KTH Electrical Engineering

Limited Feedback Information in Wireless Communications: Transmission Schemes and Performance Bounds

TÙNG T. KIM

Doctoral Thesis in Telecommunications
Stockholm, Sweden 2008

TRITA-EE 2008:022
ISSN 1653-5146
ISBN 978-91-7178-950-1

KTH School of Electrical Engineering
SE-100 44 Stockholm
SWEDEN

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen i telekommunikation fredagen den 30 maj 2008 klockan 14.00 i Hörsal D3, Lindstedtsvägen 5, Stockholm.

Tryck: Universitetsservice US AB

Abstract

This thesis studies some fundamental aspects of wireless systems with partial channel state information at the transmitter (CSIT), with a special emphasis on the high signal-to-noise ratio (SNR) regime. The first contribution is a study on multi-layer variable-rate communication systems with quantized feedback, where the expected rate is chosen as the performance measure. Iterative algorithms exploiting results in the literature of parallel broadcast channels are developed to design the system parameters. Necessary and sufficient conditions for single-layer coding to be optimal are derived. In contrast to the ergodic case, it is shown that a few bits of feedback information can improve the expected rate dramatically.

The next part of the thesis is devoted to characterizing the tradeoff between diversity and multiplexing gains (D–M tradeoff) over slow fading channels with partial CSIT. In the multiple-input multiple-output (MIMO) case, we introduce the concept of minimum guaranteed multiplexing gain in the forward link and show that it influences the D–M tradeoff significantly. It is demonstrated that power control based on the feedback is instrumental in achieving the D-M tradeoff, and that rate adaptation is important in obtaining a high diversity gain even at high rates.

Extending the D–M tradeoff analysis to decode-and-forward relay channels with quantized channel state feedback, we consider several different scenarios. In the relay-to-source feedback case, it is found that using just one bit of feedback to control the source transmit power is sufficient to achieve the multiantenna upper bound in a range of multiplexing gains. In the destination-to-source-and-relay feedback scenario, if the source-relay channel gain is unknown to the feedback quantizer at the destination, the diversity gain only grows linearly in the number of feedback levels, in sharp contrast to an exponential growth for MIMO channels.

We also consider the achievable D–M tradeoff of a relay network with the compress-and-forward protocol when the relay is constrained to make use of standard source coding. Under a short-term power constraint at the relay, using source coding without side information results in a significant loss in terms of the D–M tradeoff. For a range of multiplexing gains, this loss can be fully compensated for by using power control at the relay.

The final part of the thesis deals with the transmission of an analog Gaussian source over quasi-static fading channels with limited CSIT, taking the SNR exponent of the end-to-end average distortion as performance measure. Building upon results from the D–M tradeoff analysis, we develop novel upperbounds on the distortion exponents achieved with partial CSIT. We show that in order to achieve the optimal scaling, the CSIT feedback resolution must grow logarithmically with the bandwidth ratio for MIMO channels. The achievable distortion exponent of some hybrid schemes with heavily quantized feedback is also derived. As for the half-duplex fading relay channel, combining a simple feedback scheme with separate source and channel coding outperforms the best known no-feedback strategies even with only a few bits of feedback information.

Acknowledgement

I would like to express my gratitude to Professor Mikael Skoglund, who has taken great effort to walk me into the world of science, and beyond (as evident from our long drive to Death Valley). I am grateful that Mikael has given me an incredible opportunity to pursue my research, and that he has provided invaluable guidance and limitless support during my wonderful time at KTH.

I would also like to thank Professor Giuseppe Caire for hosting me with his group at University of Southern California. Despite his incredibly busy schedule, Giuseppe has managed to spend a large amount of time helping me attack research problems of common interest.

It has been a great pleasure to work alongside my colleagues in the Communication Theory and Signal Processing groups. Especially, I would like to thank Dr. George Jöngren for introducing me to the field of limited feedback, Dr. Joakim Jaldén for posing thoughtful questions that inspired deeper studies on some problems in this thesis, and Lei Bao, for going out of her way to help me on countless occasions. Thanks to the excellent administrative personnel, among them are Annika Augustsson and Karin Demin, and to the computer support team, the working environment at KTH has been running so smoothly that I could totally focus on my work. I also thank my fellow students at USC, especially Ozgun Bursalioglu and Raj Kumar, for making my stay in Los Angeles a pleasant one.

I have greatly enjoyed working with my co-authors, who have given me the luxury of leaving out a couple of journal papers from the thesis. In particular, Professor Mats Bengtsson and Professor Erik G. Larsson's contribution has been extremely helpful, for which I am grateful.

Thanks are due to Professor Hesham El Gamal from Ohio State University for acting as opponent for the thesis. I also thank the committee members: Professor Eduard Jorswieck from TU Dresden, Professor Ulf Jönsson from KTH, and Professor Ralf Müller from NTNU.

Last but definitely not least, I feel deeply indebted to my family for their absolute support and patience during my seemingly endless academic training. I especially thank my wife Hải Yến for quietly nourishing love and keeping happiness under our small roof.

Contents

Chapter 1

Introduction

The telecommunications industry has been continuing to develop new advanced technologies to support emerging mobile applications and reduce the performance gaps between wired and wireless systems, making wireless communications one of the most rapidly developing research areas over the last couples of decades. Recent advances in the field not only opened new opportunities to approach these ambitious goals but also gave rise to many new challenging problems in both practice and theory.

Using feedback information is common in wireless communications. Traditionally this has been widely used for example in the form of power control in spread-spectrum communication systems, or in adaptive modulation and coding. In the more modern view of wireless communications, feedback information also plays a central role. In addition to the traditional uses, the introduction of novel sophisticated transmission techniques such as linear precoding, beamforming, scheduling, and pre-cancelling of interference has further emphasized the importance of accurate feedback information in future wireless systems. Practical constraints however prevent the transmitter side in wireless communications from obtaining feedback information of arbitrarily high quality; and thus in most cases we have to deal with scenarios where the amount of feedback is strictly limited.

This thesis aims at a better understanding of the theoretical limitations of wireless communication systems with limited feedback information. A particular emphasis is placed on applications sensitive to *delay*, motivating the investigation of the so-called quasi-static fading channels. Focusing on this channel model, we will identify and investigate fundamental tradeoffs in many different communication systems. Our work has some interesting implications in the design of multiple-antenna systems, cooperative communications, and source–channel coding.

1.1 Communication Systems

We begin with a brief introduction to some pivotal information-theoretic concepts and properties of communications over wireless channels. The relations to the topics treated in the thesis will be discussed when appropriate.

To introduce some important concepts, we consider the simple but quite general discrete-time model of a communication system depicted in Fig. 1.1. The mission of the system is to send a message from a transmitter to a receiver over a *channel*. The channel is characterized by a conditional probability density of the output given the input, $p(y_1^T|x_1^T)$. This essentially represents the randomness (uncertainty) added to the transmitted signals, which may come from e.g., thermal noise, interference, and the physical medium. To protect the message from the possibly detrimental effect of the channel, some redundancy is added to the actually transmitted signals in the form of *channel coding*.

More precisely, at the transmitter, an integer *message* m, assumed to take equally likely values on the set $\{1, \ldots, 2^{RT}\}$, is mapped (encoded) into a sequence of symbols of length T to be transmitted over the channel, x_1^T. We say that the transmission consumes T *channel uses*. Such a sequence is referred to as a *codeword*, and the integer T is the *codeword length*. The set of all possible 2^{RT} codewords is called a *codebook*, which is known to both sides of the communication link. Normally some cost functions are associated with the codewords to represent the physical limitations of the transmission. An example of a cost function is the typical power constraint that keeps either the average or the peak transmit power below a certain threshold. At the receiver, a decoder attempts to detect which message has been sent, based on its received sequence y_1^T, and the result is the decoded message $\hat{m} \in \{1, \ldots, 2^{RT}\}$. The communication system attempts to convey RT bits of information through the channel after T channel uses, thus the rate of the code is said to be R bits per channel use. This definition of code rate should clearly be distinguished from other definitions used elsewhere, e.g. in [Pro95] where the code rate, a quantity less than unity, is used to indicate the level of redundancy of a code. The simple model Fig. 1.1 indeed includes the basic building blocks of a quite general communication system using channel coding (but perhaps too general to be actually implemented).

The above model is perhaps the most classical way to represent a point-to-point communication system, but is not the only one. In Chapter 2, we will deal with a system employing *multi-layer coding* where multiple messages $m_1, \ldots, m_L$ are mapped into a single sequence to be transmitted and then successively decoded at the receiver. At a first look, it seems that such an approach is a special case of the above model because multiple messages can be combined into a single message with a larger codebook, and therefore the chance of a transmission failure (error) can only increase. This is not necessarily true, however, because "failure" can be measured in different ways depending on the applications and characteristics of the channel considered. Such information-theoretic performance measures will be discussed in more details in Section 1.5.

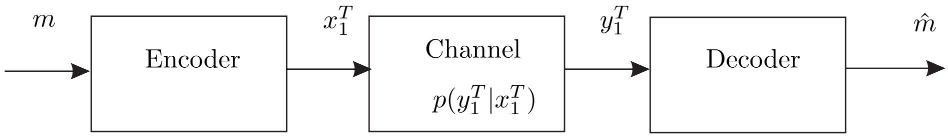

Figure 1.1: A communication system.

1.2 Fading Channels

We will now review the typical characteristics of the physical medium in a wireless environment and relate these to a more specific model of the "channel" in Fig. 1.1. In particular, the additive Gaussian noise model where the transmitted signals, after travelling through a medium, is corrupted by the addition of white Gaussian noise at the receiver, will be considered. However, one of the most distinguishing features of a wireless channel does not come from the properties of the noise, but from the time-varying nature of the underlying physical media.

The time-varying nature of wireless channels is generally governed by two dominating terms. The so-called large-scale fading term is caused by path loss and shadowing as the transmit signals travel over distance and get obstructed by large obstacles [TV05]. This however happens in a much larger time scale (i.e., changing much slower) than the duration of a symbol or a codeword. In this work we are more interested in smaller-scale effects, described as follows.

In a wireless environment, the transmitted signals normally propagate to the receiver via many different paths. For example, the transmitted signals from a mobile station can be reflected from buildings, cars and other obstacles before reaching the receiver. At the receiver, these signal components may add destructively, as they undergo different attenuations and arrive at different delays. The fluctuation of received signal strength due to multi-path is known as (small-scale) *fading*. If the bandwidth of the transmitted signal is much smaller than the coherence bandwidth B_{m} [Pro95] of the channel, all the frequency components of the transmitted signal will suffer almost the same attenuation and phase shift. Therefore, in this case the channel is called frequency-nonselective or *flat fading*. A flat fading channel is well modelled as an equivalent time-varying one-tap filter with complex-valued coefficient, illustrated in Fig. 1.2. We usually encounter the case when this coefficient, or *channel gain*, is modeled as a zero-mean complex Gaussian random variable. This represents a rich scattering environment with a lot of reflection paths and no direct line-of-sight component. Such a channel is called a Rayleigh fading one because the amplitude of the channel gain is Rayleigh distributed.

On the other hand, when the bandwidth of the transmitted signal is larger than the coherence bandwidth, the components that separate more than B_{m} in frequency will suffer almost uncorrelated gain and phase offset. The channel in this case is called *frequency-selective*, and is usually modeled as a time-varying tapped delay line with complex-valued coefficients. Transmission over a frequency-selective channel

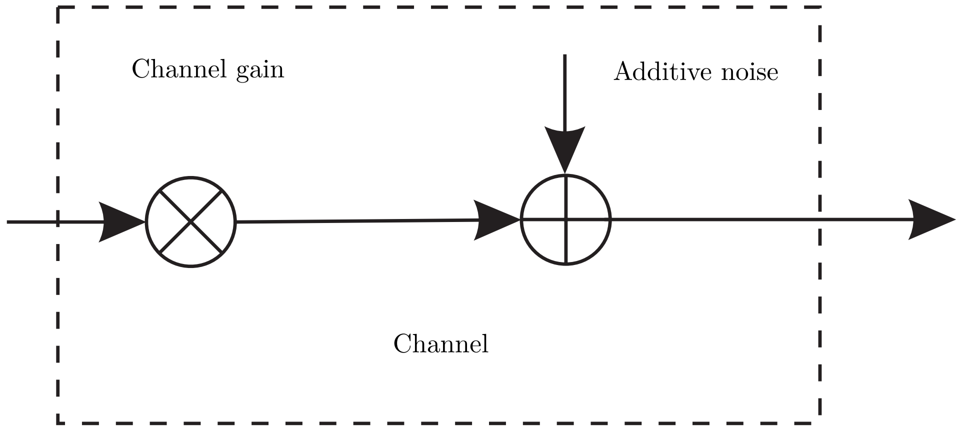

Figure 1.2: An additive-noise flat fading channel.

results in inter-symbol interference (ISI) and may require complicated equalization in the time domain. Nowadays, it is generally agreed that a common technique known as orthogonal frequency division multiplexing (OFDM) can often be applied to convert a frequency-selective channels into a set of parallel narrow-band, flat-fading channels (as usual, under some optimistic assumptions). The thesis therefore focuses only on the flat fading case.

Fading is traditionally seen as problematic for communication systems, as it may cause significant degradation in the performance, especially in deep fades (when the received signal power drops too low to be useful). A conventional method to combat fading is through *diversity* techniques. The basic idea is to provide several independent copies of the transmitted signal to the receiver so that the probability that all these copies simultaneously suffer deep fades is very small. Common diversity techniques include time, frequency and space diversity, as well as the combinations of these methods. With frequency diversity, the signals carrying the same information are transmitted on several carrier frequencies. If the separation between any two carrier frequencies exceeds the coherence bandwidth B_{m}, then each received version can be considered to undergo independent fades. With time diversity, the signals carrying the same information can be transmitted at different time instants such that the time separation between any two copies exceeds the coherence time T_{d} of the channel. One way to achieve this in wireless communications is by *interleaving* a codeword before its transmission.

1.3 Slow Fading

The classification of fading channels into fast and slow ones is critical in order to determine a suitable information-theoretic performance measure for the system of interest, as will be elaborated in Section 1.5. Throughout this work, the term "slow fading" does not necessarily reflect the speed of change of the underlying physical medium like in e.g., [Pro95], but relates to the *delay constraints* of the transmission. For applications completely insensitive to delay constraint, the receiver can, in principle, wait for an unlimited amount of time before attempting to decode. A codeword therefore can be assumed to span an infinite number of independent fading blocks. In practice that models a system with a relaxed delay constraint so that it enjoys near perfect interleaving, and thus a codeword can span a large number of independent fading states to exploit a significant amount of time diversity. One can think of downloading a large file for several hours, even days. We will refer to such a channel as an ergodic one, or less technically, a *fast* fading channel. On the other hand, for applications that require a stricter delay constraint such as real-time voice and video transmission, a codeword can only span a finite, typically *small*, number of fading blocks. The length of each fading block, where the channel gain remains constant, is typically large enough to average out the effect of noise, thus studying the system behavior in the limit of infinite block length still makes sense, even though this is seemingly contradicting to the delay-limited assumption [BCT01]. We refer to this kind of channel as a *slow* (or slowly) fading one to distinguish this from the fast fading case. Furthermore, in this thesis, we exclusively focus on the extreme case where a codeword spans a *single* fading block, i.e., the so-called quasi-static fading channel.

1.4 Channel-state Information

The performance of a communication system is greatly influenced by the assumptions on the available channel-state information (CSI) at both sides of the links. The term CSI in this thesis refers to the possibly imperfect information about the *realization* of the channel gain (or a channel matrix in a multi-antenna channel, as presented later in Section 1.6). We then distinguish between CSI at the transmitter (CSIT) and CSI at the receiver (CSIR).

In both theory and practice, CSIR is considered "easier" to acquire. A typical way to obtain CSIR is by sending a training sequence known a priori to the receiver so that the it can estimate the channel gain with a certain level of accuracy (assuming that the channel gain does not change significantly until the next training sequence is sent). A thorough analysis of training schemes is presented in [HH03]. On the other hand, CSIT is more difficult to obtain. In time division duplex (TDD) systems where uplink and downlink transmission takes place in the same frequency, the reciprocity of the channel can be exploited to estimate the reverse channel gain (provided that the time separation between uplink and downlink slots is considerably smaller than the coherence time). The reciprocal properties generally cannot be exploited in frequency division duplex (FDD) systems, where obtaining CSIT requires some form of *feedback*.

To simplify the analysis and highlight the effect of partial CSIT, throughout the thesis perfect CSIR is always assumed. Of course, in practice, the imperfectness of CSIR must also be taken into careful consideration because this may lead to remarkable changes in the behavior of some information-theoretic measures [LM03]. This thesis exclusively focuses on an explicit quantized feedback model where CSIT is obtained via a noiseless, zero-delay dedicated feedback link, depicted in Fig. 1.3. In particular, given a channel gain, the receiver employs an index mapping to obtain an integer feedback index belonging to a finite set $\{1, \ldots, K\}$ and sends it back to the transmitter prior to the transmission of a codeword. The constant K is referred to as the *feedback resolution*. Clearly for that approach to work the transmitter and receiver must agree on a common strategy with the parameters designed *off-line*. More realistic assumptions regarding the feedback link should be taken into account for any practical system, e.g., the case of noisy feedback link is treated in [JS04]. Nevertheless, the feedback resolution that we considered in this work is generally low (corresponding to 1-2 bits of feedback per fading block) so that the feedback delay can be considered insignificant and low-complexity forms of channel coding in the feedback link are also possible, making the zero-delay noiseless feedback link a relatively reasonable assumption.

The explicit quantized feedback model in Fig. 1.3 is not the only model for limited feedback. Other models may impose different assumptions, for example, that a noisy estimate of the channel is available at the transmitter [JSO02]. Another line of thought assumes that only the long-term statistics of the channel gain is available at the transmitter, for example, the case of correlated channels with covariance matrix known at the transmitter is studied thoroughly in e.g., [VM01, JB04, JG04, VP06]. A hybrid model combining both long-term statistics and short-term information regarding the channel gain realization is presented in [KBLS06]. There are also the interesting and challenging cases when CSI is *not* known by any party of the communication link. Such noncoherent communication systems, see e.g., [MH99, ZT02, GS07], are outside the scope of this thesis.

1.5 Some Information-theoretic Performance Measures on Fading Channels

Perhaps the most important information-theoretic limitation of a communication channel is the *channel capacity*, introduced in Shannon's seminal work [Sha48] (see also [CT91]). Roughly speaking, the capacity C of a channel lets us know the upper limit on the rate of reliable communication over that channel. That is, with an error defined as the event that the transmitter and receiver disagree on what has been sent, i.e., $\hat{m} \neq m$ in the model in Fig. 1.1, then for any positive number $R < C$, it is possible to find codes of rate R that yield arbitrarily small probability of error, $\Pr(\hat{m} \neq m)$, provided that the codeword length T is sufficiently large. For a memoryless channel, i.e., if $p(y_1^T|x_1^T) = \prod_{i=1}^{T}(y_i|x_i)$, the capacity is given by Shannon's famous maximum-mutual-information formula.

Fast fading channels belong to a general class of information-stable channels,

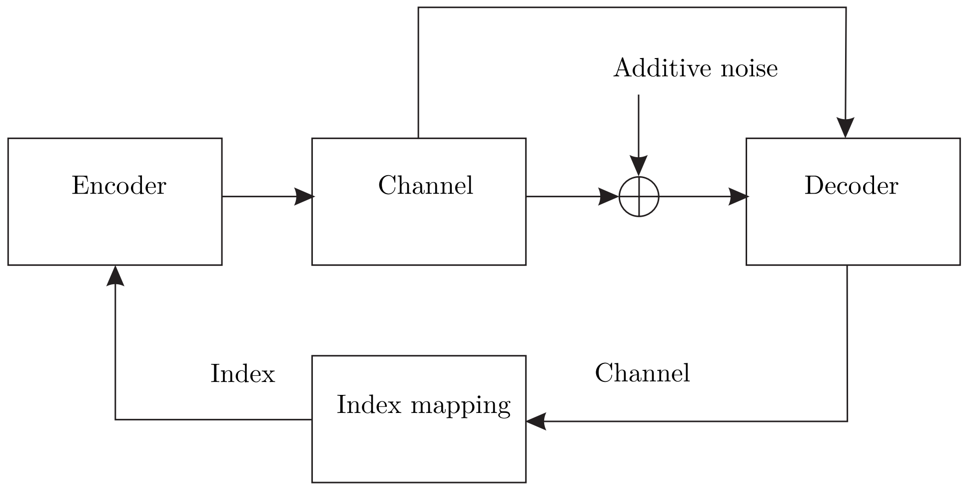

Figure 1.3: A quantized feedback model.

that is, roughly speaking, channels for which a capacity-achieving input and the resulting output behave ergodically [VH94]. Naturally, in the literature, the capacity of a fast fading channel may also be explicitly termed the ergodic capacity.

On the other hands, for slow fading channels the ergodicity assumption does not hold. To compute the capacity of a slow fading channel, one therefore should use the general formula introduced in [VH94], which holds for an arbitrary channel. Slow fading channels are often discussed in the framework of *compound channels*, see e.g., [BBT59][RV68], where the transition distribution is parameterized by some θ, i.e., $p(y_1^T|x_1^T;\theta)$. A compound channel where the parameter θ is associated with an a priori distribution is sometimes called a composite channel [EG98, BPS98]. For example, in the flat-fading model in Fig. 1.2, the amplitude of the channel gain may take the role of θ and be associated with e.g., a Rayleigh distribution.

However, in slow fading channels, the channel capacity generally does not give a useful and complete picture. For example, a slow fading Rayleigh channel with perfect CSIR and no CSIT has a pessimistic *zero capacity.* This therefore motivates the framework of capacity versus outage, first mentioned in [OSW94]. The instantaneous mutual information between the input and the output of a slow fading channel is a *random variable,* depending on the actual realization of the channel gain. Given a code rate, there is a probability that the current realization of the mutual information is strictly smaller than the code rate, thus reliable communication is not possible no matter how large the codeword length T is and how good the codes are. In such a situation, the system is said to be in *outage.* There is obviously a tradeoff between the code rate and the probability of an outage event: High code rates lead to a high outage probability (more unreliable communication), while a lower code rate increases reliability but also reduces overall throughput.

The capacity versus outage framework, however, is not the only performance measure for a slow fading channel. For certain applications, it may be better to split a message into several ones so that any of them, if successfully decoded at the receiver, can improve the performance. For example, a coarse version of an image can be obtained if some messages are correctly decoded, and a finer, higher-quality image can be reproduced if more information is available. It should be clearly emphasized that such a *multi-layer* approach is *not* suitable for many applications. For instance, in data communications an error is declared whenever *any* layer is incorrectly decoded, thus adding extra layers is generally not an appealing choice.

A frequent performance measure of multi-layer coding is the *expected rate*. Herein we avoid the term "expected capacity" as used in e.g., [BPS98] and adopt the more moderate term "expected rate" from [VH94, Cov72] instead. This is both to avoid confusion with capacity in the traditional sense of the word and to emphasize that, to our knowledge, the multi-layer coding approach has not been shown to be optimal in any sense. Expected rate can be seen as the rate that can be correctly *received*, averaged over the randomness of the channel and the noise. This is therefore also called reliably received rate in [EG98]. Interestingly, one of the main motivations of Cover's seminal work on broadcast channels [Cov72] was to improve the expected rate over a compound channel. This interesting concept has reemerged recently in [Sha97, SS03], where the asymptotic case with a continuum of layers using differential rate and power is studied. Later, Liu et al. showed that most of the gain of infinitely many layers of codes can be realized by a simple two-layer coding scheme for many common channel distributions [LLTF02]. Multi-layer coding is closely related to the study of the capacity region for a general broadcast channel, a long standing problem in information theory [VH94].

All the aforementioned work assumes perfect CSIR and no CSIT. The presence of CSIT changes the picture dramatically. For fast fading channels, Goldsmith and Varaiya studied a scalar Gaussian channel with perfect CSI at both sides of the link and showed that allocating power in a water-filling manner is optimal in a capacity sense [GV97]. However, for most common channel statistics, the benefit of CSIT in terms of capacity is not significant, especially at high signal-to-noise ratio (SNR). The achievability part in [GV97] relies on the multiplexing of *multiple* codebooks. It was later clarified that a simpler combination of a *single* codebook and a CSIT-dependent power amplifier is sufficient to achieve capacity [CS99]. Furthermore, such a separated structure is optimal even when CSIT is causal and imperfect, under certain assumptions. This holds also for multiple-antenna scenarios, where CSIT-dependent "transmit weighting" and coding are separated [SJ03].

The presence of CSIT in slow fading channels gives rise to the interesting concept of power control. If the transmit power can be varied according to the current channel gain, outage can be completely avoided even for a strictly positive code rate. Under certain assumptions, this is possible with a *finite average* transmit power over infinitely many codewords. In other words, the capacity of such a channel is strictly positive, even though it is a slow fading one. To distinguish this notion from the (ergodic) capacity in the fast fading case, this is referred to as *delay-*

limited capacity in [HT98, CTB99]. Of course, in connection to the discussion in this section, delay-limited capacity is precisely the capacity in a traditional (Shannon's) sense, applied to a special channel model.

1.6 Multiple-antenna Systems

Using multiple antennas is identified as a promising approach to improve the performance of wireless communications in fading environments. Although space diversity has long been utilized by means of multiple receive antennas, only recently has knowledge about communications with multiple antennas placed at both the transmitter and the receiver reached a new level of maturity. Such so-called multiple-input multiple-output (MIMO) systems have been an extremely active research topic over the last decade.

Seminal work by Telatar [Tel99] (see also the work by Foschini [Fos96]) showed that in a system with only CSIR, using N_t transmit antennas and N_r receive antennas, where the components of the channel matrix are independent and identically distributed (i.i.d.) zero-mean complex Gaussian, the ergodic capacity at high SNR is approximately

$$C \approx \min(N_\mathrm{r}, N_\mathrm{t}) \log \mathrm{SNR}.$$

That is, in terms of capacity a significant gain of $\min(N_\mathrm{r}, N_\mathrm{t})$ can be expected at high SNR compared to a single-antenna system. Unsurprisingly, their promising results have sparked great interests in MIMO communications.

A codeword in a MIMO communication system is a matrix, with both spatial and temporal dimensions to be exploited, giving rise to the term *space-time coding*. In [TSC98], a sufficient condition for a space-time code to achieve "full diversity" is presented. The developed criterion is quite mild, requiring all codeword difference matrices to be full rank. A surprisingly simple but extremely powerful space-time block code for two transmit antennas is introduced by Alamouti in [Ala98]. Among the attractive properties of Alamouti's code are its simplicity in combining and decoding and its ability to extract the full diversity of the channel. Later, it is shown in [TJC99] that Alamouti's codes belong to a general class of orthogonal space-time block codes (OSTBC). Unfortunately, in [TJC99], it is also shown that "full rate" OSTBC's using symbols drawn from a complex constellation (such as QAM) do not exist for more than two transmit antennas. Some extensions of OSTBC's are also proposed in [Jaf01], compromising receiver complexity and performance. However, except for the setting of two transmit and one receive antennas, OSTBC's display a performance loss compared to the more general linear dispersion codes designed to maximize mutual information in [HH02], over certain ranges of SNR. Decoding linear dispersion codes generally requires a complicated maximum likelihood receiver (assuming equally likely codewords), or some near-maximum likelihood such as the sphere decoder.

The aforementioned space-time codes are of relatively short length. Combining multiple-antenna and more sophisticated error-correcting codes such as trellis codes

[TSC98], turbo codes [BGT93], low-density parity check codes [Gal62] and variations such as repeat-accumulate codes also provides significant extra gains. For fast fading MIMO channels, very close to capacity performance can be achieved with long random-like codes and joint iterative detection-decoding, see e.g., [SD01, HtB03, tKA04, tK03].

Let us briefly review some work in MIMO channels with some forms of CSIT. In the frontier of fundamental limits, for a constant channel matrix with full CSI at both sides, a singular value decomposition converts the channel matrix into a set of parallel spatial channels, and therefore power allocation in a water-filling manner is optimal in a capacity sense [Tel99]. This is readily extendable to fast fading channels with full CSI, where water-filling over both time and space is optimal. With limited feedback, capacity results for fast fading channels are reported in [SJ03, LLC04b, LLC04a]. For slow fading channels, the probabilistic power control framework in [CTB99] is extended to the MIMO case in [BCT01]. Their scheme is not suitable for exploiting time diversity because of a noncausality assumption. The causal case is solved under a dynamic programming framework in [NC02]. The concept of minimum rates is independently proposed in [LLYS03] for a single-user channel and in [JG03] for a broadcast channel, which leads to an interesting solution combining both water-filling and channel inversion.

More practical use of partial CSIT in multi-antenna systems has attracted a great deal of attention recently. Given that a large number of complex channel coefficients needed to be quantized, it becomes more difficult to “imagine” what to send back to the transmitter with e.g., 1 bit. Early studies include the design of a precoding matrix influenced by possibly impaired CSIT to improve the performance of OSTBC's [JSO02, JS04]. Vector quantization techniques are applied in [NLTW98] to design feedback schemes under different optimization criteria. Limited feedback design using an elegant geometrical framework is pursued in [LHS03, LH05, MSEA03].

Cooperative Communications

In certain scenarios, deploying multiple antennas at some parties in a communication link may be difficult or even infeasible due to practical constraints. Interestingly, even under such restrictive conditions, it is possible to form *virtual* antenna arrays by letting these parties *cooperate* in an intelligent way. While seminal work in this area appeared a long time ago in the context of relay channels [van71, CE79], interests in cooperative communications have only renewed with the series of papers [SEA03a, SEA03b, NBK04, LW03, LTW04, KGG05, HZ05].

Herein we exclusively focus on the classical three-node model of the relay channel [van71, CE79], where a transmitter (source node) communicates with a receiver (destination node) with the help of a relay node that can assume a transmitting and/or a receiving role. Despite the simplicity of the model, the capacity of such a general relay channel is unknown. In the literature, relaying systems can be categorized as either half-duplex (the relay cannot receive and transmit at the same time) or full-duplex (the relay can transmit and receive simultaneously) ones. In this thesis, we exclusively focus on half-duplex channels.

1.7 Diversity–Multiplexing Tradeoff

Most early work on space-time coding either tried to extract a "full" diversity gain [Ala98, TJC99, TSC98], or to achieve "high rates," e.g., the vertical Bell Labs layered space-time (BLAST) structure [TV05] and the linear dispersion codes [HH02]. A new line of thought is pursued in [ZT03], where it is shown that both types of gains can be *simultaneously* achieved over a slow fading channel, with a fundamental tradeoff between them. Such an elegantly characterized tradeoff is referred to as the diversity-multiplexing tradeoff, and has sparked a great deal of attention, even if it is rather *coarse* (defined in the limit of $\mathrm{SNR} \to \infty$). Roughly speaking, the diversity gain d lets us know about the asymptotic slope of the error probability while the multiplexing gain r reflects how large the code rate is compared to the capacity of a single-antenna channel at high SNR. At high SNR, given a code rate $R = r \log \mathrm{SNR}$, an error probability in the order of SNR^{-d} can be achieved with "good" codes. In other words, this is a high-SNR tradeoff between reliability and throughput of a multi-antenna system. Notice that the notion of diversity gain in [ZT03] should not be interpreted (in a traditional way) as the number of independently faded copies of the transmit signals as seen at the receiver.

The diversity-multiplexing tradeoff is closely related to the theory of error exponents [Gal65, Gal68]. Error exponent techniques however involve an optimization over all probability distributions that is very difficult to solve in general. By restricting to a Gaussian distribution and letting the SNR grow unbounded, Zheng and Tse have been able to characterize exactly the asymptotic SNR exponent of an error event. The key idea is to analyze the asymptotic behavior of the joint probability density function of the singular values of an i.i.d. complex Gaussian matrix, under a powerful large-deviations framework.

In the original work [ZT03], it is shown that there exist codes with *finite length* that can achieve the optimal diversity-multiplexing tradeoff. In particular, for a channel matrix of size $N_\mathrm{r} \times N_\mathrm{t}$, the codeword length $T \geq N_\mathrm{t} + N_\mathrm{r} - 1$ is sufficient to achieve an *error* probability that decays as fast as the *outage* probability does. This is rather surprising, as one may have expected that it is only asymptotically achievable with infinitely long codewords. However, that conclusion is based on a random coding argument, with the only practical coding scheme known to be diversity-multiplexing optimal at that time was (again, somewhat surprisingly) the simple Alamouti's scheme with a QAM constellation for the 2×1 channel. As an example (from [ZT03]), the diversity-multiplexing tradeoffs achieved by some space-time codes together with the optimal one over a 2×2 channel are plotted in Fig. 1.4. As can be seen Alamouti's codes are strictly better in a tradeoff sense than are the simple repetition codes, even though both can achieve "full diversity," i.e., the diversity at very small rates compared to $\log \mathrm{SNR}$. However, none of these

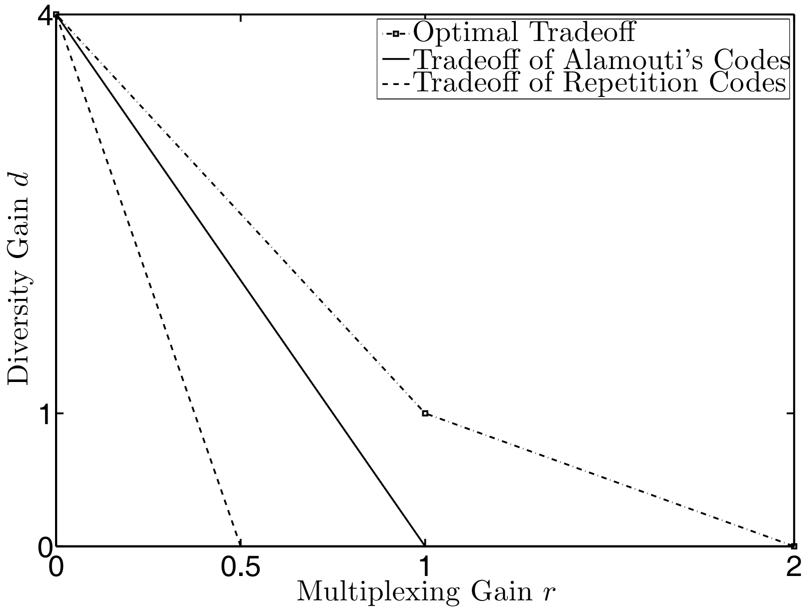

Figure 1.4: Diversity–multiplexing tradeoff over a 2×2 channel.

schemes are tradeoff optimal, especially at high multiplexing gains (they cannot be used to achieve "high rates").

Subsequently, the design of other short-length space-time codes that achieve the *entire* diversity-multiplexing tradeoff has then quickly become a very active research area. Among the first codes designed towards that goal are the lattice space-time (LAST) codes [ECD04] and their variants [ECD06]. To be precise, LAST is still a random ensemble, albeit is more structured than a Gaussian ensemble and thus allows for more efficient algorithms than a maximum likelihood search such as the sphere decoding, see e.g., [AEVZ02, JO05]. Even randomly generated LAST codes are shown to perform very well. In [YW03], Yao and Wornell explicitly constructed a family of codes for the 2-transmit-antenna case ($N_{\mathrm{t}} = 2$), using a carefully chosen rotation matrix and symbols taken from QAM constellations. Interestingly, they showed that there exist codes of length $T = 2$ that can achieve the entire tradeoff, for any number of receive antennas $N_{\mathrm{r}} \geq 2$, while the Gaussian coding argument in [ZT03] can only show the existence of codes with length $T \geq N_{\mathrm{t}} + N_{\mathrm{r}} - 1 = N_{\mathrm{r}} + 1$ in similar settings. One of the key ideas in [YW03] is to find a sequence of codes so that all codeword difference matrices have a nonvanishing or sufficiently slow decaying determinant as the code rate grows. Explicit code design based on that nonvanishing determinant criterion is studied extensively, see e.g., [BRV05, RBV04, EKP$^+$06]. The D–M tradeoff optimality of these codes can be explained in the framework of approximately universal codes [TV06], which characterizes codes having all pairwise error probabilities decay exponentially as SNR $\rightarrow \infty$, as long as the channel is not in outage. An alternative, geometric interpretation of such a class of codes together with their applications in MIMO channels with feedback are presented in [KS07].

The tradeoff between throughput and reliability naturally exists in slow fading relay channels. In [LTW04], some basic relaying strategies including AF and DF are described and investigated. The D–M tradeoffs of the schemes studied in [LTW04] show that they are not efficient in the high multiplexing gain regime. Furthermore, it is later clear that these simple schemes are sub-optimal except at zero multiplexing gain (i.e., they can achieve the maximum possible diversity gain). In the context of multiple-relay AF channels, a novel scheme termed as "slotted AF," which is proposed and analyzed in [YB07b], can outperform other known AF protocols in the literature. An intelligent scheme called dynamic DF is proposed in [AES05]. DDF relaying uses rateless codes at the source node and one acknowledgement bit from the relay node to inform the source *when* the decoder at the relay succeeds. It turns out that the D–M tradeoff of this strategy is *strictly optimal* for all multiplexing gains less than $\frac{1}{2}$. For higher multiplexing gains, while it is unclear whether DDF is still optimal or not, there is no known scheme operating under the same CSI assumptions that can outperforms DDF. Under a much more relaxed assumption on the CSI that the relay knows both the relay-destination and source-destination channel gains, it is shown in [YE07] that compress-and-forward relaying is D–M tradeoff optimal at *all* multiplexing gains.

1.8 Distortion Exponent

Inspired by the concept of diversity gains in [ZT03], some researchers have recently applied the main ideas of the D–M tradeoff to the problem of transmitting an *analog source* over slow fading channels [CN05, GE05, HG05]. It is noticeable that even over a simple scalar point-to-point slow fading channel, the celebrated separation theorem [CT91, Chapter 8] does not hold when the CSI is not fully known at the transmitter. That is, designing source and channel coding modules separately does not necessarily lead to optimal performance. For example, sophisticated hybrid digital–analog (HDA) joint source–channel coding schemes have been shown to outperform separate coding significantly [MP02, SPA02].

Focusing on source transmission over slow fading MIMO channels with no CSIT, the works in [CN05, CN07, GE05, GE08] quantified the asymptotic behavior of the end–to–end average distortion achieved by different joint source–channel coding schemes in the regime of high SNR. At high SNR, the average distortion behaves as $\text{SNR}^{-d'}$, bearing a clear similarity to the behavior of the outage probability (and also the probability of error) when a *message* is transmitted over the channel. The quantity d' is often referred to as the (SNR) distortion exponent, which measures the *slope* of the average end–to–end distortion on a log-log scale at high SNR.

There is a performance tradeoff between the distortion exponent and the so-called bandwidth ratio b, which is defined as the ratio between the channel bandwidth and the source bandwidth. The bandwidth ratio b measures the spectral efficiency of the system, with a smaller b reflecting a more efficient schemes that consumes less channel bandwidth to transmit a given source. Note that the bandwidth ratio b does not have any direct connection to the multiplexing gain r in the D–M tradeoff analysis.

The distortion exponent analysis in [CN05, CN07, GE08] provided some fresh insight into the problem of source transmission over slow fading channels. For example, in [CN07], it is shown that over MIMO channels a simple HDA scheme is *optimal* in the sense that it achieves the best tradeoff between distortion exponent and bandwidth ratio. This optimality however holds only for a range of *sufficiently small* bandwidth ratios, i.e., for highly-compressive systems. In [GE08], a broadcast strategy where the transmitter sends a layer of codes – as considered in [Sha97] and also in Chapter 2 – is shown to be optimal in a distortion exponent sense for a range of *sufficiently high* bandwidth ratios. The transmission schemes for MIMO channels [GE08] are later extended to the relay settings in [GE07b].

1.9 Contributions and Outline

The common theme of the thesis is the design and analysis of certain communications systems with limited feedback over quasi-static fading channels. The thesis is divided into three parts, with each part treating a different performance metric.

Part I: Chapter 2

This part deals with the optimization of the expected rate over slowly fading scalar channels with quantized side information. In particular, we consider a multiple-layer variable-rate system employing quantized feedback to maximize the expected rate over a single-input single-output slow fading Gaussian channel. The transmitter utilizes partial channel-state information, which is obtained via an optimized resolution-constrained feedback link, to adapt the power and to assign code layer rates, subject to different power constraints. To systematically design the system parameters, we develop a simple iterative algorithm that successfully exploits results in parallel broadcast channels [Tse97]. We present the necessary and sufficient condition for single-layer coding to be optimal, irrespective of the number of code layers that the system can afford. The key observation in this chapter is that unlike in the ergodic case [GV97], even coarsely quantized feedback can improve the expected rate considerably. Our results also indicate that with as few as one bit of feedback information, the role of multi-layer coding reduces significantly.

The material in this chapter has been published as

- T. T. Kim and M. Skoglund. On the expected rate of slowly fading channels with quantized side information. In *IEEE Transactions on Communications.*, volume 55, pp. 820-829, April 2007.

A shorter version also appeared in

- T. T. Kim and M. Skoglund. On the expected rate of slowly fading channels with quantized side information. In *Proc. 39th Asilomar Conference on Signals, Systems, Computers*, Pacific Grove, CA, October-November 2005.

Part II: Chapters 3, 4, and 5

In Chapter 3, we study a slow fading MIMO channel where the transmitter has access to partial CSIT, which takes the form of $\log_2 K$ noiseless feedback bits. It is assumed that the code rate grows as the long-term average transmit power increases, but does not adapt to the channel state, i.e., a single-rate system is considered. We first characterize the entire tradeoff between the diversity and multiplexing gains that can be simultaneously achieved over this channel. Partial power control is shown to be instrumental in achieving the optimal tradeoff over such a system. Our results indicate that the diversity gain can be increased considerably even with coarsely quantized channel state information, especially at low multiplexing gains. For example, as long as at least one side of the communication link has more than one antenna, the maximum diversity gain of the system grows exponentially in the number of feedback regions.

We then carry out the D–M tradeoff analysis for a *variable-rate* MIMO system with quantized feedback. To make "reliability" more meaningful in this variable-rate setting, the concept of minimum guaranteed multiplexing gain in the forward link is introduced and shown to influence the tradeoff remarkably. The results suggest that the optimal D–M tradeoff can be achieved by using just two codebooks: one high-rate codebook that determines the multiplexing gain of the system and the other low-rate codebook that provides the minimum level of quality of service. This holds even if the number of feedback regions is greater than two. Partial power control allows for a superior diversity gain, which is possible even in the high-multiplexing-gain regime.

We then discuss the achievability of the optimal D–M tradeoff by finite-length codes that exist in the literature. In particular, codes that satisfy the approximately universal criterion [TV06] are shown to be also D–M tradeoff optimal in our partial-CSIT scenario. We also present a useful geometrical interpretation of the approximately universal criterion.

Finally, we present two lower bounds to the optimal D–M tradeoff using Gaussian random coding argument. Unlike in the original setting (no feedback information) studied in [ZT03], these lower bounds are only asymptotically tight in the limit of large block (codeword) lengths. Nevertheless, we show that the new achievable bounds can approach the optimal D–M tradeoff closely even with moderate codeword lengths.

The material in this chapter has been published in

- [KS07] T. T. Kim and M. Skoglund. Diversity–multiplexing tradeoff in MIMO channels with partial CSIT. In *IEEE Transactions on Information Theory.*, volume 53, pages 2743-2759, August 2007.

Conference versions of this work have also appeared in

- [KS06a] T. T. Kim and M. Skoglund. Diversity–multiplexing tradeoff of MIMO systems with partial power control. In *Proc. 2006 Zurich Seminar on Communications*, Zurich, Switzerland, February 2006.
- [KS06c] T. T. Kim and M. Skoglund. Partial power control. In *Proc. 2006 IEEE International Conference on Communications*, Istanbul, Turkey, June 2006.
- [KS06b] T. T. Kim and M. Skoglund. Outage behavior of MIMO channels with partial feedback and minimum multiplexing gains. In *Proc. IEEE Symposium on Information Theory,* Seattle, WA, July 2006.

Chapters 4 and 5 investigate the D–M tradeoff of three-node scalar relay channels under different assumption of CSI at the source and the relay. Chapter 4 considers a relay channel with quantized CSI at the relay and the source. We present a rather exhaustive study considering many different possible scenarios with quantized (channel state) feedback from the relay to source, from destination to relay, and from destination to both source and relay. We show that using one bit from the relay to control the source transmit power is sufficient to achieve the multiantenna upperbound in a range of multiplexing gains. Systems with feedback from destination to control relay transmit power slightly outperform DDF at high multiplexing gains, even with one bit of feedback. Finally, we show that with feedback from destination, if the source-relay channel gain is unknown to the feedback quantizer at the destination, the diversity gain only grows linearly in the number of feedback levels, in sharp contrast to an exponential growth for MIMO channels as shown in Chapter 3.

In Chapter 5 we study a more idealistic channel where the relay node has perfect knowledge about the channel gains of the source-destination and relay-destination links. The motivation of this study is the optimality of the compress-forward relaying protocol using Wyner-Ziv (WZ) coding under the same CSI assumptions [YE07]. We pose the fundamental question: "How much of the gain in the CF scheme [YE07] comes from the perfect CSIT, and how much comes from WZ coding?" To answer this question, we quantify the asymptotic loss of compress-forward relaying with simple quantization at the relay (i.e., without using source coding with side information as in [YE07]). It turns out that in terms of the D–M tradeoff, the loss of not using WZ coding is dramatic. However, we also obtain a more optimistic result that using power control at the relay can fully compensate for this loss, as long as the multiplexing gain is not greater than $\frac{2}{3}$.

The material in these chapters has been submitted for possible publication in

- [KCS07a] T. T. Kim, G. Caire, and M. Skoglund. Decode-and-forward relay channels with quantized channel state feedback: An outage exponent analysis. Submitted to *IEEE Transactions on Information Theory,* 2007; revised

2008.

- [KSC07b] T. T. Kim, M. Skoglund, and G. Caire. Quantifying the loss of compress-forward relaying without Wyzer-Ziv coding. Submitted to *IEEE Transactions on Information Theory,* 2007.

A short version has been published in

- [KCS07b] T. T. Kim, G. Caire, and M. Skoglund. On the outage exponent of fading relay channels with partial channel state information. In *Proc. IEEE Information Workshop,* Lake Tahoe, CA, September 2007.

Part III: Chapters 6 and 7

This part considers the transmission of a continuous-amplitude source over a slow fading channel. We are exclusively interested in the high-SNR regime and the optimization of the end-to-end expected distortion over the channel.

The main contribution of this part is the investigation of the distortion exponent in the case of limited feedback. Since the distortion exponent analysis essentially investigates how fast the end-to-end mean square distortion decays to zero and SNR grows, the study in these chapters is closely related to D–M analysis. There are fundamental differences though. In particular, the end-to-end distortion can be improved even under a *short-term* power constraint (i.e., using only rate adaptation). This is generally not the case for the outage minimization problem. Furthermore, combining power control with rate adaptation yields a superior distortion performance compared to existing schemes in the literature.

Chapter 6 deals with MIMO channels. We derive upper bounds on the distortion exponents achieved with partial CSIT under a long-term power constraint. It is shown that the exponent achieved with any feedback link of fixed, finite resolution is bounded above by a polynomial of the product between the number of transmit and number of receive antennas. This behavior can be explained in connection with the D–M tradeoff results in Chapter 3. The achievable distortion exponent of some hybrid schemes with heavily quantized feedback is then derived. The results show that dramatic performance improvement over the case of no CSIT can be achieved by combining simple schemes with a very coarse CSIT feedback.

Chapter 6 treats the DF relay channels. It is shown that under a short-term power constraint, combining a simple feedback scheme with separate source and channel coding outperforms the best known no-feedback strategies even with only a few bits of feedback information. Partial power control is shown to be instrumental in achieving a very fast decaying average distortion, especially in the regime of high bandwidth ratios. Performance limitation due to the lack of full channel state information at the destination is also investigated, where the degradation in terms of the distortion exponent is shown to be significant. However, even in such restrictive scenarios, using partial feedback still yields distortion exponents superior to any no-feedback schemes.

The material in these chapters has been submitted for publication as:

- [KSC08b] T. T. Kim, M. Skoglund, and G. Caire "On source transmission over MIMO channels with limited feedback," submitted to *IEEE Transactions on Signal Processing,* 2008.
- [KSC08a] T. T. Kim, M. Skoglund, and G. Caire. On cooperative source transmission with partial rate and power control. Accepted for publication in *IEEE Journal of Selected Area in Communications*, 2008.

A short version has been published in

- [KSC07a] T. T. Kim, M. Skoglund, and G. Caire. Distortion exponents over fading MIMO channels with quantized feedback. In *Proc. IEEE International Symposium on Information Theory*, Nice, France, June 2007.

Contributions Outside the Scope of the Thesis

In addition to the material reported herein, some contributions that are not formally included in the thesis are summarized below.

Combining Linear Precoding and Outer Coding

We propose a simple linear structure to exploit CSIT in a single-user multi-antenna system. When combined with turbo-coded modulation, the proposed scheme performs very close to the capacity limits. With only a few bits per channel use to feedback CSIT, we can achieve a substantial portion of the possible gain with perfect CSIT. The converge behavior of the proposed scheme is then analyzed using extrinsic information transfer charts. Our results show that with the proposed technique, a fixed outer code can interact efficiently with the inner detector under different assumptions about the quality of CSIT.

This work has been presented in

- [KJS04b] T. T. Kim, G. Jöngren, and M. Skoglund. Weighted space-time bit-interleaved coded modulation. In *Proc. IEEE Information Theory Workshop*, San Antonio, TX, October 2004.
- [KJS04a] T. T. Kim, G. Jöngren, and M. Skoglund. On the convergence behavior of weighted space-time bit-interleaved coded modulation. In *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2004.

Limited Feedback Design for Fast Fading MIMO Channels

We propose a transmission scheme combining both short-term and long-term channel state information at the transmitter of a single-user MIMO communication system. Partial short-term CSIT in the form of a weighting matrix is obtained via a resolution-constrained feedback link, combined with a unitary transformation based on the long-term channel statistics. The feedback link is optimized under different power constraints, using vector quantization techniques. Simulations indicate the benefits of the proposed scheme in all scenarios considered.

We later extend the vector-quantization-based approach to the case of the downlink (broadcast) channel, to jointly design the scheduler, the (finite) set of precoding matrices, and the feedback link.

These works have been published in

- [KBLS08] T. T. Kim, M. Bengtsson, E. G. Larsson, and M. Skoglund. Combining long-term and low rate short-term channel state information over correlated MIMO channels. To appear in *IEEE Transactions on Wireless Communications,* 2008.
- [KBLS06] T. T. Kim, M. Bengtsson, E. G. Larsson, and M. Skoglund. Combining short-term and long-term channel state information over correlated MIMO channels. In *Proc. IEEE Conference on Acoustic, Speech, Signal Processing,* Toulouse, France, May 2006.
- [KBS07] T. T. Kim, M. Bengtsson, and M. Skoglund. Quantized feedback design for MIMO broadcast channels. In *Proc. IEEE International Conference on Acoustics, Speech, Signal Processing,* Honolulu, HI, May 2007.

1.10 Notation and Acronyms

In this section we clarify some notation and acronyms used throughout this work.

Notation

$\mathcal{A}$	A calligraphic uppercase letter denotes a set.
$\mathbf{x}$	A boldface lowercase letter denotes a vector.
$\mathbf{X}$	A boldface uppercase letter denotes a matrix.
$\mathbf{I}_N$	Identity matrix of size N.
$\mathbf{x}^{\mathrm{T}}$	The transpose of a vector $\mathbf{x}$.
$\mathbf{x}^{\mathrm{H}}$	The conjugate transpose of a vector $\mathbf{x}$.
$\mathrm{tr}(\mathbf{X})$	The trace of a matrix $\mathbf{X}$.
$\det \mathbf{X}$	The determinant of a matrix $\mathbf{X}$.
$\\|\mathbf{X}\\|_{\mathrm{F}}$	The Frobenius norm of a matrix $\mathbf{X}$.
$\mathbf{X}^{-1}$	The inverse of a nonsingular matrix $\mathbf{X}$.
$\doteq$	The exponential equality, cf. Chapter 3, Section 3.2.
$\lceil x \rceil$	The smallest integer that is not smaller than a (real) scalar x.
$\lfloor x \rfloor$	The largest integer that is not larger than a (real) scalar x.
$(x)^+$	Denotes $\max(x, 0)$.
$\vert\mathcal{A}\vert$	The cardinality of a set $\mathcal{A}$.
$\mathcal{A} \times \mathcal{B}$	The Cartesian product of two sets $\mathcal{A}$ and $\mathcal{B}$.
$\mathrm{E}[x]$	The expected value of a random variable x.

Acronyms

AF	amplify-and-forward
ARQ	automatic retransmission request
AWGN	additive white Gaussian noise
BLAST	Bell Labs layered space-time
CF	compress-and-forward
CSF	channel-state feedback
CSI	channel-state information
CSIR	channel-state information at the receiver
CSIT	channel-state information at the transmitter
DDF	dynamic decode-and-forward
DF	decode-and-forward
D–M	diversity-multiplexing
FDD	frequency division duplex
HDA	hybrid digital-analog
i.i.d.	independent and identically distributed
ISI	inter-symbol interference
KKT	Karush-Kuhn-Tucker
LAST	lattice space-time
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
MMSE	minimum mean-square error
OFDM	orthogonal frequency division multiplexing
OSTBC	orthogonal space-time block codes
p.d.f.	probability density function
QAM	quadrature amplitude modulation
SIMO	single-input multiple-output
SISO	single-input single-output
SNR	signal-to-noise ratio
TDD	time division duplex
WZ	Wyner-Ziv

Part I

Expected Rate

Chapter 2

Expected Rate Maximization

In this chapter, we will show how a scalar measure of performance over the slow fading channels that takes into account both the outage probability and the transmission rate can be improved using partial channel state feedback. We will study a multiple-layer variable-rate system employing quantized feedback to maximize the expected rate over a single-input single-output slowly fading Gaussian channel. The transmitter utilizes partial channel-state information, which is obtained via an optimized resolution-constrained feedback link, to adapt the power and to assign code layer rates, subject to different power constraints. To systematically design the system parameters, we develop a simple iterative algorithm that successfully exploits results in the study of parallel broadcast channels. We present the necessary and sufficient conditions for single-layer coding to be optimal, irrespective of the number of code layers that the system can afford. Unlike in the ergodic case, even coarsely quantized feedback is shown to improve the expected rate considerably. Our results also indicate that with as few as one bit of feedback information, the role of multi-layer coding reduces significantly.

2.1 Introduction

Consider coded data transmission over a slowly fading frequency-flat wireless link. One of the most important performance criteria in this scenario is "throughput" versus "cost" of transmission. Throughput can be measured in many different ways. In this chapter we consider an information-theoretic approach, and will investigate the "achievable expected rate" over a large number of blocks transmitted at variable rates. The cost of transmission will be measured as either "short-term" or "long-term" average power.

To study capacity and related notions over slowly fading channels one needs to be specific about the assumed delay-sensitivity of the applications considered. In applications completely insensitive to delay, a transmitted codeword can be assumed to span infinitely many independent fading blocks, even over very slowly varying channels. In such delay-unconstrained cases, the *ergodic capacity* [GV97, Tel99] is a valid performance limit. However, many wireless applications require a strict constraint on transmission delay. This motivates the block fading Gaussian channel model [OSW94], where a transmitted codeword is assumed to span a fixed and finite number of independent fading blocks. In such scenarios, capacity in the traditional sense of the term [CT91] is generally not a useful performance measure. For example, a block fading Rayleigh channel has capacity zero, since no positive rates are achievable over this channel [BPS98]. Therefore, other ways to characterize the channel, for example in terms of throughput versus *outage probability* [OSW94, CTB99], are often considered in these cases.

When characterizing the achievable performance over a fading channel, one needs also to be specific about the available *channel-state information.* As in most previous related works, we will assume perfect CSI at the receiver, motivated by the separate transmission, at negligible rate-loss, of a training sequence [BPS98]. When available, CSI at the transmitter can be utilized to adapt resources and the transmission strategy and can greatly improve the performance over a slowly fading channel. A fixed-rate system with non-causal and perfect CSIT, employing power control based on the CSIT to minimize the outage probability, was studied in [CTB99, BCT01]. These results were then later extended to the causal-CSIT case using dynamic programming in [NC02]. A great deal of research has also focused on systems where the amount of CSIT is positive but strictly limited, the case we will refer to as *partial* CSIT. The paper [BSA02] considers a fixed-rate system and deals with power control to minimize the outage probability based on partial CSIT. On the other hand, [MSEA03] focuses on quantizing the direction of the beamforming vector of a multiple-input single-output system without performing power control. Some specific adaptive digital modulation and coding schemes are studied in [GC97, GC98, VG03, LF00, LYS03, GØH05]. Intelligent use of imperfect feedback information is also shown to improve various performance measures of multiple-antenna systems in e.g., [NLTW98, VM01, JSO02, JS04, LHS03].

While outage probability is a valid measure of the performance of a fixed-rate system over slowly fading channels, for certain applications it may be more reasonable to consider the achievable *expected rate* over multiple fading blocks [Cov72, BPS98]. Expected rate can also be seen as a measure of *reliably decodable rate*, from the receiver's perspective [EG98]. With perfect CSIR the receiver knows whether the transmission of the present block is in outage, and it can therefore disregard unreliable blocks. Hence, from the receiver's perspective, *loss* of data may occur while there will never be any transmission *errors.* Consequently, all codewords, at rates allocated by the transmitter, are either supported without errors or lost. It therefore makes sense to discuss expected rate in the sense of the average number of reliably received bits, per channel use over a large number of transmitted codewords. Interestingly, the traditional outage approach has been shown to be suboptimal in an expected-rate sense, as higher rates can be achieved using a *broadcast strategy* or *multi-layer coding* [Sha97, SS03, LLTF02]. This idea was first proposed in Cover's seminal work on broadcast channels [Cov72]. The multi-layer approach is particularly appealing for, e.g., successive refinement systems, which can produce a coarse version of a source such as an image, when some information is available and gradually improve the quality of the reproduced source as more information is received.

In this chapter we consider a slowly fading frequency non-selective link with perfect CSIR, and quantized channel feedback information. Our aim is to study the properties of adaptive systems optimal in an expected-rate sense based on some particular coding strategies, and with optimized quantizers in the feedback link. In contrast to a fixed-rate system considered in some previous related works on limited feedback, we study a *multi-layer variable-rate* coding scheme under different power constraints. We assume a fairly general framework, valid for any continuous channel distribution, and we explicitly formulate the feedback design problem. Essentially, quantized CSIT transforms the original channel, which can be viewed as a *composite* one [BPS98], into a finite number of parallel composite channels. By exploiting the inherent connection between our design and problems in parallel broadcast channels [Tse97], we develop simple iterative algorithms to optimize the feedback and the power allocation, as well as some strategy-dependent parameters. This is vastly different from feedback design for an ergodic channel [SJ03, LLC04b] where there is only a single centroid, namely the power, associated with each quantization region. Furthermore, unlike in the ergodic channel [GV97, LLC04b] where even perfect CSIT improves the ergodic capacity only marginally, our results show that coarsely quantized CSIT provides a significant improvement on the expected rate. With as few as one bit of feedback information, the role of multi-layer coding is shown to decrease dramatically. Finally, we develop the necessary and sufficient conditions for single-layer coding over a conditional channel to be optimal, irrespective of the number of code layers that the system can afford.

2.2 System Model

Consider the discrete-time complex-baseband model of a flat-fading single-input single-output communication system illustrated in Fig. 2.1, where the complex-valued channel gain is assumed to be random but constant during one fading block consisting of N channel uses. The received signal at time instant t within fading block m, $m = 1, 2, \ldots$, can be written as

$$y_m(t) = h_m s_m(t) + w_m(t),\ t = 1, \ldots, N, \tag{2.1}$$

where h_m denotes the channel gain and the $s_m(t)$'s are the transmitted symbols. The noise samples $w_m(t)$ are i.i.d. complex Gaussian with zero mean and unit variance. We assume that the h_m's are i.i.d. according to some distribution. Let $\gamma_m \triangleq |h_m|^2$, that is the resulting i.i.d. *channel powers*. In this chapter, we exclusively consider the case that any transmitted codeword spans only a *single* fading block. As pointed out in [CTB99, BCT01], it is reasonable to study the case

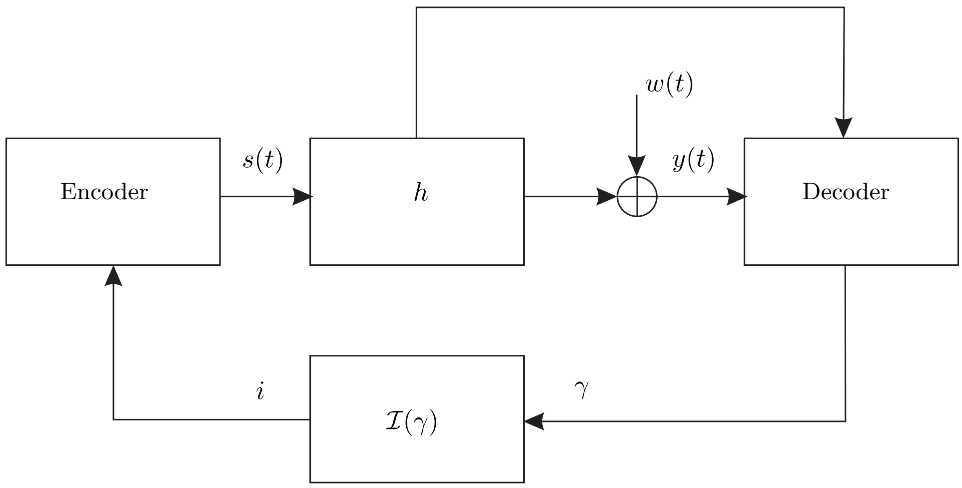

Figure 2.1: System model.

$N \to \infty$, modeling a scenario with very slow fading and a delay constraint on the transmitted codeword.

For brevity, the fading block index m will be omitted in the following discussion whenever this does not cause any confusion. With a slight abuse of notation, both the random variable representing the channel power and its realization will be denoted by γ. We assume that γ is a continuous random variable. Denote the cumulative distribution function and the probability density function of γ as $F(\gamma)$ and $f(\gamma)$, respectively. Furthermore, assume that $F(\gamma)$ and $f(\gamma)$ are continuous and $f(\gamma)$ takes on positive values over the entire region $(0, \infty)$.

The channel coefficient h is assumed to be known perfectly at the receiver. Given $\gamma = |h|^2$, the receiver employs a deterministic index mapping $\mathcal{I}(\gamma)$ that partitions the non-negative real line into K quantization regions

$$\mathcal{I}(\gamma) = i, \text{ if } \gamma \in [\gamma_i^{\mathrm{b}}, \gamma_{i+1}^{\mathrm{b}}),\ i = 0, \ldots, K-1, \tag{2.2}$$

where the γ_i^{b}'s denote the boundary points of the quantization regions. For convenience, we use the convention $\gamma_K^{\mathrm{b}} = \infty$ and $\gamma_0^{\mathrm{b}} = 0$. Herein K is a given positive integer, i.e., we consider a resolution-constrained quantizer. The index $i = \mathcal{I}(\gamma)$ is sent to the transmitter via a noiseless, zero-delay feedback channel. Conditioned on a feedback index i, any transmitted sequence $\{s(0), \ldots, s(N-1)\}$ is constrained to satisfy

$$\frac{1}{N} \sum_{t=0}^{N-1} |s(t)|^2 \leq \mathcal{P}(i) \tag{2.3}$$

where $\mathcal{P}(i)$ is a deterministic mapping from an integer index to power allocated.

Denote $P_i = \mathcal{P}(i)$, $i = 0, \ldots, K-1$. We then consider two different types of power constraint [CTB99]. The *short-term* power constraint requires that the power allocated cannot exceed P, independently of the feedback index, i.e.,

$$P_i \leq P, \, \forall i \in \{0, \ldots, K-1\}. \tag{2.4}$$

Under the more relaxed *long-term* power constraint, the transmitter can adapt the power based on the feedback index, such that the average power over multiple blocks does not exceed P,

$$\lim_{M\to\infty} \frac{1}{M} \sum_{m=1}^{M} \mathcal{P}(\mathcal{I}(\gamma_m)) = \mathrm{E}_\gamma[\mathcal{P}(\mathcal{I}(\gamma))] \leq P, \tag{2.5}$$

where the first equality holds with probability one. This is, in the scenario considered, equivalent to

$$\sum_{i=0}^{K-1} \left[F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma_i^{\mathrm{b}})\right] P_i \leq P. \tag{2.6}$$

Due to our assumption of non-zero and continuous density, the channel conditioned on a feedback index is still a composite one [BPS98], as in the case of no CSIT, however with a smaller support. That is, the range of uncertainty of the only partially known γ decreases with the feedback resolution. The coding scheme to maximize the expected rate over such a conditionally composite channel is still unknown in general. We focus our attention on two specific strategies: the traditional outage approach, which is also referred to as single-layer coding, and the broadcast strategy or multiple-layer coding. Clearly, single-layer coding is a special case of the more general multi-layer coding. We, however, consider this special case separately because the problem is more analytically tractable and therefore, more instructive. The main challenge is to optimize the index mapper $\mathcal{I}(\gamma)$, the allocated power $\mathcal{P}(i)$ and some strategy-dependent parameters jointly. In the discussion, we often refer to the set of all parameters to be designed as a *feedback scheme*.

2.3 Single-layer Coding

With the single-layer coding approach, given an index i, the transmitter selects a codeword from a rate-R_i capacity-achieving codebook where R_i, $i = 0, \ldots, K-1$ are design parameters. The system is in outage if the instantaneous mutual information of the channel is smaller than the operating rate R_i [OSW94].

It is convenient to review some results obtained under the assumption of perfect and no CSIT respectively, providing performance bounds to the quantized-CSIT system of interest. Without any CSIT, the transmitter selects an operating rate of $R_0 = \log(1 + \gamma_0 P)$ for some γ_0. (All logarithms in this chapter are natural, unless otherwise stated.) Since a codeword is successfully decoded only if $\gamma \geq \gamma_0$, maximizing the expected rate becomes

$$\max_{\gamma_0 \geq 0} [1 - F(\gamma_0)] \log(1 + \gamma_0 P). \tag{2.7}$$

Setting the first derivative to zero yields a necessary condition for a γ_0^* to be optimal, which we write in the following form for a later reference

$$1 = F(\gamma_0^*) + f(\gamma_0^*) \frac{1 + \gamma_0^* P}{P} \log(1 + \gamma_0^* P). \tag{2.8}$$

When perfect CSIT is available, assuming a short-term power constraint, it is clear that for a realization γ, the transmitter should match its rate to the current realization of the channel mutual information, i.e., $\log(1 + \gamma P)$, resulting in the following maximum expected rate

$$R^*_{\mathrm{S}-\infty} = \int_0^\infty \log(1 + \gamma P) f(\gamma) d\gamma, \tag{2.9}$$

Under the long-term power constraint, the maximum expected rate of the slowly fading channel is equal to the capacity of an ergodic channel with perfect channel side information [GV97],

$$R^*_{\mathrm{L}-\infty} = \int_{\lambda^*}^\infty \log\left(\frac{\gamma}{\lambda^*}\right) f(\gamma) d\gamma, \tag{2.10}$$

which is obtained by allocating power in a water-filling manner over multiple blocks. The water level $\frac{1}{\lambda^*}$ satisfies $\int_{\lambda^*}^\infty \left(\frac{1}{\lambda^*} - \frac{1}{\gamma}\right) f(\gamma) d\gamma = P$. Intuitively, the transmitter does not waste power on weak channel realizations and spends the saved power on strong channel realizations. The gain by optimally allocating power compared the short-term case is, however, insignificant for many frequently encountered channel distributions, especially at high SNR [GV97].

Feedback Design Under a Short-term Power Constraint

Let us now focus on the problem of limited feedback design under the short-term power constraint. It is clear that with a short-term power constraint, the optimal power $P_i^* = P$, $\forall i$, as there is no cost incurred with increasing the power allocated to each fading block up to the upper limit. Given an index i, the transmitter chooses an operating rate R_i, which is associated with a *reconstruction point* γ_i via the relation $R_i = \log(1 + \gamma_i P)$. Since the transmitter knows that the channel can support at least a rate of $\log(1 + \gamma_i^{\mathrm{b}} P)$ but cannot support any rates larger than $\log(1 + \gamma_{i+1}^{\mathrm{b}} P)$, to maximize the expected rate it is necessary that $\gamma_i \in [\gamma_i^{\mathrm{b}}, \gamma_{i+1}^{\mathrm{b}})$. If the actual channel power $\gamma \geq \gamma_i$, the codeword will be successfully decoded. On the other hand, if $\gamma_i^{\mathrm{b}} \leq \gamma < \gamma_i$, the system is in outage. Designing a feedback scheme optimal in the sense of expected rate is, therefore, equivalent to solving the following optimization problem

$$\begin{aligned} \max_{\{\gamma_i,\,\gamma_i^{\mathrm{b}}\}} \quad & \sum_{i=0}^{K-1} \left[F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma_i)\right] \log(1+\gamma_i P) \\ \text{s.t.} \quad & \gamma_{i+1}^{\mathrm{b}} - \gamma_i \geq 0,\; \gamma_i - \gamma_i^{\mathrm{b}} \geq 0. \end{aligned} \tag{2.11}$$

By a direct investigation of the activeness of the linear constraints, we can simplify the Karush-Kuhn-Tucker (KKT) conditions for a scheme $\{\gamma_i^*,\, \gamma_i^{\mathrm{b}*}\}$ to be optimal to:

$$\gamma_i^{\mathrm{b}*} = \gamma_i^*, \qquad i = 1, \ldots, K-1 \tag{2.12a}$$

$$F(\gamma_{i+1}^*) = F(\gamma_i^*) + f(\gamma_i^*)\frac{1+\gamma_i^* P}{P} \log \frac{1+\gamma_i^* P}{1+\gamma_{i-1}^* P}, i = 0, \ldots, K-1 \tag{2.12b}$$

with the convention $\gamma_{-1}^* = 0$, $\gamma_K^* = \infty$. In the special case of $K = 1$, i.e., the no-CSIT case, the necessary condition reduces to (2.8). The intuition behind (2.12a) can be explained as follows. Given a fixed set $\{\gamma_i^*\}$, the system is in outage $\forall \gamma \in [\gamma_i^{\mathrm{b}}, \gamma_i^*)$. By increasing γ_i^{b} up to γ_i^*, we effectively replace the expected rate of zero in the outage region with $R_{i-1}^* = \log(1 + \gamma_{i-1}^* P) > 0$ and hence, strictly increase the objective function of (2.11). A direct consequence of (2.12a) is that an outage, defined as the event that the channel cannot support the operating rate, can only occur if the zero index ($i = 0$) is received at the transmitter.

It is relatively simple to solve for $\{\gamma_i^*\}$ from (2.12b) since one can express $\gamma_1^*, \ldots, \gamma_{K-1}^*$, as a function of γ_0^*. (Recall that $F(\gamma)$ is invertible due to our assumption of non-zero density.) Therefore, (2.12b) with $i = K - 1$ can be expressed as an equation with a single unknown γ_0^*, which can be solved numerically. Given a γ_0^* (the solution may not be unique), one can successively compute $\gamma_1^*, \ldots, \gamma_{K-1}^*$ (in that order) using (2.12b). We observe that for many common channel distributions, (2.12b) appears to have a unique solution. Furthermore, our experiments show that solving (2.12b) directly is much more efficient than using standard optimization methods, which generally require a large number of initial random seeds due to the non-concavity of (2.11).

Feedback Design Under a Long-term Power Constraint

Under the more relaxed long-term power constraint, given an index i, the transmitter selects a codeword from a codebook of rate $R_i = \log(1 + \gamma_i P_i)$. In this case, the operating rate R_i depends not only on the reconstruction point γ_i but also on the power allocated P_i. The feedback design problem can thus be formulated as

$$\begin{aligned} \max_{\{\gamma_i^{\mathrm{b}},\,\gamma_i,\,P_i\}} \quad & \sum_{i=0}^{K-1} \left[F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma_i)\right] \log(1+\gamma_i P_i) \\ \text{s.t.} \quad & P - \sum_{i=0}^{K-1} \left[F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma_i^{\mathrm{b}})\right] P_i \geq 0, \\ & P_i \geq 0,\ \gamma_{i+1}^{\mathrm{b}} - \gamma_i \geq 0,\ \gamma_i - \gamma_i^{\mathrm{b}} \geq 0, \end{aligned} \tag{2.13}$$

which is more challenging than (2.11) due to the non-linear power constraint. Throughout this section, we assume that a *constraint qualification* holds at the maximizers of (2.13), so that the KKT conditions are necessary-optimality conditions [FGW02, BV04]. Let us introduce the Lagrange multiplier $\lambda \geq 0$ associated with the power constraint.

Let $\{\gamma_i^*, \gamma_i^{\mathrm{b}*}, P_i^*\}_{i=0}^{K-1}$ be an optimal scheme and λ^* be the corresponding optimal Lagrange multiplier. Note that for any $i \geq 1$, $\gamma_i^{\mathrm{b}*}$ solves the linearly-constrained maximization problem

$$\begin{aligned} \max_{x\in[\gamma_{i-1}^*,\gamma_i^*]} & \left[F(x) - F(\gamma_{i-1}^*)\right] \log(1+\gamma_{i-1}^* P_{i-1}^*) \\ & - \lambda^* \left([F(x) - F(\gamma_{i-1}^{\mathrm{b}})]P_{i-1}^* + [F(\gamma_{i+1}^{\mathrm{b}*}) - F(x)]P_i^*\right). \end{aligned}$$

The sign of the first derivative of the objective function does not depend on x since by assumption, $f(x) > 0$, $\forall x > 0$. Therefore, we either have $\gamma_i^{\mathrm{b}*} = \gamma_{i-1}^*$ or $\gamma_i^{\mathrm{b}*} = \gamma_i^*$. But if $\gamma_i^{\mathrm{b}*} = \gamma_{i-1}^*$, the region $[\gamma_{i-1}^{\mathrm{b}*}, \gamma_i^{\mathrm{b}*})$ contributes nothing to the expected rate and neither does the outage region $[\gamma_i^{\mathrm{b}*}, \gamma_i^*)$. In this case, we can merge those two regions forming a new scheme with $\gamma_{i-1}^{\mathrm{b}\star} = \gamma_{i-1}^\star = \gamma_{i-1}^*$, $\gamma_i^{\mathrm{b}\star} = \gamma_i^\star = \gamma_i^*$ that achieves the same expected rate without violating the power constraint. This means that we can consider $\gamma_i^{\mathrm{b}} = \gamma_i$, $\forall i \geq 1$ without loss of optimality and focus on the following dual problem

$$\begin{aligned} \min_{\lambda} \max_{\{\gamma_i,\,P_i\}} & \sum_{i=0}^{K-1} \left[F(\gamma_{i+1}) - F(\gamma_i)\right] \log(1+\gamma_i P_i) \\ & - \lambda \left(F(\gamma_1)P_0 + \sum_{i=1}^{K-1} \left[F(\gamma_{i+1}) - F(\gamma_i)\right] P_i \right) \end{aligned} \tag{2.14}$$

A simple iterative, Lloyd-like [GG92] algorithm can be developed to obtain a sequence of feedback schemes *and* dual variable $\{\gamma_i^{(k)}, P_i^{(k)}, \lambda^{(k)}\}$. For simplicity, we will omit the iteration index k whenever this does not cause any confusion. Given a set $\{\gamma_i\}$ so that $\gamma_{K-1} > \cdots > \gamma_0$, solving the dual problem (2.14) is equivalent to allocating power over a set of *parallel* scalar additive white Gaussian noise (AWGN) channels to maximize a linear combination of the achievable rates. The solution is readily obtained by the following water-filling algorithm

$$P_0 = \left(\frac{F(\gamma_1) - F(\gamma_0)}{F(\gamma_1)} \frac{1}{\lambda} - \frac{1}{\gamma_0} \right)^+ \tag{2.15a}$$

$$P_i = \left(\frac{1}{\lambda} - \frac{1}{\gamma_i} \right)^+ \qquad i = 1, \ldots, K-1, \tag{2.15b}$$

where $(x)^+ \triangleq \max(x, 0)$ and λ is chosen such that the power constraint is active. Since $\gamma_{K-1} > \cdots > \gamma_0$, there exists an i_0 such that $P_i > 0$, $\forall i \geq i_0$ and $P_i = 0$, $\forall i < i_0$. In the next step, we fix λ and $\{P_i\}$ and solve (2.14) for $\{\gamma_i\}$. Setting the first partial derivatives to zero and simplifying leads to

$$F(\gamma_{i+1}) = F(\gamma_i) + f(\gamma_i) \frac{1 + \gamma_i P_i}{P_i} \left[\log \frac{1 + \gamma_i P_i}{1 + \gamma_{i-1} P_{i-1}} + \lambda (P_i - P_{i-1}) \right], \; i \geq i_0 \tag{2.16}$$

which can be solved with the same technique used to solve (2.12b). For $i < i_0$, we can choose some values arbitrarily so that $\gamma_i > \gamma_{i-1}$, $\forall i$. The two basic steps described above are iterated until convergence. The procedure is summarized in Algorithm 1. A natural termination condition for the algorithm is to check whether

$$\frac{\bar{R}^{(k+1)} - \bar{R}^{(k)}}{\bar{R}^{(k+1)}} \leq \epsilon$$

where $\bar{R}^{(k)}$ is the value of the objective function in (2.14) evaluated at $\{\gamma_i^{(k)}, P_i^{(k)}\}$ and $\lambda^{(k)}$, and ϵ is a small positive number. Selecting the initial values is also an important issue. It is clear from (2.16) that by choosing $P_i^{(0)} = P$, $\forall i$, we can arbitrarily select $\lambda^{(0)}$. Moreover, with this particular choice of initial values, (2.16) reduces to (2.12b) and the first step in the algorithm is to solve the short-term power constraint problem. This prevents the algorithm from converging to a local optimum that is smaller than that obtained under a short-term power constraint. While we do not claim global optimality of the solution due to the non-concavity of (2.13), numerical results indicate that among a large number of random initial values, this judicious choice always yields the highest expected rate.

