BROADCAST STRATEGY FOR DELAY-LIMITED COMMUNICATION OVER FADING CHANNELS

A Dissertation

by

JAE WON YOO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Tie Liu
Committee Members,	Chanan Singh
	Henry D. Pfister
	Anxiao (Andrew) Jiang
Department Head,	Chanan Singh

May 2013

Major Subject: Electrical and Computer Engineering

ABSTRACT

Delay is an important quality-of-service measure for the design of next-generation wireless networks. This dissertation considers the problem of delay-limited communication over block-fading channels, where the channel state information is available at the receiver but not at the transmitter. For this communication scenario, the difference between the ergodic capacity and the maximum achievable expected rate (the expected capacity) for coding over a finite number of coherent blocks represents a fundamental measure of the penalty incurred by the delay constraint.

This dissertation introduces a notion of worst-case expected-capacity loss. Focusing on the slow-fading scenario (one-block delay), the worst-case additive and multiplicative expected-capacity losses are precisely characterized for the point-to-point fading channel. Extension to the problem of writing on fading paper is also considered, where both the ergodic capacity and the additive expected-capacity loss over one-block delay are characterized to within one bit per channel use.

The problem with multiple-block delay is considerably more challenging. This dissertation presents two partial results. First, the expected capacity is precisely characterized for the point-to-point two-state fading channel with two-block delay. Second, the optimality of Gaussian superposition coding with indirect decoding is established for a two-parallel Gaussian broadcast channel with three receivers. Both results reveal some intrinsic complexity in characterizing the expected capacity with multiple-block delay.

DEDICATION

To my parents, wife, and sister

ACKNOWLEDGEMENTS

First and foremost, I would like to express my heartfelt gratitude to my advisor Professor Tie Liu for the patient guidance, encouragement, and support that he has provided me throughout my Ph.D. career. Without his contributions, completing this dissertation would not have been possible. Many valuable things that I have learned from Dr. Liu including his philosophies of research and life will be a great help for my future career. It has been a great pleasure having him as a supervisor.

I would also like to thank my dissertation committee, Professor Chanan Singh, Professor Henry D. Pfister, and Professor Anxiao (Andrew) Jiang for their support and guidance. I am deeply grateful to Professor Shlomo Shamai (Shitz) at Technion-Israel Institute of Technology, Dr. Chao Tian at AT&T Labs-Research, Professor Yingbin Liang at Syracuse University, and Dr. Feng Xue at Intel Research for working with me during my Ph.D. research.

I would like to extend my thanks to Jinjing Jiang, Hung D. Ly, Neeharika Marukala, Richa Dixit, Mustafa El-Halabi, Shuo Li, Amir Salimi, Jerry Huang, Sang Woo Park, Byung-Hak Kim, Dae Hyun Choi, Lucia Sui, Luoyi Fu, Xinyi Cai, Sabit Ekin, Lili Zhang, and Armin Banaei, who have made my experience at Texas A&M enjoyable. I am indebted to all my friends in Colleges Station and South Korea who have kept me sane and happy during my Ph.D. journey. Special thanks to Sungwook Hong, Jaewon Park, and Hyun Soo Kim for the valuable memories we shared.

Last but not least, I would like to give my sincere thanks to my family. Without their constant support and unconditional love, I would have never been able to complete my Ph.D. degree. I dedicate this dissertation to them.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

1. INTRODUCTION

1.1 Motivation

In recent years, there has been an explosive increase in demands on the wireless services with stringent *quality-of-service* requirements along with the rapid evolution of wireless access technologies. This trend can be found in various wireless applications in our lives such as the real-time multimedia streaming on mobile devices. One of the key measures of quality-of-service is *delay*. With delay limitation, wireless channels may be faced with the capacity loss, which is mainly due to the time-varying nature of wireless channels, so called fading. Understanding the impact of delay constraints on the overall performance of wireless channels is an interesting subject in information theory.

Consider the discrete-time baseband representation of the single-user flat-fading channel:

$$Y[t] = \sqrt{G[t]}X[t] + Z[t] \tag{1.1}$$

where $\{X[t]\}$ are the channel inputs which are subject to a unit average power constraint, $\{G[t]\}$ are the power gains of the channel fading which we assume to be be *unknown* to the transmitter but known at the receiver, $\{Z[t]\}$ are the additive white circularly symmetric complex Gaussian noise with zero means and unit variances, and $\{Y[t]\}$ are the channel outputs. As often done in the literature, we shall consider the so-called *block-fading* model [1, Ch. 5.4.5] where $\{G[t]\}$ are assumed to be *constant* within each coherent block and change *independently* across different blocks according to a known distribution $F_G(\cdot)$. The coherent time of the channel is assumed to be large so that the additive noise $\{Z[t]\}$ can be "averaged out" within each coherent block. Since both the power constraint and the noise variances are normalized to one, the power gain $G[t]$ also represents the instantaneous *receive* signal-to-noise ratio of the channel.

The focus of this dissertation is on *delay-limited* communication for which communication is only allowed to span (at most) a total of L coherent blocks where L is a *finite* integer. In this setting, the Shannon capacity is a very pessimistic measure as it is dictated by the *worst* realization of the power-gain process and hence equals zero when the realization of the power gain can be arbitrarily close to zero. An often-adopted measure in the literature is the *expected capacity* [2–6], which is defined as the maximum *expected* reliably decoded rate where the expectation is over the distribution of the power-gain process.

The problem of characterizing the expected capacity is closely related to the problem of broadcasting over linear Gaussian channels [2–6]. The case with $L = 1$ represents the most stringent delay requirement known as *slow fading* [1, Ch. 5.4.1]. For slow-fading channels, the problem of characterizing the expected capacity is equivalent to the problem of characterizing the capacity region of a *scalar* Gaussian broadcast channel, which is well understood based on the classical works of Cover [7] and Bergmans [8], and then finding an optimal rate allocation based on the power-gain distribution. For $L > 1$, the expected capacity can be improved by treating each realization of the power-gain process as a user in an *L-parallel* Gaussian broadcast channel and coding the information bits across different sub-channels [3, 9, 10]. In the limit as $L \to \infty$, by the ergodicity of the power-gain process each "typical" realization of the power-gain process can support a reliable rate of communication which is arbitrarily close to

$$C_{erg}(F_G) = \mathbb{E}_G[\log(1 + G)]. \tag{1.2}$$

Thus, $C_{erg}(F_G)$ is both the Shannon capacity (appropriately known as the *ergodic capacity* [1, Ch. 5.4.5]) and the expected capacity in the limit as $L \to \infty$.

Formally, let us denote by $C_{exp}(F_G, L)$ the expected capacity of the block-fading channel (1.1) for which the power-gain distribution is $F_G(\cdot)$, and communication is allowed to span (at most) a total of L coherent blocks. Then, as mentioned previously, the expected capacity $C_{exp}(F_G, L) \to C_{erg}(F_G)$ in the limit as $L \to \infty$. As such, the "gap" between the ergodic capacity $C_{erg}(F_G)$ and the expected capacity $C_{exp}(F_G, L)$ represents a fundamental measure of the penalty incurred by imposing a delay constraint of L coherent blocks. Such gaps, naturally, would depend on the underlying power-gain distribution. To be more general, we are interested in characterizing the *worst-case* gaps over all possible power-gain distributions (including both the power-gain realizations and the probabilities for each realization) with a fixed number of different possible realizations of the power gain in each coherent block.

In this dissertation, the impact of delay is investigated in several channel settings.

- *Worst-cast expected capacity loss for one-block delay.* Motivated by the recent trend on wireless applications, the most stringent delay constraint $L = 1$ is considered. For this slow-fading scenario ($L = 1$), the *broadcast strategy* [3] provides the expected capacity $C_{exp}(F_G, 1)$ as a power allocation problem. Investigating the power allocation problem, our focus is on precise characterizations of the *worst-case additive* and *multiplicative* gaps between the ergodic capacity $C_{erg}(F_G)$ and the expected capacity $C_{exp}(F_G, 1)$.

- *Writing on block-fading paper.* Here, an extension of the result for slow-fading scenario to the problem of writing on block-fading paper [11–13] is considered. For block-fading paper setting, the ergodic capacity remains unknown. Our goal here is to characterize the ergodic capacity within a finite number of bits per channel usage via an appropriate coding structure and to approximate the *worst-case* capacity loss.

- *Two-block delay.* When the delay requirement is more than one coherent block ($L > 1$), the expected capacity $C_{exp}(F_G, L)$ is in general unknown. The main challenge there is on characterizing the capacity region of the L-parallel Gaussian broadcast channel with a *general* message set configuration. To shed some light on the problem with multiple-block delay, two different scenarios with two-block delay are considered. One is the point-to-point two-state block fading channel with two-block delay, which is considered in [9]. Our focus here is to establish a precise characterization of the expected capacity of the channel. Next, we consider a two-parallel Gaussian broadcast channel with three receivers, which is related to multiple-state fading channels with two-block delay. We focus on characterizing the entire capacity region by establishing the optimality of Gaussian signaling along with the indirect decoding [14].

1.2 Dissertation Outline

The rest of the dissertation is organized as follows. Next in Chapter 2, the worst-case gaps for one-block delay are precisely characterized. Key to the proof of the worst-case gap results is an explicit characterization of an optimal power allocation for characterizing the expected capacity $C_{exp}(F_G, 1)$, obtained via the marginal utility functions introduced by Tse [15]. In Chapter 3, we extend the setting from the point-to-point fading channel to the problem of writing on fading paper [11–13], and provide a characterization of the ergodic capacity and the additive expected-capacity loss over one-block delay to within one bit per channel use. In Chapter 4, the expected capacity of the point-to-point two-state fading channel with two-block delay is precisely characterized with an optimal power allocation. In Chapter 5, the capacity region of a two-parallel Gaussian broadcast channel with degraded message sets is precisely characterized. The characterization is based on optimality of Gaussian signaling along with the indirect decoding [14]. In Chapter 6, we conclude the dissertation with some remarks.

2. WORST-CASE EXPECTED CAPACITY LOSS FOR ONE-BLOCK DELAY*

2.1 Introduction

Consider the *block-fading* model (1.1) with the delay constraint of L coherent blocks. As described in Chapter 1, the gap between the ergodic capacity $C_{erg}(F_G)$ and the expected capacity $C_{exp}(F_G, L)$ represents a fundamental measure of the penalty incurred by imposing a delay constraint of L coherent blocks. Obviously, such gaps have strong dependencies on the underlying power-gain distribution. To have more general understanding on the penalty, we consider the *worst-case* gaps over all possible power-gain distributions with a fixed number of different possible realizations of the power gain in each coherent block.

More specifically, for the block-fading channel (1.1) with the power-gain distribution $F_G(\cdot)$, let us define the *additive* and the *multiplicative* gap between the ergodic capacity and the expected capacity under the delay constraint of L coherent blocks as

$$A(F_G, L) := C_{erg}(F_G) - C_{exp}(F_G, L) \tag{2.1}$$

and

$$M(F_G, L) := \frac{C_{erg}(F_G)}{C_{exp}(F_G, L)} \tag{2.2}$$

respectively. Focusing on the slow-fading scenario ($L = 1$), we have the following precise characterization of the *worst-case* additive and multiplicative gaps between the ergodic capacity and the expected capacity.

*Reprinted, with permission, from J. W. Yoo, T. Liu, S. Shamai (Shitz), and C. Tian,“Worst-Case Expected-Capacity Loss of Slow-Fading Channels,” *IEEE Transactions on Information Theory*, to appear, Copyright 2013 IEEE.

Theorem 1.

$$\sup_{F_G} A(F_G, L) = \log K \tag{2.3}$$

and

$$\sup_{F_G} M(F_G, L) = K \tag{2.4}$$

where the supremes are over all power-gain distribution $F_G(\cdot)$ with K different possible realizations of the power gain in each coherent block.

The above results have both positive and negative engineering implications, which are summarized below.

- On the positive side, note that both the ergodic capacity $C_{erg}(F_G)$ and the expected capacity $C_{exp}(F_G, 1)$ will generally grow *unboundedly* in the limit as the realizations of the power gain all tend to infinity. The difference between them, however, will remain *bounded* for any *finite-state* fading channels (where K is *finite*). Similarly, both the ergodic capacity $C_{erg}(F_G)$ and the expected capacity $C_{exp}(F_G, 1)$ will *vanish* in the limit as the realizations of the power gain all tend to zero. However, the expected capacity $C_{exp}(F_G, 1)$ (under the most stringent delay constraint of $L = 1$ coherent block) can account, at least, for a *non-vanishing* fraction of the ergodic capacity $C_{erg}(F_G)$.

- On the negative side, in the worst-case scenario both the additive gap $A(F_G, 1)$ and the multiplicative gap $M(F_G, 1)$ will grow *unboundedly* in the limit as the number of different realizations of the power gain in each coherent block $K \to \infty$. Therefore, when K is large, delay-limited communication may incur a large expected-rate loss relative to the ergodic scenario where there is no delay constraint on communication. For *continuous-fading* channels where the sample space of $F_G(\cdot)$ is infinite and uncountable, it is also possible that the expected-rate loss incurred by delay constraints is *unbounded*.

On the other hand, one should not be *overly* pessimistic when attempt to interpret the worst-case gap results (2.1) and (2.2). First, the above worst-case gap results are derived under the assumption that the transmitter does not know the realization of the channel fading at all. In practice, however, it is entirely possible that some information on the channel fading realization is made available to the transmitter (via finite-rate feedback, for example). This information can be potentially used to reduce the gap between the ergodic capacity and the expected capacity [16, 17]. Second, for specific fading distributions the gap between the ergodic capacity and the expected capacity can be much smaller. For example, it is known [3] that for Rayleigh fading, the additive gap between the ergodic capacity and the expected capacity over one-block delay is only 1.649 nats per channel use in the high signal-to-noise ratio limit, and the multiplicative gap is only 1.718 in the low signal-to-noise ratio limit, even though in this case the power-gain distribution is continuous.

2.2 Optimal Power Allocation via Marginal Utility Functions

To prove the worst-case gap results (2.1) and (2.2) as stated in Theorem 1, let us fix the transmit signal-to-noise ratio 1 and the power-gain distribution $F_G(\cdot)$ with K different possible realizations of the power gain in each coherent block. Let $g_1, \ldots, g_K$ be the collection of the possible realizations of the power gain, and let $p_k := Pr(G = g_k) > 0$. Without loss of generality, let us assume that the possible realizations of the power gain are ordered as

$$g_1 > g_2 > \cdots > g_K \geq 0. \tag{2.5}$$

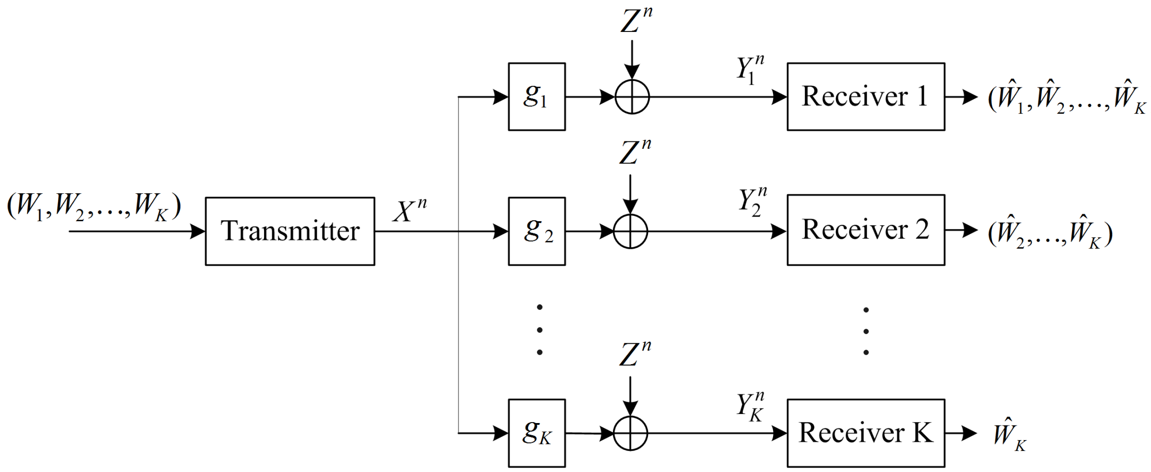

Figure 2.1: A scalar Gaussian broadcast channel with degraded massage set.

With the above notations, the expected capacity $C_{exp}(F_G, 1)$ (under the delay constraint of $L = 1$ coherent block) is given by the maximum weighted sum-rate of the scalar Gaussian broadcast channel with degraded massage sets in Figure 2.1 [3]:

$$\begin{array}{rl} \max_{(\beta_1,\ldots,\beta_K)} & \sum_{k=1}^{K} F_k \log\left(\frac{1+\beta_k g_k}{1+\beta_{k-1} g_k}\right) \\ \text{subject to} & 0 = \beta_0 \leq \beta_1 \leq \cdots \leq \beta_K \leq 1 \end{array} \tag{2.6}$$

where

$$F_k := \sum_{j=1}^{k} p_j. \tag{2.7}$$

Note that the optimization program (2.6) with respect to the cumulative power fractions $(\beta_1, \ldots, \beta_K)$ is *not* convex. However, the program can be convexified via the following simple change of variable [15, 18]

$$r_k := \log\left(\frac{1+\beta_k g_k}{1+\beta_{k-1} g_k}\right), \quad k = 1, \ldots, K. \tag{2.8}$$

In the preliminary version of this work [19], this venue was further pursued to obtain an *implicit* characterization of the optimal power allocation via the standard Karush-

Kuhn-Tucker conditions. Below we shall consider an alternative and more direct approach which provides an *explicit* characterization of an optimal power allocation via the *marginal utility functions (MUFs)* introduced by Tse [15].

Assume that $g_K > 0$ (which implies that $g_k > 0$ for all $k = 1, \ldots, K$), and let $n_k := 1/g_k$ for $k = 1, \ldots, K$. Given the assumed ordering (2.5) for the power-gain realizations $\{g1, \ldots, gK\}$, we have

$$0 < n_1 < \cdots < n_K. \tag{2.9}$$

Following [15], let us define the MUFs and the *dominating* MUF as

$$u_k(z) := \frac{F_k}{n_k + z}, \quad k = 1, \ldots, K \tag{2.10}$$

and

$$u^*(z) := \max_{k=1,\ldots,K} u_k(z) \tag{2.11}$$

respectively. Note that for any $k = 1, \ldots, K$, $u_k(z) > 0$ if and only if $z > -n_k$. Also, for any two distinct integers k and l such that $1 \leq k \leq l \leq K$, the MUFs $u_k(z)$ and $u_l(z)$ has a unique intersection at $z = z_{k,l}$ where

$$\frac{F_k}{n_k + z_{k,l}} = \frac{F_l}{n_l + z_{k,l}} \Longleftrightarrow z_{k,l} = \frac{F_k n_l - F_l n_k}{F_l - F_k}. \tag{2.12}$$

Note that $F_k < F_l$ and $n_k < n_l$, so we have $z_{k,l} > n_k$. Furthermore, it is straightforward to verify that $u_k(z) > u_l(z) > 0$ if and only if $-n_k < z < z_{k,l}$, and $u_l(z) > u_k(z) > 0$ if and only if $z > z_{k,l}$ (see Figure 2.2 for an illustration). For the rest of this dissertation, the above property will be frequently referred to as the *single crossing point property* of the MUFs.

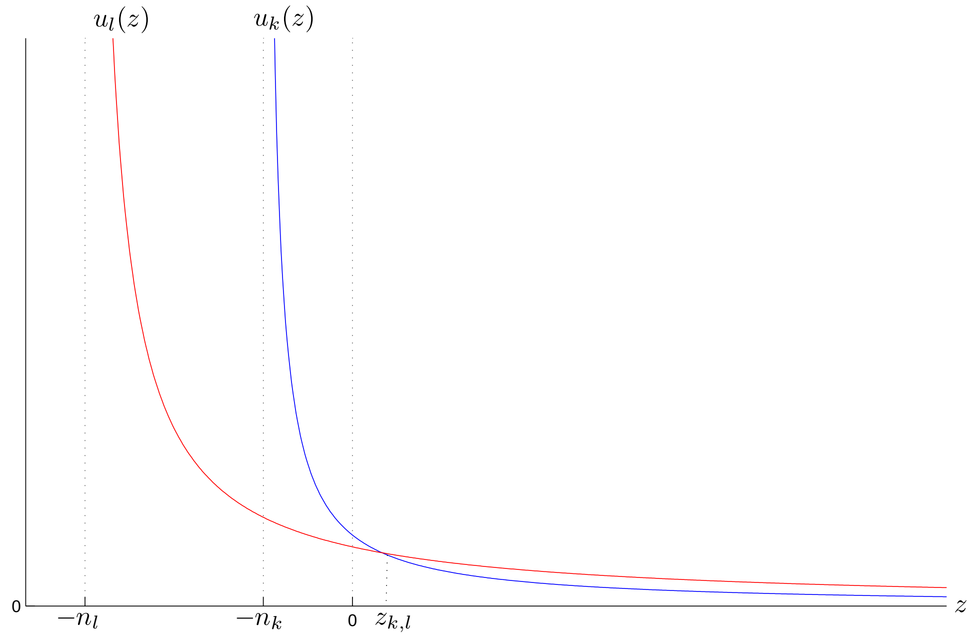

Figure 2.2: The single crossing point property between the MUFs $u_k(z)$ and $u_l(z)$ for $k < l$.

We emphasize here that the aforementioned single crossing point property relies on the fact that both sequences $\{n_k\}$ and $\{F_k\}$ increase monotonically with the subscript k. Since this particular ordering was not specifically considered in the MUFs defined in [15, Eq. (7)], next, instead of building on the results from [15], we shall borrow the concept of MUF and establish our results from first principles. Let us begin by defining a sequence of integers $\{\pi_1, \ldots, \pi_I\}$ recursively as follows.

Definition 1. *First, let* $\pi_1 = 1$. *Then, define*

$$\pi_{i+1} := \max \left[\arg \min_{l=\pi_i+1,\ldots,K} z_{\pi_i,l} \right], \quad i = 1, \ldots, I-1 \tag{2.13}$$

where I *is the total number of integers* $\{\pi_i\}$ *defined through the above recursive procedure.*

Note that in the above definition, a “max” is used to break the ties for achieving the “min” inside the brackets, so there is no ambiguity in defining the integer sequence $\{\pi_1, \ldots, \pi_I\}$. Clearly, we have

$$1 = \pi_1 < \pi_2 < \cdots < \pi_I = K. \tag{2.14}$$

Furthermore, we have the following properties for the sequence $\{z_{\pi_1,\pi_2}, z_{\pi_2,\pi_3}, \ldots, z_{\pi_{I-1},\pi_I}\}$, which are direct consequences of the recursive definition (2.13) and the single crossing point property of the MUFs.

Lemma 1. *1. For any $i = 1, \ldots, I-1$ and any $l = \pi_i + 1, \ldots, K$, we have*

$$z_{\pi_i,\pi_{i+1}} \leq z_{\pi_i,\pi_l}. \tag{2.15}$$

2. For any $i = 1, \ldots, I-2$, we have

$$z_{\pi_i,\pi_{i+1}} \leq z_{\pi_{i+1},\pi_{i+2}}. \tag{2.16}$$

3. For any $i = 1, \ldots, I-1$ and any $l = 1, \ldots, \pi_{i+1} - 1$, we have

$$z_{\pi_i,\pi_{i+1}} \geq z_{\pi_l,\pi_{i+1}}. \tag{2.17}$$

Proof. Property 1) follows directly from the recursive definition (2.13).

