptyset$ for the case $l = 1$. Substituting $t = \frac{b}{2L}$, i.e., allocating equal channel accesses per layer, equation (4.16) follows.

D.5.2 Distortion exponent achievable by of LS-JDS

In this section we obtain the distortion exponent for LS-JDS. Let us define $\bar{R}_1^l \triangleq \sum_{i=1}^{l} R_i$. First, we consider the outage event. For the successive refinement codebook we have that the l.h.s. of (4.17) is given by

$$\begin{aligned}
I(S;W_l|W_1^{l-1},T) &\overset{(a)}{=} I(S;W_l|T) - I(S;W_{l-1}|T) \\
&\overset{(b)}{=} H(W_l|T) - H(\overline{Q}_l) - H(W_{l-1}|T) + H(\overline{Q}_{l-1}) \\
&= \frac{1}{2}\log\left(\frac{\sum_{i=l-1}^{L}\sigma_i^2}{\sum_{i=l}^{L}\sigma_i^2}\frac{1+(1+\gamma\rho_s)\sum_{j=l}^{L}\sigma_j^2}{1+(1+\gamma\rho_s)\sum_{j=l-1}^{L}\sigma_j^2}\right),
\end{aligned}$$

where $\overline{Q}_l \triangleq \sum_{i=l}^{L} Q_l$, (a) is due to the Markov chain $T-S-W_L-...-W_1$, and (b) is due to the independence of $\bar{Q}_i$ with S and T, and $H(W_l|T)=\frac{1}{2}\log\left(\sum_{i=l}^{L}\sigma_i^2+\frac{1}{1+\gamma\rho_s}\right)$. We also have

$$I(S;W_1|T) = \frac{1}{2}\log\left(1+\frac{1}{(1+\gamma\rho_s)\sum_{i=1}^{L}\sigma_i^2}\right).$$

Substituing (4.16) with $R_0=0$, we have

$$I(S;W_l|W_1^{l-1},T) = \frac{1}{2}\log\left(\frac{2^{\sum_{i=1}^{l}\frac{b}{L}R_i}+\gamma\rho_s}{2^{\sum_{i=1}^{l-1}\frac{b}{L}R_i}+\gamma\rho_s}\right).$$

Then, the outage condition in (4.17) is given by

$$\log\left(\frac{2^{\sum_{i=1}^{l}\frac{b}{L}R_i}+\gamma\rho_s}{2^{\sum_{i=1}^{l-1}\frac{b}{L}R_i}+\gamma\rho_s}\right) \le \frac{b}{L}\log\prod_{i=1}^{M_*}\left(1+\frac{\rho}{M_*}\lambda_i\right). \tag{D.28}$$

Therefore, in the high SNR regime, we have, for $l=1,...,L$

$$\begin{aligned}
\frac{2^{\sum_{i=1}^{l}(\frac{b}{L}R_i-\epsilon)}+\gamma\rho_s}{2^{\sum_{i=1}^{l-1}(\frac{b}{L}R_i-\epsilon)}+\gamma\rho_s} &\doteq \frac{\rho^{\sum_{i=1}^{l}\frac{b}{L}r_i}+\rho^{x-\beta}}{\rho^{\sum_{i=1}^{l-1}\frac{b}{L}r_i}+\rho^{x-\beta}} \\
&\doteq \frac{\rho^{\sum_{i=1}^{l}\frac{b}{L}r_i-(x-\beta)}+1}{\rho^{\sum_{i=1}^{l-1}\frac{b}{L}r_i-(x-\beta)}+1} \\
&\doteq \frac{\rho^{(\sum_{i=1}^{l}\frac{b}{L}r_i-(x-\beta))^+}}{\rho^{(\sum_{i=1}^{l-1}\frac{b}{L}r_i-(x-\beta))^+}},
\end{aligned} \tag{D.29}$$

and

$$\frac{b}{L}\log\prod_{i=1}^{M_*}\left(1+\frac{\rho}{M_*}\lambda_i\right) \doteq \rho^{\frac{b}{L}\sum_{i=1}^{M_*}(1-\alpha_i)^+}.$$

The outage set (4.17) in the high SNR is equivalent to

$$\mathcal{A}_{l'}^{ls} \triangleq \left\{ (\boldsymbol{\alpha}, \beta) : \frac{b}{L}\sum_{i=1}^{M_*}[(1-\alpha_i)^+ \right.$$
$$\left. < \left(\sum_{i=1}^{l}\frac{b}{L}r_i - (x-\beta)\right)^+ - \left(\sum_{i=1}^{l-1}\frac{b}{L}r_i - (x-\beta)\right)^+ \right\}. \quad \text{(D.30)}$$

Now, we study the high SNR behavior of the expected distortion. It is not hard to see that (4.18) is given by

$$ED_{ls}(\mathbf{R}) = \sum_{l=0}^{L} \mathrm{E}_{\mathcal{O}_{l+1}^{ls}}\left[D_d\left(\frac{b}{2L}\bar{R}_1^l, \gamma\right)\right] - \mathrm{E}_{\mathcal{O}_l^{ls}}\left[D_d\left(\frac{b}{2L}\bar{R}_1^l, \gamma\right)\right], \quad \text{(D.31)}$$

where $\mathcal{O}_0^{ls} \triangleq \emptyset$ and $\mathcal{O}_{L+1}^{ls} \triangleq \mathbb{R}^{M_*+1}$. For each summing term in (D.31), we have

$$\mathrm{E}_{\mathcal{O}_{l'}^{ls}}\left[D_d\left(\frac{b}{2L}\bar{R}_1^l, \gamma\right)\right] \doteq \int_{\mathcal{A}_{l'}^{ls}} \rho^{-\max\{\frac{b}{L}\sum_{i=1}^{l} r_l, (x-\beta)^+\}}\rho^{-S_A(\boldsymbol{\alpha})}\rho^{-\beta} d\boldsymbol{\alpha} d\beta,$$

where the outage set in the high SNR regime is given by (D.30).

Applying Varadhan's lemma to (D.32), the exponential behavior for $l = 0, ..., L-1$ and $l' = l, l+1$, is found as the solution to

$$\Delta^{ls}(l, l') \triangleq \inf_{\mathcal{A}_{l'}^{ls}} \max\{b/L\bar{r}_1^l, (x-\beta)^+\} + S_A(\boldsymbol{\alpha}) + \beta,$$

where we define $\bar{r}_1^l \triangleq \sum_{i=1}^{l} r_i$. Note that since $r_1 \le r_2 \le ...$ we have $\mathcal{A}_l^{ls} \subseteq \mathcal{A}_{l+1}^{ls}$ and therefore $\Delta^{ls}(l,l) \ge \Delta^{ls}(l, l+1)$. Then, we have from (D.31)

$$ED_{ls}(\mathbf{R}) \doteq \sum_{l=0}^{L} \rho^{-\Delta^{ls}(l,l+1)} - \rho^{-\Delta^{ls}(l,l)} \doteq \sum_{l=0}^{L} \rho^{-\Delta^{ls}(l,l+1)}.$$

We define $\Delta_l^{ls}(\mathbf{r}) \triangleq \Delta^{ls}(l, l+1)$, where $\mathbf{r} \triangleq [r_1, ..., r_L]$. The distortion exponent of LS-JDS is given as follows:

$$\Delta_{ls}^*(b, x) = \max_{\mathbf{r}} \min \Delta_l^{ls}(\mathbf{r}).$$

For $l=0$, i.e., no codeword is successfully decoded, we have

$$\Delta_0^{ls}(\mathbf{r}) = \inf (x-\beta)^+ + \beta + S_A(\boldsymbol{\alpha}) \qquad \text{s.t. } \frac{b}{L}\sum_{i=1}^{M_*}(1-\alpha_i)^+ < \left(\frac{b}{L}r_1 - (x-\beta)\right)^+.$$

The infimum is achieved by $\beta = x$ and using the DMT in (4.1), we have

$$\Delta_0^{ls}(\mathbf{r}) = x + d^*\left(r_1\right).$$

The distortion exponent for l layers decoded is found as

$$\begin{aligned}\Delta_l^{ls}(\mathbf{r}) \;&=\; \inf\max\left\{\frac{b}{L}\bar{r}_1^l, (x-\beta)^+\right\} + \beta + S_A(\boldsymbol{\alpha}) \\ \text{s.t. } &\frac{b}{L}\sum_{i=1}^{M_*}(1-\alpha_i)^+ < \left(\frac{b}{L}\bar{r}_1^{l+1} - (x-\beta)\right)^+ - \left(\frac{b}{L}\bar{r}_1^l - (x-\beta)\right)^+.\end{aligned} \tag{D.32}$$

If $\frac{b}{L}\bar{r}_1^l \geq x$, the infimum of (D.32) is obtained for $\beta^* = 0$ and

$$\Delta_l^{ls}(\mathbf{r}) = \inf \frac{b}{L}\bar{r}_1^l + S_A(\boldsymbol{\alpha}) \qquad \text{s.t. } \sum_{i^1}^{M_*}(\xi_l - \alpha_i)^+ < r_{k+1}. \tag{D.33}$$

Using the DMT in (4.1), (D.33) is minimized as

$$\Delta_l^{ls}(\mathbf{r}) = \frac{b}{L}\bar{r}_1^l + d^*\left(r_{l+1}\right).$$

If $\frac{b}{L}\bar{r}_1^l \leq x$, we have that the minimum of (D.32) is achieved by $\beta^* = \left(x - \frac{b}{L}\bar{r}_1^l\right)^+$ if $\frac{b}{L}\bar{r}_1^l > (x-\beta)$ and is given by

$$\Delta_l^{ls}(\mathbf{r}) = x + d^*\left(r_{l+1}\right).$$

If $\frac{b}{L}\bar{r}_1^l \leq (x-\beta) < \frac{b}{L}\bar{r}_1^{l+1}$, problem (D.32) is equivalent to

$$\begin{aligned}\Delta_l^{ls}(\mathbf{r}) = \inf (x-\beta)^+ + \beta + S_A(\boldsymbol{\alpha}) \\ \text{s.t. } \frac{b}{L}\sum_{i=1}^{M_*}(1-\alpha_i)^+ < \left(\frac{b}{L}r_1^{l+1} - (x-\beta)\right)^+, \\ \frac{b}{L}\bar{r}_1^l \leq (x-\beta) < \frac{b}{L}\bar{r}_1^{l+1}.\end{aligned} \tag{D.34}$$

The infimum of (D.34) is achieved by the largest β, since the range of $\boldsymbol{\alpha}$ increases. Then, $\beta^* = (x - \frac{b}{L}\bar{r}_1^l)^+$, and we have,

$$\Delta_l^{ls}(\mathbf{r}) = x + d^*\left(r_{l+1}\right).$$

Finally, if $\frac{b}{L}\bar{r}_1^{l+1} \leq (x-\beta)$, there are no feasible solutions for (D.32). Therefore, putting all together we have

$$\Delta_l^{ls}(\mathbf{r}) = \inf \max \left\{ \frac{b}{L} \bar{r}_1^l, x \right\} + d^*(r_{l+1}).$$

Similarly, at layer L, the infimum is achieved by $\boldsymbol{\alpha}^* = 0$ and $\beta^* = 0$ and is given by

$$\Delta_L^{ls}(\mathbf{r}) = \max \left\{ \frac{b}{L} \bar{r}_1^L, x \right\}, \quad \text{for } r_L \leq M_*.$$

Note that the condition on r_L always holds.

D.5.3 Solution of the distortion exponent

Assume that for a given layer $\hat{l}$ we have $\bar{r}_1^{\hat{l}-1} \frac{b}{L} \leq x \leq \bar{r}_1^{\hat{l}} \frac{b}{L}$. Then, $\Delta_l^{ls}(\mathbf{r}) = x + d(r_{l+1})$ for $l = 0, ..., \hat{l} - 1$. Using the KKT conditions, the maximim distortion exponent is given when all the distortion exponents are equal.

From $\Delta_0^{ls}(\mathbf{r}) = ... = \Delta_{\hat{l}-1}^{ls}(\mathbf{r})$ we have $r_1 = ... = r_{\hat{l}}$ and thus, $\bar{r}_1^{\hat{l}} = \hat{l} r_1$. Then, the exponents are given by

$$\begin{aligned}
\Delta_0^{ls}(\mathbf{r}) &= x + d^*(r_1) \\
\Delta_{\hat{l}}^{ls}(\mathbf{r}) &= b\frac{\hat{l}}{L} r_1 + d^*(r_{\hat{l}+1}) \\
&\cdots \\
\Delta_{L-1}^{ls}(\mathbf{r}) &= b\frac{\hat{l}}{L} r_1 + b\frac{1}{L} \bar{r}_{\hat{l}+1}^{L-1} + d^*(r_L) \\
\Delta_L^{ls}(\mathbf{r}) &= b\frac{\hat{l}}{L} r_1 + b\frac{1}{L} \bar{r}_{\hat{l}+1}^{L}.
\end{aligned}$$

Equaling all these exponents, we have

$$\begin{aligned}
b\frac{1}{L}\bar{r}_L &= d^*(r_L) \\
d(r_L) + b\frac{1}{L} r_{L-1} &= d^*(r_{L-1}) \\
&\cdots \\
b\frac{1}{L} r_{l+1} + d^*(r_{l+2}) &= d^*(r_{l+1}) \\
b\frac{l}{L} r_1 + d^*(r_{l+1}) &= x + d^*(r_1).
\end{aligned}$$

Next, we adapt Lemma 3 in [15] to our setup. Let q be a line with equation $y = -\alpha(t - M)$ for some $\alpha > 0$ and $M > 0$ and let $q_i = 1, ..., L$ be the set of lines defined recursively from L to 1 as $y = (b/L)t + d_{i+1}$, where $b > 0$, $d_{L+1} = 0$ and d_i is the y component of the intersection of q_i with q. Then, sequentially solving the intersection points for $i = \hat{l} + 1, ..., L$ we have:

$$d_i - d_{i+1} = M\frac{b}{L}\left(\frac{\alpha}{\alpha + b/L}\right)^{L-i+1}.$$

Summing all the terms for $i = \hat{l} + 1, ..., L$ we obtain

$$d_i = M\alpha\left[1 - \left(\frac{\alpha}{\alpha + b/L}\right)^{L-i+1}\right].$$

In the following we consider a continuum of layers, i.e., $L \to \infty$. We let $\hat{l} = \kappa L$ be the numbers of layers needed so that $b\hat{l}/Lr_1 = b\kappa r_1 = x$, that is, from $l = 1$ to $l = \kappa L$.

