Information Theoretic Aspects of Wireless Networks with Coverage Constraint

by

Seyed Reza Mirghaderi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2007

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Seyed Reza Mirghaderi

Abstract

A wireless multicast network with a stringent decoding delay constraint and a minimum coverage requirement is characterized when the fading channel state information is available only at the receiver side. In the first part, the optimal expected rate achievable by a random user in the network is derived in a single antenna system in terms of the minimum multicast requirement in two scenarios: hard coverage constraint and soft coverage constraint. In the first case, the minimum multicast requirement is expressed by multicast outage capacity while in the second case, the expected multicast rate should satisfy the minimum requirements. Also, the optimum power allocation in an infinite layer superposition code, achieving the highest expected typical rate, is derived. For the MISO case, a suboptimal coding scheme is proposed, which is shown to be asymptotically optimal, when the number of transmit antennas grows at least logarithmically with the number of users in the network. In the second part, a joint source-channel coding scheme is motivated, where a multi-resolution Gaussian source code is mapped to a multi-level channel code. In this part, the hard and soft coverage constraints are defined as maximum outage multicast distortion and maximum expected multicast distortion, respectively. In each scenario, the minimum expected distortion of a typical user is derived in terms of the corresponding coverage constraint. The minimization is first performed for the finite state fading channels and then is extended to the continuous fading channels.

Acknowledgements

Special thanks to Prof. Khandani, from whom I learnt a lot not only about Information Theory, but also about life. I also appreciate unforgettable helps from my friend Alireza Bayesteh.

To Shirin
and my parents

Contents

List of Figures

Chapter 1

Introduction

1.1 Multicast Networks with Coverage Constraint

Wireless networks have recently received a considerable attention. The widespread applications of these networks, along with the specific circumstances of wireless communication, have motivated efficient transmission strategies for each application. One of these applications is data multicasting. In a wireless multicast system, a common source is transmitted to N users, through a fading channel. In such a network, the transmitter should ideally provide coverage to all the users. However, this will decrease the *average* quality of service received per user since the coverage is highly correlated to the user with the worst channel condition. Therefore, two issues can be studied as a measure of performance: network coverage and quality of service. In the first case, the objective is to cover all the nodes in the network at least with a basic service, regardless of their channel qualities. From this point of view, all the users have basically the same opportunity to receive data. However, in the second case, the average quality of service is the main objective. Therefore, users with better channel status should receive higher data rates and consequently, better quality of services. A good example for such networks is a TV broadcasting system [16]. In this system, all the subscribers are supposed to receive a basic video signal, while users with higher channel qualities might get additional services like high definition TV signal. The coverage issue in such systems is generally addressed as multicast minimum requirement.

Multicast applications (e.g. video multicasting) usually have a decoding delay constraint. Hence, the transmission time is smaller than the length of fading block and each user experiences a single realization of the channel during the transmission time. Assuming no Channel State Information at the Transmitter (no CSIT), the stringent decoding delay constraint leads to non-ergodicity of the fading channels.

1.2 Maximum Achievable Rates in a Multicast Network

Multicasting has been recently studied as a special scenario in broadcasting, where all the users are listening to a common source. In [2], the system challenges in lossy broadcasting of a common source are studied from information theoretical point of view. For an analog Gaussian source with a bandwidth equal to the channel bandwidth, analog transmission achieves the minimum average end-to-end distortion. The scenario in which the source has a larger bandwidth is studied in [3], where different methods of digital transmission are investigated. In [4], a different approach to source broadcasting, called static broadcasting, is proposed. It is assumed that all the users receive the same amount of data from a common source but with different number of channel uses, according to their channel qualities. However, the actual transmission time in this definition depends on the user with the lowest channel gain, and hence, the transmission rate might be very low for large number of users.

Since the performance of a multicast network is strongly affected by the user with the worst channel condition, we are motivated to define a more *fair* approach. We consider a wireless multicast network in a slowly fading Gaussian environment. The objective is to maximize the average performance while a multicast constraint is satisfied. Average performance is defined as the service received by a randomly chosen user (typical user) in the network, while the multicast requirement is the service received by all the users. These two requirements in a multicast network define a tradeoff, since the first one deals with a typical user of the network while the second depends on the worst channel state in the system. We assume the transmission block is large enough to yield a reliable communication. However, averaging over time is not possible because of the delay constraint. In other words, all the symbols within a transmission block experience the same channel gain. The channel state information (CSI) of each user is assumed to be known only at the receiver end. In this case, the ergodic capacity is not defined since the channel dose not have an ergodic behavior. The outage capacity [1] is defined for such channels as the maximum rate of single layered data, decodable with a high probability. In [10], a broadcast approach for a single user channel with these assumptions is proposed which optimizes the expected decodable rate. We will apply both "outage capacity" and "expected rate" definitions to characterize our network. Outage capacity is exploited when we have a hard coverage constraint on multicast data. In this case, we want to assure that a specific amount of data is conveyed within one transmission block to all the users, with a high probability. We relax the coverage constraint by stating it in terms of *expected* delivered rate to all the users within one block. In both cases we maximize the expected typical rate.

This minimum-service based approach has been studied in [6] for a single user fading channel, assuming CSI is known at the transmitter. In that work, given a service outage constraint for a real time application, the average rate is maximized for a non real time application sent on top of it. An adaptive variable rate code is proposed and shown to be optimum in that scenario. Similarly, a minimum rate constrained capacity measure is defined for broadcast channels in [5]. It is shown that the minimum rate capacity region is the ergodic capacity region of a broadcast channel, with an effective noise determined by the minimum rate requirements.

We will investigate the proposed multicast system in both SISO and MISO case. The MISO multicast asymptotical capacity limits are examined in [8], when the CSI is available at the transmitter. It is shown that the adverse effect of large number of users could be compensated by increasing the number of transmitter antennas. We will study similar scenario in our network and explore the effect of using multiple antennas.

1.3 Layered Joint Source-Channel Code for a Multicast Network

When we have an analog source, the quality of service can be measured in terms of end-to-end distortion of the reconstructed source at the receiver. In a SISO ergodic channel, the source-channel separation theorem [20], implies that minimum end-to-end distortion is achieved when the source is compressed at the rate equal to the channel capacity and sent over the channel with asymptotically zero probability of error. In our network, we assume a stringent decoding delay constraint with no CSIT. Therefore, the channels are non-ergodic and source-channel separation is not necessarily optimal. However, assuming the separation of source and channel code, one can reduce the end-to-end distortion by optimizing the mapping of source code to the channel code based on the source and channel characteristics. It is shown in [21], that a superposition code maximizes the expected typical rate in terms of multicast requirement in such networks. In this work, we consider a multi-resolution Gaussian source code mapped to such a multi-level channel code and optimize the expected typical distortion given the multicast coverage constraint. Each layer of the multi-resolution source code successively refines the description in its previous layer. The transmitter allocates any layer of source code an appropriate power level and superimposes them as a multilevel channel code. This multi-level code is sent to all the users in the network. Each receiver decodes the layers of code supported by its channel condition. The power allocation of the transmitter should minimize the expected distortion of a typical user while satisfying the coverage constraint.

As in [21], we define two different coverage constraint scenarios. In the first scenario, we have a hard coverage constraint where all the users should reconstruct the source with a distortion less than a specific level, with a high probability. In fact, the multicast requirement is expressed in terms of multicast *outage* distortion. In the second case, we relax the constraint to the *expected* distortion received by all the users. We first assume the channels have finite number of states and explore the optimal power allocation for this case. We then extend our results for the finite state channels to the continuous fading channels.

For a continuous fading channel with a stringent decoding delay constraint, an infinite layer superposition coding (layered broadcasting) scheme was first proposed in [10], where the optimal power allocation was derived for a SISO system. However, reference [10] dose not consider the end-to-end distortion but the channel expected rate. It is shown in [22] that in the transmission of a Gaussian source over a Gaussian channel, when the source bandwidth is equal to the channel bandwidth, uncoded transmission is optimal [23]. For other bandwidth ratios, hybrid digital-analog joint source-channel codes are designed in [24] to be optimal for a specific target SNR threshold below which the performance of the codes degrades drastically. Using the similar broadcast strategy as in [10], a linear-time power allocation algorithm is proposed in [25] to minimize the expected distortion of a Gaussian source over a fading channel. In [26], a delay-limited system is characterized in terms of minimum expected distortion when layered broadcasting is used with successive refinement of the source. We consider a system similar to [26] but with multiple number of users each with an independent fading channel when all are receiving a common message.

1.4 Thesis Outline

The rest of this thesis is organized in two parts. The first part (Chapter 2) considers the problem of maximum achievable expected rates in a wireless multicast network with a coverage constraint. In the second part (Chapter 3), we propose a layered joint source-channel coding scheme for transmission of a Gaussian source to N users in a multicast network with a coverage constraint.

Chapter 2 is organized as follows: in section 2.1, the system model is elaborated. Section 2.2 focuses on the Virtual Broadcast model for an unknown fading channel when the network is delay limited. Sections 2.3 and 2.4 are specified to characterization of multicast network when we have a single antenna at the transmitter and at each receiver. In section 2.3, we evaluate the optimal performance of the network in terms of the achievable expected typical rate and the multicast outage capacity. In other words, this section describes the hard multicast coverage constraint scenario. Section 2.4 corresponds to a soft multicast coverage constraint, where the expected multicast rate decoded in a block should satisfy the minimum requirement. In this scenario, we will explore the achievable expected typical rate. Section 2.5, investigates the MISO case, where we derive the asymptotical capacity limits for the multicast network. In section 2.6, we extend the our results to the case where multiple sources transmit data each to a group of users through a shared wireless channel.

Chapter 3 is is organized as follows: the system model is presented in Section 3.1. This section also focuses on virtual broadcast approach for channel coding used for the transmission of a source with successive refinement. Section 3.2 is specified to the characterization of the networks with finite state channels. In this section, we first preview the unconstrained problem [26] and then propose the optimal power allocation with two scenarios of multicast coverage constraint, namely hard coverage constraint and soft coverage constraint. We extend our results for discrete state channel to a continuous fading channel in section 3.3.

In Chapter 4 we conclude the thesis and propose the future work related to this subject.

Chapter 2

Maximum Achievable Rates in a Multicast Network

2.1 System Model

In this chapter, we consider a common message broadcasting network, where a single-antenna transmitter sends a common data to N single-antenna receivers. The received signal at the jth receiver, denoted by y_j can be written as

$$y_j = s_j x + n_j, \tag{2.1}$$

where $\{x\}$ is the transmitted signal with the total average power constraint $E[x^2] \leq \mathscr{P}$, $\{n_j\} \sim \mathcal{CN}(0,1)$ is the Additive White Gaussian Noise (AWGN) at this receiver, and $s_j \sim \mathcal{CN}(0,1)$ is the channel coefficient from the transmitter to the jth receiver. Therefore, the channel gain $h_j = |s_j|^2$ has the following CDF:

$$F_j(h) = 1 - e^{-h},$$

and is assumed to be constant during the transmission block. The typical (average) channel of the multicast network is defined as the channel of a randomly selected user. Since all the channels are i.i.d., the typical channel gain distribution is identical to that of each channel, i.e.,

$$F_{typ}(h) = F_j(h) = F(h). \tag{2.2}$$

Since all the N channels are Gaussian and they receive a common signal, the multicast channel is equivalent to the worst channel in the network. Due to statistical independence of the channels, the gain of that user has the following distribution:

$$\Pr\left\{\min_i(h_i) > h\right\} = \left(\Pr\left\{h_i > h\right\}\right)^N = e^{-Nh}.$$

As a result, we have:

$$F_{mul}(h) = F_{\min_i(h_i)}(h) = 1 - e^{-Nh}.$$

In this part, we are dealing with three measures defined in our network, as follows:

- the multicast outage rate, R_ϵ, the rate decodable at the multicast channel with probability $(1-\epsilon)$,
- the expected multicast rate, $R_{mul} = E_{h_{mul}}[R(h)]$, where $h_{mul} = \min_i(h_i)$, and $R(h)$ is the decodable data rate for the channel state h,
- the expected typical rate, $R_{ave} = E_{h_{typ}}[R(h)]$.

2.2 Broadcast Model for an Unknown Fading Channel

In [10], it is shown that the expected rate for a receiver with a block fading channel, unknown at the transmitter, and a stringent delay constraint, is equivalent to a weighted sum rate of a degraded broadcast channel with infinite number of receivers, each corresponding to a realization of the channel. In this chapter, we exploit the same model in a more general fashion. Regarding the frequent use of this model in our network characterization, we will study it in detail.

The ergodic capacity of a block fading Gaussian channel when the CSI is known at receiver is

$$C_{ergodic} = E_h[\frac{1}{2}\log(1 + h\mathscr{P})]$$

The capacity is achieved by a single Gaussian codebook with power $\mathscr{P}$. The codebook should be long enough to let the channel experience almost all its possible states. In fact, the assumption of ergodicity makes averaging over fading blocks possible. However, in our scenario the fading block is infinitely large. In other words, each receiver is experiencing a single fading level during the whole period of transmission. Hence, for any coding scheme we have a function $R(h)$ which determines the data rate decoded in channel state h. Regarding the degraded nature of the Gaussian channels, this function is increasing. Therefore

$$R(h) - R(h - \mathrm{d}h) = \mathrm{d}R_h \geq 0$$

Consider an infinite number of differently indexed virtual receivers, such that receiver r_h is experiencing a fading level between h and $h + \mathrm{d}h$. With these settings, r_h is receiving all the data received by $r_{h-\mathrm{d}h}$, in addition to $\mathrm{d}R_h$. The virtual receivers introduce a degraded broadcast network in which the rate associated with user r_h is $\mathrm{d}R_h$. The actual user selects receiver r_h with probability $f(h)\mathrm{d}h$, where $f(h)$ is the channel gain distribution function and uses that receiver for the whole transmission.

With this interpretation, for a given coding scheme, the distinction between different channels introduced in the previous section is their different probability distribution of virtual receiver selection. Given the fact that both multicast and typical channels deal with the same signalling, all the measures defined in the previous section could be written as follows:

$$\begin{aligned} R_{ave} &= \int_0^\infty R(h)f(h)\mathrm{d}h = \int_0^\infty (1 - F(h))\mathrm{d}R_h, \\ R_{mul} &= \int_0^\infty (1 - F_{mul}(h))\mathrm{d}R_h, \\ R_\epsilon &= R(h_\epsilon) = \int_0^{h_\epsilon} \mathrm{d}R_h, \end{aligned} \tag{2.3}$$

where $h_\epsilon = F_{mul}^{-1}(\epsilon)$. The first two derivations are statistical rate averaging over different selected receivers. In the case of multicast channel, the selected receiver has a channel gain lower than h_ϵ with probability ϵ and hence, the highest decodable rate is $R(h_\epsilon)$, with probability $1 - \epsilon$. As seen above, the performance measures in our network are three different positive weighted sum rate of the virtual broadcast network which forms a performance vector. More precisely, in the hard coverage constraint scenario (section IV), the performance vector is defined by the couple $[R_\epsilon, R_{ave}]$, and in the soft coverage constraint scenario (section V), by $[R_{mul}, R_{ave}]$. In the following, we will propose a search space for the virtual broadcast rate vector which results in the optimal performance vector. Before that, we should give a definition for the optimality of a performance vector.

Deffinition 1 *The boundary set B_1 of a closed convex region $R_1 \subset {R^+}^n$, is defined as*

$$B_1 = \{x \in R_1 | \nexists x' \in {R^{++}}^n, x' \neq \mathbf{0}, x + x' \in R_1\}$$

where R^+ and R^{++} are the set of nonnegative and strictly positive real numbers, respectively.