To prove property 2), let us consider proof by contradiction. Assume that $z_{\pi_i,\pi_{i+1}} > z_{\pi_{i+1},\pi_{i+2}}$ for some $i \in \{1, \ldots, I-2\}$. By property 1), we have $z_{\pi_i,\pi_{i+2}} \geq z_{\pi_i,\pi_{i+1}} > z_{\pi_{i+1},\pi_{i+2}}$. Following the single crossing point property, we have $0 < u_{\pi_{i+1}}(z_{\pi_i,\pi_{i+2}}) < u_{\pi_{i+2}}(z_{\pi_i,\pi_{i+2}}) = u_{\pi_i}(z_{\pi_i,\pi_{i+2}})$. Using again the single crossing point property, we may conclude that $-n_{\pi_i} < z_{\pi_i,\pi_{i+2}} < z_{\pi_i,\pi_{i+1}}$. But this contradicts the fact that $z_{\pi_i,\pi_{i+2}} \geq z_{\pi_i,\pi_{i+1}}$ as mentioned previously. This proves that for any $i = 1, \ldots, I-2$, we must have $z_{\pi_i,\pi_{i+1}} \leq z_{\pi_{i+1},\pi_{i+2}}$.

To prove property 3), let us fix $i \in \{1, \ldots, I-1\}$. Note that the desired inequality (2.17) holds trivially with equality for $l = \pi_i$, so we only need to consider the cases where $l \in \{\pi_i + 1, \ldots, \pi_{i+1} - 1\}$ and $l \in \{1, \ldots, \pi_i - 1\}$.

For the case where $l \in \{\pi_i + 1, \ldots, \pi_{i+1} - 1\}$, by property 1) we have $-n_{\pi_i} < z_{\pi_i,\pi_{i+1}} \leq z_{\pi_i,\pi_l}$. Following the single crossing point property we have $0 < u_l(z_{\pi_i,\pi_{i+1}}) \leq u_{\pi_i}(z_{\pi_i,\pi_{i+1}}) = u_{\pi_{i+1}}(z_{\pi_i,\pi_{i+1}})$, which in turn implies that $z_{\pi_i,\pi_{i+1}} \geq z_{l,\pi_{i+1}}$.

For the case where $l \in \{1, \ldots, \pi_i - 1\}$, let us assume, without loss of generality, that $l \in \{\pi_m, \ldots, \pi_{m+1} - 1\}$ for some $m \in \{1, \ldots, i-1\}$. By the previous case we have $z_{\pi_m,\pi_{m+1}} \geq z_{l,\pi_{m+1}}$ and hence

$$0 < u_l(z) \leq u_{\pi_{m+1}}(z) \quad \forall z \geq z_{\pi_m,\pi_{m+1}}. \tag{2.18}$$

Also note that

$$u_{\pi_{m+1}}(z) \leq u_{\pi_{m+2}}(z) \leq \cdots \leq u_{\pi_{i+1}}(z) \quad \forall z \geq \max_{m+1 \leq j \leq i} z_{\pi_j,\pi_{j+1}}. \tag{2.19}$$

By property 2) we have

$$\max_{m+1 \leq j \leq i} z_{\pi_j,\pi_{j+1}} = z_{\pi_i,\pi_{i+1}} \geq z_{\pi_m,\pi_{m+1}}. \tag{2.20}$$

Combining (2.18)-(2.20) gives $0 < u_l(z_{\pi_i,\pi_{i+1}}) \leq u_{\pi_{i+1}}(z_{\pi_i,\pi_{i+1}})$, which in turn implies that $z_{\pi_i,\pi_{i+1}} \geq z_{l,\pi_{i+1}}$.

Combing the above two cases completes the proof of property 3) and hence the entire lemma. □

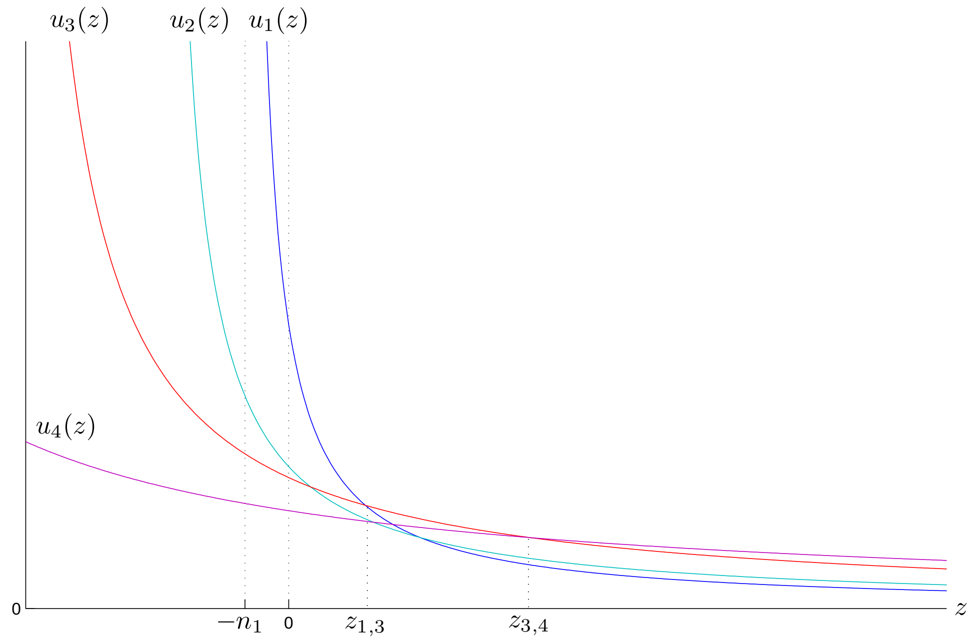

Figure 2.3: An illustration of the dominating MUF. In this example, we have $K = 4$ and $z_{1,3} < z_{1,2} < z_{1,4}$. Therefore, we have $I = 3$, $\pi_1 = 1$, $\pi_2 = 3$, and $\pi_3 = 4$. The dominating MUF $u^*(z) = u_1(z)$ for $z \in (-n_1, z_{1,3})$, $u^*(z) = u_3(z)$ for $z \in (z_{1,3}, z_{3,4})$, and $u^*(z) = u_4(z)$ for $z \in (z_{3,4}, \infty)$.

The following proposition provides an explicit characterization of the dominating MUF (see Figure 2.3 for an illustration).

Proposition 1 (Dominating marginal utility function)**.** *For any* $i = 1, \ldots, I$ *and any* $z \in (z_{\pi_{i-1},\pi_i}, z_{\pi_i,\pi_{i+1}})$*, the dominating MUF*

$$u^*(z) = u_{\pi_i}(z). \tag{2.21}$$

where we define $z_{\pi_0,\pi_1} := -n_1$ *and* $z_{\pi_I,\pi_{I+1}} := \infty$ *for notational convenience (even though* π_0 *and* π_{I+1} *will not be explicitly defined).*

Proof. Fix $i \in \{1, \ldots, I\}$. Let us show that $u_{\pi_i}(z) \geq u_l(z)$ for any $z \in (z_{\pi_{i-1},\pi_i}, z_{\pi_i,\pi_{i+1}})$ by considering the cases $l > \pi_i$ and $l < \pi_i$ separately.

For $l > \pi_i$, by the single crossing point property we have $0 < u_l(z) \leq u_{\pi_i}(z)$ for any $-n_{\pi_i} < z \leq z_{\pi_i,l}$. By property 1) of Lemma 1, for any $l > \pi_i$ we have $z_{\pi_i,\pi_{i+1}} \leq z_{\pi_i,l}$. Combined with the fact that $z_{\pi_{i-1},\pi_i} \geq -n_{\pi_i}$ (the equality holds only when $i = 1$ by the definition of z_{π_0,π_1} and the fact that $\pi_1 = 1$), we may conclude that for $l > \pi_i$, $u_{\pi_i}(z) \geq u_l(z)$ for any $z \in (z_{\pi_{i-1},\pi_i}, z_{\pi_i,\pi_{i+1}}]$.

For $l < \pi_i$, by property 3) of Lemma 1 we have $z_{\pi_{i-1},\pi_i} \geq z_{l,\pi_i}$ and hence $0 < u_l(z) \leq u_{\pi_i}(z)$ for any $z \geq z_{\pi_{i-1},\pi_i}$.

Combining the above two cases completes the proof of the proposition. □

Now, let $(\beta_1^*, \ldots, \beta_K^*)$ be an optimal solution to the optimization program (2.6). Then, the expected capacity $C_{exp}(\text{SNR}, F_G, 1)$ can be bounded from above using the dominating MUF as follows:

$$\begin{aligned}
C_{exp}(F_G, 1) &= \sum_{k=1}^{K} F_k \log\left(\frac{n_k + \beta_k^*}{n_k + \beta_{k-1}^*}\right) && (2.22)\\
&= \sum_{k=1}^{K} \int_{\beta_{k-1}^*}^{\beta_k^*} u_k(z)dz && (2.23)\\
&\leq \sum_{k=1}^{K} \int_{\beta_{k-1}^*}^{\beta_k^*} u^*(z)dz && (2.24)\\
&= \int_{\beta_0^*}^{\beta_K^*} u^*(z)dz && (2.25)\\
&\leq \int_0^1 u^*(z)dz && (2.26)
\end{aligned}$$

where (2.24) follows from the fact that for any $k = 1, \ldots, K$ we have $\beta_{k-1}^* \leq \beta_k^*$ and $u_k(z) \leq u^*(z)$ for all z, and (2.26) follows from the fact that $\beta_0^* = 0$, $\beta_K^* \leq 1$, and $u^*(z) > 0$ for all $z \geq 0$. The equalities hold if $(\beta_1^*, \ldots, \beta_K^*)$ satisfies

$$u^*(z) = u_k(z) \quad \forall z \in (\beta_{k-1}^*, \beta_k^*) \tag{2.27}$$

for any $k = 1, \ldots, K$ and $\beta_K^* = 1$.

Note that by property 3) of Lemma 1, we have

$$-n_1 =: z_{\pi_0,\pi_1} < z_{\pi_1,\pi_2} \leq \ldots \leq z_{\pi_{I-1},\pi_I} < z_{\pi_I,\pi_{I+1}} := \infty. \tag{2.28}$$

To proceed, let us define two integers s and w as follows.

Definition 2. *Let s be the largest index $i \in \{1, \ldots, I\}$ such that $z_{\pi_{i-1},\pi_i} \leq 0$ and let w be the largest index $i \in \{1, \ldots, I\}$ such that $z_{\pi_{i-1},\pi_i} < 1$.*

Clearly, we have $1 \leq s \leq w \leq I$. Furthermore if $s = w$, we have

$$\cdots \leq z_{\pi_{s-1},\pi_s} \leq 0 < 1 \leq z_{\pi_s,\pi_{s+1}} \leq \cdots \tag{2.29}$$

and if $s < w$, we have

$$\cdots \leq z_{\pi_{s-1},\pi_s} \leq 0 < z_{\pi_s,\pi_{s+1}} \leq \cdots \leq z_{\pi_{w-1},\pi_w} < 1 \leq z_{\pi_w,\pi_{w+1}} \leq \cdots \tag{2.30}$$

Using the definition of s and w, we have the following explicit characterization of an optimal power allocation.

Proposition 2 (An optimal power allocation)**.** *Assume that $g_K > 0$. Then, an optimal solution $(\beta_1^*, \ldots, \beta_K^*)$ to the optimization program (2.4) is given by*

$$\beta_k^* = \begin{cases} 0, & for \quad 1 \leq k < \pi_s \\ z_{\pi_i,\pi_{i+1}}, & for \quad \pi_i \leq k < \pi_{i+1} \text{ and } i = s, \ldots, w-1 \\ 1, & for \quad \pi_w \leq k \leq K \end{cases} \tag{2.31}$$

Proof. Note that we always have $\beta_K^* = 1$. Therefore, in light of the previous discussion, it is sufficient to show that the choice of $(\beta_1^*, \ldots, \beta_K^*)$ as given by (2.31) satisfies

(2.27) for any $k = 1, \ldots, K$. Also note that for the choice of (2.31), we only need to consider the cases where $k = \pi_i$ for $i = s, \ldots, w$. Otherwise, we have $\beta^*_{k-1} = \beta^*_k$ so the open interval $(\beta^*_{k-1}, \beta^*_k)$ is empty and hence there is nothing to prove.

Let us first assume that $s = w$. In this case, we only need to consider $k = \pi_s$, for which $\beta^*_{k-1} = 0$ and $\beta^*_k = 1$. By Proposition 1, $u^*(z) = u_{\pi_s}(z)$ for any $z \in (z_{\pi_{s-1},\pi_s}, z_{\pi_s,\pi_{s+1}})$. By (2.29), $z_{\pi_{s-1},\pi_s} \leq 0$ and $z_{\pi_s,\pi_{s+1}} \geq 1$. We thus conclude that $u^*(z) = u_{pi_s}(z)$ for any $z \in (0, 1)$.

Next, let us assume that $s < w$. We shall consider the following three cases separately.

Case 1: $k = \pi_s$. In this case, $\beta^*_{k-1} = 0$ and $\beta^*_k = z_{\pi_s,\pi_{s+1}}$. By Proposition 1, $u^*(z) = u_{\pi_s}(z)$ for any $z \in (z_{\pi_{s-1},\pi_s}, z_{\pi_s,\pi_{s+1}})$. By (2.30), $z_{\pi_{s-1},\pi_s} \leq 0$. We thus conclude that $u^*(z) = u_{\pi_s}(z)$ for any $z \in (0, z_{\pi_s,\pi_{s+1}})$.

Case 2: $k = \pi_i$ for some $i \in \{s+1, \ldots, w-1\}$. In this case, $\beta^*_{k-1} = z_{\pi_{i-1},\pi_i}$ and $\beta^*_k = z_{\pi_i,\pi_{i+1}}$. By Proposition 1, $u^*(z) = u_{\pi_i}(z)$ for any $z \in (z_{\pi_{i-1},\pi_i}, z_{\pi_i,\pi_{i+1}})$.

Case 3: $k = \pi_w$. In this case, $\beta^*_{k-1} = z_{\pi_{w-1},\pi_w}$ and $\beta^*_k = 1$. By Proposition 1, $u^*(z) = u_{pi_w}(z)$ for any $z \in (z_{\pi_{w-1},\pi_w}, z_{\pi_w,\pi_{w+1}})$. By (2.30), $z_{\pi_w,\pi_{w+1}} \geq 1$. We thus conclude that $u^*(z) = u_{pi_w}(z)$ for any $z \in (z_{\pi_{w-1},\pi_w}, 1)$.

We have thus completed the proof of the proposition. □

Note from (2.4) that the power allocated to the fading state g_k is given by $\beta_k - \beta_{k-1}$. Thus for the optimal power allocation given by (2.31), the "active" fading states g_k that are assigned to nonzero power (i.e., $\beta^*_k > \beta^*_{k-1}$) are $\pi_s, \pi_{s+1}, \ldots, \pi_w$, i.e., g_{π_s} is the strongest active fading state, and g_{π_w} is the weakest active fading state (see Figure 2.4 for an illustration). This provides an operational meaning for the integer sequence $\{\pi_1, \ldots, \pi_I\}$ and the integers s and w defined earlier.

Building on Proposition 2, we have the following characterization of the expected

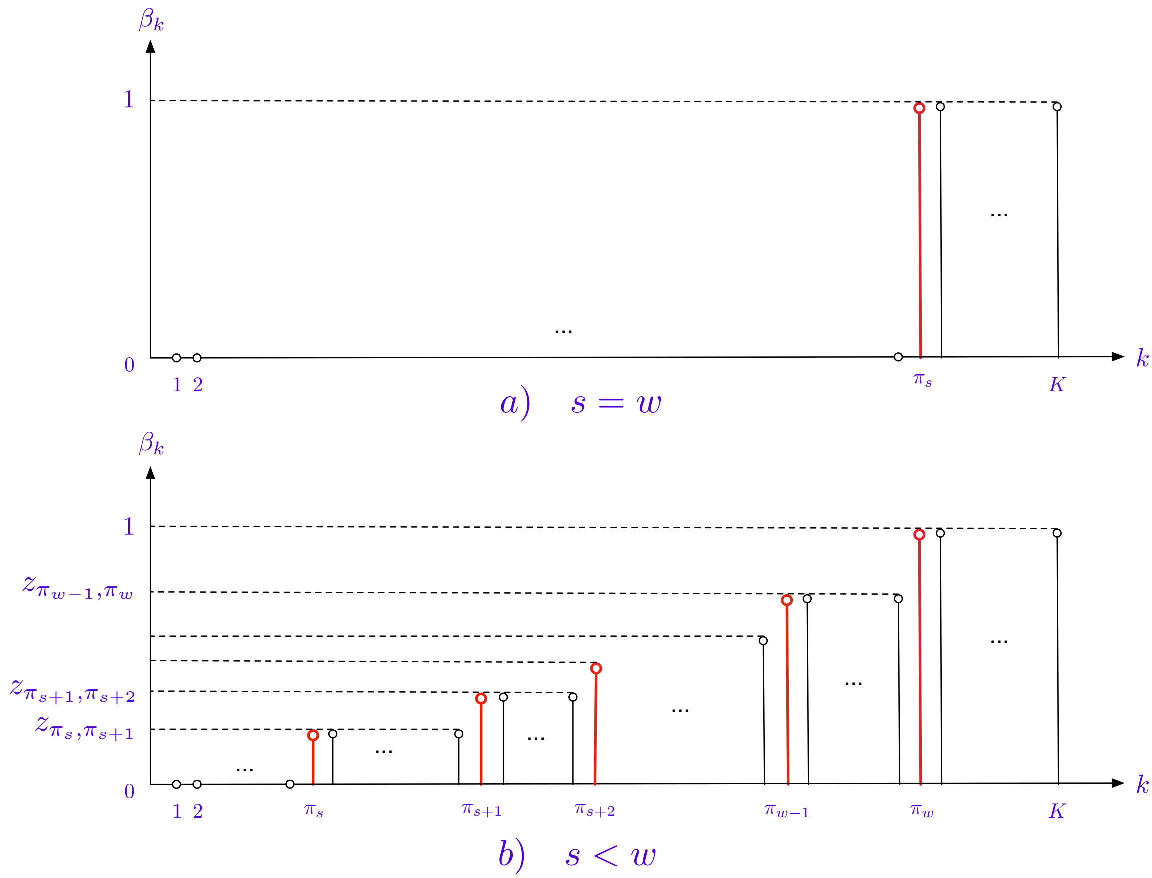

Figure 2.4: An optimal power allocation obtained via the dominating MUF.

capacity $C_{exp}(F_G, 1)$, which will play a key role in proving the desired worst-case gap results (2.1) and (2.2). The proof mainly involves some straightforward calculations and hence is deferred to Appendix A.

Proposition 3 (Expected capacity over one-block delay)**.** *Assume that $g_K > 0$ and let*

$$\Lambda_k := \begin{cases} \frac{n_{\pi_w}+1}{n_{\pi_s}} \frac{F_{\pi_s}}{F_{\pi_w}} & for \quad 1 \leq k \leq \pi_s \\ \frac{n_{\pi_w}+1}{n_{\pi_m}-n_{\pi_{m-1}}} \frac{F_{\pi_m}-F_{\pi_{m-1}}}{F_{\pi_1}} & for \quad \pi_{m-1} \leq k \leq \pi_m \text{ and } m = s+1, \ldots, 1 \\ 1 & for \quad \pi_1 \leq k \leq K. \end{cases} \tag{2.32}$$

Then, the expected capacity $C_{exp}(F_G, 1)$ can be written as

$$C_{exp}(F_G, 1) \tag{2.33}$$

$$= \sum_{k=1}^{K} p_k \log \Lambda_k \tag{2.34}$$

$$= F_{\pi_s} \log\left(\frac{F_{\pi_s}}{n_{\pi_s}}\right) + \sum_{m=s+1}^{w} (F_{\pi_m} - F_{\pi_{m-1}}) \log\left(\frac{F_{\pi_m} - F_{\pi_{m-1}}}{n_{\pi_m} - n_{\pi_{m-1}}}\right) + F_{\pi_w} \log\left(\frac{n_{\pi_w} + 1}{F_{\pi_w}}\right). \tag{2.35}$$

2.3 Two Asymptotic Regimes

Before we formally prove the worst-case gap results (2.1) and (2.2), let us first take a look at the nature of the optimal power allocation (2.31) in two asymptotic regimes. As we shall see, these analyses provides some insight into why the worst-case additive and multiplicative gaps are $\log K$ and K, respectively.

Our first asymptotic analysis is in the *high* receive signal-to-noise ratio regime and is motivated by the concept of *generalized degree of freedom* [20, 21]. Consider

$$g_k = \text{SNR}^{r_k}, \quad k = 1, \dots, K \tag{2.36}$$

for some

$$r_1 > r_2 > \cdots > r_K > 0 \tag{2.37}$$

where SNR can be made arbitrarily large. Fix $\{r_k\}$ and $\{p_k\}$. For sufficiently large SNR, by (2.12) we have

$$z_{k,l} = \frac{F_k \text{SNR}^{-r_l} - F_l \text{SNR}^{-r_k}}{F_l - F_k} \approx \frac{F_k}{F_l - F_k} \text{SNR}^{-r_l} \tag{2.38}$$

for any $1 \leq k < l \leq K$. By the ordering (2.37), we have for sufficiently large SNR

$$\frac{F_k}{F_l - F_k}\mathrm{SNR}^{-r_l} < \frac{F_k}{F_{l+1} - F_k}\mathrm{SNR}^{-r_{l+1}} \tag{2.39}$$

and hence

$$z_{k,l} < z_{k,l+1} \tag{2.40}$$

for any $1 \leq k < l \leq K-1$. By the definition (2.13), we have $I = K$ and $\pi_i = i$ for all $i = 1, \ldots, K$. Furthermore, by (2.38) we have $0 < z_{k,l} < 1$ for sufficiently large SNR. Hence, by Definition 2 we have $s = 1$ and $w = K$. We thus conclude that for sufficiently large SNR all fading states g_k, $k = 1, \ldots, K$, are active fading states that are assigned to nonzero power. By (2.35) the expected capacity over one-block delay

$$\begin{aligned} C_{exp} &= F_1 \log(F_1 \mathrm{SNR}^{r_1}) + \sum_{m=2}^{K}(F_m - F_{m-1}) \log\left(\frac{F_m - F_{m-1}}{\mathrm{SNR}^{-r_m} - \mathrm{SNR}^{-r_{m-1}}}\right) + \\ & \quad F_K \log\left(\frac{\mathrm{SNR}^{-r_K} + 1}{F_K}\right) &(2.41) \\ &\approx \left(\sum_{m=1}^{K} p_m r_m\right) \log \mathrm{SNR} + \sum_{m=1}^{K} p_m \log p_m &(2.42) \end{aligned}$$

and by (1.2) the ergodic capacity

$$C_{erg}(F_G) = \sum_{k=1}^{K} p_k \log(1 + \mathrm{SNR}^{r_k}) \approx \left(\sum_{k=1}^{K} p_k r_k\right) \log \mathrm{SNR} \tag{2.43}$$

for sufficiently large SNR. Thus, for sufficiently large SNR the additive gap

$$A(F_G, 1) \approx -\sum_{m=1}^{K} p_m \log p_m =: H(F_G) \leq \log K \tag{2.44}$$

for any $\{r_k\}$ and $\{p_k\}$, where $H(F_G)$ denotes the entropy of the power-gain distribution $F_G(\cdot)$, and the last inequality follows from the well-known fact that a uniform distribution maximizes the entropy subject to the cardinality constraint. This suggests that the *worst-case* additive gap may be $\log K$.