When $M_* = 1$, the DMT curve is composed of a single line with $\alpha = M^*$ and $M = 1$. In that case we have that, with layers from $\kappa L + 1$ to L the distortion increases up to

$$d(r_{L\kappa+1}) = M\alpha\left[1 - \left(\frac{\alpha}{\alpha + b/L}\right)^{L(1-\kappa)}\right].$$

In the limit of infinite layers, we obtain

$$d_{L\kappa+1} \triangleq \lim_{L\to\infty} d(r_{L\kappa+1}) = M\alpha\left(1 - e^{-\frac{b(1-\kappa)}{\alpha}}\right).$$

We still need to determine the distortion achieved due to the climb with layers from $l = 1$ to $l = \kappa L$ by determining r_1. Value, r_1 is found as the solution to $\Delta_0^{ls}(\mathbf{r}) = \Delta_l^{ls}(\mathbf{r})$, i.e.,

$$b\kappa r_1 + d^*(r_{l+1}) = x - \alpha(r_1 - M), \tag{D.35}$$

Since $x = b\kappa r_1$, then $r_1 = x/b\kappa$ and (D.35) is equal to

$$d^*(r_{l+1}) = -\alpha\left(\frac{x}{b\kappa} - M\right),$$

which solves for

$$\kappa^* = \frac{M^*}{b}\mathcal{W}\left(\frac{e^{\frac{b}{M^*}x}}{M^*}\right),$$

where $\mathcal{W}(z)$ is the function W of Lambert, which gives the principal solution for w in $z = we^w$. The distortion exponent in the MISO/SIMO case is then found as

$$\Delta_{ls}^*(b, x) = x + M^*\left(1 - e^{-\frac{b(1-\kappa^*)}{M^*}}\right).$$

For MIMO channels, the DMT curve is formed by M_* linear intervals $k = 1, ..., M_*$

between $M_* - k$ and $M_* - k + 1$. From the value of the DMT at $M * -k$ to the value at $M_* - k + 1$, there is a gap of $M^* - M_* + 2k - 1$ in the y abscise. The increase in each curve can be characterized by $y = -\alpha(t - M)$ where for the k-th interval we have $\alpha = \alpha_k$ and $M = M_k$ as in (4.19). Note that at $t = M_* - i + 1$ we have $y = 0$, that is, each curve is shifted and each interval starts at $y = 0$.

We consider a continuum of layers, i.e., $L \to \infty$ and we let $l = L\kappa$ be the number of lines required to have $b\kappa r_1 = x$. Then, from the remaining lines from $l + 1$ to L, let $L(1 - \kappa)\kappa_i$ be the number of lines with slope b/L required to climb all the interval i. So that the whole interval is climbed with $L(1 - \kappa)\kappa_i$ lines we need

$$d_{L-L(1-\kappa)\kappa_k} = M^* - M_* + 2k - 1,$$

where

$$d_{L-L(1-\kappa)\kappa_k} = M\alpha \left[1 - \left(\frac{\alpha}{\alpha + b/L}\right)^{L(1-\kappa)\kappa_k+1}\right].$$

In the limit we have

$$\lim_{L\to\infty} d_{L-L(1-\kappa)\kappa_k} = M\alpha \left[1 - e^{-\frac{b(1-\kappa)\kappa_k}{\alpha}}\right].$$

Then, each required proportion is found as

$$\kappa_k = \frac{M^* - M_* + 2k - 1}{b(1 - \kappa)} \ln\left(\frac{M_* - k + 1}{M_* - k}\right).$$

This gives the proportion of lines required to climb up the k-th segment of the DMT curve. In the MIMO case, to be able to go up exactly to the k-th segment with lines from $l+1$ to L we need to have $\sum_{j=1}^{k-1} \kappa_j < 1 \leq \sum_{j=1}^{k} \kappa_j$. This is equivalent to the requirement $c_{k-1} < b(1 - \kappa) \leq c_k$ using c_i as defined in the theorem. To climb up each line segment we need $\kappa_k(1-\kappa)L$ lines (layers) for $k = 1, ..., M_*-1$ and for the last segment climbed we have $(1-\sum_{j=1}^{k-1} \kappa_j)L$ lines remaining, which gives an extra ascent of

$$M\alpha\left(1 - e^{-\frac{b(1-\kappa)(1-\sum_{j=1}^{k-1}\kappa_j)}{\alpha}}\right).$$

Then, we have that the distortion exponent has climbed up to

$$\begin{aligned} d_{L\kappa+1} = &\sum_{i=1}^{k-1}(M^* - M_* + 2i - 1) \\ &+ (M_* - k + 1)(M^* - M_* + 2k - 1)\left(1 - e^{-\frac{b(1-\kappa)(1-\sum_{j=1}^{k-1}\kappa_j)}{M^*-M_*+2k-1}}\right). \end{aligned}$$

With the remaining lines, i.e., from $l = 1$ to $l = \kappa L$, the extra climb is given by solving $\Delta_0^{ls}(\mathbf{r}) = \Delta_{\kappa L}^{ls}(\mathbf{r})$, i.e.,

$$x + d^*(r_1) = b\kappa r_1 + d_{L\kappa+1},$$

The DMT curve $d^*(r_1)$ is given at segment k by

$$d^*(r_1) = -\alpha(r_1 - M) + \sum_{i=1}^{k-1}(M^* - M_* + 2i - 1).$$

Since we have $b\kappa r_1 = x$, then this equation simplifies to

$$d^*\left(\frac{x}{b\kappa}\right) = d_{L\kappa+1}.$$

Therefore, using $c_{k-1} \triangleq b(1-\kappa)\sum_{j=1}^{k-1}\kappa_j$, we solve κ from

$$-\alpha\left(\frac{x}{b\kappa} - M\right) = M\alpha\left(1 - e^{-\frac{b(1-\kappa)-c_{k-1}}{\alpha}}\right)$$

which is given by

$$\kappa^* = \frac{\alpha}{b}\mathcal{W}\left(\frac{e^{\frac{b-c_{k-1}}{\alpha}}x}{M\alpha}\right).$$

The range of validity for each k is given by $c_{k-1} < b(1-\kappa) \leq c_k$. Since for a given c, the solution to $c = b(1-\kappa^*)$ is found as

$$b = \frac{xe^{c_{k-1}-c}}{M} + c,$$

when $c = c_{k-1}$, we have

$$b > \frac{x}{M} + c_{k-1} = c_{k-1} + \frac{x}{M_* - k + 1}.$$

When $c = c_k$, since $c_{k-1} - c_k = \alpha\ln(M/(M_* - k))$ we have

$$b \leq \frac{xe^{c_{k-1}-c_k}}{M} + c_k = c_k + \frac{x}{M_* - k}.$$

Putting all together, we obtain the condition on the theorem and the distortion exponent.

D.6 Distortion exponent achievable by BS-JDS

Here, we derive the distortion exponent for BS-JDS. We consider the usual change of variables, $\lambda_i = \rho^{-\alpha_i}$, and $\gamma = \rho^{-\beta}$. Let r_l be the multiplexing gain of the l-th layer and $\mathbf{r} \triangleq [r_1, ..., r_L]$ such that $R_i = r_i \log \rho$, and define $\bar{r}_1^l \triangleq \sum_{i=1}^{l} r_i$.

First, we derive the outage set $\mathcal{O}_l^{ml}$ at high SNR regime of for each layer, which we call $\mathcal{L}_l$, and is found as

$$\mathcal{L}_l \triangleq \Bigg\{(\boldsymbol{\alpha}, \beta) : b \sum_{i=1}^{M_*} [(\xi_{l-1} - \alpha_i)^+ - (\xi_l - \alpha_i)^+] < \left(\sum_{i=1}^{l} br_i - (x-\beta)\right)^+ - \left(\sum_{i=1}^{l-1} br_i - (x-\beta)\right)^+ \Bigg\}.$$

For the power allocation $\rho_l = \rho^{\xi_{l-1}} - \rho^{\xi_l}$, we have that the l.h.s. of $\mathcal{O}_l^{ml}$ is then given as follows

$$\begin{aligned} I(\mathbf{X}_l; \mathbf{Y}|\mathbf{X}_1^{l-1}) &= I(\mathbf{X}_l^L; \mathbf{Y}|\mathbf{X}_1^{l-1}) - I(\mathbf{X}_{l+1}^L; \mathbf{Y}|\mathbf{X}_1^{l-1}) \\ &= \log \frac{\det\left(\mathbf{I} + \frac{\rho^{\xi_{l-1}}}{M_*}\mathbf{H}\mathbf{H}^H\right)}{\det\left(\mathbf{I} + \frac{\rho^{\xi_l}}{M_*}\mathbf{H}\mathbf{H}^H\right)} \\ &= \log \prod_{i=1}^{M_*} \frac{1 + \frac{\rho^{\xi_{l-1}}}{M_*}\lambda_i}{1 + \frac{\rho^{\xi_l}}{M_*}\lambda_i} \\ &\doteq \rho^{\sum_{i=1}^{M_*}(\xi_{l-1}-\alpha_i)^+ - (\xi_l - \alpha_i)^+}. \end{aligned} \tag{D.36}$$

The r.h.s. of the condition in $\mathcal{O}_l^{ml}$ can be calculated as in (D.29). Then, from (D.36) and (D.29), $\mathcal{L}_l$ follows. Since $\mathcal{O}_l^{ml}$ are mutually exclusive, in the high SNR we have

$$\begin{aligned} ED_{ml}(\mathbf{R}, \boldsymbol{\xi}) &= \sum_{l=0}^{L} \int_{\mathcal{O}_{l+1}^{ml}} D_d\left(\sum_{i=0}^{l} b/2R_i, \gamma\right) p_h(\mathbf{H}) p_\Gamma(\gamma) d\mathbf{H} d\gamma \\ &\doteq \sum_{l=0}^{L} \int_{\mathcal{L}_{l+1}} \rho^{-(\max\{\sum_{i=0}^{l} br_i, (x-\beta)^+\} + \beta + S_A(\boldsymbol{\alpha}))} d\boldsymbol{\alpha} d\beta \\ &\doteq \sum_{l=0}^{L} \rho^{-\Delta_l(\mathbf{r}, \boldsymbol{\xi})} \\ &\doteq \rho^{-\Delta_{ml}^L(\mathbf{r}, \boldsymbol{\xi})}, \end{aligned} \tag{D.37}$$

where applying Varadhan's lemma, the exponent for each integral term is given by

$$\Delta_l^{ml}(\mathbf{r}, \boldsymbol{\xi}) = \inf_{\mathcal{L}_{l+1}} \max\left\{b\bar{r}_0^l, (x-\beta)^+\right\} + \beta + S_A(\boldsymbol{\alpha}). \tag{D.38}$$

Then, the distortion exponent is found as

$$\Delta_{ml}^{L}(b,x)=\max_{\mathbf{r},\boldsymbol{\xi}}\min_{l=0,\dots,L}\left\{\Delta_l^{ml}(\mathbf{r},\boldsymbol{\xi})\right\}. \tag{D.39}$$

Similarly to the DMT, we consider the *successive decoding diversity gain*, defined in [15], as the solution to the probability of outage in the successive decoding at each layer, given by

$$d_{ds}(r_l,\xi_{l-1},\xi_l)\triangleq\inf_{\boldsymbol{\alpha}^+}S_A(\boldsymbol{\alpha})\qquad \text{s.t. } r_l>\sum_{i=1}^{M_*}[(\xi_{l-1}-\alpha_i)^+-(\xi_l-\alpha_i)^+]. \tag{D.40}$$

Without loss in generality, consider the multiplexing gain at layer, r_l, to be given by $r_l=k(\xi_{l-1}-\xi_l)+\delta_l$ where $k\in[0,1,...,M_*-1]$ and $0\leq\delta_l<\xi_{l-1}-\xi_l$. Then, the infimum for (D.40) is found as

$$d_{ds}(r_l,\xi_{l-1},\xi_l)=\Phi_k\xi_{l-1}-\Upsilon_k\delta_l, \tag{D.41}$$

for

$$\alpha_i=\begin{cases}\xi_{l-1}, & 1\leq i<M_*-k,\\ \xi_{l-1}-\delta_l, & i=M_*-k,\\ 0, & M_*-k<i\leq M_*.\end{cases}$$

Now, we solve (D.38) using (D.41) for each layer in function of the power allocation ξ_{l-1} and ξ_l and the rate r_l.

When no layer is successfully decoded, i.e., $l=0$, we have

$$\begin{aligned}\Delta_0^{ml}(\mathbf{r},\boldsymbol{\xi})&=\inf(x-\beta)^++\beta+S_A(\boldsymbol{\alpha})\\ \text{s.t. } b&\sum_{i=1}^{M_*}\left[(\xi_0-\alpha_i)^+-(\xi_1-\alpha_i)^+\right]<(br_1-(x-\beta))^+.\end{aligned}$$

The infimum is achieved by $\beta^*=x$ and using (D.40), we have

$$\Delta_0^{ml}(\mathbf{r},\boldsymbol{\xi})=x+d_{ds}\left(r_1,\xi_0,\xi_1\right).$$

At layer l, the distortion exponent is given by the solution to

$$\begin{aligned}\Delta_l^{ml}(\mathbf{r},\boldsymbol{\xi}) &= \inf\max\{b\bar{r}_1^l,(x-\beta)^+\}+\beta+S_A(\boldsymbol{\alpha})\\ &\text{s.t. } b\sum_{i=1}^{M_*}\left[(\xi_l-\alpha_i)^+-(\xi_{l+1}-\alpha_i)^+\right]<(b\bar{r}_1^{l+1}-x+\beta)^+-(b\bar{r}_1^l-x+\beta)^+.\end{aligned}$$

If $b\bar{r}_1^l \geq x$, the infimum is obtained for $\beta^* = 0$ and

$$\Delta_l^{ml}(\mathbf{r}, \boldsymbol{\xi}) = \inf \max\{b\bar{r}_1^l, x\} + S_A(\boldsymbol{\alpha}) \quad \text{s.t.} \sum_{i^1}^{M_*} \left[(\xi_l - \alpha_i)^+ - (\xi_{l+1} - \alpha_i)^+\right] < r_{k+1}.$$

using (D.40) it solves as

$$\Delta_l^{ml}(\mathbf{r}, \boldsymbol{\xi}) = \max\{x, b\bar{r}_1^l\} + d_{ds}\left(r_{l+1}, \xi_l, \xi_{l+1}\right).$$

If $b\bar{r}_1^l \leq x$, the infimum is given by $\beta^* = (x - b\bar{r}_1^l)^+$ and again, we have a version of (D.40) with the distortion exponent

$$\Delta_l^{ml}(\mathbf{r}, \boldsymbol{\xi}) = x + d_{ds}\left(r_{l+1}, \xi_l, \xi_{l+1}\right).$$

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

$$\begin{aligned}\Delta_L^{ml}(\mathbf{r}, \boldsymbol{\xi}) = &\inf \max\left\{b\bar{r}_1^L, (x-\beta)^+\right\} + \beta + S_A(\boldsymbol{\alpha}) \\ &\text{s.t.} b\sum_{i=1}^{M_*}[(\xi_{L-1} - \alpha_i)^+ - (\xi_L - \alpha_i)^+] \geq \left(b\bar{r}_1^L - (x-\beta)\right)^+ - \left(b\bar{r}_1^{L-1} - (x-\beta)\right)^+.\end{aligned}$$

The infimum is achieved by $\boldsymbol{\alpha}^* = 0$ and $\beta^* = 0$ and given by

$$\Delta_L^{ml}(\mathbf{r}, \boldsymbol{\xi}) = \max\left\{b\bar{r}_1^L, x\right\}, \quad \text{for } r_L \leq M_*(\xi_{L-1} - \xi_L).$$

Note that the condition on r_L always holds.

