Capacity and Optimal Resource Allocation for Fading Broadcast Channels—Part II: Outage Capacity

Lifang Li, *Member, IEEE*, and Andrea J. Goldsmith, *Senior Member, IEEE*

***Abstract*—In this two-part paper, we study three capacity regions for fading broadcast channels and obtain their corresponding optimal resource allocation strategies: the ergodic (Shannon) capacity region, the zero-outage capacity region, and the capacity region with outage. In Part I, we obtained the ergodic capacity region of the fading broadcast channel under different spectrum-sharing techniques. In this paper, we derive the outage capacity regions of fading broadcast channels, assuming that both the transmitter and the receivers have perfect channel side information. These capacity regions and the associate optimal resource allocation policies are obtained for code division (CD) with and without successive decoding, for time division (TD), and for frequency division (FD). We show that in an M-user broadcast system, the outage capacity region is implicitly obtained by deriving the outage probability region for a given rate vector. Given the required rate of each user, we find a strategy which bounds the outage probability region for different spectrum-sharing techniques. The corresponding optimal power allocation scheme is a multiuser generalization of the threshold-decision rule for a single-user fading channel. Also discussed is a simpler minimum common outage probability problem under the assumption that the broadcast channel is either not used at all when fading is severe or used simultaneously for all users. Numerical results for the different outage capacity regions are obtained for the Nakagami-m fading model.**

***Index Terms*—Broadcast channels, capacity region, fading channels, optimal resource allocation, outage probability.**

I. Introduction

In mobile wireless communications, the channel characteristics vary with time. By applying optimal dynamic power and rate allocation strategies, the ergodic (Shannon) capacities with channel side information (CSI) at both the transmitter and the receiver of a single-user fading channel, a fading multiple access channel (MAC), and a fading broadcast channel under different spectrum-sharing techniques are obtained in [1], [2], and Part I of this paper,[1] respectively. This kind of capacity is a measurement of the long-term achievable rate averaged over the time-varying channel. For real-time applications that cannot tolerate the variable delays exhibited by the coding strategy that achieves the ergodic capacity, we have to consider the information rate that can be maintained in all fading conditions through optimal power control. In order to maintain a constant rate during severe fading, much power is needed. Therefore, given an average power constraint, the channel fading may be so severe that no constant rate greater than zero is possible. For example, the maximum instantaneous mutual information rate that can be supported continuously on the single-user Rayleigh fading channel with a finite average transmit power constraint is zero [5]. However, if we allow some transmission outage under severe fading conditions, the maximum instantaneous mutual information rate that can be maintained during nonoutage will increase. Finding the optimal resource allocation strategy that achieves the outage capacity with a given outage probability is tantamount to deriving the strategy that minimizes the outage probability for a given rate vector. In [6], the minimum outage probability problem is solved for the single-user fading channel. In addition, it is shown that under a long-term average power constraint, the optimal power allocation depends on the fading statistics through a threshold-decision rule: no transmission is allowed in a fading state where the required power is above a threshold value.

For an M-user flat-fading broadcast channel and a given rate vector $\boldsymbol{R}$, we consider a similar minimum common outage probability problem under the assumption that the broadcast channel is either not used at all when fading is severe or is used simultaneously for all users when fading is tolerable. Such a common outage transmission strategy might be used in systems trying to minimize probability of detection or systems where all users must obtain information simultaneously for it to be useful (e.g., for a coordinated mission of the users). Under the more complex assumption that an outage can be declared for each user individually, we obtain an optimal power allocation policy that achieves boundaries of outage probability regions for time division (TD), frequency division (FD) and code division (CD) with and without successive decoding. This optimal power allocation strategy is a multiuser generalization of the single-user threshold-decision rule. Such a decision rule is a simple and intuitive method to implement optimal resource allocation in practice.

As a special case, if no outage is allowed during the transmission, the outage capacity with a given outage probability becomes the zero-outage capacity. In [5], with an average power constraint for each user, under the assumption that CSI is available at both the transmitters and the receiver, the

Manuscript received February 11, 1999; revised April 25, 2000. This work was supported by NSF under CAREER Award NCR-9501452 and under a Grant from Pacific Bell, while L. Li pursued the Ph.D. degree under the supervision of A. J. Goldsmith at the California Institute of Technology, Pasadena, CA. The material in this paper was presented in part at the IEEE International Communications Conference (ICC'99)—Eighth Communications Theory Mini-Conference, Vancouver, BC, Canada, June 1999.

L. Li is with the Exeter Group, Inc., Los Angeles, CA 90013 USA (e-mail: lifang@systems.caltech.edu).

A. J. Goldsmith is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9515 USA (email: andrea@systems.stanford.edu).

Communicated by M. L. Honig, Associate Editor for Communications.

Publisher Item Identifier S 0018-9448(01)01352-9.

[1]We see from Part I that the optimal power allocation problem under code division (CD) with successive decoding for the fading broadcast channel is very similar to that for the parallel Gaussian broadcast channel discussed in [3], [4].

zero-outage capacity region[2] and the optimal power allocation scheme are derived for the fading MAC by exploiting the special polymatroidal structure of the region. It is shown that the boundary of this capacity region can be achieved through successive decoding and applying a greedy optimal power allocation scheme. The successive decoding order depends on both the current fading state and the power price for each user.

In this paper, we first obtain directly the zero-outage capacity regions and the associate optimal resource allocation strategies of an M-user flat-fading broadcast channel for TD, FD, and CD with and without successive decoding. These results will later be used in our derivation of the more general outage capacity regions. For CD with successive decoding, we will show that the superposition coding and successive decoding order depends only on the current fading state. For the Nakagami-m fading model [7] we prove that the limiting zero-outage capacity region converges to that of the Gaussian broadcast channel for CD with and without successive decoding when $m \to \infty$. These results about the zero-outage capacity region and the outage capacity region are then extended to frequency-selective fading channels.

Part II of this paper is organized as follows: the flat-fading broadcast channel model is briefly described in Section II. In Section III, the zero-outage capacity regions are derived for each of the different spectrum-sharing techniques. We derive strategies to minimize the common outage probability and achieve the boundary of the outage probability region for TD, FD, and CD with or without successive decoding in Section IV. In Section V, we extend our flat-fading model to the case of frequency-selective fading. Section VI shows numerical results, followed by our conclusions in the last section.

II. The Fading Broadcast Channel

We consider the same discrete-time M-user flat-fading broadcast channel model as in Part I, where the signal source $X[i]$ is composed of M independent information sources and the broadcast channel consists of M independent fading subchannels. The time-varying subchannel gains are denoted as

$$\sqrt{g_1[i]},\ \sqrt{g_2[i]},\ \ldots,\ \sqrt{g_M[i]}$$

and the Gaussian noises of these subchannels are denoted as $z_1[i], z_2[i], \ldots, z_M[i]$. Let $\overline{P}$ be the total average transmit power, B the received signal bandwidth, and ν_j the noise density of $z_j[i]$, $j = 1, 2, \ldots, M$. Since the time-varying received signal-to-noise ratio (SNR) $\gamma_j[i] = \overline{P}g_j[i]/(\nu_j B)$, $j = 1, 2, \ldots, M$, by denoting[3] $n_j[i] = \nu_j/g_j[i]$, we have $\gamma_j[i] = \overline{P}/(n_j[i]B)$.

For a slowly time-varying broadcast channel, we assume that the $n_j[i]$, $j = 1, 2, \ldots, M$, are known to the transmitter and all M receivers at time i. Thus, the transmitter can vary the transmit power $P_j[i]$ for each user relative to the noise density vector

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

subject only to the average power constraint $\overline{P}$. For TD or FD, it can also vary the fraction of transmission time or bandwidth $\tau_j[i]$ assigned to each user, subject to the constraint $\sum_{j=1}^{M} \tau_j[i] = 1$ for all i. For CD, the superposition code can be varied at each transmission. Since every receiver knows the noise density vector $\boldsymbol{n}[i]$, they can decode their individual signals by successive decoding based on the known resource allocation strategy given the M noise densities. In practice, it is necessary to send the transmitter strategy to each receiver through either a header on the transmitted data or a pilot tone. We call $\boldsymbol{n}[i]$ the joint fading process and denote $\mathcal{N}$ as the set of all possible joint fading states. $F(\boldsymbol{n})$ denotes a given cumulative distribution function (cdf) on $\mathcal{N}$.

III. Zero-Outage Capacity Region

For an M-user flat-fading broadcast channel with stationary distribution $\mathcal{Q}$ and a total average power constraint $\overline{P}$, we give the following definition for the zero-outage capacity region $C_{\text{zero}}(\overline{P})$, which is similar to that of the delay-limited capacity region for the MAC in [5].

Definition 1: For a given rate vector $\boldsymbol{R} = (R_1, R_2, \ldots, R_M)$, if $\forall\, \epsilon > 0$, there exists a coding delay T such that for every fading process with stationary distribution $\mathcal{Q}$, there exist codebooks and a decoding scheme with probability of error $P_e^{(T)} < \epsilon$, then $\boldsymbol{R} \in C_{\text{zero}}(\overline{P})$. Moreover, the codewords can be chosen as a function of the realization of the fading processes.

In this section, the zero-outage capacity region of an M-user flat-fading broadcast channel is obtained for CD with and without successive decoding and for TD. For FD, using the same argument as in [8], it can be easily shown that the zero-outage capacity region is the same as for TD and the optimal power and bandwidth allocation policy for FD can be derived directly from that of TD. We will discuss extensions of the results obtained in this section to the case of frequency-selective fading channels in Section V.

A. CD

For an M-user broadcast system, we first consider superposition coding and successive decoding where, in each joint fading state, the channel can be viewed as a degraded Gaussian broadcast channel with noise densities $n_1[i], n_2[i], \ldots, n_M[i]$ and the multiresolution signal constellation is optimized relative to these instantaneous noises. In this case, the users with smaller noise densities will subtract the interference from the users with larger noise densities. Given a power allocation policy $\mathcal{P}$, let $P_j(\boldsymbol{n})$ be the transmit power allocated to User j for the joint fading state $\boldsymbol{n} = (n_1, n_2, \ldots, n_M)$ and denote $\mathcal{F}$ as the set of all possible power policies satisfying the average power constraint $E_{\boldsymbol{n}}\left[\sum_{j=1}^{M} P_j(\boldsymbol{n})\right] \le \overline{P}$, where $E[\cdot]$ denotes the expectation function. For simplicity, assume that the stationary distributions of the fading processes have continuous densities,[4] i.e., $\Pr\{n_i = n_j\} = 0, \forall\, i \ne j$.

[2]The zero-outage capacity is called "delay-limited capacity" in [5], since the coding strategy that achieves the zero-outage capacity has a delay that is independent of the channel variation.

[3]See Part I for a discussion of the case $g_j[i] = 0$.

[4]If $\Pr\{n_i = n_j\} \ne 0$ for some i, j then, in state $\boldsymbol{n}$, User i and User j can be viewed as a single user and superposition coding and successive decoding are applied to $M - 1$ users. The information for User i and User j are then transmitted by time-sharing the channel.

Theorem 1: When the transmitter and all the receivers have perfect CSI, the zero-outage capacity region for the fading broadcast channel is given by

$$\mathcal{C}_{\text{zero}}(\overline{P}) = \bigcup_{\mathcal{P}\in\mathcal{F}} \bigcap_{\boldsymbol{n}\in\mathcal{N}} \mathcal{C}_{\text{CD}}(\boldsymbol{n}, \mathcal{P}) \tag{1}$$

where $\mathcal{C}_{\text{CD}}(\boldsymbol{n}, \mathcal{P})$ is the capacity region of the time-invariant Gaussian broadcast channel. That is

$$\mathcal{C}_{\text{CD}}(\boldsymbol{n},\mathcal{P}) = \left\{ \boldsymbol{R}\text{: } R_j \le B\log\left(1+\frac{P_j(\boldsymbol{n})}{n_jB+\sum_{i=1}^{M} P_i(\boldsymbol{n})\mathbf{1}[n_j>n_i]}\right), \quad 1\le j\le M \right\} \tag{2}$$

where $\mathbf{1}[\cdot]$ denotes the indicator function ($\mathbf{1}[x]=1$ if x is true and zero otherwise).

Proof: See the Appendix, Section A. □

For a given rate vector $\boldsymbol{R}$ and a fading state $\boldsymbol{n}$, from (2) we can calculate the minimum required power $P_j^{\min}(\boldsymbol{n})$ $(j=1, 2, \ldots, M)$ that can support the rate vector $\boldsymbol{R}$. Specifically, let $\pi(\cdot)$ be the permutation such that

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

Then according to (2), we have

$$\begin{cases} R_{\pi(1)} \le B\log\left(1+\dfrac{P_{\pi(1)}(\boldsymbol{n})}{n_{\pi(1)}B}\right) \\ R_{\pi(i)} \le B\log\left(1+\dfrac{P_{\pi(i)}(\boldsymbol{n})}{n_{\pi(i)}B+\sum_{j=1}^{i-1}P_{\pi(j)}(\boldsymbol{n})}\right), \quad 2\le i\le M. \end{cases}$$

Thus, to support rate vector $\boldsymbol{R}$, we require

$$\begin{cases} P_{\pi(1)}(\boldsymbol{n}) \ge n_{\pi(1)}B\left(2^{R_{\pi(1)}/B}-1\right) \\ P_{\pi(i)}(\boldsymbol{n}) \ge \left(n_{\pi(i)}B+\sum_{j=1}^{i-1}P_{\pi(j)}(\boldsymbol{n})\right)\left(2^{R_{\pi(i)}/B}-1\right), \\ \qquad 2\le i\le M. \end{cases}$$

The minimum power required to support $\boldsymbol{R}$ for each user is

$$\begin{cases} P_{\pi(1)}^{\min}(\boldsymbol{n}) = n_{\pi(1)}B\left(2^{R_{\pi(1)}/B}-1\right) \\ P_{\pi(i)}^{\min}(\boldsymbol{n}) = \left(n_{\pi(i)}B+\sum_{j=1}^{i-1}P_{\pi(j)}^{\min}(\boldsymbol{n})\right)\left(2^{R_{\pi(i)}/B}-1\right), \\ \qquad 2\le i\le M. \end{cases}$$

Consequently, the minimum required total power $P^{\min}(\boldsymbol{R}, \boldsymbol{n})$ that can support $\boldsymbol{R}$ in fading state $\boldsymbol{n}$ is

$$\begin{aligned} P^{\min}(\boldsymbol{R},\boldsymbol{n}) &= \sum_{i=1}^{M} P_{\pi(i)}^{\min}(\boldsymbol{n}) \\ &= \sum_{i=1}^{M-1}\left[2^{\sum_{j=i+1}^{M} R_{\pi(j)}/B}\left(2^{R_{\pi(i)}/B}-1\right)n_{\pi(i)}B\right] \\ &\quad + \left(2^{R_{\pi(M)}/B}-1\right)n_{\pi(M)}B. \end{aligned} \tag{3}$$

For a given $\boldsymbol{R}$, if $\boldsymbol{R}\in\mathcal{C}_{\text{zero}}(\overline{P})$, then by (1), the minimum required average power $E_{\boldsymbol{n}}[P^{\min}(\boldsymbol{R}, \boldsymbol{n})]$ satisfies the total average power constraint

$$E_{\boldsymbol{n}}[P^{\min}(\boldsymbol{R}, \boldsymbol{n})] \le \overline{P} \tag{4}$$

where $P^{\min}(\boldsymbol{R}, \boldsymbol{n})$ is given by (3). If $\boldsymbol{R}$ is on the boundary surface of $\mathcal{C}_{\text{zero}}(\overline{P})$, then the equality in (4) is achieved. Note that for the single-user case $(M=1)$, if R_1 is on the boundary of $\mathcal{C}_{\text{zero}}(\overline{P})$, from (3) and (4) we have

$$E_{n_1}\left[\left(2^{R_1/B}-1\right)n_1B\right] = \overline{P}.$$

Thus,

$$R_1 = B\log\left(1+\frac{\overline{P}}{E_{n_1}[n_1B]}\right)$$

which is the same as derived in [2].