With the above definition, a performance vector is optimal if it is in the boundary set of all possible performance vectors. In the following theorem, we show that the optimal performance vector for each case is achieved by super-position coding in which the rate of the virtual receiver r_h is given by

$$\mathrm{d}R_h = \log\left(1 + \frac{h\rho(h)\mathrm{d}h}{1 + h\int_h^\infty \rho(u)\mathrm{d}u}\right),$$

where $\rho(h)$ is the power allocation function.

Theorem 1 *The boundary set of $[R_{w_1}, ..., R_{w_k}]$, where $R_{w_i} = \int_0^\infty w_i(h)\mathrm{d}r_h$, are positive weighted sum rate of the underlying virtual broadcast channel, is achievable by a super-position coding scheme.*

Proof: In order to prove the theorem, we first state and prove the following lemma:

Lemma 1 *Consider a mapping function $g(.)$ from a closed region $R_1 \subset {R^+}^n$ to $R_2 \subset {R^+}^k$, such that $g(x) = Mx$, where $M \in {R^+}^k \times {R^+}^n$. Denote B_1 and B_2 the boundary sets of regions R_1 and R_2, respectively. We have*

$$B_2 \subset g(B_1)$$

Proof: Assume this is not true. Hence there must exist $x_2 \in B_2$ such that $x_2 \notin g(B_1)$ and $x_1 \in R_1$, such that $x_2 = g(x_1)$. Since $x_1 \notin B_1$, there exists $x_1' \in B_1$ such that $x_1' - x_1 \in R^{n++}$. Defining $x_2' = g(x_1') \in R_2$,

$$x_2' - x_2 = M(x_1' - x_1) \subset R^{k+} \tag{2.4}$$

which contradicts the fact that x_2 is in the boundary set of R_2 and the lemma is proved. ■

In conclusion of Lemma 1, if we let n tend to infinity, the Matrix transform will tend to k weighted sums of an infinite dimension vector $x(h)$ as follows:

$$Mx \to \left[\int_0^\infty w_1(h)x(h)\mathrm{d}h, \cdots, \int_0^\infty w_k(h)x(h)\mathrm{d}h\right].$$

Setting $x(h) = \mathrm{d}R_h$, we can conclude that the boundary region of $[R_{w_1}, \cdots, R_{w_k}]$ is a subset of the transformation of the boundary set of rate vector $[\mathrm{d}R_h]$, which is achieved, as shown in [16], by the superposition coding. In other words, for any vector $\mathbf{v}$ in the boundary set of $[R_{w_1}, \cdots, R_{w_k}]$, there exists a scalar positive function $\rho^{\mathbf{v}}(h)$ such that

$$\mathrm{d}R_h^{\mathbf{v}} = \log\left(1 + \frac{h\rho^{\mathbf{v}}(h)\mathrm{d}h}{1 + h\int_h^\infty \rho^{\mathbf{v}}(u)\mathrm{d}u}\right),$$

in which $\mathrm{d}R_h^{\mathbf{v}}$ satisfies $\mathbf{v} = [\int w_1(h)\mathrm{d}R_h^{\mathbf{v}}, \cdots, \int w_k(h)\mathrm{d}R_h^{\mathbf{v}}]$, and $\int_0^\infty \rho^{\mathbf{v}}(u)\mathrm{d}u = \mathscr{P}$. This completes the proof of Theorem 1. ■

Using the above theorem, it easily follows that the optimal performance vectors $[R_\epsilon, R_{ave}]$ and $[R_{mul}, R_{ave}]$, defined for the sections IV and V, respectively, are achieved by superposition coding.

2.3 Hard Coverage Constraint

In this section, we consider a scenario where the multicast data has a high priority. Hence, it should be delivered to all the users in the network with a high probability $(1-\epsilon)$, where ϵ is the outage probability of the system. In this case, any loss of the multicast data by any user is defined as a coverage outage. Given this constraint, we want to maximize the average rate received by a randomly chosen user in the network. This average rate includes the expected rate of all data received by a typical user, even if the user is in outage. However, we will show that for a small enough outage probability, the users in the outage do not contribute to the expected average rate (it is optimum not to allocate them any power). In this scenario, we deal with two channels: (i) a multicast channel for which we want to guarantee an outage rate R_ϵ, and (ii) an average channel for which the highest expected rate R_{ave} is desired.

Setting $w_1(h) = 1_{\{h \le h_\epsilon\}}$ and $w_2(h) = 1 - F(h)$, Theorem 1 states that the boundary set of $[R_\epsilon, R_{ave}]$ is achieved by superposition coding, in which

$$\mathrm{d}R_h = \log\left(1 + \tfrac{h\rho(h)\mathrm{d}h}{1+hI(h)}\right) = \int_{I(h)}^{I(h)+\rho(h)\mathrm{d}h} \tfrac{h\mathrm{d}p}{1+hp}, \tag{2.5}$$

and $I(h) = \int_h^\infty \rho(u)\mathrm{d}u$. Note that $\mathrm{d}R_h$ is not necessarily very small since our power allocation function might have some impulses in the general case. As stated before, we want to jointly optimize the weighted sum of these rates according to the weighting functions $w_1(h)$ and $w_2(h)$. The optimization is on the function $I(h)$ and $\rho(h)$. However, we can simplify our optimization problem to a point optimization. Let us define $s(p)$ as

$$s(p) = \max\left\{h \mid I(h) \ge p\right\}.$$

In fact the above function is the inverse of the interference function in in terms of the channel level and could be called *channel gain-interference function*. It is evident that it is a decreasing function of p. According to (2.5), we can write the expected rate as

$$R_{ave}^{s(.)} = \int_0^\infty (1 - F(h))\mathrm{d}R_h = \int_0^{\mathscr{P}} g(p, s(p))\mathrm{d}p, \tag{2.6}$$

where $g(x, y) = (1 - F(y))\frac{y}{1+xy}$. Differentiating this function with respect to y, we get

$$\frac{\partial}{\partial y} g(x, y) = \frac{1 - F(y) - yf(y)(1 + xy)}{(1 + xy)^2}. \tag{2.7}$$

Since $g(x, y)$ is a concave function of x,

$$\arg\max(g(x, y)|_{x=p}) = I_0^{-1}(p), \tag{2.8}$$

where $I_0(h) = \frac{(1-F(h))-hf(h)}{h^2 f(h)}$. Moreover, $g(x, y)|_{x=p}$ is increasing for $y < I_0^{-1}(p)$, and decreasing elsewhere.

Let us define $P_\epsilon^{s(.)}$ for the function $s(.)$ as

$$P_\epsilon^{s(.)} = \min\left\{p | s(p) \le h_\epsilon\right\}.$$

For simplicity, we assume $h_\epsilon \le 1$.With the above definitions, our problem is translated to find

$$\max{}_{s(.)} R_{ave}^{s(.)} = \max_{s(.)} \int_0^{\mathscr{P}} g(p, s(p))\mathrm{d}p, \tag{2.9}$$

subject to

$$R_\epsilon^{s(.)} = \int_{P_\epsilon^{s(.)}}^{\mathscr{P}} m(p, s(p))\mathrm{d}p \geq R_\epsilon,$$

where $m(x,y) = \frac{y}{1+xy}$ and $s(.)$ is a decreasing positive function. For any chosen x, $m(x,y)$ is an increasing function of y. Hence, we can write

$$R_\epsilon \leq \int_{P_\epsilon^{s(.)}}^{\mathscr{P}} m(p, h_\epsilon)\mathrm{d}p = \log\left(\frac{1+h_\epsilon \mathscr{P}}{1+h_\epsilon P_\epsilon^{s(.)}}\right) = C(P_\epsilon^{s(.)}).$$

Since $C(p)$ is a decreasing function of p,

$$P_\epsilon^{s(.)} \leq C^{-1}(R_\epsilon). \tag{2.10}$$

Lemma 2 *Denoting the optimizer of the problem (2.9) as* $s^*(.)$, *we have* $P_\epsilon^{s^*(.)} \leq I_0(h_\epsilon)$.

Proof: Assume $P_\epsilon^{s^*(.)} > I_0(h_\epsilon)$. Define $s^{**}(p)$ as

$$s^{**}(p) = \begin{cases} I_0^{-1}(p) & p < I_0(h_\epsilon) \\ h_\epsilon & I_0(h_\epsilon) \leq p \leq P_\epsilon^{s^*(.)} \\ s^*(p) & p > P_\epsilon^{s^*(.)} \end{cases}.$$

We can write

$$\begin{aligned} R_\epsilon^{s^{**}(.)} &= \int_{I_0(h_\epsilon)}^{P_\epsilon^{s^*(.)}} m(p, s^{**}(p))\mathrm{d}p + \int_{P_\epsilon^{s^*(.)}}^{P} m(p, s^*(p))\mathrm{d}p \\ &\geq R_\epsilon^{s^*(.)}. \end{aligned}$$

Moreover, we have

$$\begin{aligned} R_{ave}^{s^*(.)} &= \int_0^{I_0(h_\epsilon)} g(p, s^*(p))\mathrm{d}p + \int_{I_0(h_\epsilon)}^{P_\epsilon^{s^*(.)}} g(p, s^*(p))\mathrm{d}p \\ &+ \int_{P_\epsilon^{s^*(.)}}^{\mathscr{P}} g(p, s^{**}(p))\mathrm{d}p \leq \int_0^{I_0(h_\epsilon)} g(p, I_0^{-1}(p))\mathrm{d}p \\ &+ \int_{I_0(h_\epsilon)}^{P_\epsilon^{s^*(.)}} g(p, h_\epsilon)\mathrm{d}p + \int_{P_\epsilon^{s^*(.)}}^{\mathscr{P}} g(p, s^{**}(p))\mathrm{d}p \\ &= R_{ave}^{s^{**}(.)}, \end{aligned}$$

where the inequality is concluded from (2.7), (2.8), and the fact that $s^*(p) > h_\epsilon$, for $p \leq P_\epsilon^{s^*(.)}$. Therefore, our assumption of $s^*(.)$ being optimal is not valid and the lemma is proved. ∎

The above lemma states the fact that, applying the multicast outage constraint, more power will be allocated to the channel gains lower than the outage threshold, compared to the unconstrained scenario [10], where $I_0(.)$ is the interference term which leads to the optimal expected rate.

Lemma 3 *Given $P_\epsilon^{s(.)} = \alpha$, the optimizer of (2.9) is given by*

$$s^*_\alpha(p) = \eta(\lambda, p) = \begin{cases} I_0^{-1}(p) & p < \alpha \\ h_\epsilon & \alpha \leq p \leq I_\lambda(h_\epsilon) \\ I_\lambda^{-1}(p) & p > I_\lambda(h_\epsilon) \end{cases}, \tag{2.11}$$

where $I_\lambda(h) = \frac{(\lambda+1-F(h))-hf(h)}{h^2 f(h)}$, and

$$\lambda = \begin{cases} 0, \int_\alpha^{\mathscr{P}} m(p, \eta(0,p))\mathrm{d}p > R_\epsilon \\ \arg\left(\int_\alpha^{\mathscr{P}} m(p, \eta(\lambda,p))\mathrm{d}p = R_\epsilon\right), otherwise \end{cases}.$$

Proof: It can be concluded directly from (2.8) that,

$$\int_0^\alpha g(p, s^*_\alpha(p))\mathrm{d}p \leq \int_0^\alpha g(p, I_0^{-1}(p))\mathrm{d}p. \tag{2.12}$$

Moreover, regarding the outage constraint of our problem,

$$\int_\alpha^{\mathscr{P}} g(p, s^*_\alpha(p))\mathrm{d}p = R_{max}(R_\epsilon, \alpha),$$

where $R_{max}(R_\epsilon, \alpha) = \max_{s(p)\leq h_\epsilon, p\geq\alpha} \int_\alpha^{\mathscr{P}} g(p, s(p))\mathrm{d}p$, subject to $\int_\alpha^{\mathscr{P}} m(p, s(p))\mathrm{d}p \geq R_\epsilon$. Writing K.K.T. condition, we have

$$R_{max}(R_\epsilon, \alpha) = \max_{s(p)\leq h_\epsilon, p\geq\alpha} \int_\alpha^{\mathscr{P}} T_\lambda(p, s(p))\mathrm{d}p, \tag{2.13}$$

where, $T_\lambda(x,y) = (g(x,y) + \lambda m(x,y))$. λ is 0, if the outage constraint is not limiting; otherwise, it could be obtained through the outage constraint $\int_\alpha^{\mathscr{P}} m(p,s(p))\mathrm{d}p = R_\epsilon$. Differentiating the function $T_\lambda(x,y)$ with respect to y, we get

$$\frac{\partial}{\partial y}T_\lambda(x,y) = \frac{\lambda + 1 - F(y) - yf(y)(1+xy)}{(1+xy)^2}. \tag{2.14}$$

Since $T_\lambda(x,y)$ is a concave function of y,

$$\arg\max(T_\lambda(x,y)|_{x=p}) = I_\lambda^{-1}(p). \tag{2.15}$$

Moreover, $T_\lambda(x,y)|_{x=p}$ is increasing for $y < I_\lambda^{-1}(p)$, and decreasing elsewhere. Hence, for any function $s(p)$ such that $s(p) < h_\epsilon$ for $P > \alpha$, we can write

$$\int_\alpha^{\mathscr{P}} T_\lambda(p,s(p))\mathrm{d}p \le \int_\alpha^{\mathscr{P}} T_\lambda(p,s(\lambda,p))\mathrm{d}p. \tag{2.16}$$

Therefore,

$$R_{ave}^{s_\alpha^*(.)} = \int_0^{\mathscr{P}} g(p,s_\alpha^*(p))\mathrm{d}p \le \int_0^{\mathscr{P}} g(p,s(\lambda,p))\mathrm{d}p, \tag{2.17}$$

and the proof of lemma is complete. ■

Theorem 2 *The solution to the optimization problem (2.9) can be written as*

$$\max_{s(.)} R_{ave}^{s(.)}|R_\epsilon = \max_{0\le\alpha\le\min(C^{-1}(R_\epsilon),I_0(h_\epsilon))} \int_0^{\mathscr{P}} g(p,s_\alpha^*(p))\mathrm{d}p.$$

Proof: The proof is directly concluded from Lemma 2, Lemma 3, and inequality (2.10). ■

Corollary 1 *The capacity region of a Rayleigh fading multicast network (R_ϵ, R_{ave}), is bounded by (C_ϵ, C_{ave}), such that*

$$C_\epsilon = \log\left(1 + \frac{h_\epsilon\beta\mathscr{P}}{1+h_\epsilon(1-\beta)\mathscr{P}}\right),$$

where β changes from 0 to 1 and

$$C_{ave} = 2(E_i(\theta(\beta)) - E_i(1)) - (e^{-\theta(\beta)} - e^{-1}) + e^{-h_\epsilon}C_\epsilon,$$

where $\theta(\beta) = \frac{2}{1+\sqrt{1+4(1-\beta)\mathscr{P}}}$, and $E_i(x) = \int_x^\infty \frac{e^{-t}}{t}\mathrm{d}t$, for any $\epsilon > 0$ such that $h_\epsilon \le I_0^{-1}(\mathscr{P})$.