It is also possible to introduce a peak power constraint that limits the power allocated to any fading block. Such a constraint arises frequently in practice due to e.g., hardware limitations. In such a case, the design problem can be stated as follows: Solve (2.13) with the additional constraints $P_i \leq P^{\mathrm{m}}$, $i = 0, \ldots, K-1$ where $P^{\mathrm{m}} > P$. (If $P^{\mathrm{m}} \leq P$, we return to the short-term power constraint case.) By introducing additional Lagrange multipliers, it can be shown that the optimal power allocation scheme given $\{\gamma_i^{(k)}\}$ is a modified version of (2.15a)-(2.15b) where we redefine $(x)^+ \triangleq \max(\min(x, P^{\mathrm{m}}), 0)$, i.e., water-filling up to P^{m} is performed. The partitioning step (2.16) still applies in this case.

Algorithm 1: Single-layer Coding with a Long-term Power Constraint

---

Initialize $k = 0$, $P_i^{(0)} = P$, $\forall i$, arbitrary $\lambda^{(0)}$;
repeat
  Fix $\{P_i^{(k)}\}$ and $\lambda^{(k)}$, solve for $\{\gamma_i^{(k)}\}$ using (2.16);
  Fix $\{\gamma_i^{(k)}\}$, determine $\{P_i^{(k+1)}\}$ and $\lambda^{(k+1)}$ by the water-filling algorithm (2.15a), (2.15b) ;
  $k \longleftarrow k + 1$;
until *Convergence* ;

2.4 Multiple-layer Coding

A higher expected rate over a composite channel can be achieved by means of superposition coding. This approach exploits the degradedness of scalar AWGN broadcast channels. The transmitter sends the superposition of L codewords taken from L different codebooks, hence the term multi-layer coding. The receiver employs a successive decoder and the amount of data that can be successfully decoded depends on the actual realization of the channel. The no-CSIT case has been considered in [Sha97, SS03, LLTF02]. Notice that conditioned on perfect CSIT, the transmitter no longer sees a composite channel, thus the single-layer results (2.9) and (2.10) apply.

Feedback Design Under a Short-term Power Constraint

Assume that L-layer coding is employed over each quantization region. Given a feedback index i, the transmitter sends the superposition of L codewords taken from L capacity-achieving codebooks. The rate of codebook j, $j = 0, \ldots, L-1$, is designed so that

$$R_{ij} = \log\left(1 + \frac{\gamma_{ij} P_{ij}}{1 + \gamma_{ij}\sum_{k=j+1}^{L-1} P_{ik}}\right),$$

where γ_{ij}'s are referred to as the *reconstruction points*, and P_{ij} is the power constraint of codebook j. Without loss of generality, assume that $\gamma_{ij} < \gamma_{i(j+1)}$, $\forall i, j$.

Note that the mapping from feedback index to reconstruction points $\{\gamma_{ij}\}$ is one-to-many, meaning that each code layer utilizes the index differently. Herein γ_{ij}'s can be interpreted as the channel powers of L users among infinitely many users of an imaginative broadcast channel that the transmitters chooses to communicate with, and P_{ij}'s can be seen as the power allocated to these users [LLTF02, Sha97, SS03].

The receiver performs successive decoding, i.e., it decodes layer j treating all other layers k, $k > j$ as AWGN. This is possible due to the fact that the distribution of the codewords of a capacity-achieving codebook can be considered to be Gaussian, in the limit $N \to \infty$ [Cov72, SV97]. Due to the degradedness of the scalar AWGN broadcast channel [CT91] and by definition of R_{ij}, code layer j is successfully decoded and subtracted from the received signals if and only if $\gamma > \gamma_{ij}$. The short-term power problem is, therefore, explicitly formulated as

$$\begin{aligned} \max_{\{\gamma_i^{\mathrm{b}},\, \gamma_{ij},\, P_{ij}\}} \quad & \sum_{i=0}^{K-1} \sum_{j=0}^{L-1} \left[F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma_{ij})\right] \log\left(1 + \frac{\gamma_{ij} P_{ij}}{1 + \gamma_{ij} \sum_{k=j+1}^{L-1} P_{ik}}\right) \\ \text{s.t.} \quad & P \geq \sum_{j=0}^{L-1} P_{ij},\ P_{ij} \geq 0,\ \gamma_{ij} \geq \gamma_{i(j-1)},\ \gamma_{i+1}^{\mathrm{b}} \geq \gamma_{i(L-1)},\ \gamma_{i0} \geq \gamma_i^{\mathrm{b}}. \end{aligned} \tag{2.17}$$

Similar to the single-layer coding case, it is necessary that $\gamma_i^{\mathrm{b}*} = \gamma_{i0}^*$, $\forall i > 0$ for a scheme to be optimal. However, unlike the single-layer case, the necessary conditions in general cannot be solved directly. Herein we focus on a low-complexity iterative algorithm that successfully exploits results in parallel broadcast channels [Tse97]. We first fix the boundaries of the quantization regions $\{\gamma_{i0}^{(k)}\}$ and the power levels $\{P_{ij}^{(k)}\}$ to find the optimal $\{\gamma_{ij}^{(k)}\}$, $j > 0$. Next $\{\gamma_{ij}^{(k)}\}$ are fixed to find the optimal power levels. Finally, the newly obtained set of powers and reconstruction points is fixed to find new boundaries $\{\gamma_{i0}^{(k+1)}\}$. The procedure is summarized in Algorithm 2 and the details are presented in the following. For clarity, from now on we omit the iteration indices.

First, given $\{\gamma_{i0}\}$ and $\{P_{ij}\}$, the joint optimization in (2.17) over $\{\gamma_{ij}\}$, $j > 0$, decouples into the optimization of each individual γ_{ij}, which can be done efficiently using numerical methods [GMW81]. Next, given $\{\gamma_{ij}\}$, the optimal P_{ij}'s can be found separately for each quantization region. Over the quantization region i, the optimization problem is equivalent to allocating a total power of P to maximize a linear combination of the achievable rates of an L-user scalar Gaussian broadcast channel, where user j has channel power γ_{ij} and *rate reward* $\left[F(\gamma_{(i+1)0}) - F(\gamma_{ij})\right]$. This can be solved by a simple algorithm [Tse97]. In particular, consider the function

$$J(z) = \arg\max_j \frac{F(\gamma_{(i+1)0}) - F(\gamma_{ij})}{\frac{1}{\gamma_{ij}} + z} - \lambda_i \tag{2.18}$$

where $z \in [0, P]$ and

$$\lambda_i = \max_j \frac{F(\gamma_{(i+1)0}) - F(\gamma_{ij})}{\frac{1}{\gamma_{ij}} + P} \tag{2.19}$$

For any j, the set of all $z \in [0, P]$ such that $J(z) = j$ is shown to be either empty or a single interval [Tse97]. The length of such an interval is equal to the optimal power allocated to the code layer associated with γ_{ij}.

Finally, we need to find optimal $\{\gamma_{i0}\}$ given $\{\gamma_{ij}\}$, $j > 0$ and $\{P_{ij}\}$. The necessary conditions can be simplified to

$$F(\gamma_{(i+1)0}) = F(\gamma_{i0}) + f(\gamma_{i0}) \frac{(1 + \gamma_{i0} P)(1 + \gamma_{i0}(P - P_{i0}))}{P_{i0}} (R_{i0} - R_{i-1}) \tag{2.20}$$

where $R_i \triangleq \sum_{j=0}^{L-1} R_{ij}$. Clearly, (2.12b) is a special case of (2.20) where $P_{i0} = P$, $\forall i$. Again, we solve (2.20) with standard non-derivative numerical techniques.

Algorithm 2: Multi-layer Coding with a Short-term Power Constraint

Initialize $k = 0$, $\{\gamma_{i0}^{(0)}\}$, $\{P_{ij}^{(0)}\}$ s.t. $\sum_j P_{ij}^{(0)} = P$;
repeat
  Fix $\{P_{ij}^{(k)}\}$, $\{\gamma_{i0}^{(k)}\}$, find optimal $\{\gamma_{ij}^{(k)}\}$, $\forall j > 0$;
  Fix $\{\gamma_{ij}^{(k)}\}$, $\forall j$, find optimal $\{P_{ij}^{(k+1)}\}$ using (2.18);
  Fix $\{P_{ij}^{(k+1)}\}$, $\forall j$ and $\{\gamma_{ij}^{(k)}\}$, $\forall j > 0$, find optimal $\{\gamma_{i0}^{(k+1)}\}$ using (2.20);
  $k \longleftarrow k + 1$;
until *Convergence* ;

Feedback Design Under a Long-term Power Constraint

In this section, we consider the most general case, when multi-layer coding is employed and temporal power control is also possible. The design problem has the following form

$$\begin{aligned}
\max_{\{\gamma_i^{\mathrm{b}},\, \gamma_{ij},\, P_{ij}\}} \quad & \sum_{i=0}^{K-1} \sum_{j=0}^{L-1} \left[F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma_{ij})\right] \log\left(1 + \frac{\gamma_{ij} P_{ij}}{1 + \gamma_{ij} \sum_{k=j+1}^{L-1} P_{ik}}\right) \\
\text{s.t.} \quad & P \geq \sum_{i=0}^{K-1} \left[F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma_i^{\mathrm{b}})\right] \sum_{j=0}^{L-1} P_{ij}, \\
& P_{ij} \geq 0,\ \gamma_{ij} \geq \gamma_{i(j-1)},\ \gamma_{i+1}^{\mathrm{b}} \geq \gamma_{i(L-1)},\ \gamma_{i0} \geq \gamma_i^{\mathrm{b}}.
\end{aligned} \tag{2.21}$$

Assuming a constraint qualification at the optimal point, we can extend the algorithm in Section 2.4 to take into account the Lagrange multiplier λ associated with the power constraint. As in Section 2.3, we consider only $\gamma_i^{\mathrm{b}} = \gamma_{i0}$, $\forall i > 0$. Given $\{\gamma_{i0}\}$, $\{P_{ij}\}$, the set $\{\gamma_{ij}\}$ (for $j > 0$) can be found similarly to the short-term power constraint case. The solution does not depend on λ.

Next, for a fixed $\{\gamma_{ij}\}$, the optimal power allocation and corresponding Lagrange multiplier λ can be found with a greedy algorithm for parallel scalar AWGN broadcast channels [Tse97]. In the currently investigated scenario, we first need to find the optimal power allocated to each quantization region

$$P_i = \left(\max_j \left(\frac{F(\gamma_{(i+1)0}) - F(\gamma_{ij})}{F(\gamma_{(i+1)0}) - F(\gamma_{i0})} \frac{1}{\lambda} - \frac{1}{\gamma_{ij}}\right)\right)^+ \tag{2.22}$$

where λ is chosen such that the power constraint is active. We can then apply (2.18), where λ_i, $i = 0, \ldots, K-1$, is replaced by λ, to find the individual P_{ij}'s. (The total power assigned to each conditional channel is no longer P.) Note that a peak power constraint can be directly incorporated into (2.22). In this case, each quantization region is allocated power up to some P^{m}.

Finally, given the set $\{\gamma_{ij}\}$ for $j > 0$, $\{P_{ij}\}$ and λ, the optimal $\{\gamma_{i0}\}$ can be found by solving

$$\begin{aligned} F(\gamma_{(i+1)0}) = F(\gamma_{i0}) + f(\gamma_{i0}) & \frac{(1+\gamma_{i0}P_i)(1+\gamma_{i0}(P_i - P_{i0}))}{P_{i0}} \\ & \cdot \left[R_{i0} - R_{i-1} + \lambda(P_i - P_{i-1})\right], \end{aligned} \tag{2.23}$$

which is a generalization of (2.16). The entire procedure is outlined in Algorithm 3. For the same reasons as in the single-layer coding case, the initial power levels are chosen so that $\sum_j P_{ij}^{(0)} = P$, $\forall i$.

Algorithm 3: Multi-layer Coding with a Long-term Power Constraint

Initialize $k = 0$, $\{\gamma_{i0}^{(0)}\}$, $\{P_{ij}^{(0)}\}$ s.t. $\sum_j P_{ij}^{(0)} = P \ \forall i$, arbitrary $\lambda^{(0)}$;
repeat
  Fix $\{P_{ij}^{(k)}\}$, $\{\gamma_{i0}^{(k)}\}$, $\lambda^{(k)}$, find optimal $\{\gamma_{ij}^{(k)}\}$, $j > 0$;
  Fix $\{\gamma_{ij}^{(k)}\}$, $\forall j$, find optimal $\{P_{ij}^{(k+1)}\}$, $\lambda^{(k+1)}$ using (2.18), (2.22);
  Fix $\lambda^{(k+1)}$, $\{P_{ij}^{(k+1)}\}$, $\forall j$ and $\{\gamma_{ij}^{(k)}\}$, $j > 0$, find optimal $\{\gamma_{i0}^{(k+1)}\}$ using (2.23);
  $k \longleftarrow k + 1$;
until *Convergence* ;

Single-layer Optimality Conditions

The special structure of the feedback problem allows us to obtain some interesting results. Clearly, it is not necessary that all the coding levels are assigned non-zero power. The following proposition states the necessary and sufficient condition for *single-layer* coding to be optimal, independent of the number of code layers L that the system can afford. More interestingly, we show that if the single-layer optimality condition is satisfied, then the single reconstruction point must be the left boundary of the quantization region, i.e., the conditionally worst-case channel realization.

Proposition 2.1. *Consider a quantization region $\left[\gamma_i^{\mathrm{b}}, \gamma_{i+1}^{\mathrm{b}}\right)$. Suppose that L-layer coding is employed. For any $L \geq 2$, allocating all the available power to a single reconstruction point is optimal in an expected-rate sense if and only if*

$$\gamma_i^{\mathrm{b}} \in \arg \max_{\gamma \in \left[\gamma_i^{\mathrm{b}}, \gamma_{i+1}^{\mathrm{b}}\right)} \left[F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma)\right] \gamma. \tag{2.24}$$

If (2.24) holds, the optimal reconstruction point that receives all the available power is $\gamma_{i0}^* = \gamma_i^{\mathrm{b}}$.

Proof. See Appendix A. □

Intuitively, (2.24) states that the layer corresponding to γ_i^{b} is the first layer to be allocated power. However, γ_i^{b} also corresponds to the weakest layer and according to [Tse97], this must also be the last layer to be allocated power. But the power allocated to each layer corresponds to a single interval, implying that the layer γ_i^{b} receives *all* the available power. This is indeed single-layer coding. The condition (2.24) is rather interesting because it only depends on the nature of the channel distribution. As an example, consider a Rayleigh channel with a mean channel power of $\bar{\gamma}$, i.e., $F(\gamma) = 1 - \exp(-\frac{\gamma}{\bar{\gamma}})$. One can verify that for any quantization region such that $\gamma_i^{\mathrm{b}} \geq \bar{\gamma}$, (2.24) is satisfied independently of the upper boundary $\gamma_{i+1}^{\mathrm{b}}$ of the region considered. Hence, upon receiving a feedback index corresponding to such a region, the transmitter can employ the traditional single-layer coding without any loss in expected rate.

The condition (2.24) is particularly useful in our proposed iterative procedures. Since the boundaries are known a priori in the iteration steps, the region where single-layer coding is optimal can be quickly determined. The feedback design, however, requires joint optimization of the boundaries and other parameters. Finding stronger conditions that hold even if the boundaries are not known a priori remains an interesting open problem.

2.5 Numerical Results

In Fig. 2.2, we plot the expected rate achieved by several feedback schemes with different numbers of quantization regions and different numbers of code layers over a Rayleigh channel with unit mean power, i.e., $F(\gamma) = 1 - \exp(-\gamma)$. A short-term power constraint is assumed. The average signal-to-noise ratio is defined as $\mathrm{SNR} \triangleq P/\sigma_{\mathrm{n}}^2 = P$ since the noise variance σ_{n}^2 is assumed to be unit. Significant gains can be observed even with coarsely quantized systems. For example, to achieve a target expected rate of 2 nats per channel use, feedback schemes with 2 and 4 quantization regions require a power of roughly 3 and 5 dB less than a no-CSIT system does, respectively. Most of the gain of multi-layer coding is observed in the high-SNR regime. Furthermore, the benefit of multi-layer coding appears to be more pronounced as the SNR increases. However, as the quality of partial CSIT improves, the role of multi-layer coding reduces substantially. For instance, in a system with $K = 4$ quantization regions, there is practically no benefit of using 2-layer coding over single-layer coding for any SNR smaller than 30 dB. This can be attributed to at least two factors. Firstly, as the feedback resolution increases, the support of a conditional channel becomes smaller, hence the "users" of the virtual broadcast channel experience almost the same channel power. There is therefore little benefit in optimally allocating power among those users. Secondly, we experimentally observed that more quantization regions satisfy the single-layer optimality condition (2.24) as the feedback resolution increases.

The difference between the expected rates achieved by different schemes and that achieved by a single-layer coding system under a short-term power constraint

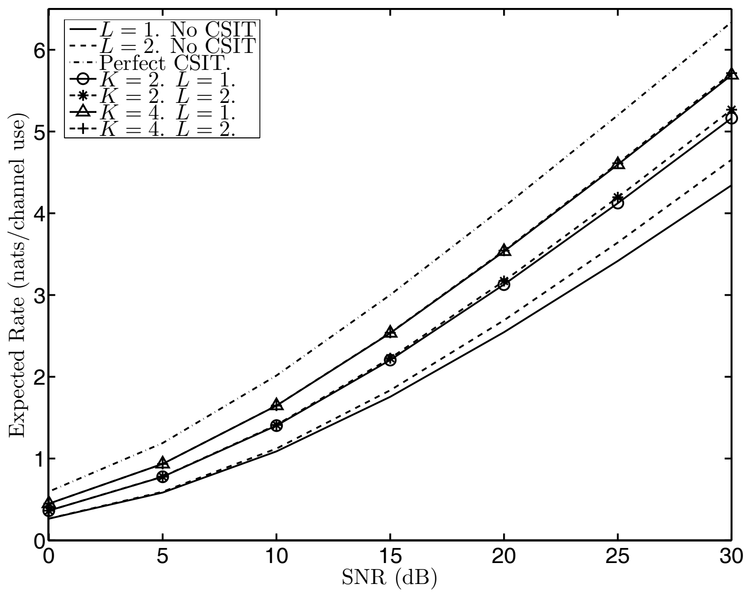

Figure 2.2: Expected rate achieved with different feedback schemes over a Rayleigh channel. A short-term power constraint is assumed.

is plotted in Fig. 2.3. We will refer to this difference as the *absolute gain*. As can be seen, under a short-term power constraint, multiple-layer coding provides an *increasing* absolute gain as the SNR increases. On the other hand, long-term power control (with single-layer coding) yields a *decreasing* absolute gain as the SNR increases. Thus combining multi-layer coding and long-term power control results in an interesting effect: The absolute gain appears to be *minimum* at some intermediate SNR. Another observation is that the most significant absolute gain is obtained when the number of code layers increases from $L = 1$ to $L = 2$, a behavior that can also be observed in systems without feedback [LLTF02].

To emphasize the promising role of long-term (temporal) power control in systems with very limited power, we plot the expected rate in the low-SNR region in Fig. 2.4. It should be noted that the SNR range depicted may not be relevant for some wireless communication systems. Since multi-layer coding only provides a negligible improvement, we only plot the expected rate achieved by single-layer coding. Interestingly, a system controlling transmit power with coarsely quantized CSIT, namely $K = 2$ or 1-bit feedback, outperforms a short-term power constrained system with *perfect* CSIT for any SNR smaller than -5 dB. Although a long-term power constraint is clearly more relaxed, the big difference between the quality of

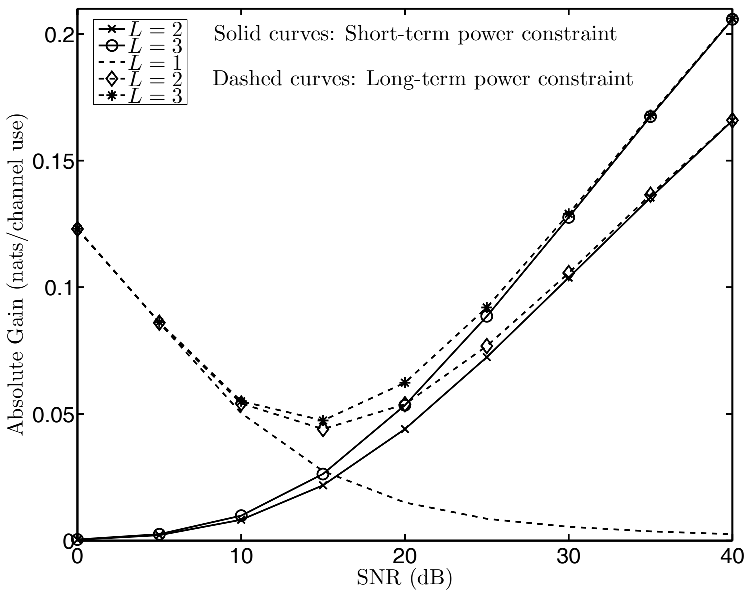

Figure 2.3: Absolute gains (i.e., the gains compared to a single-layer, short-term power constrained system) over a Rayleigh channel, $F(\gamma) = 1 - \exp(-\gamma)$. The feedback resolution $K = 2$. Solid and dashed curves correspond to a short-term and a long-term power constraint, respectively.

CSIT (1 bit vs. perfect) makes the comparison sensible. Consequently, the effects of imposing a peak power constraint on the system is also most pronounced at this region, as depicted in Fig. 2.5. As can be seen, when $K = 2$, a peak power that is 3 dB higher than the average one severely affects the expected rate. Increasing the number of quantization regions appears to reduce this effect significantly. For any K, an optimal scheme tends to allocate power more evenly in the moderate and high SNR region, hence the effect of a peak power constraint diminishes as the SNR increases.

Similar behavior is also observed in some other commonly encountered channels. The performance of various feedback schemes over the equivalent channel of some single-input multiple-output (SIMO) systems using maximum ratio combining are plotted in Fig. 2.6. (Clearly, the SISO framework we consider can also be applied to such SIMO scenarios.)

Numerical results indicate that the proposed algorithms converge relatively fast. As an example, the convergence behavior of the iterative algorithms over a Rayleigh channel is plotted in Fig. 2.7. Note that the convergence speed of Algorithms 2 and

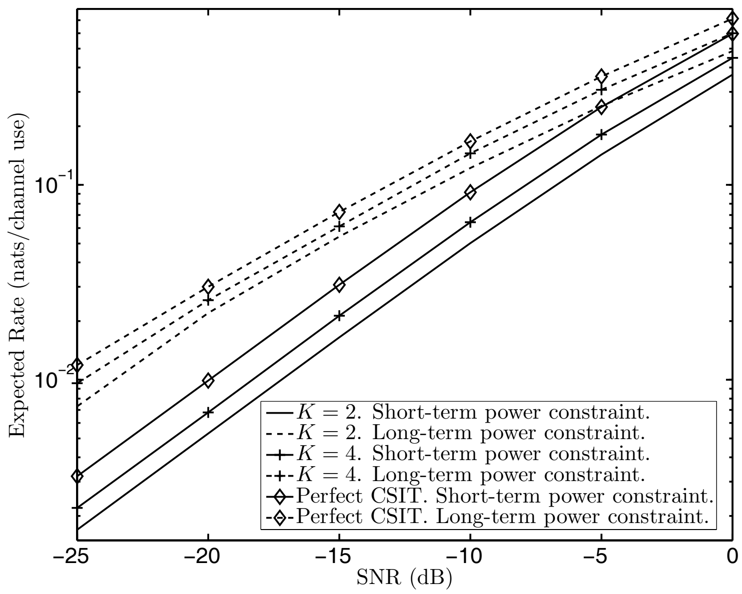

Figure 2.4: Expected rate achieved over the low-SNR region. Solid and dashed curves correspond to a short-term and a long-term power constraint, respectively. $F(\gamma) = 1 - \exp(-\gamma)$.

3 depends on the initial values $\{\gamma_{i0}^{(0)}\}$ and $\{P_{ij}^{(0)}\}$. To obtain a "good" starting point, we choose $P_{ij}^{(0)} = P/L$, $\forall i, j$ and take $\gamma_{i0}^{(0)}$ as the boundaries obtained after several iterations of Algorithm 1, which converges extremely fast. As shown in Fig. 2.7, this heuristic approach appears to be reasonably efficient. Experiments with random seeds also suggest that, at least for the channel distributions presented in this section, the algorithms lead to convergence to the global optimum. However we have not been able to prove that analytically.

2.6 Conclusion

We have studied a variable-rate system employing partial CSIT to increase the expected rate over a slowly fading channel. Our results indicate that a substantial portion of the gain with perfect CSIT can be achieved by heavily quantized CSIT. The improvement provided by the sophisticated multi-layer coding technique is shown to reduce dramatically if the resolution of the feedback quantizer increases. Moreover, in practice, multi-layer coding is also limited by other factors such as error propagation. This suggests that from an expected-rate perspective, single-

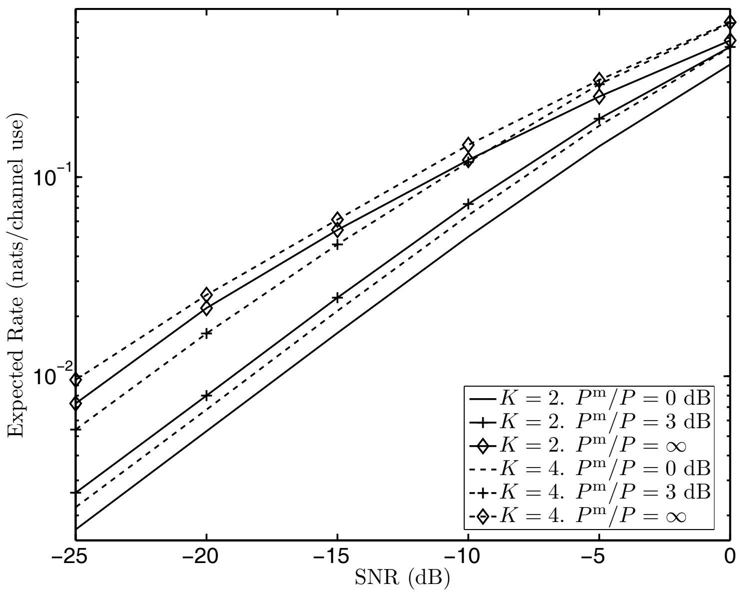

Figure 2.5: Effects of a peak power constraint on the expected rate over a Rayleigh channel. Solid and dashed curves correspond to $K = 2$ and $K = 4$, respectively.

layer coding may be a suitable choice for systems with feedback, even though some optimality is lost.

With the same number of quantization regions, optimally allocating power over time only improves the expected rate marginally at moderate to high SNR's. Hence, in this region, optimizing the feedback index mapping appears to be more important than optimizing the power allocation. On the other hand, in the low-SNR regions, most of the gain comes from temporal power control, which is based on partial CSIT. Therefore, the performance over the low-SNR region is highly affected if a peak power constraint is imposed on the system.

The design of multi-layer coding for MIMO channels is considerably more challenging because channel ordering over such systems is in general not uniquely defined. (A particular channel ordering based on majorization theory and a single-dimensional approximation is shown to be inefficient in [SS03].) Nevertheless, a SISO model still applies to a certain multiple-transmit antenna scenarios. For instance, the equivalent channel of a MIMO system employing orthogonal space-time codes is conveniently fitted into a SISO framework.

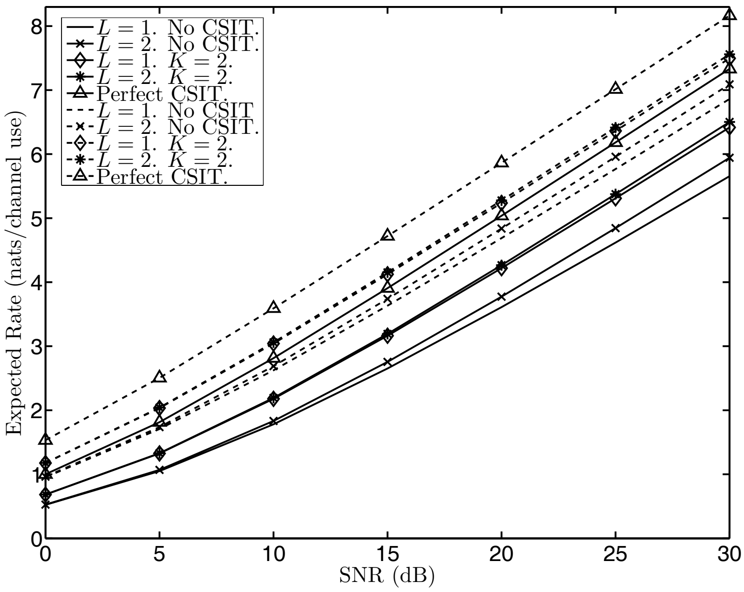

Figure 2.6: Expected rate achieved with different feedback schemes over SIMO channels. The channel coefficients are assumed to be i.i.d. zero-mean complex Gaussian with unit variance. Solid curves correspond to a 1×2 channel. Dashed curves correspond to a 1×4 channel. A short-term power constraint is assumed.

Appendix for Chapter 2

2.A Proof of Proposition 2.1

We first need the following lemma.

Lemma 2.1. *Consider a quantization region $\left[\gamma_i^{\mathrm{b}},\, \gamma_{i+1}^{\mathrm{b}}\right)$ and an arbitrary set of L reconstruction points $\{\gamma_{ij}\}_{j=0}^{L-1}$, where $\gamma_{i0} = \gamma_i^{\mathrm{b}}$. Suppose that (2.24) is satisfied. Then, the optimal power allocation in an expected-rate sense is $P_{i0}^{*} = P$, and $P_{ij}^{*} = 0,\ \forall j > 0$.*

Proof. (Lemma 2.1) Assume the contrary. Then, according to [Tse97], there exists a $\gamma \in (\gamma_i^{\mathrm{b}},\, \gamma_{i+1}^{\mathrm{b}})$ and a $z > 0$ such that

$$\frac{F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma_i^{\mathrm{b}})}{\frac{1}{\gamma_i^{\mathrm{b}}} + z} = \frac{F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma)}{\frac{1}{\gamma} + z}.$$

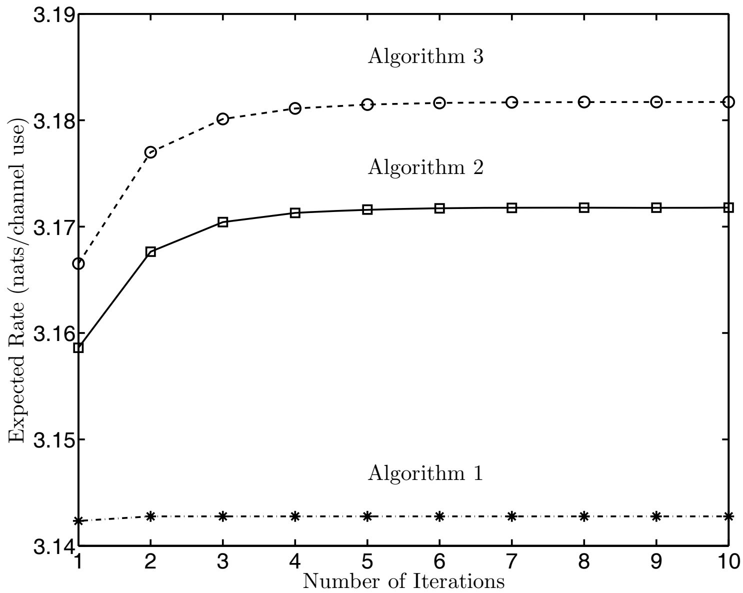

Figure 2.7: Convergence behavior of the proposed algorithms. $F(\gamma) = 1 - \exp(-\gamma)$. The number of quantization regions $K = 2$, SNR = 20 dB. The number of code layers for Algorithm 2 and 3 is $L = 2$.

However, expressing z as a function of γ leads to

$$z = \frac{\left[F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma)\right]\gamma - \left[F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma_i^{\mathrm{b}})\right]\gamma_i^{\mathrm{b}}}{F(\gamma) - F(\gamma_i^{\mathrm{b}})} \frac{1}{\gamma\gamma_i^{\mathrm{b}}} \leq 0$$

due to (2.24), which is a contradiction. □

Assume that (2.24) holds. We need to show that single-layer coding at γ_i^{b} is optimal. Let $\mathcal{S}_L^* = \{\gamma_{ij}^*\}_{j=0}^{L-1}$ be the optimal set of reconstruction points. If $\mathcal{S}_L^*$ contains γ_i^{b}, the asserted result immediately follows Lemma 2.1. Assume now that the optimal set of reconstruction points does not contain γ_i^{b}. Let R_L^* be the maximum expected rate achieved by allocating the power P over $\mathcal{S}_L^*$. Form a set of $L+1$ reconstruction points by adding γ_i^{b} to $\mathcal{S}_L^*$. Obviously, the maximum expected rate R_{L+1}^* achieved by allocating the same power P over the newly formed set can only be greater than or equal to R_L^*. But according to Lemma 2.1, R_{L+1}^* is achieved by allocating all available power to γ_i^{b}, implying that an arbitrary point other than γ_i^{b} can be removed from the newly formed $(L+1)$-point set without affecting the expected rate. Thus we have constructed a set of L construction points containing

γ_i^{b} that achieve $R_{L+1}^* \geq R_L^*$. This contradicts to the assumption that the optimal L-point set does not contain γ_i^{b}.

We complete the proof by showing that if allocating all available power to some reconstruction point $\tilde{\gamma} \in [\gamma_i^{\mathrm{b}}, \gamma_{i+1}^{\mathrm{b}})$ is optimal for any L, then (2.24) is satisfied. Assume the contrary, i.e., there exists an $\gamma^{\mathrm{m}} > \gamma_i^{\mathrm{b}}$ such that

$$\gamma^{\mathrm{m}} = \min\{\arg \max_{\gamma \in [\gamma_i^{\mathrm{b}}, \gamma_{i+1}^{\mathrm{b}})} \left[F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma)\right] \gamma\}. \tag{2.25}$$

If $\tilde{\gamma} \neq \gamma^{\mathrm{m}}$, consider optimal power allocation [Tse97] over a set of two reconstruction points $\{\tilde{\gamma}, \gamma^{\mathrm{m}}\}$. But (2.25) implies that γ^{m} is the first layer to receive power for any $P > 0$, which is a contradiction.

Now consider the case $\tilde{\gamma} = \gamma^{\mathrm{m}}$, i.e., single-layer coding at γ^{m} is optimal. Select a $\gamma \in [\gamma_i^{\mathrm{b}}, \gamma^{\mathrm{m}})$ and allocate power over $\{\gamma, \gamma^{\mathrm{m}}\}$. Let $\bar{P}(\gamma)$ be the solution to the following equation

$$\frac{F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma_i^{\mathrm{m}})}{\frac{1}{\gamma^{\mathrm{m}}} + \bar{P}(\gamma)} = \frac{F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma)}{\frac{1}{\gamma} + \bar{P}(\gamma)}.$$

Easily see that

$$\bar{P}(\gamma) = \frac{\left[F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma^{\mathrm{m}})\right] \gamma^{\mathrm{m}} - \left[F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma)\right] \gamma}{\gamma \gamma^{\mathrm{m}} \left[F(\gamma^{\mathrm{m}}) - F(\gamma)\right]} > 0. \tag{2.26}$$

This means that the power $\bar{P}(\gamma)$ is (optimally) allocated to γ^{m}, while all the remaining power $P - \bar{P}(\gamma)$ is allocated to γ. To show a contradiction, it suffices to prove that we can select γ in $[\gamma_i^{\mathrm{b}}, \gamma^{\mathrm{m}})$ so that $\bar{P}(\gamma)$ is arbitrarily close to zero, and thus $P - \bar{P}(\gamma)$ can be made positive for any $P > 0$.

To that end, note that the first derivative of the objective function in (2.25) must vanish at γ^{m}, i.e.,

$$F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma^{\mathrm{m}}) - \gamma^{\mathrm{m}} f(\gamma^{\mathrm{m}}) = 0. \tag{2.27}$$

Using (2.27) and l'Hospital's rule, we have

$$\lim_{\gamma \to \gamma^{\mathrm{m}}} \bar{P}(\gamma) = \lim_{\gamma \to \gamma^{\mathrm{m}}} \frac{-\left[F(\gamma_{i+1}^{\mathrm{b}}) - F(\gamma) - \gamma f(\gamma)\right]}{\gamma^{\mathrm{m}} \left[F(\gamma^{\mathrm{m}}) - F(\gamma) - \gamma f(\gamma)\right]} = 0.$$

(Recall that $f(\gamma)$ is assumed to be positive for all $\gamma > 0$.) This completes the proof. □

Part II

Diversity–Multiplexing Tradeoff

Chapter 3

D–M Tradeoff in MIMO Channels

In Chapter 2, we have seen that the outage probability and the transmission rate can be *combined* into a single performance measure over slow fading channels (i.e., the expected rate). In this chapter, we will take a closer look at the *two-dimensional tradeoff* between the outage probability and the transmission rate. In particular, we derive the diversity–multiplexing tradeoff over a MIMO channel with optimized resolution-constrained channel state feedback. We will introduce the concept of minimum guaranteed multiplexing gain in the forward link and show that this significantly influences the optimal D–M tradeoff. It is demonstrated that power control based on the feedback is instrumental in achieving the D–M tradeoff, and that rate adaptation is important in obtaining a high diversity gain even at high rates. A criterion to determine finite-length codes to be tradeoff optimal is presented, leading to a useful geometric characterization of the class of extended approximately universal codes. With codes from this class, the optimal D–M tradeoff is achievable by the combination of a feedback-dependent power controller and a single codebook for single-rate or two codebooks for adaptive-rate transmission. Finally, lower bounds to the optimal D–M tradeoffs based on Gaussian coding arguments are also studied. In contrast to the no-feedback case, these random coding bounds are only asymptotically tight, but can quickly approach the optimal tradeoff even with moderate codeword lengths.

3.1 Introduction

Communication over wireless MIMO channels has attracted great interest over the last decade because, in comparison to traditional single-input single-output systems, MIMO systems promise a better reliability through the use of many independent propagation paths and a higher throughput through the use of parallel spatial modes. While most works focus exclusively on one of these two gains, a new line of thought is introduced in [ZT03], where both types of gain are shown to be simultaneously achievable over a slow fading channel, and the fundamental tradeoff between these gains is elegantly characterized.

Attempts at a characterization of the D–M tradeoff over a multiple-antenna channel, under the assumption of *partial CSIT*, are reported in [KS04a, KS05] where the feedback information consists of the scalar quantized singular values of the channel matrix. This assumption is, however, restrictive and the results in [KS04a, KS05] only reflect the asymptotic outage behavior of a particular class of feedback schemes. An automatic retransmission request (ARQ) scheme for MIMO systems and its performance tradeoffs are studied in [ECD06]. Such an ARQ system can be viewed as a simple causal feedback scheme using one bit of feedback information per retransmission round. Combining ARQ with transmit power control [CTB99, BCT01, KS04b], yields a superior tradeoff compared to the no-CSIT case [ECD06].

Direct partial CSIT via resolution-constrained feedback is normally more useful than is the indirect channel knowledge provided by an ARQ flag. Furthermore, certain applications that are more sensitive to delay may exclude the possibility of retransmitting codewords in error. This motivates our present study of the optimal tradeoff between the reliability and rate that can be simultaneously achieved over a MIMO system with resolution-constrained quantized feedback. We advocate an approach that quantizes the information *needed* at the transmitter, and not the channel matrix itself.

We consider two related cases: Single-rate transmission and adaptive-rate transmission. For single-rate transmission, the information rate is kept constant and independent of the CSIT available, a good model for constant bit-rate services. For adaptive-rate systems, inspired by the notion of minimum rate [LLYS03, JG03], we propose the concept of *minimum multiplexing gain* to make "reliability" more meaningful in the limit of high SNR's. By performing joint power and rate control, and building upon the elegant framework of [ZT03], we derive the optimal D–M tradeoff in both cases in a recursive manner.

Since partial power control is a key ingredient in achieving the optimal performance over slow fading channels with feedback, it is not surprising that the characterization of the optimal D–M tradeoff presented in this chapter is similar to that obtained for the ARQ scheme with power control in [ECD06]. The results are, however, not equivalent. In particular, we demonstrate that introducing the concept of minimum multiplexing gain dramatically influences the optimal D–M tradeoff. Another difference distinguishing the current work from related works lies in the achievability part. More precisely, our achievability proof relies on the existence of a large class of codes, referred here to as *extended approximately universal codes*. This class is defined via a (nontrivial) extension of the concepts introduced in [TV06]. We present a simple and useful interpretation of these codes, giving some novel insight into approximate universality, as studied in [TV06]. Somewhat surprisingly, we show that if the power constraint imposed on every codeword is replaced by a more relaxed average power constraint over the whole codebook, then our extended approximately universal condition does not necessarily coincide with the well-known non-vanishing determinant criterion [YW03, BR03]. We do not attempt to design explicit codes, but instead show that a large class of existing codes can be used to achieve the present optimal D–M tradeoff, when combined with a carefully designed feedback link.

Our results also shed some light into the design of adaptive MIMO systems with sufficiently long codewords at very high SNR. For single-rate systems, the optimal D–M tradeoff can be achieved by a *single* codebook and a CSIT-dependent power controller. On the other hand, an adaptive MIMO system may achieve performance close to the optimal tradeoff with only *two* different codebooks, even if the resolution of the feedback link is higher.

Finally, to get additional insight, we develop lower bounds on the optimal D–M tradeoff based on random coding arguments. Interestingly, in contrary to the no-CSIT case where codes drawn from Gaussian ensemble of short length can achieve the entire optimal D–M tradeoff, we show that except for some special cases, Gaussian coding arguments do not appear to be sufficient to complete the achievability part when quantized feedback is available. This seems to hold even with carefully expurgated codes in combination with optimized feedback schemes. Our random coding bounds highlight that for a random code, the gap between the code rate and optimal feedback threshold is *not* negligible. Designing a good feedback scheme is therefore a critical task, even in the high SNR regime.

3.2 System Model

Consider the discrete-time complex-baseband model of a frequency-nonselective MIMO communication system with N_t transmit antennas and N_r receive antennas. The channel is constant during a fading block consisting of T channel uses, but changes independently from one block to the next. During a fading block l, the channel is represented by an $N_\mathrm{r} \times N_\mathrm{t}$ random matrix $\mathbf{H}_l$, and the received signal can be written in matrix form as

$$\mathbf{Y}_l = \mathbf{H}_l \mathbf{S}_l + \mathbf{W}_l. \tag{3.1}$$

The components of the temporally and spatially white noise matrix $\mathbf{W}_l$ of size $N_r \times T$ are i.i.d. complex Gaussian with zero mean and unit variance, $\mathcal{CN}(0,1)$. The elements of the channel matrix $\mathbf{H}_l$ are assumed to be i.i.d. $\mathcal{CN}(0,1)$. Thus we consider a rich-scattering, Rayleigh fading environment.

It is assumed that the receiver knows the channel matrix perfectly. Given a channel realization, the receiver sends back an index $\mathcal{I}(\mathbf{H}_l)$ via a noiseless, zero-delay feedback link to the transmitter. Herein $\mathcal{I}(\mathbf{H}_l)$ is a *deterministic* mapping from a channel matrix to an integer index. With a slight abuse of notation, we also denote the random variable representing the feedback index as $\mathcal{I}$, which takes values on the set $\{1, \ldots, K\}$ where K is a positive integer, i.e., a *resolution-constrained* feedback link is considered. In other words, the mapping $\mathcal{I}(\mathbf{H}_l)$ partitions the set of all possible channel realizations into K regions; and *priori to* the transmission the transmitter knows exactly which region the channel matrix belongs to.

A *codeword* $\mathbf{S}_l$ is assumed to span a *single* fading block. Since we do not consider coding over multiple fading blocks, the block index l will be omitted whenever this does not cause any confusion. Conditioned on an index $\mathcal{I} = i$, the codeword $\mathbf{S}$ is taken from a codebook $\mathcal{C}_i = \{\mathbf{S}_i(1), \ldots, \mathbf{S}_i(M_i)\}$ of rate R_i, where all codewords are equally likely. Herein the $\mathbf{S}_i(k)$'s are matrices of size $N_\mathrm{t} \times T$.

Define

$$P_i \triangleq \frac{1}{TM_i} \sum_{k=1}^{M_i} \|\mathbf{S}_i(k)\|_\mathrm{F}^2 \tag{3.2}$$

where $\|\mathbf{S}\|_\mathrm{F}$ denotes the Frobenius norm of matrix $\mathbf{S}$. Note that P_i can be interpreted as the average total transmit power conditioned on the event that the feedback index $\mathcal{I} = i$ is received. Since the noise has unit variance, we conveniently translate the average transmit power over multiple fading blocks to average signal-to-noise ratio (SNR) and impose a long-term power constraint [CTB99, BCT01]

$$\lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} \frac{1}{T} \|\mathbf{S}_l\|_\mathrm{F}^2 \overset{\text{a.s.}}{=} \mathrm{E}_\mathbf{H}[P_{\mathcal{I}(\mathbf{H})}] \le \mathrm{SNR} \tag{3.3}$$

where the first equality holds with probability one.

Note that the index mapping $\mathcal{I}(\mathbf{H})$, and the codebooks $\mathcal{C}_i$'s (thus P_i's and R_i's) are all SNR-dependent. In other words, we study a *sequence* of feedback schemes, one for each value of SNR (cf. [ZT03]). The dependence of the rates on the SNR is explicitly given by

$$R_i = r_i \log \mathrm{SNR}, \quad i = 1, \ldots, K,$$

where the r_i's are some real values in $(0, \min(N_\mathrm{r}, N_\mathrm{t}))$, independent of SNR. We refer to r_i's as the *individual multiplexing gains*, essentially quantifying how large the rate of each codebook is compared to the capacity of a SISO link at high SNR.

An *error* occurs when the transmitted codeword is incorrectly detected at the receiver. Let P_e be the average probability of error (over the randomness of the channel, the noise and the data). Then the system is said to have a *diversity gain* of d if

$$P_\mathrm{e} \doteq \mathrm{SNR}^{-d}, \tag{3.4}$$

where we adopt the following notation of [ZT03]

$$f \doteq \mathrm{SNR}^b \Leftrightarrow \lim_{\mathrm{SNR}\to\infty} \frac{\log f}{\log \mathrm{SNR}} = b$$

where f is a function of SNR.

Throughout this work, we focus on the *average rate* over infinitely many fading blocks

$$R \triangleq \lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} R_{\mathcal{I}(\mathbf{H}_l)} \overset{\text{a.s.}}{=} \sum_{i=1}^{K} \Pr(\mathcal{I} = i) R_i.$$

Accordingly, the system is said to have a *multiplexing gain* of r, also in an average sense, if

$$\lim_{\text{SNR}\to\infty} \frac{R}{\log \text{SNR}} = \sum_{i=1}^{K} r_i \lim_{\text{SNR}\to\infty} \Pr(\mathcal{I} = i) = r. \tag{3.5}$$

The main question we try to answer in this chapter is: Over all sequences of schemes subject to the above constraints, which are able to provide a multiplexing gain of r, what is the maximum diversity gain $d_K^*(r)$? We will consider two related problems:

- *Single-rate transmission:* For each value of SNR, the transmission rate is independent of the feedback index. In other words, it is constrained that

$$r_1 = \cdots = r_K = r. \tag{3.6}$$

  This models a system designed to support constant-rate services, such as voice or video transmission [CTB99, BCT01].
- *Adaptive-rate transmission:* In a more relaxed setting, the values of $r_1, \ldots, r_K$ can also be optimized subject to (3.5), i.e., we consider a variable-rate MIMO system. In addition, a *constraint on the individual multiplexing gains* is assumed

$$r_i \geq r_{\min}, \forall i \in \{1, \ldots, K\}, \tag{3.7}$$

  where $r_{\min}$ is a constant in $(0, \min(N_\text{r}, N_\text{t}))$, referred to as the *minimum multiplexing gain.*

The interpretation of (3.7) is that for certain applications, an acceptable quality of service is only achieved at a certain minimum rate. Any rate above this threshold will enhance the quality of service, thus rate adaptation still makes sense [LLYS03]. In the limit $\text{SNR} \to \infty$, to be meaningful, the minimum rate should be translated into a minimum multiplexing gain. Any fixed minimum rate then becomes a limiting case, $r_{\min} \downarrow 0$. Furthermore, without imposing (3.7), it is not very meaningful to discuss *outage*, i.e., the event that a channel realization cannot support the rate, because a transmitter may, upon accessing to some CSIT, switch off transmission completely. Of course, (3.7) can also be imposed for single-rate transmission, but the solution is only a simple truncated version of the optimal D–M tradeoff without assuming such a constraint. This is not the case for the adaptive-rate problem, where the value $r_{\min}$ heavily influences the optimal D–M tradeoff.

Before proceeding to the derivation of the optimal D–M tradeoff, let us clarify the notations used throughout the discussion. Throughout this chapter, we define $m \triangleq \max(N_\text{r}, N_\text{t})$ and $n \triangleq \min(N_\text{r}, N_\text{t})$. For a matrix $\mathbf{X}$, $\operatorname{tr}\mathbf{X}$ denotes the trace of $\mathbf{X}$ and $\mathbf{X}^\text{H}$ denotes transpose and conjugate. For any $\pi > 0$ and $\rho > 0$, define

$$I(\mathbf{H}, \pi) \triangleq \log\det\left(\mathbf{I}_{N_\text{r}} + \frac{\pi}{N_\text{t}}\mathbf{H}\mathbf{H}^\text{H}\right) \tag{3.8}$$

where the identity matrix of size N_r is denoted as $\mathbf{I}_{N_\text{r}}$, and

$$F(\rho, \pi) \triangleq \Pr\left(I(\mathbf{H}, \pi) < \rho\right) \tag{3.9}$$

where the probability is over the randomness of the channel matrix.

The notation $f \doteq \mathrm{SNR}^{-\infty}$ is used to indicate that the function f of SNR decays exponentially as $\mathrm{SNR} \to \infty$, for example when $f = \exp(-\mathrm{SNR}^p)$ for some constant $p > 0$. The "exponent inequalities" $\dot{\leq}$ and $\dot{\geq}$ are explicitly defined as

$$f \dot{\geq} \mathrm{SNR}^b \Leftrightarrow \liminf_{\mathrm{SNR}\to\infty} \frac{\log f}{\log \mathrm{SNR}} \geq b,$$

$$f \dot{\leq} \mathrm{SNR}^b \Leftrightarrow \limsup_{\mathrm{SNR}\to\infty} \frac{\log f}{\log \mathrm{SNR}} \leq b.$$

For two sequences f, g of SNR, $f \doteq g \Leftrightarrow \frac{f}{g} \doteq \mathrm{SNR}^0$.

3.3 Outage Upper Bound on the Optimal D–M Tradeoff

An *outage* event has the interpretation that the conditional channel cannot support the rate $R_{\mathcal{I}}$, even if "Gaussian codebooks" with infinitely long codewords are employed [OSW94]. For a given feedback scheme, the *outage probability*, which is also SNR-dependent, is thus defined as

$$P_{\mathrm{out},\,K} \triangleq \Pr\left(\log\det\left(\mathbf{I}_{N_\mathrm{r}} + \mathbf{H}\mathbf{Q}_{\mathcal{I}(\mathbf{H})}\mathbf{H}^{\mathrm{H}}\right) < R_{\mathcal{I}(\mathbf{H})}\right)$$

where the positive semi-definite matrices $\{\mathbf{Q}_i\}_{i=1}^K$ correspond to the covariance matrices of the input of the channel conditioned on a feedback index. Outage probability is relevant to our discussion on the D–M tradeoff because an application of Fano's inequality, similar to that in [ZT03, Lemma 5], yields an upper bound on the optimal D–M tradeoff

$$P_\mathrm{e} \dot{\geq} P_{\mathrm{out},\,K} \doteq \mathrm{SNR}^{-d_{\mathrm{out},K}(r)} \geq \mathrm{SNR}^{-d^*_{\mathrm{out},K}(r)},$$

for any feedback schemes and any codes of finite length T. Herein $-d^*_{\mathrm{out},\,K}(r)$ denotes the SNR exponent of the minimum outage probability P^*_{out} (over all feedback schemes with resolution K). To find $d^*_{\mathrm{out},\,K}(r)$, we will, for each given SNR, characterize the resolution-constrained feedback schemes that minimize the outage probability, and then study the asymptotic behavior of the solution.

Single-rate Transmission

Recall that single-rate transmission means that the constraint (3.6) is imposed. Finding $d^*_{\mathrm{out},\,K}(r)$ requires a joint optimization over $\mathcal{I}(\mathbf{H})$, $\{P_i\}_{i=1}^K$, and $\{\mathbf{Q}_i\}_{i=1}^K$ where $\mathrm{tr}(\mathbf{Q}_i) \leq P_i$. In the limit $\mathrm{SNR} \to \infty$, however, it suffices to consider the $\mathbf{Q}_i$'s to be scaled identity matrices, which are of course dependent on the partial CSIT and the SNR. To see this, note that for any given $\mathcal{I}(\mathbf{H})$ and P_1^K, by choosing $\underline{\mathbf{Q}}_i = \frac{P_i}{N_\mathrm{t}}\mathbf{I}_{N_\mathrm{t}}$ and $\overline{\mathbf{Q}}_i = P_i\mathbf{I}_{N_\mathrm{t}}$, $\forall i \in \{1, \ldots, K\}$, we obtain the following bounds on the outage probability,

$$\begin{aligned}
&\lim_{\mathrm{SNR}\to\infty} \frac{\log \Pr\left(\log\det\left(\mathbf{I}_{N_\mathrm{r}} + P_{\mathcal{I}}\mathbf{H}\mathbf{H}^\mathrm{H}\right) < r\log\mathrm{SNR}\right)}{\log\mathrm{SNR}} \\
&\le \lim_{\mathrm{SNR}\to\infty} \frac{\log P_{\mathrm{out},\,K}}{\log\mathrm{SNR}} \\
&\le \lim_{\mathrm{SNR}\to\infty} \frac{\log \Pr\left(\log\det\left(\mathbf{I}_{N_\mathrm{r}} + \frac{P_{\mathcal{I}}}{N_\mathrm{t}}\mathbf{H}\mathbf{H}^\mathrm{H}\right) < r\log\mathrm{SNR}\right)}{\log\mathrm{SNR}}.
\end{aligned} \tag{3.10}$$

With "regular" power allocation schemes such that $P_i \doteq \mathrm{SNR}^{p_i}$ where $0 < p_i < \infty$, $\forall i$, it then follows from (3.10) that we can restrict our analysis to the case $\mathbf{Q}_i = \frac{P_i}{N_\mathrm{t}}\mathbf{I}_{N_\mathrm{t}}$, $\forall i$. Intuitively, in the high-SNR regime of interest, allocating the transmit power evenly over all spatial directions does not affect the SNR exponent of the outage probability [1]. We can therefore focus exclusively on the asymptotic behavior of systems utilizing partial CSIT to control *only* the transmit power so as to minimize the outage probability. Such a system is completely determined by an index mapping and a *power codebook* $\{P_i\}_{i=1}^K$, characterized in the following lemma (recall that only deterministic mappings are considered).

Lemma 3.1. *For a given SNR and rate R, the outage-minimizing power codebook* $\{P_i^*\}_{i=1}^K$ *solves the following optimization problem*

$$\begin{aligned}
\max \quad & P_K \\
\text{s.t.} \quad & [F(R,P_K) + 1 - F(R,P_1)]P_1 \\
& + \sum_{i=2}^{K}[F(R,P_{i-1}) - F(R,P_i)]P_i \le SNR, \\
& 0 \le P_1 < \cdots < P_K,
\end{aligned} \tag{3.11}$$

where $F(\rho,\pi)$ *is defined as in (3.9). The optimal deterministic index mapping is given by*

$$\mathcal{I}^*(\mathbf{H}) = \begin{cases} 1 & \text{if } I(\mathbf{H},P_K^*) < R \\ \min\{i : i\in\{1,\dots,K\},\ I(\mathbf{H},P_i^*) \ge R\} & \text{otherwise,} \end{cases} \tag{3.12}$$

where $I(\mathbf{H},\pi)$ *is defined as in (3.8). The minimum outage probability is* $F(P_K^*)$.

Proof. Given in Appendix 3.A. □

[1]Indeed, quantizing the direction, i.e., the right singular vectors of the channel matrix [LHS03, MSEA03], provides an *SNR gain* (somewhat similar to the coding gain of a code) that is still important in some range of SNR. However, in the current work we are only interested in characterizing the diversity gain in the framework of the D–M tradeoff, i.e., as SNR → ∞.

Interestingly, the optimal index mapping (3.12) has a "circular" structure where the "best" and the "worst" channel realizations share a common index. An economic interpretation is that if a channel realization is too costly to invert, the system should save power for better channel condition. Unlike in the perfect-CSIT case [CTB99, BCT01], however, the optimal transmitter in the limited-feedback case does not necessarily switch off transmission, i.e., P_1^* is generally nonzero. An intuitive explanation is that switching off transmission requires a zero power level in the power codebook, which is rather costly given the finite resolution of the feedback link. This is especially true in the high SNR regime, where switching off transmission does not save power significantly. Also note that the region associated with $\mathcal{I}^*(\mathbf{H}) = 1$ is also the *only* region where an outage event may occur.

It now remains to determine the asymptotic SNR exponent of P_K^* that solves (3.11). Notice that at high SNR, most channel are not in outage, thus some quantization regions may become so unlikely that a power level with an SNR exponent strictly larger than one can be employed without violating (3.3). This motivates the following lemma, which determines the SNR exponent of the outage probability given the asymptotic SNR exponent of the transmit power.

Lemma 3.2. *For $r \in (0, n)$, let π be a function of SNR such that $\pi \doteq SNR^p$ where p is a finite constant and $p \geq 1$. We then have*

$$F(r \log SNR, \pi) \doteq SNR^{-D(r,p)}$$

where $F(r \log SNR, \pi)$ is defined as in (3.9) and

$$D(r,p) \triangleq \inf_{\alpha_1^n \in \mathcal{A}} \sum_{i=1}^{n} (2i - 1 + m - n)\, \alpha_i, \tag{3.13}$$

where

$$\mathcal{A} \triangleq \left\{ \alpha_1^n | \alpha_1 \geq \cdots \geq \alpha_n \geq 0,\ \sum_{i=1}^{n} (p - \alpha_i)^+ < r \right\}.$$

Proof. The asserted result is a natural extension of [ZT03, Theorem 4]. We have

$$F(r \log \mathrm{SNR}, \pi) = \Pr\left(\sum_{j=1}^{n} \log\left(1 + \frac{\pi}{N_\mathrm{t}} \lambda_j\right) < r \log \mathrm{SNR} \right)$$

where $\lambda_1 \leq \cdots \leq \lambda_n$ are the n largest eigenvalues of $\mathbf{H}\mathbf{H}^\mathrm{H}$. By a change of variables $\lambda_j = \mathrm{SNR}^{-\alpha_j}$, $j = 1, \ldots, n$, we obtain

$$\begin{aligned} F(r \log \mathrm{SNR}, p) &\doteq \Pr\left(\prod_{j=1}^{n} \mathrm{SNR}^{(p-\alpha_j)^+} < \mathrm{SNR}^r \right) \\ &= \Pr\left(\sum_{i=j}^{n} (p - \alpha_j)^+ < r \right) \end{aligned}$$

where $(x)^+ \triangleq \max(0, x)$. With the λ_j's following a Wishart distribution, an application of Varadhan's integral lemma [DZ98, ZT03] (noticing that outage is a *rare* event as SNR $\to \infty$) yields the asserted result. □

The region $\mathcal{A}$ only contains $\alpha_j \geq 0$, $\forall j$, because outside this region, the outage probability decays exponentially as SNR $\to \infty$. Also note that the minimizer of (3.13) can be found by a direct investigation:

$$\alpha_i^* = \begin{cases} p & \text{for } i = 1, \ldots, n - J - 1, \\ (J+1)p - r & \text{for } i = n - J, \\ 0 & \text{for } i = n - J + 1, \ldots, n, \end{cases} \tag{3.14}$$

where $J \triangleq \left\lfloor \frac{r}{p} \right\rfloor$, i.e., the largest integer that does not exceed $\frac{r}{p}$.

We are now ready to state the following theorem, which recursively characterizes the SNR exponent of the minimum outage probability of a single-rate MIMO system with quantized feedback.

Theorem 3.1. *The optimal D–M tradeoff of a single-rate MIMO system with K quantization regions in the feedback link is upper-bounded by the outage bound*

$$d^*_{out,\,K}(r) = D(r, 1 + d^*_{out,\,K-1}(r)) \tag{3.15}$$

where $d^*_{out,\,0}(r) \triangleq 0$, $\forall r$ and $D(r,p)$ is defined as in (3.13).

The formal proof, deferred to Appendix 3.B, involves the computation of a lower bound and an upper bound on $d^*_{\text{out},K}(r)$ that asymptotically match. The lower bound is obtained by choosing $\underline{P}_1 = \frac{\text{SNR}}{K}, \underline{P}_2 = \frac{\text{SNR}}{KF(R,\underline{P}_1)}, \ldots, \underline{P}_K = \frac{\text{SNR}}{KF(R,\underline{P}_{K-1})}$, implying that the following index mapping together with the power codebook $\{\underline{P}_i\}_{i=1}^K$ can be used to achieve the outage bound

$$\hat{\mathcal{I}}(\mathbf{H}) = \begin{cases} K \quad \text{if} & I(\mathbf{H}, \underline{P}_K) < r \log \text{SNR}, \\ \min\{i : i \in \{1, \ldots, K\},\ I(\mathbf{H}, \underline{P}_i) \geq R\} & \text{otherwise}. \end{cases} \tag{3.16}$$

But (3.16) is similar to the power control scheme proposed for the MIMO ARQ [ECD06] and for multiple-input single-output (MISO) channels [BSA02, KS04b] in the sense that the *highest* power level is used *even in the outage region*. Our result implies that such a non-circular power control scheme is (only) asymptotically optimal in an outage sense. The intuition is that the amount of power saved by using the truly optimal circular mapping (3.12) goes to zero as SNR $\to \infty$.

Some interesting results can be obtained as direct consequences of Theorem 3.1. The first one characterizes the limiting values on the outage bound, highlighting the fast increase in diversity gain at low rate.

Corollary 3.1. *We have*

$$\lim_{r \downarrow 0} d_K^*(r) = \sum_{k=1}^{K} (N_t N_r)^k$$

and

$$\lim_{r \uparrow n} d_{out,\ K}^*(r) = 0, \quad \forall K.$$

Proof. For r sufficiently close to zero and any $p \geq 1$, (3.13) is minimized by $\alpha_i^* = p$, $i = 1, \ldots, n-1$, and $\alpha_n^* = p - r$ leading to $\lim_{r \downarrow 0} D(r, p) = N_\mathrm{t} N_\mathrm{r} p$. Applying Theorem 3.1, we obtain $\lim_{r \downarrow 0} d_{\mathrm{out},\ K}^*(r) = \sum_{k=1}^{K} (N_\mathrm{t} N_\mathrm{r})^k$, $\forall K$. It can also be verified that $\lim_{r \uparrow n} D(r, 1) = 0$. Using Theorem 3.1 gives $\lim_{r \uparrow n} d_{\mathrm{out},\ K}^*(r) = 0$, $\forall K$. □

The results reveal the fact that in a SISO channel, the maximal diversity gain only scales *linearly* with the number of feedback levels K. On the contrary, when the system is equipped with multiple antennas, the maximal diversity gain grows *exponentially* in the feedback resolution K.

The next result explicitly characterizes the outage bound for a channel where at least one side of the communication link has a single antenna.

Corollary 3.2. *If $n = 1$, the outage bound consists of a single segment between $(0, \sum_{k=1}^{K} m^k)$ and $(1, 0)$.*

Proof. For any $k \geq 1$, because $d_{\mathrm{out},\ k}^*(r)$ is a monotonically decreasing function of r, we have $\frac{r}{1+d_{\mathrm{out},\ k}^*(r)} < 1$, $\forall r \in (0, 1)$. From (3.14) , we then obtain $\alpha_1^* = 1 + d_{\mathrm{out},\ k}^*(r) - r$, leading to $d_{\mathrm{out},\ k+1}^*(r) = D(r, 1 + d_{\mathrm{out},\ k}^*(r)) = m(1 + d_{\mathrm{out},\ k}^*(r) - r)$. But $d_{\mathrm{out},\ 1}^*(r)$ is a single segment, thus $d_{\mathrm{out},\ 2}^*(r), \ldots, d_{\mathrm{out},\ K}^*(r)$ are all single segments. □

Adaptive-rate Transmission

A few technicalities need to be taken care of in the adaptive-rate case. One may try to characterize the optimal index mapping, power and rate allocation for each SNR as in the previous section. This is however a difficult task, even in the single-antenna case with perfect CSIT [LLYS03], and becomes even less tractable in our case due to the joint optimization involving the index mapping. Thus we follow another approach and find an upper bound on the outage bound itself, which is shown to be asymptotically tight, leading to a situation where the single-rate results can be reused. The result, proved in Appendix 3.C, is formalized as follows.

Theorem 3.2. *For a given $r_{min} \in (0, n)$ and $r \in [r_{min}, n)$, the optimal D–M tradeoff of a MIMO system with $K \geq 2$ quantization regions in the feedback link and a minimum multiplexing gain r_{min} is upper-bounded by the outage bound*

$$d_{out,\ K}^*(r, r_{min}) = D(r_{min}, 1 + d_{out,\ K-1}^*(r, r_{min}))$$

where $d^*_{out,\,1}(r, r_{min}) \triangleq D(r, 1)$, $\forall r \geq r_{min}$.

Essentially, the outage bound is obtained by considering a two-rate system. The rate of one quantization region asymptotically dominates the average multiplexing gain of the system, while power control is employed over the remaining $K - 1$ regions, where the minimum rate is used. The fact that the outage bounds in Theorem 3.2 have a very similar form to those in Theorem 3.1 can be explained intuitively as follows. The solution to the adaptive-rate problem is an optimal combination of channel inversion to reduce outage and water-filling to improve throughput [LLYS03]. At high SNR, the effect of water-filling quickly diminishes and therefore, we essentially reduce to a channel inversion scheme having similar structure to the single-rate case.

In an extreme case, when $r_{\min} \downarrow 0$, a surprisingly simple outage bound can be obtained, which we loosely refer to as the case of *zero minimum multiplexing gain*.

Corollary 3.3. *The optimal D–M tradeoff of a MIMO system with $K \geq 2$ quantization regions in the feedback link and a zero minimum multiplexing gain is upper-bounded by the outage bound*

$$\lim_{r_{min}\downarrow 0} d^*_{out,K}(r, r_{min}) = (N_r N_t)^{K-1} D(r, 1) + \sum_{k=1}^{K-1} (N_r N_t)^k.$$

For a given K, the bound in Corollary 3.3 is simply a scaled version of the no-CSIT tradeoff $D(r, 1)$ plus a constant, and therefore is piece-wise linear between the points $(r = j, d = (N_\mathrm{r} N_\mathrm{t})^{K-1}(N_\mathrm{r} - j)(N_\mathrm{t} - j) + \sum_{k=1}^{K-1}(N_\mathrm{r} N_\mathrm{t})^k)$, for $j = 0, \ldots, n$. Interestingly, the so-called "full diversity gain" $N_\mathrm{r} N_\mathrm{t}$, is already achieved with $K = 2$ at the "maximal" multiplexing gain, i.e., when $r \uparrow n$. However, to obtain this bound, the multiplexing gains over $K - 1$ quantization regions must approach zero, i.e., some strictly positive, but negligible compared to $\log \mathrm{SNR}$, rates are employed.

3.4 Achievability of the Optimal D–M Tradeoff

In this section, we show that the outage bound is achievable, conditioned on the existence of a general class of codes with finite length. For brevity, we present the proof only for the single-rate case and omit the almost identical derivations of the adaptive-rate case. The analysis reveals a simple geometrical interpretation of such a class of codes. It turns out that for single-rate transmission, the optimal D–M tradeoff can be achieved by a *single* codebook and a feedback-dependent power controller, a combination known to be optimal in different scenarios [CS99, CTB99, BCT01, SJ03]. Connection to approximately universal codes [TV06] and non-vanishing determinant codes, e.g., in [BR03, EKP$^+$06] among others, will be discussed.

Extended Approximately Universal Condition

Consider a sequence of codes $\mathcal{C}$ with rate $r \log \mathrm{SNR}$ (bits per channel use) using equally likely codewords $\mathbf{X}(k)$, $k = 1, \ldots, M$ with $M \triangleq \lfloor \mathrm{SNR}^{rT} \rfloor$, which are matrices of size $N_\mathrm{t} \times T$, where $T \geq N_\mathrm{t}$. The *average* energy of a component in a codeword matrix is normalized such that

$$\frac{1}{MTN_\mathrm{t}} \sum_{k=1}^{M} \|\mathbf{X}(k)\|_\mathrm{F}^2 \leq 1. \tag{3.17}$$

For an arbitrary pair of codewords, let $\mu_1 \leq \cdots \leq \mu_n$ be the n smallest squared singular values of the codeword difference matrix, and let $\mu_j = \mathrm{SNR}^{-\beta_j}$.We constrain $\mathcal{C}$ so that[2]

$$\left(\min_{\mathcal{C}} \prod_{j=1}^{n} \mathrm{SNR}^{-(\beta_j)^+} \right) \dot{\geq} \mathrm{SNR}^{-r} \tag{3.18}$$

where the minimization is over all pairs of different codewords in $\mathcal{C}$. Recall that $(\beta_j)^+ = \max(\beta_j, 0)$. We refer to (3.18) as the *extended approximately universal criterion*, because it is related, but not necessarily equivalent to the approximately universal criterion developed in [TV06]. In the following, we show that the outage bound is achievable by combining $\mathcal{C}$ with a CSIT-dependent power amplifier and a suitable index mapping.

Let ϵ be an arbitrarily small positive number and consider the following index mapping

$$\underline{\mathcal{I}}(\mathbf{H}) = \begin{cases} 1 & \text{if } I(\mathbf{H}, \underline{P}_K) < (r+\epsilon)\log \mathrm{SNR}, \\ \min\{i : i \in \{1, \ldots, K\}, I(\mathbf{H}, \underline{P}_i) \geq (r+\epsilon)\log \mathrm{SNR}\} & \\ & \text{otherwise.} \end{cases}$$

where $\underline{P}_1 = \frac{\mathrm{SNR}}{K}$, $\underline{P}_2 = \frac{\mathrm{SNR}}{KF((r+\epsilon)\log \mathrm{SNR}, \underline{P}_1)}, \ldots, \underline{P}_K = \frac{\mathrm{SNR}}{KF((r+\epsilon)\log \mathrm{SNR}, \underline{P}_{K-1})}$. We then construct the transmit codewords as

$$\mathbf{S}_i(k) = \sqrt{\frac{\underline{P}_i}{N_\mathrm{t}}} \mathbf{X}(k), \quad i = 1, \ldots, K, \ k = 1, \ldots, M.$$

where the factor $\sqrt{\frac{\underline{P}_i}{N_\mathrm{t}}}$ is used to guarantee the power constraint (3.3).

For any $i \in \{1, \ldots, K\}$, consider an arbitrary pair of codewords in $\mathcal{C}$, say $\mathbf{X}(1)$ and $\mathbf{X}(2)$ and denote $\Delta\mathbf{X} = \mathbf{X}(1) - \mathbf{X}(2)$. Also define the ith ϵ-outage-free region as

$$\overline{\mathcal{O}}_i^\epsilon = \{\mathbf{H} : \underline{\mathcal{I}}(\mathbf{H}) = i, \ I(\mathbf{H}, \underline{P}_i) \geq (r+\epsilon)\log \mathrm{SNR}\}.$$

[2] Recall that by our definition of the exponent ("dot") inequality, the sequence on the left hand side of (3.18) does not necessarily converge.

Clearly, except for $i = 1$ the second condition defining $\overline{\mathcal{O}}_i^\epsilon$ is always redundant. The pairwise error probability, that $\mathbf{S}_i(2)$ is incorrectly detected given $\mathbf{S}_i(1)$ is transmitted, averaged over the channel, can be upper-bounded by

$$\begin{aligned}
&\Pr\left(\mathbf{S}_i(1) \to \mathbf{S}_i(2), \mathbf{H} \in \overline{\mathcal{O}}_i^\epsilon\right) \\
&\le \int_{\overline{\mathcal{O}}_i^\epsilon} \exp\left(-\frac{\underline{P}_i}{4N_\mathrm{t}}\|\mathbf{H}\Delta\mathbf{X}\|^2\right) f(\mathbf{H})d\mathbf{H} \\
&\le \int_{\overline{\mathcal{O}}_i^\epsilon} \exp\left(-\frac{\underline{P}_i}{4N_\mathrm{t}}\sum_{j=1}^{n}\lambda_{n-j+1}\mu_j\right) f(\lambda_1^n)d\lambda_1^n.
\end{aligned}$$

The second inequality is due to the worst-case rotation which aligns the ordered singular values of the codeword different matrix $\sqrt{\mu_j}$'s with those of the channel $\sqrt{\lambda_j}$'s in reverse order [KW03]. Because $\underline{\mathcal{I}}(\mathbf{H})$ and $I(\mathbf{H}, p)$ only depend on $\mathbf{H}$ via λ_1^n, any $\overline{\mathcal{O}}_i^\epsilon$ is completely characterized by λ_1^n.

We now compute the asymptotic SNR exponent of the upperbound on the pairwise error probability. By changing variables $\lambda_j = \mathrm{SNR}^{-\alpha_j}$, $\mu_j = \mathrm{SNR}^{-\beta_j}$ and assuming $\underline{P}_i \doteq \mathrm{SNR}^{p_i}$, we have

$$\begin{aligned}
&\int_{\overline{\mathcal{O}}_i^\epsilon} \exp\left(-\frac{\underline{P}_i}{4N_\mathrm{t}}\sum_{j=1}^{n}\lambda_{n-j+1}\mu_j\right) f(\lambda_1^n)d\lambda_1^n \\
&\doteq \int_{\overline{\mathcal{O}}_i^\epsilon} \exp\left(-\sum_{j=1}^{n}\mathrm{SNR}^{p_i-(\alpha_{n-j+1}+\beta_j)}\right) f(\alpha_1^n)d\alpha_1^n \\
&\doteq \int_{\mathcal{B}_i} f(\alpha_1^n)d\alpha_1^n + \mathrm{SNR}^{-\infty}
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{B}_i &\triangleq \overline{\mathcal{O}}_i^\epsilon \cap \{\alpha_1^n : \alpha_{n-j+1} + \beta_j \ge p_i, j = 1, \ldots, n\} \\
&\subset \{\alpha_1^n : \alpha_1 \ge \cdots \ge \alpha_n, \sum_{j=1}^{n}(p_i - \alpha_j)^+ \ge r + \epsilon\} \cap \\
&\qquad \cap \{\alpha_1^n : \alpha_{n-j+1} + \beta_j \ge p_i, j = 1, \ldots, n\}.
\end{aligned}$$

Thus $\mathcal{B}_i$ is essentially a subset of $\overline{\mathcal{O}}_i^\epsilon$ containing the channel realizations that make the bound on the pairwise error probability decay sub-exponentially. Note that although $\mathcal{B}_i$ can be specified precisely for the Wishart distributed λ_1^n, we bound it by a larger set that is *independent* of the distribution of α_1^n and will show that this is sufficient.