B. CD Without Successive Decoding

In CD without successive decoding, each receiver treats the signals for other users as interfering noise. For a given power allocation policy $\mathcal{P}$, let $P_j(\boldsymbol{n})$ denote the transmit power allocated to User j in the state $\boldsymbol{n}$ and let $\mathcal{F}$ denote the set of all possible power policies satisfying the average power constraint

$$E_{\boldsymbol{n}}\left[\sum_{j=1}^{M} P_j(\boldsymbol{n})\right] \le \overline{P}.$$

Then the achievable zero-outage rate region for CD without successive decoding is given by

$$\mathcal{C}_{\text{zero}}(\overline{P}) = \bigcup_{\mathcal{P}\in\mathcal{F}} \bigcap_{\boldsymbol{n}\in\mathcal{N}} \mathcal{C}_{\text{CDWO}}(\boldsymbol{n}, \mathcal{P}) \tag{5}$$

where $\mathcal{C}_{\text{CDWO}}(\boldsymbol{n}, \mathcal{P})$ is the rate region of the time-invariant Gaussian broadcast channel using CD without successive decoding:

$$\mathcal{C}_{\text{CDWO}}(\boldsymbol{n}, \mathcal{P}) = \left\{ \boldsymbol{R}\text{:} R_j \le B\log\left(1+\frac{P_j(\boldsymbol{n})}{n_jB+\sum_{i=1,\, i\neq j}^{M} P_i(\boldsymbol{n})}\right), \quad 1\le j\le M \right\}. \tag{6}$$

The proof of the achievability follows along the same lines as that for the capacity region of CD given in the Appendix, Section

A and is therefore omitted. Note that in this paper, we refer to this achievable rate region as the zero-outage capacity region for CD without successive decoding, though we do not have a converse proof since the converse only applies to the optimal transmission strategy, which is CD with successive decoding.

For a given rate vector $\boldsymbol{R}$ and a fading state $\boldsymbol{n}$, we know from (6) that

$$P_i(\boldsymbol{n}) \geq \left(n_i B + \sum_{j=1,\, j\neq i}^{M} P_j(\boldsymbol{n})\right)\left(2^{R_i/B}-1\right), \qquad i=1,\,2,\,\ldots,\,M. \tag{7}$$

Denoting $P_i^{\min}(\boldsymbol{n})$ $(i=1,\,2,\,\ldots,\,M)$ as the minimum power required for User i in order to support rate vector $\boldsymbol{R}$, by (7) we have

$$P_i^{\min}(\boldsymbol{n}) \geq \left(n_i B + \sum_{j=1,\, j\neq i}^{M} P_j^{\min}(\boldsymbol{n})\right)\left(2^{R_i/B}-1\right), \qquad i=1,\,2,\,\ldots,\,M.$$

Therefore, $P_i^{\min}(\boldsymbol{n})$ must satisfy

$$P_i^{\min}(\boldsymbol{n}) = \left(n_i B + \sum_{j=1,\, j\neq i}^{M} P_j^{\min}(\boldsymbol{n})\right)\left(2^{R_i/B}-1\right), \qquad i=1,\,2,\,\ldots,\,M. \tag{8}$$

By defining matrix $A=(a_{ij})$, $i,\, j=1,\,2,\,\ldots,\,M$, where

$$a_{ij}=\begin{cases}\frac{1}{2^{R_i/B}-1}, & \text{if } i=j\\ -1, & \text{if } i\neq j\end{cases} \tag{9}$$

we prove in the Appendix, Section B that the M linear equations in (8) have positive solutions for all $P_i^{\min}(\boldsymbol{n})$ $(1\leq i\leq M)$ in every fading state $\boldsymbol{n}$ if and only if $\det A>0$. Assuming that $\det A>0$, it is clear that the explicit solution to

$$\boldsymbol{P}^{\min}(\boldsymbol{n})=(P_1^{\min}(\boldsymbol{n}),\, P_2^{\min}(\boldsymbol{n}),\,\ldots,\, P_M^{\min}(\boldsymbol{n}))$$

is

$$\boldsymbol{P}^{\min}(\boldsymbol{n})=A^{-1}\cdot B\boldsymbol{n}^T \tag{10}$$

where A^{-1} denotes the inverse of matrix A and $\boldsymbol{n}^T$ denotes the transpose of vector $\boldsymbol{n}$. Thus, the minimum required total power $P^{\min}(\boldsymbol{R},\,\boldsymbol{n})$ is

$$P^{\min}(\boldsymbol{R},\,\boldsymbol{n})=\sum_{i=1}^{M} P_i^{\min}(\boldsymbol{n}). \tag{11}$$

For example, in the two-user case $(M=2)$, if $\det A>0$, i.e., if $2^{R_1/B}+2^{R_2/B}>2^{(R_1+R_2)/B}$, the solution for $P_1^{\min}(\boldsymbol{n})$ and $P_2^{\min}(\boldsymbol{n})$ will be

$$\begin{cases}P_1^{\min}(\boldsymbol{n})=\frac{\left(2^{R_1/B}-1\right)n_1B+\left(2^{R_1/B}-1\right)\left(2^{R_2/B}-1\right)n_2B}{2^{R_1/B}+2^{R_2/B}-2^{(R_1+R_2)/B}}\\ P_2^{\min}(\boldsymbol{n})=\frac{\left(2^{R_2/B}-1\right)n_2B+\left(2^{R_2/B}-1\right)\left(2^{R_1/B}-1\right)n_1B}{2^{R_1/B}+2^{R_2/B}-2^{(R_1+R_2)/B}}.\end{cases}$$

Thus,

$$\begin{aligned}P^{\min}(\boldsymbol{R},\,\boldsymbol{n}) &= P_1^{\min}(\boldsymbol{n})+P_2^{\min}(\boldsymbol{n})\\ &= \frac{\left(2^{(R_1+R_2)/B}-2^{R_2/B}\right)n_1B}{2^{R_1/B}+2^{R_2/B}-2^{(R_1+R_2)/B}}\\ &\quad+\frac{\left(2^{(R_1+R_2)/B}-2^{R_1/B}\right)n_2B}{2^{R_1/B}+2^{R_2/B}-2^{(R_1+R_2)/B}}.\end{aligned} \tag{12}$$

Therefore, a given $\boldsymbol{R}\in\mathcal{C}_{\text{zero}}(\overline{P})$ if the average power constraint in (4) is satisfied with $P^{\min}(\boldsymbol{R},\,\boldsymbol{n})$ given by (11). If $\boldsymbol{R}$ is on the boundary surface of $\mathcal{C}_{\text{zero}}(\overline{P})$, then the equality in (4) is achieved.

C. TD

Now we consider the TD case where, in each fading state $\boldsymbol{n}$, the information for the M users will be divided and sent in time slots which are functions of $\boldsymbol{n}$. For a given power and time allocation policy $\mathcal{P}$, let $P_j(\boldsymbol{n})$ and $\tau_j(\boldsymbol{n})$ $(0\leq\tau_j(\boldsymbol{n})\leq 1)$ be the transmit power and fraction of transmission time allocated to User j $(j=1,\,2,\,\ldots,\,M)$, respectively, for fading state $\boldsymbol{n}$, and let $\mathcal{F}$ be the set of all such possible power and time allocation policies satisfying

$$\begin{cases}E_{\boldsymbol{n}}\left[\sum_{j=1}^{M}\tau_j(\boldsymbol{n})P_j(\boldsymbol{n})\right]\leq\overline{P} \qquad \text{and}\\ \sum_{j=1}^{M}\tau_j(\boldsymbol{n})=1, \qquad \forall\,\boldsymbol{n}\in\mathcal{N}.\end{cases} \tag{13}$$

Then the achievable zero-outage capacity region for the variable power and transmission time scheme is

$$\mathcal{C}_{\text{zero}}(\overline{P})=\bigcup_{\mathcal{P}\in\mathcal{F}}\bigcap_{\boldsymbol{n}\in\mathcal{N}}\mathcal{C}_{\text{TD}}(\boldsymbol{n},\,\mathcal{P}) \tag{14}$$

where $\mathcal{C}_{\text{TD}}(\boldsymbol{n},\,\mathcal{P})$ is the rate region of the time-invariant Gaussian broadcast channel using the TD technique

$$\mathcal{C}_{\text{TD}}(\boldsymbol{n},\mathcal{P})=\left\{\boldsymbol{R}\colon R_j\leq\tau_j(\boldsymbol{n})B\log\left(1+\frac{P_j(\boldsymbol{n})}{n_jB}\right),\,1\leq j\leq M\right\}. \tag{15}$$

The proof of the achievability follows along the same lines as that for the capacity region of CD given in the Appendix, Section A and is therefore omitted. Note that as in the case of CD without successive decoding, we refer to this achievable rate region as the zero-outage capacity region for TD, though we do not have a converse proof due to the fact that the converse only holds for the optimal transmission strategy for this channel, which, according to Theorem 1, is CD with successive decoding.

For a given rate vector $\boldsymbol{R}$ and a fading state $\boldsymbol{n}$, from (15) we have

$$P_i(\boldsymbol{n})\geq n_iB\left(2^{\frac{R_i}{B\tau_i(\boldsymbol{n})}}-1\right), \qquad i=1,\,2,\,\ldots,\,M.$$

Therefore, the required total power $P(\boldsymbol{R},\,\boldsymbol{n})$ of the M users satisfies

$$\begin{aligned}P(\boldsymbol{R},\,\boldsymbol{n}) &= \sum_{i=1}^{M}\tau_i(\boldsymbol{n})P_i(\boldsymbol{n})\\ &\geq\sum_{i=1}^{M}\tau_i(\boldsymbol{n})n_iB\left(2^{\frac{R_i}{B\tau_i(\boldsymbol{n})}}-1\right).\end{aligned}$$

Let

$$\boldsymbol{\tau}(\boldsymbol{n})\triangleq[\tau_1(\boldsymbol{n}),\,\tau_2(\boldsymbol{n}),\,\ldots,\,\tau_M(\boldsymbol{n})]$$

and let $P^{\min}(\boldsymbol{R},\,\boldsymbol{n})$ be the minimum required total power of the M users for fading state $\boldsymbol{n}$, then

$$\begin{cases}P^{\min}(\boldsymbol{R},\,\boldsymbol{n})=\min_{\boldsymbol{\tau}(\boldsymbol{n})}\left\{\sum_{i=1}^{M}\tau_i(\boldsymbol{n})n_iB\left(2^{\frac{R_i}{B\tau_i(\boldsymbol{n})}}-1\right)\right\}\\ \text{subject to:}\quad \sum_{i=1}^{M}\tau_i(\boldsymbol{n})=1, \qquad \forall\,\boldsymbol{n}\in\mathcal{N}.\end{cases} \tag{16}$$

By applying the Lagrangian technique, we can find the optimal $\boldsymbol{\tau}(\boldsymbol{n})$ which achieves $P^{\min}(\boldsymbol{R}, \boldsymbol{n})$ in (16). For example, in a two-user system $(M = 2)$, let

$$P(\tau_1(\boldsymbol{n})) \triangleq \tau_1(\boldsymbol{n}) n_1 B \left(2^{\frac{R_1}{B\tau_1(\boldsymbol{n})}} - 1\right) + (1 - \tau_1(\boldsymbol{n})) n_2 B \left(2^{\frac{R_2}{B(1-\tau_1(\boldsymbol{n}))}} - 1\right)$$

and let $\tau_1^*(\boldsymbol{n})$ be the solution to the nonlinear equation

$$\frac{dP(\tau_1(\boldsymbol{n}))}{d\tau_1(\boldsymbol{n})} = 0.$$

Since it is easy to verify that for $\tau_1(\boldsymbol{n}) > 0$, $\frac{d^2 P(\tau_1(\boldsymbol{n}))}{d\tau_1^2(\boldsymbol{n})} > 0$, we have

$$P^{\min}(\boldsymbol{R}, \boldsymbol{n}) = \min_{\tau_1(\boldsymbol{n})} P(\tau_1(\boldsymbol{n})) = P(\tau_1^*(\boldsymbol{n})).$$

Therefore, $\boldsymbol{R} \in \mathcal{C}_{\text{zero}}(\overline{P})$ if $P^{\min}(\boldsymbol{R}, \boldsymbol{n})$ in (16) satisfies (4). If $\boldsymbol{R}$ is on the boundary surface of $\mathcal{C}_{\text{zero}}(\overline{P})$ then the equality in (4) is achieved.

D. The Limiting Zero-Outage Capacity Region for Nakagami Fading

The Nakagami-m fading model [9] can be used to describe different fading conditions ranging from Rayleigh $(m = 1)$ to Rician channels with strong line-of-sight components. In fact, there is a direct mapping from the Rician K factor to the Nakagami m parameter. The Nakagami distribution also has a more tractable mathematical form than the Rician distribution. As the fading parameter m $(m \geq 1/2)$ goes to infinity, the Nakagami-m fading channel converges to an additive white Gaussian noise (AWGN) channel. Therefore, it is expected that the limiting zero-outage capacity region of the Nakagami-m fading broadcast channel converges to the capacity region of an AWGN broadcast channel as $m \to \infty$. In this section, we prove this to be true for CD with and without successive decoding in a two-user system. These results can be easily extended to more users.

1) CD with Successive Decoding: For a two-user broadcast channel with fading, given rate vector $\boldsymbol{R} = (R_1, R_2)$, we know by (3) that the minimum required total power to support $\boldsymbol{R}$ in fading state $\boldsymbol{n} = (n_1, n_2)$ is

$$P^{\min}(\boldsymbol{R}, \boldsymbol{n}) = \begin{cases} 2^{R_2/B}(2^{R_1/B} - 1) n_1 B \\ \quad + (2^{R_2/B} - 1) n_2 B, & \text{if } n_1 < n_2 \\ 2^{R_1/B}(2^{R_2/B} - 1) n_2 B \\ \quad + (2^{R_1/B} - 1) n_1 B, & \text{if } n_1 > n_2. \end{cases} \tag{17}$$

If $\boldsymbol{R}$ is on the boundary surface of $\mathcal{C}_{\text{zero}}(\overline{P})$ in (1), by substituting (17) into (4) with equality, we obtain

$$\begin{aligned}\overline{P} &= E_{\boldsymbol{n}}[P^{\min}(\boldsymbol{R}, \boldsymbol{n})] \\ &= 2^{R_2/B}(2^{R_1/B} - 1) E_{n_1<n_2}[n_1 B] \\ &\quad + (2^{R_2/B} - 1) E_{n_1<n_2}[n_2 B] \\ &\quad + 2^{R_1/B}(2^{R_2/B} - 1) E_{n_1>n_2}[n_2 B] \\ &\quad + (2^{R_1/B} - 1) E_{n_1>n_2}[n_1 B]. \end{aligned} \tag{18}$$

Let $\overline{n}_1 B$ and $\overline{n}_2 B$ be the average noise variances of the channels for User 1 and User 2, respectively. Assuming that the signal power is normalized to 1, the SNR γ_i of User i $(i = 1, 2)$ for a given channel state $\boldsymbol{n}$ is

$$\gamma_i = \frac{1}{n_i B}. \tag{19}$$

We use the following lemma to show that for Nakagami-m fading, as $m \to \infty$, (18) converges to the boundary equation for the capacity region of an AWGN broadcast channel with subchannel noise variances $\overline{n}_1 B$ and $\overline{n}_2 B$.

