Capacity of Fading Broadcast Channels with Rate Constraints

Chris T. K. Ng
Dept. of Electrical Engineering
Stanford University
Stanford, CA 94305
ngctk@stanford.edu

Andrea J. Goldsmith
Dept. of Electrical Engineering
Stanford University
Stanford, CA 94305
andrea@ee.stanford.edu

Abstract

For fading broadcast channels, different capacity regions such as the ergodic, outage, minimum rate, or limited-jitter capacity region can be represented by a triplet of rate constraint parameters from each user: a maximum rate, a minimum rate, and a shortage probability over which the minimum rate needs not be guaranteed. For each fading state, the capacity region is obtained by evaluating a finite set of possible extreme points. The optimal power allocation strategy across fading states is determined by the power requirements associated with the maximum and minimum rates, combined with water-filling for the excess power when the maximum and minimum rate constraints are not tight. Shortage capacity can be obtained by optimizing the power allocation over the shortage fading states, when the minimum rate constraints are removed, and the non-shortage fading states.

1 Introduction

While the capacity region for fading broadcast channels with different constraints are known: ergodic capacity [1, 2], outage capacity [3], and minimum rate capacity [4], this paper focuses on the scenario where each user is able to specify a set of rate constraint parameters on the channel. This model is motivated by the desire to support heterogeneous applications with different rate requirements over a common wireless channel. The user specifies a maximum rate the system can handle, a minimum rate required for the application, and a shortage probability over which the minimum rate stipulation might be waived. Effectively, when in shortage, the system resorts to a best-service resource provisioning scheme, in which minimum rates are no longer guaranteed.

Each user specifies its rate constraint parameters in a triplet of maximum rate R^m, minimum rate R^n, and shortage probability q. Let (R^m_j, R^n_j, q_j) be the parameters given by user j, with $0 \leq R^n_j \leq R^m_j$ and $0 \leq q_j \leq 1$. The maximum rate may represent system bottlenecks (e.g., limited encoder/decoder processing rate, buffer access speed) such that even if additional power is available, the system still cannot transmit to the user at a higher rate. On the other hand, setting a minimum rate with shortage probability means user j must receive rate equal to or above R^n_j with probability $1 - q_j$. Shortage capacity differs from outage capacity [3] in that when the system is in shortage, while it is short of providing a minimum rate guarantee, transmission might still take place at a lower rate. If transmission power is set to zero when the system is in shortage, then shortage capacity reduces to outage capacity. For simplicity, a common shortage mode is assumed, in which all users are required to declare shortage simultaneously ($q_1 = \ldots = q_M = q$). Setting both the maximum and minimum rate constraints also limits the rate jitter of the transmission, which can be used to bound buffer size for store-and-forward relay nodes.

The rate constraint parameters can be set to represent different types of channel capacity regions. For example, having a minimum rate of zero and maximum rate of infinity leads to the ergodic capacity, while setting the minimum rate the same as the maximum rate represents zero-outage capacity. The types of capacity regions represented by the rate constraint parameters are shown in Table 1.

2 System Model

The analysis adopts the same system model as defined in [2, 3, 4]. It assumes a discrete-time flat-fading Gaussian broadcast channel, with perfect channel state information at the transmitter and at all receivers. The transmitter varies its transmit power (thus controlling the transmission rate) under an average power constraint, and superposition coding with successive decoding [5] is used.

At each time instance i, the transmitter sends a signal $X_j(i)$ with power $p_j(i)$ to user j over bandwidth B. The average total transmission power over time cannot exceed $\overline{P}$. To characterize the channel, let v_j be the noise density of user j's channel, and $\sqrt{g_j(i)}$ its time-varying channel gain at time i. Then we combine the channel gain and noise density to form the time-varying noise density $n_j(i) = v_j/g_j(i)$; since $n_j(i)$ incorporates the time-varying channel state, it will also be referred to as a fading state. Signals to all M users are sent simultaneously, hence the received signal for user j becomes $Y_j(i) = \sum_{k=1}^{M} X_k(i) + z_j(i)$, where $z_j(i)$ is a Gaussian random variable with zero mean and variance $n_j(i)B$. It is assumed that the noise density vector $\mathbf{n}(i) = (n_1(i), \ldots, n_M(i))$ is known to the transmitter and all receivers at each time instance i. Using superposition coding with successive decoding, the rate for user j at fading state $\mathbf{n}$ is

$$R_j(\mathbf{n}) = B \log \left(1 + \frac{p_j(\mathbf{n})}{n_j B + \sum_{i=1}^{M} p_i(\mathbf{n})\, \mathbf{1}[n_j > n_i]} \right) \tag{1}$$

where $\mathbf{1}[\cdot]$ is the indicator function. Since B is simply a constant scaling factor, it can be assumed to be 1 without loss of generality. This paper will focus on the case where the channel has two users ($M = 2$).