Proof: Since $h_\epsilon \leq I_0^{-1}(\mathscr{P})$, $I_\lambda(h_\epsilon) > \mathscr{P}$, for any $\lambda \geq 0$. Therefore, (2.11) leads to the optimum power distribution function

$$\rho(h) = (\mathscr{P} - \alpha)\delta(h - h_\epsilon) + A(h),$$

where

$$A(h) = \begin{cases} \frac{2}{h^3} - \frac{1}{h^2} & I_0^{-1}(\alpha) < h < 1 \\ 0 & else \end{cases}.$$

This power distribution results in the proposed capacity region. ∎

An interesting conclusion of Corollary 1 is that, the expected typical rate is maximized when the multicast rate is provided in a single layer code. In the case we have no multicast constraint, it is shown in [10] that a multilevel coding with a small rate in each level is optimal in terms of maximizing the expected rate. However, when we are constrained to distribute a fraction of power to a set of *low* channel gains $[0, h_\epsilon]$ (coverage constraint), it is optimum to allocate all the power to the highest gain (h_ϵ).

Note that the assumption $h_\epsilon \leq I_0^{-1}(\mathscr{P})$ is not hard to satisfy, since the outage probability ϵ is usually small. Moreover, the value of h_ϵ decreases significantly with the number of users, such that it could be approximated by $\frac{\epsilon}{N}$. For example, for $N = 5$ and $\mathscr{P} = 100$, the outage probability ϵ could be as high as 0.38 in order to have $h_\epsilon \leq I_0^{-1}(\mathscr{P})$. In figure (2.1) we can see the capacity region of this network when $\epsilon = 0.01$. It is evident that due to hard coverage constraint for all the users, the achievable outage rates are very small in comparison with the expected rate values.

2.4 Soft Coverage Constraint

In the previous section, we observed that a strict coverage constraint for multicasting results in very small values of multicast rate. We can relax the coverage requirement by considering the average service received by *all* the users in *one* channel block. In fact, we can replace the *outage* requirement by the *expected* multicast rate. In this case, all the users should receive a minimum rate in average and given that, we want a typical user

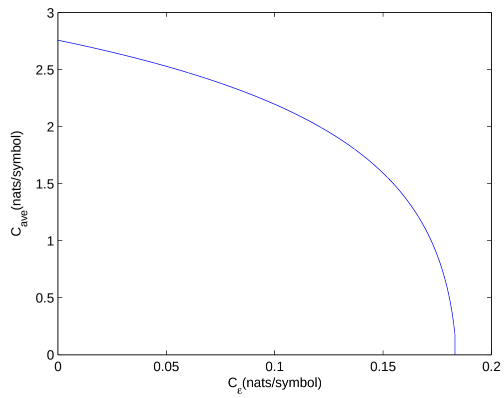

Figure 2.1: Multicast outage capacity vs. expected typical rate for $\mathscr{P} = 100$ and $N = 5$

to receive the highest expected rate. Therefore, the measures we are dealing with in this section are R_{mul} and R_{ave}.

According to (1) the optimality of superposition coding scheme is concluded for the performance vector $[R_{mul}, R_{ave}]$. In fact, we can state that the multicast constraint dose not affect the optimality of superposition coding to achieve the highest expected rate. As in [10], the transmitter can view an unknown channel as a continuum of receivers, experiencing different fading levels. However, in our scenario we have two of such channels. The objective is to design a continuum of code layers to provide the required expected rate in the multicast channel and maximize the expected rate in the typical channel.

Theorem 3 *The capacity region of a Rayleigh fading multicast network (R_{mul}, R_{ave}), is bounded by (C_{mul},C_{ave}), such that:*

$$C_{ave} = \int_0^\infty e^{-u} \frac{u\rho_\gamma(u)\mathrm{d}u}{1+uI_\gamma(u)}, \tag{2.18}$$

$$C_{mul} = \int_0^\infty e^{-Nu} \frac{u\rho_\gamma(u)\mathrm{d}u}{1+uI_\gamma(u)}, \tag{2.19}$$

where

$$I_\gamma(h) = \begin{cases} \mathscr{P} & \text{if } h < h_0 \\ \frac{e^{-h}(1-h)+\gamma e^{-Nh}(1-Nh)}{h^2(e^{-h}+\gamma N e^{-Nh})} & h_0 < h < h_1 \\ 0 & h > h_1 \end{cases},$$

$\rho_\gamma(h) = -\frac{\partial I_\gamma(h)}{\partial h}$, *and* h_0 *and* h_1 *are real numbers, such that*

$$\frac{e^{-h_0}(1-h_0)+\gamma e^{-Nh_0}(1-Nh_0)}{h_0^2(e^{-h_0}+\gamma N e^{-Nh_0})} = \mathscr{P},$$
$$\frac{e^{-h_1}(1-h_1)+\gamma e^{-Nh_1}(1-Nh_1)}{h_1^2(e^{-h_1}+\gamma N e^{-Nh_1})} = 0,$$

for different positive values of γ.

Proof: If we set $w_1(h) = 1 - F_{ave}(h)$ and $w_2 h = 1 - F_{mul}(h)$, Corollary 1 states that, in order to find the boundary set of our performance vector, we should search between different infinite layer superposition codes. Assuming $\rho(h)\mathrm{d}h$ as the power allocated to the layer associated to the channel gain h, the rate of that layer is

$$\mathrm{d}R_h = \log\left(1 + \frac{h\rho(h)\mathrm{d}h}{1+hI(h)}\right) = \frac{h\rho(h)\mathrm{d}h}{1+hI(h)}, \tag{2.20}$$

where

$$I(h) = \int_h^\infty \mathrm{d}u\rho(u),$$

and

$$I(0) = \mathscr{P}.$$

Using the above equation, the rate received at the receiver at the fading level h is

$$R(h) = \int_0^h \frac{u\rho(u)\mathrm{d}u}{1+uI(u)}.$$

Regarding to our definitions of multicast channel and average channel, the average rate in each of them can be written as follows:

$$R_{mul} = \int_0^\infty (1 - F_{mul}(u))\mathrm{d}R(u) = \int_0^\infty e^{-Nu}\frac{u\rho(u)\mathrm{d}u}{1+uI(u)}, \tag{2.21}$$
$$R_{ave} = \int_0^\infty (1 - F_{ave}(u))\mathrm{d}R(u) = \int_0^\infty e^{-u}\frac{u\rho(u)\mathrm{d}u}{1+uI(u)}. \tag{2.22}$$

Now, the problem is given $R_{mul} = r$, what is the maximum achievable R_{ave}. In other word,

$$R_{ave} = \max_{I(u)} \int_0^\infty e^{-u}\frac{u\rho(u)\mathrm{d}u}{1+uI(u)}, \tag{2.23}$$

subject to:

$$\int_0^\infty e^{-Nu} \frac{u\rho(u)\mathrm{d}u}{1+uI(u)} = r, \tag{2.24}$$
$$I(0) = \mathscr{P},$$

and

$$I(\infty) = 0.$$

In order to solve this optimization problem we define $S(x, I(x), I'(x), \gamma)$ as follows:

$$S(x, I(x), I'(x), \gamma) = e^{-x} \frac{x\rho(x)}{1+xI(x)} - \gamma e^{-Nx} \frac{x\rho(x)}{1+xI(x)}, \tag{2.25}$$

where

$$I'(x) = -\rho(x).$$

The necessary condition for $I(x)$ to maximize (2.23) with the constraint (2.24) is the zero functional variation [13] of $S(x, I(x), I'(x), \gamma)$,

$$\frac{\partial}{\partial I} S - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial}{\partial I'} S = 0, \tag{2.26}$$

where

$$\frac{\partial}{\partial I} S = (e^{-x} - \gamma e^{-Nx}) \frac{x^2 I'(x)}{(1+xI(x))^2},$$
$$\frac{\partial}{\partial I'} S = (e^{-x} - \gamma e^{-Nx}) \frac{-x}{1+xI(x)},$$
$$\frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial}{\partial I'} S = \frac{x(e^{-x} - \gamma N e^{-Nx})}{1+xI(x)} + (e^{-x} - \gamma e^{-Nx}) \frac{x^2 I'(x) - 1}{(1+xI(x))^2}.$$

Therefore, (2.26) simplifies to a linear equation which leads to the optimum interference function given in (2.20).

■

Figure (2.2) shows the achievable rate region for $N = 5$ and $\mathscr{P} = 100$. It can be observed that the maximum average rate is achieved for multicast requirement, $R_{mul} \leq 1.05$. It is shown in [12], that a good fraction of the highest expected rate with infinite

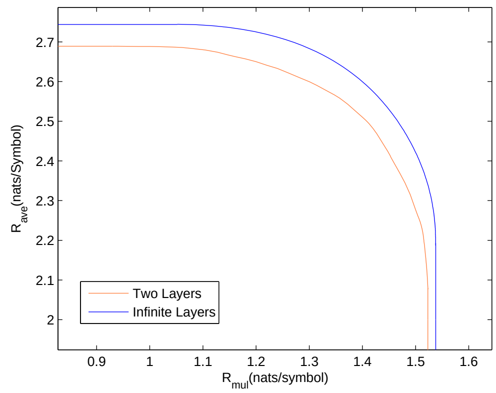

Figure 2.2: Expected multicast rate vs. expected typical rate for $\mathscr{P} = 100$ and $N = 5$

layers of code is achieved by two layers. Figure (2.2) shows that this is true for our multicast network as well. Furthermore, we can observe that the two-layer code region, gets closer to the capacity region at high multicast rate area. This can be justified by relative good performance of finite level codes for the channels with low variance power gain.

If we didn't have the muticast constraint and our objective was only to maximize the average rate, it is shown in [11] that the power distribution function would be,

$$\rho_{average}(h) = \begin{cases} \frac{2}{h^3} - \frac{1}{h^2} & s_0 < h < 1 \\ 0 & else \end{cases},$$

where

$$s_0 = \frac{2}{1 + \sqrt{1 + 4(P - P_1)}}.$$

This function is depicted in figure (2.3), and is compared with the case we have a multicast requirement $R_{mul} = 1.4$. As shown in the figure, the coverage requirement for all the users has shifted the power to lower channel gains, in order to provide service for the user with the worst channel quality in the network.

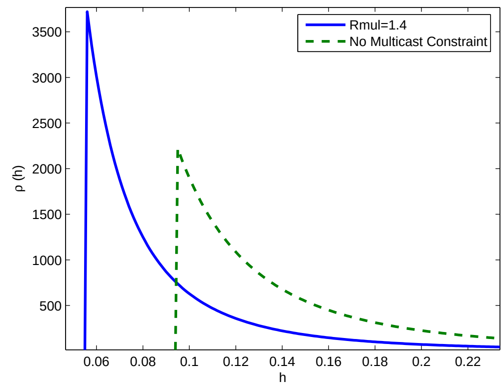

Figure 2.3: Power distribution function for no multicast requirement and for $R_{mul} = 1.4$

2.5 Extension to MISO

In the case we have multiple (M) antennas at the transmitter, we can adopt the broadcast approach proposed in [10]. In this approach, the receiver with unknown quasi-static fading MIMO channel is modeled as a continuum of receivers each associated with a channel realization. These receivers are ordered in a degraded fashion. However, since MIMO-BC is inherently non-degraded, this approach dose not necessarily lead to the optimum performance.

Assuming single antenna at each receiver side, the ordering of the modeled receivers in this approach is based on their normalized channel norm, $\frac{||HH^\dagger||}{M}$. Hence, the rate of the receiver is

$$R(\frac{||HH^\dagger||}{M}) = \log(1 + \frac{P_S \frac{||HH^\dagger||}{M}}{1+P_I \frac{||HH^\dagger||}{M}}) = C(\frac{||HH^\dagger||}{M}, P_I, P_S),$$

where P_S and P_I are the decodable and undecodable signal power levels, respectively.

Now, assume N users in this model, all receiving a common source through an infinite-layer code. We want to design this code to maximize the average rate received by a typical user, while providing a given rate for all the users. For this purpose, we should provide this rate for the worst user in our degraded broadcast model. This user has the lowest channel vector norm. The normalized channel norm of user i, denoted by $\frac{1}{M}||H_i H_i^\dagger||$, is a scaled χ_2 random variable with $2M$ degrees of freedom, whose CDF can be obtained as

$$F_{ave}(h) = F_{\frac{1}{M}||H_iH_i^\dagger||}(h) = 1 - \frac{\Gamma(M, Mh)}{\Gamma(M)}, \tag{2.27}$$

where $\Gamma(\alpha)$ is a gamma function, and $\Gamma(\alpha, \beta)$ is an upper incomplete gamma function. Since, the users' channels are statistically independent, the distribution of the norm of the weakest channel can be obtained as following:

$$\begin{aligned} \Pr\left\{\min_i \frac{||H_iH_i^\dagger||}{M} > h\right\} &= \left(\Pr\{\frac{||H_iH_i^\dagger||}{M} > h\}\right)^N \\ &= \left(\frac{\Gamma(M, Mh)}{\Gamma(M)}\right)^N. \end{aligned}$$

Hence, the cumulative distribution function for the weakest user's channel norm is

$$F_{mul}(h) = 1 - \left(\frac{\Gamma(M, Mh)}{\Gamma(M)}\right)^N. \tag{2.28}$$

Following the same approach as in section IV, the average rate and multicast rate could be written as

$$\begin{aligned} R_{mul} = \int_0^\infty (1 - F_{mul}(u))\mathrm{d}R(u) = \int_0^\infty (\tfrac{\Gamma(M,Mu)}{\Gamma(M)})^N \tfrac{u\rho(u)\mathrm{d}u}{1+uI(u)}, \\ R_{ave} = \int_0^\infty (1 - F_{ave}(u))\mathrm{d}R(u) = \int_0^\infty \tfrac{\Gamma(M,Mu)}{\Gamma(M)} \tfrac{u\rho(u)\mathrm{d}u}{1+uI(u)}, \end{aligned}$$

where $\rho(u)$ and $I(u)$ are the corresponding power allocation and interference power functions. Defining $S(x, I(x), I'(x), \lambda)$ as

$$\begin{aligned} S(x, I(x), I'(x), \lambda) &= \frac{\Gamma(M, Mx)}{\Gamma(M)} \frac{xI'(x)}{1 + xI(x)} \\ &- \lambda \left(\frac{\Gamma(M, Mx)}{\Gamma(M)}\right)^N \frac{xI'(x)}{1 + xI(x)}, \end{aligned}$$

and setting its functional variation equal to zero to maximize the average rate, similar to (2.26), we obtain the optimizer $I(x)$ as

$$I(h) = \begin{cases} \mathscr{P} & \text{if } h < h_0 \\ \frac{\Gamma(M,Mh) - \lambda \frac{\Gamma(M,Mh)^N}{\Gamma(M)^{N-1}}}{Mh^{M+1}e^{-Mh}(1 - \lambda N \frac{\Gamma(M,Mh)}{\Gamma(M)}^{N-1})} - \frac{1}{h} & h_0 < h < h_1 \\ 0 & h > h_1 \end{cases},$$

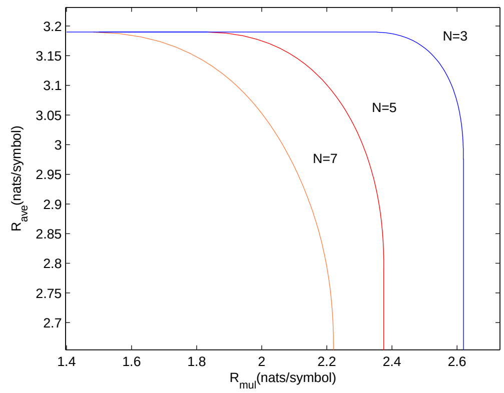

Figure 2.4: MISO expected multicast vs. typical rate for $M = 2$ and $\mathscr{P} = 100$

where h_0, h_1 and λ are obtained through the following equations, respectively:

$$I(h_0) = \frac{\Gamma(M,Mh_0)-\lambda\frac{\Gamma(M,Mh_0)^N}{\Gamma(M)^{N-1}}}{Mh_0^{M+1}e^{-Mh_0}\left(1-\lambda N(\frac{\Gamma(M,Mh_0)}{\Gamma(M)})^{N-1}\right)} - \frac{1}{h_0} = \mathscr{P},$$

$$I(h_1) = \frac{\Gamma(M,Mh_1)-\lambda\frac{\Gamma(M,Mh_1)^N}{\Gamma(M)^{N-1}}}{Mh_1^{M+1}e^{-Mh_1}\left(1-\lambda N(\frac{\Gamma(M,Mh_1)}{\Gamma(M)})^{N-1}\right)} - \frac{1}{h_1} = 0,$$

$$R_{multicast} = \int_0^\infty \left(\frac{\Gamma(M,Mu)}{\Gamma(M)}\right)^N \frac{u\rho(\lambda,u)\mathrm{d}u}{1+uI(\lambda,u)} = r.$$

The achievable rate region is shown in figure (2.4) for different values of N, when $M = 2$. As mentioned in [10], the idea of modeling the unknown fading channel by a degraded broadcast channel with infinite number of receivers is not optimal when we have multiple antennas. This is mainly because a MIMO broadcast channel is not degraded. One may claim that the same model with a general MIMO broadcast channel and the capacity region proposed in [17] might outperform our model. However, since we have a common message broadcasting, all the data decoded at a transmitter is important for us, even the part treated as the interference in the Broadcast Channel. In other words, we are utilizing the degraded characteristic of the channel as we are assuming it receives whatever a weaker receiver decodes, plus its corresponding data. As a result, there would be some limitation for applying a general MIMO Broadcast model.