Lemma 1: Given $0 < p < 1/2$ and a fixed integer r

$$\lim_{k\to\infty} \sum_{i=0}^{k+r} \binom{i+k-1}{i} p^i (1-p)^k = 1.$$

Proof: See the Appendix, Section C. □

For Nakagami-m fading, the probability density function (pdf) of γ_i in (19) is

$$p_i(\gamma_i) = \frac{(m\overline{n}_i B)^m}{(m-1)!} \gamma_i^{m-1} e^{-m\overline{n}_i B \gamma_i}, \qquad i = 1, 2. \tag{20}$$

Thus, from (18) and (19) we know that

$$\begin{aligned}\overline{P} &= 2^{R_2/B}(2^{R_1/B} - 1) D_a(m) + (2^{R_2/B} - 1) D_b(m) \\ &\quad + 2^{R_1/B}(2^{R_2/B} - 1) D_c(m) + (2^{R_1/B} - 1) D_d(m) \end{aligned} \tag{21}$$

where

$$D_a(m) = E_{n_1<n_2}[n_1 B] = \int_{\gamma_1>\gamma_2} \frac{1}{\gamma_1} p_1(\gamma_1) p_2(\gamma_2)\, d\boldsymbol{\gamma} \tag{22}$$

$$D_b(m) = E_{n_1<n_2}[n_2 B] = \int_{\gamma_1>\gamma_2} \frac{1}{\gamma_2} p_1(\gamma_1) p_2(\gamma_2)\, d\boldsymbol{\gamma} \tag{23}$$

$$D_c(m) = E_{n_1>n_2}[n_2 B] = \int_{\gamma_1<\gamma_2} \frac{1}{\gamma_2} p_1(\gamma_1) p_2(\gamma_2)\, d\boldsymbol{\gamma} \tag{24}$$

$$D_d(m) = E_{n_1>n_2}[n_1 B] = \int_{\gamma_1<\gamma_2} \frac{1}{\gamma_1} p_1(\gamma_1) p_2(\gamma_2)\, d\boldsymbol{\gamma}. \tag{25}$$

By applying *Lemma 1*, we obtain the following lemma.

Lemma 2: For Nakagami-m fading, assuming that $\overline{n}_1 B < \overline{n}_2 B$

$$\begin{aligned}\lim_{m\to\infty} D_a(m) &= \overline{n}_1 B \\ \lim_{m\to\infty} D_b(m) &= \overline{n}_2 B \\ \lim_{m\to\infty} D_c(m) &= 0 \\ \lim_{m\to\infty} D_d(m) &= 0.\end{aligned}$$

Proof: See the Appendix, Section D. □

Theorem 2: As $m \to \infty$, assuming that $\overline{n}_1 B < \overline{n}_2 B$, the boundary of the capacity region for Nakagami-m fading broadcast channel (21) becomes

$$\overline{P} = 2^{R_2/B}(2^{R_1/B} - 1)\overline{n}_1 B + (2^{R_2/B} - 1)\overline{n}_2 B \tag{26}$$

which is the same as the boundary of the capacity region for the AWGN broadcast channel using CD with successive decoding.

Proof: Applying Lemma 2 to (21) directly yields (26). For the two-user degraded AWGN broadcast channel with noise variances $\overline{n}_1 B$ and $\overline{n}_2 B$ ($\overline{n}_1 B < \overline{n}_2 B$), the capacity region for CD with successive decoding is [10], [11]

$$\mathcal{C}_{\text{CD}} = \left\{ \boldsymbol{R}\colon R_1 \le B \log\left(1 + \frac{P_1}{\overline{n}_1 B}\right), \right. \\ \left. R_2 \le B \log\left(1 + \frac{P_2}{\overline{n}_2 B + P_1}\right); \forall P_1 + P_2 \le \overline{P} \right\}. \tag{27}$$

Therefore, if $\boldsymbol{R}$ is on the boundary of the capacity region (27), i.e., all the equalities in (27) are achieved, then

$$\begin{aligned} \overline{P} &= P_1 + P_2 \\ &= (2^{R_1/B} - 1)\overline{n}_1 B \\ &\quad + (2^{R_2/B} - 1)[\overline{n}_2 B + (2^{R_1/B} - 1)\overline{n}_1 B] \\ &= 2^{R_2/B}(2^{R_1/B} - 1)\overline{n}_1 B + (2^{R_2/B} - 1)\overline{n}_2 B \end{aligned}$$

which means that $\boldsymbol{R}$ also satisfies (26). □

2) CD Without Successive Decoding: For a two-user broadcast channel with fading, given rate vector $\boldsymbol{R} = (R_1, R_2)$, we know by (12) that the minimum required total power to support $\boldsymbol{R}$ in a fading state $\boldsymbol{n} = (n_1, n_2)$ is

$$P^{\min}(\boldsymbol{R}, \boldsymbol{n}) = \frac{\left(2^{(R_1+R_2)/B} - 2^{R_2/B}\right) n_1 B}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}} \\ + \frac{\left(2^{(R_1+R_2)/B} - 2^{R_1/B}\right) n_2 B}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}}. \tag{28}$$

If $\boldsymbol{R}$ is on the boundary surface of $\mathcal{C}_{\text{zero}}(\overline{P})$ in (5), substituting (28) into (4) with equality we obtain

$$\begin{aligned} \overline{P} &= \frac{\left(2^{(R_1+R_2)/B} - 2^{R_2/B}\right) E_{n_1}[n_1 B]}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}} \\ &\quad + \frac{\left(2^{(R_1+R_2)/B} - 2^{R_1/B}\right) E_{n_2}[n_2 B]}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}} \\ &= \frac{\left(2^{(R_1+R_2)/B} - 2^{R_2/B}\right) [D_a(m) + D_d(m)]}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}} \\ &\quad + \frac{\left(2^{(R_1+R_2)/B} - 2^{R_1/B}\right) [D_b(m) + D_c(m)]}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}} \end{aligned} \tag{29}$$

where $D_a(m)$, $D_b(m)$, $D_c(m)$, and $D_d(m)$ are as defined in (22)–(25) for the Nakagami-m fading channel. From Lemma 2 we know that

$$\begin{aligned} \lim_{m\to\infty} [D_a(m) + D_d(m)] &= \overline{n}_1 B \\ \lim_{m\to\infty} [D_b(m) + D_c(m)] &= \overline{n}_2 B. \end{aligned}$$

Thus, as $m \to \infty$, (29) becomes

$$\overline{P} = \frac{\left(2^{(R_1+R_2)/B} - 2^{R_2/B}\right) \overline{n}_1 B}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}} \\ + \frac{\left(2^{(R_1+R_2)/B} - 2^{R_1/B}\right) \overline{n}_2 B}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}} \tag{30}$$

which is the same as the boundary of the capacity region for the AWGN broadcast channel using CD without successive decoding, since for the two-user degraded AWGN broadcast channel with noise variances $\overline{n}_1 B$ and $\overline{n}_2 B$ ($\overline{n}_1 B < \overline{n}_2 B$), the capacity region for CD without successive decoding is [8]

$$\mathcal{C}_{\text{CDWO}} = \left\{ \boldsymbol{R}\colon R_1 \le B \log\left(1 + \frac{P_1}{\overline{n}_1 B + P_2}\right), \right. \\ \left. R_2 \le B \log\left(1 + \frac{P_2}{\overline{n}_2 B + P_1}\right);\ \forall P_1 + P_2 \le \overline{P} \right\}.$$

IV. Outage Capacities and Minimum Outage Probability

In the previous section we have obtained the zero-outage capacity region of an M-user flat-fading broadcast channel, where the transmitter was required to maintain a constant rate for each user no matter how severe its fading. We now consider the outage capacity region for this channel, where the transmitter may suspend transmission over a subset of fading states with a given probability. Specifically, for a given average power constraint $\overline{P}$, the outage capacity regions $C_{\text{out}}(\overline{P}, \Pr)$ and $C_{\text{out}}(\overline{P}, \mathbf{Pr})$ are defined as follows.

Definition 2: Assuming that the transmission to all users is turned on or off simultaneously so that the outage probability for each user is the same (common outage probability), for a given $0 \le \Pr \le 1$, the outage capacity region $C_{\text{out}}(\overline{P}, \Pr)$ consists of all rate vectors $\boldsymbol{R} = (R_1, R_2, \ldots, R_M)$ which can be maintained with a common outage probability no larger than $\Pr$ under the power constraint $\overline{P}$.

Definition 3: Assuming that the transmission to each user is turned on or off independently so that the outage probability for each user may be different, for a given probability vector $\mathbf{Pr} = (\Pr_1, \Pr_2, \ldots, \Pr_M)$, the outage capacity region $C_{\text{out}}(\overline{P}, \mathbf{Pr})$ consists of all rate vectors $\boldsymbol{R} = (R_1, R_2, \ldots, R_M)$ which can be maintained with the outage probability for User j no larger than $\Pr_j$ ($\forall 1 \le j \le M$) under the given power constraint $\overline{P}$.

With these definitions, we wish to find: a) the optimal resource allocation strategy that achieves the boundary of the outage capacity region $C_{\text{out}}(\overline{P}, \Pr)$; b) the optimal resource allocation strategy that achieves the boundary of $C_{\text{out}}(\overline{P}, \mathbf{Pr})$. The first optimization problem is equivalent to deriving the resource allocation policy that minimizes the common outage probability for a given rate vector $\boldsymbol{R}$ and we have the following definition for the corresponding minimum common outage probability $\Pr_{\min}(\overline{P}, \boldsymbol{R})$.

Definition 4: Assuming that the transmission to all users is turned on or off simultaneously, the minimum common outage probability $\Pr_{\min}(\overline{P}, \boldsymbol{R})$ is the smallest common outage probability with which the rate vector $\boldsymbol{R}$ can be maintained under the given power constraint $\overline{P}$.

The second optimization problem is equivalent to obtaining the resource allocation policy that achieves the boundary of the outage probability region $\mathcal{O}(\overline{P}, \boldsymbol{R})$ or the usage probability region $\overline{\mathcal{O}}(\overline{P}, \boldsymbol{R})$ defined as follows.

Definition 5: Assuming that the transmission to each user is turned on or off independently, for a given rate vector $\boldsymbol{R}$, the outage probability region $\mathcal{O}(\overline{P}, \boldsymbol{R})$ consists of all outage probability vectors $\mathbf{Pr}$ for which $\boldsymbol{R}$ can be maintained for the M users under the given power constraint $\overline{P}$.

Definition 6: The usage probability region $\overline{\mathcal{O}}(\overline{P}, \boldsymbol{R})$ is the complementary region of the outage probability region $\mathcal{O}(\overline{P}, \boldsymbol{R})$, i.e., if a probability vector

$$\mathbf{Pr} = (\mathrm{Pr}_1, \mathrm{Pr}_2, \ldots, \mathrm{Pr}_M) \in \mathcal{O}(\overline{P}, \boldsymbol{R})$$

then the probability vector

$$\mathbf{Pr}^{\mathrm{on}} = (\mathrm{Pr}_1^{\mathrm{on}}, \mathrm{Pr}_2^{\mathrm{on}}, \ldots, \mathrm{Pr}_M^{\mathrm{on}}) \in \overline{\mathcal{O}}(\overline{P}, \boldsymbol{R})$$

where

$$\mathrm{Pr}_j^{\mathrm{on}} = 1 - \mathrm{Pr}_j \qquad \forall 1 \le j \le M.$$

With the above definitions, it is easily seen that given $0 \le \mathrm{Pr} \le 1$, the outage capacity region $C_{\mathrm{out}}(\overline{P}, \mathrm{Pr})$ is implicitly obtained once the minimum common outage probability $\mathrm{Pr}_{\min}(\overline{P}, \boldsymbol{R})$ for a given rate vector is calculated under the optimal resource allocation, since $\forall \boldsymbol{R}$, we can determine that $\boldsymbol{R} \in C_{\mathrm{out}}(\overline{P}, \mathrm{Pr})$ if $\mathrm{Pr}_{\min}(\overline{P}, \boldsymbol{R}) \le \mathrm{Pr}$, and $\boldsymbol{R} \notin C_{\mathrm{out}}(\overline{P}, \mathrm{Pr})$ otherwise. Similarly, given a probability vector $\mathbf{Pr}$, the outage capacity region $C_{\mathrm{out}}(\overline{P}, \mathbf{Pr})$ is implicitly obtained once the boundary of the outage probability region $\mathcal{O}(\overline{P}, \boldsymbol{R})$ [and so the whole region $\mathcal{O}(\overline{P}, \boldsymbol{R})$] for a given rate vector $\boldsymbol{R}$ is derived through the optimal resource allocation, since $\forall \boldsymbol{R}$, we can determine that $\boldsymbol{R} \in C_{\mathrm{out}}(\overline{P}, \mathbf{Pr})$ if $\mathbf{Pr} \in \mathcal{O}(\overline{P}, \boldsymbol{R})$, and $\boldsymbol{R} \notin C_{\mathrm{out}}(\overline{P}, \mathbf{Pr})$ otherwise. We now derive the minimum common outage probability $\mathrm{Pr}_{\min}(\overline{P}, \boldsymbol{R})$ and the corresponding optimal resource allocation strategy in Section IV-A. We obtain the outage probability region boundary of $\mathcal{O}(\overline{P}, \boldsymbol{R})$ as well as the optimal resource allocation strategy in Section IV-B for the case of independent outage problems.