Gathering all results, the exponent distortion problem in (D.39) is found as the minimum of each layer exponent $\Delta_l^{ml}(\mathbf{r}, \boldsymbol{\xi})$, which can be formulated as

$$\begin{aligned}\Delta_{ml}^L(b, x) = &\max_{\mathbf{r}, \boldsymbol{\xi}} t \\ \text{s.t.} \quad & t \leq x + d_{sd}\left(r_1, \xi_0, \xi_1\right), \\ & t \leq \max\{b\bar{r}_1^l, x\} + d_{sd}\left(r_{l+1}, \xi_l, \xi_{l+1}\right), \\ & \quad \text{for } l = 1 \ldots L-1, \\ & t \leq \max\{b\bar{r}_1^L, x\}.\end{aligned} \tag{D.42}$$

If $x \geq b\bar{r}_1^L$, then $\max\{x, b\bar{r}_1^l\} = x$ for all l and the minimum exponent is given by $\Delta_L^{ml}(\mathbf{r}, \boldsymbol{\xi}) = x$, which implies $\Delta_{mj}^L(b, x) = x$. If $x \leq br_1$, then $\max\{x, b\bar{r}_1^l\} = b\bar{r}_1^l$ for all l. In general, if $b\bar{r}_1^q < x \leq b\bar{r}_1^{q+1}$, $q = 0, ..., L$ and $\bar{r}_1^0 \triangleq 0$, $\bar{r}_1^{L+1} \triangleq \infty$, then (D.42) can be formulated, using $r_l = k(\xi_{l-1} - \xi_l) + \delta_l$, $\boldsymbol{\delta} \triangleq [\delta_1, \cdots, \delta_L]$ and $\boldsymbol{\xi}$, as the following linear optimization program:

$$
\begin{aligned}
\Delta_{mj}^{L}(b,x) = & \min_{\substack{1\leq q\leq L,\\ 0\leq k\leq M_*-1.}} \min_{\boldsymbol{\delta},\boldsymbol{\xi}} -t \\
\text{s.t.} \quad & t \leq x + \Phi_k \xi_0 - \Upsilon_k \delta_1, \\
& t \leq x + \Phi_k \xi_l - \Upsilon_k \delta_{l+1}, \quad \text{for } \ l = 1,\ldots,q, \\
& t \leq b \sum_{i=1}^{l} [k(\xi_{i-1} - \xi_i) + \delta_i] + \Phi_k \xi_l - \Upsilon_k \delta_{l+1}, \quad \text{for } \ l = q,\ldots,L-1, \\
& t \leq b \sum_{i=1}^{L} [k(\xi_{i-1} - \xi_i) + \delta_i], \\
& 0 \leq \delta_l < \xi_{l-1} - \xi_l, \quad \text{for } \ l = 1,\ldots,L, \\
& 0 \leq \xi_L \leq ... \leq \xi_1 \leq \xi_0 = 1, \\
& \sum_{l=1}^{l'} [bk(\xi_{l-1} - \xi_l) + \delta_l] < x.
\end{aligned} \tag{D.43}
$$

The linear program (D.43) can be efficiently solved using numerical methods. In Figure 4.4, the numerical solution is shown. However, in the following we provide an analytical result by fixing the multiplexing gains $\mathbf{r}$. We fix the multiplexing gains r_l to $\hat{r}_l$ by fixing $\delta_l \triangleq (\xi_{l-1} - \xi_l) - \epsilon_1$, such that $\hat{r}_l = [(k+1)(\xi_{l-1} - \xi_l) - \epsilon_1]$, $\epsilon_1 > 0$ for $k = 0,...,M_* - 1$ in the bandwidth regime

$$
b \in \left[\frac{\Phi_{k+1} + x}{k+1}, \frac{\Phi_k + x}{k} \right). \tag{D.44}
$$

Note that by fixing $\hat{r}_l$, the solution might be suboptimal. In fact, single layer JDS transmission in Section 4.5.2, is excluded from the set of feasible solutions.

Assume $br_1 \geq x$. Then, each distortion exponent is found as

$$
\begin{aligned}
& \Delta_0^{ml}(\mathbf{r},\boldsymbol{\xi}) = x + \Phi_k \xi_l - \Upsilon_k \delta_{l+1}, \\
& \Delta_l^{ml}(\mathbf{r},\boldsymbol{\xi}) = b\bar{r}_1^l + \Phi_k \xi_l - \Upsilon_k \delta_{l+1}, \quad \text{for } l = 1,...,L-1, \\
& \Delta_L^{ml}(\mathbf{r},\boldsymbol{\xi}) = b\bar{r}_1^L.
\end{aligned} \tag{D.45}
$$

Similarly to the other schemes, for which the distortion exponent is maximized by equaling the exponents, we look for the power allocation $\boldsymbol{\xi}$, such that all distortion exponent terms $\Delta_l^{ml}(\hat{\mathbf{r}},\boldsymbol{\xi})$ in (D.39) are equal.

Equaling all distortion exponents $\Delta_l^{ml}(\hat{\mathbf{r}},\boldsymbol{\xi})$ for $l = 2,...,L-1$, i.e., $\Delta_{l-1}^{ml}(\hat{\mathbf{r}},\boldsymbol{\xi}) = \Delta_l^{ml}(\hat{\mathbf{r}},\boldsymbol{\xi})$, we have

$$
d_{sd}(\hat{r}_l, \xi_{l-1}, \xi_l) = br_l + d_{sd}(\hat{r}_{l+1}, \xi_l, \xi_{l+1}). \tag{D.46}
$$

Since $\hat{r}_l = [(k+1)(\xi_{l-1} - \xi_l) - \epsilon_1]$, we have

$$d_{sd}\left(\hat{r}_l, \xi_{l-1}, \xi_l\right) = \Phi_k \xi_{l-1} - \Upsilon_k(\xi_{l-1} - \xi_l - \epsilon_1).$$

Substituting in (D.46) we obtain that the power allocations for $l \geq 2$ need to satisfy,

$$(\xi_l - \xi_{l+1}) = \eta_k(\xi_{l-1} - \xi_l) + \pi(\epsilon_1),$$

where η_k is defined in equation (4.20) and $\mathcal{O}(\epsilon_1) \to 0$ for $\epsilon_1 \to 0$.

Then, for $l = 2, ..., L-1$ we have the following recursion for each power allocation in terms of $(\xi_1 - \xi_2)$:

$$\xi_l - \xi_{l+1} = \eta_k^{l-1}(\xi_1 - \xi_2) + \pi(\epsilon_1), \tag{D.47}$$

and each power allocations ξ_l can be found as

$$\begin{aligned}
1 - \xi_l &= (1 - \xi_1) + \sum_{i=1}^{l-1}(\xi_i - \xi_{i+1}) \\
&= (1 - \xi_1) + \sum_{i=1}^{l-1} \eta_k^{i-1}(\xi_1 - \xi_2) + \pi(\epsilon_1) \\
&= (1 - \xi_1) + (\xi_1 - \xi_2)\frac{1 - \eta_k^{l-1}}{1 - \eta_k} + \pi(\epsilon_1).
\end{aligned}$$

From $\Delta_L^{ml}(\hat{\mathbf{r}}, \boldsymbol{\xi}) = b\bar{r}_1^L = b\sum_{i=1}^{L}(k+1)(\xi_{i-1} - \xi_i)$, we have

$$\begin{aligned}
\Delta_L^{ml}(\hat{\mathbf{r}}, \boldsymbol{\xi}) &= b(k+1)(\xi_0 - \xi_1) + b(k+1)(\xi_2 - \xi_1)\sum_{i=1}^{L} \eta_k^{i-1} + \pi(\epsilon_1) \\
&= b(k+1)\left[(\xi_0 - \xi_1) + (\xi_2 - \xi_1)\frac{1 - \eta_k^{L-1}}{1 - \eta_k}\right] + \pi(\epsilon_1).
\end{aligned} \tag{D.48}$$

Putting all together, from (D.45) we obtain the following determined linear system

$$\begin{aligned}
\Delta_0^{ml}(\hat{\mathbf{r}}, \boldsymbol{\xi}) &= x + \Phi_k \xi_0 - \Upsilon_k(\xi_0 - \xi_1 - \epsilon_1), \\
\Delta_1^{ml}(\hat{\mathbf{r}}, \boldsymbol{\xi}) &= b(k+1)(\xi_0 - \xi_1) + \Phi_k \xi_1 - \Upsilon_k(\xi_1 - \xi_2 + \epsilon_1), \\
\Delta_L^{ml}(\hat{\mathbf{r}}, \boldsymbol{\xi}) &= b(k+1)[(\xi_0 - \xi_1) + (\xi_2 - \xi_1)\Gamma_k] + \pi(\epsilon_1).
\end{aligned} \tag{D.49}$$

By solving $\Delta_{ml}^L(b, x) = \Delta_0^{ml}(\hat{\mathbf{r}}, \boldsymbol{\xi}) = \Delta_1^{ml}(\hat{\mathbf{r}}, \boldsymbol{\xi}) = \Delta_L^{ml}(\hat{\mathbf{r}}, \boldsymbol{\xi})$, the solution to the linear systems and letting $\epsilon_1 \to 0$ is given in (4.21), (4.22) and (4.23). So that this solution is feasible, the power allocation sequence has to satisfy $1 \geq \xi_1 \geq ...\xi_L \geq 0$, i.e., $\xi_l - \xi_{l+1} \geq 0$. From (D.47) we need $\eta_k \geq 0$ and $\xi_1 - \xi_2 \geq 0$. We have $\eta_k \geq 0$ if $b \geq \frac{\Phi_{k+1}}{k+1}$, which holds in the regime given by (D.44). Then, $\xi_1 - \xi_2 \geq 0$ holds if $\Upsilon_k + b(k+1) - \Phi_k - x \geq 0$ and $(\Upsilon_k + b(1+k))(\Upsilon_k + b(1+k)\Gamma_k) - b(k+1)\Phi_k\Gamma_k \geq 0$. It can be shown that $(\Upsilon_k + b(1+k))(\Upsilon_k + b(1+k)\Gamma_k) - b(k+1)\Phi_k\Gamma_k$ is monotonically increasing in $b \geq 0$ and to be positive for $k = 0, ..., M_* - 1$. Therefore, we need to check $\Upsilon_k + b(k+1) - \Phi_k - x \geq 0$, which holds since this condition is equivalent to

$$b \geq \frac{\Phi_{k+1} + x}{k+1}.$$

Note that in this regime, we have $\xi_1 \geq 0$. In addition, $\xi_l = \xi_1 + (\xi_1 - \xi_2)\Gamma_k \geq 0$. Therefore, for each k the power allocation is feasible in the regime given by (D.44). It can also be checked that $br_1 > x$ is satisfied. This completes the proof.

D.6.1 Convergence for $L \to \infty$.

In the continuum infinity of layers, i.e., by letting $L \to \infty$, this scheme converges to $\Delta_{ml}^{\infty}(b, x) = \max\{x, b(k+1)\}$ when $0 \leq \eta_k < 1$, i.e.,

$$b \in \left[\frac{\Phi_{k+1} + x}{k+1}, \frac{\Phi_k}{k+1}\right),$$

and it converges to

$$\Delta_{ml}^{\infty}(b, x) = \Phi_k + x\left(\frac{b(k+1) - \Phi_k}{b(k+1) - \Phi_{k+1}}\right),$$

when $\eta_k \geq 1$, that is, for

$$b \in \left[\frac{\Phi_k}{k+1}, \frac{\Phi_k + x}{k}\right).$$

Appendix E

Proofs for Chapter 5

E.1 Proof of Lemma 20

In the rate expression and joint pmf in Theorem 15, we set $X^n = (X_1^n, X_2^n)$, $Y^n = (Y_1^n, Y_2^n, Z^n)$, $V = \emptyset$, and generate X_R^n and X_1^n independent of the rest of the random variables with distributions $p^*(x_R)$ and $p^*(x_1)$, which maximize the mutual information terms in (5.5), respectively. Under this set of distributions we have

$$
\begin{aligned}
I(X;Y\hat{Y}_R|X_RU) &= I(X_1X_2;Y_1Y_2\hat{Y}_RZ|X_R,U) \\
&\overset{(a)}{=} I(X_1X_2;Y_2\hat{Y}_R|X_RUZ) \\
&\overset{(b)}{=} I(X_2;Y_2|Z) + I(X_1;\hat{Y}_R|UZ) \\
&= R_1 + I(X_1;\hat{Y}_R|UZ), \\
I(U;Y_R|X_RV) &= I(U;Y_R|X_R) \overset{(c)}{=} I(U;Y_R), \\
I(XX_R;Y) &= I(X_1X_2X_R;Y_1Y_2Z) \\
&\overset{(d)}{=} I(X_2X_R;Y_1Y_2|Z) \\
&\overset{(e)}{=} I(X_R;Y_1) + I(X_2;Y_2|Z) = R_0 + R_1, \\
I(\hat{Y}_R;Y_R|XX_RUY) &= I(\hat{Y}_R;Y_R|X_RX_1X_2UY_1Y_2Z) \\
&\overset{(f)}{=} I(\hat{Y}_R;Y_R|X_RX_1X_2UY_2Z) \\
&\overset{(g)}{=} I(\hat{Y}_R;Y_R|X_1UZ), \\
I(\hat{Y}_R;Y_R|YX_RU) &= I(\hat{Y}_R;Y_R|Y_1Y_2ZX_RU) \overset{(h)}{=} I(\hat{Y}_R;Y_R|UZ), \\
I(X_R;Y|V) &= I(X_R;Y_1Y_2Z) = I(X_R;Y_1) = R_0,
\end{aligned}
$$

where (a) is due to the Markov chain $(X_1X_2) - X_R - Y_1$; $(b), (c), (e), (f), (g), (h)$ are due to the independence of (U, X_1) and X_R, and (d) is due to the Markov chain $(Y_1Y_2) - (X_2X_RZ) - X_1$.

Then, (5.8) reduces to the following rate

$$\begin{aligned} R = &\sup_{\mathcal{P}} \min\{I(U;Y_R) + R_1 + I(X_1;\hat{Y}_R|UZ), R_1 + R_0 - I(\hat{Y}_R;Y_R|X_1UZ)\}, \\ &\text{s.t. } R_0 \geq I(\hat{Y}_R;Y_R|UZ). \end{aligned} \tag{E.1}$$

By denoting by the joint distributions in $\mathcal{P}$ such the the minimum in R is achieved for the first argument, i.e.,

$$R_0 - I(\hat{Y}_R;Y_R|X_1UZ) \geq I(U;Y_R) + I(X_1;\hat{Y}_R|UZ), \tag{E.2}$$

and arranging using the chain rule for the mutual information, we have that the rate achievable by pDCF is lower bounded by

$$\begin{aligned} R \geq &\sup_{\mathcal{P}} R_1 + I(U;Y_R) + I(X_1;\hat{Y}_R|UZ) \\ &\text{s.t. } R_0 \geq I(U;Y_R) + I(X_1Y_R;\hat{Y}_R|UZ), && \text{(E.3)} \\ &\qquad R_0 \geq I(\hat{Y}_R;Y_R|UZ). && \text{(E.4)} \end{aligned}$$

From (E.3), we have

$$\begin{aligned} R_0 &\geq I(U;Y_R) + I(X_1Y_R;\hat{Y}_R|UZ) \\ &\overset{(a)}{=} I(U;Y_R) + I(\hat{Y}_R;Y_R|UZ) \\ &\geq I(\hat{Y}_R;Y_R|UZ), \end{aligned} \tag{E.5}$$

where (a) is due to the Markov chain $\hat{Y}_R - (UY_R) - (X_1Z)$. Hence, (E.3) implies (E.4), i.e., the latter condition is redundant, and $R \geq \mathcal{C}$. Therefore the capacity expression $\mathcal{C}$ in (5.7) is achievable by pDCF. This concludes the proof.

