Capacity and Optimal Power Allocation for Fading Broadcast Channels With Minimum Rates

Nihar Jindal, *Student Member, IEEE,* and Andrea Goldsmith, *Senior Member, IEEE*

***Abstract*—We derive the capacity region and optimal power allocation scheme for a slowly fading broadcast channel in which minimum rates must be maintained for each user in all fading states, assuming perfect channel state information at the transmitter and at all receivers. We show that the minimum-rate capacity region can be written in terms of the ergodic capacity region of a broadcast channel with an effective noise determined by the minimum rate requirements. This allows us to characterize the optimal power allocation schemes for minimum-rate capacity in terms of the optimal power allocations schemes that maximize ergodic capacity of the broadcast channel with effective noise. Numerical results are provided for different fading broadcast channel models.**

***Index Terms*—Broadcast channel, capacity region, fading channels, minimum rates, optimal resource allocation.**

I. Introduction

THE time-varying nature of the underlying channel is one of the most significant challenges in designing wireless communication systems. Dynamic allocation of power, bandwidth, and rate can result in significant performance improvements over constant resource allocation strategies. Practical systems are beginning to incorporate more and more elements of adaptation in order to effectively utilize the time-varying channels found in most wireless systems.

In this paper, we focus on the downlink of a single cell where one base station transmits independent information to multiple receivers and each receiver suffers from time-varying flat-fading and additive Gaussian noise. We assume that the transmitter and all receivers can track the channel fade perfectly, or in other words, that the transmitter and all receivers have perfect channel state information (CSI). Furthermore, we assume the channel is *slowly fading* relative to codeword length, i.e., the channel is constant during transmission of a codeword.

Two notions of Shannon capacity have been developed for multiuser fading channels: ergodic capacity and outage capacity. Ergodic capacity is concerned with achieving long-term rates averaged over all fading states [1]–[3], while outage capacity achieves a constant rate in all non-outage fading states subject to an outage probability [4], [5]. Zero-outage capacity refers to outage capacity with zero outage probability [6].

Manuscript received October 4, 2001; revised December 13, 2002. This work was supported by the Office of Naval Research under Grants N00014-99-1-0578 and N00014-99-1-0698.

The authors are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (email: njindal@systems.stanford.edu; andrea@systems.stanford.edu).

Communicated by D. N. C. Tse, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2003.819328

The ergodic capacity of a fading broadcast channel determines the maximum achievable long-term rates averaged over all fading states. The optimal resource allocation scheme for rates in the ergodic capacity region is found in [3], [7] and corresponds to multilevel water-filling over both time (i.e., fading states) and users. As intuition would suggest, users are allocated the most power when their channels are strong, and little, if any, power when their channels are weak. Such an allocation scheme maximizes long-term average rates, but depending on the duration of channel fades, users with poor channels may not receive data for long periods of time while waiting for their channel to improve. This clearly may not be reasonable for delay-sensitive applications such as video or voice transmission.

In the outage capacity region of a broadcast channel, each user maintains a constant rate some percentage of the time and no data is transmitted (i.e., an outage is declared) the rest of the time. In essence, no data is transmitted to a user when his channel is weak because it takes a great deal of power to transmit data over a weak channel. Constant rates are maintained in all other states. The optimal power allocation scheme is essentially a multiuser extension of channel inversion. This scheme eliminates all channel variation seen by the receivers by scaling the transmitted signal to invert fading so constant rates can be maintained during nonoutage. Because constant-rate transmission requires more power in a weak channel than in a strong channel, users are allocated the most power when their channels are weak. This is in sharp contrast to the allocation scheme used to maximize ergodic rates, where users are allocated the most power when their channels are strongest. It is therefore clear that stronger channel states are not truly taken advantage of and, as a result, the outage capacity region may be significantly smaller than the ergodic capacity region. Zero-outage capacity is a special case of outage capacity in which no outage is allowed and constant rates must be maintained in all fading states.

Ergodic and outage capacity are clearly two very different performance measures, as reflected by their contrasting power allocation strategies. In ergodic capacity, the transmitter *takes advantage* of time variation in the channel by transmitting more data to users with strong channels, while in outage capacity the transmitter *equalizes* time variation by transmitting at constant rates in all non-outage states. For a system which simultaneously transmits delay-sensitive and delay-insensitive data, neither of these approaches appears optimal. It is not desirable to shut off users for long periods of time as is possible in the ergodic capacity region, but forcing constant rates to be maintained subject only to an outage probability as is done in the outage capacity region severely reduces the set of achievable rates.

In this work, we propose to combine the notions of ergodic and zero-outage capacity by maximizing the ergodic capacity subject to minimum rate requirements for all users in all fading states. Thus, some power is used to maintain the minimum rates in all fading states while the remaining power is used to maximize the average rates in excess of the minimum rates. Users are never completely cut off due to the minimum rate requirements, but time variation of the channel is still taken advantage of by transmitting to users at rates higher than the minimum rates when their channels are strong and at exactly the minimum rates when their channels are poor. Clearly, the minimum rate requirement must be in the zero-outage capacity region for the rates to be achievable in all states.

We consider a slowly fading channel that is assumed to be constant over the duration of each codeword. Thus, we associate an instantaneous rate with each user in every fading state. The minimum-rate capacity region is defined as the set of all average rates achievable subject to an average power constraint such that the instantaneous rates in each fading states do not violate a minimum rate constraint. We show that the minimum-rate capacity region is equal to the sum of the minimum-rate vector plus the *ergodic* capacity region of an effective noise channel, where the effective noise depends on the minimum rate requirements. This relationship allows us to easily characterize the boundary of the minimum-rate capacity region and the optimal power allocation policies in terms of known results for ergodic capacity [1], [3], [7].

We then extend these results to find the minimum-rate capacity region subject to a peak power constraint instead of an average power constraint, and also subject to both a peak and average power constraint. Furthermore, the problem of minimum rates with outage is also addressed. When outage is allowed, ergodic capacity is maximized with the constraint that minimum rates must be satisfied at least a certain percentage of time. This is a combination of ergodic capacity and outage capacity, as opposed to non-outage minimum-rate capacity, which is a combination of ergodic and zero-outage capacity. A similar notion of minimum-rate outage capacity was independently proposed by Luo *et al.* in [8], [9] for single-user channels.

The remainder of this paper is organized as follows. Section II describes the flat-fading broadcast channel model and Section III defines ergodic and zero-outage capacity. In Section IV, we precisely define the minimum-rate capacity region. In Section V, we characterize the minimum-rate capacity region in terms of the ergodic capacity region and find the optimal power allocation schemes. In Section VI, we find the minimum-rate capacity region with peak power constraints and in Section VII, we find the minimum-rate outage capacity region. Numerical results are presented in Section VIII, followed by our conclusions.

Notation: We use boldface to denote vectors and $\mathbb{E}_n$ to denote expectation over the random variable n.

II. The Fading Broadcast Channel

We consider a Gaussian broadcast channel with a single transmitter communicating independent information to K users over bandwidth B. The signal source $X[i]$ is composed of K independent information sources, where i represents the time index. The time-varying channel gain of the path to user j is denoted by $\sqrt{g_j[i]}$. Each receiver has additive Gaussian noise with noise density v_j. The received signal of user j then is

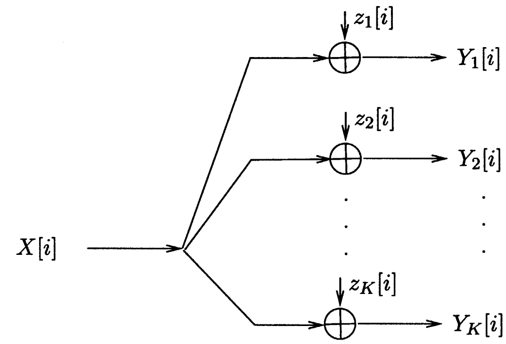

Fig. 1. Equivalent broadcast channel.

$$Y_j[i] = \sqrt{g_j[i]}X[i] + w_j[i] \tag{1}$$

where $w_j[i]$ is white Gaussian noise with power v_jB. By incorporating the channel gain into the noise term as in [3], we define an effective noise density[1] $n_j[i] = v_j/g_j[i]$ and get an equivalent form for the received signal

$$Y_j[i] = X[i] + z_j[i] \tag{2}$$

where $z_j[i]$ is Gaussian noise with power $n_j[i]B$. The equivalent channel model is shown in Fig. 1. For simplicity, we assume $B = 1$ throughout this paper.

We assume that the noise density vector

$$\boldsymbol{n}[i] = (n_1[i], n_2[i], \ldots, n_K[i])$$

is known to the transmitter and all K receivers at time instant i. The transmitter can therefore vary the power of the signal transmitted to each user $P_j[i]$ as a function of the noise vector $\boldsymbol{n}[i]$ subject to an average power constraint $\overline{P}$. Since all receivers have knowledge of $\boldsymbol{n}$, each receiver can perform successive decoding in which the decoding order depends on the ordering of $\boldsymbol{n}$. We also assume that the fading state $\boldsymbol{n}$ has some joint distribution.

As the noise density vector incorporates the effects of the channel gain, we will alternatively refer to $\boldsymbol{n}$ as the *fading state* throughout this paper.

III. Ergodic and Zero-Outage Capacity Regions

In this section, we present results from [3], [4] on the ergodic and zero-outage capacity of the fading broadcast channel.

A. Ergodic Capacity Region

The ergodic capacity region is defined as the set of all long-term average rates achievable in a fading channel with arbitrarily small probability of error. In [3], the ergodic capacity region and optimal power allocation scheme for the fading

[1]Notice that the noise density is the instantaneous power of the noise and is not the instantaneous noise sample.

broadcast channel is found by decomposing the fading channel into a parallel set of constant broadcast channels, one for each fading state $\boldsymbol{n}$. In each fading state, the channel can be viewed as a degraded Gaussian broadcast channel. Since the transmitter and all receivers know $\boldsymbol{n}[i]$, superposition coding according to the ordering of the current noise density vector can be used by the transmitter. Each receiver can perform successive decoding in which the signals of weaker users (i.e., users with larger noise power) are decoded and subtracted off before decoding the desired signal. Furthermore, the power transmitted to each user $P_j(\boldsymbol{n})$ is a function of the current fading state.

We define a power policy $\mathcal{P}$ over all possible fading states as a function that maps from any fading state $\boldsymbol{n}$ to the transmitted power $P_j(\boldsymbol{n})$ for each user. Let $\mathcal{F}$ denote the set of all power policies satisfying average power constraint $\overline{P}$

$$\mathcal{F} \equiv \left\{ \mathcal{P} : \mathbb{E}_{\boldsymbol{n}} \left[\sum_{j=1}^{K} P_j(\boldsymbol{n}) \right] \leq \overline{P} \right\}.$$

The capacity of user j assuming a constant fading state $\boldsymbol{n}$ under superposition coding and successive decoding is

$$R_j(\boldsymbol{P}(n)) = \log \left(1 + \frac{P_j(\boldsymbol{n})}{n_j + \sum_{k=1}^{K} P_k(\boldsymbol{n}) \mathbf{1}[n_j > n_k]} \right) \tag{3}$$

where $\mathbf{1}[.]$ is the indicator function.