A. Minimum Common Outage Probability

Certain systems might require an outage to be declared simultaneously for all users, either to minimize the probability of detection or in situations where users are coordinating based on the transmitted information. Under the assumption that an outage is declared for all users simultaneously, the minimum common outage probability problem for the M-user broadcast channel is similar to that of the single-user case [6]. For each joint fading state $\boldsymbol{n}$ and a given rate vector $\boldsymbol{R}$, the minimum required total power $P^{\min}(\boldsymbol{R}, \boldsymbol{n})$ for the M users using CD with or without successive decoding or using TD can be calculated as in (3), (11), or (16), respectively. Thus, $\forall s > 0$, we define the sets of fading states $\mathcal{R}(s)$ and $\tilde{\mathcal{R}}(s)$ as

$$\mathcal{R}(s) = \{\boldsymbol{n}: P^{\min}(\boldsymbol{R}, \boldsymbol{n}) < s\} \tag{31}$$

$$\tilde{\mathcal{R}}(s) = \{\boldsymbol{n}: P^{\min}(\boldsymbol{R}, \boldsymbol{n}) \le s\}. \tag{32}$$

The corresponding average power over the two sets are

$$P(s) = E_{\boldsymbol{n} \in \mathcal{R}(s)} \left[P^{\min}(\boldsymbol{R}, \boldsymbol{n})\right] \tag{33}$$

$$\tilde{P}(s) = E_{\boldsymbol{n} \in \tilde{\mathcal{R}}(s)} \left[P^{\min}(\boldsymbol{R}, \boldsymbol{n})\right]. \tag{34}$$

For a given total power $\overline{P} > 0$, let

$$s^* \triangleq \sup\{s: P(s) < \overline{P}\} \tag{35}$$

$$w^* \triangleq \frac{\overline{P} - P(s^*)}{\tilde{P}(s^*) - P(s^*)}. \tag{36}$$

By using [6, Lemma 3], for each fading state $\boldsymbol{n}$, the optimal power policy that minimizes the common outage probability is: if $\boldsymbol{n} \notin \tilde{\mathcal{R}}(s^*)$, no power is assigned to any user; if $\boldsymbol{n} \in \mathcal{R}(s^*)$, a total power of $P^{\min}(\boldsymbol{R}, \boldsymbol{n})$ is assigned to the M users and the power to each user is allocated as described in Section III; if $\boldsymbol{n} \notin \mathcal{R}(s^*)$ but $\boldsymbol{n} \in \tilde{\mathcal{R}}(s^*)$, then with probability w^*, $P^{\min}(\boldsymbol{R}, \boldsymbol{n})$ is assigned to the M users and with probability $1-w^*$, no power is assigned to any user. The minimum common outage probability $\mathrm{Pr}_{\min}(\overline{P}, \boldsymbol{R})$ is

$$\begin{aligned}\mathrm{Pr}_{\min}(\overline{P}, \boldsymbol{R}) = 1 &- \mathrm{Pr}\{\boldsymbol{n} \in \mathcal{R}(s^*)\}\\ &- w^* \mathrm{Pr}\left\{\boldsymbol{n} \in \tilde{\mathcal{R}}(s^*) \text{ and } \boldsymbol{n} \notin \mathcal{R}(s^*)\right\}\end{aligned} \tag{37}$$

where $\mathrm{Pr}\{\cdot\}$ denotes the probability function.

B. Outage Probability Region

We now consider the case where an outage can be declared independently for each user. From Definitions 5 and 6, it is clear that for a given rate vector $\boldsymbol{R}$ and an average power constraint $\overline{P}$, deriving the boundary of the outage probability region $\mathcal{O}(\overline{P}, \boldsymbol{R})$ is equivalent to deriving the boundary of the usage probability region $\overline{\mathcal{O}}(\overline{P}, \boldsymbol{R})$. We will require the following definition and lemma to derive the boundary of $\overline{\mathcal{O}}(\overline{P}, \boldsymbol{R})$ and the corresponding optimal power allocation that achieves this boundary.

Definition 7: For a given rate vector

$$\boldsymbol{R} = (R_1, R_2, \ldots, R_M)$$

assume that rate R_i is maintained with probability $\mathrm{Pr}_i^{\mathrm{on}}(\boldsymbol{R})$, $1 \le i \le M$. Denoting

$$\mathrm{Pr}^{\mathrm{on}}(\boldsymbol{R}) = [\mathrm{Pr}_1^{\mathrm{on}}(\boldsymbol{R}), \mathrm{Pr}_2^{\mathrm{on}}(\boldsymbol{R}), \ldots, \mathrm{Pr}_M^{\mathrm{on}}(\boldsymbol{R})]$$

the total usage reward $W(\boldsymbol{R})$ is

$$W(\boldsymbol{R}) = \boldsymbol{\mu}\mathbf{Pr}^{\mathrm{on}}(\boldsymbol{R}) = \sum_{i=1}^{M} \mu_i \mathrm{Pr}_i^{\mathrm{on}}(\boldsymbol{R})$$

where $\boldsymbol{\mu} \in \Re_+^M$ with $\sum_{j=1}^M \mu_j = 1$, and μ_i is the relative reward if the information for User i is transmitted.[5]

Lemma 3: The usage probability region $\overline{\mathcal{O}}(\overline{P}, \boldsymbol{R})$ of the fading broadcast channel is convex.

This lemma can be easily shown to be true by using the time-sharing technique. Therefore, since $\overline{\mathcal{O}}(\overline{P}, \boldsymbol{R})$ is convex, $\forall \boldsymbol{\mu} \in \Re_+^M$ with $\sum_{j=1}^M \mu_j = 1$, a usage probability vector $\mathbf{Pr}^{\mathrm{on}}(\boldsymbol{R})$ will be on the boundary surface of $\overline{\mathcal{O}}(\overline{P}, \boldsymbol{R})$ if it is a solution to

$$\max_{\mathbf{Pr}^{\mathrm{on}}(\boldsymbol{R}) \in \overline{\mathcal{O}}(\overline{P}, \boldsymbol{R})} W(\boldsymbol{R}) \tag{38}$$

where the total usage reward $W(\boldsymbol{R})$ is defined in Definition 7.

For any given fading state $\boldsymbol{n}$, $\sum_{i=1}^M \binom{M}{i} = 2^M - 1$ different combinations of the M users may be transmitting over the

[5]Also, μ_i can be viewed as the relative penalty if an outage is declared for User i.

channel. We will represent each of these $2^M - 1$ possible combinations of users as a vector $[\psi(k, 1), \psi(k, 2), \ldots, \psi(k, M)]$ equal to the binary expansion of k, $1 \le k \le 2^M - 1$. For each vector $[\psi(k, 1), \psi(k, 2), \ldots, \psi(k, M)]$, if $\psi(k, i) = 1$, then User i is transmitting; otherwise, User i is not. For $0 \le k \le 2^M - 1$, we define the set of active users $\mathcal{U}_k$ relative to k as

$$\mathcal{U}_k = \{j: \psi(k, j) = 1, 1 \le j \le M\}.$$

$\mathcal{U}_0$ denotes the empty set (no active users). For any fading state $\boldsymbol{n}$, suppose that the broadcast channel only transmits information to users in the nonempty set $\mathcal{U}_k$. Then, as discussed in Section III, we can calculate the minimum total power $P_k^{\min}(\boldsymbol{R}, \boldsymbol{n})$ $(1 \le k \le 2^M - 1)$ required to support a subvector of $\boldsymbol{R}$ composed of the required rates of the users in $\mathcal{U}_k$ under those different spectrum-sharing techniques. For the fading state $\boldsymbol{n}$, let $w_k(\boldsymbol{R}, \boldsymbol{n})$ denote the probability that the broadcast channel transmits information to the subset of users in $\mathcal{U}_k$. Then obviously

$$\sum_{k=1}^{2^M-1} w_k(\boldsymbol{R}, \boldsymbol{n}) = 1 - w_0(\boldsymbol{R}, \boldsymbol{n}) \le 1.$$

For a given rate vector $\boldsymbol{R}$ and a fading state $\boldsymbol{n}$, let $\Pr_i^{\rm on}(\boldsymbol{R}, \boldsymbol{n})$ be the probability that information is sent to User i

$$\Pr_i^{\rm on}(\boldsymbol{R}, \boldsymbol{n}) = \sum_{k=1}^{2^M-1} w_k(\boldsymbol{R}, \boldsymbol{n}) \mathbf{1}[i \in \mathcal{U}_k] \tag{39}$$

where $\mathbf{1}[\cdot]$ denotes the indicator function. Then the average outage probability $\Pr_i(\boldsymbol{R})$ of User i $(1 \le i \le M)$ is

$$\begin{aligned} \Pr_i(\boldsymbol{R}) &= 1 - \Pr_i^{\rm on}(\boldsymbol{R}) \\ &= 1 - E_{\boldsymbol{n}}\left[\Pr_i^{\rm on}(\boldsymbol{R}, \boldsymbol{n})\right]. \end{aligned}$$

For a given fading state $\boldsymbol{n}$, according to (39), the total usage reward $W(\boldsymbol{R}, \boldsymbol{n})$ is

$$\begin{aligned} W(\boldsymbol{R}, \boldsymbol{n}) &= \sum_{i=1}^{M} \mu_i \Pr_i^{\rm on}(\boldsymbol{R}, \boldsymbol{n}) \\ &= \sum_{i=1}^{M} \mu_i \left(\sum_{k=1}^{2^M-1} w_k(\boldsymbol{R}, \boldsymbol{n}) \mathbf{1}[i \in \mathcal{U}_k] \right) \\ &= \sum_{k=1}^{2^M-1} w_k(\boldsymbol{R}, \boldsymbol{n}) \left(\sum_{i=1}^{M} \mu_i \mathbf{1}[i \in \mathcal{U}_k] \right) \\ &= \sum_{k=1}^{2^M-1} w_k(\boldsymbol{R}, \boldsymbol{n}) \eta_k \end{aligned} \tag{40}$$

where the reward for transmitting information to the users in set $\mathcal{U}_k$ is

$$\eta_k \triangleq \sum_{i=1}^{M} \mu_i \mathbf{1}[i \in \mathcal{U}_k]. \tag{41}$$

Thus, the total usage reward averaged over the time-varying channel is

$$W(\boldsymbol{R}) = E_{\boldsymbol{n}}\left[W(\boldsymbol{R}, \boldsymbol{n})\right]. \tag{42}$$

Since in fading state $\boldsymbol{n}$, the total required minimum power to support $\boldsymbol{R}$ with usage probability $\Pr_i^{\rm on}(\boldsymbol{R}, \boldsymbol{n})$ for each user i $(\forall\, 1 \le i \le M)$ is

$$P^{\min}(\boldsymbol{R}, \boldsymbol{n}) = \sum_{k=1}^{2^M-1} P_k^{\min}(\boldsymbol{R}, \boldsymbol{n}) w_k(\boldsymbol{R}, \boldsymbol{n}) \tag{43}$$

the total required minimum average power to achieve $W(\boldsymbol{R})$ will be $E_{\boldsymbol{n}}[P^{\min}(\boldsymbol{R}, \boldsymbol{n})]$.

For a given rate vector $\boldsymbol{R}$, we wish to solve the maximization problem (38), which is equivalent to finding the optimal $w_k^*(\boldsymbol{R}, \boldsymbol{n})$ $(1 \le k \le 2^M - 1, \forall \boldsymbol{n} \in \mathcal{N})$ that maximizes $W(\boldsymbol{R})$ in (42) under the total power constraint. That is, we can rewrite the maximization problem (38) as

$$\begin{cases} \max\limits_{\boldsymbol{w}(\boldsymbol{R}, \boldsymbol{n})} E_{\boldsymbol{n}}[W(\boldsymbol{R}, \boldsymbol{n})] \quad \text{subject to:} \\ E_{\boldsymbol{n}}\left[P^{\min}(\boldsymbol{R}, \boldsymbol{n})\right] \le \overline{P}, \quad \sum\limits_{k=1}^{2^M-1} w_k(\boldsymbol{R}, \boldsymbol{n}) \le 1 \\ \text{and} \quad 0 \le w_k(\boldsymbol{R}, \boldsymbol{n}) \le 1 \end{cases} \tag{44}$$

where

$$\boldsymbol{w}(\boldsymbol{R}, \boldsymbol{n}) = [w_1(\boldsymbol{R}, \boldsymbol{n}), w_2(\boldsymbol{R}, \boldsymbol{n}), \ldots, w_N(\boldsymbol{R}, \boldsymbol{n})]$$

with $N \triangleq 2^M - 1$, $W(\boldsymbol{R}, \boldsymbol{n})$ and $P^{\min}(\boldsymbol{R}, \boldsymbol{n})$ are as given in (40) and (43), respectively, and $\overline{P}$ is the total average transmit power. The maximization problem (44) can be decomposed into the following two problems.

1) Assuming that $\forall \boldsymbol{n} \in \mathcal{N}$, $P(\boldsymbol{n})$ is the total average power assigned to the N sets of users in state $\boldsymbol{n}$, i.e.,

$$P(\boldsymbol{n}) = \sum_{k=1}^{N} w_k(\boldsymbol{R}, \boldsymbol{n}) P_k^{\min}(\boldsymbol{R}, \boldsymbol{n})$$

we must choose $\boldsymbol{w}(\boldsymbol{R}, \boldsymbol{n})$ so that the total usage reward in state $\boldsymbol{n}$ is maximized. That is, we must find

$$\begin{cases} J(P(\boldsymbol{n})) \triangleq \max\limits_{\boldsymbol{w}(\boldsymbol{R}, \boldsymbol{n})} \sum\limits_{k=1}^{N} w_k(\boldsymbol{R}, \boldsymbol{n}) \eta_k \\ \quad \text{subject to:} \\ \sum\limits_{k=1}^{N} w_k(\boldsymbol{R}, \boldsymbol{n}) P_k^{\min}(\boldsymbol{R}, \boldsymbol{n}) \le P(\boldsymbol{n}), \\ \sum\limits_{k=1}^{N} w_k(\boldsymbol{R}, \boldsymbol{n}) \le 1, \ \text{and} \ 0 \le w_k(\boldsymbol{R}, \boldsymbol{n}) \le 1 \end{cases} \tag{45}$$

where η_k is given in (41).

2) After we obtain the expression $J(\cdot)$ by solving (45), the remaining problem is how to assign the total power $P(\boldsymbol{n})$ of the N sets of users for each state $\boldsymbol{n}$ so that the total usage reward averaged over all fading states as expressed in (42) is maximized. That is

$$\begin{cases} \max\limits_{P(\boldsymbol{n})} E_{\boldsymbol{n}}[J(P(\boldsymbol{n}))] - \frac{1}{s} E_{\boldsymbol{n}}[P(\boldsymbol{n})] \\ \text{subject to } E_{\boldsymbol{n}}[P(\boldsymbol{n})] \le \overline{P} \end{cases} \tag{46}$$

where $\frac{1}{s}$ is the Lagrangian multiplier.