3 Additive White Gaussian Noise Channel

We first consider an additive white Gaussian noise (AWGN) channel where there is no fading. The capacity derived in this case, when n_1, n_2 are constants, will subsequently be used to formulate the optimal power allocation strategy for a fading channel. We assume $n_2 > n_1$ (user 2 has a noisier channel); in the case of $n_1 > n_2$, the derivation remains valid with the subscripts 1, 2 reversed. Let user 1 and user 2 have rate requirements $(R_1^m, R_1^n, 0)$ and $(R_2^m, R_2^n, 0)$, respectively (since the channel has no fading, we set shortage probability q_1 and q_2 to 0). Letting the noise density vector $\mathbf{n} = (n_1, n_2)$, we define the weighted sum of rates function

$$R(p_1, p_2, \mathbf{n}) = \mu_1 \log \left(1 + \frac{p_1}{n_1} \right) + \mu_2 \log \left(1 + \frac{p_2}{p_1 + n_2} \right), \tag{2}$$

where p_1 is the power allocated to user 1, and p_2 the power allocated to user 2. The parameters μ_1 and μ_2 allow achievable rates of user 1 and user 2 to be weighted differently;

Capacity Type	*Rate Constraint Parameters*
Ergodic [1, 2]	$(R^m, R^n, q) = (\infty, 0, 0)$
Zero-outage [3]	$R^m = R^n$, $q = 0$
Min rate [4]	$(R^m, R^n, q) = (\infty, r_{min}, 0)$
Outage [3]	$R^m = R^n$, $q = p_{out}$
Limited-jitter	$R^m - R^n = r_{jitter}$

Table 1: Types of capacity regions represented by rate constraint parameters.

Line	*Constraint Type*
ℓ_T	total avg power
ℓ_{M_1}	user 1 max rate
ℓ_{N_1}	user 1 min rate
ℓ_{M_2}	user 2 max rate
ℓ_{N_2}	user 2 min rate

Table 2: Power region boundaries of the rate constraints.

$0 < \mu_1 < 1$, and $\mu_2 = 1 - \mu_1$. To determine the capacity region, it is needed to maximize the weighted sum of rates subject to the total power and rate constraints:

$$\begin{aligned} &\max_{p_1, p_2} \; R(p_1, p_2, \mathbf{n}) \\ &\quad \text{subject to: } p_1 + p_2 \le P, \; (R_1^m, R_1^n, 0), \; (R_2^m, R_2^n, 0) \end{aligned} \tag{3}$$

Letting μ_1 range from 0 to 1 in (3) sketches out the capacity region.

The capacity region is a function of (p_1, p_2), the noise density vector $\mathbf{n}$, and the total available power P. In an AWGN channel, $\mathbf{n}$ is known and remains constant. Suppose total power P is available for transmission, then the problem is to find out how to optimally divide up P between the two users; i.e., to determine the values of p_1^* and p_2^*. Let the function $(p_1^*, p_2^*) = F_A(P, \mathbf{n})$ denote this optimal power allocation between the two users. Given P and $\mathbf{n}$, $F_A(P, \mathbf{n})$ returns the optimal p_1^* and p_2^*, which can then be substituted back into (3) to obtain the capacity region boundary under optimal power allocation.

Successive decoding allows user 1 to subtract out the signal power of user 2. Therefore, user 1 is not affected by user 2. To user 2, however, user 1's signal power p_1 is effectively noise. For user 2 to maintain a constant rate R_2^m, as derived in [4], its power p_2 needs to be accordingly increased whenever p_1 is increased: $p_2 = (e^{R_2^m} - 1)p_1 + n_2(e^{R_2^m} - 1)$. Likewise, if a constant rate of R_2^n is to be maintained: $p_2 = (e^{R_2^n} - 1)p_1 + n_2(e^{R_2^n} - 1)$. For convenience, define the following constants:

$$\begin{aligned} P_1^m &\triangleq n_1(e^{R_1^m} - 1) & P_1^n &\triangleq n_1(e^{R_1^n} - 1) \\ G_m &\triangleq e^{R_2^m} - 1 & G_n &\triangleq e^{R_2^n} - 1 \\ P_2^m &\triangleq G_m(P_1^m + n_2) & P_2^n &\triangleq G_n(P_1^n + n_2). \end{aligned}$$

Then the constraints of (3) can be expressed in terms of the following five inequalities: $p_1 + p_2 \le P$, $p_1 \le P_1^m$, $p_1 \ge P_1^n$, $p_2 \le G_m(p_1 + n_2)$, and $p_2 \ge G_n(p_1 + n_2)$. Note that all of the constraints are linear in p_1, p_2. Let Ω denote the region over p_1, p_2 where all constraints are met. Since $P \ge 0$ and $0 \le R_j^n \le R_j^m$ for each user $j = 1, 2$, this implies $P_1^m \ge P_1^n$, $G_m \ge G_n$ and $P_2^m \ge P_2^n$. These conditions lead to a contiguous and closed region Ω on the p_1-p_2 plane, as shown in Fig. 1. Let the lines ℓ_T, ℓ_{M_1}, ℓ_{N_1}, ℓ_{M_2} and ℓ_{N_2} denote the boundaries of Ω, which correspond, respectively, to the above constraints with equality being taken, as summarized in Table 2.

An extreme point is defined by the intersection of two boundaries. Note that a point (p_1, p_2) can be represented as a vector $\mathbf{p} = p_1 \mathbf{i} + p_2 \mathbf{j}$, where $\mathbf{i}$, $\mathbf{j}$ are unit vectors in the p_1, p_2 directions, respectively. Six extreme points $\mathbf{M_1}$, $\mathbf{N_1}$, $\mathbf{M_2}$, $\mathbf{N_2}$, $\mathbf{X}$ and $\mathbf{Y}$ are of particular interest; their definitions are tabulated in Table 3.