Furthermore, let $\mathcal{C}_{BC}(\mathcal{P})$ denote the set of achievable rates averaged over all fading states (i.e., long-term rates) for power policy $\mathcal{P}$

$$\mathcal{C}_{BC}(\mathcal{P}) = \{ R_j : R_j \leq \mathbb{E}_{\boldsymbol{n}} [R_j(\boldsymbol{P}(n))], \quad j = 1, 2, \ldots, K \}$$

where $R_j(\boldsymbol{P}(n))$ is defined in (3). From [3, Theorem 1], the ergodic capacity region of the broadcast channel with perfect CSI at the transmitter and receivers and power constraint $\overline{P}$ is

$$\mathcal{C}_{\text{ergodic}}(\overline{P}) = \bigcup_{\mathcal{P} \in \mathcal{F}} \mathcal{C}_{BC}(\mathcal{P}). \tag{4}$$

Additionally, the region $\mathcal{C}_{\text{ergodic}}(\overline{P})$ is convex. The optimal power allocation scheme that achieves the boundary points of the ergodic capacity region is a multilevel extension of water-filling. Because the data rate varies from state to state, a different codebook (a codebook is assumed to have codewords for all K users) is used in every joint fading state, as in the multiplexing strategy described in [3], [10]. This coding scheme works in either a slow-fading or fast-fading environment, but the decoding delay is highly dependent on the correlation time of the channel because of the multiplexing structure. An achievability proof and a converse are provided in [3].

B. Zero-Outage Capacity Region

For the K-user broadcast channel, a rate vector $\boldsymbol{R} = (R_1, R_2, \ldots, R_K)$ is in the zero-outage capacity region if and only if the rate vector can be achieved in all fading states while meeting the average power constraint. The zero-outage capacity region (also referred to as the *delay-limited capacity*) for the multiple-access channel is derived in [6]. In [4], it is shown that rates in the zero-outage capacity region of the broadcast channel can be achieved using superposition coding and successive decoding (using the same weakest to strongest decoding order used to achieve ergodic capacity).

From [4, eq. (3)], the minimum power to support a rate vector $\boldsymbol{R}$ in fading state $\boldsymbol{n}$ is

$$P^{\min}(\boldsymbol{R}, \boldsymbol{n}) = \sum_{k=1}^{K} \left[e^{\sum_{j=k+1}^{K} R_{\pi(j)}} \left(e^{R_{\pi(k)}} - 1 \right) n_{\pi(k)} \right] + \left(e^{R_{\pi(K)}} - 1 \right) n_{\pi(K)} \tag{5}$$

where $\pi(.)$ is the permutation such that

$$n_{\pi(1)} < n_{\pi(2)} < \cdots < n_{\pi(K)}.$$

Therefore, the zero-outage capacity region is the union of all rate vectors that meet the average power constraint

$$\mathcal{C}_{\text{zero}}(\overline{P}) = \bigcup_{\mathbb{E}_{\boldsymbol{n}}[P^{\min}(\boldsymbol{R}, \boldsymbol{n})] \leq \overline{P}} (R_1, R_2, \ldots, R_K). \tag{6}$$

The boundary of the capacity region is the set of all rate vectors $\boldsymbol{R}$ such that the power constraint is met with equality [4]. For the two-user broadcast channel with time-varying additive white Gaussian noise (AWGN) with powers n_1 and n_2, the boundary assumes the following form:

$$\begin{aligned} &\Pr\{n_1 < n_2\} [\mathbb{E}[n_1 | n_1 > n_2] e^{R_2} (e^{R_1} - 1) + \\ &\qquad \mathbb{E}[n_2 | n_1 < n_2] (e^{R_2} - 1)] + \\ &\Pr\{n_1 \geq n_2\} [\mathbb{E}[n_2 | n_1 \geq n_2] e^{R_1} (e^{R_2} - 1) + \\ &\qquad \mathbb{E}[n_1 | n_1 \geq n_2] (e^{R_1} - 1)] = \overline{P}. \end{aligned}$$

For a single-user channel, this reduces to

$$\mathcal{C}_{\text{zero}}(\overline{P}) = \left\{ R : R \leq \log \left(1 + \frac{\overline{P}}{\mathbb{E}[n_1]} \right) \right\}.$$

The zero-outage capacity region depends only on the expected value of the noise in the single-user case. Similarly for the two-user broadcast channel, the zero-outage region is determined solely by $\mathbb{E}[n_1 | n_1 < n_2]$, $\mathbb{E}[n_2 | n_1 < n_2]$, $\mathbb{E}[n_1 | n_1 \geq n_2]$, and $\mathbb{E}[n_2 | n_1 \geq n_2]$. This is due to the fact that the power required to achieve a rate vector is a linear function of the noise levels in each state, as seen in (5). The zero-outage capacity region depends on the conditional expectations as opposed to the unconditional expectations of the noises because every different ordering of noises leads to a different expression for the power required in each state, also seen in (5).

The zero-outage capacity region is more formally defined as the set of rate vectors for which there exist codebooks that can be decoded with a delay *independent* of the channel correlation structure (i.e., the speed of the fading) for any desired nonzero probability of error. This is in stark contrast to the ergodic capacity, in which the decoding delay is highly dependent on the channel correlation.

IV. Minimum-Rate Capacity Region

A. Definition of Capacity Region

We define the minimum-rate capacity region of a K-user broadcast channel as the region of all achievable average rate vectors subject to an average power constraint $\overline{P}$ and minimum rate constraints $\boldsymbol{R}^* = (R_1^*, R_2^*, \ldots, R_K^*)$. The minimum rate constraint forces the instantaneous rate of each user to be at least as large as its corresponding minimum rate in all fading states, i.e., we require $R_j(\boldsymbol{n}) \geq R_j^*, \ j = 1, \ldots, K, \ \forall \boldsymbol{n}$. Since we are dealing with slowly fading channels that are assumed to be constant over the length of a codeword, the notion of an instantaneous rate $R_j(\boldsymbol{n})$ in each fading state is reasonable. Moreover, the set of achievable instantaneous rates in each fading state is equal to the capacity region of the constant Gaussian broadcast channel defined by the joint fading state and the amount of power allocated to each user.

Using the previously stated notion of a power allocation scheme, let $\mathcal{C}_{\min}(\mathcal{P})$ denote the set of achievable long-term average rates in excess of the minimum rates for power policy $\mathcal{P}$

$$\mathcal{C}_{\min}(\mathcal{P}) = \{R_j : R_j^* \leq R_j \leq \mathbb{E}_{\boldsymbol{n}}[R_j(\boldsymbol{P}(n))] \quad j = 1, \ldots, K\}$$

where $R_j(\boldsymbol{P}(n))$ is defined in (3). Notice this definition is slightly different from the definition of $\mathcal{C}(\mathcal{P})$ in Section III-A. The set $\mathcal{C}_{\min}(\mathcal{P})$ does not include the rates below the minimum rates because if the average rates are less than the minimum rates, then the minimum rates must be violated in some fading states.

To ensure that the minimum rates are satisfied, we must restrict the set of feasible power policies more tightly than in the case of ergodic capacity. Let $\mathcal{F}'$ denote the set of all power policies that satisfy the minimum rate constraints in every fading state and the average power constraint $\overline{P}$

$$\mathcal{F}' \equiv \left\{ \mathcal{P} : \mathbb{E}_{\boldsymbol{n}}\left[\sum_{j=1}^{K} P_j(\boldsymbol{n})\right] \leq \overline{P}, \quad R_j(\boldsymbol{P}(n)) \geq R_j^* \quad \forall j, \boldsymbol{n} \right\}.$$

The additional constraint ensures that the minimum rates can be maintained for all K users in every fading state for any power policy in $\mathcal{F}'$.

Definition 1: The minimum-rate capacity region of a fading broadcast channel with perfect CSI at the transmitter and receivers, average power constraint $\overline{P}$, and minimum rate constraint $\boldsymbol{R}^* = (R_1^*, R_2^*, \ldots, R_K^*) \in \mathcal{C}_{\text{zero}}(\overline{P})$ is

$$\mathcal{C}_{\min}(\overline{P}, \boldsymbol{R}^*) = \operatorname{Co}\left(\bigcup_{\mathcal{P} \in \mathcal{F}'} \mathcal{C}_{\min}(\mathcal{P}) \right) \tag{7}$$

where $\operatorname{Co}(\cdot)$ denotes the convex hull operation. The achievability of this region follows from the achievability proof for ergodic capacity given in [3] and standard time sharing arguments.

B. Remarks on Coding

In the slowly fading channel model which we consider, the channel is assumed to be constant over the duration of a codeword. If the transmitter and receivers use a multiplexing strategy similar to that of [10], then a different rate vector and a different set of codebooks is associated with every joint fading state. In this context, minimum-rate capacity is the set of all achievable average rates such that the instantaneous rates in every fading state meet the minimum rate requirements. The associated decoding delay at each user is equal to the codeword length, which can be arbitrarily long due to our slow fading assumption.

Since our definition of minimum-rate capacity explicitly mentions instantaneous rates (i.e., rates associated with each fading state), no converse exists for this formulation. A more Shannon-theoretic formulation of minimum-rate capacity which would not require the slow fading assumption might consider transmitting delay-sensitive data at the minimum rate with a delay independent of the channel variation (similar to zero-outage capacity), while simultaneously maximizing transmission of delay-insensitive data with no delay requirement (similar to ergodic capacity). In this setting, it appears natural to transmit using two independent codebooks, one for the delay-sensitive data and one for the delay-insensitive data. However, as we discuss below, it appears to be quite difficult to apply this approach to the broadcast channel.

In Section V-D, we discuss a coding strategy for the single-user channel such that the minimum rate data (i.e., the codeword from the minimum rate codebook) can be decoded before the codeword from the excess rate codebook. This allows the minimum rate data to be decoded with a delay that is independent of the rate of channel variation, but the decoding delay associated with the excess rate (i.e., above the minimum rate) data can be infinite. This coding strategy works in both slow-fading and fast-fading environments. However, this scheme does not generalize to the multiuser broadcast channel because the successive decoding structure (which is capacity achieving for the broadcast channel) essentially precludes the possibility of all users having finite delays associated with their minimum rate data and infinite delays associated with their excess rate data. Since successive decoding is needed in the broadcast channel, strong users are required to decode and cancel out the codewords intended for weaker users before being able to decode their own codewords. This must include a cancellation of the minimum rate data and the excess rate data of other users. Thus, the decoding delay of the strongest user is at least as large as the maximum of the decoding delay of all other users. If users have a possibly infinite delay associated with decoding the excess rate data, then the decoding delay associated with the minimum rate codebook of the strongest user can also be infinite. One possibility is for all users to treat all excess rate codewords (including their own) as noise while decoding their minimum rate codewords, but this appears to be quite suboptimal. In this paper, we concentrate solely on the slow-fading channel in which coding can be performed in each fading state and we leave the subject of minimum-rate capacity for fast-fading channels as a topic for future research.