For any $\alpha_1^n \in \mathcal{B}_i$, we have $\beta_j \geq p_i - \alpha_{n-j+1}$ leading to $(\beta_j)^+ \geq (p_i - \alpha_{n-j+1})^+$. Thus

$$\sum_{j=1}^{n} (\beta_j)^+ \geq \sum_{j=1}^{n} (p_i - \alpha_j)^+ \geq r + \epsilon.$$

However, the condition (3.18) implies that for any $\epsilon > 0$, there exists an $\overline{\mathrm{SNR}}$ so that for all $\mathrm{SNR} > \overline{\mathrm{SNR}}$:

$$\prod_{j=1}^{n} \mathrm{SNR}^{-(\beta_j)^+} \geq \mathrm{SNR}^{-r-\frac{\epsilon}{2}},$$

leading to

$$\sum_{j=1}^{n} (\beta_j)^+ \leq r + \frac{\epsilon}{2},$$

meaning that the set $\mathcal{B}_i$ is *empty* for any $\mathrm{SNR} > \overline{\mathrm{SNR}}$, and hence the average pairwise error probability decays exponentially as $\mathrm{SNR} \to \infty$. This holds for any pair of codewords, hence the unionbound on the average error probability gives

$$\Pr(\text{error}, \mathbf{H} \in \overline{\mathcal{O}}_i^{\epsilon}) \doteq \mathrm{SNR}^{Tr} \mathrm{SNR}^{-\infty} \doteq \mathrm{SNR}^{-\infty}$$

for any i and finite codeword length T. Thus, for an extended approximately universal code, the error probability is dominated by the outage event $I(\mathbf{H}, \underline{P}_K) < (r + \epsilon) \log \mathrm{SNR}$, which by construction of $\underline{\mathcal{I}}(\mathbf{H})$ has asymptotic SNR exponent $-d^*_{\mathrm{out},K}(r + \epsilon)$, with $d^*_{\mathrm{out},K}(r + \epsilon)$ defined as in Theorem 3.1. Therefore

$$P_{\mathrm{e}} \dot{\leq} \mathrm{SNR}^{-d^*_{\mathrm{out},K}(r+\epsilon)}.$$

By continuity of $d^*_{\mathrm{out},K}(r)$, the system can achieve points arbitrarily close to the outage bound as $\epsilon \downarrow 0$. We finally have the following.

Theorem 3.3. *If there exists a sequence of codes satisfying the extended approximately universal criterion, i.e., satisfying (3.17) and (3.18), then the optimal D–M tradeoff of the MIMO channel (3.1) with feedback resolution K and codeword length T is given by*

$$d^*_K(r) = d^*_{out,\ K}(r)$$

for single-rate transmission and

$$d^*_K(r, r_{min}) = d^*_{out,\ K}(r, r_{min})$$

for adaptive-rate transmission with minimum multiplexing gain r_{min} .

The adaptive-rate case can be proved in a similar manner. Interestingly, for adaptive transmission, the optimal tradeoff can be achieved by only two codebooks, leading to an implication for practical applications: At sufficiently high SNR, an

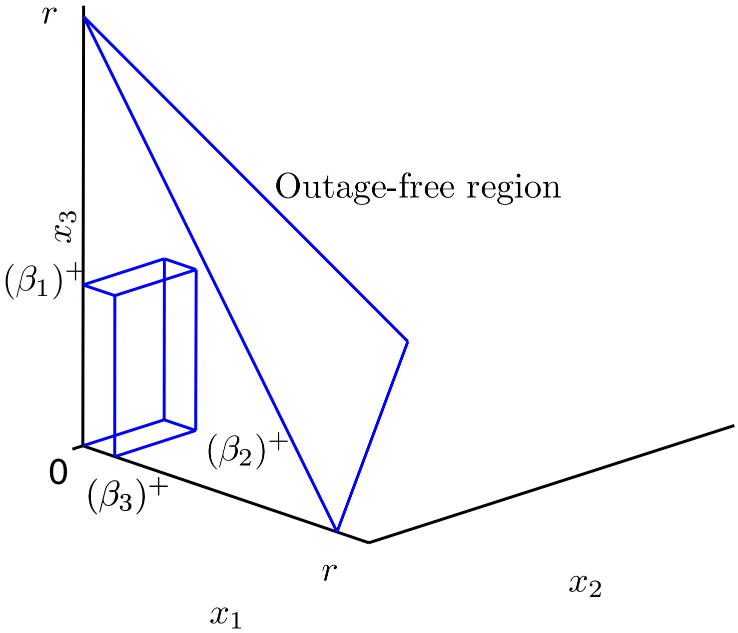

Figure 3.1: A geometrical interpretation of the condition (3.18). The channel realizations making a pairwise error probability decay polynomially are contained in the cuboid, and (3.18) keeps the corners of all the cuboids in the code below the surface of the plane $x_1 + x_2 + x_3 = r$. (The constraints $x_1 \leq x_2 \leq x_3$ are not plotted to improve readability.)

adaptive MIMO system may not need too many different code rates to achieve performance close to the optimal tradeoff. Indeed, using two code rates is sufficient. The higher rate asymptotically dominates the average multiplexing gain of the system, while the lower rate guarantees the minimum level of quality service in poor channel conditions. Notice however that the D–M tradeoff result is an asymptotic analysis, thus for low and moderate SNR, that conclusion does not necessarily hold.

Discussion

The condition (3.18) has a simple geometrical interpretation, illustrated in Fig. 3.1. Let $x_j = (p_i - \alpha_j)^+ \geq 0$. The set of all x_1^n in the nonnegative orthant so that $\sum_{j=1}^n x_j \geq r + \epsilon$ contains the ϵ-outage-free region. On the other hand, the set of all "bad" channels that make the pairwise probability decay sub-exponentially is contained in the cuboid defined by $0 \le x_j \le (\beta_j)^+$, where the β_j's correspond to the SNR exponents of the squared singular values of an arbitrary codeword difference matrix. By keeping the corners of any such cuboid below the surface of the hyperplane $\sum_{j=1}^n x_j = r$, the condition (3.18) guarantees that all pairwise error probabilities decay exponentially as SNR $\to \infty$.

Having all pairwise error probabilities decay exponentially in the limit SNR $\to \infty$ as long as the channel is not in outage, regardless of the channel distribution, is called the *approximately universal* property in [TV06]. In [TV06], the necessary and sufficient conditions for a code to be approximately universal are established, assuming no CSIT ($P_i =$ SNR) and under a power constraint on *every* codeword (an indispensable assumption for their results to hold). In this chapter, we relax this assumption and approach the problem within the framework of quantized feedback, giving some new insight. Furthermore, by tailoring the universality condition to the problem at hand, we obtain a stronger result, that the optimal D–M tradeoff can be achieved by a *single* code together with a power amplifier, which only depends on the feedback index.

For completeness, in Appendix 3.D we show that (3.18) cannot be relaxed to admit more codes that is approximately universal for *any sequence* of channel distributions, even though such a necessary condition is not the main focus of this chapter. However, this (very) strong requirement may not be necessary to achieve the optimal D–M for a *particular* channel distribution. This motivates our further study on the achievable tradeoffs of codes drawn from random ensembles in Section 3.5.

Relations Between Different Criteria

If replacing (3.17) with a more stringent power constraint imposed on every codeword

$$\frac{1}{TN_\mathrm{t}}\|\mathbf{X}(k)\|_\mathrm{F}^2 \le 1,\ \forall k, \tag{3.19}$$

then with any SNR we have $\mu_j = \mathrm{SNR}^{-\beta_j} \le TN_\mathrm{t}$, $\forall j$ for any pair of codewords. The approximately universal condition in [TV06], obtained under (3.19), can be written as

$$\left(\min_{\mathcal{C}} \prod_{j=1}^{n} \mathrm{SNR}^{-\beta_j}\right) \dot{\ge} \mathrm{SNR}^{-r}. \tag{3.20}$$

When $n = N_\mathrm{t} \le N_\mathrm{r}$, (3.20) is obviously equivalent to the non-vanishing determinant criterion (NVD) [BR03, YW03, EKP+06], which requires

$$\left(\min_{\mathcal{C}} \prod_{j=1}^{N_\mathrm{t}} \mathrm{SNR}^{-\beta_j}\right) \dot{\ge} \mathrm{SNR}^{-r}. \tag{3.21}$$

However, for $n = N_\mathrm{r} < N_\mathrm{t}$, (3.20) and (3.21) are *not* equivalent. A code satisfying the stronger NVD criterion (3.21) also satisfies (3.20), but not vice versa.

Under (3.17), the arguments in [TV06] generally do not hold because the eigenvalues of a codeword difference matrix are not bounded. Our results show that under a less restrictive power constraint, (3.20) should be further relaxed to (3.18). Furthermore, any code that has the property (3.20) under (3.19) also satisfies the weaker requirement (3.18), because

$$\left(\min_{\mathcal{C}} \prod_{j=1}^{n} \mathrm{SNR}^{-\beta_j}\right) \overset{(3.19)}{\leq} (TN_\mathrm{t})^{n-\bar{n}} \left(\min_{\mathcal{C}} \prod_{j=1}^{\bar{n}} \mathrm{SNR}^{-\beta_j}\right)$$
$$\doteq \left(\min_{\mathcal{C}} \prod_{j=1}^{n} \mathrm{SNR}^{-(\beta_j)^+}\right)$$

where $\bar{n}$ is the integer such that $\beta_{\bar{n}} \geq 0$ and $\beta_{\bar{n}+1} < 0$. Interestingly, even if $n = N_\mathrm{t}$, the product $\prod_{j=1}^{n} \mathrm{SNR}^{-(\beta_j)^+}$ is no longer equal to the determinant of a codeword difference matrix if that codeword difference has some very large singular values. Somewhat surprisingly, the condition (3.18) implies that having such an abnormally large codeword difference with energy much larger than the noise variance (in an order of magnitude sense) does not really help, as only the nonnegative parts of β_j's count. On the other hand, finding codes belonging to the larger class defined by (3.18) may be an easier task. Besides, it is more helpful to use (3.18) when working with random codes, as will be shown in Section 3.5.

A remaining and very important question is: Does such a constrained code (3.18) exist for MIMO channels? A number of codes are known to satisfy (3.20) as presented in [TV06], for example QAM for single transmit-antenna, Alamouti's scheme with QAM constellations [Ala98] for 2×1 channels and, tilted QAM for two transmit antennas [YW03], Golden codes [BRV05] for 2×2 systems. For a certain numbers of transmit antennas, codes presented in [RBV04, ORBV06, KR05] satisfy (3.21). Consequently, all the aforementioned codes can achieve the optimal D–M tradeoff presented herein. The strongest claim to date, based on an explicit construction, states that there exist codes of length $T \geq N_\mathrm{t}$ satisfying (3.21) for any N_t [EKP$^+$06]. Their existence completes the achievability part, i.e., the optimal D–M tradeoffs of MIMO systems with feedback resolution K coincide with the corresponding outage bounds for any $T \geq N_\mathrm{t}$.

Numerical Examples

The tradeoffs for single-rate and adaptive-rate (with a near zero minimum multiplexing gain) systems are compared in Fig. 3.2. As can be seen, even a few bits of feedback information can increase the diversity gain of a MIMO channel dramatically. A large improvement in the diversity gain can be observed in all cases, even with coarsely quantized feedback. This suggests that from a diversity gain

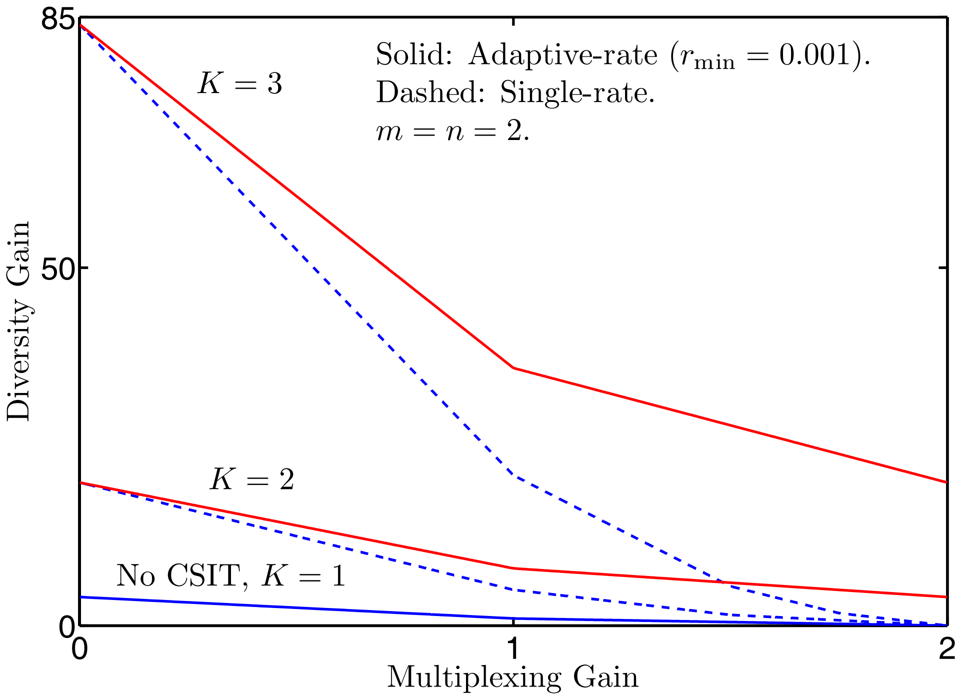

Figure 3.2: D–M tradeoffs of single-rate and adaptive-rate ($r_{\min} = 0.001$) transmission over a 2×2 channel with different feedback resolution K.

perspective, increasing the feedback resolution can be more efficient than adding antennas, provided that transmit power control is possible. The intuition behind this "power-control diversity" is that we can isolate certain "bad" channels into regions with polynomially small probability measures and employ polynomially large powers over those regions without violating the power constraint (cf. also [ECD06]). This gives "*diversity on top of diversity.*" The effect can be interpreted as *time diversity* even though we only code over a *single fading block*: Transmit power is saved over a long time period to be used in a few rare peaks.

As illustrated in Fig. 3.2, the penalty of keeping the rate independent of the feedback index is relatively large, especially at high multiplexing gain. More importantly, it is possible to achieve nonzero diversity gain with rate adaptation, even at the "maximal" average multiplexing gain. For a given resolution K, the two curves coincide at zero multiplexing gain, where the multiplexing gain is not the parameter of interest. Thus rate adaptation is essential to achieve a high throughput together with a nonzero diversity gain.

The D–M tradeoff over a 4×4 channel with different minimum multiplexing gains is plotted in Fig. 3.3. Clearly, increasing the minimum threshold on the individual

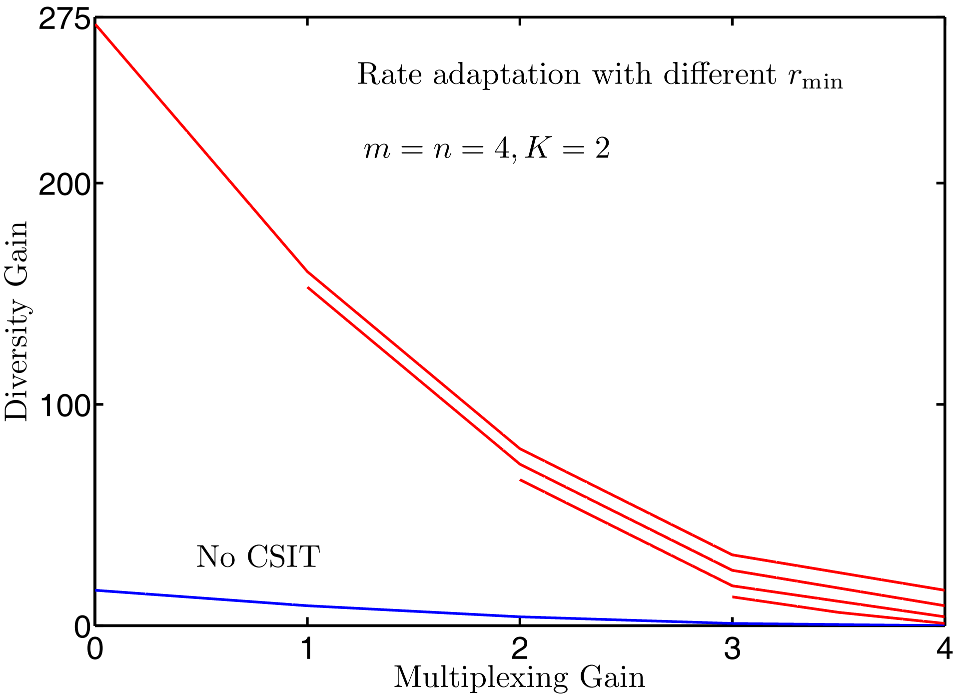

Figure 3.3: D–M tradeoffs for adaptive-rate transmission over a 4×4 channel with minimum multiplexing gain $r_{\min} = 0.001, 1, 2, 3$.

rates leads to a degradation in reliability. However, since $r_{\min}$ is dictated by higher-layer applications, one should not conclude that a small $r_{\min}$ is preferable. As can be seen, the tradeoff with very coarse feedback ($K = 2$) is still strictly better than that without CSIT. For example, the "maximal" multiplexing gain $r = 4$ can be achieved together with a diversity gain $d = 1$ even if the multiplexing gain used over any channel realization never drops below 3.

Comparison to the MIMO ARQ Channel

The numerical results above show that the D–M tradeoffs of the current work share some similarities with those of the MIMO ARQ schemes [ECD06], especially in the adaptive-rate case where the tradeoff curves appear to be "pushed" upwards even at the maximal multiplexing gain. Of course, since our assumed feedback model is fundamentally different from the one in [ECD06], it is not entirely fair to compare the corresponding performance. Nevertheless, it is of interest to understand the relation between these two cases. We discuss a specific example in this section to demonstrate that the D–M tradeoff of the ARQ scheme indeed shares common characteristics with *both* single-rate and adaptive-rate scenarios.

We will compare a power-controlled ARQ channel having an effective multiplexing gain of r and no more than L_{R} rounds of transmission (the channel matrix is unchanged during all rounds) to an adaptive-rate system with multiplexing gain r, feedback resolution L_{R} and minimum multiplexing gain $r_{\min} = \frac{r}{L_{\mathrm{R}}}$. This is because the worst channel conditions require all L_{R} rounds, and thus the smallest "instantaneous" multiplexing gain of the ARQ scheme is only $\frac{r}{L_{\mathrm{R}}}$. In addition, the power codebooks of both systems are of size L_{R}. Note that for the sake of comparison, we constrain the adaptive-rate system to have a minimum multiplexing gain $r_{\min}$ that is *dependent* on r.

The explicit form of the D–M tradeoff for a power-controlled MIMO ARQ system is difficult to compute in general [ECD06]. To get some insight, we study the case $n = 1$ (e.g., an $m \times 1$ MISO channel) where closed-form expressions can be worked out. Let the channel vector be $\mathbf{h}$. Then the outage event in the MIMO ARQ scheme [ECD06] is given by

$$\log\left(1 + \frac{P_1^{\mathrm{ARQ}}}{N_{\mathrm{t}}}\|\mathbf{h}\|_{\mathrm{F}}^2\right) + \ldots + \log\left(1 + \frac{P_{L_{\mathrm{R}}}^{\mathrm{ARQ}}}{N_{\mathrm{t}}}\|\mathbf{h}\|_{\mathrm{F}}^2\right) < r \log \mathrm{SNR}. \tag{3.22}$$

For the adaptive-rate system, the outage event is

$$\log\left(1 + \frac{P_{L_{\mathrm{R}}}^{\mathrm{adap}}}{N_{\mathrm{t}}}\|\mathbf{h}\|_{\mathrm{F}}^2\right) < \frac{r}{L_{\mathrm{R}}} \log \mathrm{SNR}. \tag{3.23}$$

Note that the SNR exponents of the power levels in the two schemes are recursively computed. While $P_1^{\mathrm{ARQ}} \doteq P_1^{\mathrm{adap}} \doteq \mathrm{SNR}$ and $P_2^{\mathrm{ARQ}} \doteq P_2^{\mathrm{adap}} \doteq \mathrm{SNR}^{1+m-mr}$, this exponential equality does not hold for an arbitrary pair P_k^{ARQ} and P_k^{adap}. Indeed the adaptive-rate system has a higher power level for any $k \geq 3$, as discussed in Observation 2 below.

From (3.22), the D–M tradeoffs of the ARQ channel, denoted as $d_{L_{\mathrm{R}}}^{\mathrm{ARQ}}(r)$, can be computed. Let us find $d_{L_{\mathrm{R}}}^{\mathrm{ARQ}}(r)$ explicitly for the first few values of L_{R}. For $L_{\mathrm{R}} = 2$ we have

$$d_2^{\mathrm{ARQ}}(r) = \begin{cases} (m+m^2)(1-r) & \text{if } r < \frac{m}{m+1}, \\ \frac{m(m+2)-m(m+1)r}{2} & \text{otherwise.} \end{cases}$$

For $L_{\mathrm{R}} = 3$ and $m > 1$ we have

$$d_3^{\mathrm{ARQ}}(r) = \begin{cases} (m+m^2+m^3)(1-r) & \text{if } r < \frac{m}{m+1}, \\ m\left(1 - r + \frac{m(m+2)-m(m+1)r}{2}\right) & \text{otherwise.} \end{cases}$$

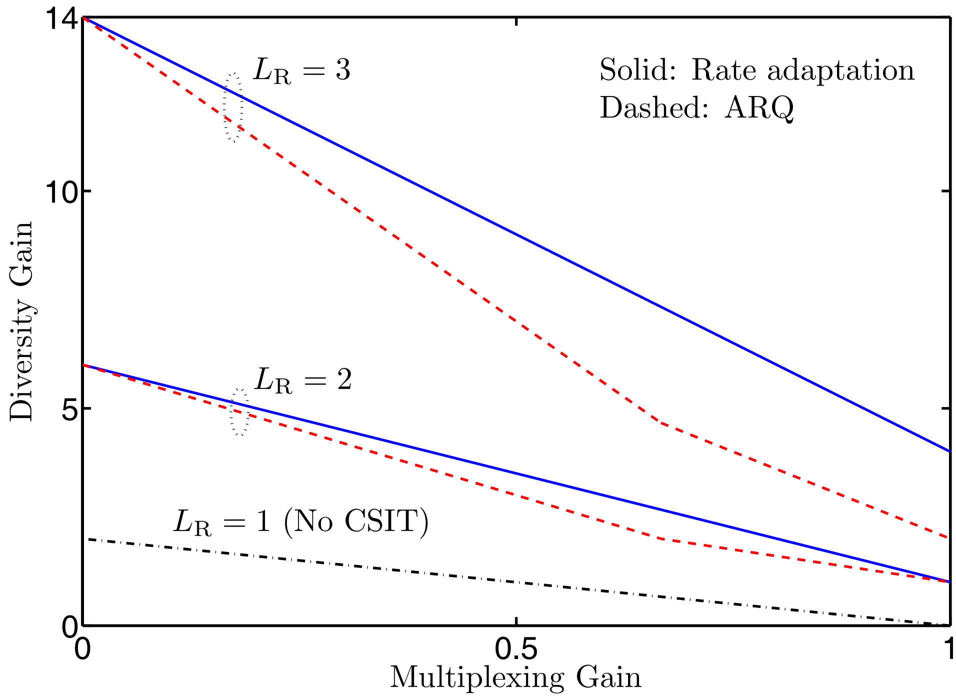

Figure 3.4: Comparison of an adaptive-rate system and the power-controlled MIMO ARQ scheme [ECD06] over a 2×1 channel.

For $m = 1$, i.e., a SISO channel, the result for $L_R = 3$ is slightly different:

$$d_3^{\text{ARQ}}(r) = \begin{cases} 3 - 3r & \text{if } r \in (0, \frac{1}{2}], \\ \frac{9-6r}{4} & \text{if } r \in (\frac{1}{2}, \frac{5}{6}], \\ \frac{11-6r}{6} & \text{otherwise.} \end{cases}$$

The D–M tradeoff of the adaptive-rate system in this $n = 1$ case is found to be (cf. Theorem 3.2)

$$d^*_{L_R}\left(r, \frac{r}{L_R}\right) = \left(m + m^2 + \ldots + m^{L_R}\right) - m^{L_R} r - \left(m + \ldots + m^{L_R - 1}\right) \frac{r}{L_R}.$$

In Fig. 3.4 the D–M tradeoffs of both the ARQ scheme and the adaptive-rate system in the case $m = 2$ are plotted.

We summarize the discussion above in the following two observations.

Observation 1: The first (leftmost) segment of the ARQ tradeoff coincides *exactly* with the D–M curve $d^*_{L_{\rm R}}(r)$ of the *single-rate* system with $L_{\rm R}$ feedback levels (cf. Theorem 3.1). The intuition behind this remarkable phenomenon is that for a sufficiently small r, the $\mathbf{h}$'s that dominate the outage exponent must be in very deep fades such that the first $L_{\rm R}-1$ rounds of transmission are completely "wiped out" by the channel in the sense that $(1+P_i^{\rm ARQ}\|\mathbf{h}\|_{\rm F}^2) \doteq {\rm SNR}^0$, $i=1,\ldots,L_{\rm R}-1$. Thus the first $L_{\rm R}-1$ transmission rounds do not contribute anything to the SNR exponent of the outage probability. Since only the last block of transmission (round $L_{\rm R}$) makes an impact on the outage exponent, it is then clear from (3.22) that the D–M tradeoff of the ARQ system is equivalent to that of the single-rate system investigated in Section 3.3. We notice that this fact is also consistent with a numerical observation in [ECD06] when lower bounds to the optimal D–M tradeoff are computed.[3]

Observation 2: The adaptive-rate system has a higher diversity gain at all multiplexing gains. This can be explained intuitively by comparing (3.22) and (3.23). Because the causal-feedback ARQ scheme keeps trying higher power levels and longer codeword lengths (i.e., lower rates), in the outage event (3.22) each part of the ARQ codeword experiences a different power level. On the other hand, with noncausal knowledge about the channel state, the adaptive-rate scheme can *immediately* switch to a small-rate codebook and a high power. Therefore in the outage event the whole codeword experiences the largest power level as in (3.23), leading to a higher diversity gain.

Since the power levels are computed recursively, this effect is magnified as $L_{\rm R}$ grows. That is, the performance gap between rate adaptation and ARQ widens as the maximum number of transmission rounds increases. On a final note, for a given $L_{\rm R}$, both ARQ and rate adaptation provide the same maximal diversity gain. Given the discussion in Observation 1, we can say that this is simply due to the fact that single-rate and adaptive-rate curves coincide in the limit $r \downarrow 0$.

3.5 Lower Bounds on the Optimal D–M Tradeoff: Gaussian Coding Bounds

There are several reasons to study the achievable D–M tradeoff with codes drawn from a random ensemble. First, the strong codes satisfying the extended approximately universal criterion (3.18) may not be necessary to achieve the optimal D–M for a particular channel distribution. Furthermore, in the no-CSIT case, even Gaussian codes of very short lengths can achieve the optimal D–M tradeoff, inspiring some important classes of low-complexity codes such as Lattice Space-Time codes

[3] This suggests that for sufficiently low multiplexing gains, a simple repetitive power-controlled ARQ scheme, where the transmitter keeps sending the same codeword with higher and higher power levels until success or deadline and where the receiver decodes based on the latest round only, is D–M tradeoff optimal. Intuitively, the gain from power control at low multiplexing gains is so large that it masks out the effects of reducing the code rate (which is already small from the first round).

and their variations for the ARQ channels [ECD04, ECD06]. In this section, we develop two lower bounds on the optimal D–M tradeoff with Gaussian random codes. These bounds are nontrivial generalization of their counterparts in the no-CSIT case [ZT03], and their derivations are considerably different due to the presence of quantized CSIT. It turns out that, except for some special cases, the bounds are only asymptotically tight, which is quite surprising given the result of [ZT03]. Nevertheless, the derived lower bounds quickly approach the outage upper bound even for very moderate codeword lengths. Our results give some insight into the approximate universality of codes drawn from Gaussian ensembles, as well as into the D–M performance of random codes in the presence of CSIT.

Back-off Bounds

The key idea of back-off bounds is to feedback when the channel is good enough to support a strictly larger rate than the transmission rate, hence the term "back-off." By exploiting the gap between the code rate and the instantaneous mutual information of the channel, we gain in terms of error exponent. Interestingly, the feedback thresholds can be optimized for any given codeword length T and multiplexing gain r. For simplicity we first discuss the case with $K = 2$ and a single rate, then generalize the results to $K > 2$ and adaptive rates.

Consider the following sequence of index mappings and power codebooks

$$\mathcal{I}(\mathbf{H}) = \begin{cases} 1 & \text{if } I(\mathbf{H}, P_2) < (r+\rho_2)\log \mathrm{SNR}, \\ \min\{i : i \in \{1,\ 2\},\ I(\mathbf{H}, P_i) \geq (r+\rho_i)\log \mathrm{SNR}\} & \text{otherwise,} \end{cases}$$

and

$$P_1 = \frac{\mathrm{SNR}}{2},\ P_2 = \frac{\mathrm{SNR}}{2F((r+\rho_1)\log \mathrm{SNR}, P_1)}.$$

The *back-off multiplexing gains* ρ_i's are nonnegative functions of r and T, but do not depend on SNR. By construction $P_1 \doteq \mathrm{SNR} \equiv \mathrm{SNR}^{p_1}$ and $P_2 \doteq \mathrm{SNR}^{1+D(r+\rho_1,1)} \equiv \mathrm{SNR}^{p_2}$.

Conditioned on a feedback index i, the transmit signals are constructed as

$$\mathbf{S}_i(k) = \sqrt{\frac{P_i}{N_\mathrm{t}}}\mathbf{X}_i(k),$$

where $\mathbf{X}_i(k)$ are codewords of a random codebook with i.i.d. components $\sim \mathcal{CN}(0,1)$. The error probabilities, averaged over the codebook, the channel, and the code ensembles are [ZT03, Lemma 6]

$$\Pr(\text{error}, \mathcal{I}(\mathbf{H}) = 1, I(\mathbf{H}, P_1) \geq (r+\rho_1)\log \mathrm{SNR}) \dot{\leq} \mathrm{SNR}^{-d_{\mathrm{B},1}(r)}$$

and

$$\Pr(\text{error}, \mathcal{I}(\mathbf{H}) = 2) \dot{\leq} \mathrm{SNR}^{-d_{\mathrm{B},2}(r)}$$

where

$$d_{\mathrm{B},i}(r) = \inf_{\alpha_1^n \in \mathcal{B}_i} \sum_{j=1}^{n} (2j-1+m-n)\alpha_j + T\left(\sum_{j=1}^{n}(p_i-\alpha_j)^+ - r\right),$$

and

$$\mathcal{B}_i \triangleq \{\alpha_1^n : \alpha_1 \geq \cdots \geq \alpha_n \geq 0, \sum_{j=1}^{n}(p_i-\alpha_j)^+ \geq r+\rho_i\}.$$

For $T \geq N_\mathrm{t}+N_\mathrm{r}-1$, the optimum α_j^*'s always satisfy $\sum_{j=1}^{n}(p_i-\alpha_j^*)^+ = r+\rho_i$ and thus we have

$$d_{\mathrm{B},1}(r) = D(r+\rho_1, p_1) + T\rho_1 = D(r+\rho_1, 1) + T\rho_1$$

and

$$d_{\mathrm{B},2}(r) = D(r+\rho_2, p_2) + T\rho_2 = D(r+\rho_2, 1+D(r+\rho_1,1)) + T\rho_2.$$

Furthermore, by construction,

$$\Pr(\text{outage}) \doteq \mathrm{SNR}^{D(r+\rho_2,p_2)} = \mathrm{SNR}^{D(r+\rho_2,1+D(r+\rho_1,1))} = \mathrm{SNR}^{-d_{\mathrm{out}}(r)}.$$

The error probability, which we would like to minimize, is dominated by the error event that decays the slowest as SNR grows. Thus we can optimize the back-off multiplexing gains, for each r, as follows

$$\sup_{\rho_1^2 \in [0,n-r)^2} \min\left(d_{\mathrm{B},1}(r), d_{\mathrm{B},2}(r), d_{\mathrm{out}}(r)\right).$$

Notice that $d_{\mathrm{B},1}(r)$ does not depend on ρ_2. Also,

$$d_{\mathrm{B},2}(r) = d_{\mathrm{out}}(r) + T\rho_2 \geq d_{\mathrm{out}}(r)$$

and

$$d_{\mathrm{out}}(r) = D(r+\rho_2, 1+D(r+\rho_1,1)) \leq D(r, 1+D(r+\rho_1,1))$$

where both inequalities become equalities if $\rho_2 = 0$. We conclude that $\rho_2^* = 0$ is optimal[4] and rewrite the optimization as

$$\sup_{\rho_1 \in [0,n-r)} \min\left\{D(r+\rho_1,1)+T\rho_1, D(r,1+D(r+\rho_1,1))\right\}. \tag{3.24}$$

For $T \geq N_\mathrm{t}+N_\mathrm{r}-1$ and any $r \in (0,n)$, $d_{\mathrm{B},1}(r)$ is an increasing function of ρ_1 while $d_{\mathrm{B},2}(r) = d_{\mathrm{out}}(r)$ is a decreasing function of ρ_1. The optimization over ρ_1 can thus

[4]That can be seen as a generalization of the no-CSIT case, where a back-off multiplexing gain $\rho_1^* = 0$ makes the outage and outage-free regions have the same SNR exponent, solving $\sup_{\rho_1 \in [0,n-r)} \min(d_{\mathrm{B},1}(r), d_{\mathrm{out}}(r))$.

be interpreted as an attempt to *balance* the SNR exponents of the error events over the quantization regions. Choosing ρ_1 too small leads to $d_{\mathrm{B},1}(r) < d_{\mathrm{out}}(r)$, and most errors occur when $I(\mathbf{H}, P_1) \geq r + \rho_1$ (channel is not in outage), while setting $\rho_1 > \rho_1^*$ enlarges the outage region too much.

Generalization to $K > 2$ is straightforward, and the final result is summarized in the following.

Proposition 3.1. *For $T \geq N_t + N_r - 1$ and $K \geq 2$, recursively define*

$$d_{B,k}(r) = D(r + \rho_k, 1 + d_{B,k-1}(r) - T\rho_{k-1}) + T\rho_k, \quad k = 1, \ldots, K$$

where $d_{B,0}(r) \triangleq 0$ and $\rho_0 \triangleq 0$, $\rho_K \triangleq 0$. Then, the optimal D–M tradeoff of a single-rate MIMO system with feedback resolution K is lower-bounded by

$$d^*_{B,\ K}(r) \triangleq \sup_{\rho_1^{K-1} \in [0, n-r)^{K-1}} \min\{d_{B,1}(r), \ldots, d_{B,K}(r)\}.$$

For adaptive-rate systems, we have

Proposition 3.2. *For $T \geq N_t + N_r - 1$ and $K \geq 2$, recursively define*

$$d_{B,k}(r, r_{min}) = D(r_{min} + \rho_k, 1 + d_{B,k-1}(r, r_{min}) - T\rho_{k-1}) + T\rho_k, \quad k = 2, \ldots, K$$

where $d_{B,1}(r, r_{min}) \triangleq D(r, 1)$ and $\rho_0 \triangleq 0$, $\rho_K \triangleq 0$. Then, the optimal D–M tradeoff of an adaptive-rate MIMO system with feedback resolution K is lower-bounded by

$$d^*_{B,\ K}(r, r_{min}) \triangleq \sup_{\rho_1^{K-1} \in [0, n-r) \times [0, n-r_{min})^{K-2}} \min\{d_{B,1}(r, r_{min}), \ldots, d_{B,K}(r, r_{min})\}.$$

We plot in Fig. 3.5 the back-off bounds in a 2×2 single-rate system with different codeword lengths and feedback resolution $K = 2$. In this simple example, the optimal back-off multiplexing gain and the corresponding bound can be found in closed-form. In particular, if $3 \leq T \leq 7$, the lower bound consists of two segments

$$d^*_{\mathrm{B},\ 2}(r) = \begin{cases} 2 - r + \frac{T-1}{T+3}(10 - 6r) & \text{if } r \in \left(0, 1.5 - \frac{1}{2T}\right], \\ 2 - r + \frac{T-1}{T+1}(4 - 2r) & \text{if } r \in \left(1.5 - \frac{1}{2T}, 2\right). \end{cases}$$

For $T \geq 8$, the bound consists of three segments

$$d^*_{\mathrm{B},\ 2}(r) = \begin{cases} 4 - 3r + \frac{T-3}{T+9}(16 - 12r) & \text{if } r \in (0, 1 - \frac{4}{T-3}], \\ 2 - r + \frac{T-1}{T+3}(10 - 6r) & \\ & \text{if } r \in (1 - \frac{4}{T-3}, 1.5 - \frac{1}{2T}], \\ 2 - r + \frac{T-1}{T+1}(4 - 2r) & \text{if } r \in (1.5 - \frac{1}{2T}, 2). \end{cases}$$

For higher resolution K, it is more convenient to compute the back-off bound numerically.

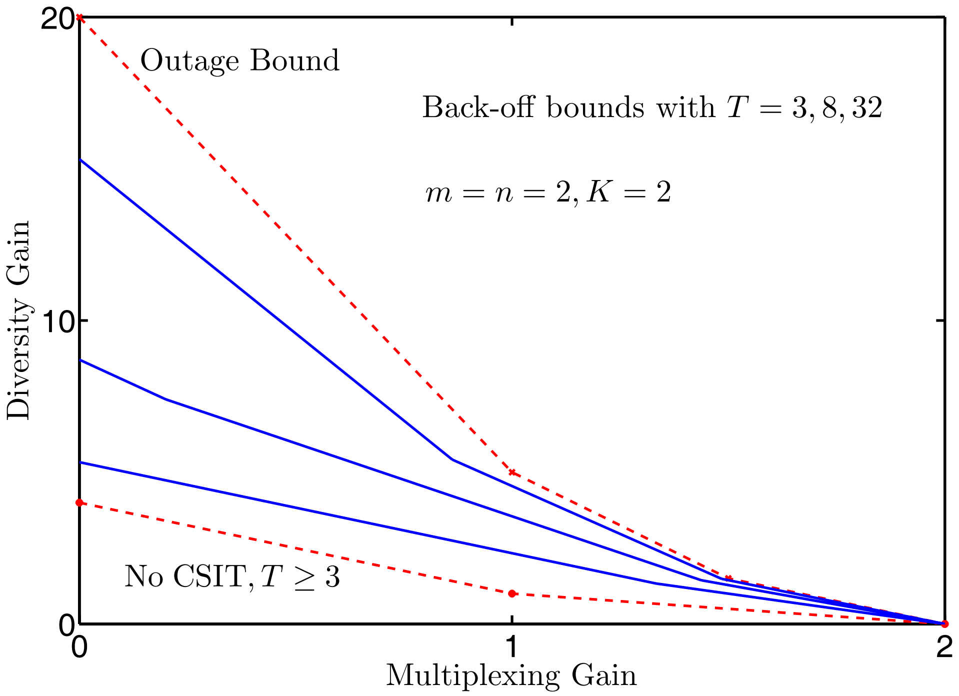

Figure 3.5: Back-off bounds over a 2×2 channel with feedback resolution $K = 2$.

The back-off bound is asymptotically tight as $T \to \infty$. We illustrate this fact with the $K = 2$ example. Recall that in this case, the back-off bound is given by (3.24), i.e.

$$d^*_{\mathrm{B},2}(r) = \sup_{\rho_1 \in [0, n-r)} \min \{D(r + \rho_1, 1) + T\rho_1, D(r, 1 + D(r + \rho_1, 1))\}.$$

By choosing a particular value $\tilde{\rho}_1 = \frac{d^*_{\mathrm{out},2}(r) - D(r,1)}{T} = \frac{D(r,1+D(r,1)) - D(r,1)}{T}$ (with a sufficiently large T so that $\tilde{\rho}_1 < n - r$) we obtain

$$d^*_{\mathrm{B},2}(r) \geq \min \left\{D(r + \tilde{\rho}_1, 1) + d^*_{\mathrm{out},2}(r) - D(r, 1), D(r, 1 + D(r + \tilde{\rho}_1, 1))\right\}.$$

But $\lim_{T\to\infty} \tilde{\rho}_1 = 0$, thus $\lim_{T\to\infty} d^*_{\mathrm{B},2} \geq \min \left\{d^*_{\mathrm{out},\ 2}(r), D(r, 1 + D(r, 1))\right\} = d^*_{\mathrm{out},\ 2}(r)$. But the lower bound $d^*_{\mathrm{B},2}(r)$ cannot exceed the outage bound $d^*_{\mathrm{out},\ 2}(r)$ for any T and we conclude that

$$\lim_{T\to\infty} d^*_{\mathrm{B},2}(r) = d^*_{\mathrm{out},\ 2}(r), \quad \forall r \in (0, n).$$

Even with moderate values of T, the back-off bounds are very tight at high rates and quickly approach the outage bound. We will tighten the bound in the low-rate region using expurgation techniques [Gal65, ZT03] in the next section.

Expurgated Bounds

Instead of backing off and trying to balance the SNR exponents of the error probability in each quantization region, we now pursue another approach inspired by the extended approximately universal condition and expurgation techniques [Gal65, ZT03]. We begin with codes drawn from a Gaussian ensemble and then expurgate bad codewords that do not satisfy the extended approximately universal condition (3.18)[5]. It turns out that expurgating all bad codewords results in lower-rate codes with high probability, thus the expurgated codes are no longer approximately universal. However, due to the presence of quantized CSIT, a combination of the expurgated codes with rate back-off yields a much improved bound, especially at low multiplexing gains.

Consider codes drawn from a random ensemble where each codeword is a matrix of size $N_{\mathrm{t}} \times T$, $T \geq N_{\mathrm{t}}$ with components $\mathcal{CN}(0,1)$. The number of codeword is $T\hat{r}\log \mathrm{SNR}$ thus the rate is $\hat{r}\log \mathrm{SNR}$ bits per use. For a given codeword $\mathbf{X}$, if there exists at least another codeword in the codebook so that $\sum_{j=1}^{n}(\beta_j)^+ > r$ where $\mu_j = \mathrm{SNR}^{-\beta_j}$ are the n smallest squared singular values of the codeword difference $\mathbf{\Delta X}$ then $\mathbf{X}$ is expurgated.

Let $q = \Pr(\sum_{j=1}^{n}(\beta_j)^+ \geq r)$ where the probability is over the code ensemble. Because $\Delta\mathbf{X}$ is a matrix of size $N_{\mathrm{t}} \times T$ with i.i.d. zero-mean complex Gaussian elements with variance 2, we readily have

$$q \doteq \mathrm{SNR}^{-\inf_{\mathcal{B}} \sum_{j=1}^{N_{\mathrm{t}}}(2j-1+T-N_{\mathrm{t}})\beta_j}$$

where

$$\mathcal{B} = \{\beta_1^{N_{\mathrm{t}}} : \beta_1 \geq \cdots \geq \beta_{N_{\mathrm{t}}} \geq 0,\ \sum_{j=1}^{n}\beta_j \geq r\}.$$

This gives $q \doteq \mathrm{SNR}^{-(T-N_{\mathrm{t}}+1)r}$ with minimizers $\beta_1^* = r$, $\beta_2^* = \cdots = \beta_{N_{\mathrm{t}}}^* = 0$, implying that the probability of a random code drawn from Gaussian ensemble not being approximately universal is dominated by the event that the smallest squared singular value of the codeword different matrix is too small.

Over the ensemble, the probability that a codeword is expurgated can be union-bounded by

$$\Pr(\mathbf{X}\text{ expurgated}) \leq \mathrm{SNR}^{T\hat{r}}q.$$

If $T\hat{r} = (T - N_{\mathrm{t}} + 1)r - \epsilon$ for any arbitrarily small $\epsilon > 0$, then

$$\Pr(\mathbf{X}\text{ expurgated}) \dot{\leq} \mathrm{SNR}^{-\epsilon}$$

meaning that we obtain a code with multiplexing gain arbitrarily close to $\hat{r} = (1 - \frac{N_{\mathrm{t}}-1}{T})r$ such that $\sum_{j=1}^{n}(\beta_j)^+ \leq r$ for all pairs of codewords.

[5]Because the codeword matrix size is fixed ($N_{\mathrm{t}} \times T$) while the number of codewords grows unbounded, codes drawn from a Gaussian ensemble will violate the power constraint imposed on *every* codeword almost surely as $\mathrm{SNR} \to \infty$. Thus one cannot discuss the approximate universality of such a random code outside our average power constraint framework, i.e., (3.17)-(3.18)

Some observations can be made from the result. First, for $N_\mathrm{t} = 1$, there exists at least an expurgated Gaussian code of length $T \geq 1$ that is approximately universal. Second, as $T \to \infty$, the expurgated codes become closer to universal. However, for any finite T and $N_\mathrm{t} \geq 2$, we obtain a code with strictly smaller rate, i.e., $\hat{r} < r$, which will not be of much use without CSIT. To combine the expurgated codes with CSIT, consider the following feedback scheme

$$\mathcal{I}_\mathrm{E}(\mathbf{H}) = \begin{cases} 1 & \text{if } I(\mathbf{H}, P_K) < \frac{r}{1-\frac{N_\mathrm{t}-1}{T}} \log \mathrm{SNR} \\ \min\{i : i \in \{1, \dots, K\},\ I(\mathbf{H}, P_i) \geq \frac{r}{1-\frac{N_\mathrm{t}-1}{T}} \log \mathrm{SNR}\} & \text{otherwise.} \end{cases}$$

Clearly, we need to constrain $\frac{r}{1-\frac{N_\mathrm{t}-1}{T}} < n$, and the intuition is that it is not possible to have expurgated codes of very high rates. Herein

$$P_1 = \frac{\mathrm{SNR}}{K},\ P_2 = \frac{\mathrm{SNR}}{KF\left(\frac{r}{1-\frac{N_\mathrm{t}-1}{T}} \log \mathrm{SNR}, P_1\right)}, \dots,$$
$$P_K = \frac{\mathrm{SNR}}{KF\left(\frac{r}{1-\frac{N_\mathrm{t}-1}{T}} \log \mathrm{SNR}, P_{K-1}\right)}.$$

The transmit signals are constructed as

$$\mathbf{S}_i(k) = \sqrt{\frac{P_i}{N_\mathrm{t}}} \mathbf{X}_\mathrm{E}(k)$$

where $\mathbf{X}_\mathrm{E}(k)$'s are the codewords of the expurgated code.

By construction $\Pr(\text{error}, \mathcal{I} = i, I(\mathbf{H}, P_i) \geq \frac{r}{1-\frac{N_\mathrm{t}-1}{T}} \log \mathrm{SNR}) \doteq \mathrm{SNR}^{-\infty}$, $\forall i$. Therefore

$$P_\mathrm{e} \dot{\leq} \mathrm{SNR}^{-d_\mathrm{out}(r)} \equiv \mathrm{SNR}^{-d_{\mathrm{E},\,K}(r)}.$$

We can define recursively

$$d_{\mathrm{E},\,k}(r) = D\left(\frac{r}{1-\frac{N_\mathrm{t}-1}{T}}, 1 + d_{\mathrm{E},\,k-1}(r)\right)$$

where $d_{\mathrm{E},\,0}(r) = 0$.

For $T \geq N_\mathrm{t} + N_\mathrm{r} - 1$, the bound can be slightly tightened. From the discussion of back-off bounds, we know that there exist codes of length $T \geq N_\mathrm{t} + N_\mathrm{r} - 1$ in the Gaussian ensemble such that $\Pr(\text{error}, \mathcal{I} = K) \doteq \mathrm{SNR}^{-d_\mathrm{out}(r)}$, $\forall r$. Thus there is no need to sacrifice rate, i.e., no need to use expurgated codes, in the region corresponding to $\mathcal{I} = K$. We summarize the derivations as follows.

Proposition 3.3. *For $k = 1, \dots, K-1$, define recursively*

$$d_{E,\,k}(r) = D\left(\frac{r}{1-\frac{N_t-1}{T}}, 1 + d_{E,\,k-1}(r)\right)$$

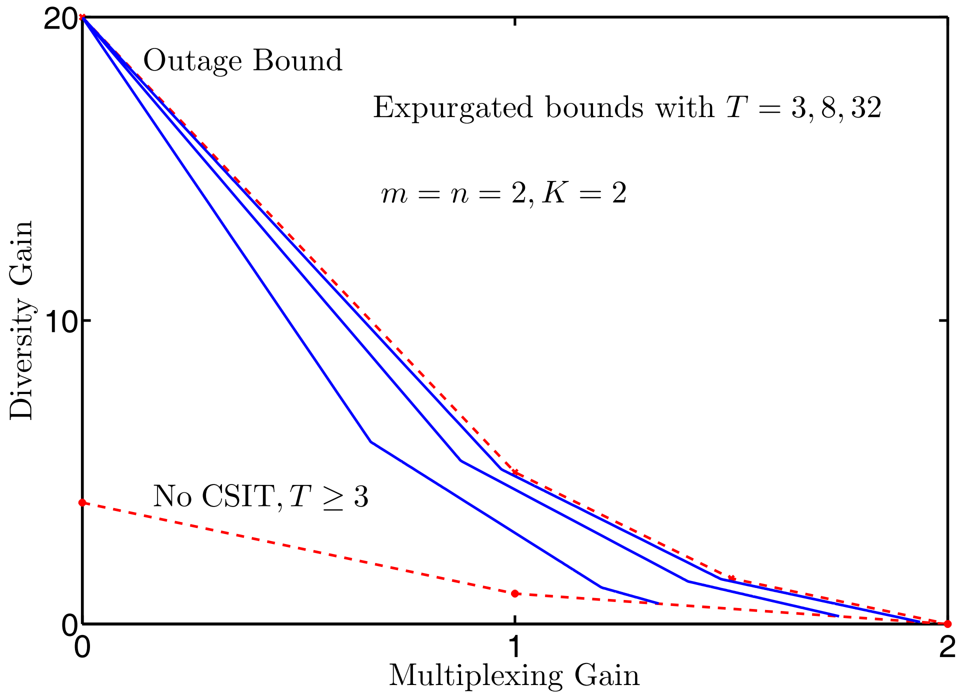

Figure 3.6: Expurgated bounds over a 2×2 channel with feedback resolution $K = 2$.

where $d_{E,\,0}(r) = 0$. *For* $r < \left(1 - \frac{N_t - 1}{T}\right) n$*, the optimal D–M tradeoff of a single-rate system with feedback resolution* K *is lower-bounded by*

$$D\left(\frac{r}{1 - \frac{N_t - 1}{T}}, 1 + d_{E,\,K-1}(r)\right), \text{ for } T \geq N_t$$

and

$$D\left(r, 1 + d_{E,\,K-1}(r)\right), \text{ for } T \geq N_t + N_r - 1.$$

We plot in Fig. 3.6 the expurgated bounds with different codeword lengths over a 2×2 channel. It is not surprising that these bounds are generally much tighter than the back-off bounds, as we optimize both the codes and the feedback link with expurgation techniques. The advantage of the back-off bound, however, is that it is defined over the entire $(0, n)$, whereas the expurgated bounds only exist for sufficiently small multiplexing gains. Notice that at low multiplexing gains, expurgating a large number of bad codeword does not cause much degradation because the code rate is already small anyway, thus the expurgated bounds are very tight even with small T. (Not surprisingly, Gallager's expurgation techniques [Gal65] were also proposed for low-rate codes.) As can be seen, for moderate codeword length T, the lower bounds quickly approach for outage bound over the entire $(0, n)$.

The same idea can be applied for adaptive-rate transmission. For brevity, we omit the derivations and summarize the results as follows.

Proposition 3.4. *For $k = 2, \ldots, K-1$, define recursively*

$$d_{E,\,k}(r, r_{min}) = D\left(\frac{r_{min}}{1 - \frac{N_t - 1}{T}}, 1 + d_{E,\,k-1}(r, r_{min})\right)$$

where $d_{E,\,1}(r, r_{min}) = D\left(\frac{r}{1-\frac{N_t-1}{T}}, 1\right)$, $\forall r_{min}$. *For* $r \in \left[r_{min}, \left(1 - \frac{N_t-1}{T}\right) n\right)$, *the optimal D–M tradeoff of an adaptive-rate system with feedback resolution K is lower-bounded by*

$$D\left(\frac{r_{min}}{1 - \frac{N_t - 1}{T}}, 1 + d_{E,\,K-1}(r, r_{min})\right), \ \textit{for } T \geq N_t$$

and

$$D\left(r_{min}, 1 + d_{E,\,K-1}(r, r_{min})\right), \ \textit{for } T \geq N_t + N_r - 1.$$

We plot both the back-off and expurgated bounds for adaptive transmission with $N_{\rm t} = N_{\rm r} = 2$, $r_{\rm min} = 0.5$ in Fig. 3.7. The back-off bounds consist of three sections connected at $r = 1 - \frac{5.5}{T}, 2 - \frac{2.5}{T}$ if $T \geq 6$, and two sections connected at $r = 2 - \frac{2.5}{T}$ for $T = 3, 4, 5$. The last segment of the back-off bound can be interpreted as the range where attempts to balance the SNR exponent in the outage and outage-free region fail. Indeed, if $r \geq 2 - \frac{2.5}{T}$ then $d^*_{{\rm B},2}(r, 0.5) = T(2-r)$ and most errors occur in the *outage-free* region. At the multiplexing gain $r = 2 - \frac{2.5}{T}$ the diversity gain is $T(2 - (2 - \frac{2.5}{T})) = 2.5$, for any $T \geq 3$. It is interesting to observe that the right-most points of expurgated bounds also seem to have this diversity gain. This can be explained as follows. At sufficiently high rate, expurgating results in a back-off threshold close to n, thus the probability that $\mathcal{I} = 2$ is in the order of ${\rm SNR}^{-\epsilon}$ for some small $\epsilon > 0$. The power P_2 therefore is in the order of ${\rm SNR}^{1+\epsilon}$ and thus the diversity of a system with expurgated Gaussian codes is $D(r_{\rm min}, 1 + \epsilon) \to 2.5$ as $\epsilon \downarrow 0$.

We may draw some conservative conclusions, keeping in mind that results derived are only lower bounds. Unlike in the no-CSIT case [ZT03], Gaussian coding arguments do not appear to be sufficient to complete the D–M tradeoff analysis. It is likely that many other randomly drawn codes also suffer from this phenomenon and therefore are not immediately suitable for the feedback model considered herein. Nevertheless, such a random code with moderate codeword lengths and/or careful expurgation may quickly approach outage bounds, when combined with a properly designed feedback scheme.

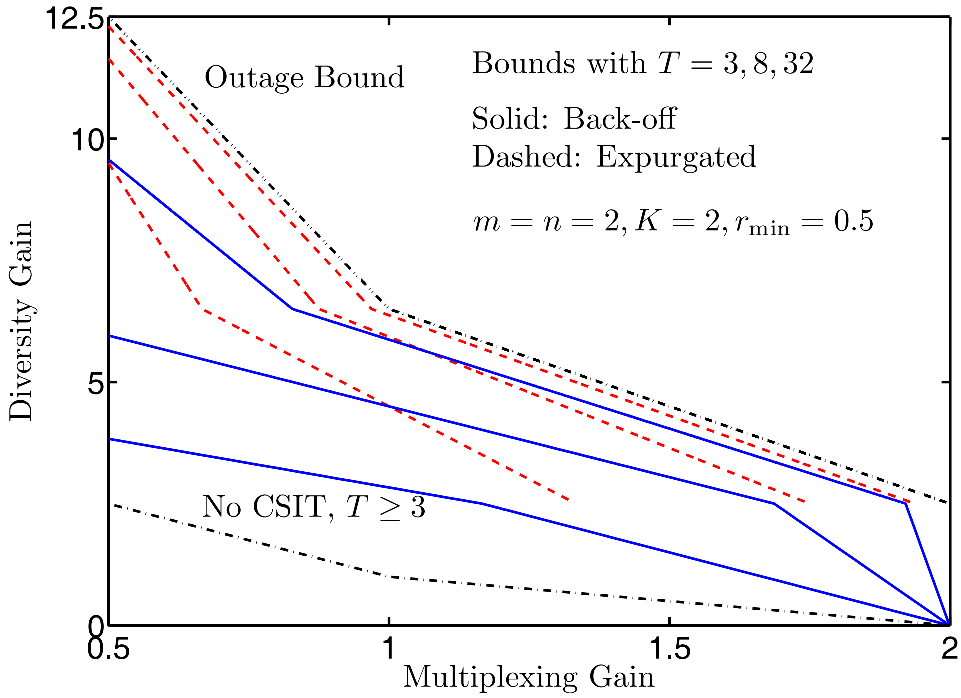

Figure 3.7: Gaussian coding bounds for adaptive-rate transmission over a 2×2 channel with feedback resolution $K = 2$, $r_{\min} = 0.5$.

3.6 Conclusion

In this chapter, we have analyzed the asymptotic behavior of MIMO systems with slow fading and quantized CSIT under the framework of the D–M tradeoff. While this is clear that an error event is very unlikely in such a high-SNR scenario, we have been able to characterize exactly the *level of unlikelihood* of that event. We have shown that, unlike in the no-CSIT case, even carefully constructed feedback schemes with Gaussian codes are not sufficient to complete the achievability part of the optimal D–M tradeoff. Instead, the optimal D–M tradeoff is shown to be achievable by combining a certain class of finite-length codes with a suitable feedback scheme. Our extended approximately universal condition gives some novel insight into the class of codes that are D–M tradeoff optimal over any channel distribution. Altogether, the results provide a better understanding on the behavior of adaptive MIMO systems in the asymptotically high SNR regime.

Appendices for Chapter 3

3.A Proof of Lemma 3.1

We first show that the optimal index mapping must have the form (3.12). Notice that for an optimal scheme, the power constraint is active, because the joint p.d.f. of the singular values of the channel matrix are continuous and takes on positive values over the entire positive orthant.

Let $\{P_i\}_{i=1}^K$ be an arbitrary power codebook and $\mathcal{I}(\mathbf{H})$ be a deterministic index mapping such that $0 \leq P_1 < \ldots < P_K$ and $\sum_{i=1}^K \Pr(\mathcal{I}(\mathbf{H}) = i)P_i \leq \mathrm{SNR}$. Consider another feedback scheme using the same power codebook and the following index mapping

$$\hat{\mathcal{I}}(\mathbf{H}) \triangleq \begin{cases} 1 & \text{if } I(\mathbf{H}, P_{\mathcal{I}(\mathbf{H})}) < R, \\ \min\{i : i \in \{1, \ldots, \mathcal{I}(\mathbf{H})\},\ I(\mathbf{H}, P_i) \geq R\} & \text{otherwise.} \end{cases}$$

This gives exactly the same outage probability as $\mathcal{I}(\mathbf{H})$ does. However, by construction, the average transmit power of the newly constructed scheme is

$$\sum_{i=1}^{K} \Pr(\hat{\mathcal{I}}(\mathbf{H}) = i)P_i \leq \sum_{i=1}^{K} \Pr(\mathcal{I}(\mathbf{H}) = i)P_i.$$

This means that we can restrict our attention to $\hat{\mathcal{I}}(\mathbf{H})$ with the following properties. Firstly, all channel realizations that are in outage are mapped to $\mathcal{I} = 1$. Secondly, a channel realization $\mathbf{H}$, if not in outage, is mapped to the smallest power in the power codebook that can "invert" $\mathbf{H}$, i.e., when this power level is applied at the transmitter, the mutual information is greater than R.

It remains to show that no set of channel realizations with strictly positive probability measure that can be inverted by some P_i, $i \geq 2$, is mapped to $\mathcal{I} = 1$. Assume the contrary, i.e., there exists a set $\mathcal{S}$ and an index $j \geq 2$ such that $\Pr(\mathbf{H} \in \mathcal{S}) = \pi_{\mathcal{S}} > 0$ and $I(\mathbf{H}, P_j) \geq R$, $\mathcal{I}(\mathbf{H}) = 1$, $\forall \mathbf{H} \in \mathcal{S}$. Then there exists a $\hat{P}_j \in (P_{j-1}, P_j)$ such that $\Pr\left(P^{\mathrm{m}}(\mathbf{H}; R_j) \in (\hat{P}_j, P_j)\right) = \pi_{\mathcal{S}}$ where $P^{\mathrm{m}}(\mathbf{H}; R)$ satisfies $I\left(\mathbf{H}, P^{\mathrm{m}}(\mathbf{H}; R)\right) = R$, i.e., the minimum power required to invert the channel matrix $\mathbf{H}$. Since all choices of $\mathcal{S}$ subject to $\Pr(\mathbf{H} \in \mathcal{S}) = \pi_{\mathcal{S}}$ are equivalent in terms of both average power and outage probability, we can consider $\mathcal{S} = \left\{\mathbf{H} : P^{\mathrm{m}}(\mathbf{H}; R) \in (\hat{P}_j, P_j)\right\}$.

Now consider another feedback scheme using the index mapping $\hat{\mathcal{I}}(\mathbf{H})$ and the power codebook $\{P_1, \ldots, P_{j-1}, \hat{P}_j, P_{j+1}, \ldots, P_K\}$. By construction, this gives the same outage probability as that obtained by the codebook $\{P_i\}_{i=1}^K$ together with $\hat{\mathcal{I}}(\mathbf{H})$. However the newly constructed scheme uses less average power, because $\hat{P}_j < P_j$, i.e. the power constraint is inactive. Thus, the optimal index mapping must have the form (3.12), which yields an outage probability of $F(R, P_K^*)$.

Finally, because $P_1^* < \cdots < P_K^*$, the event $I(P_{i-1}^*, \mathbf{H}) > R$ also implies $I(P_i^*, \mathbf{H}) > R$. Thus

$$\begin{aligned}\Pr\left(I(P_i^*, \mathbf{H}) > R, I(P_{i-1}^*, \mathbf{H}) < R\right) &= \Pr\left(I(P_{i-1}^*, \mathbf{H}) < R\right) - \Pr\left(I(P_i^*, \mathbf{H}) < R\right) \\ &= F\left(R, P_{i-1}^*\right) - F\left(R, P_i^*\right).\end{aligned}$$

Therefore the average power of the system is given by

$$\left[F\left(R, P_K^*\right) + 1 - F\left(R, P_1^*\right)\right] P_1^* + \sum_{i=2}^{K} [F(R, P_{i-1}^*) - F(R, P_i^*)] P_i^*.$$

Since $F(R, P_K)$ is a monotonically decreasing function of P_K for any given $R > 0$, the optimal power codebook is the solution to (3.11). □

3.B Proof of Theorem 3.1

First we derive an upper bound on the SNR exponent of the largest power level in the optimal power codebook, i.e., P_K^*. Let $\{\overline{P}_i\}$ be the solution to the following optimization problem, which is a relaxed version of (3.11),

$$\begin{aligned} \max \quad & P_K \\ \text{s.t.} \quad & [F(R, P_K) + 1 - F(R, P_1)] P_1 \le \text{SNR}, \\ & [F(R, P_{i-1}) - F(R, P_i)] P_i \le \text{SNR}, \quad i \ge 2, \\ & 0 \le P_1 < \cdots < P_K. \end{aligned} \tag{3.25}$$

Clearly $\overline{P}_K \ge P_K^*$ due to relaxation. Note that the constraints of (3.25) imply $\sum_{i=1}^{K} \frac{\text{SNR}}{\overline{P}_i} \ge 1$. We must have $\overline{P}_1 \le K\text{SNR}$ otherwise $\sum_{i=1}^{K} \frac{\text{SNR}}{\overline{P}_i} < K\frac{1}{K} = 1$. Because K is a finite constant, we have $\overline{P}_1 \dot{\le} \text{SNR}$. An application of Lemma 3.2 leads to $F(R, \overline{P}_1) \dot{\ge} \text{SNR}^{-D(r,1)} = \text{SNR}^{-d^*_{\text{out},\,1}(r)}$.

The constraints of (3.25) also require $\frac{\text{SNR}}{\overline{P}_2} + F(R, \overline{P}_2) \ge F(R, \overline{P}_1)$, leading to

$$\frac{\text{SNR}}{\overline{P}_2} + F(\overline{P}_2) \dot{\ge} \text{SNR}^{-d^*_{\text{out},\,1}(r)}. \tag{3.26}$$

For any $\epsilon > 0$, if $\overline{P}_2 \doteq \text{SNR}^{1+d^*_{\text{out},\,1}(r)+\epsilon}$ then

$$\frac{\text{SNR}}{\overline{P}_2} + F(\overline{P}_2) \doteq \text{SNR}^{-d^*_{\text{out},\,1}(r)-\epsilon} + \text{SNR}^{-D(r,1+d^*_{\text{out},\,1}(r)+\epsilon)},$$

which contradicts to (3.26) because $D(r, 1 + d^*_{\text{out},\,1}(r) + \epsilon) > D(r, 1) = d^*_{\text{out},\,1}(r)$. Therefore we must have $\overline{P}_2 \dot{\le} \text{SNR}^{1+d^*_{\text{out},\,1}(r)}$, and thus

$$F(R, \overline{P}_2) \dot{\ge} \text{SNR}^{-D(r,1+d^*_{\text{out},\,1}(r))} = \text{SNR}^{-d^*_{\text{out},\,2}(r)}.$$

By induction, we eventually obtain $\overline{P}_K \dot{\leq} \mathrm{SNR}^{1+d^*_{\mathrm{out},\,K-1}(r)}$ and

$$F(R, P^*_K) \geq F(R, \overline{P}_K) \dot{\geq} \mathrm{SNR}^{-D(r, 1+d^*_{\mathrm{out},\,K-1}(r))} = \mathrm{SNR}^{-d^*_{\mathrm{out},\,K}(r)}.$$

Finally, a lower bound on P^*_K is obtained by choosing $\underline{P}_1 = \frac{\mathrm{SNR}}{K}$, $\underline{P}_2 = \frac{\mathrm{SNR}}{KF(\underline{P}_1)}$, …, $\underline{P}_K = \frac{\mathrm{SNR}}{KF(\underline{P}_{K-1})}$. Since these $\underline{P}_i$'s satisfy the constraints of (3.11), we have $P^*_K \geq \underline{P}_K$. Because K is a finite constant and by construction, $\underline{P}_1 \doteq \mathrm{SNR}$, $\underline{P}_2 \doteq \mathrm{SNR}^{1+d^*_{\mathrm{out},\,1}(r)}$, …, $\underline{P}_K \doteq \mathrm{SNR}^{1+d^*_{\mathrm{out},\,K-1}(r)}$. Thus

$$F(R, P^*_K) \leq F(R, \underline{P}_K) \doteq \mathrm{SNR}^{-d^*_{\mathrm{out},\,K}(r)}.$$

This concludes the proof. □

3.C Proof of Theorem 3.2

Similarly to the single-rate case, $\mathbf{Q}_i = \frac{P_i}{N_t}\mathbf{I}_{N_t}$, $\forall i \in \{1, \ldots, K\}$, can be assumed without loss of generality.

Consider an arbitrary sequence of deterministic feedback schemes $\mathcal{F}$ that provides a multiplexing gain of r. Let the outage probability $P_{\mathrm{out},\mathcal{F}} \doteq \mathrm{SNR}^{-d_{\mathrm{out},\mathcal{F}}(r)}$. We now derive an upper bound on $d_{\mathrm{out},\mathcal{F}}(r)$. To that end, let p_i, $i = 1, \ldots, K$, be positive real numbers such that $P_i \doteq \mathrm{SNR}^{p_i}$. From (3.3), we have

$$\sum_{i=1}^{K} \Pr(\mathcal{I} = i)\mathrm{SNR}^{p_i} \dot{\leq} \mathrm{SNR}. \tag{3.27}$$

Without loss of generality, assume that $0 < p_1 \leq \cdots \leq p_K$. Let $\bar{K}$ be the integer in $\{1, \ldots, K\}$ such that $p_{\bar{K}} \leq 1$ and $p_{\bar{K}+1} > 1$. Herein we use the convention $p_{K+1} = \infty$. Such a $\bar{K}$ must exist, otherwise (3.27) cannot be satisfied. Note that (3.27) implies $\Pr(\mathcal{I} = i)\mathrm{SNR}^{p_i} \dot{\leq} \mathrm{SNR}$, $\forall i$. Thus $\Pr(\mathcal{I} = i) \dot{\leq} \mathrm{SNR}^{1-p_i}$, and

$$\lim_{\mathrm{SNR}\to\infty} \frac{\Pr(\mathcal{I} = i)R_i}{\log \mathrm{SNR}} = r_i \lim_{\mathrm{SNR}\to\infty} \Pr(\mathcal{I} = i) = 0, \forall i \geq \bar{K} + 1,$$

meaning that all regions using a power whose SNR exponent is strictly larger than 1, if any, contribute nothing to the multiplexing gain (3.5). In our context, $\bar{K}$ can be interpreted as the number of regions that contribute considerably to the overall throughput, while $K - \bar{K}$ can be seen as the number of regions that are added to improve the overall reliability.

Consider the first $\bar{K}$ regions of $\mathcal{F}$, with power levels having dominant SNR exponent less than or equal to 1, and let $r_{\max} = \max(r_1, \ldots, r_{\bar{K}})$. By definition (3.5), $r \leq r_{\max}$. Let us consider another sequence of feedback schemes $\hat{\mathcal{F}}$ using $\hat{r}_1 = r < r_{\max}$, $\hat{r}_i = r_{\min} \leq r_i$, $\forall i \geq 2$ and having $\hat{P}_1 \doteq \mathrm{SNR}$. Let $P_{\mathrm{out},\hat{\mathcal{F}}} \doteq \mathrm{SNR}^{-d_{\hat{\mathcal{F}}}(r)}$. Outage probability is an increasing function of rate and decreasing function of power, thus *maximizing* $d_{\hat{\mathcal{F}}}(r)$ yields an upper bound to $d_{\mathcal{F}}(r)$.

If a multiplexing gain of r is achievable even with an outage-minimizing $\hat{\mathcal{F}}$, then the lower bound is tight.

We claim that to minimize the SNR exponent of outage, it suffices to consider $\hat{\mathcal{F}}$ such that

$$\hat{p}_i > 1, \quad \forall i \geq 2 \tag{3.28}$$

where $\hat{P}_i \doteq \mathrm{SNR}^{\hat{p}_i}$. Assume the contrary, then at least $\hat{P}_2 \doteq \mathrm{SNR}$. Then $d_{\hat{\mathcal{F}}}(r)$ can be upper-bounded by $d^*_{K-1}(r_{\min})$, i.e., the outage exponent of a single-rate system with rate $r_{\min} \log \mathrm{SNR}$ and a *coarser* resolution of $K-1$. It is easy to show that such an upper bound is strictly below the achievable $d_{\hat{\mathcal{F}}}(r)$ when assuming (3.28), derived below.

Assume (3.28) holds, i.e., $\bar{K} = 1$. For sufficiently high SNR, (3.28) implies $\hat{P}_1 < \hat{P}_i$, $\forall i > 2$. We can characterize the optimal index mapping and power codebook with the following lemma, which is a straightforward generalization of Lemma 3.2.

Lemma 3.3. *Consider a system employing some given rates $R_1 > R_2 = \cdots = R_K$. It is constrained that $0 \leq P_1 < \cdots < P_K$. The outage-minimizing power codebook $\{P_i^*\}_{i=1}^K$ solves*

$$\begin{aligned} \max \quad & P_K \\ \textit{s.t.} \quad & [F(R_K, P_K) + 1 - F(R_1, P_1)]\, P_1 + \sum_{i=2}^{K} [F(R_{i-1}, P_{i-1}) - F(R_i, P_i)]\, P_i \leq \mathit{SNR}. \end{aligned} \tag{3.29}$$

The optimal index mapping is given by

$$\mathcal{I}^*(\mathbf{H}) = \begin{cases} 1 & \textit{if}\, I(\mathbf{H}, P_K^*) < R_K, \\ \min\{i : i \in \{1, \ldots, K\},\ I(\mathbf{H}, P_i^*) \geq R_i\} & \textit{otherwise.} \end{cases} \tag{3.30}$$

The minimum outage probability is

$$P_{out}^* = F(R_K, P_K^*).$$

The proof is almost identical to that of Lemma 3.2 and therefore omitted for brevity. Similarly to Appendix 3.B, computing the SNR exponent of P_K^* that solves (3.29) gives the optimal $d^*_{\hat{\mathcal{F}}}(r) = d^*_{\mathrm{out},K}(r, r_{\min})$.

Recall that $d^*_{\hat{\mathcal{F}}}(r)$ is only an upper bound on $d_{\mathcal{F}}(r)$ because the multiplexing gain of $\hat{\mathcal{F}}$ may actually be smaller than r due to construction. To show the tightness of the bound, consider the following power levels for $\hat{\mathcal{F}}$

$$\underline{P}_1 = \frac{\mathrm{SNR}}{K}, \underline{P}_2 = \frac{\mathrm{SNR}}{KF(R_1, \underline{P}_1)}, \ldots, \underline{P}_K = \frac{\mathrm{SNR}}{KF(R_{K-1}, \underline{P}_{K-1})},$$

which satisfy (3.28), and also achieve $d^*_{\hat{\mathcal{F}}}(r)$. It is not difficult to verify that

$$\lim_{\mathrm{SNR}\to\infty} \Pr(\mathcal{I} = 1) = \lim_{\mathrm{SNR}\to\infty} [F(R_K, \underline{P}_K) + 1 - F(R_1, \underline{P}_1)] = 1,$$

and $\lim_{\text{SNR}\to\infty} \Pr(\mathcal{I} = i) = 0$, $\forall i > 1$. By definition (3.5), the multiplexing gain of such a scheme is exactly r. □

3.D Towards the Necessity of (3.18)

In this section, we will show that, if a sequence of codes $\mathcal{C}$ satisfies

$$\liminf_{\text{SNR}\to\infty} \frac{\left(\min_{\mathcal{C}} \prod_{j=1}^{n} \text{SNR}^{-(\beta_j)^+}\right)}{\log \text{SNR}} = -\hat{r} \tag{3.31}$$

for some $\hat{r} > r$, then for a certain sequence of channel distributions, there exist channels in the ϵ-outage-free region that make the maximum pairwise error probability bounded away from zero even as SNR $\to \infty$. Furthermore, the probability measure of such bad channels set decays only polynomially with SNR.

To that end, consider

$$\overline{\mathcal{O}}^{\epsilon} = \{\mathbf{H} : \ I(\mathbf{H}, \text{SNR}) \geq (r + \epsilon) \log \text{SNR}\}.$$

If the condition (3.31) holds, then for arbitrarily small $\delta > 0$, we can find $\overline{\text{SNR}}$ so that for any SNR $> \overline{\text{SNR}}$, there exists a codeword differences $\boldsymbol{\Delta}\mathbf{X}_{\text{B}}$ with squared singular values $\mu_j = \text{SNR}^{-\beta_j}$ such that $\sum_{j=1}^{n} (\beta_j)^+ \geq \hat{r} - \delta$. For sufficiently small δ, then $\hat{r} - \delta > r > 0$, thus at least $\beta_1 > 0$.