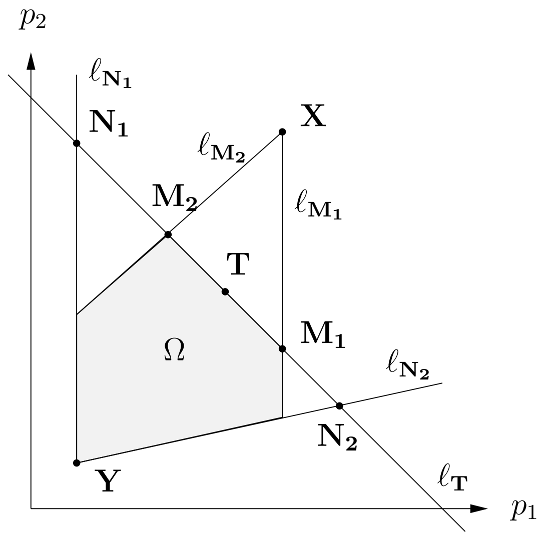

Figure 1: Rate constraints on power allocation for user 1, 2.

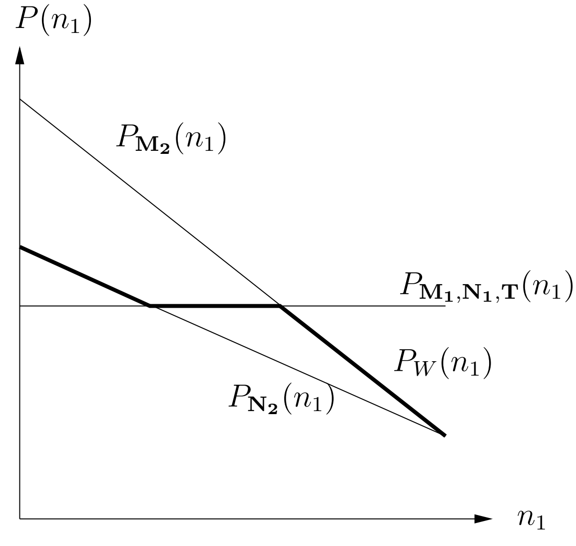

Figure 2: Power allocation for different noise density states with respect to n_1, with the combined water-filling power allocation $P_W(n_1)$ shown in bold.

Point	*Intersection*	*Constraint Type*
$\mathbf{M_1}$	ℓ_T, ℓ_{M_1}	maximum rate for user 1
$\mathbf{N_1}$	ℓ_T, ℓ_{N_1}	minimum rate for user 1
$\mathbf{M_2}$	ℓ_T, ℓ_{M_2}	maximum rate for user 2
$\mathbf{N_2}$	ℓ_T, ℓ_{N_2}	minimum rate for user 2
$\mathbf{X}$	ℓ_{M_1}, ℓ_{M_2}	minimum rate for both users
$\mathbf{Y}$	ℓ_{N_1}, ℓ_{N_2}	minimum rate for both users

Table 3: Constraint boundary extreme points.

For a fixed noise vector $\mathbf{n}$, the weighted sum of rates function

$$R(\mathbf{p}) = \mu_1 \log \left(1 + \frac{p_1}{n_1}\right) + \mu_2 \log \left(1 + \frac{p_2}{p_1 + n_2}\right), \tag{4}$$

being differentiable everywhere, is differentiable on the closed region Ω. Therefore, $R(\mathbf{p})$ must take on an absolute maximum on Ω, which can be found as follows:

1. First calculate the gradient of $R(\mathbf{p})$:

$$\nabla R(\mathbf{p}) = \left(\frac{\mu_1}{p_1 + n_1} - \frac{\mu_2 p_2}{(p_1 + n_2)(p_1 + p_2 + n_2)}\right) \mathbf{i} + \frac{\mu_2}{p_1 + p_2 + n_2} \mathbf{j}. \tag{5}$$

   The gradient, defined everywhere, is never $\mathbf{0}$. In particular, since $\mu_2 > 0$, its $\mathbf{j}$-component will always be positive. Therefore, the maximum point of $R(\mathbf{p})$ is on one of the boundaries of Ω.

2. Consider the boundaries ℓ_{M_2} and ℓ_{N_2}. Define unit vectors $\mathbf{u_m}$, $\mathbf{u_n}$, respectively, to have the same directions as ℓ_{M_2}, ℓ_{N_2} (in the increasing p_1-p_2 polarity):

$$\mathbf{u_m} = \frac{\mathbf{i} + G_m \mathbf{j}}{\sqrt{G_m^2 + 1}}, \qquad \mathbf{u_n} = \frac{\mathbf{i} + G_n \mathbf{j}}{\sqrt{G_n^2 + 1}}. \tag{6}$$

The directional derivatives of $R(\mathbf{p})$ with respect to $\mathbf{u_m}$ and $\mathbf{u_n}$, respectively, are given by:

$$R'_{\mathbf{u_m}}(\mathbf{p}) = \frac{1}{\sqrt{G_m^2+1}}\frac{\mu_1}{p_1+n_2}, \qquad R'_{\mathbf{u_n}}(\mathbf{p}) = \frac{1}{\sqrt{G_n^2+1}}\frac{\mu_1}{p_1+n_2}. \tag{7}$$

Since both $R'_{\mathbf{u_m}}(\mathbf{p})$ and $R'_{\mathbf{u_n}}(\mathbf{p})$ are always positive, $R(\mathbf{p})$ is monotonically increasing in the directions of $\mathbf{u_m}$ and $\mathbf{u_n}$. Therefore, if the maximum point lies on ℓ_{M_2} or ℓ_{N_2}, it has to be one of the extreme points $\mathbf{M_2}$, $\mathbf{N_2}$, or $\mathbf{X}$.