C. Relationship With Ergodic and Zero-Outage Capacity Regions

The minimum-rate capacity region is closely related to the zero-outage and ergodic capacity regions because min-

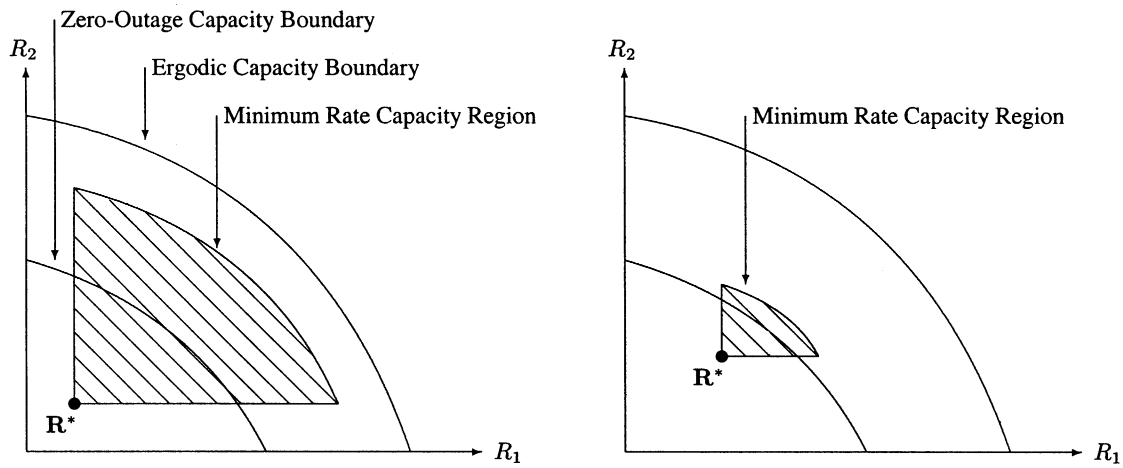

Fig. 2. Ergodic, zero-outage, and minimum-rate capacity regions for small (left) and large (right) minimum rates.

imum-rate capacity is essentially a combination of these two capacities. Some fraction of the available power is used to achieve the minimum rates in all fading states, while the remaining power is used to maximize the long-term rates achievable in excess of the minimum rates. For the minimum rate problem to be feasible, the minimum rate vector must be in the zero-outage capacity region of the channel in order for the rates to be achievable in all fading states. For any $\boldsymbol{R}^* \in \mathcal{C}_{\text{zero}}(\overline{P})$, the boundary of the minimum-rate capacity region lies between the boundaries of the zero-outage capacity region and the ergodic capacity region

$$\mathcal{C}_{\text{zero}}(\overline{P}) \subseteq \textbf{Boundary}\{\mathcal{C}_{\min}(\overline{P}, \boldsymbol{R}^*)\} \subseteq \mathcal{C}_{\text{ergodic}}(\overline{P}). \quad (8)$$

This follows from the definition of zero-outage capacity as the set of rates achievable in all fading states and from the definition of ergodic capacity as the set of all achievable average rates, without any minimum rate constraints. If the minimum rates of all users are zero, the minimum-rate capacity region is the same as the ergodic capacity region. If the minimum rate vector $\boldsymbol{R}^*$ is on the boundary of the zero-outage capacity region, achieving the minimum rate vector in all states consumes all available power and rates in excess of the minimum rates are not possible. In this situation, the minimum-rate capacity region consists of only one point, $\boldsymbol{R}^*$. When $\boldsymbol{R}^*$ is nonzero and not on the boundary of the zero-outage capacity region, the boundary of the minimum-rate capacity region lies strictly between $\mathcal{C}_{\text{zero}}(\overline{P})$ and $\mathcal{C}_{\text{ergodic}}(\overline{P})$.

To illustrate the relationship between the different capacity regions, Fig. 2 shows the ergodic, zero-outage, and minimum-rate capacity regions for two different minimum rate constraints. The corner point of the minimum-rate capacity region corresponds to $\boldsymbol{R}^*$. In the graph on the left, the minimum rate vector is well within the zero-outage capacity region and, as a result, the minimum-rate capacity region extends significantly past the zero-outage capacity region. In the second graph, the minimum rate vector is close to the boundary of the zero-outage capacity region and, therefore, a large fraction of the power is used to simply achieve the minimum rates. Thus, there is little power left over to exceed the minimum rates and, as a result, the boundary of the minimum-rate capacity region does not extend much further out than the boundary of the zero-outage capacity region. Notice that in all cases the minimum-rate capacity region does not extend to the axes due to the minimum rate constraints.

Since the minimum rate boundary lies between the ergodic and zero-outage boundaries, the difference between the ergodic and zero-outage capacity regions is a good indicator of the degradation in capacity (i.e., the difference between $\mathcal{C}_{\text{ergodic}}(\overline{P})$ and $\mathcal{C}_{\min}(\overline{P}, \boldsymbol{R}^*)$) due to minimum rates. If the zero-outage capacity region is much smaller than the ergodic capacity region, the minimum-rate capacity region is generally much smaller than the ergodic capacity region. Alternatively, if the zero-outage capacity region is not much smaller than the ergodic capacity region, the minimum-rate capacity region is generally quite close to the ergodic capacity region.

V. Explicit Characterization of Minimum-Rate Capacity Region

In this section, we explicitly characterize the boundary of the minimum-rate capacity region of a K-user broadcast channel and find the corresponding optimal power-allocation scheme. Directly characterizing the minimum-rate capacity region appears to yield a rather nonintuitive solution, but we show that the minimum-rate capacity region can be written in terms of the *ergodic* capacity region of a related broadcast channel. This characterization is intuitively easy to understand and allows the minimum-rate capacity region to be calculated using only the ergodic capacity techniques of [3].

A. Derivation of Minimum-Rate Capacity Region

Due to the convexity of the minimum-rate capacity region, for any $\boldsymbol{R}^* \in \mathcal{C}_{\text{zero}}(\overline{P})$ and power constraint $\overline{P}$, the boundary of the region can be traced out by the following maximization:

$$\max_{\boldsymbol{R}} \boldsymbol{\mu} \cdot \boldsymbol{R} \quad \text{subject to} \quad \boldsymbol{R} \in \mathcal{C}_{\min}(\overline{P}, \boldsymbol{R}^*) \quad (9)$$

over all priority vectors $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_K)$ such that $\sum_{i=1}^{K}\mu_i=1$. By the definition of $\mathcal{C}_{\min}(\overline{P},\boldsymbol{R}^*)$, the following is an equivalent maximization:

$$\max_{\boldsymbol{P}(n)} \mathbb{E}_{\boldsymbol{n}}\left[\sum_{i=1}^{K}\mu_i R_i(\boldsymbol{P}(n))\right] \tag{10}$$
$$\text{subject to:}\quad \mathbb{E}_{\boldsymbol{n}}\left[\sum_{i=1}^{K}P_i(\boldsymbol{n})\right]\le\overline{P}$$
$$R_i(\boldsymbol{P}(n))\ge R_i^*,\forall i,\boldsymbol{n}$$

where $R_i(\boldsymbol{P}(n))$ is defined in (3).

For each fading state $\boldsymbol{n}$, let $\pi(.)$ be the permutation such that

$$n_{\pi(1)}\le n_{\pi(2)}\le\cdots\le n_{\pi(K)}.$$

Since successive decoding is performed at each receiver in which the weakest user (i.e., the user with the largest noise power), or User $\pi(K)$) is decoded first, $R_i(\boldsymbol{P}(n))$ can be defined as

$$R_{\pi(i)}(\boldsymbol{P}(n))=\log\left(1+\frac{P_{\pi(i)}(\boldsymbol{n})}{n_{\pi(i)}+\sum_{j<i}P_{\pi(j)}(\boldsymbol{n})}\right)$$
$$=C\left(\frac{P_{\pi(i)}(\boldsymbol{n})}{n_{\pi(i)}+\sum_{j<i}P_{\pi(j)}(\boldsymbol{n})}\right) \tag{11}$$

where $C(\cdot)$ is defined as $C(x)=\log(1+x)$.

In order for each user to achieve their respective minimum rates in each state, a minimum amount of power must be allocated to each user in each fading state. We use $P_i^*(\boldsymbol{n})$ to denote the minimum power that User i must be allocated in fading state $\boldsymbol{n}$ in order to exactly achieve R_i^*. We define the minimum powers such that if all users are allocated their minimum powers in a fading state, then all users will exactly achieve their respective minimum rates. From the definition of $R_i(\boldsymbol{P}(\boldsymbol{n}))$ in (11) it follows that the minimum power of each user is given by

$$P_{\pi(i)}^*(\boldsymbol{n})\triangleq\left(n_{\pi(i)}+\sum_{j<i}P_{\pi(j)}^*(\boldsymbol{n})\right)\left(e^{R_{\pi(i)}^*}-1\right). \tag{12}$$

We define $\hat{P}_i(\boldsymbol{n})$ as the power allocated to User i in excess of the minimum power. The total power allocated to each user in fading state $\boldsymbol{n}$ is thus $P_i(\boldsymbol{n})=P_i^*(\boldsymbol{n})+\hat{P}_i(\boldsymbol{n})$. The minimum rate constraints clearly imply $P_i(\boldsymbol{n})\ge P_i^*(\boldsymbol{n})$, which implies $\hat{P}_i(\boldsymbol{n})\ge 0$.

Since the rates are direct functions of the power allocation, we can replace the rate constraints in (10) with a power constraint to result in the following equivalent maximization:

$$\max_{\hat{\boldsymbol{P}}(n)} \mathbb{E}_{\boldsymbol{n}}\left[\sum_{i=1}^{K}\mu_i C\left(\frac{P_{\pi(i)}^*(\boldsymbol{n})+\hat{P}_{\pi(i)}(\boldsymbol{n})}{n_{\pi(i)}+\sum_{j<i}P_{\pi(j)}(\boldsymbol{n})}\right)\right] \tag{13}$$
$$\text{subject to :}\quad \mathbb{E}_{\boldsymbol{n}}\left[\sum_{i=1}^{K}\hat{P}_i(\boldsymbol{n})\right]\le P'$$
$$\hat{P}_{\pi(i)}(\boldsymbol{n})\ge(e^{R_i^*}-1)\sum_{j<i}\hat{P}_{\pi(j)}(\boldsymbol{n})\ \ \forall i,\boldsymbol{n}$$
$$\hat{P}_{\pi(i)}(\boldsymbol{n})\ge 0\ \ \forall i,\boldsymbol{n}$$

where $P'\triangleq\overline{P}-\mathbb{E}_{\boldsymbol{n}}\left[\sum_{i=1}^{K}P_i^*(\boldsymbol{n})\right]$ is the total excess power. Notice that the maximization is over the excess power allocation $\hat{\boldsymbol{P}}(n)$ only. The minimum rate constraints make this problem more difficult than maximizing ergodic capacity. However, with some algebraic manipulation we will see that the minimum-rate capacity maximization is equivalent to a related ergodic capacity maximization.

Using the rate-splitting identity (i.e., $C(\frac{a+b}{n})=C(\frac{a}{n})+C(\frac{b}{a+n})$), we can simplify the rate equation in (11). We have omitted the dependence on the fading state $\boldsymbol{n}$ for brevity

$$\begin{aligned}
&R_{\pi(i)}(\boldsymbol{P}(n))\\
&=C\left(\frac{P_{\pi(i)}^*+\hat{P}_{\pi(i)}}{n_{\pi(i)}+\sum_{j<i}P_{\pi(j)}}\right)\\
&=C\left(\frac{P_{\pi(i)}^*+(e^{R_{\pi(i)}^*}-1)\sum_{j<i}\hat{P}_{\pi(j)}}{n_{\pi(i)}+\sum_{j<i}P_{\pi(j)}}\right)\\
&\quad+C\left(\frac{\hat{P}_{\pi(i)}-(e^{R_{\pi(i)}^*}-1)\sum_{j<i}\hat{P}_{\pi(j)}}{n_{\pi(i)}+\sum_{j<i}P_{\pi(j)}+P_{\pi(i)}^*+(e^{R_{\pi(i)}^*}-1)\sum_{j<i}\hat{P}_{\pi(j)}}\right)\\
&=C\left(\frac{P_{\pi(i)}^*+(e^{R_{\pi(i)}^*}-1)\sum_{j<i}\hat{P}_{\pi(j)}}{n_{\pi(i)}+\sum_{j<i}P_{\pi(j)}^*+\sum_{j<i}\hat{P}_{\pi(j)}}\right)\\
&\quad+C\left(\frac{\hat{P}_{\pi(i)}-(e^{R_{\pi(i)}^*}-1)\sum_{j<i}\hat{P}_{\pi(j)}}{n_{\pi(i)}+\sum_{j\le i}P_{\pi(j)}^*+e^{R_{\pi(i)}^*}\sum_{j<i}\hat{P}_{\pi(j)}}\right)\\
&=R_{\pi(i)}^*+C\left(\frac{\hat{P}_{\pi(i)}-(e^{R_{\pi(i)}^*}-1)\sum_{j<i}\hat{P}_{\pi(j)}}{n_{\pi(i)}+\sum_{j\le i}P_{\pi(j)}^*+e^{R_{\pi(i)}^*}\sum_{j<i}\hat{P}_{\pi(j)}}\right)
\end{aligned}$$

where we have used the definition of $P_i^*(\boldsymbol{n})$ to obtain the final step. From this simplification it should be clear that power

$$P_{\pi(i)}^*+(e^{R_i^*}-1)\sum_{j<i}\hat{P}_{\pi(j)}$$

maintains the minimum rate of each user, while power

$$\hat{P}_{\pi(i)}-(e^{R_i^*}-1)\sum_{j<i}\hat{P}_{\pi(j)}$$

(which is nonnegative by the power constraint in (13)) increases the rate above the minimum rate. Let us introduce the following effective noise and power terms (for each joint fading state), denoted by $P_i^e(\boldsymbol{n})$ and n_i^e:

$$n_{\pi(i)}^e\triangleq\left(n_{\pi(i)}+\sum_{j\le i}P_{\pi(j)}^*(\boldsymbol{n})\right)e^{\sum_{j>i}R_{\pi(j)}^*} \tag{14}$$

$$P_{\pi(i)}^e(\boldsymbol{n})\triangleq\left(\hat{P}_{\pi(i)}(\boldsymbol{n})-(e^{R_{\pi(i)}^*}-1)\sum_{j<i}\hat{P}_{\pi(j)}(\boldsymbol{n})\right)\times e^{\sum_{j>i}R_{\pi(j)}^*}. \tag{15}$$