Our second asymptotic analysis is in the *low* receive signal-to-noise ratio regime and is motivated by the concept of *channel capacity per unit cost* [22]. Consider

$$g_k = \alpha_k \mathrm{SNR}, \quad k = 1, \dots, K \tag{2.45}$$

for some

$$\alpha_1 > \alpha_2 > \cdots > \alpha_K > 0 \tag{2.46}$$

where SNR can be made arbitrarily close to zero. Fix $\{\alpha_k\}$ and $\{p_k\}$. For sufficiently small SNR, by (2.12) we have

$$z_{k,l} = \frac{F_k \alpha_l^{-1} - F_l \alpha_k^{-1}}{F_l - F_k} \frac{1}{\mathrm{SNR}} \tag{2.47}$$

for any $1 \leq k < l \leq K$. Note that for sufficiently small SNR we have $z_{k,l} > 1$ whenever it is positive. Thus, by Definition 2 we have $w = s$, i.e., the only active fading state is g_{π_s}, for sufficiently small SNR. By (2.35) the expected capacity over one-block delay

$$C_{exp}(F_G, 1) = F_{\pi_s} \log(1 + \alpha_{\pi_s} \mathrm{SNR}) \approx F_{\pi_s} \alpha_{\pi_s} \mathrm{SNR} \tag{2.48}$$

and by (1.2) the ergodic capacity

$$C_{erg}(F_G) = \sum_{k=1}^{K} p_k \log(1 + \alpha_k \mathrm{SNR}) \approx \left(\sum_{k=1}^{K} p_k \alpha_k \right) \mathrm{SNR} \tag{2.49}$$

for sufficiently small SNR. By Lemma 1 and the fact that $w = s$ we have

$$z_{k,\pi_s} \leq z_{\pi_{s-1},\pi_s} \leq 0 < 1 < z_{\pi_s,\pi_{s+1}} \leq z_{\pi_s,l} \tag{2.50}$$

for any $1 \leq k < \pi_s < l \leq K$, which implies that

$$F_{\pi_s}\alpha_{\pi_s} \geq F_k\alpha_k, \quad \forall k = 1, \ldots, K. \tag{2.51}$$

Thus, for sufficiently small SNR the multiplicative gap

$$M(F_G, 1) \approx \frac{\sum_{k=1}^{K} p_k\alpha_k}{F_{\pi_s}\alpha_{\pi_s}} \leq \sum_{k=1}^{K} 1 = K \tag{2.52}$$

for any $\{\alpha_k\}$ and $\{p_k\}$, suggesting that the *worst-case* multiplicative gap may be K.

2.4 Additive Gap

To prove the worst-case additive gap result (2.1), we shall prove that

$$\sup_{F_G} A(F_G, 1) \leq \log K \tag{2.53}$$

and

$$\sup_{F_G} A(F_G, 1) \geq \log K \tag{2.54}$$

separately.

Proposition 4 (Worst-case additive gap, converse part)**.** *For any power-gain distribution $F_G(\cdot)$ with K different realizations of the power gain in each coherent block, we have*

$$A(F_G, 1) \leq \log K \tag{2.55}$$

Proof. Let us first prove the desired inequality (2.55) for the case where $g_K > 0$. In this case, by Proposition 3 the additive gap $A(F_G, 1)$ can be written as

$$A(F_G, 1) = \sum_{k=1}^{K} p_k \log\left(\frac{n_k + 1}{n_k}\right) - \sum_{k=1}^{K} p_k \log \Lambda_k \tag{2.56}$$

$$= \sum_{k=1}^{K} p_k \log\left(\frac{n_k + 1}{n_k \Lambda_k}\right). \tag{2.57}$$

We have the following lemma, whose proof is rather technical and hence is deferred to Appendix B.

Lemma 2. *For any* $k = 1, \ldots, K$*, we have*

$$\frac{n_k + 1}{n_k \Lambda_k} \leq \frac{1}{p_k}. \tag{2.58}$$

Substituting (2.58) into (2.57), we have

$$A(F_G, 1) \leq \sum_{k=1}^{K} p_k \log\left(\frac{1}{p_k}\right) =: H(F_G) \leq \log K. \tag{2.59}$$

This proves the desired inequality (2.55) for the case where $g_K > 0$.

For the case where $g_K = 0$, let us consider a modified power-gain distribution $F'_G(\cdot)$ with probabilities $p'_k = p_k$ for all $k = 1, \ldots, K$ and $g'_k = g_k$ for all $k = 1, \ldots, K - 1$. While we have $g_K = 0$ for the original power-gain distribution $F_G(\cdot)$, we shall let $g'_K = \epsilon$ for some

$$0 < \epsilon < \min_{k=1,\ldots,K-1} \left[\frac{F_k}{(1 - F_k) + n_k}\right]. \tag{2.60}$$

By 2.12, this will ensure that

$$z'_{k,K} = \frac{F_k/\epsilon - n_k}{1 - F_k} > 1, \quad \forall k = 1, \ldots, K-1. \tag{2.61}$$

By the definition of w', $z'_{\pi'_{w'-1},\pi'_{w'}} < 1$ so we must have $\pi'_{w'} \neq K$ and hence $\pi'_{w'} < K$. By Proposition 2, this implies that $\beta'^*_K = \beta'^*_{K-1}$ so the fading state g'_K are assigned to zero power for the given power allocation $(\beta'^*_1, \ldots, \beta'^*_K)$. Hence, the given power allocation $(\beta'^*_1, \ldots, \beta'^*_K)$ achieves the *same* expected rate for both power-gain distributions $F_G(\cdot)$ and $F'_G(\cdot)$. Since $(\beta'^*_1, \ldots, \beta'^*_K)$ is optimal for the power-gain distribution $F'_G(\cdot)$ but not necessarily so for $F_G(\cdot)$, we have

$$C_{exp}(F_G, 1) \geq C_{exp}(F'_G, 1) \tag{2.62}$$

On the other hand, improving the realizations of the power-gain can only improve the channel capacity[1], so we have

$$C_{erg}(F_G) \leq C_{erg}(F'_G). \tag{2.63}$$

Combining (2.62) and (2.63) gives

$$\begin{aligned} A(F_G, 1) &= C_{erg}(F_G) - C_{exp}(F_G, 1) && (2.64) \\ &\leq C_{erg}(F'_G) - C_{exp}(F'_G, 1) && (2.65) \\ &= A(F'_G, 1) && (2.66) \\ &\leq \log K && (2.67) \end{aligned}$$

[1]By the same argument, we also have $C_{exp}(F_G, 1) \leq C_{exp}(F'_G, 1)$ and hence $C_{exp}(F_G, 1) = C_{exp}(F'_G, 1)$, even though this direction of the inequality is not needed in the proof.

where the last inequality follows from the previous case for which $g'_K = \epsilon > 0$. This completes the proof for the case where $g_K = 0$.

Combing the above two cases completes the proof of Proposition 4. $\square$

Proposition 5 (Worst-case additive gap, forward part)**.** *Fix K, and consider the power-gain distribution $F_G^{(d)}(\cdot)$ with*

$$g_k = \sum_{j=1}^{K-k+1} d^j = \frac{d(d^{K-k+1}-1)}{d-1} \tag{2.68}$$

for some $d > \max[K-1, 2]$ and uniform probabilities $p_k = 1/K$ for all $k = 1, \ldots, K$. For this particular parameter family of power-gain distributions, we have

$$\lim_{d\to\infty} A(F_G^{(d)}, 1) = \log K. \tag{2.69}$$

Proof. For the given power-gain distribution $F_G^{(d)}$, it is straightforward to calculate that for any $1 \le k < l < K$

$$\frac{n_k + z_{k,l}}{n_k + z_{k,l+1}} = \frac{l-k+1}{l-k} \frac{d^{K-l}-1}{d^{K-l+1}-1} \frac{d^{l-k}-1}{d^{l-k+1}-1} \tag{2.70}$$

$$< \frac{l-k+1}{l-k} \frac{d^{l-k}-1}{d^{l-k+1}-1} \tag{2.71}$$

where the last inequality follows from the fact that $d > 1$. Since $l - k \ge 1$ and $d > 2$, we have

$$(l-k+1)(d^{l-k}-1) - (l-k)(d^{l-k+1}-1)$$

$$= [1-(l-k)(d-1)]d^{l-k} - 1 \tag{2.72}$$

$$< 0. \tag{2.73}$$

Substituting (2.73) into (2.71) gives

$$\frac{n_k + z_{k,l}}{n_k + z_{k,l+1}} < 1 \tag{2.74}$$

which immediately implies that $z_{k,l} < z_{k,l+1}$ for any $1 \leq k < l < K$. We thus have $I = K$ and $\pi_i = i$ for all $i = 1, \ldots, K$. Since $d > \max\{K-1, 2\}$, we have

$$z_{1,2} = \frac{(d-2)d^K + d}{(d-1)g_1 g_2} > 0 \tag{2.75}$$

and

$$z_{K-1,K} = \frac{(K-1)(d+d^2) - Kd}{d(d+d^2)} < \frac{K-1}{d} < 1 \tag{2.76}$$

so by definition we have $s = 1$ and $w = K$. Thus, by the expression of Λ_k from (2.32) we have

$$\Lambda_k = \begin{cases} \frac{\left(\sum_{j=1}^{K} d^j\right)(1+d)}{K \cdot d}, & k = 1 \\ \frac{\left(\sum_{j=1}^{K-k+1} d^j\right)\left(\sum_{j=1}^{K-k+2} d^j\right)(1+d)}{K \cdot d^{K-k+3}}, & k = 2, \ldots, K. \end{cases} \tag{2.77}$$

It follows that

$$\frac{n_1 + 1}{n_1 \Lambda_1} = K \cdot \frac{\left(1 + \sum_{j=1}^{K} d^j\right) d}{\left(\sum_{j=1}^{K} d^j\right)(1+d)} \tag{2.78}$$

$$= K \cdot \frac{d^{K+1} + O\left(d^K\right)}{d^{K+1} + O\left(d^K\right)} \tag{2.79}$$

$$\rightarrow K \tag{2.80}$$

in the limit as $d \to \infty$ and

$$\frac{n_k+1}{n_k\Lambda_k} = K \cdot \frac{\left(1+\sum_{j=1}^{K-k+1} d^j\right) d^{K-k+3}}{\left(\sum_{j=1}^{K-k+1} d^j\right)\left(\sum_{j=1}^{K-k+2} d^j\right)(1+d)} \tag{2.81}$$

$$= K \cdot \frac{d^{2(K-k)+4} + O\left(d^{2(K-k)+3}\right)}{d^{2(K-k)+4} + O\left(d^{2(K-k)+3}\right)} \tag{2.82}$$

$$\to K \tag{2.83}$$

in the limit as $d \to \infty$ for any $k = 2 \ldots, K$. A numerical example illustrating the convergence of (2.80) and (2.83) is provided in Figure 2.5- 2.7. By (2.57), the additive gap

$$A(F_G^{(d)}, 1) = \sum_{k=1}^{K} p_k \log\left(\frac{n_k+1}{n_k\Lambda_k}\right) \tag{2.84}$$

$$\to \sum_{k=1}^{K} \frac{1}{K} \log K \tag{2.85}$$

$$= \log K \tag{2.86}$$

in the limit as $d \to \infty$. This completes the proof of Proposition 5. □

Combining Propositions 4 and 5 completes the proof of the desired worst-case additive gap result (2.1).

2.5 Multiplicative Gap

Similar to the additive case, to prove the worst-case multiplicative gap result (2.2) we shall prove that

$$sup_{F_G} M(F_G, 1) \leq K \tag{2.87}$$

and

$$sup_{F_G} M(F_G, 1) \geq K \tag{2.88}$$

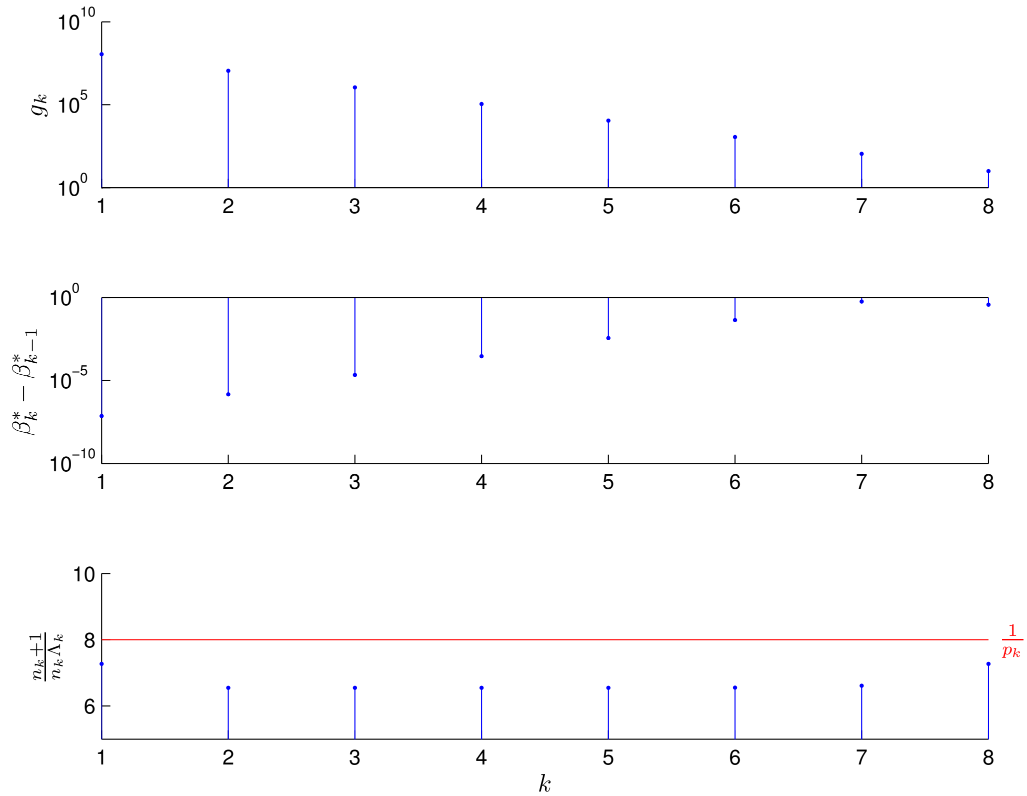

Figure 2.5: A numerical example illustrating the convergence of (2.80) and (2.83). In this example, $K = 8$ and $d = 10$.

separately.

Proposition 6 (Worst-case multiplicative gap, converse part)**.** *For any power-gain distribution $F_G(\cdot)$ with K different realizations of the power gain in each coherent block, we have*

$$M(F_G, 1) \le K. \tag{2.89}$$

Proof. Let us first prove the desired inequality (2.89) for the case where $g_K > 0$. By

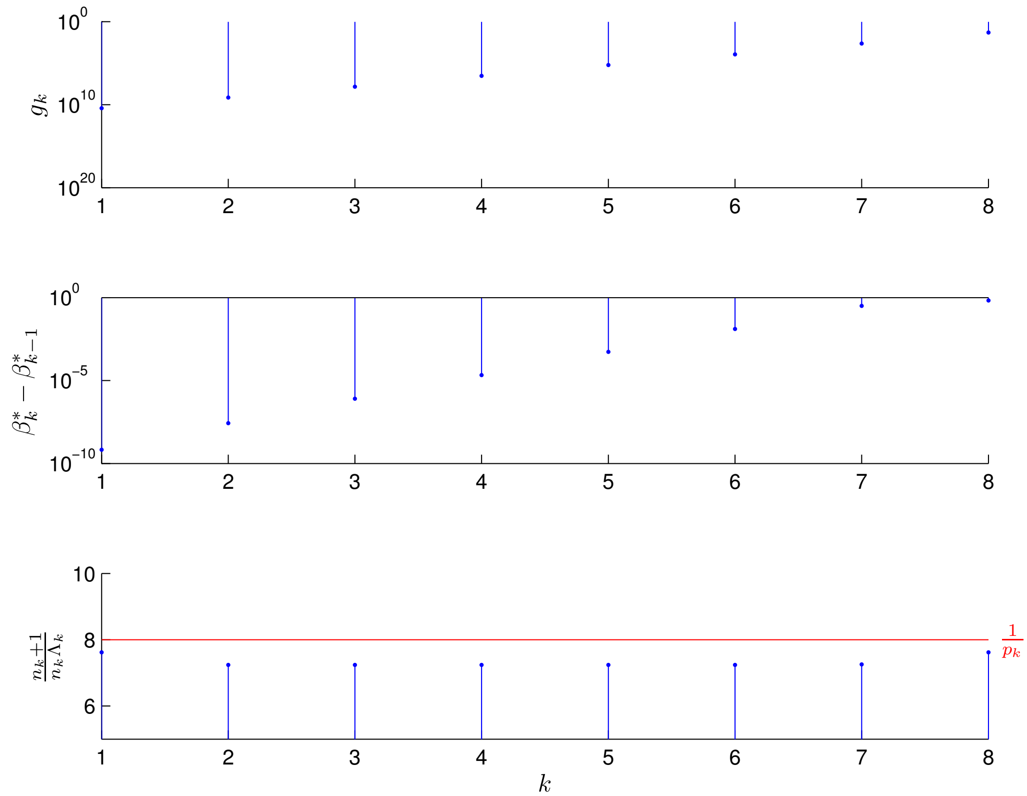

Figure 2.6: A numerical example illustrating the convergence of (2.80) and (2.83). In this example, $K = 8$ and $d = 20$.

definition the multiplicative gap $M(F_G, 1)$ can be written as

$$M(F_G, 1) = \sum_{k=1}^{K} \frac{p_k \log\left(\frac{n_k+1}{n_k}\right)}{C_{exp}(F_G, 1)}. \tag{2.90}$$

We have the following lemma, whose proof is deferred to Appendix C.

Lemma 3. *For any $k = 1, \ldots, K$, we have*

$$\frac{p_k \log\left(\frac{n_k+1}{n_k}\right)}{C_{exp}(F_G, 1)} \leq 1. \tag{2.91}$$

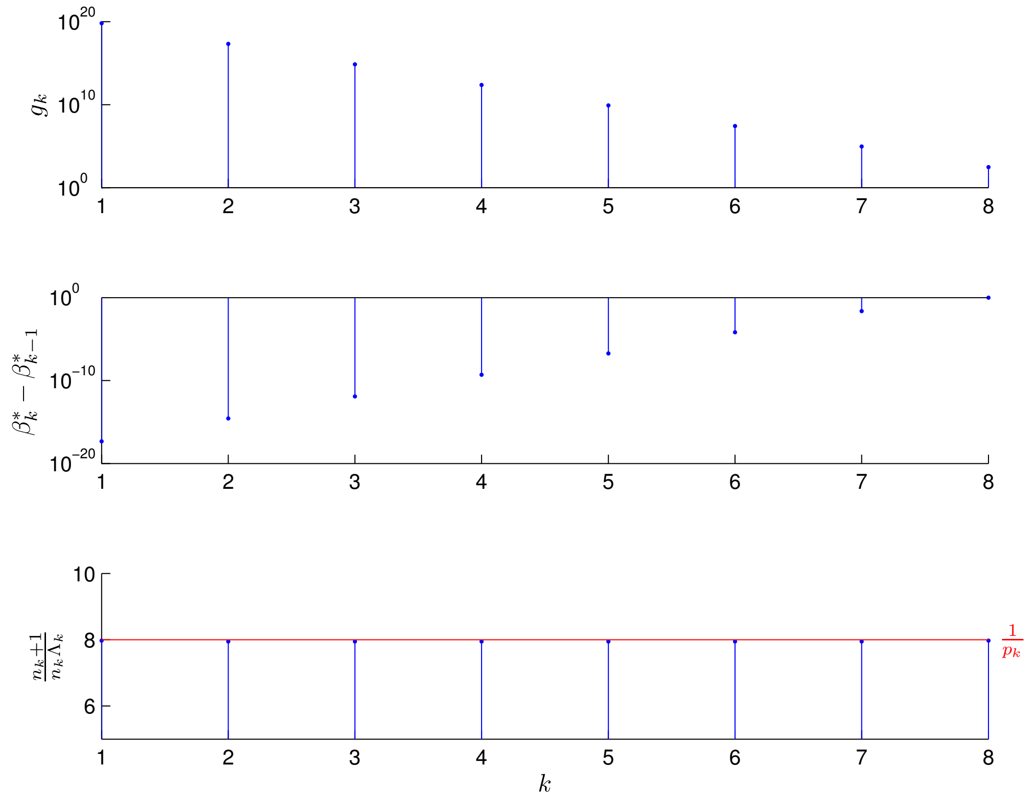

Figure 2.7: A numerical example illustrating the convergence of (2.80) and (2.83). In this example, $K = 8$ and $d = 300$.

Substituting (2.91) into (2.90), we have

$$M(F_G, 1) \leq \sum_{k=1}^{K} 1 = K. \tag{2.92}$$

This proves the desired inequality (2.89) for the case where $g_K > 0$.

For the case where $g_K = 0$, we can use the same argument as for the additive case. More specifically, a modified power-gain distribution $F'_G(\cdot)$ can be found such that $g'_K > 0$, $C_{exp}(F'_G, 1) = C_{exp}(F_G, 1)$, and $C_{erg}(F'_G) \geq C_{erg}(F_G)$. Thus, the multiplicative gap

$$
\begin{aligned}
M(F_G, 1) &= \frac{C_{erg}(F_G)}{C_{exp}(F_G, 1)} && (2.93)\\
&\leq \frac{C_{erg}(F'_G)}{C_{exp}(F'_G, 1)} && (2.94)\\
&= M(F'_G, 1) && (2.95)\\
&\leq K && (2.96)
\end{aligned}
$$

where the last inequality follows from the previous case for which $g'_K > 0$. This completes the proof for the case where $g_K = 0$.