We solve the maximization problem (45) by first defining the permutation $\pi(\cdot)$ such that

$$0 < \eta_{\pi(1)} \le \eta_{\pi(2)} \le \cdots \le \eta_{\pi(N)}.$$

For simplicity, we denote the reward and power needed for transmitting information to the users in set $\mathcal{U}_{\pi(i)}$ as λ_i and v_i, respectively, where

$$\begin{cases} \lambda_i \triangleq \eta_{\pi(i)} \\ v_i \triangleq P^{\min}_{\pi(i)}(\boldsymbol{R}, \boldsymbol{n}) \end{cases} \quad \forall 1 \le i \le N. \tag{47}$$

Note that the power v_i's are all functions of rate vector $\boldsymbol{R}$ and fading state $\boldsymbol{n}$. For a given state $\boldsymbol{n}$, $\forall 1 \le k \le N$, if $\exists j$ that satisfies

$$\frac{\lambda_k}{v_k} \le \frac{\lambda_j}{v_j}, \qquad k < j \le N \tag{48}$$

or satisfies

$$\begin{cases} \lambda_j = \lambda_k, \\ \frac{\lambda_j}{v_j} > \frac{\lambda_k}{v_k}, \end{cases} \qquad j = k - 1 \tag{49}$$

then we will get a larger reward by assigning the same power $P(\boldsymbol{n})$ to set $\mathcal{U}_{\pi(j)}$ instead of set $\mathcal{U}_{\pi(k)}$, $\forall P(\boldsymbol{n}) > 0$. Specifically, if (48) is true, since $\lambda_k \le \lambda_j$ for $k < j$, if $v_k \ge v_j$, then obviously by transmitting information to users in set $\mathcal{U}_{\pi(k)}$, we need more power and get less reward than transmitting information to users in set $\mathcal{U}_{\pi(j)}$. If (48) is true and $v_k < v_j$ then, assuming that $w^*_{\pi(k)}(\boldsymbol{R}, \boldsymbol{n}) \ne 0$ $(0 < w^*_{\pi(k)}(\boldsymbol{R}, \boldsymbol{n}) \le 1)$, the reward we get from assigning power v_k to set $\mathcal{U}_{\pi(k)}$ with a fraction $w^*_{\pi(k)}(\boldsymbol{R}, \boldsymbol{n})$ of the transmission time in state $\boldsymbol{n}$ is $\lambda_k w^*_{\pi(k)}(\boldsymbol{R}, \boldsymbol{n})$ and the power needed is $P(\boldsymbol{n}) = v_k w^*_{\pi(k)}(\boldsymbol{R}, \boldsymbol{n})$, while the reward we get from the same power $P(\boldsymbol{n})$ by assigning power v_j to set $\mathcal{U}_{\pi(j)}$ with a fraction

$$\frac{P(\boldsymbol{n})}{v_j} = \frac{v_k w^*_{\pi(k)}(\boldsymbol{R}, \boldsymbol{n})}{v_j}$$

of the transmission time will be $\lambda_j \cdot \frac{v_k w^*_{\pi(k)}(\boldsymbol{R}, \boldsymbol{n})}{v_j}$, which is larger than $\lambda_k w^*_{\pi(k)}(\boldsymbol{R}, \boldsymbol{n})$ by (48). Moreover, the fraction of transmission time needed for set $\mathcal{U}_{\pi(j)}$ is less than that for set $\mathcal{U}_{\pi(k)}$, since when $v_k < v_j$

$$\frac{v_k w^*_{\pi(k)}(\boldsymbol{R}, \boldsymbol{n})}{v_j} < w^*_{\pi(k)}(\boldsymbol{R}, \boldsymbol{n}).$$

If (49) is true, it is obvious that $v_j < v_k$. Thus, to obtain the same reward $\lambda_j = \lambda_k$, the set $\mathcal{U}_{\pi(j)}$ requires less power than the set $\mathcal{U}_{\pi(k)}$. Therefore, in order to get the largest reward under the given power constraint, we do not consider assigning any power to those sets $\mathcal{U}_{\pi(k)}$ for which $\exists j$ satisfying either (48) or (49). That is, we remove them from further consideration and set $w^*_{\pi(k)}(\boldsymbol{R}, \boldsymbol{n}) = 0$.

For example, when $N = 3$ and the relative values of v_1, v_2, v_3, and λ_1, λ_2, λ_3 are as shown in Fig. 1, where λ_i and v_i correspond to the reward and power needed for set $\mathcal{U}_{\pi(i)}$ $(i = 1, 2, 3)$, respectively. It is obvious that $\frac{\lambda_1}{v_1} < \frac{\lambda_2}{v_2}$. Thus, if the available power $P(\boldsymbol{n}) \le v_1$ and we assign it to set $\mathcal{U}_{\pi(1)}$, the reward we can get is the straight line OA; if $v_1 < P(\boldsymbol{n}) < v_2$ and we assign it to sets $\mathcal{U}_{\pi(1)}$ and $\mathcal{U}_{\pi(2)}$ by time sharing, the reward we can get is the straight line AB. However, in both cases, if we assign the power $P(\boldsymbol{n})$ to set $\mathcal{U}_{\pi(2)}$, we get a larger reward which is indicated by the straight line OB. Thus, no power should be assigned to set $\mathcal{U}_{\pi(1)}$ and $J(P(\boldsymbol{n}))$ defined in (45) is as shown by the solid curve in Fig. 1.

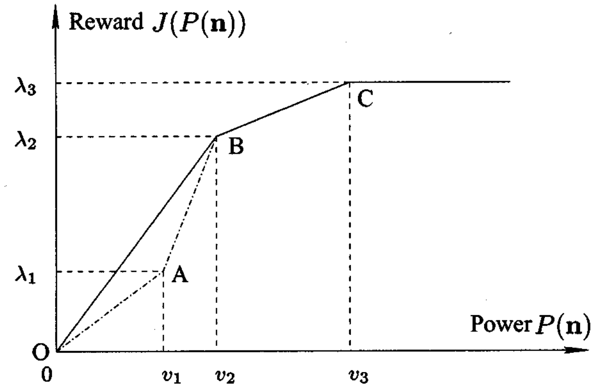

Fig. 1. Power $P(\boldsymbol{n})$ versus reward $J(P(\boldsymbol{n}))$ for $N = 3$.

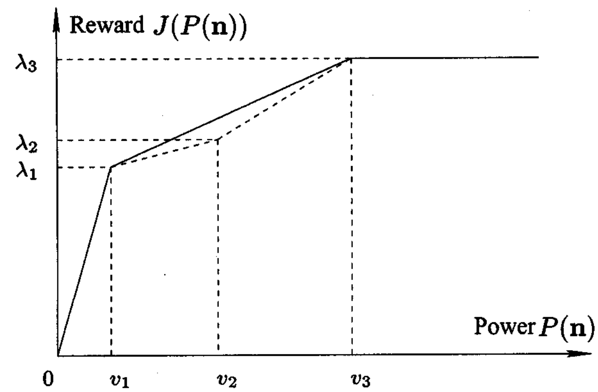

Fig. 2. Power $P(\boldsymbol{n})$ versus reward $J(P(\boldsymbol{n}))$ for three remaining sets of users.

Generally, for the remaining sets of users, it is possible that there are still some sets $\mathcal{U}_{\pi(k)}$ to which no power should be assigned in order to get the largest reward. For example, assume that the remaining sets are $\mathcal{U}_{\pi(1)}, \mathcal{U}_{\pi(2)}$, and $\mathcal{U}_{\pi(3)}$, and the relative values of v_1, v_2, v_3, and λ_1, λ_2, λ_3 are as shown in Fig. 2. From this figure we see that neither (48) nor (49) is satisfied for any $k = 1, 2, 3$, since $\lambda_1 < \lambda_2 < \lambda_3$ and

$$\frac{\lambda_1}{v_1} > \frac{\lambda_2}{v_2} > \frac{\lambda_3}{v_3}.$$

However, it is obvious that no power should be assigned to set $\mathcal{U}_{\pi(2)}$, because a larger reward can be obtained when the same power is time-shared by sets $\mathcal{U}_{\pi(1)}$ and $\mathcal{U}_{\pi(3)}$ instead of by sets $\mathcal{U}_{\pi(1)}$ and $\mathcal{U}_{\pi(2)}$, or by sets $\mathcal{U}_{\pi(2)}$ and $\mathcal{U}_{\pi(3)}$.

In the following, we use an iterative procedure to find all the sets $\mathcal{U}_{\pi(k)}$ in the remaining sets that should be assigned no power and remove them from further consideration [i.e., let $w^*_{\pi(k)}(\boldsymbol{R}, \boldsymbol{n}) = 0$]. An interpretation of this procedure based on Fig. 2 will be given shortly.

Initialization: Let $m = 1$.

Step 1) Denote the number of remaining sets as G_u and let the permutation $\rho(\cdot)$ be defined such that for the remaining G_u sets

$$\lambda_{\rho(1)} < \lambda_{\rho(2)} < \cdots < \lambda_{\rho(G_u)}. \tag{50}$$

Due to the removal criterion, it must be true that

$$\frac{\lambda_{\rho(1)}}{v_{\rho(1)}} > \frac{\lambda_{\rho(2)}}{v_{\rho(2)}} > \cdots > \frac{\lambda_{\rho(G_u)}}{v_{\rho(G_u)}}. \tag{51}$$

Step 2) Let $\rho_0(m) = \rho(1)$ and $z_{\rho_0(m)} = \frac{\lambda_{\rho(1)}}{v_{\rho(1)}}$. If $G_u \leq 2$, all the sets that should be assigned no power have been removed and the procedure terminates; if $G_u > 2$, go to Step 3).

Step 3) For $2 \leq k \leq G_u$, decrease $\lambda_{\rho(k)}$ and $v_{\rho(k)}$ by $\lambda_{\rho(1)}$ and $v_{\rho(1)}$, respectively. Do not assign any power to those sets of users $\mathcal{U}_{\pi[\rho(k)]}$ for which $\exists\, j$ that satisfies

$$\frac{\lambda_{\rho(k)}}{v_{\rho(k)}} \leq \frac{\lambda_{\rho(j)}}{v_{\rho(j)}}, \qquad k < j \leq G_u$$

and remove them from further consideration [i.e., let $w^*_{\pi[\rho(k)]}(\boldsymbol{R}, \boldsymbol{n}) = 0$]. Also remove set $\mathcal{U}_{\pi[\rho(1)]}$. Increase m by 1 and return to Step 1).

In this procedure, we observe that in the first iteration, from (50) and (51) it is clear that

$$v_{\rho(1)} < v_{\rho(2)} < \cdots < v_{\rho(G_u)}.$$

Since $\frac{\lambda_{\rho(1)}}{v_{\rho(1)}} > \frac{\lambda_{\rho(i)}}{v_{\rho(i)}}$ for all $1 < i \leq G_u$, when the total average power $P(\boldsymbol{n}) \leq v_{\rho(1)}$, we get the largest reward $\frac{\lambda_{\rho(1)}}{v_{\rho(1)}} \cdot P(\boldsymbol{n})$ by assigning it to set $\mathcal{U}_{\pi[\rho(1)]}$. That is, in (45), as shown in the example in Fig. 2[6]

$$J(P(\boldsymbol{n})) = \frac{\lambda_{\rho(1)}}{v_{\rho(1)}} \cdot P(\boldsymbol{n}), \qquad \text{if } P(\boldsymbol{n}) \leq v_{\rho(1)}$$

and we store the index $\rho(1)$ of set $\mathcal{U}_{\pi[\rho(1)]}$ and the tangent $\frac{\lambda_{\rho(1)}}{v_{\rho(1)}}$ in $\rho_0(1)$ and $z_{\rho_0(1)}$, respectively. Next we wish to identify those sets $\mathcal{U}_{\pi[\rho(k)]}$ $(1 < k \leq G_u)$ for which $\exists\, j$ that satisfies

$$\frac{\lambda_{\rho(k)} - \lambda_{\rho(1)}}{v_{\rho(k)} - v_{\rho(1)}} < \frac{\lambda_{\rho(j)} - \lambda_{\rho(1)}}{v_{\rho(j)} - v_{\rho(1)}}, \qquad k < j \leq G_u \tag{52}$$

and we do not assign any power to them since, for $k < j$, $v_{\rho(k)} < v_{\rho(j)}$, $\lambda_{\rho(k)} < \lambda_{\rho(j)}$, and if (52) is true, on the Power-Reward plane as shown in Fig. 2, point $(v_{\rho(k)}, \lambda_{\rho(k)})$ will be under the straight line formed by connecting point $(v_{\rho(1)}, \lambda_{\rho(1)})$ and point $(v_{\rho(j)}, \lambda_{\rho(j)})$.[7] Therefore, in (45)

$$J(P(\boldsymbol{n})) = \frac{\lambda_{\rho(i^*)} - \lambda_{\rho(1)}}{v_{\rho(i^*)} - v_{\rho(1)}} \times [P(\boldsymbol{n}) - v_{\rho(1)}], \qquad \text{if } v_{\rho(1)} < P(\boldsymbol{n}) \leq v_{\rho(i^*)}$$

where the index i^* is given by

$$i^* = \arg \max_{1 < i \leq G_u} \left\{ \frac{\lambda_{\rho(i)} - \lambda_{\rho(1)}}{v_{\rho(i)} - v_{\rho(1)}} \right\}. \tag{53}$$

After removing those sets $\mathcal{U}_{\pi[\rho(k)]}$ for which (52) holds [i.e., let $w^*_{\pi[\rho(k)]}(\boldsymbol{R}, \boldsymbol{n}) = 0$] and also removing set $\mathcal{U}_{\pi[\rho(1)]}$, the index $\rho(i^*)$ in the first iteration becomes $\rho(1)$ of a new permutation $\rho(\cdot)$ in the second iteration [otherwise, (53) cannot be true] and is stored in $\rho_0(2)$, and the corresponding tangent $\frac{\lambda_{\rho(i^*)} - \lambda_{\rho(1)}}{v_{\rho(i^*)} - v_{\rho(1)}}$ is stored in $z_{\rho_0(2)}$. In the second iteration, similarly, the new index $\rho(i^*)$ that satisfies (53) will be identified, and new set(s) $\mathcal{U}_{\pi[\rho(k)]}$ for which (52) holds will be removed, since the point(s) $(v_{\rho(k)}, \lambda_{\rho(k)})$ will be under the straight line formed by connecting points $(v_{\rho(1)}, \lambda_{\rho(1)})$ and $(v_{\rho(j))}, \lambda_{\rho(j)})$, where j satisfies (52). The new index $\rho(i^*)$ will become $\rho(1)$ in the third iteration and be stored in $\rho_0(3)$, and the corresponding tangent $\frac{\lambda_{\rho(i^*)} - \lambda_{\rho(1)}}{v_{\rho(i^*)} - v_{\rho(1)}}$ will be stored in $z_{\rho_0(3)}$. The iterative procedure continues until all the sets that should be assigned no power have been removed and the curve of $J(P(\boldsymbol{n}))$ in (45) is obtained by connecting the origin $(0, 0)$ and points $(v_{\rho_0(1)}, \lambda_{\rho_0(1)})$–$(v_{\rho_0(m_0)}, \lambda_{\rho_0(m_0)})$, where m_0 is the value of m when the iteration stops $(1 \leq m_0 \leq N)$. That is,

$$J(P(\boldsymbol{n})) = \begin{cases} z_{\rho_0(1)} \cdot P(\boldsymbol{n}), & 0 < P(\boldsymbol{n}) \leq v_{\rho_0(1)} \\ \lambda_{\rho_0(m_0)}, & P(\boldsymbol{n}) > v_{\rho_0(m_0)} \\ z_{\rho_0(j)} \cdot [P(\boldsymbol{n}) - v_{\rho_0(j-1)}], & v_{\rho_0(j-1)} < P(\boldsymbol{n}) \leq v_{\rho_0(j)} \\ & \text{for some } j,\ 1 < j \leq m_0. \end{cases} \tag{54}$$

Note that from the iterative procedure, it is clear that

$$\lambda_{\rho_0(1)} < \lambda_{\rho_0(2)} < \cdots < \lambda_{\rho_0(m_0)} \tag{55}$$

$$v_{\rho_0(1)} < v_{\rho_0(2)} < \cdots < v_{\rho_0(m_0)} \tag{56}$$

$$z_{\rho_0(1)} > z_{\rho_0(2)} > \cdots > z_{\rho_0(m_0)} \tag{57}$$

where

$$z_{\rho_0(j)} = \begin{cases} \frac{\lambda_{\rho_0(1)}}{v_{\rho_0(1)}}, & j = 1 \\ \frac{\lambda_{\rho_0(j)} - \lambda_{\rho_0(j-1)}}{v_{\rho_0(j)} - v_{\rho_0(j-1)}}, & 1 < j \leq m_0. \end{cases} \tag{58}$$

For example, in Fig. 2, if we execute the above three-step procedure, we will have $m_0 = 2$, $\rho_0(1) = 1$, $\rho_0(2) = 3$, $z_{\rho_0(1)} = \frac{\lambda_1}{v_1}$, and $z_{\rho_0(2)} = \frac{\lambda_3 - \lambda_1}{v_3 - v_1}$. Therefore, from (54) we obtain

$$J(P(\boldsymbol{n})) = \begin{cases} \frac{\lambda_1}{v_1} \cdot P(\boldsymbol{n}), & 0 < P(\boldsymbol{n}) \leq v_1 \\ \frac{\lambda_3 - \lambda_1}{v_3 - v_1} \cdot [P(\boldsymbol{n}) - v_1], & v_1 < P(\boldsymbol{n}) \leq v_3 \\ \lambda_3, & P(\boldsymbol{n}) > v_3 \end{cases}$$

which is exactly as shown by the solid curve in Fig. 2.