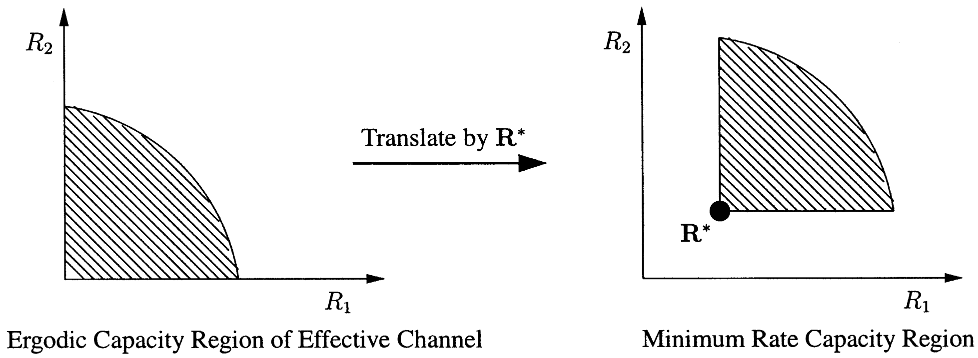

Fig. 3. Ergodic capacity region of effective channel and minimum-rate capacity region.

Substituting these terms into our previous expression, we get

$$R_{\pi(i)}(\boldsymbol{P}(n)) = R^*_{\pi(i)} + C\left(\frac{P^e_{\pi(i)}(\boldsymbol{n})}{n^e_{\pi(i)} + e^{\sum_{j\geq i} R^*_{\pi(j)}} \sum_{j<i} \hat{P}_{\pi(j)}(\boldsymbol{n})}\right).$$

In Appendix A we show that

$$\sum_{j<i} P^e_{\pi(j)}(\boldsymbol{n}) = e^{\sum_{j\geq i} R^*_{\pi(j)}} \sum_{j<i} \hat{P}_{\pi(j)}(\boldsymbol{n}).$$

Thus, we can finally rewrite the rate expression as

$$R_{\pi(i)}(\boldsymbol{P}(n)) = R^*_{\pi(i)} + C\left(\frac{P^e_{\pi(i)}(\boldsymbol{n})}{n^e_{\pi(i)} + \sum_{j<i} P^e_{\pi(j)}(\boldsymbol{n})}\right) \tag{16}$$

which is identical to the rate equations for ergodic capacity for a channel with noises $n^e_1, \ldots, n^e_K$. Since the rate of each user can be written explicitly in terms of effective power and effective noise, we can, in fact, maximize the weighted sum rate as a function of only the effective noises and effective powers. In Appendix C, we show that every set of excess powers satisfying the minimum rate constraints in (13) maps uniquely to a set of nonnegative effective powers, and *vice versa*. In Appendix D, we show that the the mapping from noise state to effective noise state is one-to-one for a fixed minimum rate vector and strictly unequal noise powers (which is true with probability 1 for a continuous fading distribution). Thus, we can write the effective power allocation as a function of the joint effective noise state instead of the joint noise state. Furthermore

$$\sum_{i=1}^{k} P^e_i(\boldsymbol{n}) = \sum_{i=1}^{K} \hat{P}_i(\boldsymbol{n})$$

by Appendix A. Therefore, the maximization in (13) is equivalent to

$$\sum_{i=1}^{K} \mu_i R^*_i + \max_{\boldsymbol{P}^e(n)} \mathbb{E}_{\boldsymbol{n}}\left[\sum_{i=1}^{K} \mu_i C\left(\frac{P^e_{\pi(i)}(\boldsymbol{n})}{n^e_{\pi(i)} + \sum_{j<i} P^e_{\pi(j)}(\boldsymbol{n})}\right)\right]$$

$$\text{subject to:} \quad \mathbb{E}_{\boldsymbol{n}}\left[\sum_{i=1}^{K} P^e_i(\boldsymbol{n})\right] \leq P',\ P^e_i(\boldsymbol{n}) \geq 0 \quad \forall i, \boldsymbol{n}. \tag{17}$$

In Appendix B, we show that the ordering of the effective noises is the same as the ordering of the actual noises, i.e., $n^e_{\pi(1)} \leq \cdots \leq n^e_{\pi(K)}$. Thus, the preceding maximization is identical to the problem of maximizing $\boldsymbol{\mu} \cdot \boldsymbol{R}$ in the ergodic capacity region of the channel with noises defined as in (14) and power P'. We refer to the channel with noises n^e_i and power P' as the *effective channel*. The joint distribution of $\boldsymbol{n}^e$ can be derived from the mapping in (14).

Without the constant term $\sum_{i=1}^{K} \mu_i R^*_i$, (17) is identical to the ergodic capacity maximization expression of the effective broadcast channel [3], [7]. Therefore, the average rates achievable *in excess* of the minimum rates are equal to the rates achievable in the effective channel, or to the ergodic capacity region of the effective channel. The minimum-rate capacity region is therefore equal to the ergodic capacity region of the effective channel plus the minimum rates[2]

$$\boxed{\mathcal{C}_{\min}(\overline{P}, \boldsymbol{R}^*) = \boldsymbol{R}^* + \mathcal{C}_{\text{ergodic}}(P'; n^e_1, \ldots, n^e_K)} \tag{18}$$

where $\mathcal{C}_{\text{ergodic}}(P'; n^e_1, \ldots, n^e_K)$ refers to the ergodic capacity of the effective channel. In Fig. 3, the ergodic capacity region of the effective channel and the minimum-rate capacity region are plotted as an example of this relationship.

B. Optimal Power Allocation Policies

The optimal power allocation scheme to achieve the boundary of the minimum-rate capacity region can be found by finding the optimal power allocation to achieve the boundary of the ergodic capacity region of the effective channel. The allocation of minimum power is predetermined by the minimum rate requirements and the noise powers, while the optimal allocation of excess power is related to the optimal power allocation to achieve the ergodic capacity region of the effective channel. More specifically, to find the optimal power allocation policy that maximizes $\boldsymbol{\mu} \cdot \boldsymbol{R}$ in $\mathcal{C}_{\min}(\overline{P}, \boldsymbol{R}^*)$ for some fixed priority vector $\boldsymbol{\mu}$, we define the optimal allocation of effective power (i.e., $P^e_i(\boldsymbol{n}^e)$) to be the optimal power allocation policy that maximizes $\boldsymbol{\mu} \cdot \boldsymbol{R}$ in $\mathcal{C}_{\text{ergodic}}(P'; n^e_1, \ldots, n^e_K)$ for the same priority vector $\boldsymbol{\mu}$. We can then transform the effective power allocation $P^e_i(\boldsymbol{n}^e)$ to the excess power allocation $\hat{P}_i(\boldsymbol{n})$ by the relationships given in (14) and (15). The minimum power allocation

[2]The sum here refers to the set found by adding $\boldsymbol{R}^*$ to every element in $\mathcal{C}_{\text{ergodic}}(P'; n^e_1, \ldots, n^e_K)$.

$P_i^*(\boldsymbol{n})$ is defined in (12), and the total power allocated to each user in every fading state is $P_i(\boldsymbol{n}) = P_i^*(\boldsymbol{n}) + \hat{P}_i(\boldsymbol{n})$.

The optimal power allocation scheme for ergodic capacity maximization is described in [3, Sec. III]. We briefly discuss the power allocation here, but we defer the reader to [3] for a more complete description. The optimal power allocation is a more complicated version of the single-user water-filling algorithm derived in [10]. In each fading state, power can be allocated to any of the K users, or none at all. The total amount of power allocated to each fading state can be described in the following compact form:

$$\sum_{i=1}^{K} P_i^e(\boldsymbol{n}) = \left[\max_{i=1,\ldots,K}\left(\frac{\mu_i}{\lambda} - n_i^e\right)\right]^+ \tag{19}$$

where $[x]^+ \triangleq \max(x, 0)$ and $\frac{1}{\lambda}$ is the water-filling level chosen such that the power constraint P' is met with equality. This is akin to water-filling to the "best" user in each fading state, where the notion of best user depends not only on the noise power but also on the user-by-user priorities. Notice, however, that this is only the allocation of *total* power to each fading state. The actual distribution of power between users in each fading state is rather involved and we defer the reader to [3] for more details. A greedy algorithm to find the optimal power allocation policy (over fading states and users) can also be found in [1], [3]

If the maximum sum rate of the minimum-rate capacity region is being found (i.e., $\mu_1 = \cdots = \mu_K$), then from results on ergodic capacity we know that it is optimal to only allocate effective power to the user with the smallest noise power. Thus, at most one user per fading state strictly exceeds his minimum rate requirement. However, for general priorities this is not true. Note that we are discussing only the allocation of effective power, which relates directly to the excess power. Of course, each user must be allocated the minimum power in every fading state, so all K users are active in every fading state.

Fig. 4 illustrates the optimal amount of effective power in a two-user system that is allocated to each fading state for a discrete, four-state fading distribution where $\mu_2 > \mu_1$. Note that the breakdown of power between the two users, which requires the iterative algorithm of [3], is not indicated in this figure. Water-level $\frac{\mu_1}{\lambda}$ is used for channels that are allocated excess power on n_1^e and $\frac{\mu_2}{\lambda}$ is used for channels allocated excess power on n_2^e. Water-filling is done on the effective noise level that corresponds to the largest power allocation in that state. In the first state, water-filling is done on n_2^e because although $n_2^e > n_1^e$, the higher water-level of n_2^e compensates for this difference. Because $\mu_2 > \mu_1$ in the figure, water-filling is done on n_1^e only when $n_1^e \ll n_2^e$, as in state 2. In states 3 and 4, water-filling is done on n_2^e.