Choose a sequence of distributions that always align the singular values λ_j of the channel matrix to those of $\boldsymbol{\Delta}\mathbf{X}_{\text{B}}$ in the worst order, i.e., $\|\mathbf{H}\boldsymbol{\Delta}\mathbf{X}_{\text{B}}\|_{\text{F}}^2 = \sum_{j=1}^{n} \lambda_j \mu_{n+j-1}$, $\forall \mathbf{H}$. For convenience, also choose λ_1^n to be Wishart distributed (which can be made independent of the choice of the worst-case rotation). Then

$$\begin{aligned}
&\max_{\mathcal{C}} \Pr(\text{pairwise error}, \mathbf{H} \in \overline{\mathcal{O}}^{\epsilon}) \\
&\geq \Pr(\text{pairwise error}, \boldsymbol{\Delta}\mathbf{X}_{\text{B}}, \mathbf{H} \in \overline{\mathcal{O}}^{\epsilon}) \\
&= \int_{\overline{\mathcal{O}}^{\epsilon}} Q\left(\sqrt{\frac{\text{SNR}}{2N_{\text{t}}} \sum_{j=1}^{n} \lambda_j \mu_{n-j+1}}\right) f(\lambda_1^n) d\lambda_1^n \\
&\geq \frac{1}{2\sqrt{\pi}} \int_{\overline{\mathcal{O}}^{\epsilon}} \frac{\exp\left(-\frac{\text{SNR}}{4N_{\text{t}}} \sum_{j=1}^{n} \lambda_j \mu_{n-j+1}\right)}{1 + \sqrt{\frac{\text{SNR}}{2N_{\text{t}}} \sum_{j=1}^{n} \lambda_j \mu_{n-j+1}}} f(\lambda_1^n) d\lambda_1^n \\
&\doteq \int_{\overline{\mathcal{O}}^{\epsilon}} \frac{1}{1 + \sqrt{\sum_{j=1}^{n} \text{SNR}^{1-\alpha_j-\beta_{n-j+1}}}} \exp\left(-\sum_{j=1}^{n} \text{SNR}^{1-\alpha_j-\beta_{n-j+1}}\right) f(\lambda_1^n) d\lambda_1^n \\
&\doteq \int_{\mathcal{B}} f(\alpha_1^n) d\alpha_1^n + \text{SNR}^{-\infty}
\end{aligned}$$

where the second inequality is due to the fact that

$$Q(x) \geq \frac{1}{2\sqrt{\pi}} \frac{\exp(-x^2/2)}{1+x}, \ \forall x \geq 0.$$

The bad channels set $\mathcal{B}$ where the pairwise error probability is in the order of SNR^0 is given as

$$\begin{aligned}\mathcal{B} = &\{\alpha_1^n : \alpha_1 \geq \cdots \geq \alpha_n \geq 0\} \cap \{\alpha_1^n : 1 - \alpha_j < \beta_{n-j+1}\} \\ &\cap \left\{ \alpha_1^n : \sum_{j=1}^{n} (1-\alpha_j)^+ \geq r + \epsilon \right\}.\end{aligned}$$

Let $\bar{r} = \min(1, \hat{r} - \delta)$. Since $\beta_1 > 0$ $\forall \text{SNR} > \overline{\text{SNR}}$, the function $\sum_{j=1}^{n}(1-\alpha_j)^+$ is continuous and assumes all values on $[0, \bar{r})$ for the α_1^n's in the intersection of the first two sets defining $\mathcal{B}$. Thus for any $r \in (0, \bar{r})$, $\mathcal{B}$ can be made nonempty by choosing, e.g.,

$$\epsilon = \frac{\bar{r} - r}{2}.$$

With that choice of r and ϵ, over $\mathcal{B}$, the function $\sum_{j=1}^{n}(2j-1+m-n)\alpha_j$ is continuous and bounded below, and Varadhan's integral lemma [DZ98] can be applied to show that the probability that $\mathbf{H} \in \mathcal{B}$ is dominated by the term

$$\text{SNR}^{-\inf_{\mathcal{B}} \sum_{j=1}^{n}(2j-1+m-n)\alpha_j}$$

as $\text{SNR} \to \infty$. Thus there exists a set of "bad" channels in the ϵ-outage-free region that make the largest pairwise error probability in the order of SNR^0 (bounded below by a positive constant). Furthermore the probability measure of that set does not decay exponentially. This makes the largest pairwise error probability averaged over $\overline{\mathcal{O}}^\epsilon$ decay sub-exponentially.

Chapter 4

D–M Tradeoff in Decode–and–Forward Relay Channels

This chapter studies the problem of resource allocation to maximize the outage exponent over a decode-and-forward fading relay channel with quantized channel state feedback (CSF). Three different scenarios are considered: relay-source, destination-relay, and destination-source CSF. It is found that using just one bit of CSF from the relay to control the source transmit power is sufficient to achieve the multi-antenna upper bound in a range of multiplexing gains, with fixed-length codes, i.e., with coding schemes significantly simpler than dynamic DF. Systems with CSF from destination to control relay transmit power slightly outperform DDF at high multiplexing gains, even with one bit of feedback. Finally, with CSF from destination, if the source-relay channel gain is unknown to the feedback quantizer at the destination, the diversity gain only grows linearly in the number of feedback levels K, in sharp contrast to an exponential growth for multiantenna channels. In this last scenario, a simple scheme is shown to perform close to the corresponding upper bound.

4.1 Introduction

Motivated by the potential of having simple single-antenna communication nodes cooperate and approach the promising performance of multiantenna systems, there has been a renewed interest in the classical relay channels [van71, CE79], for example in the recent work [SEA03a, SEA03b, LW03, LTW04, NBK04, KGG05]. Resource allocation for relay channels is known to enhance the performance significantly in many different scenarios [HA04, MY04, LC05, GE07a, HZ05, LV05, LVP07]. Most previous work on resource allocation, however, assumed perfect network state information at both the source and the relay. In [AKSA06], power control for AF relaying with quantized feedback from the destination is considered, but a diversity analysis is not pursued.

For the relay channel, the D–M tradeoff curves of some baseline schemes are obtained in [LTW04, PV04]. A sophisticated scheme named dynamic decode-and-forward is proposed in [AES05] and shown to achieve the multi-antenna upper bound at all multiplexing gains less than one-half. The D–M tradeoff over AF relay channels is extensively treated in [AES05, YB07a, YB07b]. From a diversity-multiplexing tradeoff viewpoint, CF relaying is shown to be optimal [YE07], under the critical assumption that the relay knows the *full* CSI. In the recent work [EVAK06], optimizing the dimension allocation for DF using only the statistical knowledge of the channel is considered.

This chapter considers a three-node half-duplex cooperative communication channel subject to very slowly-varying, but random, channel gains (quasi-static fading), under different forms of *heavily quantized* CSF. The partial CSF is used to allocate the number of channel uses in the two phases of the DF protocol, i.e., dimension allocation, and also to control the transmit power *across fading states*. Both orthogonal (source and relay do not transmit simultaneously) and nonorthogonal (source and relay can transmit at the same time) schemes are considered.

Three different feedback scenarios are considered: relay-to-source CSF, destination-to-source-and-relay CSF, and destination-to-relay CSF. The last case is motivated by the fact that in certain scenarios, the feedback link from the destination to the source is of significantly lower quality than that from the destination to the relay (which inspires the relaying model in the first place). The possibilities that we do *not* study are the destination-to-source CSF case and the scenario of *joint feedback* from both the relay and the destination. The former case is omitted because of the relative distances between the nodes in practice. It is unrealistic that the feedback from the destination is reliably received by the source but not by the relay, due to the broadcast nature of wireless communications. As for the latter case, we chose to study only separate feedback from either the relay or the destination to simplify the analysis and also to get insight into the actual value of each individual feedback type.

As in Chapter 3, we study the asymptotically high SNR regime in terms of the D–M tradeoff. A summary of our findings in this chapter is given in the following. Over a statistically symmetric Gaussian decode-and-forward relay channel (the channel is described more precisely later) with

Relay-to-source CSF:

- With dimension allocation only, performance gains over no-CSF schemes can be achieved even with heavily quantized CSF from the relay to the source. Nonorthogonal schemes provide additional, but not very significant gains over an orthogonal system.
- With dimension allocation and power control at the source, even *one* bit of CSF is sufficient to achieve the fully cooperative upper bound over a wide range of multiplexing gains.

- Relay-to-source CSF systems quickly approach the performance of DDF [AES05] as the quality of CSF increases. Our results suggest the use of an orthogonal DF scheme with as few as one bit of CSF from the relay, in combination with power control at the source.

Destination-to-relay CSF: Our proposed nonorthogonal scheme with power control at the relay slightly *outperforms DDF* for large values of multiplexing gain. Furthermore, the performance of one bit CSF and perfect CSF is virtually indistinguishable. The results first imply that with perfect CSF, this DF and power control scheme is much less efficient than a compress-and-forward approach [YE07]. Second, with limited CSF, our results suggest the use of just one bit to control relay transmit power for *high-rate* DF systems.

Destination-to-source-and-relay CSF:

- By developing a novel upper bound to the diversity gain for channels with *restricted* CSF (see Section 4.6), we show that under the relatively realistic assumption that the source-relay channel gain is unknown to the feedback quantizer at the destination, the diversity gain only grows *linearly* in the number of feedback levels (also referred to as feedback resolution). In contrast, the diversity gain of a MISO channel grows *exponentially* in the feedback resolution K.
- A proposed scheme is shown to perform relatively close to the new upper bound.

The results above are obtained in the limit of infinitely large block lengths. We then discuss the achievable tradeoffs with *finite-length* codes, arguing that in the orthogonal case, "good" codes designed for parallel channels are also suitable for the DF relay scenario. Inspired by the concept of approximate universality [TV06] we next present sufficient conditions for finite-length codes to achieve the optimal exponent. By expurgating codes drawn from Gaussian ensemble to match the new sufficient conditions, we show the existence of codes that approach the outage exponent for all multiplexing gains even at moderate block lengths. In this scenario, the loss in performance of finite block length schemes comes from the discreteness of the possible dimension allocation ratios, rather than from the length of the codes.

4.2 System Model

Consider the complex baseband model of a frequency-nonselective fading relay channel in Fig. 4.1. The channel is assumed to be slowly fading, i.e., the channel gains are constant during a fading block consisting of T channel uses, but changes independently from one block to the next. We exclusively consider the case when a transmission codeword spans a single fading block to identify the gain of spatial cooperation. The channel is assumed to be statistically symmetric. In particular, the channel gains between source-destination, source-relay and relay-destination h, h_1, and h_2 are i.i.d. complex Gaussian random variables with zero mean and unit variance. Let $g = |h|^2$, $\gamma_1 = |h_1|^2$ and $\gamma_2 = |h_2|^2$. Assume perfect channel state

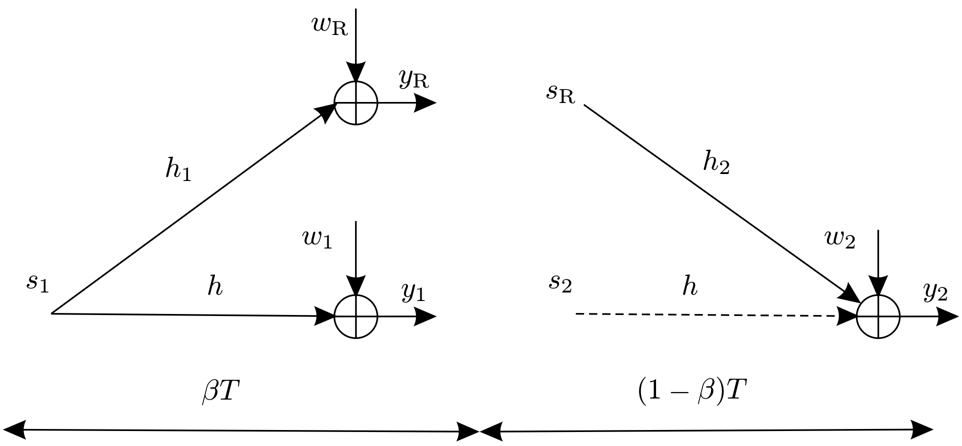

Figure 4.1: System model.

information at the receiver of each communication link. We also assume perfect synchronization.

We consider a half-duplex channel, where the relay cannot transmit and receive simultaneously. Communication between source and destination takes place in two phases. In the learning phase, the source transmits, the relay and the destination listen. In the relaying phase the relay transmits based on what it has learned and the source may transmit more symbols but no new message (in nonorthogonal schemes), or remain silent (in orthogonal schemes). The destination attempts to decode based on the entire signals received during both phases.

The received signals at time instant n during the learning phase are given by

$$\begin{aligned} y_1(n) &= hs_1(n) + w_1(n) \\ y_R(n) &= h_1 s_1(n) + w_R(n). \end{aligned} \tag{4.1}$$

In the relaying phase, the received signal at the destination is

$$y_2(n) = hs_2(n) + h_2 s_R(n) + w_2(n) \tag{4.2}$$

where for an orthogonal scheme $s_2 = 0$ (this will be discussed in more detail later on). The noises w_1, w_2, w_R are mutually independent temporally white complex Gaussian with zero mean and unit variance.

Assume individual power constraints at the source and the relay. We consider both short-term and long-term power constraints [CTB99], which are specified in more detail later. We will also refer to systems under a short-term (respectively long-term) power constraint as ones without power control (respectively, with power control).

At each value of SNR, the system is designed to serve a fixed rate of $r \log \text{SNR}$ bits per channel use (which is independent of the CSI at the transmitter), where

$r \in (0, 1)$ is the multiplexing gain. That is, no rate control is possible. Assume that all three links in the model have the same SNR. This is not too restrictive because, as long as the SNR's of the three links scale in the same order, the diversity–multiplexing tradeoff does not change. Recall from Chapter 3 that the system has a diversity gain of d if the probability of error at the destination satisfies

$$\Pr(\varepsilon) \doteq \mathrm{SNR}^{-d}.$$

We are interested in characterizing d as a function of r for different feedback schemes over the channel. In Sections 4.3-4.6, we focus on the SNR exponent of the outage probability $P_{\mathrm{out}}(r \log \mathrm{SNR})$, also referred to as the *outage exponent*. Herein outage is defined as the event that the relevant instantaneous mutual information (given a particular scheme) is smaller than the data rate $r \log \mathrm{SNR}$. One should exercise some extra care when interpreting the meaning of the outage exponents in the current work. The outage exponents discussed herein do not serve as a universal upper bound to the diversity-multiplexing tradeoff as in the multi-antenna case, as finding the capacity of the general relay channel is a long-standing open problem.[1] Rather, they represent the achievable tradeoffs of specific schemes. We can achieve these tradeoffs by using standard Gaussian coding arguments, *first* letting the codeword length $T \to \infty$ for each SNR and *then* considering a sequence of schemes with increasing SNR. Achievability results with finite-length codes will be presented and discussed later in Section 4.7.

4.3 Decode-and-Forward without CSF

When the source and relay have no information about their corresponding instantaneous forward channel gains, the system can still optimize the SNR exponent of the outage probability using the knowledge about the statistics of the channel.

This optimization for DF with multiple relays has been solved in [EVAK06]. We summarize the results herein along with a sketch of the proof in order to introduce some important concepts and help facilitate the presentation of more sophisticated CSF schemes. In addition, we will later present novel results on *finite-length* codes in Section 4.7, and show that this practical constraint changes the picture considerably.

In this simple setting, the learning phase uses $T_1 = \beta T$ channel uses and the relaying phase uses $T_2 = (1 - \beta)T$ channel uses where $0 < \beta < 1$ is the fraction of dimension allocated that we want to optimize over. The source encodes an equally likely message $m \in \{1, \ldots, 2^{rT \log \mathrm{SNR}}\}$ to a sequence $\{s_1(1), \ldots, s_1(T_1), s_2(1), \ldots, s_2(T_2)\}$ belonging to a codebook $\mathcal{C}_1$ where $s_2(1) = \cdots = s_2(T_2) = 0$ for

[1] With an exception for the cooperative bounds, which are universal upper bounds for any possible relaying scheme.

an orthogonal scheme. The power constraint over the codebook at the source reads

$$\frac{1}{|\mathcal{C}_1|T}\sum_{\mathcal{C}_1}\left(\sum_{n=1}^{T_1}|s_1(n)|^2+\sum_{n=1}^{T_2}|s_2(n)|^2\right)\leq \mathrm{SNR}.$$

If the source-relay link is not in outage, the relay attempts to decode m in full[2], producing $m_{\mathrm{R}}\in\{1,\ldots,2^{rT\log\mathrm{SNR}}\}$ and then re-encodes m_{R} to $\{s_{\mathrm{R}}(1),\ldots,s_{\mathrm{R}}(T_2)\}$ belonging to a codebook $\mathcal{C}_{\mathrm{R}}$, under the constraint

$$\frac{1}{|\mathcal{C}_{\mathrm{R}}|T}\sum_{\mathcal{C}_{\mathrm{R}}}\sum_{n=1}^{T_2}|s_{\mathrm{R}}(n)|^2\leq \mathrm{SNR}.$$

Note that $|\mathcal{C}_1|=|\mathcal{C}_{\mathrm{R}}|=2^{rT\log\mathrm{SNR}}$. The destination decodes to obtain $\hat{m}\in\{1,\ldots,2^{rT\log\mathrm{SNR}}\}$ based on $\{y_1(1),\ldots,y_1(T_1),\ y_2(1),\ldots,y_2(T_2)\}$. If the source-relay link is in outage, the relay outputs nothing and the destination decodes based on $\{y_1(1),\ldots,y_1(T_1)$. Note that in the nonorthogonal case, no new *message* is communicated from the source in the relaying phase. We assume that the destination knows if the relay transmits or not in the relaying phase. (This is implicitly assumed in most previous work on DF protocols and can be done, for example, by using an all-zero training sequence to mimic the event $h_2=0$.)

Proposition 4.1. *[EVAK06] With dimension allocation, the optimal outage exponent of an orthogonal DF system with no CSF is*

$$D_O^{NF}(r)=\begin{cases}2-3r & \text{if } r<\frac{1}{3},\\ \frac{2-2r}{1+r} & \text{otherwise.}\end{cases}\tag{4.3}$$

With dimension allocation, the optimal outage exponent of a nonorthogonal DF system with no CSF is

$$D_{NO}^{NF}(r)=\begin{cases}2-\frac{3+\sqrt{5}}{2}r & \text{if } r<\frac{3-\sqrt{5}}{2},\\ (1-r)(2-r) & \text{otherwise.}\end{cases}\tag{4.4}$$

We now present a sketch of the proof. The following convenient lemma will be used throughout this chapter. The proof is straightforward and thus omitted.

Lemma 4.1. *For $0<r,\beta<1$ and $p\geq 1$, $q\geq 1$ satisfying $\beta p\geq r$, let*

$$\mathcal{A}=\left\{(\alpha_1,\alpha_2)\in\mathbb{R}_+^2:\beta(p-\alpha_1)^+ +(1-\beta)(q-\alpha_2)^+<r\right\}.$$

Then

$$\inf_{\mathcal{A}}(\alpha_1+\alpha_2)=\begin{cases}p+q-\frac{r}{\beta} & \text{if } \beta<\frac{1}{2},\\ p+q-\frac{r}{1-\beta} & \text{if } \frac{1}{2}\leq\beta<1-\frac{r}{q},\\ p-q+\frac{q-r}{\beta} & \text{if } \beta\geq\max\left(\frac{1}{2},1-\frac{r}{q}\right).\end{cases}\tag{4.5}$$

[2] That excludes, for example, the use of symbol-by-symbol decoding or multi-layer coding.

We now briefly present the proof for the orthogonal part of Proposition 4.1 (the nonorthogonal case can be proved similarly). The event that the relay is in outage *and* the source-destination link is in outage is given by

$$\mathcal{O}_1 = \{(\beta \log(1+\mathrm{SNR}\gamma_1) < r \log \mathrm{SNR}) \cap (\beta \log(1+\mathrm{SNR}g) < r \log \mathrm{SNR})\}\,.$$

The event that the relay is not in outage, but the effective channel from source to destination is still in outage is given by

$$\begin{aligned}\mathcal{O}_2 = \{&(\beta \log(1+\mathrm{SNR}\gamma_1) \geq r \log \mathrm{SNR}) \\ &\cap\ (\beta \log(1+\mathrm{SNR}g) + (1-\beta)\log(1+\mathrm{SNR}\gamma_2) < r \log \mathrm{SNR})\}\,.\end{aligned}$$

The overall outage probability is thus

$$P_{\mathrm{out}} = \Pr(\mathcal{O}_1 \cup \mathcal{O}_2) = \Pr(\mathcal{O}_1) + \Pr(\mathcal{O}_2)$$

since $\mathcal{O}_1$ and $\mathcal{O}_2$ are mutually exclusive. The outage exponent is

$$P_{\mathrm{out}}(r \log \mathrm{SNR}) \doteq \mathrm{SNR}^{-D_1} + \mathrm{SNR}^{-D_2} \tag{4.6}$$

where D_1 and D_2 are the SNR exponents of $\mathcal{O}_1$ and $\mathcal{O}_2$ respectively.

Let us perform the standard change of variables $\alpha_i = -\log \gamma_i / \log \mathrm{SNR}$ and $a = -\log g / \log \mathrm{SNR}$ [ZT03, AES05]. Then, the set of channel realizations corresponding to $\mathcal{O}_1$ is

$$\mathcal{A}_1 = \left\{(\alpha_1, \alpha_2, a) \in \mathbb{R}_+^3 : (1-\alpha_1)^+ < \frac{r}{\beta}, (1-a)^+ < \frac{r}{\beta}\right\}$$

and

$$D_1 = \inf_{\mathcal{A}_1}(\alpha_1 + \alpha_2 + a) = \left(2 - \frac{2r}{\beta}\right)^+.$$

For an optimal system, the condition $r \leq \beta$ must be satisfied otherwise $D_1 = 0$, i.e., the outage probability will not decay to zero as the SNR increases. Thus from now on, we constrain $\beta \geq r$.

Consider the second SNR exponent $D_2 = \inf_{\mathcal{A}_2}(\alpha_1 + \alpha_2 + a)$ where

$$\mathcal{A}_2 = \{(\alpha_1, \alpha_2, a) \in \mathbb{R}_+^3 : \beta(1-\alpha_1)^+ \geq r, \beta(1-a)^+ + (1-\beta)(1-\alpha_2)^+ < r\}.$$

The event $\beta(1-\alpha_1)^+ \geq r$ has a probability in the order of SNR^0 and is independent of the event $\beta(1-a)^+ + (1-\beta)(1-\alpha_2)^+ < r$. Thus by invoking Lemma 4.1, we obtain

$$D_2 = \begin{cases} 2 - \frac{r}{\beta} & \text{if } r < \beta < \frac{1}{2}, \\ \frac{1-r}{\beta} & \text{if } r \in (1-\beta, \beta), \beta \geq \frac{1}{2}, \\ 2 - \frac{r}{1-\beta} & \text{if } r < 1-\beta, \beta \geq \frac{1}{2}. \end{cases} \tag{4.7}$$

The slowest decayed term on the right hand side of (4.6) is the dominating one as SNR $\to \infty$, thus we have $D_{\text{out}}(r) = \min(D_1, D_2)$. The outage exponent corresponding to an *optimal* dimension allocation is given by

$$D_{\text{O}}^{\text{NF}}(r) = \sup_{\beta \in [r,1)} \min(D_1, D_2).$$

Due to the symmetry of D_2, for a given $r \leq 1/2$ and any $\beta \in (r, 1/2]$ there exists a $\hat{\beta} = (1-\beta) \in [1/2, 1-r)$ that results in $\hat{D}_2 = D_2$. But $\hat{\beta} \geq \beta$ thus $D_1 \leq \hat{D}_1$ and $\min(D_1, D_2) \leq \min(\hat{D}_1, \hat{D}_2)$. This means there is no loss of optimality by considering $\beta \geq 1/2$ and

$$D_{\text{O}}^{\text{NF}}(r) = \sup_{\beta \in [\max(r,1/2),1)} \min(D_1, D_2).$$

The results then follow a straightforward investigation.

Both the fixed DF schemes with no CSF in Proposition 4.1 are unsurprisingly outperformed by the DDF scheme [AES05] for any $r \in (0,1)$. Recall that in the DDF scheme, the source keeps transmitting until the accumulated mutual information at the relay is sufficient for decoding (the relay informs the source about such an event by an acknowledgement bit). Therefore DDF requires rateless codes, which are more complicated than codes with fixed lengths needed for the fixed DF approaches. We will further improve the simple fixed DF schemes with the help of quantized channel state information in the next sections.

4.4 Relay-to-Source CSF

In this section, we assume that given the channel h_1, the relay sends an index $\mathcal{I}(h_1) \in \{1, \ldots, K\}$ back to the source via an error-free zero-delay feedback link. The positive integer K is referred to as the *resolution* of the feedback link. The source allocates dimension (under a short-term power constraint) and both dimension and power (under a long-term power constraint). Since the two channel gains to the destination are unknown at the source and relay, the outage exponent of this scheme can be upper-bounded by that of a MISO 2×1 channel with no-CSIT (i.e., the fully cooperative transmitters bound), $D_{\text{MISO}}(r) = 2 - 2r$.

Short-term Power Constraint

First consider a short-term power constraint case, i.e., no power control at the source. Let β_k be the fraction of dimension allocated for the learning phase given feedback index k, $\beta_1 < \beta_2 < \cdots < \beta_K$. Recall that the information rate $r \log \text{SNR}$ is independent of the feedback index. Consider the index mapping

$$\mathcal{I}(\gamma_1) = \begin{cases} K & \text{if } \beta_K \log(1+\gamma_1 \text{SNR}) < r \log \text{SNR}, \\ \min\{k \in \{1, \ldots, K\} : \beta_k \log(1+\gamma_1 \text{SNR}) \geq r \log \text{SNR}\} & \text{otherwise.} \end{cases}$$

That is, from a finite set of $\{\beta_k\}$ the source allocates the minimum dimension so that the rate can be supported by the source-relay link. If the rate is not supportable even with the largest dimension β_K, then the index $\mathcal{I} = K$ is fed back so that the direct link can benefit the most in the orthogonal case (in the nonorthogonal case, whatever index sent back when the relay fails does not matter because the direct transmission always uses all the available dimension).

We quantify the outage exponent of the scheme described above in the following proposition. The proof is deferred to Appendix 4.A.

Proposition 4.2. *An orthogonal scheme without power control and with relay-to-source CSF resolution K can achieve an outage exponent of*

$$D_{O\text{-}NPC}^{RF-K}(r) = \sup_{\max(r,0.5)\leq\beta_1<\cdots<\beta_K<1} \min\left(D_1,\ldots,D_K, 2\left(1-\frac{r}{\beta_K}\right)\right) \tag{4.8}$$

where

$$D_k = \begin{cases} 3 - \frac{r}{1-\beta_k} - \frac{r}{\beta_{k-1}} & \textit{if } \beta_k < 1-r, \\ \frac{1-r}{\beta_k} + 1 - \frac{r}{\beta_{k-1}} & \textit{otherwise}, \end{cases} \tag{4.9}$$

with the convention $\beta_0 = r$.

A nonorthogonal scheme without power control and relay-to-source CSF resolution K can achieve an outage exponent of

$$D_{NO\text{-}NPC}^{RF-K}(r) = \sup_{\max(r,0.5)\leq\beta_1<\cdots<\beta_K<1} \min\left(D_1,\ldots,D_K, 2-r-\frac{r}{\beta_K}\right) \tag{4.10}$$

where D_k is defined as in (4.9).

Notice that for $r \geq 1/2$, then $D_k = \frac{1-r}{\beta_k} + 1 - \frac{r}{\beta_{k-1}}$, $\forall k$, due to the constraint $\beta_k \geq \max(r, 0.5)$. But in general, the D_k's in (4.9) are neither concave or convex, making the evaluation of the maximin difficult. We next present tight upper bounds and lower bounds to the optimized outage exponent in Proposition 4.2, which are simpler to compute. The proof is in Appendix 4.B.

Proposition 4.3. *The outage exponent $D_{O\text{-}NPC}^{RF-K}(r)$ is lower-bounded by*

$$D_{LB\text{-}O\text{-}NPC}^{RF-K}(r) = \begin{cases} \frac{2(1-r)-2(1-r)\left(\frac{r}{1-r}\right)^K}{1-2r\left(\frac{r}{1-r}\right)^K} & \textit{if } r \neq 1/2 \\ \frac{2K}{1+2K} & \textit{if } r = 1/2. \end{cases} \tag{4.11}$$

For $r \geq 1/2$, we have $D_{O\text{-}NPC}^{RF-K}(r) = D_{LB\text{-}O\text{-}NPC}^{RF-K}(r)$.

For $r < 1/2$, $D_{O\text{-}NPC}^{RF-K}(r)$ is upper-bounded by

$$D_{UB\text{-}O\text{-}NPC}^{RF-K}(r) = \sup_{1/2\leq\beta_1<\cdots<\beta_K<1} \min\left(\bar{D}_1,\ldots,\bar{D}_K, 2\left(1-\frac{r}{\beta_K}\right)\right) \tag{4.12}$$

where

$$\bar{D}_k = \begin{cases} 3 - \frac{r}{1-\beta_k} - \frac{r}{\beta_{k-1}} & \text{if } \beta_k < 1-r, \\ 3 - r - \beta_k - \frac{r}{\beta_{k-1}} & \text{otherwise,} \end{cases} \tag{4.13}$$

with the convention $\beta_0 = r$;

The outage exponent $D^{RF-K}_{NO\text{-}NPC}(r)$ *is lower-bounded by*

$$D^{RF-K}_{LB\text{-}NO\text{-}NPC}(r) = \begin{cases} \frac{2(1-r)^2-(1-r)\left(\frac{r}{1-r}\right)^K}{1-r-r\left(\frac{r}{1-r}\right)^K} & \text{if } r \neq 1/2 \\ \frac{1+2K}{2+2K} & \text{if } r = 1/2. \end{cases} \tag{4.14}$$

For $r \geq 1/2$, *we have* $D^{RF-K}_{NO\text{-}NPC}(r) = D^{RF-K}_{LB\text{-}NO\text{-}NPC}(r)$.

For $r < 1/2$, $D^{RF-K}_{NO\text{-}NPC}(r)$ *is upper-bounded by*

$$D^{RF-K}_{UB\text{-}NO\text{-}NPC}(r) = \sup_{1/2 \leq \beta_1 < \cdots < \beta_K < 1} \min\left(\bar{D}_1, \ldots, \bar{D}_K, 2 - r - \frac{r}{b_K}\right) \tag{4.15}$$

where $\bar{D}_k$ *is defined in (4.13).*

The lower bounds, which coincide with the actual outage exponents for $r \geq 1/2$, are explicitly given. As for the upper bounds, which are only necessary when $r < 1/2$, we consider all possible cases

$$\beta_1^K \in \underbrace{[1/2, 1-r) \times \cdots \times [1/2, 1-r)}_{l \text{ times}} \times \underbrace{[1-r, 1) \times \cdots \times [1-r, 1)}_{K-l \text{ times}},$$

for $l = 0, \ldots, K$ where $\mathcal{A} \times \mathcal{B}$ denotes the Cartesian product. Due to the constraint $\beta_1 < \cdots < \beta_K$, we only need to consider $K+1$ such regions (and not 2^K). Over each of the $K+1$ regions we can efficiently solve for the optimum using convex optimization methods. The problem is convex because over either $\{\beta_{k-1}, \beta_k\} \in [1/2, 1) \times [1/2, 1-r)$ or $\{\beta_{k-1}, \beta_k\} \in [1/2, 1) \times [1-r, 1)$, the function $\bar{D}_k$ in (4.13) is *concave*; and the point-wise minimum of a family of concave function is concave [BV04]. Finally, the maximum of the $K+1$ solutions gives the desired upper bound.

We plot in Fig. 4.2 the upper and lower bounds in Proposition 4.3 for orthogonal schemes. A relatively large improvement over the no-CSF case can be achieved, but this gain diminishes as K increases from 2 to 3. As can be seen, the proposed CSF scheme is outperformed by DDF, but gradually approaches the performance of DDF as the CSF quality improves. We formalize this observation in the following.

Corollary 4.1. *We have*

$$\lim_{K\to\infty} D^{RF-K}_{O\text{-}NPC}(r) = \lim_{K\to\infty} D^{RF-K}_{NO\text{-}NPC}(r) = \begin{cases} 2 - 2r & \text{if } r < \frac{1}{2} \\ \frac{1-r}{r} & \text{if } r \geq \frac{1}{2}. \end{cases} \tag{4.16}$$

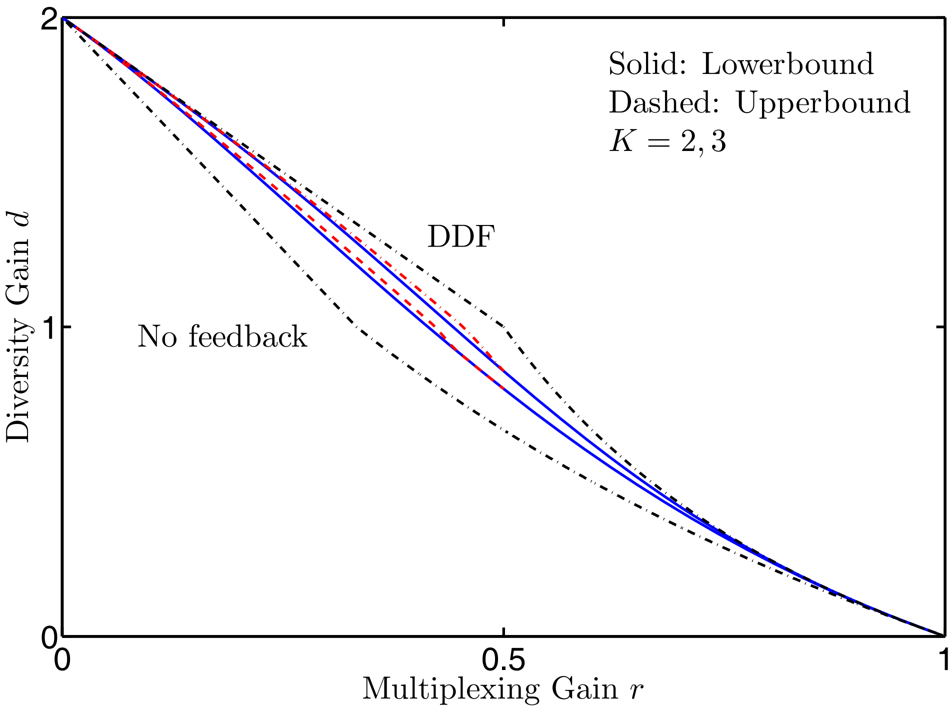

Figure 4.2: Outage exponents with relay CSF, no power control: Orthogonal schemes with different feedback resolution K.

Proof. Taking the limit of the lower bounds in Proposition 4.3 as $K \to \infty$ and combining with the cooperative upper bound $D^{\text{RF}-K}_{\text{O-NPC}}(r) \le D^{\text{RF}-K}_{\text{NO-NPC}}(r) \le 2 - 2r$ give the stated result. Note that for $r \ge 1/2$, the lower bounds of Proposition 4.3 coincide with the exact outage exponent. $\square$

The result essentially says that nonorthogonality is not necessary to achieve the DDF bounds when perfect CSIT is available and the systems fully adapts the available dimension to the channel condition. This is practically the case for low-rate CSIT feedback as well, as illustrated in Fig. 4.4. For K as low as 2 (i.e., 1 bit of CSF) there is insignificant gain by using nonorthogonal schemes. Our result suggests the use of an orthogonal, low-rate relay CSF scheme to achieve a substantial portion of the cooperative gain.

Long-term Power Constraint

We now relax the power constraint at the source, so that temporal power control is possible. Let $r < \beta_1 < \ldots < \beta_K < 1$ be the set of dimension fractions, and

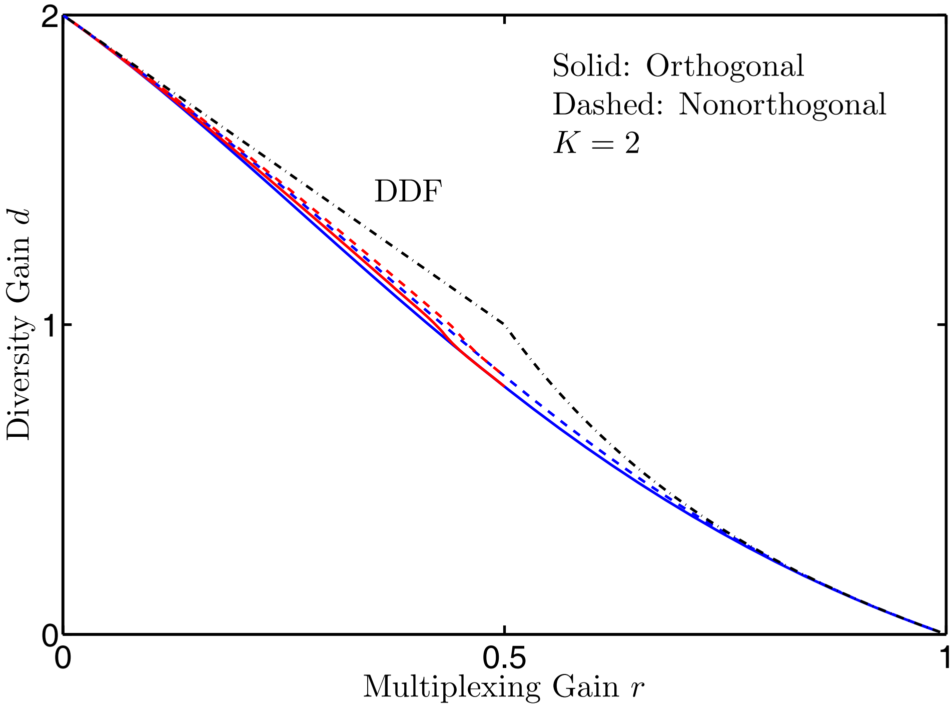

Figure 4.3: Outage exponents with relay CSF, no power control: Orthogonal vs. nonorthogonal schemes, $K = 2$.

$P_1 < P_2 < \cdots < P_K$ be the set of power levels used at the source. We consider the following index mapping

$$\mathcal{I}(\gamma_1) = \begin{cases} K & \text{if } \beta_K \log(1+\gamma_1 P_K) < r \log \mathrm{SNR}, \\ \min\{k \in \{1,\ldots,K\} : \beta_k \log(1+\gamma_1 P_k) \geq r \log \mathrm{SNR}\} & \text{otherwise.} \end{cases}$$

That is, the system not only increases the dimension but also allocates more power in poor source-relay channel conditions. A long-term power constraint is imposed so that

$$\lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} P_{\mathcal{I}(\gamma_1(l))} \overset{\text{a.s.}}{=} \mathrm{E}_{\gamma_1} P_{\mathcal{I}(\gamma_1)} \leq \mathrm{SNR}$$

where $\gamma_1(l)$ denotes the source-relay channel power gain during fading block l. Even in this power control scenario, the outage exponent is still upper-bounded by the cooperative bound $D_{\mathrm{MISO}}(r) = 2 - 2r$ since neither the source nor the relay knows the channel to the destination. We state the optimal outage exponent of the proposed scheme in the following.

Proposition 4.4. *The proposed orthogonal scheme with relay-to-source feedback resolution K and source power control can achieve an outage exponent of*

$$D_{O\text{-}PC}^{RF-K}(r) = \sup_{\max(r,1/2)\leq\beta_1<\ldots<\beta_K<1} \min\left(D_1,\ldots,D_K, 2\left(K-\sum_{k=1}^{K}\frac{r}{\beta_k}\right)\right) \quad (4.17)$$

where

$$D_k = \begin{cases} k+1-\frac{r}{1-\beta_k}-\sum_{i=1}^{k-1}\frac{r}{\beta_i} & \text{if } \beta_k < 1-r, \\ k-1+\frac{1-r}{\beta_k}-\sum_{i=1}^{k-1}\frac{r}{\beta_i} & \text{otherwise.} \end{cases} \quad (4.18)$$

The proposed nonorthogonal scheme with relay-to-source feedback resolution K and source power control can achieve an outage exponent of

$$D_{NO\text{-}PC}^{RF-K}(r) = \sup_{\max(r,1/2)\leq\beta_1,\ldots,\beta_K<1} \min\left(D_1,\ldots,D_K, 2\left(K-\sum_{k=1}^{K-1}\frac{r}{\beta_k}\right)-r-\frac{r}{\beta_K}\right) \quad (4.19)$$

where D_k is defined as in (4.18).

Proof. See Appendix 4.C. □

As in the no power control case, it is difficult to compute the entire outage exponent curves of the proposed schemes due to the nonconvexity of the optimization problem. We summarize computable bounds to $D_{\text{O-PC}}^{\text{RF}-K}(r)$ and $D_{\text{NO-PC}}^{\text{RF}-K}(r)$ as follows. The proof is deferred to Appendix 4.D.

Proposition 4.5. *The outage exponent $D_{O\text{-}PC}^{RF-K}(r)$ is lower-bounded by*

$$D_{LB\text{-}O\text{-}PC}^{RF-K}(r) = \begin{cases} 2-2r & \text{if } (1-r)^{K+1} < 1-2r, \\ \frac{(1-r)\left(1-(1-r)^K\right)}{r\left(1-\frac{1}{2}(1-r)^K\right)} & \text{otherwise.} \end{cases} \quad (4.20)$$

For $\{r : (1-r)^{K+1} < 1-2r\}$ and for $r \geq 1/2$, we have $D_{O\text{-}PC}^{RF-K}(r) = D_{LB\text{-}O\text{-}PC}^{RF-K}(r)$.

For $\{r : (1-r)^{K+1} \geq 1-2r, r < 1/2\}$, $D_{O\text{-}PC}^{RF-K}(r)$ is upper-bounded by

$$D_{UB\text{-}O\text{-}PC}^{RF-K}(r) = \sup_{1/2\leq\beta_1<\cdots<\beta_K<1} \min\left(\bar{D}_1,\ldots,\bar{D}_K, 2\left(K-\sum_{k=1}^{K}\frac{r}{\beta_k}\right)\right) \quad (4.21)$$

where

$$\bar{D}_k = \begin{cases} k+1-\frac{r}{1-\beta_k}-\sum_{i=1}^{k-1}\frac{r}{\beta_i} & \text{if } \beta_k < 1-r, \\ k+1-r-\beta_k-\sum_{i=1}^{k-1}\frac{r}{\beta_i} & \text{otherwise.} \end{cases} \quad (4.22)$$

The outage exponent $D^{RF-K}_{NO\text{-}PC}(r)$ is lower-bounded by

$$D^{RF-K}_{LB\text{-}NO\text{-}PC}(r) = \begin{cases} 2-2r & \text{if } (2-3r)(1-r)^K < (2-r)(1-2r), \\ \frac{(1-r)(2-r)\left(1-(1-r)^K\right)}{r(2-r-(1-r)^K)} & \text{otherwise.} \end{cases} \tag{4.23}$$

For $\{r : (2-3r)(1-r)^K < (2-r)(1-2r)\}$ and for $r \geq 1/2$, we have $D^{RF-K}_{NO\text{-}PC}(r) = D^{RF-K}_{LB\text{-}NO\text{-}PC}(r)$.

For $\{r : r : (2-3r)(1-r)^K \geq (2-r)(1-2r), r < 1/2\}$, $D^{RF-K}_{NO\text{-}PC}(r)$ is upper-bounded by

$$D^{RF-K}_{UB\text{-}NO\text{-}PC}(r) = \sup_{1/2 \leq \beta_1 < \cdots < \beta_K < 1} \min \left(\bar{D}_1, \ldots, \bar{D}_K, 2\left(K - \sum_{k=1}^{K-1} \frac{r}{\beta_k}\right) - r - \frac{r}{\beta_K} \right) \tag{4.24}$$

where $\bar{D}_k$ is defined as in (4.22).

We plot in Fig. 4.4 the bounds of Proposition 4.5 for an orthogonal scheme. Allowing nonorthogonality only provides some insignificant additional gains (not plotted here). A very pleasing fact is that the proposed source power control scheme is *strictly optimal* in an outage exponent sense for *any* finite feedback resolution $K \geq 2$ over a wide range of multiplexing gain, as it achieves the MISO cooperative upper bound $D_{\text{MISO}}(r) = 2 - 2r$. For example, when $K = 2$ an orthogonal scheme is optimal for any $r < \frac{3-\sqrt{5}}{2} \approx 0.38$.

As can be seen, even with a modest $K = 4$, the difference between the proposed scheme and DDF is marginal. It is not difficult to see that as K increases, using source power control also approaches the DDF performance for all r. We state this fact in the following.

Corollary 4.2. *We have*

$$\lim_{K\to\infty} D^{RF-K}_{O\text{-}PC}(r) = \lim_{K\to\infty} D^{RF-K}_{NO\text{-}PC}(r) = \begin{cases} 2-2r & \text{if } r < \frac{1}{2}, \\ \frac{1-r}{r} & \text{otherwise.} \end{cases} \tag{4.25}$$

While this result says that with high-rate CSF, power control at the source does not really help over dimension allocation only, we can see in Fig. 4.4 that power control is quite beneficial in the very low rate CSF regime.

That perfect power control at the source cannot achieve the MISO bound over the entire range of multiplexing gains may come as a surprise. A closer look suggests that this is due to the nature of the decode-and-forward approach. Consider the following *heuristic* arguments. Even with perfect power control, most of the time (i.e., with probability in the order of SNR^0), a power in the order of SNR^1 is applied at the source. Fully conveying the message from the source to the destination therefore typically requires a fraction $\beta \geq r$ of the dimension. For $r < 1/2$ let us

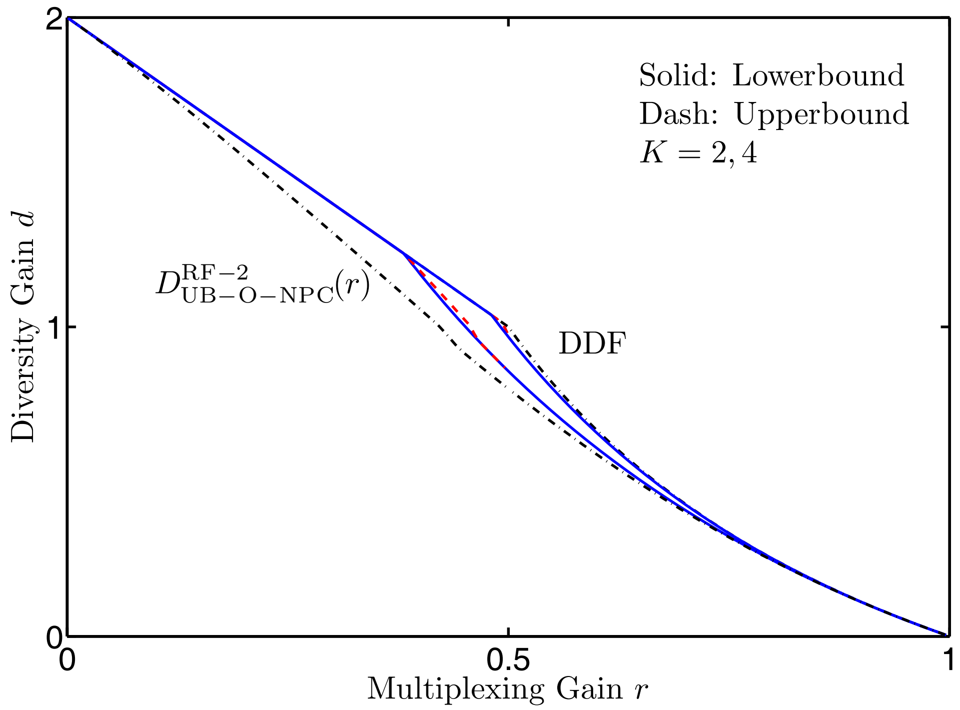

Figure 4.4: Outage exponent of an orthogonal scheme with relay CSF and power control. The upper bound to the outage exponent of an orthogonal scheme with no power control and relay feedback resolution $K = 2$ is also plotted.

choose $\beta = 1/2 > r$ and for $r \geq 1/2$, choose $\beta = r$. Assume that after the learning phase the message is known perfectly at the relay so that we actually have a MISO at the relaying phase. Then an application of Lemma 4.1 gives the outage exponent of such an idealized system, which is exactly what is dictated by Corollary 4.2. In other words, for $r > 1/2$, once the virtual antenna array has been successfully formed, we no longer have sufficient dimension for fully-cooperative diversity.

4.5 Destination-to-Relay CSF

In this section, we study the outage exponent over a channel with destination-relay limited CSF. Since only the relay knows partially about its forward channel gain, dimension allocation is not possible. We can only adapt the power, assuming a long-term power constraint at the relay. Note that the outage exponent of such a scheme is upper-bounded by that of a 1×2 SIMO channel with no CSIT: $D_{\mathrm{SIMO}}(r) = 2 - 2r$ (i.e., the fully cooperative receivers bound, which follows from the cut-set bound) because the source does not know its two forward channel gains.

Let $\beta \in (\max(r, 1/2), 1)$ be the fraction of dimension allocated for the learning phase, and K be the resolution of the destination-relay CSF link. At the destination, given g, γ_2, an index mapping $\mathcal{I}(g, \gamma_2)$ produces an index that is sent back to the relay. Note that the destination does *not* know the source-relay channel gain. The relay employs an index-to-power mapping

$$\mathcal{P} : \{1, \dots, K\} \to \{P_1, \dots, P_K\}$$

where $P_1 < \cdots < P_K$. We consider the following index mapping.

$$\mathcal{I}(g, \gamma_2) = \begin{cases} 1 & \text{if} \quad \beta \log(1 + g\mathrm{SNR}) + (1-\beta)\log(1+\gamma_2 P_K) < r \log \mathrm{SNR}, \\ \min k \in \{1, \dots, K\} & \\ \text{s.t.} \quad \beta \log(1 + g\mathrm{SNR}) + (1-\beta)\log(1+\gamma_2 P_k) \geq r \log \mathrm{SNR} & \text{otherwise.} \end{cases}$$

That is, the destination informs the relay to use the smallest power level possible to help the transmission, assuming pessimistically that the source-relay link is not in outage. If the channel is too costly (in a too deep fade) to compensate for, then the smallest index is sent back to save power. If the source-relay link is in outage then the relay, which knows about this event since it has perfect receiver CSI, simply ignores the feedback index $\mathcal{I}(g, \gamma_2)$.

The outage exponent of the proposed scheme is given in the following. The proof is presented in Appendix 4.E.

Proposition 4.6. *The optimal outage exponent of the proposed orthogonal scheme with destination-to-relay resolution feedback K is*

$$D^{DRF\text{-}K}_{O\text{-}PC}(r) = \sup \min_{\beta \in [\max(1/2, r), 1)} \left(2\left(1 - \frac{r}{\beta}\right), D_K \right) \tag{4.26}$$

where

$$D_k = \begin{cases} 2 + D_{k-1} - \frac{r}{1-\beta} & \text{if } \frac{1}{2} \leq \beta < 1 - \frac{r}{1+D_{k-1}} \\ \frac{1-\beta}{\beta} D_{k-1} + \frac{1-r}{\beta} & \text{if } \beta \geq \max\left(\frac{1}{2}, 1 - \frac{r}{1+D_{k-1}}\right) \end{cases} \tag{4.27}$$

with the convention $D_0 = 0$.

The optimal outage exponent of the proposed nonorthogonal scheme is

$$D^{DRF\text{-}K}_{NO\text{-}PC}(r) = \sup \min_{\beta \in [\max(1/2, r), 1)} \left(2 - r - \frac{r}{\beta}, D_K \right) \tag{4.28}$$

with D_k defined as in (4.27).

A major difficulty in finding a closed-form expression for the outage exponent in this case is the dependency on k and β of the point $1 - \frac{r}{1+D_{k-1}}$ where the function D_k changes its behavior, in contrast to the fixed value $1 - r$ when only the source controls its transmit power (cf. Proposition 4.4). Nevertheless, for small values of K, it is possible to find explicit expressions for the outage exponent. For example, in the case $K = 2$ we have

$$D_{\text{O-PC}}^{\text{DRF-2}}(r) = \begin{cases} 2 - \frac{2r(1+2r)}{1+r} & \text{if } r < \frac{\sqrt{17}-1}{8}, \\ 2 - \frac{2r\left(\sqrt{r^2-2r+2}-r\right)}{1-r} & \text{otherwise.} \end{cases}$$

As we will see even for K as small as 2, performance very close to perfect CSF can be achieved. Furthermore, the outage exponent for the full-CSI feedback case can be found explicitly as follows.

Proposition 4.7. *We have*

$$\lim_{K\to\infty} D_{\textit{O-PC}}^{\textit{DRF-K}}(r) = \begin{cases} 2\left(1 - \frac{2r}{2-r}\right) & \textit{if } r \le \frac{2}{5}, \\ 2\left(1 - \frac{8r}{3r+3+\sqrt{9r^2-14r+9}}\right) & \textit{otherwise,} \end{cases} \tag{4.29}$$

and

$$\lim_{K\to\infty} D_{\textit{NO-PC}}^{\textit{DRF-K}}(r) = \begin{cases} 2 - r - \frac{2r}{2-r} & \textit{if } r \le \frac{5-\sqrt{17}}{2}, \\ 2 - r - \frac{4r(2-r)}{3+\sqrt{8r^2-16r+9}} & \textit{otherwise.} \end{cases} \tag{4.30}$$

We herein present a proof using an indirect approach based on the observation that $K \to \infty$ indeed means the relay knows both h and h_2 perfectly. An alternative proof is presented in Appendix 4.E where we directly apply Proposition 4.6 and then take the limit $K \to \infty$.

Proof. For the sake of brevity, we present only the orthogonal case. Note that $K \to \infty$ is equivalent to a direct mapping from channel gains to the power allocated at the relay $P(g, \gamma_2) \doteq \text{SNR}^{\pi(g,\gamma_2)}$. A change of variables gives $a = -\log g/\log \text{SNR}$, $\alpha_2 = -\log \gamma_2/\log \text{SNR}$. Asymptotically the long-term power constraint at the relay reads

$$\sup_{a\ge 0, \alpha_2 \ge 0} (\pi(a, \alpha_2) - a - \alpha_2) \le 1.$$

Since outage probability is a nonincreasing function of $\pi^*(a, \alpha_2)$, we conclude that the optimal power exponent $\pi^*(a, \alpha_2) = 1 + a + \alpha_2$.

The outage event is

$$P_{\text{out}}(r \log \text{SNR}) = \Pr(\mathcal{O}_1) + \Pr(\mathcal{O}_2)$$

where $\mathcal{O}_1$ is the event that the relay fails to decode the message and the destination cannot decode the direct transmission.

$$\begin{aligned} \Pr(\mathcal{O}_1) &= \Pr\left(\beta \log(1 + \gamma_1 \text{SNR}) < r \log \text{SNR}, \beta \log(1 + g\text{SNR}) < r \log \text{SNR}\right) \\ &\doteq \text{SNR}^{-2\left(1-\frac{r}{\beta}\right)}. \end{aligned}$$

The event $\mathcal{O}_2$ happens when the relay succeeds to decode but the combined direct and relayed signals cannot be decoded by the destination.

$$\begin{aligned}\Pr(\mathcal{O}_2) &= \Pr\Big(\beta\log(1+\gamma_1\mathrm{SNR}) \geq r\log\mathrm{SNR},\\ &\qquad \beta\log(1+g\mathrm{SNR}) + (1-\beta)\log(1+\gamma_2\mathrm{SNR}^{\pi^*(g,\gamma_2)}) < r\log\mathrm{SNR}\Big)\\ &= \mathrm{SNR}^{-D_\infty}.\end{aligned}$$

Then

$$\begin{aligned}D_\infty &= \inf_{a,\alpha_2\geq 0}\{a+\alpha_2\} \quad \text{s.t.} \quad \beta(1-a)^+ + (1-\beta)(1+a+\alpha_2-\alpha_2)^+ < r\\ &= \inf_{a\geq 0}\{a\} \quad \text{s.t.} \quad \beta(1-a)^+ + (1-\beta)(1+a) < r\end{aligned} \tag{4.31}$$

Since $\min_{a\geq 0}\beta(1-a)^+ + (1-\beta)(1+a) = 2(1-\beta)$ (the minimum is attained at $a=1$), the optimization problem finding D_∞ is infeasible if $r < 2(1-\beta) \Leftrightarrow \beta < \frac{2-r}{2}$. Thus $\Pr(\mathcal{O}_2)$ decays *exponentially*[3] in SNR if $\beta < \frac{2-r}{2}$ because the set of "bad" channels making $\Pr(\mathcal{O}_2)$ decay polynomially in SNR is *empty*. We now compute the outage exponent for the two cases $\beta < \frac{2-r}{2}$ and $\beta \geq \frac{2-r}{2}$ separately and then take the maximum one to be $D^{\mathrm{DRF}-\infty}_{\mathrm{O\text{-}PC}}(r)$.

Case 1: $\beta < \frac{2-r}{2}$. Due to the constraint $\beta \geq \max(r, 1/2)$, this case never happens if $r > \frac{2}{3}$. For $r \leq \frac{2}{3}$, since $\Pr(\mathcal{O}_2) \doteq \mathrm{SNR}^{-\infty}$ (the notation means exponentially decayed in SNR) we have

$$\underline{D}_1 = \sup_{\beta\in[r,\frac{2-r}{2})}\min\left(2-\frac{2r}{\beta},\infty\right) = \sup_{\beta\in[r,\frac{2-r}{2})} 2-\frac{2r}{\beta} = 2-\frac{4r}{2-r}.$$

Case 2: $\beta \geq \frac{2-r}{2}$. Solving (4.31), we obtain the explicit outage exponent for $\Pr(\mathcal{O}_2)$

$$D_\infty = \frac{1-r}{2\beta-1}.$$

Noting that $\frac{2-r}{2} > \frac{1}{2}, \forall r \in (0,1)$, we next solve

$$\underline{D}_2 = \sup_{\beta\in[\max(r,\frac{2-r}{2}),1)}\min\left(2-\frac{2r}{\beta},\frac{1-r}{2\beta-1}\right).$$

For $r < \frac{2}{5}$, we have $\max\left(\frac{1}{2}, r, \frac{2-r}{2}\right) = \frac{2-r}{2}$, then $2-2\frac{r}{\beta} > \frac{1-r}{2\beta-1}$, $\forall\beta\in\left[\frac{2-r}{2},1\right)$. This gives

$$\underline{D}_2 = \frac{1-r}{2\frac{2-r}{2}-1} = 1.$$

[3] Our notion of exponential decay is extended to also include the case of zero probability $\Pr(\mathcal{O}_2) = 0$.

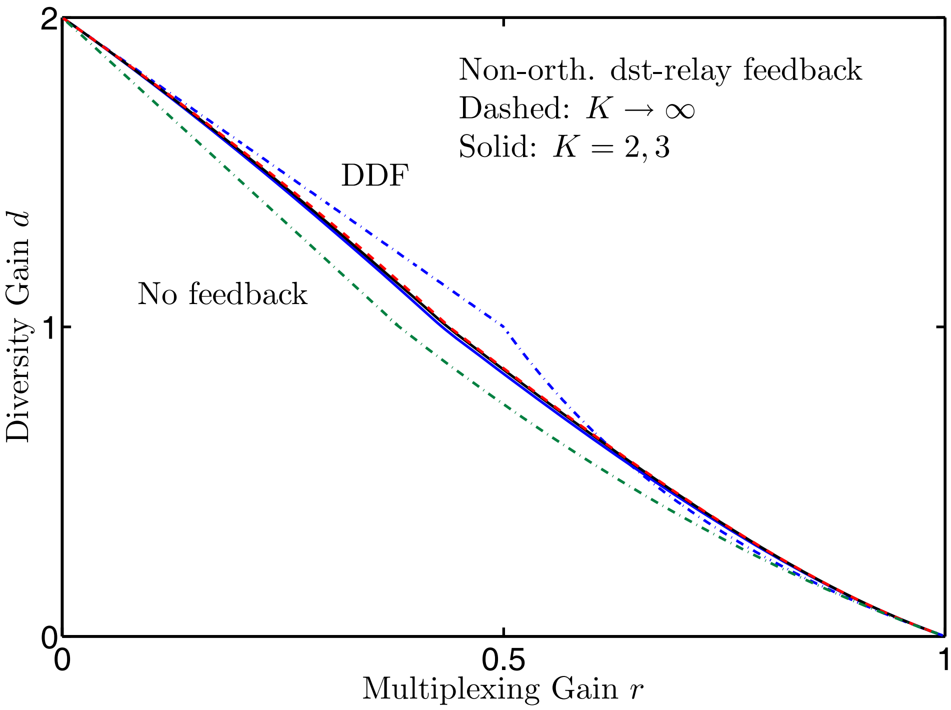

Figure 4.5: Outage exponent of a nonorthogonal scheme with destination-relay CSF.

For $r \geq \frac{2}{5}$, the maximin is always the intersection of the two component functions that lies inside $\left[\max\left(\frac{2-r}{2}, r\right), 1\right)$, which turns out to be the largest root of $f(\beta) = 4\beta^2 - 3(1+r)\beta + 2r$. This gives

$$\underline{D}_2 = 2 - \frac{16r}{3 + 3r + \sqrt{9r^2 - 14r + 9}}.$$

Finally, combining Case 1 and Case 2 we have

$$\begin{aligned}\lim_{K\to\infty} D_{\text{O-PC}}^{\text{DRF-}K}(r) &= \begin{cases}\max(\underline{D}_1, \underline{D}_2) & \text{if } r < \frac{2}{3},\\ \underline{D}_2 & \text{otherwise}\end{cases}\\ &= \begin{cases} 2 - \frac{4r}{2-r} & \text{if } r < \frac{2}{5},\\ 2 - \frac{16r}{3+3r+\sqrt{9r^2-14r+9}} & \text{otherwise.}\end{cases}\end{aligned}$$

□

In Fig. 4.5, the outage exponents of the proposed nonorthogonal schemes are plotted. Indeed the performance in the three cases $K = 2$, $K = 3$ and $K \to \infty$ is

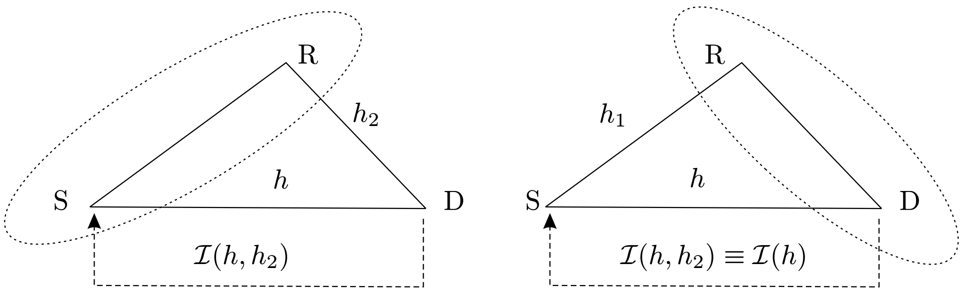

Figure 4.6: The two cooperative upper bounds in the destination-to-source-and-relay CSF scenario. Since the feedback quantizer at destination does not know h_1, the feedback index $\mathcal{I}$ only depends on h and h_2.

practically indistinguishable. This suggests the use of a *single* bit of feedback ($K = 2$) for power control in the proposed DF scheme to achieve a near optimal diversity-multiplexing tradeoff with very low complexity. A more important and interesting observation is that the performance is *strictly better than DDF* for sufficiently large r. While the gain is marginal, it is worth noting that this is realizable even with $K = 2$. With $K = 2$, over the range $r > 0.65$ (rounded up value) the proposed nonorthogonal CSF scheme dominates DDF. Thus for low-rate-CSF, high-data-rate systems, destination-relay feedback is more beneficial than relay-source feedback in a diversity-multiplexing tradeoff sense.

With perfect CSF, however, Proposition 4.7 also shows that the DF with power control at the relay cannot achieve the multiantenna upper bound. In this perfect-CSF scenario, it is better to use compress-and-forward to achieve the optimal diversity-multiplexing tradeoff of the relay channel [YE07]. It is not yet clear how the performance of the CF scheme will be affected by the assumption of heavily quantized CSF considered herein. This remains an interesting topic for future work.

4.6 Destination-to-Source-and-Relay CSF

We now consider the final scenario where the destination broadcasts a feedback index to the source and the relay so that both nodes can control their individual powers. We assume that , the index mapper at the destination depends on the channel gains h, h_2 but *not* on the source-relay gain h_1. Assume a long-term power constraint at both the source and the relay.

Before proceeding to the derivation of the achievable outage exponent of any particular scheme, it is useful to find an upper bound on the performance of such a CSF setting. One can naturally think of a fully cooperative MISO channel with feedback resolution K. Unfortunately, such an upper bound, which is given by $D_{\text{MISO}}^{K}(r) = \left(2 + \cdots + 2^{K}\right)(1 - r) = 2\left(2^{K} - 1\right)(1 - r)$ turns out to be quite loose. In fact, taking the worst-case of the two cut-set cooperative bounds, we notice that since the CSF index depends only on g and γ_2, the SIMO case where the source transmits and the relay acts as a perfectly cooperating additional receiver is the most restrictive from a diversity exponent viewpoint, as illustrated in Fig. 4.6. We present a tighter upper bound in the following, obtained by considering a SIMO channel with K feedback levels where a restricted feedback quantizer knows only *one* of the two scalar channel gains.

Lemma 4.2. *Assume destination-to-source-and-relay CSF and power control at both the source and the relay. If the feedback quantizer at the destination does not depend on the source–relay gain, then the outage exponent with K feedback levels is upper-bounded by*

$$D^{DSRF\text{-}K}_{UB\text{-}SPC\text{-}RPC}(r) = 2K(1-r). \tag{4.32}$$

Proof. To get an upper bound, we let the relay and the destination fully cooperate and compute the outage exponent $D^{\text{DSF-}K}_{\text{UB-SPC-RPC}}(r)$ of the resulting equivalent channel. The equivalent 1×2 SIMO channel (see Fig. 4.6) has the following properties:

- Squared magnitudes of channel gains $\lambda_1 = \text{SNR}^{-a_1}$ and $\lambda_2 = \text{SNR}^{-a_2}$ are independent exponentially distributed random variables with unit variance.
- Perfect CSI at the receiver.
- An index mapping $\mathcal{I}(\lambda_1)$ produces a feedback index depending on λ_1 but *independent* of λ_2. The feedback link is noiseless, and has zero delay.
- A long-term power constraint at the transmitter.

Let the K possible transmit power levels be $P_k = \text{SNR}^{p_k}$ with $1 = p_1 < p_2 < \cdots < p_K < p_{K+1} = \infty$. Let $\mathcal{R}_k$ be the kth quantization region, i.e., the set of of all channel realizations that are mapped to index k. Because the power levels used at the transmitter depend on λ_1 and not on λ_2, the long-term power constraint (asymptotically) becomes

$$\sup_{\{a_1 \geq 0\} \cap \mathcal{R}_k} (p_k - a_1) \leq 1.$$

Since outage probability is a nonincreasing function of power, the quantization region $\mathcal{R}_k$ is given by the set of all a_1 where SNR^{p_k} can be applied but not $\text{SNR}^{p_{k+1}}$. That is,

$$\mathcal{R}_k = \{a_1 : p_k - 1 \leq a_1 < p_{k+1} - 1\}. \tag{4.33}$$

The constraint $a_1 \geq 0$ always satisfies since $p_k \geq 1$, $\forall k$. Let $\Pr(\text{outage}, \mathcal{I} = k) \doteq \text{SNR}^{-D_k}$, then the overall outage exponent is given by the slowest decaying term $\min(D_1, \ldots, D_K)$. We then have

$$\Pr(\text{outage}, \mathcal{I} = k) \doteq \Pr(\log(1 + \text{SNR}^{p_k}\lambda_1 + \text{SNR}^{p_k}\lambda_2) < r \log \text{SNR}, \mathcal{I} = k).$$

This leads to

$$D_k = \inf_{a_1 \in \mathcal{R}_k, a_2 \geq 0} (a_1 + a_2) \quad \text{s.t.} \quad \max\left((p_k - a_1)^+, (p_k - a_2)^+\right) < r. \tag{4.34}$$

We distinguish two cases. First, if $p_{k+1}-1 < p_k - r$ then (4.34) is infeasible because of (4.33). In other words, the set of "bad" channels making $\Pr(\text{outage}, \mathcal{I}=k)$ decay polynomially in SNR is empty, and by convention $D_k = \infty$. Second, if $p_{k+1}-1 \geq p_k - r$ then $D_k = 2p_k - 2r$. Since $p_1 < \cdots < p_K$, the second case does not happen for any $k \in \{1, \ldots, K-1\}$, otherwise the regions $R_{k+1}, \ldots, \mathcal{R}_K$ are completely redundant and increasing K does *not* increase diversity gain. Therefore we must have $D_1 = \infty, \ldots, D_{K-1} = \infty$ and $D_K = 2p_K - 2r$. Then $\min(D_1, \ldots, D_K) = 2p_K - 2r$ is maximized when p_K is maximized. This happens when $p_{k+1} = 1 + p_k - r - \epsilon_k$, $k = 1, \ldots, K-1$ where $\epsilon_k > 0$ is arbitrarily small. Recalling that $p_1 = 1$, we let $\epsilon_k \downarrow 0$ and recursively obtain $p_k = k - (k-1)r$, $k = 2, \ldots, K$. Finally

$$D_{\text{UB-SPC}}^{\text{DSF-}K}(r) = 2p_K - 2r = 2K - 2(K-1)r - 2r = 2K(1-r).$$

□

A key observation from Lemma 4.2 is that the diversity gain of a relay channel with destination-to-source-and-relay CSF increases no faster than *linearly* in K, which is similar to the behavior of a SISO channel, $D_{\text{SISO}}^{K}(r) = K(1-r)$. On the contrary, the diversity gain of a MISO or SIMO channel increases *exponentially* in the feedback resolution K, $D_{\text{SIMO}}^{K}(r) = 2\left(2^K - 1\right)(1-r)$. Thus not knowing the source-relay gain at the destination, which is a relatively realistic assumption, severely affects the diversity performance of the relay channel with CSF.

We now show that the bound $D_{\text{UB-SPC-RPC}}^{\text{DSRF-}K}(r)$ can be approached by studying a particular strategy. Going back to the original relay channel, we consider the following index mapping

$$\mathcal{I}(g,\gamma_2) \equiv \mathcal{I}(g) = \begin{cases} K & \text{if } \log(1+gP_{K-1}) < r\log\text{SNR}, \\ \min\{k \in \{1,\ldots,K-1\} : \log(1+gP_k) \geq r\log\text{SNR}\} & \text{otherwise.} \end{cases}$$

That is, the source uses the smallest power in $\{P_1, \ldots, P_K\}$ so that the direct transmission succeeds (using full dimension $\beta_k = 1$, $k = 1, \ldots, K-1$, i.e. no relaying is necessary). If direct transmission would fail even with the largest power P_K, then relaying is used. In this rare event, a fraction of dimension $\beta_K = \beta$ is assigned to the learning phase.

Optimizing over β leads to the following result.

Proposition 4.8. *Assuming both power control at the source and at the relay, the proposed orthogonal scheme with destination-to-source-and-relay feedback resolution K has an outage exponent of*

$$D_{\textit{O-SPC-RPC}}^{\textit{DSRF-K}}(r) = \begin{cases} 2K - (2K+1)r & \textit{if } r < \frac{K}{K+2}, \\ \frac{2K(1-r)[K-(K-1)r]}{K-(K-2)r} & \textit{otherwise.} \end{cases} \tag{4.35}$$

The nonorthogonal scheme has an outage exponent of

$$D_{\textit{NO-SPC-RPC}}^{\textit{DSRF-K}}(r) = \begin{cases} 2K - \left(2K + \frac{\sqrt{5}-1}{2}\right)r & \textit{if } r < \frac{(3-\sqrt{5})K}{(3-\sqrt{5})K + (\sqrt{5}-1)}, \\ \frac{K(1-r)[2K-(2K-1)r)]}{K-(K-1)r} & \textit{otherwise.} \end{cases} \tag{4.36}$$

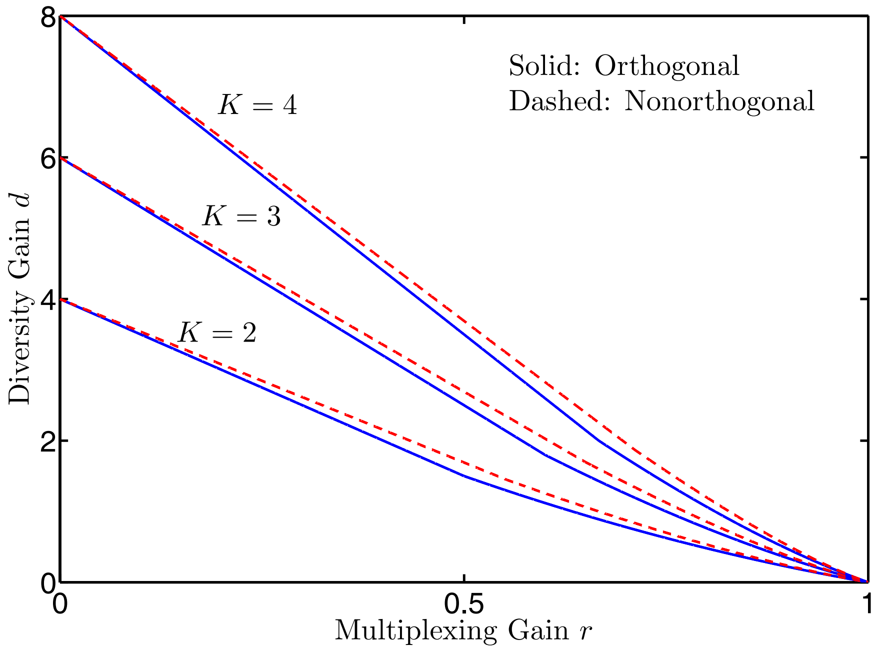

Figure 4.7: Achievable outage exponents of the proposed scheme with destination-to-source-and-relay CSF, where the feedback quantizer at the destination does not know the source–relay gain..

The proof is deferred to Appendix 4.F.

Note that in the special case $K = 1$, the outage exponents in Proposition 4.8 reduce to the corresponding ones in Proposition 4.1. This is because the proposed destination-to-source-and-relay CSF scheme essentially mimics the nonfeedback case. However the feedback scheme can use a much higher power level at both the source and the relay in order to combat outage. It is possible because at high SNR's, the relay is only *rarely* used by this scheme.

In Fig. 4.7, we plot the diversity gains of the scheme with different destination-to-source-and-relay feedback resolutions K. With any K, the achievable outage exponents are relatively close to the upper bound $2K(1-r)$ of Lemma 4.2. Clearly, this form of CSF allows for a large performance gain, as in the MIMO channels considered in Chapter 3. However, the assumption that there is a reliable direct feedback link between the destination and the source is a matter of debate, as it is exactly the lack of such a good direct link that motivates the study of the relay channel.

4.7 On Finite-length Codes

A major obstacle in finding codes and schemes with finite length that can achieve an *error probability* that decays in SNR as fast as the *outage probability* does is that with a given finite T, the dimension cannot be divided arbitrarily. In fact for a fixed T, we can only allocate the fraction of dimension from a finite, discrete set of rational numbers. In the following, we present some results towards the achievability of the outage exponents in this work with finite-length codes.

For simplicity of presentation, we first consider the orthogonal DF scheme with no CSF. The techniques used in the proof are readily extendable to the CSF cases (see Chapter 3), which we omit as they do not offer any additional insight into the problem.

Proposition 4.9. *With no CSF, there exist orthogonal codes using $T \geq 2$ channel uses that achieve the following diversity gain*

$$\max_{\beta \in \mathcal{B}} \min \left(2\left(1 - \frac{r}{\beta}\right), D_2 \right)$$

where

$$D_2 = \begin{cases} 2 - \frac{r}{\beta} & \text{if } \beta < \frac{1}{2} \\ \frac{1-r}{\beta} & \text{if } r > 1-\beta, \beta \geq \frac{1}{2}, \\ 2 - \frac{r}{1-\beta} & \text{if } r < 1-\beta, \beta \geq \frac{1}{2} \end{cases}$$

and

$$\mathcal{B} = \left\{ \frac{T_1}{T} : T_1 \in \{\max(\lceil Tr \rceil, \lfloor T/2 \rfloor), \max(\lceil Tr \rceil, \lfloor T/2 \rfloor) + 1, \ldots, T-1\} \right\}.$$

Proof. **Step 1 -** Preliminaries: Let $T_2 = T - T_1$ and $\beta = \frac{T_1}{T}$. For convenience we repeat the communication protocol herein. An encoder is a mapping from a message to a codeword in codebook $\mathcal{C}$, $m \to (\sqrt{\text{SNR}}\mathbf{x}_1, \sqrt{\text{SNR}}\mathbf{x}_\text{R})$ where $\mathbf{x}_1 = [x_1(1) \quad \cdots \quad x_1(T_1)]$ and $\mathbf{x}_\text{R} = [x_\text{R}(1) \quad \cdots \quad x_\text{R}(T_2)]$ are the normalized codewords. Note that $\mathcal{C}$ is simply the combination of $\mathcal{C}_1, \mathcal{C}_\text{R}$ in Section 4.3, normalized by $\sqrt{\text{SNR}}$. The individual power constraints read

$$\frac{1}{|\mathcal{C}|T} \sum_{\mathcal{C}} \sum_{n=1}^{T_1} |x_1(n)|^2 \leq 1,$$

$$\frac{1}{|\mathcal{C}|T} \sum_{\mathcal{C}} \sum_{n=1}^{T_2} |x_\text{R}(n)|^2 \leq 1.$$

Recall that the destination knows whether the source-relay link is in outage or not, i.e., if $\beta \log(1 + \gamma_1 \text{SNR}) \geq r \log \text{SNR}$. If $\beta \log(1 + \gamma_1 \text{SNR}) \geq r \log \text{SNR}$ then the entire received signals (a sequence of length T) are used for decoding at the destination, otherwise only the received signals in the learning phase (length T_1) are used for decoding.