In figure (2.4), we can see that as the number of users decreases, the proposed achievable rate region expands more. It is also evident by comparing the region of MISO and SISO (figure(2.2)) channels with $N = 5$ users, that using multiple antennas improves the achievable rates. However, its effect on the achievable rates for the multicast channel is more considerable than for the average channel. This prominent gain for multicast channel is sensible, since we are using multiple independent paths to convey the data, so the probability of having very low channel gains for all paths (which mainly corresponds to multicast channel) significantly decreases. In fact, we will show that we can compensate the adverse effect of number of users by increasing the number of transmit antennas. More specifically, if both N and M tend to infinity and M grows highly enough with respect to N, we will show that the multicast rate could reach the average rate and our scheme gives the optimal solution, although it is not for small number of transmit antennas. The following theorem states this fact.

Theorem 4 *For large values of M and N, the proposed infinite layer superposition coding will provide R_{mul}, such that*

$$R_{mul} \geq R_{opt} - \sigma, \tag{2.29}$$

if

$$M > \frac{\mathscr{P}^2 \log(N) + \omega(1)}{(1+\mathscr{P})^2 \sigma^2}, \tag{2.30}$$

where R_{opt} is the highest achievable average rate for a randomly selected user in the network and σ is an arbitrarily small positive number.

Proof: First of all, we propose an upper bound for the achievable average rate for a randomly selected user, by assuming no stringent delay constraint, meaning that the transmission block can be chosen as long as the fading block. In this case, the channel has an ergodic behavior, so that ergodic capacity is defined and is shown to be:

$$C_{erg} = E\left[\log\left(1 + \frac{||HH^{\dagger}||}{M}\mathscr{P}\right)\right]. \tag{2.31}$$

As a result,

$$R_{opt} \leq C_{erg}. \tag{2.32}$$

Regarding the central limit theorem [19], the distribution of $\frac{||HH^{\dagger}||}{M}$, where

$$\frac{1}{M}||HH^{\dagger}|| = \frac{{h_1}^2 + {h_2}^2 + ... + {h_M}^2}{M} \tag{2.33}$$

and h_i's are independent rayleigh distributions with unit variance and unit mean, approaches to a Gaussian distribution with the CDF:

$$F_{\frac{||HH^{\dagger}||}{M}}(h) = Q\left(\frac{h-1}{\frac{1}{\sqrt{M}}}\right), \tag{2.34}$$

and consequently the CDF of multicast channel will be

$$F_{mul}(h) = 1 - Q\left(\frac{h-1}{\frac{1}{\sqrt{M}}}\right)^N. \tag{2.35}$$

Using the concavity of log function, and having the fact that $E\left[\frac{||HH^{\dagger}||}{M}\right] = 1$, we have

$$C_{erg} \leq \log(1 + \mathscr{P}). \tag{2.36}$$

We will show that our scheme provides a multicast rate arbitrarily close to this upper bound, if we use enough number of transmit antennas. Since this upper bound is larger than the average rate the theorem will be proved. For this purpose, we use a single-layer coding. We know that our scheme outperforms this scheme, as the single-layer coding is a special case of superposition coding. Using a single-layer code with power $\mathscr{P}$ and rate R_σ, where

$$R_\sigma = \log(1 + \mathscr{P}(1 - \sigma')), \tag{2.37}$$

and

$$\sigma' = \frac{(1 + \mathscr{P})\sigma}{\mathscr{P}},$$

the average multicast rate in our network will be

$$R_{mul} = Pr\left\{\frac{||HH^{\dagger}||_{mul}}{M} > 1 - \epsilon'\right\} R_{\sigma}, \tag{2.38}$$

where $||HH^{\dagger}||_{mul} = \min_i ||H_i H_i^{\dagger}||$. Regarding (2.35), the above equation can be written as

$$R_{mul} = Q(-\sqrt{M}\epsilon')^N R_{\sigma} = \left[1 - Q(\sqrt{M}\sigma')\right]^N R_{\sigma}. \tag{2.39}$$

Assuming M large enough to have $\sqrt{M}\sigma' >> 1$, and consequently $Q(\sqrt{N}\sigma') << 1$, we can rewrite the above equation as

$$R_{mul} = e^{-NQ(\sqrt{M}\sigma')} R_{\sigma}. \tag{2.40}$$

Now, using the approximation

$$Q(x) \approx \frac{1}{\sqrt{2\pi x}} e^{-\frac{x^2}{2}} \tag{2.41}$$

for large values of x, we can write

$$Q(\sqrt{M}\sigma') \leq e^{-M\sigma'^2}. \tag{2.42}$$

Therefore, having

$$M \sim \frac{\log(N) + \omega(1)}{\sigma'^2}, \tag{2.43}$$

incurs

$$NQ(\sqrt{M}\sigma') \sim o(1), \tag{2.44}$$

and as a result,

$$\lim_{N\to\infty} R_{mul} - R_{\sigma} = 0. \tag{2.45}$$

Moreover, assuming $\sigma \ll 1$, (2.37) can be written as,

$$\begin{aligned} R_{\sigma} &\backsimeq \log(1 + \mathscr{P}) - \frac{\mathscr{P}\sigma'}{1 + \mathscr{P}} \\ &\geq C_{erg} - \sigma, \end{aligned} \tag{2.46}$$

where the second line results from (2.36). Combining (2.32), (2.45), and (2.46), the result of Theorem 6 easily follows. ∎

2.6 Multiple Sources

So far we have assumed a common source and multiple receivers. However, in many applications of common message broadcasting, we have more than one sources each having a group of interested users. For example in TV broadcasting we have several TV channels and each channel has its own group of viewers. In this section we extend our results to such cases. However, regarding the nature of these networks which are usually large, we have to make a slight change in our performance measures.

For a network with large number of users (e.g. TV broadcasting), the data rate delivered to all the users is very small. Therefore, we should be less ambitious about the multicasting. Hence, the coverage constraint is usually expressed in terms of the data rate decoded for a high percentage of the users. In this case, the definition of outage rate (R_ϵ) changes to the rate decoded by $100(1-\epsilon)$ percent of the total users. Since the users fading levels are i.i.d., this rate is equal to the outage rate of the typical user in the network. Hence the new outage rate is the same as what is defined in (2.3), except that

$$h_\epsilon = F_{typ}^{-1}(\epsilon).$$

The objective is to maximize the typical expected rate while providing a basic outage rate.

In the case of multiple sources, the latter interpretation of coverage constraint is more meaningful. Assume k sources and N users in the network. In the general case user i selects the j^{th} channel with probability $p_{i,j}$. For the sake of simplicity, we assume all the probabilities equal: $p_{i,j} = \frac{1}{k}$. Data from source i must be delivered to most of its users with rate R_ϵ^i. The expected rate decoded by a typical user of source i is denoted by R_{typ}^i. In the following we will consider the case with two sources and state the achievable rate couple (R_{typ}^1, R_{typ}^2), given the above coverage constraint.

Theorem 5 *Denote the optimizer of problem (2.9) for* $R_\epsilon = R_\epsilon^1 + R_\epsilon^2$ *by* $s^*(.)$. *Define* P_i, *for* $i = 1, 2$, *such that* $\int_{P_i}^{\mathscr{P}} m(p, s^*(p))\mathrm{d}p = R_\epsilon^i$. *Also, define* $R_i^{min} = \int_{P_i}^{P} g(p, s^*(p))\mathrm{d}p$ *and* $R_i^{max} = \int_0^{P_{j,j\neq i}} g(p, s^*(p))\mathrm{d}p$. *The boundary set of the rate couple* $(R_{typ}^1, R_{typ}^2) \in {R^+}^2$, *subject to the coverage constraint couple* $(R_\epsilon^1, R_\epsilon^2)$, *is given by,*

$$C_{1,2} = \{(R_1, R_2) | R_i = \alpha R_i^{max} + (1-\alpha) R_i^{min}, 0 \le \alpha \le 1\}$$

Proof: First, we show that this set of rate couples are achievable when the multicast coverage constraint couple is $(R_\epsilon^1, R_\epsilon^2)$. Let us define $\mathscr{C}_i$, for $i = 1, 2$, as an infinite layer superposition code with the total power $\mathscr{P}$ and the power distribution $s^*(.)$, where the code layers associated to the interference levels higher than P_i. In fact $\mathscr{C}_i$ is the code which maximizes the sum expected rate of received from both sources while it holds the multicast data rate constraint and among all such codes it leads to the minimum expected rate received from the source i. It is evident from the definitions that the rate couples produced by the codes $\mathscr{C}_1$ and $\mathscr{C}_2$ are (R_1^{min}, R_2^{max}) and (R_1^{max}, R_2^{min}), respectively. Since both codes achieve provide the multicast minimum requirement, a time sharing between them also satisfies the coverage constraint. The different time sharing factors $0 \leq \alpha \leq 1$, will give rise to the rate couple set $C_{1,2}$.

In order to prove the optimality of $C_{1,2}$, we assume that this set is not the boundary set of all achievable rate couples. According to this assumption, there should exist a couple rate $(C_1, C_2) \in C_{1,2}$ and a coding scheme $\mathscr{C}'$ which satisfies the multicast constraint and leads to the expected rate couple (C_1', C_2') such that $C_1' C_1$ and $C_2' > C_2$. Therefore we have

$$C_1' + C_2' > C_1 + C_2 = R_1^{min}, R_2^{max}.$$

However this contradicts the optimality of the codes $\mathscr{C}_1$ and $\mathscr{C}_2$ in terms of producing the maximum sum rate of both sources while satisfying the multicast constraint. Hence, there is no such coding scheme and the proof is complete. ∎

In fact, the above theorem states that the capacity region for multiple sources could be achieved by time sharing between the schemes prioritizing one source over another one while satisfying the minimum multicast rate for both of them.

Chapter 3

Layered Joint Source-Channel Code for a Multicast Network

3.1 System Model

In this chapter, we consider a common source broadcasting network, where a single-antenna transmitter wishes to send a Gaussian source over a wireless channel to N single-antenna receivers. Let the source be denoted by s, which is a sequence of independent identically distributed (iid) zero-mean circularly symmetric complex Gaussian (ZMCSCG) random variables with unit variance: $s \in \mathbb{C} \sim \mathcal{CN}(0,1)$. The received signal at the jth receiver, denoted by y_j can be written as

$$y_j = t_j x + n_j, \tag{3.1}$$

where $x \in \mathbb{C}$ is the transmitted signal, $\{n_j\} \in \mathbb{C} \sim \mathcal{CN}(0,1)$ is the Additive White Gaussian Noise (AWGN) at this receiver, and $t_j \in \mathbb{C} \sim \mathcal{CN}(0,1)$ is the channel coefficient from the transmitter to the jth receiver.

Suppose the distribution of the channel power gain for user j is described by the probability density function (pdf) $f(h_j)$, where $h_j \triangleq |t_j|^2$. We first consider fading distributions with a finite number of discrete fading states; subsequently we generalize to continuous fading distributions. The receiver has perfect CSI but the transmitter has only channel distribution information (CDI), i.e., the transmitter knows the pdf $f(h_j)$, $j = 1, \cdots, N$, but not their instantaneous realizations. The channels are modeled by a quasi-static block fading process: t_j is realized iid at the onset of each fading block and remains unchanged over the block duration. We assume decoding at the receiver is *delay-limited*; namely, delay constraints preclude coding across fading blocks but dictate that the receiver decodes at the end of each block. Hence, the channel is non-ergodic. Suppose each fading block spans n channel uses, over which the transmitter describes k of the source symbols. We define the bandwidth ratio as $b = n/k$, which relates the number of channel uses per source symbol. At the transmitter, there is a power constraint on the transmit signal $E[|x|^2] \leq \mathscr{P}$, where the expectation is taken over each fading block. We assume a short-term power constraint and do not consider power allocation across fading blocks. We assume k is large enough to consider the source as ergodic, and n is large enough to design codes that achieve the instantaneous channel capacity of a given fading state with negligible probability of error. At the receivers, the channel output y_j is used to reconstruct an estimate of the source denoted by $\hat{s}_j$. The distortion D_j is measured by the mean squared error $E[|s - \hat{s}_j|^2]$. The instantaneous distortion of the reconstruction depends on the fading realization of the channel.

The typical (average) channel of the multicast network is defined as the channel of a randomly selected user. Since all the channels are i.i.d., we have

$$F_{typ}(h) = F_j(h) = F(h). \tag{3.2}$$

Since all the N channels are Gaussian and they receive a common signal, the multicast channel is equivalent to the worst channel in the network. Due to statistical independence of the channels, the gain of that user has the following distribution:

$$\Pr\left\{\min_j(h_j) > h\right\} = (\Pr\{h_j > h\})^N.$$

As a result, we have

$$F_{mul}(h) = F_{\min_j(h_j)}(h)$$

In this chapter, we consider three measures of performance for our network, as follows:

- the multicast outage distortion, D_ϵ, the minimum distortion of the multicast channel with probability $(1 - \epsilon)$,

- the expected multicast distortion, $D_{mul} = E_{h_{mul}}[D(h)]$, where $h_{mul} = \min_j(h_j)$, and $D(h)$ is the distortion for the channel state h,

- the expected typical distortion, $D_{ave} = E_{h_{typ}}[D(h)]$.

In this chapter, we consider two cases associated with two different performance vectors for the network: i) $[D_\epsilon, D_{ave}]$, ii) $[D_{mul}, D_{ave}]$. In the first case, we want to jointly optimize expected distortion for a typical user in the network, while providing a distortion less than a given threshold for all the users with a high probability $(1-\epsilon)$. In the second case, we relax the coverage constraint to a given expected distortion for all the users. For this purpose, we first review the unconstrained problem studied in [26], in which the optimal expected distortion (D_{ave}) is derived.