Combing the above two cases completes the proof of Proposition 6. □

Proposition 7 (Worst-case multiplicative gap, forward part)**.** *Fix K, and consider the power-gain distributions $F_G^{(d)}(\cdot)$ with*

$$n_k = \sum_{j=1}^{k} d^j \tag{2.97}$$

for some $d > 0$ and

$$p_k = \frac{d^k}{\sum_{j=1}^{K} d^j} \tag{2.98}$$

for all $k = 1, \ldots, K$. For this particular parameter family of power-gain distributions, we have

$$\lim_{d \to \infty} M(F_G^{(d)}, 1) = K. \tag{2.99}$$

Proof. Note that for the given power-gain distribution $F_G^{(d)}$,

$$F_k = \sum_{j=1}^{k} p_j = \frac{\sum_{j=1}^{k} d^j}{\sum_{j=1}^{K} d^j} \tag{2.100}$$

so

$$z_{k,l} = \frac{F_k n_l - F_l n_k}{F_l - F_k} = 0, \quad \forall 1 \le k < l \le K. \tag{2.101}$$

We thus have $I = 2$, $\pi_1 = 1$, $\pi_2 = K$, and $s = w = 2$. By the expression of Λ_k from (2.32), we have

$$\Lambda_k = \frac{n_K + 1}{n_K}, \quad \forall k = 1, \ldots, K. \tag{2.102}$$

It follows that the expected capacity

$$C_{exp}(F_G^{(d)}, 1) = \sum_{k=1}^{K} p_k \log \Lambda_k = \log \frac{n_K + 1}{n_K}. \tag{2.103}$$

We thus have

$$\frac{p_k \log\left(\frac{n_k+1}{n_k}\right)}{C_{exp}(F_G^{(d)}, 1)} = \frac{p_k \log\left(\frac{n_k+1}{n_k}\right)}{\log\left(\frac{n_K+1}{n_K}\right)} \tag{2.104}$$

$$\ge \frac{p_k n_K}{n_k + 1} \tag{2.105}$$

$$= \frac{d^k}{\sum_{j=1}^{k} d^j + 1} \tag{2.106}$$

$$= \frac{d^k}{d^k + O\left(d^{k-1}\right)} \tag{2.107}$$

$$\to 1 \tag{2.108}$$

in the limit as $d \to \infty$ for any $k = 1, \ldots, K$, where (2.90) follows from the well-known inequalities

$$\frac{x}{1+x} \le \log(1+x) \le x, \quad \forall x \ge 0, \tag{2.109}$$

so we have $\log\left(\frac{n_k+1}{n_k}\right) \ge \frac{1}{n_k+1}$ and $\log\left(\frac{n_K+1}{n_K}\right) \le \frac{1}{n_K}$. On the other hand, by Lemma 3

$$\frac{p_k \log\left(\frac{n_k+1}{n_k}\right)}{C_{exp}(F_G^{(d)}, 1)} \le 1 \tag{2.110}$$

for any $k = 1, \ldots, K$. Combining (2.108) and (2.110) proves that

$$\frac{p_k \log\left(\frac{n_k+1}{n_k}\right)}{C_{exp}(F_G^{(d)}, 1)} \to 1 \tag{2.111}$$

in the limit as $d \to \infty$ for all $k = 1, \ldots, K$. A numerical example illustrating the convergence of (2.111) is illustrated in Figure 2.8- 2.10. By (2.90), the multiplicative gap

$$M(F_G^{(d)}, 1) = \sum_{k=1}^{K} \frac{p_k \log\left(\frac{n_k+1}{n_k}\right)}{C_{exp}(F_G^{(d)}, 1)} \to \sum_{k=1}^{K} 1 = K \tag{2.112}$$

in the limit as $d \to \infty$. This completes the proof of Proposition 7. $\square$

Combining Propositions 6 and 7 completes the proof of the desired worst-case multiplicative gap result (2.2).

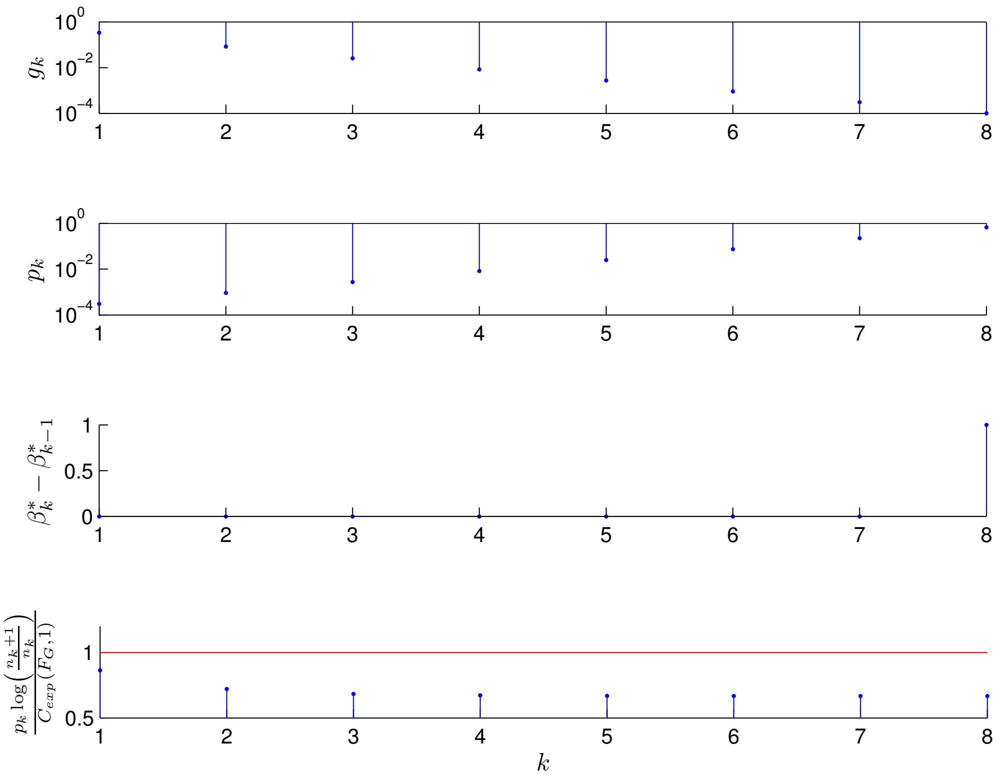

Figure 2.8: A numerical example illustrating the convergence of (2.111) is illustrated in Figure 2.8- 2.10. In this example, $K = 8$ and $d = 3$.

Figure 2.9: A numerical example illustrating the convergence of (2.111) is illustrated in Figure 2.8- 2.10. In this example, $K = 8$ and $d = 6$.

Figure 2.10: A numerical example illustrating the convergence of (2.111) is illustrated in Figure 2.8- 2.10. In this example, $K = 8$ and $d = 60$.

3. WRITING ON BLOCK-FADING PAPER*

3.1 Introduction

Consider the problem of writing on block-fading paper [11–13]:

$$Y[t] = G[t](X[t] + S[t]) + Z[t] \tag{3.1}$$

where $\{X[t]\}$ are the channel inputs which are subject to a unit average power constraint, $\{G[t]\}$ are the power gains of the channel fading which are assumed to be *constant* within each coherent block and change *independently* across different blocks according to a known distribution $F_G(\cdot)$, $\{S[t]\}$ and $\{Z[t]\}$ are independent additive white circularly symmetric complex Gaussian interference and noise with zero means and variance INR and 1 respectively, and $\{Y[t]\}$ are the channel outputs. The power gains $\{G[t]\}$ are assume to be *unknown* to the transmitter but known at the receiver and the interference signal $\{S[t]\}$ are assumed to be non-causally known at the transmitter but not to the receiver. Note here that the instantaneous power gain $G[t]$ applies to both the channel input $X[t]$ and the known interference $S[t]$, so this model is particularly relevant to the problem of precoding for multiple-input multiple-output fading broadcast channels.

As for the point-to-point fading channel (1.1) in Chapter 1, we are interested in characterizing the *worst-case* expected-rate loss for the slow-fading scenario. However, unlike for the point-to-point fading channel (1.1), the ergodic capacity of the fading-paper channel (3.1) is unknown. We first characterize the ergodic capacity

*Reprinted, with permission, from J. W. Yoo, T. Liu, S. Shamai (Shitz), and C. Tian,"Worst-Case Expected-Capacity Loss of Slow-Fading Channels," *IEEE Transactions on Information Theory*, to appear, Copyright 2013 IEEE.

of the fading-paper model (3.1) to within in one bit per channel use. As we will see, this will also lead to a characterization of the additive expected-capacity loss to within one bit per channel use for the slow-fading scenario.

3.2 Ergodic Capacity to within One Bit

Denote by $C_{erg}^{fp}(\mathrm{INR}, F_G)$ the ergodic capacity of the fading-paper channel (3.1) with transmit interference-to-noise ratio INR, and power-gain distribution $F_G(\cdot)$. We have the following characterization of $C_{erg}^{fp}(\mathrm{INR}, F_G)$ to within one bit.

Theorem 2. *For any transmit interference-to-noise ratio* INR*, and any power-gain distribution* $F_G(\cdot)$*, we have*

$$C_{erg}(F_G) - \log 2 \leq C_{erg}^{fp}(\mathrm{INR}, F_G) \leq C_{erg}(F_G) \tag{3.2}$$

where $C_{erg}(F_G)$ *is the ergodic capacity of the point-to-point fading channel* (1.1) *of the same signal-to-noise ratio and power-gain distribution as the fading-paper channel* (3.1).

Proof. To show that $C_{erg}^{fp}(\mathrm{INR}, F_G) \leq C_{erg}(F_G)$, let us assume that the interference signal $\{S[t]\}$ are also known at the receiver. When the receiver knows both the power gain $\{G[t]\}$ and the interference signal $\{S[t]\}$, it can subtract $\{\sqrt{G[t]}S[t]\}$ from the received signal $\{Y[t]\}$. This will lead to an interference-free point-to-point fading channel (1), whose ergodic capacity is given by $C_{erg}(F_G)$. Since giving additional information to the receiver can only improve the ergodic capacity, we conclude that $C_{erg}^{fp}(\mathrm{INR}, F_G) \leq C_{erg}(F_G)$.

To show that $C_{erg}^{fp}(\mathrm{INR}, F_G) \geq C_{erg}(F_G) - \log 2$, we shall show that

$$R = \mathbb{E}_G\left[(\log G)^+\right] \tag{3.3}$$

is an achievable ergodic rate for the fading-paper channel (3.1), where $x^+ := \max(x, 0)$. Since

$$(\log G)^+ \geq \log(1+G) - \log 2 \tag{3.4}$$

for every possible realization of G, we will have

$$\begin{aligned} C_{erg}^{fp}(\mathrm{INR}, F_G) &\geq \mathbb{E}_G\left[(\log G)^+\right] &(3.5)\\ &\geq \mathbb{E}_G[\log(1+G)] - \log 2 &(3.6)\\ &= C_{erg}(F_G) - \log 2. &(3.7)\end{aligned}$$

To prove the achievability of the ergodic rate (3.3), we shall consider a communication scheme which is motivated by the following thought experiment. Note that with *ideal* interleaving, the block-fading channel (3.1) can be converted to a *fast-fading* one [1, Ch. 5.4.5] for which the power gains $\{G[t]\}$ are independent across different time index t. Now that the channel is memoryless, by the well-known result of Gel'fand and Pinsker [23] the following ergodic rate is achievable:

$$R = \max_{(X,U)} \left[I(U; \sqrt{G}(X+S) + Z|G) - I(U;S) \right] \tag{3.8}$$

where U is an auxiliary variable which must be independent of (G, Z). An optimal choice of the input-auxiliary variable pair (X, U) is unknown [11, 12]. Motivated by the recent work [24], let us consider

$$U = X + S \tag{3.9}$$

where X is circularly symmetric complex Gaussian with zero mean and variance P and is independent of S. For this choice of the input-auxiliary variable pair (X, U), we have

$$I(U;\sqrt{G}(X+S)+Z|G)-I(U;S) \tag{3.10}$$

$$=\mathbb{E}_G[\log(1+G(1+\mathrm{INR}))]-\log(1+\mathrm{INR}) \tag{3.11}$$

$$\geq \mathbb{E}_G[\log(G(1+\mathrm{INR}))]-\log(1+\mathrm{INR}) \tag{3.12}$$

$$\geq \mathbb{E}_G[\log G]. \tag{3.13}$$

This proves that

$$R=\{\mathbb{E}_G[\log G]\}^+ \tag{3.14}$$

is an achievable ergodic rate for the fading-paper channel (3.1).

Note that even though the achievable ergodic rate (3.14) is independent of the transmit interference-to-noise ratio INR, it is *not* always within one bit of the interference-free ergodic capacity $C_{erg}(F_G)$. This is because when $G<1$, we have $\log G<0$, i.e., the realizations of the power gain which are less than 1 contribute *negatively* to the achievable rate (3.14). By comparison, the realizations of the power gain *never* contribute negatively (but possibly zero) to the achievable rate (3.3). Next, motivated by the secure multicast code construction proposed in [25], we shall consider a *separate-binning* scheme which allows *opportunistic* decoding at the receiver to boost the the achievable ergodic rate from (3.14) to (3.3).

Fix $\epsilon>0$ and let (U,X) be chosen as in (3.9). Consider communicating a message $W\in\{1,\ldots,e^{LT_cR}\}$ over L coherent blocks, each of a block length T_c which we assume to be sufficiently large.

Codebook generation. Randomly generate L codebooks, each for one coherent block and consisting of $e^{T_c(LR+I(U;S)+\epsilon)}$ codewords of length T_c. The entries of the codewords are independently generated according to P_U. Randomly partition each

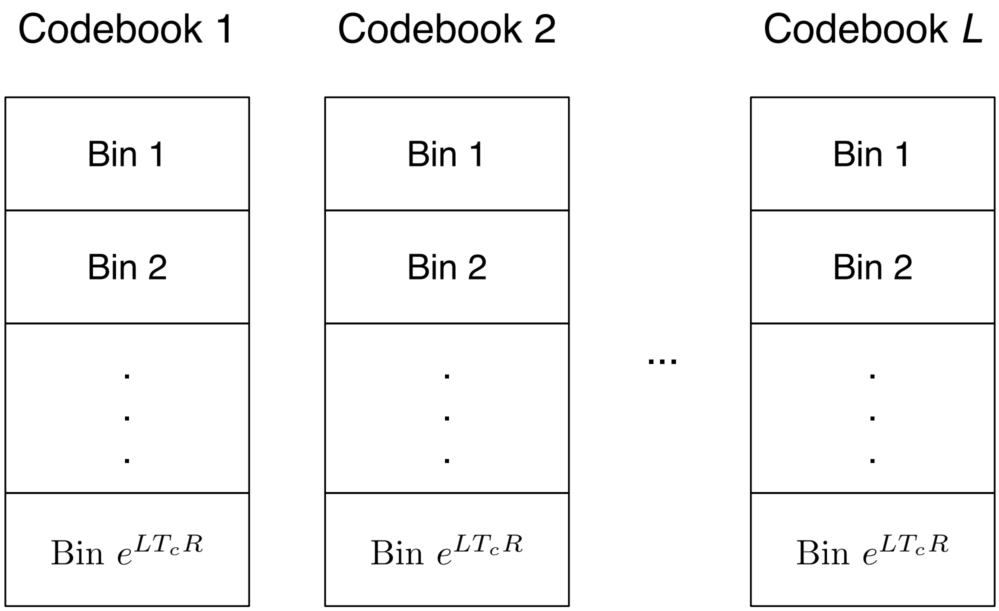

Figure 3.1: The codebook structure for achieving the ergodic rate (3.3). Each codeword bin in the codebooks contains codewords.

codebook into e^{LT_cR} bins, so each bin contains $e^{T_c(I(U;S)+\epsilon)}$ codewords. See Fig. 3.1 for an illustration of the codebook structure.

Encoding. Given the message W and the interference signal $S^{LT_c} := (S[1], \ldots, S[LT_c])$, the encoder looks into the Wth bin in each codebook l and tries to find a codeword that is jointly typical with $S_l^{T_c}$, where $S_l^{T_c} := (S[(l-1)T_c+1], \ldots, S[lT_c])$ represents the segment of the interference signal S^{LT_c} transmitted over the lth coherent block. By assumption, T_c is sufficiently large so with high probability such a codeword can be found in each codebook [26]. Denote by $U_l^{T_c} := (U[(l-1)T_c+1], \ldots, U[lT_c])$ the codeword chosen from the lth codebook. The transmit signal $X_l^{T_c} := (X[(l-1)T_c + 1], \ldots, X[lT_c])$ over the lth coherent block is given by $X_l^{T_c} = U_l^{T_c} - S_l^{T_c}$.

Decoding. Let G_l be the realization of the power gain during the lth coherent block, and let

$$\mathcal{L} := \{l : I(U; \sqrt{G_l}(X+S)+Z) - I(U;S) > 0\}. \tag{3.15}$$

Given the received signal $Y^{LT_c} := (Y[1], \ldots, Y[LT_c])$, the decoder looks for a codeword bin which contains for *each* coherent block $l \in \mathcal{L}$, a codeword that is jointly typical with the segment of $Y^{LT_c}(\mathcal{L})$ received over the lth coherent block. If only one such codeword bin can be found, the estimated message $\hat{W}$ is given by the index of the codeword bin. Otherwise, a decoding error is declared.

Performance analysis. Note that averaged over the codeword selections and by the union bound, the probability that an incorrect bin index is declared by the decoder is no more than

$$\begin{aligned}&\prod_{l\in\mathcal{L}} e^{T_c(I(U;S)+\epsilon)} \cdot e^{-T_c(I(U;\sqrt{G_l}(X+S)+Z)-\epsilon)} \\ &\quad = e^{-T_c \sum_{l\in\mathcal{L}}\left[I(U;\sqrt{G_l}(X+S)+Z)-I(U;S)-2\epsilon\right]}. \end{aligned}\tag{3.16}$$

Thus, by the union bound again, the probability of decoding error is no more than

$$\begin{aligned}&e^{T_cLR} \cdot e^{-T_c \sum_{l\in\mathcal{L}}\left[I(U;\sqrt{G_l}(X+S)+Z)-I(U;S)-2\epsilon\right]} \\ &\quad = e^{-T_c\left\{\sum_{l\in\mathcal{L}}\left[I(U;\sqrt{G_l}(X+S)+Z)-I(U;S)-2\epsilon\right]-LR\right\}}. \end{aligned}\tag{3.17}$$

It follows that the transmit message W can be reliably communicated (with exponentially decaying error probability for sufficiently large T_c) as long as

$$\sum_{l\in\mathcal{L}} \left[I(U;\sqrt{G_l}(X+S)+Z) - I(U;S) - 2\epsilon\right] - LR > 0 \tag{3.18}$$

or equivalently

$$R < \frac{1}{L}\sum_{l\in\mathcal{L}}\left[I(U;\sqrt{G_l}(X+S)+Z) - I(U;S) - 2\epsilon\right]. \tag{3.19}$$

Note that

$$\begin{aligned}
&\frac{1}{L}\sum_{l\in\mathcal{L}}\left[I(U;\sqrt{G_l}(X+S)+Z) - I(U;S) - 2\epsilon\right] \\
&\quad = \frac{1}{L}\sum_{l\in\mathcal{L}}\left[I(U;\sqrt{G_l}(X+S)+Z) - I(U;S)\right] - \frac{2|\mathcal{L}|}{L}\epsilon &(3.20)\\
&\quad \geq \frac{1}{L}\sum_{l\in\mathcal{L}}\left[I(U;\sqrt{G_l}(X+S)+Z) - I(U;S)\right] - 2\epsilon &(3.21)\\
&\quad = \frac{1}{L}\sum_{l=1}^{L}\left[I(U;\sqrt{G_l}(X+S)+Z) - I(U;S)\right]^+ - 2\epsilon &(3.22)\\
&\quad \geq \frac{1}{L}\sum_{l=1}^{L}(\log G_l)^+ - 2\epsilon &(3.23)
\end{aligned}$$

where (3.21) follows from the fact that $|\mathcal{L}| \leq L$, (3.22) follows from the definition of $\mathcal{L}$ from (3.15), and (3.23) follows from (3.13). Finally, by the weak law of large numbers,

$$\frac{1}{L}\sum_{l=1}^{L}(\log G_l)^+ \to \mathbb{E}_G\left[(\log G)^+\right] \tag{3.24}$$

in probability in the limit as $L \to \infty$. We thus conclude that (3.3) is an achievable ergodic rate for the fading-paper channel (3.1).

We have thus completed the proof of Theorem 2. □

The following observations are now in place. First, the boost of the achievable rate from (3.14) to (3.3) is mainly due to opportunistic decoding used by the receiver, which ensures that the realizations of the power gain which are less than 1 do not contribute negatively to the achievable rate. Second, the separate-binning scheme takes advantage of the *block-fading* nature and does not apply to the fast-fading scenario. Finally, the nature of the separate-binning scheme is such that the interference signal $S[t]$ within each coherent block only needs to be made available to the transmitter at the beginning of the block and not necessarily at the start of the entire communication.

3.3 Additive Expected-Capacity Loss to within One Bit

Let $C_{exp}^{fp}(\mathrm{INR}, F_G, L)$ be the expected capacity of the fading-paper channel (3.1) under the delay constraint of L coherent blocks, and let

$$A^{fp}(\mathrm{INR}, F_G, L) := C_{erg}^{fp}(\mathrm{INR}, F_G) - C_{exp}^{fp}(\mathrm{INR}, F_G, L) \tag{3.25}$$

be the additive gap between the ergodic capacity $C_{erg}^{fp}(\mathrm{INR}, F_G)$ and the expected capacity $C_{exp}^{fp}(\mathrm{INR}, F_G, L)$. We have the following results.

Theorem 3. *For any transmit interference-to-noise ratio* INR *and any power-gain distribution* $F_G(\cdot)$*, we have*

$$A(F_G, 1) - \log 2 \leq A^{fp}(\mathrm{INR}, F_G, 1) \leq A(F_G, 1). \tag{3.26}$$

Proof. We claim that for any transmit interference-to-noise ratio $\mathrm{INR} > 0$ and any power-gain distribution $F_G(\cdot)$, we have

$$C_{exp}^{fp}(\mathrm{INR}, F_G, 1) = C_{exp}(F_G, 1). \tag{3.27}$$

Then, the desired inequalities in (3.26) follow immediately from the above claim and Theorem 2.

To prove (3.27), let us consider the following K-user memoryless Gaussian broadcast channel:

$$Y_k = \sqrt{g_k}(X + S) + Z, \quad k = 1, \ldots, K \tag{3.28}$$

where X is the channel input which is subject an average power constraint, S and Z are independent additive white circularly symmetric complex Gaussian interference and noise, and g_k and Y_k are the power gain and the channel output of user k, respectively. The interference S is assumed to be non-causally known at the transmitter but not to the receivers. Similar to the interference-free (scalar) Gaussian broadcast channel, the broadcast channel (3.28) is also (stochastically) degraded. Furthermore, Steinberg [27] showed that through *successive* Costa precoding [26] at the transmitter, the capacity region of the broadcast channel (3.28) is the same as that of the interference-free Gaussian broadcast channel. We may thus conclude that the expected capacity $C_{exp}^{fp}(\text{INR}, F_G, 1)$ of the fading-paper channel (3.1) is the same as the expected capacity $C_{exp}(F_G, 1)$ of the interference-free point-to-point fading channel (1.1) of the same power-gain distribution $F_G(\cdot)$. This completes the proof of Theorem 3. □

Combining Theorems 1 and 3 immediately leads to the following corollary.