Let us define three events $\mathcal{A} = \{\beta \log(1+\mathrm{SNR}g) < r \log \mathrm{SNR}\}$, $\mathcal{A}_1 = \{\beta \log(1+\mathrm{SNR}\gamma_1) < r \log \mathrm{SNR}\}$, and $\mathcal{A}_2 = \{\beta \log(1+\mathrm{SNR}g) + (1-\beta)\log(1+\mathrm{SNR}\gamma_2) < r \log \mathrm{SNR}\}$. The probability of *error*, over the randomness of the channel gains, the noise, and the (uniformly distributed) messages can be written as

$$
\begin{aligned}
\Pr(\varepsilon) &= \Pr(\varepsilon, \mathcal{A}_1) + \Pr(\varepsilon, \bar{\mathcal{A}}_1) \\
&= \Pr(\varepsilon, \mathcal{A}_1, \mathcal{A}) + \Pr(\varepsilon, \mathcal{A}_1, \bar{\mathcal{A}}) + \Pr(\varepsilon, \bar{\mathcal{A}}_1, \mathcal{A}_2) + \Pr(\varepsilon, \bar{\mathcal{A}}_1, \bar{\mathcal{A}}_2) \\
&\le \Pr(\mathcal{A}_1, \mathcal{A}) + \Pr(\varepsilon, \mathcal{A}_1, \bar{\mathcal{A}}) + \Pr(\bar{\mathcal{A}}_1, \mathcal{A}_2) + \Pr(\varepsilon, \bar{\mathcal{A}}_1, \bar{\mathcal{A}}_2, \bar{\varepsilon}_{\mathrm{R}}) \\
&\qquad\qquad + \Pr(\varepsilon, \bar{\mathcal{A}}_1, \bar{\mathcal{A}}_2, \varepsilon_{\mathrm{R}}) \\
&\le \Pr(\mathcal{A}_1, \mathcal{A}) + \Pr(\varepsilon, \bar{\mathcal{A}} | \Delta_{\mathrm{D}}) \Pr(\mathcal{A}_1) + \Pr(\bar{\mathcal{A}}_1, \mathcal{A}_2) \\
&\qquad\qquad + \Pr(\varepsilon, \bar{\mathcal{A}}_2 | \Delta_{\mathrm{C}}, \bar{\varepsilon}_{\mathrm{R}}) \Pr(\bar{\mathcal{A}}_1, \bar{\varepsilon}_{\mathrm{R}}) + \Pr(\bar{\mathcal{A}}_1, \varepsilon_{\mathrm{R}})
\end{aligned}
$$

where ε denotes the error event at the destination $\hat{m} \ne m$, Δ_{D} denotes the event that the destination uses only the directly transmitted sequence to decode, Δ_{C} is the event that the combined received sequences are used for detection. Note that $\varepsilon = \{\varepsilon, \Delta_{\mathrm{D}}\} \cup \{\varepsilon, \Delta_{\mathrm{C}}\}$. Finally ε_{R} is the event that the relay makes an incorrect decision $m_{\mathrm{R}} \ne m$. In the last inequality we use $\Pr(\varepsilon, \bar{\mathcal{A}} | \mathcal{A}_1) = \Pr(\varepsilon, \bar{\mathcal{A}} | \Delta_{\mathrm{D}})$ and $\Pr(\varepsilon, \bar{\mathcal{A}}_2 | \bar{\mathcal{A}}_1, \bar{\varepsilon}_{\mathrm{R}}) = \Pr(\varepsilon, \bar{\mathcal{A}}_2 | \Delta_{\mathrm{C}}, \bar{\varepsilon}_{\mathrm{R}})$, which follow the protocol of the scheme.

Inspired by the concept of approximately universal codes [TV06], we will next find sufficient conditions on the (sequence of) codes $\mathcal{C}$ so that $\Pr(\varepsilon, \bar{\mathcal{A}} | \Delta_{\mathrm{D}}) \doteq \mathrm{SNR}^{-\infty}$, $\Pr(\varepsilon_{\mathrm{R}}, \bar{\mathcal{A}}_1) \doteq \mathrm{SNR}^{-\infty}$, and $\Pr(\varepsilon, \bar{\mathcal{A}}_2 | \Delta_{\mathrm{C}}, \bar{\varepsilon}_{\mathrm{R}}) \doteq \mathrm{SNR}^{-\infty}$. Recall that $f(\mathrm{SNR}) \doteq \mathrm{SNR}^{-\infty}$ denotes an exponentially decaying function of SNR. If such a sequence of codes exists, then

$$\Pr(\varepsilon) \,\dot{\le}\, \Pr(\mathcal{A}_1, \mathcal{A}) + \Pr(\bar{\mathcal{A}}_1, \mathcal{A}_2)$$

and by a similar argument as in the outage analysis we readily obtain the results in Proposition 4.9, which essentially formulate the maximization of the dominant SNR exponents of $\Pr(\mathcal{A}_1, \mathcal{A}) + \Pr(\bar{\mathcal{A}}_1, \mathcal{A}_2)$.

At this point, we notice that a simple "shortcut" argument may conclude the proof. The two-phase orthogonal transmission resembles T parallel scalar Gaussian channels with one channel use. The first T_1 scalar channels have a common channel gain h while the remaining T_2 channels have a common channel gain h_2. But there exist approximately universal codes $\mathcal{C}_{\mathrm{P}}$ of length 1 for parallel scalar Gaussian channels [TV06], thus by definition $\Pr(\varepsilon, \bar{\mathcal{A}}_2 | \Delta_{\mathrm{C}}, \bar{\varepsilon}_{\mathrm{R}}) \doteq \mathrm{SNR}^{-\infty}$ for those codes. Furthermore, $\Pr(\varepsilon_{\mathrm{R}}, \bar{\mathcal{A}}_1) \doteq \mathrm{SNR}^{-\infty}$ also holds for the class of codes $\mathcal{C}_{\mathrm{P}}$, which can be seen by considering the special case $h_2 = 0$ (and $\Pr(\varepsilon, \bar{\mathcal{A}} | \Delta_{\mathrm{D}}) \doteq \mathrm{SNR}^{-\infty}$ holds with $\mathcal{C}_{\mathrm{P}}$ due to symmetry).

In practical terms, one way of achieving the tradeoff in Proposition 4.9 is to use a code of length 1 that is approximately universal for T parallel SISO channels.

In the learning phase, the source sends the first T_1 symbols (which now span in time rather than in different parallel channels). In the relaying phase, the relay transmits the remaining T_2 symbols (if the source-relay link is not in outage). An interesting feature of this approach is that the relay and destination can employ identical decoders. This is especially useful in scenarios where any single node can alternatively act either as relay or destination.

Note however that the above arguments do not directly extend to the nonorthogonal case, where the two-phase transmission resembles a mixture of SISO and MISO parallel channels. Furthermore, using existing codes for parallel channels is not necessarily the only way to achieve the tradeoffs in Proposition 4.9. Thus in the following, we present a more general approach to Proposition 4.9 using standard Gaussian arguments.

Step 2 - Universal conditions: Let $(\Delta\mathbf{x}_1, \Delta\mathbf{x}_\mathrm{R})$ be a nonzero codeword difference vector and perform a change of variable so that $\|\Delta\mathbf{x}_1\|^2 = \mathrm{SNR}^{-\delta_1}$ and $\|\Delta\mathbf{x}_\mathrm{R}\|^2 = \mathrm{SNR}^{-\delta_\mathrm{R}}$ with the convention $\delta_1 = \infty$ if $\Delta\mathbf{x}_1 = \mathbf{0}$ and $\delta_\mathrm{R} = \infty$ if $\Delta\mathbf{x}_\mathrm{R} = \mathbf{0}$.

We claim that if the following two conditions are satisfied

$$\left(\min_{\mathcal{C}} \mathrm{SNR}^{-\beta(\delta_1)^+}\right) \dot{\geq} \mathrm{SNR}^{-r} \tag{4.37}$$

$$\left(\min_{\mathcal{C}} \mathrm{SNR}^{-\beta(\delta_1)^+}\mathrm{SNR}^{-(1-\beta)(\delta_\mathrm{R})^+}\right) \dot{\geq} \mathrm{SNR}^{-r} \tag{4.38}$$

then $\Pr(\varepsilon, \bar{\mathcal{A}}|\Delta_\mathrm{D}) \doteq \mathrm{SNR}^{-\infty}$, $\Pr(\varepsilon_\mathrm{R}, \bar{\mathcal{A}}_1) \doteq \mathrm{SNR}^{-\infty}$, and $\Pr(\varepsilon, \bar{\mathcal{A}}_2|\Delta_\mathrm{C}, \bar{\varepsilon}_\mathrm{R}) \doteq \mathrm{SNR}^{-\infty}$. Recall that by convention $0 \doteq \mathrm{SNR}^{-\infty}$, thus we must have $T_1 \geq 1, T - T_1 \geq 1$ otherwise (4.38) cannot be satisfied.

We now prove the above claim. Consider maximum likelihood decoding at both relay and the destination, and the following conditional pairwise error probability

$$\begin{aligned}\Pr(\text{pairwise}_\text{relay}|h_1) = \Pr(\mathbf{x}_1 \to \hat{\mathbf{x}}_1|h_1) &\leq \exp\left(-\frac{\mathrm{SNR}}{2}|h_1|^2\|\Delta\mathbf{x}_1\|^2\right)\\ &= \exp\left(-\frac{\mathrm{SNR}}{2}\mathrm{SNR}^{-\alpha_1}\mathrm{SNR}^{-\delta_1}\right)\\ &= \exp\left(-\frac{1}{2}\mathrm{SNR}^{(1-\alpha_1)-\delta_1}\right).\end{aligned}$$

Averaging over $\alpha_1 \in \bar{\mathcal{A}}_1$ gives

$$\Pr(\text{pairwise}_\text{relay}, \bar{\mathcal{A}}_1) = \int_{\alpha_1 \in \bar{\mathcal{A}}_1} \Pr(\text{pairwise}_\text{relay}|\alpha_1) f(\alpha_1) d\alpha_1$$

where $f(\alpha_1)$ is the p.d.f. of α_1. The subset of channel realizations in $\bar{\mathcal{A}}_1$ making $\Pr(\text{pairwise}_\text{relay}, \bar{\mathcal{A}}_1)$ decay polynomially in SNR is

$$\mathcal{H}_1 = \{\alpha_1 \geq 0 : (1-\alpha_1) \leq \delta_1\} \cap \bar{\mathcal{A}}_1 = \{\alpha_1 \geq 0 : (1-\alpha_1) \leq \delta_1, \beta(1-\alpha_1)^+ \geq r + \epsilon\}.$$

But due to condition (4.37), there exists a $\bar{\text{SNR}} > 0$ so that $\forall \text{SNR} > \bar{\text{SNR}}$ we have

$$\text{SNR}^{-\beta(\delta_1)^+} \geq \text{SNR}^{-r-\frac{\epsilon}{2}}$$

and thus $\beta(\delta_1)^+ \leq r + \frac{\epsilon}{2} < \beta(1-\alpha_1)^+$. This however cannot be satisfied because $\forall \alpha_1 \in \mathcal{H}_1$, $(1-\alpha_1) \leq \delta_1$ leading to $(1-\alpha_1)^+ \leq (\delta_1)^+$. We conclude that $\mathcal{H}_1$ is empty, and thus every pairwise error probability, averaged over $\bar{\mathcal{A}}_1$, decays exponentially in SNR. Since the number of codewords SNR^{rT} grow only polynomially in SNR, applying the union bound still leads to an exponentially vanishing average error probability, $\Pr(\varepsilon_{\text{R}}, \bar{\mathcal{A}}_1) \doteq \text{SNR}^{-\infty}$. Due to the statistical symmetry of the channel, (4.37) also guarantees that $\Pr(\varepsilon, \bar{\mathcal{A}}|\Delta_{\text{D}}) \doteq \text{SNR}^{-\infty}$.

Similarly, by considering the pairwise error probability at the destination

$$\begin{aligned}\Pr(\text{pairwise}_{\text{dst}}|\Delta_{\text{C}}, \bar{\varepsilon}_{\text{R}}, \gamma, \gamma_2) &\leq \exp\left(-\frac{\text{SNR}}{2}\left(|h|^2\|\Delta\mathbf{x}_1\|^2 + |h_2|^2\|\Delta\mathbf{x}_{\text{R}}\|^2\right)\right) \\ &= \exp\left(-\frac{1}{2}\text{SNR}^{(1-\alpha)-\delta_1}\right)\exp\left(-\frac{1}{2}\text{SNR}^{(1-\alpha_2)-\delta_{\text{R}}}\right)\end{aligned}$$

and averaging over the nonoutage region $\bar{\mathcal{A}}_2 = \{\alpha, \alpha_2 : (1-\beta)(1-\alpha)^+ + \beta(1-\alpha_2)^+ \geq r+\epsilon\}$, we find that the subset of bad channel realizations in $\bar{\mathcal{A}}_2$ is

$$\mathcal{H}_2 = \{\alpha \geq 0, \alpha_2 \geq 0 : (1-\alpha_1) \leq \delta_1, (1-\alpha_2) \leq \delta_{\text{R}}, (1-\beta)(1-\alpha)^+ + \beta(1-\alpha_2)^+ \geq r+\epsilon\}.$$

The condition (4.38) means that $\forall \text{SNR} > \tilde{\text{SNR}}$ for a certain finite $\tilde{\text{SNR}}$, we have $\beta(\delta_1)^+ + (1-\beta)(\delta_{\text{R}})^+ \leq r + \frac{\epsilon}{2}$. This leads to a contradiction of the equations defining $\mathcal{H}_2$ and thus the subset $\mathcal{H}_2$ making the average error probability $\Pr(\text{pairwise}_{\text{destination}}, \bar{\mathcal{A}}_2|\Delta_{\text{C}})$ decay only polynomially in SNR is empty. This leads to $\Pr(\varepsilon, \bar{\mathcal{A}}_2|\Delta_{\text{C}}, \bar{\varepsilon}_{\text{R}}) \doteq \text{SNR}^{-\infty}$.

This concludes the proof of our claim regarding the sufficient conditions (4.37) and (4.38).

Step 3 - Expurgation: We now draw the normalized codewords $(\mathbf{x}_1, \mathbf{x}_{\text{R}})$ from an i.i.d. zero-mean unit-variance complex Gaussian ensemble. The codebook size is $|\mathcal{C}| = \text{SNR}^{(r-\epsilon)T}$ with an arbitrarily small $\epsilon > 0$. Then we expurgate all bad codewords that lead to the violation of at least one of the approximately universal conditions [Gal65, ZT03].

In particular, we fix the first codeword $\mathbf{x}^0$ and compute all codeword differences $\mathbf{x} - \mathbf{x}^0 = (\Delta\mathbf{x}_1, \Delta\mathbf{x}_{\text{R}}) = (\text{SNR}^{-\delta_1}, \text{SNR}^{-\delta_{\text{R}}})$. We expurgate $\mathbf{x}^0$ if there exists another codeword so that $(\delta_1, \delta_{\text{R}}) \in \mathcal{B}_1 \cup \mathcal{B}_2$ where

$$\mathcal{B}_1 = \left\{\delta_1, \delta_{\text{R}} : \text{SNR}^{-\beta(\delta_1)^+} \leq \text{SNR}^{-r}\right\},$$

$$\mathcal{B}_2 = \{\delta_1, \delta_{\text{R}} : \text{SNR}^{-\beta(\delta_1)^+ - (1-\beta)(\delta_{\text{R}})^+} \leq \text{SNR}^{-r}\}.$$

We then repeat the process for $\mathbf{x}^1, \ldots, \mathbf{x}^{|\mathcal{C}|-1}$.

Over the ensemble, the codeword difference $\Delta\mathbf{x}_1$, $\Delta\mathbf{x}_{\mathrm{R}}$ are i.i.d. zero-mean variance-2 complex Gaussian random vectors of length T_1 and T_2 respectively. We then have $\Pr\left((\delta_1, \delta_{\mathrm{R}}) \in \mathcal{B}_1 \cup \mathcal{B}_2\right) \doteq \mathrm{SNR}^{-D_{\mathrm{e}}}$ where

$$D_{\mathrm{e}} = \inf_{(\delta_1 \geq 0, \delta_{\mathrm{R}} \geq 0) \in \{\mathcal{B}_1 \cup \mathcal{B}_2\}} T_1\delta_1 + T_2\delta_{\mathrm{R}} = \inf_{(\delta_1 \geq 0, \delta_{\mathrm{R}} \geq 0) \in \{\mathcal{B}_1 \cup \mathcal{B}_2\}} T(\beta\delta_1 + (1-\beta)\delta_{\mathrm{R}})$$

However for $\mathrm{SNR} > 1$, we have

$$\mathcal{B}_1 = \{\delta_1, \delta_{\mathrm{R}} : \beta(\delta_1)^+ \geq r\},$$

$$\mathcal{B}_2 = \{\delta_1, \delta_{\mathrm{R}} : \beta(\delta_1)^+ + (1-\beta)(\delta_{\mathrm{R}})^+ \geq r\}.$$

Clearly $\mathcal{B}_1 \cup \mathcal{B}_2 = \mathcal{B}_2$, leading to $D_{\mathrm{e}} = Tr$. The probability that we expurgate $\mathbf{x}^0$ is then union-bounded by

$$\Pr(\mathbf{x}^0 \text{ expurgated}) \dot{\leq} \mathrm{SNR}^{(r-\epsilon)T}\mathrm{SNR}^{-D_{\mathrm{e}}} = \mathrm{SNR}^{-T\epsilon}.$$

Let the indicator function be $\mathbf{1}(\cdot)$. Then the average (over the ensemble) number of codewords after expurgation is

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=0}^{|\mathcal{C}|-1}\left(1 - \mathbf{1}(\mathbf{x}^i \text{ expurgated})\right)\right] &= |\mathcal{C}| - \sum_{i=0}^{|\mathcal{C}|-1} \mathrm{E}\left[\mathbf{1}(\mathbf{x}^i \text{ expurgated})\right] \\
&= |\mathcal{C}| - \sum_{i=0}^{|\mathcal{C}|-1} \Pr\left(\mathbf{x}^i \text{ expurgated}\right) \\
&\geq |\mathcal{C}| - |\mathcal{C}| \Pr\left(\mathbf{x}^0 \text{ expurgated}\right) \\
&\doteq |\mathcal{C}|\left(1 - \mathrm{SNR}^{-T\epsilon}\right) \\
&\doteq |\mathcal{C}|
\end{aligned}$$

where the inequality is due to the fact that the first codeword $\mathbf{x}^0$ is more likely to be expurgated than any other codeword.

The derivations above show that the expurgation progress does not lead to a loss of multiplexing gain (rate) of at least one code in the ensemble. Because ϵ can be made arbitrarily close to zero, we conclude that in codes drawn from Gaussian ensemble there exists at least a code of rate $r \log \mathrm{SNR}$ that satisfies both sufficient conditions, provided that $T_1 \geq 1$ and $T_2 = T - T_1 \geq 1$. □

Since the two-phase transmission resembles Gaussian channels in parallel, we can readily extend the above expurgation process to the case of an arbitrary number of parallel channels and even multiple receive antennas. We then obtain an interesting by-product:

Corollary 4.3. *For parallel SIMO Gaussian channels, there exists at least a sequence of approximately universal codes with length 1. The codes can be obtained by expurgating codes drawn from an i.i.d. Gaussian ensemble.*

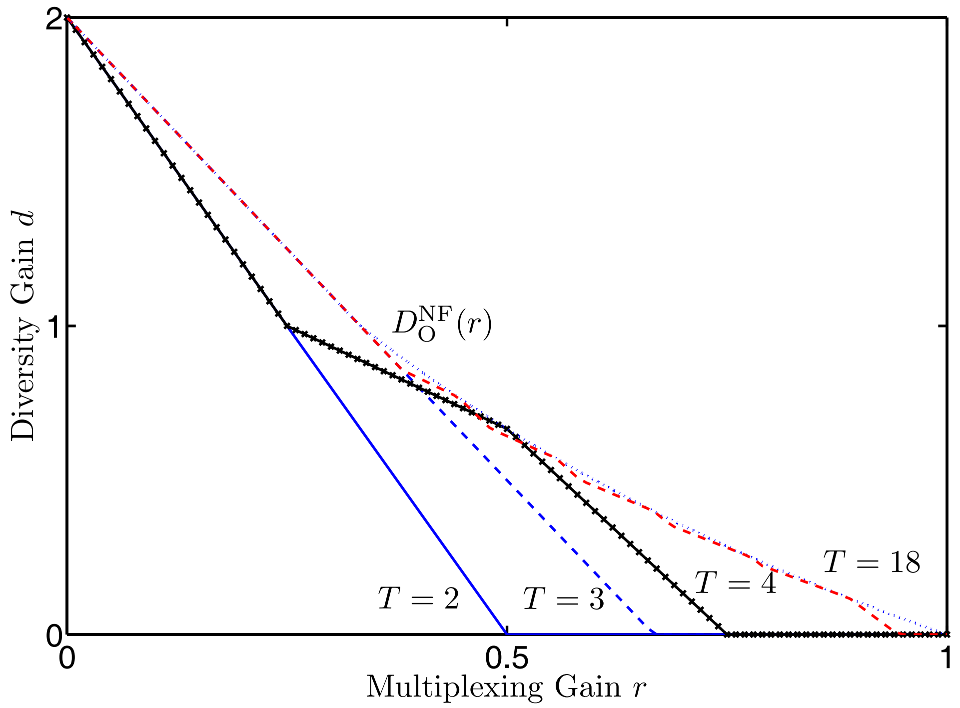

Figure 4.8: Achievable diversity-multiplexing tradeoff of the orthogonal scheme with no CSF and different finite codeword lengths T. The dotted curve is the outage exponent $D_{\mathrm{O}}^{\mathrm{NF}}(r)$ of Proposition 4.1.

Note that the fact that approximately universal codes of length 1 exist has been established in [TV06] by a QAM permutation argument.

In Fig. 4.8, we plot the achievable diversity-multiplexing tradeoff curves with different codeword length T. Some useful conclusions can be drawn from the results. First, there exist codes achieving the linear segment $2 - 3r$ of the optimal outage exponent $D_{\mathrm{O}}^{\mathrm{NF}}(r)$, given that $T = 3N$, $\forall N \in \{1, 2, \ldots\}$. And even with a moderate length $T = 18$, very close to outage performance at all r can be achieved. Note that due to the discreteness of the set $\mathcal{B}$, increasing the length T does not necessarily lead to a better performance at *all* multiplexing gains. Finally, since the relay is required to decode the message m completely, we cannot achieve a nonzero diversity gain for any $r \geq \frac{T-1}{T}$.

It is possible to apply the above expurgation process to the nonorthogonal DF case. However, it turns out that expurgating "bad" codewords from a Gaussian ensemble leads to a loss of rate, i.e. the number of remaining codewords is not large enough to carry out the information rate $r \log \mathrm{SNR}$ of the original codebook.

This is due to the fact that the relaying phase resembles a multiple transmit antenna system, where the expurgation process is known to incur loss of rate. Thus at any multiplexing gain, the Gaussian coding bound in this case is only asymptotically tight as $T \to \infty$. We state the result in the following, and present a sketch of the proof in Appendix 4.G.

Proposition 4.10. *With no CSF, there exist nonorthogonal codes using $T \geq 3$ channel uses that achieve the following diversity gain*

$$\max_{\beta \in \mathcal{B}} \min\left(2 - \bar{r} - \frac{\bar{r}}{\beta}, D_2\right)$$

where

$$D_2 = \begin{cases} 2 - 2\bar{r} & \text{if } \beta < \frac{1}{2} \\ \frac{1-\bar{r}}{\beta} & \text{if } \bar{r} > 1 - \beta, \beta \geq \frac{1}{2} \\ 2 - \frac{\bar{r}}{1-\beta} & \text{if } \bar{r} < 1 - \beta, \beta \geq \frac{1}{2} \end{cases}$$

with $\bar{r} = \frac{r}{1 - \frac{1}{(1-\beta)T}}$ and

$$\mathcal{B} = \left\{ \frac{T_1}{T} : T_1 \in \{\max(\lceil T\bar{r} \rceil, \lfloor T/2 \rfloor), \max(\lceil T\bar{r} \rceil, \lfloor T/2 \rfloor), \ldots, \min\left(T - 2, \left\lfloor T - \frac{1}{1-r} \right\rfloor\right)\right\} \right\}.$$

Unlike in the orthogonal case, an achievable curve with a finite T never "touches" the outage upper bound $D_{\mathrm{NO}}^{\mathrm{NF}}(r)$ because of the loss of rate in expurgation. In this case, it is potential to get close to the outage upper bound and better than the orthogonal upper bound $D_{\mathrm{O}}^{\mathrm{NF}}(r)$) with relatively large codeword lengths ($T \approx 100$). For short codes the achievable bound is loose, and it is likely that using more structured codes will give a much better performance than relying on the Gaussian coding arguments herein, as evident in the MIMO case [TV06, EKP$^+$06].

4.8 Conclusion

We provide a comprehensive study of decode-and-forward relay channels with different forms of quantized channel state feedback. Our results suggest that with the help of severely limited channel state feedback, the performance of a three-node cooperative communication system can be considerably improved even with relatively simple decode-and-forward schemes. The construction of explicit codes that achieve or approach the performance of the outage bounds presented in this work, especially in the nonorthogonal case, remains a topic for future study.

Appendices for Chapter 4

4.A Proof of Proposition 4.2

For simplicity of presentation, let us consider $K = 2$. Generalizing to $K > 2$ is straightforward. Furthermore, we consider only an orthogonal scheme, as the proof for a nonorthogonal scheme is completely similar.

The outage probability is

$$P_{\text{out}}(r \log \text{SNR}) = \Pr(\mathcal{O}_1) + \Pr(\mathcal{O}_2) + \Pr(\mathcal{O}_2)$$

where

$$\begin{aligned}\Pr(\mathcal{O}_1) &= \Pr(\mathcal{I} = 1, \beta_1 \log(1 + g\text{SNR}) + (1 - \beta_1) \log(1 + \gamma_2 \text{SNR}) < r \log \text{SNR}) \\ &\doteq \Pr(\beta_1 (1 - \alpha_1)^+ \geq r, \beta_1 (1 - a)^+ + (1 - \beta_1)(1 - \alpha_2)^+ < r) \\ &\doteq \text{SNR}^{-D_1}.\end{aligned}$$

Invoking Lemma 4.1 yields

$$D_1 = \begin{cases} 2 - \frac{r}{\beta_1} & \text{if } r < \beta_1 < \frac{1}{2}, \\ \frac{1-r}{\beta_1} & \text{if } r \in (1 - \beta_1, \beta_1), \beta_1 \geq \frac{1}{2}, \\ 2 - \frac{r}{1-\beta_1} & \text{if } r < 1 - \beta_1, \beta_1 \geq \frac{1}{2}. \end{cases} \tag{4.39}$$

Also,

$$\begin{aligned}\Pr(\mathcal{O}_2) &= \Pr(\mathcal{I} = 2, \beta_2 \log(1 + \gamma_1 \text{SNR}) \geq r \log \text{SNR}, \\ &\qquad \beta_2 \log(1 + g\text{SNR}) + (1 - \beta_2) \log(1 + \gamma_2 \text{SNR}) < r \log \text{SNR}) \\ &\doteq \Pr(\beta_1 (1 - \alpha_1)^+ < r, \beta_2 (1 - \alpha_1)^+ \geq r, \\ &\qquad \beta_2 (1 - a)^+ + (1 - \beta_2)(1 - \alpha_2)^+ < r),\end{aligned}$$

leading to

$$D_2 = \begin{cases} 2 - \frac{r}{\beta_2} + 1 - \frac{r}{\beta_1} & \text{if } r < \beta_2 < \frac{1}{2}, \\ \frac{1-r}{\beta_2} + 1 - \frac{r}{\beta_1} & \text{if } r \in (1 - \beta_2, \beta_2), \beta_2 \geq \frac{1}{2}, \\ 2 - \frac{r}{1-\beta_2} + 1 - \frac{r}{\beta_1} & \text{if } r < 1 - \beta_2, \beta_2 \geq \frac{1}{2}. \end{cases} \tag{4.40}$$

Finally

$$\begin{aligned}\Pr(\mathcal{O}_3) &= \Pr(\mathcal{I} = 2, \beta_2 \log(1 + \gamma_1 \text{SNR}) < r \log \text{SNR}, \beta_2 \log(1 + g\text{SNR}) < r \log \text{SNR}) \\ &\doteq \Pr(\beta_1 (1 - \alpha_1)^+ < r, \beta_2 (1 - \alpha_1)^+ < r, \beta_2 (1 - \alpha)^+ < r).\end{aligned}$$

This leads to

$$D_3(r) = 2\left(1 - \frac{r}{\beta_2}\right). \tag{4.41}$$

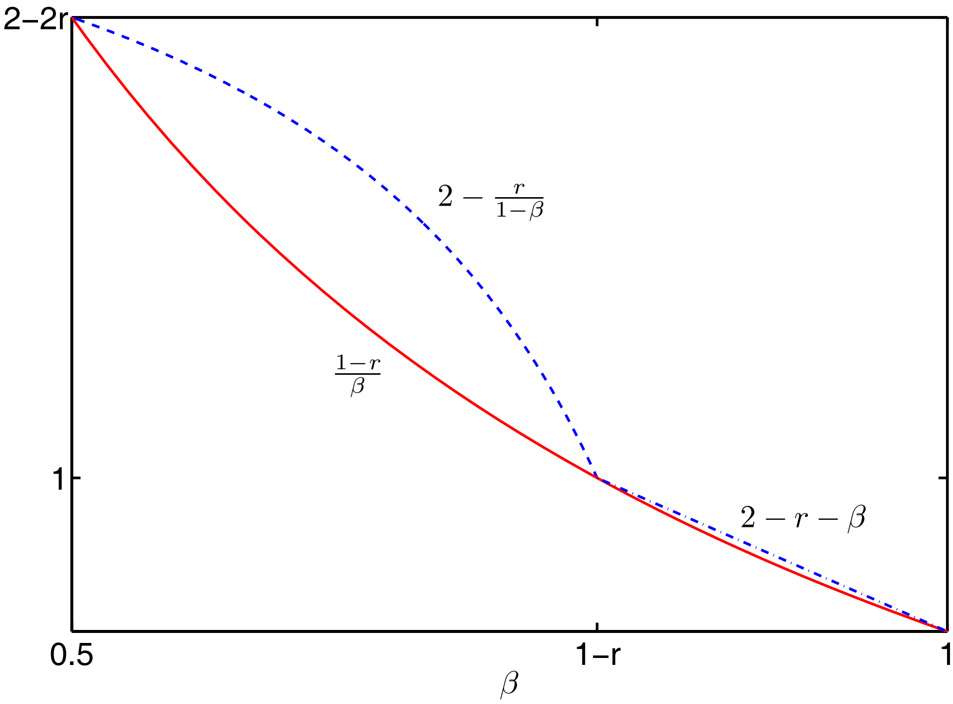

Figure 4.9: Bounding techniques in Proposition 4.2.

We again only need to consider $\beta_k > 1/2$ (cf. Section 4.3). The outage exponent corresponding to an *optimal* dimension allocation is therefore given by

$$D^*_{\text{out}}(r) = \sup_{\max(r,1/2)\leq\beta_1<\beta_2<1} \min(D_1, D_2, D_3).$$

□

4.B Proof of Proposition 4.3

The upper bound is obtained by showing that

$$2 - r - \beta \geq \frac{1-r}{\beta}$$

for all $\beta \in [1-r, 1)$. To that end, noticing that $\frac{1-r}{\beta}$ is convex. Thus over $\beta \in [1-r, 1)$, $\frac{1-r}{\beta}$ is upper-bounded by the linear segment between the two end points $(1-r, 1)$ and $(1, 1-r)$. This linear segment is given by $2-r-\beta$ (cf. Fig. 4.9).

As for the lower bound, we begin with the following lemma.

Lemma 4.3. *The outage exponent* $D^{RF-K}_{O\text{-}NPC}(r)$ *is lower-bounded by*

$$D^{RF-K}_{LB\text{-}O\text{-}NPC}(r) = \sup_{\min(1/r,2)>x_1>\cdots>x_K\geq 1} \min\{(1-r)x_1+1-rx_0,\ldots,\\ (1-r)x_K+1-rx_{K-1}, 2(1-rx_K)\}$$

with the convention $rx_0 = 1$. *For* $r \geq 1/2$*, this lower bound is tight.*

The outage exponent $D^{RF-K}_{NO\text{-}NPC}(r)$ *is lower-bounded by*

$$D^{RF-K}_{LB\text{-}NO\text{-}NPC}(r) = \sup_{\min(1/r,2)>x_1>\cdots>x_K\geq 1} \min\{(1-r)x_1+1-rx_0,\ldots,\\ (1-r)x_K+1-rx_{K-1}, 2-r-rx_K)\}$$

with the convention $rx_0 = 1$. *For* $r \geq 1/2$ *this lower bound is tight.*

Proof. Consider the orthogonal case. We lower-bound the piecewise function

$$f(\beta) = \begin{cases} 2 - \frac{r}{1-\beta_k} - & \text{if } \beta < 1-r, \\ \frac{1-r}{\beta} & \text{otherwise,} \end{cases}$$

by *extending* the function $\frac{1-r}{\beta}$ over the entire support $[\max(r, 0.5), 1)$, as illustrated in Fig. 4.9. By construction, the bound is tight when $r \geq 1/2$. As for $r < 1/2$, we need to show that

$$f_1(\beta) \equiv 2 - \frac{r}{1-\beta} \geq \frac{1-r}{\beta} \equiv f_2(\beta), \forall \beta \in (1/2, 1-r).$$

To that end, note that both $f_1(\beta)$ and $f_2(\beta)$ pass through the points $(1/2, 2-2r)$ and $(1-r, 1)$. But $f_1(\beta)$ is concave and $f_2(\beta)$ is convex over $[1/2, 1-r)$, we readily have $f_1(\beta) \geq f_2(\beta)$. Finally a change of variable $x_k = 1/b_k$ gives the claimed bound.

The bounds to $D^{\mathrm{RF}-K}_{\mathrm{NO\text{-}NPC}}(r)$ are obtained in a similar manner. □

We now compute the explicit solution to the optimization problems posed in Lemma 4.3. Consider the orthogonal case. The problem can in fact be recast as a linear program. The *global* optimum of the maximin is given by the intersection of all $K+1$ affine functions, provided that the intersection satisfies the linear constraints $\min(1/r, 2) > x_1 > \cdots > x_K \geq 1$. Let us find this balancing point. Introduce a variable Δ so that

$$\begin{aligned} (1-r)x_1 &= \Delta \\ (1-r)x_2 + 1 - rx_1 &= \Delta \\ &\cdots \\ (1-r)x_K + 1 - rx_{K-1} &= \Delta \\ 2(1-rx_K) &= \Delta. \end{aligned}$$

From the last equation we have

$$x_K = \frac{2-\Delta}{2r}. \tag{4.42}$$

From the first K equations, we have

$$x_2 = \frac{\Delta-1}{1-r} + \frac{r}{1-r}x_1 = \frac{\Delta-1}{1-r} + \frac{r}{(1-r)^2}\Delta = \frac{\Delta-1}{1-r}\left(1+\frac{r}{1-r}\right) + \frac{r}{(1-r)^2}$$

$$x_3 = \frac{\Delta-1}{1-r}\left(1+\frac{r}{1-r}+\left(\frac{r}{1-r}\right)^2\right) + \frac{r^2}{(1-r)^3}$$

$$\ldots$$

$$x_K = \frac{\Delta-1}{1-r}\sum_{k=0}^{K-1}\left(\frac{r}{1-r}\right)^k + \frac{r^{K-1}}{(1-r)^K}.$$

Combining with (4.42) leads to

$$\Delta = \frac{\frac{1}{r} + \frac{1}{1-r}\sum_{k=0}^{K-1}\left(\frac{r}{1-r}\right)^k - \frac{r^{K-1}}{(1-r)^K}}{\frac{1}{2r} + \frac{1}{1-r}\sum_{k=0}^{K-1}\left(\frac{r}{1-r}\right)^k}$$

$$= \begin{cases} \frac{2(1-r)-2(1-r)\left(\frac{r}{1-r}\right)^K}{1-2r\left(\frac{r}{1-r}\right)^K} & \text{if } r \neq 1/2 \\ \frac{2K}{1+2K} & \text{if } r = 1/2. \end{cases}$$

It can be verified that with this value of Δ, all the linear constraints are satisfied. Thus $D^{\text{RF}-K}_{\text{LB-O-NPC}}(r) = \Delta$.

Using a similar technique, we can show that

$$D^{\text{RF}-K}_{\text{LB-NO-NPC}}(r) = \frac{\frac{2-r}{r} + \frac{1}{1-r}\sum_{k=0}^{K-1}\left(\frac{r}{1-r}\right)^k - \frac{r^{K-1}}{(1-r)^K}}{\frac{1}{2r} + \frac{1}{1-r}\sum_{k=0}^{K-1}\left(\frac{r}{1-r}\right)^k}$$

$$= \begin{cases} \frac{2(1-r)^2-(1-r)\left(\frac{r}{1-r}\right)^K}{1-r-r\left(\frac{r}{1-r}\right)^K} & \text{if } r \neq 1/2 \\ \frac{1+2K}{2+2K} & \text{if } r = 1/2. \end{cases}$$

□

4.C Proof of Proposition 4.4

We essentially follow the same line of arguments as in the short-term case. The key difference is the fact that under a long-term power constraint, a power level in the order of

$$\frac{\text{SNR}}{\Pr(\mathcal{I}(\gamma_1) = k)}$$

can be applied given that the index $\mathcal{I}(\gamma_1) = k$ is received at the source [ECD06].

Again we only present the case $K = 2$ to improve readability. Consider an orthogonal scheme. Let p_k be the SNR exponent of power level P_k, i.e. $P_k \doteq \mathrm{SNR}^{p_k}$, with $1 = p_1 < p_2 < \infty$. We have $P_{\mathrm{out}}(r \log \mathrm{SNR}) = \Pr(\mathcal{O}_1) + \Pr(\mathcal{O}_2) + \Pr(\mathcal{O}_3)$. Herein

$$\Pr(\mathcal{O}_1) = \Pr(\mathcal{I} = 1, \beta_1 \log(1+gP_1)+(1-\beta_1)\log(1+\gamma_2\mathrm{SNR}) < r\log\mathrm{SNR}) \doteq \mathrm{SNR}^{-d1}$$

where

$$D_1 = \begin{cases} \frac{1-r}{\beta_1} & \text{if } r \in (1-\beta_1, \beta_1), \beta_1 \geq \frac{1}{2}, \\ 2 - \frac{r}{1-\beta_1} & \text{if } r < 1-\beta_1, \beta_1 \geq \frac{1}{2}. \end{cases}$$

Notice that $P_1 \doteq \mathrm{SNR}^1$ and that we do not consider $\beta_1 < 1/2$ for the same reason as presented in Section 4.3.

We now have

$$\begin{aligned} \Pr(\mathcal{I} = 2) &= \Pr(\beta_1 \log(1+\gamma_1\mathrm{SNR}) < r\log\mathrm{SNR}, \beta_2\log(1+\gamma_1 P_2) \geq r\log\mathrm{SNR}) \\ &\doteq \mathrm{SNR}^{-\left(1-\frac{r}{\beta_1}\right)} \end{aligned}$$

where the exponent equality is due to $p_2 - \frac{r}{\beta_2} > 1 - \frac{r}{\beta_1}$, making the probability of the event $\{\beta_2 \log(1+\gamma_1 P_2) < r\log\mathrm{SNR}\}$ decays to zero faster than that of $\{\beta_1\log(1+\gamma_1\mathrm{SNR}) < r\log\mathrm{SNR}\}$. Then, we have

$$P_2 \doteq \frac{\mathrm{SNR}}{\Pr(\mathcal{I} = 2)} = \mathrm{SNR}^{2-\frac{r}{\beta_1}}.$$

Let us next compute the SNR exponent D_2 of

$$\begin{aligned} \Pr(\mathcal{O}_2) &= \Pr(\mathcal{I} = 2, \beta_2\log(1+\gamma_1 P_2) \geq r\log\mathrm{SNR}, \\ &\qquad \beta_2\log(1+gP_2) + (1-\beta_2)\log(1+\gamma_2\mathrm{SNR}) < r\log\mathrm{SNR}) \\ &\doteq \Pr\left(\beta_1(1-\alpha_1)^+ < r, \beta_2\left(2-\frac{r}{\beta_1}-a\right)^+ + (1-\beta_2)(1-\alpha_2)^+ < r\right) \end{aligned}$$

Using Lemma 4.1, we obtain

$$D_2 = \begin{cases} 1 - \frac{r}{\beta_1} + \frac{1-r}{\beta_2} & \text{if } r \in (1-\beta_2, \beta_2), \\ 3 - \frac{r}{\beta_1} - \frac{r}{1-\beta_2} & \text{if } r < 1-\beta_2. \end{cases}$$

Finally

$$\begin{aligned} \Pr(\mathcal{O}_3) &= \Pr(\beta_2\log(1+\gamma_1 P_2) < r\log\mathrm{SNR}, \beta_2\log(1+gP_2) < r\log\mathrm{SNR}) \\ &\doteq \mathrm{SNR}^{-2\left(p_2-\frac{r}{\beta_2}\right)} \\ &\doteq \mathrm{SNR}^{-2\left(2-\frac{r}{\beta_1}-\frac{r}{\beta_2}\right)}. \end{aligned}$$

□

4.D Proof of Proposition 4.5

We present only the orthogonal case as the nonorthogonal case follows exactly the same line of arguments. The proof of lower bounds are similar to that in the short-term power constraint case and thus omitted.

For the lower bound, we first lower-bound the exponent $D_k \geq k - 1 + \frac{1-r}{\beta_k} - \sum_{i=1}^{k-1} \frac{r}{\beta_i}$ and change variable $x_k = (\beta_k)^{-1}$ (cf. Appendix 4.B). This leads to

$$D_{\text{O-PC}}^{\text{RF}-K}(r) \geq \sup \min \Big((1-r)x_1, 1 + (1-r)x_2 - rx_1, \ldots, \\ K - 1 + (1-r)x_K - r\sum_{k=1}^{K-1} x_k, 2\left(K - r\sum_{k=1}^{K} x_k\right)\Big)$$

where the supremum is over $\min(1/r, 2) \geq x_1 > x_2 > \cdots > x_K > 1$. We will balance the $K+1$ terms and verify if the solution satisfies the linear constraints. To that end, let

$$\begin{aligned}(1-r)x_1 &= 1 + (1-r)x_2 - rx_1 = \cdots \\ &= K - 1 + (1-r)x_K - r\sum_{k=1}^{K-1} x_k \\ &= 2\left(K - r\sum_{k=1}^{K} x_k\right) \\ &= \Delta.\end{aligned}$$

Balancing the first K terms leads to

$$\begin{aligned} x_1 &= 1 + (1-r)x_2, \\ x_2 &= 1 + (1-r)x_3, \\ &\cdots \\ x_K &= 1 + (1-r)x_{K-1} \end{aligned}$$

and thus

$$\begin{aligned} x_1 &= \frac{\Delta}{1-r}, \\ x_2 &= \frac{1}{1-r}\left(\frac{\Delta}{1-r} - 1\right), \\ x_3 &= \frac{1}{1-r}\left(\frac{\Delta}{(1-r)^2} - 1 - \frac{1}{1-r}\right), \\ &\cdots \\ x_K &= \frac{1}{1-r}\left(\frac{\Delta}{(1-r)^{K-1}} - \sum_{k=0}^{K-2} \frac{1}{(1-r)^k}\right). \end{aligned} \tag{4.43}$$

Inserting these values into $2\left(K - r\sum_{k=1}^{K} x_k\right) = \Delta$ and simplifying give

$$\Delta = \frac{(1-r)\left(1-(1-r)^K\right)}{r\left(1-\frac{1}{2}(1-r)^K\right)}.$$

With this Δ, we can readily verify that $x_1 > \cdots > x_K > 1$. However the constraint $\min(1/r, 2) \geq x_1$ does not always satisfy. In particular,

$$x_1 = \frac{\Delta}{1-r} = \frac{\left(1-(1-r)^K\right)}{r\left(1-\frac{1}{2}(1-r)^K\right)} \leq \min(1/r, 2) = \frac{1}{r}, \forall r \geq 1/2$$

but when $r < 1/2$,

$$x_1 = \frac{\Delta}{1-r} = \frac{\left(1-(1-r)^K\right)}{r\left(1-\frac{1}{2}(1-r)^K\right)} \leq \min(1/r, 2) = 2 \quad \text{iff } (1-r)^{K+1} \geq 1-2r.$$

That is, over $\{r : (1-r)^{K+1} < 1-2r\}$ the intersection is *not* the solution of the maximin. Since $D_{\text{LB-O-PC}}^{\text{RF}-K}(r) \leq 2-2r$, it suffices to shows that $D_{\text{LB-O-PC}}^{\text{RF}-K}(r) = 2-2r$ over this range of r. Let $\hat{K} \in \{2, \ldots, K\}$ so that $(1-r)^{\hat{K}} \geq 1-2r$ and $(1-r)^{\hat{K}+1} < 1-2r$. Such a $\hat{K}$ always exists because $(1-r)^2 \geq 1-2r, \forall r$ and $(1-r)^{K+1} < 1-2r$ by assumption.

Inserting $\Delta = 2-2r$ into (4.43) we obtain

$$\begin{aligned}
\hat{x}_1 &= 2, \\
\hat{x}_2 &= \frac{1}{1-r}, \\
&\ldots \\
\hat{x}_K &= \frac{1}{1-r}\left(\frac{2}{(1-r)^{K-2}} - \sum_{k=0}^{K-2}\frac{1}{(1-r)^k}\right).
\end{aligned}$$

Note that $2 \geq \hat{x}_1 > \hat{x}_2 > \cdots > \hat{x}_K$. Furthermore, with the given $\hat{K}$, we can easily check that

$$\hat{x}_{\hat{K}} = \frac{1}{1-r}\left(\frac{2}{(1-r)^{\hat{K}-2}} - \sum_{k=0}^{\hat{K}-2}\frac{1}{(1-r)^k}\right) \geq 1$$

and

$$\hat{x}_{\hat{K}+1} < 1.$$

We now show that $\min\left(D_1, \ldots, D_K, 2\left(K - r\sum_{k=1}^{K} x_k\right)\right) = 2-2r$ with the choice

$$x_1 = \hat{x}_1, x_2 = \hat{x}_2, \ldots, x_{\hat{K}} = \hat{x}_{\hat{K}}, x_{\hat{K}+1} = \cdots = x_K = 1.$$

Note that this choice of $\{x_k\}$ satisfies all the linear constraints.

By construction we have $D_1 = \cdots = D_{\hat{K}} = 2 - 2r$. Furthermore,

$$D_{\hat{K}+1} - D_{\hat{K}} = (1-r)x_{\hat{K}+1} + 1 - \hat{x}_{\hat{K}} > (1-r)\hat{x}_{\hat{K}+1} + 1 - \hat{x}_{\hat{K}} = 0$$

thus we have $D_{\hat{K}+1} > D_{\hat{K}} = 2 - 2r$. For $l > \hat{K} + 1$ we have

$$D_l - D_{l-1} = (1-r) + 1 - 1 = 1 - r > 0$$

thus $D_l > D_{l-1} > 2 - 2r$.

It remains to show that $2\left(K - r\sum_{k=1}^{K} x_k\right) \geq 2 - 2r \Leftrightarrow K - r\sum_{k=1}^{K} x_k \geq 1 - r$. But

$$K - r\sum_{k=1}^{K} x_k = \hat{K} - r\sum_{k=1}^{\hat{K}} \hat{x}_k + (K - \hat{K})(1-r) \geq \hat{K} - r\sum_{k=1}^{\hat{K}} \hat{x}_k.$$

With the given $\{\hat{x}_k\}$, after some straightforward manipulation we have

$$\hat{K} - r\sum_{k=1}^{\hat{K}} \hat{x}_k \geq 1 - r \Leftrightarrow (1-2r) > (1-r)^{\hat{K}+1}.$$

But this always holds given our choice of $\hat{K}$.

We conclude that over $\{r : (1-r)^{K+1} < 1 - 2r\}$,

$$\sup_{2 \geq x_1 > \cdots > x_K > 1} \min\left(D_1, \ldots, D_K, 2\left(K - r\sum_{k=1}^{K} x_k\right)\right) = 2 - 2r.$$

□

4.E Destination-Relay CSF

Proof of Proposition 4.6

For any K, the outage event is dominated by two terms

$$P_{\text{out}}(r \log \text{SNR}) \doteq \Pr(\mathcal{O}_1) + \Pr(\mathcal{O}_2)$$

where $\mathcal{O}_1$ is the event that the relay fails to decode the message and the destination cannot decode the direct transmission either.

$$\begin{aligned}\Pr(\mathcal{O}_1) &= \Pr\left(\beta \log(1 + \gamma_1 \text{SNR}) < r \log \text{SNR}, \beta \log(1 + g\text{SNR}) < r \log \text{SNR}\right) \\ &\doteq \text{SNR}^{-2\left(1 - \frac{r}{\beta}\right)}.\end{aligned}$$

The event $\mathcal{O}_2$ happens when the relay succeeds to decode but the combined direct and relayed signals cannot be decoded by the destination. Due to the construction of $\mathcal{I}(g,\gamma_2)$, this only happens when $\mathcal{I}=K$, i.e. when largest power level P_K is applied at the relay.

$$\Pr(\mathcal{O}_2) = \Pr\left(\mathcal{I}=K, \beta\log(1+\gamma_1\mathrm{SNR}) \geq r\log\mathrm{SNR},\right.$$
$$\left.\log(1+g\mathrm{SNR}) + (1-\beta)\log(1+\gamma_2 P_K) < r\log\mathrm{SNR}\right).$$

To apply Lemma 4.1 to find the SNR exponent of $\Pr(\mathcal{O}_2)$, we first recursively compute the exponent of P_K.

We have $P_1 \doteq \mathrm{SNR}$, and $\Pr(\mathcal{I}=2) \doteq \mathrm{SNR}^{-D_1}$ where D_1 is given by Lemma 4.1

$$D_1 = \begin{cases} 2-\frac{r}{1-\beta} & \text{if } \frac{1}{2} \leq \beta < 1-r \\ \frac{1-r}{\beta} & \text{if } \beta \geq \max(1/2, 1-r). \end{cases}$$

Then we have $P_2 \doteq \mathrm{SNR}^{1+D_1}$. Invoking Lemma 4.1 again gives $\Pr(\mathcal{I}=3) \doteq \mathrm{SNR}^{-D_2}$ where

$$D_2 = \begin{cases} 2+D_1-\frac{r}{1-\beta} & \text{if } \frac{1}{2} \leq \beta < 1-\frac{r}{1+D_1} \\ -D_1+\frac{1+D_1-r}{\beta} & \text{if } \beta \geq \max\left(\frac{1}{2}, 1-\frac{r}{1+D_1}\right). \end{cases}$$

Continuing this line of arguments leads to the claimed result. □

Alternative Proof of Proposition 4.7

We have

$$P_{\mathrm{out}}(r\log\mathrm{SNR}) \doteq \mathrm{SNR}^{-2(1-\frac{r}{\beta})} + \mathrm{SNR}^{-D_\infty}$$

where $D_\infty = \lim_{k\to\infty} D_K$ with D_K defined as in (4.27), i.e.

$$D_k = \begin{cases} 2+D_{k-1}-\frac{r}{1-\beta} & \text{if } \frac{1}{2} \leq \beta < 1-\frac{r}{1+D_{k-1}} \\ \frac{1-\beta}{\beta}D_{k-1}+\frac{1-r}{\beta} & \text{if } \beta \geq \max\left(\frac{1}{2}, 1-\frac{r}{1+D_{k-1}}\right). \end{cases}$$

Notice that for any given r,β the sequence (indexed by k) $1-\frac{r}{1+D_k}$ is increasing because the sequence D_k is increasing. We consider two cases.

Case 1: Assume $\beta > 1-\frac{r}{1+D_\infty}$, implying that $\beta > 1-\frac{r}{1+D_k}$, $\forall k$. Then by definition

$$\begin{aligned}
D_1 &= \frac{1-r}{\beta}, \\
D_2 &= \frac{1-r}{\beta} + \left(\frac{1}{\beta}-1\right)\frac{1-r}{\beta}, \\
D_3 &= \frac{1-r}{\beta} + \left(\frac{1}{\beta}-1\right)\left(\frac{1-r}{\beta} + \left(\frac{1}{\beta}-1\right)\frac{1-r}{\beta}\right), \\
&\dots \\
D_K &= \frac{1-r}{\beta}\left(1+\frac{1-\beta}{\beta}+\dots+\left(\frac{1-\beta}{\beta}\right)^{K-1}\right).
\end{aligned}$$

We can exclude the point $\beta = 1/2$ because of the following. With $\beta = 1/2$ we have $D_K = 2K(1-r) \to \infty$ as $K \to \infty$. But we require $\beta = \frac{1}{2} > 1 - \frac{r}{1+D_\infty} = 1 - 0 = 1$, which is a contradiction. Thus we only consider $\beta > 1/2$ meaning that $\frac{1-\beta}{\beta} < 1$. Then

$$D_K = \frac{(1-r)\left(1 - \left(\frac{1-\beta}{\beta}\right)^K\right)}{2\beta - 1}$$

and

$$D_\infty = \lim_{K\to\infty} D_K = \frac{1-r}{2\beta-1}.$$

To satisfy $\beta > 1 - \frac{r}{1+D_\infty}$ we must have

$$\beta > 1 - \frac{r}{1 + \frac{1-r}{2\beta-1}} \Leftrightarrow \beta > \frac{2-r}{2}.$$

The above condition always holds when $r > 2/3$, given the constraint $\beta \geq \max(1/2, r)$. From the indirect proof (cf. Section 4.5), we know that

$$\begin{aligned} \underline{D}_2 &= \sup \min_{\beta \in [\max(r, \frac{2-r}{2}), 1)} \left(2\left(1 - \frac{r}{\beta}\right), \frac{1-r}{2\beta-1}\right) \\ &= \begin{cases} 1 & \text{if } r < \frac{2}{5}, \\ 2 - \frac{16r}{3+3r+\sqrt{9r^2-14r+9}} & \text{otherwise.} \end{cases} \end{aligned}$$

Case 2: $\beta < 1 - \frac{r}{1+D_\infty} \Leftrightarrow \beta < \frac{2-r}{2}$, which only happens when $r < \frac{2}{3}$. There exists an integer $\bar{K} \in [0, \infty)$ so that

$$1 - \frac{r}{1+D_{\bar{K}}} > \beta \geq 1 - \frac{r}{1+D_{\bar{K}-1}}$$

with the convention $1 - \frac{r}{1+D_{-1}} = \max(r, 1/2)$. Then $D_{\bar{K}} > 0$ and by definition, we have

$$\begin{aligned} D_{\bar{K}+1} &= 2 + D_{\bar{K}} - \frac{r}{1-\beta}, \\ D_{\bar{K}+2} &= 2 + D_{\bar{K}+1} - \frac{r}{1-\beta} = D_{\bar{K}} + 2\left(2 - \frac{r}{1-\beta}\right), \\ &\cdots \\ D_{\bar{K}+L} &= D_{\bar{K}} + L\left(2 - \frac{r}{1-\beta}\right) \end{aligned}$$

Since $\beta < \frac{2-r}{2}$, we have $2 - \frac{r}{1-\beta} > 0$ thus

$$D_\infty = \lim_{L\to\infty} D_{\bar{K}+L} = \infty.$$

This leads to

$$\underline{D}_1 = \sup_{\beta\in[\max(\frac{1}{2},r),\frac{2-r}{2})} \min\left(2-\frac{2r}{\beta}, D_\infty\right) = \sup_{\beta\in[\max(\frac{1}{2},r)\frac{2-r}{2})} \left(2-\frac{2r}{\beta}\right) = 2-\frac{4r}{2-r}.$$

Combining $\underline{D}_1$ and $\underline{D}_2$ leads to the desired result. □

4.F Proof of Proposition 4.8

We present only the proof for an orthogonal scheme. Recall that we use the following index mapping

$$\mathcal{I}(g,\gamma_2) \equiv \mathcal{I}(g) = \begin{cases} K & \text{if } \log(1+gP_{K-1}) < r\log\mathrm{SNR}, \\ \min\{k\in\{1,\dots,K-1\} : \log(1+gP_k) \geq r\log\mathrm{SNR}\} & \text{otherwise} \end{cases}$$

and that β is the fraction of dimension assigned to the learning phase (only) when $\mathcal{I}=K$.

We now show that

$$D_{\text{O-SPC}}^{\text{DSF-}K}(r) = \sup_{\beta\in\left[\max\left(\frac{r}{K-(K-1)r},\frac{1}{2}\right),1\right)} \min\left(2K-2(K-1)r-\frac{2r}{\beta}, D_K\right) \tag{4.44}$$

where

$$D_K = \begin{cases} 2K-2(K-1)r-\frac{r}{1-\beta} & \text{if } \max\left(\frac{r}{K-(K-1)r},\frac{1}{2}\right) \leq \beta < 1-\frac{r}{K-(K-1)r} \\ \frac{K(1-r)}{\beta} & \text{if } \beta \geq \max\left(\frac{r}{K-(K-1)r},\frac{1}{2}, 1-\frac{r}{K-(K-1)r}\right). \end{cases}$$

To that end, let p_k be the SNR exponent of the power level used at the source (and also the relay when $\mathcal{I}=K$) given that $\mathcal{I}=k$. We readily see that $p_1=1$, $p_2=1+p_1-r=2-r,\dots,p_K=1+p_{K-1}-r=K-(K-1)r$. Due to construction, outage can only happen when $\mathcal{I}=K$ thus

$$P_{\text{out}}(r\log\mathrm{SNR}) \doteq \Pr(\mathcal{O}_1)+\Pr(\mathcal{O}_2)$$

where

$$\begin{aligned}\Pr(\mathcal{O}_1) &= \Pr\big(\mathcal{I}=K, \beta\log(1+\gamma_1P_K) \geq r\log\mathrm{SNR}, \\ &\qquad \beta\log(1+gP_K)+(1-\beta)\log(1+\gamma_2P_K) < r\log\mathrm{SNR}\big) \\ &\doteq \mathrm{SNR}^{-D_K}.\end{aligned}$$

If $\beta p_K < r$ then the event $\beta\log(1+\gamma_1P_K) < r\log\mathrm{SNR}$ happens with probability in the order of SNR^0, meaning that the relay is completely redundant in the proposed scheme. Therefore we constrain $\beta \geq \frac{r}{p_K} = \frac{r}{K-(K-1)r}$. Note that the event $\beta\log(1+$

$gP_K) + (1-\beta)\log(1+P_K\text{SNR}) < r\log\text{SNR}$ implies $\beta\log(1+gP_{K-1}) < r\log\text{SNR}$ or $\mathcal{I} = K$. Invoking Lemma 4.1 yields

$$D_K = \begin{cases} 2K - 2(K-1)r - \frac{r}{1-\beta} & \text{if } \max\left(\frac{r}{K-(K-1)r}, \frac{1}{2}\right) \le \beta < 1 - \frac{r}{K-(K-1)r} \\ \frac{K(1-r)}{\beta} & \text{if } \beta \ge \max\left(\frac{r}{K-(K-1)r}, \frac{1}{2}, 1 - \frac{r}{K-(K-1)r}\right). \end{cases}$$

Similarly

$$\mathcal{O}_2 = \{\mathcal{I} = K, \beta\log(1+\gamma_1 P_K) < r\log\text{SNR}, \beta\log(1+gP_K) < r\log\text{SNR}\},$$

which gives

$$\Pr(\mathcal{O}_2) \doteq \text{SNR}^{-2\left(p_K - \frac{r}{\beta}\right)} = \text{SNR}^{-2\left(K-(K-1)r-\frac{r}{\beta}\right)}.$$

Inserting into (4.44) and solving for the optimal β^* yield $\beta^* = 2/3$ for $r < \frac{K}{K+2}$ and $\beta^* = \frac{K-(K-2)r}{2K-2(K-1)r}$ for $r \ge \frac{K}{K+2}$. This leads to the claimed results. □

4.G Proof of Proposition 4.10

Since the proof follows closely that of Proposition 4.9, we briefly summarize the main steps herein.

Step 1 - Preliminaries. Let $\beta = T_1/T \ge 1$, $T_2 = (1-\beta)T \ge 2$. An encoder is a mapping $m \to \sqrt{\text{SNR}}(\mathbf{x}_1, \mathbf{X}_2)$ where X_2 is a $T_2 \times 2$ matrix with the first row $\mathbf{x}_{21}$ being the sequence sent from the relay and the second row $\mathbf{x}_{22}$ being the sequence sent from the source during the relaying phase. The power constraints are

$$\frac{1}{|\mathcal{C}|T}\sum_{\mathcal{C}}(\|\mathbf{x}_1\|^2 + \|\mathbf{x}_{22}\|^2) \le 1,$$

$$\frac{1}{|\mathcal{C}|T}\sum_{\mathcal{C}}\|\mathbf{x}_{21}\|^2 \le 1.$$

Step 2 - Universal conditions. Consider a codeword difference $(\Delta\mathbf{x}_1, \Delta\mathbf{X}_2)$. Let $\|\Delta\mathbf{x}_1\|^2 = \text{SNR}^{-\delta_1}$ and let $\lambda_2 = \text{SNR}^{-\delta_2}$ be the smallest squared singular value of $\Delta\mathbf{X}_2$. Applying again the Chernoff bound and using the worst-case rotation argument [TV06], we end up with the conditions

$$\left(\min_{\mathcal{C}} \text{SNR}^{-\beta(\delta_1)^+}\right) \dot{\ge} \text{SNR}^{-r}, \tag{4.45}$$

$$\left(\min_{\mathcal{C}} \text{SNR}^{-\beta(\delta_1)^+}\text{SNR}^{-(1-\beta)(\delta_2)^+}\right) \dot{\ge} \text{SNR}^{-r}. \tag{4.46}$$

Step 3 - Expurgation. Draw $(\mathbf{x}_1, \mathbf{X}_2)$ from an i.i.d. zero-mean unit-variance complex Gaussian ensemble. The probability that a codeword difference is too small leading to the expurgation of a codeword, $\Pr((\Delta\mathbf{x}_1, \Delta\mathbf{X}_2) \in \mathcal{B})$, has an SNR exponent of (recall that $\Delta\mathbf{X}_2$ in this case is a $T_2 \times 2$ matrix)

$$\begin{aligned} \inf_{\delta_1 \geq 0, \delta_2 \geq 0} \quad & T_1\delta_1 + (T_2 - 1)\delta_2 \\ \text{s.t.} \quad & T_1\delta_1 + T_2\delta_2 \geq Tr, \end{aligned}$$

which is optimized at $\delta_1^* = 0$, $\delta_2^* = \frac{T}{T_2}r$, leading to

$$\Pr((\Delta\mathbf{x}_1, \Delta\mathbf{X}_2) \in \mathcal{B}) \doteq \mathrm{SNR}^{-\frac{T_2-1}{T_2}Tr}.$$

To make the average number of expurgated codewords insignificant compared to the number of codewords before expurgation, we require the original codebook size to be $\mathrm{SNR}^{\hat{r}T}$, where

$$\hat{r} = \frac{T_2 - 1}{T_2}r - \epsilon = \left(1 - \frac{1}{(1-\beta)T}\right)r - \epsilon$$

for an arbitrarily small $\epsilon > 0$. The expurgated code itself is therefore not approximately universal because its multiplexing gain strictly is less than r (even if $\epsilon = 0$). However we can conclude that when this expurgated code of rate $\hat{r}\log\mathrm{SNR}$ is used, an SNR exponent of the error probability $\min(2-r-r/\beta, D_2)$ is achievable, where $\hat{r} < r$ (strict inequality). In other words, as long as the mutual information is greater than $r\log\mathrm{SNR} > \hat{r}\log\mathrm{SNR}$, then the expurgated code has $\Pr(\varepsilon) \doteq \mathrm{SNR}^{-\infty}$. This leads to the claimed result.

On a final note, the constraint $T_1 \leq \min\left(T-2, \left\lfloor T - \frac{1}{1-r}\right\rfloor\right)$ in the maximin of Proposition 4.10 is due to the fact that we constrain $T_2 \geq 2$ and that we require the multiplexing gain $\frac{r}{1-\frac{1}{(1-\beta)T}} < 1$. Clearly, as in Chapter 3, this expurgation procedure is only applicable for sufficiently small r. □

Chapter 5

D–M Tradeoff in Compress–and–Forward Relay Channels

We continue the D–M tradeoff analysis with another relaying protocol in this chapter. It is shown in [YE07] that the compress-and-forward strategy achieves the full cooperative D–M tradeoff bound of a three-node wireless relay network. This is obtained under the possibly unrealistic assumption that the relay has perfect knowledge of all three channel coefficients (source-destination, source-relay and relay-destination), and that the destination has perfect knowledge of the source-relay channel coefficient. Also, in the optimal CF strategy the relay makes use of Wyner-Ziv source coding with side information. This chapter investigates the achievable D-M tradeoff of the same network, under the same assumptions of perfect channel state information, when the relay is constrained to make use of standard (non-WZ) source coding. It is shown that under a short-term power constraint at the relay, using source coding without side information results in a significant loss in terms of the D–M tradeoff. For multiplexing gains $r \leq \frac{2}{3}$, this loss can be fully compensated for by relaxing the power constraint to a long-term one, and by using power control at the relay. On the contrary, for large multiplexing gain $r \in \left(\frac{2}{3}, 1\right)$ the loss with respect to WZ coding remains strict.

5.1 Introduction

In [YE07], it is shown that the multiple-antenna relay network achieves the cooperative upper bound (fully cooperative transmit antennas) by using a compress and forward strategy, where the relay sends to the destination a source-encoded (lossy) version of its received signal, which is then treated by the destination as an additional observation and used to decode the source information message. The optimal CF strategy makes use of Wyner-Ziv source coding with side information at the relay, and its optimality is shown under the *critical* assumption that the relay knows both the channel matrix from relay to destination and that from source to destination. It is also necessary that the destination knows the channel matrix from source to relay.

The practical implementation of WZ source coding with side information is generally much more complex than standard source coding. Since WZ coding requires a channel decoding operation at the decoder, approaching the rate-distortion limit in this case requires large block length and significant encoding and decoding complexity (see for example [WO01, MB02, CPR03, YCXZ03, CX05]).

At this point, it is legitimate to ask the following question: "How much of the optimality of the CF strategy in [YE07] comes from WZ coding, and how much comes from CSI at the relay?."

In this chapter, we quantify the D-M tradeoff loss incurred by not using WZ coding at the relay. We thus indirectly show that source coding with side information as in [HZ05, YE07] is *instrumental* in achieving a superior performance over quasi-static fading relay channels. However, the picture is not as bleak as it may appear. By relaxing the power constraint at the relay from peak (short-term) to average (long-term) and allowing power control, we show that the non-WZ CF strategy achieves the optimal cooperative D–M tradeoff for all multiplexing gain $r \leq \frac{2}{3}$ and generally outperforms the best known AF and DF strategies [AES05, LTW04, KCS07b].

Our results indicate that simple CF strategies based on scalar quantization and power control at the relay, coupled with clever protocols that distribute the required channel state information to the relay and to the destination, can indeed achieve very good practical throughput and reliability performance with low complexity.

5.2 System Model

Consider a three-node slowly fading wireless relay channel. All terminals have a single antenna, and are constrained to operate in half-duplex mode [LTW04]. The source-destination, source-relay, and relay-destination channel gains are denoted by h, h_1, and h_2 respectively. The channel is statistically symmetric, with h, h_1, and h_2 being mutually independent complex Gaussian random variables with zero mean and unit variance. The channel is constant during a fading block of T channel uses, and changes independently from one block to the next. Again, we are interested in the high-SNR performance of this channel when the allowed decoding delay is equal to T. Since the fading is constant over blocks of T symbols, the error probability performance is dominated by the outage probability. Furthermore, we consider the case of large block length T, for which the information theoretic limits of channel coding and rate-distortion source coding are achievable by standard random Gaussian coding. Since we look at both large SNR and large T, the limit ordering becomes important. Conceptually, we may think of an increasing sequence of operating SNRs, and for each SNR we look at the system performance when $T \to \infty$. All three channels are affected by additive white noises, mutually independent and identically distributed, with with complex circularly symmetric Gaussian components $\sim \mathcal{CN}(0,1)$. It is well-known that in the quasi-static large T regime each receiver can estimate with vanishing loss its own channel coefficient [BPS98]. Therefore, with no loss of generality, we assume perfect CSI at all receivers (relay and destination).

We study a CF scheme under the assumptions made in [YE07], that the relay knows h and h_2 perfectly, and that the destination has perfect knowledge of the source-relay gain h_1. First, we consider the case of a peak (or short-term) power constraint, where the relay codewords are constrained to satisfy

$$\frac{1}{T|\mathcal{C}|}\sum_{\mathbf{s}_{\rm R}\in\mathcal{C}}\|\mathbf{s}_{\rm R}\|^2 \leq {\rm S}NR$$

in each fading block. Then, we relax this to an average (or long-term) power constraint, given by

$$\frac{1}{T|\mathcal{C}|}\sum_{\mathbf{s}_{\rm R}\in\mathcal{C}}{\rm E}\left[\|\mathbf{s}_{\rm R}\|^2\right] \leq {\rm S}NR$$

where the expectation is with respect to the fading statistics. This relaxed constraint allows for the use of power control at the relay, where in certain fading blocks a power much larger than average can be used, provided that the average constraint is satisfied.

The information rate of the source is $r\log{\rm SNR}$ bit per channel use, which is fixed for a given SNR. Recall that an *outage* event occurs when the mutual information between source and destination is smaller than $r\log{\rm SNR}$. The corresponding probability of outage is denoted as $P_{\rm out}(r\log{\rm SNR})$. We recall that the system achieves an outage exponent of d when

$$P_{\rm out}(r\log{\rm SNR}) \doteq {\rm SNR}^{-d}.$$

Due to the half-duplex constraint, the transmission is divided into two phase. Phase 1 uses a fraction $\beta\in(0,1)$ of the available degrees of freedom (i.e., it uses βT channel uses), in which the source encodes an equally likely message $m \in \{1,\dots,2^{rT\log{\rm SNR}}\}$ to a codeword $\mathbf{s}^m$ and then transmits the first βT symbols of $\mathbf{s}^m$. The received signals at the destination and at the relay are

$$\begin{aligned} y_1 &= hs_1 + w_1 \\ y_{\rm R} &= h_1 s_1 + w_{\rm R} \end{aligned} \tag{5.1}$$

respectively, where we omit the time index for brevity.

In Phase 2, the source transmits the remaining $(1-\beta)T$ symbols $\mathbf{s}_2$. The relay vector-quantizes the signal vector received during Phase 1, denoted as $\mathbf{y}_{\rm R}$, and encodes the resulting quantization index into a codeword $\mathbf{s}_{\rm R}$ of length $(1-\beta)T$, transmitted over the relay-destination channel. Notice that the source–channel coding used by the relay is a purely digital "tandem" encoding [MP02] approach. The channel code rate is (smaller than but arbitrarily close to)

$$R_0 = \log\left(1 + \frac{|h_2|^2\text{SNR}}{1+|h|^2\text{SNR}}\right),$$

The channel bandwidth to source bandwidth ratio for this tandem source-channel coding scheme is $\frac{1-\beta}{\beta}$.

The destination in Phase 2 observes the signal

$$y_2 = hs_2 + h_2 s_\text{R} + w_2$$

The decoder at the destination first decodes the quantization index sent by the relay by decoding $\mathbf{s}_\text{R}$ from $\mathbf{y}_2$, treating the source codeword as additive Gaussian noise. Then, assuming that the relay codeword is correctly decoded, it decodes the source information message m based on the joint observation $(\mathbf{y}_1, \mathbf{y}'_2, \hat{\mathbf{y}}_\text{R})$ where $\mathbf{y}'_2 = \mathbf{y}_2 - h_2\mathbf{s}_\text{R}$ and $\hat{\mathbf{y}}_\text{R}$ is the reconstructed (quantized) version of the relay received signal during Phase 1.

Since this relaying scheme does not use Wyner–Ziv source coding with side information [WZ76, Wyn78] as in [HZ05, YE07] we refer to it as the non-WZ CF scheme.

The vector quantization codebook at the relay is randomly generated, according to the probability distribution of

$$\hat{y}_\text{R} = y_\text{R} - w_\text{Q} = y_\text{R} - (cy_\text{R} + u)$$

where $c = 2^{-\frac{1-\beta}{\beta}R_0}$ and u is a zero-mean complex Gaussian, independent of y_R, with variance $\sigma_\text{u}^2 = (1+|h_1|^2\text{SNR})\,(1-c)\,c$. With this choice, the test (backward) channel [CT91] is realized and the rate-distortion bound can be achieved. The quantization error

$$w_\text{Q} = y_\text{R} - \hat{y}_\text{R}$$

is then i.i.d. complex Gaussian with zero mean and variance

$$\begin{aligned}
\sigma_\text{Q}^2 &= c^2\,\text{E}[|y_\text{R}|^2] + \sigma_\text{u}^2 \\
&= 2^{-\frac{1-\beta}{\beta}R_0}\,\text{E}[|y_\text{R}|^2] \\
&= (1+|h_1|^2\text{SNR})\left(1 + \frac{|h_2|^2\text{SNR}}{1+|h|^2\text{SNR}}\right)^{-\frac{1-\beta}{\beta}}.
\end{aligned}$$

Note that w_Q is uncorrelated with $\hat{y}_\text{R}$.

Rewriting

$$\hat{y}_\text{R} = (1-c)y_\text{R} - u = (1-c)h_1 s_1 + (1-c)w_1 - u,$$

we then determine the achievable mutual information of the strategy as

$$
\begin{aligned}
& I(s; y_1, y_2, \hat{y}_{\mathrm{R}}) \\
&= \beta \log\left(1 + |h|^2\mathrm{SNR} + \frac{|h_1|^2\mathrm{SNR}}{1 + \frac{\sigma_{\mathrm{u}}^2}{(1-c)^2}}\right) + (1-\beta)\log(1 + |h|^2\mathrm{SNR}) \\
&= \beta \log\left(1 + |h|^2\mathrm{SNR} + \frac{|h_1|^2\mathrm{SNR}}{1 + \frac{(1+|h_1|^2\mathrm{SNR})c}{1-c}}\right) + (1-\beta)\log(1 + |h|^2\mathrm{SNR}) \\
&= \beta \log\left(1 + |h|^2\mathrm{SNR} + \frac{|h_1|^2\mathrm{SNR}(1-c)}{1 + c|h_1|^2\mathrm{SNR}}\right) + (1-\beta)\log(1 + |h|^2\mathrm{SNR}).
\end{aligned} \tag{5.2}
$$

Recall that

$$
c = 2^{-\frac{1-\beta}{\beta}R_0} = \left(1 + \frac{|h_2|^2\mathrm{SNR}}{1 + |h|^2\mathrm{SNR}}\right)^{-\frac{1-\beta}{\beta}}. \tag{5.3}
$$

Note that in order to achieve this rate, the destination needs to know full CSI, i.e., h, h_1, and h_2. Computing the outage exponent of the scheme then gives the following result, the detailed proof of which is presented in Appendix 5.3.

Proposition 5.1. *Assume the relay knows h_2 and h. Then under a short-term power control at the relay, the non-WZ CF scheme achieves the outage exponent*

$$
D_{NPC}^{CF-\infty}(r) = \begin{cases} 2 - \frac{3+\sqrt{5}}{2}r & \text{if } r < \frac{3-\sqrt{5}}{2}, \\ (1-r)(2-r) & \text{otherwise.} \end{cases}
$$

It is remarkable that the tradeoff of the CF system coincides *exactly* with the outage exponent of a nonorthogonal *decode-and-forward* relaying scheme with no CSIT. Compared to the D–M tradeoff of the CF scheme using Wyner-Ziv source coding, which is shown to be $2 - 2r$ in [YE07], this simple CF with no side information exhibits a large degradation at all multiplexing gains. It is now clear that WZ coding is the key ingredient in achieving the superior performance of the CF schemes in [HZ05, YE07].

Where does this superior gain come from? Recall from [HZ05] that the achievable mutual information of CF using WZ coding is

$$
\beta\left(1 + |h|^2\mathrm{SNR} + \frac{|h_1|^2\mathrm{SNR}}{1 + \sigma_{\mathrm{WZ}}^2}\right) + (1-\beta)\log(1 + |h|^2\mathrm{SNR}) \tag{5.4}
$$

where the compression noise

$$
\begin{aligned}
\sigma_{\mathrm{WZ}}^2 &= \frac{(1 + |h_1|^2\mathrm{SNR} + |h|^2\mathrm{SNR})c}{(1 + |h|^2\mathrm{SNR})(1-c)} \\
&= \frac{\left(1 + \frac{|h_1|^2\mathrm{SNR}}{1+|h|^2\mathrm{SNR}}\right)c}{1-c}
\end{aligned}
$$

with c given by (5.3). Comparing the two achievable rates (5.2) and (5.4), we see that the only subtle difference lies in their compression noise variances (given the same value of β). Recall that the compression noise without WZ coding in (5.2) is

$$\sigma^2_{\mathrm{NWZ}} = \frac{(1+|h_1|^2\mathrm{SNR})c}{1-c}.$$

In terms of the SNR exponent, most of the time, i.e. when $|h_1|^2 \doteq \mathrm{SNR}^0$ and $|h|^2 \doteq \mathrm{SNR}^0$, the ratio between the quantization noise variances is $\frac{\sigma^2_{\mathrm{NWZ}}}{\sigma^2_{\mathrm{WZ}}} \doteq \mathrm{SNR}^1$. This relatively large difference in turn influences the optimization of β, leading to the performance loss of the non-WZ CF scheme.