Once the curve $J(P(\boldsymbol{n}))$ is obtained, from (46) we know that $\forall\, \frac{1}{s} > 0$ fixed, the optimal power $P^*(\boldsymbol{n})$ satisfies $J'(P^*(\boldsymbol{n})) = \frac{1}{s}$ if the tangent of $J(P(\boldsymbol{n}))$ is continuous. However, in our case, the tangent of $J(P(\boldsymbol{n}))$ is discrete and $P^*(\boldsymbol{n})$ cannot be determined directly. Therefore, we will use the following theorem to find the optimal $w^*_{\pi[\rho_0(j)]}(\boldsymbol{R}, \boldsymbol{n})$ $(1 \leq j \leq m_0)$ for the remaining m_0 sets $\{\mathcal{U}_{\pi[\rho_0(j)]}\}_{j=1}^{m_0}$.

Before stating the theorem, we first define some additional notations and parameters. In (55)–(58), the indexes $\{\rho_0(i)\}_{i=1}^{m_0}$ are all functions of $\boldsymbol{n}$. Therefore, we will refer to them as $\{\rho_0(i, \boldsymbol{n})\}_{i=1}^{m_0}$ and for simplicity, $\forall\, 1 \leq i \leq m_0$, we denote[8]

$$\begin{cases} \lambda_i(\boldsymbol{n}) \triangleq \lambda_{\rho_0(i)} \\ v_i(\boldsymbol{n}) \triangleq v_{\rho_0(i)} \\ z_i(\boldsymbol{n}) \triangleq z_{\rho_0(i)}. \end{cases} \tag{59}$$

[6]Note that in this example, $G_u = 3$ and $\rho(i) = i$, $i = 1, 2, 3$.

[7]In Fig. 2, point (v_2, λ_2) is under the straight line formed by connecting point (v_1, λ_1) and point (v_3, λ_3).

[8]Note that $\{\rho_0(i)\}_{i=1}^{m_0}$ are also functions of the given rate vector $\boldsymbol{R}$. However, we omit the explicit dependence on $\boldsymbol{R}$ in their notations for simplicity.

Thus, (55)–(58) become

$$\lambda_1(\boldsymbol{n}) < \lambda_2(\boldsymbol{n}) < \cdots < \lambda_{m_0}(\boldsymbol{n}) \tag{60}$$

$$v_1(\boldsymbol{n}) < v_2(\boldsymbol{n}) < \cdots < v_{m_0}(\boldsymbol{n}) \tag{61}$$

$$z_1(\boldsymbol{n}) > z_2(\boldsymbol{n}) > \cdots > z_{m_0}(\boldsymbol{n}) \tag{62}$$

and

$$z_j(\boldsymbol{n}) = \begin{cases} \frac{\lambda_1(\boldsymbol{n})}{v_1\boldsymbol{n})}, & j=1 \\ \frac{\lambda_j(\boldsymbol{n})-\lambda_{j-1}(\boldsymbol{n})}{v_j(\boldsymbol{n})-v_{j-1}(\boldsymbol{n})}, & 1 < j \le m_0. \end{cases} \tag{63}$$

Moreover, we define Ω_{m_0} as the set of all fading states $\boldsymbol{n}$ for which the final value of the loop parameter m is m_0 $(1 \le m_0 \le N)$ when the three-step iteration procedure terminates. By denoting $z_{m_0+1}(\boldsymbol{n}) \triangleq 0, \forall s > 0$, for $j = 1, 2, \ldots, m_0$, we define sets $L_j(m_0, s)$ and $\tilde{L}_j(m_0, s)$ as

$$L_j(m_0, s) = \left\{\boldsymbol{n}: \boldsymbol{n} \in \Omega_{m_0}, \, z_j(\boldsymbol{n}) > \frac{1}{s} > z_{j+1}(\boldsymbol{n})\right\}$$

$$\tilde{L}_j(m_0, s) = \left\{\boldsymbol{n}: \boldsymbol{n} \in \Omega_{m_0}, \, \frac{1}{s} = z_j(\boldsymbol{n})\right\}.$$

Thus, $\forall s > 0$

$$\mathcal{N} = \bigcup_{1 \le m_0 \le N} \Omega_{m_0} = \bigcup_{1 \le m_0 \le N} \bigcup_{1 \le j \le m_0} \left(L_j(m_0, s) \cup \tilde{L}_j(m_0, s)\right)$$

where $\mathcal{N}$ is the set of all possible fading states. For a given total average power constraint $\overline{P} > 0$, define s^* as

$$s^* \triangleq \sup\{s: P(s) < \overline{P}\}$$

where

$$P(s) \triangleq \sum_{m_0=1}^{N} \sum_{j=1}^{m_0} \int_{\boldsymbol{n} \in L_j(m_0, s)} v_j(\boldsymbol{n}) \, dF(\boldsymbol{n}).$$

Therefore, $\forall \boldsymbol{n} \in \Omega_{m_0}$, $1 \le m_0 \le N$, finding the optimal $w^*_{\pi[\rho_0(j)]}(\boldsymbol{R}, \boldsymbol{n})$ $(1 \le j \le m_0)$ in (44) for the remaining m_0 sets $\{\mathcal{U}_{\pi[\rho_0(j)]}\}_{j=1}^{m_0}$ after the iterative procedure is equivalent to solving the maximization problem

$$\begin{cases} \max_{\boldsymbol{u}(\boldsymbol{n})} \sum_{m_0=1}^{N} E_{\boldsymbol{n} \in \Omega_{m_0}} \left[\sum_{j=1}^{m_0} \lambda_j(\boldsymbol{n}) u_j(\boldsymbol{n})\right] \\ \quad \text{subject to:} \\ \sum_{m_0=1}^{N} E_{\boldsymbol{n} \in \Omega_{m_0}} \left[\sum_{j=1}^{m_0} v_j(\boldsymbol{n}) u_j(\boldsymbol{n})\right] \le \overline{P}, \\ 0 \le \sum_{j=1}^{m_0} u_j(\boldsymbol{n}) \le 1, \text{ and } 0 \le u_j(\boldsymbol{n}) \le 1 \end{cases} \tag{64}$$

where $\boldsymbol{u}(\boldsymbol{n}) = [u_1(\boldsymbol{n}), u_2(\boldsymbol{n}), \ldots, u_{m_0}(\boldsymbol{n})]$.

Theorem 3: $\forall \boldsymbol{n} \in \Omega_{m_0}$, $1 \le m_0 \le N$, by denoting $u_j^*(\boldsymbol{n})$ $(1 \le j \le m_0)$ as the solution to the maximization problem (64), we have

a) if $\frac{1}{s^*} > z_1(\boldsymbol{n})$, then $u_j^*(\boldsymbol{n}) = 0, \forall 1 \le j \le m_0$;

b) if $\exists j \in \{1, 2, \ldots, m_0\}$, $z_j(\boldsymbol{n}) > \frac{1}{s^*} > z_{j+1}(\boldsymbol{n})$, then $u_j^*(\boldsymbol{n}) = 1$, $u_i^*(\boldsymbol{n}) = 0, \forall i \ne j, 1 \le i \le m_0$;

c) if $\exists j \in \{1, 2, \ldots, m_0\}$, $\frac{1}{s^*} = z_j(\boldsymbol{n})$, then $u_j^*(\boldsymbol{n}) = \tau^*$, $u_{j-1}^*(\boldsymbol{n}) = 1 - \tau^*$, $u_i^*(\boldsymbol{n}) = 0, \forall i \ne j, j-1, 1 \le i \le m_0$, where τ^* satisfies

$$\sum_{m_0=1}^{N} \sum_{j=1}^{m_0} [v_j(\boldsymbol{n})\tau^* + v_{j-1}(\boldsymbol{n})(1-\tau^*)] \Pr\left\{\tilde{L}_j(m_0, s^*)\right\} + P(s^*) = \overline{P}.$$

Proof: See the Appendix, Section E. □

Note that this theorem is a generalization of [6, Lemma 3], which corresponds to $N = 1$. Therefore, $\forall \boldsymbol{n} \in \Omega_{m_0}$ $(1 \le m_0 \le N)$, given the remaining m_0 sets $\{\mathcal{U}_{\pi[\rho_0(j)]}\}_{j=1}^{m_0}$ after the iterative procedure, Theorem 3 determines which set(s) of users should be chosen for transmission by solving (64), since after removing those users to which no power should be assigned, the maximization problems (44) and (64) are equivalent and

$$w^*_{\pi[\rho_0(j)]}(\boldsymbol{R}, \boldsymbol{n}) = u_j^*(\boldsymbol{n}), \qquad 1 \le j \le m_0.$$

In particular, the theorem indicates that, based on the total power constraint, there is a threshold power level s^* which is important in determining the optimal set(s) of users. Moreover, in fading states of set $\tilde{L}_j(m_0, s)$ $(\forall 1 \le m_0 \le N, \forall 1 \le j \le m_0)$, at most two sets of users are chosen and the information for the selected two sets are sent by time-sharing the channel. In each of the other fading states, at most one set of users is chosen. Therefore, if the cdf $F(\boldsymbol{n})$ is continuous, with probability 1, at most one set of users is chosen in each state, since $\forall 1 \le m_0 \le N, \forall 1 \le j \le m_0$, $\Pr[\tilde{L}_j(m_0, s)] = 0$. If $F(\boldsymbol{n})$ is discontinuous, $\Pr[\tilde{L}_j(m_0, s)]$ may be larger than zero for some j and m_0, and the probability that two sets of users are chosen in some fading states may be larger than zero.

C. Multimedia Outage Probability Region

In an M-user broadcast system, some users may require constant-rate transmission without any outage (e.g., voice users), while other users allow certain outages in the transmission of their information (e.g., data users). Let M_0 be the number of those users allowing no outage. Then, the M-user outage probability region contracts to an $(M - M_0)$-user outage probability region. Since in each fading state, the channel can be used for $\sum_{i=1}^{M-M_0} \binom{M-M_0}{i} = 2^{M-M_0} - 1$ different sets of users, by applying the same optimal strategy discussed in Section IV-B, we can obtain the boundary of the outage probability region for the $M - M_0$ users.

For example, in a two-user system where one user (say, User 1) allows some outage and the other user (say, User 2) requires no outage $(M = 2, M_0 = 1)$, the minimum outage probability problem for User 1 is a modified threshold-decision rule similar to that of the single-user case.

For each joint fading state $\boldsymbol{n} = (n_1, n_2)$ and a given rate vector $\boldsymbol{R} = (R_1, R_2)$, if the information for User 1 is not transmitted, we denote the minimum required total power as $P_{\text{off}}(\boldsymbol{R}, \boldsymbol{n})$ and it is just the power needed to support rate R_2 for User 2; if the information for User 1 is transmitted, we denote the minimum required total power as $P_{\text{on}}(\boldsymbol{R}, \boldsymbol{n})$ and it is $P^{\min}(\boldsymbol{R}, \boldsymbol{n})$ given in (3), (11), and (16) for CD with or without successive decoding and for TD, respectively. Let $\overline{P}$ be the total average power and assume that in fading state $\boldsymbol{n}$, the channel transmits the information for User 1 with probability $w(\boldsymbol{n}, \boldsymbol{R})$ and for User 2 with probability 1 (no outage). The maximization problem

$$\begin{cases} \max E_{\boldsymbol{n}}[w(\boldsymbol{R}, \boldsymbol{n})] \quad \text{subject to:} \\ E_{\boldsymbol{n}}\{P_{\rm on}(\boldsymbol{R}, \boldsymbol{n})w(\boldsymbol{R}, \boldsymbol{n}) + P_{\rm off}(\boldsymbol{R}, \boldsymbol{n})[1 - w(\boldsymbol{R}, \boldsymbol{n})]\} = \overline{P} \end{cases} \tag{65}$$

is equivalent to

$$\begin{cases} \max E_{\boldsymbol{n}}[w(\boldsymbol{R}, \boldsymbol{n})] \quad \text{subject to:} \\ E_{\boldsymbol{n}}\{[P_{\rm on}(\boldsymbol{R}, \boldsymbol{n}) - P_{\rm off}(\boldsymbol{R}, \boldsymbol{n})]w(\boldsymbol{R}, \boldsymbol{n})\} \\ \qquad\qquad = \overline{P} - E_{\boldsymbol{n}}\{P_{\rm off}(\boldsymbol{R}, \boldsymbol{n})\}. \end{cases}$$

Therefore, by substituting $P^{\min}(\boldsymbol{R}, \boldsymbol{n})$ with $P_{\rm on}(\boldsymbol{R}, \boldsymbol{n}) - P_{\rm off}(\boldsymbol{R}, \boldsymbol{n})$ and $\overline{P}$ with $\overline{P} - E_{\boldsymbol{n}}[P_{\rm off}(\boldsymbol{R}, \boldsymbol{n})]$ into the definitions of $\mathcal{R}(s)$, $\tilde{\mathcal{R}}(s)$, $P(s)$, $\tilde{P}(s)$, s^*, and w^* in (31)–(36), we obtain the solution to (65): if $\boldsymbol{n} \notin \tilde{\mathcal{R}}(s^*)$, $w(\boldsymbol{R}, \boldsymbol{n}) = 0$, i.e., only the information of User 2 is transmitted; if $\boldsymbol{n} \in \mathcal{R}(s^*)$, $w(\boldsymbol{R}, \boldsymbol{n}) = 1$, i.e., the information of User 1 is transmitted together with that of User 2; if $\boldsymbol{n} \notin \mathcal{R}(s^*)$ but $\boldsymbol{n} \in \tilde{\mathcal{R}}(s^*)$, then $w(\boldsymbol{R}, \boldsymbol{n}) = w^*$, i.e., with probability w^*, the information of User 1 is transmitted with that of User 2. The minimum outage probability $\Pr_{\min}(\overline{P}, \boldsymbol{R})$ for User 1 is as given in (37).