Corollary 1.

$$\log(K/2) \le \sup_{\text{INR}, F_G} A^{fp}(\text{INR}, F_G, 1) \le \log K. \tag{3.29}$$

where the supreme is over all transmit interference-to-noise ratio INR *and all power-gain distribution $F_G(\cdot)$ with K different possible realizations of the power gain in each coherent block.*

4. TWO-STATE BLOCK-FADING WITH TWO-BLOCK DELAY

4.1 Introduction

Consider the block-fading channel with a delay constraint of two blocks and two possible power gain realizations at each coherent block given by

$$Y_i[t] = X_i[t] + Z_i[t], \quad i = 1, 2 \tag{4.1}$$

where $\{X_1[t]\}$ and $\{X_2[t]\}$ are the channel inputs during block 1 and 2, which are subject to the unit average total power constraint

$$\frac{1}{2N} \sum_{t=1}^{N} \left(|X_1[t]|^2 + |X_2[t]|^2 \right) \leq 1 \tag{4.2}$$

and $\{Z_1[t]\}$ and $\{Z_2[t]\}$ are independent additive (complex) white Gaussian noise. For each block, there are two possible realizations for the noise variance: with probability p the noise variance is σ_H^2 and with probability $1-p$ the noise variance is σ_L^2 where $0 < \sigma_H^2 \leq \sigma_L^2$.

In [9], Whiting and Yeh characterized the expected capacity of the channel (4.1). However, their result on the expected capacity is, in fact, incorrect due to a wrong expression for the expected rate in their converse proof. This issue in [9] was noticed in [10]. Here, we characterize the expected capacity of the channel, which is strictly greater than the expected rate provided in [9]. An optimal power allocation is also characterized via *marginal utility functions (MUFs)* [15].

Let Z_{iH} and Z_{iL} denote Z_i with the noise variance σ_H^2 and σ_L^2 respectively. As described in [9], there are four possible states of the received signal at the receiver, which are (Y_{1H}, Y_{2H}), (Y_{1H}, Y_{2L}), (Y_{1L}, Y_{2H}), and (Y_{1L}, Y_{1L}) where

$$Y_{iH} \;:=\; X_i + Z_{iH} \tag{4.3}$$

$$Y_{iL} \;:=\; X_i + Z_{iL} \tag{4.4}$$

for $i = 1, 2$. Note here that there are relationships between the possible states as

$$\begin{aligned} (Y_{1L}, Y_{2L}) \;\rightarrow\; (Y_{1H}, Y_{2L}) \;\rightarrow\; (Y_{1H}, Y_{2H}) \\ (Y_{1L}, Y_{2L}) \;\rightarrow\; (Y_{1L}, Y_{2H}) \;\rightarrow\; (Y_{1H}, Y_{2H}). \end{aligned} \tag{4.5}$$

From the relationship (4.5), without loss of generality, we may consider a set of five independent messages $\{W_{LL}, W_0, W_{HL}, W_{LH}, W_{HH}\}$ where W_{LL} is intended for (Y_{1L}, Y_{1L}), $\{W_{LL}, W_0, W_{HL}\}$ are intended for (Y_{1H}, Y_{2L}), $\{W_{LL}, W_0, W_{LH}\}$ is intended for (Y_{1L}, Y_{2H}), and $\{W_{LL}, W_0, W_{HL}, W_{HL}, W_{HH}\}$ are intended for (Y_{1H}, Y_{2H}). We thus have a parallel Gaussian broadcast channel with the message set $\{W_{LL}, W_0, W_{HL}, W_{LH}, W_{HH}\}$, which is equivalent to the channel (4.1) as illustrated in Figure 4.1. Based on the equivalent broadcast channel, the expected capacity of the channel (4.1) across two coherent blocks is characterized in the following theorem.

Theorem 4. *The expected capacity C_{exp} of the Gaussian block fading channel* (4.1) *across two coherent blocks is given by*

$$\begin{aligned} C_{exp} \;=\; \max_{0 \le \beta_1 \le \beta_2 \le 1} \Bigg[& p \log\left(\frac{\beta_1 + \sigma_H^2}{\sigma_H^2}\right) + \log\left(\frac{1 + \sigma_L^2}{\beta_2 + \sigma_L^2}\right) + \\ & \left(p - \frac{1}{2}p^2\right)\left(\log\left(\frac{\beta_2 + \sigma_H^2}{\beta_1 + \sigma_H^2}\right) + \log\left(\frac{\beta_2 + \sigma_L^2}{\beta_1 + \sigma_L^2}\right)\right)\Bigg]. \end{aligned} \tag{4.6}$$

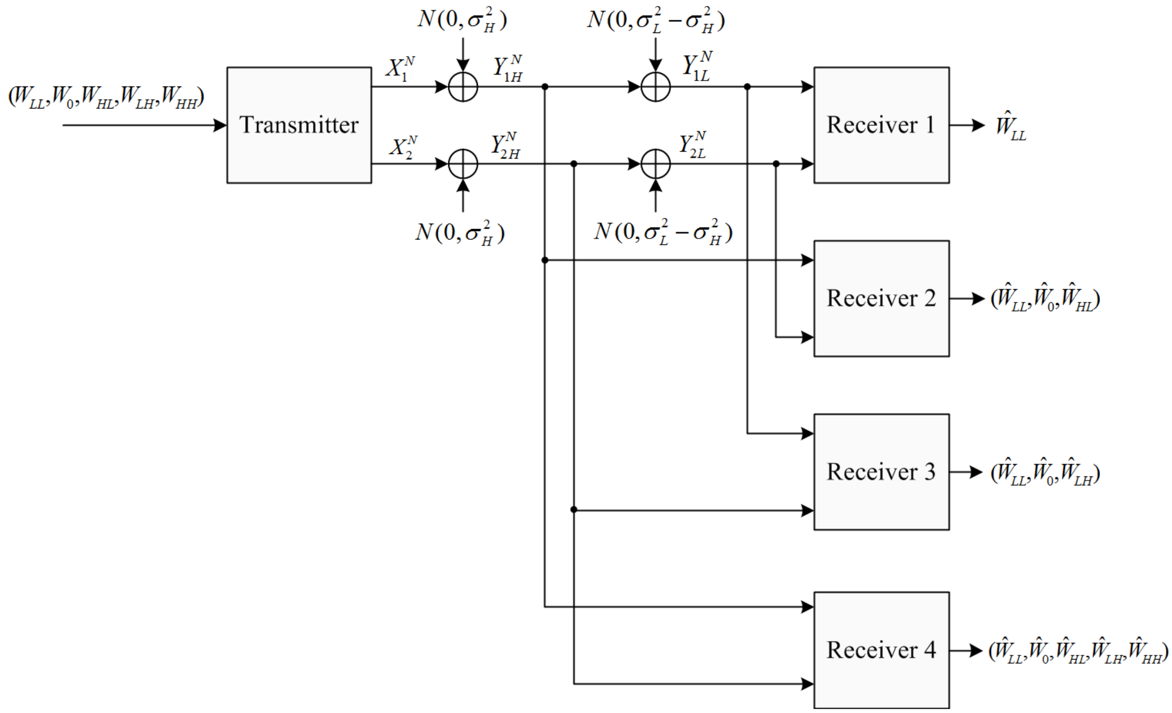

Figure 4.1: An equivalent Gaussian product broadcast channel.

Compared to the expected rate in [9], we can achieve

$$(p-p^2)\log\left(\frac{\beta_1+\sigma_H^2}{\sigma_H^2}\right) \tag{4.7}$$

extra expected bits per channel use.

4.2 Proof of the Main Result

Note that the expected achievable rate $\mathbb{E}[R]$ of the channel (4.1) for coding over two blocks is given by:

$$\begin{aligned}
\mathbb{E}[R] &= p^2(R_{HH} + R_{HL} + R_{LH} + R_0 + R_{LL}) + p(1-p)(R_{HL} + R_{LL} + R_0) + \\
&\quad p(1-p)(R_{LH} + R_{LL} + R_0) + (1-p)^2 R_{LL} \\
&= p^2(R_{HH} + R_{HL} + R_{LH}) + p(1-p)(R_{HL} + R_{LH}) + \\
&\quad (p^2 + 2p(1-p))R_0 + R_{LL} \\
&\leq (p^2 + p(1-p))(R_{HH} + R_{HL} + R_{LH}) + (p^2 + 2p(1-p))R_0 + R_{LL} \\
&= p(R_{HH} + R_{HL} + R_{LH}) + (2p - p^2)R_0 + R_{LL}. \qquad (4.8)
\end{aligned}$$

where R_{HH}, R_{HL}, R_{LH}, R_0, and R_{LL} denote achievable rates of the messages W_{HH}, W_{HL}, W_{LH}, W_0, and W_{LL} respectively.

4.2.1 The Converse

By Fano's inequality, we can bound each term on the right-hand side of (4.8) as:

$$\begin{aligned}
&2N(R_{HH} + R_{HL} + R_{LH} - \epsilon) \\
&\quad = H(W_{HH}, W_{HL}, W_{LH}) - 2N\epsilon \\
&\quad \leq I(W_{HH}, W_{HL}, W_{LH}; Y_{1H}^N, Y_{2H}^N | W_0, W_{LL}) \\
&\quad = h(Y_{1H}^N, Y_{2H}^N | W_0, W_{LL}) - h(Y_{1H}^N, Y_{2H}^N | W_{HH}, W_{HL}, W_{LH}, W_0, W_{LL}),
\end{aligned}$$

$$
\begin{aligned}
2N(R_0 - \epsilon) &= H(W_0) - 2N\epsilon \\
&\leq \min\left\{ I(W_0; Y_{1H}^N, Y_{2L}^N | W_{LL}), I(W_0; Y_{1L}^N, Y_{2H}^N | W_{LL}) \right\} \\
&\leq \frac{1}{2}\left[I(W_0; Y_{1H}^N, Y_{2L}^N | W_{LL}) + \frac{1}{2} I(W_0; Y_{1L}^N, Y_{2H}^N | W_{LL}) \right] \\
&= \frac{1}{2} h(Y_{1H}^N, Y_{2L}^N | W_{LL}) - \frac{1}{2} h(Y_{1H}^N, Y_{2L}^N | W_0, W_{LL}) + \\
&\quad \frac{1}{2} h(Y_{1H}^N, Y_{2L}^N | W_{LL}) - \frac{1}{2} h(Y_{1H}^N, Y_{2L}^N | W_0, W_{LL})
\end{aligned}
$$

and

$$
\begin{aligned}
2N(R_{LL} - \epsilon) &= H(W_{LL}) - 2N\epsilon \\
&\leq I(W_{LL}; Y_{1L}^N, Y_{2L}^N) \\
&= h(Y_{1L}^N, Y_{2L}^N) - h(Y_{1L}^N, Y_{2L}^N | W_{LL})
\end{aligned}
$$

where $\epsilon \to 0$ in the limit as $N \to \infty$. Thus, from (4.8), the expected rate $\mathbb{E}[R]$ can be bounded from above as

$$
\begin{aligned}
&2N\left(\mathbb{E}[R] - (3p - p^2 + 1)\epsilon\right) \\
&\quad \leq p\left[h(Y_{1H}^N, Y_{2H}^N | W_0, W_{LL}) - h(Y_{1H}^N, Y_{2H}^N | W_{HH}, W_{HL}, W_{LH}, W_0, W_{LL}) \right] + \\
&\quad \left(p - \frac{1}{2}p^2\right)\left[h(Y_{1H}^N, Y_{2L}^N | W_{LL}) - h(Y_{1H}^N, Y_{2L}^N | W_0, W_{LL}) + \right. \\
&\quad \left. h(Y_{1L}^N, Y_{2H}^N | W_{LL}) - h(Y_{1L}^N, Y_{2H}^N | W_0, W_{LL}) \right] + \\
&\quad \left[h(Y_{1L}^N, Y_{2L}^N) - h(Y_{1L}^N, Y_{2L}^N | W_{LL}) \right]. \qquad (4.9)
\end{aligned}
$$

Equivalently, we have

$$
\begin{aligned}
&2N\left(\mathbb{E}[R]-(3p-p^2+1)\epsilon\right)\\
&\quad\le\ p\left[h(Y_{1H}^N|W_0,W_{LL})+h(Y_{2H}^N|Y_{1H}^N,W_0,W_{LL})-\right.\\
&\qquad\left.h(Y_{1H}^N,Y_{2H}^N|W_{HH},W_{HL},W_{LH},W_0,W_{LL})\right]+\\
&\qquad\left(p-\frac{1}{2}p^2\right)\left[h(Y_{1H}^N|W_{LL})+h(Y_{2L}^N|Y_{1H}^N,W_{LL})-\right.\\
&\qquad\left(h(Y_{1H}^N|W_0,W_{LL})+h(Y_{2L}^N|Y_{1H}^N,W_0,W_{LL})\right)+\\
&\qquad\left(h(Y_{1L}^N|W_{LL})+h(Y_{2H}^N|Y_{1L}^N,W_{LL})\right)-\\
&\qquad\left.\left(h(Y_{1L}^N|W_0,W_{LL})+h(Y_{2H}^N|Y_{1L}^N,W_0,W_{LL})\right)\right]+\\
&\qquad\left[h(Y_{1L}^N,Y_{2L}^N)-\left(h(Y_{1L}^N|W_{LL})+h(Y_{2L}^N|Y_{1L}^N,W_{LL})\right)\right]. \qquad (4.10)
\end{aligned}
$$

Note from the Markov relationship

$$Y_{1L}^N - Y_{1H}^N - \{X_1^N, X_2^N\} - Y_{2H}^N - Y_{2L}^N$$

that

$$h(Y_{2L}^N|W_{LL},Y_{1H}^N) \le h(Y_{2L}^N|W_{LL},Y_{1L}^N)$$

and

$$h(Y_{2H}^N|W_0,W_{LL},Y_{1H}^N) \le h(Y_{2H}^N|W_0,W_{LL},Y_{1L}^N).$$

We thus have

$$
\begin{aligned}
&2N\left(\mathbb{E}[R]-(3p-p^2+1)\epsilon\right)\\
&\leq\ h(Y_{1L}^N,Y_{2L}^N)-ph(Y_{1H}^N,Y_{2H}^N|W_{HH},W_{HL},W_{LH},W_0,W_{LL})+\\
&\quad\left(p-\frac{1}{2}p^2\right)h(Y_{1H}^N|W_{LL})-\left(1-p+\frac{1}{2}p^2\right)h(Y_{1L}^N|W_{LL})+\\
&\quad\frac{1}{2}p^2h(Y_{1H}^N|W_0,W_{LL})-\left(p-\frac{1}{2}p^2\right)h(Y_{1L}^N|W_0,W_{LL})+\\
&\quad\left(p-\frac{1}{2}p^2\right)h(Y_{2H}^N|W_{LL},Y_{1L}^N)-\left(1-p+\frac{1}{2}p^2\right)h(Y_{2L}^N|W_{LL},Y_{1L}^N)+\\
&\quad\frac{1}{2}p^2h(Y_{2H}^N|W_0,W_{LL},Y_{1H}^N)-\left(p-\frac{1}{2}p^2\right)h(Y_{2L}^N|W_0,W_{LL},Y_{1H}^N). \qquad (4.11)
\end{aligned}
$$

Note that

$$h(Y_{1H}^N,Y_{2H}^N|W_{HH},W_{HL},W_{LH},W_0,W_{LL})=2N\log\left(\pi e\sigma_H^2\right). \qquad (4.12)$$

Furthermore, by the power constraint (4.2) we have

$$\mathbb{E}\left[\frac{1}{2N}\sum_{t=1}^{N}\left(|X_1[t]|^2+|X_2[t]|^2\right)\right]\leq 1.$$

Hence, without loss of generality we may assume that

$$\frac{1}{2N}\sum_{t=1}^{N}|X_i[t]|^2=\theta_i, \quad i=1,2$$

where $\theta_1\geq 0$, $\theta_2\geq 0$, and $\theta_1+\theta_2\leq 1$. It follows that

$$h(Y_{1L}^N,Y_{2L}^N)\leq\sum_{i=1}^{2}h(Y_{iL}^N)\leq\sum_{i=1}^{2}N\log\left(\pi e(2\theta_i+\sigma_L^2)\right). \qquad (4.13)$$

Further note that

$$
\begin{aligned}
N \log\left(\pi e \sigma_L^2\right) &= h(Y_{1L}^N | W_{HH}, W_{HL}, W_{LH}, W_0, W_{LL}) \\
&\leq h(Y_{1L}^N | W_0, W_{LL}) \\
&\leq h(Y_{1L}^N | W_{LL}) \\
&\leq h(Y_{1L}^N) \\
&\leq N \log\left(\pi e (2\theta_1 + \sigma_L^2)\right),
\end{aligned}
$$

so there exist $\beta_1^{(1)}$ and $\beta_2^{(1)}$ such that $0 \leq \beta_1^{(1)} \leq \beta_2^{(1)} \leq 1$,

$$
h(Y_{1L}^N | W_{LL}) = N \log\left(\pi e (2\theta_1 \beta_2^{(1)} + \sigma_L^2)\right) \tag{4.14}
$$

and

$$
h(Y_{1L}^N | W_0, W_{LL}) = N \log\left(\pi e (2\theta_1 \beta_1^{(1)} + \sigma_L^2)\right). \tag{4.15}
$$

By the conditional entropy power inequality, we have

$$
\begin{aligned}
N \log\left(\pi e (2\theta_1 \beta_2^{(1)} + \sigma_L^2)\right) &= h(Y_{1L}^N | W_{LL}) \\
&\geq N \log\left(2^{N h(Y_{1H}^N | W_{LL})} + \pi e (\sigma_L^2 - \sigma_H^2)\right)
\end{aligned}
$$

and

$$
\begin{aligned}
N \log\left(\pi e (2\theta_1 \beta_1^{(1)} + \sigma_L^2)\right) &= h(Y_{1L}^N | W_0, W_{LL}) \\
&\geq N \log\left(2^{N h(Y_{1H}^N | W_0, W_{LL})} + \pi e (\sigma_L^2 - \sigma_H^2)\right).
\end{aligned}
$$

Equivalently,

$$
h(Y_{1H}^N | W_{LL}) \leq N \log\left(\pi e (2\theta_1 \beta_2^{(1)} + \sigma_H^2)\right) \tag{4.16}
$$

and

$$h(Y_{1H}^N|W_0, W_{LL}) \leq N \log\left(\pi e(2\theta_1\beta_1^{(1)} + \sigma_H^2)\right). \tag{4.17}$$

Similarly, note that

$$\begin{aligned}
N\log\left(\pi e\sigma_L^2\right) &= h(Y_{2L}^N|W_{HH}, W_{HL}, W_{LH}, W_0, W_{LL}, Y_{1H}^N) \\
&\leq h(Y_{2L}^N|W_0, W_{LL}, Y_{1H}^N) \\
&\leq h(Y_{2L}^N|W_0, W_{LL}, Y_{1L}^N) \\
&\leq h(Y_{2L}^N|W_{LL}, Y_{1L}^N) \\
&\leq h(Y_{2L}^N) \\
&\leq N\log\left(\pi e(2\theta_2 + \sigma_L^2)\right)
\end{aligned}$$

so there exist $\beta_1^{(2)}$ and $\beta_2^{(2)}$ such that $0 \leq \beta_1^{(2)} \leq \beta_2^{(2)} \leq 1$

$$h(Y_{2L}^N|W_{LL}, Y_{1L}^N) = N\log\left(\pi e(2\theta_2\beta_2^{(2)} + \sigma_L^2)\right) \tag{4.18}$$

and

$$h(Y_{2L}^N|W_0, W_{LL}, Y_{1H}^N) = N\log\left(\pi e(2\theta_2\beta_1^{(2)} + \sigma_L^2)\right). \tag{4.19}$$

By the conditional entropy power inequality, we have

$$h(Y_{2H}^N|W_{LL}, Y_{1L}^N) \leq N\log\left(\pi e(2\theta_2\beta_2^{(2)} + \sigma_H^2)\right) \tag{4.20}$$

and

$$h(Y_{2H}^N|W_0, W_{LL}, Y_{1H}^N) \leq N\log\left(\pi e(2\theta_2\beta_1^{(2)} + \sigma_H^2)\right). \tag{4.21}$$

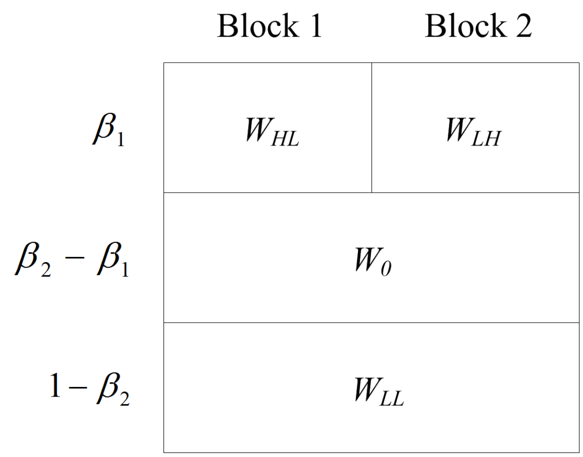

Figure 4.2: A three-layer superposition coding scheme for broadcasting over two coherent blocks

Substituting (4.12)–(4.21) ino (4.11), dividing both sides of the inequality by $2N$, and letting $N \to \infty$, we have

$$\mathbb{E}[R] \leq \max_{\left(\theta_1,\theta_2,\beta_1^{(1)},\beta_2^{(1)},\beta_1^{(2)},\beta_2^{(2)}\right)} R_1 + R_2 \tag{4.22}$$

where

$$\begin{aligned} R_i \;&:=\; \frac{1}{2}\left[p\log\left(\frac{2\theta_i\beta_1^{(i)}+\sigma_H^2}{\sigma_H^2}\right) + \log\left(\frac{2\theta_i+\sigma_L^2}{2\theta_i\beta_2^{(i)}+\sigma_L^2}\right) + \right. \\ &\left.\left(p-\frac{1}{2}p^2\right)\left(\log\left(\frac{2\theta_i\beta_2^{(i)}+\sigma_H^2}{2\theta_i\beta_1^{(i)}+\sigma_H^2}\right) + \log\left(\frac{2\theta_i\beta_2^{(i)}+\sigma_L^2}{2\theta_i\beta_1^{(i)}+\sigma_L^2}\right)\right)\right] \end{aligned} \tag{4.23}$$

for $i = 1, 2$. Using the "super code" argument provided in [9], it can be shown that without loss of optimality, we may assume that $\theta_1 = \theta_2 = 1/2$. Once we fix $\theta_1 = \theta_2 = 1/2$, by the symmetry of R_1 and R_2 we may assume without loss of optimality that $\beta_1^{(1)} = \beta_1^{(2)} = \beta_1$ and $\beta_2^{(1)} = \beta_2^{(2)} = \beta_2$. This completes the proof of the converse part of the theorem.

