DELAY CONSTRAINED MULTIMEDIA COMMUNICATIONS: COMPARING SOURCE-CHANNEL APPROACHES FOR QUASI-STATIC FADING CHANNELS

A Thesis

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Master of Science in Electrical Engineering

by

Brian P. Dunn, B.S.

J. Nicholas Laneman, Director

Graduate Program in Electrical Engineering
Notre Dame, Indiana
August 2005

DELAY CONSTRAINED MULTIMEDIA COMMUNICATIONS: COMPARING SOURCE-CHANNEL APPROACHES FOR QUASI-STATIC FADING CHANNELS

Abstract

by

Brian P. Dunn

Real-time multimedia communication over a wireless link presents many challenges that require non-traditional methods to ensure good performance. A strict delay constraint prevents averaging over variations in the channel's fading coefficient, resulting in a channel with zero capacity in the Shannon sense. Without knowledge of the channel realization at the transmitter, separate source and channel coding is no longer optimal, and we must consider joint source-channel coding techniques.

In this thesis we examine the performance of several schemes that attempt to mitigate the effects of non-ergodic fading on the end-to-end mean-square distortion. We derive an upper bound on the rate at which the expected distortion decays for high SNR, and the performance of each scheme is analytically characterized using this metric, the distortion exponent. Limitations of this distortion metric are also discussed and illustrated. We analyze the performance of uncoded and rate-optimized digital transmission over both a single channel and parallel channels. We consider successive refinement source coding utilizing superposition channel coding and show that in the high SNR limit it offers significantly improved performance relative to standard digital techniques. We present a hybrid digital-analog scheme as a simple form of multiple descriptions and show that it outperforms the other techniques considered for parallel channels.

CONTENTS

FIGURES

SYMBOLS

a	Fading coefficient, a complex Gaussian random variable
$\mathbf{x}$	Channel input, a random vector
$\mathbf{y}$	Channel output, a random vector
$x_{1,i}$	The ith input on channel 1, a random variable
$x_{2,i}$	The ith input on channel 2, a random variable
$y_{1,i}$	The ith output on channel 1, a random variable
$y_{2,i}$	The ith output on channel 2, a random variable
$\mathbf{s}$	Source, a random vector
$\hat{\mathbf{s}}$	Source reconstruction, a random vector
$\asymp$	For asymptotically high SNR

CHAPTER 1

INTRODUCTION

In recent years it has become of increasing interest to send multimedia information over a wireless link in real-time, such as sensor data over a sensor network, voice over a cellular network, digital radio broadcasts, or audio to wireless speakers. Over the past 50 years, fundamental techniques in digital communications have relied on concepts outlined by Shannon in his landmark 1948 paper [1]. Unfortunately, these techniques require infinite delays and impractical complexity for optimal performance, and therefore do not directly apply to real-time communications. Furthermore, there is currently no complete characterization of the achievable performance when a finite delay constraint is imposed. The result is we do not know what level of performance is possible, and it is unclear what technique should be used for communication.

This thesis serves to clearly illustrate the sub-optimality of separate source and channel coding for the block-fading channel, present an analytical characterization of various schemes on the single-input single-output (SISO) and certain multiple-input multiple-output (MIMO) channels, and provide intuition into the types of systems whose distortion approaches that of a known lower bound. We begin by providing a brief synopsis of digital communication theory along with a more technical summary of relevant background material in Chapter 2. Chapter 3 introduces a framework for the comparison of different schemes using a single metric, and presents a clear motivation for the study of joint source channel coding schemes through an analytical comparison of classic digital and analog communication over a block-fading SISO channel. This analysis is carried out for several more advanced schemes in Chapter 4 for block-fading parallel channels. Finally, concluding remarks along with some potential ideas for future research are given in Chapter 5.

CHAPTER 2

BACKGROUND

In this chapter we present a general overview of communication theory and a summary of related research. We begin by introducing a basic channel model and the concepts of source and channel coding. We then discuss how to analyze a system's performance, followed by generalizing our channel model. Finally, we give an overview of related research, from both theoretical and practical viewpoints, focusing on various forms of joint source-channel coding techniques such as multiple descriptions, successive refinement, and hybrid digital-analog transmission.

■ 2.1 General Overview of Wireless Communications

The general goal of communications is to transmit information from one location to another. More specifically, we consider transmitting a continuous time source $\mathsf{s}(t)$, such as audio or video, to a destination through some non-ideal channel. Although the sources and channels are often continuous in time, in many circumstances we can consider discrete-time equivalents. Without loss of generality [2], we consider the equivalent discrete-time signal s_i, which is a sampled version of the band-limited random process, $\mathsf{s}(t)$. Our channel can then be described as

$$\mathbf{y} = \mathbf{x} + \mathbf{w} \tag{2.1}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the channel input and output, respectively, and $\mathbf{w}$ is additive white Gaussian noise (AWGN) with power spectral density $N_0/2 = \sigma_{\mathsf{w}}^2$. It is of

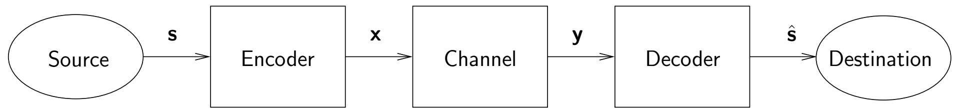

Figure 2.1. Block diagram of a general communication system.

fundamental interest to know what level of performance can be guaranteed, and how best to achieve it. Finding answers to these questions is the main goal of wireless communications. Figure 2.1 shows a block diagram of a general communication system, depicting that we encode the signal before transmission in order to ensure a certain level of performance is achieved.

Throughout this work, we model our source as Gaussian such that s_i are independent identically distributed (i.i.d.) zero-mean Gaussian random variables with variance σ_s^2, i.e., $s_i \sim \mathcal{N}(0, \sigma_s^2)$. In order to evaluate the performance of a system, we introduce a distortion measure between the source s_i and its reconstruction $\hat{s}_i$: $D_{\mathbf{s}} = f(s_i, \hat{s}_i)$. For simplicity of exposition we almost exclusively utilize mean-square error as our distortion measure, i.e.,

$$D_{\mathbf{s}} = |s_i - \hat{s}_i|^2. \tag{2.2}$$

We extend (2.2) additively to blocks so that

$$|\mathbf{s} - \hat{\mathbf{s}}|^2 = \frac{1}{N} \sum_{i=0}^{N-1} |s_i - \hat{s}_i|^2. \tag{2.3}$$

These assumptions are practical, for example, when considering the transmission of i.i.d. Gaussian sensor data and the error signal's power is of interest. For correlated or nonwhite sources such as speech or video, practical systems could do at least as well as those considered in this thesis.

The problem of how best to encode a source for reliable transmission was greatly simplified for certain scenarios by Shannon [1] in 1948. He proved that, under certain

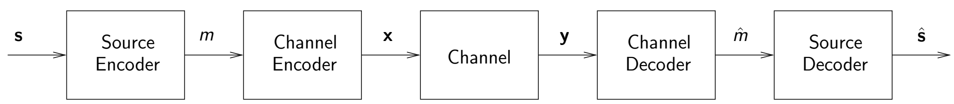

Figure 2.2. Block diagram of a communication system depicting separate source and channel coding.

conditions, the challenge of communicating a source over a particular channel could be broken down into two simpler problems, with each considered independently of the other. A block diagram depicting this simplification is shown in Figure 2.2. The first problem, *source coding*, involves compressing the source to a lossy, but finite-valued, representation, preferably so that all possible representations are equally likely. The second component to the transmission process, *channel coding*, adds redundancy to the compressed representation in a controlled manner, so that the source encoder's output is faithfully reproduced at the source decoder input. The source decoder can then use this information to create a reconstruction $\hat{\mathbf{s}}$ of the original source signal. In this manner the source encoder/decoder pair needs no information about the channel, and the channel encoder/decoder pair can likewise be designed without regard for specific properties of the source.

■ 2.1.1 Channel Coding

Another significant contribution of Shannons work was the discovery that for any channel there is a limit to the amount of information we can reliably communicate over it. This fundamental maximum rate of communication is called the channel capacity, C. Conversely, if we try to communicate information at a rate $R > C$, there is a non-zero probability of error associated with the decoded bit stream. Therefore, the channel cannot support reliable communication at rates above capacity. The capacity of the additive white Gaussian noise (AWGN) channel can be shown to be

[3]

$$C = \frac{1}{2} \log(1 + \text{SNR}) \text{ (nats per channel use)}, \tag{2.4}$$

where $\text{SNR} = P/N_0$ is the channels signal to noise ratio. Thus, for a given SNR we can use (2.4) to compute the highest rate at which the channel will support reliable communication. The source encoder should then be designed to output a binary representation of the source at a rate less than capacity.

Mathematically, this error-free communication is guaranteed only when we encode the entire infinite duration source sequence at once. In practice, the source is encoded in chunks of block length N. Using advanced channel coding techniques [4], bit error rates on the order of 10^{-5} can be achieved at SNRs within 1 dB of capacity for block lengths of only a few thousand bits.

■ 2.1.2 Source Coding

Whenever we describe a continuous valued source with a finite alphabet, there will be some loss of information. Thus, for the type of sources under consideration, *lossy* source coding is often used. The goal of lossy source coding is to create the best possible description of the source for a given rate. Yet another result of Shannon's is that for a given source coding rate R_s (bits per source sample)[1], there is a limit to how low the distortion incurred can be. The function that describes the trade-off between the rate of the code and the resultant distortion is called the distortion-rate function. For a Gaussian source with mean-square distortion, the distortion-rate function is [3]

$$D(R_s) = \sigma_s^2 \cdot 2^{-2R_s}. \tag{2.5}$$

[1] At times we will alternatively express the channel capacity and the rate-distortion function in terms of nats/source sample.

Equivalently,

$$R_s(D) = \begin{cases} \frac{1}{2}\log\frac{\sigma_s^2}{D}, & 0 \le D \le \sigma_s^2 \\ 0, & D > \sigma_s^2 \end{cases}. \quad (2.6)$$

expresses the rate required to guarantee that a specific distortion is achieved.

As in the case of realizing the channel capacity, distortion-rate function can be achieved only for infinitely long block lengths and using vector quantization [5]. In practice, a source coder's performance can approach the distortion-rate function for reasonably short block lengths using vector quantization. In the event that the distortion-rate function cannot be achieved, (2.6) can still be used as a lower bound on the performance of any source coder.

■ 2.1.3 Evaluating End-to-End System Performance

Combining the notions of channel capacity and rate-distortion, we can plot the end-to-end distortion as a function of the rate, as shown in Figure 2.3. For rates less than capacity, the source encoder's description is available error-free at the source decoder, and thus the distortion is simply the distortion-rate function. As the rate goes beyond the capacity of the channel, the probability of error exponentially approaches 1, and the distortion will approach the variance of the source. Figure 2.3 shows that the performance improves as the rate approaches the capacity of the channel. Therefore, not only does the separation theorem provide a tractable means for designing communication systems, it also yields a way to compute the end-to-end distortion. Since, with high probability, the channel decoder reproduces the source encoders description perfectly, the end-to-end distortion is found by evaluating the distortion-rate function of the source at the capacity of the channel. For transmitting

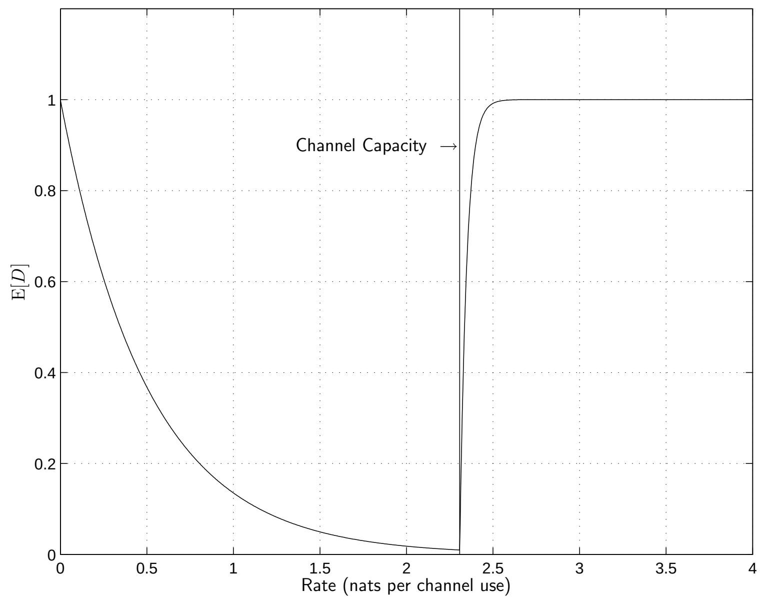

Figure 2.3. Expected mean-square distortion as a function of rate for an AWGN channel with SNR = 20 dB.

a Gaussian source over an AWGN channel, this yields

$$\begin{aligned} \mathrm{E}[D] &= e^{-2R}\Big|_{R=\frac{1}{2}\log(1+\mathrm{SNR})} \\ &= \frac{1}{1+\mathrm{SNR}} \quad (2.7) \\ &\approx \frac{1}{\mathrm{SNR}} \text{ for SNR} \gg 1. \quad (2.8) \end{aligned}$$

■ 2.2 Wireless Fading Channels

For most wireless settings, the simple AWGN channel model does not capture all of the effects of propagation through the communication medium. When there are multiple paths for electromagnetic radiation to propagate from the transmitter to the receiver, there will be several copies of the original signal at the destination. Each of these signals will have a different delay, τ_i, and attenuation, α_i, associated with them. The baseband equivalent received signal can be expressed in the form

$$y(t) = \sum_i \alpha_i e^{-j2\pi f_c \tau_i} x(t - \tau_i). \tag{2.9}$$

When these signals add constructively, the received signal will have greater power than if only a single copy of $x(t)$ was present. Alternatively, when the signals arrive at the destination such that they add destructively, the received signal can be extremely small, or essentially zero. When each path has essentially the same delay when compared to the symbol duration, this process is called *multiplicative fading* and introduces a significant challenge to the design of digital communication systems. When there are a large number of propagation paths, the central limit theorem can be applied, and the multiplicative fading can be modeled as a zero-mean circularly symmetric complex Gaussian (ZMCG) random process. The signal amplitude for each transmission will then be scaled by a Rayleigh distributed random variable (RV). The channel model can be adjusted to reflect this multiplicative factor:

$$y = a \cdot x + w. \tag{2.10}$$

A significant challenge with multimedia communication over fading channels is that the quality of the channel is continuously varying, making it difficult to ensure reliable communication at all times. An attractive means for improving performance is to spread the signal over space, time, or frequency so that with high probability at least some of the transmission will be successful. This concept, termed *diversity*, is discussed in detail by Proakis [2].

When the block length is long relative to variations in the channel, we can average over realizations of the fading coefficients, and guarantee reliable communication at a rate near the capacity of the channel given in (2.4). If the fading is too slow for this, however, separate source and channel coding is no longer optimal. For example, this is the case when the fading coefficient remains constant over an entire block length. Although we are often at liberty to choose the block length, in real-time communications there are stringent delay requirements that potentially prohibit us from increasing the block length beyond the duration of a single channel realization. We refer to this type of channel as a Rayleigh block-fading (BF) channel, or quasi-static Rayleigh fading. This can be expressed for the transmission of a block as

$$y_i = a \cdot x_i + w_i. \tag{2.11}$$

The BF channel is one of many communication environments of current interest that do not lend themselves to the Shannon theoretic separation of source and channel coding. More specifically, certain broadcast scenarios, packet based or network communications, real-time or delay-constrained communications, and many other settings require the encoder pair and decoder pair be designed jointly for optimal performance. The majority of work until the mid-90's was done either in source *or* channel coding, presenting a new challenge in wireless communication theory, which has since motivated many practical implementations of joint source-channel codes (see [6] for a thorough overview).

From a higher level, the failure of separate source and channel coding is due to the inherent nature of the separation. The source coder is designed under the assumption that its output will be available to the source decoder with no errors; a condition that may be impossible to meet for certain channels. When errors are present in the decoded bit stream the source decoder may fail completely, resulting in a mean-square error equal to the source variance. The basic solution is to design encoding schemes that degrade more gracefully as the quality of the channel decreases, a topic that has received recent attention.

■ 2.3 Related Research

An important problem of real-time communications over a slowly fading channel arises if the realized SNR of the channel is not known at the transmitter. Classic separation of source and channel coding results in optimal performance when the channel's SNR is known at the transmitter. However, the performance of the system, digital systems in particular, can degrade drastically if the actual SNR falls only slightly below the designed SNR. Additionally, any improvement in SNR does not result in a corresponding improvement in system performance. A code is said to be *robust* if it can perform optimally over a wide range of channel conditions, similar to the case of quasi-static fading. In order to design systems that perform well on these types of channels we must look at techniques that inherently offer some form of robustness. We now discuss several approaches that do this to a certain extent.

■ 2.3.1 Multiple Descriptions

An example of a source encoding scheme that offers multiple levels of performance consists of creating two (or more) distinct, yet complimentary, descriptions of the source, such that a lower quality reconstruction of the source can be made when any single description is available, and the quality of the reconstruction can be improved by additional descriptions. Gersho, Witsenhausen, Wolf, Wyner, Ziv, and Ozarow introduced this type of encoding, referred to as *multiple descriptions* (MD), at the 1979 IEEE Information Theory Workshop. Their initial worked contained in [7, 8, 9, 10] formalized the problem. More specifically, consider transmitting a source s using two descriptions of rate R_1 and R_2 over a channel that introduces some uncertainty into the received signals. The encoding is done such that: if only description 1 is decoded, the receiver can reconstruct s with distortion D_1; if only description 2 is decoded, the receiver can reconstruct s with distortion D_2; and if both descriptions are successfully decoded the receiver reconstructs s with distortion D_0. The main questions are: (1) what quintuples $(R_1, R_2, D_0, D_1, D_2)$ are achievable for a given distortion measure, and (2) how can we achieve a certain point in this set?