V. Frequency-Selective Fading Channels

In previous sections, we have derived implicitly the zero-outage capacity region and the outage capacity region of a flat-fading broadcast channel. The zero-outage capacity and the ergodic capacity, discussed in Part I of this paper, characterize two very different aspects of a fading channel. That is, under a given average power constraint, the ergodic capacity is the maximum average rate over all fading states with no delay constraint; the zero-outage capacity, on the other hand, is the maximum common rate that can be achieved in every fading state with the given delay constraint satisfied. In order to extend the concepts of the zero-outage capacity and the outage capacity of a narrow-band fading channel discussed in previous sections to that of a frequency-selective wide-band fading channel, we first relax the delay-limited requirement and consider a multiple time-scale fading channel characterized by both fast fading (e.g., due to multipath) and slow fading (e.g., due to shadowing) [5]. For this channel, assuming that the fast fading is fast enough to average out over the tolerable delay, we define the zero-outage capacity with respect to the slow fading as the maximum common rate over all subsets of fading states, where each subset is associated with a slow-fading state. Within each subset, dynamic rate allocation is allowed for the fast-fading states under a given common average rate constraint for each subset.

Specifically, for an M-user fading broadcast channel, let $\mathcal{S}$ be the set of all joint slow states of the M users, and $\mathcal{N}$ the set of all joint fast-fading states. Let $(\boldsymbol{s}, \boldsymbol{n})$ be a joint slow- and fast-fading state, with $\boldsymbol{n}$ having stationary distribution $\mathcal{Q}$. When conditional on a given slow state $\boldsymbol{s}$, $\boldsymbol{n}$ has stationary distribution $\mathcal{Q}_{\boldsymbol{s}}$. For a given power allocation policy $\mathcal{P}$, let $P_j(\boldsymbol{s}, \boldsymbol{n})$ denote the power assigned to User j $(1 \le j \le M)$ in a joint state $(\boldsymbol{s}, \boldsymbol{n})$. Therefore, for CD with successive decoding, it can be similarly shown as for Theorem 1 that the zero-outage capacity region with respect to slow fading is

$$\mathcal{C}_{\rm zero}(\overline{P}) = \bigcup_{\mathcal{P} \in \mathcal{F}} \bigcap_{\boldsymbol{s} \in \mathcal{S}} \mathcal{C}_{\rm CD}(\boldsymbol{s}, \mathcal{P})$$

where $\mathcal{C}_{\rm CD}(\boldsymbol{s}, \mathcal{P})$ is the achievable average rate (averaged over the fast-fading states) region for the given slow-fading state $\boldsymbol{s}$ and power allocation policy $\mathcal{P}$, and $\mathcal{F}$ is the set of all power policies satisfying $E_{\boldsymbol{s},\boldsymbol{n}}\left[\sum_{j=1}^{M} P_j(\boldsymbol{s}, \boldsymbol{n})\right] \le \overline{P}$. We obtain region $\mathcal{C}_{\rm zero}(\overline{P})$ implicitly. That is, for each given rate vector $\tilde{\boldsymbol{R}}$, if there exists a power allocation policy $\mathcal{P} \in \mathcal{F}$ such that the average rate vector for any slow-fading state $\boldsymbol{s}$ is $\tilde{\boldsymbol{R}}$, then $\tilde{\boldsymbol{R}} \in \mathcal{C}_{\rm zero}(\overline{P})$. Therefore, in order to determine whether $\tilde{\boldsymbol{R}} \in \mathcal{C}_{\rm zero}(\overline{P})$ or not, we have to compute the minimum average total power $P^{\min}(\boldsymbol{s})$ required for the M users to support $\tilde{\boldsymbol{R}}$ in each slow state $\boldsymbol{s}$, i.e., solving the minimization problem

$$\min \tilde{P} \quad \text{subject to:} \quad \tilde{\boldsymbol{R}} \in \mathcal{C}(\boldsymbol{s}, \tilde{P}) \tag{66}$$

where $\tilde{P}$ denotes the total average power of the M users in the slow state $\boldsymbol{s}$, and $\mathcal{C}(\boldsymbol{s}, \tilde{P})$ is the ergodic capacity region for the slow state $\boldsymbol{s}$ under the power constraint $\tilde{P}$. That is

$$\mathcal{C}(\boldsymbol{s}, \tilde{P}) = \bigcup_{\mathcal{P} \in \mathcal{F}_{\boldsymbol{s}}} \left\{ \boldsymbol{R}: R_j \le E_{\boldsymbol{n}|\boldsymbol{s}}\left[B \log\left(1 + \frac{P_j(\boldsymbol{s}, \boldsymbol{n})}{n_j B + \sum_{i=1}^{M} P_i(\boldsymbol{s}, \boldsymbol{n}) \mathbf{1}[n_j > n_i]}\right)\right],\ j = 1, 2, \ldots, M \right\}$$

where $\mathcal{F}_{\boldsymbol{s}}$ is the set of all power policies satisfying $E_{\boldsymbol{n}|\boldsymbol{s}}\left[\sum_{j=1}^{M} P_j(\boldsymbol{s}, \boldsymbol{n})\right] \le \tilde{P}$.

As shown in [4], the minimization problem in (66) is equivalent to

$$\min_{(\boldsymbol{R}, \tilde{P})} \left[\tilde{P} - \boldsymbol{\mu} \cdot \boldsymbol{R}\right] \quad \text{subject to:} \quad \boldsymbol{R} \in \mathcal{C}(\boldsymbol{s}, \tilde{P})$$

where $\boldsymbol{\mu}$ is the Lagrange multiplier vector (rate reward vector) chosen such that the target rate vector $\tilde{\boldsymbol{R}}$ is met. For a given slow state $\boldsymbol{s}$, from Part I of this paper we know that given a rate reward vector $\boldsymbol{\mu}$ and a water-filling power level $s^* = \frac{1}{\lambda}$, the optimal power allocation strategy is determined and we can calculate the required total average power $\tilde{P}$ of the M users and obtain the corresponding boundary vector $\boldsymbol{R}$ of region $\mathcal{C}(\boldsymbol{s}, \tilde{P})$. By fixing $\lambda = 1$, the following algorithm is proposed in [4] to find the appropriate $\boldsymbol{\mu}$ such that $\tilde{\boldsymbol{R}}$ is met:

Algorithm 1 [4]: Start the iteration at $\boldsymbol{\mu}(0) = \mathbf{0}$. Given the nth iteration $\boldsymbol{\mu}(n)$, the $(n+1)$th iteration $\boldsymbol{\mu}(n+1)$ is given by the following: for each j, $\mu_j(n+1)$ is a rate reward for the jth user such that $R_j(\boldsymbol{\mu}) = \tilde{R}_j$, when the rate rewards of the other users remain fixed at $\boldsymbol{\mu}(n)$ while the reward for the jth user is adjusted.

Therefore, given the target rate vector $\tilde{\boldsymbol{R}}$, once the appropriate rate reward vector $\boldsymbol{\mu}$ is determined, we can obtain the minimum average total power $P^{\min}(\boldsymbol{s})$ in (66) easily. Thus, if $E_{\boldsymbol{s}}[P^{\min}(\boldsymbol{s})] \leq \overline{P}$, then $\tilde{\boldsymbol{R}} \in \mathcal{C}_{\text{zero}}(\overline{P})$. Otherwise, $\tilde{\boldsymbol{R}} \notin \mathcal{C}_{\text{zero}}(\overline{P})$. In addition, if we allow some transmission outage in certain slow states, given the average outage probability constraint for each user, we can also determine whether rate vector $\tilde{\boldsymbol{R}}$ is inside the outage capacity region (with respect to slow states) or not by using the techniques developed in Section IV.

Note that similar reasoning as that for CD with successive decoding can be applied to TD, FD, and CD without successive decoding as well. Since TD, FD, and CD without successive decoding have the same ergodic capacity region for a given slow state, it is obvious that they will also have the same zero-outage capacity region and outage capacity region when these capacity regions are defined with respect to slow states. This is quite different from that of the narrow-band fading channel discussed in the previous sections.

By applying the ideas developed for the time-varying channel characterized by both fast fading and slow fading, we can easily obtain the zero-outage capacity region and the outage capacity region of the frequency-selective fading broadcast channel discussed in Part I of this paper. Since the frequency-selective fading channel can be viewed as a time-varying channel, where the frequency responses of the M users are different in different fading states, we can define the zero-outage capacity and the outage capacity with respect to the fading states. That is, while a common average rate is maintained for all fading states, dynamic rate allocation is allowed for each user at different frequencies in each fading state. The resulting optimization problem is identical to the one studied for the time-varying channel characterized by both fast fading and slow fading, and so is the optimal dynamic power allocation. Therefore, we see that wide-band systems provide the possibility of performing dynamic power allocation over different frequencies in addition to over different fading states, which is an advantage over narrow-band systems.

VI. Numerical Results

In this section, we present numerical results for zero-outage capacity regions, outage capacity regions, and outage probability regions of narrow-band fading broadcast channels under different spectrum-sharing techniques. The Nakagami-m flat-fading model is used for its mathematical tractability. The total average transmit power in the figures below is denoted as $\overline{P}$, and the average noise density of the ith subchannel is denoted as $\overline{n}_i$, $i = 1, 2$. We refer to the CD without successive decoding technique as CDWO. Since TD and FD are equivalent in the sense that they have the same capacity region of any kind, all results for TD in the figures also apply for FD.

In Fig. 3, the two-user zero-outage capacity region for the Nakagami-m fading broadcast channel is shown for $m = 2, 3, 4$, and ∞. The SNR difference between the two users is 20 dB and the total average power $\overline{P} = 25$ dB. Similar to the ergodic (Shannon) capacity region comparison in Part I of this paper, optimal CD results in a much larger zero-outage capacity region than optimal TD. But the zero-outage capacity

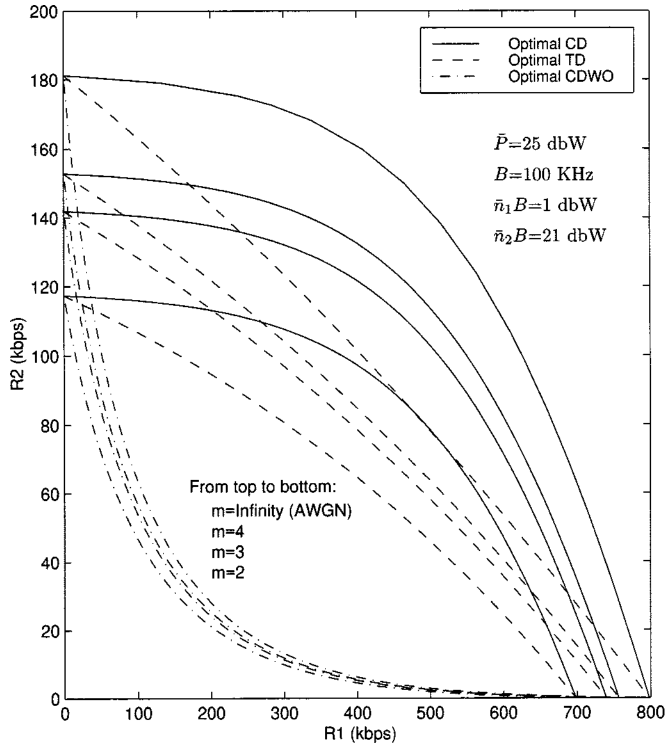

Fig. 3. Two-user zero-outage capacity region in Nakagami fading: 20-dB SNR difference.

region of optimal TD is now much larger than that of the optimal CDWO,[9] the boundary of which is convex. Note that the zero-outage capacity region increases as m increases for all of the three types of spectrum-sharing techniques, since smaller m corresponds to more severe fading. However, unlike using optimal CD or TD, the capacity region using CDWO does not increase much with the increase of m. Also note that for the Rayleigh fading channel $(m = 1)$, the zero-outage capacity region is zero, which is why it is not shown. When $m \to \infty$, the Nakagami-m fading channel approaches the Gaussian channel and as proved in Section III-D, the limiting zero-outage capacity region of the Nakagami-m fading channel is the same as that of the AWGN channel for CD or CDWO.

Fig. 4 shows the case where the SNR difference between the two users is 3 dB and the total average power is 10 dB. Since the SNR difference between the two users is relatively small, the differences of the zero-outage capacity region between using CD, TD, or CDWO is not so dramatic as in the previous case. When m increases, the capacity region of CDWO now increases faster than it does in Fig. 3.

In Figs. 5–11 , as discussed in Section IV-A, for any fading state, the broadcast fading channel is either used for all users or not used for any user. The resulting common outage probability is denoted as $\Pr$. Given a common outage probability of $\Pr = 0.1$, the two-user capacity region for the Nakagami-m fading channel $(m = 2, 3, 4)$ using optimal CD is shown for 3- and 20-dB SNR differences between the users in Figs. 5 and 6, respectively. In both cases, notice that the increase in capacity obtained by allowing a nonzero outage probability is larger for smaller m. This is because a smaller m corresponds

[9]As shown in Part I, TD and CDWO have the same ergodic capacity region.

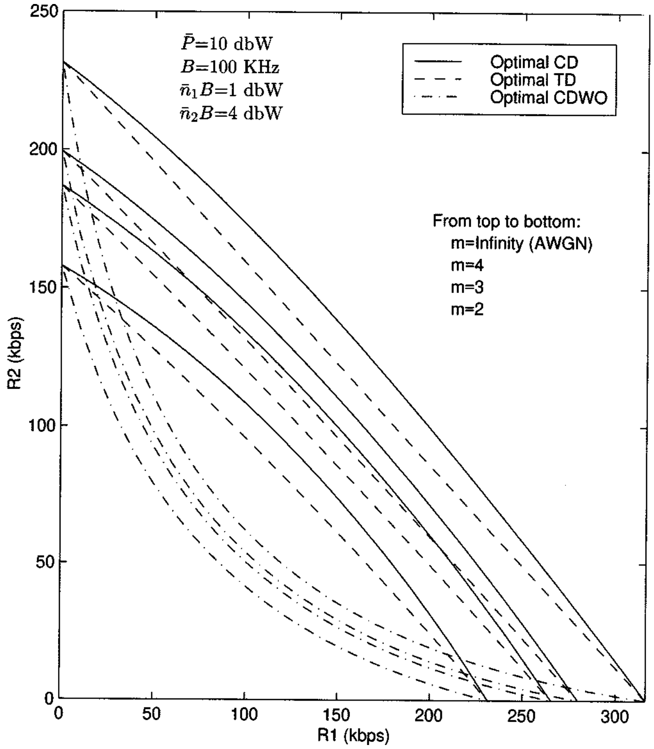

Fig. 4. Two-user zero-outage capacity region in Nakagami fading: 3-dB SNR difference.