We now relax the short-term power constraint at the relay, allowing the relay to control its transmit power over time based on (perfect) CSIT. In this case, the channel coding rate from relay to destination becomes

$$R_0^{\mathrm{PC}} = \log\left(1 + \frac{|h_2|^2\mathrm{SNR}^{\pi^*(h,h_2)}}{1+|h|^2\mathrm{SNR}}\right)$$

where $\mathrm{SNR}^{\pi^*(h,h_2)}$ is the optimal power allocated to the tuple of channel states (h, h_2) fed back from the destination to the relay. Thus the achievable rate is similar to the no power control case, with a new parameter $c^{\mathrm{PC}} = 2^{-\frac{1-\beta}{\beta}R_0^{\mathrm{PC}}}$. That is, in this case the noise variance due to quantization decays to zero faster as SNR grows. We characterize the achievable diversity-multiplexing tradeoff of this scheme by the following result, proved in Appendix 5.4.

Proposition 5.2. *Assume the relay knows h_2 and h. Then under a long-term power control at the relay, the non-WZ CF scheme achieves the outage exponent*

$$D_{PC}^{CF-\infty}(r) = \begin{cases} 2-2r & \text{if } r < \frac{2}{3}, \\ \frac{(1-r)(2-r)}{r} & \text{otherwise.} \end{cases}$$

With power control, the simple non-WZ CF approach is optimal for all multiplexing gains $r \leq \frac{2}{3}$. What is the intuition behind this phenomenon? The analysis in Appendix 5.4 reveals that for β larger than an r-dependent threshold, the quantization noise becomes too large, rendering useless the quantized signal sent from the relay. In such cases, no cooperative gain can be achieved (cf. (5.2)). For $r > \frac{2}{3}$, this threshold on β is strictly less than $\frac{1}{2}$, implying that the relay can contributes with its own (quantized) received signal observation for less than 1/2 of the available degrees of freedom (as evident from (5.2)), and fully cooperative diversity gain cannot be achieved. The proof in Appendix 5.4 also shows that the probability of outage is dominated by the event that the source–destination link alone is in a deep fade. Thus in the high-multiplexing gain regime the bottleneck of non-WZ CF relaying with power control is the source–destination link (unfortunately, in

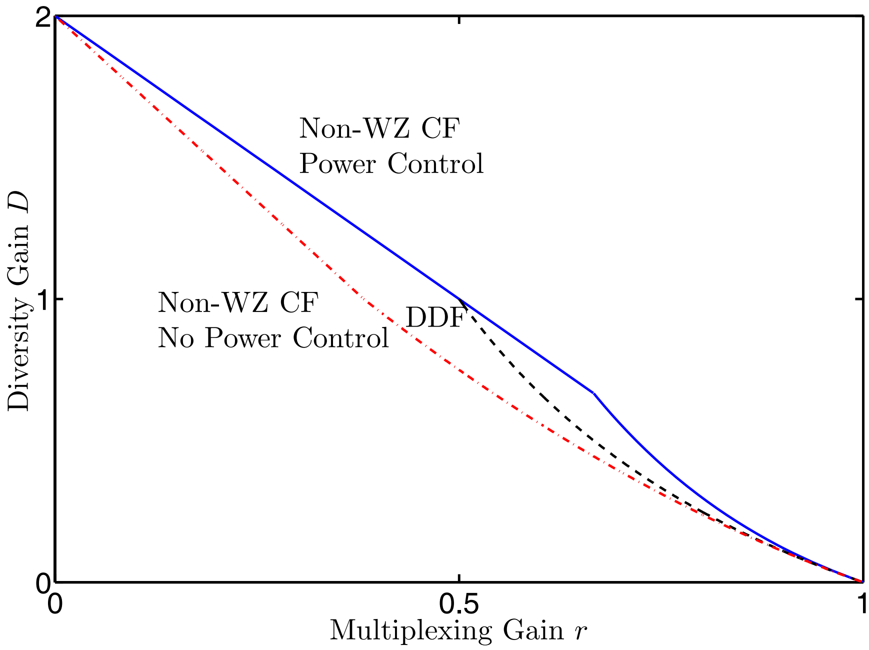

Figure 5.1: Outage exponents of simple CF schemes with perfect CSIT available to the relay under different power constraints. The dashed curve is the outage exponent of the dynamic decode-and-forward [AES05] scheme.

practice, it is likely that the quality of the direct link is indeed worse than that of the links connecting to the relay.)

An interesting implication of the analysis in the appendices is that no loss is incurred by restricting to $\beta \leq \frac{1}{2}$. Because the channel bandwidth to source bandwidth for the source–channel tandem encoder at the relay is $\frac{1-\beta}{\beta}$, this implies that it is not necessary to consider *bandwidth compression* [MP02] at the relay.[1] This can be seen as a converse property to that of DF with quantized CSIT, where it is shown in Chapter 4 that considering $\beta \geq \frac{1}{2}$ is sufficient.

We plot the tradeoff curves of both CF schemes in Fig. 5.1. For comparison, we also plot the performance of the DDF scheme [AES05], which is uniformly better than non-WZ CF without power control but is inferior to non-WZ CF with power control for all $r > \frac{1}{2}$. Note however that the assumptions on CSIT of DDF and the CF analyzed here as well as the WZ CF scheme in [YE07] are fundamentally

[1]Therefore the traditional term "compress-and-forward" used herein is not entirely correct, because an optimal system either does bandwidth expansion or keeps the channel bandwidth equal to the source bandwidth.

different.

As in [YE07], this work assumes perfect CSIT at the relay, which is a too optimistic assumption in most practical scenarios. Evaluating the performance of more realistic CF relaying schemes with *limited* feedback remains an interesting topic for further study.

Appendices for Chapter 5

5.3 Proof of Proposition 5.1

We first state the following convenient lemma without proof.

Lemma 5.1. *For $\beta \in (0,1)$, the linear programming problem*

$$D^* = \inf_{\alpha,\alpha_1 \geq 0} \{\alpha + \alpha_1\} \quad s.t. \quad \beta(1-\alpha, 1-\alpha_1)^+ + (1-\beta)(1-\alpha)^+ < r$$

where $(x_1, \ldots, x_n)^+ \triangleq \max(x_1, \ldots, x_n, 0)$, has solution

$$D^* = \begin{cases} 2 - 2r & \text{if } \beta \geq \frac{1}{2}, \\ \frac{1-r}{1-\beta} & \text{if } \beta < \min\left(\frac{1}{2}, r\right), \\ 2 - \frac{r}{\beta} & \text{if } r \leq \beta < \frac{1}{2}. \end{cases}$$

with the corresponding optimizers

$$(\alpha^*, \alpha_1^*) = \begin{cases} (1-r, 1-r) & \text{if } \beta \geq \frac{1}{2}, \\ (\frac{1-r}{1-\beta}, 0) & \text{if } \beta < \min\left(\frac{1}{2}, r\right), \\ (1, 1-\frac{r}{\beta}) & \text{if } r \leq \beta < \frac{1}{2}. \end{cases}$$

We now prove Proposition 5.1. Perform the standard change of variable $a = -\log|h|^2/\log\text{SNR}$, $\alpha_1 = -\log|h_1|^2/\log\text{SNR}$, $\alpha_2 = -\log|h_2|^2/\log\text{SNR}$ (cf. [ZT03, AES05]). Since

$$P_{\text{out}} = \Pr\Big(\beta\log\Big(1+|h|^2\text{SNR} + \frac{|h_1|^2\text{SNR}(1-c)}{1+c|h_1|^2\text{SNR}}\Big) + (1-\beta)\log(1+|h|^2\text{SNR}) < r\log\text{SNR}\Big),$$

to compute the outage exponent as $\text{SNR} \to \infty$ we can focus on the set [ZT03]

$$\mathcal{O} = \Big\{a, \alpha_1, \alpha_2 \in \mathbb{R}_+^3 : \beta\left(1-a, 1-\alpha_1 - \left(1-\alpha_1 - \frac{1-\beta}{\beta}\left(1-\alpha_2-(1-a)^+\right)^+\right)^+\right)^+ + (1-\beta)(1-a)^+ < r\Big\}.$$

Herein

$$c = \left(1 + \frac{|h_2|^2\text{SNR}}{1+|h|^2\text{SNR}}\right)^{-\frac{1-\beta}{\beta}} \doteq \text{SNR}^{-\frac{1-\beta}{\beta}\left(1-\alpha_2-(1-a)^+\right)^+}.$$

The outage exponent is then given by

$$D_{\text{NPC}}^{\text{CF-}\infty}(r) = \min_{a,\alpha_1,\alpha_2\in\mathcal{O}} (a+\alpha_1+\alpha_2). \tag{5.5}$$

We partition $\mathcal{O}$ into two disjoint regions and solve (5.5) over each region, and then take the minimum of the two solutions.

Case 1: $a \geq 1$. The equation defining $\mathcal{O}$ reduces to

$$\beta\left(1-\alpha_1-\left(1-\alpha_1-\frac{1-\beta}{\beta}(1-\alpha_2)^+\right)^+\right)^+ < r.$$

Let D_1 be the minimum of the objective function in (5.5) over this region. We further divide **Case 1** into two cases.

Case 1.1: $1-\alpha_1 < \frac{1-\beta}{\beta}(1-\alpha_2)^+$. The optimizers of (5.5) over this subset are readily found to be $(a^*,\alpha_1^*,\alpha_2^*) = \left(1, \left(1-\frac{r}{\beta}, 1-\frac{1-\beta}{\beta}\right)^+, 0\right)$, where we can extend

$$\alpha_1^* = \begin{cases} 0 & \text{if } 0<\beta<\min(r,1/2), \\ 1-\frac{1-\beta}{\beta} & \text{if } \max(1/2,1-r)<\beta<1, \\ 1-\frac{r}{\beta} & \text{otherwise.} \end{cases} .$$

Case 1.2: $1-\alpha_1 \geq \frac{1-\beta}{\beta}(1-\alpha_2)^+$. Similarly we have

$$(a^*,\alpha_1^*,\alpha_2^*) = \left(1, 0, \left(1-\frac{r}{\beta}, 1-\frac{1-\beta}{\beta}\right)^+\right)$$

and

$$\alpha_2^* = \begin{cases} 0 & \text{if } \max(1-r,1/2)<\beta<1, \\ 1-\frac{\beta}{1-\beta} & \text{if } 0<\beta<\min(r,1/2), \\ 1-\frac{r}{1-\beta} & \text{otherwise.} \end{cases} .$$

Combining **Case 1.1** and **Case 1.2** leads to

$$D_1 = \begin{cases} 1+\min\left(1-\frac{r}{\beta}, 1-\frac{r}{1-\beta}\right) & \text{if } r<\beta<1-r, \\ 1 & \text{otherwise.} \end{cases}$$

Notice that $D_1 = 1$, $\forall r \geq \frac{1}{2}$.

Case 2: $a < 1$. We now consider

$$\beta\max\left(1-a, 1-\alpha_1-\left(1-\alpha_1-\frac{1-\beta}{\beta}(a-\alpha_2)^+\right)^+\right) + (1-\beta)(1-a) < r. \tag{5.6}$$

This case is divided into two sub-cases.

Case 2.1: $1-\alpha_1 < \frac{1-\beta}{\beta}(a-\alpha_2)^+$. Then (5.6) becomes

$$\beta\max(1-a, 1-\alpha_1) + (1-\beta)(1-a) < r \tag{5.7}$$

which does not depend on α_2. We easily see that the optimizer $\alpha_2^* = 0$ and the condition defining **Case 2.1** reduces to $1-\alpha_1 < \frac{1-\beta}{\beta}a$. The solution to (5.5) under *only* the constraint (5.7) is then given by Lemma 5.1 (note that the condition of **Case 2**, $a < 1$, is always satisfied by the solution in Lemma 5.1). It turns out that the optimizers in Lemma 5.1 satisfy $1-\alpha_1^* < \frac{1-\beta}{\beta}a^*$ iff $r+\beta \le 1$.

In case $r+\beta > 1$, a simple modified version of Lemma 5.1 leads to

$$(a^*, \alpha_1^*) = \begin{cases} \left(1-r, 1-\frac{1-\beta}{\beta}(1-r)\right) & \text{if } \beta \ge 1/2, \\ \left(\frac{\beta}{1-\beta}, 0\right) & \text{if } \beta < \min(r, 1/2). \end{cases}$$

In summary, in **Case 2.1**, the minimum of the objective function in (5.5) is

$$D_2^{(1)} = \begin{cases} 2-2r & \text{if } \frac{1}{2} \le \beta \le 1-r, \\ \frac{1-r}{1-\beta} & \text{if } \beta < \min(1/2, r, 1-r), \\ 2-\frac{r}{\beta} & \text{if } r < \beta < 1/2, \\ 1+\frac{2\beta-1}{\beta}(1-r) & \text{if } \beta \ge \max(1/2, 1-r), \\ \frac{\beta}{1-\beta} & 1-r < \beta < \min(r, 1/2). \end{cases}$$

Case 2.2: $1-\alpha_1 \ge \frac{1-\beta}{\beta}(a-\alpha_2)^+$. In this case, (5.6) becomes

$$\beta\max\left(1-a, \frac{1-\beta}{\beta}(a-\alpha_2)^+\right) + (1-\beta)(1-a) < r. \tag{5.8}$$

Again we note that $\alpha_1^* = 0$. Now consider the sub-case $\frac{1-\beta}{\beta}(a-\alpha_2) > 1-a \Leftrightarrow \frac{1}{\beta}a - \frac{1-\beta}{\beta}\alpha_2 > 1$, when (5.8) reduces to $(1-\beta)(1-\alpha_2) < r$. This leads to the optimizers of (5.5) $(a^*, \alpha_2^*) = \left(1-r, \left(1-\frac{r}{1-\beta}\right)^+\right)$. In the sub-case $\frac{1}{\beta}a - \frac{1-\beta}{\beta}\alpha_2 \le 1$, (5.8) becomes $(1-a) < r$ and we also obtain $(a^*, \alpha_2^*) = \left(1-r, \left(1-\frac{r}{1-\beta}\right)^+\right)$. Thus we have

$$D_2^{(2)} = 1-r+\left(1-\frac{r}{1-\beta}\right)^+.$$

Then $D_2 = \min\left(D_2^{(1)}, D_2^{(2)}\right)$. After some tedious but straightforward manipulation, we have

$$D_2 = \begin{cases} \frac{1-r}{1-\beta} & \text{if } \beta < \min\left(r, \frac{1-r}{2-r}\right), \\ 2-\frac{r}{\beta} & \text{if } r < \beta < \frac{3-\sqrt{5}}{2}, \\ 1-r & \text{if } \beta \ge 1-r, \\ 2-r-\frac{r}{1-\beta} & \text{otherwise.} \end{cases}$$

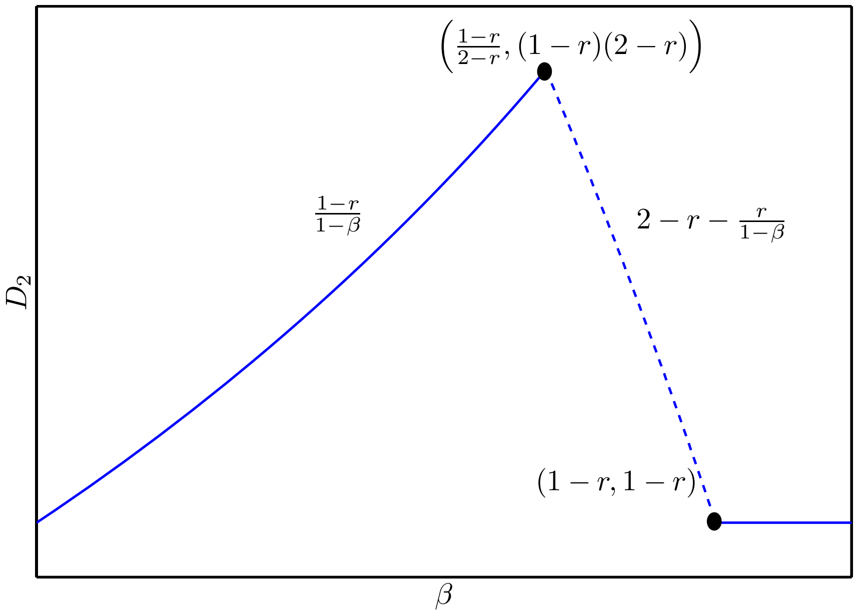

Figure 5.2: The computation of the optimal β in the no-power-control case, $r \geq \frac{3-\sqrt{5}}{2}$.

Notice from the result that to realize any cooperative gains, we need to use $\beta < 1-r$.

Finally combining **Case 1** and **Case 2** we have the solution to (5.5) $D = \min(D_1, D_2)$. Recall that D depends on the dimension fraction β, which can be optimized over, i.e.

$$D_{\text{NPC}}^{\text{CF}-\infty}(r) = \sup_{\beta \in (0,1)} \min(D_1, D_2).$$

If $r \geq 1/2$, we have

$$\begin{aligned} D_{\text{NPC}}^{\text{CF}-\infty}(r) &= \sup_{\beta \in (0,1-r)} \min(1, D_2) \\ &= \sup_{\beta \in (0,1-r)} D_2 \\ &= \sup_{\beta \in (0,1-r)} \min\left(\frac{1-r}{1-\beta}, 2-r-\frac{r}{1-\beta}\right). \end{aligned}$$

Solving this gives the optimizer $\beta^* = \frac{1-r}{2-r} < \frac{1}{2}$ (illustrated in Fig. 5.2), resulting in $D_{\text{NPC}}^{\text{CF}-\infty}(r) = (1-r)(2-r)$. With this optimal dimesion allocation the minimum

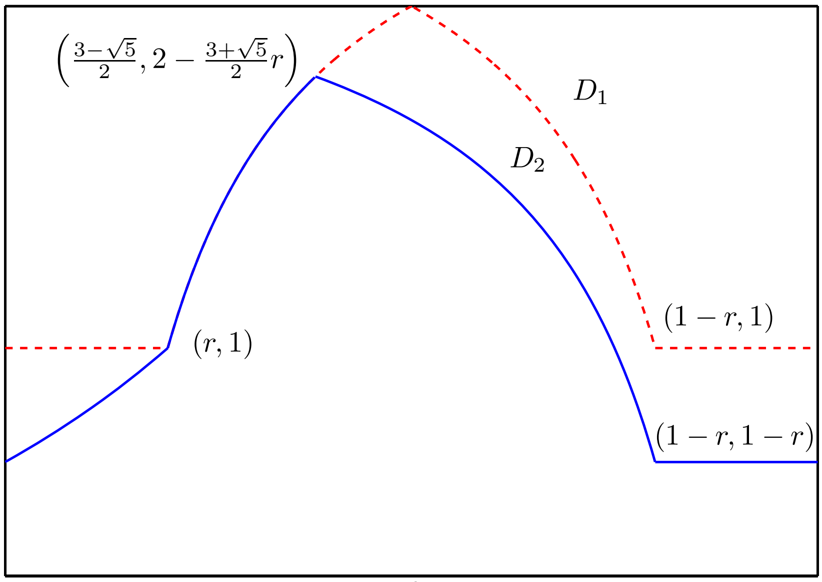

Figure 5.3: The computation of the optimal β in the no-power-control case, $r < \frac{3-\sqrt{5}}{2}$.

of (5.5) is attained at either $(a^*, \alpha_1^*, \alpha_2^*) = \left(\frac{1-r}{1-\beta^*}, 0, 0\right) = ((1-r)(2-r), 0, 0)$ or $(a^*, \alpha_1^*, \alpha_2^*) = \left(1-r, 0, 1-\frac{r}{1-\beta^*}\right) = (1-r, 0, 1-r(2-r))$.

If $r \in \left[\frac{3-\sqrt{5}}{2}, \frac{1}{2}\right)$, we have $D_2 \leq (1-r)(2-r) < 1 \leq D_1$ thus $\sup_\beta \min(D_1, D_2) = \sup_\beta D_2$, which also results in $D_{\text{NPC}}^{\text{CF}-\infty}(r) = (1-r)(2-r)$.

If $r < \frac{3-\sqrt{5}}{2}$, it again turns out that $\sup_\beta \min(D_1, D_2) = \sup_\beta D_2$, as illustrated in Fig. 5.3. This results in $\beta^* = \frac{3-\sqrt{5}}{2}$ and $D_{\text{NPC}}^{\text{CF}-\infty}(r) = 2 - \frac{3+\sqrt{5}}{2}r$. The minimum in (5.5) is attained at $(a^*, \alpha_1^*, \alpha_2^*) = \left(1, 1-\frac{r}{\beta^*}, 0\right)$ or $(a^*, \alpha_1^*, \alpha_2^*) = \left(1-r, 0, 1-\frac{r}{1-\beta^*}\right)$. □

5.4 Proof of Proposition 5.2

The proof is quite similar to the no power control case; and the key steps are summarized as follows. The power constraint at the relay asymptotically reads

$$\sup \{\pi(a, \alpha_2) - a - \alpha_2\} \leq 1$$

which leads to the optimal SNR exponent of the relay transmit power

$$\pi^*(a,\alpha_2) = 1 + a + \alpha_2.$$

Recall

$$c^{\mathrm{PC}} = \left(1 + \frac{|h_2|^2 \mathrm{SNR}^{\pi^*(a,\alpha_2)}}{1+|h|^2\mathrm{SNR}}\right)^{-\frac{1-\beta}{\beta}} \doteq \mathrm{SNR}^{-\frac{1-\beta}{\beta}(\pi^*(a,\alpha_2)-\alpha_2-(1-a)^+)^+}.$$

We now have $D_{\mathrm{PC}}^{\mathrm{CF}-\infty}(r) = \inf_{a,\alpha_1,\alpha_2\in\mathcal{O}}\{a+\alpha_1+\alpha_2\}$ where

$$\begin{aligned}\mathcal{O} = \Big\{&a,\alpha_1,\alpha_2 \in \mathbb{R}_+^3 : \\ &\beta\left(1-a, 1-\alpha_1 - \left(1-\alpha_1 - \frac{1-\beta}{\beta}\left(1+a-(1-a)^+\right)^+\right)^+\right)^+ \\ &+(1-\beta)\left(1-a\right)^+ < r\Big\}.\end{aligned}$$

Case 1: $a \geq 1$. Then we need to consider the subset defined by

$$\beta\left(1-\alpha_1-\left(1-\alpha_1-\frac{1-\beta}{\beta}(1+a)\right)^+\right)^+ < r. \tag{5.9}$$

Case 1.1: $1-\alpha_1-\frac{1-\beta}{\beta}(1+a) < 0$ or $2\beta-1 < (1-\beta)a+\beta\alpha_1$. Then $\alpha_1^* = \left(1-\frac{r}{\beta}\right)^+$, $\alpha_2^*=0$, $a^* = \max\left(1, \frac{2\beta-1-(1-\frac{r}{\beta})^+}{1-\beta}\right)$, and

$$\begin{aligned}D_1^{(1)} &= \max\left(1, \frac{2\beta-1-(1-\frac{r}{\beta})^+}{1-\beta}\right) + \left(1-\frac{r}{\beta}\right)^+ \\ &= \begin{cases} 2-\frac{r}{\beta} & \text{if } r < \min\left(\beta, \frac{3\beta-4\beta^2}{1-\beta}\right), \\ 1 & \text{if } \beta < \min\left(r, \frac{2-r}{2}\right), \\ \frac{2\beta-1}{1-\beta} & \text{if } \frac{(3\beta-4\beta^2)^+}{1-\beta} < r < \beta, \\ \frac{r+\beta-1}{1-\beta} & \text{if } \frac{2-r}{2} < \beta < r. \end{cases}\end{aligned}$$

Case 1.2: $2\beta-1 \geq (1-\beta)a+\beta\alpha_1$ then from (5.9), $1+a<\frac{r}{1-\beta}$. Since $a\geq 1$, this happens only if $\beta \geq \max\left(\frac{2-r}{2},\frac{2}{3}\right)$, resulting in $D_1^{(2)}=1$. If $\beta < \max\left(\frac{2-r}{2},\frac{2}{3}\right)$ then the problem is infeasible, and we write $D_1^{(2)}=\infty$.

Thus

$$D_1 = \min\left(D_1^{(1)}, D_1^{(2)}\right) = \begin{cases} 2-\frac{r}{\beta} & \text{if } r\leq\beta\leq\frac{2}{3} \text{ or } \beta>2/3, r<\frac{3\beta-4\beta^2}{1-\beta}, \\ \frac{2\beta-1}{\beta} & \text{if } \beta>\frac{2}{3}, \frac{(3\beta-4\beta^2)^+}{1-\beta} < r < 2-2\beta, \\ \frac{r+\beta-1}{1-\beta} & \text{if } \frac{2-r}{2}<\beta<\frac{2}{3}, \\ 1 & \text{otherwise.}\end{cases}$$

Case 2: $0 \le a < 1$. We need to consider the constraint

$$\beta\left(1-a, 1-\alpha_1-\left(1-\alpha_1-\frac{1-\beta}{\beta}2a\right)^+\right)^+ + (1-\beta)(1-a) < r. \qquad (5.10)$$

Case 2.1: $1-\alpha_1-\frac{1-\beta}{\beta}2a \ge 0$ or $\beta \ge 2(1-\beta)a+\beta\alpha_1$ then (5.10) becomes $\max(\beta(1-a),(1-\beta)2a)+(1-\beta)(1-a) < r$. In this case

$$D_2^{(1)} = \begin{cases} 1-r & \text{if } \beta \ge \frac{2-2r}{2-r}, \\ \infty & \text{otherwise.} \end{cases}$$

We constrain $\beta < \frac{2-2r}{2-r}$ from now on, in order to realize any cooperative diversity gains.

Case 2.2: $\beta < 2(1-\beta)a+\beta\alpha_1$. Then (5.10) reduces to $\beta\max(1-a,1-\alpha_1)+(1-\beta)(1-a) < r$. Applying Lemma 5.1 yields (recall that we constrain $\beta < \frac{2-2r}{2-r}$)

$$D_2^{(2)} = \begin{cases} 2-2r & \text{if } 1/2 \le \beta < \frac{2-2r}{2-r}, \\ \frac{1-r}{1-\beta} & \text{if } \beta < \min\left(r, \frac{1}{2}, \frac{2-2r}{2-r}\right), \\ 2-\frac{r}{\beta} & r < \beta \le \min\left(\frac{1}{2}, \frac{2-2r}{2-r}\right). \end{cases}.$$

Within the range of $\beta < \frac{2-2r}{2-r}$, $D_2 = \min\left(D_2^{(1)}, D_2^{(2)}\right) = D_2^{(2)}$.

Finally we optimize the dominant exponent of the outage probability:

$$D_{\text{PC}}^{\text{CF}-\infty}(r) = \sup_{\beta\in\left(0,\frac{2-2r}{2-r}\right)} \min\left(D_1, D_2\right).$$

For $r < \frac{2}{3}$, we readily have $D_{\text{PC}}^{\text{CF}-\infty}(r) = 2-2r$ with the optimizers $\beta^* = \frac{1}{2}$. The minimum of $a+\alpha_1+\alpha_2$ is attained at $(1-r,1-r,0)$ and at $(1,1-2r)$ when $r < \frac{1}{2}$. When $r \in \left[\frac{1}{2},\frac{2}{3}\right)$, the minimum is attained at $(1-r,1-r,0)$ and at $(2-2r,0,0)$.

For $r \ge \frac{2}{3}$, some manipulation leads to

$$D_{\text{PC}}^{\text{CF}-\infty}(r) = \sup_{\beta\in\left(0,\frac{2-2r}{2-r}\right)} \frac{1-r}{1-\beta},$$

which has the optimizer $\beta^* = \frac{2-2r}{2-r} \le \frac{1}{2}$. With this β^*, the optimal $(a^*,\alpha_1^*,\alpha_2^*) = \left(\frac{(1-r)(2-r)}{r},0,0\right)$. □

Part III

End-to-end Distortion Exponent

Chapter 6

Distortion Exponent over MIMO Channels

In this chapter, the problem of source-channel coding over a multiple-antenna channel with quantized CSIT is considered. As in Chapters 3-5, we focus on the asymptotically high SNR regime. Upper bounds on the SNR exponent of the end-to-end distortion achieved with partial CSIT under a long-term power constraint are developed. It is shown that the distortion exponent with perfect CSIT grows unbounded as the ratio between the channel and source bandwidth increases, while the exponent achieved with any feedback link of fixed, finite resolution is bounded above by a polynomial of the product between the number of transmit and number of receive antennas. The resolution of the feedback link should grow with the bandwidth ratio to make the distortion exponent scale as fast as that in the perfect-CSIT case. We show that in order to achieve the optimal scaling the CSIT feedback resolution must grow logarithmically with the bandwidth ratio for MIMO channels, and faster than linear for the single-input single-output channel. The achievable distortion exponent of some hybrid schemes with heavily quantized feedback is also derived. The results demonstrate that dramatic performance improvement over the case of no CSIT can be achieved by combining simple schemes with a very coarse CSIT feedback.

6.1 Introduction

Consider the problem of source-channel coding over a multiple-antenna fading channel. We focus on the high SNR regime, and adopt the distortion exponent [LMWA05] as the performance measure. Studying the distortion exponent is useful in understanding the relation between the average end-to-end distortion and the spectral efficiency of the system at high SNR. The analysis is generally carried out under the elegant large-deviation framework of [ZT03], which is originally used to characterize the diversity-multiplexing tradeoff over multiantenna channels.

The distortion exponent problem has attracted a great deal of interest recently. A simple hybrid digital-analog scheme is proposed and shown to be optimal for high-compression systems in [CN05, CN07]. Such HDA joint source-channel coding schemes are first proposed for the broadcast scenario in [MP02]. The advantages predicted by the theoretical results in [MP02] are later realized by some practical schemes in e.g., [SPA02, SPA06]. Extensive studies of different layered source-channel coding schemes can be found in [GE05, GE08, BNC06]. Interestingly, layering with superposition coding is shown to be asymptotically optimal in the limit of infinite superimposed digital layers over a wide range of the ratio between the channel and source bandwidth [GE05, GE08, BNC06]. In practice, however, the complexity and other phenomena inherent to successive decoding such as error propagation will likely limit the potentials of such a superposition approach. The optimal source-channel rate allocation is also considered in [HG05].

Most previous work assumed only CSIR. While perfect channel knowledge at the transmitter is generally difficult to obtain, partial CSIT e.g., in the form of a few feedback bits, is often available. This motivates the present work where we consider a MIMO channel with quantized CSIT, under either a short- or a long-term power constraint.

A remarkable feature of the distortion SNR exponent setting is that the presence of CSIT yields an improvement even in the case of *short-term* power constraint. This is generally not the case for outage. This is because the outage problem, by its original formulation [OSW94], does not allow rate adaptation[1], while the distortion problem allows instantaneous adaptation of the rate and still results in some improved distortion exponent compared to a no-CSIT system.

In this chapter, we develop upper bounds on the achievable distortion over channels with partial CSIT. Interestingly, it turns out that with perfect CSIT, under a long-term power constraint, the distortion exponent grows linearly as a function of the bandwidth ratio. This is in contrast with the short-term power constraint case, where the distortion exponent is “saturated” if the bandwidth ratio is sufficiently large. However, our results also show that with any feedback link of fixed, finite number of quantization regions (referred to as the feedback “resolution” hereafter), the distortion exponent is bounded above by a polynomial of the product between the number of transmit and number receive antennas. It is necessary to let the feedback resolution grow with the bandwidth ratio to overcome this issue. We determine the rate at which the feedback resolution must grow with the bandwidth ratio in order to achieve the same distortion exponent scaling of the perfect CSIT bound. In particular, for MIMO systems we show that the resolution must grow logarithmically with the bandwidth ratio, i.e., the number of feedback bits grows only doubly-logarithmically.

We also study the performance of certain suboptimal hybrid digital-analog source-channel coding schemes when combined with *heavily* quantized CSIT. It

[1]Rate adaptation under the constraint of a minimum rate or a minimum multiplexing gain still results in a modified outage problem, as presented in Chapter 3.

turns out that over a wide range of practical bandwidth ratios, achieving a significant portion of the perfect-CSIT scheme is possible by combining simple schemes with a few bits of feedback information. While these schemes are generally suboptimal, we emphasize that very high distortion exponents are achievable with significantly lower complexity than many other no-CSIT techniques. Our results also highlight that even heavily quantized CSIT yields excellent performance when temporal power control is available.

At this point it is useful to provide a brief summary of our results, classified in terms of the bandwidth ratio b (see definition in Section 6.2). In a system with N_{t} transmit, N_{r} receive antennas, and feedback resolution K, let $m = \max(N_{\mathrm{t}}, N_{\mathrm{r}})$, $n = \min(N_{\mathrm{t}}, N_{\mathrm{r}})$. Then:

1. For $0 < b \leq \frac{m-n+1}{n}$: Proposition 6.1 shows that with perfect CSIT the distortion exponents with and without power control coincide $d_{\mathrm{PC}-\infty}(b) = d_{\mathrm{NPC}-\infty}(b)$. Thus in terms of distortion exponent, even the combination of perfect CSIT and power control does *not* help in this very high spectral efficiency regime. Furthermore, this distortion exponent is achievable even *without* CSIT by a simple HDA scheme in the range $b < \frac{1}{n}$ [CN05], and by layering and superposition coding in the range $b < \frac{m-n+1}{n}$ [BNC06]. Thus the distortion exponent in this regime is completely characterized and we have the noteworthy result that that CSIT feedback does not increase the distortion exponent in this range of b.
2. For $\frac{m-n+1}{n} \leq b < m-n+1$: We show that the two exponents with perfect CSIT still coincide, i.e., $d_{\mathrm{PC}-\infty}(b) = d_{\mathrm{NPC}-\infty}(b)$, meaning that in terms of distortion exponent, only rate control is necessary in this regime if full CSIT is available. Achievable exponents of simple HDA schemes with quantized (finite resolution) feedback are derived in Propositions 6.4–6.6. These schemes generally do not achieve the upper bounds.
3. For $m-n+1 \leq b < mn$: We have $d_{\mathrm{PC}-\infty}(b) > d_{\mathrm{UB-PC}-K}(b) > d_{\mathrm{NPC}-\infty}(b)$ where $d_{\mathrm{UB-PC}-K}(b)$ is an upper bound on the distortion exponent of any power-controlled system with feedback resolution K. This implies that no fixed-resolution feedback schemes can achieve $d_{\mathrm{PC}-\infty}(b)$. The HDA schemes considered in this work generally fall short of achieving $d_{\mathrm{UB-PC}-K}(b)$, but yield large improvement compared to no-CSIT schemes, even for coarsely quantized feedback.
4. For $b \geq mn$: We have $d_{\mathrm{PC}-\infty}(b) > d_{\mathrm{UB-PC}-K}(b) > d_{\mathrm{NPC}-\infty}(b)$. It is known that layering and superposition coding achieves $d_{\mathrm{NPC}-\infty}(b)$, even without CSIT [GE08]. The achievable distortion exponents under a *short-term* power constraint in Propositions 6.4–6.6 are therefore somewhat redundant. The achievability of the scheme in [GE08] however comes at the price of high complexity in the form of infinitely many superimposed digital layers. On the other hand, we show that the use of quantized CSIT feedback allows for *low-complexity* schemes that can achieve exponents close to $d_{\mathrm{NPC}-\infty}(b)$. The achievable exponents in Propositions 6.4–6.6 under a *long-term* power con-

straint are novel. They improve upon $d_{\text{NPC}-\infty}(b)$ and are close to $d_{\text{PC}-\infty}(b)$ over a wide range of b, even with very coarse feedback resolution.

6.2 System Model

Consider the transmission of an i.i.d. complex Gaussian source with zero mean and unit variance over a wireless flat-fading channel. The communication system uses N_{t} transmit and N_{r} receive antennas.

A complex source vector $\mathbf{s}_l$ of size N_{s} is generated every T channel uses. Herein, $b = T/N_{\text{s}}$ is the *channel bandwidth to source bandwidth* ratio, and l is the block index. The source vector is mapped to a joint source-channel codeword $\mathbf{X}_l$ of size $N_{\text{t}} \times T$. We consider the case where source blocks are mapped onto codewords spanning a single fading block, and coding across the fading blocks is forbidden (e.g., because of some strict decoding delay requirement). However, the transmit power may or may not be averaged over a long sequence of fading blocks (see later).

The complex-baseband received signal during fading block l can be written as

$$\mathbf{Y}_l = \mathbf{H}_l \mathbf{X}_l + \mathbf{N}_l. \tag{6.1}$$

The components of the $N_{\text{r}} \times N_{\text{t}}$ channel matrix $\mathbf{H}_l$ are i.i.d. complex Gaussian with zero mean and unit variance. The channel matrix is constant during a block of T channel uses, but changes independently from one block to another. The components of the temporally and spatially white Gaussian noise matrix $\mathbf{N}_l$ have zero mean and unit variance. For brevity, we omit the block fading index l whenever there is no ambiguity.

Assume perfect knowledge of $\mathbf{H}$ at the receiver. Given $\mathbf{H}$, the receiver employs a deterministic index mapping $\mathcal{I}(\mathbf{H}) \in \{1, \dots, K\}$ from channel matrix to feedback index, where the number of quantization regions K defines the feedback resolution. The feedback index is sent back to the transmitter via a noiseless, zero-delay dedicated feedback link, so that at the beginning of each fading block the corresponding index $\mathcal{I}(\mathbf{H})$ is known by the transmitter.

At the transmitter, the power allocation function

$$\mathcal{P} : \{1, \dots, K\} \to \{P_1, \dots, P_K\} \tag{6.2}$$

maps the feedback index onto the corresponding power level $\mathcal{P}(\mathcal{I}(\mathbf{H}))$, taking on values in a discrete set of possible power levels $\{P_1, \dots, P_K\}$. This means that the transmitted codeword satisfies

$$\frac{1}{T} \mathrm{E}\left[\|\mathbf{X}\|_{\text{F}}^2 | \mathbf{H}\right] \le \mathcal{P}(\mathcal{I}(\mathbf{H})),$$

where the expectation is over the source sequence and where $\|\cdot\|_{\text{F}}$ denotes the Frobenius norm. As usual, we consider two types of power constraint: a *long-term* power constraint (as in Chapter 3) requires that

$$\lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} \mathcal{P}(\mathcal{I}(\mathbf{H}_l)) \stackrel{\text{a.s.}}{=} \mathrm{E}_{\mathbf{H}}\, \mathcal{P}(\mathcal{I}(\mathbf{H})) \leq \mathrm{SNR}. \tag{6.3}$$

A short-term power constraint requires that

$$\mathcal{P}(\mathcal{I}(\mathbf{H})) \leq \mathrm{SNR}, \quad \text{with probability 1.} \tag{6.4}$$

The long-term power constraint models the case where power control is possible, and power saved in good fading conditions can be used over bad fading conditions. The short-term power constraint models the case where power control is impossible. Hence, we also refer to a system subject to (6.4) as one *without power control.*

Let $\hat{\mathbf{s}}$ be the reconstructed vector at the receiver corresponding to an input source vector $\mathbf{s}$. The mean squared error is adopted as the distortion measure,

$$\bar{\Delta} = \frac{1}{N_{\mathrm{s}}} \mathrm{E}\, \|\mathbf{s} - \hat{\mathbf{s}}\|_{\mathrm{F}}^2,$$

where the expectation is over randomness of the channel, the noise, and the source.

For later use, we indicate the mutual information corresponding to i.i.d. Gaussian inputs and a fixed channel matrix $\mathbf{H}$ by

$$C(\mathbf{H}) = \log\det\left(\mathbf{I}_{N_{\mathrm{r}}} + \frac{\mathrm{SNR}}{N_{\mathrm{t}}}\mathbf{H}\mathbf{H}^{\mathrm{H}}\right).$$

As in Chapter 3, we are interested in the asymptotic behavior of the system in the high-SNR regime. To that end, consider a *sequence* of joint source-channel codes and feedback schemes for fixed b and operating at increasing levels of SNR. Each code provides a mean squared error $\bar{\Delta}(\mathrm{SNR})$. The system is then said to achieve a distortion exponent of $d(b)$ if

$$\bar{\Delta}(\mathrm{SNR}) \doteq \mathrm{SNR}^{-d(b)}.$$

Clearly, the above definition of the distortion exponent bears a similarity to the definition of the diversity gain in Chapter 3.

6.3 Upper Bounds on Partial-CSIT Distortion Exponent

We first study the distortion exponent in the perfect-CSIT case under a long-term power constraint, characterized in the following proposition. The result gives an absolute upper bound on the distortion exponent of any system with perfect CSIR and any form of CSIT.

Proposition 6.1 (Perfect-CSIT distortion exponent)**.** *With perfect CSIT and a long-term power constraint, the optimal distortion exponent is given by*

$$d_{PC-\infty}(b) = bn.$$

Proof. Let $\lambda_n \geq \cdots \geq \lambda_1$ be the n largest eigenvalues of $\mathbf{HH}^{\mathrm{H}}$ and perform the change of variables $\alpha_i = -\log \lambda_i / \log \mathrm{SNR}$ [ZT03]. Let $P(\mathbf{H})$ be the total transmit power allocated given a channel $\mathbf{H}$, and let $\pi(\mathbf{H})$ be the SNR exponent of $P(\mathbf{H})$, i.e.

$$P(\mathbf{H}) \doteq \mathrm{SNR}^{\pi(\mathbf{H})}.$$

With this power allocation we have

$$C(\mathbf{H}) = \log\det\left(\mathbf{I}_{N_\mathrm{r}} + \frac{\mathrm{SNR}^{\pi(\mathbf{H})}}{N_\mathrm{t}}\mathbf{HH}^{\mathrm{H}}\right) = \sum_{i=1}^{n}\log\left(1 + \frac{\mathrm{SNR}^{\pi(\mathbf{H})-\alpha_i}}{N_\mathrm{t}}\right),$$

which only depends on $\mathbf{H}$ through the α_i's. We thus write $\pi(\alpha_1^n) \equiv \pi(\mathbf{H})$. Since we are only interested in the asymptotic behavior as $\mathrm{SNR} \to \infty$, restricting our attention to the class of power allocations satisfying

$$\int_{\alpha_1^n} \mathrm{SNR}^{\pi(\alpha_1^n)} f(\alpha_1^n) d\alpha_1^n \dot{\leq} \mathrm{SNR}$$

involves no loss of generality.[2] Herein $f(\alpha_1^n)$ denotes the joint p.d.f. of α_1^n. Following in the footsteps of the large-deviation analysis of [ZT03, DZ98], the long term power constraint yields the condition

$$\sup_{\alpha_1^n \geq 0}\left\{\pi(\alpha_1^n) - \sum_{i=1}^{n}(2i-1+m-n)\alpha_i\right\} \leq 1,$$

where we neglect the set of channel matrices with exponentially small probability measure by restricting $\alpha_i \geq 0$ [ZT03]. We use the notation $\alpha_1^n \geq 0$ to denote the set $\{\alpha_1^n : \alpha_1 \geq \cdots \geq \alpha_n \geq 0\}$. Since the instantaneous mutual information, and therefore the average distortion are non-decreasing functions of the allocated power, we conclude that the optimal $P(\mathbf{H})$ has exponent $\pi^*(\alpha_1^n) = 1 + \sum_{i=1}^{n}(2i-1+m-n)\alpha_i$.

With perfect CSIT the transmitter can adapt both its transmit power and the coding rate according to the resulting instantaneous mutual information. Assuming large block length N_s, such that the rate-distortion limit can be approached, the resulting instantaneous mean square error is given by [MP02, CN05]

$$\bar{\Delta}(\mathrm{SNR}, \mathbf{H}) = \exp\left(-bC(\mathbf{H})\right) = \det\left(\mathbf{I}_{N_\mathrm{r}} + \frac{\mathrm{SNR}^{\pi(\mathbf{H})}}{N_\mathrm{t}}\mathbf{HH}^{\mathrm{H}}\right)^{-b}$$

[2] Consider all power allocations such that $\mathrm{E}[P(\mathbf{H})] \doteq \mathrm{SNR}^{1-2\epsilon}$ for arbitrarily small $\epsilon > 0$. By definition, $\lim_{\mathrm{SNR}\to\infty} \frac{\log \mathrm{E}[P(\mathbf{H})]}{\log \mathrm{SNR}} = 1 - 2\epsilon$ thus there exists an $\bar{\mathrm{SNR}} < \infty$ so that for all $\mathrm{SNR} > \bar{\mathrm{SNR}}$, $1 - 3\epsilon \leq \frac{\log \mathrm{E}[P(\mathbf{H})]}{\log \mathrm{SNR}} \leq 1 - \epsilon$, leading to $\mathrm{E}[P(\mathbf{H})] \leq \mathrm{SNR}^{1-\epsilon} < \mathrm{SNR}$. This means the long-term power constraint is satisfied for all $\mathrm{SNR} > \bar{\mathrm{SNR}}$.

Using the optimal power allocation determined (asymptotically) above and averaging with respect to $\mathbf{H}$ we obtain the optimal mean square error as

$$\bar{\Delta}^*(\mathrm{SNR}) \doteq \int_{\alpha_1^n} \mathrm{SNR}^{-b\sum_{i=1}^n (\pi^*(\alpha_1^n)-\alpha_i)^+} f(\alpha_1^n) d\alpha_1^n$$

and the corresponding optimal distortion exponent is given by

$$d_{\mathrm{PC}-\infty}(b) = \inf_{\alpha_1^n \geq \mathbf{0}} \left\{ b\sum_{i=1}^n (\pi^*(\alpha_1^n) - \alpha_i)^+ + \sum_{i=1}^n (2i-1+m-n)\alpha_i \right\}.$$

The optimal values $\alpha_i^* = 0$, $\forall i$, and it follows that $d_{\mathrm{PC}-\infty}(b) = bn$. □

Recall from [CN05] that under a short-term power constraint, the distortion exponent with perfect CSIT (which also serves as the best known upper bound for the no-CSIT case) is given by

$$d_{\mathrm{NPC}-\infty}(b) = \sum_{i=1}^n \min(2i-1+m-n, b), \tag{6.5}$$

which "saturates" at $d_{\mathrm{NPC}-\infty}(b) = mn$ for $b \geq m+n-1$. In contrast, the distortion exponent in a power-controlled system $d_{\mathrm{PC}-\infty}(b)$ grows unbounded as $b \to \infty$. An intuitive explanation is that the distortion exponent is outage-limited in the high bandwidth expansion region. This limitation is overcome by using power control at the transmitter. Interestingly, the two bounds coincide iff $b \leq m-n+1$ meaning that, in terms of distortion exponent, power control is not useful unless the bandwidth ratio b is sufficiently large.

Next we develop the upper bound on the achievable distortion exponent given a *finite* feedback resolution K. For $0 < b \leq m-n+1$, the bound is trivially given by bn, $\forall K$. Hence, we restrict to the case $b \geq m-n+1$. Interestingly, our result shows that the achievable distortion with any finite-resolution feedback system is *bounded* above, for all b. However, the exponent upper bound increases with the feedback resolution K. This implies that increasing the resolution of the feedback link always yields advantages, provided that the bandwidth ratio is sufficient large.

Before proceeding to the derivation of the bounds, let us first define the two-variable function $D^{mn}(r,p)$, which is the diversity gain corresponding to a multiplexing gain r and power in the order of SNR^p, $p \geq 1$. This is given by

$$D^{mn}(r,p) = \inf_{\alpha_1^n \geq 0} \sum_{i=1}^n (2i-1+m-n)\alpha_i \quad \text{s.t.} \quad \sum_{i=1}^n (p-\alpha_i)^+ < r.$$

Notice that $D^{mn}(r,p) = pD^{mn}(r/p,1)$, where $D^{mn}(r,1)$ is the classical multiplexing-diversity tradeoff exponent of MIMO channels given explicitly in [ZT03]. For simplicity of notation, we will omit the superscript mn whenever this does not cause any confusion, i.e., $D(r,p) \equiv D^{mn}(r,p)$.

For convenience, we also recursively define

$$D_k \triangleq D(r_k, 1 + D_{k-1})$$

where $D_0 \triangleq 0$. Note that D_k, $k \geq 1$, is a function of k variables $r_1, \ldots, r_k$. We will use the notation $D_k(r)$ in the special case where

$$r_1 = \cdots = r_k \equiv r.$$

Indeed, $D_k(r)$ is the diversity gain corresponding to a multiplexing gain r of a rate-nonadaptive MIMO system with feedback resolution k.

The following bound is obtained by restricting the transmitter power allocation in the form (6.2), but allowing for perfect rate adaptation. This yields clearly an upper bound on the performance of a system based on feedback with resolution K.

Proposition 6.2 (K-power-level upper bound)**.** *For $b \geq m - n + 1$, let*

$$J \triangleq \max j \in \{1, 2, \ldots, n\} \quad s.t. \quad b \geq 2j - 1 + m - n.$$

Then the achievable distortion exponent of a MIMO channel with feedback resolution K under a long-term power constraint is upper-bounded by

$$d_{UB-PC-K}(b) = \sup_{n-J<r_1,\ldots,r_{K-1}<n} \min\left\{(1 + D_{K-1})\, d_{NPC-\infty}(b), br_1 + D_1, \ldots, br_{K-1} + D_{K-1}\right\}.$$

Proof. See Appendix 6.A. □

As discussed in Appendix 6.A, even though the system employs K power levels and perfect rate control, we can characterize the distortion exponent in terms of only $K - 1$ parameters $r_1, \ldots, r_{K-1}$. Interestingly, these r_k's can be interpreted as the *multiplexing gains* at the *boundaries* of the quantization regions in the feedback link. Also in Appendix 6.A, we provide a discussion on the computation of the $d_{\text{UB-PC-}K}(b)$, which involves a *nonconvex optimization.*

As a sanity check, with the convention $D_0 = 0$ we have $d_{\text{UB-PC-1}}(b) = d_{\text{NPC-}\infty}(b)$, since $d_{\text{NPC-}\infty}(b)$ is also obtained by assuming ideal rate adaptation and a single power level, $K = 1$.

We plot in Fig. 6.1 the upper bounds of Proposition 6.2 over a 2×2 channel. The bounds are relatively close to each other in the low-bandwidth-ratio regime, but quickly separate as b increases. That is, feedback resolution is a critical parameter in the low-spectral-efficiency regime.

Furthermore, as the bandwidth ratio grows, the K-level upper bound will converge to a finite limit. This is given by the following

Corollary 6.1. *For all b, $d_{UB-PC-K}(b) \leq \sum_{k=1}^{K}(mn)^k$. Furthermore, this value is attained asymptotically for $b \to \infty$.*

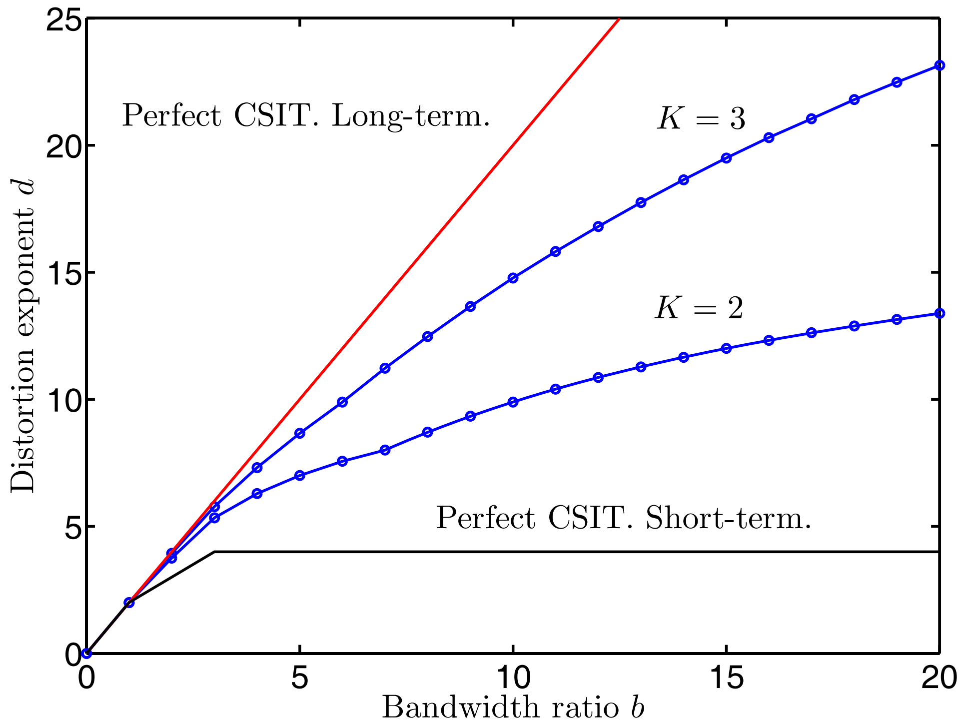

Figure 6.1: Upper bounds on the distortion exponents over a 2×2 channel.

Proof. For $b > m + n - 1$ we have $J = n$ and $d_{\text{NPC}-\infty} = mn$, thus

$$d_{\text{UB}-\text{PC}-K}(b) \leq \sup_{0<r_1,\ldots,r_{K-1}<n} (1 + D_{K-1})mn = \sum_{k=1}^{K} (mn)^k \tag{6.6}$$

where the supremum is attained at $r_1 = \cdots = r_{K-1} = 0$. Because $d_{\text{UB}-\text{PC}-K}(b)$ is a nondecreasing function of b, this implies

$$d_{\text{UB}-\text{PC}-K}(b) \leq \sum_{k=1}^{K} (mn)^K, \ \forall b.$$

In addition, by choosing $r_1 = \ldots = r_{K-1} = \min\left(\frac{\sum_{k=1}^{K}(mn)^k}{b}, n - \epsilon\right)$ where $\epsilon > 0$ is arbitrarily small, then $br_1 + D_1 < br_2 + D_2 < \cdots < br_{K-1} + D_{K-1}$ and $\lim_{b\to\infty} r_1 = 0$. For sufficiently large b we have

$$d_{\text{UB}-\text{PC}-K}(b) \geq \min\left(\sum_{k=1}^{K} (mn)^k + D(r_1, 1), (1 + D_{K-1}(r_1))mn\right).$$

Because

$$\sum_{k=1}^{K}(mn)^k + \lim_{r_1\downarrow 0} D(r_1,1) = \sum_{k=1}^{K}(mn)^k + mn > \sum_{k=1}^{K}(mn)^k = \lim_{r_1\downarrow 0}(1+D_{K-1}(r_1))mn,$$

we conclude that

$$\lim_{b\to\infty} d_{\mathrm{UB-PC-}K}(b) \geq (mn) + (mn)^2 + \cdots + (mn)^K.$$

This together with the upper bound (6.6) give

$$\lim_{b\to\infty} d_{\mathrm{UB-PC-}K}(b) = (mn) + (mn)^2 + \cdots + (mn)^K. \tag{6.7}$$

□

The limit in (6.7) is equal to the "maximum" diversity gain achieved with K feedback levels. Intuitively, with any finite-resolution feedback link, outage will eventually become the dominant factor in the low spectral efficiency regime.

One may expect that if we let the feedback resolution *grow* with the bandwidth ratio, $K \equiv K(b)$, the distortion exponent will also grow unbounded. It is then natural to investigate the rate of increase of the feedback resolution $K(b)$ with b such that the same behavior of the upper bound of Proposition 6.1 is achieved. In particular, we define

$$\eta = \lim_{b\to\infty} \frac{d_{\mathrm{UB-PC-}K(b)}(b)}{b}. \tag{6.8}$$

This can be interpreted as an upper bound to the asymptotic efficiency of a feedback scheme with resolution $K(b)$. Since $d_{\mathrm{UB-PC-}K(b)}(b)$ is only an upper bound, (6.8) is a necessary condition for the distortion of a scheme with resolution $K(b)$ to behave like $\mathrm{SNR}^{-b\eta}$ for large b. Recall that with ideal CSIT and power control the distortion is in the order of SNR^{-bn}. Hence, $d_{\mathrm{UB-PC-}K(b)}(b)$ has the same behavior of the perfect CSIT upper bound for large b if $\eta = n$. In this case we say that the finite resolution scheme is *asymptotically efficient*.

Let us first take a look at the special case of a single-input single-output (SISO), where exponents can be derived in closed form. With $m = n = 1$, a direct investigation given in Appendix 6.B shows that, for $b > 1$,

$$d_{\mathrm{UB-PC-}K(b)}(b) = b - b\left(1 - \frac{1}{b}\right)^{K(b)} \tag{6.9}$$

thus

$$\eta = \lim_{b\to\infty} 1 - \left(1 - \frac{1}{b}\right)^{K(b)}.$$

For example, by letting the resolution grow as $K(b) = \lceil b \rceil$ where $\lceil x \rceil$ is the smallest integer that is not smaller than x, we have

$$\eta = 1 - \frac{1}{e} < 1.$$

Thus using $\log_2\lceil b\rceil$ bits of feedback is not sufficient to achieve the "full" potential of perfect CSIT.

Indeed, the condition

$$\lim_{b\to\infty}\left(1-\frac{1}{b}\right)^{K(b)}=0 \tag{6.10}$$

also guarantees the existence of a scheme with distortion behaving similarly to SNR^{-bn}. To see this sufficiency, consider a simple separate source-channel coding system with optimized rate allocation using a feedback resolution of $K(b)$ where the transmitter, given i, allocates a fixed rate $r_i \log \mathrm{SNR}$ for all $\mathbf{H}\in\mathcal{R}_i$. Herein r_i is the *maximum* multiplexing gain that *all* channel realizations in $\mathcal{R}_i$ can support. It is shown in Appendix 6.B that the achievable distortion exponent of this approach, referred to as the single-layer coding scheme, is

$$d_{\mathrm{SL\text{-}PC\text{-}}K(b)}(b)=b-b\left(1-\frac{1}{b+1}\right)^{K(b)}. \tag{6.11}$$

Given that the condition (6.10) holds, we have

$$\lim_{b\to\infty}\frac{d_{\mathrm{SL\text{-}PC\text{-}}K(b)}(b)}{b}=1-\lim_{b\to\infty}\left(1-\frac{1}{b+1}\right)^{K(b)}=1,$$

meaning that an actual system is asymptotically efficient. An example of such asymptotically efficient feedback resolution functions is $K(b)=\lceil b\log b\rceil$.

In Fig. 6.2, we plot the upper bounds $d_{\mathrm{UB-PC-}K(b)}(b)$ for two different functions $K(b)$. Interestingly, even though the analysis is asymptotical in b, it reflects quite accurately the behavior of the system even with moderate b.

The generalization of the results obtained in the SISO case to a MIMO channel is a non-trivial problem, and the closed-form expression of η does not appear to be tractable in general. Nevertheless, we were able to find bounds and obtain simple conditions for necessity and sufficiency of the rate of increase of $K(b)$ such that the optimal $\eta=n$ can be achieved.

Proposition 6.3. *If $m=\max(N_t,N_r)>1$, a sufficient condition on the feedback resolution $K(b)$ such that the system is asymptotically efficient is given by*

$$\lim_{b\to\infty}\frac{b}{[(m-n+1)n]^{K(b)}}=0,$$

and a necessary condition is given by

$$\lim_{b\to\infty}\frac{b}{(mn)^{K(b)}}\leq\frac{m}{mn-1}.$$

Proof. See Appendix 6.C. □

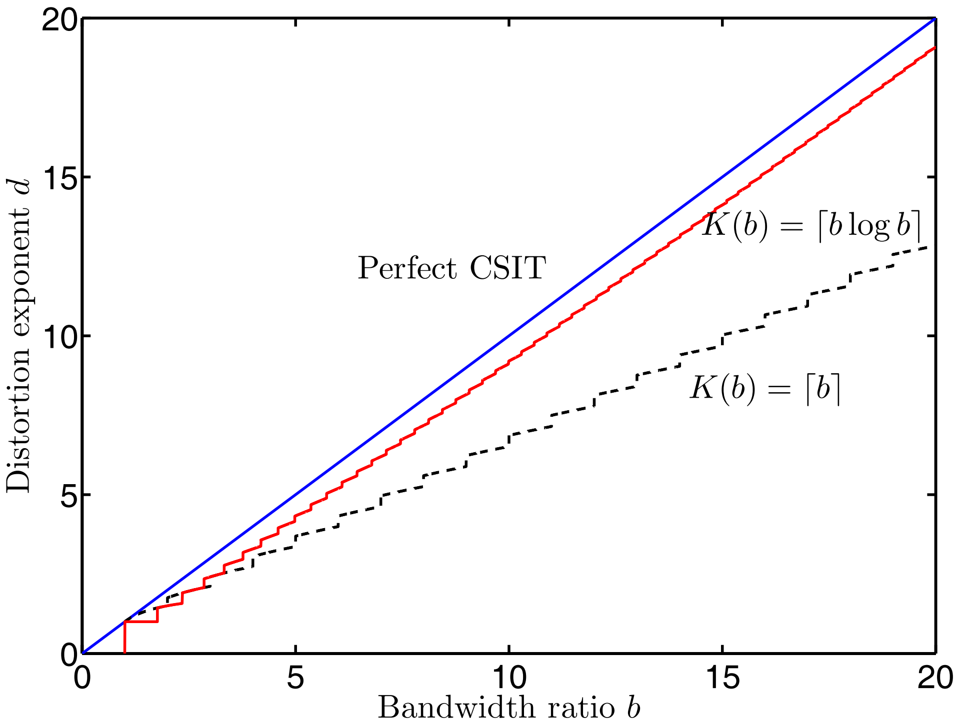

Figure 6.2: Upper bounds on the distortion exponents over a SISO channel where the feedback resolution is bandwidth ratio dependent.

Proposition 6.3 essentially provides upper and lower bounds on the *minimum* growth rate of the feedback resolution to make the distortion exponent behave similarly to the perfect-CSIT case. Clearly the bounds are not tight in a strict sense. However, in terms of order of growth, we can conclude that with $m > 1$, having a resolution $K(b) = \Theta(\log b)$ as $b \to \infty$ is necessary and sufficient for the system to behave like an ideal-CSIT.[3]

Furthermore, Proposition 6.3 implies that the more antennas we have, the slower $K(b)$ needs to be scaled. This can be attributed to the "channel hardening" effect [HMT04], i.e., the instantaneous mutual information becomes more and more deterministic as the number of antennas increases. Given a fixed n, while the perfect-CSIT distortion exponent only depends on n, it is required less feedback to achieve that upper bound with a larger m. We also see that the SISO channel is the worst case channel in the sense that, to behave like a perfect-CSIT system, a SISO

[3]Notice the pleasing fact that, in the MIMO case, the number of bits per feedback message must grow only doubly-logarithmically in b in order to have the same behavior of the optimal distortion exponent.

system requires a faster than linearly scaled $K(b)$, in contrast to a logarithmically scaled $K(b)$ when $m > 1$.

Note that the above result is asymptotic in the bandwidth ratio b. However, the bandwidth ratio b is small in practice. Furthermore, for most practical systems, the number of feedback bits are typically fixed and small. Therefore, in the following sections we will study the performance of certain hybrid schemes when combined with a fixed, low-resolution feedback link. We consider the case where the block length T can be made arbitrarily large. More precisely, we *first* let T and N_s grow unbounded (with a fix ratio $T/N_\mathrm{s} = b$) for a fixed SNR, and then consider a sequence of such schemes operating at increasing values of SNR.

6.4 Achievable Distortion Exponents: HDA with Dimension Splitting

In this section, we present the distortion exponent achieved by a hybrid digital-analog scheme with dimension (bandwidth or time) splitting. The HDA approach [MP02, CN05] together with the corresponding proposed feedback schemes are described as follows.

First consider the case $N_\mathrm{r} \geq N_\mathrm{t} = n$. Conditioned on a feedback index i, the transmitter uses a digital part (tandem encoder [MP02]) with channel rate $r_i \log \mathrm{SNR}$ and power in the order SNR^{p_i}, which occupies a fraction $1 - \frac{1}{bn}$ of the channel bandwidth. The output of the source encoder is subtracted from the source vector $\mathbf{s}$, suitably scaled so that its power is SNR^{p_i}, and then transmitted directly from the antennas (N_t symbols at each time instant). The analog part therefore occupies $\lceil N_\mathrm{s}/n \rceil$ channel uses, or a fraction $\frac{1}{bn}$ of the channel bandwidth.

The receiver decodes the digital part and reconstructs $\hat{\mathbf{s}} = \mathbf{0}$ if the decoding fails.[4] If the decoding is successful, a linear minimum mean square error (MMSE) filter is applied to estimate the analog part, and the output of the filter is added to the digital part.

The index mapping is explicitly described as follows

$$\mathcal{I}(\mathbf{H}) = \begin{cases} K & \text{if } C(\mathbf{H}) < r_K \log \mathrm{SNR} \\ \max\{i \in \{1, \dots, K\} : C(\mathbf{H}) \geq r_i \log \mathrm{SNR}\} & \text{otherwise,} \end{cases}$$

where $r_1 > \cdots > r_K$. That is, the receiver feeds back the maximum rate from a finite set that the channel can support. If the channel cannot support the smallest rate $r_K \log \mathrm{SNR}$, an arbitrary index can be sent back, which in this case is set to be K.

[4] When the digital part fails, then even if the analog part $\mathbf{s}_\mathrm{A}$ (quantization error) is known exactly at the receiver so that $\hat{\mathbf{s}} = \mathbf{s}_\mathrm{A}$, we still have $\mathrm{E}\,\|\mathbf{s} - \mathbf{s}_\mathrm{A}\|^2 = \mathrm{E}\,\|\mathbf{s}_\mathrm{D}\|^2 \doteq \mathrm{SNR}^0$. The exponent equality is due to the fact that $\mathrm{E}\,\|\mathbf{s}\|^2 = \mathrm{E}\,\|\mathbf{s}_\mathrm{D}\|^2 + \mathrm{E}\,\|\mathbf{s}_\mathrm{A}\|^2$ (from the Gaussianity and orthogonality of $\mathbf{s}_\mathrm{D}$ and $\mathbf{s}_\mathrm{A}$) and that $\mathrm{E}\,\|\mathbf{s}_\mathrm{A}\|^2 \dot{\leq} \mathrm{SNR}^0$.

Setting $\hat{\mathbf{s}} = \mathbf{0}$ therefore has no effect on the distortion exponent.

Proposition 6.4. *For $N_r \geq N_t \equiv n$ and $b \geq \frac{1}{n}$, an HDA scheme with dimension splitting, and feedback resolution $K \geq 2$ can achieve a distortion exponent of*

$$d_{HDA-DIM-NPC-K}(b) = \sup_{0<r_K<\ldots<r_1<n} \min \Big\{ D(r_K, 1), \left(b - \frac{1}{n}\right) r_1 + 1, \left(b - \frac{1}{n}\right) r_2 + D(r_1, 1) + (r_1 + 1 - n)^+, \ldots, \left(b - \frac{1}{n}\right) r_K + D(r_{K-1}, 1) + (r_{K-1} + 1 - n)^+ \Big\}$$

under a short-term power constraint, and

$$d_{HDA-DIM-PC-K}(b) = \sup_{0<r_K<\ldots<r_1<n} \min \Big\{ D_K, \left(b - \frac{1}{n}\right) r_1 + 1, \left(b - \frac{1}{n}\right) r_2 + 2D_1 - D_0 + (r_1 + (1-n)(1+D_0))^+, \ldots, \left(b - \frac{1}{n}\right) r_K + 2D_{K-1} - D_{K-2} + (r_{K-1} + (1-n)(1+D_{K-2}))^+ \Big\}$$

under a long-term power constraint. Herein $D_k \triangleq D(r_k, 1 + D_{k-1})$ with $D_0 \triangleq 0$.

An interesting feature of this dimension-splitting approach is that its distortion exponent is achievable even with a *finite* block length T, as shown in the no-CSIT case in [CN07]. In particular, by employing approximately universal codes for the digital part [TV06, EKP$^+$06] and applying results in high-rate vector quantization [NN95], we can show that having $T \geq N_t + \lceil \frac{N_s}{n} \rceil$ is sufficient to achieve the distortion in Proposition 6.4.

The proof of Proposition 6.4, presented in Appendix 6.D, is outlined as follows. For clarity, we consider $K = 2$ under a short-term power constraint. The distortion consists of three terms: the distortion in quantization region 1, that in region 2 when the digital part is successfully decoded and the outage probability (i.e., when the digital part fails leading to a distortion of order SNR^0). The inner minimization is to find the dominating (slowest decayed) term among these three terms as $\text{SNR} \to \infty$ and the outer maximization is to find the best possible rate allocation for each quantization region. The intuition behind the presence of the term $(r_1 + 1 - n)^+$ is that, when the multiplexing gain r_1 is sufficiently large, the most likely channels to cause an outage event are still "good enough" to contribute to an "extra exponent" when the MMSE filter is applied.

The long-term power constraint case can be proved similarly. In this case, a power in the order of $\mathrm{SNR}^{1+D(r_1,1)}$ can be applied to the second quantization region $\mathcal{R}_2$ without violating the power constraint, leading to a recursive characterization as in the second part of Proposition 6.4.

The achievable distortion exponent in Proposition 6.4 can be numerically evaluated by equating the $K+1$ terms, then expressing $r_2, \ldots, r_K$ as functions of r_1, and solving for r_1. Interestingly, in the SISO case, i.e., $m = n = 1$, explicit expressions of the achievable distortion exponent are found to be $D_{\mathrm{HDA-DIM-NPC-}K}(b) = 1$, $\forall b \geq 1, \forall K \geq 1$ and

$$D_{\mathrm{HDA-DIM-PC-}K}(b) = b - \frac{(b-1)^K}{b^{K-1}}, \quad b \geq 1$$

which coincides with the upper bound $D_{\mathrm{UB-PC-}K}(b)$ in (6.9). Thus in the SISO case, the HDA with dimension splitting scheme is *optimal* for any feedback resolution K. This however does not hold for an arbitrary MIMO channel.

For the case $N_\mathrm{t} > N_\mathrm{r}$, the scheme described at the beginning of this section is not directly applicable since the MMSE filter will *not* produce an MSE that decays with SNR (we have more unknowns than equations at the receiver). We thus consider using a fixed *subset* N_r out of N_t possible transmit antennas. Unfortunately, in such an approach, the event that $\mathbf{H}$ is not in outage does not exclude the event that the $N_\mathrm{r} \times N_\mathrm{r}$ sub-matrix $\hat{\mathbf{H}}$ is in outage. Therefore, in this case we have only been able to derive lower bounds on the achievable distortion exponents. The results are summarized in the following proposition and proofs are outlined in Appendix 6.E.

Proposition 6.5. *For $N_t > N_r \equiv n$, and $b \geq \frac{1}{n}$, an HDA scheme with dimension splitting, and feedback resolution $K \geq 2$ can achieve a distortion exponent not smaller than*

$$\begin{aligned} d_{HDA-DIM-NPC-K}(b) = \sup_{0<r_K<\cdots<r_1<n} \min \{ & D^{mn}(r_K, 1), \\ & \left(b - \frac{1}{n}\right) r_1 + 1, \\ & \left(b - \frac{1}{n}\right) r_2 + D^{nn}(r_1, 1) + (r_1 + 1 - n)^+, \cdots, \\ & \left(b - \frac{1}{n}\right) r_K + D^{nn}(r_{K-1}, 1) + (r_{K-1} + 1 - n)^+ \Big\} \end{aligned}$$

under a short-term power constraint, and

$$
\begin{aligned}
&d_{HDA-DIM-PC-K}(b) = \\
&\sup_{0<r_K<\ldots<r_1<n} \min\{D_K, \\
&\left(b-\frac{1}{n}\right) r_1 + 1, \\
&\left(b-\frac{1}{n}\right) r_2 + D_1 - D_0 + D^{nn}(r_1, 1+D_0) + (r_1 + (1-n)(1+D_0))^+, \ldots, \\
&\left(b-\frac{1}{n}\right) r_K + D_{K-1} - D_{K-2} + D^{nn}(r_{K-1}, 1+D_{K-2}) \\
&\qquad\qquad +(r_{K-1} + (1-n)(1+D_{K-2}))^+\}
\end{aligned}
$$

under a long-term power constraint. Herein $D_k \triangleq D^{mn}(r_k, 1+D_{k-1})$ *with* $D_0 \triangleq 0$.

6.5 Achievable Distortion Exponents: HDA with Power Splitting

In this section, we study another HDA scheme when combined with quantized feedback. Again first consider $N_\mathrm{r} \geq N_\mathrm{t}$. Conditioned on an index i, the transmitter encodes $\mathbf{s}$ with a tandem encoder using a channel code with rate $r_i \log \mathrm{SNR}$. The digital part is assigned power in the order of $\mathrm{SNR}^{p_{\mathrm{D}i}}$ (as usual, we assume that the digital part equally allocates power over the antennas). The output of the the source coder is subtracted from $\mathbf{s}$. The outcome is then properly scaled so that its power is in the order of $\mathrm{SNR}^{p_{\mathrm{A}i}}$ where $\mathrm{SNR}^{p_{\mathrm{A}i}} \leq \mathrm{SNR}^{p_{\mathrm{D}i}}$, mapped directly to the antennas, and finally *superimposed* onto the digital codewords. The analog part utilizes only $\lceil \frac{N_\mathrm{s}}{n} \rceil$ channel uses. Thus the digital codewords are *interference-free* for a fraction $1 - \frac{1}{bn}$ of the time. Therefore, we say that the codewords are *partially* superimposed. The receiver decodes the digital part treating the analog interference as Gaussian noise. If the decoding is successful, an MMSE filter estimates the analog part and adds the outcome to the digital part. The corresponding outage exponent is characterized in the following lemma. Herein the outage event is the event that the mutual information, with the analog part treated as noise, is below the rate of the digital layer.

Lemma 6.1 (Outage exponent of partially superimposed codewords)**.** *The outage exponent of a digital layer with rate* $r \log SNR$ *and power* SNR^{p_D}*, with partially superimposed analog layer (interference) with analog power* SNR^{p_A}*, is denoted by* $D_{SP}(r, p_D, p_A)$ *and it is given by the piecewise linear function joining the points*

$$
\left(j\left(p_D - \frac{p_A}{bn}\right), p_D(m-j)(n-j)\right),
$$

and

$$\Big(j\left(p_D - \frac{p_A}{bn}\right) + p_D - p_A,$$
$$p_D(m-j-1)(n-j-1) + (m+n-2j-1)p_A\Big)\,,$$

and the points

$$\Big(j\left(p_D - \frac{p_A}{bn}\right) + p_D - p_A,$$
$$p_D(m-j-1)(n-j-1) + (m+n-2j-1)p_A\Big)\,,$$

and

$$\left((j+1)\left(p_D - \frac{p_A}{bn}\right), p_D(m-j-1)(n-j-1)\right)$$

for $j = 0, \ldots, n-1$.

Proof. In the following we only consider the case $p_\mathrm{D} \geq p_\mathrm{A}$ because in the case $p_\mathrm{A} > p_\mathrm{D}$, the outage event has the same exponent as the event

$$\left(1 - \frac{1}{bn}\right)\sum_{i=1}^{n}(p_\mathrm{D} - \alpha_i)^+ < r,$$

i.e., the portion of the digital part superimposed with the analog part is completely useless. Such a scheme coincides with a dimension-splitting scheme where we allocate a fraction $1 - \frac{1}{bn}$ of the dimensions to the digital part and encode at a rate $\frac{r}{1-\frac{1}{bn}}$. Therefore, only the case $p_\mathrm{D} \geq p_\mathrm{A}$ matters.