C. Interpretation of Effective Channel

The effective channel encapsulates how power allocated to one user manifests itself into additional required power for other users due to the minimum rate requirements. Consider the power allocated to each user as consisting of two components: a part that achieves the minimum rate, and the part that leads to excess rate above the minimum rate. The minimum power $P_i^*(\boldsymbol{n})$ allocated to each user leads to the minimum rates of each user only if all other users are allocated exactly their minimum power levels. The minimum power does not take into account excess power allocated to users who are seen as interference. Every increment of power δP allocated to User $\pi(i)$ forces User $\pi(i+1)$ to allocate $\delta P \cdot (e^{R^*_{\pi(i+1)}} - 1)$ to maintain his minimum rate. User $\pi(i+2)$ must then compensate for power δP and power $\delta P \cdot (e^{R^*_{\pi(i+1)}} - 1)$. This forces User $\pi(i+2)$ to allocate $\delta P \cdot e^{R^*_{\pi(i+1)}}(e^{R^*_{\pi(i+2)}} - 1)$ to maintain his minimum rate. This process continues up to the weakest user. In total, every increment of power δP allocated to User $\pi(i)$ corresponds to a total allocation of power $\delta P \cdot e^{\sum_{j>i} R^*_{\pi(j)}}$ to Users $\pi(i), \ldots, \pi(K)$. Thus, allocation of excess power must capture two elements. First, excess power allocated to stronger users (i.e., $\sum_{j<i} \hat{P}_{\pi(j)}$) must be compensated for. The leftover excess power of User $\pi(i)$ after compensating for the excess power of stronger users thus is

$$\hat{P}_{\pi(i)}(\boldsymbol{n}) - (e^{R^*_{\pi(i)}} - 1)\sum_{j<i} \hat{P}_{\pi(j)}.$$

However, this leftover excess power must be multiplied by the factor $e^{\sum_{j>i} R^*_{\pi(j)}}$ to account for the fact that weaker users must compensate for any leftover excess power allocated to User $\pi(i)$. Therefore, the effective power of User i is

$$\left(\hat{P}_{\pi(i)}(\boldsymbol{n}) - (e^{R^*_{\pi(i)}} - 1)\sum_{j<i} \hat{P}_{\pi(j)}(\boldsymbol{n})\right) e^{\sum_{j>i} R^*_{\pi(j)}}.$$

It seems that the effective noise of each user should be equal to the actual noise plus the minimum power allocated to stronger users. However, the actual effective noise is multiplied by the factor $e^{\sum_{j>i} R^*_{\pi(j)}}$ to compensate for the fact that the effective power of User i is multiplied by the same factor.

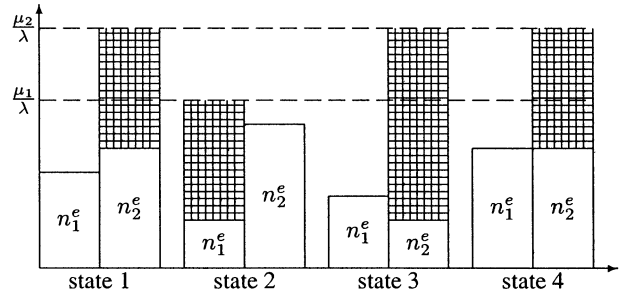

Fig. 4. Water-filling diagram for two-user channel with min rates.

D. Single-User Channel

A single-user channel can be viewed as the broadcast channel described in Section II with $K = 1$. Thus, the characterization of minimum-rate capacity derived in Section V-A can be applied to the single-user channel as well. Clearly, the minimum power for each state is defined as $P^*(n) \triangleq n(e^{R^*} - 1)$. As before, the minimum-rate capacity can be found by solving the ergodic capacity of the effective channel. From the expressions in (14) and (15), we see that $n^e = n + P^*(n)$ and $P^e(n) = \hat{P}(n)$. The power constraint of the effective channel is $P' = \overline{P} - \mathbb{E}_n[P^*(n)]$. Since water-filling over time achieves ergodic

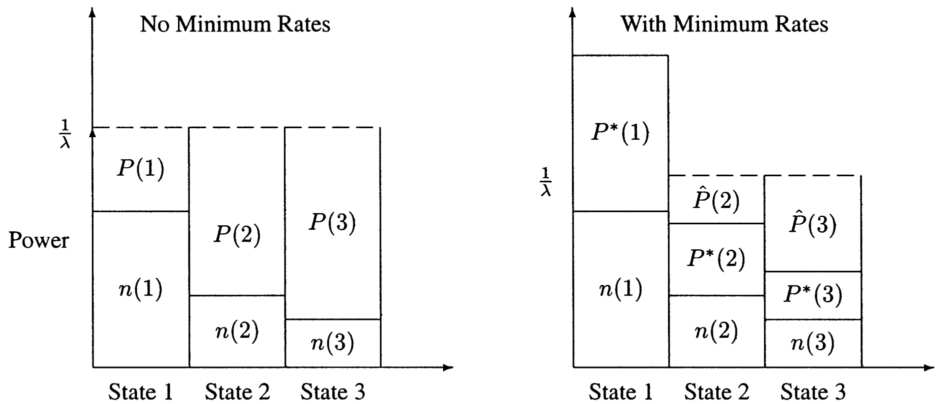

Fig. 5. Water-filling diagram for a single-user with zero and nonzero minimum rates.

capacity of a single-user fading channel [10], the optimal allocation of effective power is found by water-filling over the effective noise n^e

$$\hat{P}(n) = \begin{cases} \frac{1}{\lambda} - (n + P^*(n)), & n + P^*(n) \leq \frac{1}{\lambda} \\ 0, & n + P^*(n) > \frac{1}{\lambda} \end{cases}$$

where $\frac{1}{\lambda}$ is the water-filling level satisfying the excess power constraint P'.

This simple power allocation scheme yields a closed-form expression for the capacity of a single-user channel with power constraint $\overline{P}$ and minimum rate R^*

$$\begin{aligned}&\mathcal{C}_{\min}(\overline{P}, R^*)\\ &\quad = R^* \Pr\left[n \geq \frac{1}{\lambda}e^{-R^*}\right] + \int_0^{\frac{1}{\lambda}e^{-R^*}} \log\left(\frac{1}{\lambda n}\right) p(n)dn.\end{aligned}$$

In this expression we use the fact that $n + P^*(n) = ne^{R^*}$.

Fig. 5 illustrates the water-filling procedure for zero and nonzero minimum rates for a single-user three-state channel. State 1 is the weakest of the three channels. The graph on the left shows the power allocation scheme without minimum rates. We see that all three channels are allocated power, but the rate achieved in states 1 and 2 may be quite small. When minimum rates are applied, the minimum power P^* becomes an additional source of noise. Because $P^*(n)$ is an increasing function of n, the effective noise term of state 1 becomes much larger than the other two terms. When water-filling is done on the effective noise terms, additional power is only allocated to states 2 and 3 because the effective noise term of state 1 is too large and because much of the power was used to simply achieve the minimum rates in all three states. In state 1, transmission will be done at exactly R^*, whereas the minimum rate will be exceeded in the other two states due to the excess power allocated to those states.

As briefly mentioned earlier, in a single-user channel data can be transmitted at the minimum rate with a decoding delay that is independent of the rate of channel variation while simultaneously transmitting delay-insensitive data which takes advantage of the ergodic nature of the channel. This can be accomplished through the use of a separate minimum rate codebook and an ergodic rate codebook and the idea of rate splitting [11]. Notice that the rate in each fading state can be expanded as

$$\begin{aligned}R(n) &= C\left(\frac{P^*(n) + \hat{P}(n)}{n}\right)\\ &= C\left(\frac{\hat{P}(n)e^{-R^*}}{n}\right) + C\left(\frac{P^*(n) + \hat{P}(n)(1 - e^{-R^*})}{n + \hat{P}(n)e^{-R^*}}\right)\\ &= \hat{R}(n) + R^*(n)\end{aligned}$$

where the excess rate is

$$\hat{R}(n) = C\left(\frac{\hat{P}(n)e^{-R^*}}{n}\right) = C\left(\frac{\hat{P}(n)}{n + P^*(n)}\right).$$

A minimum rate codebook of size $2^{n_{\min}R^*}$ with block length $n_{\min}$ can be used to transmit data at the minimum rate, while a codebook of size $2^{n\mathbb{E}[\hat{R}(n)]}$ with block length n which is an integer multiple of $n_{\min}$ can be used to transmit data at the excess rate. Codewords from both codebooks are simultaneously sent. The minimum rate codeword is scaled by the quantity $P^*(n) + \hat{P}(1 - e^{-R^*})$, while the ergodic codeword is scaled by $\hat{P}(n)e^{-R^*}$. Treating the ergodic codeword as interference, it is easy to show that the received signal-to-interference-noise ratio (SINR) of the minimum rate codeword is exactly $e^{R^*} - 1$, as required to transmit at rate R^*. Thus, the minimum rate codeword can be successfully decoded while treating the ergodic codeword as interference. After decoding and subtracting out the minimum rate codewords, the ergodic codeword can be decoded at the end of the ergodic block length, since only the actual noise remains in the channel.

This two-codebook strategy cannot be used for the broadcast channel because the strongest user must decode both the ergodic and minimum rate codeword of every other weaker user before being able to decode his own minimum rate codeword. This eliminates the possibility of decoding the minimum rate codewords before the ergodic codewords.

VI. Alternative Constraints on Transmitted Power

We have derived the minimum-rate capacity region of a broadcast channel subject to an average power constraint. The optimal transmitted power is a function of the joint fading state and can be quite large in some fading states. In practical broadcast situations, there is generally a peak power constraint and there may or may not be an average power constraint. In this section, we characterize the minimum-rate capacity region of a K-user broadcast channel subject to two different constraint sets: a peak power constraint only, and both a peak and an average power constraint.

A. Peak Power Constraint

We now consider the problem of maximizing minimum-rate capacity subject to only a peak power constraint P_{peak} in each fading state. The capacity region can then be defined as the set of all achievable average rates subject to minimum rate, peak, and power constraints as it was for the average power constraint case in Section IV. We let $\mathcal{F}''$ denote the set of feasible power policies satisfying the peak power constraint and the minimum rate constraint in all fading states

$$\mathcal{F}'' \equiv \left\{ \mathcal{P} : \sum_{j=1}^{K} P_j(\boldsymbol{n}) \le P_{\text{peak}}, \ \ R_j(\boldsymbol{P}(n)) \ge R_j^* \ \ \forall j, \boldsymbol{n} \right\}.$$

The capacity region subject to peak power constraint P_{peak} then is

$$\mathcal{C}_{\min}^{\text{peak}}(P_{\text{peak}}, \boldsymbol{R}^*) = \text{Co}\left(\bigcup_{\mathcal{P} \in \mathcal{F}''} \mathcal{C}_{\min}(\mathcal{P}) \right). \tag{20}$$

To find the boundary of the capacity region, we perform a maximization similar to (10), except with a peak power constraint replacing the average power constraint.

Since the weighted sum of the rates is an increasing function of the total power allocated to each fading state, each fading state should be allocated the peak power. Clearly, the minimum rates must be achievable in each state under the peak power constraint P_{peak} which implies $P_{\text{peak}} \ge \sum_{i=1}^{K} P_i^*(\boldsymbol{n}), \ \forall \boldsymbol{n}$. Given that each fading state is allocated the peak power, the remaining task is to optimally allocate P_{peak} between the K users in each fading state. We may first allocate the minimum power required to achieve the minimum rates in each state, leaving excess power $P_{\text{peak}} - \sum_{i=1}^{K} P_i^*(\boldsymbol{n})$ in each fading state. The excess power must then be optimally distributed between the K users to maximize the weighted sum of their rates in excess of the minimum rates. The set of achievable excess rates is equal to the capacity region of the effective broadcast channel, which takes the form of a constant broadcast channel in each fading state. However, maximizing weighted sum rate for a constant channel turns out to be nearly as difficult as maximizing weighted sum rate for a fading channel. First, a different water-filling level $\frac{1}{\lambda(\boldsymbol{n})}$ must be chosen for *each* fading state to satisfy

$$P_{\text{peak}} - \sum_{i=1}^{K} P_i^*(\boldsymbol{n}) = \max_{i=1,\dots,K} \left(\frac{\mu_i}{\lambda(\boldsymbol{n})} - n_i^e \right).$$

The effective power $P_{\text{peak}} - \sum_{i=1}^{K} P_i^*(\boldsymbol{n})$ is then allocated to the K users in each fading state according to the procedure detailed in [3, Sec. III]. As before, the actual excess power allocation policy can be inferred from the allocation of effective power by the relationship in (14) and (15).