E.2 Proof of Lemma 21

Consider any sequence of $(2^{nR}, n, \nu_n)$ codes such that $\lim_{n\to\infty} \nu_n \to 0$. We need to show that $R \leq R_{up}$.

Let us define $U_i \triangleq (Y_{R1}^{i-1} X_{1i+1}^n Z^{n\backslash i})$ and $\hat{Y}_{Ri} \triangleq (Y_{1i+1}^n)$. For such $\hat{Y}_{Ri}$ and U_i, the following Markov chain holds

$$\hat{Y}_{Ri} - (U_i, Y_{Ri}) - (X_{1i}, X_{2i}, Z_i, Y_{1i}, Y_{2i}, X_{Ri}). \tag{E.6}$$

From Fano's inequality, we have

$$H(W|Y_1^n Y_2^n Z^n) \leq n\epsilon_n, \tag{E.7}$$

such that $\epsilon_n \to 0$ as $n \to \infty$.

First, we derive the following set of inequalities related to the capacity of the source-destination channel.

$$\begin{aligned} nR &= H(W) \\ &\overset{(a)}{=} I(W; Y_1^n Y_2^n | Z^n) + H(W | Y_1^n Y_2^n Z^n) \\ &\overset{(b)}{\leq} I(X_1^n X_2^n; Y_1^n Y_2^n | Z^n) + n\epsilon_n, \end{aligned} \tag{E.8}$$

where (a) follows from the independence of Z^n and W and (b) follows from Fano's inequality in (E.7).

We also have the following inequalities:

$$\begin{aligned} I(X_2^n; Y_2^n | Z^n) &= \sum_{i=1}^{n} H(Y_{2i} | Z^n Y_{21}^{i-1}) - H(Y_{2i} | Z^n Y_{21}^{i-1} X_2^n) \\ &\overset{(a)}{\leq} \sum_{i=1}^{n} H(Y_{2i} | Z_i) - H(Y_{2i} | Z_i X_{2i}) \\ &= \sum_{i=1}^{n} I(X_{2i}; Y_{2i} | Z_i) \\ &\overset{(b)}{=} nI(X_{2Q'}; Y_{2Q'} | Q') \\ &\overset{(c)}{\leq} nI(X_{2Q'}; Y_{2Q'}) \\ &\overset{(d)}{\leq} nR_1, \end{aligned} \tag{E.9}$$

where (a) follows since conditioning reduces entropy, (b) follows by defining Q' as a uniformly distributed random variable over $\{1, ..., n\}$ and $(X_{2Q'}, Y_{2Q'})$ as a pair of random variables satisfying $\Pr\{X_{2i} = x_2, Y_{2i} = y_2\} = \Pr\{X_{2Q'} = x_2, Y_{2Q'} = y_2 | Q = i\}$ for $i = 1, ..., n$, (c) follows from the Markov chain relation $Q' - X_{2Q'} - Y_{2Q'}$ and (d) follows from the definition of R_1 in (5.5).

Then, we can bound the achievable rate as,

$$\begin{aligned} nR &= I(W; Y_1^n Y_2^n Z^n) + H(W | Y_1^n Y_2^n Z^n) \\ &\overset{(a)}{\leq} I(W; Y_1^n Y_2^n Z^n) + n\epsilon_n \\ &\overset{(b)}{=} I(W; Y_2^n | Z^n) + I(W; Y_1^n | Y_2^n Z^n) + n\epsilon_n \\ &\overset{(c)}{\leq} I(X_2^n; Y_2^n | Z^n) + I(W; Y_1^n | Y_2^n Z^n) + n\epsilon_n \\ &\overset{(d)}{\leq} nR_1 + H(Y_1^n | Y_2^n Z^n) - H(Y_1^n | W Z^n) + n\epsilon_n \\ &\overset{(e)}{\leq} nR_1 + H(Y_1^n | Z^n) - H(Y_1^n | W X_1^n Z^n) + n\epsilon_n \end{aligned}$$

$$
\begin{aligned}
&\overset{(f)}{=} nR_1 + I(X_1^n;Y_1^n|Z^n) + n\epsilon_n \\
&\overset{(g)}{\leq} nR_1 + H(X_1^n) - H(X_1^n|Y_1^nZ^n) + n\epsilon_n \\
&= nR_1 + \sum_{i=1}^n H(X_{1i}|X_{1i+1}^n) - H(X_1^n|Y_1^nZ^n) + n\epsilon_n \\
&\overset{(h)}{\leq} nR_1 + \sum_{i=1}^n \Big[I(Y_{R1}^{i-1}Z^{n\setminus i};Y_{Ri}) + H(X_{1i}|X_{1i+1}^n) \Big] - H(X_1^n|Y_1^nZ^n) + n\epsilon_n \\
&= nR_1 + \sum_{i=1}^n \Big[I(Y_{R1}^{i-1}Z^{n\setminus i}X_{1i+1}^n;Y_{Ri}) - I(X_{1i+1}^n;Y_{Ri}|Y_{R1}^{i-1}Z^{n\setminus i}) \\
&\quad + H(X_{1i}|Y_{R1}^{i-1}Z^{n\setminus i}X_{1i+1}^n) + I(X_{1i};Y_{R1}^{i-1}Z^{n\setminus i}|X_{1i+1}^n) \Big] - H(X_1^n|Y_1^nZ^n) + n\epsilon_n \\
&\overset{(i)}{=} nR_1 + \sum_{i=1}^n \Big[I(Y_{R1}^{i-1}Z^{n\setminus i}X_{1i+1}^n;Y_{Ri}) + H(X_{1i}|Y_{R1}^{i-1}Z^{n\setminus i}X_{1i+1}^n) \Big] \\
&\quad - H(X_1^n|Y_1^nZ^n) + n\epsilon_n \\
&= nR_1 + \sum_{i=1}^n \Big[I(U_i;Y_{Ri}) + H(X_{1i}|U_i) \Big] - H(X_1^n|Y_1^nZ^n) + n\epsilon_n \\
&\overset{(j)}{\leq} nR_1 + \sum_{i=1}^n \Big[I(U_i;Y_{Ri}) + H(X_{1i}|U_i) - H(X_{1i}|U_iZ_i\hat{Y}_{Ri}) \Big] + n\epsilon_n \\
&\overset{(k)}{=} nR_1 + \sum_{i=1}^n \Big[I(U_i;Y_{Ri}) + I(X_{1i};\hat{Y}_{Ri}|U_iZ_i) \Big] + n\epsilon_n.
\end{aligned}
$$

where (a) is due to Fano's inequality; (b) is due to the chain rule and the independence of Z^n from W; (c) is due to the data processing inequality, (d) is due to the Markov chain relation $Y_2^n - (W, Z^n) - Y_1^n$ and (E.9), (e) is due to the fact that conditioning reduces entropy, and that X_1^n is a deterministic function of W; (f) is due to the Markov chain relation $Y_1^n - X_1^n - W$; (g) is due to the independence of Z^n and X_1^n; (i) follows because

$$
\begin{aligned}
\sum_{i=1}^n I(X_{1i};Y_{R1}^{i-1}Z^{n\setminus i}|X_{1i+1}^n) &\overset{(l)}{=} \sum_{i=1}^n I(X_{1i};Y_{R1}^{i-1}|X_{1i+1}^nZ^{n\setminus i}) \\
&\overset{(m)}{=} \sum_{i=1}^n I(X_{1i+1}^n;Y_{Ri}|Y_{R1}^{i-1}Z^{n\setminus i}),
\end{aligned} \tag{E.10}
$$

where (l) is due to the independence of Z^n and X_1^n; and (m) is the conditional version of Csiszár's equality [25]. Inequality (j) is due to the following bound,

$$
\begin{aligned}
H(X_1^n|Y_1^nZ^n) &= \sum_{i=1}^n H(X_{1i}|X_{1i+1}^nZ^nY_1^n) \\
&\geq \sum_{i=1}^n H(X_{1i}|Y_{R1}^{i-1}X_{1i+1}^nZ^nY_1^n)
\end{aligned}
$$

$$
\begin{aligned}
&\stackrel{(n)}{=} \sum_{i=1}^{n} H(X_{1i}|Y_{R1}^{i-1}X_{1i+1}^{n}Z^{n}Y_{1i+1}^{n}) \\
&= \sum_{i=1}^{n} H(X_{1i}|U_i Z_i \hat{Y}_{Ri}), \qquad \text{(E.11)}
\end{aligned}
$$

where (n) is follows from the Markov chain relation $X_{1i} - (Y_{R1}^{i-1}X_{1i+1}^{n}Z^{n}Y_{1i+1}^{n}) - Y_{11}^{i}$, and noticing that $X_{Ri} = f_{r,i}(Y_{R1}^{i-1})$. Finally, (k) is due to the fact that Z_i independent of (X_{1i}, U_i).

We can also obtain the following sequence of inequalities

$$
\begin{aligned}
nR_0 + nR_1 &\stackrel{(a)}{\geq} I(X_R^n;Y_1^n) + I(X_2^n;Y_2^n|Z^n) \\
&\stackrel{(b)}{\geq} H(Y_2^n|Z^n) - H(Y_2^n|X_2^nZ^n) + H(Y_1^n|Y_2^nZ^n) - H(Y_1^n|X_R^n) \\
&= H(Y_1^nY_2^n|Z^n) - H(Y_2^n|X_2^nZ^n) - H(Y_1^n|X_R^n) \\
&\stackrel{(c)}{=} H(Y_1^nY_2^n|Z^n) - H(Y_2^n|X_1^nX_2^nY_R^nZ^n) - H(Y_1^n|X_R^nX_1^nX_2^nY_R^nY_2^nZ^n) \\
&\stackrel{(d)}{=} H(Y_1^nY_2^n|Z^n) - H(Y_1^nY_2^n|X_1^nX_2^nY_R^nZ^n) \\
&= I(Y_1^nY_2^n;X_1^nX_2^nY_R^n|Z^n) \\
&= I(X_1^nX_2^n;Y_1^nY_2^n|Z^n) + I(Y_1^nY_2^n;Y_R^n|X_1^nX_2^nZ^n) \\
&\stackrel{(e)}{\geq} nR + I(Y_1^nY_2^n;Y_R^n|X_1^nX_2^nZ^n) - n\epsilon_n \\
&\stackrel{(f)}{=} nR + I(Y_1^n;Y_R^n|X_1^nZ^n) - n\epsilon_n \\
&= nR + \sum_{i=1} I(Y_1^n;Y_{Ri}|X_1^nY_{R1}^{i-1}Z^n) - n\epsilon_n \\
&\geq nR + \sum_{i=1} I(Y_{1i+1}^n;Y_{Ri}|X_1^nY_{R1}^{i-1}Z^n) - n\epsilon_n \\
&\stackrel{(g)}{\geq} nR + \sum_{i=1} I(Y_{1i+1}^n;Y_{Ri}|X_{1i}^nY_{R1}^{i-1}Z^n) - n\epsilon_n \\
&= nR + \sum_{i=1} I(\hat{Y}_{Ri};Y_{Ri}|X_{1i}U_iZ_i) - n\epsilon_n,
\end{aligned}
$$

where (a) follows from the definitions of R_0 and R_1 in (5.5); (b) is due to the fact that conditioning reduces entropy; (c) is due to the Markov chains $Y_2^n - (X_2^nZ^n) - (X_1^nY_R^n)$ and $Y_1^n - X_R^n - (X_1^nX_2^nY_R^nY_2^nZ^n)$; (d) is follows since X_R^n is a deterministic function of Y_R^n; (e) is due to the expression in (E.8); (f) is due to the Markov chain $(Y_R^nY_1^n) - (X_1^nZ^n) - (X_2^nY_2^n)$ and; (g) is due to the Markov chain $(Y_{1i+1}^n) - (X_{1i}^nY_{R1}^{i-1}Z^n) - X_{11}^{i-1}$.

A single letter expression can be obtained by using the usual time-sharing random variable arguments. Let Q be a time sharing random variable uniformly distributed over $\{1, ..., n\}$, independent of all the other random variables. Also, define a set of random variables $(X_{1Q}, Y_{RQ}, U_Q, \hat{Y}_{RQ}, Z_Q)$ satisfying

$$\begin{aligned}
&\Pr\{X_{1Q} = x_1, Y_{RQ} = y_R, U_Q = u, \hat{Y}_{RQ} = \hat{y}_R, Z_Q = z|Q = i\} \\
&= \Pr\{X_{1i} = x_1, Y_{Ri} = y_R, U_i = u, \hat{Y}_{Ri} = \hat{y}_D, Z_i = z\} \qquad \text{for } i = 1, ..., n.
\end{aligned}$$

Define $U = (U_Q, Q)$, $\hat{Y}_R = \hat{Y}_{RQ}$, $X_1 = X_{1Q}$, $Y_{RQ} = Y_R$ and $Z = Z_Q$. We note that the pmf of the tuple $(X_1, Y_R, U, \hat{Y}_R, Z)$ belongs to $\mathcal{P}$ in (5.6) as follows:

$$\begin{aligned}
p(u, x_1, y_R, z, \hat{y}_R) &= p(q, u_Q, x_{1Q}, y_{RQ}, z_Q, \hat{y}_{RQ}) \\
&= p(q, u_Q, x_{1Q})p(z_Q y_{RQ} \hat{y}_{RQ}|q, u_Q x_{1Q}) \\
&= p(q, u_Q, x_{1Q})p(z_Q|q, u_Q, x_{1Q})p(y_{RQ}, \hat{y}_{RQ}|q, u_Q, x_{1Q}, z_Q) \\
&\overset{(a)}{=} p(q, u_Q, x_{1Q})p(z)p(y_{RQ}|q, u_Q, x_{1Q}, z_Q)p(\hat{y}_{RQ}|q, u_Q, x_{1Q}, z_Q, y_{RQ}) \\
&\overset{(b)}{=} p(q, u_Q, x_{1Q})p(z)p(y_R|x_1, z)p(\hat{y}_{RQ}|q, u_Q, x_{1Q}, z_Q, y_{RQ}) \\
&\overset{(c)}{=} p(q, u_Q, x_{1Q})p(z)p(y_R|x_1, z)p(\hat{y}_{RQ}|q, u_Q, y_{RQ}) \\
&= p(u, x_1)p(z)p(y_R|x_1, z)p(\hat{y}_R|u, y_R),
\end{aligned}$$

where (a) follows since the channel state Z^n is i.i.d. and thus $p(z_Q|q, u_Q, x_{1Q}) = p(z_Q|q) = p(z)$, (b) follows since $p(y_{RQ}|q, u_Q, x_{1Q}, z_Q) = p(y_{RQ}|q, x_{1Q}, z_Q) = p(y_R|x_1, z)$, (c) follows from the Markov chain in (E.6).