3. Next consider the boundaries ℓ_{M_1} and ℓ_{N_1}. Both of them are parallel to the p_2 axis, and the derivative of $R(\mathbf{p})$ in the direction of $\mathbf{j}$ is

$$R'_{\mathbf{j}}(\mathbf{p}) = \nabla R(\mathbf{p}) \cdot \mathbf{j} = \frac{\mu_2}{p_1+p_2+n_2}. \tag{8}$$

Similarly, since the directional derivative $R'_{\mathbf{j}}(\mathbf{p})$ is always positive, $R(\mathbf{p})$ is monotonically increasing in the direction of $\mathbf{j}$. Therefore, if the maximum point lies on ℓ_{M_1} or ℓ_{N_1}, it has to be one of the extreme points $\mathbf{M_1}$ or $\mathbf{N_1}$.

4. If the maximum point does not lie on ℓ_{M_2}, ℓ_{N_2}, ℓ_{M_1}, or ℓ_{N_1}, then it has to be on the boundary ℓ_T within the open interval between points $\mathbf{M_1}$ and $\mathbf{M_2}$. Denote this point as $\mathbf{T}$. Then optimization (3) reduces to a standard one-dimensional maximization problem. The derivative of the capacity region along ℓ_T with respect to p_1 is

$$\frac{dR_{\ell_T}}{dp_1}(p_1) = \frac{(\mu_1-\mu_2)p_1+\mu_1 n_2-\mu_2 n_1}{(p_1+n_1)(p_1+n_2)}. \tag{9}$$

Since $n_2 > n_1$, the derivative is always positive for $\mu_1 \geq \mu_2$. In this case $R_{\ell_T}(p_1)$ is monotonically increasing in the direction of p_1, so $\mathbf{M_1}$ or $\mathbf{N_2}$ are the possible maximum points rather than $\mathbf{T}$. For $\mu_1 < \mu_2$, $\mathbf{T}$ is given by

$$\mathbf{T} = \left(\frac{\mu_1 n_2-\mu_2 n_1}{\mu_2-\mu_1}, P - \frac{\mu_1 n_2-\mu_2 n_1}{\mu_2-\mu_1}\right). \tag{10}$$

5. Finally, when available total power equals the minimum power requirement $P = P_1^n + P_2^n$, all of the above operating points close in to meet at the minimum rate extreme point $\mathbf{Y}$.

Therefore, there are only seven potential maximum point candidates: $\mathbf{M_1}$, $\mathbf{N_1}$, $\mathbf{M_2}$, $\mathbf{N_1}$, $\mathbf{T}$, $\mathbf{X}$ and $\mathbf{Y}$, of which the ones that lie outside of Ω need not be considered. Thus the maximum of $R(\mathbf{p})$ can be obtained by evaluating the candidate points that lie on the boundaries of Ω, and choosing the one that yields the largest value, i.e.,

$$\begin{gathered} F_A(P,\mathbf{n}) = \mathbf{v} \text{ such that } R(\mathbf{v}) \geq R(\mathbf{v}') \ \forall\ \mathbf{v}' \neq \mathbf{v}, \\ \text{where } \mathbf{v},\mathbf{v}' \in \{\mathbf{M_1},\mathbf{N_1},\mathbf{M_2},\mathbf{N_2},\mathbf{T},\mathbf{X},\mathbf{Y}\} \cap \Omega. \end{gathered} \tag{11}$$

The coordinates, i.e., the power allocation p_1 and p_2, of the possible maximum points are presented in Table 4. The power allocation p_1 and p_2 at each possible maximum point can then be substituted in (4) to obtain the capacity achieved.

Point	p_1	p_2	*Constraint Type*
$\mathbf{M_1}$	P_1^m	$P - P_1^m$	maximum rate for user 1
$\mathbf{N_1}$	P_1^n	$P - P_1^n$	minimum rate for user 1
$\mathbf{M_2}$	$\frac{1}{G_m+1}(P - G_m n_2)$	$\frac{G_m}{G_m+1}(P + n_2)$	maximum rate for user 2
$\mathbf{N_2}$	$\frac{1}{G_n+1}(P - G_n n_2)$	$\frac{G_n}{G_n+1}(P + n_2)$	minimum rate for user 2
$\mathbf{T}$	$\frac{\mu_1 n_2 - \mu_2 n_1}{\mu_2 - \mu_1}$	$P - \frac{\mu_2 n_1 - \mu_1 n_2}{\mu_1 - \mu_2}$	total average power
$\mathbf{X}$	P_1^m	P_2^m	maximum rate for both users
$\mathbf{Y}$	P_1^n	P_2^n	minimum rate for both users

Table 4: Possible maximum points.