Since the problem of multiple descriptions was initially posed, there has been significant progress in both characterizing the achievable rate-distortion region and developing practical implementations for certain sources and channels. Initially, El Gamal & Cover [11] presented an achievable rate region for a discrete memoryless source with two descriptions and mean-square distortion measure, which Ozarow later proved to be optimal for Gaussian sources in [9]. The problem of multiple descriptions for the binary symmetric source with Hamming distortion was studied by Berger & Zhang [12, 13, 14], Ahlswede [15], Witsenhausen & Wyner [7], and Wolf, Wyner, and Ziv [8]. The problem of multiple descriptions for more general sources, distortion measures, and descriptions is yet to be solved.

$$R_1 > I(\mathsf{x}; \hat{\mathsf{x}}_1) \tag{2.12a}$$

$$R_2 > I(\mathsf{x}; \hat{\mathsf{x}}_2) \tag{2.12b}$$

$$R_1 + R_2 > I(\mathsf{x}; \hat{\mathsf{x}}_1, \hat{\mathsf{x}}_2, \hat{\mathsf{x}}_0) + I(\hat{\mathsf{x}}_1; \hat{\mathsf{x}}_2) \tag{2.12c}$$

Around the same time practical implementations of multiple description source coding were being developed [16, 17], a novel setting for its application emerged. The community realized that systems employing multiple transmit and receive antennas could be used to greatly improve the performance of a wireless system by providing significant diversity and multiplexing gains. Telatar derived the capacity for a multi-antenna Gaussian channel in [18], and showed that the multiple-input multiple-output (MIMO) system can be decomposed into independent parallel channels.

There have been several practical schemes that realize some of the capacity gains promised in [18] by exploiting either improved diversity, or increased degrees of freedom (spatial multiplexing). Zheng & Tse [19] showed that there exists a fundamental tradeoff between diversity and multiplexing gains. The optimal balance between diversity and multiplexing gains depends on the specific end-to-end metric of interest. For example, a MIMO system exploiting full diversity gains will support very reliable communication at a lower rate. Alternatively, optimal spatial multiplexing, by utilizing increased degrees of freedom, can support significantly higher rates at the cost of transmission reliability.

Another important question arises from the decomposition of a multi-antenna channel into independent parallel channels: What is the best way to exploit diversity using parallel channels when an end-to-end metric is of interest? Laneman et al. first addressed this question in [20] for both on-off channels and those exhibiting a continuous fading distribution. Prior to this work, most work studying the performance of multiple description source coding considered only on-off channel models in which each description is available at the receiver either error free or not at all. Laneman et al. examined the performance of multiple descriptions as a form of source coding diversity, using a more general framework encompassing a variety of fading models.

In [21], Laneman et al. established a simple means to compare system performance by considering how the expected distortion behaves at high SNR. Using this structure for the analysis, they showed that when source and channel decoding are done independently, the optimal form of diversity (source coding diversity vs. channel coding diversity) depends on the specific fading characteristics of the channel. Surprisingly, when decoding is done jointly, a system using only source coding diversity can perform as well as any of the other schemes analyzed, for all fading models considered. The examination of joint decoding is a first step towards an information theoretic understanding of the important synergy between source and channel coding. The performance of systems using complete joint encoding and decoding is not yet understood, an important problem which could provide valuable insights to understanding how to best merge source and channel coders. These notions will be examined further in this thesis.

■ 2.3.2 Successive Refinement

The analysis introduced by Laneman et al. was used by Gunduz [22] to introduce a protocol that offers some trade-off between spectral efficiency and diversity. This protocol relies on a special case of multiple descriptions called successive refinement (SR). Also referred to as layered or superposition coding, a dual-layered SR code can be considered the special case of MD with $D_2 = \sigma_s^2$, the source variance. Equivalently, SR source coding consists of breaking down the source descriptions into multiple stages, or layers, such that decoding each additional layer reduces the distortion. Furthermore, each layer beyond the first provides no useful information about the source without successful decoding of all lower level layers. Gunduz's SR protocol is analogous to sending the base layer over one channel, and transmitting the enhancement layer on another independent channel. Gunduz's results rely on the successive refinability property of a Gaussian source and are therefore not as general as those presented by Laneman et al. [21], where a wider class of sources are considered. Furthermore, there has been no characterization of the end-to-end distortion achievable using a successive refinement strategy over a single channel, or sending both base and enhancement layers over parallel channels.

A source is said to be *successively refinable* if a description exists as above, that also achieves the optimal distortion as each layer is decoded. Equitz & Cover [23] derived necessary and sufficient conditions for a source to be successively refinable.

They also gave several types of sources/distortion measures that meet these conditions, including a Gaussian source with mean-square distortion — a property that will be used extensively in this thesis. Rimoldi [24] generalized the results in [23] by finding the achievable rate region for a given pair of distortions, along with an interpretation of Equitz and Cover's successive refinability condition.

■ 2.3.3 Hybrid Digital-Analog

Another approach to improving performance through graceful degradation, called *systematic* communication, involves transmitting both uncoded and coded versions of the source. When the digital data cannot be decoded, a noisy version of the source is always available, reducing the threshold effect present in non-systematic communication. Shamai et al. [25] derived necessary and sufficient conditions for when systematic methods perform optimally. Mittal & Phamdo [26] designed nearly robust joint source-channel codes using systematic hybrid digital-analog (HDA) techniques. Although their results were presented in the context of broadcasting and robust communication, the general concepts can be applied to certain fading scenarios, as will be done in Chapter 4. Also, previous analysis of HDA systems has not considered a quasi-static fading channel, or independent parallel channels.

CHAPTER 3

COMMUNICATION OVER A SINGLE CHANNEL

As mentioned in Chapter 2, although the model given by (2.10) permits straight forward analysis, for many real communication systems it is overly simplified. For example, a delay constraint may prevent the block length N from increasing large enough to code over variations in the channel. Similarly, the fading may be too slow to model each coefficient as an i.i.d. random variable. In order to incorporate this into our model, we now consider the fading coefficient, a, to remain fixed over a single block, and to be chosen independently from a complex Gaussian distribution in separate blocks. We refer to this model as a quasi-static Rayleigh fading channel, with corresponding channel model given by (2.11). Since the fading is now a non-ergodic random process, separate source and channel coding may no longer be optimal, and it is of interest to consider alternative techniques. It should be noted that, in the general case of non-ergodic fading, the best method for transmission is unknown, so we turn to analyzing several schemes and then comparing their performance at high SNR.

This channel can be thought of in an alternative context. For each block the fading coefficient is a constant but unknown random variable (RV), thus the channel becomes an AWGN channel with unknown SNR. Although the realized channel SNR is unknown, we do know the PDF of the SNR, and can exploit this fact to minimize the average distortion over all possible channel realizations. This is equivalent to a

Gaussian broadcast channel, with a continuum of users, and the users' SNR profile is Rayleigh distributed. The goal here is not to characterize the achievable distortions for each user, as is standard for broadcast channels, but to minimize the expected distortion averaged over all users. Ideally we would be interested in a complete characterization of the distortion, such as its PDF. If the block length N is small enough such that the user's perception of distortion is related to the average over each realization, considering only the average distortion may be sufficient.

For all of the systems studied we obtain closed form expressions for the expected distortion. Unfortunately, the evaluation of these equations often require some form of numerical optimization or integration, limiting the potential for purely analytical comparison. To facilitate a visual comparison of each scheme's performance, we perform the compuations and plot the average distortions for a range of SNRs.

In order to facilitate a tractable analytical comparison between systems, we can consider how the expected distortion behaves for large SNR. To do this we consider the expected distortion for asymptotically high SNR, where it behaves as $\mathrm{E}[D] = C \cdot \mathrm{SNR}^{-\Delta}$. In this regime we can partially characterize a scheme's performance with a single metric, the distortion exponent defined as

$$\Delta := -\lim_{\mathrm{SNR}\to\infty} \frac{\log \mathrm{E}[D]}{\log \mathrm{SNR}}. \tag{3.1}$$

The notion of the distortion exponent as used in this context was introduced by Laneman et al. in [21], and has been further utilized in [22]. High SNR approximations may not completely describe a system's performance, but as is evident in Figure 3.7 and Figure **??**, the systems under consideration begin to display asymptotic behavior even at moderate values of SNR. In order to account for the channel's bandwidth, it is assumed that the encoder maps K source samples to N real channel inputs, or $N/2$ complex channel inputs. The bandwidth expansion ratio N/K is denoted as $L := N/K$. Therefore, a bandwidth expansion ratio of $L = 1$ corresponds to mapping each real source sample to a real channel input, or equivalently mapping each pair of real source samples to a single complex channel use.

■ 3.1 Analog Transmission

We begin by analyzing two obvious ways to communicate on this channel. The first is uncoded, or analog, transmission (SISO − A). Analog transmission can be considered the simplest form of communication in the sense that it requires no encoding or decoding. It does, however, require knowledge of the channel's SNR at the receiver, and the estimate of $\mathbf{s}$ that minimizes the mean-square distortion must also be computed. Strictly speaking, analog transmission is a form of joint-source channel coding because there is no intermediate mapping of source samples onto a finite alphabet prior to the construction of channel symbols. Likewise, received vectors are never mapped onto a finite alphabet prior to performing the final source reconstruction. Formally, for $L = 1$ this can be expressed as

$$x = s \tag{3.2a}$$

$$\hat{s} = \mathsf{D}_{\hat{s} \leftarrow y}(y) \tag{3.2b}$$

For $L > 1$ the encoder does not use the additional bandwidth available; correspondingly, the decoder ignores the unused bandwidth when forming the source reconstruction.

The receiver's goal is to form an estimate of the source symbol s such that the mean-square distortion between the original source sample, s, and its reconstruction, $\hat{s}$, is minimized. This is done using minimum mean-square error (MMSE) estimation of s as a function of the received data, y, e.g., $\hat{s} = \hat{s}_{\text{MMSE}}(y)$. Since y is a linear combination of independent Gaussian RVs (s and w), y and s are jointly Gaussian. It is known that for the special case of estimating a RV that is jointly Gaussian with the observation, the MMSE estimate is a linear function of the data, i.e.,

$$\hat{s}_{\mathrm{MMSE}}(y) = \alpha \cdot y + \beta \tag{3.3}$$

where α and β are constants. This means the the MMSE estimate of s, $\hat{s}_{\mathrm{MMSE}}(y)$, coincides with the linear least-squares estimate, $\hat{s}_{\mathrm{LLS}}(y)$, for which closed form expressions for both the estimate and the resulting distortion exist. These are given by

$$\hat{s}_{\mathrm{LLS}}(y) = \mu_s + \frac{\Lambda_{sy}(y - \mu_y)}{\Lambda_y} \tag{3.4}$$

$$\Lambda_{\mathrm{LLS}} = \Lambda_s - \frac{\Lambda_{sy}^2}{\Lambda_y} \tag{3.5}$$

Considering a unit-variance source, for each transmitted source sample the received signal is

$$y = a\sqrt{\mathrm{SNR}} \cdot x + w, \tag{3.6}$$

where a is the complex Gaussian fading parameter and $w \sim \mathcal{N}(0,1)$ is additive Gaussian noise. Note that we have normalized w to be unit-variance so that the SNR is equal to the available power, P. Substituting

$$\mu_s = 0 \tag{3.7}$$

$$\mu_y = 0 \tag{3.8}$$

$$\Lambda_{sy} = a^*\sqrt{\mathrm{SNR}} \tag{3.9}$$

$$\Lambda_y = |a|^2\mathrm{SNR} + 1 \tag{3.10}$$

in (3.4), we have

$$\hat{s}_{\mathrm{LLS}}(y) = \frac{a^*\sqrt{\mathrm{SNR}}}{|a|^2\mathrm{SNR} + 1} \cdot y. \tag{3.11}$$

The resulting conditional distortion is then found to be

$$\Lambda_{\mathrm{LLS}}(a) = \frac{1}{|a|^2\mathrm{SNR} + 1}. \tag{3.12}$$

Note that (3.12) is the distortion for a specific realization of the fading coefficient, i.e. it is a function of the channel realization. In order to obtain the expected value of the distortion, we must now average over all possible channel realizations. Since a is complex Gaussian, $|a|^2$ is an exponential random variable (RV), and the average can be found as follows:

$$\begin{aligned} \mathrm{E}[D] &= \mathrm{E}_{a}\left[\frac{1}{1+|a|^2\mathrm{SNR}}\right] && (3.13) \\ &= \int_0^\infty \frac{e^{-\lambda}}{1+\lambda\mathrm{SNR}}d\lambda. && (3.14) \end{aligned}$$

The integral in (3.14) can be computed numerically for a specific value of SNR, and is plotted in Figure 3.7. Notice that uncoded transmission does not rely on knowledge of the channel's average SNR at the transmitter, a characteristic unique to this scheme. As will be shown in Section 3.4, analog transmission achieves the lowest distortion possible on this channel. This is a direct extension of the classic results for uncoded transmission being optimal on an AWGN channel with matched bandwidths. This is not the case if the bandwidth of the source differs from that of the channel, or for parallel block-fading channels considered in Section 4.2.

In order to perform a high SNR analysis of uncoded transmission, we begin by rewriting (3.14) using the substitution $t = \frac{1+\lambda\mathrm{SNR}}{\mathrm{SNR}}$.

$$\begin{aligned} \mathrm{E}[D] &= \int_0^\infty \frac{e^{-\lambda}}{1+\lambda\mathrm{SNR}}d\lambda && (3.15) \\ &= \int_{1/\mathrm{SNR}}^\infty \frac{1}{t\mathrm{SNR}}\exp\left[\frac{1}{\mathrm{SNR}}-t\right]dt && (3.16) \\ &= \frac{1}{\mathrm{SNR}}e^{1/\mathrm{SNR}}E_1\left(\frac{1}{\mathrm{SNR}}\right), && (3.17) \end{aligned}$$

where $E_1(\cdot)$ is the exponential integral

$$E_1(x) := \int_x^\infty \frac{e^{-t}}{t}dt. \qquad (3.18)$$

Using the inequalities

$$\frac{1}{2x}\ln(1+2x) < \frac{1}{x}e^{1/x}E_1\left(\frac{1}{x}\right) < \frac{1}{x}\ln(1+x) \tag{3.19}$$

found as Eq. 5.1.20 in [27], we have an upper and lower bound on the high SNR approximation of the distortion. Computing the distortion exponent for the lower bound yields

$$\Delta_{\mathrm{SISO-A}} < -\lim_{\mathrm{SNR}\to\infty}\frac{\log\left[\frac{1}{2\mathrm{SNR}}\ln(1+2\mathrm{SNR})\right]}{\log\mathrm{SNR}} \tag{3.20}$$

$$= 1. \tag{3.21}$$

Using the upper bound in (3.19) we have

$$\Delta_{\mathrm{SISO-A}} > -\lim_{\mathrm{SNR}\to\infty}\frac{\log\left[\frac{1}{\mathrm{SNR}}\ln(1+\mathrm{SNR})\right]}{\log\mathrm{SNR}} \tag{3.22}$$

$$= 1. \tag{3.23}$$

Therefore

$$\boxed{\Delta_{\mathrm{SISO-A}} = 1}. \tag{3.24}$$

■ 3.2 Separate Source and Channel Coding

To illustrate the sub-optimality of separate source and channel coding (SISO − D), we now consider the simple case of using a source encoder/decoder $\mathsf{E}_{m\leftarrow s}(\cdot)/\mathsf{D}_{\hat{s}\leftarrow\hat{m}}(\cdot)$ designed independently of the channel encoder/decoder $\mathsf{E}_{x\leftarrow m}(\cdot)/\mathsf{D}_{\hat{m}\leftarrow y}(\cdot)$. The overall encoder and decoder are given by

$$x = \mathsf{E}_{x\leftarrow s}(s) = \mathsf{E}_{x\leftarrow m}(\mathsf{E}_{m\leftarrow s}(s)) \tag{3.25a}$$

$$\hat{s} = \mathsf{D}_{\hat{s}\leftarrow y}(y) = \begin{cases} \mathsf{D}_{\hat{s}\leftarrow\hat{m}}(\mathsf{D}_{\hat{m}\leftarrow y}(y)), & \mathsf{D}_{\hat{m}\leftarrow y}(y) \neq 0 \\ \mathrm{E}[s], & \text{otherwise} \end{cases}. \tag{3.25b}$$

Recall that when the fading is ergodic, this architecture performs as well as if the source and channel encoder/decoder are designed jointly. Furthermore, in the ergodic case, the average distortion could be computed by evaluating the source's distortion rate function $D_s(R)$ at $R = C$, where C is the channel capacity. When the fading is non-ergodic, the mutual information $I(x;y)$ is a random variable, and the Shannon capacity of the channel is zero. We must therefore turn to alternative techniques to compute the average distortion. To facilitate this computation we adopt the notion of outage probability and wish to find the probability that the mutual information falls below the chosen coding rate R, i.e.,

$$\Pr[\text{outage}] := \Pr\left[I(x;y) < R\right]. \tag{3.26}$$

For digital communication on the channel under consideration we compute P_{out} as follows:

$$\begin{aligned}
\mathrm{P}_{\text{out}}(R, \text{SNR}) &= \Pr\left[I(x;y) < R\right] && (3.27)\\
&= \Pr\left[\frac{L}{2}\log\left(1 + |a|^2\text{SNR}\right) < R\right] && (3.28)\\
&= \Pr\left[|a|^2 < \frac{e^{2R/L} - 1}{\text{SNR}}\right] && (3.29)\\
&= \int_0^{\frac{e^{2R/L}-1}{\text{SNR}}} e^{-\lambda} d\lambda && (3.30)\\
&= 1 - \exp\left(-\frac{e^{2R/L} - 1}{\text{SNR}}\right). && (3.31)
\end{aligned}$$

Note that the outage probability, and hence the expected distortion, is a function of both R and SNR.

When a given channel realization prohibits us from decoding the received codeword, which will occur with probability P_{out}, we reconstruct to the source mean, and thus $\mathrm{E}[D|\text{outage}] = \sigma_s^2$. With probability $1 - \mathrm{P}_{\text{out}}$ we will be able to decode the received codeword, resulting in a distortion of $\mathrm{E}[D|\overline{\text{outage}}] = \sigma_s^2 \cdot e^{-2R}$. Using the total probability law we can compute the average distortion as

$$\mathrm{E}[D] = \mathrm{E}[D|\text{outage}] \cdot \mathrm{P}_{\text{out}} + \mathrm{E}[D|\overline{\text{outage}}] \cdot (1 - \mathrm{P}_{\text{out}}). \tag{3.32}$$

For the system described by (3.25), the expected distortion can be expressed as

$$\mathrm{E}[D(R,\mathrm{SNR})] = \sigma_s^2 \cdot \left[1 - \exp\left(-\frac{e^{2R/L}-1}{\mathrm{SNR}}\right)\right] + \sigma_s^2 \cdot e^{-2R} \cdot \exp\left(-\frac{e^{2R/L}-1}{\mathrm{SNR}}\right). \quad (3.33)$$

The performance achieved by the above digital scheme is a function of the rate at which we choose to communicate; therefore, it makes sense to choose R so as to minimize the expected distortion for a given SNR. This leads to the final expression for the average distortion of separate source and channel coding:

$$\mathrm{E}[D] = \min_R \left\{ \sigma_s^2 \cdot \left[1 - \exp\left(-\frac{e^{2R/L}-1}{\mathrm{SNR}}\right)\right] + \sigma_s^2 \cdot e^{-2R} \cdot \exp\left(-\frac{e^{2R/L}-1}{\mathrm{SNR}}\right) \right\}. \quad (3.34)$$

The minimization in (3.34) is performed numerically for specific values of SNR and a unit variance source. The optimal rate as a function of SNR is shown in Figure 3.1, and the minimum distortion in Figure 3.7. It is clear that the average distortion for separate source and channel coding is strictly greater than that of uncoded transmission for all values of SNR shown in Figure 3.7.