3.1.1 Virtual Broadcast Approach with Successive Refinement

Virtual broadcast approach was first proposed in [10] to model a block slowly fading channel with stringent decoding delay constraint with no CSIT. In this approach, any channel state of the unknown fading channel is associated with a virtual receiver. The original receiver's rate at a realization is equal to the rate received by the virtual receiver associated with that realization. The average rate when the averaging time tends to infinity will become the expectation of the virtual receivers' rate according to the channel probability distribution. Regarding the degrading nature, the virtual receivers could be ordered based on their decodable rates. This will introduce a degraded broadcast (BC) network. The expected rate of the original receiver is the weighted sum rate of this BC network. It is shown in [21] that any positive weighted sum of such a virtual broadcast channel will be maximized, using superposition coding and successive decoding. For any virtual receiver the transmitter designates a layer and superimposes this layer above the layers associated with lower virtual receivers. Any virtual receiver decodes its assigned code layer in addition to all the layers below it. It was shown in [21] that this broadcast approach leads to the optimal performance for a multicast network in terms of achievable expected rate for a typical user of the network given the coverage constraint.

According to [9] and [27], a Gaussian source is successively refinable. Successive refinability implies that the distortion incurred using a description of a source at rate R_1

first, and subsequently refining it at rate R_2 is equal to the distortion when the source is described at rate $R_1 + R_2$ in the first place. We can utilize the successive refinability of a Gaussian source to transmit it using a broadcast approach. In this method, the source is described in multiple layers where each layer of refinement is carried by a level of channel code in the broadcast approach. Hence, the layer associated with a virtual receiver refines the information received by the receivers with lower channel gains.

Cascading a multi-resolution source code to a superposition channel code is proposed in [14] and [28] and shown to be optimal in terms of distortion exponent in high SNRs. We will apply the virtual broadcast approach with successive refinement as a joint source-channel coding scheme and optimize its performance according to the source and channel characteristics.

3.2 Discrete States

3.2.1 Preview

In this scenario, the fading realization has M states denoted by $\{h_i\}_{i=1}^{M}$ with probabilities $\{\eta_i\}_{i=1}^{M}$. In other words, the channel power gain realization is h_i with probability η_i, for $i = 1, \cdots, M$. According to the virtual broadcast model in [3] and [4], there are M virtual receivers and the transmitter sends the sum of M layers of codewords. Let layer i denote the layer of the codeword intended for virtual receiver i, and we order the layers as $h_M > h_{M-1} > \cdots > h_1$. We refer to layer M as the highest layer and layer 1 as the lowest layer. Each layer successively refines the description of the source s from the layer below it, and the codewords in different layers are independent. Let $\mathscr{P}_i$ be the transmit power allocated to layer i, then the transmit symbol x can be written as

$$x = \sqrt{\mathscr{P}_1}x_1 + \sqrt{\mathscr{P}_2}x_2 + \cdots + \sqrt{\mathscr{P}_M}x_M, \tag{3.3}$$

where $x_1, \cdots, x_M$ are iid ZMCSCG random variables with unit variance.

With successive decoding, each virtual receiver first decodes and cancels the lower layers before decoding its own layer; the undecodable higher layers are treated as noise. Thus, the rate R_i intended for virtual receiver i is

$$R_i = \log\left(1 + \frac{\mathscr{P}_i h_i}{1 + h_i \sum_{j=i+1}^{M} \mathscr{P}_j}\right), \tag{3.4}$$

where the term $h_i \sum_{j=i+1}^{M} \mathscr{P}_j$ represents the interference power from the higher layers. Suppose h_k is the realized channel power gain, then the original receiver can decode layer k and all the layers below it. Hence the realized rate $R_{rlz}(k)$ at the original receiver is $R_1 + \cdots + R_k$. Using the rate distortion function of a complex Gaussian source [16], the mean squared distortion is 2^{-bR} when the source is described at a rate of bR per symbol. Thus, the realized distortion $D_{rlz}(k)$ of the reconstructed source $\hat{s}$ is

$$D_{rlz}(k) = 2^{-bR_{rlz}(k)} = 2^{-b(R_1+\cdots+R_k)}, \tag{3.5}$$

where the last equality follows from successive refinability. The expected distortion $E_h[D]$ is obtained by averaging over the fading distribution pmf:

$$E_h[D] = \sum_{i=1}^{M} \eta_i D_{rlz}(i) = \sum_{i=1}^{M} \eta_i 2^{-b\sum_{j=1}^{i} R_j}. \tag{3.6}$$

In [26], the optimal power allocation $\mathscr{P}_1^*, \cdots, \mathscr{P}_M^*$ among the layers is derived to find the minimum expected distortion $E_h[D]^*$. To this end, they first considered a two-state channel and showed that under optimal power allocation, with respect to the minimum expected distortion, the two layers can be represented by a single *aggregate* layer. They used this idea to solve the optimization problem in the general M-state case in a *reccursive* fashion. For this purpose, the expected distortion is written as

$$E_h[D] = \sum_{i=1}^{M} \eta_i \left(\prod_{j=1}^{i} \frac{1 + h_j T_j}{1 + h_j T_{j+1}}\right)^{-b}, \tag{3.7}$$

where $T_j \triangleq \sum_{l=j}^{M} \mathscr{P}_l$. Hence, the expected distrotion can be written as a set of recurrence relations:

$$D_M = (1 + h_M \mathscr{P}_M)^{-b} \eta_M \tag{3.8}$$

$$D_i = \left(\frac{1 + h_i T_i}{1 + h_i T_{i+1}}\right)^{-b} (\eta_i + D_{i+1}), \tag{3.9}$$

with $D_1 = E_h[D]$. Note that D_i depends on only two adjacent power allocation variables T_i and T_{i+1}; therefore, in each recurrence step i, we solve for the optimal T^*_{i+1} in terms of T_i:

$$D^*_M \triangleq D_M, \tag{3.10}$$

$$D^*_i = \left(\frac{1+h_iT_i}{1+h_iT_{i+1}}\right)^{-b}\left(\eta_i + D^*_{i+1}\right). \tag{3.11}$$

The minimum expected distortion is given by $E_h[D]^* = D^*_1$, which is computed at the last step of recurrence equation ($i = 1$). In the first recurrence step, ($i = M-1$), the power allocation between the topmost two layers is considered. The minimal distortion D^*_M is found by obtaining the optimum value of T^*_M in terms of T_{M-1} in the following optimization problem:

$$D^*_{M-1} = \min_{0\le T_M \le T_{M-1}} \left(\frac{1+\gamma_{M-1}T_{M-1}}{1+\gamma_{M-1}T_M}\right)^{-b}\left(u_{M-1} + \left(1+\beta_{M-1}T_M\right)^{-b} w_{M-1}\right), \tag{3.12}$$

with the parameters:

$$\begin{array}{ll} w_{M-1} = \eta_M, & \beta_{M-1} = h_M, \\ u_{M-1} = \eta_{M-1}, & \gamma_{M-1} = h_{M-1}, \end{array} \tag{3.13}$$

where the subscripts on the layer parameters w, u, β, γ designate the recurrence step. In general, in recurrence step i, the power allocation between layer i and layer $i+1$ can be found by the optimization:

$$D^*_i = \min_{0\le T_{i+1} \le T_i} \left(\frac{1+\gamma_iT_i}{1+\gamma_iT_{i+1}}\right)^{-b}\left(u_i + \left(1+\beta_iT_{i+1}\right)^{-b} w_i\right), \tag{3.14}$$

the solution to which is given by

$$D^*_i = \begin{cases} (1+\gamma_iT_i)^{-b}W_i & U_{i+1} \le T_i \\ u_i + (1+\beta_iT_i)^{-b}w_i & else \end{cases}, \tag{3.15}$$

where $W_i \triangleq (1+\gamma_iU_{i+1})^b\left[u_i + (1+\beta_iU_{i+1})^{-b}w_i\right]$ and

$$U_{i+1} \triangleq \begin{cases} 0 & \beta_i/\gamma_i \le 1 + u_i/w_i \\ \frac{1}{\beta_i}\left(\left[\frac{w_i}{u_i}\left(\frac{\beta_i}{\gamma_i} - 1\right)\right]^{\frac{1}{1+b}} - 1\right) & else \end{cases}. \tag{3.16}$$

There are two cases to the solution of D_i^* . In the first case, the power allocation is not constrained by the available power T_i, and the minimum distortion in the recurrence step $i-1$ is obtained as:

$$D_{i-1}^* = \min_{0 \le T_i \le T_{i-1}} \left(\frac{1 + h_{i-1}T_{i-1}}{1 + h_{i-1}T_i} \right)^{-b} \left(\eta_{i-1} + (1 + \gamma_i T_i)^{-b} W_i \right). \tag{3.17}$$

Hence, the minimization in (3.17) has the same form as the one in (3.14), but with the following parameters:

$$\begin{aligned} w_{i-1} &= W_i, & \beta_{i-1} &= \gamma_i, \\ u_{i-1} &= \eta_{i-1}, & \gamma_{i-1} &= h_{i-1}. \end{aligned} \tag{3.18}$$

Hence the minimization can be solved the same way as in the last recurrence step. In the second case, the power allocation is constrained by the available power T_i, and we have

$$D_{i-1}^* = \min_{0 \le T_i \le T_{i-1}} \left(\frac{1 + h_{i-1}T_{i-1}}{1 + h_{i-1}T_i} \right)^{-b} \left(\eta_{i-1} + u_i + (1 + \beta_i T_i)^{-b} w_i \right), \tag{3.19}$$

which again has the same form as in (3.14), with the following parameters:

$$\begin{aligned} w_{i-1} &= W_i, & \beta_{i-1} &= \beta_i, \\ u_{i-1} &= \eta_{i-1} + u_i, & \gamma_{i-1} &= h_{i-1}. \end{aligned} \tag{3.20}$$

Therefore, in each recurrence step, the two-layer optimization procedure can be used to find the minimum distortion and the optimal power allocation between the current layer and the aggregate higher layer.

3.2.2 Hard Coverage Constraint

In this section, we consider the following optimization problem:

$$\begin{aligned} &\min_{\substack{\mathbf{P} \\ 0 \le \mathscr{P}_i \le \mathscr{P}}} && E_{h_{typ}}[D] \\ &\textit{subject to} && D_{\max} \le D_\epsilon \quad \textit{with probability}\, (1-\epsilon), \end{aligned} \tag{3.21}$$

where $\mathbf{P} \triangleq [\mathscr{P}_1, \cdots, \mathscr{P}_M]$, and $D_{\max} \triangleq \max_{1 \le n \le N} D_n$. $D_{\max}$ is a random variable which takes M values $\{2^{-b\sum_{j=1}^{i} R_i}\}_{i=1}^{M}$ with the corresponding probabilities $\{\phi_i\}_{i=1}^{M}$, where

$$R_i = \log \left(1 + \frac{h_i \mathscr{P}_i}{1 + h_i \sum_{j=i+1}^{M} \mathscr{P}_j} \right), \tag{3.22}$$

and ϕ_i's are obtained as follows:

$$\begin{aligned}\phi_i &= \Pr\{h_{\min} = h_i\} \\ &= \Pr\{h_{\min} > h_{i-1}\} - \Pr\{h_{\min} > h_i\} \\ &= \left(\Pr\{h_{\min} > h_{i-1}\}\right)^N - \left(\Pr\{h_{\min} > h_i\}\right)^N \\ &= \left(\sum_{j=i}^{M} \eta_j\right)^N - \left(\sum_{j=i+1}^{M} \eta_j\right)^N . \end{aligned} \tag{3.23}$$

In order to simplify the derivations, we consider two cases based on the value of ϵ, as follows:

1. $\epsilon < \min_i \phi_i$
2. $\exists k, \sum_{i=1}^{k} \phi_i \leq \epsilon < \sum_{i=1}^{k+1} \phi_i$ [1]

For each of the above cases, we find the solution of (3.21).

Case 1

In this case, in order to satisfy the condition $D_{\max} \leq D_\epsilon$ with probability $1-\epsilon$, we must have

$$2^{-bR_1} \leq D_\epsilon \Longrightarrow R_1 \geq R_\epsilon \triangleq \frac{\log D_\epsilon}{b}. \tag{3.24}$$

Therefore, (3.21) can be written as

$$\begin{aligned}\min_{\substack{\mathbf{P} \\ 0 \leq \mathscr{P}_i \leq \mathscr{P}}} \quad & E_{h_{typ}}[D], \\ \textit{subject to} \quad & R_1 \geq R_\epsilon. \end{aligned} \tag{3.25}$$

Let us define $R_1^{unconst.}$ as the value of R_1 in the solution of the unconstrainted problem. Two situations may occur here:

- $R_1^{unconst.} \geq R_\epsilon$; in this case, the solution of the unconstrainted problem already satisfies the constraint $R_1 \geq R_\epsilon$. In this situation, the solution of (3.25) is exactly the solution of the unconstrained problem obtained in the previous section.

[1] Note that $k \leq M-1$.

- $R_1^{unconst.} < R_\epsilon$; in this situation, we have

$$\begin{aligned} R_1 = \log\left(1 + \frac{\mathscr{P}_1 h_1}{1 + h_1(\mathscr{P} - \mathscr{P}_1)}\right) &\geq R_\epsilon \\ \Longrightarrow \mathscr{P}_1 &\geq \frac{(1 + h_1\mathscr{P})(1 - 2^{-R_\epsilon})}{h_1}. \end{aligned} \tag{3.26}$$

In the case that $\frac{(1+h_1\mathscr{P})(1-2^{-R_\epsilon})}{h_1} > \mathscr{P}$, or equivalently, $R_\epsilon > \log(1+\mathscr{P}h_1)$, (3.25) does not have any solution. In fact, it is not possible to satsify the coverage constraint even if all the available power is allocated to the lowermost layer. In the case that $R_\epsilon \leq \log(1 + \mathscr{P}h_1)$, it is easy to show that it is optimal to have $\mathscr{P}_1 = \mathscr{P}_1^* \triangleq \frac{(1+h_1\mathscr{P})(1-2^{-R_\epsilon})}{h_1}$. Having this, the expected typical distortion can be written as

$$\begin{aligned} E_{h_{typ}}[D] &= 2^{-bR_\epsilon}\eta_1 + 2^{-bR_\epsilon}\sum_{i=2}^{M}\eta_i 2^{-b\sum_{j=2}^{i} R_j} \\ &= 2^{-bR_\epsilon}(\eta_1 + \sum_{i=2}^{M}\eta_i 2^{-b\sum_{j=2}^{i} R_j}) \\ &= D_\epsilon \times E_{\widehat{h}_{typ}}[D], \end{aligned} \tag{3.27}$$

where $\widehat{h}_{typ}$ is the modified fading process which takes the values $\{0, h_2, \cdots, h_M\}$ with the probabilities $\{\eta_1, \eta_2, \cdots, \eta_M\}$, correspondingly. The above equation implies that by defining the modified fading process $\widehat{h}_{typ}$, the original constrainted problem is converted to solving the *unconstrainted* problem for $\widehat{h}_{typ}$, when the total power is constrainted to $\widehat{\mathscr{P}} = \mathscr{P} - \mathscr{P}_1^*$. Denoting the minimum distortion of the unconstrainted problem with the total power $\mathscr{P}$ and the fading process characterized by $(\mathbf{h}, \boldsymbol{\eta})$, where $\boldsymbol{\eta} \triangleq (\eta_1, \cdots, \eta_M)$, and $\mathbf{h} \triangleq (h_1, \cdots, h_M)$, as $d^*(\mathscr{P}, \boldsymbol{\eta}, \mathbf{h})$, the solution to (3.25) can be expressed as $D_\epsilon d^*(\widehat{\mathscr{P}}, \boldsymbol{\eta}, \widehat{\mathbf{h}})$, where $\widehat{\mathbf{h}} \triangleq (0, h_2, \cdots, h_M)$.