Then, we get the single letter expression,

$$\begin{aligned}
R &\le R_1 + \frac{1}{n}\sum_{i=1}^{n}[I(U_i; Y_{Ri}) + I(X_{1i}; \hat{Y}_{Ri}|U_i Z_i)] + \epsilon_n \\
&= R_1 + I(U_Q; Y_{RQ}|Q) + I(X_{1Q}; \hat{Y}_{RQ}|U_Q Z_Q Q) + \epsilon_n \\
&\le R_1 + I(U_Q Q; Y_{RQ}) + I(X_{1Q}; \hat{Y}_{RQ} Q|U_Q Z_Q) + \epsilon_n \\
&= R_1 + I(U; Y_R) + I(X_1; \hat{Y}_R|UZ) + \epsilon_n,
\end{aligned}$$

and

$$\begin{aligned}
R_0 + R_1 &\ge R + \frac{1}{n}\sum_{i=1} I(\hat{Y}_{Ri}; Y_{Ri}|X_{1i} U_i Z_i) - n\epsilon_n \\
&= R + I(\hat{Y}_{RQ}; Y_{RQ}|X_{1Q} U_Q Z_Q Q) - n\epsilon_n \\
&= R + I(\hat{Y}_R; Y_R|X_1 U Z) - n\epsilon_n.
\end{aligned}$$

The cardinality of the bounds on the alphabets of U and $\hat{Y}_R$ can be found using the usual techniques [25]. This completes the proof.

E.3 Proof of Lemma 22

Now, we will show that the expression of R_{up} in (5.9) is equivalent to the expression $\mathcal{C}$ in (5.7). First we will show that $\mathcal{C} \leq R_{up}$ as follows. Consider the subset of pmfs in $\mathcal{P}$ such that

$$R_0 + R_1 - I(\hat{Y}_R; Y_R|X_1UZ) \geq R_1 + I(U; Y_R) + I(X_1; \hat{Y}_R|UZ) \tag{E.12}$$

holds. Then, similarly to (E.5) in Appendix E this condition is necessitates

$$R_0 \geq I(U; Y_R) + I(Y_R; \hat{Y}_R|UZ). \tag{E.13}$$

Hence, we have $\mathcal{C} \leq R_{up}$.

Then, it remains to show that $\mathcal{C} \geq R_{up}$. As R_1 can be extracted from the supremum, it is enough to show that, for each $(X_1, U, Z, Y_R, \hat{Y}_R)$ tuple with a joint pmf $p_e \in \mathcal{P}$ satisfying

$$\begin{aligned} &R(p_e) \leq I(U; Y_R) + I(X_1; \hat{Y}_R|UZ), \\ &\text{where } R(p_e) \triangleq R_0 - I(\hat{Y}_R; Y_R|X_1UZ), \end{aligned} \tag{E.14}$$

there exist random variables $(X_1^*, U^*, Z, Y_R^*, \hat{Y}_R^*)$ with joint pmf $p_e^* \in \mathcal{P}$ that satisfy

$$\begin{aligned} &R(p_e) = I(U^*; Y_R) + I(X_1^*; \hat{Y}_R^*|U^*Z) \text{ and} \\ &R(p_e) \leq R_0 - I(\hat{Y}_R^*; Y_R|X_1^*U^*Z). \end{aligned} \tag{E.15}$$

This argument is proven next.

Let B denote a Bernoulli random variable with parameter $\lambda \in [0, 1]$, i.e., $B = 1$ with probability λ, and $B = 0$ with probability $1 - \lambda$. We define the triplets of random variables:

$$(U^{'}, X_1^{'}, \hat{Y}_R^{'}) = \begin{cases} (U, X_1, \hat{Y}_R) & \text{if } B = 1, \\ (X_1, X_1, \emptyset) & \text{if } B = 0, \end{cases} \tag{E.16}$$

and

$$(U^{''}, X_1^{''}, \hat{Y}_R^{''}) = \begin{cases} (X_1, X_1, \emptyset) & \text{if } B = 1, \\ (\emptyset, \emptyset, \emptyset) & \text{if } B = 0. \end{cases} \tag{E.17}$$

We first consider the case $R(p_e) > I(X_1; Y_R)$. Let $U^* = (U^{'}, B)$, $X_1^* = X_1^{'}$, $\hat{Y}_R^* =$

$(\hat{Y}_R^{'}, B)$. For $\lambda = 1$,

$$I(U^*; Y_R) + I(X_1^*; \hat{Y}_R^* | U^* Z) = I(U; Y_R) + I(X_1; \hat{Y}_R | UZ) > R(p_e),$$

and for $\lambda = 0$,

$$I(U^*; Y_R) + I(X_1^*; \hat{Y}_R^* | U^* Z) = I(X_1; Y_R) < R(p_e).$$

As $I(U^*; Y_R) + I(X_1^*; \hat{Y}_R^* | U^* Z)$ is a continuous function of λ, by the intermediate value theorem, there exists a $\lambda \in [0, 1]$ such that $I(U^*; Y_R) + I(X_1^*; \hat{Y}_R^* | U^* Z) = R(p_e)$. We denote the corresponding joint distribution by p_e^*.

We have

$$\begin{aligned} I(\hat{Y}_R^*; Y_R | X_1^* U^* Z) &= I(\hat{Y}_R^{'}; Y_R | X_1^{'} U^{'} Z B) \\ &= \lambda I(\hat{Y}_R; Y_R | X_1 U Z) \\ &\leq I(\hat{Y}_R; Y_R | X_1 U Z), \end{aligned} \tag{E.18}$$

which implies that p_e^* satisfies (E.15) since

$$\begin{aligned} R(p_e) &= R_0 - I(\hat{Y}_R; Y_R | X_1 U Z) \\ &\leq R_0 - I(\hat{Y}_R^*; Y_R | X_1^* U^* Z). \end{aligned} \tag{E.19}$$

Next we consider the case $R(p_e) \leq I(X_R; Y_1)$. We define $U^* = (U^{''}, B)$, $X_1^* = X_1^{''}$ and $\hat{Y}_R^* = (\hat{Y}_R^{''}, B)$. Then, for $\lambda = 1$,

$$I(U^*; Y_R) + I(X_1^*; \hat{Y}_R^* | U^* Z) = I(X_1; Y_R) \geq R(p_e),$$

and for $\lambda = 0$,

$$I(U^*; Y_R) + I(X_1^*; \hat{Y}_R^* | U^* Z) = 0 < R(p_e). \tag{E.20}$$

Once again, as $I(U^*; Y_R) + I(X_1^*; \hat{Y}_R^* | U^* Z)$ is a continuous function of λ, by the intermediate value theorem, there exists a $\lambda \in [0, 1]$ such that $I(U^*; Y_R) + I(X_1^*; \hat{Y}_R^* | U^* Z) = R(p_e)$. Again, we denote this joint distribution by p_e^*. On the other hand, we have $I(\hat{Y}_R^*; Y_R | X_1^* U^* Z) = 0$, which implies that

$$\begin{aligned} R(p_e) &= R_0 - I(\hat{Y}_R; Y_R | X_1 U Z) \\ &\leq R_0 \\ &= R_0 - I(\hat{Y}_R^*; Y_R | X_1^* U^* Z). \end{aligned} \tag{E.21}$$

That is, p_e^* also satisfies (E.15).

We have shown that for any joint pmf $p_e \in \mathcal{P}$ satisfying ($E.14$), there exist another joint pmf, p_e^*, that satisfies ($E.15$). For a distribution satisfying ($E.15$) we can write

$$\begin{aligned}R_0 &> I(U^*;Y_R) + I(X_1^*;\hat{Y}_R^*|U^*Z) + I(\hat{Y}_R^*;Y_R|X_1^*U^*Z)\\ &= I(U^*;Y_R) + I(Y_R X_1^*;\hat{Y}_R^*|U^*Z)\\ &\overset{(a)}{=} I(U^*;Y_R) + I(\hat{Y}_R^*;Y_R|U^*Z)\end{aligned}$$

where (a) is due to Markov chain $X_1^* - (Y_R Z U^*) - \hat{Y}_R^*$. This concludes the proof.

E.4 Proof of Lemma 23

Before deriving the maximum achievable rate by CF in Lemma 23, we provide some definitions that will be used in the proof.

Let X and Y be a pair of discrete random variables, where $\mathcal{X} = \{1, 2, ..., n\}$ and $\mathcal{Y} = \{1, 2, ..., m\}$, for $n, m < \infty$. Let $\mathbf{p}_Y \in \Delta_m$ denote the distribution of Y, where Δ_k denotes the $(k-1)$-dimensional simplex of probability k-vectors. We define T_{XY} as the $n \times m$ stochastic matrix with entries $T_{XY}(j,i) = \Pr\{X = j|Y = i\}$. Note that the joint distribution $p(x,y)$ is characterized by T_{XY} and $\mathbf{p}_Y$.

Next, we state the conditional entropy bound from [96], which lower bounds the conditional entropy between two variables. Note the relabeling of the variables to fit our model.

Definition 4 (Conditional Entropy Bound)**.** *Let $\mathbf{p}_Y \in \Delta_m$ be the distribution of Y and T_{XY} denote the channel matrix relating X and Y. Then, for $\mathbf{q} \in \Delta_m$ and $0 \le s \le H(Y)$, define the function*

$$F_{T_{XY}}(\mathbf{q}, s) \triangleq \inf_{\substack{p(w|y):\, X-Y-W,\\ H(Y|W)=s,\, \mathbf{p}_Y=\mathbf{q}.}} H(X|W). \tag{E.22}$$

That is, $F_{T_{XY}}(\mathbf{q}, s)$ is the infimum of $H(X|W)$ given a specified distribution $\mathbf{q}$ and the value of $H(Y|W)$. Many properties of $F_{T_{XY}}(\mathbf{q}, s)$ are derived in [96], such as its convexity on $(\mathbf{q}, s)$ [96, Theorem 2.3] and its non-decreasing monotonicity in s [96, Theorem 2.5].

Consider a sequence of N random variables $\mathbf{Y} = (Y_1, ..., Y_N)$ and denote by $\mathbf{q}_i$ the distribution of Y_i, for $i = 1, ..., N$, by $\mathbf{q}^{(N)}$ the joint distribution of $\mathbf{Y}$ and by $\mathbf{q} = \frac{1}{N}\sum_{i=1}^N \mathbf{q}_i$ the average distribution. Note that $Y_1, ..., Y_N$ can have arbitrary correlation. Define the sequence $\mathbf{X} = (X_1, ..., X_N)$, in which X_i, $i = 1, ..., N$, is jointly distributed with each Y_i through the stochastic matrix T_{XY} and denote by $T_{XY}^{(N)}$ the Kronecker product of N copies of the stochastic matrix T_{XY}.

Then, the theorem given in [96, Theorem 2.4] can be straightforwardly generalized to non i.i.d. sequences as given in the following lemma.

Lemma 35. *For $N = 1, 2, ...,$ and $0 \leq Ns \leq H(\mathbf{Y})$, we have*

$$F_{T_{XY}^{(N)}}(\mathbf{q}^{(N)}, Ns) \geq NF_{T_{XY}}(\mathbf{q}, s), \tag{E.23}$$

where equality holds for i.i.d. Y_i components following $\mathbf{q}$.

Proof. Let $W, \mathbf{X}, \mathbf{Y}$ be a Markov chain, such that $H(\mathbf{Y}|W) = Ns$. Then, using the standard identity we have

$$H(\mathbf{Y}|W) = \sum_{k=1}^{N} H(Y_k|\mathbf{Y}_1^{k-1}, W), \tag{E.24}$$

$$H(\mathbf{X}|W) = \sum_{k=1}^{N} H(X_k|\mathbf{X}_1^{k-1}, W). \tag{E.25}$$

Letting $s_k = H(Y_k|\mathbf{Y}_1^{k-1}, W)$, we have

$$\frac{1}{N}\sum_{k=1}^{N} s_k = s. \tag{E.26}$$

Also, from the Markov chain $X_k - (\mathbf{Y}_1^{k-1}, W) - \mathbf{X}_1^{k-1}$, we have

$$H(X_k|\mathbf{X}_1^{k-1}, W) \geq H(X_k|\mathbf{Y}_1^{k-1}, \mathbf{X}_1^{k-1}, W) \tag{E.27}$$

$$= H(X_k|\mathbf{Y}_1^{k-1}, W). \tag{E.28}$$

Applying the conditional entropy bound in (E.22) we have

$$H(X_k|\mathbf{Y}_1^{k-1}, W) \geq F_{T_{XY}}(\mathbf{q}_k, s_k). \tag{E.29}$$

Combining (E.25), (E.27) and (E.29) we have

$$H(\mathbf{X}|W) \geq \sum_{k=1}^{N} F_{T_{XY}}(\mathbf{q}_k, s_k) \geq NF_{T_{XY}}(\mathbf{q}, s),$$

where the last inequality follows from the convexity of $F_T(\mathbf{q}, s)$ in $\mathbf{q}$ and s and (E.26).

If we let $W, \mathbf{Y}$, $\mathbf{X}$ be N independent copies of the random variables W, X, Y, that achieve $F_{T_{XY}}(\mathbf{q}, s)$, we have $H(\mathbf{Y}|W) = Ns$ and $H(\mathbf{X}|W) = F_{T_{XY}^{(N)}}(\mathbf{q}^N) = NF_{T_{XY}}(\mathbf{q}, s)$. Hence, $F_{T_{XY}^{(N)}}(\mathbf{q}^N) \leq NF_{T_{XY}}(\mathbf{q}, s)$ and the equality holds for i.i.d. components of $\mathbf{Y}$. □

Now, we look into the binary symmetric channel $Y = X \oplus N$ where $N \sim \text{Ber}(\delta)$. Due to the binary modulo-sum operation, we have $X = Y \oplus N$, and we can characterize the channel T_{XY} of this model as

$$T_{XY} = \begin{bmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{bmatrix}. \tag{E.30}$$

When Y and X are related through channel T_{XY} in (E.30), $F_{T_{XY}}(\mathbf{q}, s)$ is characterized as follows [96].

Lemma 36. *Let $Y \sim Ber(q)$, i.e., $\mathbf{q} = [q, 1-q]$, and T_{XY} be given as in (E.30). Then the conditional entropy bound is*

$$F_{T_{XY}}(q, s) = h_2(\delta \star h_2^{-1}(s)), \quad \textit{for } 0 \leq s \leq h_2(q).$$

In the following, we use the properties of $F_{T_{XY}}(\mathbf{q}, s)$ to derive the maximum rate achievable by CF in the parallel binary symmetric MRC-D. From (5.16), we have

$$\begin{aligned} I(Y_R^1, Y_R^2; \hat{Y}_R|Z) &= I(X_1^1 \oplus N_1 \oplus Z, X_1^2 \oplus N_2; \hat{Y}_R|Z) \\ &= I(X_1^1 \oplus N_1, X_1^2 \oplus N_2; \hat{Y}_R|Z). \end{aligned}$$

Let us define $\bar{Y}_R^1 \triangleq X_1^1 \oplus N_1$ and $\bar{\mathbf{Y}}_R \triangleq (\bar{Y}_R^1, Y_R^2)$, and the channel input $\mathbf{X} \triangleq (X_1^1, X_1^2)$. Note that the distribution of $\bar{\mathbf{Y}}_R$, given by $\mathbf{q}^{(2)}$, determines the distribution of $\mathbf{X}$ via $T_{XY}^{(2)}$, the Kronecker product of T_{XY} in (E.30). Then, we can rewrite the achievable rate for CF in (5.16) as follows

$$\begin{aligned} R_{CF} = &\max_{p(\mathbf{x})p(z)p(\bar{\mathbf{y}}_R|\mathbf{x})p(\hat{y}_R|\bar{\mathbf{y}}_R,z)} I(\mathbf{X}, \hat{Y}_R|Z) \\ &\text{s.t. } R_0 \geq I(\bar{\mathbf{Y}}_R; \hat{Y}_R|Z) \end{aligned} \tag{E.31}$$

Next, we derive a closed form expression for R_{CF}. First, we note that if $R_0 \geq 2$, we have $H(\bar{\mathbf{Y}}_R) \leq R_0$ and $R_{CF} = 2(1-h(\delta))$, i.e., CF meets the cut-set bound.