In addition to the rate-optimized digital scheme's inferior performance relative to uncoded transmission, the optimal source coding rate is a specific function of the channel's average SNR. Therefore, in order to achieve the performance given by (3.33), the source coder must operate at different rates for different average SNRs, significantly increasing complexity and requiring knowledge of the channel's average SNR at the transmitter. The cost of operating at a fixed rate over a range of average SNRs can be considerable for values of SNR more than about 8 dB from the designed SNR, as shown in Figure 3.2. For an actual SNR within 5 dB of the designed SNR, the incurred distortion is typically less than 1 dB. This offers the designer a range of SNRs of about 10 dB, over which the performance is still nearly optimal.

Figure 3.2 also illustrates the notion of rate and outage limited regimes. For SNRs below the designed SNR, an outage occurs with higher probability than is

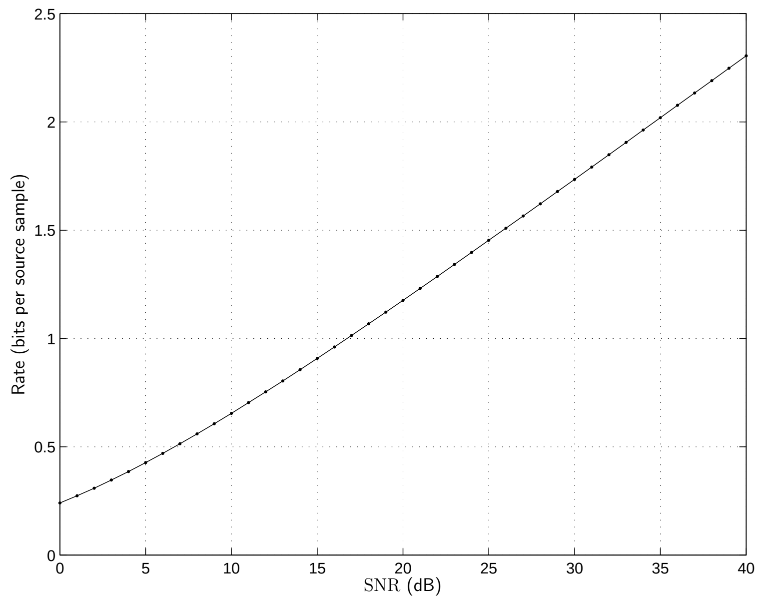

Figure 3.1. Optimal channel coding rate for separate source and channel coding (Section 3.2) as a function of SNR, found numerically ($L = 1$).

optimal. Alternatively, for SNRs above the designed SNR, there is rarely an outage event, but the rate is lower than what the channel could usually support. We refer to the range of SNR where the low rate dominates the system's performance as the *rate-limited* regime. If outage is the dominating contributor to source distortion, we are operating in the *outage-limited* regime.

We now compute the distortion exponent for separate source and channel coding. As can be seen in Figure 3.1, the optimal rate scales linearly with $\log \text{SNR}$ for large SNR, thus the optimal rate can be approximated as

$$R_{opt} = r \log \text{SNR}, \tag{3.35}$$

where r is the multiplexing gain [28], a constant independent of SNR yet to be determined. If $R = r \log \text{SNR}$, the outage probability is

$$\text{P}_{\text{out}} = 1 - \exp\left(-\frac{\text{SNR}^{2r/L} - 1}{\text{SNR}}\right). \tag{3.36}$$

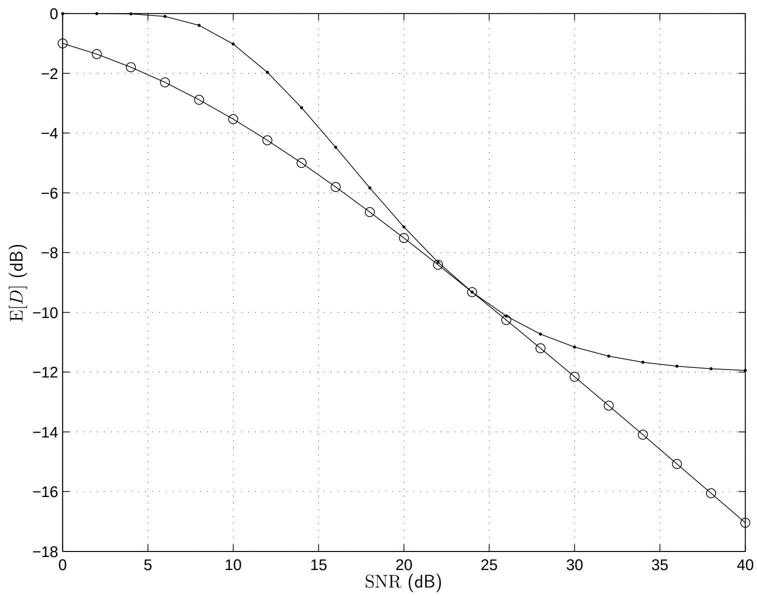

Figure 3.2. Average distortion as a function of SNR for fixed rate ($\cdot$) and rate optimized ($\circ$) separate source and channel coding (Section 3.2) ($L = 1$).

To ensure $\mathrm{E}[D] \to 0$ as $\mathrm{SNR} \to \infty$, the probability of outage must also go to zero. We account for this by imposing the constraint that $r \in [0, L/2)$. Next we use the well known inequality

$$1 - e^{-x} < x, \tag{3.37}$$

which is asymptotically tight for small x (large SNR), to approximate $\mathrm{P}_{\mathrm{out}}$ as

$$\mathrm{P}_{\mathrm{out}} \approx \frac{\mathrm{SNR}^{2r/L} - 1}{\mathrm{SNR}}. \tag{3.38}$$

Finally, using (3.38) in (3.34) along with the fact that $(1 - \mathrm{P}_{\mathrm{out}}) \to 1$, we have

$$\begin{aligned} \mathrm{E}[D(r, \mathrm{SNR})] &= \frac{\mathrm{SNR}^{2r/L} - 1}{\mathrm{SNR}} + \mathrm{SNR}^{-2r} &(3.39)\\ &= \mathrm{SNR}^{2r/L-1} + \mathrm{SNR}^{-2r}, &(3.40) \end{aligned}$$

where (3.40) follows because either $\mathrm{SNR}^{2r/L-1}$ or SNR^{-2r} will decay slower than SNR^{-1} for $r \in [0, L/2)$. At high SNR the largest exponent will dominate, thus we wish to choose r to minimize the maximum exponent in (3.40). More explicitly:

$$\begin{aligned}
\Delta_{\text{SISO-D}} &= -\lim_{\text{SNR}\to\infty} \frac{\log\left(\text{SNR}^{2r/L-1} + \text{SNR}^{-2r}\right)}{\log \text{SNR}} && (3.41)\\
&= \min_r \max\left(2r/L - 1, -2r\right) && (3.42)\\
&= \frac{L}{L+1}, && (3.43)
\end{aligned}$$

where the optimal multiplexing gain is

$$r = \frac{L}{2(L+1)}. \tag{3.44}$$

The above digital scheme's sub-optimal performance is a result of its having two performance regimes: for certain channel realizations we are unable to decode the received codeword at all, and for all other channel realizations we are transmitting at a rate lower than the channel realization can support, e.g., at a rate below the realized mutual information. In other words, $\text{E}[D|\mathsf{a}]$ can take on only two possible values, compared to the continuum of values possible with analog transmission.

■ 3.3 Successive Refinement

In order to partially combat the characteristics of rate-optimized digital transmission that result in suboptimal performance, we consider a successive refinement scheme. Since successive refinement coding is a *layered* scheme, the $\text{E}[D|\mathsf{a}]$ can take on more than two values, giving it the potential to decrease the average distortion. We first consider a dual-layer successive refinement code, where the refinement layer is superimposed on the base layer and power allocation between the layers is optimized to minimize the expected distortion. The base layer is encoded at a rate R_B with power $\alpha \cdot \text{SNR}$, and the enhancement layer is encoded at rate R_E with power $(1 -$

$\alpha) \cdot \mathrm{SNR}$. This scheme's encoder/decoder pair is defined as

$$x = \mathsf{E}_{x\leftarrow s}(s) = \mathsf{E}_{x_B\leftarrow m_B}(\mathsf{E}_{m_B\leftarrow s}(s)) + \mathsf{E}_{x_E\leftarrow m_E}(\mathsf{E}_{m_E\leftarrow s}(s)) \tag{3.45a}$$

$$\hat{s} = \mathsf{D}_{\hat{s}\leftarrow y}(y) = \begin{cases} \mathsf{D}_{\hat{m}\leftarrow y}(\mathsf{D}_{\hat{m}\leftarrow y}(s)), & \mathsf{D}_{\hat{m}\leftarrow y}(y) \neq 0 \\ \mathrm{E}[s], & \text{otherwise} \end{cases}. \tag{3.45b}$$

The received signal is

$$y_i = a\sqrt{\mathrm{SNR}} \cdot [\sqrt{\alpha} x_{B,i} + \sqrt{1-\alpha} x_{E,i}] + z_i \tag{3.46}$$

The decoding is performed as follows: The receiver first attempts to decode the base layer treating the refinement layer as additive noise. If the base layer is successfully decoded, the receiver subtracts its estimate of the transmitted codeword from the received signal and attempts to decode the refinement layer. The average distortion as a function of α, R_B, and R_E can be expressed as

$$\begin{aligned} \mathrm{E}[D(R_B, R_E, \alpha)] &= \Pr[\mathcal{B}_{\text{out}}] + e^{-2R_B} \cdot \Pr[\overline{\mathcal{B}_{\text{out}}}, \mathcal{E}_{\text{out}}] + e^{-2(R_B+R_E)} \cdot \Pr[\overline{\mathcal{B}_{\text{out}}}, \overline{\mathcal{E}_{\text{out}}}] \\ &= \Pr[\mathcal{B}_{\text{out}}] + e^{-2R_B} \cdot \Pr[\overline{\mathcal{B}_{\text{out}}}] \Pr[\mathcal{E}_{\text{out}} | \overline{\mathcal{B}_{\text{out}}}] \\ &\quad + e^{-2(R_B+R_E)} \cdot \Pr[\overline{\mathcal{B}_{\text{out}}}] \Pr[\overline{\mathcal{E}_{\text{out}}} | \overline{\mathcal{B}_{\text{out}}}], \end{aligned} \tag{3.47}$$

where $\mathcal{B}_{\text{out}}$ and $\mathcal{E}_{\text{out}}$ denote the events of a base layer and enhancement layer outage, respectively. In order to compute $\Pr[\mathcal{B}_{\text{out}}]$ we must first find $\Pr[I(\mathbf{x}_B; \mathbf{y}) < R_B]$. Since the received base layer power is $\alpha |a|^2 P$ and the received noise power is $(1 - \alpha)|a|^2 P + N_0$, we can express the base layer's effective SNR as

$$\mathrm{SNR}_B = \frac{\alpha |a|^2 P}{(1-\alpha)|a|^2 P + N_0} \tag{3.48}$$

and thus

$$\begin{aligned}
\Pr[\mathcal{B}_{\text{out}}] &= \Pr\left[I(\mathbf{x}_B;\mathbf{y}) < R_B\right] && (3.49)\\
&= \Pr\left\{\frac{L}{2}\log\left[1+\frac{\alpha|a|^2P}{(1-\alpha)|a|^2P+N_0}\right] < R_B\right\} && (3.50)\\
&= \Pr\left\{|a|^2 < \frac{e^{2R_B/L}-1}{\text{SNR}\left[1-(1-\alpha)e^{2R_B/L}\right]}\right\}\\
&= 1-\exp\left\{-\frac{e^{2R_B/L}-1}{\text{SNR}\left[1-(1-\alpha)e^{2R_B/L}\right]}\right\}. && (3.51)
\end{aligned}$$

Note that (3.51) is only valid for

$$\begin{aligned}
1-(1-\alpha)e^{2R_B/L} &> 0\\
\Rightarrow \alpha &> 1-e^{-2R_B/L}. && (3.52)
\end{aligned}$$

We must ensure the condition given in (3.52) is met, because for

$$\alpha < 1-e^{-2R_B/L} \qquad (3.53)$$

$\Pr[\mathcal{B}_{\text{out}}] = 1$.

In order to evaluate (3.47) we must also find $\Pr[\mathcal{E}_{\text{out}}|\overline{\mathcal{B}_{\text{out}}}]$, which is done as follows:

$$\begin{aligned}
\Pr\left[\mathcal{E}_{\text{out}}|\overline{\mathcal{B}_{\text{out}}}\right] &= \Pr\left[I(\mathbf{x}_E;\mathbf{y}|\mathbf{x}_B) < R_E\,|I(\mathbf{x}_B;\mathbf{y}) > R_B\right] && (3.54)\\
&= \Pr\left\{\frac{L}{2}\log\left[1+(1-\alpha)|a|^2\text{SNR}\right] < R_E\right|\\
&\qquad \left.\frac{L}{2}\log\left[1+\frac{\alpha|a|^2P}{(1-\alpha)|a|^2P+N_0}\right] > R_B\right\} && (3.55)\\
&= \Pr\left\{|a|^2 < \frac{e^{2R_E/L}-1}{(1-\alpha)\text{SNR}}\right|\\
&\qquad \left.|a|^2 > \frac{e^{2R_B/L}-1}{\text{SNR}\left[1-(1-\alpha)e^{2R_B/L}\right]}\right\} && (3.56)\\
&= \Pr\left\{|a|^2 < \frac{e^{2R_E/L}-1}{(1-\alpha)\text{SNR}} - \frac{e^{2R_B/L}-1}{\text{SNR}\left[1-(1-\alpha)e^{2R_B/L}\right]}\right\} && (3.57)\\
&= 1-\exp\left\{\frac{e^{2R_B/L}-1}{\text{SNR}\left[1-(1-\alpha)e^{2R_B/L}\right]} - \frac{e^{2R_E/L}-1}{(1-\alpha)\text{SNR}}\right\}. && (3.58)
\end{aligned}$$

Note that (3.57) follows from (3.56) by exploiting the memoryless property of an exponential RV, and is only valid for

$$\alpha > \frac{e^{2R_E/L}(1 - e^{2R_B/L})}{1 - e^{2(R_B+R_E)/L}}; \tag{3.59}$$

otherwise $\Pr\left[\mathcal{E}_{\text{out}}|\overline{\mathcal{B}_{\text{out}}}\right] = 1$.

The final expression for the expected distortion is given as

$$\begin{aligned}
\mathrm{E}[D] = \min_{\alpha, R_B, R_E} \quad & 1 - \exp\left\{-\frac{e^{2R_B/L} - 1}{\mathrm{SNR}\left[1 - (1-\alpha)e^{2R_B/L}\right]}\right\} \\
& + e^{-2R_B} \cdot \exp\left\{-\frac{e^{2R_B/L} - 1}{\mathrm{SNR}\left[1 - (1-\alpha)e^{2R_B/L}\right]}\right\} \\
& \cdot \left(1 - \exp\left\{\frac{e^{2R_B/L} - 1}{\mathrm{SNR}\left[1 - (1-\alpha)e^{2R_B/L}\right]} - \frac{e^{2R_E/L} - 1}{(1-\alpha)\mathrm{SNR}}\right\}\right) \\
& + e^{-2(R_B+R_E)} \cdot \exp\left\{-\frac{e^{2R_B/L} - 1}{\mathrm{SNR}\left[1 - (1-\alpha)e^{2R_B/L}\right]}\right\} \\
& \cdot \exp\left\{\frac{e^{2R_B/L} - 1}{\mathrm{SNR}\left[1 - (1-\alpha)e^{2R_B/L}\right]} - \frac{e^{2R_E/L} - 1}{(1-\alpha)\mathrm{SNR}}\right\}.
\end{aligned} \tag{3.60}$$

The optimal α, R_B, and R_E are found numerically for a range of SNRs. Figure 3.3 shows the optimal power allocation factor; Figure 3.4 shows the optimal rates, and Figure 3.7 shows the resultant distortion using the optimal α, R_B, and R_E.

It is interesting to note that the optimal rates scale linearly with $\log \mathrm{SNR}$, as was the case for rate-optimized digital communication. Using this fact, we develop a high SNR approximation to (3.60). For high SNR, the optimal rates obey

$$R_B = r_B \log \mathrm{SNR} \tag{3.61}$$

$$R_E = r_E \log \mathrm{SNR}. \tag{3.62}$$

As can be seen in Figure 3.5, for high SNR the optimal α satisfies

$$\alpha = 1 - \mathrm{SNR}^{-\hat{\alpha}}, \tag{3.63}$$

where the constant $\hat{\alpha}$ determines the exponential rate at which more power is allo-

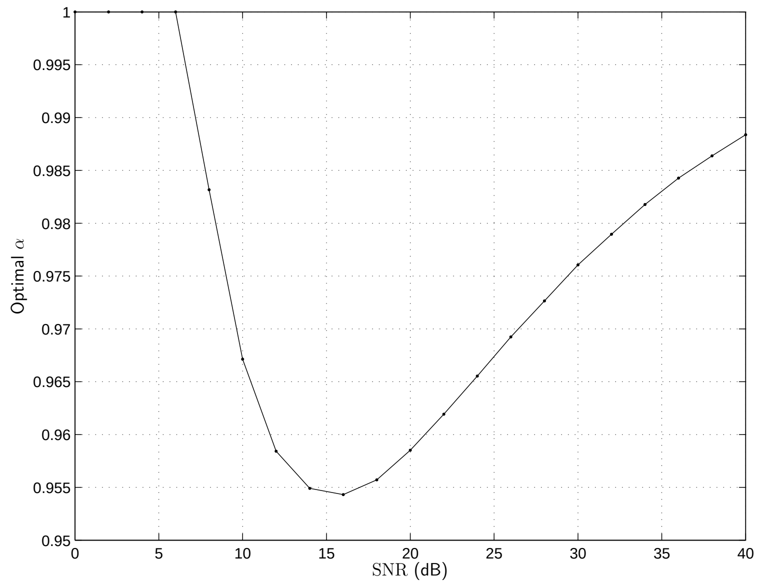

Figure 3.3. Optimal power allocation factor, α, for superposition successive refinement coding (Section 3.3) with $L = 1$.