Case 2

In this case, the constraint $D_{\max} \leq D_\epsilon$, with probability $1 - \epsilon$ is translated to

$$2^{-b\sum_{i=1}^{k+1} R_i} \leq D_\epsilon, \tag{3.28}$$

or equivalently,

$$\sum_{i=1}^{k+1} R_i \geq R_\epsilon. \tag{3.29}$$

Similar to the previous case, we can consider two situations:

- $\sum_{i=1}^{k+1} R_i^{unconst.} \geq R_\epsilon$; in this case, the solution to the unconstrainted problem already satisfies the constraint $\sum_{i=1}^{k+1} R_i \geq R_\epsilon$. In this situation, the solution to (3.25) is exactly the solution of the unconstrained problem, obtained in the previous section.

- $\sum_{i=1}^{k+1} R_i^{unconst.} < R_\epsilon$; in this case, we have the following optimization problem:

$$\begin{array}{ll} \min & f(\mathbf{R}) \\ \textit{subject to} & \mathbf{R}\mathbf{I}_{k+1} \geq R_\epsilon, \ \textit{and} \ \sum_{i=1}^{M} \mathscr{P}_i \leq \mathscr{P}, \end{array} \tag{3.30}$$

where $\mathbf{R} \triangleq (R_1, \cdots, R_M)$, $\mathbf{I}_{k+1} \triangleq \sum_{i=1}^{k+1} \mathbf{e}_i$, in which $\mathbf{e}_i$ is an $M \times 1$ vector whose ith element is 1 and the rest are zero, and $f(.)$ is the function of the average distortion in terms of the rate vector $\mathbf{R}$. Using Karush-Kuhn-Tucker theorem, if $\mathbf{R}^*$ is the solution of the above problem, there exists some $\lambda \geq 0$ and $\mu \geq 0$ such that

$$\nabla f(\mathbf{R}^*) + \lambda \nabla \left(\mathbf{R}^* \mathbf{I}_{k+1}\right) + \mu \nabla h(\mathbf{R}^*) = 0, \tag{3.31}$$

$$\lambda \left(\mathbf{R}^* \mathbf{I}_{k+1} - R_\epsilon\right) = 0, \tag{3.32}$$

$$\mu \left(h(\mathbf{R}^*) - \mathscr{P}\right) = 0, \tag{3.33}$$

where $h(.)$ is the function of sum-power $(\sum_{i=1}^{M} \mathscr{P}_i)$ in terms of $\mathbf{R}$. From the second condition (complementary slackness), it follows that either $\mathbf{R}^*$ is the solution to the unconstrainted problem or we have $\mathbf{R}^* \mathbf{I}_{k+1} = R_\epsilon$. Hence, in the case that $\sum_{i=1}^{k+1} R_i^{unconst.} < R_\epsilon$, the optimization problem is equivalent to

$$\begin{array}{ll} \min & E_{h_{typ}}[D], \\ \textit{Subject to} & \sum_{i=1}^{k+1} R_i = R_\epsilon. \end{array} \tag{3.34}$$

To solve the above problem, we write $E_{h_{typ}}[D]$ as follows:

$$E_{h_{typ}}[D] = \sum_{i=1}^{k} \eta_i 2^{-b\sum_{j=1}^{i} R_j} + 2^{-bR_\epsilon} E_g[D], \tag{3.35}$$

where g is a fading process which takes the values $(0, h_{k+2}, \cdots, h_M)$ with probabilities $(\eta_{k+1}, \cdots, \eta_M)$ [2], which results in $E_g[D] = \eta_{k+1} + \sum_{i=k+2}^{M} \eta_i 2^{-b\sum_{j=k+2}^{M} R_j}$ [3]. Suppose that α portion of the available power is allocated to the first k levels. In this case, the expected distortion can be expressed as a function of α, denoted by $d(\alpha)$, which can be written from (3.35) as

$$\begin{aligned} E_{h_{typ}}[D] &= d(\alpha) \\ &= d_1(\alpha) + e^{-bR_\epsilon} d_2(\alpha), \end{aligned} \tag{3.36}$$

where $d_1(\alpha) \triangleq \sum_{i=1}^{k} \eta_i 2^{-b\sum_{j=1}^{i} R_j}$ and $d_2(\alpha) \triangleq E_g[D]$. Note that in the above equation, the minimization of $d(\alpha)$ can be performed by minimizing $d_1(\alpha)$ and $d_2(\alpha)$ separately. This is because the only dependency between the distortion of the lower layers (indexed from 1 to k) and the distortion of the upper layers (indexed from $k+1$ to M) is through the portion of the power allocated to each layer (and not *exact values of allocated power levels*) and also the sum of the rates of the lower layers (the former is set to α and the later is equal to R_ϵ). Minimization of $d_2(\alpha)$ is the unconstrained minimization of $E_g[D]$, when the total power equals $(1-\alpha)\mathscr{P}$, and its solution can be expressed as $d^*((1-\alpha)\mathscr{P}, \boldsymbol{\eta}_g, \mathbf{g})$, where $\boldsymbol{\eta}_g \triangleq (\eta_{k+1}, \cdots, \eta_M)$, and $\mathbf{g} \triangleq (0, h_{k+2}, \cdots, h_M)$. Minimization of $d_1(\alpha)$ is equivalent to minimization of $E_w[D]$, where w is a fading process taking the values $\{h_i\}_{i=1}^{k}$ with the probabilities $\{\eta_i\}_{i=1}^{k}$, with the constraint that $\sum_{i=1}^{k} R_i \leq R_\epsilon$. This constraint is because of the fact that $\sum_{i=1}^{k+1} R_i = R_\epsilon$. Hence, the minimum expected distortion can be found by

[2] Here, we assumed that $k \leq M-2$. In the case of $k = M-1$, $g = 0$ and therefore, $E_g[D] = \eta_{k+1}$.

[3] Note that here the probabilities $\{\eta_i\}_{i=k+2}^{M}$ do not add to one. However, this does not affect the validity of the results.

minimizing $d(\alpha)$ over α. More precisely,

$$E_{h_{typ}}[D]^* = \min_{\alpha} \left(\min_{\substack{\mathscr{P}_i, i=1,\cdots,k \\ \sum_{i=1}^{k} \mathscr{P}_i = \alpha\mathscr{P} \\ \sum_{i=1}^{k} R_i \le R_\epsilon}} d_1(\alpha) + D_\epsilon \min_{\substack{\mathscr{P}_i, i=k+2,\cdots,M \\ \sum_{i=k+2}^{M} \mathscr{P}_i = (1-\alpha-\alpha_\epsilon)\mathscr{P}}} d_2(\alpha) \right), \quad (3.37)$$

where α_ϵ is the portion of power allocated to the layer $k+1$ ($\mathscr{P}_{k+1} = \alpha_\epsilon$) and is determined by the constraint $\sum_{i=1}^{k+1} R_i = R_\epsilon$. The following lemma shows that the constraint $\sum_{i=1}^{k} R_i \le R_\epsilon$ in the minimization of $d_1(\alpha)$ is not required:

Lemma 4 *The minimization given by (3.37) is equivalent to*

$$E_{h_{typ}}[D]^* = \min_{\alpha} \left(\min_{\substack{\mathscr{P}_i, i=1,\cdots,k \\ \sum_{i=1}^{k} \mathscr{P}_i = \alpha\mathscr{P}}} d_1(\alpha) + D_\epsilon \min_{\substack{\mathscr{P}_i, i=k+2,\cdots,M \\ \sum_{i=k+2}^{M} \mathscr{P}_i = (1-\alpha-\alpha_\epsilon)\mathscr{P}}} d_2(\alpha) \right). \quad (3.38)$$

Proof - In order to prove the lemma, it suffices to show that there does not exist a power allocation $\{\mathscr{P}_i\}_{i=1}^{k}$ achieving the minimum distortion in (3.38), while having $\sum_{i=1}^{k} R_i > R_\epsilon$. Assume that there exists such a power allocation. Hence, the minimum total achievable distortion using this power allocation, denoted by $\mathcal{D}^*$, is lower-bounded by

$$\begin{aligned} \mathcal{D}^* &\overset{(a)}{>} \mathcal{D}^{**} \\ &\triangleq \min_{\alpha} \left(\min_{\substack{\mathscr{P}_i, i=1,\cdots,k \\ \sum_{i=1}^{k} \mathscr{P}_i = \alpha\mathscr{P}}} d_1(\alpha) + 2^{-b\sum_{j=1}^{k} R_j} \min_{\substack{\mathscr{P}_i, i=k+1,\cdots,M \\ \sum_{i=k+1}^{M} \mathscr{P}_i = (1-\alpha)\mathscr{P}}} d_2(\alpha) \right) \\ &= \min_{\substack{\mathscr{P}_i, i=1,\cdots,M \\ \sum \mathscr{P}_i = \mathscr{P} \\ \sum_{i=1}^{k} R_i > R_\epsilon}} \sum_{i=1}^{M} \eta_i 2^{-b\sum_{j=1}^{i} R_j}. \end{aligned} \quad (3.39)$$

where (a) results from the facts that $2^{-b\sum_{j=1}^{k} R_j} < D_\epsilon$. From the complementary slackness condition expressed in (3.31), it follows that $\mathcal{D}^{**} > E_{h_{typ}}[D]^*$. Combining this fact with the above equation, we have $\mathcal{D}^* > E_{h_{typ}}[D]^*$, which completes the proof of Lemma 1.

■

Note that in (3.38), the constrained optimization of $E_{h_{typ}}[D]$ is decomposed into *two unconstrained* optimization problems, for which we know the solutions. The solution for the minimization of $d_2(\alpha)$, as expressed earlier, is equal to $d^*((1-\alpha-\alpha_\epsilon)\mathscr{P}, \boldsymbol{\eta}_g, \mathbf{g})$. However, the minimization of $d_1(\alpha)$ is slightly different from the considered earlier unconstrainted minimization problem considered in [26]. The reason is that in the minimization of $d_1(\alpha)$, we should consider the fact that the lower layers experience an interference of at least $(1-\alpha)\mathscr{P}$ from the upper layers. In other words, in the reccursive algorithm explained in the previous section, the cumulative power variables $T_i = \sum_{j=i}^{M} \mathscr{P}_j$, $i = 1, \cdots, k$ are constrained to be larger than $(1-\alpha)\mathscr{P}$. Therefore, in each recurrence step, when solving the two-layer optimization problem, three situations may occur:

1) $U_{i+1} < (1-\alpha)\mathscr{P}$; in this case, $T^*_{i+1} = (1-\alpha)\mathscr{P}$. The optimal distortion of the ith layer can be obtained as

$$D^*_i = (1+\gamma_i T_i)^{-b} W_i, \tag{3.40}$$

where $W_i = (1+\gamma_i(1-\alpha)\mathscr{P})^b \left[u_i + (1+\beta_i(1-\alpha)\mathscr{P})^{-b} w_i\right]$. Also, for the $i-1$th layer, we have

$$\begin{aligned} w_{i-1} &= W_i, & \beta_{i-1} &= \beta_i, \\ u_{i-1} &= \eta_{i-1} + u_i, & \gamma_{i-1} &= h_{i-1}. \end{aligned} \tag{3.41}$$

2) $(1-\alpha)\mathscr{P} \le U_{i+1} \le T_i$; in this case, $T^*_{i+1} = U_{i+1}$.

3) $U_{i+1} > T_i$; in this case, $T^*_{i+1} = T_i$.

For the second and third situations, the optimal distortion and and the reccursive equations relating $(w_{i-1}, u_{i-1}, \beta_{i-1}, \gamma_{i-1})$ to $(w_i, u_i, \beta_i, \gamma_i)$ are exactly the same as the unconstrained problem. Denoting the minimum value of $d_1(\alpha)$, in terms of the standard unconstrainted problem, as $d^*(\alpha\mathscr{P}, (1-\alpha)\mathscr{P}, \boldsymbol{\eta}_{\mathbf{g}^C}, \mathbf{h}_{\mathbf{g}^C})$, where $\boldsymbol{\eta}_{\mathbf{g}^C} \triangleq \{\eta_i\}_{i=1}^k$ and $\mathbf{h}_{\mathbf{g}^C} \triangleq \{h_i\}_{i=1}^k$, and $d^*(\mathcal{P}, \mathcal{I}, \boldsymbol{p}, \mathbf{h})$ denotes the general solution of the unconstrained problem when the total available power is $\mathcal{P}$, the intereference from the upper levels is $\mathcal{I}$, and the fading levels and their corresponding probabilities are $\mathbf{h}$

and $\mathbf{p}$. Finally, the solution of (3.21) can be written as the following point optimization problem:

$$E_{h_{typ}}[D]^* = \min_{0\leq\alpha\leq 1} \left[d^*(\alpha\mathscr{P}, (1-\alpha)\mathscr{P}, \boldsymbol{\eta}_{\mathbf{g}^C}, \mathbf{h}_{\mathbf{g}^C})) + D_\epsilon d^*((1-\alpha)\mathscr{P}, \boldsymbol{\eta}_{\mathbf{g}}, \mathbf{h}_{\mathbf{g}})\right]. \quad (3.42)$$

3.2.3 Soft Coverage Constraint

In this section, we consider a *soft coverage constraint*, meaning that the expected distortion of the worst user in the network must be upper-bounded by a given threshold level, D_{mul}. Having this constraint, we would like to minimize the expected distortion of a typical user in the network. In other words, the solution of the following optimization problem is desired

$$\min_{\substack{\boldsymbol{P} \\ \sum_{i=1}^M \mathscr{P}_i \leq \mathscr{P}}} \quad E_{h_{typ}}[D], \quad (3.43)$$

$$\textit{Subject to} \quad E_{h_{mul}}[D] \leq D_{mul}. \quad (3.44)$$

The above problem is equivalent to

$$\min_{\substack{\boldsymbol{P} \\ \sum_{i=1}^M \mathscr{P}_i \leq \mathscr{P}}} \quad \boldsymbol{\eta}.\mathbf{D}, \quad (3.45)$$

$$\textit{Subject to} \quad \boldsymbol{\phi}.\mathbf{D} \leq D_{mul}, \quad (3.46)$$

where $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_M)$, $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_M)$, and $\mathbf{D}$ is the distortion vector defined as

$$\mathbf{D} \triangleq \left(2^{-bR_1}, \cdots, 2^{-b\sum_{j=1}^{i} R_j}, \cdots, 2^{-b\sum_{i=1}^{M} R_i}\right)^T.$$

Two situations can occur here:

- The solution of the unconstrained problem $\min_{\mathbf{P}} \boldsymbol{\eta}.\mathbf{D}$, denoted by $\mathbf{D}^*$, already satisfies $\boldsymbol{\phi}.\mathbf{D}^* \leq D_{mul}$. In this case, the solution to (3.45) is exactly equal to the solution to the unconstrained problem, i.e., $\mathbf{D}^*$.