For fixed $\mathbf{q}^{(2)}$, if $H(\bar{\mathbf{Y}}_R) \leq R_0 \leq 2$, the constraint in (E.31) is satisfied by any $\hat{Y}_R$, and can be ignored. Then, due to the Markov chain $\mathbf{X} - \bar{\mathbf{Y}}_R - \hat{Y}_R Z$, and the data processing inequality, the achievable rate is upper bounded by

$$R_{CF} \leq I(\mathbf{X}, \bar{\mathbf{Y}}_R) = H(\bar{\mathbf{Y}}_R) - 2h(\delta) \leq R_0 - 2h(\delta). \tag{E.32}$$

For $R_0 \leq H(\bar{\mathbf{Y}}_R) \leq 2$, the achievable rate by CF is upper bounded as follows.

$$\begin{aligned} R_{CF} &\stackrel{(a)}{=} \max_{p(\mathbf{x})p(z)p(\bar{\mathbf{y}}_R|\mathbf{x})p(\hat{y}_R|\bar{\mathbf{y}}_R,z)} H(\mathbf{X}) - H(\mathbf{X}|Z\hat{Y}_R) \\ &\quad\ \text{s.t. } H(\bar{\mathbf{Y}}_R|Z\hat{Y}_R) \geq H(\bar{\mathbf{Y}}_R) - R_0 \\ &\stackrel{(b)}{\leq} \max_{p(\mathbf{x})p(\bar{\mathbf{y}}_R|\mathbf{x})p(w|\bar{\mathbf{y}}_R)} H(\mathbf{X}) - H(\mathbf{X}|W) \end{aligned}$$

$$\begin{aligned}
&\text{s.t. } H(\bar{\mathbf{Y}}_R|W) \geq H(\bar{\mathbf{Y}}_R) - R_0 \\
&= \max_{p(\mathbf{x})p(\bar{\mathbf{y}}_R|\mathbf{x})} [H(\mathbf{X}) - \min_{p(w|\bar{\mathbf{y}}_R)} H(\mathbf{X}|W)] \\
&\quad \text{s.t. } H(\bar{\mathbf{Y}}_R|W) \geq H(\bar{\mathbf{Y}}_R) - R_0 \\
&\stackrel{(c)}{=} \max_{p(\mathbf{x})p(\bar{\mathbf{y}}_R|\mathbf{x}), 0 \leq s \leq H(\bar{\mathbf{Y}}_R)} [H(\mathbf{X}) - F_{T_{XY}^{(2)}}(\mathbf{q}^{(2)}, s)] \\
&\quad \text{s.t. } s \geq H(\bar{\mathbf{Y}}_R) - R_0 \\
&\stackrel{(d)}{=} \max_{p(\mathbf{x})p(\bar{\mathbf{y}}_R|\mathbf{x})} [H(\mathbf{X}) - F_{T_{XY}^{(2)}}(\mathbf{q}^{(2)}, H(\bar{\mathbf{Y}}_R) - R_0)] \\
&\stackrel{(e)}{\leq} \max_{p(\mathbf{x})p(\bar{\mathbf{y}}_R|\mathbf{x})} [H(\mathbf{X}) - 2F_{T_{XY}}(\mathbf{q}, (H(\bar{\mathbf{Y}}_R) - R_0)/2)], \\
&\stackrel{(f)}{=} \max_{p(\mathbf{x})p(\bar{\mathbf{y}}_R|\mathbf{x})} [H(\mathbf{X}) - 2h_2(\delta \star h_2^{-1}(H(\bar{\mathbf{Y}}_R) - R_0)/2))] \\
&\quad \text{s.t. } 0 \leq (H(\bar{\mathbf{Y}}_R) - R_0)/2 \leq h_2(q)\} \\
&\stackrel{(g)}{\leq} \max_{p(\mathbf{x})p(\bar{\mathbf{y}}_R|\mathbf{x})} [H(\mathbf{X}) - 2h_2(\delta \star h_2^{-1}(H(\bar{\mathbf{Y}}_R) - R_0)/2))] \\
&\quad \text{s.t. } R_0 \leq H(\bar{\mathbf{Y}}_R) \leq 2 + R_0
\end{aligned}$$

where (a) follows from the independence of Z from $\mathbf{X}$ and $\bar{\mathbf{Y}}_R$, (b) follows since optimizing over W can only increase the value compared to optimizing over $(Z, \hat{Y}_R)$, (c) follows from the definition of the conditional entropy bound in (E.22), (d) follows from the nondecreasing monotonicity of $F_{T_{XY}^{(2)}}(\mathbf{q}^{(2)}, s)$ in s, and (e) follows from Lemma 35, and $\mathbf{q} \triangleq [q, 1-q] = \frac{1}{2}(\mathbf{q}_1 + \mathbf{q}_2)$ is the average distribution of $\mathbf{Y}$. Equality (f) follows from the definition of $F_{T_{XY}}(q, s)$ for the binary symmetric channel, and (g) follows since we are increasing the optimization domain since $h_2(q) \leq 1$.

Now, we lower bound $H(\bar{\mathbf{Y}}_R)$. Since conditioning reduces entropy, we have $H(\bar{\mathbf{Y}}_R) \geq H(\bar{\mathbf{Y}}_R|N_1 N_2) = H(\mathbf{X})$, and then we can lower bound $H(\bar{\mathbf{Y}}_R)$ as follows:

$$\max\{H(\mathbf{X}), R_0\} \leq H(\bar{\mathbf{Y}}_R) \leq 2. \tag{E.33}$$

Then, we have

$$\begin{aligned}
R_{CF} &\stackrel{(a)}{=} \max_{p(\mathbf{x})} [H(\mathbf{X}) - 2h_2(\delta \star h_2^{-1}(H(\bar{\mathbf{Y}}_R) - R_0)/2))] \\
&\quad \text{s.t. } \max\{H(\mathbf{X}), R_0\} \leq H(\bar{\mathbf{Y}}_R) \leq 2 \\
&\stackrel{(b)}{\leq} \max_{p(\mathbf{x})} [H(\mathbf{X}) - 2h_2(\delta \star h_2^{-1}((\max\{H(\mathbf{X}), R_0\} - R_0)/2))] \\
&\quad \text{s.t. } \max\{H(\mathbf{X}), R_0\} \leq H(\bar{\mathbf{Y}}_R) \leq 2 \\
&\stackrel{(c)}{=} \max_{0 \leq \alpha \leq 1} [2\alpha - 2h_2(\delta \star h_2^{-1}((\max\{2\alpha, R_0\} - R_0)/2))] \\
&\quad \text{s.t. } \max\{R_0, 2\alpha\} \leq 2,
\end{aligned}$$

where (a) follows since there is no loss in generality by introducing (E.33) since it is satisfied by any $(\mathbf{X}, \bar{\mathbf{Y}}_R)$ following $p(\mathbf{x}, \bar{\mathbf{y}}_R)$, (b) follows from (E.33) and $F_{T_{XY}}(\mathbf{q}, s)$ being non-decreasing in s, and (c) follows from defining $H(\mathbf{X}) \triangleq 2\alpha$, for $0 \leq \alpha \leq 1$.

Then, for $2\alpha \leq R_0$, we have

$$R_{CF} \leq \max_{0 \leq \alpha \leq R_0/2} [2\alpha - 2h_2(\delta)] = R_0 - 2h_2(\delta), \tag{E.34}$$

and for $2\alpha > R_0$, we have

$$R_{CF} \leq \max_{R_0/2 < \alpha \leq 1} [2\alpha - 2h_2(\delta \star h_2^{-1}(\alpha - R_0/2))]. \tag{E.35}$$

Now, we solve (E.35). Let us define $f(u) \triangleq h_2(\delta \star h_2^{-1}(u))$ for $0 \leq u \leq 1$. Then, we have the following lemma from [97].

Lemma 37 ([97])**.** *Function $f(u)$ is convex for $0 \leq u \leq 1$.*

Then, we define $g(\alpha) \triangleq \alpha - h_2(\delta \star h_2^{-1}(\alpha - R_0/2))$, such that $R_{CF} \leq \max_{R_0/2<\alpha\leq 1} 2g(\alpha)$. We have that $g(\alpha)$ is concave in α, since is a shifted version by α, which is linear with the composition of the concave function $-f(u)$ and the affine function $\alpha - R_0/2$.

Proposition 2. *$g(\alpha)$ is monotonically increasing for $R_0/2 \leq \alpha \leq 1 + R_0/2$.*

Proof. Using the chain rule for composite functions, we have

$$\frac{d^2 g(\alpha)}{d\alpha^2} = -f''(\alpha - R_0/2), \tag{E.36}$$

where $f''(u) \triangleq d^2 f/du^2(u)$.

Since $g(\alpha)$ is convex and is defined over a convex region, it follows that its unique maximum is achieved either for $f''(\alpha - R_0/2) = 0$, or at the boundaries of the region. It is shown in [97, Lemma 2] that $f''(u) > 0$ for $0 < u < 1$. That means that the maximum is achieved either at $u = 0$ or at $u = 1$, or equivalently, for $\alpha = R_0/2$ or $\alpha = 1 + R_0/2$. Since $g(R_0/2) = R_0/2 - h_2(\delta)$ and $g(1 + R_0/2) = R_0/2$, i.e., $g(R_0/2) < g(1 + R_0/2)$, it follows that $g(\alpha)$ is monotonically increasing in α for $R_0/2 \leq \alpha \leq 1 + R_0/2$. □

From Proposition 2 if follows that for $R_0/2 \leq \alpha \leq 1$, $g(\alpha)$ achieves its maximum at $\alpha = 1$. Then, for $2\alpha > R_0$, we have

$$R_{CF} \leq 2(1 - h_2(\delta \star h_2^{-1}(1 - R_0/2))). \tag{E.37}$$

Thus, from (E.34) and (E.37), we have that for $R_0 \leq H(\bar{\mathbf{Y}}_R)$

$$\begin{aligned} R_{CF} &\leq 2\max\{R_0/2 - h_2(\delta), 1 - h_2(\delta \star h_2^{-1}(1 - R_0/2))\} \\ &= 2(1 - h_2(\delta \star h_2^{-1}(1 - R_0/2))), \end{aligned} \tag{E.38}$$

where the equality follows from Proposition 2 by noting that the first element in the maximum coincides with $g(R_0/2) = R_0/2 - h_2(\delta)$, and the second one coincides $g(1)$.

Finally, R_{CF} is upper bounded by the maximum over the joint distributions satisfying $H(\bar{\mathbf{Y}}_R) \leq R_0$ given in (E.32) and the upper bound for the joint distributions satisfying $R_0 \leq H(\bar{\mathbf{Y}}_R)$ given in (E.38). Since (E.32) coincides with $g(R_0/2)$, R_{CF} is upper bounded when $R_0 \leq H(\bar{\mathbf{Y}}_R)$ as in (E.38).

Next, we show that the upper bound on the rate in (E.38) is achievable by considering the following variables

$$\begin{aligned}
&X_1^1 \sim \text{Ber}(1/2), && X_1^2 \sim \text{Ber}(1/2), && \hat{Y}_R = (\hat{Y}_R^1, \hat{Y}_R^2),\\
&\hat{Y}_R^1 = Y_R^1 \oplus Q_1, && Q_1 \sim \text{Ber}(h_2^{-1}(1 - R_0/2)), &&\\
&\hat{Y}_R^2 = Y_R^2 \oplus Q_2, && Q_2 \sim \text{Ber}(h_2^{-1}(1 - R_0/2)). &&
\end{aligned}$$

Let $Q_i \sim \text{Ber}(\nu)$ for $i = 1, 2$. Then from the constraint in (5.16) we have

$$\begin{aligned}
I(Y_R^1, Y_R^2; \hat{Y}_R|Z) &= H(\hat{Y}_R|Z) - H(\hat{Y}_R|Y_R^1 Y_R^2 Z)\\
&= H(X_1^1 \oplus N_1 \oplus Q_1, X_1^2 \oplus N_2 \oplus Q_2) - H(Q_1, Q_2)\\
&\overset{(a)}{=} 2 - 2h_2(\nu),
\end{aligned}$$

where (a) follows since $X_1^i \sim \text{Ber}(1/2)$, $i = 1, 2$ and from the independence of Q_1 and Q_2. We have $2h_2(\nu) \geq 2 - R_0$, and thus, $\nu \geq h_2^{-1}(1 - R_0/2)$.

Then, the achievable rate in (5.16) is given by

$$\begin{aligned}
I(\mathbf{X}; \hat{Y}_R|Z) &= H(\hat{Y}_R|Z) - H(\hat{Y}_R|\mathbf{X}Z)\\
&= H(X_1^1 \oplus N_1 \oplus Q_1, X_1^2 \oplus N_2 \oplus Q_2) - H(N_1 \oplus Q_1, N_2 \oplus Q_2)\\
&= 2 - 2h(\delta \star \nu)\\
&\leq 2 - 2h_2(\delta \star h_2^{-1}(1 - R_0/2)),
\end{aligned}$$

where the last inequality follows from the bound on ν. This completes the proof.

E.5 Proof of Lemma 24

From (5.7), the achievable rate for the proposed pDCF scheme is given by

$$\begin{aligned}
R_{pDCF} = \sup\, & I(X_1^1; Y_R^1) + I(X_1^2; \hat{Y}_R|Z)\\
\text{s.t. } & R_0 \geq I(X_1^1; Y_R^1) + I(Y_R^2; \hat{Y}_R|Z).
\end{aligned}$$

First, we note that the constraint is always satisfied for the choice of variables:

$$\begin{aligned} I(X_1^1;Y_R^1)+I(Y_R^2;\hat{Y}_R|Z) &= H(Y_R^1)-H(N_1)+H(X_1^2\oplus N_2\oplus Q)-H(Q)\\ &= 1-h_2(\delta)+1-h_2(h_2^{-1}(2-h(\delta)-R_0))\\ &= R_0, \end{aligned} \tag{E.39}$$

where $H(Y_R^1)=1$ since $X_1^1\sim \text{Ber}(1/2)$ and $H(X_1^2\oplus N_2\oplus Q)=1$ since $X_1^2\sim\text{Ber}(1/2)$. Then, similarly the achievable rate is given by

$$\begin{aligned} R_{pDCF} &= I(X_1^1;Y_R^1)+I(X_1^2;\hat{Y}_R|Z)\\ &= H(Y_R^1)-H(N_1)+H(X_1^2\oplus N_2\oplus Q)-H(V\oplus Q)\\ &= 1-h_2(\delta)+1-h_2(\delta\star h_2^{-1}(2-h(\delta)-R_0)), \end{aligned}$$

which completes the proof.