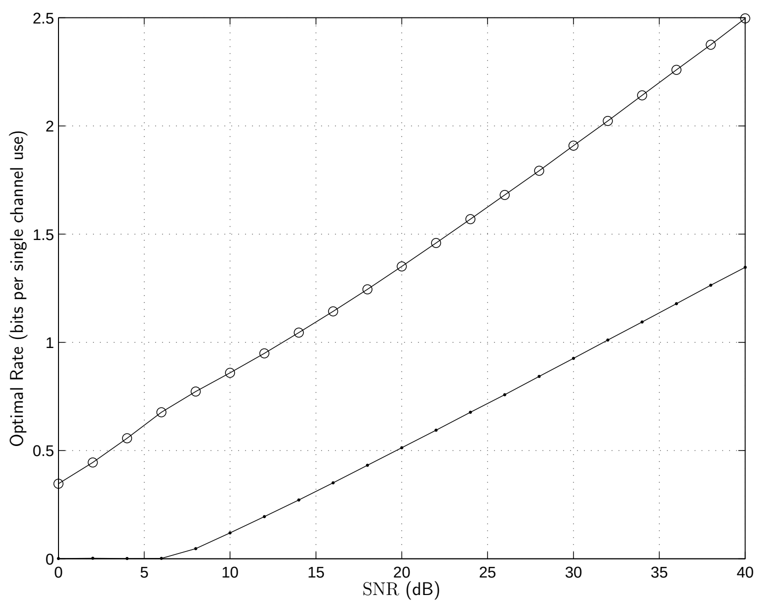

Figure 3.4. Optimal rates for superposition successive refinement coding (Section 3.3) found numerically with $L = 1$. R_B is shown with ($\circ$), R_E is shown with ($\cdot$).

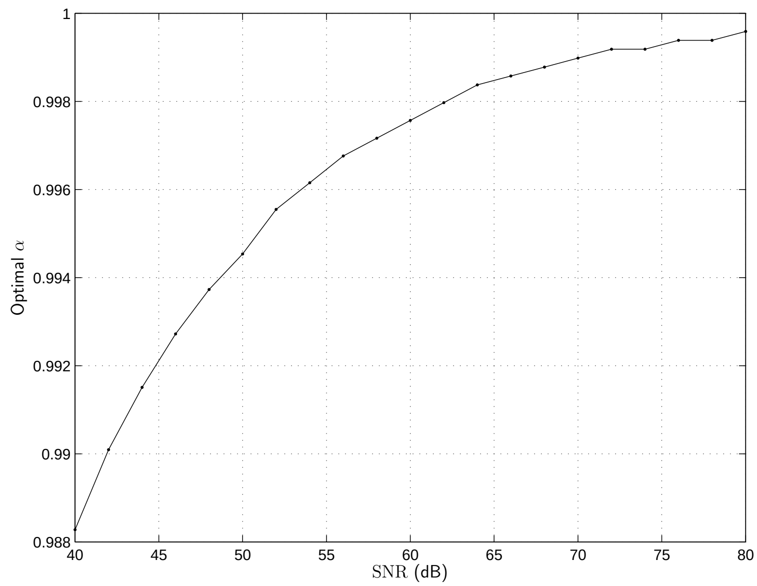

Figure 3.5. Optimal power allocation factor, α, for superposition successive refinement coding (Section 3.3) with $L = 1$ in the high SNR regime.

cated to the base layer. Then (3.52) becomes

$$\alpha > \frac{2r_B}{L}. \tag{3.64}$$

Using (3.61), (3.63), and (3.37) we can approximate $\Pr[\mathcal{B}_{\text{out}}]$ as

$$\begin{aligned} \Pr[\mathcal{B}_{\text{out}}] &= \frac{\text{SNR}^{2r_B/L} - 1}{\text{SNR}\left(1 - \text{SNR}^{-\hat{\alpha}}\text{SNR}^{2r_B/L}\right)} && (3.65) \\ &= \frac{\text{SNR}^{2r_B/L-1}}{1 - \text{SNR}^{2r_B/L-\hat{\alpha}}} && (3.66) \\ &= \text{SNR}^{2r_B/L-1} && (3.67) \end{aligned}$$

Similarly,

$$\begin{aligned} \Pr\left[\mathcal{E}_{\text{out}}|\overline{\mathcal{B}_{\text{out}}}\right] &= \frac{\text{SNR}^{2r_E/L} - 1}{\text{SNR}^{-\hat{\alpha}}\text{SNR}} - \frac{\text{SNR}^{2r_B/L} - 1}{\text{SNR}\left(1 - \text{SNR}^{-\hat{\alpha}}\text{SNR}^{2r_B/L}\right)} && (3.68) \\ &= \text{SNR}^{2r_E/L-1+\hat{\alpha}} - \text{SNR}^{2r_B/L-1} && (3.69) \\ &= \text{SNR}^{2r_E/L-1+\hat{\alpha}}, && (3.70) \end{aligned}$$

where (3.70) follows (3.69) because (3.52) implies $2r_E/L - 1 + \hat{\alpha} > 2r_B/L - 1$, and the largest exponent will dominate at high SNR. Finally, the expected distortion for asymptotically high SNR is

$$\mathrm{E}[D(r_B, r_E, \hat{\alpha})] = \mathrm{SNR}^{2r_B/L-1} + \mathrm{SNR}^{-2r_B}\mathrm{SNR}^{2r_E/L-1+\hat{\alpha}} + \mathrm{SNR}^{-2(r_B+r_E)} \qquad (3.71)$$

We must now minimize (3.71) over $r_B \in [0, 1)$, $r_E \in [0, 1)$, and $\hat{\alpha} \in [0, 1)$, subject to $\hat{\alpha} > 2r_B/L$. This is done by equating exponents and solving, resulting in

$$r_B = \frac{L}{2(L^2 + L + 1)} \qquad (3.72)$$

$$r_E = \frac{L^2}{2(L^2 + L + 1)} \qquad (3.73)$$

$$\hat{\alpha} = \frac{1}{L^2 + L + 1}. \qquad (3.74)$$

The optimal values for α, R_B, and R_E give a distortion exponent of

$$\boxed{\Delta_{SR} = \frac{L(L+1)}{L^2 + L + 1}}, \qquad (3.75)$$

versus $L/(L + 1)$ for standard digital transmission.

■ 3.4 A Lower Bound on the Achievable Distortion

In order to assess the quality of any system, it is convenient to have a sense of the fundamental limits on performance. We now derive a lower bound on the achievable distortion by considering the case in which the transmitter has complete knowledge of each realized fading coefficient. A system that does not have channel state information at the transmitter can do no better than one that does, and so this assumption provides us with a simple bound. When the transmitter has knowledge of a for each block, the channel becomes a conditionally AWGN channel with known SNR. This allows the transmitter to compute the realized mutual information

$$I(\mathbf{x}; \mathbf{y}) = \frac{L}{2} \log \left(1 + |\mathsf{a}|^2 \mathrm{SNR}\right), \qquad (3.76)$$

and hence the maximal rate, R, at which the channel realization can support reliable communication. For a specific a, the expected distortion can be found by evaluating the rate distortion function of the source at a rate equal to the realized mutual information, to obtain

$$\begin{aligned}\mathrm{E}[D|\mathsf{a}] &= e^{-2R}\Big|_{R=\frac{L}{2}\log(1+|\mathsf{a}|^2\mathrm{SNR})} \\ &= \frac{1}{(1+|\mathsf{a}|^2\mathrm{SNR})^L}. \end{aligned} \tag{3.77}$$

Averaging over all possible channel realizations leads to

$$\mathrm{E}[D] = \int_0^\infty \frac{e^{-\lambda}}{(1+\lambda\mathrm{SNR})^L} d\lambda \tag{3.78}$$

$$= \frac{e^{1/\mathrm{SNR}}}{\mathrm{SNR}} \int_1^\infty \frac{e^{-t/\mathrm{SNR}}}{t^L} dt. \tag{3.79}$$

Note that (3.79) follows from (3.78) using the substitution $t = 1+\lambda\mathrm{SNR}$. For $L = 1$ (3.78) is the same as (3.17), and therefore

$$\Delta_{\mathrm{SISO-CSIT}} = 1 \text{ for } L = 1. \tag{3.80}$$

By limiting ourselves to integer values of L, we can also express (3.79) as

$$\mathrm{E}[D] = \frac{e^{1/\mathrm{SNR}}}{\mathrm{SNR}} E_L\left(\frac{1}{\mathrm{SNR}}\right) \text{ for } L = 1, 2, 3, ... \tag{3.81}$$

where [27]

$$E_n(z) := \int_1^\infty \frac{e^{-zt}}{t^n} dt \text{ for } n = 0, 1, 2, ... \tag{3.82}$$

Using Eq. 5.1.23 from [27],

$$E_n(0) = \frac{1}{n-1} \ (n > 1), \tag{3.83}$$

we can compute the distortion exponent for $L = 2, 3, 4, ...$ as

$$\begin{aligned}\Delta_{\mathrm{SISO-CSIT}} &= -\lim_{\mathrm{SNR}\to\infty} \frac{\log\left[\frac{e^{1/\mathrm{SNR}}}{\mathrm{SNR}} E_L\left(\frac{1}{\mathrm{SNR}}\right)\right]}{\log \mathrm{SNR}} \\ &= -\lim_{\mathrm{SNR}\to\infty} \frac{\log\left(\frac{e^{1/\mathrm{SNR}}}{\mathrm{SNR}} \cdot \frac{1}{L-1}\right)}{\log \mathrm{SNR}} \\ &= 1 \text{ for } L = 2, 3, 4, ... \end{aligned} \tag{3.84}$$

Additional bandwidth will not increase a system's distortion; therefore the distortion exponent is always a non-decreasing function of the bandwidth expansion ratio. This leads to conclusion that

$$\Delta_{\text{SISO}-\text{CSIT}} = 1 \text{ for } L \geq 1 \tag{3.85}$$

where L need not be an integer.

We now derive a lower bound on the distortion exponent for $0 < L < 1$, by first finding an upper bound on the distortion. Since the integrand of (3.79) is positive valued for $t \in (0, \infty)$,

$$\mathrm{E}[D] \leq \frac{e^{1/\text{SNR}}}{\text{SNR}} \int_0^\infty \frac{e^{-t/\text{SNR}}}{t^L} dt \tag{3.86}$$

Using Euler's Integral (Eq. 6.1.1 in [27]) we have

$$\text{SNR}^{1-L} \int_0^\infty t^{-L} e^{-t} dt = \int_0^\infty t^{-L} e^{-t/\text{SNR}} dt, \tag{3.87}$$

which leads to

$$\begin{aligned} \mathrm{E}[D] &\leq \frac{e^{1/\text{SNR}}}{\text{SNR}} \text{SNR}^{1-L} \int_0^\infty \frac{e^{-t}}{t^L} dt & (3.88) \\ &= \text{SNR}^{-L} e^{1/\text{SNR}} \int_0^\infty \frac{e^{-t}}{t^L} dt & (3.89) \\ &= \text{SNR}^{-L} e^{1/L} \Gamma(1-L). & (3.90) \end{aligned}$$

For asymptotically high SNR, the first term in (3.90) will dominate and so we have

$$\Delta_{\text{SISO}-\text{CSIT}} \geq L \text{ for } 0 < L < 1. \tag{3.91}$$

As was found in Section 3.1, $\Delta_{\text{SISO}-\text{A}} = 1$ and therefore uncoded transmission achieves the optimal distortion exponent for $L > 1$. Although in the special case of $L = 1$ the expected distortion of uncoded transmission is optimal for all SNR, this is not the case for $L \neq 1$.

3.5 System Comparison

The distortion exponents for three systems on a single channel have been found: analog transmission, rate-optimized digital, and dual-layer successive refinement. An upper bound on the distortion exponent was also derived by assuming perfect CSIT is available. These exponents are summarized in Table 3.1, and are plotted as a function of the bandwidth expansion ratio, L, in Figure 3.6. The average distortion as a function of SNR are shown in Figure 3.7 and Figure 3.8 for $L = 1$ and $L = 10$, respectively. Again, in this special case of a single channel matched to the bandwidth of the source, analog transmission remains optimal and we achieve the bound derived in Section 3.4 for $L = 1$. When we consider parallel channels exhibiting independent fading, uncoded transmission is no longer optimal, and we must find other benchmarks with which to compare system performance. Note that the distortion exponent for dual-layer successive refinement is half way between the rate-optimized digital transmission and the optimal system. As we add more layers of refinement we should approach $\Delta_{OPT} = 1$ [29], without using analog transmission. This fact may be useful for extending these results to channels of bandwidth larger than the source, or for parallel channels.

TABLE 3.1. DISTORTION EXPONENTS FOR A SINGLE CHANNEL WITH $L \geq 1$

System	Δ
Analog	1
Rate Optimized Digital	$\frac{L}{L+1}$
Successive Refinement	$\frac{L(L+1)}{L^2+L+1}$
CSIT	1

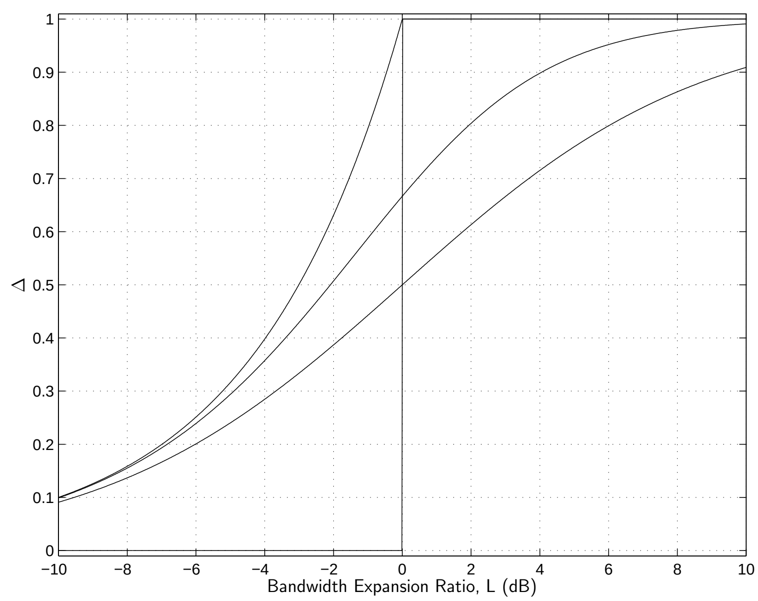

Figure 3.6. Distortion Exponents as a function of the bandwidth expansion ratio, L, in dB. At $L = 2$ dB from top to bottom the curves correspond to: the CSIT upper bound, successive refinement, optimal separate source and channel coding, and uncoded transmission.

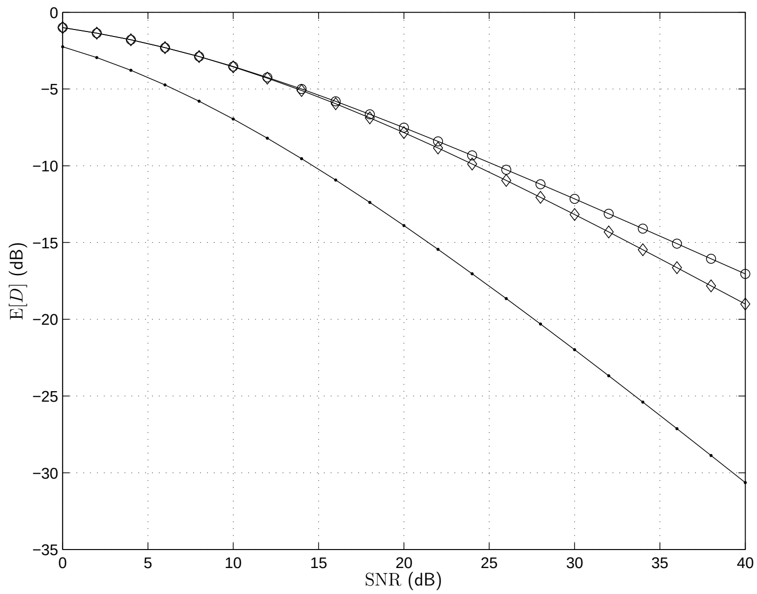

Figure 3.7. Average distortion on a quasi-static Rayleigh fading AWGN channel with bandwidth expansion ratio $L = 1$. Uncoded transmission (Section 3.1) and the lower bound (Section 3.4) are shown with $(\cdot)$, optimal separate source and channel coding (Section 3.2) with $(\circ)$, and successive refinement (Section 3.3) with $(\diamond)$.

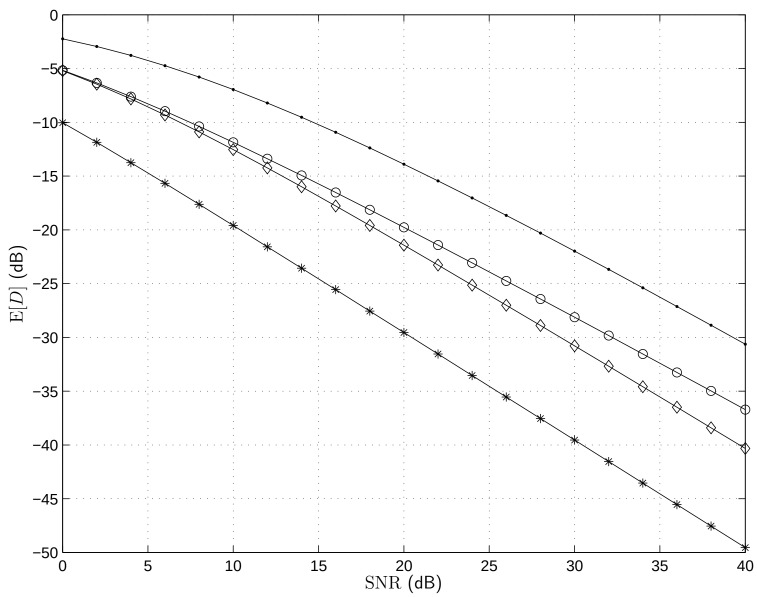

Figure 3.8. Average distortion on a quasi-static Rayleigh fading AWGN channel with bandwidth expansion ratio $L = 10$. The lower bound (Section 3.4) on distortion is shown with ($*$), uncoded transmission (Section 3.1) with ($\cdot$), optimal separate source and channel coding (Section 3.2) with ($\circ$), and successive refinement (Section 3.3) with ($\diamond$).