4.2.2 Achievability

The expected rate on the right-hand side of (4.6) can be achieved by a three-layer superposition coding as illustrated in Figure 4.2. The power allocations, from the bottom to the top layers, are given by $1-\beta_2$, $\beta_2-\beta_1$, and β_1, respectively. The bottom layer is used to encode the message W_{LL} and is of a total rate

$$\log\left(\frac{1+\sigma_L^2}{\beta_2+\sigma_L^2}\right).$$

It can be either a joint codebook or two separate codebooks, each for one coherent block. The middle layer is used to encode the message W_0 and is of a rate

$$\frac{1}{2}\log\left(\frac{\beta_2+\sigma_H^2}{\beta_1+\sigma_H^2}\right)+\frac{1}{2}\log\left(\frac{\beta_2+\sigma_L^2}{\beta_1+\sigma_L^2}\right).$$

It must be a joint codebook as it needs to be decodable when the received signals are either (Y_{1H}, Y_{2L}) or (Y_{1L}, Y_{2H}). The top layer consists of two separate codebooks, one for each coherent block and of a rate

$$\frac{1}{2}\log\left(\frac{\beta_1+\sigma_H^2}{\sigma_H^2}\right).$$

One of them is used to encode the message W_{HL}, and the other is used to encode the message W_{LH}. No power is allocated to encode the message W_{HH}. It is straightforward to verify that the achievable expected rate of the above scheme is indeed given by the right-hand side of (4.6). This completes the proof of the achievability part and hence the entire theorem.

Remark 1. *By comparison, the coding scheme considered in [9] is also a three-layer superposition coding scheme for which the bottom and the middle layers are the same* *as those considered in this dissertation. The difference is in the top layer, which uses a joint codebook to encode the message W_{HH}. Such a coding scheme is apparently inferior to the coding scheme considered in this dissertation, as the top layer is not decodable when the received signals are either (Y_{1H}, Y_{2L}) or (Y_{1L}, Y_{2H}), resulting a net loss of*

$$(p - p^2) \log \left(\frac{\beta_1 + \sigma_H^2}{\sigma_H^2} \right)$$

bits in expected rate.

4.3 Optimal Power Allocation

In the following proposition, an optimal power allocation for the expected capacity (4.6) is explicitly characterized.

Proposition 8 (An optimal power allocation)**.** *An optimal solution (β_1^*, β_2^*) to the optimization problem* (4.6) *is given by*

$$\beta_1^* = \min \left\{ \left(\frac{(1-(1-p))\sigma_L^2 - (1+(1-p))\sigma_H^2}{2(1-p)} \right)^+, 1 \right\} \tag{4.24}$$

and

$$\beta_2^* = \min \left\{ \left(\frac{(1-(1-p)^2)\sigma_L^2 - (1+(1-p)^2)\sigma_H^2}{2(1-p)^2} \right)^+, 1 \right\}. \tag{4.25}$$

Proof. Let us rewrite the expected capacity the expected capacity (4.6) as

$$\begin{aligned} C_{exp} = \max_{0 \le \beta_1 \le \beta_2 \le 1} & \left[p \log(1 + g_H \beta_1) + \log \left(\frac{1 + g_L}{1 + g_L \beta_2} \right) + \right. \\ & \left. \left(p - \frac{1}{2} p^2 \right) \left(\log \left(\frac{1 + g_H \beta_2}{1 + g_H \beta_1} \right) + \log \left(\frac{1 + g_L \beta_2}{1 + g_L \beta_1} \right) \right) \right] \end{aligned} \tag{4.26}$$

where $g_H := 1/\sigma_H^2$ and $g_L := 1/\sigma_L^2$ and thus $g_L < g_H$. The optimal solution (β_1^*, β_2^*) of the maximization problem in (4.26) can be obtained by considering MUFs [15].

Following [9], let us define the MUFs as

$$u_1(z) \;:=\; p \cdot \frac{g_H}{1+g_H z} \tag{4.27}$$

$$u_2(z) \;:=\; \left(p - \frac{p^2}{2}\right)\left(\frac{g_H}{1+g_H z} + \frac{g_L}{1+g_L z}\right) \tag{4.28}$$

$$u_3(z) \;:=\; \frac{g_L}{1+g_L z} \tag{4.29}$$

and the dominating MUF as

$$u^*(z) := \max_{l=1,2,3} u_l(z). \tag{4.30}$$

It is easy to see that MUFs $u_1(z)$, $u_2(z)$, and $u_3(z)$ have the *single crossing point property* as MUFs defined in Chapter 2. Clearly, $u_1(z)$ and $u_2(z)$ have a unique intersection at $z = z_1$ where

$$z_1 = \frac{(1-(1-p))g_H - (1+(1-p))g_L}{2(1-p)g_H g_L}. \tag{4.31}$$

Providing that $z_1 \geq 0$, $u_1(z) > u_2(z)$ if and only if $0 \leq z < z_1$ and $u_1(z) \leq u_2(z)$ if and only if $z \geq z_1$. Similarly, $u_2(z)$ and $u_3(z)$ have a unique intersection at $z = z_2$ where

$$z_2 = \frac{(1-(1-p)^2)g_H - (1+(1-p)^2)g_L}{2(1-p)^2 g_H g_L}. \tag{4.32}$$

Assuming that $z_2 \geq 0$, $u_2(z) > u_3(z)$ if and only if $0 \leq z < z_2$ and $u_2(z) \leq u_3(z)$ if and only if $z \geq z_2$. Note here that each of the MUFs can dominate the others on a single interval at most within $[0, 1]$ and that $z_1 \leq z_2$ since $g_L \leq g_H$. Thus, we can rewrite the dominating MUF $u^*(z)$ for $z \in [0,1]$ as

$$u^*(z) = \begin{cases} u_1(z), & \text{for} \quad 0 \le z < \min\{z_1^+, 1\} \\ u_2(z), & \text{for} \quad \min\{z_1^+, 1\} \le z < \min\{z_2^+, 1\} \\ u_3(z), & \text{for} \quad \min\{z_2^+, 1\} \le z < 1 \end{cases} \tag{4.33}$$

where $(\cdot)^+ := \max\{\cdot, 0\}$.

Now, we have

$$\begin{aligned} C_{exp} &= p \log\left(1 + g_H \beta_1^*\right) + \log\left(\frac{1+g_L}{1+g_L\beta_2^*}\right) + \\ &\quad \left(p - \frac{p^2}{2}\right)\left(\log\left(\frac{1+g_H\beta_2^*}{1+g_H\beta_1^*}\right) + \log\left(\frac{1+g_L\beta_2^*}{1+g_L\beta_1^*}\right)\right) && (4.34) \\ &= \int_0^{\beta_1^*} p \cdot \frac{g_H}{1+g_H z} dz + \int_{\beta_2^*}^1 \frac{g_L}{1+g_L z} dz + \\ &\quad \int_{\beta_1^*}^{\beta_2^*} \left(p - \frac{p^2}{2}\right)\left(\frac{g_H}{1+g_H z} + \frac{g_L}{1+g_L z}\right) dz && (4.35) \\ &= \sum_{l=1}^{3} \int_{\beta_{l-1}^*}^{\beta_l^*} u_l(z) dz && (4.36) \\ &\le \int_0^1 u^*(z) dz && (4.37) \end{aligned}$$

where $\beta_0 := 0$ and $\beta_3 := 1$. Note here that the inequality (4.37) holds with equality if and only if $\beta_1^* = \min\{z_1^+, 1\}$ and $\beta_2^* = \min\{z_2^+, 1\}$. Therefore, we have

$$C_{exp} = \int_0^1 u^*(z) dz \tag{4.38}$$

with the optimal power allocation

$$\beta_1^* = \min\left\{\left(\frac{(1-(1-p))g_H - (1+(1-p))g_L}{2(1-p)g_H g_L}\right)^+, 1\right\} \tag{4.39}$$

and

$$\beta_2^* = \min\left\{\left(\frac{(1-(1-p)^2)g_H - (1+(1-p)^2)g_L}{2(1-p)^2 g_H g_L b}\right)^+, 1\right\}. \tag{4.40}$$

This completes the proof of Proposition 8. □

5. THE CAPACITY REGION OF A PRODUCT GAUSSIAN BROADCAST CHANNEL WITH DEGRADED MESSAGE SETS*

5.1 Introduction

Broadcast is a fundamental nature of wireless communication: any receiver within the transmission range can listen to the source and potentially decode some of the messages. With appropriate coding architecture, the broadcast nature of wireless communication can be used to the advantage of simultaneously transmitting to several receivers at high rates. Understanding the limits and the appropriate coding architectures that can harness the broadcast advantage of wireless communication is an important subject of network information theory [28].

Most of the previous work focused on one of the following two scenarios:

1. to deliver the *same* messages to each of the receivers, usually known as the *multicast* problem; and

2. to deliver completely *distinct* messages to different receivers, namely the *private message* problem.

Formally, the distinction between these two broadcast scenarios can be identified by the configurations of the message sets associated with each of the receivers. For the multicast problem, the intended message sets for each of the receivers are *identical*. For the private message problem, the intended message sets for each of the receivers are *mutually exclusive*. Clearly, the appropriate coding architecture depends on the configurations of the message sets.

---

*Reprinted, with permission, from J. W. Yoo, T. Liu, and Y. Liang, "The capacity region of a product Gaussian broadcast channel with degraded message sets," in *Proc. 5th International ICST Conference on Communications and Networking in China*, Beijing, China, August 2010, Copyright ICST.

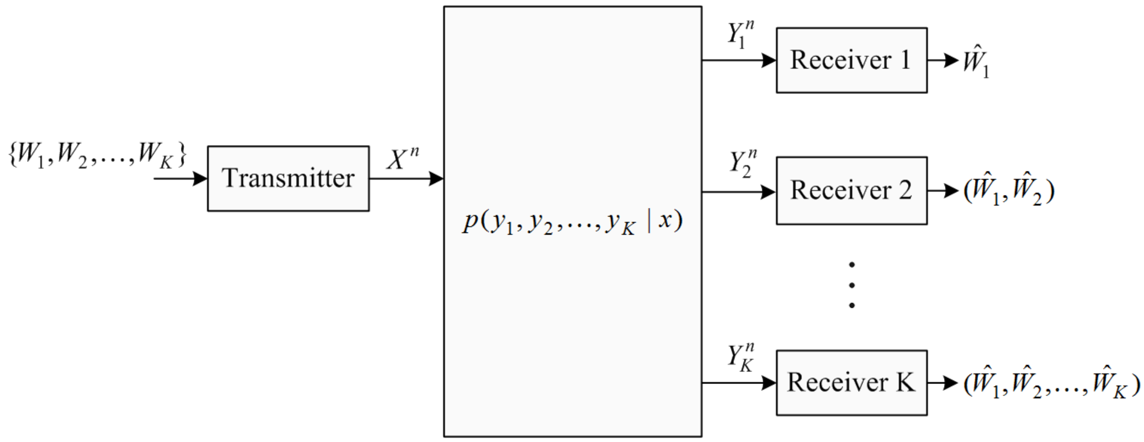

Figure 5.1: Broadcast channel with degraded message sets.

Between these two "extreme" broadcast scenarios, the multicast and the private message problems, there is a rich collection of "intermediate" problems with message sets of interesting configurations and significant engineering appeal. A good example is the *degraded message set* problems first considered in [29], which can be used to model broadcast scenarios with a progressively encoded source and receivers of different quality-of-service requirement.

Fig. 5.1 illustrates a general discrete memoryless broadcast channel with degraded message sets. The transmitter has a total of K independent messages $(M_1, M_2, \ldots, M_K)$. Each of the K receivers demands a subset of messages from the transmitter. The message set $\mathcal{S}_k$ intended for receiver k is given by

$$\mathcal{S}_k = \{M_1, M_2, \ldots, M_k\}, \quad k = 1, 2, \ldots, K.$$

Clearly, we have

$$\mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq \cdots \subseteq \mathcal{S}_K$$

and hence the name "degraded message sets".

For the degraded message set problem, there is a natural communication strategy based on superposition coding [30] and direct decoding. With K independent messages at the transmitter and K receivers, an K-layer superposition code can be built with the kth layer from the bottom representing message M_k. Receiver k decodes messages $(W_1, W_2, \ldots, W_k)$ by *directly* decoding all the bottom layers up to the kth. For $K = 2$, it was shown in [29] that this natural strategy is also optimal in achieving the capacity region of the channel. For $K \geq 3$, however, finding the capacity region of the discrete memoryless broadcast channel with degraded message sets remains an open problem in network information theory.

In an excellent contribution [14], Nair and El Gamal considered a special three-receiver discrete memoryless broadcast channel with degraded message sets and presented a precise single-letter characterization of the capacity region. Specifically, in [14], it was assumed that:

1. receiver 2 is degraded with respect to receiver 1, i.e., $X - Y_1 - Y_2$ forms a Markov for any input distribution $p(x)$; and

2. the rate of message M_2 is set to be zero so in defacto, there are only two independent messages M_1 and M_3 at the transmitter.

Under these two assumptions, Nair and El Gamal [14] proved a surprising result that the natural scheme that uses direct decoding is, in general, *suboptimal*. Instead, a coding scheme that uses *indirect decoding* [14] can always achieve the capacity region of the channel.

Building on the result of [14], we consider a specific product Gaussian broadcast channel with degraded message sets and provide an *explicit* characterization of the capacity region. The main tools used in this characterization are Lagrangian theory [31] and an extremal entropy inequality of Liu and Viswanath [32]. It is worth mentioning that the exact same product Gaussian model was also considered in the original work of Nair and El Gamal [14], and characterizing the capacity region was posted as an open problem.

5.2 Channel Model

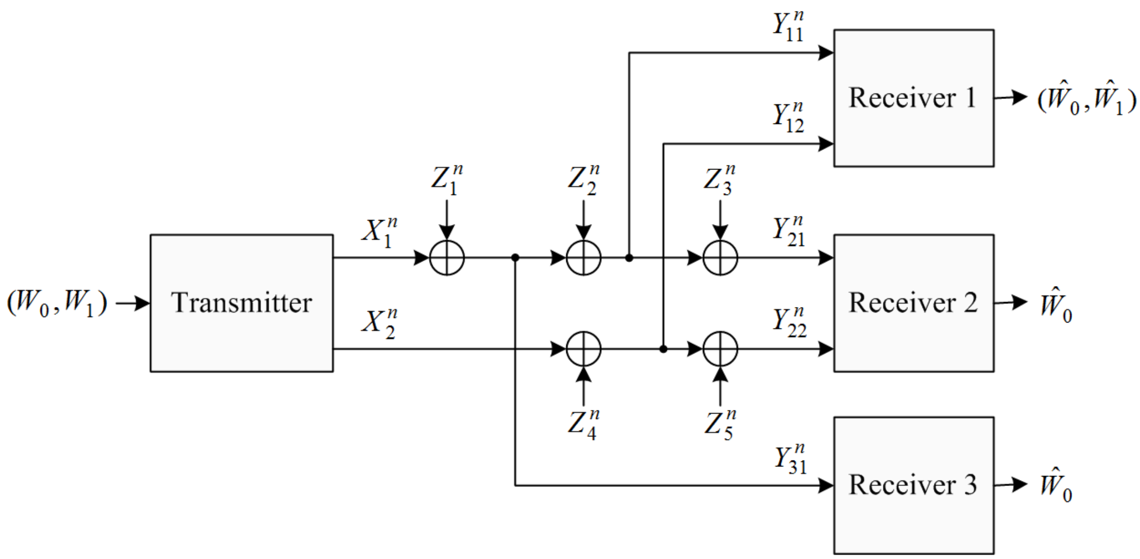

Figure 5.2: Product Gaussian broadcast channel with degraded message sets.

As shown in Fig. 5.2, consider a discrete-time memoryless product Gaussian broadcast channel with three receivers. At each time sample, the received signals at receivers 1, 2 and 3 are given by $Y_1 = (Y_{11}, Y_{12})$, $Y_2 = (Y_{21}, Y_{22})$ and $Y_3 = Y_{31}$, respectively, where

$$\begin{aligned} &Y_{31} = X_1 + Z_1, \quad Y_{11} = Y_{31} + Z_2, \quad Y_{21} = Y_{11} + Z_3 \\ &Y_{12} = X_2 + Z_4, \quad Y_{22} = Y_{12} + Z_5. \end{aligned} \tag{5.1}$$

Here, $X = (X_1, X_2)$ is the channel input, and Z_i, $i = 1, 2, 3, 4, 5$, are Gaussian noise with zero means with covariance N_i, respectively, and are assumed to be mutually independent of each other. We consider two different types of power constraints on the channel input X: an average total power constraint

$$E[X_1^2 + X_2^2] \leq P \tag{5.2}$$

and an individual per-subchannel power constraint

$$E[X_i^2] \leq P_i, \quad i = 1, 2. \tag{5.3}$$

The transmitter has two independent messages M_0 and M_1, where M_0 is a common message intended for all three receivers and M_1 is a private message intended only for receiver 1. The capacity region $\mathcal{C}(P)$ is given by the set of nonnegative rate pairs (R_0, R_1) that can be achieved by any coding scheme under the average total power constraint (5.2). Likewise, the capacity region $\mathcal{C}(P_1, P_2)$ is given by the set of nonnegative rate pairs (R_0, R_1) that can be achieved by any coding scheme under the individual per-subchannel power constraint (5.3).

From the channel model (5.1), it is clear that $X - Y_1 - Y_2$ forms a Markov for any distribution on the channel input X. In this case, a single-letter characterization of the capacity region was obtained in [14, Prop. 2] and is given by the set of nonnegative rate tuples (R_0, R_1) such that

$$\begin{aligned} R_0 &\leq I(U_1; Y_{21}) + I(U_2; Y_{22}) \\ R_0 &\leq I(V_1; Y_{31}) \\ R_1 &\leq I(X_1; Y_{11}|U_1) + I(X_2; Y_{12}|U_2) \\ R_0 + R_1 &\leq I(V_1; Y_{31}) + I(X_1; Y_{11}|V_1) + I(X_2; Y_{12}|U_2) \end{aligned} \tag{5.4}$$

for some joint distributions on (U_1, V_1, X_1) and (U_2, X_2) such that $U_1 - V_1 - X_1$ forms a Markov chain. The main goal is to evaluate the rate region (5.4) for the specific product Gaussian model (5.1) under both average total and individual per-subchannel power constraints.

5.3 Main Result

The main result of this chapter is an explicit characterization of the capacity region of the product Gaussian broadcast channel (5.1) under the individual per-subchannel power constraint (5.3), summarized in the following theorem.

Theorem 5. *The capacity region $\mathcal{C}(P_1, P_2)$ of the three-receiver product Gaussian broadcast channel* (5.1) *under the individual per-subchannel power constraint* (5.3) *is given by the set of nonnegative rate tuple (R_0, R_1) such that*

$$\begin{aligned}
R_0 &\leq C\left(\tfrac{P_1-Q_1}{Q_1+N_1+N_2+N_3}\right) + C\left(\tfrac{P_2-Q_2}{Q_2+N_4+N_5}\right) \\
R_0 &\leq C\left(\tfrac{P_1}{N_1}\right) \\
R_1 &\leq C\left(\tfrac{Q_1}{N_1+N_2}\right) + C\left(\tfrac{Q_2}{N_4}\right) \\
R_0 + R_1 &\leq C\left(\tfrac{P_1}{N_1}\right) + C\left(\tfrac{Q_2}{N_4}\right)
\end{aligned} \tag{5.5}$$

for some $0 \leq Q_1 \leq P_1$ and $0 \leq Q_2 \leq P_2$, where $C(x) := \frac{1}{2}\log(1+x)$.

As a corollary, we have the following characterization of the capacity region of the product Gaussian broadcast channel (5.1) under the average total power constraint (5.2).

Corollary 2. *The capacity region $\mathcal{C}(P)$ of the three-receiver product Gaussian broadcast channel* (5.1) *under the average total power constraint* (5.2) *is given by the set* *of nonnegative rate tuple* (R_0, R_1) *such that*

$$\begin{aligned} R_0 &\leq C\left(\tfrac{Q_3}{Q_1+N_1+N_2+N_3}\right) + C\left(\tfrac{Q_4}{Q_2+N_4+N_5}\right) \\ R_0 &\leq C\left(\tfrac{Q_1+Q_3}{N_1}\right) \\ R_1 &\leq C\left(\tfrac{Q_1}{N_1+N_2}\right) + C\left(\tfrac{Q_2}{N_4}\right) \\ R_0 + R_1 &\leq C\left(\tfrac{Q_1+Q_3}{N_1}\right) + C\left(\tfrac{Q_2}{N_4}\right) \end{aligned} \tag{5.6}$$

for some $Q_i \geq 0$, $i = 1, 2, 3, 4$, *and* $Q_1 + Q_2 + Q_3 + Q_4 \leq P$.

Proof. This is a simple consequence of Theorem 5 and the well-known fact that

$$\mathcal{C}(P) = \bigcup_{P_1+P_2 \leq P} \mathcal{C}(P_1, P_2).$$

□

5.4 Proof of the Main Result

The achievability of the rate region (5.5) follows from that of (5.4) by setting $X_i = U_i + W_i$ for $i = 1, 2$ and $V_1 = X_1$, where U_i and W_i are two independent Gaussian variables with zero means and variances $P_i - Q_i$ and Q_i, respectively. (Note that for such a choice of (U_1, V_1, X_1), $U_1 - V_1 - X_1$ forms a trivial Markov chain.) We therefore concentrate on proving the converse part of the theorem.

To prove the converse part of the theorem, we shall need the following extremal entropy inequality which first appeared in [32, Th. 8].

Lemma 4 ([32]). *Let* P *and* μ *be two nonnegative real numbers, and let* Z_1, Z_2 *be two Gaussian variables with zero means and variances* N_1 *and* N_2*, respectively.*

Assume that $0 < N_1 \le N_2$. If there exists a nonnegative real number P^* satisfying

$$\begin{aligned} (P^* + N_1)^{-1} + M_1 &= \mu(P^* + N_2)^{-1} + M_2 \\ M_1 P^* &= 0 \\ M_2(P - P^*) &= 0 \end{aligned}$$

for some nonnegative real numbers M_1 and M_2, then

$$\begin{aligned} & h(X + Z_1|U) - \mu h(X + Z_2|U) \\ & \quad \le \frac{1}{2}\log 2\pi e(P^* + N_1) - \frac{\mu}{2}\log 2\pi e(P^* + N_2) \end{aligned}$$

for any (X, U) independent of (Z_1, Z_2) and such that $E[X^2] \le P$.