E.6 Proof of Lemma 27

By evaluating (5.7) with the considered Gaussian random variables, we get

$$\begin{aligned} R &= \frac{1}{2}\log\left(1+\frac{\alpha P}{\bar{\alpha}P+1}\right)\left(1+\frac{\bar{\alpha}P}{(1-\rho^2)+\sigma_q^2}\right)\\ &\text{s.t. } R_0\geq\frac{1}{2}\log\left(1+\frac{\alpha P}{\bar{\alpha}P+1}\right)\left(1+\frac{\bar{\alpha}P+(1-\rho^2)}{\sigma_q^2}\right). \end{aligned}$$

We can rewrite the constraint on R_0 as,

$$\sigma_q^2\geq f(\alpha)\triangleq\frac{(P+1)(\bar{\alpha}P+1-\rho^2)}{2^{2R_0}(\bar{\alpha}P+1)-(P+1)}. \tag{E.40}$$

Since R is increasing in σ_q^2, it is clear that the optimal σ_q^2 is obtained by $\sigma_q^2=f(\alpha)$, where α is chosen such that $f(\alpha)\geq 0$. It is easy to check that $f(\alpha)\geq 0$ for

$$\alpha\in\left[0,\min\left\{(1-2^{-2R_0})\left(1+\frac{1}{P}\right),1\right\}\right]. \tag{E.41}$$

Now, we substitute $\sigma_q^2=f(\alpha)$ in (E.40), and write the achievable rate as a function of α as

$$R(\alpha)=\frac{1}{2}\log G(\alpha), \tag{E.42}$$

where

$$G(\alpha) \triangleq \left(1 + \frac{\alpha P}{\bar{\alpha} P + 1}\right)\left(1 + \frac{\bar{\alpha} P}{(1-\rho^2) + f(\alpha)}\right)$$
$$= \frac{2^{2R_0}(1+P)(1-\rho^2+\bar{\alpha}P)}{(1-\rho^2)2^{2R_0}(1+\bar{\alpha}P)+\bar{\alpha}P(1+P)}. \tag{E.43}$$

We take the derivative of $G(\alpha)$ with respect to α:

$$G'(\alpha) \triangleq \frac{2^{2R_0}P(1+P)\left(1-\rho^2\right)\left(P+1-2^{2R_0}\rho^2\right)}{\left[P(1+P)\bar{\alpha}+2^{2R_0}(1+\bar{\alpha}P)\left(1-\rho^2\right)\right]^2}.$$

We note that if $\rho^2 \geq 2^{-R_0}(P+1)$, then $G'(\alpha) < 0$, and hence, $G(\alpha)$ is monotonically decreasing. Achievable rate R is maximized by setting $\alpha^* = 0$. When $\rho^2 < 2^{-R_0}(P+1)$, we have $G'(\alpha) > 0$, and hence $\alpha^* = \min\left\{\left(1-2^{-R_0}\right)\left(1+\frac{1}{P}\right), 1\right\} = 1$, since we have $\left(1-2^{-R_0}\right)\left(1+\frac{1}{P}\right) \geq \left(1+\frac{1-\rho^2}{P}\right) > 1$.

E.7 Proof of Lemma 28

In order to characterize the capacity of the binary symmetric MRC-D, we find the optimal distribution of $(U, X_1, \hat{Y}_R)$ in Theorem 14 for $Z \sim \text{Ber}(1/2)$. First, we note that U is independent of Y_R since

$$I(U; Y_R) \leq I(X_1; Y_R) = 0, \tag{E.44}$$

where the inequality follows from the Markov chain $U - X_1 - Y_R$, and the equality follows since for $Z \sim \text{Ber}(1/2)$ the channel output of the binary channel $Y_R = X_1 \oplus N \oplus Z$ is independent of the channel input X_1 [24]. Then, the capacity region in (5.7) is given by

$$\mathcal{C} = \sup\ \{I(X_1; \hat{Y}_R|UZ) : R_0 \geq I(Y_R; \hat{Y}_R|UZ)\},$$
$$\text{over } p(u, x_1)p(z)p(y_R|x_1, z)p(\hat{y}_R|y_R, u). \tag{E.45}$$

Let us define $\bar{Y} \triangleq X_1 \oplus N$. The capacity is equivalent to

$$\mathcal{C} = \sup\ \{I(X_1; \hat{Y}_R|UZ) : H(\bar{Y}|\hat{Y}UZ) \geq H(\bar{Y}|U) - R_0\},$$
$$\text{over } p(u, x_1)p(z)p(\bar{y}|x_1)p(\hat{y}_R|\bar{y}, u, z), \tag{E.46}$$

where we have used the fact that $\bar{Y}$ is independent from Z.

For any joint distribution for which $0 \leq H(\bar{Y}|U) \leq R_0$, the constraint in (E.46) is also satisfied. In that case, we can find the following upper bound on the capacity. It follows from the Markov chain $X_1 - \bar{Y} - \hat{Y}_R$ given U, Z, and the data processing inequality, that

$$\begin{aligned}\mathcal{C} &\leq \max_{p(u,x_1)} \{I(X_1;\bar{Y}|ZU) : H(\bar{Y}|U) \leq R_0\} \\ &= \max_{p(u,x_1)} \{H(\bar{Y}|U) - h_2(\delta) : H(\bar{Y}|U) \leq R_0\} \\ &\leq R_0 - h_2(\delta).\end{aligned} \tag{E.47}$$

We next consider the joint distributions for which $R_0 \leq H(\bar{Y}|U)$. Let $p(u) = \Pr[U = u]$ for $u = 1, .., |\mathcal{U}|$, and we can write

$$I(X_1;\hat{Y}_R|UZ) = H(X_1|U) - \sum_u p(u)H(X_1|\hat{Y}_R Zu), \tag{E.48}$$

and

$$I(Y_R;\hat{Y}_R|UZ) \overset{(a)}{=} I(\bar{Y};\hat{Y}_R|UZ) \overset{(b)}{=} H(\bar{Y}|U) - \sum_u p(u)H(\bar{Y}|\hat{Y}_R Zu),$$

where (a) follows from the definition of $\bar{Y}$, and (b) follows from the independence of Z from $\bar{Y}$ and U.

For each u, the channel input X_1 corresponds to a binary random variable $X_u \sim \mathrm{Ber}(\nu_u)$, where $\nu_u \triangleq \Pr[X_1 = 1|U = u] = p(1|u)$ for $u = 1, ..., |\mathcal{U}|$. The channel output for each X_u is given by $\bar{Y}_u = X_u \oplus N$. We denote by $q_u \triangleq \Pr[Y_u = 1] = \Pr[Y_R = 1|U = u]$. Similarly, we define $\hat{Y}_u$ as $\hat{Y}_R$ for each u value. Note that for each u, $X_u - \bar{Y}_u - \hat{Y}_u$ form a Markov chain. Then, we have $H(X_1|u) = h_2(\nu_u)$ and $H(\bar{Y}|u) = h_2(\delta \star \nu_u)$. We define $s_u \triangleq H(\bar{Y}|\hat{Y}_R Zu)$, such that $0 \leq s_u \leq H(\bar{Y}_u)$. Substituting (E.48) and (E.49) in (E.46) we have

$$\begin{aligned}\mathcal{C} \;=\; & \max_{p(u,x_1)p(\hat{y}_R|y_R,u)} [H(X_1|U) - \sum_u p(u)H(X_1|\hat{Y}_R Zu)] \\ & \text{s.t. } R_0 \geq H(\bar{Y}|U) - \sum_u p(u)H(\bar{Y}|\hat{Y}_R Zu) \\ \overset{(a)}{=}\; & \max_{p(u,x_1),} [H(X_1|U) - \sum_u p(u)F_{T_{XY}}(q_u, s_u)] \\ & \text{s.t. } R_0 \geq H(\bar{Y}|U) - \sum_u p(u)s_u,\ 0 \leq s_u \leq H(\bar{Y}_u) \\ \overset{(b)}{=}\; & \max_{p(u,x_1)} [H(X_1|U) - \sum_u p(u)h_2(\delta \star h_2^{-1}(s_u))] \\ & \text{s.t. } R_0 \geq H(\bar{Y}|U) - \sum_u p(u)s_u,\ 0 \leq s_u \leq H(\bar{Y}_u), \\ \overset{(c)}{\leq}\; & \max_{p(u,x_1)} H(X_1|U) - h_2\left(\delta \star h_2^{-1}\left(\sum_u p(u)s_u\right)\right) \\ & \text{s.t. } \sum_u p(u)s_u \geq H(\bar{Y}|U) - R_0,\end{aligned}$$

where (a) follows from the definition of $F_{T_{XY}}(q,s)$ for channel $\bar{Y}_u = X_u \oplus N$, which for each u has a matrix T_{XY} as in (E.30), (b) follows from the expression of $F_{T_{XY}}(q,s)$ for the binary channel T_{XY} in Lemma 36, (c) follows from noting that $-h_2(\delta \star h_2^{-1}(s_u))$ is concave on s_u from Lemma 37 and applying Jensen's inequality. We also drop the conditions on s_u, which can only increase $\mathcal{C}$.

Then, similarly to the proof of Lemma 23, we have $H(\bar{Y}|U) \geq H(\bar{Y}|UV) = H(X_1|U)$, and we can upper bound the capacity as follows

$$\begin{aligned}\mathcal{C} \leq \max_{p(x_1,u)} & \left[H(X_1|U) - h_2\left(\delta \star h_2^{-1}\left(\sum_u p(u)s_u\right)\right)\right] \\ & \text{s.t. } \sum_u p(u)s_u \geq \max\{H(X_1|U), R_0\} - R_0 \\ \leq & \max_{0\leq\alpha\leq 1} \alpha - h_2(\delta \star h_2^{-1}(\max\{\alpha, R_0\} - R_0)), \end{aligned} \tag{E.49}$$

where we have defined $\alpha \triangleq H(X_1|U)$.

The optimization problem can be solved similarly to the proof in Appendix E as follows. If $0 \leq \alpha \leq R_0$, we have $\bar{s} \geq 0$ and

$$\mathcal{C} \leq \max_{0\leq\alpha\leq R_0} \alpha - h_2(\delta) = R_0 - h_2(\delta). \tag{E.50}$$

For $R_0 \leq \alpha \leq 1$, we have

$$\mathcal{C} \leq \max_{R_0\leq\alpha\leq 1} \alpha - h_2(\delta \star h_2^{-1}(\alpha - R_0)). \tag{E.51}$$

Then, it follows from a scaled version of Proposition 2 that the upper bound is maximized for $\alpha = 1$. Then, by noticing that (E.50) corresponds to the value of the bound in (E.51) for $\alpha = R_0$, it follows that

$$\mathcal{C} \leq 1 - h_2(\delta \star h_2^{-1}(1 - R_0)). \tag{E.52}$$

This bound is achievable by CF. This completes the proof.

E.8 Proof of the Cut-Set Bound Optimality Conditions

Cases 1 and 2 are straightforward since under these assumptions, the ORC-D studied here becomes a particular case of the channel models in [91] and [77], respectively.

To prove case 3 we use the following arguments. For any channel input distribution to the ORC-D, we have

$$\begin{aligned} I(X_1; Y_R|Z) &= H(X_1|Z) - H(X_1|Y_R, Z) \\ &\geq H(X_1) - H(X_1|Y_R) \\ &= I(X_1; Y_R), \end{aligned} \tag{E.53}$$

where we have used the independence of X_1 and Z and the fact that conditioning reduces entropy. Then, the condition $\max_{p(x_1)} I(X_1; Y_R) \geq R_0$, implies $\max_{p(x_1)} I(X_1; Y_R|Z) \geq R_0$; and hence, the cut-set bound is given by $R_{CS} = R_1 + R_0$, which is achievable by DF scheme.

In case 4, the cut-set bound is given by $R_1 + \min\{R_0, I(\bar{X}_1; \bar{Y}_R|Z)\} = R_1 + I(\bar{X}_1; \bar{Y}_R|Z)$ since $R_0 \geq H(\bar{Y}_R|Z)$. CF achieves the capacity by letting X_1 be distributed with $\bar{p}(x_1)$, and choosing $\hat{Y}_R = \bar{Y}_R$. This choice is always possible as the CF constraint

$$R_0 \geq I(\hat{Y}_R; \bar{Y}_R|Z) = H(\bar{Y}_R|Z) - H(\bar{Y}_R|Z, \hat{Y}_R) = H(\bar{Y}_R|Z),$$

always holds. Then, the achievable rate for CF is $R_{CF} = R_1 + I(\bar{X}_1; \hat{Y}_R|Z) = R_1 + I(\bar{X}_1; \bar{Y}_R|Z)$, which is the capacity.

Bibliography

[1] "Visualetworking index," *White paper at Cisco.com*, Feb 2014.

[2] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?," *IEEE Journal on Selected Areas in Communications*, vol. PP, pp. 1–1, May 2014.

[3] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, July and October 1948.

[4] A. Wyner, "The rate-distortion function for source coding with side information at the decoder," *Information and Control*, vol. 38, pp. 60–80, Jan. 1978.

[5] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. on Information Theory*, vol. 31, pp. 727 – 734, Nov. 1985.

[6] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Information Theory*, vol. 19, pp. 471–480, Jul. 1973.

[7] T. Berger and R. Yeung, "Multiterminal source encoding with one distortion criterion," *IEEE Trans. on Information Theory*, vol. 35, pp. 228–236, Mar. 1989.

[8] A. Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. on Information Theory*, vol. 28, pp. 851–857, Nov. 1982.

[9] V. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, pp. 74–93, Sep. 2001.

[10] A. El Gamal and T. Cover, "Multiple user information theory," *Proc. of the IEEE*, vol. 68, pp. 1466–1483, Dec. 1980.

[11] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Stat. Prob.*, vol. 1, (Berkeley, CA), pp. 611–644, 1961.

[12] D. Gündüz, E. Erkip, A. Goldsmith, and H. V. Poor, "Source and channel coding for correlated sources over multiuser channels," *IEEE Trans. on Information Theory*, vol. 55, pp. 3927–3944, Sep. 2009.

[13] A. Lapidoth and S. Tinguely, "Sending a bivariate Gaussian over a Gaussian MAC," *IEEE Trans. on Information Theory*, vol. 56, pp. 2714–2752, Jun. 2010.

[14] M. Gastpar and M. Vetterli, "Power, spatio-temporal bandwidth, and distortion in large sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 745–754, Apr. 2005.

[15] D. Gündüz and E. Erkip, "Joint source–channel codes for MIMO block-fading channels," *IEEE Trans. on Information Theory*, vol. 54, pp. 116–134, Jan. 2008.

[16] G. Caire and K. Narayanan, "On the distortion SNR exponent of hybrid digital–analog space–time coding," *IEEE Trans. on Information Theory*, vol. 53, pp. 2867–2878, Aug. 2007.

[17] U. Mittal and N. Phamdo, "Hybrid digital–analog (HDA) joint source–channel codes for broadcasting and robust communications," *IEEE Trans. on Information Theory*, vol. 48, May 2002.

[18] Y. Polyanskiy, H. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. on Information Theory*, vol. 56, pp. 2307 –2359, May 2010.

[19] V. Kostina and S. Verdu, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Trans. on Information Theory*, vol. 58, pp. 3309 –3338, Jun. 2012.