State	$A_{\mathbf{v}}$	$B_{\mathbf{v}}$	$C_{\mathbf{v}}$	$D_{\mathbf{v}}$
$S_{\mathbf{M_1}}$	μ_2	P_1^m	$P_1^m + n_2$	$\mu_1 R_1^m$
$S_{\mathbf{N_1}}$	μ_2	P_1^n	$P_1^n + n_2$	$\mu_1 R_1^n$
$S_{\mathbf{M_2}}$	μ_1	$G_m n_2$	$(G_m + 1)n_1$	$\mu_2 R_2^m$
$S_{\mathbf{N_2}}$	μ_1	$G_n n_2$	$(G_n + 1)n_1$	$\mu_2 R_2^n$
$S_{\mathbf{T}}$	μ_2	$\frac{\mu_2 n_1 - \mu_1 n_2}{\mu_1 - \mu_2}$	$\frac{\mu_2}{\mu_1 - \mu_2}(n_1 - n_2)$	$\mu_1 \log\left(1 + \frac{\mu_2 n_1 - \mu_1 n_2}{(\mu_1 - \mu_2)n_1}\right)$

Table 5: Rate function parameters for each noise density state.

4 Broadcast Fading Channel

In this section, the capacity region with rate constraints of a broadcast fading channel will be studied. It is assumed that the shortage probability is zero; effects of shortage are analyzed in Section 5. For a given total power $P(\mathbf{n})$ and noise density $\mathbf{n}$, optimally allocating the power between the users is solved in Section 3: the power allocation between user 1 and user 2 is given by (11) and the capacity region boundary is given by (3). Hence, the remaining step is to determine the optimal power allocation across fading states $\mathbf{n}$.

In a fading channel, n_1, n_2 are random variables with known joint probability distribution. Based on the noise density vector $\mathbf{n} = (n_1, n_2)$, the total power $P(\mathbf{n})$ can be varied. We partition the noise density vector into states $\{S_{\mathbf{M_1}}, S_{\mathbf{N_1}}, S_{\mathbf{M_2}}, S_{\mathbf{N_2}}, S_{\mathbf{T}}, S_{\mathbf{X}}, S_{\mathbf{Y}}\}$ based on the optimal power allocation:

$$\mathbf{n} \in S_{\mathbf{v}} \text{ if } F_A(P, \mathbf{n}) = \mathbf{v}, \text{ where } \mathbf{v} \in \{\mathbf{M_1}, \mathbf{N_1}, \mathbf{M_2}, \mathbf{N_2}, \mathbf{T}, \mathbf{X}, \mathbf{Y}\}. \tag{12}$$

Namely, if the optimal power allocation for a noise vector $\mathbf{n}$ is $\mathbf{v}$, where $\mathbf{v}$ is one of the seven possible maximum points with power allocation given in Table 4, then $\mathbf{n}$ is in $S_{\mathbf{v}}$. For states $S_{\mathbf{X}}$ and $S_{\mathbf{Y}}$, since the rates are limited by the maximum or the minimum rate constraints instead of the available transmit power, the channel has constant rates $R_1^m + R_2^m$ and $R_1^n + R_2^n$, respectively. For the other states, the maximum weighted sum of rates as a function of the total transmit power $P(\mathbf{n})$ and noise vector $\mathbf{n}$ has the following form:

$$R_P(P(\mathbf{n}), \mathbf{n}) = A_{\mathbf{v}} \log\left(1 + \frac{P(\mathbf{n}) - B_{\mathbf{v}}}{C_{\mathbf{v}}}\right) + D_{\mathbf{v}}, \tag{13}$$

where $A_{\mathbf{v}}$, $B_{\mathbf{v}}$, $C_{\mathbf{v}}$ and $D_{\mathbf{v}}$ are parameters specific to the noise density state $S_{\mathbf{v}}$. The values of these parameters are tabulated in Table 5.

To find the optimal power allocation strategy for each $\mathbf{n}$, i.e., to determine $P(\mathbf{n})$, it is needed to maximize the following:

$$\max_{P(\mathbf{n})} \ \mathrm{E}_{\mathbf{n}}\left[R_P(P(\mathbf{n}), \mathbf{n})\right] \tag{14}$$

$$\text{subject to: } \mathrm{E}_{\mathbf{n}}[P(\mathbf{n})] \leq \overline{P}. \tag{15}$$

State	$A_{\mathbf{v}}$	$n^*_{\mathbf{v}}(n_1, n_2)$
$S_{\mathbf{M_1}}$	μ_2	n_2
$S_{\mathbf{N_1}}$	μ_2	n_2
$S_{\mathbf{M_2}}$	μ_1	$(G_m+1)n_1 - G_m n_2$
$S_{\mathbf{N_2}}$	μ_1	$(G_n+1)n_1 - G_n n_2$
$S_{\mathbf{T}}$	μ_2	n_2

Table 6: Water-filling parameters for each noise density state.