B. Peak and Average Power Constraint

In this subsection, we find the minimum-rate capacity subject to average power constraint $\overline{P}$ and peak power constraint P_{peak}. We assume $P_{\text{peak}} > \overline{P}$. If this condition is not satisfied, the average power constraint is meaningless. The capacity region can be defined as it was for the average power constraint case in Section IV. We let $\mathcal{F}'''$ denote the set of feasible power policies

$$\mathcal{F}''' \equiv \left\{ \mathcal{P} : \mathbb{E}_{\boldsymbol{n}}\left[\sum_{j=1}^{K} P_j(\boldsymbol{n}) \right] \le \overline{P}, \ \ \sum_{j=1}^{K} P_j(\boldsymbol{n}) \le P_{\text{peak}} \ \ \forall \boldsymbol{n}, \right.$$
$$\left. R_j(\boldsymbol{P}(n)) \ge R_j^* \ \ \forall j, \boldsymbol{n} \right\}.$$

The capacity region subject to peak power constraint P_{peak} can then be characterized as

$$\mathcal{C}_{\min}^{\text{peak+avg}}(\overline{P}, P_{\text{peak}}, \boldsymbol{R}^*) = \text{Co}\left(\bigcup_{\mathcal{P} \in \mathcal{F}'''} \mathcal{C}_{\min}(\mathcal{P}) \right). \tag{21}$$

To find the boundary of the K-user capacity region, we perform a maximization similar to (10) with the addition of a state-by-state peak power constraint. We can therefore allocate minimum power to both users and reduce the problem to an ergodic capacity maximization problem. As stated before, the minimum power required in each state to meet the minimum rate requirements must not violate the peak power constraint. However, we must maximize the ergodic capacity of the effective channel subject to an average power constraint

$$P' \triangleq \overline{P} - \mathbb{E}_{\boldsymbol{n}}\left[\sum_{i=1}^{K} P_i^*(\boldsymbol{n}) \right]$$

and peak power constraint $P_{\text{peak}} - \sum_{i=1}^{K} P_i^*(\boldsymbol{n})$ in each fading state. The optimal power allocation with both average and peak power constraints is simply a truncated version of the optimal power allocation policy with only an average power constraint. This is easiest to see by considering the greedy algorithm [1, Sec. 3.2], [3, Sec. III-A] to allocate power with only an average power constraint. In the greedy algorithm, each user is represented via a utility function which is a function of the amount of power allocated in each fading state. The peak power constraint effectively truncates the utility functions of all users at $z = P_{\text{peak}} - \sum_{i=1}^{K} P_i^*(\boldsymbol{n})$ in each fading state. Then it is easy to show that the total effective power allocated to each fading state is given by

$$\sum_{i=1}^{K} P_i^e(\boldsymbol{n}) = \min\left(P_{\text{peak}} - \sum_{i=1}^{K} P_i^*(\boldsymbol{n}), \left[\max_{i=1,\dots,K} \left(\frac{\mu_i}{\lambda} - n_i^e \right) \right]^+ \right).$$

The only difference between this scheme and the optimal excess power allocation scheme without the peak power constraint is that the excess power allocated to a state is truncated at $P_{\text{peak}} - \sum_{i=1}^{K} P_i^*(\boldsymbol{n})$, which in turn affects the optimal water-filling level $\frac{1}{\lambda}$. The distribution of excess power to the K users within each fading state follows the procedure detailed in [3, Sec. III], with the simple caveat that the total effective power allocated to each fading state cannot be larger than $P_{\text{peak}} - \sum_{i=1}^{K} P_i^*(\boldsymbol{n})$.

VII. Minimum-Rate Outage Capacity

In this section, we discuss minimum-rate capacity with outage subject to an average power constraint. In minimum-rate capacity, minimum rates must be maintained in all fading states. With outage, however, this constraint is loosened slightly and the minimum rate of every user must only be met subject to outage probabilities $\boldsymbol{P}^{\text{out}} = (P_1^{\text{out}}, \ldots, P_K^{\text{out}})$. In other words, ergodic capacity is maximized subject to the constraint that the minimum rate of user k must be met with at least probability $(1 - P_k^{\text{out}})$ for $k = 1, \ldots, K$. Minimum rate outage allows minimum rate transmission to be suspended to users when their channels are very poor. Transmission is allowed during outages, but minimum rates are not required to be met during these times. In more practical terms, delay-sensitive data must be transmitted at the minimum rates a certain percentage of the time, whereas delay-insensitive data has no such constraint. This is different than the definition of outage capacity [4], [5] in which no data is transmitted during outages and the only concern is the constant channel achievable during nonoutages.

In certain severe fading distributions (i.e., Rayleigh fading), it is not possible to maintain a constant data rate at all times with an average power constraint. In other words, channels with certain severe fading distributions have no zero-outage capacity region. These channels therefore have no minimum-rate capacity region. However, all fading channels can support a constant rate with outage. Therefore, all fading channels do have a minimum-rate outage capacity region.

In this section, we analyze the scenario where outage is declared on a user-by-user basis as opposed to declaring a common outage during which no user is required to meet his minimum rate [4]. We will see that the case of common outage is a special case of the more general independent outage formulation.

A. Characterization of Minimum-Rate Outage Capacity Region With Independent Outage

To find the minimum rate outage capacity, we first define the outage function $\boldsymbol{w}(n) = (w_1(\boldsymbol{n}), \ldots, w_K(\boldsymbol{n}))$ over all fading states where $w_k(\boldsymbol{n}) = 1$ for fading states in which the minimum rate of User k must be satisfied and 0 otherwise.[3] Due to the outage constraints, the outage function must satisfy $\mathbb{E}_{\boldsymbol{n}}[w_k(\boldsymbol{n})] \geq (1 - P_k^{\text{out}})$ for each user. The outage function is an indicator function which determines which states are required to maintain the minimum rates of the different users. Maximizing ergodic capacity given outage function $\boldsymbol{w}(n)$ is very similar to finding non-outage minimum-rate capacity, except with *time-varying* minimum rates $\boldsymbol{R}^*(\boldsymbol{n}, w(\boldsymbol{n}))$. We define $\boldsymbol{R}^*(\boldsymbol{n}, w_k(\boldsymbol{n}))$ as

$$R_k^*(\boldsymbol{n}, w_k(\boldsymbol{n})) = \begin{cases} 0, & w_k(\boldsymbol{n}) = 0 \\ R_k^*, & w_k(\boldsymbol{n}) = 1 \end{cases} \tag{22}$$

where R_k^* is assumed to be the actual desired minimum rate of User k. We then write the time-varying minimum rates as

$$\boldsymbol{R}^*(\boldsymbol{n}, \boldsymbol{w}(n)) = (R_1^*(\boldsymbol{n}, w_1(\boldsymbol{n})), \ldots, R_K^*(\boldsymbol{n}, w_K(\boldsymbol{n}))).$$

[3] We need not consider $0 < w_k(\boldsymbol{n}) < 1$ since we are only concerned with continuous fading distributions.

Though the minimum rates were assumed to be constant in the original minimum-rate capacity formulation, time-varying minimum rates can be handled using almost the identical solution. To achieve the minimum-rate capacity with time-varying minimum rates $\boldsymbol{R}^*(\boldsymbol{n}, \boldsymbol{w}(n))$, we simply need to replace $\boldsymbol{R}^*$ with $\boldsymbol{R}^*(\boldsymbol{n}, \boldsymbol{w}(n))$ in the optimal power allocation scheme derived in Section V. The fact that the fading broadcast channel was decomposed into a parallel set of constant broadcast channels, one for each fading state, allows us to optimally deal with time-varying minimum rates using this simple substitution.

With this in mind, we define $\mathcal{C}_{\min}(\overline{P}, \boldsymbol{R}^*(\boldsymbol{n}, \boldsymbol{w}(n)))$ to be the minimum-rate capacity of the broadcast channel with time-varying minimum rates $\boldsymbol{R}^*(\boldsymbol{n}, \boldsymbol{w}(n))$. For each outage function $\boldsymbol{w}(n)$ satisfying the outage constraints, $\mathcal{C}_{\min}(\overline{P}, \boldsymbol{R}^*(\boldsymbol{n}, \boldsymbol{w}(n))$ defines an achievable rate region that satisfies both the average power constraint and the outage constraints.

Definition 2: The minimum-rate outage capacity of a fading broadcast channel with perfect CSI at the transmitter and receivers, average power constraint $\overline{P}$, minimum rate constraint

$$\boldsymbol{R}^* = (R_1^*, R_2^*, \ldots, R_K^*) \in \mathcal{C}_{\text{out}}(\overline{P}, \boldsymbol{P}^{\text{out}})$$

and outage probabilities $\boldsymbol{P}^{\text{out}} = (P_1^{\text{out}}, \ldots, P_K^{\text{out}})$ is

$$\mathcal{C}_{\min}^{\text{outage}}(\overline{P}, \boldsymbol{R}^*, \boldsymbol{P}^{\text{out}}) = \text{Co}\left(\bigcup_{\boldsymbol{w}(n)} \mathcal{C}_{\min}(\overline{P}, \boldsymbol{R}^*(\boldsymbol{n}, \boldsymbol{w}(n)))\right)$$

where the union is over all $\boldsymbol{w}(n)$ satisfying

$$\mathbb{E}_{\boldsymbol{n}}[w_k(\boldsymbol{n})] \geq (1 - P_k^{\text{out}}) \; \forall k = 1, \ldots, K.$$

Notice that the minimum-rate vector $\boldsymbol{R}^*$ must be in the independent outage capacity region [4], i.e., $\boldsymbol{R}^* \in \mathcal{C}_{\text{out}}(\overline{P}, \boldsymbol{P}^{\text{out}})$, for the minimum rates to be achievable with the given outage probability.

B. Characterization of Minimum-Rate Outage Capacity Region With Common Outage

The minimum-rate outage capacity with common outage can be characterized using the expression for minimum-rate outage capacity with independent outage. With common outage, the outage function $\boldsymbol{w}(n)$ must satisfy the additional constraint $w_1(\boldsymbol{n}) = w_2(\boldsymbol{n}) = \cdots = w_K(\boldsymbol{n}) \; \forall \boldsymbol{n}$. In addition, the vector outage constraint becomes a scalar outage probability P^{out}. The capacity region then is

$$\mathcal{C}_{\min}^{\text{outage}}(\overline{P}, \boldsymbol{R}^*, P^{\text{out}}) = \text{Co}\left(\bigcup_{\boldsymbol{w}(n)} \mathcal{C}_{\min}(\overline{P}, \boldsymbol{R}^*(\boldsymbol{n}, \boldsymbol{w}(n)))\right)$$

where the union is over all $\boldsymbol{w}(n)$ satisfying

$$\mathbb{E}_{\boldsymbol{n}}[w_1(\boldsymbol{n})] \geq (1 - P^{\text{out}}).$$

Notice that the minimum rate vector $\boldsymbol{R}^*$ must be in the common outage capacity region [4], i.e., $\boldsymbol{R}^* \in \mathcal{C}_{\text{out}}(\overline{P}, P^{\text{out}})$, for the minimum rates to be achievable with common outage and with the given outage probability.