Noticing that $\mathrm{SNR}^{p_\mathrm{D}} + \mathrm{SNR}^{p_\mathrm{A}} \doteq \mathrm{SNR}^{p_\mathrm{D}}$ for any $p_\mathrm{A} \leq p_\mathrm{D}$, the outage probability can be written as

$$\begin{aligned}
&\Pr\Bigg(\frac{1}{bn}\log\det\left(\mathbf{I}_{N_\mathrm{r}} + \left(\mathbf{I}_{N_\mathrm{r}} + \mathrm{SNR}^{p_\mathrm{A}}\mathbf{H}\mathbf{H}^\mathrm{H}\right)^{-1}\mathrm{SNR}^{p_\mathrm{D}}\mathbf{H}\mathbf{H}^\mathrm{H}\right)\\
&\qquad + \left(1 - \frac{1}{bn}\right)\log\det\left(\mathbf{I}_{N_\mathrm{r}} + \mathrm{SNR}^{p_\mathrm{D}}\mathbf{H}\mathbf{H}^\mathrm{H}\right) < r\log\mathrm{SNR}\Bigg)\\
&\doteq \Pr\left(\sum_{i=1}^{n}\log\mathrm{SNR}^{(p_\mathrm{D}-\alpha_i)^+} - \frac{1}{bn}\sum_{i=1}^{n}\log\mathrm{SNR}^{(p_\mathrm{A}-\alpha_i)^+} < \log\mathrm{SNR}^r\right)\\
&\doteq \Pr\left(\sum_{i=1}^{n}\left((p_\mathrm{D}-\alpha_i)^+ - \frac{1}{bn}(p_\mathrm{A}-\alpha_i)^+\right) < r\right).
\end{aligned}$$

Then the outage exponent is given by

$$\begin{aligned}
D_\mathrm{SP}(r, p_\mathrm{D}, p_\mathrm{A}) = \inf_{\alpha_1^n \geq 0} &\sum_{i=1}^{n}(2i-1+m-n)\alpha_i\\
\text{s.t.}\quad &\sum_{i=1}^{n}\left((p_\mathrm{D}-\alpha_i)^+ - \frac{1}{bn}(p_\mathrm{A}-\alpha_i)^+\right) < r
\end{aligned}$$

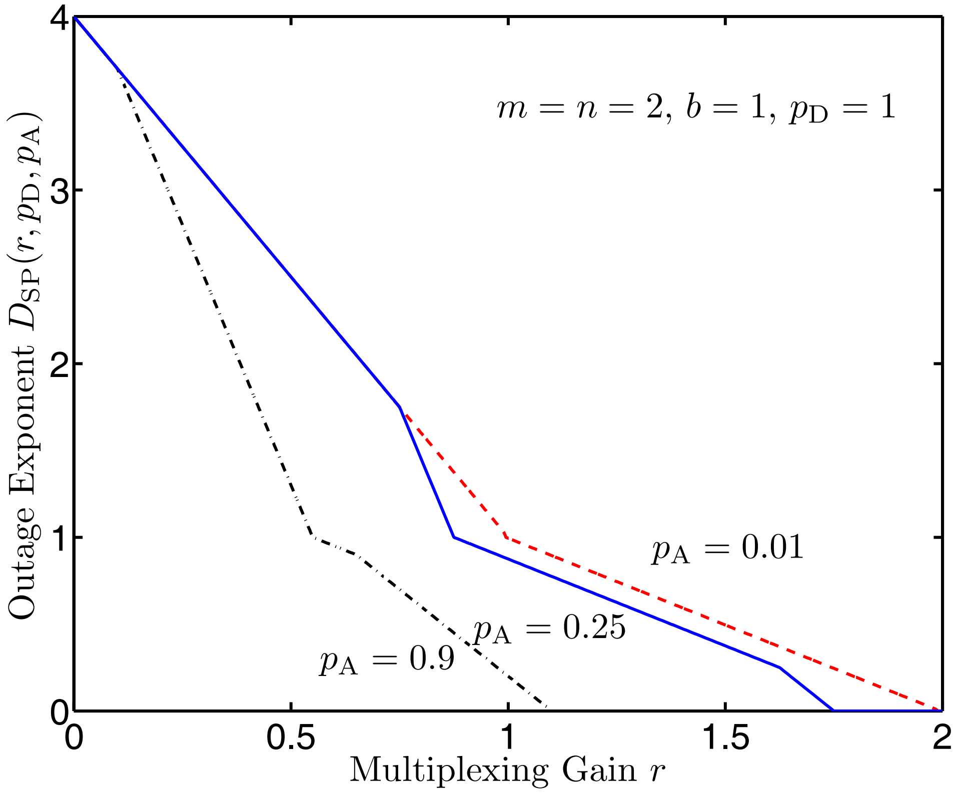

Figure 6.3: The outage exponents in Lemma 6.1 for a 2×2 channel.

The asserted result is the explicit solution of the above linear programming problem. □

We notice that the outage exponent $D_{\mathrm{SP}}(r, p_{\mathrm{D}}, p_{\mathrm{A}})$, illustrated in Fig. 6.3, generally consists of more segments than in the case of no interference, i.e., when $p_{\mathrm{D}} = 1$ and $p_{\mathrm{A}} = 0$ [ZT03]. Furthermore, $D_{\mathrm{SP}}(r, p_{\mathrm{D}}, p_{\mathrm{A}})$ is generally neither convex nor concave.

Using the Lemma 6.1, we obtain the following result characterizing the achievable distortion of an HDA scheme with power splitting and quantized feedback.

Proposition 6.6. *For $N_r \geq N_t \equiv n$, an HDA scheme with power splitting and* *feedback resolution $K \geq 2$ can achieve a distortion exponent of*

$$d_{HDA-POW-NPC-K}(b) = \sup_{0<r_1,\dots,r_K<n, 0<p_1,\dots,p_K<1} \min\{$$

$$D_{SP}(r_K, 1, p_K),$$

$$br_1 + p_1,$$

$$br_2 + D_{SP}(r_1, 1, p_1) + \max\left(p_2 - \frac{n - \frac{p_1}{b} - r_1}{1 - \frac{1}{bn}}, p_2 - 1 + r_1 - (n-1)\left(1 - \frac{p_1}{bn}\right), 0\right),$$

$$\dots$$

$$br_K + D_{SP}(r_{K-1}, 1, p_{K-1})$$
$$+ \max\left(p_K - \frac{n - \frac{p_{K-1}}{b} - r_{K-1}}{1 - \frac{1}{bn}}, p_K - 1 + r_{K-1} - (n-1)\left(1 - \frac{p_{K-1}}{bn}\right), 0\right)\Big\}$$

under a short-term power constraint, and

$$d_{HDA-POW-PC-K}(b) = \sup_{0<r_1,\dots,r_K<n, \{p_k<1+D_{k-1}\}} \min\{$$

$$D_K, br_1 + p_1,$$

$$br_2 + \max\left(0, p_2 - p_{D1}, p_2 - \frac{np_{D1} - \frac{p_1}{b} - r_1}{1 - \frac{1}{bn}}, p_2 - p_{D1} + r_1 - (n-1)\left(p_{D1} - \frac{p_1}{bn}\right)\right)$$
$$+ D_1,$$

$$\dots$$

$$br_K + \max\left(0, p_K - p_{DK-1}, p_K - \frac{np_{DK-1} - \frac{p_{K-1}}{b} - r_{K-1}}{1 - \frac{1}{bn}} p_K - p_{DK-1}\right.$$
$$\left. + r_{K-1} - (n-1)\left(p_{DK-1} - \frac{p_{K-1}}{bn}\right)\right) + D_{K-1}\Big\}$$

under a long-term power constraint. Herein $D_k \triangleq D_{SP}(r_k, 1 + D_{k-1}, p_k)$, $p_{Dk} \triangleq 1 + D_{k-1}$ with $D_0 \triangleq 0$.

The p_i's herein refer to the SNR exponent of the power allocated to the analog layer. In the short-term power constraint case, the digital layer has $p_{\mathrm{D}} = 1$ for all feedback indices. In the long-term constraint case, the power exponent of the digital layer is denoted by $p_{\mathrm{D}k}$.

The proof of Proposition 6.6, deferred to Appendix 6.F, essentially follows that of Proposition 6.4. The presence of more terms inside the $\max(\cdot)$ is due to the fact that $g(x) = (p_{\mathrm{D}} - x)^+ - \frac{1}{bn}(p_{\mathrm{A}} - x)^+$ is piece-wise linear in $0 < x < p_{\mathrm{D}}$. We omit a lower bound on the achievable distortion in the case $N_{\mathrm{t}} > N_{\mathrm{r}}$, which is similar to Proposition 6.5.

Unlike the dimension splitting case, even numerically evaluating the distortion exponent in Proposition 6.6 is difficult due to the presence of the parameters

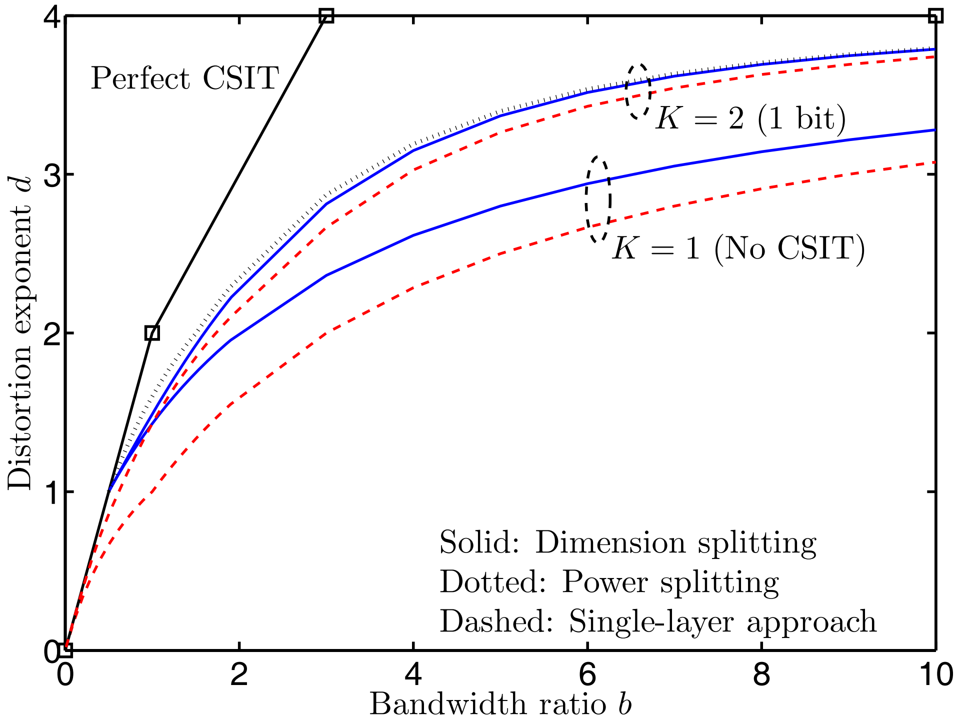

Figure 6.4: Distortion exponent of different schemes over a 2×2 channel. A short-term power constraint is assumed.

$p_1, \ldots, p_K$. In the numerical section, we resorted to a grid search on the $(p_1, \ldots, p_K)$ space. This is of course prohibitive even for a moderate K.

The achievability of power-splitting schemes is much more difficult to analyze for finite-length source and channel codes. The quantization error for a finite source vector length N_s is generally not Gaussian, making the analysis complicated. Even if it were possible to replace the quantization error with i.i.d. Gaussian noise, then we still need to show the existence of channel codes with finite-length having the approximately universal property, taking the partially overlapping behavior of the system into account. This is a very difficult task even for a traditional no-interference model [TV06, EKP$^+$06]. Finite-length coding analysis of the proposed scheme remains an interesting topic for future work.

6.6 Numerical Examples

We plot in Fig. 6.4 the achievable distortion of different schemes under a short-term power constraint. As can be seen, even one bit of feedback information improves

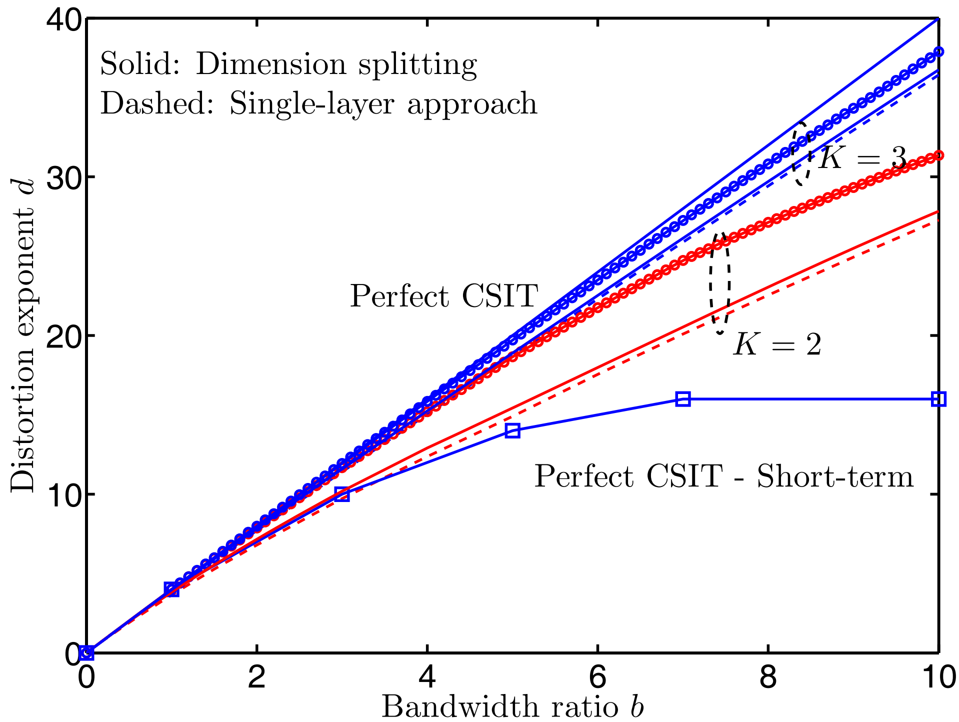

Figure 6.5: Achievable distortion exponents of different schemes over a 4×4 channel. A long-term power constraint is assumed. The K-power-level upper bounds, marked by circles, are also plotted.

the performance significantly. For comparison, we also plot the performance of the optimal joint source-channel rate allocation, see e.g., [HG05], (i.e., the single-layer coding scheme). For the sake of clarity, we recall here that a "single-layer" coding scheme is a scheme that for each feedback index $\mathcal{I}(\mathbf{H}) = i$ allocates a power level P_i and a channel coding rate R_i, and makes use of an all digital "tandem" code with source coding rate bR_i and channel coding rate R_i. This differs from the proposed HDA schemes that make use of hybrid digital-analog strategies for each value of the feedback index, and also from the upper bounds, for which we have assumed a finite and discrete set of power levels $P_1, \ldots, P_K$ but a continuum of tandem coding schemes such that the source coding rate can be made equal to the instantaneous channel mutual information. Notice also that single-layer schemes can be obtained as special cases of the proposed power-splitting HDA scheme by allocating no power to the analog layer.

It is demonstrated in Fig. 6.5, where the distortion exponents over a 4×4 channel are plotted, that temporal power control is instrumental in achieving a large distortion exponent in the high bandwidth ratio regime. For example, a partial power controlled system with a *single bit* of feedback outperforms a no power control system with *perfect* CSIT by a wide margin for all $b \geq 5$. In the scenarios considered, one can achieve performance very close to perfect-CSIT with a very coarse feedback resolution K. Recall however from Proposition 6.2 that the performance gap between the finite-feedback case and the perfect-CSIT case grows as b increases, unless the feedback resolution also grows (logarithmically) with b.

In Fig. 6.5 the upper- and lower bounds corresponding to the same value of K will eventually coincide as the bandwidth ratio becomes large. This is because even the simple single-layer coding scheme can achieve the upper bound $D_{\text{UB-PC-}K}(b)$ as $b \to \infty$:

$$\begin{aligned}
\lim_{b\to\infty} d_{\text{SL-PC-}K}(b) &= \lim_{b\to\infty} \sup_{r_1,\ldots,r_K} \min\left(br_1, br_2 + D_1, \ldots, br_K + D_{K-1}, D_K\right) \\
&\geq \lim_{b\to\infty} \min\left(\sum_{k=1}^{K}(mn)^k, D_K\left(\frac{\sum_{k=1}^{K}(mn)^k}{b}\right)\right) \\
&= \sum_{k=1}^{K}(mn)^k
\end{aligned}$$

where the inequality is obtained by choosing $r_1 = \cdots = r_K = \frac{\sum_{k=1}^{K}(mn)^k}{b}$ and the last equality is due to $\lim_{\epsilon\downarrow 0} D_K(\epsilon) = \sum_{k=1}^{K}(mn)^k$.

6.7 Conclusion

We have studied the end-to-end distortion exponent of single-user fading MIMO channels with limited feedback. One remarkable conclusion is that the optimized single-layer approach achieves performance very close to that of the HDA schemes even for a moderate feedback resolution. Our single-layer results also suggest that in a practical scenario, the schemes with a library of source and channel codes that adapt rate and power based on partial feedback is a practically sound approach, provided that power and rate control are *jointly* optimized. This highlights the importance of a careful cross-layer design in multimedia transmission over wireless fading channels. Finding joint-source channel coding schemes with properly designed feedback links that achieve the upper bound $d_{\text{UB}-\text{PC}-K}(b)$ for any feedback resolution K (similar to the HDA scheme with dimension splitting over a SISO channel) remains an open problem.

Appendices for Chapter 6

6.A Partial-CSIT Upper Bound

Proof of Proposition 6.2

The upper bound is obtained by assuming that perfect CSIT is available, but the system is only allowed to use K power levels.

The set of all channel realizations that use the same power level is referred to as a quantization region. Let $\mathcal{R}_k, p_k \geq 1$ be the kth quantization region and the corresponding SNR exponent of the power level. Without loss of generality, assume $1 = p_1 < \cdots < p_K < p_{K+1} = \infty$. The long-term power constraint leads to

$$\sup_{\alpha_1^n \in \mathcal{R}_k} p_k - \sum_{i=1}^{n} (2i - 1 + m - n)\alpha_i \leq 1,$$

or equivalently

$$\inf_{\alpha_1^n \in \mathcal{R}_k} \sum_{i=1}^{n} (2i - 1 + m - n)\alpha_i \geq p_k - 1.$$

Achievable rates and therefore distortion exponents are non-decreasing functions of power. Thus the optimal kth quantization region simply consists of all channel realizations where the highest power level can be applied without violating the power constraint:

$$\mathcal{R}_k = \left\{ \alpha_1^n \geq 0 : p_k - 1 \leq \sum_{i=1}^{n} (2i - 1 + m - n)\alpha_i < p_{k+1} - 1 \right\}. \tag{6.12}$$

Let d_k be the distortion exponent over the region $\mathcal{R}_k$. Then $d_{\mathrm{UB-PC-}K}(b) = \sup \min(d_1, \ldots, d_K)$ where the supremum is over all parameters used to characterize the $\mathcal{R}_k$'s. First consider

$$d_K = \inf_{\alpha_1^n \in \mathcal{R}_K} \sum_{i=1}^{n} \left(b(p_K - \alpha_i)^+ + (2i - 1 + m - n)\alpha_i \right). \tag{6.13}$$

A *relaxed* version of (6.13) has a simple solution:

$$\inf_{\alpha_1^n \geq 0} \sum_{i=1}^{n} \left[b(p_K - \alpha_i)^+ + (2i - 1 + m - n)\alpha_i \right] = p_K d_{\mathrm{NPC-}\infty}(b). \tag{6.14}$$

Recall that $d_{\mathrm{NPC-}\infty}(b)$ is the distortion exponent for the ideal-CSIT case *without* power control. With $b > m - n + 1$, the optimizers α_i^* of (6.14) satisfy

$$\sum_{i=1}^{n} (2i - 1 + m - n)\alpha_i^* \geq (m - n + 1)p_K > p_K - 1$$

meaning that they belong to $\mathcal{R}_K$. Thus solutions of (6.13) and (6.14) coincide, and

$$d_K = p_K d_{\text{NPC}-\infty}(b).$$

Now consider the region $\mathcal{R}_1$, with

$$\begin{aligned} d_1 = \inf_{\alpha_1^n \geq 0} & \sum_{i=1}^{n} \left[b(1-\alpha_i)^+ + (2i-1+m-n)\alpha_i \right] \\ \text{s.t.} & \sum_{i=1}^{n} (2i-1+m-n)\alpha_i < p_2 - 1. \end{aligned} \tag{6.15}$$

Consider a relaxed version of (6.15)

$$\inf_{\alpha_1^n \geq 0} \sum_{i=1}^{n} \left[b(1-\alpha_i)^+ + (2i-1+m-n)\alpha_i \right], \tag{6.16}$$

with solution $\alpha_1^{\text{rlx}} = \cdots = \alpha_J^{\text{rlx}} = 1$, $\alpha_{J+1}^{\text{rlx}} = \cdots = \alpha_n^{\text{rlx}} = 0$ where

$$J \triangleq \max j \in \{1, \ldots, n\} \quad \text{s.t.} \quad b \geq 2j - 1 + m - n.$$

Note that $\sum_{i=1}^{n}(2i-1+m-n)\alpha_i^{\text{rlx}} = D(n-J, 1)$, i.e., the diversity gain corresponding to multiplexing gain $n - J$. Recall that we only consider $b \geq m - n + 1$ so that J is well defined.

If the minimum of (6.15) coincides with that of (6.16) then $d_1 = d_{\text{NPC}-\infty}(b) < d_K$. Thus $d_{\text{UB}-\text{PC}-K}(b) \leq d_{\text{NPC}-\infty}(b)$, leading to a contradiction, as $d_{\text{NPC}-\infty}(b)$ can be seen as $d_{\text{UB}-\text{PC}-K}(b)$ with *one* power level $K = 1$. The optimizers of (6.15) therefore must have the form $\alpha_1^* = \cdots = \alpha_{I-1}^* = 1, 0 \leq \alpha_I^* < 1, \alpha_{I+1}^* = \cdots = \alpha_n^* = 0$ for some $I \in \{1, 2, \ldots, J\}$ so that the constraint in (6.15) is active, i.e.,

$$\sum_{i=1}^{n} (2i-1+m-n)\alpha_i^* = p_2 - 1$$

Let $r_1 = \sum_{i=1}^{n}(1-\alpha_i^*)^+ = n - I + 1 - \alpha_I^* \in (n-J, n)$ then $d_1 = br_1 + D(r_1, 1)$ and $p_2 = D(r_1, 1) + 1$.

Repeating this argument over $\mathcal{R}_2, \ldots, \mathcal{R}_{K-1}$ and optimizing over $r_1, \ldots, r_{K-1}$ lead to the asserted result. □

Some Interpretations

Even though the bounding technique of Proposition 6.2 considers a system that uses K power levels and a continuum of multiplexing gains (rates), we are able to characterize the upper bound using only $K - 1$ variables $r_1, \ldots, r_{K-1}$. The variables r_k's can be interpreted as the multiplexing gains on the *boundaries* of the quantization regions in the feedback link. To see that, note that by definition of r_k, the power exponents p_k can be recursively characterized via r_k via the relation:

$$p_k = 1 + D(r_{k-1}, p_{k-1}) \tag{6.17}$$

with the convention $r_0 = 0$, $p_0 = 0$. Inserting these into the definition of the quantization region k (6.12), we have

$$\mathcal{R}_k = \left\{ \alpha_1^n \geq 0 : D(r_{k-1}, p_{k-1}) < \sum_{i=1}^{n} (2i - 1 + m - n)\alpha_i < D(r_k, p_k) \right\}.$$

Now consider the dual problem of the diversity gain problem

$$\inf_{\alpha_1^n \geq 0} \sum_{i=1}^{n} (p_k - \alpha_i)^+ \quad \text{s.t.} \quad \sum_{i=1}^{n} (2i - 1 + m - n)\alpha_i < D(r_k, p_k),$$

which has minimum r_k. We conclude that

$$\inf_{\alpha_1^n \in \mathcal{R}_k} \sum_{i=1}^{n} (p_k - \alpha_i)^+ = r_k.$$

That is, when applying a power in the order of SNR^{p_k} to any channel realization in $\mathcal{R}_k$, a multiplexing gain of at least r_k can be achieved.

Similarly, by considering

$$\sup_{\alpha_1^n \geq 0} \sum_{i=1}^{n} (p_{k-1} - \alpha_i)^+ \quad \text{s.t.} \quad \sum_{i=1}^{n} (2i - 1 + m - n)\alpha_i > D(r_{k-1}, p_{k-1})$$

we have

$$\sup_{\alpha_1^n \in \mathcal{R}_k} \sum_{i=1}^{n} (p_{k-1} - \alpha_i)^+ = r_{k-1}.$$

That is, all channel realizations in $\mathcal{R}_k$, when excited by a power of order $\text{SNR}^{p_{k-1}}$, cannot support the multiplexing gain r_{k-1} (resulting in an outage event with respect to power $\text{SNR}^{p_{k-1}}$ and rate $r_{k-1} \log \text{SNR}$). This is consistent with (6.17) because the probability that the channel realization belongs to $\mathcal{R}_k$ is given by the outage probability

$$\Pr(\alpha_1^n \in \mathcal{R}_k) \doteq \text{SNR}^{-D(r_{k-1}, p_{k-1})},$$

and thus the power exponent

$$p_k = 1 + D(r_{k-1}, p_{k-1})$$

can be applied over $\mathcal{R}_k$ without violating the long-term power constraint.

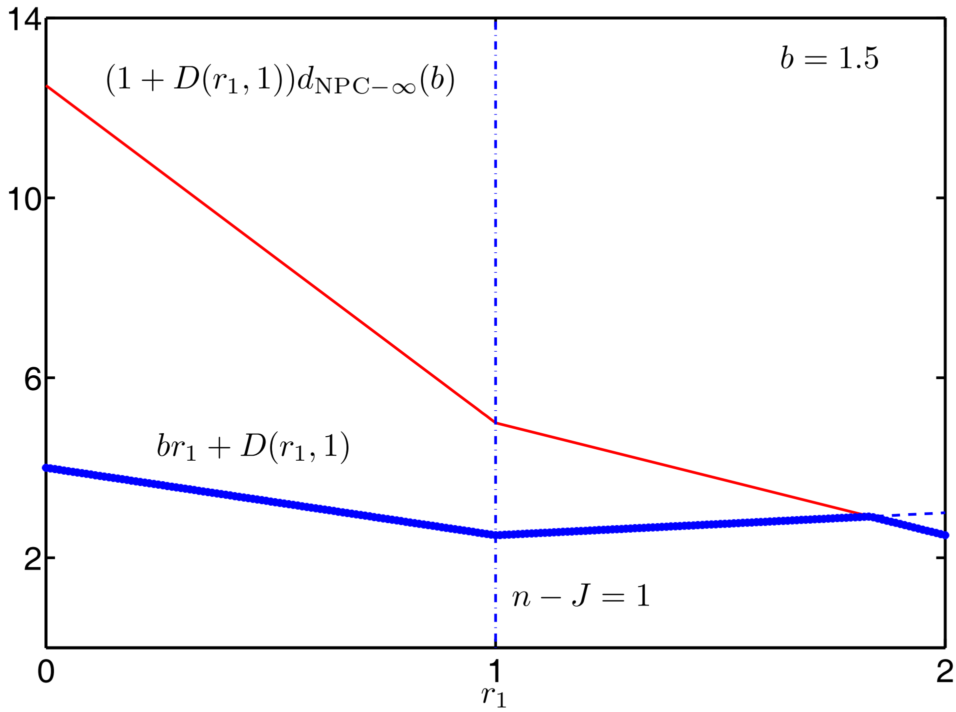

Figure 6.6: Computation of the bound $D_{\text{UB-PC}-K}(b)$ over a 2×2 channel, feedback resolution $K = 2$. Case 1: $J = 1$.

Computation of $d_{\text{UB-PC}-K}(b)$

Since $d_{\text{UB-PC}-K}(b)$ is an upper bound, finding the *global* optimum of the maximin in Proposition 6.2 is a critical task. In this section, we discuss some difficulties when computing this bound and give a conjecture on the optimal solution.

We first illustrate the K-level upper bound by studying a specific example. Consider a MIMO 2×2 channel with feedback resolution $K = 2$. By definition we have

$$d_{\text{UB-PC}-2} = \sup_{r_1 \in [n-J, n)} \min \{(1 + D(r_1, 1))d_{\text{NPC}-\infty}(b), br_1 + D(r_1, 1))\} \tag{6.18}$$

for $b \geq m - n + 1 = 1$, where

$$D(r_1, 1) = \begin{cases} 4 - 3r_1 & \text{if } r_1 \in (0, 1), \\ 2 - r_1 & \text{if } r_1 \in [1, 2). \end{cases}$$

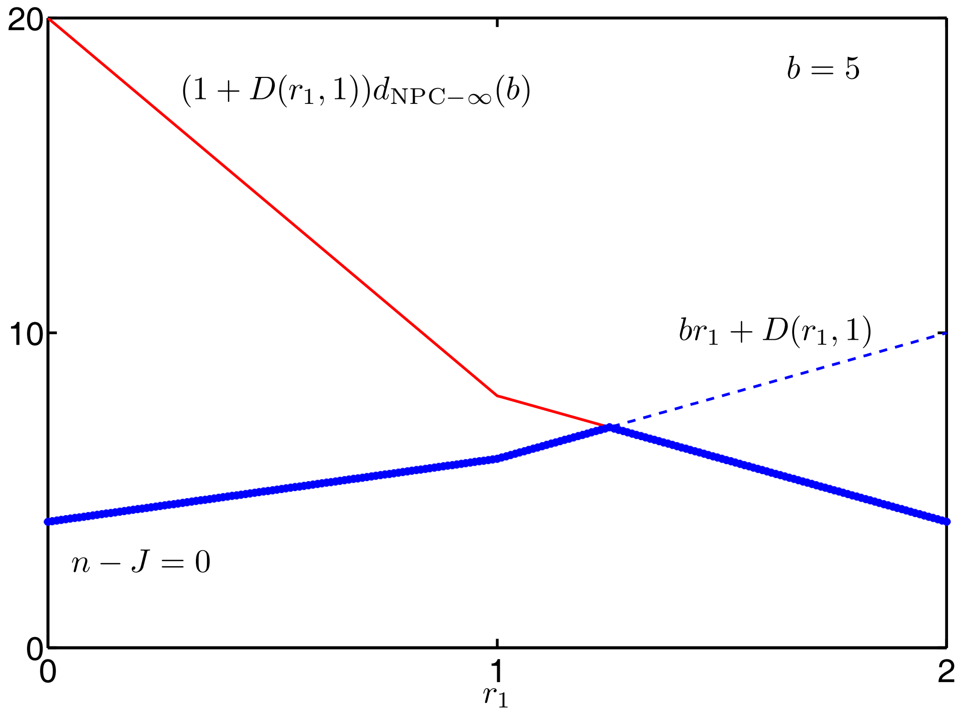

Figure 6.7: Computation of the bound $D_{\text{UB-PC}-K}(b)$ over a 2×2 channel, feedback resolution $K = 2$. Case 2: $J = 2$.

and the upper bound without power control is given by [CN05]

$$d_{\text{NPC}-\infty}(b) = \begin{cases} 2b & \text{if } b \in [0, 1), \\ b + 1 & \text{if } b \in [1, 3), \\ 4 & \text{otherwise.} \end{cases}$$

We plot in Figs. 6.6 and 6.7 the two component functions in (6.18) and their pointwise minimum. The problem is clearly a *nonconvex optimization*. The parameter J is b-dependent, in particular $J = 1$ for $b < 3$ and $J = 2$ for $b \geq 3$.

For $J = 1$, due to the constraint $r_1 > n - J = 1$, the optimum is given by the intersection between the two linear segments (the intersection always exists), leading to $d_{\text{UB-PC}-K}(b) = \frac{(3b-1)(b+1)}{2b}$. The constraint $r_1 > n - J = 1$ essentially removes all the segments of the piecewise linear function $br_1 + D(r_1, 1)$ that have a negative slope.

For $J = 2$ or $b \geq 3$, $br + D(r_1, 1)$ is nondecreasing in b, and the optimum is again given by the intersection of two piecewise linear functions. Because of the piecewise linearity, we need to distinguish between the case when the optimizer r_1^*

is in $(0,1]$ and when it is in $(1,2)$. We eventually obtain

$$d_{\text{UB-PC}-K}(b) = \begin{cases} \frac{(3b-1)(b+1)}{2b} & \text{if } b \in [1,3), \\ \frac{12b-4}{b+3} & \text{if } b \in [3,7), \\ \frac{20b-12}{b+9} & \text{if } b \in [7,\infty). \end{cases}$$

Unfortunately, computing $d_{\text{UB-PC}-K}(b)$ for a higher feedback resolution K is much more complicated. The component functions in the maximin are generally not concave, thus their point-wise minimum is not a concave function. Interestingly, in all the cases that we observed, equating all K components always gives a unique solution in $(n-J,n)^K$, and this intersection yields the largest value of the point-wise minimum (as confirmed by a grid search). This suggests that the *global* optimum may indeed be achieved by the intersection of all K component functions. However we have not been able to prove this conjecture analytically.

The computation of other achievable bounds in this chapter also gives rise to similar difficulties. However, finding the global one for an achievable bound of the maximin form is less critical because any local optimum still serves as a lower bound.

6.B Derivation of K-level Upper Bound and Achievable Distortion Exponent for SISO Channels

Partial-CSIT Upper Bound

We need to prove the equality (6.9). For $m = n = 1$, we have $d_{\text{NPC}-\infty}(b) = 1$, $\forall b \geq 1$. Furthermore, from the diversity-multiplexing results in Chapter 3 we have

$$\begin{aligned} D_1 &= D(r_1) = 1 - r_1, \\ D_2 &= D(r_2, 1 + D_1) = 2 - r_1 - r_2, \\ &\cdots \\ D_{K-1} &= K - 1 - r_1 - \ldots - r_{K-1}. \end{aligned}$$

Applying these results to Proposition 6.2 gives

$$\begin{aligned} d_{\text{UB}-\text{PC}-K}(b) = \sup_{0<r_1,\ldots,r_{K-1}<1} \min \big(& 1 + K - 1 - r_1 - \ldots - r_{K-1}, \\ & br_1 + 1 - r_1, \\ & br_2 + 2 - r_1 - r_2, \\ & \cdots \\ & br_{K-1} + K - 1 - r_1 - \ldots - r_{K-1} \big) . \end{aligned}$$

The point-wise minimum of a family of affine functions is concave [BV04]. The *global* optimum of the maximin in this particular case is obtained when all the exponents are equal, provided that the solution of such a system of equations exists and is strictly inside $(0,1)^{K-1}$. Balancing the K terms inside the brackets yields

$$\begin{aligned} r_{K-1} &= \frac{1}{b}, \\ r_{K-2} &= \frac{1}{b} + \frac{b-1}{b} r_{K-1} = \frac{1}{b} + \frac{b-1}{b^2}, \\ &\cdots \\ r_1 &= \frac{1}{b} + \frac{b-1}{b^2} + \ldots + \frac{(b-1)^{K-2}}{b^{K-1}} = 1 - \left(\frac{b-1}{b}\right)^{K-1}. \end{aligned}$$

We indeed have $r_k \in (0,1), \forall k$. Finally, for a SISO channel with $b \geq 1$ we have

$$d_{\mathrm{UB-PC-}K}(b) = 1 + (b-1)\left[1 - \left(\frac{b-1}{b}\right)^{K-1}\right] = b - b\left(1 - \frac{1}{b}\right)^{K}.$$

□

Achievable Distortion Exponent

In this section, we show that the exponent (6.11) is achievable for the SISO case. Consider a simple source-channel coding system with optimized rate allocation where the transmitter allocates a fixed rate $r_i \log \mathrm{SNR}$ for all $\alpha_1 \in \mathcal{R}_i$ where r_i the *maximum* multiplexing gain that *all* channel realizations in $\mathcal{R}_i$ can support, i.e.

$$r_{i+1} \log \mathrm{SNR} > \log\left(1 + \mathrm{SNR}^{p_i - \alpha_1}\right) \geq r_i \log \mathrm{SNR}, \forall \alpha_1 \in \mathcal{R}_i.$$

For such a system, we can achieve a distortion exponent of

$$\min(d_1, d_2, \ldots, d_K, d_{\mathrm{out}})$$

where d_i is the distortion exponent over $\mathcal{R}_i$ and d_{out} is the outage exponent. Herein outage is the event

$$\log(1 + \mathrm{SNR}^{p_K - \alpha_1}) < r_K \log \mathrm{SNR}.$$

We then have $p_1 = 1$, $\Pr(\alpha_1 \in \mathcal{R}_1) \doteq \mathrm{SNR}^0$, and $\Pr(\alpha_1 \notin \mathcal{R}_1) \doteq \mathrm{SNR}^{-D(r_1,1)} = \mathrm{SNR}^{-(1-r_1)}$. The average exponent over $\mathcal{R}_1$ has the same exponent as

$$\mathrm{SNR}^{-br_1} \Pr(\alpha_1 \in \mathcal{R}_1) \doteq \mathrm{SNR}^{-br_1}.$$

Hence $d_1 = br_1$. Then the power level in $\mathcal{R}_2$ is in the order of

$$\mathrm{SNR}^{p_2} \doteq \frac{\mathrm{SNR}}{\Pr(\alpha_1 \in \mathcal{R}_2)} \doteq \mathrm{SNR}^{2-r_1}$$

and the average exponent over $\mathcal{R}_2$ has the same exponent as

$$\mathrm{SNR}^{-br_2} \Pr(\alpha_1 \in \mathcal{R}_2) \doteq \mathrm{SNR}^{-br_2 - D(r_1,1)} = \mathrm{SNR}^{-br_2-(1-r_1)}.$$

This gives $d_2 = br_2 + 1 - r_1$. Repeating the arguments yields

$$\begin{aligned} d_3 &= br_3 + (1 + 1 - r_1) - r_2 = br_3 + 2 - r_1 - r_2, \\ &\cdots \\ d_K &= br_K + K - 1 - r_1 - \cdots - r_{K-1}. \end{aligned}$$

Finally the outage exponent is given by

$$d_{\text{out}} = D(r_K, p_K) = K - r_1 - \cdots - r_K.$$

Imposing $d_1 = \cdots = d_K = d_{\text{out}}$ gives

$$\begin{aligned} r_K &= \frac{1}{b+1} \\ r_{K-1} &= \frac{1}{b+1} + \frac{b}{b+1} r_K = \frac{1}{b+1} + \frac{b}{(b+1)^2} \\ &\cdots \\ r_1 &= \frac{1}{b+1} + \frac{b}{(b+1)^2} + \cdots + \frac{b^{K-1}}{(b+1)^K} = 1 - \left(\frac{b}{b+1}\right)^K. \end{aligned}$$

Finally,

$$d_{\text{SL-PC-}K} = b - \frac{b^{K+1}}{(b+1)^K},$$

which is (6.11) that we want to prove. □

6.C Proof of Proposition 6.3

We need the following lemma.

Lemma 6.2 (Linear lower bounds on the D-M curves)**.** *Define* $D_k(r) \triangleq D(r, 1 + D_{k-1}(r))$ *where* $D_0(r) \triangleq 0, \forall r$*, then*

$$D_K(r) \geq \left[(m - n + 1) \sum_{k=0}^{K-1} [(m - n + 1)n]^k\right] (n - r).$$

Proof. The lower bound is obtained by extending the last (rightmost) segment of the D-M curve $D_k(r)$ to the entire $(0, n)$. We claim that the last segment of $D_k(r)$ is given by

$$(m - n + 1) \sum_{k=0}^{L-1} [(m - n + 1)n]^k (n - r)$$

and show that by induction.

For $k = 1$ the claim holds. Assume the claim holds for $k = l \geq 1$, we will show that it also holds for $l+1$. For simplicity, first denote $\kappa = (m-n+1)n$. Note that for $r \in [(n-1)p, np)$ we have $D(r,p) = (m-n+1)(np-r)$. Thus, the last segment of $D_{l+1}(r)$ is given by

$$\begin{aligned}(m-n+1)(n(1+D_l(r))-r) &= \kappa + \kappa(m-n+1)\sum_{k=0}^{l-1}\kappa^k(n-r) - (m-n+1)r \\ &= \kappa\sum_{k=0}^{l}\kappa^k - \sum_{k=0}^{l}\kappa^k(m-n+1)r \\ &= \left[(m-n+1)\sum_{k=0}^{l}\kappa^k\right](n-r).\end{aligned}$$

□

It now remains to show that $D_k(r)$ (which is piecewise linear) is convex so that the extension of the last segment serves as a lower bound. To that end, we will show that the *negative slope* of $D_k(r)$ is non-increasing in r.

To see that, recall from [ZT03] an important property of $D(r,p)$: If we let $I = \lfloor \frac{r}{p} \rfloor$ then the negative slope of $D(r,p)$ at r is $(m-I)(n-I)-(m-I-1)(n-I-1)$, independent of p. Note that $(m-I)(n-I)-(m-I-1)(n-I-1) = m+n-1-2I$ is a decreasing function of I.

Now for any k, let $I(x) = \frac{x}{1+D_{k-1}(x)}$. If $x < y$ then $D_{k-1}(x) > D_{k-1}(y)$, thus

$$I(x) = \frac{x}{1+D_{k-1}(x)} < I(y) = \frac{y}{1+D_{k-1}(y)}.$$

This means the negative slope of $D(x, 1+D_{k-1}(x))$ is decreasing in x, and thus the piece-wise linear function $D_k(x) = D(x, 1+D_{k-1}(x))$ is convex, $\forall k$.

We now prove Proposition 6.3. Recall that we assume $m > 1$ throughout.

Proof of Sufficiency

Assume

$$\lim_{b\to\infty} \frac{b}{[(m-n+1)n]^{K(b)}} = 0. \tag{6.19}$$

Again consider an optimized single-layer coding system. Note that the distortion exponent of such a system with $K(b)$ levels of feedback is given by

$$d_{\text{SL–PC–}K(b)}(b) = \sup_{0<r_1,\ldots,r_{K(b)}<n} \min\left(br_1, br_2 + D_1, \ldots, br_{K(b)} + D_{K(b)-1}, D_{K(b)}\right).$$

Choosing $r_1 = r_2 = \cdots = r_{K(b)} = n - \frac{Cb}{[(m-n+1)n]^{K(b)}}$ where C is a finite positive constant specified later, we obtain

$$\frac{d_{\text{SL–PC–}K(b)}}{b} \geq \min\left(n - \frac{Cb}{[(m-n+1)n]^{K(b)}}, \frac{D_{K(b)}(r_1)}{b}\right).$$

Using the lower bound in Lemma 6.2, we have

$$\frac{D_{K(b)}(r_1)}{b} \geq \frac{m-n+1}{b} \frac{[(m-n+1)n]^{K(b)}-1}{(m-n+1)n-1} \frac{Cb}{[(m-n+1)n]^{K(b)}}.$$

Note that $(m-n+1)n > 1$ since $m > 1$. Choosing

$$C = n\frac{(m-n+1)n-1}{(m-n+1)}$$

and taking the limit $b \to \infty$, we then obtain

$$\lim_{b\to\infty} \frac{d_{\text{SL-PC-}K(b)}}{b} \geq \min\left(n, n \lim_{b\to\infty} \frac{[(m-n+1)n]^{K(b)}-1}{(m-n+1)n]^{K(b)}}\right) = n,$$

given that (6.19) holds. □

Proof of Necessity

Assume there exists at least one scheme that is asymptotically efficient, then we must have $\eta = n$. From the definition of $d_{\text{UB-PC-}K(b)}$, for $b > m+n-1$ we have

$$\begin{aligned} d_{\text{UB-PC-}K(b)} &\leq \sup_{0<r_1,\dots,r_{K(b)}<n} mn(1+D_{K(b)-1}) \\ &= mn\left(1+mn+\ldots+(mn)^{K(b)-1}\right) \\ &= \frac{mn}{mn-1}\left((mn)^{K(b)}-1\right). \end{aligned}$$

Note that $mn > 1$ because $m > 1$. Thus

$$\begin{aligned} n &= \lim_{b\to\infty} \frac{d_{\text{UB-PC-}K(b)}}{b} \\ &\leq \frac{mn}{mn-1} \lim_{b\to\infty} \frac{(mn)^{K(b)}}{b}, \end{aligned}$$

leading to

$$\lim_{b\to\infty} \frac{b}{(mn)^{K(b)}} \leq \frac{m}{mn-1},$$

which is the asserted condition.

Note that in the $m = n = 1$ case, we can use exactly the same arguments but the results slightly change because $\sum_{k=1}^{K(b)}(mn)^k = K(b)$. In particular we obtain a sufficient condition

$$\lim_{b\to\infty} \frac{b}{K(b)} = 0$$

and a necessary condition

$$\lim_{b\to\infty} \frac{b}{K(b)} \leq 1.$$

This is redundant because it is known from a direct investigation in Section 6.3 that the *exact* necessary and sufficient condition is (6.10), i.e.,

$$\lim_{b\to\infty}\left(1-\frac{1}{b}\right)^{K(b)}=0.$$

□

6.D Proof of Proposition 6.4

Short-term Power Constraint

Before proceeding to the main proof, we note that when transmitting an i.i.d. zero-mean unit-variance Gaussian vector of length $N_\mathrm{t}\le N_\mathrm{r}$, the MMSE averaged over a subset $\mathcal{A}$ of channel matrices is given by

$$\begin{aligned}\int_{\mathcal{A}}\mathrm{MMSE}(\mathbf{H})f(\mathbf{H})d\mathbf{H} &\doteq \int_{\mathcal{A}}\mathrm{tr}\left(\mathbf{I}_{N_\mathrm{r}}+\mathrm{SNR}\mathbf{H}\mathbf{H}^\mathrm{H}\right)^{-1}f(\mathbf{H})d\mathbf{H}\\ &=\int_{\mathcal{A}}\sum_{i=1}^{n}\frac{1}{1+\mathrm{SNR}^{1-\alpha_i}}f(\alpha_1^n)d\alpha_1^n.\end{aligned}$$

The dominating terms then have a negative SNR exponent of

$$\inf_{\{\alpha_1^n\ge 0\}\cap\mathcal{A}}\left[(1-\alpha_1)^+ +\sum_{i=1}^{n}(2i-1+m-n)\alpha_i\right]. \tag{6.20}$$

This fact will be used later in the proof.

Let $n>r_1>\cdots>r_K>0$ be the multiplexing gains of the digital part and let $d_1,\ldots,d_K$ be the distortion exponents over the K corresponding quantization regions. Then the scheme can achieve a distortion exponent of

$$\min(d_\mathrm{out},d_1,\ldots,d_K),$$

where d_out is the outage exponent.

By definition, we have

$$d_\mathrm{out}=D(r_K,1).$$

We now compute d_1. Let $\sigma_{\mathrm{e}1}^2\doteq\mathrm{SNR}^{-d_{\mathrm{e}1}}$ be the variance of the quantization error (i.e., the difference between the source vector $\mathbf{s}$ and the output of the tandem encoder)

$$\mathrm{SNR}^{-d_1}\doteq\mathrm{SNR}^{-d_{\mathrm{e}1}}\mathrm{SNR}^{-\mathrm{mmse}_1}=\mathrm{SNR}^{-(d_{\mathrm{e}1}+\mathrm{mmse}_1)}.$$

Since the analog part occupies only a fraction of the block, the *effective* bandwidth ratio is

$$\frac{T-\frac{N_\mathrm{s}}{n}}{N_\mathrm{s}}=b-\frac{1}{n}$$

The resulting distortion exponent is obtained from the rate-distortion limit as follows. We have $\sigma_{\mathrm{e}1}^2 = \exp(-R_s)$, with the source coding rate R_s satisfying $R_{\mathrm{s}}N_{\mathrm{s}} = (T - N_{\mathrm{s}}/n)r_1 \log \mathrm{SNR}$. It follows that the variance of the error has an SNR exponent given by

$$d_{\mathrm{e}1} = \left(b - \frac{1}{n}\right) r_1.$$

Furthermore, solving (6.20) with the non-outage region

$$\mathcal{A} = \mathcal{R}_1 = \left\{ \alpha_1^n : \sum_{i=1}^{n} (1-\alpha_i)^+ \geq r_1 \right\},$$

we readily obtain the optimizers $\alpha_1^* = \cdots = \alpha_n^* = 0$ and thus

$$\mathrm{mmse}_1 = 1.$$

This leads to

$$d_1 = d_{\mathrm{e}1} + \mathrm{mmse}_1 = \left(b - \frac{1}{n}\right) r_1 + 1.$$

We next compute $d_2 = d_{\mathrm{e}2} + \mathrm{mmse}_2$. Again we have

$$d_{\mathrm{e}2} = \left(b - \frac{1}{n}\right) r_2.$$

To compute mmse_2, we solve the optimization (6.20) with

$$\mathcal{A} = \mathcal{R}_2 = \left\{ \alpha_1^n : r_2 \leq \sum_{i=1}^{n} (1-\alpha_i)^+ < r_1 \right\}.$$

With this $\mathcal{A}$, the optimizers α_i^* of (6.20) always coincide with those of the problem:

$$\inf_{\{\alpha_1^n \geq 0\} \cap \mathcal{A}} \sum_{i=1}^{n} (2i - 1 + m - n)\alpha_i.$$

But the above minimum is exactly the diversity gain $D(r_1, 1)$. This means that $\sum_{i=1}^n (1-\alpha_i^*)^+ = r_1$, and $\alpha_1^* = 1$ if $r_1 \leq n-1$ and $\alpha_1^* = n - r_1$ if $r_1 \geq n-1$, leading to

$$\mathrm{mmse}_2 = (r_1 + 1 - n)^+ + D(r_1, 1).$$

Finally

$$d_2 = \left(b - \frac{1}{n}\right) r_2 + D(r_1, 1) + (r_1 + 1 - n)^+.$$

Repeating the above steps for $d_3, \ldots, d_K$ and optimizing over the rate $r_1, \ldots, r_K$ gives the claimed result. □

Long-term Power Constraint

The proof follows exactly the same arguments as those of the short-term power constraint case. The main difference comes from the fact that with a long-term power constraint, a power in the order $P_k \doteq \frac{\text{SNR}}{\Pr(\mathbf{H}\in\mathcal{R}_k)}$ can be applied to the region $\mathcal{R}_k$.

Let $d_1, \dots, d_K$ be the distortion exponents corresponding to $\mathcal{R}_1, \dots, \mathcal{R}_K$. We still have

$$d_1 = d_{e1} + \text{mmse}_1 = \left(b - \frac{1}{n}\right) r_1 + 1.$$

As for

$$\mathcal{R}_2 = \left\{ \alpha_1^n : \sum_{i=1}^{n} (1-\alpha_i)^+ < r_1, \sum_{i=1}^{n} (p_2 - \alpha_i)^+ \geq r_2 \right\},$$

we have $d_{e2} = \left(b - \frac{1}{n}\right) r_2$. The power level applied to this region $P_2 \doteq \text{SNR}^{1+D(r_1,1)} = \text{SNR}^{1+D_1} \equiv \text{SNR}^{p_2}$ and thus

$$\begin{aligned}
\text{mmse}_2 &= \inf_{\{\alpha_1^n \geq 0\}\cap\mathcal{R}_2} \left[(1+D_1-\alpha_1)^+ + \sum_{i=1}^{n} (2i-1+m-n)\alpha_i \right] \\
&= D_1 + (r_1+1-n)^+ + D_1,
\end{aligned}$$

which eventually leads to

$$d_2 = \left(b - \frac{1}{n}\right) r_2 + 2D_1 + (r_1+1-n)^+.$$

At the (possibly locally) optimal rate allocation, impose $d_1 = d_2$ or

$$\left(b - \frac{1}{n}\right) r_1 + 1 = \left(b - \frac{1}{n}\right) r_2 + 2D_1 + (r_1+1-n)^+. \tag{6.21}$$

The function $2D_1 + (r_1+1-n)^+$ is monotonically non-increasing in r_1 thus $2D_1 + (r_1+1-n)^+ > 1$ for $0 < r_1 < n$. If $r_2 \geq r_1$ then (6.21) cannot be satisfied. Thus $r_2 < r_1$, and therefore $D_2 = D(r_2, 1+D_1) > D(r_1, 1+D_1) > D(r_1, 1) = D_1$.

Now consider

$$\mathcal{R}_3 = \left\{ \alpha_1^n : \sum_{i=1}^{n} (1+D_1-\alpha_i)^+ < r_2, \sum_{i=1}^{n} (1+D_2-\alpha_i)^+ \geq r_3 \right\},$$

where we have

$$\begin{aligned}
\text{mmse}_3 &= \inf_{\{\alpha_1^n \geq 0\}\cap\mathcal{R}_3} \left[(1+D_2-\alpha_1)^+ + \sum_{i=1}^{n} (2i-1+m-n)\alpha_i \right] \\
&= (1+D_2-(1+D_1)+(r_2-(n-1)(1+D_1))^+)^+ + D_2 \\
&= 2D_2 - D_1 + (r_2+(1-n)(1+D_1))^+,
\end{aligned}$$

where the second equality is due to the fact that the optimizer $\alpha_1^* = n(1+D_1) - r_2$ if $r_2 > (n-1)(1+D_1)$ and $\alpha_1^* = 1 + D_1$ otherwise, and the last equality is due to $D_1 < D_2$. This leads to

$$d_3 = \left(b - \frac{1}{n}\right) r_3 + 2D_2 - D_1 + (r_2 + (1-n)(1+D_1))^+.$$

Repeating the arguments leads to the asserted result. □

6.E Proof of Proposition 6.5

The proof follows the same pattern of the proof of Proposition 6.4. To avoid repetition, we only present the computation of a bound to the MMSE exponent, which is the key difference from Proposition 6.4.

To further simplify the presentation, consider the short-term power constraint, and a feedback resolution $K = 2$. Let $\hat{\mathbf{H}}$ be the $n \times n$ sub-matrix channel used for the transmission of the analog part. Let $\mu_i = \text{SNR}^{-\beta_i}$ be the corresponding eigenvalues. In the first region $\mathcal{R}_1$, $\text{MMSE}(\mathbf{H} \in \mathcal{R}_1) \doteq \text{SNR}^{-\text{mmse}_1}$ where

$$\begin{aligned}
\text{mmse}_1 &= \inf_{\beta_1^n \geq 0 \cap \mathcal{R}_1} (1-\beta_1)^+ + \sum_{i=1}^{n} (2i-1)\beta_i \\
&\geq \inf_{\beta_1^n \geq 0} (1-\beta_1)^+ + \sum_{i=1}^{n} (2i-1)\beta_i \\
&= 1.
\end{aligned}$$

This leads to

$$d_1 = \left(b - \frac{1}{n}\right) r_1 + 1.$$

Consider $\mathcal{R}_2 = \{\alpha_1^n : r_2 \leq \sum_{i=1}^n (1-\alpha_i)^+ < r_1\}$. The event that $\log\det(\mathbf{I}_{N_\text{r}} + \text{SNR}\mathbf{H}\mathbf{H}^\text{H}) < R$ implies that $\log\det(\mathbf{I}_{N_\text{r}} + \text{SNR}\hat{\mathbf{H}}\hat{\mathbf{H}}^\text{H}) < R$ due to the monotonicity of $\log\det(\cdot)$ on the semi-definite cone. Equivalently, the event $\alpha_1^n \in \mathcal{R}_2$ also implies $\sum_{i=1}^n (1-\beta_i)^+ < r_1$. Let

$$\hat{\mathcal{R}}_2 = \left\{\beta_1^n : \sum_{i=1}^{n} (1-\beta_i)^+ < r_1\right\} \supset \mathcal{R}_2,$$

then $\text{MMSE}(\mathbf{H} \in \mathcal{R}_2) \doteq \text{SNR}^{-\text{mmse}_2}$ where

$$\begin{aligned}
\text{mmse}_2 &= \inf_{\beta_1^n \geq 0 \cap \mathcal{R}_2} (1-\beta_1)^+ + \sum_{i=1}^{n} (2i-1)\beta_i \\
&\geq \inf_{\beta_1^n \geq 0 \cap \hat{\mathcal{R}}_2} (1-\beta_1)^+ + \sum_{i=1}^{n} (2i-1)\beta_i \\
&= (1 + r_1 - n)^+ + D^{nn}(r_1, 1).
\end{aligned}$$

This leads to

$$d_2 = \left(b - \frac{1}{n}\right) r_2 + D^{nn}(r_1, 1) + (1 + r_1 - n)^+.$$

$\square$

6.F Proof of Proposition 6.6

Consider a short-term power constraint and the following index mapping, which is characterized indirectly via $\alpha_1^n \geq 0$,

$$\mathcal{I}\left(\alpha_1^n\right) = \begin{cases} K & \text{if } \sum_{i=1}^n (1-\alpha_i)^+ - \frac{1}{bn}(p_K - \alpha_i)^+ < r_K, \\ k : r_{k-1} + \sum_{i=1}^n \frac{1}{bn}(p_{k-1} - \alpha_i)^+ > \sum_{i=1}^n (1-\alpha_i)^+ \\ \qquad \geq r_k + \sum_{i=1}^n \frac{1}{bn}(p_k - \alpha_i)^+ \quad \text{otherwise.} \end{cases}$$

We first compute the distortion exponent over the quantization region $\mathcal{R}_1$ corresponding to $\mathcal{I} = 1$

$$d_1 = br_1 + \text{mmse}_1.$$

Notice that there is no loss in bandwidth since the analog part is superimposed onto the digital codeword, resulting in the quantization error (at the output of the tandem encoder) of order SNR^{-br_1}. The analog part however can only use a power in the order of SNR^{p_1}, thus (similarly to (6.20))

$$\text{mmse}_1 = \inf_{(\alpha_1^n \geq 0) \cap \mathcal{R}_1} \left\{ (p_1 - \alpha_1)^+ + \sum_{i=1}^n (2i - 1 + m - n)\alpha_i \right\}$$

where

$$\mathcal{R}_1 = \left\{ \alpha_1^n : \sum_{i=1}^n \left[(1-\alpha_i)^+ - \frac{1}{bn}(p_1 - \alpha_i)^+ \right] \geq r_1 \right\}.$$

This optimization gives $\text{mmse}_1 = p_1$, $\forall r_1 \in \left(0, n - \frac{p_1}{b}\right)$ and thus $d_1 = br_1 + p_1$.

Now consider

$$\mathcal{R}_2 = \left\{ \alpha_1^n : \sum_{i=1}^n \left[(1-\alpha_i)^+ - \frac{1}{bn}(p_1 - \alpha_i)^+ \right] < r_1, \right.$$
$$\left. \sum_{i=1}^n \left[(1-\alpha_i)^+ - \frac{1}{bn}(p_2 - \alpha_i)^+ \right] \geq r_2 \right\}$$

and

$$\text{mmse}_2 = \inf_{(\alpha_1^n \geq 0) \cap \mathcal{R}_2} \left\{ (p_2 - \alpha_1)^+ + \sum_{i=1}^n (2i - 1 + m - n)\alpha_i \right\}.$$

Since the term $(p_2 - \alpha_1)^+$ can only reduce the weight corresponding to α_1, the optimizers α_i^*'s of the above optimization coincide with these of

$$\inf_{(\alpha_1^n \geq 0) \cap \mathcal{R}_2} \sum_{i=1}^{n} (2i - 1 + m - n)\alpha_i,$$

meaning that

$$\alpha_1^* = \begin{cases} 1 & \text{if } r_1 < (n-1)\left(1 - \frac{p_1}{bn}\right) \\ 1 - r_1 + (n-1)\left(1 - \frac{p_1}{bn}\right) & \text{if } (n-1)\left(1 - \frac{p_1}{bn}\right) \leq r_1 \\ & \qquad < (n-1)\left(1 - \frac{p_1}{bn}\right) + 1 - p_1, \\ \frac{n - \frac{p_1}{b} - r_1}{1 - \frac{1}{bn}} & \text{if } (n-1)\left(1 - \frac{p_1}{bn}\right) + 1 - p_1 \leq r_1 < n\left(1 - \frac{p_1}{bn}\right) \end{cases}$$

and that $\sum_{i=1}^{n}(2i - 1 + m - n)\alpha_i^* = D_{\text{SP}}(r_1, 1, p_1)$. This eventually leads to

$$\begin{aligned} \text{mmse}_2 = \max &\left(0, p_2 - 1 + r_1 - (n-1)\left(1 - \frac{1}{bn}\right), p_2 - \frac{n - \frac{p_1}{b} - r_1}{1 - \frac{1}{bn}}\right) \\ &+ D_{\text{SP}}(r_1, 1, p_1) \end{aligned}$$

and $d_2 = br_2 + \text{mmse}_2$.

Continuing this line of arguments, we obtain the achievable distortion exponent of Proposition 6.6 under a short-term power constraint. The exponent for the long-term power constraint case can be derived in a completely similar manner. □

Chapter 7

Distortion Exponent over Relay Channels

This chapter continues on the distortion exponent problem, with a new focus on the relay channels with limited feedback. Building upon results from Chapters 4 and 6, we show that under a short-term power constraint, combining a simple feedback scheme with separate source and channel coding outperforms the best known no-feedback strategies even with only a few bits of feedback information. Partial power control is shown to be instrumental in achieving a very fast decaying average distortion, especially in the regime of high bandwidth ratios. Performance limitation due to the lack of full CSI at the destination feedback quantizer is also investigated, where the degradation in terms of the distortion exponent is shown to be significant. However, even in such restrictive scenarios, using partial feedback still yields distortion exponents superior to any no-feedback schemes.

7.1 Introduction

Despite the large amount of work in the literature addressing various performance measures in cooperative systems, many challenging problems remain open. One particular scenario is the transmission of a source over a slow fading channel. In this problem the separation theorem [CT91] does not hold when the channel state information is not fully known at the transmitter. Indeed, even if the transmitter in a relay channel knows the full CSI, the optimal strategy for source transmission is still unknown, as the capacity of a completely general relay channel is a long standing open problem.

In the context of source–channel coding over relay channels, various protocols are proposed and their achievable distortion exponents are analyzed in [GE07b] under the assumption that the source and the relay do not have any knowledge of their corresponding forward channel gains. However, for many practical scenarios, limited channel state feedback is present at the transmitter side, allowing for partial

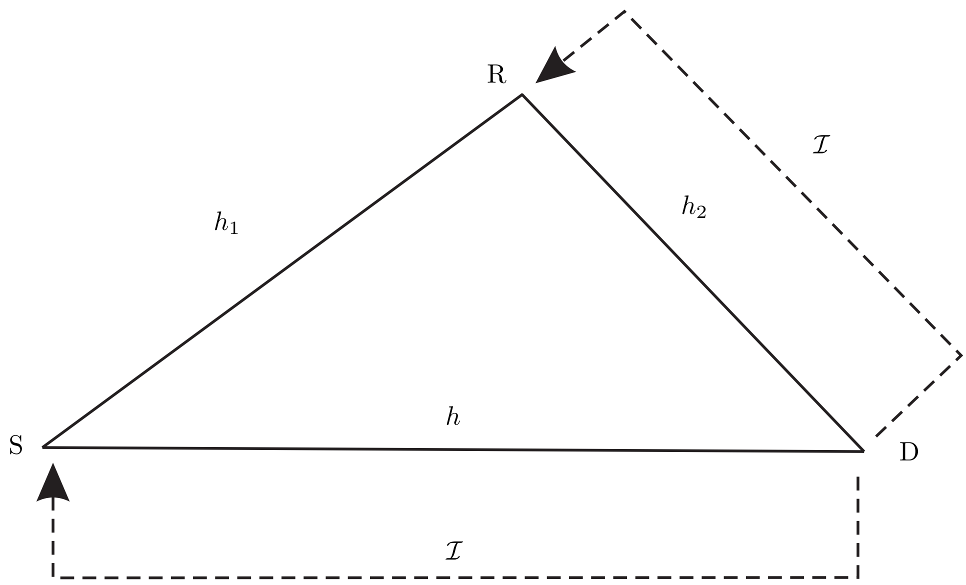

Figure 7.1: System model. Note that in the feedback model of Section 7.3, we have $\mathcal{I} = \mathcal{I}(h, h_1, h_2)$. On the other hand, the feedback index in Section 7.4 does not depend on h_1, i.e., $\mathcal{I} = \mathcal{I}(h, h_2)$

.

rate and/or power adaptation.

This chapter studies the asymptotic performance of the average end-to-end distortion over a three-node single-antenna relay channel *in the presence of partial CSIT*. We derive upper bounds on the optimal distortion exponent of any possible relaying and feedback strategy, given the number of feedback levels. For the achievability part, we exclusively focus on the decode-and-forward strategy [LTW04]. We show that even with the separation of source and channel coding, partial rate control allows for an improved performance over the best layering schemes in [GE07b]. Thus the separation of source–channel coding provides an effective yet simple solution in practical systems, even if some optimality may be lost. Temporal power control with limited CSIT provides additional gains which cannot be obtained by any no-feedback system. Finally, upper bounds and achievable bounds for the case of channel states partially known at the destination are derived. Even in such a restrictive scenario, using partial rate and power control still provides superior gains, especially in the low spectral efficiency regime.

7.2 System Model

Consider the complex baseband model of a frequency-nonselective half-duplex fading relay channel, depicted in Fig. 7.1. The three single-antenna nodes in the systems are the source, the relay, and the destination. The channel is assumed to be slowly fading, i.e., the channel gains are constant during a fading block consisting of T channel uses, but changes independently from one block to the next. We exclusively consider the case when a transmission codeword spans a single fading block to study the gain from spatial cooperation. The channel is assumed to be statistically symmetric. In particular, h, h_1, and h_2 are i.i.d. complex Gaussian random variables with zero mean and unit variance. That is, we consider a quasi-static Rayleigh fading scenario. Assume perfect CSI at the receiver of each communication link in the system, and perfect synchronization.

We consider *individual* power constraints at the source and at the relay. Since we deal with partial-CSIT systems, both short-term and long-term power constraints are considered, as in Chapter 4.

At the source, an i.i.d. memoryless Gaussian source produces N_s complex symbols $\mathbf{s}$ every T time instant. The source symbols are assumed to have zero mean and unit variance. Let $b = \frac{N_s}{T}$ be the channel bandwidth to source bandwidth ratio (bandwidth ratio).

At each SNR, let the mean squared error between the source vector and the reconstructed vector be

$$\bar{\Delta} = \frac{1}{N_s} \mathrm{E}\left[\|\mathbf{s} - \hat{\mathbf{s}}\|_2^2\right]$$

where the expectation is over the randomness of the source symbols, the noise and the channel gains. We consider a sequence of schemes at increasingly high SNR, and obtain a sequence of $\bar{\Delta}$ indexed by SNR. Completely similarly to Chapter 6, the system is then said to achieve a *distortion exponent* of d if

$$\bar{\Delta} \doteq \mathrm{SNR}^{-d}.$$

We are interested in characterizing the distortion exponent d as a function of the bandwidth ratio b. This provides a coarse (in the limit of asymptotically high SNR) tradeoff between spectral efficiency (via b) and distortion (represented by d), giving useful insight into the performance of different source transmission schemes.

7.3 Distortion Exponents with Channel State Feedback

In this section we consider the scenario where the channel states h, h_1 and h_2 are *fully known* at the destination. The destination then maps the tuple of channel states into an integer index $\mathcal{I}(h, h_1, h_2) \in \{1, \ldots, K\}$ and broadcasts this index to both the source and the relay via a noiseless zero-delay dedicated feedback link (cf. Fig. 7.1).

Upper Bounds to the Distortion Exponent

We begin by deriving relevant upper bounds to the distortion exponent that can be achieved by any source–channel coding scheme over the relay channel of interest.

First we have the cooperative transmitters upper bound, obtained by letting the source and the relay fully cooperate resulting in a MISO 2×1 channel:

$$d^{\mathrm{UB}}_{\mathrm{MISO\text{–}NPC}}(b) = \min(b, 2)$$

under a short-term power constraint; and

$$d^{\mathrm{UB}}_{\mathrm{MISO\text{–}PC}}(b) = b$$

under a long-term power constraint. Recall that the proof for the no power control case can be found in [GE08, CN07], and the power control case is presented in Chapter 6. Both bounds assume perfect CSIT. Note that for $b < 1$, even using a simple HDA scheme in [CN07] over a SISO channel can achieve the MISO upper bound. The regime of $b < 1$ is therefore not of our interest.

For the power control case, in Chapter 6 an upper bound for the MISO channel with K power levels is given. For $b < 2$, the upper bound for a MISO channel with K levels of feedback trivially coincides with that of the perfect-CSIT case, i.e. $d^{\mathrm{UB\text{–}}K}_{\mathrm{MISO\text{–}PC}}(b) = b$, $\forall K$. Otherwise, we have:

Proposition 7.1. *For $b \geq 2$, the distortion exponent of a 2×1 MISO channel with K levels of feedback is upper-bounded by*

$$d^{UB\text{–}K}_{MISO\text{–}PC}(b) = \frac{4b - 2(b-2)\left(\frac{b-2}{2b}\right)^{K-1}}{4 + (b-2)\left(\frac{b-2}{2b}\right)^{K-1}}.$$

Note that the upper bound $d^{\mathrm{UB\text{–}}K}_{\mathrm{MISO\text{–}PC}}(b)$ is given in the form of a general maximin problem in Chapter 6. To obtain the explicit form in Proposition 7.1, the proof in Appendix 7.A deviates from that in Chapter 6, expressing the upper bound in the form of an optimization problem with respect to the *power levels* instead of the *multiplexing gains.*

Rewriting

$$d^{\mathrm{UB\text{–}}K}_{\mathrm{MISO\text{–}PC}}(b) = \frac{2^{K+1}b^K - 2(b-2)^K}{2^{K+1}b^{K-1} + (b-2)^K}$$

we readily obtain

Corollary 7.1.

$$\lim_{b\to\infty} d^{UB\text{–}K}_{MISO\text{–}PC}(b) = 2^{K+1} - 2 = \sum_{k=1}^{K} 2^k.$$

This is consistent with the limiting results in multiple-antenna channels, obtained by an indirect approach in Chapter 6. In short, $d^{\mathrm{UB}-K}_{\mathrm{MISO-PC}}(b)$ equals to the maximal diversity gain of the 2×1 channel with K feedback levels, because at low spectral efficiency, even if the multiplexing gain can be made close to zero, the average distortion of the system is still *outage-limited*.

Achievable Distortion Exponents without Power Control

In this section, we derive certain achievable bounds to the optimal distortion exponent over relay channels with feedback. We exclusively study decode–forward relaying [LTW04] as this strategy naturally fits into our quantized feedback framework.

In short, the DF protocol that we consider consists of two communication phases. Phase 1 uses βT channel uses with $\beta \in (0, 1)$. The source in Phase 1 encodes a message m to a codeword of length βT and transmits. The relay attempts to decode m based on its received signal in Phase 1. In Phase 2, if the source–relay link is not in outage the relay re-encodes m and transmits. Otherwise the relay outputs nothing. The source may transmit $(1 - \beta)T$ additional symbols in Phase 2, in which case we have a non-orthogonal scheme; or it may remain silent (in orthogonal schemes). The destination decodes based on the received signals in both phases.

Notice that for a given multiplexing gain, the numbers of channel uses allocated to Phase 1 and Phase 2 in DF relaying can be optimized over (which we often refer to as *dimension allocation*) so as to maximize the diversity gain. The diversity gain of DF relaying with optimized dimension allocation is given in the following lemma. This is a natural extension of Proposition 4.1, which deals with the no-power control case (i.e., when the transmit power is SNR^1, to the power control case). The proof is relatively similar to that of Proposition 4.1, and thus omitted.

Lemma 7.1. *Let the transmit power at both the source and the relay be* SNR^p *where* $p \geq 1$*, then the outage exponent of DF relaying with optimized dimension allocation is*

$$D_O^{DF}(r,p) = \begin{cases} 2p - 3r & \text{if } r < \frac{p}{3}, \\ \frac{2p(p-r)}{p+r} & \text{otherwise.} \end{cases}$$

for orthogonal schemes, and

$$D_{NO}^{DF}(r,p) = \begin{cases} 2p - \frac{3+\sqrt{5}}{2}r & \text{if } r < \frac{3-\sqrt{5}}{2}p, \\ \frac{(2p-r)(p-r)}{p} & \text{otherwise} \end{cases}$$

for non-orthogonal schemes. The optimal allocation of the available channel uses is

$$\beta^*(r,p) = \begin{cases} \frac{2}{3} & \text{if } r < \frac{p}{3}, \\ \frac{p+r}{2p} & \text{otherwise} \end{cases}$$

in the orthogonal case, and

$$\beta^*(r,p) = \begin{cases} \frac{\sqrt{5}-1}{2} & \textit{if } r < \frac{3-\sqrt{5}}{2}p, \\ \frac{p}{2p-r} & \textit{otherwise} \end{cases}$$

in the non-orthogonal case.

For convenience, we denote the diversity gain in the no power control case as $D_{\mathrm{O}}^{\mathrm{DF}}(r) \triangleq D_{\mathrm{O}}^{\mathrm{DF}}(r,1)$ and $D_{\mathrm{NO}}^{\mathrm{DF}}(r) \triangleq D_{\mathrm{NO}}^{\mathrm{DF}}(r,1)$. Similarly $\beta^*(r) \triangleq \beta^*(r,1)$.

The proposed feedback scheme works as follows. For brevity we only describe the orthogonal case, as the non-orthogonal scheme is based on a similar idea.

The system is equipped with a library of K tandem encoders [MP02], i.e. K pairs of source encoder and channel encoder, with channel code rates $\{r_i \log \mathrm{SNR}\}$ bits per channel use, where $1 > r_1 > \cdots > r_K > 0$. Herein K is the number of feedback levels, also known as the *feedback resolution*. Upon receiving the index $\mathcal{I}(h,h_1,h_2) = i$ fed back from the destination, the source node encodes the source symbols with the tandem code whose channel code rate is $r_i \log \mathrm{SNR}$. Given $\mathcal{I} = i$, orthogonal DF relaying with optimal allocation $\beta^*(r_i)$ will be used.

Let $\mu_{\mathrm{O}}(h,h_1,h_2;r)$ be the mutual information of the orthogonal scheme given the channel states h, h_1, h_2 and a certain multiplexing gain r, i.e.

$$\mu_{\mathrm{O}}(h,h_1,h_2;r) \triangleq \begin{cases} \beta^*(r)\log\left(1+|h|^2\mathrm{SNR}\right) & \text{if } \log\left(1+|h_1|^2\mathrm{SNR}\right) < \frac{r\log\mathrm{SNR}}{\beta^*(r)}, \\ \beta^*(r)\log\left(1+|h|^2\mathrm{SNR}\right) + (1-\beta^*(r))\log\left(1+|h_2|^2\mathrm{SNR}\right), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise.} \end{cases}$$

The destination employs the index mapping

$$\mathcal{I}(h,h_1,h_2) = \begin{cases} K & \text{if } \mu_{\mathrm{O}}(h,h_1,h_2;r_K) < r_K \log\mathrm{SNR}, \\ \max i \in \{1,\ldots,K\} \text{ s.t. } \mu_{\mathrm{O}}(h,h_1,h_2;r_i) \geq r_i \log\mathrm{SNR} & \text{otherwise.} \end{cases} \tag{7.1}$$

That is, the destination informs the source node to use the largest channel code rate possible in the library of codes so that the transmission will not be in outage. This is equivalent to using the quantizer with the highest resolution possible to encode the source. In case the channel is in a too bad condition and no reliable communication is possible with any rate in the library, we can send back an arbitrary index (which is set to K in (7.1)) without changing the results.

We are now ready to state the following achievable distortion exponents. The proof is deferred to Appendix 7.B.

Proposition 7.2. *Under a short-term power constraint at both the source and the relay, the proposed orthogonal scheme can achieve the distortion exponent*

$$d_{O\text{-}NPC}^{K}(b) = \sup_{1>r_1>\cdots>r_K>0} \min\left(D_O^{DF}(r_K), br_1, \ldots, br_K + D_O^{DF}(r_{K-1})\right).$$

The non-orthogonal scheme can achieve

$$d^K_{NO\text{–}NPC}(b) = \sup_{1>r_1>\cdots>r_K>0} \min\left(D^{DF}_{NO}(r_K), br_1, \ldots, br_K + D^{DF}_{NO}(r_{K-1})\right).$$

Unfortunately, the optimization problems in Proposition 7.2 are not convex, because $D^{\mathrm{DF}}_{\mathrm{O}}(r)$ and $D^{\mathrm{DF}}_{\mathrm{NO}}(r)$ are *not* concave functions. However, for sufficiently high b, these maximin problems can be reduced to convex ones. In such cases, we can find the closed-form expression of the distortion exponents in Proposition 7.2.

Corollary 7.2. *For $b \geq 6$*

$$d^K_{O\text{–}NPC}(b) = 2\frac{\left(\frac{b}{3}\right)^{K+1} - \frac{b}{3}}{\left(\frac{b}{3}\right)^{K+1} - 1}.$$

For $b \geq 3 + \sqrt{5}$

$$d^K_{NO\text{–}NPC}(b) = 2\frac{\left(\frac{3-\sqrt{5}}{2}b\right)^{K+1} - \frac{3-\sqrt{5}}{2}b}{\left(\frac{3-\sqrt{5}}{2}b\right)^{K+1} - 1}.$$

Proof. We need the following lemma. The proof is straightforward and thus omitted.