CHAPTER 4

COMMUNICATION ON PARALLEL CHANNELS

An attractive means to combat multipath fading is the use of multiple antennas, frequency bands, or time slots to provide the user access to several independent channels. In this chapter we will propose and analyze the performance of several methods to transmit multimedia information over parallel Rayleigh block-fading AWGN channels. The system model depicted in Figure 4.1 shows that the encoder has access to two (or more) independent channels. As in Chapter 3, a delay constraint prevents us from increasing the block length long enough to average over variations in the channel, but we assume the rate-distortion function of the source is essentially achieved.

The corresponding channel model becomes

$$y_{1,i} = a_1 \cdot x_{1,i} + w_{1,i} \tag{4.1a}$$

$$y_{2,i} = a_2 \cdot x_{2,i} + w_{2,i}, \tag{4.1b}$$

where a_1 and a_2 remain constant over the entire block. To simplify notation, the average power per channel is defined as P, as opposed to $P/2$ as in [**?**]. This

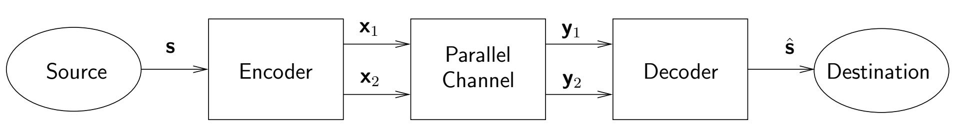

Figure 4.1. Block diagram of a communication system depicting parallel channels.

should be kept in mind to ensure a fair comparison between two-channel and one-channel systems, although it does not affect the distortion exponent. Furthermore, the encoder maps K source samples to N *pairs* of real channel inputs, or $N/2$ pairs of complex channel inputs. As for a single channel, the bandwidth expansion ratio N/K is denoted as $L := N/K$[1]. Therefore, a bandwidth expansion ratio of $L = 1$ corresponds to mapping two real source samples to a single pair of complex channel inputs (equivalently, two pairs of real channel inputs). If each source sample is mapped to a pair of complex channel inputs, the bandwidth expansion will be $L = 2$.

In this chapter we derive a lower bound on the achievable performance for communication on parallel channels, and the corresponding upper bound on the distortion exponent. We also analyze the performance of simple analog and digital repetition schemes, followed by successive refinement source coding on parallel channels. Finally, we introduce a novel hybrid digital-analog scheme and discuss its performance.

■ 4.1 A Lower Bound on the Achievable Distortion

We now determine a lower bound on the expected mean-square distortion, and the corresponding upper bound on the distortion exponent. As was done in Section 3.4, by assuming perfect CSIT is available at the transmitter, the transmitter can compute the mutual information between the inputs and outputs of the channel,

$$I(\mathbf{x};\mathbf{y}) = \frac{L}{2}\log\left(1 + |a_1|^2\mathrm{SNR}\right) + \frac{L}{2}\log\left(1 + |a_2|^2\mathrm{SNR}\right), \tag{4.2}$$

[1] In [21] the encoder maps K *real* source samples to N pairs of *complex* channel inputs. Since [21] defines the bandwidth expansion ratio as $\omega = N/K$, we can relate the two quantities as $\omega = L/2$.

and determine the the maximum rate that reliable communication can be supported. The expected distortion is then

$$\begin{aligned} \mathrm{E}[D|a_1, a_2] &= e^{-2R}\Big|_{R=\frac{L}{2}\log(1+|a_1|^2\mathrm{SNR})+\frac{L}{2}\log(1+|a_2|^2\mathrm{SNR})} \\ &= \frac{1}{(1+|a_1|^2\mathrm{SNR})^L}\cdot\frac{1}{(1+|a_2|^2\mathrm{SNR})^L}. \end{aligned} \tag{4.3}$$

Averaging over channel realizations gives

$$\mathrm{E}[D] = \int_0^\infty \frac{e^{-\lambda_1}}{(1+\lambda_1\mathrm{SNR})^L}d\lambda_1 \int_0^\infty \frac{e^{-\lambda_2}}{(1+\lambda_2\mathrm{SNR})^L}d\lambda_2 \tag{4.4}$$

$$= \left(\mathrm{E}[D_{\mathrm{SISO-CSIT}}]\right)^2. \tag{4.5}$$

It immediately follows that

$$\Delta_{\mathrm{MIMO-CSIT}} = 2\cdot\Delta_{\mathrm{SISO-CSIT}} \tag{4.6}$$

and so we have

$$\boxed{\Delta_{\mathrm{MIMO-CSIT}} = 2 \text{ for } L \geq 1} \tag{4.7}$$

and

$$\Delta_{\mathrm{MIMO-CSIT}} \geq 2L \text{ for } 0 < L < 1. \tag{4.8}$$

Although the distortion exponent derived for a single channel with perfect CSIT is achievable for $L > 1$ using uncoded transmission, this is not the case for parallel channels. A scheme that can achieve the distortion exponent given by (4.7) without CSIT likely requires some form of joint source-channel coding. We now analyze several practical schemes that do not require CSIT and compare their performance to the benchmark derived in this section.

■ 4.2 Analog Repetition

The natural extension of uncoded transmission considered in Section 3.1 to parallel channels consists of simply transmitting the source uncoded on each component channel. We will refer to this technique as Analog Repetition Coding (MIMO − A). The received signals can be modeled as

$$y_{1,i} = a_1\sqrt{\mathrm{SNR}} \cdot s_i + w_{1,i} \quad (4.9a)$$

$$y_{2,i} = a_2\sqrt{\mathrm{SNR}} \cdot s_i + w_{2,i}. \quad (4.9b)$$

Since we wish to minimize the end-to-end mean-square error, a natural choice is to perform vector Minimum Mean-Square Error (MMSE) estimation at the receiver. Because s and $\mathbf{y}$ are jointly Gaussian RVs, the MMSE estimate of s is a linear function of $\mathbf{y}$. Therefore, the MMSE estimate coincides with the linear least squares (LLS) estimate of s, i.e. $\hat{s}_{\mathrm{MMSE}}(\mathbf{y}) = \hat{s}_{\mathrm{LLS}}(\mathbf{y})$. This allows us to exploit the closed form expression for the LLS estimator [30] given by

$$\hat{s}_{\mathrm{LLS}}(\mathbf{y}) = \mu_s + \mathbf{\Lambda}_{s\mathbf{y}}\mathbf{\Lambda}_{\mathbf{y}}^{-1} \cdot (\mathbf{y} - \mu_{\mathbf{y}}) \quad (4.10)$$

$$\Lambda_{\mathrm{LLS}} = \Lambda_s - \mathbf{\Lambda}_{s\mathbf{y}}\mathbf{\Lambda}_{\mathbf{y}}^{-1}\mathbf{\Lambda}_{s\mathbf{y}}^{\mathrm{T}} \quad (4.11)$$

Using

$$\mathbf{\Lambda}_{s\mathbf{y}} = \begin{bmatrix} a_1\sqrt{\mathrm{SNR}} \\ a_2\sqrt{\mathrm{SNR}} \end{bmatrix} \quad (4.12)$$

and

$$\mathbf{\Lambda}_{\mathbf{y}} = \begin{bmatrix} |a_1|^2\mathrm{SNR} & a_1 a_2\mathrm{SNR} \\ a_1 a_2\mathrm{SNR} & |a_2|^2\mathrm{SNR} \end{bmatrix} \quad (4.13)$$

along with the fact that s and $\mathbf{y}$ are zero-mean, we have

$$\hat{s}_{\mathrm{LLS}}(\mathbf{y}) = \frac{a_1}{\sqrt{\mathrm{SNR}}(|a_1|^2 + |a_2|^2) + \sqrt{2}} \cdot y_1 + \frac{a_2}{\sqrt{\mathrm{SNR}}(|a_1|^2 + |a_2|^2) + \sqrt{2}} \cdot y_2 \quad (4.14)$$

The resulting conditional distortion is found by evaluating (4.11) to obtain

$$\mathrm{E}[D|a_1, a_2] = \frac{1}{1 + \mathrm{SNR}(|a_1|^2 + |a_2|^2)}. \quad (4.15)$$

In order to compute the expected distortion we must average (4.15) over the fading coefficients. Since a_1 and a_2 are complex Gaussian RVs, the sum of their squares will be χ^2 distributed with 4 degrees of freedom [2]. This means $\lambda = |a_1|^2 + |a_2|^2$ will be an Erlang RV, e.g.,

$$p(\lambda) = \lambda e^{-\lambda}. \tag{4.16}$$

Thus

$$\mathrm{E}[D] = \int_0^\infty \frac{\lambda e^{-\lambda}}{1 + \lambda \mathrm{SNR}} d\lambda. \tag{4.17}$$

To compare the performance of analog repetition coding to other schemes, we now compute its distortion exponent, which is found as follows:

$$\begin{aligned}
\mathrm{E}[D] &= \int_0^\infty \frac{\lambda e^{-\lambda}}{1 + \lambda \mathrm{SNR}} d\lambda && (4.18)\\
&= \frac{1}{\mathrm{SNR}} \int_0^\infty \frac{\lambda e^{-\lambda}}{\frac{1}{\mathrm{SNR}} + \lambda} d\lambda && (4.19)\\
&= \frac{1}{\mathrm{SNR}} \int_{1/\mathrm{SNR}}^\infty \frac{(t - \frac{1}{\mathrm{SNR}}) e^{-(t - \frac{1}{\mathrm{SNR}})}}{t} dt && (4.20)\\
&= \mathrm{SNR}^{-1} - \mathrm{SNR}^{-2} e^{1/\mathrm{SNR}} E_1\left(\frac{1}{\mathrm{SNR}}\right) && (4.21)\\
&\asymp \mathrm{SNR}^{-1} + \mathrm{SNR}^{-2} && (4.22)\\
&\asymp \mathrm{SNR}^{-1}, && (4.23)
\end{aligned}$$

where $E_1(\cdot)$ is the exponential integral (3.18). We note that (4.20) results from the change of variables $t = 1/\mathrm{SNR} + \lambda$, (4.22)[2] is obtained using the inequality

$$\frac{1}{2x^2} \ln(1 + 2x) < \frac{1}{x^2} e^{-1/x} E_1\left(\frac{1}{x}\right) < \frac{1}{x^2} \ln(1 + x), \tag{4.24}$$

which comes from Eq. 5.1.20 in [27], and (4.23) follows from the fact that the largest exponent dominates at high SNR. It is interesting to note that $\Delta_{\mathrm{MIMO-A}} = 1$, which is the same as in uncoded transmission on the SISO channel found in Section 3.1.

[2]The symbol $\asymp$ denotes an approximation to the rate of decay for asymptotically high SNR, i.e. ignoring constant multipliers.

■ 4.3 Digital Transmission with Selection Combining

In order to demonstrate the sub-optimality of separate source and channel coding, as well as provide a reference scheme that is easily implemented, we now analyze the performance of a system which provides diversity by sending two separate descriptions over each channel. This is the simplest form of digital transmission over parallel channels because it uses separate source and channel coding, along with a basic selection decoder. At the receiver the realized fading coefficients are known, and are used to determine if either description can be reliably decoded. The selection combiner uses the best description available, and discards the other. If neither channel realization can support reliable transmission of the information, the receiver uses the source mean as its estimate, and thus incurs a distortion equal to the source variance. We allow the rates of each description to be chosen independently, a condition slightly more general than that considered in [21] where the rates are constrained to be equal.

We define this system as one with the encoder and decoder given by

$$(\mathbf{x}_1, \mathbf{x}_2) = \mathsf{E}_{\mathbf{x}_1,\mathbf{x}_2 \leftarrow \mathbf{s}}(\mathbf{s}) = (\mathsf{E}_{\mathbf{x} \leftarrow m}(\mathsf{E}_{m \leftarrow \mathbf{s}}(\mathbf{s})), \mathsf{E}_{\mathbf{x} \leftarrow m}(\mathsf{E}_{m \leftarrow \mathbf{s}}(\mathbf{s}))) \tag{4.25a}$$

$$\hat{\mathbf{s}} = \mathsf{D}_{\hat{\mathbf{s}} \leftarrow \mathbf{y}_1,\mathbf{y}_2}(\mathbf{y}_1, \mathbf{y}_2) = \begin{cases} \mathsf{D}_{\hat{\mathbf{s}} \leftarrow \hat{m}}(\mathsf{D}_{\hat{m} \leftarrow \mathbf{y}}(\mathbf{y}_1)) & \mathsf{D}_{\hat{m} \leftarrow \mathbf{y}}(\mathbf{y}_1) \neq 0 \\ \mathsf{D}_{\hat{\mathbf{s}} \leftarrow \hat{m}}(\mathsf{D}_{\hat{m} \leftarrow \mathbf{y}}(\mathbf{y}_2)), & \mathsf{D}_{\hat{m} \leftarrow \mathbf{y}}(\mathbf{y}_1) = 0 \text{ and} \\ & \mathsf{D}_{\hat{m} \leftarrow \mathbf{y}}(\mathbf{y}_2) \neq 0 \\ \mathrm{E}[\mathbf{s}], & \text{otherwise} \end{cases} \tag{4.25b}$$

where $\mathsf{E}_{\mathbf{x}_i \leftarrow m}(\cdot)/\mathsf{D}_{\hat{m} \leftarrow \mathbf{y}_i}(\cdot)$ correspond to a single digital channel's encoder/decoder and $\mathsf{E}_{m \leftarrow \mathbf{s}}(\cdot)/\mathsf{D}_{\hat{\mathbf{s}} \leftarrow \hat{m}}(\cdot)$ correspond to the digital channel's source encoder/decoder. Note that without loss of generality, we assume that the higher rate code is sent on channel 1. This implies that when both descriptions are reliably decoded, description 1 is selected.

Since each channel carries a standard single description digital transmission, just as that in Section 3.2, the outage probability for a single channel is given by (3.31). The expected distortion for a unit-variance source is then

$$\begin{aligned} \mathrm{E}[D] &= \left[1-\exp\left(-\frac{e^{2R_1/L}-1}{\mathrm{SNR}}\right)\right]\cdot\left[1-\exp\left(-\frac{e^{2R_2/L}-1}{\mathrm{SNR}}\right)\right] \\ &\quad + e^{-2R_2}\cdot\exp\left(-\frac{e^{2R_2/L}-1}{\mathrm{SNR}}\right)\cdot\left[1-\exp\left(-\frac{e^{2R_1/L}-1}{\mathrm{SNR}}\right)\right] \\ &\quad + e^{-2R_1}\cdot\exp\left(-\frac{e^{2R_1/L}-1}{\mathrm{SNR}}\right) \end{aligned} \tag{4.26}$$

The optimal rates for (4.26) are evaluated numerically for a range of SNR and plotted in Figure 4.2. Interestingly, the optimal R_1 and R_2 are not equal. This is due to the inherent nature of a selection decoder — when both channels are successfully decoded, one of the descriptions is simply discarded, a very bandwidth inefficient operation. When $R_1 = R_2$ there are only two possible instantaneous distortions, σ_s^2 and $\sigma_s^2 \cdot e^{-2R}$. By allowing for unequal rates we increase this number to three, a result of the two channels carrying information corresponding to a different distortion. Furthermore, when it is necessary to ignore a successfully decoded transmission, i.e., when both channels are decoded, we achieve a lower distortion than if the rates are equal.

Since it is apparent from Figure 4.2 that both rates scale linearly with $\log \mathrm{SNR}$ for high SNR, we consider

$$R_1 = r_1 \log \mathrm{SNR} \tag{4.27}$$

$$R_2 = r_2 \log \mathrm{SNR}. \tag{4.28}$$

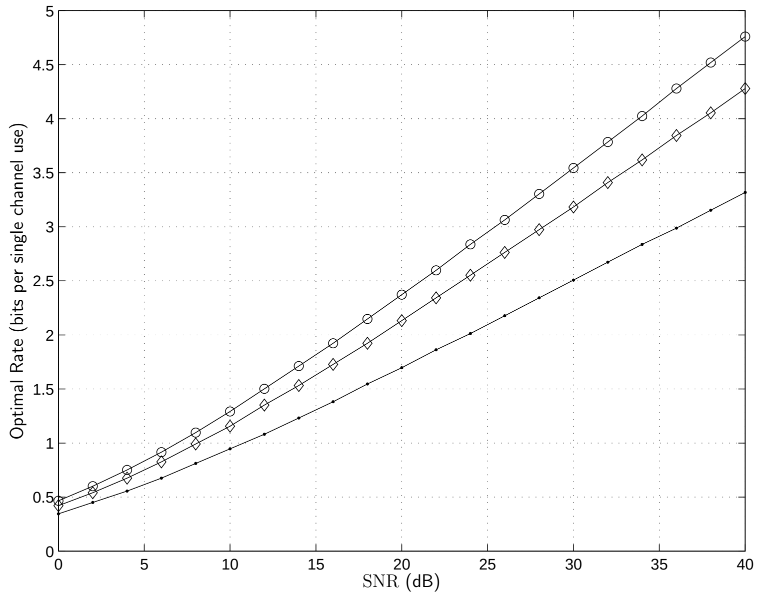

Figure 4.2. Optimal rates for rate-optimized digital transmission (Section 4.3) found numerically with $L = 1$. Equal-rate repetition coding is shown with ($\diamond$), and for multi-rate R_1 is shown with ($\circ$) and R_2 is shown with ($\cdot$).