C. Characterization of Minimum-Rate Outage Capacity for a Single-User Channel

The definition of minimum-rate outage capacity given in Theorem 2 applies to single-user channels as well, but the expression can be simplified significantly in the single-user case. For a single-user channel, the outage function $w(n)$ is only a function of the fading state because there is only one user and the capacity region is one-dimensional. Finding the largest achievable rate subject to the power and outage constraint therefore is equivalent to finding the outage function that corresponds to the largest achievable rate. In [8], the concept of minimum-rate capacity with outage was independently proposed and the optimal outage function $w^*(n)$ was found to be

$$w^*(n) = \begin{cases} 1, & n < s^* \\ 0, & n \geq s^* \end{cases} \tag{23}$$

where the threshold s^* is chosen to satisfy

$$\mathbb{E}_n[w^*(n)] = \Pr\{n < s^*\} = 1 - P^{\text{out}}.$$

The optimal scheme is therefore seen to be a threshold policy: minimum rates must be maintained in all states better (i.e., smaller noise values) than the threshold, while minimum rates need not be maintained in states worse than the threshold. This is very similar to the solution to the minimum outage probability problem under a long-term average power constraint for a single-user channel solved in [12]. When maximizing outage capacity, all power available goes toward maintaining a constant rate in non-outage states. In minimum-rate outage capacity, however, some fraction of the power maintains the minimum rate in non-outage states. The excess power, however, is water-filled over the fading states with respect to the effective channel to maximize rates achieved in excess of the minimum rates.

Unfortunately, the multiuser broadcast channel does not appear to have such a simple solution for either common outage or independent outage because the relationship between the minimum power allocation, effective noise terms, and the effectiveness of each fading state and user is much more complicated than the single-user case.

VIII. Numerical Results

In this section, we present numerical results on the capacity of a two-user broadcast channel with minimum-rate constraints with an average power constraint and no outage. In all plots, the total transmitted power is 10 mW, the bandwidth is 100 kHz, and the noise distribution is symmetric. Furthermore, the minimum rates are symmetric in Figs. 6–9.

In Fig. 6, the capacity region of a two-user channel with very different noise levels is plotted. In one fading state, n_1 is 40 dB less than n_2 (i.e., the signal-to-noise ratio (SNR) of user 1 would be 40 dB larger than the SNR of user 2 assuming each user was allocated the same power), and *vice versa* in the second fading state. Without minimum rates, capacity is achieved by allocating almost all power to the better of the two users in each channel state. This causes the capacity region to be highly convex. When minimum-rate constraints are applied, however, power must also be allocated to the weaker user in every fading state to satisfy the minimum rates, leading to a large capacity reduction. It is clear from Fig. 6 that the minimum-rate capacity region is significantly smaller than the ergodic capacity region, especially for large minimum rates.

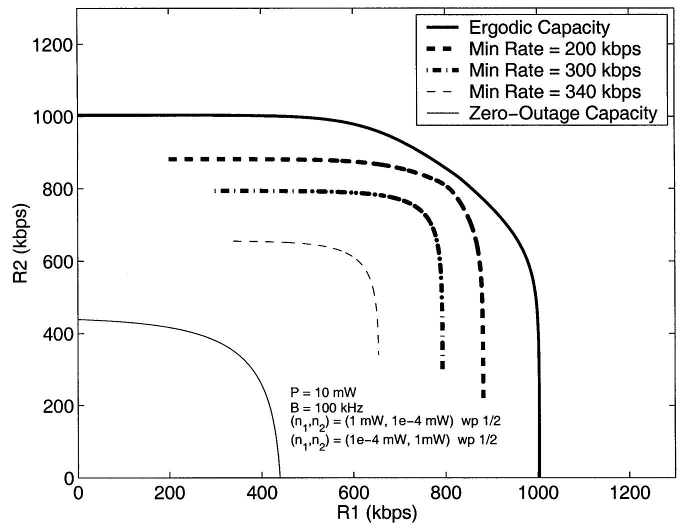

Fig. 6. Capacity of symmetric channel with 40-dB difference in SNR.

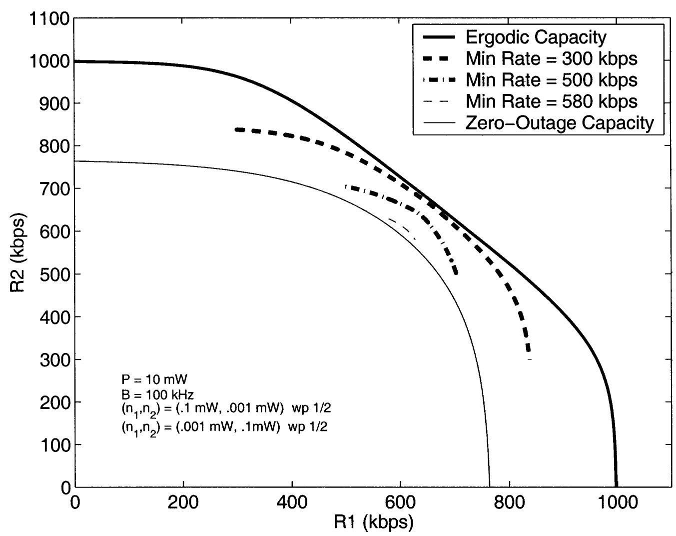

Fig. 7. Capacity of symmetric channel with 20-dB difference in SNR.

Because the minimum-rate boundary always lies between the zero-outage capacity region boundary and the ergodic capacity region boundary, the zero-outage capacity region is generally a good approximation for the minimum-rate capacity region. Channels in which the ergodic capacity region is much larger than the zero-outage capacity region will be significantly affected by minimum rate requirements, and *vice versa* for channels with zero-outage capacity regions that are not much smaller than the ergodic capacity region.

The zero-outage capacity region in Fig. 6 is significantly smaller than the ergodic capacity region. As expected, the minimum-rate capacity region is significantly smaller than the ergodic capacity region. We will see a similar relationship between the zero-outage and minimum-rate capacity regions for the other channel models.

The capacity region of a channel where n_1 and n_2 differ by 20 dB in each fading state is plotted in Fig. 7. The ergodic capacity region is much less convex than in Fig. 6 because the channels of the two users are more similar in each state. This is because

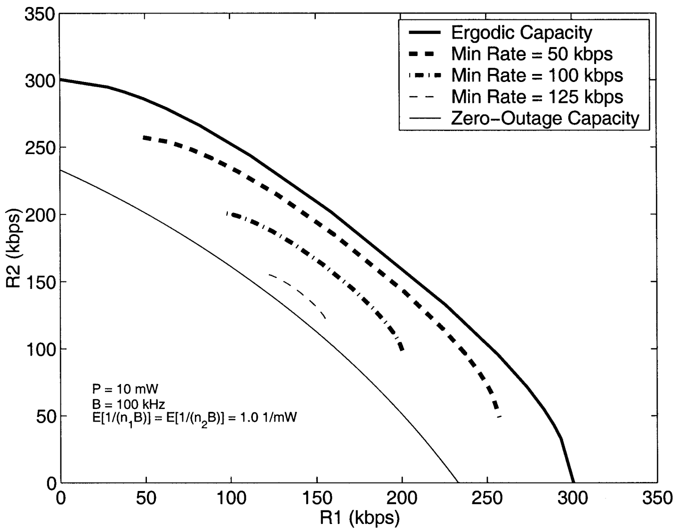

Fig. 8. Rician fading with $K = 1$, Average SNR =10 dB.

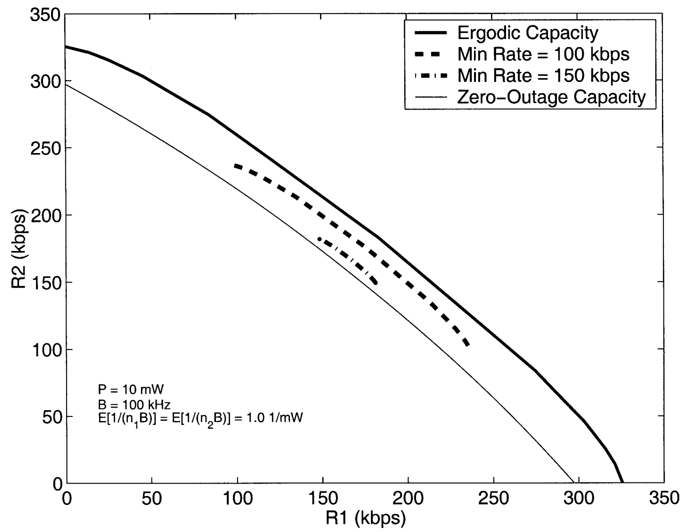

Fig. 9. Rician fading with $K = 5$, Average SNR = 10 dB.

the optimal power allocation scheme is not so heavily weighted toward the better user in each state so even the poorer user is allocated significant power in each state. Minimum-rate constraints force allocation of additional power to the poorer user in each fading state, but this is not as suboptimal as it is for the first example. We see in Fig. 7 that the minimum-rate capacity region is smaller than the ergodic capacity region, but not by as much as in Fig. 6. This result could have been predicted from the fact that the zero-outage capacity region of this channel is not much smaller than the ergodic capacity region due to the similarity of the users' channels.

In the subsequent two plots, results for more realistic channel models are presented. Independent fading is assumed for both receivers and the channel gain is incorporated into the noise power, as described in Section II. Rician fading with $K = 1$ is modeled in Fig. 8. This is not as severe as Rayleigh fading (which has no zero-outage capacity region), but the power of the multipath component is equal to the power of the line-of-sight component. The noise levels take on a wide range of values, as they do in the channel plotted in Fig. 6. As expected by our earlier results, minimum rates reduce capacity significantly. Once again we see that the zero-outage capacity region is much smaller than the ergodic capacity region.

In Fig. 9, Rician fading with $K = 5$ is modeled. Because the power of the line-of-sight component is five times as strong as the multipath component, both users generally have strong channels and this channel resembles the channel plotted in Fig. 7. As expected, minimum rates do not reduce capacity significantly.

Finally, in Fig. 10, the capacity regions of a Rician fading channel with $K = 1$ and asymmetric minimum rates are plotted. In the graph the capacity regions for minimum rates of (100 kb/s, 100 kb/s), (100 kb/s, 50 kb/s), and (100 kb/s, 0 kb/s) are shown. This relates to a scenario where one user has stricter requirements than the other or only one of the two users requires a minimum rate. We see that the capacity region for the asymmetric minimum-rate pair is considerably larger than the capacity region for the symmetric-rate pair. Notice that reducing the minimum rate of user 2 increases the capacity of both users, not just user 2, because reducing R_2^* frees up power that can be allocated to either user.

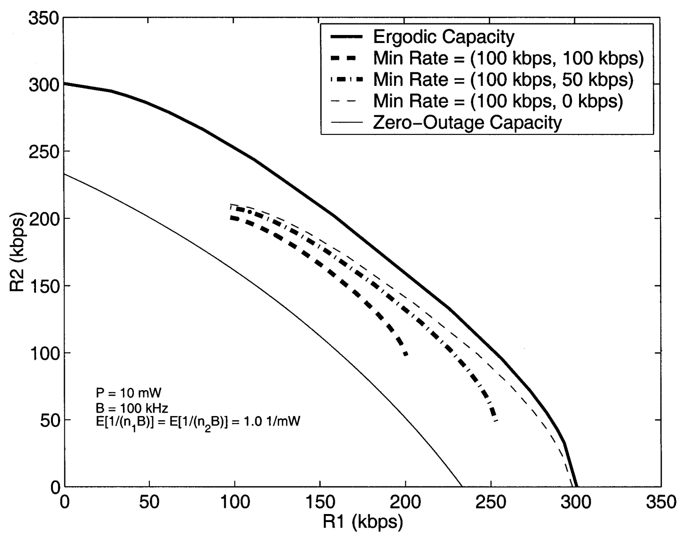

Fig. 10. Rician fading with $K = 1$, Average SNR = 10 dB, asymmetric minimum rates.