[20] A. Fuldseth and T. Ramstad, "Bandwidth compression for continuous amplitude channels based on vector approximation to a continuous subset of the source signal space," in *IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, pp. 3093–3096 vol.4, Apr. 1997.

[21] V. Vaishampayan and S. I. R. Costa, "Curves on a sphere, shift-map dynamics, and error control for continuous alphabet sources," *IEEE Trans. on Information Theory*, vol. 49, pp. 1658–1672, Jul. 2003.

[22] F. Hekland, P. Floor, and T. Ramstad, "Shannon-Kotelnikov mappings in joint source-channel coding," *IEEE Trans. on Communications*, vol. 57, pp. 94–105, Jan. 2009.

[23] I. Estella, M. Varasteh, and D. Gündüz, "Zero-delay joint source-channel coding," in *Iran Workshop on Communication and Information Theory (IWCIT)*, pp. 1–6, May 2014.

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

[25] A. E. Gamal and Y.-H. Kim, *Network Information Theory.* Cambridge University Press, 2011.

[26] T. J. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Information Theory*, vol. 11, pp. 558–567, November 1965.

[27] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. on Information Theory*, vol. 49, pp. 1147–1158, May 2003.

[28] S. Shamai, S. Verdú, and R. Zamir, "Systematic lossy source-channel coding," *IEEE Trans. Information Theory*, vol. 44, pp. 564–579, Mar. 1998.

[29] M. Gastpar, "Uncoded Transmission Is Exactly Optimal for a Simple Gaussian Sensor Network," *IEEE Trans. on Information Theory*, vol. 54, pp. 5247–5251, Nov. 2008.

[30] V. Kostina and S. Verdu, "To code or not to code: Revisited," in *Proc. IEEE Information Theory Workshop (ITW)*, pp. 5–9, Sep. 2012.

[31] O. Dabeer, A. Roumy, and C. Guillemot, "Linear transceivers for sending correlated sources over the Gaussian MAC," in *National Communication Conference*, (Kanpur, India), Jan. 2007.

[32] S. Bross, A. Lapidoth, and S. Tinguely, "Superimposed Coded and Uncoded Transmissions of a Gaussian Source over the Gaussian Channel," in *Proc. IEEE Int'l Symposium on Information Theory Proceedings (ISIT)*, pp. 2153–2155, Jul. 2006.

[33] J. Nayak, E. Tuncel, and D. Gunduz, "Wyner–Ziv Coding Over Broadcast Channels: Digital Schemes," *IEEE Trans. on Information Theory*, vol. 56, pp. 1782–1799, Apr. 2010.

[34] M. Wilson, K. Narayanan, and G. Caire, "Joint source channel coding with side information using hybrid digital analog codes," *IEEE Trans. on Information Theory*, vol. 56, pp. 4922–4940, Oct. 2010.

[35] S. H. Lim, P. Minero, and Y.-H. Kim, "Lossy communication of correlated sources over multiple access channels," in *Proc. 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton).*, pp. 851 –858, Oct. 2010.

[36] P. Minero, S. H. Lim, and Y.-H. Kim, "Joint source-channel coding via hybrid coding," in *Proc. IEEE Int'l Symposium on Information Theory Proceedings (ISIT)*, pp. 781 –785, Aug. 2011.

[37] Y.-H. Kim, S. Lim, and P. Minero, "Relaying via hybrid coding," in *Proc. IEEE Int'l Symposium on Information Theory (ISIT)*, pp. 1881–1885, Jul. 2011.

[38] W. Equitz and T. Cover, "Successive refinement of information," *IEEE Trans. on Information Theory*, vol. 37, pp. 269–275, Mar. 1991.

[39] K. Bhattad, K. Narayanan, and G. Caire, "On the distortion SNR exponent of some layered transmission schemes," *IEEE Trans. on Information Theory*, vol. 54, no. 7, pp. 2943–2958, 2008.

[40] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Transactions on Information Theory*, vol. 43, pp. 1912–1923, Nov. 1997.

[41] J. Laneman, E. Martinian, G. W. Wornell, and J. Apostolopoulos, "Source-channel diversity for parallel channels," *IEEE Trans. on Information Theory*, vol. 51, pp. 3518–3539, Oct. 2005.

[42] C. Ng, D. Gündüz, A. Goldsmith, and E. Erkip, "Distortion minimization in Gaussian layered broadcast coding with successive refinement," *IEEE Trans. on Information Theory*, vol. 55, pp. 5074–5086, Nov. 2009.

[43] C. Ng, C. Tian, A. Goldsmith, and S. Shamai, "Minimum expected distortion in Gaussian source coding with uncertain side information," in *Proc. IEEE Information Theory Workshop (ITW)*, pp. 454–459, Sep. 2007.

[44] T. Berger, *Multi-terminal Source Coding.* Chapter in The Information Theory Approach to Communications (G. Longo, ed.), Springer-Verlag, 1978.

[45] H. Behroozi and M. Soleymani, "Sending correlated Gaussian sources over a Gaussian MAC: To code, or not to code," in *Proc. IEEE Int'l Conf. on Communication (ICC)*, (Glasgow, Scotland), Jun. 2007.

[46] W. Liu and B. Chen, "Communicating correlated Gaussian sources over Gaussian Z interference channels," in *Proc. IEEE Global Telecommunications Conference (Globecom)*, (Honolulu, Hawaii), Dec. 2009.

[47] D. Gündüz, J. Nayak, and E. Tuncel, "Wyner-Ziv coding over broadcast channels using hybrid digital/analog transmission," in *Proc. IEEE Int'l Symposium on Information Theory Proceedings (ISIT)*, (Toronto, ON, Canada), Jul. 2008.

[48] Y. Gao and E. Tuncel, "New hybrid digital/analog schemes for transmission of a gaussian source over a gaussian channel," *IEEE Trans. on Information Theory*, vol. 56, pp. 6014 –6019, Dec. 2010.

[49] M. H. M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. on Information Theory*, vol. 29, pp. 439–441, May 1983.

[50] I. Estella and D. Gündüz, "Hybrid digital-analog transmission for the gaussian one-helper problem," in *Proc. IEEE Global Telecommunications Conference (Globecom)*, pp. 1 –5, Dec. 2010.

[51] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge Univ Pr, 2004.

[52] A. Sutivong, M. Chiang, T. Cover, and Y.-H. Kim, "Channel capacity and state estimation for state-dependent gaussian channels," *IEEE Trans. on Information Theory*, vol. 51, pp. 1486–1495, Apr. 2005.

[53] Y.-C. Huang and K. Narayanan, "Joint source-channel coding with correlated interference," *IEEE Trans. on Communications*, vol. 60, pp. 1315–1327, May 2012.

[54] M. Varasteh and H. Behroozi, "On the transmission of a gaussian source over an awgn channel with correlated interference," in *Proc. 12th Canadian Workshop on Information Theory (CWIT)*, pp. 19–23, May 2011.

[55] C. Tian, S. Diggavi, and S. Shamai, "The achievable distortion region of sending a bivariate gaussian source on the gaussian broadcast channel," *IEEE Trans. on Information Theory*, vol. 57, pp. 6419–6427, Oct. 2011.

[56] E. Tuncel, "Lossless joint source-channel coding across broadcast channels with decoder side information," in *Proc. IEEE Int'l Symposium on Information Theory Proceedings (ISIT)*, (Adelaide, Australia), Sep. 2005.

[57] E. Tuncel, "Slepian-Wolf coding over broadcast channels," *IEEE Trans. on Information Theory*, vol. 52, pp. 1469–1482, Apr. 2006.

[58] J. Nayak, E. Tuncel, and D. Gunduz, "Wyner-ziv coding over broadcast channels: Digital schemes," *IEEE Trans. on Information Theory*, vol. 56, pp. 1782–1799, Apr. 2010.

[59] F. Etemadi and H. Jafarkhani, "Optimal layered transmission over quasi-static fading channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jul. 2006.

[60] C. Tian, A. Steiner, S. Shamai, and S. Diggavi, "Successive refinement via broadcast: Optimizing expected distortion of a Gaussian source over a Gaussian fading channel," *IEEE Trans. on Information Theory*, vol. 54, pp. 2903 –2918, Jul. 2008.

[61] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, “Cooperative diversity in wireless networks: Efficient protocols and outage behavior,” *IEEE Trans. on Information Theory*, vol. 50, pp. 3062–3080, Dec. 2004.

[62] D. Gündüz and E. Erkip, “Distortion exponent of parallel fading channels,” in *Proc. IEEE Int'l Symposium on Information Theory Proceedings (ISIT)*, pp. 694–698, Jul. 2006.

[63] D. Gündüz and E. Erkip, “Source and channel coding for cooperative relaying,” *IEEE Trans. on Information Theory*, vol. 53, pp. 3454–3475, Oct. 2007.

[64] D. Gündüz, A. Goldsmith, and H. Poor, “Distortion exponent in mimo channels with feedback,” in *Proc. IEEE Information Theory Workshop (ITW)*, pp. 293–297, Oct. 2009.

[65] J. Wang and J. Liang, “Distortion exponents of two-way relay networks,” *IEEE Trans. on Signal Processing*, vol. 59, pp. 4424–4437, Sep. 2011.

[66] S. Zhao and D. Tuninetti, “On the distortion exponent of block-fading gaussian interference channels,” in *Proc. 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton).*, pp. 1266–1272, Oct. 2010.

[67] S. Zhao, R. Timo, T. Chan, A. Grant, and D. Tuninetti, “The impact of side-information on gaussian source transmission over block-fading channels,” in *Proc. IEEE Int'l Conference on Communications (ICC)*, pp. 1–5, May 2010.

[68] S. Zhao, R. Timo, T. Chan, A. Grant, and D. Tuninetti, “The impact of side-information on Gaussian source transmission over block-fading channels,” in *Proc. IEEE Int'l Conference on Communications (ICC)*, pp. 1–5, May 2010.

[69] E. Koken and E. Tuncel, “Gaussian HDA coding with bandwidth expansion and side information at the decoder,” in *Proc. IEEE Int'l Symposium on Information Theory (ISIT)*, pp. 11–15, Jul. 2013.

[70] I. Estella and D. Gündüz, “Distortion exponent with side-information diversity,” in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 941–944, Dec. 2013.

[71] Y. Steinberg and N. Merhav, “On hierarchical joint source-channel coding with degraded side information,” *IEEE Trans. on Information Theory*, vol. 52, pp. 886–903, Mar. 2006.

[72] M. Shaqfeh, W. Mesbah, and H. Alnuweiri, “Utility maximization for layered transmission using the broadcast approach,” *IEEE Trans. on Wireless Communications*, vol. 11, pp. 1228–1238, Mar. 2012.

[73] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. on Information Theory*, vol. 49, pp. 1073–1096, May. 2003.

[74] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. on Information Theory*, vol. 25, pp. 572 – 584, Sep. 1979.

[75] A. El Gamal and S. Zahedi, "Capacity of a class of relay channels with orthogonal components," *IEEE Trans. on Information Theory*, vol. 51, pp. 1815 –1817, May 2005.

[76] A. Gamal and M. Aref, "The capacity of the semideterministic relay channel (corresp.)," *IEEE Trans. on Information Theory*, vol. 28, p. 536, May 1982.

[77] T. M. Cover and Y.-H. Kim, "Capacity of a class of deterministic relay channels," in *Proc. IEEE Int'l Symposium on Information Theory Proceedings (ISIT)*, pp. 591 –595, Jun. 2007.

[78] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channels," *IEEE Trans. on Information Theory*, vol. 55, pp. 921 –930, Mar. 2009.

[79] W. Kang and S. Ulukus, "Capacity of a class of diamond channels," *IEEE Trans. on Information Theory*, vol. 57, pp. 4955 –4960, Aug. 2011.

[80] M. Khormuji and M. Skoglund, "On cooperative downlink transmission with frequency reuse," in *IEEE Int'l Symposium on Information Theory Proceedings (ISIT)*, pp. 849–853, Jun. 2009.

[81] A. Zaidi, S. Shamai, P. Piantanida, and L. Vandendorpe, "Bounds on the capacity of the relay channel with noncausal state information at source," in *Proc. IEEE Int'l Symposium on Information Theory Proceedings (ISIT)*, pp. 639 –643, Jun. 2010.

[82] B. Akhbari, M. Mirmohseni, and M.-R. Aref, "Achievable rate regions for relay channel with private messages with states known at the source," in *Proc. IEEE Int'l Conf. on Information Theory and Information Security (ICITIS)*, pp. 1080–1087, Dec. 2010.

[83] A. Zaidi, S. Kotagiri, J. Laneman, and L. Vandendorpe, "Cooperative relaying with state available noncausally at the relay," *IEEE Trans. on Information Theory*, vol. 56, pp. 2272 –2298, May 2010.

[84] B. Akhbari, M. Mirmohseni, and M. Aref, "Compress-and-forward strategy for relay channel with causal and non-causal channel state information," *IET Communications*, vol. 4, pp. 1174–1186, Jul. 2010.

[85] M. Mirmohseni, B. Akhbari, and M.-R. Aref, “Compress-and-forward strategy for the relay channel with causal state information,” in *Proc. IEEE Information Theory Workshop (ITW)*, pp. 426–430, Oct. 2009.

[86] M. Khormuji, A. El Gamal, and M. Skoglund, “State-dependent relay channel: Achievable rate and capacity of a semideterministic class,” *IEEE Trans. on Information Theory*, vol. 59, pp. 2629–2638, May 2013.

[87] O. Simeone, D. Gunduz, and S. Shamai, “Compound relay channel with informed relay and destination,” in *Proc. 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton).*, pp. 692 –699, Oct. 2009.

[88] A. Behboodi and P. Piantanida, “On the simultaneous relay channel with informed receivers,” in *Proc. IEEE Int’l Symposium on Information Theory Proceedings (ISIT)*, pp. 1179 –1183, Jul. 2009.

[89] K. Bakanoglu, E. Erkip, O. Simeone, and S. Shamai, “Relaying under structured interference,” in *Proc. 46th Annual Conference on Information Sciences and Systems (CISS),*, pp. 1–6, Mar. 2012.

[90] K. Bakanoglu, E. Erkip, and O. Simeone, “Half-duplex gaussian diamond relay channel with interference known at one relay,” in *Proc. 45th Asilomar Conference on Signals, Systems and Computers (Asilomar)*, pp. 1353 –1357, Nov. 2011.

[91] A. El Gamal and S. Zahedi, “Capacity of a class of relay channels with orthogonal components,” *IEEE Trans. on Information Theory*, vol. 51, pp. 1815–1817, May 2005.

[92] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, 1981.

[93] L. Ozarow, “The capacity of the white gaussian multiple access channel with feedback,” *IEEE Trans. on Information Theory*, vol. 30, pp. 623 – 629, Jul. 1984.

[94] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications.* Springer, New York, 2nd ed. ed., 1998.

[95] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables.* Courier Dover Publications, 2012.

[96] H. S. Witsenhausen and A. D. Wyner, “A conditional entropy bound for a pair of discrete random variables,” *IEEE Trans. on Information Theory*, vol. 21, pp. 493–501, Sep. 1975.

[97] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications–i," *IEEE Trans. on Information Theory*, vol. 19, pp. 769–772, Nov. 1973.