Substituting into (4.26) we obtain the high SNR approximation for the distortion

$$\begin{aligned} \mathrm{E}[D] \quad = \quad & \left[1 - \exp\left(-\frac{\mathrm{SNR}^{2r_1/L} - 1}{\mathrm{SNR}}\right)\right] \cdot \left[1 - \exp\left(-\frac{\mathrm{SNR}^{2r_2/L} - 1}{\mathrm{SNR}}\right)\right] \\ & + \mathrm{SNR}^{-2r_2} \cdot \exp\left(-\frac{\mathrm{SNR}^{2r_2/L} - 1}{\mathrm{SNR}}\right) \cdot \left[1 - \exp\left(-\frac{\mathrm{SNR}^{2r_1/L} - 1}{\mathrm{SNR}}\right)\right] \\ & + \mathrm{SNR}^{-2r_1} \cdot \exp\left(-\frac{\mathrm{SNR}^{2r_1/L} - 1}{\mathrm{SNR}}\right). \end{aligned} \tag{4.29}$$

Using (3.37) and ignoring terms that approach unity as SNR $\to \infty$, (4.29) becomes

$$\mathrm{E}[D] = \mathrm{SNR}^{2r_1/L + 2r_2/L - 2} + \mathrm{SNR}^{2r_1/L - 2r_2 - 1} + \mathrm{SNR}^{-2r_1} \tag{4.30}$$

Taking the partial derivative of (4.30) with respect to r_1 and r_2 and equating to zero yields the optimal multiplexing gains of

$$r_1 = \frac{L(2L+1)}{2(L+1)^2} \tag{4.31a}$$

$$r_2 = \frac{L}{2(L+1)}. \tag{4.31b}$$

Using the optimal multiplexing gains of (4.31), we arrive at the distortion exponent of

$$\boxed{\Delta_{MR} = \frac{L(2L+1)}{(L+1)^2}}. \tag{4.32}$$

■ 4.3.1 Comparison to Equal Rate Repetition Coding

As previously mentioned, a simpler version of digital transmission with selection combining was considered in [21]. By choosing $R_1 = R_2$, the source coder only needs to produce a single source description, and therefore uses half as many computations as the source coder for the multi-rate scheme considered above. The optimal multiplexing gain is

$$r = \frac{L}{2+L}, \tag{4.33}$$

resulting in [20]

$$\Delta = \frac{2L}{L+2}. \tag{4.34}$$

As can be seen in Figure 4.3, this simplification becomes more costly as the bandwidth expansion ratio increases. The reason for this becomes clear when we compare the optimal multiplexing gains for each system, as shown in Figure 4.4. When bandwidth is plentiful, we operate in the outage-limited regime [21]. By allowing each channel to operate at different rates we can decrease the probability of a complete outage event significantly, and still provide a low-distortion description most of the time. Diversity is improved by ensuring the lower rate code will be successfully decoded with high probability. The increased distortion incurred is more than offset by the high-fidelity description on the other channel. Even so, at

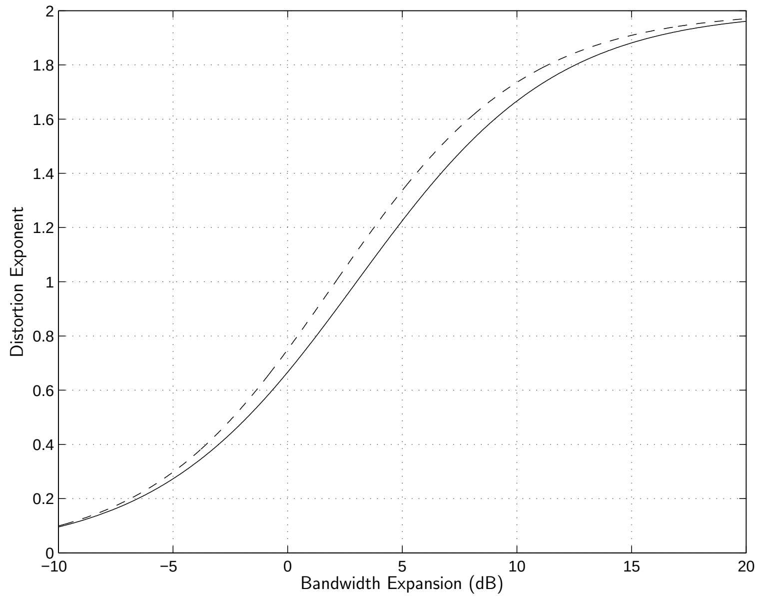

Figure 4.3. Distortion exponents as a function of bandwidth expansion factor L in decibels. The dashed curve corresponds to multi-rate coding and the solid curve corresponds to equal rate repetition coding (Section 4.3).

$L = 10$ dB the distortion exponent for the multi-rate system is only 0.069 better, which corresponds to a performance improvement of 2.5 dB at an SNR of 30 dB. Ultimately, the cost of implementing two different rate source coders simultaneously may not be worth the small improvement in system performance. For a visual comparison, the average distortion for both scenarios is shown in Figure 4.5 for $L = 1$.

■ 4.4 Naive Successive Refinement

We now present a simple extension to the successive refinement scheme considered in Section 3.3 to parallel channels. In this protocol, the encoder generates a dual-layer successive refinement code consisting of a base layer description, $\mathbf{s}_B$, at rate R_B nats per source sample and an enhancement layer description, $\mathbf{s}_E$, at rate R_E nats per source sample. As before, we assume the block length is long enough to

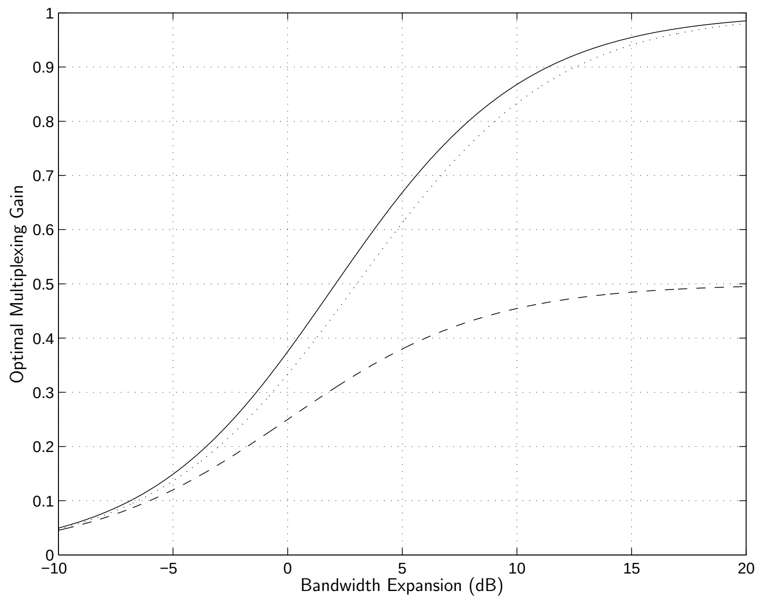

Figure 4.4. Optimal multiplexing gains as a function of bandwidth expansion factor L in decibels. The solid and dashed curves correspond to r_1 and r_2 of multi-rate coding (Section 4.3), respectively. The dotted curve corresponds to equal rate repetition coding (Section 4.3).

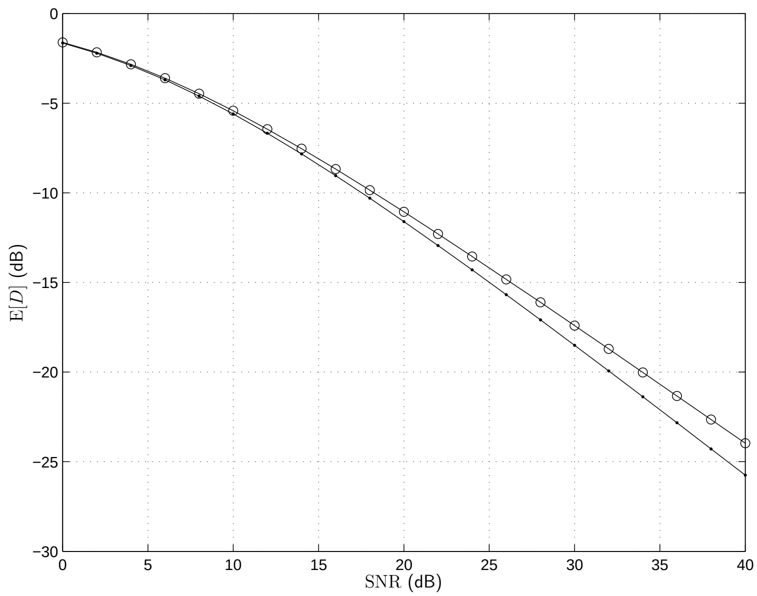

Figure 4.5. Average distortion on parallel quasi-static Rayleigh fading AWGN channels with bandwidth expansion ratio $L = 1$. Both lines are for rate-optimized digital transmission with selection combining (Section 4.3). Equal rate transmission is shown with $(\cdot)$ and non-equal rate is shown with $(\circ)$.

ensure the source coder's performance is, with high probability, equal to that of the rate-distortion function. To further generalize, we allow power allocation between each component channel using the power allocation factor $\alpha \in (0, 2)$, maintaining the convention that the *average* power per channel is P. This gives three parameters that must be chosen for a specific average SNR: R_E, R_E, and α. The base layer description is sent on component channel 1 with power $\alpha \cdot \mathrm{SNR}$, and the enhancement layer is sent on component channel 2 with power $(2-\alpha) \cdot \mathrm{SNR}$. The received signals are then given by

$$y_{1,i} = a_1 \sqrt{\alpha \mathrm{SNR}} \cdot x_{B,i} + w_{1,i} \tag{4.35a}$$

$$y_{2,i} = a_2 \sqrt{(2-\alpha)\mathrm{SNR}} \cdot x_{E,i} + w_{1,i}. \tag{4.35b}$$

The mathematical characterization of this system is defined as

$$(\mathbf{x}_1, \mathbf{x}_2) = \mathsf{E}_{\mathbf{x}_1,\mathbf{x}_2 \leftarrow \mathbf{s}}(\mathbf{s}) = (\mathsf{E}_{\mathbf{x}_1 \leftarrow m_1}(\mathsf{E}_{m_1 \leftarrow \mathbf{s}}(\mathbf{s}), \mathsf{E}_{\mathbf{x}_2 \leftarrow m_2}(\mathsf{E}_{m_2 \leftarrow \mathbf{s}}(\mathbf{s}))) \tag{4.36a}$$

$$\hat{\mathbf{s}} = \mathsf{D}_{\hat{\mathbf{s}} \leftarrow \mathbf{y}_1,\mathbf{y}_2}(\mathbf{y}_1, \mathbf{y}_2) = \begin{cases} \mathsf{D}_{\hat{\mathbf{s}}_{1,2} \leftarrow m_{1,2}}(\mathsf{D}_{\hat{m}_1 \leftarrow \mathbf{y}_1}(\mathbf{y}_1), \mathsf{D}_{\hat{m}_2 \leftarrow \mathbf{y}_2}(\mathbf{y}_2)), & \mathsf{D}_{\hat{m}_1 \leftarrow \mathbf{y}_1}(\mathbf{y}_1) \neq 0 \text{ and} \\ & \mathsf{D}_{\hat{m}_2 \leftarrow \mathbf{y}_2}(\mathbf{y}_2) \neq 0 \\ \mathsf{D}_{\hat{\mathbf{s}}_1 \leftarrow m_1}(\mathsf{D}_{\hat{m}_1 \leftarrow \mathbf{y}_1}(\mathbf{y}_1)), & \mathsf{D}_{\hat{m}_1 \leftarrow \mathbf{y}_1}(\mathbf{y}_1) \neq 0 \text{ and} \\ & \mathsf{D}_{\hat{m}_2 \leftarrow \mathbf{y}_2}(\mathbf{y}_2) = 0 \\ \mathrm{E}[\mathbf{s}], & \mathsf{D}_{\hat{m}_1 \leftarrow \mathbf{y}_1}(\mathbf{y}_1) = 0 \text{ and} \\ & \mathsf{D}_{\hat{m}_2 \leftarrow \mathbf{y}_2}(\mathbf{y}_2) = 0 \end{cases} . \tag{4.36b}$$

Following the analysis of Section 3.2 for the outage probability of a single channel, a base layer outage occurs if the mutual information between $\mathbf{x}_1$ and $\mathbf{y}_1$ falls below the transmission rate, R_B. $\Pr[\mathcal{B}_{\text{out}}]$ is therefore

$$\Pr[\mathcal{B}_{\text{out}}] = 1 - \exp\left(-\frac{e^{2R_B/L} - 1}{\alpha \cdot \text{SNR}}\right). \tag{4.37}$$

Similarly, an enhancement layer outage occurs with probability

$$\Pr[\mathcal{E}_{\text{out}}] = 1 - \exp\left[-\frac{e^{2R_E/L} - 1}{(2-\alpha) \cdot \text{SNR}}\right]. \tag{4.38}$$

At the receiver, the encoder attempts to decode both descriptions independently. As suggested by (4.36), there are three cases to consider: both $\mathbf{x}_1$ and $\mathbf{x}_2$ are decoded; $\mathbf{x}_1$ is decoded but $\mathbf{x}_2$ is not; or neither $\mathbf{x}_1$ nor $\mathbf{x}_2$ are decoded. In the best case, both descriptions are successfully decoded and the reconstruction will incur distortion $D(R_B + R_E)$, i.e.,

$$\mathrm{E}[D|\overline{\mathcal{B}_{\text{out}}}, \overline{\mathcal{E}_{\text{out}}}] = \sigma_{\mathbf{s}}^2 \cdot e^{-2(R_B+R_E)}, \tag{4.39}$$

which occurs with probability

$$\Pr[\overline{\mathcal{B}_{\text{out}}}, \overline{\mathcal{E}_{\text{out}}}] = \exp\left(-\frac{e^{2R_B/L} - 1}{\text{SNR}}\right) \cdot \exp\left(-\frac{e^{2R_E/L} - 1}{\text{SNR}}\right). \tag{4.40}$$

If the base layer is successfully decoded and the enhancement layer is not, the resultant distortion is given as

$$\mathrm{E}[D|\overline{\mathcal{B}_{\mathrm{out}}}, \mathcal{E}_{\mathrm{out}}] = \sigma_s^2 \cdot e^{-2 \cdot R_B}, \tag{4.41}$$

which occurs with probability

$$\Pr[\overline{\mathcal{B}_{\mathrm{out}}}, \mathcal{E}_{\mathrm{out}}] = \exp\left(-\frac{e^{2R_B/L}-1}{\mathrm{SNR}}\right) \cdot \left[1 - \exp\left(-\frac{e^{2R_E/L}-1}{\mathrm{SNR}}\right)\right]. \tag{4.42}$$

Finally, when there is a base-layer outage, we reconstruct to the source mean and incur distortion

$$\mathrm{E}[D|\mathcal{B}_{\mathrm{out}}] = \sigma_s^2. \tag{4.43}$$

Even if we can decode the enhancement layer, for the system described by (4.36), we assume it is useless without the base layer.

Combining (4.39)-(4.43) using the total probability law, we arrive at a formula for the average distortion (4.44) for a unit-variance source as a function of R_B, R_E, and α, which we are free to choose, and SNR, which is considered fixed. The result is

$$\begin{aligned} \mathrm{E}[D(R_B, R_E, \alpha)] &= e^{-2(R_B+R_E)} \cdot \exp\left(-\frac{e^{2R_B/L}-1}{\alpha \cdot \mathrm{SNR}}\right) \cdot \exp\left[-\frac{e^{2R_E/L}-1}{(2-\alpha)\cdot \mathrm{SNR}}\right] \\ &\quad + e^{-2R_B} \cdot \exp\left(-\frac{e^{2R_B/L}-1}{\alpha \cdot \mathrm{SNR}}\right) \cdot \left\{1 - \exp\left[-\frac{e^{2R_E/L}-1}{(2-\alpha)\cdot \mathrm{SNR}}\right]\right\} \\ &\quad + 1 - \exp\left(-\frac{e^{2R_B/L}-1}{\alpha \cdot \mathrm{SNR}}\right). \end{aligned} \tag{4.44}$$

The optimal R_B, R_E, and α can be found numerically for various SNR. Figure 4.6 shows the optimal α. The resulting distortion, along with that achieved without using power allocation ($\alpha = 1$), can be seen in Figure 4.10. Although there is some benefit to power allocation for all SNR, the gain is slight — at most 0.18 dB.

As before, for asymptotically high SNR we can use the approximation (3.37) and ignore terms that approach unity to arrive at the asymptotic expression

$$\mathrm{E}[D] \asymp \mathrm{SNR}^{-2(r_E+r_B)} + \mathrm{SNR}^{-2r_B+2r_E/L-1} + \mathrm{SNR}^{2r_B/L-1}. \tag{4.45}$$

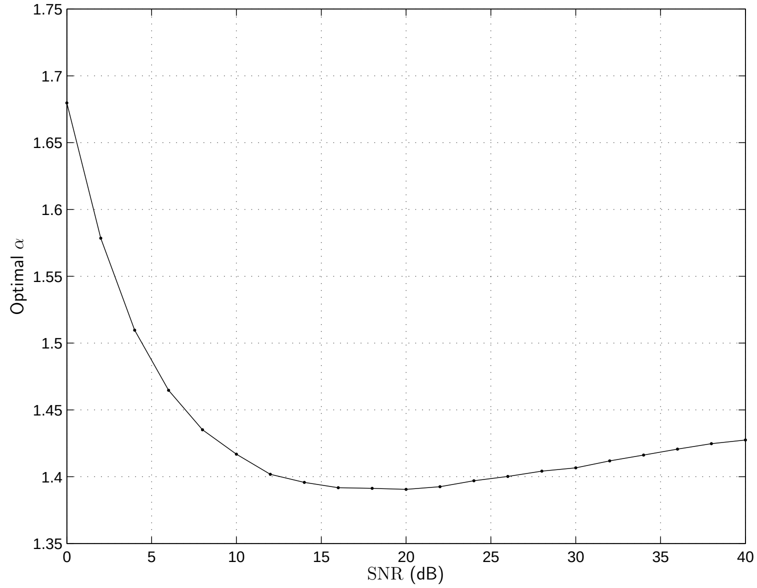

Figure 4.6. Optimal power allocation factor, α, for naive successive refinement (Section 4.4).