- In the case of $\boldsymbol{\phi}.\mathbf{D}^* > D_{mul}$, we know from KKT conditions that the solution of (3.45) is equal to the solution of the following problem:

$$\min_{\substack{\boldsymbol{P} \\ \sum_{i=1}^M \mathscr{P}_i \leq \mathscr{P}}} \quad \boldsymbol{\eta}.\mathbf{D} - \lambda\boldsymbol{\phi}.\mathbf{D}, \quad (3.47)$$

for some $\lambda > 0$, with the constraint $\boldsymbol{\phi}.\mathbf{D}^*_\lambda = D_{mul}$, in which $\mathbf{D}^*_\lambda$ denotes the solution of the above problem. Defining $\boldsymbol{\eta}_\lambda \triangleq \boldsymbol{\eta} - \lambda\boldsymbol{\phi}$, the above optimization problem can be expressed as the following unconstrained problem:

$$\min_{\substack{\boldsymbol{P} \\ \sum_{i=1}^{M} \mathscr{P}_i \leq \mathscr{P}}} \quad \boldsymbol{\eta}_{\lambda^*}.\mathbf{D}, \tag{3.48}$$

where λ^* is the solution of

$$\boldsymbol{\phi}.\mathbf{D}^*_{\lambda^*} = D_{mul}. \tag{3.49}$$

Hence, the solution of (3.43) can be expressed as $d^*(\mathscr{P}, \boldsymbol{\eta}_{\lambda^*}, \mathbf{h})$, where λ^* satisfies the above equation.

3.3 Continuous Fading Distribution

3.3.1 Preview

As stated before, the typical channel can be treated as a single point-to-point channel with the distribution identical to any of the channels in the network. The minimum expected distortion for this case is derived in [26]. In fact, we can generalize the results for channels with discrete states to the continuous fading channels. For this purpose, we should assume infinite number of channel levels with even spacing Δh and derive the results when Δh tends to zero. In this setting, the channel level $h - \Delta h$ is realized with the probability $f(h)\Delta h$, where $f(h)$ is the pdf of the continuous fading channel. Using a continuous indexing for the same values defined in the discrete case and following (3.15), we can rewrite the cumulative distortion from layers h and above as

$$D^{unconst.}(h) = (1 + hT^{unconst.}(h))^{-b} W(h) \tag{3.50}$$

where $T^{unconst.}(h)$ is the total available power for layers h and above and $W(h)$ is interpreted as an equivalent probability weight summarizing the aggregate effect of the layers h and above. Following (3.15) and (3.16), we can find the optimal power allocation as

$$T^{unconst.*}(h) = \begin{cases} U^{unconst.}(h) & if\ U^{unconst.}(h) \leq T^{unconst.*}h - \Delta h) \\ T^{unconst.*}(h - \Delta h) & else \end{cases}, \tag{3.51}$$

where

$$U(h) \triangleq \begin{cases} 0 & h \geq W(h)/f(h) + \Delta h \\ \frac{1}{h}\left(\left[\frac{W(h)}{f(h)(h-\Delta h)} - 1\right]^{\frac{1}{1+b}} - 1\right) & else. \end{cases} \tag{3.52}$$

In the region where $T(h)$ is given by the unconstrained minimizer $UT^{unconst.}(h)$, $W(h)$ is derived from the following recurrence equation

$$W(h-\Delta h) = (1 + (h-\Delta h)U^{unconst.}(h))^b . \left[f(h)\Delta h + (1 + (h)U^{unconst.}(h))^{-b} W(h)\right] \tag{3.53}$$

As Δh tends to zero, the above recurrence equations give rise to a set of linear first order differential equations. Those differential equations are solved in [26] and the power allocation for a Rayleigh Fading channel with expected channel gain $\bar{h}$ is shown to be as follows:

$$T^{unconst.*}(h) = \begin{cases} 0 & h > h_0 \\ U^{unconst.}(h) & h_{\mathscr{P}} \leq h \leq h_0) \\ \mathscr{P} & h < h_{\mathscr{P}} \end{cases}, \tag{3.54}$$

where,

$$U^{unconst.}(h) = \frac{\int_{\bar{h}}^{h} \left(\frac{1}{\bar{h}} - \frac{2}{s}\right) \left[s^{1-b} e^{-\frac{s}{\bar{h}}}\right]^{\frac{1}{b+1}} \mathrm{d}s}{(1+b)\left[h^2 e^{-h/\bar{h}}\right]^{\frac{1}{b+1}}}, \tag{3.55}$$

and h_0 and $h_{\mathscr{P}}$ are obtained applying the continuity conditions of $T^*(h)$. Also, in the region $h_{\mathscr{P}} \leq h \leq h_0$, the cumulative distortion function $D(h)$, for a Rayleigh Fading channel is shown to be

$$D^{unconst.}(h) = \frac{-\frac{1}{\bar{h}} \int_{\bar{h}}^{h} e^{-s/\bar{h}} [(\frac{s}{\bar{h}})^2 e^{-(s-\bar{h})/\bar{h}}]^{-\frac{b}{b+1}} \mathrm{d}s + e^{-1}}{[(\frac{h}{\bar{h}})^2 e^{-(h-\bar{h})/\bar{h}}]^{-\frac{b}{b+1}}}. \tag{3.56}$$

As seen above, no power is allocated to the layers $h \leq h_{\mathscr{P}}$. Therefore, the minimum expected distortion of a point to point Rayleigh Fading channel is given as

$$E_h[D] = D^{unconst.}(0) = F(h_{\mathscr{P}}) + D^{unconst.}(h_{\mathscr{P}}), \tag{3.57}$$

where $F(h)$ is the CDF of the fading channel.

3.3.2 Hard Coverage Constraint

According to the definitions in the system model, the multicast outage distortion is the distortion seen by a multicast user with channel level h_ϵ, where $h_\epsilon = F_{multicast}^{-1}(\epsilon)$. Using the notation of the continuous case, the minimum expected typical rate with hard coverage constraint is as follows

$$\begin{aligned} \min_{T(.)} \quad & E_{h_{typ}}[D], \\ \textit{subject to} \quad & D_{\max} \leq D_\epsilon \quad \textit{with probability}\,(1-\epsilon), \end{aligned} \tag{3.58}$$

where $T(h)$ is the available power for the layers h and above and $D_{max} =\triangleq \max_{1\leq n \leq N} D_n$ is the instantaneous multicast distortion. Since $D_{max}(.)$ is a decreasing function of the channel gain, the constraint is satisfied if and only if $D_{max}(h_\epsilon) \leq D_\epsilon$. We can translate this distortion constraint to a rate constraint:

$$D_{max}(h_\epsilon) = e^{-b\int_0^{h_\epsilon} \mathrm{d}R_h} \leq D_\epsilon \Longrightarrow \int_0^{h_\epsilon} \mathrm{d}R_h \geq -\frac{\log D_\epsilon}{b} = R_\epsilon. \tag{3.59}$$

Similar to the second case for the discrete channel case, we can classify the constraint rate (R_ϵ) to two regions. Denoting the rate set given by the unconstrained optimization by $\{\mathrm{d}R_h^{unconst.}\}$, the solution to the constrained case when $R_\epsilon \leq \int_0^{h_\epsilon} \mathrm{d}R_h^{unconst.}$ is exactly the same as the unconstrained solution. In fact, in this case the multicast distortion requirement is not limiting in terms of the expected distortion.

On the other hand, for $\int_0^{h_\epsilon} \mathrm{d}R_h^{unconst.} \leq R_\epsilon$, it is shown as aresult of the K.K.T. conditions that our optimization problem is simplified to

$$\begin{aligned} \min_{T(.)} \quad & E_{h_{typ}}[D], \\ \textit{subject to} \quad & \int_0^{h_\epsilon} \mathrm{d}R_h = R_\epsilon. \end{aligned} \tag{3.60}$$

We can rewrite the expected distortion as follows

$$E_{h_{typ}}[D] = \int_0^\infty f(h) e^{-b\int_0^h \mathrm{d}R_u} \mathrm{d}h = \int_0^{h_\epsilon} f(h) e^{-b\int_0^h \mathrm{d}R_u} \mathrm{d}h + D_\epsilon \int_{h_\epsilon}^\infty f(h) e^{-b\int_{h_\epsilon^+}^h \mathrm{d}R_u} \mathrm{d}h \tag{3.61}$$

Similar to the unconstrained case, we can approach this optimization problem as a problem with a discrete channel with infinite number of states with even spacing Δh between adjacent channel gains. For such a channel, we can use the result of Lemma 4 which implies that, given the total power allocated to the layers lower than h_ϵ is $\alpha\mathscr{P}$, the power allocation for these layers is independent of the power allocation to the layers higher than h_ϵ and vice versa when we want to minimize the expected distortion with multicast constraint. Denoting the interference seen by level h as $T(h)$, the minimization problem will be simplified to

$$E_{h_{typ}}[D]^* = \tag{3.62}$$
$$\min_\alpha \left(\min_{\substack{T(h), 0\le h<h_\epsilon \\ T(h_\epsilon)=(1-\alpha)\mathscr{P}, T(0)=\mathscr{P}}} d_1(\alpha) + f(h_\epsilon)\Delta h D_\epsilon + D_\epsilon \min_{\substack{T(h), h_\epsilon<h \\ T(h_\epsilon+\Delta h)=(1-\alpha-\alpha_\epsilon)\mathscr{P}, T(\infty)=\mathscr{P}}} d_2(\alpha) \right)$$

where α_ϵ is the power allocated to the layer h_ϵ to fulfil the condition of sum rate of the layers not above h_ϵ being equal to R_ϵ, $d_1(\alpha)$ is the weighted distortion of the layers lower than h_ϵ according to the weighting function $f(h)$, when all of them see at least $(1-\alpha\mathscr{P})$ interference power and $d_2(\alpha)$ is the same weighted sum for the layers higher than h_ϵ when their total power is at most $(1-\alpha-\alpha_\epsilon)\mathscr{P}$. In fact, in the above derivation, we have divided the layers to three groups:

1. Layers lower than h_ϵ for which we optimize its contribution to the expected distortion with the total power $\alpha\mathscr{P}$. It is justified in the previous section that we can remove the constraint $\int_0^{h_\epsilon} \mathrm{d}R_h < R_\epsilon$.

2. Layer $h = h_\epsilon$. The output of the first optimization is the set of rates $\{\mathrm{d}R_h\}_0^{h_\epsilon-\Delta h}$. In order to fulfill the the outage multicast constraint, the rate of this layer should be $R_\epsilon - \int_0^{h_\epsilon-\Delta h} \mathrm{d}R_h$. This requires a power level equal to
$$\alpha_\epsilon\mathscr{P} = (1-\alpha)\mathscr{P} - \left(\frac{1+h_\epsilon(1-\alpha)\mathscr{P}}{e^{R_\epsilon - \int_0^{h_\epsilon-\Delta h}\mathrm{d}R_h}} - 1 \right) / h_\epsilon. \tag{3.63}$$

3. Layers higher than h_ϵ which are allocated the total power $(1-\alpha-\alpha_\epsilon)\mathscr{P}$ and their contribution to the total expected distortion is optimized. As mentioned before, the optimal power distribution is independent of power allocation for the lower levels.

Denoting the optimal interference function for our problem by $T^*(.)$, according to [26], for $h > h_\epsilon + \Delta h$, we have

$$T^*(h) = \begin{cases} T^{unconst.*}(h) & if\ T^{unconst.*}(h) < (1-\alpha-\alpha_\epsilon)\mathscr{P} \\ (1-\alpha-\alpha_\epsilon)\mathscr{P} & otherwise \end{cases}. \tag{3.64}$$

Therefore, the power level associated with the layer $h_\epsilon + \Delta h$ is given by

$$P_\alpha(h_\epsilon + \Delta h) = [(1-\alpha-\alpha_\epsilon)\mathscr{P} - T^{unconst.*}(h)]^+ \tag{3.65}$$

As $\Delta h \to 0$, this power level merges to the layer h_ϵ, while the layer h_ϵ is already decoding R_ϵ from the layers lower or equal to h_ϵ. Therefore, in the limit of $\Delta h \to 0$, if $P_\alpha(h_\epsilon + \Delta h) > 0$, then $\int_0^{h_\epsilon} \mathrm{d}R_h > R_\epsilon$, which contradicts the equality in (3.60). Hence, denoting the optimizer of (3.62) by α^*, we conclude that

$$T^{unconst.*}(h_\epsilon) > (1-\alpha-\alpha_\epsilon^*)\mathscr{P}. \tag{3.66}$$

Consequently, we can rewrite the second minimization term in (3.62) as follows

$$\min_{\substack{T(h), h_\epsilon < h \\ T(h_\epsilon + \Delta h) = (1-\alpha-\alpha_\epsilon)\mathscr{P}, T(\infty)=0}} d_2(\alpha) = D^{unconst.}(h_{(1-\alpha-\alpha_\epsilon)\mathscr{P}}) + F(h_{(1-\alpha-\alpha_\epsilon)\mathscr{P}}) - F(h_\epsilon), \tag{3.67}$$

where $T^{unconst.*}(h_{(1-\alpha-\alpha_\epsilon)\mathscr{P}}) = (1-\alpha-\alpha_\epsilon)\mathscr{P}$.