For $S_{\mathbf{X}}$ and $S_{\mathbf{Y}}$, power allocation is dictated by the maximum and minimum rates:

$$P_{\mathbf{X}}(\mathbf{n}) = P_1^m + P_2^m \qquad\qquad P_{\mathbf{Y}}(\mathbf{n}) = P_1^n + P_2^n \tag{16}$$

where $P_{\mathbf{X}}(\mathbf{n})$, $P_{\mathbf{Y}}(\mathbf{n})$ is the power allocation strategy for states $S_{\mathbf{X}}$, $S_{\mathbf{Y}}$, respectively. For the other states, we substitute (13) in (14) to obtain:

$$\begin{gathered}\max_{P(\mathbf{n})} \mathrm{E}_{\mathbf{n}}\left[A_{\mathbf{v}} \log\left(1 + \frac{P(\mathbf{n}) - B_{\mathbf{v}}}{C_{\mathbf{v}}}\right) + D_{\mathbf{v}}\right] \\ \text{subject to: } \mathrm{E}_{\mathbf{n}}[P(\mathbf{n})] \le \overline{P}.\end{gathered} \tag{17}$$

Using the Lagrangian method, we form

$$J(P(\mathbf{n})) = A_{\mathbf{v}} \log\left(1 + \frac{P(\mathbf{n}) - B_{\mathbf{v}}}{C_{\mathbf{v}}} + D_{\mathbf{v}}\right) - \lambda(\mathrm{E}_{\mathbf{n}}[P(\mathbf{n})] - \overline{P}), \tag{18}$$

then set $\frac{\partial J}{\partial P(\mathbf{n})}$ to zero to obtain:

$$P(\mathbf{n}) = \begin{cases} A_{\mathbf{v}}\frac{1}{\lambda} - (C_{\mathbf{v}} - B_{\mathbf{v}}) & A_{\mathbf{v}}\frac{1}{\lambda} \ge C_{\mathbf{v}} - B_{\mathbf{v}} \\ 0 & \text{else} \end{cases} \tag{19}$$

where $\frac{1}{\lambda}$ is the water-filling level. We define effective noise $n^*_{\mathbf{v}}(n_1, n_2) = C_{\mathbf{v}} - B_{\mathbf{v}}$. The values of $A_{\mathbf{v}}$ and $n^*_{\mathbf{v}}$ for each state are summarized in Table 6. Notice that states S_{M_1}, S_{N_2} and S_T all have $A_{\mathbf{v}} = \mu_2$ and the same effective noise $n^*_{\mathbf{v}} = n_2$. Therefore, there are only three water-filling equations:

$$P_{\mathbf{M_1},\mathbf{N_1},\mathbf{T}}(\mathbf{n}) = \lfloor \mu_2 \frac{1}{\lambda} - n_2 \rfloor_+ \tag{20}$$

$$P_{\mathbf{M_2}}(\mathbf{n}) = \lfloor \mu_1 \frac{1}{\lambda} - (G_m + 1)n_1 + G_m n_2 \rfloor_+ \tag{21}$$

$$P_{\mathbf{N_2}}(\mathbf{n}) = \lfloor \mu_1 \frac{1}{\lambda} - (G_n + 1)n_1 + G_n n_2 \rfloor_+ \tag{22}$$

Next, for a given noise vector $\mathbf{n}$, it is needed to determine its state $S_{\mathbf{v}}$ and choose the corresponding water-filling equation (20), (21), or (22). Consider the power allocation $P(\mathbf{n})$ in the direction n_1 for each state. Observe that as given in Table 4, as n_1 increases, $\mathbf{T}$ moves from $\mathbf{N_2}$ towards $\mathbf{M_2}$; therefore, in the direction of increasing n_1 the states can only appear in this order: $\{S_{\mathbf{N_2}}, S_{\mathbf{T}}, S_{\mathbf{M_2}}\}$. Coupled with the fact that, as illustrated in Fig. 2 with respect to n_1 , the slope of $P_{\mathbf{M_1},\mathbf{N_1},\mathbf{T}}(n_1)$ is greater than or equal to the slope of $P_{\mathbf{N_2}}(n_1)$, which is in turn greater than or equal to the slope of $P_{\mathbf{M_2}}(n_1)$; hence, the water-filling power allocation can be represented compactly by

$$P_W(\mathbf{n}) = \min\{\max\{P_{\mathbf{N_2}}(\mathbf{n}), P_{\mathbf{M_1},\mathbf{N_1},\mathbf{T}}(\mathbf{n})\}, P_{\mathbf{M_2}}(\mathbf{n})\}. \tag{23}$$

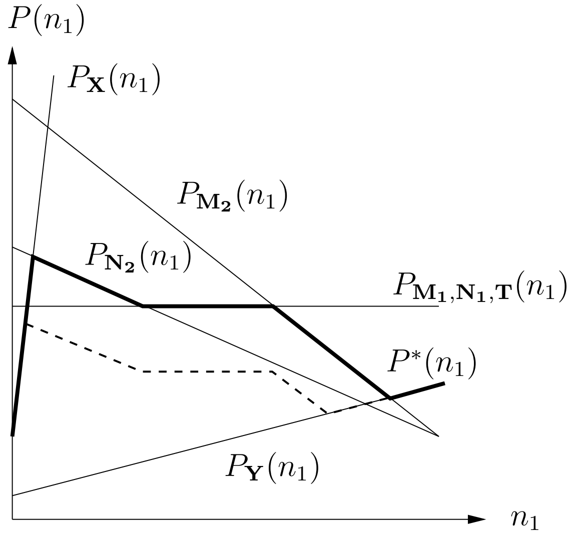

Figure 3: Water-filling combined with constant rate power allocation. The dashed line denotes a lower water-filling level than the one in bold.