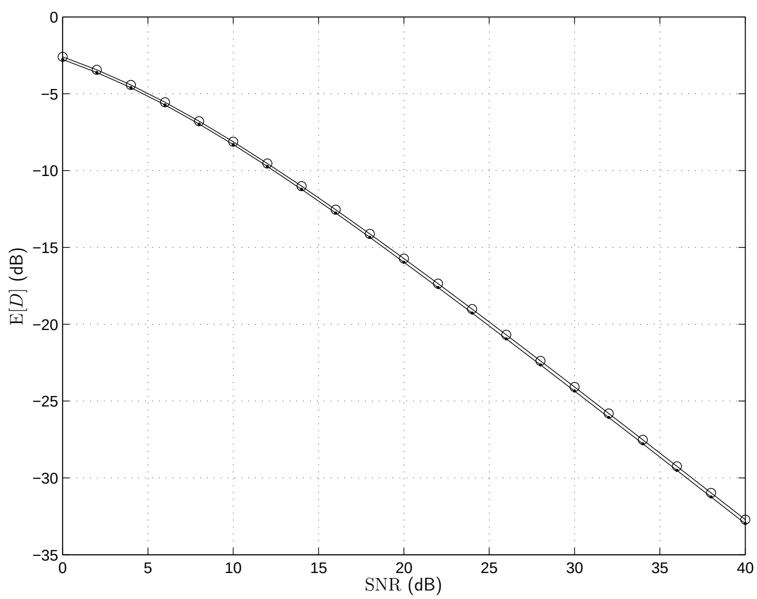

Figure 4.7. Expected distortion for naive successive refinement (Section 4.4) as a function of SNR (dB) for $L = 1$. The curve corresponding to the expected distortion with power allocation is below the curve corresponding to no power allocation.

Note that (4.45) is not a function of α. The power allocation factor does not affect the slope of the expected distortion at high SNR, and so it is ignored in the asymptotic analysis. Equating exponents gives the optimal multiplexing gains of

$$r_1 = \frac{1}{2(L+1)^2} \tag{4.46}$$

$$r_2 = \frac{1}{2(L+1)}. \tag{4.47}$$

Evaluating (4.45) at these optimal multiplexing gains gives the distortion exponent for MIMO − SR (4.48).

$$\boxed{\Delta_{\text{MIMO}-\text{SR}} = \frac{L(L+2)}{(L+1)^2}}. \tag{4.48}$$

■ 4.5 Hybrid Digital-Analog

We now propose a scheme that outperforms those considered above, and does as well in terms of its distortion exponent as optimal channel coding diversity [21]. The technique consists of transmitting both an analog and a digital version of the source, one on each component channel, and is therefore a hybrid digital-analog (HDA) scheme, similar to that proposed in [25]. On the first channel we transmit the source uncoded, using some form of analog modulation. As was discussed earlier, this is optimal in the sense that it achieves the lowest distortion possible on a single component channel for $L = 1$. The distortion on this channel will be independent of the source, and is simply a result of additive noise. We will also quantize the source and add channel coding to produce a rate-optimized digital description on the second channel. An important feature is that the distortion on channel two, a result of AWGN and quantization noise, is completely independent of the distortion on the analog channel. Note that this scheme can be thought of as a form of multiple descriptions coding in the sense that we are providing a complete, yet different, description of the source on each channel.

Since the source is transmitted uncoded on channel one, the bandwidth expansion ratio must be at least 1. For $L > 1$, the analysis is still valid, but the scheme does not utilize the additional bandwidth available on the analog channel. Even so, MIMO – HDA still performs as well as optimal channel coding diversity, which makes full use of the available bandwidth. It is likely that a more complex scheme utilizing the leftover bandwidth would outperform simple optimal channel coding in terms of its distortion exponent.

At the receiver we have perfect knowledge of the fading coefficients, and thus know if the digital channel can be reliably decoded. If this is possible, we decode the transmission and then combine the digital reconstruction with the demodulated analog version. If we are unable to decode the digital transmission, we simply use the analog description as our final source reconstruction.

We define an HDA system as one with the encoder and decoder given by

$$(\mathbf{x}_1, \mathbf{x}_2) = \mathsf{E}_{\mathbf{x}_1,\mathbf{x}_1 \leftarrow \mathbf{s}}(\mathbf{s}) = (\mathbf{s}, \mathsf{E}_{\mathbf{x}_2 \leftarrow m}(\mathsf{E}_{m \leftarrow \mathbf{s}}(\mathbf{s}))) \tag{4.49a}$$

$$\hat{\mathbf{s}} = \mathsf{D}_{\hat{\mathbf{s}} \leftarrow \mathbf{y}_1,\mathbf{y}_2}(\mathbf{y}_1, \mathbf{y}_2) = \begin{cases} \mathsf{COM}_{\hat{\mathbf{s}} \leftarrow \hat{\mathbf{s}}_1, \hat{m}}(\mathsf{EST}_{\hat{\mathbf{s}}_1 \leftarrow \mathbf{y}_1}(\mathbf{y}_1), \mathsf{D}_{\hat{\mathbf{s}}_2 \leftarrow m}(\mathsf{D}_{\hat{m} \leftarrow \mathbf{y}_2}(\mathbf{y}_2))) & \mathsf{D}_{\hat{m} \leftarrow \mathbf{y}_2}(\mathbf{y}_2) \neq 0 \\ \mathsf{EST}_{\hat{\mathbf{s}}_1 \leftarrow \mathbf{y}_1}(\mathbf{y}_1), & \text{otherwise} \end{cases}. \tag{4.49b}$$

where $\mathsf{E}_{\mathbf{x}_2 \leftarrow m}(\cdot)/\mathsf{D}_{\hat{m} \leftarrow \mathbf{y}_2}(\cdot)$ correspond to the digital channel's encoder/decoder, $\mathsf{E}_{m \leftarrow \mathbf{s}}(\cdot)/\mathsf{D}_{\hat{\mathbf{s}}_2 \leftarrow m}(\cdot)$ correspond to the digital channel's source encoder/decoder, $\mathsf{EST}_{\hat{\mathbf{s}}_1 \leftarrow \mathbf{y}_1}(\cdot)$ forms an estimate of $\mathbf{s}$ given $\mathbf{y}_1$, and $\mathsf{COM}_{\hat{\mathbf{s}} \leftarrow \hat{\mathbf{s}}_1, \hat{m}}(\cdot, \cdot)$ combines two source reconstructions using a weighted average.

For low to moderate SNR it may be beneficial to allocate power between the digital and analog transmissions unequally. This case is considered by using the power allocation factor $\alpha \in [0, 2]$ such that the power for the analog channel is $\alpha \cdot \text{SNR}$, and $(2 - \alpha) \cdot \text{SNR}$ for the digital transmission. The received signals are given by

$$y_{1,i} = a_1 \cdot \alpha\sqrt{\mathrm{SNR}} \cdot s_i + w_{1,i} \tag{4.50a}$$

$$y_{2,i} = a_2 \cdot (2-\alpha)\sqrt{\mathrm{SNR}} \cdot x_{2,i} + w_{2,i}. \tag{4.50b}$$

Power allocation does not offer improved performance in terms of the rate at which distortion decays for asymptotically high SNR, thus the power allocation factor is ignored in distortion exponent analysis.

The receiver first computes the minimum mean-square estimate of $\mathbf{s}$ given $\mathbf{y}_1$. Each source reconstruction, $\hat{s}_i$, $i = 1, \ldots, N$, is found using

$$\hat{s}_{1,i}(y_{1,i}) = \frac{a_1\sqrt{\alpha \cdot \mathrm{SNR}}}{|a_1|^2 \cdot \alpha \cdot \mathrm{SNR} + 1} \cdot y_{1,i}. \tag{4.51}$$

The receiver must then decide to use this as the final source reconstruction, or if the digital transmission can be reliably decoded and a better estimate of the source can be found. This is done by computing the channel's mutual information, which is a function of the fading coefficient, a_2. If

$$|a_2| \leq \sqrt{\frac{e^{2R/L} - 1}{(2-\alpha)\mathrm{SNR}}} \tag{4.52}$$

the digital transmission cannot be reliably decoded and we must rely completely on the analog reconstruction, i.e.,

$$\hat{\mathbf{s}} = \hat{\mathbf{s}}_1. \tag{4.53}$$

This will occur with probability

$$\mathrm{P_{out}}(R, \mathrm{SNR}) = \Pr\left[I(\mathbf{x}_2; \mathbf{y}_2) < R\right] \tag{4.54}$$

$$= 1 - \exp\left[-\frac{e^{2R/L} - 1}{(2-\alpha)\mathrm{SNR}}\right] \tag{4.55}$$

When an outage event on channel two occurs the resulting distortion will be

$$\begin{aligned} \mathrm{E}[D|\text{outage}] &= \int_0^\infty \frac{e^{-\lambda}}{1+\lambda\cdot\alpha\cdot\mathrm{SNR}}d\lambda & (4.56) \\ &= \int_{1/\alpha\cdot\mathrm{SNR}}^\infty \frac{1}{t\cdot\mathrm{SNR}}\exp\left(\frac{1}{\alpha\cdot\mathrm{SNR}}-t\right)dt & (4.57) \\ &= \frac{1}{\alpha\cdot\mathrm{SNR}}\cdot\exp\left(\frac{1}{\alpha\cdot\mathrm{SNR}}\right)\cdot E_1\left(\frac{1}{\alpha\cdot\mathrm{SNR}}\right). & (4.58) \end{aligned}$$

Note that (4.56) is found directly from the formula for expected distortion of a linear least-squares estimator, and (4.57) follows from (4.56) using the substitution

$$t = \frac{1+\lambda\cdot\alpha\cdot\mathrm{SNR}}{\alpha\cdot\mathrm{SNR}}. \tag{4.59}$$

If, however,

$$|a_2| > \sqrt{\frac{e^{2R/L}-1}{(2-\alpha)\mathrm{SNR}}} \tag{4.60}$$

we decode $\mathbf{x}_2$ to form the digital source reconstruction $\hat{\mathbf{s}}_2$. We now wish to combine our digital and analog source reconstructions such that the resultant distortion is minimized. Ideally, this would require computing the minimum mean-square error estimate of $\mathbf{s}$ from $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$. Since the distortion on the digital channel is a highly non-linear function of the source, this estimation problem is difficult to solve. To simplify the combiner, we constrain our estimate of $\mathbf{s}$ to be a linear function of $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$. To compute this we first rewrite each component of $\hat{s}_2$ in the form of

$$\hat{s}_2 = \beta\cdot s + z_2 \tag{4.61}$$

using the test channel shown in Figure 4.8. We constrain s and z_2 to be uncorrelated, which implies

$$\beta = \frac{\sigma_s^2 - D}{\sigma_s^2} \tag{4.62}$$

and

$$\sigma_z^2 = \frac{D}{\sigma_s^2}(\sigma_s^2 - D) \tag{4.63}$$

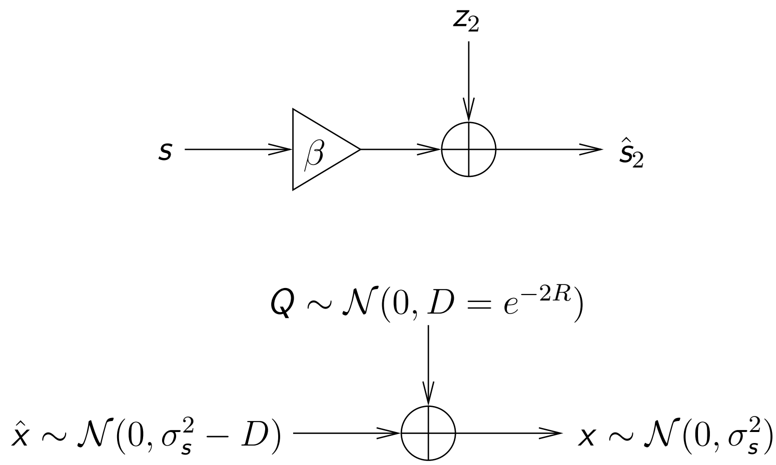

Figure 4.8. Test channel for computation of the expected distortion for hybrid digital-analog transmission (Section 4.5).

The effective SNR of the digital channel is then

$$\gamma_2 = \frac{\beta^2}{\sigma_z^2} = \frac{1}{D} - \frac{1}{\sigma_s^2} = \frac{1}{\sigma_s^2}(e^{2R} - 1) \tag{4.64}$$

Equivalently,

$$\tilde{y}_2 = \hat{s}_2 = \sqrt{\gamma_2} \cdot s + w_2. \tag{4.65}$$

The analog reconstruction can be expressed in a similar form:

$$\tilde{y}_1 = \hat{s}_1 = \sqrt{\gamma_1} \cdot s + w_1, \tag{4.66}$$

where s, w_1, and w_2 are zero-mean, unit-variance, and uncorrelated. The effective SNR of the analog channel is simply

$$\gamma_1 = |a_1|^2 \cdot \alpha \cdot \mathrm{SNR} \tag{4.67}$$

Computing

$$\mathbf{\Lambda}_{s\tilde{\mathbf{y}}} = \begin{bmatrix} \sqrt{\gamma_1} \\ \sqrt{\gamma_2} \end{bmatrix} \tag{4.68}$$

and

$$\mathbf{\Lambda}_{\tilde{\mathbf{y}}} = \begin{bmatrix} \gamma_1 + 1 & \sqrt{\gamma_1\gamma_2} \\ \sqrt{\gamma_1\gamma_2} & \gamma_2 + 1 \end{bmatrix}, \tag{4.69}$$

we arrive at the formula for the LLS combiner

$$\begin{aligned} \hat{s}_{\mathrm{LLS}}(\tilde{\mathbf{y}}) &= \frac{1}{1+\gamma_1+\gamma_2}(\sqrt{\gamma_1}\cdot\tilde{y}_1 + \sqrt{\gamma_2}\cdot\tilde{y}_2) && (4.70) \\ &= \frac{1}{|a_1|^2\alpha\cdot\mathrm{SNR}+e^{2R}}(a_1\sqrt{\alpha\cdot\mathrm{SNR}}\cdot\tilde{y}_1 + \sqrt{e^{2R}-1}\cdot\tilde{y}_2). && (4.71) \end{aligned}$$

When there is not an outage, using (4.71) as the combiner results in a distortion of

$$\begin{aligned} \mathrm{E}[D|\overline{\mathrm{outage}}, a_1] &= \frac{1}{1+\gamma_1+\gamma_2} && (4.72) \\ &= \frac{1}{|a_1|^2\cdot\alpha\cdot\mathrm{SNR}+e^{2R}}. && (4.73) \end{aligned}$$

Averaging over a_1 yields

$$\begin{aligned} \mathrm{E}[D|\overline{\mathrm{outage}}] &= \int_0^\infty \frac{e^{-\lambda}}{\lambda\cdot\alpha\cdot\mathrm{SNR}+e^{2R}}d\lambda && (4.74) \\ &= e^{-2R}\int_0^\infty \frac{e^{-\lambda}}{1+\lambda\cdot\alpha\cdot\mathrm{SNR}e^{-2R}}d\lambda && (4.75) \\ &= \frac{1}{\alpha\cdot\mathrm{SNR}}\exp\left(\frac{e^{2R}}{\alpha\cdot\mathrm{SNR}}\right)E_1\left(\frac{e^{2R}}{\alpha\cdot\mathrm{SNR}}\right). && (4.76) \end{aligned}$$

We have now found expressions for the distortion under outage (4.58) and no outage (4.76), along with the probability that these events occur. Using the total probability law, the average distortion as a function of the power allocation factor, α, the rate, R, and the average SNR, is given by

$$\begin{aligned} \mathrm{E}[D(R,\alpha)] = &\left\{1-\exp\left[-\frac{e^{2R/L}-1}{(2-\alpha)\mathrm{SNR}}\right]\right\}\frac{1}{\alpha\mathrm{SNR}}\exp\left(\frac{1}{\alpha\cdot\mathrm{SNR}}\right)E_1\left(\frac{1}{\alpha\cdot\mathrm{SNR}}\right) \\ &+\exp\left[-\frac{e^{2R/L}-1}{(2-\alpha)\mathrm{SNR}}\right]\frac{1}{\alpha\cdot\mathrm{SNR}}\exp\left(\frac{e^{2R}}{\alpha\cdot\mathrm{SNR}}\right)E_1\left(\frac{e^{2R}}{\alpha\cdot\mathrm{SNR}}\right) \end{aligned}$$

It is assumed that the transmitter knows the channel's average SNR, and must optimize the transmission over α and R. This optimization can be performed numerically, and the optimal rate is shown in Figure 4.9. shows the distortion achieved

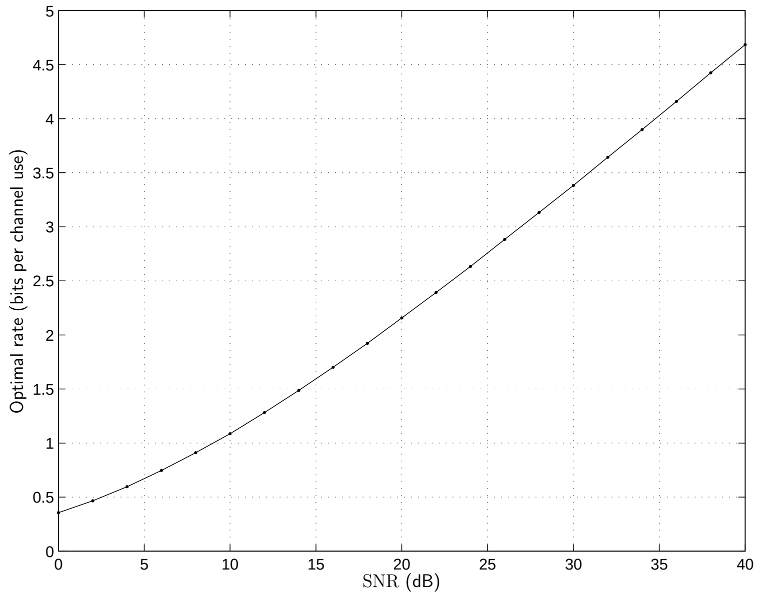

Figure 4.9. Optimal rate in bits per channel use for hybrid digital-analog transmission (Section 4.5) with $L = 1$.

using the optimal rate and $\alpha = 1$, along with the performance of other schemes for comparison.