Denoting the optimal cumulative distortion function with $(1-\alpha)\mathscr{P}$ of interference power by $D^*_{(1-\alpha)\mathscr{P}}$, as in the discrete case, we can write

$$D^*_{(1-\alpha)\mathscr{P}}(h) = (1 + hT^*(h))^{-b} W(h), \tag{3.68}$$

where $W(h)$ is the probability weight capturing the aggregate effect of the layers h and above. We can adopt the results of the discrete channel case to optimize $d_1(\alpha)$ and conclude that for $h < h_\epsilon$

$$T^*(h) = \begin{cases} U(h) & if\ U(h) \le T^*(h-\Delta h) \\ T^*(h-\Delta h) & else \end{cases}, \tag{3.69}$$

where

$$U(h) \triangleq \begin{cases} (1-\alpha)\mathscr{P} & if\ \frac{1}{h}\left(\left[\frac{W(h)}{f(h)(h-\Delta h)}\right]^{\frac{1}{1+b}} - 1\right) < (1-\alpha)\mathscr{P} \\ \frac{1}{h}\left(\left[\frac{W(h)}{f(h)(h-\Delta h)}\right]^{\frac{1}{1+b}} - 1\right) & else. \end{cases} \tag{3.70}$$

The cumulative distortion function in the region which unconstrained minimizer applies, can be written as following

$$D^*_{(1-\alpha)\mathscr{P}}(h-\Delta h) = \left(\frac{1 + (h-\Delta h)T^*(h-\Delta h)}{1 + (h-\Delta h)U(h)}\right) \cdot \left[f(h)\Delta h + (1 + hU(h))^{-b} W(h)\right]. \tag{3.71}$$

where $W(h)$ is derived through a recurrence equation

$$W(h-\Delta h) = (1+(h-\Delta h)U(h))^b.[f(h)\Delta h + (1+hU(h))^{-b}W(h)]. \tag{3.72}$$

As $\Delta h \to 0$, we can rewrite (3.69) as follows

$$T^*(h) = \begin{cases} (1-\alpha)\mathscr{P} & h_0 \le h \le h_\epsilon \\ U(h) & h_{\mathscr{P}} \le h \le h_0 \\ \mathscr{P} & h < h_{\mathscr{P}} \end{cases}, \tag{3.73}$$

where h_0 is the solution of $\frac{1}{h}\left(\left[\frac{W(h)}{f(h)h}\right]^{\frac{1}{1+b}} - 1\right) = (1-\alpha)\mathscr{P}$ and $h_{\mathscr{P}}$ is the solution of $\frac{1}{h}\left(\left[\frac{W(h)}{f(h)h}\right]^{\frac{1}{1+b}} - 1\right) = \mathscr{P}$. As seen above, no power is allocated to the region $h_0 \le h \le h_\epsilon$. Hence, in this region $W(h) = F(h_\epsilon) - F(h)$. Substituting $W(h)$ in the boundary condition for h_0, we conclude

$$\frac{1}{h_0}\left(\left[\frac{F(h_\epsilon) - F(h_0)}{f(h_0)h_0}\right]^{\frac{1}{1+b}} - 1\right) = (1-\alpha)\mathscr{P}. \tag{3.74}$$

Also, when the spacing Δh approaches zero, it is shown in [26] that (3.72) will give rise to the following first order linear differential equation

$$U'(h) = -\left(\frac{2/h + f'(h)/f(h)}{1+b}\right)[U(h) + \frac{1}{h}]. \tag{3.75}$$

With the initial condition $U(h_0) = (1-\alpha)\mathscr{P}$, the solution will be

$$U(h) = \frac{-\int_{h_0}^{h} \frac{1}{s}\left(\frac{2}{s} + \frac{f'(s)}{f(s)}\right)[s^2 f(s)]^{\frac{1}{b+1}} \mathrm{d}s + (1-\alpha)\mathscr{P}(1+b)[{h_0}^2 f(h_0)]^{\frac{1}{b+1}}}{(1+b)[h^2 f(h)]^{\frac{1}{b+1}}}, \tag{3.76}$$

and $h_{\mathscr{P}}$ is obtained solving $U(h_{\mathscr{P}}) = \mathscr{P}$. In the limit of $\Delta h \to 0$, equation (3.71) also leads to a differential equation

$$D^{*\prime}_{(1-\alpha)\mathscr{P}}(h) = -\frac{bhU'(h)}{1+hU(h)} D^*_{(1-\alpha)\mathscr{P}}(h) - f(h) \tag{3.77}$$

$$= \left[\frac{b}{1+b}\left(\frac{2}{h} + \frac{f'(h)}{f(h)}\right)\right] D^*_{(1-\alpha)\mathscr{P}}(h) - f(h). \tag{3.78}$$

The solution of the above equation with the initial condition $D^*_{(1-\alpha)\mathscr{P}}(h_0) = W(h_0) = F(h_\epsilon) - F(h_0)$ will be

$$D^*_{(1-\alpha)\mathscr{P}}(h) = \frac{-\int_{h_0}^{h} f(s)[s^2 f(s)]^{\frac{-b}{b+1}}\mathrm{d}s + (F(h_\epsilon) - F(h_0)).[h_0^2 f(h_0)]^{\frac{-b}{b+1}}}{[h^2 f(h)]^{\frac{-b}{b+1}}}. \tag{3.79}$$

For a Rayleigh Fading channel with the unit mean, (3.76) and (3.79) are simplified to

$$U(h) = \frac{\int_{h_0}^{h_\epsilon}(1-\frac{2}{s})[s^{1-b}e^{-h}]^{\frac{1}{b+1}} + (1-\alpha)\mathscr{P}(b+1)[h_0^2 e^{-h_0}]^{\frac{1}{b+1}}}{(b+1)[h^2 e^{-h}]^{\frac{1}{b+1}}}, \tag{3.80}$$

$$D^*_{(1-\alpha)\mathscr{P}}(h) = \frac{-\int_{h_0}^{h} e^{-s}[s^2 e^{-s}]^{\frac{-b}{b+1}}\mathrm{d}s + (e^{h_0} - e^{h_\epsilon}).[h_0^2 e^{-h_0}]^{\frac{-b}{b+1}}}{[h^2 e^{-h}]^{\frac{-b}{b+1}}}, \tag{3.81}$$

and h_0 is the solution of $h(h(1-\alpha)\mathscr{P}+1)^{b+1} + e^{h-h_\epsilon} - 1 = 0$. For $h < h_{\mathscr{P}}$, there is no power allocated to the layers and $W(h) = F(h_{\mathscr{P}}) - F(h)$, therefore

$$\min_{\substack{T(h), 0 \le h < h_\epsilon \\ T(h_\epsilon)=(1-\alpha)\mathscr{P}, T(0)=\mathscr{P}}} d_1(\alpha) = D^*_{(1-\alpha)\mathscr{P}}(0) = F(h_{\mathscr{P}}) + D^*_{(1-\alpha)\mathscr{P}}(h_{\mathscr{P}}) \tag{3.82}$$

Equations (3.62), (3.67) and (3.82) lead to

$$E_{h_{typ}}[D]^* = \min_{\alpha}\left(F(h_{\mathscr{P}}) + D^*_{(1-\alpha)\mathscr{P}}(h_{\mathscr{P}}) + D_\epsilon[D^{unconst.}(h_{(1-\alpha-\alpha_\epsilon)\mathscr{P}}) + F(h_{(1-\alpha-\alpha_\epsilon)\mathscr{P}}) - F(h_\epsilon)]\right) \tag{3.83}$$

3.3.3 Soft Coverage Constraint

Similar to the discrete case, we can consider the expected multicast distortion as a coverage measure. In this case, the optimization problem is equivalent to

$$\min_{T(0)=\mathscr{P}, T(.) \textbf{ is decreasing}} E_{h_{typ}}[D] \tag{3.84}$$

$$\textit{Subject to} \quad E_{h_{mul}}[D] \le D_{mul}. \tag{3.85}$$

As in the discrete channel scenario, there are two cases:

1. The solution to the unconstrained problem already satisfies the multicast requirement. In this case, the unconstrained solution is optimal.

2. The solution to the unconstrained problem dose not satisfy the multicast requirement. In this case, we know that according to the KKT conditions, the optimization problem is equivalent to

$$\min_{T(0)=\mathscr{P},T(.)\ \textbf{is decreasing}} \quad E_{h_{typ}}[D] + \lambda E_{h_{mul}}[D] = \tag{3.86}$$

$$\min_{T(0)=\mathscr{P},T(.)\ \textbf{is decreasing}} \quad \int_0^\infty (f(h) - \lambda f_{mul}(h)) e^{-b\int_0^h \mathrm{d}R_u} \mathrm{d}h, \tag{3.87}$$

which is equivalent to the unconstrained problem for a channel with pdf $f_\lambda(h) = f(h) - \lambda f_{mul}(h)$. Therefore, the corresponding interference function is as follows

$$T^*(h) = \begin{cases} 0 & h_0 \leq h \\ U(h) & h_{\mathscr{P}} \leq h \leq h_0 \\ \mathscr{P} & h < h_{\mathscr{P}} \end{cases}, \tag{3.88}$$

where

$$U(h) = \frac{-\int_{h_0}^h \frac{1}{s}\left(\frac{2}{s} + \frac{f'_\lambda(s)}{f_\lambda(s)}\right)[s^2 f_\lambda(s)]^{\frac{1}{b+1}} \mathrm{d}s}{(1+b)[h^2 f_\lambda(h)]^{\frac{1}{b+1}}}, \tag{3.89}$$

h_0 and $h_{\mathscr{P}}$ are obtained by continuity conditions. In this case, the cumulative typical distortion in the region $h_{\mathscr{P}} \leq h \leq h_0$ is

$$D_{typ}(h) = \frac{-\int_{h_0}^h f_\lambda(s)[s^2 f_\lambda(s)]^{\frac{-b}{b+1}} \mathrm{d}s + (1 - F(h_0)).[h_0^2 f_\lambda(h_0)]^{\frac{-b}{b+1}}}{[h^2 f_\lambda(h)]^{\frac{-b}{b+1}}}, \tag{3.90}$$

and the cumulative multicast distortion in the same region, is given by

$$D_{mul}(h) = \frac{-\int_{h_0}^h f_\lambda(s)[s^2 f_\lambda(s)]^{\frac{-b}{b+1}} \mathrm{d}s + (1 - F_{mul}(h_0)).[h_0^2 f_\lambda(h_0)]^{\frac{-b}{b+1}}}{[h^2 f_\lambda(h)]^{\frac{-b}{b+1}}}, \tag{3.91}$$

where λ is computed through the following equation

$$E_{h_{mul}} = D_{mul}(0) = F_{mul}(h_{\mathscr{P}}) + D_{mul}(h_{\mathscr{P}}) = D_{mul}. \tag{3.92}$$

Having obtained λ, we can write the expected typical distortion as

$$E_{h_{typ}} = D_{typ}(0) = F(h_{\mathscr{P}}) + D_{typ}(h_{\mathscr{P}}). \tag{3.93}$$

Chapter 4

Conclusion and Future Work

We have considered a multicast network, where a common data is transmitted from a sender to several users. It is assumed that a minimum service must be provided for all the users. For this setup, we have optimized the average service received by a typical user in the network. Two scenarios are considered for the coverage constraint. In the case of hard coverage constraint, the minimum multicast requirement is stated in terms of an outage rate received by all the users in a single transmission block. For small enough outage probabilities, it is shown that the optimal rate region is achieved by providing the required multicast rate in a single layer code, and designing an infinite-layer code as in [10], on top of it. In the case of soft coverage constraint, the multicast requirement is expressed in terms of the expected multicast rate received by all the users. An infinite layer superposition coding is shown to achieve the capacity region (C_{mul}, C_{ave}). We have also proposed a suboptimal coding scheme for the MISO multicast channel. This scheme is shown to be asymptotically optimal, when the number of transmit antennas grows at least logarithmically with the number of users. Finally we have extended our results to the case where multiple sources are sharing the same channel each to transmit to a group of users.

For the proposed constrained multicast network, we have also considered the problem of minimum expected typical distortion for the transmission of a Gaussian source. We have proposed a multi-resolution source code mapped to a multi-level channel code as a joint source-channel coding scheme and optimized it based on the characteristics of channel and source. We first solved the problem for a finite state fading channel for each user and then extended it to the case of continuous fading channels. The output of the optimization problem is the power allocation for different channel code layers each carrying a refinement of the source.

Since the joint source-channel coding scheme proposed in our work is not globally optimal, there is a motivation to find the optimal code to minimize the expected distortion in a constrained multicast network. Schemes like analog coding could be designed and optimized for such a network.

We can also generalize the proposed setup in this work to the multi-relay networks when all the relays are listening to a common source.

Bibliography

[1] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information theoretic and communication aspects," *IEEE Trans. Inform. Theory,* vol. 44, pp. 26192692, Oct. 1998.

[2] S. Sesia, G. Caire, and G. Vivier, "Broadcasting a common source: Information theoretic results and system challenges," in Proc. *IEEE Int. Symp. Information Theory(ISIT03)*, Monte Verita, Switzerland, Feb. 2427, 2003

[3] S. Sesia, G. Caire, and G. Vivier, "Lossy Transmission over Slow-Fading AWGN Channels: a Comparison of Progressive, Superposition and Hybrid Approaches," in Proc. *IEEE Int. Symp. Information Theory(ISIT05)*, Adelaide, Australia, Sep. 4-9, 2005,p. 224-228

[4] Nadav Shulman and Meir Feder. "Static broadcasting," In Proc. *IEEE Int. Symp. Information Theory(ISIT2000)*, page 23, 2000.

[5] N. Jindal and A. J. Goldsmith, "Capacity and optimal power allocation for fading broadcast channels with minimum rates," *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 28952909, Nov. 2003

[6] Jianghong Luo, Roy Yates and Predrag Spasojevic, "Service Outage Based Power and Rate Allocation for Parallel Fading Channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 25942611, Jul. 2005.

[7] C. T. K. Ng and A. J. Goldsmith, "Capacity of Fading Broadcast Channels with Rate Constraints," *Forty-Second Annual Allerton Conference on Communication, Control and Computing*, September 29 October 1, 2004, Monticello, IL.

[8] N. Jindal and Z. Q. Luo, "Capacity Limits of Multiple Antenna Multicast ," in *Proc. IEEE Int. Symp. Information Theory (ISIT06)*, Seatle, USA, June July 9-14 2006.

[9] W. Equitz and T. Cover, "Successive refinement of information," *IEEE Trans. on Inform. Theory*, 37(2): 269-275, Mar. 1991.

[10] S. Shamai (Shitz) and A. Steiner, "A broadcast approach for a single-user slowly fading MIMO channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2617-2635, Oct. 2003.

[11] S. Shamai (Shitz), "A broadcast strategy for the Gaussian slowly fading channel," in *Proc. IEEE Int. Symp. Information Theory (ISIT97)*, Ulm, Germany, June 29July 4 1997, p. 150.

[12] Y. Liu, K. N. Lau, C. Y. Takeshita, and M. P. Fitz, "Optimal rate allocation for superposition coding in quasistatic fading channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT02)*, Lausanne, Switzerland, June 30July 5 2002, p. 111.

[13] I. M. Geldfand and S. V. Fomin, *Calculus of Variations*, Englewood Cliffs, NJ: Prentice-Hall, 1963.

[14] D. Gunduz, E. Erkip, "Source and channel coding for quasi-static fading channels," In *Proc. of 39th Asilomar Conf.*, Pacific Grove, CA, Nov. 2005.

[15] B. Dunn, J. N. Laneman, "Characterizing source-channel diversity approaches beyond the distortion exponent," in *Proc. Allerton Conf.*,Monticello, IL, Sept. 2005.

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

[17] H.Weingarten,Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, Volume 52, Issue 9, Sept. 2006 Page(s):3936 - 3964.

[18] P. Bergmans and T. Cover, "Cooperative Broadcasting," *IEEE Trans. Inf. Theory*, Volume 20, Issue 3, May 1974 Page(s):317-324.

[19] A. Papoulis and S. U. Pialli, *Probability, Random Variables and Stochastic Processes.* New York: McGraw-Hill,2002.

[20] C. E. Shannon,"Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, part 4, pp. 142163, 1959.

[21] Reza Mirghaderi, Alireza Bayesteh and Amir K. Khandani, "On the Maximum Achievable Rates in Wireless Multicast Networks," in *Proc. 2007 IEEE International Symposium on Information Theory* (Nice, France, June 24-29 2007)

[22] M. Gastpar, B. Rimoldi, and M. Vetterli,"To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 11471158, May 2003.

[23] T. J. Goblick, Jr.,"Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inform. Theory*, vol. 11, no. 4, pp. 558567, Oct. 1965.

[24] U. Mittal and N. Phamdo, "Hybrid digital-analog (HDA) joint source-channel codes for broadcasting and robust communications," *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 10821102, May 2002.

[25] C. Tian, A. Steiner, S. Shamai (Shitz), and S. Diggavi, "Expected distortion for Gaussian source with a broadcast transmission strategy over a fading channel," submitted to *IEEE Inform. Theory Workshop*, Sept. 2007.

[26] Chris T. K. Ng, Deniz Gndz, Andrea J. Goldsmith and Elza Erkip, "Optimal Power Distribution and Minimum Expected Distortion in Gaussian Layered Broadcast Coding with Successive Refinement," submitted to *IEEE Transactions on Information Theory*, 2007

[27] B. Rimoldi, "Successive refinement of information: Characterization of the achievable rates," *IEEE Trans. Inform. Theory*, vol. 40, no. 1, pp. 253259, Jan. 1994.

[28] D. Gunduz and E. Erkip, "Joint source-channel codes for MIMO block fading channels," *IEEE Trans. Inform. Theory*, submitted.