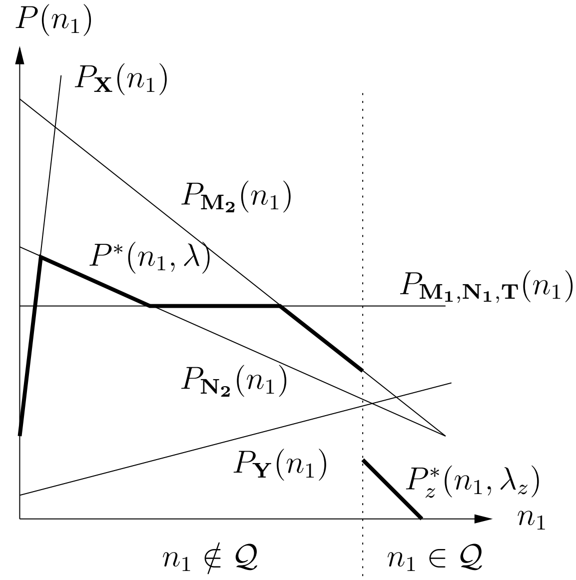

Figure 4: Power allocation with shortage probability.

Finally, combining (23) with the maximum and minimum rate constraints, the optimal power allocation strategy is:

$$P^*(\mathbf{n}) = \min\{\max\{P_W(\mathbf{n}), P_{\mathbf{Y}}(\mathbf{n})\}, P_{\mathbf{X}}(\mathbf{n})\}. \tag{24}$$

The water-filling level $\frac{1}{\lambda}$ can be determined by choosing the maximum water-filling level allowed by the average total power constraint:

$$\lambda = \lambda^* \text{ such that } \max_{\lambda^*} \ \mathrm{E}_{\mathbf{n}}[P(\mathbf{n},\lambda^*)] \le \overline{P}, \tag{25}$$

and the capacity region achieved is simply the weighted average of the capacity region over the fading states. The combined power allocation and water-filling level are illustrated in Fig. 3. Note that the water-filling level $\frac{1}{\lambda}$ does not affect the slopes of the power allocation segments with respect to n_1, but only introduces a vertical offset. For example, the dashed line represents a lower water-filling level $\frac{1}{\lambda}$ than the one in bold; they have the same slopes but different vertical offsets.

5 Shortage Probability

In the common shortage mode, when the system declares shortage for a certain fading state $\mathbf{n}$, the minimum rate constraints for all M users are removed. Effectively, this is same optimization problem as before with minimum rates $R_j^n = 0\ \forall j = 1,\ldots,M$. Therefore, similar procedures from Section 3 can be used to deduce the optimal power allocation.

1. First calculate the power allocation function as previously derived when the shortage probability is zero. Write $P^*(\mathbf{n},\lambda)$ to denote power allocation in terms of its water-filling level $\frac{1}{\lambda}$, and $P^*(\mathbf{n})$ when λ is determined by the average total power constraint (25).

2. Repeat the same calculations with the minimum rate constraints removed, i.e., set $R_j^n = 0\ \forall j$; denote the power allocation function obtained as $P_z^*(\mathbf{n},\lambda_z)$, and when the water-filling level $\frac{1}{\lambda_z}$ is determined, as $P_z^*(\mathbf{n})$.

3. Select the set of shortage fading states $\mathcal{Q}$ to contain the fading states $\mathbf{n}$ that results in maximum power savings while not exceeding the specified shortage probability q:

$$\max_{\mathcal{Q}} \sum_{\mathbf{n}\in\mathcal{Q}} P^*(\mathbf{n}) - P_z^*(\mathbf{n}) \text{ such that } \sum_{\mathbf{n}\in\mathcal{Q}} \Pr(\mathbf{n}) \leq q \tag{26}$$

4. As shown in Fig. 4, the final shortage capacity power allocation strategy $P_{\mathcal{Q}}(\mathbf{n})$ is obtained from combining $P^*(\mathbf{n},\lambda)$ and $P_z^*(\mathbf{n},\lambda_z)$ over their respective associated domain of $\mathbf{n}$.

$$P_{\mathcal{Q}}^*(\mathbf{n},\lambda,\lambda_z) = \begin{cases} P^*(\mathbf{n},\lambda) & \mathbf{n} \notin \mathcal{Q}, \\ P_z^*(\mathbf{n},\lambda_z) & \mathbf{n} \in \mathcal{Q}. \end{cases} \tag{27}$$

The water-filling levels $\frac{1}{\lambda}$, $\frac{1}{\lambda_z}$ need to be calculated to satisfy the average total power constraint (25). However, since the removal of minimum rate constraints represents a discontinuous change of capacity region with respect to $\mathbf{n}$, the water-filling levels λ and λ_z are independent. After constraint (25) is applied, there is still an extra degree of freedom, namely, the division of power between the shortage fading states and the non-shortage fading states. Unfortunately, as λ and λ_z have no closed form analytic expressions, standard optimization techniques cannot be applied. Notwithstanding, in practice when the shortage probability q specified is small, $\frac{1}{\lambda_z}$ is often zero (i.e., assign all power to non-shortage fading states). This is because small q limits $\mathcal{Q}$ to only contain fading states with large noise density $\mathbf{n}$, over which the optimal power allocation strategy simply is to suspend transmission. In this case shortage capacity reduces to outage capacity.