We are now ready to prove the converse part of the theorem. Consider proof by contradiction. Let (R_0^o, R_1^o) be an *achievable* rate pair that lies *outside* the rate region (5.5). From [33], we have $R_0^o \le R_0^{max}$ where

$$R_0^{max} := \min\left\{ C\left(\tfrac{P_1}{N_1+N_2+N_3}\right) + C\left(\tfrac{P_2}{N_4+N_5}\right), C\left(\tfrac{P_1}{N_1}\right)\right\}.$$

Note that when $R_1^o = 0$, R_0^{max} can be achieved by letting $Q_1 = Q_2 = 0$ in (5.5). Thus, we may assume that $R_1^o > 0$ and write $R_1^o = R_1^* + \delta$ for some $\delta > 0$, where R_1^*

$$
\begin{aligned}
T_2(Q_1^* + N_1 + N_2)^{-1} + M_1 &= T_1(Q_1^* + N_1 + N_2 + N_3)^{-1} + M_2 && (5.7)\\
(Q_2^* + N_4)^{-1} + M_3 &= T_1(Q_2^* + N_4 + N_5)^{-1} + M_4 && (5.8)\\
T_2 + T_3 &= 1 && (5.9)\\
T_1 R_0^o &= T_1 \left[C\left(\frac{P_1 - Q_1^*}{Q_1^* + N_1 + N_2 + N_3} \right) + \right. \\
&\qquad \left. C\left(\frac{P_2 - Q_2^*}{Q_2^* + N_4 + N_5} \right) \right] && (5.10)\\
T_2 R_1^* &= T_2 \left[C\left(\frac{Q_1^*}{N_1 + N_2} \right) + C\left(\frac{Q_2^*}{N_4} \right) \right] && (5.11)\\
T_3(R_0^o + R_1^*) &= T_3 \left[C\left(\frac{P_1}{N_1} \right) + C\left(\frac{Q_2^*}{N_4} \right) \right] && (5.12)\\
M_1 Q_1^* &= 0 && (5.13)\\
M_2(P_1 - Q_1^*) &= 0 && (5.14)\\
M_3 Q_2^* &= 0 && (5.15)\\
M_4(P_2 - Q_2^*) &= 0 && (5.16)
\end{aligned}
$$

is given by

$$
\begin{aligned}
\max \quad & R_1 \\
\text{s.t.} \quad & R_0^o \leq C\left(\tfrac{P_1 - Q_1}{Q_1 + N_1 + N_2 + N_3}\right) + C\left(\tfrac{P_2 - Q_2}{Q_2 + N_4 + N_5}\right) \\
& R_1 \leq C\left(\tfrac{Q_1}{N_1 + N_2}\right) + C\left(\tfrac{Q_2}{N_4}\right) \\
& R_0^o + R_1 \leq C\left(\tfrac{P_1}{N_1}\right) + C\left(\tfrac{Q_2}{N_4}\right) \\
& Q_1 \leq P_1 \\
& Q_2 \leq P_2 \\
& -Q_1 \leq 0 \\
& -Q_2 \leq 0.
\end{aligned}
$$

Let (R_1^*, Q_1^*, R_2^*) be an optimal solution to the above optimization problem. Then, (R_1^*, Q_1^*, R_2^*) must satisfy the Karush-Kuhn-Tucker (KKT) conditions [31] as shown in the top of next page, where T_i, $i = 1, 2, 3, 4$, and M_i, $i = 1, 2, 3, 4$, are nonnegative

Lagrangian multipliers. From the KKT conditions (5.9)–(5.12), we have

$$
\begin{aligned}
&(T_1+T_3)R_0^o+R_1^o \\
&\quad= (T_1+T_3)R_0^o+R_1^*+\delta \\
&\quad= (T_1+T_3)R_0^o+(T_2+T_3)R_1^*+\delta \\
&\quad= T_1R_0^o+T_2R_1^*+T_3(R_0^o+R_1^*)+\delta \\
&\quad= T_1\left[C\left(\frac{P_1-Q_1^*}{Q_1^*+N_1+N_2+N_3}\right)+C\left(\frac{P_2-Q_2^*}{Q_2^*+N_4+N_5}\right)\right]+ \\
&\qquad T_2\left[C\left(\frac{Q_1^*}{N_1+N_2}\right)+C\left(\frac{Q_2^*}{N_4}\right)\right]+T_3\left[C\left(\frac{P_1}{N_1}\right)+C\left(\frac{Q_2^*}{N_4}\right)\right]+\delta \\
&\quad= T_1\left[C\left(\frac{P_1-Q_1^*}{Q_1^*+N_1+N_2+N_3}\right)+C\left(\frac{P_2-Q_2^*}{Q_2^*+N_4+N_5}\right)\right]+ \\
&\qquad T_2C\left(\frac{Q_1^*}{N_1+N_2}\right)+T_3C\left(\frac{P_1}{N_1}\right)+C\left(\frac{Q_2^*}{N_4}\right)+\delta.
\end{aligned}
\tag{5.17}
$$

On the other hand, by the KKT condition (5.9) and the assumption that (R_0^o, R_1^o) is achievable, we have

$$
\begin{aligned}
&(T_1+T_3)R_0^o+R_1^o \\
&\quad= (T_1+T_3)R_0^o+R_1^o \\
&\quad= (T_1+T_3)R_0^o+(T_2+T_3)R_1^o \\
&\quad= T_1R_0^o+T_2R_1^*+T_3(R_0^o+R_1^o) \\
&\quad\le T_1\left[I(U_1;Y_{21})+I(U_2;Y_{22})\right]+T_2\left[I(X_1;Y_{11}|U_1)+I(X_2;Y_{12}|U_2)\right]+ \\
&\qquad T_3\left[I(V_1;Y_{31})+I(X_1;Y_{11}|V_1)+I(X_2;Y_{12}|U_2)\right] \\
&\quad= T_1h(Y_{21})+T_1h(Y_{22})+T_3h(Y_{31})-\left[h(Y_{11}|X_1)+h(Y_{12}|X_2)\right]+ \\
&\qquad [T_2h(Y_{11}|U_1)-T_1h(Y_{21}|U_1)]+[h(Y_{12}|U_2)-T_1h(Y_{22}|U_2)]+ \\
&\qquad T_3[h(Y_{11}|V_1)-h(Y_{31}|V_1)]
\end{aligned}
\tag{5.18}
$$

for some joint distributions on (U_1, V_1, X_1) and (U_2, X_2) such that $U_1 - V_1 - X_1$ forms a Markov chain and $E[X_i^2] \leq P_i$ for $i = 1, 2$.

The terms on the right-hand side of the above equation can be further bounded/evaluated as follows.

1. It is well known [28] that Gaussian maximizes differential entropy for a given power, so we have

$$\begin{aligned} h(Y_{21}) &\leq \tfrac{1}{2}\log 2\pi e(P_1 + N_1 + N_2 + N_3) \\ h(Y_{22}) &\leq \tfrac{1}{2}\log 2\pi e(P_2 + N_4 + N_5) \\ h(Y_{31}) &\leq \tfrac{1}{2}\log 2\pi e(P_1 + N_1). \end{aligned} \tag{5.19}$$

2. The channel inputs (X_1, X_2) are independent of the Gaussian noise $(Z_1, Z_2, Z_3, Z_4, Z_5)$, so we have

$$\begin{gathered} h(Y_{11}|X_1) = h(Z_1 + Z_2) = \tfrac{1}{2}\log 2\pi e(N_1 + N_2) \\ h(Y_{12}|X_1) = h(Z_4) = \tfrac{1}{2}\log 2\pi e N_4. \end{gathered} \tag{5.20}$$

3. Putting together the KKT conditions (5.7), (5.13) and (5.14), we have

$$\begin{aligned} T_2(Q_1^* + N_1 + N_2)^{-1} + M_1 &= T_1(Q_1^* + N_1 + N_2 + N_3)^{-1} + M_2 \\ M_1 Q_1^* &= 0 \\ M_2(P_1 - Q_1^*) &= 0 \end{aligned}$$

where M_1, M_2, T_1 and T_2 are nonnegative real numbers. By Lemma 4[1], we

[1] If $T_2 = 0$, we have either $T_1 = 0$ or $Q_1^* = 0$. In either case, inequality (5.21) holds trivially.

have

$$
\begin{aligned}
&T_2 h(Y_{11}|U_1) - T_1 h(Y_{21}|U_1) \\
&= T_2 h(X_1 + Z_1 + Z_2|U_1) - T_1 h(X_1 + Z_1 + Z_2 + Z_3|U_1) \\
&\leq \frac{T_2}{2}\log 2\pi e(Q_1^* + N_1 + N_2) - \frac{T_1}{2}\log 2\pi e(Q_1^* + N_1 + N_2 + N_3). \quad (5.21)
\end{aligned}
$$

4. Similarly, putting together the KKT conditions (5.8), (5.15) and (5.16), we have

$$
\begin{aligned}
(Q_2^* + N_4)^{-1} + M_3 &= T_1(Q_2^* + N_4 + N_5)^{-1} + M_4 \\
M_3 Q_2^* &= 0 \\
M_4(P_2 - Q_2^*) &= 0
\end{aligned}
$$

where M_3, M_4 and T_1 are nonnegative real numbers. Again, by Lemma 4, we have

$$
\begin{aligned}
&h(Y_{12}|U_2) - T_1 h(Y_{22}|U_2) \\
&= h(X_2 + Z_4|U_2) - T_1 h(X_2 + Z_4 + Z_5|U_2) \\
&\leq \frac{1}{2}\log 2\pi e(Q_2^* + N_4) - \frac{T_1}{2}\log 2\pi e(Q_2^* + N_4 + N_5). \quad (5.22)
\end{aligned}
$$

5. Finally, note that

$$
\begin{aligned}
&h(Y_{11}|V_1) - h(Y_{31}|V_1) \\
&\quad = h(X_1 + Z_1 + Z_2|V_1) - h(X_1 + Z_1|V_1) \\
&\quad = I(Z_2; X_1 + Z_1 + Z_2|V_1) \\
&\quad = h(Z_2) - h(Z_2|X_1 + Z_1 + Z_2, V_1) &(5.23)\\
&\quad \leq h(Z_2) - h(Z_2|X_1 + Z_1 + Z_2, V_1, X_1) &(5.24)\\
&\quad = h(Z_2) - h(Z_2|Z_1 + Z_2, V_1, X_1) \\
&\quad = h(Z_2) - h(Z_2|Z_1 + Z_2) &(5.25)\\
&\quad = I(Z_2; Z_1 + Z_2) \\
&\quad = \frac{1}{2}\log 2\pi e(N_1 + N_2) - \frac{1}{2}\log 2\pi e N_1 &(5.26)
\end{aligned}
$$

where (5.23) is due to the independence of Z_2 and V_1; (5.24) is due to the fact that conditioning reduces differential entropy [28]; and (5.25) is due to the independence of (Z_1, Z_2) and (V_1, X_1).

Substitute (5.19)–(5.22) and (5.26) into (5.18). With some rearranging of terms, we may obtain

$$
\begin{aligned}
&(T_1 + T_3)R_0^o + R_1^o \\
&\quad \leq T_1\left[C\left(\frac{P_1 - Q_1^*}{Q_1^* + N_1 + N_2 + N_3}\right) + C\left(\frac{P_2 - Q_2^*}{Q_2^* + N_4 + N_5}\right)\right] + \\
&\quad\quad T_2 C\left(\frac{Q_1^*}{N_1 + N_2}\right) + T_3 C\left(\frac{P_1}{N_1}\right) + C\left(\frac{Q_2^*}{N_4}\right) &(5.27)
\end{aligned}
$$

Note that $\delta > 0$, so this is a contradiction to (5.17). We therefore conclude that any achievable rate pair (R_0^o, R_1^o) must also be *inside* the rate region (5.5). This completes the proof of the converse part of the theorem.

6. CONCLUSION

Delay is an important quality-of-service measure for the design of next-generation wireless networks. For delay-limited communication over block-fading channels, the difference between the ergodic capacity and the maximum achievable expected rate for coding over a finite number of coherent blocks represents a fundamental measure of the penalty incurred by the delay constraint.

This dissertation introduced a notion of worst-case expected-capacity loss. Focusing on the slow-fading scenario (one-block delay), it was shown that the worst-case additive expected-capacity loss is precisely $\log K$ nats per channel use and the worst-case multiplicative expected-capacity loss is precisely K, where K is the total number of different possible realizations of the power gain in each coherent block. Extension to the problem of writing on fading paper was also considered, where both the ergodic capacity and the additive expected-capacity loss over one-block delay were characterized to within one bit per channel use.

The problem with multiple-block delay is considerably more challenging. The main difficulty there is that the capacity region of the parallel Gaussian broadcast channel with a general message set configuration remains unknown. This dissertation presents two partial results. First, the expected capacity is precisely characterized for the point-to-point two-state fading channel with two-block delay. Second, the optimality of Gaussian superposition coding with indirect decoding is established for a two-parallel Gaussian broadcast channel with three receivers. Both results reveal some intrinsic complexity in characterizing the expected capacity with multiple-block delay.

Many research problems are open along the line of broadcasting over fading channels. Unlike for the case of one-block delay, the expected capacity of the point-to-point fading channel over multiple-block delay is unknown except for the case with two-block delay and two different possible realizations of the power gain in each coherent block, which is considered in Chapter 4 and in [9, 10]. With multiple transmit antennas, the expected capacity of the point-to-point fading channel is unknown even for one-block delay [3]. Another interesting and challenging scenario is the mixed-delay setting, where there are multiple messages of different delay requirement at the transmitter. Some preliminary results can be found in [34]. With known interference at the transmitter, one may also consider the setting where the channel fading applies only to the known interference (the fading-dirt problem) [35] or, more generally, different channel fading applies to the input signal and the known interference separately.

REFERENCES

[1] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[2] S. Shamai (Shitz), “A broadcast strategy for the Gaussian slowly fading channel,” in *Proc. IEEE Int. Symp. Inf. Theory*, Ulm, Germany, Jun.-Jul. 1997, p. 150.

[3] S. Shamai (Shitz) and A. Steiner, “A broadcast approach for a single-user slowly fading MIMO channel,” *IEEE Trans. Inf. Theory*, vol. 49, pp. 2617–2635, Oct. 2003.

[4] M. Effros and A. Goldsmith, “Capacity definitions and coding strategies for general channels with receiver side information,” in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Aug. 1998, p. 39.

[5] M. Effros, A. Goldsmith, and Y. Liang, “Generalizing capacity: New definitions and capacity theorems for composite channels,” *IEEE Trans. Inf. Theory*, vol. 56, pp. 3069–3087, Jul. 2010.

[6] S. Verdú and S. Shamai (Shitz), “Variable-rate channel capacity,” *IEEE Trans. Inf. Theory*, vol. 56, pp. 2651–2667, Jun. 2010.

[7] T. M. Cover, “Broadcast channels,” *IEEE Trans. Inf. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.

[8] P. P. Bergmans, “A simple converse for broadcast channels with additive white Gaussian noise,” *IEEE Trans. Inf. Theory*, vol. IT-20, pp. 279–280, Mar. 1974.

[9] P. A. Whiting and E. M. Yeh, "Broadcasting over uncertain channels with decoding delay constraints," *IEEE Trans. Inf. Theory*, vol. 52, pp. 904–921, Mar. 2006.

[10] A. Steiner, "On the broadcast approach over parallel channels," *Technical Report*, Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel, Oct. 19, 2006.

[11] A. Bennatan and D. Burstein, "On the fading paper achievable region of the fading MIMO broadcast channel," *IEEE Trans. Inf. Theory*, vol. 54, pp. 100–115, Jan. 2008.

[12] W. Zhang, S. Kotagiri and J. N. Laneman, "Writing on dirty paper with resizing and its application to quasi-static fading broadcast channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, Jun. 2007, pp. 381–385.

[13] S. Borade and L. Zheng, "Writing on fading paper and causal transmitter CSI," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, Jul. 2006, pp. 744–748.

[14] C. Nair and A. El Gammal, "The capacity region of a class of three-receiver broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 55, pp. 4479–4493, Oct. 2009.

[15] D. N. C. Tse, "Optimal power allocation over parallel Gaussian broadcast channels," *U.C. Berkeley Tech. Rep.*, UCB/ERL M99/7, 1999. Available online at http://www.eecs.berkeley.edu/Pubs/TechRpts/1999/3578.html

[16] A. Steiner and S. Shamai (Shitz), "Achievable rates with imperfect transmitter side information using a broadcast transmission strategy," *IEEE Trans. Wireless Communications*, vol. 7, pp. 1043–1051, Mar. 2008.

[17] A. Steiner and S. Shamai (Shitz), "Multi-layer broadcasting hybrid-ARQ strategies for block fading channels," *IEEE Trans. Wireless Communications*, vol. 7, pp. 2640–2650, Jul. 2008.

[18] C. Tian, A. Steiner, S. Shamai (Shitz), and S. N. Diggavi, "Successive refinement via broadcast: Optimizing expected distortion of a Gaussian source over a Gaussian fading channel," *IEEE Trans. Inf. Theory*, vol. 54, pp. 2903–2908, Jul. 2008.

[19] J. W. Yoo, T. Liu, and S. Shamai (Shitz), "Worst-case expected-rate loss of slow-fading channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 2012.

[20] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1073–1096, May 2003.

[21] R. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, pp. 5534–5562, Dec. 2008.

[22] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, pp. 1019–1030, Sep. 1990.

[23] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. Inf. Theory*, vol. 9, pp. 19–31, 1980.

[24] M. El-Halabi, T. Liu, C. Georghiades, and S. Shamai (Shitz), "Secret writing on dirty paper: A deterministic view," *IEEE Trans. Inf. Theory*, vol. 58, pp. 3419–3429, Jun. 2012.

[25] A. Khisti, A. Tchamkerten, and G. W. Wornell, "Secure broadcasting over fading channels," *IEEE Trans. Inf. Theory*, vol. 54, pp. 2453–2469, Jun. 2008.

[26] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, pp. 439–441, May 1983.

[27] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information," *IEEE Trans. Inf. Theory*, vol. 51, pp. 2867–2877, Aug. 2005.

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

[29] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. IT-23, pp. 60–64, Jan. 1977.

[30] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inf. Theory*, vol. IT-19, pp. 197–207, Mar. 1973.

[31] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

[32] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information-theoretic problems," *IEEE Trans. Inf. Theory*, vol. 53, pp. 1839–1851, May 2007.

[33] D. Blackwell, L. Breiman, and A. Thomasian, "The capacity of a class of channels," *Ann. Math. Statist.*, vol. 30, no. 4, pp. 1229–1241, Dec. 1959.

[34] K. Cohen, A. Steiner, and S. Shamai (Shitz), "The broadcast approach under mixed delay constraints," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 2012.

[35] A. Khina and U. Erez, "On the robustness of dirty paper coding," *IEEE Trans. Wireless Communications*, vol. 58, pp. 1437–1446, May 2010.

APPENDIX A

PROOF OF PROPOSITION 3

Let us first rewrite the expression (2.22) for the expected capacity $C_{exp}(F_G, 1)$ as follows:

$$\begin{aligned} C_{exp}(F_G, 1) &= \sum_{j=1}^{K} \left(\sum_{k=1}^{j} p_j \right) \log \left(\frac{n_j + \beta_j^*}{n_j + \beta_{j-1}^*} \right) && \text{(A.1)} \\ &= \sum_{k=1}^{K} p_k \left[\sum_{j=k}^{K} \log \left(\frac{n_j + \beta_j^*}{n_j + \beta_{j-1}^*} \right) \right] && \text{(A.2)} \\ &= \sum_{k=1}^{K} p_k \log \Lambda_k && \text{(A.3)} \end{aligned}$$

where

$$\Lambda_k = \prod_{j=k}^{K} \frac{n_j + \beta_j^*}{n_j + \beta_{j-1}^*} \tag{A.4}$$

and $(\beta_1^*, \ldots, \beta_K^*)$ is given by (2.31).

To show that Λ_k as given by (A.4) equals the right-hand side of (2.32), let us first assume that $s = w$. For this case, by (2.31) we have $\beta_j^* = \beta_{j-1}^*$ for every $j \neq \pi_s$. Thus, substituting (2.31) into (A.4) gives

$$\Lambda_k = \begin{cases} \frac{n_{\pi_s}+1}{n_{\pi_s}}, & \text{for} \quad 1 \leq k \leq \pi_s \\ 1, & \text{for} \quad \pi_s < k \leq K. \end{cases} \tag{A.5}$$

Next, let us assume that $s < w$. We shall consider the following three cases separately.

Case 1: $k \le \pi_s$. For this case, substituting (2.31) into (A.4) gives

$$\Lambda_k = \frac{n_{\pi_s} + z_{\pi_s,\pi_{s+1}}}{n_{\pi_s}} \left(\prod_{j=s+1}^{w-1} \frac{n_{\pi_j} + z_{\pi_j,\pi_{j+1}}}{n_{\pi_j} + z_{\pi_{j-1},\pi_j}} \right) \frac{n_{\pi_w} + 1}{n_{\pi_w} + z_{\pi_{w-1},\pi_w}} \tag{A.6}$$

$$= \frac{n_{\pi_w} + 1}{n_{\pi_s}} \prod_{j=s}^{w-1} \frac{n_{\pi_j} + z_{\pi_j,\pi_{j+1}}}{n_{\pi_{j+1}} + z_{\pi_j,\pi_{j+1}}} \tag{A.7}$$

$$= \frac{n_{\pi_w} + 1}{n_{\pi_s}} \prod_{j=s}^{w-1} \frac{F_{\pi_j}}{F_{\pi_{j+1}}} \tag{A.8}$$

$$= \frac{n_{\pi_w} + 1}{n_{\pi_s}} \frac{F_{\pi_s}}{F_{\pi_w}} \tag{A.9}$$

where (A.8) follows from the fact that the MUFs $u_{\pi_j}(z)$ and $u_{\pi_{j+1}}(z)$ intersect at $z = z_{\pi_j,\pi_{j+1}}$ so we have

$$\frac{n_{\pi_j} + z_{\pi_j,\pi_{j+1}}}{F_{\pi_j}} = \frac{n_{\pi_{j+1}} + z_{\pi_j,\pi_{j+1}}}{F_{\pi_{j+1}}} \iff \frac{n_{\pi_j} + z_{\pi_j,\pi_{j+1}}}{n_{\pi_{j+1}} + z_{\pi_j,\pi_{j+1}}} = \frac{F_{\pi_j}}{F_{\pi_{j+1}}}. \tag{A.10}$$

Case 2: $\pi_{m-1} < k \le \pi_m$ for some $m \in \{s+1, \ldots, w\}$. For this case, substituting substituting (2.31) into (A.4) gives

$$\Lambda_k = \left(\prod_{j=m}^{w-1} \frac{n_{\pi_j} + z_{\pi_j,\pi_{j+1}}}{n_{\pi_j} + z_{\pi_{j-1},\pi_j}} \right) \frac{n_{\pi_w} + 1}{n_{\pi_w} + z_{\pi_{w-1},\pi_w}} \tag{A.11}$$

$$= \frac{n_{\pi_w} + 1}{n_{\pi_m} + z_{\pi_{m-1},\pi_m}} \prod_{j=m}^{w-1} \frac{n_{\pi_j} + z_{\pi_j,\pi_{j+1}}}{n_{\pi_{j+1}} + z_{\pi_j,\pi_{j+1}}} \tag{A.12}$$

$$= \frac{n_{\pi_w} + 1}{n_{\pi_s} n_{\pi_m} + z_{\pi_{m-1},\pi_m}} \prod_{j=m}^{w-1} \frac{F_{\pi_j}}{F_{\pi_{j+1}}} \tag{A.13}$$

$$= \frac{n_{\pi_w} + 1}{n_{\pi_m} + z_{\pi_{m-1},\pi_m}} \frac{F_{\pi_m}}{F_{\pi_w}} \tag{A.14}$$

$$= \frac{n_{\pi_w} + 1}{n_{\pi_m} - n_{\pi_{m-1}}} \frac{F_{\pi_m} - F_{\pi_{m-1}}}{F_{\pi_w}} \tag{A.15}$$

where (A.13) follows from (A.10), and (A.15) follows from the fact that the MUFs

$u_{\pi_{m-1}}(z)$ and $u_{\pi_m}(z)$ intersect at $z = z_{\pi_{m-1},\pi_m}$ so by (2.12) we have

$$z_{\pi_{m-1},\pi_m} = \frac{F_{\pi_{m-1}} n_{\pi_m} - F_{\pi_m} n_{\pi_{m-1}}}{F_{\pi_m} - F_{\pi_{m-1}}} \iff \frac{F_{\pi_m}}{n_{\pi_m} + z_{\pi_{m-1},\pi_m}} = \frac{F_{\pi_m} - F_{\pi_{m-1}}}{n_{\pi_m} - n_{\pi_{m-1}}}. \quad \text{(A.16)}$$

Case 3: $k > \pi_w$. For this case, we have $\beta_j^* = \beta_{j-1}^* = 1$ for any $j \geq k$. Hence, by (2.31) we have

$$\Lambda_k = 1. \quad \text{(A.17)}$$

Finally, substituting (2.32) into (2.34) gives

$$C_{exp}(F_G, 1) \quad \text{(A.18)}$$

$$= \sum_{k=1}^{\pi_s} p_k \log \Lambda_{\pi_s} + \sum_{m=s+1}^{w} \left(\sum_{k=\pi_{m-1}+1}^{\pi_m} p_k \right) \log \Lambda_{\pi_m} \quad \text{(A.19)}$$

$$= F_{\pi_s} \log \Lambda_{\pi_s} + \sum_{m=s+1}^{w} \left(F_{\pi_m} - F_{\pi_{m-1}} \right) \log \Lambda_{\pi_m} \quad \text{(A.20)}$$

$$= F_{\pi_s} \log \left(\frac{n_{\pi_w} + 1}{n_{\pi_s}} \frac{F_{\pi_s}}{F_{\pi_w}} \right) + \sum_{m=s+1}^{w} \left(F_{\pi_m} - F_{\pi_{m-1}} \right) \log \left(\frac{n_{\pi_w} + 1}{n_{\pi_m} - n_{\pi_{m-1}}} \frac{F_{\pi_m} - F_{\pi_{m-1}}}{F_{\pi_w}} \right) \quad \text{(A.21)}$$

$$= F_{\pi_s} \log \left(\frac{F_{\pi_s}}{n_{\pi_s}} \right) + \sum_{m=s+1}^{w} \left(F_{\pi_m} - F_{\pi_{m-1}} \right) \log \left(\frac{F_{\pi_m} - F_{\pi_{m-1}}}{n_{\pi_m} - n_{\pi_{m-1}}} \right) + F_{\pi_w} \log \left(\frac{n_{\pi_w} + 1}{F_{\pi_w}} \right) \quad \text{(A.22)}$$

This completes the proof of Proposition 3.