From these results, it is clear that minimum rates decrease the capacity regions of fading channels in which the noise levels of the users differ significantly in many channel states, i.e., when either of the two users has a significantly larger channel gain than the other user in many channel states. When the channels of the users do not differ significantly, minimum rates do not reduce the capacity region significantly.

IX. Conclusion

We defined the minimum-rate capacity region as the set of all achievable average rates subject to minimum rate requirements for each user in every fading state. By decomposing the power allocated to each user in every fading state into a portion which achieves the minimum rate and a portion which exceeds the minimum rate, we were able to specify the minimum-rate capacity region in terms of the ergodic capacity region of an effective broadcast channel. The effective channel incorporates the effect of the minimum rate requirements into the joint fading state and into the amount of total power available.

By analyzing several different channel models, we determined that for severely fading channels, the minimum-rate capacity region is significantly smaller than the ergodic capacity region. On the other hand, benign fading environments are able to support large minimum rates with little reduction in the capacity region. Furthermore, we saw that the difference between the zero-outage capacity region and the ergodic capacity region approximated the difference between the minimum-rate capacity region and the ergodic capacity region.

Additionally, it can be shown that a duality exists between the minimum-rate capacity region of the Gaussian broadcast and multiple-access channels [13]. Using this duality, [13] uses the results found in this paper to find the minimum-rate capacity region for the Gaussian multiple-access channel as well.

Appendix A
Proof of Excess and Effective Power Relationship

In this appendix, we prove the following result:

$$\sum_{j<i} P^e_{\pi(j)} = e^{\sum_{j\geq i} R^*_{\pi(j)}} \sum_{j<i} \hat{P}_{\pi(j)} \tag{24}$$

for $i = 2, \ldots, K+1$.. First, notice that for $i = 2$, we have $P^e_{\pi(1)} = \hat{P}_{\pi(1)} e^{\sum_{i=2}^K R^*_i}$ by the definition of $P^e_{\pi(i)}$ in (15). Assume (24) holds for i. We will show it holds for $i+1$ as well.

$$\begin{aligned}\sum_{j\leq i} P^e_{\pi(j)} &= P^e_{\pi(i)} + \sum_{j<i} P^e_{\pi(j)} \\ &= e^{\sum_{j>i} R^*_{\pi(j)}} \Bigg(\hat{P}_{\pi(i)}(\boldsymbol{n}) - \left(e^{R^*_{\pi(i)}} - 1\right) \\ &\qquad \sum_{j<i} \hat{P}_{\pi(j)}(\boldsymbol{n}) \Bigg) \\ &\quad + e^{\sum_{j\geq i} R^*_{\pi(j)}} \sum_{j<i} \hat{P}_{\pi(j)} \\ &= e^{\sum_{j>i} R^*_{\pi(j)}} \Bigg(\hat{P}_{\pi(i)}(\boldsymbol{n}) - \left(e^{R^*_{\pi(i)}} - 1\right) \\ &\qquad \sum_{j<i} \hat{P}_{\pi(j)}(\boldsymbol{n}) + e^{R^*_{\pi(i)}} \sum_{j<i} \hat{P}_{\pi(j)} \Bigg) \\ &= e^{\sum_{j>i} R^*_{\pi(j)}} \sum_{j\leq i} \hat{P}_{\pi(j)} \\ &= e^{\sum_{j\geq i+1} R^*_{\pi(j)}} \sum_{j<i+1} \hat{P}_{\pi(j)}.\end{aligned}$$

Notice that for $i = K+1$, this implies

$$\sum_{j=1}^{K} P^e_{\pi(j)} = \sum_{j=1}^{K} \hat{P}_{\pi(j)}$$

or that the sum of effective powers equals the sum of excess powers.

Appendix B
Proof of Effective Noise Ordering Equivalence

In this appendix, we prove that the effective noise terms $n^e_1, \ldots, n^e_K$ have the same ordering as the original noises $n_1, \ldots, n_K$. Since $n_{\pi(1)} \leq n_{\pi(2)} \leq \cdots \leq n_{\pi(K)}$ by the definition of $\pi(\cdot)$, we wish to show that

$$n^e_{\pi(1)} \leq n^e_{\pi(2)} \leq \cdots \leq n^e_{\pi(K)}$$

or that $n^e_{\pi(i)} \leq n^e_{\pi(i+1)}$. We can expand $n^e_{\pi(i+1)}$ as

$$\begin{aligned} n^e_{\pi(i+1)} &= \left(\sum_{j\leq i+1} P^*_{\pi(j)} + n_{\pi(i+1)} \right) e^{\sum_{j>i+1} R^*_{\pi(j)}} \\ &= \left(\sum_{j\leq i} P^*_{\pi(j)} + n_{\pi(i+1)} + P^*_{\pi(i+1)} \right) e^{\sum_{j>i+1} R^*_{\pi(j)}} \\ &= \left(\sum_{j\leq i} P^*_{\pi(j)} + n_{\pi(i+1)} \right) e^{R^*_{\pi(i+1)}} e^{\sum_{j>i+1} R^*_{\pi(j)}} \\ &= \left(\sum_{j\leq i} P^*_{\pi(j)} + n_{\pi(i+1)} \right) e^{\sum_{j>i} R^*_{\pi(j)}} \\ &= n^e_{\pi(i)} + \left(n_{\pi(i+1)} - n_{\pi(i)}\right) e^{\sum_{j>i} R^*_{\pi(j)}}. \end{aligned}$$

Since $n_{\pi(i+1)} \geq n_{\pi(i)}$ by our choice of $\pi(\cdot)$, we have $n^e_{\pi(i+1)} \geq n^e_{\pi(i)}$.

Appendix C
Proof That Excess Power to Effective Power Transformation Is One-to-One

In this appendix, we show that every set of nonnegative *effective* powers $(P^e_1(\boldsymbol{n}), \ldots, P^e_K(\boldsymbol{n}))$ corresponds (uniquely) to a valid (i.e., powers that meet or exceed all minimum-rate constraints) set of excess powers $(\hat{P}_1(\boldsymbol{n}), \ldots, \hat{P}_K(\boldsymbol{n}))$, and *vice versa*. This property is required so that the maximization over nonnegative effective powers in (17) is equivalent to the original maximization over excess powers in (13).

First, by the definition of effective power given in (15) it is easy to see that any set of excess powers $(\hat{P}_1(\boldsymbol{n}), \ldots, \hat{P}_K(\boldsymbol{n}))$ that meet the constraints of (13) map to nonnegative effective powers. Note also that the transformation preserves sum power in each fading state (Appendix A), and thus preserves average power as well.

To show equivalence in the other direction, first note that the effective power transformation in (15) can be written in matrix form as $\boldsymbol{P}^e(\boldsymbol{n}) = \boldsymbol{A}\hat{\boldsymbol{P}}(\boldsymbol{n})$ where $\boldsymbol{A}$ is a $K \times K$ matrix with

$$\boldsymbol{A}(i,i) = e^{\sum_{j>i} R^*_{\pi(j)}} > 0$$

and

$$\boldsymbol{A}(i,j) = -e^{\sum_{j>i} R^*_{\pi(j)}} (e^{R^*_{\pi(i)}} - 1) \leq 0$$

for all $j < i$. Thus, $\boldsymbol{A}$ is lower-triangular with strictly positive diagonal entries (which ensures invertibility) and nonnegative entries below the diagonal. It is straightforward to show that the inverse of such a matrix is lower-triangular with all nonnegative entries. Thus, by using $\boldsymbol{A}^{-1}$, we can map (uniquely) from nonnegative effective powers to nonnegative excess powers. Furthermore, since the powers satisfy (15) by definition, for all i and $\boldsymbol{n}$ we have

$$\hat{P}_{\pi(i)}(\boldsymbol{n}) - (e^{R^*_{\pi(i)}} - 1)\sum_{j<i}\hat{P}_{\pi(j)}(\boldsymbol{n}) = \frac{P^e_{\pi(i)}(\boldsymbol{n})}{e^{\sum_{j>i} R^*_{\pi(j)}}} \geq 0$$

since $\boldsymbol{P}^e(\boldsymbol{n}) \geq 0$ by assumption. Also, as noted earlier, the sum of the effective powers equals the sum of the excess powers in each fading state. Thus, the excess powers corresponding to any nonnegative set of effective powers satisfy all constraints in the original rate maximization in (13).

APPENDIX D
PROOF THAT NOISE TO EFFECTIVE NOISE TRANSFORMATION IS ONE-TO-ONE

Here, we show that the transformation from noise state $\boldsymbol{n}$ to effective noise state $\boldsymbol{n}^e$ is a one-to-one transformation by showing that the map from $\boldsymbol{n}$ to $\boldsymbol{n}^e$ is an invertible linear transformation from $\Re^K$ to $\Re^K$. The effective noise is defined in (14) as

$$n^e_{\pi(i)} \triangleq \left(n_{\pi(i)} + \sum_{j\leq i} P^*_{\pi(j)}(\boldsymbol{n})\right) e^{\sum_{j>i} R^*_{\pi(j)}}.$$

One can inductively show that

$$\sum_{j\leq i} P^*_{\pi(j)}(\boldsymbol{n}) = \sum_{k=1}^{i} n_{\pi(k)}(e^{R^*_{\pi(k)}} - 1)e^{\sum_{l=k+1}^{i} R^*_{\pi(l)}}. \tag{25}$$

Substituting this expression into the definition of effective noise, we get

$$n^e_{\pi(i)} = n_{\pi(i)} e^{\sum_{j\geq i} R^*_{\pi(j)}} + \sum_{k=1}^{i-1} n_{\pi(k)}(e^{R^*_{\pi(k)}} - 1)e^{\sum_{j>k} R^*_{\pi(j)}}. \tag{26}$$

In matrix terms, we can write the effective noise as $\boldsymbol{n}^e = \boldsymbol{B}\boldsymbol{n}$ where $\boldsymbol{B}$ is a lower-triangular $K \times K$ matrix defined by the coefficients given in (26). Notice that

$$\boldsymbol{B}(i,i) = e^{\sum_{j\geq i} R^*_{\pi(j)}} > 0$$

which implies that the matrix $\boldsymbol{B}$ is invertible and thus the transformation is one-to-one.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their suggestions and insights, which greatly improved the manuscript.

REFERENCES

[1] D. N. C. Tse, "Optimal power allocation over parallel Gaussian broadcast channels," *IEEE Trans. Inform. Theory*, submitted for publication.

[2] D. N. C. Tse and S. Hanly, "Multiaccess fading channels-Part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2796–2815, Nov. 1998.

[3] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels-Part I: Ergodic capacity," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1083–1102, Mar. 2001.

[4] ——, "Capacity and optimal resource allocation for fading broadcast channels-Part II: Outage capacity," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1103–1127, Mar. 2001.

[5] L. Li, N. Jindal, and A. Goldsmith, "Outage capacities and optimal power allocation for fading multiple-access channels," *IEEE Trans. Inform. Theory*, submitted for publication.

[6] S. Hanly and D. N. C. Tse, "Multiaccess fading channels-Part II: Delay-limited capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2816–2831, Nov. 1998.

[7] D. Hughes-Hartog, "The capacity of the degraded spectral Gaussian broadcast channel," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1975.

[8] J. Luo, L. Lin, R. Yates, and P. Spasojević, "Service outage based power and rate allocation," *IEEE Trans. Inform. Theory*, vol. 49, pp. 323–330, Jan. 2003.

[9] J. Luo, R. Yates, and P. Spasojević, "Service outage based power and rate allocation for parallel fading channels," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, June/July 2002, p. 108.

[10] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.

[11] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 42, pp. 364–375, Mar. 1996.

[12] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1468–1489, July 1999.

[13] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," *IEEE Trans. Inform. Theory*, to be published.