Lemma 7.2. *Let $(x^*_1, \ldots, x^*_K)$ be the solutions to the system of linear equations*

$$bx_1 = bx_2 + \gamma - \theta x_1 = \cdots = bx_K + \gamma - \theta x_{K-1} = \gamma - \theta x_K$$

where $\gamma > 0, \theta > 0$ are real parameters then

$$\Delta \triangleq bx^*_1 = \gamma\frac{\left(\frac{b}{\theta}\right)^{K+1} - \frac{b}{\theta}}{\left(\frac{b}{\theta}\right)^{K+1} - 1}$$

and

$$x^*_k = \frac{\Delta - \gamma}{b}\frac{1 - \left(\frac{\theta}{b}\right)^K}{1 - \frac{\theta}{b}} + \left(\frac{\theta}{b}\right)^{K-1}\frac{\gamma}{b}.$$

We now prove Corollary 7.2. Consider only the orthogonal case. Since $d^K_{\mathrm{O\text{–}NPC}}(b) \leq 2$, it suffices to consider r_1 such that $br_1 \leq 2$ or $r_1 \leq \frac{2}{b}$. But $\frac{2}{b} < \frac{1}{3}$, $\forall b > 6$, thus $r_K < \cdots < r_1 < \frac{1}{3}$ and $D^{\mathrm{DF}}_{\mathrm{O}}(r_k) = 2 - 3r_k$, $\forall k$. In this case, the maximin problem finding $d^K_{\mathrm{O\text{–}NPC}}(b)$ is of the form

$$\sup_{\frac{2}{b}>r_1>\cdots>r_K>0} \min\left\{2 - 3r_K, br_1, br_2 + 2 - 3r_1, \ldots, br_K + 2 - 3r_{K-1}\right\}.$$

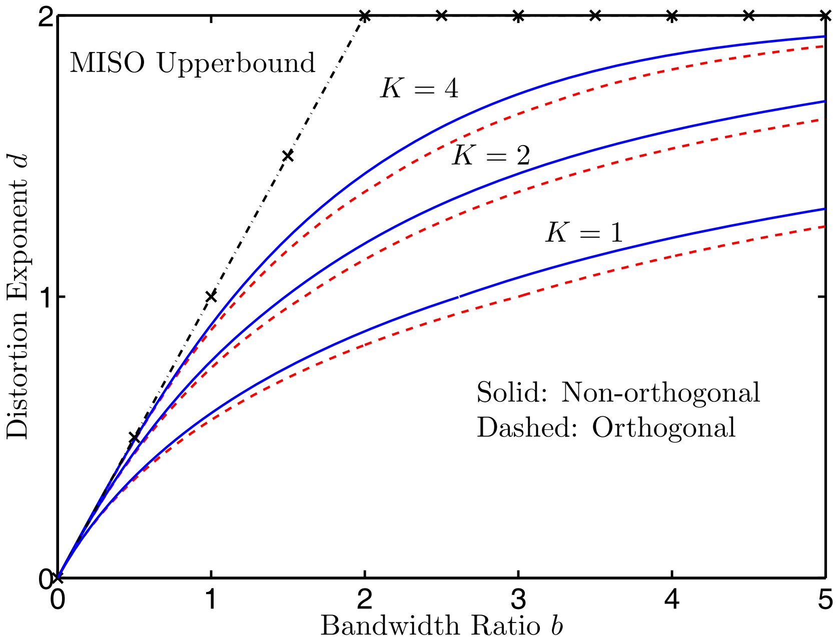

Figure 7.2: Achievable distortion exponents with different feedback resolution K. A short-term power constraint is assumed.

This is a convex optimization, since the pointwise minimum of linear functions is concave. Equating all the terms and applying Lemma 7.2 yields the claimed $d^K_{\text{O-NPC}}(b)$. It can then be verified that with this $d^K_{\text{O-NPC}}(b)$, the linear constraints $\frac{2}{b} > r_1 > \cdots > r_K > 0$ are fulfilled. □

For b outside the range specified in Corollary 7.2, we obtain *achievable* bounds by equating all the terms inside the mininum of Proposition 7.2 and solving these systems of equations numerically.

We plot in Fig. 7.2 the achievable distortion exponents of both orthogonal and non-orthogonal schemes without power control. Clearly, a few bits of feedback information provides an excellent performance, even with the very simple single-layer coding scheme that we consider. Indeed, over a certain range of b, the proposed feedback schemes with K as low as 4 (2 bits) outperform even the best known no-feedback strategies in [GE07b], which require a very high complexity in terms of infinitely many code layers (cf. Fig. 7.3). We also observe that in this low feedback rate regime, the effect of adding one bit of CSIT is much more pronounced than switching from orthogonal to non-orthogonal schemes. Thus from a practical point of view, coupling a few bits of channel state feedback with a simple orthogonal scheme seems to be an appealing approach.

At a first glance, one may tend to conjecture that as the feedback resolution K grows, the distortion exponents of the schemes will converge to the MISO upper bound $d^{\mathrm{UB}}_{\mathrm{MISO\text{-}NPC}}(b) = \min(b, 2)$. This is far from obvious though, as the capacity of the relay channel is unknown in general. The optimality of the proposed schemes is therefore not guaranteed. Indeed, the following results state that even with a continuum of multiplexing gains (rates) at the source (i.e., when the feedback levels $K \to \infty$), the MISO upper bound $d^{\mathrm{UB}}_{\mathrm{MISO\text{-}NPC}}(b)$ cannot be fully realized by the proposed strategy.

Corollary 7.3. *With a continuum of multiplexing gains we have*

$$d^{\infty}_{O\text{-}NPC}(b) = \begin{cases} b & \text{if } b \leq 1, \\ 4\sqrt{b} - b - 2 & \text{if } 1 < b \leq \frac{9}{4}, \\ 1 + \frac{b}{3} & \text{if } \frac{9}{4} < b \leq 3, \\ 2 & \text{if } 3 < b. \end{cases}$$

For non-orthogonal schemes

$$d^{\infty}_{NO\text{-}NPC}(b) = \begin{cases} b & \text{if } b \leq 1, \\ \frac{-b^2+6b-1}{4} & \text{if } 1 < b \leq \sqrt{5}, \\ 1 + \frac{3-\sqrt{5}}{2} b & \text{if } \sqrt{5} < b \leq \frac{3+\sqrt{5}}{2}, \\ 2 & \text{if } 3 < b. \end{cases}$$

Proof. We only present the orthogonal case. With a continuum of multiplexing gains at the source, we index the feedback quantization regions in terms of r instead of feedback index i. The SNR distortion exponent associated with feedback region r is given by $br + D^{\mathrm{DF}}_{\mathrm{O}}(r)$. The minimum distortion exponent is the dominant one

$$\begin{aligned} d^{\infty}_{\mathrm{O\text{-}NPC}}(b) &= \inf_{r \in (0,1)} \left\{ br + D^{\mathrm{DF}}_{\mathrm{O}}(r) \right\} \\ &= \min\left(\inf_{r \in (0,1/3)} \{ br + 2 - 3r \}, \inf_{r \in (1/3,1)} \left\{ br + \frac{2-2r}{1+r} \right\} \right). \end{aligned}$$

Combining

$$\inf_{r \in (0,1/3)} \{ br + 2 - 3r \} = \begin{cases} 2 & \text{if } b \geq 3, \\ 1 + \frac{b}{3} & \text{if } b < 3 \end{cases}$$

and

$$\inf_{r \in (1/3,1)} \left\{ br + \frac{2-2r}{1+r} \right\} = \begin{cases} b & \text{if } b \leq 1, \\ 4\sqrt{b} - b - 2 & \text{if } 1 < b < \frac{9}{4}, \\ 1 + \frac{b}{3} & \text{if } b \geq \frac{9}{4} \end{cases}$$

gives the claimed results. □

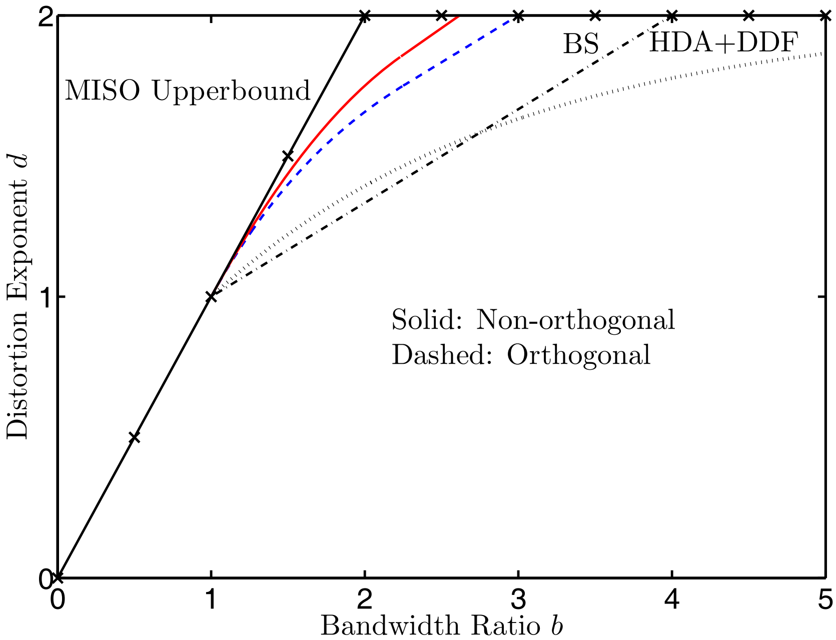

Figure 7.3: Achievable distortion exponents of the proposed schemes with a continuum of multiplexing gains (no power control). The distortion exponent of the broadcast strategy (BS) with mixed DF and direct transmission [GE07b], and that of the HDA scheme with dynamic DF [GE07b], are also plotted, both in the limit of infinitely many layers.

In Fig. 7.3, the distortion exponents achieved by using a continuum of rates at the source are plotted. Clearly both studied schemes cannot achieve the MISO upper bound over an intermediate range of the bandwidth ratio b. We also see that even in the high feedback resolution regime, the non-orthogonal scheme still displays some advantage over the orthogonal one for certain values of b. On a final note, the best known source–channel coding schemes without CSIT in [GE07b], namely the HDA scheme coupled with dynamic DF and the broadcast strategy with mixed DF and direct transmission, do not approach $d^{\infty}_{\mathrm{O-NPC}}(b)$ even when equipped with infinitely many code layers, as illustrated in Fig. 7.3.

Achievable Distortion Exponents with Power Control

We now relax the power constraint, allowing the source and the relay to control their transmit power based on channel state feedback from the destination. The scheme that we study is described as follows. Again, we only describe the orthogonal scheme.

The system is equipped with a library of K tandem codes with channel code rates $\{r_i \log \mathrm{SNR}\}$ where $1 > r_1 > \cdots > r_K > 0$ and a power codebook consisting of K predetermined power levels $\{P_1 < \cdots < P_K\}$. The destination broadcasts a common feedback index $\mathcal{I}(h, h_1, h_2)$ to both the source and the relay. Upon receiving $\mathcal{I} = i$, the source and the relay properly scale their transmit codewords so that each individual transmit power is P_i. Given $\mathcal{I} = i$, DF relaying with optimal dimension allocation is used.

The following index mapping is used by the destination

$$\mathcal{I}(h, h_1, h_2) = \begin{cases} 1 & \text{if } \mu_{\mathrm{O}}(h, h_1, h_2; P_K, r_K) < r_K \log \mathrm{SNR}, \\ \max i \in \{1, \ldots, K\} \text{ s.t. } \mu_{\mathrm{O}}(h, h_1, h_2; P_i, r_i) \geq r_i \log \mathrm{SNR} & \text{otherwise.} \end{cases}$$

Similarly to the no power control case, $\mu_{\mathrm{O}}(h, h_1, h_2; P_k, r_k)$ is the mutual information (given that power P_k is used). In essence, the scheme first tries the highest code rate possible. If that fails, lower code rates in conjunction with higher power will be attempted.

Proposition 7.3. *With power control at both source and relay, the proposed orthogonal scheme can achieve*

$$d^K_{O\text{-}PC}(b) = \sup_{1 > r_1 > \ldots > r_K > 0} \min\left(D^K_O, br_1, br_2 + D^1_O, \ldots, br_K + D^{K-1}_O\right)$$

where $D^k_O \triangleq D^{DF}_O(r_k, 1 + D^{k-1}_O)$, *with* $D^0_O = 0$.

The non-orthogonal scheme can achieve

$$d^K_{NO\text{-}PC}(b) = \sup_{1 > r_1 > \ldots > r_K > 0} \min\left(D^K_{NO}, br_1, br_2 + D^1_{NO}, \ldots, br_K + D^{K-1}_{NO}\right)$$

where $D^k_{NO} \triangleq D^{DF}_{NO}(r_k, 1 + D^{k-1}_{NO})$, *with* $D^0_{NO} = 0$.

The proof closely follows that of Proposition 7.2 and is thus omitted. The presence of the terms D^k_{O} and D^k_{NO} is due to the recursive nature of the DMT with quantized CSIT.

We demonstrate the potential of power control with limited feedback in Fig. 7.4. As can be seen, the gain of using (even partial) power control is significant. The effect is particularly pronounced in the high bandwidth ratio regime. For example, an orthogonal scheme using only two levels of power control provides a better performance than *any* no power control strategy as long as $b \geq 3.7$. Note that the achievable bounds however fall short of getting close to the corresponding K-level upper bounds, leaving room for future improvement.

In the limit of large bandwidth ratios, the distortion exponents of both orthogonal and no-orthogonal schemes converge to the upper bound $d^{\mathrm{UB}\text{-}K}_{\mathrm{MISO\text{-}PC}}(b)$ for any given K:

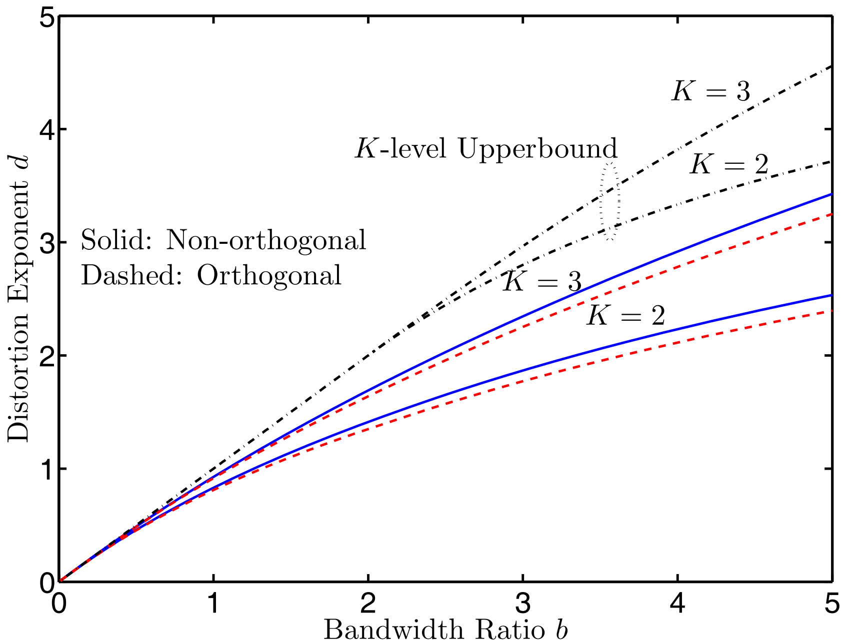

Figure 7.4: Achievable distortion exponents with power control and different K. The corresponding K-level upper bounds are also plotted.

Corollary 7.4.

$$\lim_{b\to\infty} d^K_{O\text{-}PC}(b) = \lim_{b\to\infty} d^{UB\text{-}K}_{NO\text{-}PC}(b) = \sum_{k=1}^{K} 2^k.$$

Proof. It suffices to consider the orthogonal case. By choosing $r_1 = \cdots = r_K = \hat{r} \triangleq \min\left(\frac{\sum_{k=1}^K 2^k}{b}, 1\right)$ we obtain

$$\begin{aligned}
\lim_{b\to\infty} d^K_{\text{O-PC}}(b) &\geq \lim_{b\to\infty} \min\left(D^K_{\text{O}}, b\hat{r}, b\hat{r} + D^1_{\text{O}}, \ldots, b\hat{r} + D^{K-1}_{\text{O}}\right) \\
&= \lim_{b\to\infty} \min\left(D^K_{\text{O}}, b\hat{r}\right) \\
&= \sum_{k=1}^{K} 2^k.
\end{aligned}$$

The last equality is due to $\lim_{b\to\infty} \hat{r} = 0$, and the fact that $D^K_{\text{O}} \to \sum_{k=1}^K 2^k$ when the multiplexing gain tends to zero. $\square$

7.4 Distortion Exponents with a Restricted Feedback Quantizer

In practice, while the destination can learn h and h_2 via training, acquiring accurate knowledge about the source–relay channel gain h_1 at the destination is a difficult task. This motivates the studies presented in this section where we consider the scenario that h_1 is unknown to the feedback quantizer at the destination. To derive a tighter upper bound than $D_{\text{MISO}}^{\text{UB}-K}(b)$, we let the relay and the destination fully cooperate, resulting in a 1×2 channel with K levels of feedback. However the feedback quantizer depends only on a *single* scalar gain, since the destination quantizer does not know h_1 by assumption. Computing an upper bound to the distortion exponent over such a channel with *restricted feedback quantizer* gives the following. The proof is deferred to Appendix 7.C.

Proposition 7.4. *Assume that the index mapping at the destination does not know the source-relay gain h_1. With power control at both the source and the relay, and K feedback levels from the destination, the distortion exponent is upper-bounded by*

$$\tilde{d}_{PC}^{UB-K}(b) = b - \frac{(b-2)^K}{b^{K-1}}.$$

In the limit of large bandwidth ratios b, i.e. at very low spectral efficiency, we have

Corollary 7.5. *For any $K \geq 2$*

$$\lim_{b \to \infty} \tilde{d}_{PC}^{UB-K}(b) = 2K.$$

Comparing the above limit with that of $d_{\text{MISO-PC}}^{\text{UB}-K}(b)$, we can see that not knowing h_1 incurs a large degradation in the distortion exponent. In such scenarios, the rate at which the end-to-end distortion decays to zero will increase much slower (as a function of the feedback resolution K) than in the case of full-CSI at the destination. This is due to the degradation of the *diversity gain* in this scenario, cf. Chapter 4.

We now describe a particular scheme operating under such a constraint on the CSI at the destination and study its achievable distortion exponents.

The proposed index mapping is

$$\mathcal{I}(h,h_2) = \begin{cases} K & \text{if } \log(1+|h|^2 P_{K-1}) < r_{K-1}\log \text{SNR}, \\ \max i \in \{1,\dots,K-1\} \text{ s.t. } \log(1+|h|^2 P_i) \geq r_i \log\text{SNR} & \text{otherwise.} \end{cases}$$

The source uses direct transmission for $=1,\dots,K-1$, i.e. $\beta_1 = \cdots \beta_{K-1} = 1$ and uses (either orthogonal or non-orthogonal) DF with rate $r_K \log \text{SNR}$ and optimized fraction $\beta_K^*(r_K)$ when $\mathcal{I} = K$. That is, the proposed scheme *rarely* makes use of the relay.

Proposition 7.5. *If* $b \in \{b : \left(b^2-3\right)\left(\frac{1+b}{b}\right)^{K-1} > b^2 - b\}$ *then the proposed orthogonal scheme can achieve*

$$\tilde{d}^K_{O\text{-}PC}(b) = \frac{(2b+3)\left(\frac{1+b}{b}\right)^{K-1} - 2b - 1}{\left(2+\frac{3}{b}\right)\left(\frac{1+b}{b}\right)^{K-1} - 1}.$$

If $b \in \{b : \left(b^2-3\right)\left(\frac{1+b}{b}\right)^{K-1} \leq b^2 - b\}$ *then* $\tilde{d}^K_{O\text{-}PC}(b) = \frac{-B-\sqrt{B^2-4AC}}{2A}$ *where*

$$A = 2(1+b)^{2K-1} - 3b^{K-1}(1+b)^K + b^{2K-1},$$

$$B = b\left(3b^K - 4(1+b)^K\right)\left((1+b)^{K-1} - b^{K-1}\right) - (b+2)b^{K-1}(1+b)^{K-1},$$

and

$$C = 2b^2\left((1+b)^K - b^K\right)\left(b^{K-2} + (1+b)^{K-1} - b^{K-1}\right).$$

For non-orthogonal schemes, if

$$b \in \left\{b : \left(b^2 - (3-\sqrt{5})b - \frac{5+\sqrt{5}}{2}\right)\left(\frac{1+b}{b}\right)^{K-1} > b^2 - \frac{5-\sqrt{5}}{2}b\right\}$$

then

$$\tilde{d}^K_{NO\text{-}PC}(b) = \frac{\left(2b+\frac{3+\sqrt{5}}{2}\right)\left(\frac{1+b}{b}\right)^{K-1} - 2b - \frac{\sqrt{5}-1}{2}}{\left(2+\frac{3+\sqrt{5}}{2b}\right)\left(\frac{1+b}{b}\right)^{K-1} - 1}$$

If $b \in \left\{b : \left(b^2 - (3-\sqrt{5})b - \frac{5+\sqrt{5}}{2}\right)\left(\frac{1+b}{b}\right)^{K-1} \leq b^2 - \frac{5-\sqrt{5}}{2}b\right\}$ *then* $\tilde{d}^K_{NO\text{-}PC}(b) = \frac{-B-\sqrt{B^2-4AC}}{2A}$ *where*

$$A = b^{2K} - 3b^K(1+b)^K + (2b+1)(1+b)^{2K-1},$$

$$B = -3b^{2K+1} - 2b(2b+1)(1+b)^{2K-1} + (7b^2+5b-1)b^K(1+b)^{K-1},$$

and

$$C = b^2\left((1+b)^K - b^K\right)\left((2b+1)(1+b)^{K-1} - (2b-1)b^{K-1}\right).$$

In Fig. 7.5, we plot the achievable distortion exponents with restrictive feedback quantizers. Even these restricted feedback schemes yield significant gains over a no-CSIT system, but they stay short of getting close to the upper bound except at the high spectral efficiency regime (very small b) and the very low spectral efficiency regime (very high b - not plotted herein). We can also conclude that the high-SNR gain of non-orthogonal schemes is negligible.

Finally, for completeness, we state the asymptotic optimality of both schemes in the high-b regime. The proof is similar to that of Corollary 7.4.

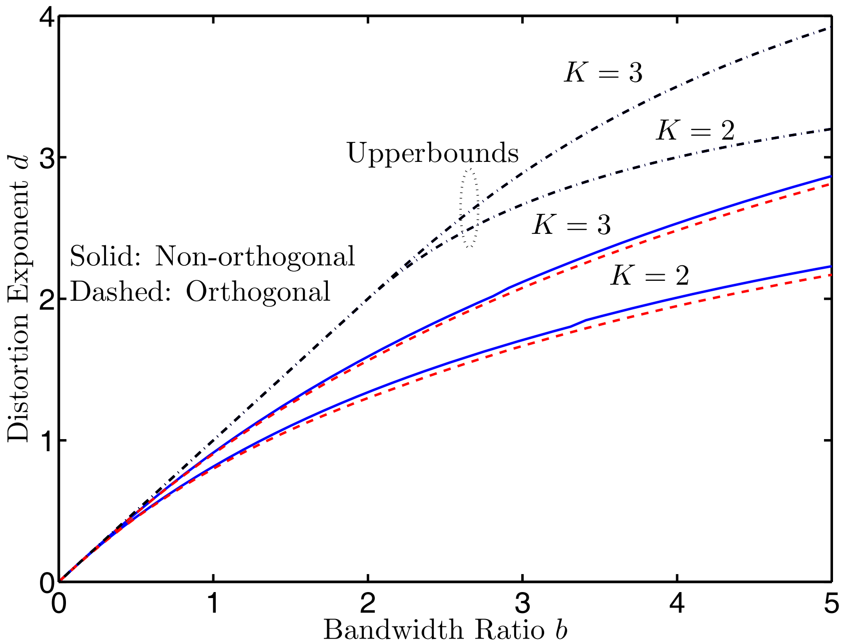

Figure 7.5: Achievable distortion exponents with restricted feedback quantizers and different K.

Corollary 7.6. *For any given* $K \geq 2$

$$\lim_{b\to\infty} \tilde{d}^K_{O-PC}(b) = \lim_{b\to\infty} \tilde{d}^K_{NO-PC}(b) = 2K.$$

As a final remark, note that the achievable distortion exponents in Propositions 7.2, 7.3, and 7.5 are obtained under the assumption of infinitely large block length T, so that the rate–distortion bound and the outage bound are both asymptotically achievable. How does the performance change with a fixed, finite block length T? Since all the proposed schemes in the current work rely on the *separation* of source and channel coding, it suffices that each individual component (source and channel) code has performance that decays with the appropriate SNR exponent. It is known that scalar quantizers applied to a Gaussian source yield the same exponent as that of the optimal rate–distortion function. Hence, in terms of the source code, *scalar quantization* is good enough for our purposes. On the other hand, for a finite T, there exist only a restricted number of discrete dimension splitting values β (i.e., the fraction of channel uses assigned for Phase 1 of the DF relaying protocol), instead of a continuum (i.e., when β can take on any values in $(0, 1)$). This entails a certain loss in the achievable *error exponent* of the channel code compared to the *outage exponent*. Consequently, for any finite T the achievable distortion exponents are generally less than those obtained for infinite T presented in this work, due to the *channel coding* part. This loss however asymptotically (and rather quickly, as studied in Chapter 4) vanishes as T increases.

7.5 Conclusion

We have investigated the problem of rate and power allocation using partial CSIT to maximize the decay rate of the end-to-end expected distortion in a single-antenna relay channel. Even a few levels of rate control allows a simple separate source and channel coding scheme to outperform the best known complex no-feedback strategies. The performance can be further improved with an appropriate partial power control policy. Our results suggest that combining simple relaying schemes with low-rate channel state feedback can be an effective approach for the transmission of analog sources over cooperative channels.

Appendices for Chapter 7

7.A Proof of Proposition 7.1

Consider a 1×2 slow fading channel with i.i.d. zero-mean unit-variance complex Gaussian channel gains a_1, a_2. Let $\alpha = -\frac{\log\left(|a_1|^2+|a_2|^2\right)}{\log \mathrm{SNR}}$. To obtain an upper bound to the distortion exponent, we assume perfect CSIT so that the transmitter can perform perfect rate adaptation. However the transmitter is *constrained* to use only K power levels.

Let $P_k \doteq \mathrm{SNR}^{p_k}$ be the K power levels at the transmitter, with $1 = p_1 < \cdots < p_K < p_{K+1} = \infty$. We refer to the set of all α's that are mapped to a certain power level P_k as the kth quantization region $\mathcal{R}_k$. We only consider $\alpha \geq 0$ as the region $a < 0$ has a probability measure that exponentially decays in SNR and thus does not affect the SNR exponent of interest [ZT03]. The long-term power constraint leads to

$$\int_{\mathcal{R}_k} \mathrm{SNR}^{p_k} f(\alpha) d\alpha \,\dot{\leq}\, \mathrm{SNR}$$

where $f(\alpha)$ is the p.d.f. of α. Some standard large-deviation arguments [ZT03] then lead to

$$\sup_{\alpha \in \mathcal{R}_k} \{p_k - 2\alpha\} \leq 1$$

Since the end-to-end distortion is a decreasing function of transmit power, we conclude that

$$\mathcal{R}_k = \left\{ \alpha \geq 0 : \frac{p_k - 1}{2} \leq \alpha < \frac{p_{k+1} - 1}{2} \right\}.$$

Because the transmitter can do perfect rate adaptation, the multiplexing gain of the transmit rate corresponding to a channel state α is the same as that of $\log(1+\mathrm{SNR}^{p_k-\alpha})$. Assume codes achieving the rate-distortion function, the average distortion over $\mathcal{R}_k$ is then

$$\bar{\Delta}_k \doteq \int_{\mathcal{R}_k} 2^{-b\log(1+\mathrm{SNR}^{p_k-\alpha})} f(\alpha) d\alpha$$

Thus for a fixed set of power levels, the distortion exponent over $\mathcal{R}_k$ is given by

$$d_k = \inf_{\alpha\in\mathcal{R}_k} \left\{ b(p_k-\alpha)^+ + 2\alpha \right\}.$$

Since $p_{K+1} = \infty$, over $\mathcal{R}_K$ the optimization readily gives $d_K = 2p_K$ with the optimizer $\alpha^* = p_K$. For any $k < K$,

$$d_k = \begin{cases} 2p_k & \text{if } p_k < \frac{p_{k+1}-1}{2}, \\ bp_k - \frac{(b-2)(p_{k+1}-1)}{2} & \text{if } p_k \geq \frac{p_{k+1}-1}{2}. \end{cases}$$

However, since $d = \min(d_1, \ldots, d_K)$, the case $d_k = 2p_k$ is not of interest, otherwise $2p_k < d_K = 2p_K$ meaning that adding the quantization region $\mathcal{R}_K$ is redundant. Thus we only consider $p_k \geq \frac{p_{k+1}-1}{2}$, and optimize the power levels to obtain

$$\begin{aligned} D^{\mathrm{UB}-K}_{\mathrm{MISO-PC}}(b) = \sup_{p_2,\ldots,p_K} \min \Bigg(& 2p_K, bp_1 - \frac{b-2}{2}p_2 + \frac{b-2}{2}, \ldots, \\ & bp_{K-1} - \frac{b-2}{2}p_K + \frac{b-2}{2} \Bigg) \\ \text{s.t.} \quad & 2p_1 \geq p_2 - 1, \ldots, 2p_{K-1} \geq p_K - 1, p_K > \cdots > p_1. \end{aligned}$$

Recall that $p_1 = 1$. Since the functions inside the point-wise minimum are linear in p_k's, the solution to this maximin is the intersection of all the hyperplanes, i.e. at $d_1 = \cdots = d_K \triangleq d^*$, provided that the intersection satisfies all the linear constraints. We now have

$$\begin{aligned} p_K &= \frac{d^*}{2}, \\ p_{K-1} &= \frac{d^*}{b}\left(1 + \frac{b-2}{4}\right) - \frac{b-2}{2b}, \\ &\cdots \\ p_{K-k} &= \frac{d^*}{b}\left(1 + \frac{b-2}{2b} + \cdots + \left(\frac{b-2}{2b}\right)^{k-1} + \frac{(b-2)^k}{2^{k+1}b^{k-1}}\right) - \sum_{i=1}^{k}\left(\frac{b-2}{2b}\right)^i, \\ &\cdots \\ p_1 &= \frac{d^*}{b}\left(\frac{1-\left(\frac{b-2}{2b}\right)^{K-1}}{\frac{b+2}{2b}} + \frac{(b-2)^{K-1}}{2^K b^{K-2}}\right) - \frac{b-2}{2b}\frac{1-\left(\frac{b-2}{2b}\right)^{K-1}}{\frac{b+2}{2b}}. \end{aligned}$$

Then, using the fact that $p_1 = 1$, we have

$$d^* = \frac{1 + \frac{b-2}{b+2}\left(1 - \left(\frac{b-2}{2b}\right)^{K-1}\right)}{\frac{1-\left(\frac{b-2}{2b}\right)^{K-1}}{\frac{b+2}{2}} + \frac{1}{2}\left(\frac{b-2}{2b}\right)^{K-1}} = \frac{4b - 2(b-2)\left(\frac{b-2}{2b}\right)^{K-1}}{4 + (b-2)\left(\frac{b-2}{2b}\right)^{K-1}}.$$

We can readily verify that with this d^*, the constraints $2p_k \geq p_{k+1}$ are all fulfilled. Thus $d_{\text{MISO–PC}}^{\text{UB–}K}(b) = d^*$. □

7.B Proof of Proposition 7.2

For brevity, we only present the orthogonal case. Let $a = -\log|h|^2/\log \text{SNR}$, $\alpha_1 = -\log|h_1|^2/\log \text{SNR}$, $\alpha_2 = -\log|h_2|^2/\log \text{SNR}$ and thus we can write the mutual information $\mu_{\text{O}}(a, \alpha_1, \alpha_2; r) = \mu_{\text{O}}(h, h_1, h_2; r)$. The set of all channel state tuples that are mapped to an index $\mathcal{I} = k$ is referred to as quantization region $\mathcal{R}_k$. Let the distortion exponent over $\mathcal{R}_k$ be d_k. Define

$$\begin{aligned}\mathcal{O}_k \triangleq & \{a, \alpha_1, \alpha_2 \in \mathbb{R}_+^3 : \beta^*(r_k)(1-\alpha_1)^+ < r_k, \beta^*(r_k)(1-a)^+ < r_k\} \\ & \cup \{a, \alpha_1, \alpha_2 \in \mathbb{R}_+^3 : \beta^*(r_k)(1-\alpha_1)^+ \geq r_k, \\ & \quad \beta^*(r_k)(1-a)^+ + (1-\beta^*(r_k))(1-\alpha_2)^+ < r_k\}\end{aligned}$$

which is essentially the asymptotic outage set conditioned on $\mathcal{I} = k$.

We readily have $\Pr(\mathcal{I} = 1) \doteq \text{SNR}^0$ (intuitively, most channel uses do not result in outage at high SNR), thus

$$\int_{\mathcal{R}_1} 2^{-b_1 r_1 \log \text{SNR}} da d\alpha_1 d\alpha_2 \doteq \text{SNR}^{-br_1} = \text{SNR}^{-d_1}.$$

The exponent equality is due to an application of Bennett's integral for high-rate vector quantization [NN95]. We next have

$$\Pr(\mathcal{I} = 2) = \Pr(\bar{\mathcal{O}}_2 \cap \mathcal{O}_1) \dot{\leq} \Pr(\mathcal{O}_1). \tag{7.2}$$

Let $\Pr(\mathcal{I} = 2) \doteq \text{SNR}^{-D_2}$ then from (7.2): $D_2 \geq D_{\text{O}}^{\text{DF}}(r_1) = 2 - \frac{2r_1}{\beta_1^*(r_1)}$. However by choosing a particular tuple $\hat{a} = 1 - \frac{r_1}{\beta_1^*(r_1)} + \epsilon$, $\hat{\alpha}_1 = \hat{a}$, $\hat{\alpha}_2 = 0$ where $\epsilon > 0$ is arbitrarily small, we then have

$$D_2 = \inf_{\bar{\mathcal{O}}_2 \cap \mathcal{O}_1} \{\hat{a} + \hat{\alpha}_1 + \alpha_2\} \leq 2 - \frac{2r_1}{\beta_1^*(r)} + 2\epsilon. \tag{7.3}$$

For (7.3) to hold we have to show that $(\hat{a}, \hat{\alpha}_1, \hat{\alpha}_2) \in \{\bar{\mathcal{O}}_2 \cap \mathcal{O}_1\}$. To that end, notice that $\beta_1^*(r_1)(1-\hat{a})^+ = \beta_1^*(r_1)(1-\hat{\alpha}_1)^+ = r_1 - \beta_1^*(r_1)\epsilon < r_1$ thus $(\hat{a}, \hat{\alpha}_1, \hat{\alpha}_2) \in \mathcal{O}_1$ (since both source–relay and source–destination links are in outage). In addition,

$\beta_2^*(r_2)(1-\hat{a})^+ = \beta_2^*(r_2)(1-\hat{\alpha}_1)^+ = \beta_2^*(r_2)\frac{r_1}{\beta_1^*(r_1)} - \beta_2^*(r_2)\epsilon > r_2$ for sufficiently small ϵ. This is because

$$\frac{r}{\beta^*(r)} = \begin{cases} \frac{3r}{2} & \text{if } r < \frac{1}{3}, \\ \frac{2r}{1+r} & \text{otherwise.} \end{cases}$$

is a monotonically increasing function of r, and thus for any $r_1 > r_2$ we have $\frac{r_1}{\beta_1^*(r_1)} > \frac{r_2}{\beta_2^*(r_2)}$ with strict inequality. Thus $(\hat{a}, \hat{\alpha}_1, \hat{\alpha}_2) \in \bar{\mathcal{O}}_2$ and (7.3) holds. Finally, letting $\epsilon \downarrow 0$ we conclude that $\Pr(\mathcal{I} = 2) \doteq \mathrm{SNR}^{-D_{\mathrm{O}}^{\mathrm{DF}}(r_1)}$ and consequently $d_2 = b_2 r_2 + D_{\mathrm{O}}^{\mathrm{DF}}(r_1)$. Continuing this line of arguments leads to the claimed results.

7.C Proof of Proposition 7.4

Let the relay and destination fully cooperate, resulting in a 1×2 multiple-receive antenna channel. Let a_1, a_2 be the two channel gains. Let $\alpha_1 = -\log|a_1|^2/\log \mathrm{SNR}$ and $\alpha_2 = -\log|a_2|^2/\log \mathrm{SNR}$. The feedback index (thus the transmit power) only depends on a_1 and not on a_2. Thus we have $\sup_{\alpha_1 \in \mathcal{R}_k} \{p_k - \alpha\} \leq 1$ where $\mathcal{R}_k$ is the kth quantization region. Then similarly to Appendix 7.A

$$\mathcal{R}_k = \{\alpha_1 : p_k - 1 \leq \alpha_1 < p_{k+1} - 1\}.$$

The code rate associated with the channel state α_1, α_2 has the same multiplexing gain as $\log\left(1 + \mathrm{SNR}^{p_k - \alpha_1} + \mathrm{SNR}^{p_k - \alpha_2}\right)$. Thus the distortion exponent over $\mathcal{R}_k$ is

$$d_k = \inf_{\alpha_1 \in \mathcal{R}_k, \alpha_2 \geq 0} b \max\left((p_k - \alpha_1), (p_k - \alpha_2), 0\right) + \alpha_1 + \alpha_2.$$

We begin with $\mathcal{R}_K$ where it is readily found that $d_K = 2p_K$. For $k < K$, it can be shown that

$$d_k = \begin{cases} 2p_k & \text{if } p_k < p_{k+1} - 1, \\ b(p_k - p_{k+1} + 1) + 2(p_{k+1} - 1) & \text{if } p_k \geq p_{k+1} - 1. \end{cases}$$

Again the case $d_k = 2p_k < 2p_K$ is not of interest. Thus we end up with the maximin problem

$$\begin{aligned} \tilde{d}_{\mathrm{PC}}^{\mathrm{UB}-K}(b) = \sup_{p_2,\ldots,p_K} & \min\left(2p_K, bp_1 - (b-2)p_2 + b - 2, \ldots, \right. \\ & \left. bp_{K-1} - (b-2)p_K + b - 2\right) \\ \text{s.t.} \quad & p_1 > p_2 - 1, \ldots, p_{K-1} > p_K - 1, p_K > \ldots > p_2 > p_1 = 1. \end{aligned}$$

This is again a convex optimization problem. We will find the intersection of all the hyperplanes and verify that it satisfies all the linear constraints.

Let $d_1 = \cdots = d_K \triangleq d^*$. We have

$$\begin{aligned}
p_K &= \frac{d^*}{2}, \\
p_{K-1} &= \frac{d^*}{b}\left(1 + \frac{b-2}{2}\right) - \frac{b-2}{b}, \\
&\cdots \\
p_{K-k} &= \frac{d^*}{b}\left(1 + \frac{b-2}{b} + \cdots + \left(\frac{b-2}{b}\right)^{k-1} + \frac{(b-2)^k}{2b^{k-1}}\right) - \sum_{i=1}^{k}\left(\frac{b-2}{b}\right)^i, \\
&\cdots \\
p_1 &= \frac{d^*}{b}\left(\frac{1-\left(\frac{b-2}{b}\right)^{K-1}}{\frac{2}{b}} + \frac{(b-2)^{K-1}}{2b^{K-2}}\right) - \frac{b-2}{b}\frac{1-\left(\frac{b-2}{b}\right)^{K-1}}{\frac{2}{b}}
\end{aligned}$$

But $p_1 = 1$ thus we obtain

$$\begin{aligned}
d^* &= \frac{1 + \frac{(b-2)\left(1-\left(\frac{b-2}{b}\right)^{K-1}\right)}{2}}{\frac{1-\left(\frac{b-2}{b}\right)^{K-1}}{2} + \frac{(b-2)^{K-1}}{2b^{K-1}}} \\
&= 2 + (b-2)\left(1 - \left(\frac{b-2}{b}\right)^{K-1}\right) \\
&= b - \frac{(b-2)^K}{b^{K-1}}.
\end{aligned}$$

Again all the linear constraints are fulfilled with this d^* and thus $\tilde{d}_{\text{PC}}^{\text{UB}-K}(b) = d^*$. □

7.D Proof of Proposition 7.5

We only present the orthogonal case. Let p_k be the SNR exponent of the power levels, i.e. $P_k \doteq \text{SNR}^{p_k}$. Denote d_k as the distortion exponent over the quantization $\mathcal{R}_k$ (the set of all channel state tuples that are mapped to $\mathcal{I} = k$).

We have $\Pr(\mathcal{I} = 1) \doteq \text{SNR}^0$ and thus $d_1 = br_1$. Then

$$\Pr(\mathcal{I} = 2) \doteq \text{SNR}^{-D_{\text{SISO}}(r_1,1)}.$$

Herein $D_{\text{SISO}}(r,p) = p - r$ is the diversity gain of a SISO channel corresponding to a multiplexing gain r and a transmit power SNR^p. We then obtain $d_2 = br_2 + D_{\text{SISO}}(r_1, 1) = br_2 + 1 - r_1$. The power applied to $\mathcal{R}_2$ thus have the SNR exponent $p_2 = 1 + D_{\text{SISO}}(r_1, 1) = 1 + 1 - r_1 = 2 - r_1$. Then $\Pr(\mathcal{I} = 3) \doteq \text{SNR}^{-D_{\text{SISO}}(r_2, 2-r_1)} = \text{SNR}^{-(2-r_1-r_2)}$. We next have $d_3 = br_3 + 2 - r_1 - r_2$.

Repeating the above arguments finally leads to $d_k = br_k + k - 1 - r_1 - \cdots - r_{k-1}$, $k = 1, \ldots, K-1$. Notice that due to construction, an outage event (the event that the channel does not support the code rate) can only occur when $\mathcal{I} = K$. Over the region $\mathcal{R}_K$ we have

$$d_K = \min(br_K + K - 1 - r_1 - \cdots - r_{K-1}, D_{\mathrm{O}}^{\mathrm{DF}}(r_K, p_K)).$$

The optimization finding $\tilde{d}_{\mathrm{O-PC}}^{K}(b)$ can thus be written as

$$\sup_{1>r_1>\ldots>r_K>0} \min\big(br_1, br_2 + 1 - r_1, br_3 + 2 - r_1 - r_2, \ldots, \\ br_K + K - 1 - r_1 - \cdots - r_{K-1}, D_{\mathrm{O}}^{\mathrm{DF}}(r_K, K - r_1 - \ldots - r_{K-1})\big)$$

Note that this maximin is not a convex problem. However, since we are only interested in an achievable distortion exponent, we can choose $r_1, \ldots, r_K$ to be the solutions to

$$\begin{aligned} d^* &\triangleq D_{\mathrm{O}}^{\mathrm{DF}}(r_K, K - r_1 - \ldots - r_{K-1}) = br_1 = br_2 + 1 - r_1 = \cdots \\ &= br_K + K - 1 - r_1 - \cdots - r_{K-1}. \end{aligned}$$

That is we find the point where all the terms are equal (and verify that such a point exists). From $br_{k-1} + k - 2 - r_1 - \cdots - r_{k-2} = br_k + k - 1 - r_1 - \cdots - r_{k-1}$ we obtain

$$(1+b)r_{k-1} = 1 + br_k.$$

This leads to the recursive relation

$$r_k = \frac{1+b}{b} r_{k-1} - \frac{1}{b}$$

with the initial value $r_1 = \frac{d^*}{b}$. This recursion then gives

$$r_k = 1 - \left(\frac{1+b}{b}\right)^{k-1} \frac{b - d^*}{b}. \tag{7.4}$$

We now consider the case when $r_K < \frac{p_K}{3} = \frac{K - r_1 - \ldots - r_{K-1}}{3}$. From Lemma 7.1, in this case we have $D_{\mathrm{O}}^{\mathrm{DF}}(r_k, K - r_1 - \ldots - r_{K-1}) = 2(K - r_1 - \ldots - r_{K-1}) - 3r_K$. From (7.4) we have

$$\sum_{i=1}^{K-1} r_i = K - 1 - \frac{\left(\frac{1+b}{b}\right)^{K-1} - 1}{\frac{1}{b}} \frac{b - d^*}{b} = K - 1 - \left[\left(\frac{1+b}{b}\right)^{K-1} - 1\right](b - d^*).$$

This leads to

$$\begin{aligned} d^* &= 2(K - r_1 - \ldots - r_{K-1}) - 3r_K \\ &= 2\left(1 + \left[\left(\frac{1+b}{b}\right)^{K-1} - 1\right](b - d^*)\right) - 3\left(1 - \left(\frac{1+b}{b}\right)^{K-1} \frac{b - d^*}{b}\right) \\ &= -1 + (b - d^*)\left[2\left(\frac{1+b}{b}\right)^{K-1} - 2 + \frac{3}{b}\left(\frac{1+b}{b}\right)^{K-1}\right]. \end{aligned}$$

Finally we obtain

$$d^* = \frac{(2b+3)\left(\frac{1+b}{b}\right)^{K-1} - 2b - 1}{\left(2+\frac{3}{b}\right)\left(\frac{1+b}{b}\right)^{K-1} - 1}.$$

Recall that this happens iff $r_K < \frac{p_K}{3} = \frac{K - r_1 - \ldots - r_{K-1}}{3}$, or

$$\begin{aligned} & 3 - 3\left(\frac{1+b}{b}\right)^{K-1}\frac{b-d^*}{b} < 1 + \left[\left(\frac{1+b}{b}\right)^{K-1} - 1\right](b-d^*) \\ & \Leftrightarrow 2 < (b-d^*)\left[\left(\frac{1+b}{b}\right)^{K-1} - 1 + \frac{3}{b}\left(\frac{1+b}{b}\right)^{K-1}\right] \\ & \Leftrightarrow 2 < \left(b - \frac{(2b+3)\left(\frac{1+b}{b}\right)^{K-1} - 2b - 1}{\left(2+\frac{3}{b}\right)\left(\frac{1+b}{b}\right)^{K-1} - 1}\right)\left[\left(1+\frac{3}{b}\right)\left(\frac{1+b}{b}\right)^{K-1} - 1\right] \\ & \Leftrightarrow \left(b^2 - 3\right)\left(\frac{1+b}{b}\right)^{K-1} > b^2 - b. \end{aligned}$$

We now consider the remaining case $r_K \geq \frac{p_K}{3}$, when Lemma 7.1 yields

$$d^* = D_{\mathrm{O}}^{\mathrm{DF}}(r_K, p_K) = \frac{2p_K(p_K - r_K)}{p_K + r_K}.$$

Inserting (7.4) into the above expression gives

$$\begin{aligned} & d^*\left(2 + \left[\frac{b-1}{b}\left(\frac{1+b}{b}\right)^{K-1} - 1\right](b-d^*)\right) \\ & = 2\left(1 + \left[\left(\frac{1+b}{b}\right)^{K-1} - 1\right](b-d^*)\right)\left[\left(\frac{1+b}{b}\right)^{K} - 1\right](b-d^*). \end{aligned} \tag{7.5}$$

This results in a quadratic equation of the form $f(d^*) = 0$ where $f(d^*) \triangleq Ad^{*2} + Bd^* + C$ with the coefficients $A > 0, B$, and C are given by Proposition 7.5. Using (7.5), we can readily verify that $f(b) < 0$ and $f(0) > 0$, $\forall b, K$ so that the equation $f(d^*) = 0$ has two real roots d_1^*, d_2^* satisfying $0 < d_1^* < b < d_2^*$. This gives the claimed result. □

Chapter 8

Conclusion

8.1 Concluding Remarks

In this thesis, we have investigated the roles of partial CSIT in slow fading channels, including multiple-antenna and relaying systems. We considered many different performance metrics, leading to the introduction of radically different transmission schemes. These schemes may utilize partial CSIT to adapt their transmission power, rate and even the time spent on each transmission phase. Analytical performance bounds are derived in each case, suggesting that very promising performance improvement may be achieved even if the CSIT is heavily quantized.

One important remark is that the impact of partial CSIT on fast and slow fading channels can be strikingly different, even if the performance criteria are quite similar. For instance, as we have seen from the study of expected rate, a couple of feedback bits can improve the expected rate over a scalar slow fading channel significantly. In contrast, the influence of quantized feedback on the ergodic capacity of scalar fast fading channels is typically very small, especially at moderate and high SNR's.

Another conclusion is the significant influence of rate and power adaptation over slow fading channels, even with coarsely quantized CSIT. In particular, with intelligent adaptation based on partial CSIT, the frame error probability of multiple-antenna and relaying systems can be made to decay extremely fast at high SNR. Remarkably, this improvement in reliability does not require any sacrifice in the asymptotic throughput. In relaying systems, we can additionally control the fraction of channel uses allocated to different transmission phases, which also leads to significant performance gains. This conclusion holds even when we transmit an analog source over slow fading channels and take the decaying rate of the end–to–end expected distortion as performance measure.

8.2 Future Work

Of course, many interesting questions still need answering, especially in relaying systems. A particularly intriguing problem is to find relaying schemes that are D–M tradeoff optimal when the source and the relay have absolutely no CSIT. Recall that with full CSIT, it is known that compress–and–forward relaying is optimal in the D–M tradeoff sense. Generalizing the results in the thesis to multiple-relay systems with multiple-antenna nodes is also an important extension.

Designing and analyzing more sophisticated joint source–channel coding schemes that can achieve the partial-CSIT upper bounds derived in the thesis is an interesting open problem. A better understanding of the effects of finite-length codes on the end–to–end distortion exponent also requires more work. We may also think of an investigation on the distortion outage, i.e., the event that the end–to–end distortion drops below an acceptable threshold.

A critical problem for practical wireless communication systems that adapt their resources based on CSIT is the processing delay and the presence of noise in the feedback link. Within the scope of this thesis, we have not paid enough attention to these effects. Further work is needed before we can fully understand the detrimental effects that delayed and noisy feedback may cause to the D–M tradeoff and the distortion exponent–bandwidth ratio tradeoff presented in the thesis. Designing intelligent schemes that are robust to noisy and delayed feedback would also be of practical relevance.

Apart from the expected rate maximization problem, we have mostly focused on the asymptotically high SNR regime. While this asymptotic analysis provides a "clean" and compact characterization, thus giving useful insight into complicated problems, it does not necessarily give a complete picture about the system performance. In particular, the D–M tradeoff (as well as the distortion exponent analysis) totally ignores any constant gains in SNR, which are definitely important for any practical communication systems. Therefore, a finer characterization of the system performance in a more practical range of the SNR would be a worthwhile and useful direction for future work.

Bibliography

[AES05] K. Azarian, H. El Gamal, and P. Schniter. On the achievable diversity-multiplexing tradeoffs in half-duplex cooperative channels. *IEEE Transactions on Information Theory*, 51:4152–4172, December 2005.

[AEVZ02] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger. Closest point search in lattices. *IEEE Transactions on Information Theory*, 48:2201–2214, August 2002.

[AKSA06] N. Ahmed, M. A. Khojastepour, A. Sabharwal, and B. Aazhang. Outage minimization with limited feedback for the fading relay channel. *IEEE Transactions on Communications*, 54:659–669, April 2006.

[Ala98] S. M. Alamouti. A simple transmitter diversity scheme for wireless communications. *IEEE Journal on Selected Areas in Communications*, 16:1451–1458, October 1998.

[BBT59] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacity of a class of channels. *Ann. Math. Statist.*, 30:1229–1241, December 1959.

[BCT01] E. Biglieri, G. Caire, and G. Taricco. Limiting performance of block-fading channels with multiple antennas. *IEEE Transactions on Information Theory*, 47:1273–1289, May 2001.

[BGT93] C. Berrou, A. Glavieux, and P. Thitimajshima. Near Shannon limit error-correcting coding and decoding: Turbo-codes. In *Proc. IEEE Int. Conf. Communications*, pages 1064–1070, May 1993.

[BNC06] K. Bhattad, K. Narayanan, and G. Caire. On the distortion exponent of some layered transmission schemes. In *Proc. Asilomar Conf. Signals, Systems, Computers*, pages 211–215, Pacific Grove, CA, October 2006.

[BPS98] E. Biglieri, J. Proakis, and S. Shamai (Shitz). Fading channels: Information-theoretic and communications aspects. *IEEE Transactions on Information Theory*, 44:2619–2692, October 1998.

[BR03] J.-C. Belfiore and G. Rekaya. Quaternionic lattices for space-time coding. In *Proc. Information Theory Workshop*, pages 267–270, April 2003.

[BRV05] J.-C. Belfiore, G. Rekaya, and E. Viterbo. The Golden code: A 2×2 full-rate space-time code with non-vanishing determinants. *IEEE Transactions on Information Theory*, 51:1432–1436, April 2005.

[BSA02] S. Bhashyam, A. Sabharwal, and B. Aazhang. Feedback gain in multiple antenna systems. *IEEE Transactions on Communications*, 50:785–798, May 2002.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge Univ. Press, Cambridge, UK, 2004.

[CE79] T. M. Cover and A. A. El Gamal. Capacity theorems for the relay channel. *IEEE Transactions on Information Theory*, 25:572–584, September 1979.

[CN05] G. Caire and K. Narayanan. On the SNR exponent of hybrid digital analog space-time coding. In *Proc. Annu. Allerton Conf. Communication, Control, Computing*, Monticello, IL, September 2005.

[CN07] G. Caire and K. Narayanan. On the distortion SNR exponent of hybrid digital-analog space time coding. *IEEE Transactions on Information Theory*, 53:2867–2878, August 2007.

[Cov72] T. Cover. Broadcast channels. *IEEE Transactions on Information Theory*, 18:2–14, January 1972.

[CPR03] J. Chou, S. S. Pradhan, and K. Ramchandran. Turbo and trellis-based constructions for source coding with side information. In *Proc. Data Compression Conf.*, pages 33–42, Snowbird, UT, March 2003.

[CS99] G. Caire and S. Shamai (Shitz). On the capacity of some channels with channel state information. *IEEE Transactions on Information Theory*, 45:2007–2019, September 1999.

[CT91] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley & Sons, New York, 1991.

[CTB99] G. Caire, G. Taricco, and E. Biglieri. Optimum power control over fading channels. *IEEE Transactions on Information Theory*, 45:1468–1489, July 1999.

[CX05] S. Cheng and Z. Xiong. Successive refinement for the Wyner-Ziv problem and layered code design. *IEEE Transactions on Signal Processing*, 53:3269–3281, August 2005.

[DZ98] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications.* Springer, New York, 1998.

[ECD04] H. El Gamal, G. Caire, and M. O. Damen. Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels. *IEEE Transactions on Information Theory*, 50:968–985, June 2004.

[ECD06] H. El Gamal, G. Caire, and M. O. Damen. The MIMO ARQ channel: Diversity-multiplexing-delay tradeoff. *IEEE Transactions on Information Theory*, 52:3601–3621, August 2006.

[EG98] M. Effros and A. Goldsmith. Capacity definitions and coding strategies for general channels with receiver side information. In *Proc. IEEE Int. Symp. Information Theory*, page 39, August 1998.

[EKP+06] P. Elia, K. R. Kumar, S. A. Pawar, P. V. Kumar, and H.-F. Lu. Explicit space-time codes achieving the diversity-multiplexing gain tradeoff. *IEEE Transactions on Information Theory*, 52:3869–3884, September 2006.

[EVAK06] P. Elia, K. Vinodh, M. Anand, and P. V. Kumar. D-MG tradeoff and optimal codes for a class of AF and DF cooperative communication protocols. *IEEE Transactions on Information Theory*, November 2006. submitted for publication.

[FGW02] A. Forsgren, P. E. Gill, and M. H. Wright. Interior methods for nonlinear optimization. *SIAM Review*, 44:525–597, October 2002.

[Fos96] G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. Technical report, Bell Labs. Tech. J., 1996.

[Gal62] R. Gallager. Low-density parity-check codes. *IEEE Transactions on Information Theory*, 8:21–28, January 1962.

[Gal65] R. G. Gallager. A simple derivation of the coding theorem and some applications. *IEEE Transactions on Information Theory*, 11:3–18, January 1965.

[Gal68] R. G. Gallager. *Information Theory and Reliable Communication*. John Wiley & Sons, New York, NY, 1968.

[GC97] A. J. Goldsmith and S.-G. Chua. Variable-rate variable-power MQAM for fading channels. *IEEE Transactions on Communications*, 45:595–602, October 1997.

[GC98] A. J. Goldsmith and S.-G. Chua. Adaptive coded modulation for fading channels. *IEEE Transactions on Communications*, 46:595–602, May 1998.

[GE05] D. Gündüz and E. Erkip. Source and channel coding for quasi-static fading channels. In *Proc. Asilomar Conf. Signals, Systems, Computers*, pages 18–22, Pacific Grove, CA, November 2005.

[GE07a] D. Gündüz and E. Erkip. Opportunistic cooperation by dynamic resource allocation. *IEEE Transactions on Wireless Communications*, 6:1446–1454, April 2007.

[GE07b] D. Gündüz and E. Erkip. Source and channel coding for cooperative relaying. *IEEE Transactions on Information Theory*, 53:3454–3475, October 2007.

[GE08] D. Gündüz and E. Erkip. Joint source channel codes for MIMO block fading channels. *IEEE Transactions on Information Theory*, 54:116–134, January 2008.

[GG92] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression.* Kluwer Academic, Norwell, MA, 1992.

[GMW81] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization.* Academic Press, London, UK, 1981.

[GØH05] A. Gjendemsjø, G. E. Øien, and H. Holm. Optimal power control for discrete-rate link adaptation schemes with capacity-approaching coding. In *Proc. IEEE Globecom*, pages 3498–3502, November 2005.

[GS07] J. Giese and M. Skoglund. Space-time constellation design for partial CSI at the receiver. *IEEE Transactions on Information Theory*, 53:2715–2731, August 2007.

[GV97] A. J. Goldsmith and P. P. Varaiya. Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory*, 43:1986–1992, November 1997.

[HA04] M. O. Hasna and M.-S. Alouini. Optimal power allocation for relayed transmissions over Rayleigh-fading channels. *IEEE Transactions on Wireless Communications*, 6:1999–2004, November 2004.

[HG05] T. Holliday and A. J. Goldsmith. Optimizing end-to-end distortion in MIMO systems. In *Proc. IEEE Int. Symp. Information Theory*, pages 1671–1675, Adelaide, Australia, September 2005.

[HH02] B. Hassibi and B. Hochwald. High-rate codes that are linear in space and time. *IEEE Transactions on Information Theory*, 48:18 04–1824, July 2002.

[HH03] B. Hassibi and B. M. Hochwald. How much training is needed in multiple-antenna wireless links? *IEEE Transactions on Information Theory*, 49:951–963, April 2003.

[HMT04] B. M. Hochwald, T. L. Marzetta, and V. Tarokh. Multiple-antenna channel hardening and its implications for rate feedback and scheduling. *IEEE Transactions on Information Theory*, 50:1893–1909, September 2004.

[HT98] S. V. Hanly and D. N. C. Tse. Multiacess fading channels - Part II: Delay-limited capacities. *IEEE Transactions on Information Theory*, 44:2816–2831, November 1998.

[HtB03] B. M. Hochwald and S. ten Brink. Achieving near-capacity on a multiple-antenna channel. *IEEE Transactions on Communications*, 51:389–399, March 2003.

[HZ05] A. Høst-Madsen and J. Zhang. Capacity bounds and power allocation for wireless relay channels. *IEEE Transactions on Information Theory*, 51:2020–2040, June 2005.

[Jaf01] H. Jafarkhani. A quasi-orthogonal space-time block code. *IEEE Transactions on Communications*, 49:1–4, January 2001.

[JB04] E.A. Jorswieck and H. Boche. Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback. *IEEE Transactions on Wireless Communications*, 3(5):1543–1553, September 2004.

[JG03] N. Jindal and A. Goldsmith. Capacity and optimal power allocation for fading broadcast channels with minimum rates. *IEEE Transactions on Information Theory*, 49:2895–2909, November 2003.

[JG04] S. A. Jafar and A. Goldsmith. Transmitter optimization and optimality of beamforming for multiple antenna systems. *IEEE Transactions on Wireless Communications*, 3(4):1165–1175, July 2004.

[JO05] J. Jaldén and B. Ottersten. On the complexity of sphere decoding in digital communications. *IEEE Transactions on Signal Processing*, 53:1474–1484, April 2005.

[JS04] G. Jöngren and M. Skoglund. Quantized feedback information in orthogonal space-time block coding. *IEEE Transactions on Information Theory*, 50:2473–2486, October 2004.

[JSO02] G. Jöngren, M. Skoglund, and B. Ottersten. Combining beamforming and orthogonal space-time block coding. *IEEE Transactions on Information Theory*, 48:611–627, March 2002.

[KBLS06] T. T. Kim, M. Bengtsson, E. G. Larsson, and M. Skoglund. Combining short-term and long-term channel state information over correlated MIMO channels. In *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pages IV.13–IV.16, Toulouse, France, May 2006.

[KBLS08] T. T. Kim, M. Bengtsson, E. G. Larsson, and M. Skoglund. Combining long-term and low rate short-term channel state information over correlated MIMO channels. *IEEE Transactions on Wireless Communications*, 2008. To appear.

[KBS07] T. T. Kim, M. Bengtsson, and M. Skoglund. Quantized feedback design for MIMO broadcast channels. In *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pages III.1–III.4, Honolulu, HI, May 2007.

[KCS07a] T. T. Kim, G. Caire, and M. Skoglund. Decode-and-forward relay channels with quantized channel state feedback: An outage exponent analysis. *IEEE Transactions on Information Theory*, May 2007. submitted for publication.

[KCS07b] T. T. Kim, G. Caire, and M. Skoglund. On the outage exponent of fading relay channels with partial channel state information. In *Proc. IEEE Information Theory Workshop*, pages 620–625, Lake Tahoe, CA, September 2007.

[KGG05] G. Kramer, M. Gastpar, and P. Gupta. Cooperative strategies and capacity theorems for relay networks. *IEEE Transactions on Information Theory*, 51:3037–3063, September 2005.

[KJS04a] T. T. Kim, G. Jöngren, and M. Skoglund. On the convergence behavior of weighted space-time bit-interleaved coded modulation. In *Proc. Asilomar Conf. Signals, Systems, Computers*, pages 1257–1261, San Antonio, TX, November 2004.

[KJS04b] T. T. Kim, G. Jöngren, and M. Skoglund. Weighted space-time bit-interleaved coded modulation. In *Proc. IEEE Information Theory Workshop*, pages 375–380, Pacific Grove, CA, October 2004.

[KR05] Kiran T. and B. S. Rajan. STBC-schemes with nonvanishing determinants for certain number of transmit antennas. *IEEE Transactions on Information Theory*, 51:2984–2992, August 2005.

[KS04a] A. Khoshnevis and A. Sabharwal. On diversity and multiplexing gain of multiple antennas systems with transmitter channel information. In *Proc. Annu. Allerton Conf. Communication, Control, Computing*, September 2004.

[KS04b] A. Khoshnevis and A. Sabharwal. Performance of quantized power control in multiple antenna systems. In *Proc. IEEE Int. Conf. Communications*, pages 803–807, Paris, France, June 2004.

[KS05] A. Khoshnevis and A. Sabharwal. Achievable diversity and multiplexing in multiple antenna systems with quantized power control. In *Proc.*

IEEE Int. Conf. Communications, pages 463–467, Seoul, Korea, May 2005.

[KS06a] T. T. Kim and M. Skoglund. Diversity-multiplexing tradeoff of MIMO systems with partial power control. In *Proc. Int. Zurich Seminar on Communications*, pages 106–109, Zurich, Switzerland, February 2006.

[KS06b] T. T. Kim and M. Skoglund. Outage behavior of MIMO channels with partial feedback and minimum multiplexing gains. In *Proc. IEEE Int. Symp. Information Theory*, pages 2694–2698, Seattle, WA, July 2006.

[KS06c] T. T. Kim and M. Skoglund. Partial power control for slowly fading MIMO channels. In *Proc. IEEE Int. Conf. Communications*, Istanbul, Turkey, June 2006.

[KS07] T. T. Kim and M. Skoglund. Diversity-multiplexing tradeoff in MIMO channels with partial CSIT. *IEEE Transactions on Information Theory*, 53:2743–2759, August 2007.

[KSC07a] T. T. Kim, M. Skoglund, and G. Caire. Distortion exponents over fading MIMO channels with quantized feedback. In *Proc. IEEE Int. Symp. Information Theory*, pages 2236–2240, Nice, France, June 2007.

[KSC07b] T. T. Kim, M. Skoglund, and G. Caire. Quantifying the loss of compress-forward relaying without Wyner-Ziv coding. *IEEE Transactions on Information Theory*, October 2007. submitted for publication.

[KSC08a] T. T. Kim, M. Skoglund, and G. Caire. On cooperative source transmission with partial rate and power control. *IEEE Journal on Selected Areas in Communications*, 26, October 2008. to appear.

[KSC08b] T. T. Kim, M. Skoglund, and G. Caire. On source transmission over MIMO channels with limited feedback. *IEEE Transactions on Signal Processing*, March 2008. submitted for publication.

[KW03] C. Köse and R. D. Wesel. Universal space-time trellis codes. *IEEE Transactions on Information Theory*, 49:2717–2727, October 2003.

[LC05] E. G. Larsson and Y. Cao. Collaborative transmit diversity with adaptive radio resource and power allocation. *IEEE Communications Letters*, 9:511–513, June 2005.

[LF00] P. Ligdas and N. Farvardin. Optimizing the transmit power for slow fading channels. *IEEE Transactions on Information Theory*, 46:565–576, March 2000.

[LH05] D. J. Love and R. W. Heath, Jr. Limited feedback unitary precoding for orthogonal space-time block codes. *IEEE Transactions on Signal Processing*, 53:64–73, January 2005.

[LHS03] D. J. Love, R. W. Heath, Jr., and T. Strohmer. Grassmannian beamforming for multiple-input multiple-output wireless systems. *IEEE Transactions on Information Theory*, 49:2735–2747, October 2003.

[LLC04a] V. Lau, Y. Liu, and T.-A. Chen. On the design of MIMO block-fading channels with feedback-link capacity constraint. *IEEE Transactions on Communications*, 52:62–70, January 2004.

[LLC04b] V. K. N. Lau, Y. Liu, and T.-A. Chen. Capacity of memoryless channels and block-fading channels with designable cardinality-constrained channel state feedback. *IEEE Transactions on Information Theory*, 50:2038–2049, September 2004.

[LLTF02] Y. Liu, K. N. Lau, O. Y. Takeshita, and M. P. Fitz. Optimal rate allocation for superposition coding in quasi-static fading channels. In *Proc. IEEE Int. Symp. Information Theory*, page 111, June 2002.

[LLYS03] J. Luo, L. Lin, R. Yates, and P. Spasojević. Service outage based power and rate allocation. *IEEE Transactions on Information Theory*, 49:323–330, January 2003.

[LM03] A. Lapidoth and S. M. Moser. Capacity bounds by duality with applications to multiple-antenna systems on flat-fading channels. *IEEE Transactions on Information Theory*, 49:2426–2467, October 2003.

[LMWA05] J. N. Laneman, E. Martinian, G. W. Wornell, and J. G. Apostolopoulos. Source-channel diversity for parallel channels. *IEEE Transactions on Information Theory*, 51:3518–3539, October 2005.

[LTW04] J. N. Laneman, D. N. C. Tse, and G. W. Wornell. Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Transactions on Information Theory*, 50:3062–3080, December 2004.

[LV05] Y. Liang and V. V. Veeravalli. Gaussian orthogonal relay channels: Optimal resource allocation and capacity. *IEEE Transactions on Information Theory*, 51:3284–3289, September 2005.

[LVP07] Y. Liang, V. V. Veeravalli, and H. V. Poor. Resource allocation for wireless fading relay channels: Max-min solution. *IEEE Transactions on Information Theory*, 53:3432–3453, October 2007.

[LW03] J. N. Laneman and G. W. Wornell. Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks. *IEEE Transactions on Information Theory*, 49:2415–2425, October 2003.

[LYS03] L. Lin, R. Yates, and P. Spasojević. Adaptive transmission with discrete code rates and power levels. *IEEE Transactions on Communications*, 51:2115–2125, December 2003.

[MB02] P. Mitran and J. Bajcsy. Coding for the Wyner-Ziv problem with turbo-like codes. In *Proc. IEEE Int. Symp. Information Theory*, page 91, Lausanne, Switzerland, June 2002.

[MH99] T. L. Marzetta and B. M. Hochwald. Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading. *IEEE Transactions on Information Theory*, 45:139–157, January 1999.

[MP02] U. Mittal and N. Phamdo. Hybrid digital-analog (HDA) joint source-channel codes for broadcasting and robust communications. *IEEE Transactions on Information Theory*, 48:1082–1102, May 2002.

[MSEA03] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang. On beam-forming with finite rate feedback in multiple-antenna systems. *IEEE Transactions on Information Theory*, 49:2562–2579, October 2003.

[MY04] I. Marić and R. D. Yates. Bandwidth and power allocation for cooperative strategies in Gaussian relay networks. In *Proc. Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, November 2004.

[NBK04] R. U. Nabar, H. Bölcskei, and F. W. Kneubuhler. Fading relay channels: Performance limits and space-time signal design. *IEEE Journal on Selected Areas in Communications*, 22:1099–1109, August 2004.

[NC02] R. Negi and J. Cioffi. Delay-constrained capacity with causal feedback. *IEEE Transactions on Information Theory*, 48:2478–2494, September 2002.

[NLTW98] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell. Efficient use of side information in multiple-antenna data transmission over fading channels. *IEEE Journal on Selected Areas in Communications*, 16:1423–1436, October 1998.

[NN95] S. Na and D. L. Neuhoff. Bennett's integral for vector quantizers. *IEEE Transactions on Information Theory*, 41:886–900, July 1995.

[ORBV06] F. Oggier, G. Rekaya, J C. Belfiore, and E. Viterbo. Perfect space-time block codes. *IEEE Transactions on Information Theory*, 52:3885–3902, September 2006.

[OSW94] L. H. Ozarow, S. Shamai (Shitz), and A. D. Wyner. Information theoretic considerations for cellular mobile radio. *IEEE Transactions on Vehicular Technology*, 43:359–378, May 1994.

[Pro95] J. G. Proakis. *Digital Communications*. McGraw-Hill, New York, NY, 1995.

[PV04] N. Prasad and M. K. Varanasi. Diversity and multiplexing tradeoff bounds for cooperative diversity protocols. In *Proc. IEEE Int. Symp. Information Theory*, page 268, Chicago, IL, 2004.

[RBV04] G. Rekaya, J.-C. Belfiore, and E. Viterbo. Algebraic 3×3, 4×4 and 6×6 space-time codes with non-vanishing determinants. In *Proc. Int. Symp. Information Theory and its Applications*, pages 325–329, October 2004.

[RV68] W. L. Root and P. P. Varaiya. Capacity of classes of Gaussian channels. *SIAM J. Appl. Math.*, 16:1350–1393, November 1968.

[SD01] A. Stefanov and T. M. Duman. Turbo-coded modulation for systems with transmit and receive antenna diversity over block fading channels: System model, decoding approaches, and practical considerations. *IEEE Journal on Selected Areas in Communications*, 19:958–968, May 2001.

[SEA03a] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity - Part I: System description. *IEEE Transactions on Communications*, 51:1927–1938, November 2003.

[SEA03b] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity - Part II: Implementation aspects and performance analysis. *IEEE Transactions on Communications*, 51:1939–1948, November 2003.

[Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical J.*, 27:379–423 and 623–656, Jul. and Oct. 1948.

[Sha97] S. Shamai (Shitz). A broadcast strategy for the Gaussian slowly fading channel. In *Proc. IEEE Int. Symp. Information Theory*, page 150, June 1997.

[SJ03] M. Skoglund and G. Jöngren. On the capacity of a multiple-antenna communication link with channel side information. *IEEE Journal on Selected Areas in Communications*, 21:395–405, April 2003.

[SPA02] M. Skoglund, N. Phamdo, and F. Alajaji. Design and performance of VQ-based hybrid digital-analog joint source-channel codes. *IEEE Transactions on Information Theory*, 48(3):708–720, March 2002.

[SPA06] M. Skoglund, N. Phamdo, and F. Alajaji. Hybrid digital-analog source-channel coding for bandwidth compression/expansion. *IEEE Transactions on Information Theory*, 52(8):3757–3763, August 2006.

[SS03] S. Shamai (Shitz) and A. Steiner. A broacast approach for a single-user slowly fading MIMO channel. *IEEE Transactions on Information Theory*, 49:2617–2635, October 2003.

[SV97] S. Shamai (Shitz) and S. Verdú. The empirical distribution of good codes. *IEEE Transactions on Information Theory*, 43:836–846, May 1997.

[Tel99] İ. E. Telatar. Capacity of multiple-antenna Gaussian channels. *Eur. Trans. Telecommun.*, 10:585–595, November 1999.

[TJC99] V. Tarokh, H. Jafarkhani, and A. R. Calderbank. Space-time block codes from orthogonal designs. *IEEE Transactions on Information Theory*, 45:1456–1467, July 1999.

[tK03] S. ten Brink and G. Kramer. Design of repeat-accumulate codes for iterative detection and decoding. *IEEE Transactions on Signal Processing*, 51:2764–2772, November 2003.

[tKA04] S. ten Brink, G. Kramer, and A. Ashikhmin. Design of low-density parity-check codes for modulation and detection. *IEEE Transactions on Communications*, 52:670–678, April 2004.

[TSC98] V. Tarokh, N. Seshadri, and A. R. Calderbank. Space-time codes for high data rate wireless communication: Performance criterion and code construction. *IEEE Transactions on Information Theory*, 44:744–765, March 1998.

[Tse97] D. N. C. Tse. Optimal power allocation over parallel Gaussian broadcast channels. In *Proc. IEEE Int. Symp. Information Theory*, page 27, June 1997.

[TV05] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge Univ. Press, Cambridge, UK, 2005.

[TV06] S. Tavildar and P. Viswanath. Approximately universal codes over slow fading channels. *IEEE Transactions on Information Theory*, 52:3233–3258, July 2006.

[van71] E. C. van der Meulen. Three-terminal communication channels. *Adv. Appl. Prob.*, 3:120–154, 1971.

[VG03] S. Vishwanath and A. Goldsmith. Adaptive turbo-coded modulation for flat-fading channels. *IEEE Transactions on Communications*, 51:964–972, June 2003.

[VH94] S. Verdú and T. S. Han. A general formula for channel capacity. *IEEE Transactions on Information Theory*, 40:1147–1157, July 1994.

[VM01] E. Visotsky and U. Madhow. Space-time transmit precoding with imperfect feedback. *IEEE Transactions on Information Theory*, 47:2632–2639, September 2001.

[VP06] M. Vu and A. Paulraj. Optimal linear precoders for MIMO wireless correlated channels with non-zero mean in space-time coded systems. *IEEE Transactions on Signal Processing*, 54:2318–2332, June 2006.

[WO01] X. Wang and M. T. Orchard. Design of trellis codes for source coding with side information at the decoder. In *Proc. Data Compression Conf.*, pages 361–370, Snowbird, UT, March 2001.

[Wyn78] A. Wyner. The rate-distortion function for source coding with side information at the decoder-II: General sources. *Inf. Control*, 38:60–80, 1978.

[WZ76] A. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, 22:1–10, January 1976.

[YB07a] S. Yang and J.-C. Belfiore. Optimal space-time codes for the MIMO amplify-and-forward cooperative channel. *IEEE Transactions on Information Theory*, 53:647–663, February 2007.

[YB07b] S. Yang and J.-C. Belfiore. Towards the optimal amplify-and-forward cooperative diversity scheme. *IEEE Transactions on Information Theory*, 53:3114–3126, September 2007.

[YCXZ03] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao. Wyner-Ziv coding based on TCQ and LDPC codes. In *Proc. Asilomar Conf. Signals, Systems, Computers*, pages 825–829, Pacific Grove, CA, November 2003.

[YE07] M. Yuksel and E. Erkip. Multiple-antenna cooperative wireless systems: A diversity-multiplexing tradeoff perspective. *IEEE Transactions on Information Theory*, 53:3371–3393, October 2007.

[YW03] H. Yao and G. W. Wornell. Achieving the full MIMO diversity-multiplexing frontier with rotation-based space-time codes. In *Proc. Annu. Allerton Conf. Communication, Control, Computing*, October 2003.

[ZT02] L. Zheng and D. N. C. Tse. Communication on the Grassman manifold: A geometric approach to the noncoherent multiantenna channels. *IEEE Transactions on Information Theory*, 48:359–383, February 2002.

[ZT03] L. Zheng and D. N. C. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels. *IEEE Transactions on Information Theory*, 49:1073–1096, May 2003.