APPENDIX B

PROOF OF LEMMA 2

Let us consider the following three cases separately.

Case 1: $k \le \pi_s$. For such k, by property 3) of Lemma 1 and the definition of s we have

$$\frac{F_k n_{\pi_s} - F_{\pi_s} n_k}{F_{\pi_s} - F_k} = z_{k,\pi_s} \le z_{\pi_{s-1},\pi_s} \le 0 \tag{B.1}$$

which implies that

$$\frac{n_{\pi_s}}{F_{\pi_s}} \le \frac{n_k}{F_k}. \tag{B.2}$$

By the expression of Λ_k from (2.32), for $k \le \pi_s$ we have

$$\begin{aligned} \frac{n_k+1}{n_k \Lambda_k} &= \frac{n_k+1}{n_{\pi_w}+1} \frac{F_{\pi_w} n_{\pi_s}}{F_{\pi_s} n_k} && \text{(B.3)} \\ &\le \frac{n_k+1}{n_{\pi_w}+1} \frac{F_{\pi_w}}{F_k} && \text{(B.4)} \\ &\le \frac{1}{p_k} && \text{(B.5)} \end{aligned}$$

where (B.4) follows from (B.2), and (B.5) follows from the fact that $n_k+1 \le n_{\pi_s}+1 \le n_{\pi_w}+1$, $F_{\pi_w} \le 1$, and $F_k \ge p_k$.

Case 2: $\pi_{m-1} < k \le \pi_m$ for some $m \in \{s+1, \ldots, w\}$. For such k, by (2.32) we have

$$\frac{n_k+1}{n_k \Lambda_k} = \frac{n_k+1}{n_{\pi_w}+1} \frac{n_{\pi_m} - n_{\pi_{m-1}}}{F_{\pi_m} - F_{\pi_{m-1}}} \frac{F_{\pi_w}}{n_k}. \tag{B.6}$$

By property 1) of Lemma 1 we have $z_{\pi_{m-1},\pi_m} \le z_{\pi_{m-1},k}$ which implies that

$$\frac{n_{\pi_m} - n_{\pi_{m-1}}}{F_{\pi_m} - F_{\pi_{m-1}}} = \frac{n_{\pi_{m-1}} + z_{\pi_{m-1},\pi_m}}{F_{\pi_{m-1}}} \le \frac{n_{\pi_{m-1}} + z_{\pi_{m-1},k}}{F_{\pi_{m-1}}} = \frac{n_k - n_{\pi_{m-1}}}{F_k - F_{\pi_{m-1}}}. \tag{B.7}$$

Substituting (B.7) into (B.6) gives

$$\frac{n_k+1}{n_k\Lambda_k} \le \frac{n_k+1}{n_{\pi_w}+1}\frac{n_k-n_{\pi_{m-1}}}{F_k-F_{\pi_{m-1}}}\frac{F_{\pi_w}}{n_k} \le \frac{1}{p_k} \tag{B.8}$$

where the last inequality follows from the fact that $n_k+1 \le n_{\pi_m}+1 \le n_{\pi_w}+1$, $n_k - n_{\pi_{m-1}} \le n_k$, $F_{\pi_w} \le 1$, and $F_k - F_{\pi_{m-1}} \ge p_k$.

Case 3: $k > \pi_w$. For such k, by (2.32) we have $\Lambda_k = 1$ and hence

$$\frac{n_k+1}{n_k\Lambda_k} = \frac{n_k+1}{n_k}. \tag{B.9}$$

By property 1) of Lemma 1 and the definition of w, we have $1 \le z_{\pi_w,\pi_{w+1}} \le z_{\pi_w,k}$, which implies that

$$n_k+1 \le n_k + z_{\pi_w,k} = \frac{F_k(n_k-n_{\pi_w})}{F_k-F_{\pi_w}}. \tag{B.10}$$

Substituting (B.10) into (B.9) gives

$$\frac{n_k+1}{n_k\Lambda_k} \le \frac{F_k}{F_k-F_{\pi_w}}\frac{n_k-n_{\pi_w}}{n_k} \le \frac{1}{p_k} \tag{B.11}$$

where the last inequality follows from the fact that $n_k - n_{\pi_w} \le n_k$, $F_k \le 1$, and $F_k - F_{\pi_w} \ge p_k$.

Combining the above three cases completes the proof of Lemma 2.

APPENDIX C

PROOF OF LEMMA 3

Let us begin by establishing a simple lower bound on the expected capacity $C_{exp}(F_G, 1)$. Applying the long-sum inequality

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{C.1}$$

we have

$$F_{\pi_s} \log\left(\frac{F_{\pi_s}}{n_{\pi_s}}\right) + \sum_{m=s+1}^{w} \left(F_{\pi_m} - F_{\pi_{m-1}}\right) \log\left(\frac{F_{\pi_m} - F_{\pi_{m-1}}}{n_{\pi_m} - n_{\pi_{m-1}}}\right) \geq F_{\pi_w} \log\left(\frac{F_{\pi_w}}{n_{\pi_w}}\right). \tag{C.2}$$

Substituting (C.2) into the expression of $C_{exp}(F_G, 1)$ from (2.35), we have

$$\begin{aligned} C_{exp}(F_G, 1) &\geq F_{\pi_w} \log\left(\frac{F_{\pi_w}}{n_{\pi_w}}\right) + F_{\pi_w} \log\left(\frac{n_{\pi_w} + 1}{F_{\pi_w}}\right) && \text{(C.3)} \\ &= F_{\pi_w} \log\left(\frac{n_{\pi_w} + 1}{n_{\pi_w}}\right). && \text{(C.4)} \end{aligned}$$

Next we shall prove the desired inequality (2.91) by considering the following four cases separately.

Case 1: $k > \pi_w$. For such k, by property 1) of Lemma 1 and the definition of w we have $z_{\pi_w,k} \geq z_{\pi_w,\pi_{w+1}} \geq 1$ and hence

$$\frac{n_{\pi_w} + 1}{F_{\pi_w}} \leq \frac{n_{\pi_w} + z_{\pi_w,k}}{F_{\pi_w}} = \frac{n_k - n_{\pi_w}}{F_k - F_{\pi_w}}. \tag{C.5}$$

Thus

$$\frac{p_k \log\left(\frac{n_k+1}{n_k}\right)}{C_{exp}(F_G,1)} \leq \frac{p_k \log\left(\frac{n_k+1}{n_k}\right)}{F_{\pi_w} \log\left(\frac{n_{\pi_w}+1}{n_{\pi_w}}\right)} \quad \text{(C.6)}$$

$$\leq \frac{p_k}{F_{\pi_w}} \frac{n_{\pi_w}+1}{n_k} \quad \text{(C.7)}$$

$$\leq \frac{p_k}{n_k} \frac{n_k - n_{\pi_w}}{F_k - F_{\pi_w}} \quad \text{(C.8)}$$

$$\leq 1 \quad \text{(C.9)}$$

where (C.6) follows from (C.4), (C.7) is due to the well-know inequalities (2.109) so $\log\left(\frac{n_k+1}{n_k}\right) \leq \frac{1}{n_k}$, and $\log\left(\frac{n_{\pi_w}+1}{n_{\pi_w}}\right) \geq \frac{1}{n_{\pi_w}+1}$, (C.8) follows from (C.5), and (C.9) is due to the fact that $n_k - n_{\pi_w} \leq n_k$ and $F_k - F_{\pi_w} \geq p_k$.

Case 2: $k = \pi_w$. For such k, by (C.4) we have

$$\frac{\log\left(\frac{n_k+1}{n_k}\right)}{C_{exp}(F_G,1)} \leq \frac{\log\left(\frac{n_{\pi_w}+1}{n_{\pi_w}}\right)}{F_{\pi_w} \log\left(\frac{n_{\pi_w}+1}{n_{\pi_w}}\right)} = \frac{1}{F_{\pi_w}} \quad \text{(C.10)}$$

and hence

$$\frac{p_k \log\left(\frac{n_k+1}{n_k}\right)}{C_{exp}(F_G,1)} \leq \frac{p_{\pi_w}}{F_{\pi_w}} \leq 1. \quad \text{(C.11)}$$

Case 3: $k = \pi_m$ for some $m \in \{s, \ldots, w-1\}$. For this case, we shall show that for any $m \in \{s, \ldots, w-1\}$

$$\frac{\log\left(\frac{n_{\pi_m}+1}{n_{\pi_m}}\right)}{C_{exp}(F_G,1)} \leq \frac{1}{F_{\pi_m}} \quad \text{(C.12)}$$

and hence

$$\frac{p_{\pi_m} \log\left(\frac{n_{\pi_m}+1}{n_{\pi_m}}\right)}{C_{exp}(F_G,1)} \leq \frac{p_{\pi_m}}{F_{\pi_m}} \leq 1. \quad \text{(C.13)}$$

To prove (C.12), let us define $g(z) := N(z)/D(z)$ where

$$N(z) = \log\left(\frac{n_{\pi_m}+z}{n_{\pi_m}}\right) \tag{C.14}$$

$$D(z) = F_{\pi_s}\log\left(\frac{F_{\pi_s}}{n_{\pi_s}}\right) + \sum_{i=s+1}^{w}\left(F_{\pi_i}-F_{\pi_{i-1}}\right)\log\left(\frac{F_{\pi_i}-F_{\pi_{i-1}}}{n_{\pi_i}-n_{\pi_{i-1}}}\right) + F_{\pi_w}\log\left(\frac{n_{\pi_w}+z}{F_{\pi_w}}\right). \tag{C.15}$$

By Lemma 1 and the definition of s and w, we have

$$0 < z_{\pi_s,\pi_{s+1}} \le z_{\pi_m,\pi_{m+1}} \le z_{\pi_m,\pi_w} \le z_{\pi_{w-1},\pi_w} < \text{SNR}. \tag{C.16}$$

By the expression of $C_{exp}(F_G,1)$ from (2.35), we have

$$\frac{\log\left(\frac{n_{\pi_m}+1}{n_{\pi_m}}\right)}{C_{exp}(F_G,1)} = g(1) \le \sup_{z\ge z_{\pi_m,\pi_w}} g(z) \tag{C.17}$$

where the last inequality follows from the fact that $z_{\pi_m,\pi_w} < 1$ as mentioned in (C.16). Next, we shall show that $g(z) \le 1/F_{\pi_m}$ at the boundary points $z = z_{\pi_m,\pi_w}$ and $z = \infty$, and for any *local maximum* $z^* > z_{\pi_m,\pi_w}$. We may then conclude that

$$\sup_{z\ge z_{\pi_m,\pi_w}} g(z) \le 1/F_{\pi_m}. \tag{C.18}$$

First, since $m < w$ we have

$$g(\infty) = 1/F_{\pi_w} \le 1/F_{\pi_m}. \tag{C.19}$$

Next, to show that $g(z_{\pi_m,\pi_w}) \le 1/F_{\pi_m}$, let us apply the log-sum inequality (152) to obtain

$$F_{\pi_s} \log\left(\frac{F_{\pi_s}}{n_{\pi_s}}\right) + \sum_{i=s+1}^{m} \left(F_{\pi_i} - F_{\pi_{i-1}}\right) \log\left(\frac{F_{\pi_i} - F_{\pi_{i-1}}}{n_{\pi_i} - n_{\pi_{i-1}}}\right) \geq F_{\pi_m} \log\left(\frac{F_{\pi_m}}{n_{\pi_m}}\right) \quad \text{(C.20)}$$

and

$$\sum_{i=m+1}^{w} \left(F_{\pi_i} - F_{\pi_{i-1}}\right) \log\left(\frac{F_{\pi_i} - F_{\pi_{i-1}}}{n_{\pi_i} - n_{\pi_{i-1}}}\right) \geq (F_{\pi_w} - F_{\pi_m}) \log\left(\frac{F_{\pi_w} - F_{\pi_m}}{n_{\pi_w} - n_{\pi_m}}\right). \quad \text{(C.21)}$$

Substituting (C.20) and (C.21) into (C.15) gives

$$\begin{aligned}
D(z_{\pi_m,\pi_w}) &\geq F_{\pi_m} \log\left(\frac{F_{\pi_m}}{n_{\pi_m}}\right) + (F_{\pi_w} - F_{\pi_m}) \log\left(\frac{F_{\pi_w} - F_{\pi_m}}{n_{\pi_w} - n_{\pi_m}}\right) + \\
&\quad F_{\pi_w} \log\left(\frac{n_{\pi_w} + z_{\pi_m,\pi_w}}{F_{\pi_w}}\right) & \text{(C.22)} \\
&= F_{\pi_m} \log\left(\frac{F_{\pi_m}}{n_{\pi_m}} \frac{n_{\pi_w} - n_{\pi_m}}{F_{\pi_w} - F_{\pi_m}}\right) + \\
&\quad F_{\pi_w} \log\left(\frac{F_{\pi_w} - F_{\pi_m}}{n_{\pi_w} - n_{\pi_m}} \frac{n_{\pi_w} + z_{\pi_m,\pi_w}}{F_{\pi_w}}\right) & \text{(C.23)} \\
&= F_{\pi_m} \log\left(\frac{n_{\pi_m} + z_{\pi_m,\pi_w}}{F_{\pi_m}}\right) & \text{(C.24)} \\
&= F_{\pi_m} N(z_{\pi_m,\pi_w}) & \text{(C.25)}
\end{aligned}$$

where (C.24) follows from the fact that the MUFs $u_{\pi_m}(z)$ and $u_{\pi_w}(z)$ intersect at $z = z_{\pi_m,\pi_w}$ so we have

$$\frac{F_{\pi_m}}{n_{\pi_m} + z_{\pi_m,\pi_w}} = \frac{F_{\pi_w}}{n_{\pi_w} + z_{\pi_m,\pi_w}} = \frac{F_{\pi_w} - F_{\pi_m}}{n_{\pi_w} - n_{\pi_m}}. \quad \text{(C.26)}$$

It follows immediately from (C.25) that

$$g(z_{\pi_m,\pi_w}) = N(z_{\pi_m,\pi_w})/D(z_{\pi_m,\pi_w}) \leq 1/F_{\pi_m}. \quad \text{(C.27)}$$

Finally, to show that $g(z^*) \le 1/F_{\pi_m}$ for any local maximum $z^* > z_{\pi_m,\pi_w}$, let us note that $g(z)$ is continuous and differentiable for all $z > z_{\pi_m,\pi_w}$ so z^* must satisfy

$$\left.\frac{d}{dz} g(z)\right|_{z^*} = 0 \tag{C.28}$$

or equivalently

$$\left.\frac{dN(z)}{dz} D(z)\right|_{z^*} = \left.\frac{dD(z)}{dz} N(z)\right|_{z^*}. \tag{C.29}$$

We thus have

$$\begin{aligned}
g(z^*) &= \frac{N(z^*)}{D(z^*)} && \text{(C.30)}\\
&= \left.\frac{dN(z)/dz}{dD(z)/dz}\right|_{z^*} && \text{(C.31)}\\
&= \frac{1}{F_{\pi_w}} \frac{n_{\pi_w} + z^*}{n_{\pi_m} + z^*} && \text{(C.32)}\\
&\le \frac{1}{F_{\pi_w}} \frac{n_{\pi_w} + z_{\pi_m,\pi_w}}{n_{\pi_m} + z_{\pi_m,\pi_w}} && \text{(C.33)}\\
&= \frac{1}{F_{\pi_m}} && \text{(C.34)}
\end{aligned}$$

where (C.33) follows from the facts that $n_{\pi_w} > n_{\pi_m}$ so $\frac{n_{\pi_w}+z}{n_{\pi_m}+z}$ is a monotone decreasing function of z for $z \ge 0$ and that $z^* \ge z_{\pi_m,\pi_w} > 0$, and (C.34) follows from (C.26).

Substituting (C.18) into (C.17) completes the proof of the desired inequality (C.12) for Case 3.

Case 4: $k < \pi_w$ but $k \neq \pi_i$ for any $i = s, \ldots, w-1$. For such k, let m be the *smallest* integer from $\{s, \ldots, w\}$ such that $k < \pi_m$. Note that

$$\frac{p_k \log\left(\frac{n_k+1}{n_k}\right)}{C_{exp}(F_G, 1)} = \frac{p_k \log\left(\frac{n_k+1}{n_k}\right)}{\log\left(\frac{n_m+1}{n_m}\right)} \frac{\log\left(\frac{n_m+1}{n_m}\right)}{C_{exp}(F_G, 1)} \tag{C.35}$$

$$\leq \frac{p_k \log\left(\frac{n_k+1}{n_k}\right)}{F_{\pi_m} \log\left(\frac{n_{\pi_m}+1}{n_{\pi_m}}\right)} \tag{C.36}$$

$$= \frac{p_k}{F_{\pi_m}} f(1) \tag{C.37}$$

where (C.36) follows from (C.10) for $m = w$ and from (C.12) for $m = s, \ldots, w-1$, and

$$f(z) := \frac{\log\left(\frac{n_k+z}{n_k}\right)}{\log\left(\frac{n_{\pi_m}+z}{n_{\pi_m}}\right)}. \tag{C.38}$$

Since $n_k < n_{\pi_m}$, $f(z)$ is a monotone decreasing function for $z > 0$. By Lemma 1 and the definition of w, we have

$$z_{k,\pi_m} \leq z_{\pi_{m-1},\pi_m} \leq z_{\pi_{w-1},\pi_w} < 1. \tag{C.39}$$

We shall consider the following two sub-cases separately.

Sub-case 4.1: $z_{k,\pi_m} > 0$. By the monotonicity of $f(z)$ and the fact that SNR $> z_{k,\pi_m} > 0$ as mentioned in (C.39), we have

$$f(1) \leq f(z_{k,\pi_m}) = \frac{\log\left(\frac{n_k+z_{k,\pi_m}}{n_k}\right)}{\log\left(\frac{n_{\pi_m}+z_{k,\pi_m}}{n_{\pi_m}}\right)} \leq \frac{n_{\pi_m} + z_{k,\pi_m}}{n_k} \tag{C.40}$$

where the last inequality follows from the inequalities (2.109) so we have $\log\left(\frac{n_k+z_{k,\pi_m}}{n_k}\right) \leq \frac{z_{k,\pi_m}}{n_k}$ and $\log\left(\frac{n_{\pi_m}+z_{k,\pi_m}}{n_{\pi_m}}\right) \geq \frac{z_{k,\pi_m}}{n_{\pi_m}+z_{k,\pi_m}}$. By Lemma 1 and the fact that $k < \pi_m$, we have $z_{\pi_{m-1},\pi_m} \geq z_{k,\pi_m} > 0$ and hence $m \geq s+1$. Therefore, $k \neq \pi_{m-1}$ and we must have $k > \pi_{m-1}$. Again, by Lemma 1 we have $z_{k,\pi_m} \leq z_{\pi_{m-1},\pi_m} \leq z_{\pi_{m-1},k}$ and hence

$$\frac{n_{\pi_m} + z_{k,\pi_m}}{F_{\pi_m}} = \frac{n_k + z_{k,\pi_m}}{F_k} \leq \frac{n_k + z_{\pi_{m-1},k}}{F_k} = \frac{n_k - n_{\pi_{m-1}}}{F_k - F_{\pi_{m-1}}}. \tag{C.41}$$

Substituting (C.41) into (C.40) gives

$$f(1) \leq \frac{F_{\pi_m}(n_k - n_{\pi_{m-1}})}{n_k(F_k - F_{\pi_{m-1}})} \leq \frac{F_{\pi_m}}{F_k - F_{\pi_{m-1}}}. \tag{C.42}$$

Further substituting (C.42) into (C.37) gives

$$\frac{p_k \log\left(\frac{n_k+1}{n_k}\right)}{C_{exp}(F_G, 1)} \leq \frac{p_k}{F_k - F_{\pi_{m-1}}} \leq 1. \tag{C.43}$$

Sub-case 4.2: $z_{k,\pi_m} \leq 0$. In this case, $z_{k,\pi_m} = \frac{F_k n_{\pi_m} - F_{\pi_m} n_k}{F_{\pi_m} - F_k} \leq 0$ so we have $F_k n_{\pi_m} \leq F_{\pi_m} n_k$. By the monotonicity of $f(z)$ and the fact that SNR > 0, we have

$$f(1) \leq \lim_{z \downarrow 0} f(z) = \frac{n_{\pi_m}}{n_k} \leq \frac{F_{\pi_m}}{F_k}. \tag{C.44}$$

Substituting (C.44) into (C.37) gives

$$\frac{p_k \log\left(\frac{n_k+1}{n_k}\right)}{C_{exp}(F_G, 1)} \leq \frac{p_k}{F_k} \leq 1. \tag{C.45}$$

Combining the above two sub-cases completes the proof for Case 4. We have thus completed the proof of Lemma 3.