6 Numerical Results

In this section numerical results are presented. In all plots it is assumed that the average total transmission power $\overline{P}$ is 10 mW, and the channel has a bandwidth B of 100 KHz. The shortage probability for both users is zero. In Fig. 5, the fading state distribution is: $(n_1B,\ n_2B) = (1 \times 10^{-4}$ mW, 1 mW) with probability 1/2, and $(n_1B,\ n_2B) =$ (1 mW, 1×10^{-4} mW) with probability 1/2. The channel has a large 40 dB SNR fluctuation between the fading states for each user. As expected, imposing a minimum rate constraint of 300 Kbps for each of the users reduces the capacity region. When a shortage probability of 0.1 is specified, the minimum rate constraints are relaxed for 10% of the time, and the capacity region is slightly expanded. When each user specifies a maximum rate constraint of 1200 Kbps in addition to the minimum rate of 300 Kbps, the capacity region is further reduced compared to the minimum rate region. Note that the 1200 Kbps maximum rate actually lies outside of the ergodic capacity region. However, it still represents a limitation on the system since the transmitter can no longer send at arbitrarily high rates (within the transmit power constraint) in a favorable fading state. In Fig. 6, the channel has the fading distribution: $(n_1B,\ n_2B) = (0.001$ mW, 0.1 mW) with probability 1/2, and $(n_1B,\ n_2B) = (0.1$ mW, 0.001 mW) with probability 1/2. In this case, the channel has a relatively smaller 20 dB fluctuation in SNR between the fading states for each user. Therefore, imposing a minimum rate of 300 Kbps does not reduce the capacity region drastically. When a more restrictive additional maximum rate constraint of 900 Kbps is imposed, it can be observed that only then the capacity region is reduced moderately.

7 Conclusion

In this paper it has been shown that different types of channel capacity regions, e.g., ergodic, outage, minimum rate, or a heterogeneous combination of them, can be repre-

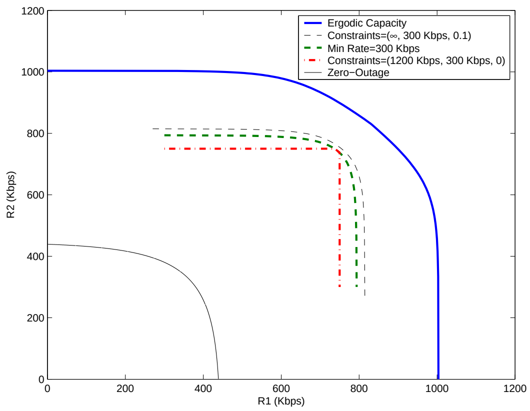

Figure 5: Capacity regions of a symmetric two-user fading broadcast channel, with 40 dB SNR difference between fading states.

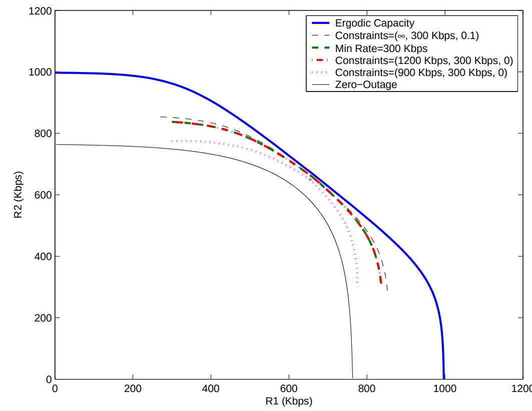

Figure 6: Capacity regions of a symmetric two-user fading broadcast channel, with 20 dB SNR difference between fading states.

sented by a triplet of rate constraint parameters from each user. Optimal power allocation between the users can be readily determined by evaluating a finite set of possible extreme points. The resulting capacity region with respect to total power is either constant, or has a common form that is logarithmically proportional to the total power relative to an effective noise. This common capacity region expression allows power allocation across fading states to be optimized, which is shown to be a modified water-filling with rates determined by the effective noise of the corresponding fading states, then combined with constant rate power allocation. Finally, it is shown that the shortage capacity region can be similarly obtained by removing the minimum rate constraints when the system is in one of the shortage fading states. Therefore, the rate constraint parameter model provides a unifying framework in which channel capacity regions with heterogeneous requirements from different users can be analyzed to obtain the corresponding capacity region and optimal power allocation strategy.

References

[1] D. Hughes-Hartog, "The capacity of the degraded spectral gaussian broadcast channel," Ph.D. dissertation, Stanford University, 1975.

[2] L. Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels—Part I: Ergodic capacity," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 1083–1102, March 2001.

[3] ——, "Capacity and optimal resource allocation for fading broadcast channels—Part II: Outage capacity," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 1103–1127, March 2001.

[4] 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. 2895–2909, Nov. 2003.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 1991.