In order to simplify the comparison of $\text{MIMO} - \text{HDA}$ transmission to other schemes, we are interested in characterizing it's performance in the high SNR (low distortion) regime. As previously mentioned, we can ignore α in deriving the distortion exponent, greatly simplifying the analysis. As in other rate optimized systems studied, for asymptotically high SNR, the optimal rate grows linear with $\log \text{SNR}$, as can be seen in Figure 4.9. Using $R = r \log \text{SNR}$ in (4.77) we arrive at

$$\text{E}[D] = \frac{\text{SNR}^r - 1}{\text{SNR}} \cdot \frac{1}{\text{SNR}} e^{1/\text{SNR}} E_1 \left(\frac{1}{\text{SNR}} \right) + \frac{1}{\text{SNR}} \exp \left(\text{SNR}^{2r-1} \right) E_1 \left(\text{SNR}^{2r-1} \right). \tag{4.77}$$

The first term in (4.77) behaves as $\text{SNR}^{2r/L-2}$ for high SNR, and the second term behaves as SNR^{-2r}, which follows from (4.77) using (3.19). Equating the exponents

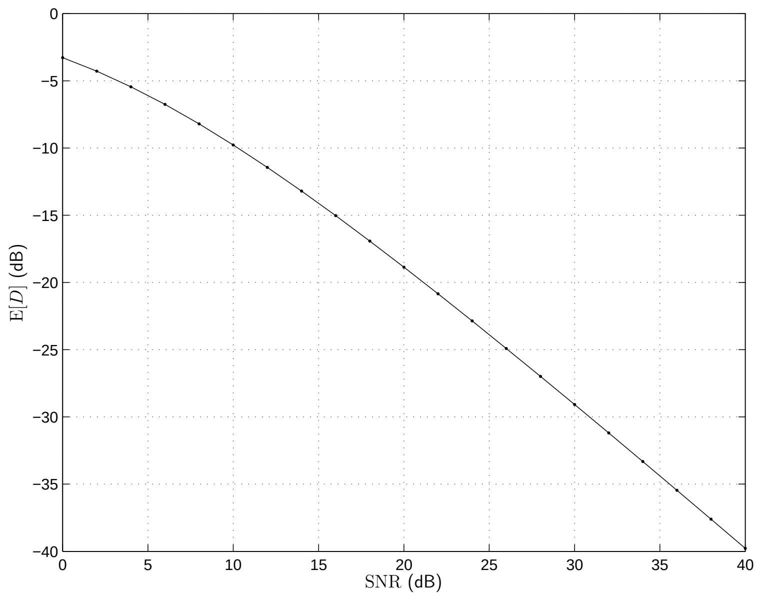

Figure 4.10. Expected distortion for hybrid digital-analog transmission (Section 4.5) with $L = 1$ as a function of SNR (dB).

and solving for r yields the optimal multiplexing gain of

$$r = \frac{L}{L+1}, \tag{4.78}$$

and an overall distortion exponent of

$$\boxed{\Delta_{\text{MIMO}-\text{HDA}} = \frac{2L}{L+1}}. \tag{4.79}$$

In the limit of high SNR the digital distortion can be modeled as additive Gaussian noise independent of the source with variance $D_{\mathsf{s}}(R)$ [5]. In this case, the MMSE estimate of s is a linear function of $\tilde{y}_1$ and $\tilde{y}_2$. This means for asymptotically high SNR, the LLS estimate will converge to the MMSE estimate. Since the distortion exponent is a property of how a system behaves at asymptotically high SNR, (4.79) is therefore also valid for MIMO − HDA transmission utilizing the significantly more complicated MMSE combiner.

In addition to good performance at high SNR the HDA scheme offers the advantage that there is never a complete outage event. It is always possible to provide an estimate of the source that is better than the naive choice of reconstructing to the source mean, resulting in the distortion having a more desirable PDF than one in which a complete outage event occurs with non-zero probability. Although this distinction is not captured by comparing the first moment of the distortion, for practical systems, a higher order moment performance criterion may be more relevant.

■ 4.6 System Comparison

In this chapter we derived a lower bound on the achievable distortion, and an upper bound on the distortion exponent. The average distortion for analog and digital repetition schemes was also found. We considered a rate-optimized digital scheme with selection combining, and showed that better performance can be achieved if the rates are not constrained to be equal. Finally, we presented a hybrid digital-analog scheme and characterized its performance for asymptotically high SNR. The distortion exponent for all of these schemes were found analytically and are summarized in Table 4.1 for $L > 1$, and plotted in Figure 4.11.

TABLE 4.1. DISTORTION EXPONENTS FOR PARALLEL CHANNELS WITH $L \geq 1$

System	Δ
Analog Repetition	1
Multi-Rate	$\frac{L(2L+1)}{(L+1)^2}$
Naive Successive Refinement	$\frac{L(L+2)}{(L+1)^2}$
Hybrid Digital-Analog	$\frac{2L}{L+1}$
CSIT	2

The average distortion performance for each scheme with $L = 1$ and $L = 10$ are shown in Figure 4.12 and Figure 4.13, respectively. This helps to facilitate a visual

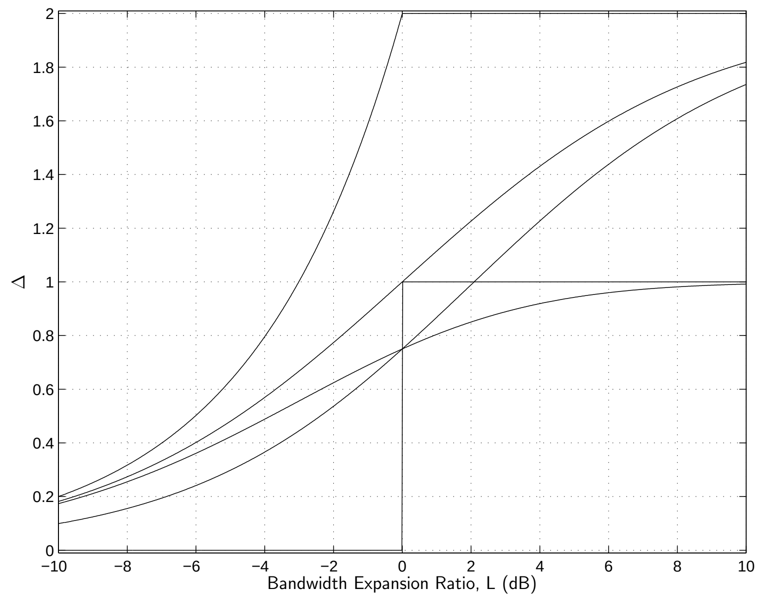

Figure 4.11. Distortion Exponents as a function of the bandwidth expansion ratio, L, in dB. At $L = 4$ dB from top to bottom the curves correspond to: the CSIT upper bound, hybrid digital-analog transmission, multi-rate digital transmission, analog repetition, and naive successive refinement.

comparison for low to moderate SNR, and shows that the distortion exponent is a relevant metric for comparing different schemes in some regimes, but in others it is an not an adequate characterization.

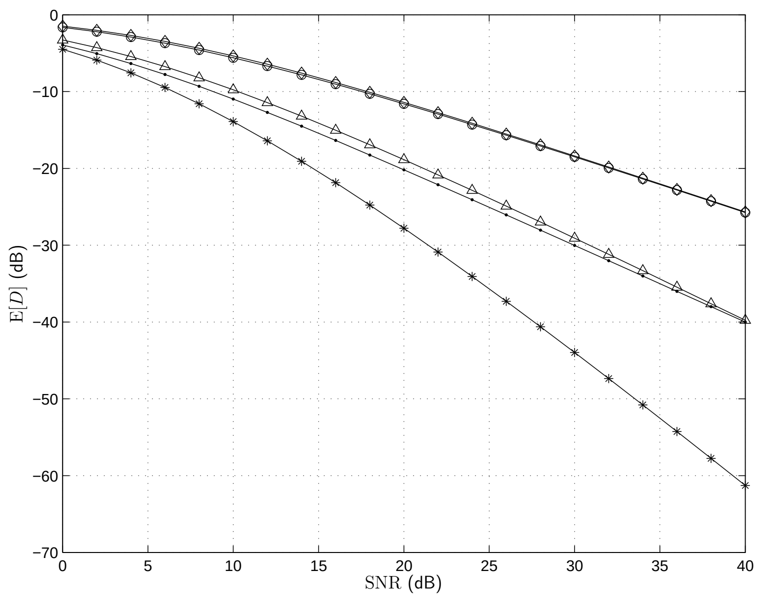

Figure 4.12. Average distortion on a quasi-static Rayleigh fading AWGN channel with bandwidth expansion ratio $L = 1$. The lower bound (Section 3.4) on distortion is shown with ($*$), analog repetition (Section 4.2) with ($\cdot$), rate-optimized digital (Section 4.3) with ($\circ$), naive successive refinement with ($\diamond$), and hybrid digital-analog (Section 4.5) with ($\triangle$).

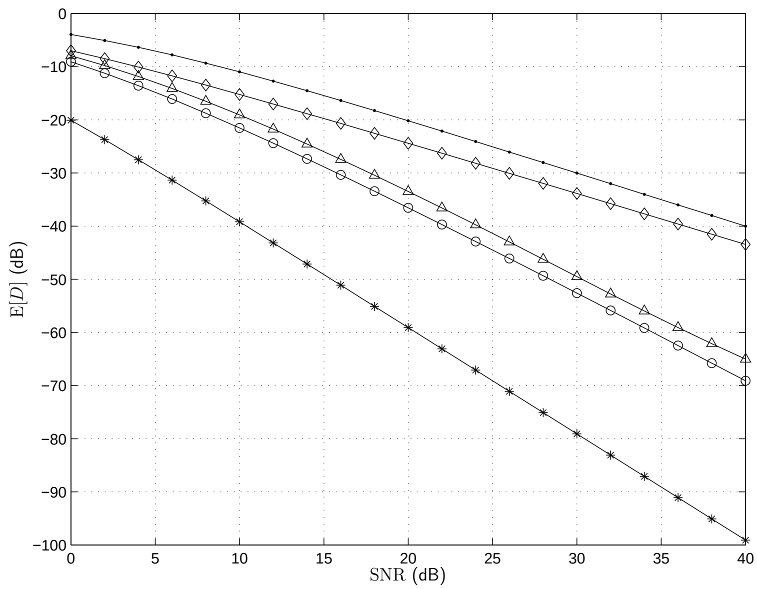

Figure 4.13. Average distortion on a quasi-static Rayleigh fading AWGN channel with bandwidth expansion ratio $L = 10$. The lower bound (Section 3.4) on distortion is shown with ($*$), analog repetition (Section 4.2) with ($\cdot$), rate-optimized digital (Section 4.3) with ($\circ$), naive successive refinement with ($\diamond$), and hybrid digital-analog (Section 4.5) with ($\triangle$).

CHAPTER 5

CONCLUSIONS

■ 5.1 Insights

This thesis serves to illustrate a fundamental theme in delay constrained multimedia communications: the need for graceful degradation in received signal quality. The step-like performance of standard digital communication results in average distortion performance significantly below other schemes, especially relative to the CSIT lower bound. The lower distortion achieved by successive refinement for a single channel (Section 3.3) suggests that allowing more distortion levels translates to a decrease in end-to-end distortion.

Power allocation was also considered as a means to decrease distortion for schemes that operate with multiple layers or coding rates simultaneously. It was shown that power allocation does not offer improved performance in terms of a systems distortion exponent. Although power allocation did improve performance for most schemes, this improvement was often negligible and may not outweigh the added system complexity and cost. The one exception to this was the superposition successive refinement coding scheme considered in (Section 3.3). Since the enhancement layer was treated as additive noise, it was essential that as SNR increased, more power was allocated to the base layer.

For parallel channels, in terms of its distortion exponent, the hybrid digital-analog scheme outperformed analog and and digital repetition, naive successive refinement, and multi-rate digital, but its performance still fell well below the lower bound derived in Section 4.1. The naive successive refinement scheme did not perform well because the dominate event of a base layer outage occurs with the same probability as an outage for rate-optimized digital transmission over a single channel.

5.2 Future Research

Although this thesis showed that the CSIT bound on the expected distortion is achievable for a single channel in terms of the rate of decay of the mean-square distortion for asymptotically high SNR, it does not present an achievable bound on the distortion itself. For $L > 1$, uncoded transmission does not do as well as the lower bound on distortion, except in terms of the distortion exponent. For parallel channels, it is still unknown if the lower bound on distortion is achievable, and if so, how best to achieve it. Furthermore, we only considered a mean-square distortion measure, which may not be a realistic performance criterion for many practical sources. It would be interesting to study the correlation between mean-square distortion and perceptual distortion models when considering the types of channels studied in this thesis. It may also be relevant to consider higher order, or more general, distortion measures.

There are also several variations on the schemes presented in this thesis that may improve performance. Two interesting possibilities apply to successive refinement and hybrid digital-analog transmission.

Over parallel channels, a superposition successive refinement code would likely benefit the same way it did for a single channel. One possible implementation would be to split the base and enhancement layers equally between the two channels, taking advantage of the sum of each channels' realized mutual information. For both a single and parallel channels, more layers could be used to decrease distortion.

It was mentioned in Section 4.5 that for MIMO − HDA transmission, combining the digital and analog source reconstructions ideally consists of finding $\hat{\mathbf{s}}_{\mathrm{MMSE}}(\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2)$, where $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$ are the analog and digital source reconstructions, respectively. This architecture implies that the analog and digital estimates are formed separately, and then passed on to the combiner - along with some reliability information for the analog channel, the fading coefficient a_1. Theoretically, we could further reduce the average distortion by performing joint decoding of the received vector by forming an estimate of $\mathbf{s}$ from the channel outputs $\mathbf{x}_1$ and $\mathbf{x}_2$. Improved performance could also be achieved for $L > 1$ by utilizing the additional bandwidth available on the analog channel for digital communication.

Some other interesting ideas include determining the distortion exponents of practical schemes, using a simulation based aproach. Also, a low SNR distortion metric could serve to provide a more complete characterization of the average distortion. Ultimately, there is still a wide variety of questions relating to communicating over parallel channels, and even more regarding the more general MIMO channel.

REFERENCES

[1] C. E. Shannon, “A Mathematical Theory of Communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.

[2] J. G. Proakis, *Digital Communications*, Third ed. New York: McGraw-Hill, Inc., 1995.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: John Wiley & Sons, Inc., 1991.

[4] S. Lin and D. J. Costello, *Error Control Coding*, 2nd ed. Prentice Hall, Apr. 2004.

[5] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression.* Boston, MA: Kluwer Academic Publishers, 1992.

[6] V. K. Goyal, “Multiple Description Coding: Compression Meets the Network,” vol. 18, no. 5, pp. 74–93, Sept. 2001.

[7] H. Witsenhausen, “On Source Networks with Minimal Breakdown Degradation,” *Bell Syst. Tech. J.*, vol. 59, no. 6, pp. 1083–1087, July-Aug. 1980.

[8] J. Wolf, A. Wyner, and J. Ziv, “Source Coding for Multiple Descriptions,” *Bell Syst. Tech. J.*, vol. 59, no. 8, pp. 1417–1426, Oct. 1980.

[9] L. Ozarow, “On a Source Coding Problem with Two Channels and Three Receivers,” *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, Dec. 1980.

[10] H. Witsenhausen and A. D. Wyner, “Source Coding for Multiple Descriptions II: A Binary Source,” Dec. 1980, Bell Labs Tech. Rept. TM-80-1217.

[11] A. A. El Gamal and T. M. Cover, “Achievable Rates for Multiple Descriptions,” *IEEE Trans. Inform. Theory*, vol. 28, no. 6, pp. 851–857, Nov. 1982.

[12] T. Berger and Z. Zhang, “Minimum Breakdown Degradation in Binary Source Encoding,” *IEEE Trans. Inform. Theory*, vol. 29, no. 6, pp. 807–814, Nov. 1983.

[13] Z. Zhang and T. Berger, “New Results in Binary Multiple Descriptions,” *IEEE Trans. Inform. Theory*, vol. 33, no. 4, pp. 502–521, July 1987.

[14] ——, “Multiple Description Source Coding with No Excess Marginal Rate,” *IEEE Trans. Inform. Theory*, vol. 41, no. 2, pp. 349–357, Mar. 1995.

[15] R. F. Ahlswede, "The Rate-Distortion Region for Multiple Descriptions without Excess Rate," *IEEE Trans. Inform. Theory*, vol. 31, no. 6, pp. 721–726, Nov. 1985.

[16] V. A. Vaishampayan, "Design of Multiple Description Scalar Quantizers," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 821–834, May 1993.

[17] V. K. Goyal and J. Kovacevic, "Generalized Multiple Description Coding with Correlating Transforms," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2199–2224, Sept. 2001.

[18] I. E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," Bell Labs, Lucent Technologies, Tech. Rep., 1995. [Online]. Available: http://mars.bell-labs.com/papers/proof/proof.pdf

[19] L. Zheng and D. N. C. Tse, "Optimal Diversity-Multiplexing Tradeoff in Multiple Antenna Channels," in *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Oct. 2001.

[20] J. N. Laneman, E. Martinian, G. W. Wornell, J. G. Apostolopoulos, and S. J. Wee, "Comparing Application- and Physical-Layer Approaches to Diversity on Wireless Channels," in *Proc. IEEE Int. Conf. Communications (ICC)*, May 2003. [Online]. Available: http://www.nd.edu/ jnl/pubs/icc2003.pdf

[21] J. N. Laneman, E. Martinian, G. W. Wornell, and J. G. Apostolopoulos, "Source-Channel Diversity Approaches for Multimedia Communication," *IEEE Trans. Inform. Theory*, Nov. 2004, revised and accepted for publication. [Online]. Available: http://www.nd.edu/ jnl/pubs/sc-div.pdf

[22] D. Gunduz and E. Erkip, "Joint Source-Channel Cooperation: Diversity versus Spectral Efficiency," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Chicago, IL, July 2004.

[23] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. 37, no. 2, pp. 269–275, March 1991.

[24] B. Rimoldi, "Successive Refinement of Information: Characterization of the Achievable Rates," *IEEE Trans. Inform. Theory*, vol. 40, no. 1, pp. 253–259, Jan. 1994.

[25] S. Shamai (Shitz), S. Verdú, and R. Zamir, "Systematic Lossy Source/Channel Coding," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 564–579, Mar. 1998.

[26] U. Mittal and N. Phamdo, "Hybrid Digital-Analog (HDA) Joint Source-Channel Codes for Broadcasting and Robust Communications," *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1082–1102, May 2002.

[27] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions.* New York: Dover, June 1974.

[28] L. Zheng and D. N. C. Tse, "Diversity and Multiplexing: A Fundamental Tradeoff in Multiple-Antenna Channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[29] D. Gunduz and E. Erkip, "Source and Channel Coding for Quasi-Static Fading Channels," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 2005.

[30] H. L. V. Trees, *Detection, Estimation, and Modulation Theory.* New York: Wiley, 1968.