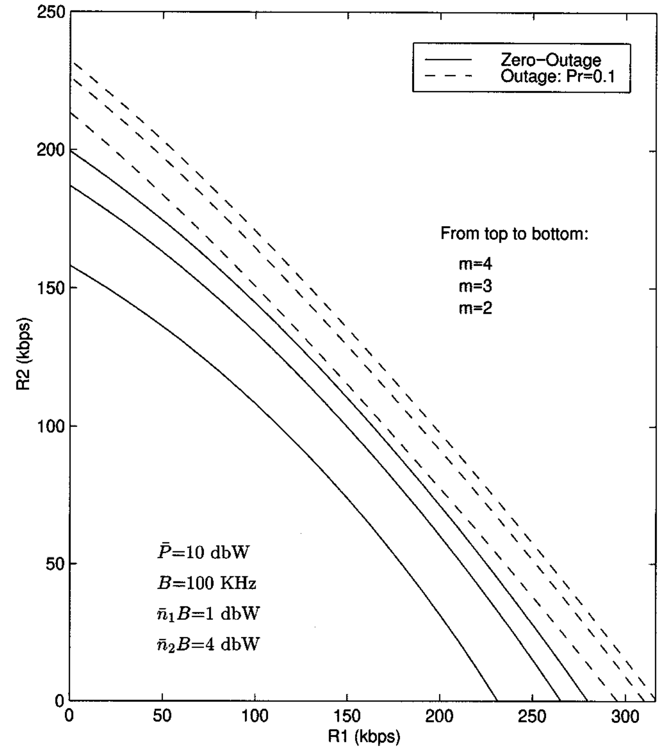

Fig. 5. Two-user capacity region for a given common outage probability in Nakagami fading using CD: 3-dB SNR difference.

to more severe fading, which is difficult to compensate for in the zero-outage case. In Fig. 5, since the SNR difference between the users is small, the increase of R_1 obtained by allowing outage is pretty much independent of R_2. However, when the SNR difference is large, we notice in Fig. 6 that in the region where R_2 is large, we obtain a large increase in User 1's rate R_1 by allowing some outage. This increase is much smaller in the region where R_2 is small. Since User 2 has much more noise on the average, for R_2 large and no outage, most of the total transmit power is used to send the information to User 2 and allowing some common outage probability will then save relatively more power for User 1 than in the case where R_2 is small.

In Fig. 7, the capacity regions using CD in Nakagami fading $(m = 2)$ with common outage probability $\mathrm{Pr} = 0.02$, $\mathrm{Pr} = 0.05$, and $\mathrm{Pr} = 0.1$ are compared when the SNR difference between the two users is 3 dB and the total average power is 10 dB. We see that by allowing even a small outage probability, we obtain a significant capacity increase relative to the zero-outage case. Fig. 8 shows the minimum common outage probability Pr as a function of the total average power $\overline{P}$ at a given rate pair $(R_1, R_2) = (100, 130)$ kb/s using CD under the same channel conditions as in Fig. 7. According to (3) and (4), this rate vector is on the boundary of the zero-outage capacity region for a total average power $\overline{P} \approx 10.9$ dB, as is shown in the figure.

Figs. 9 and 10 show the two-user capacity region of CDWO in Nakagami-m fading for a common outage probability of $\mathrm{Pr} = 0.1$. The SNR differences between the two users in these figures are 3 and 20 dB, respectively. Similar to the zero-outage capacity region, when the SNR difference between the users is small, the capacity regions with a given common outage probability increase faster with the increase of the Nakagami channel parameter m than when the SNR difference is large. However, in both cases, the increase of the capacity region from zero outage to an outage of 0.1 for each m is not that much and the differences between the outage capacity regions with different m

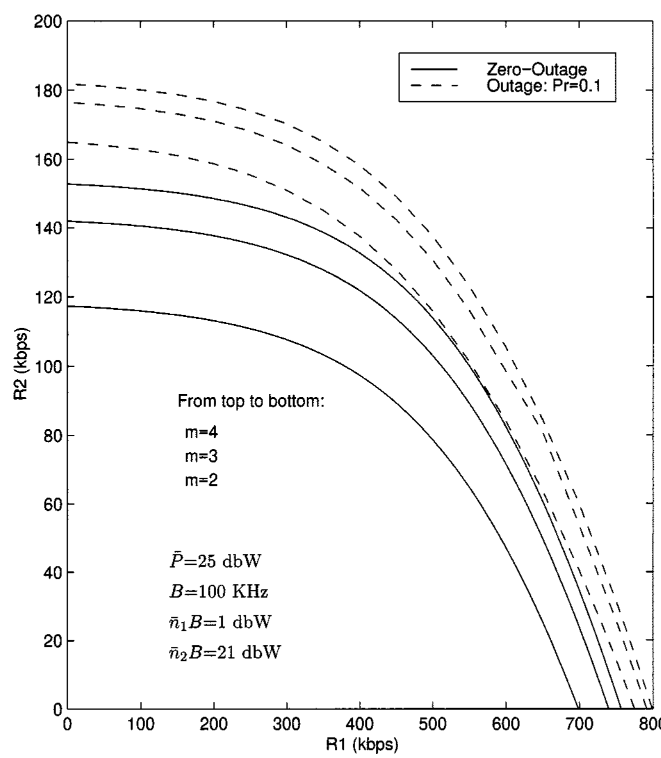

Fig. 6. Two-user capacity region for a given common outage probability in Nakagami fading using CD: 20-dB SNR difference.

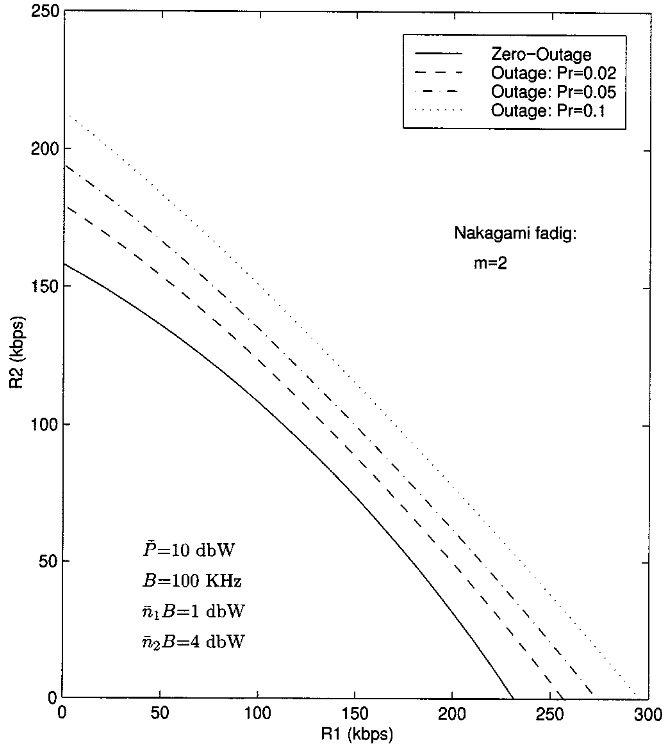

Fig. 7. Two-user capacity region comparison of different common outage probabilities in Nakagami fading using CD: 3-dB SNR difference.

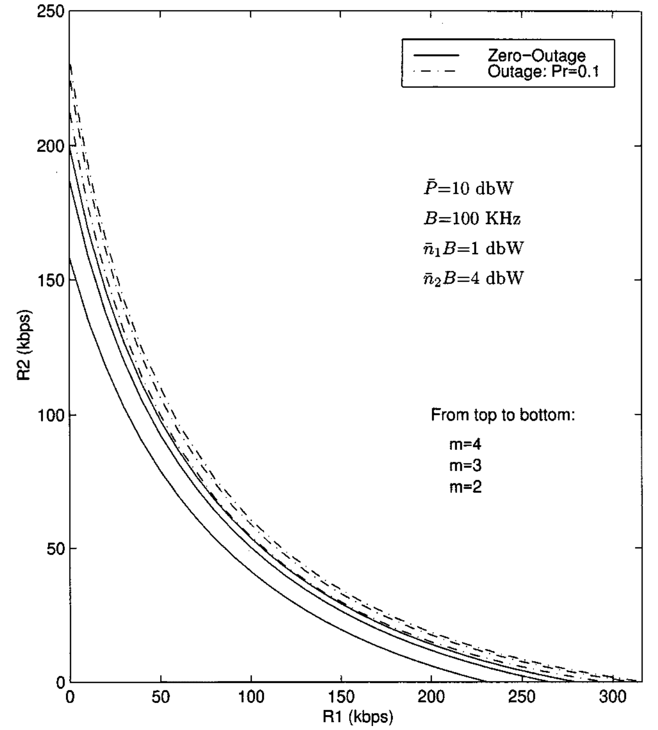

Fig. 9. Two-user capacity region for a given common outage probability in Nakagami fading using CDWO: 3-dB SNR difference.

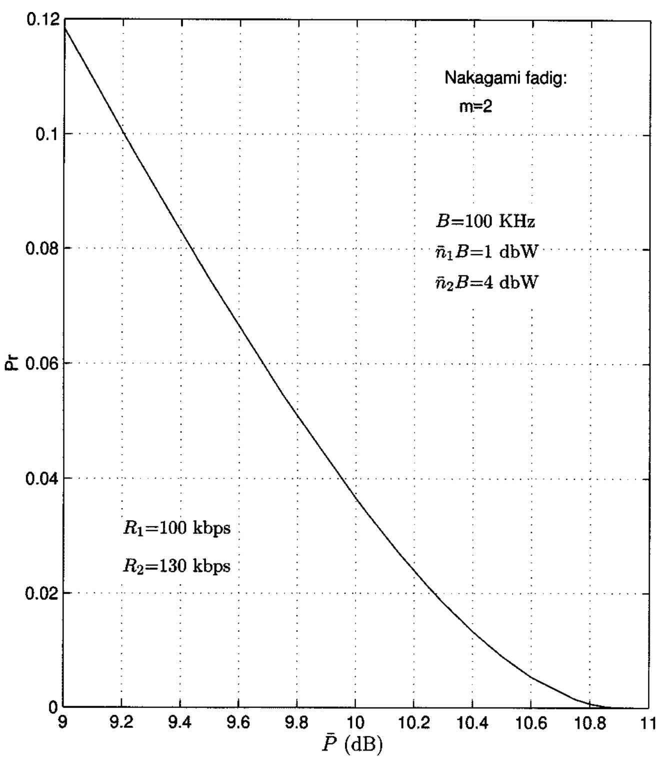

Fig. 8. Minimum common outage probability for a given rate vector versus average transmit power in Nakagami fading using CD.

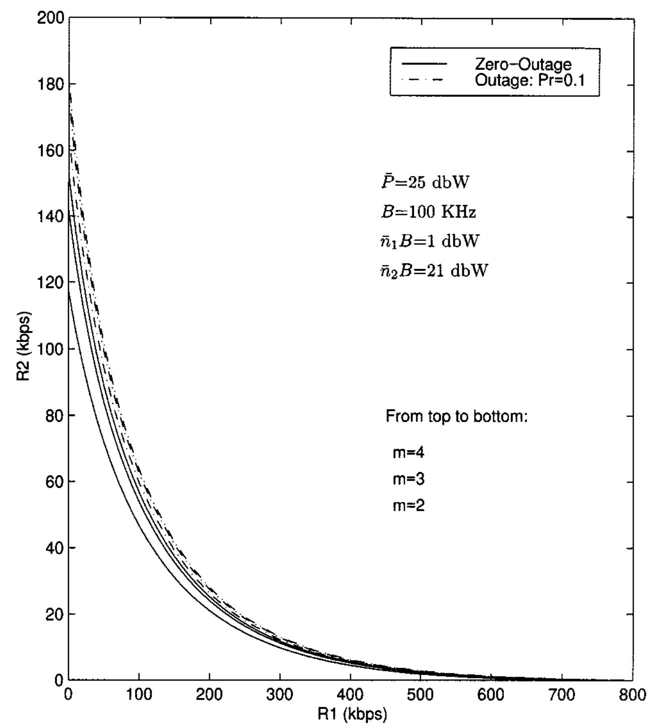

Fig. 10. Two-user capacity region for a given common outage probability in Nakagami fading using CDWO: 20-dB SNR difference.

are even smaller than that of the zero-outage capacity regions. This means that the optimal power policy that allows a certain common outage probability does not help much in increasing the capacity region of CDWO, especially when there is a great difference between the average channel conditions of the two users.

In Fig. 11, the capacity region with a common outage probability $\Pr = 0.1$ using optimal TD for the Nakagami fading channel $(m = 2)$ is shown and compared to that of the CD

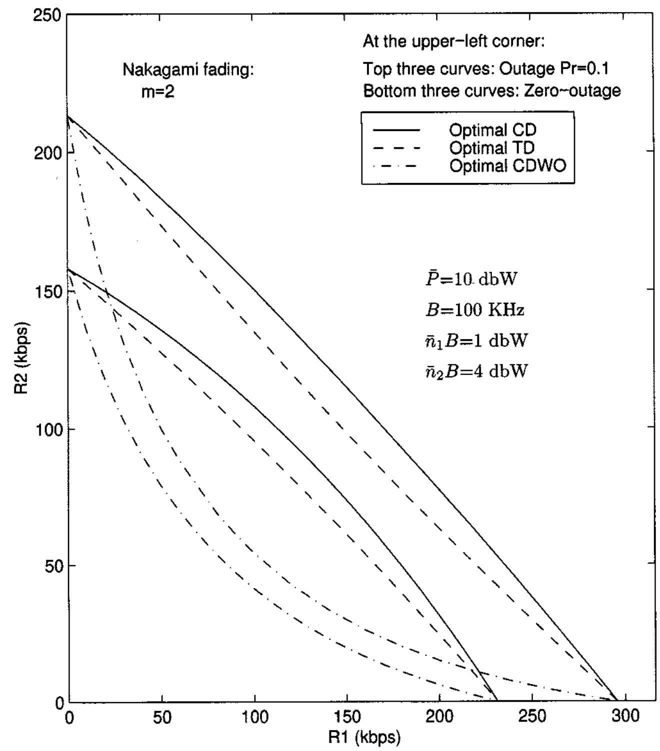

Fig. 11. Two-user capacity region comparison for a given common outage probability in Nakagami fading using CD, TD, and CDWO: 3-dB SNR difference.

and CDWO techniques. The SNR difference between the two users is 3 dB and the total power $\overline{P} = 10$ dB. As in the case where there is no outage, the capacity region with a common outage probability using TD is smaller than that of CD but is much larger than that of CDWO. Note that by allowing some common outage probability, there is a large increase of the capacity region for both CD and TD, but the increase is relatively small for CDWO.

In Figs. 12–15, we assume that a different outage probability can be declared for each user. The corresponding optimal power policy is obtained by applying the three-step procedure described in Section IV-B and Theorem 3, which is then used to calculate either the capacity region for a given outage probability vector or the outage probability region for a given rate vector. We obtain the capacity regions with outage or the outage probability regions for CD only, since the relative behavior of TD or CDWO is similar to that of CD in these figures.

In a two-user system, let Pr_1 and Pr_2 denote the outage probabilities for User 1 and User 2, respectively. Given $(\mathrm{Pr}_1, \mathrm{Pr}_2) = (0.02, 0.03)$, the two-user capacity regions with this outage for the Nakagami fading channels $(m = 2, 3, 4)$ are shown in Figs. 12 and 13 for SNR difference between the users of 3 and 20 dB, respectively. In both cases, as was true for the capacity region with a common outage probability, allowing some outage probability for each user results in a capacity increase that is larger for smaller m. Thus, the optimal power policy is more effective in increasing the capacity region when the overall broadcast channel fading is more severe. For different m, the differences between the capacity regions with outage are smaller than those between the capacity regions with no outage.

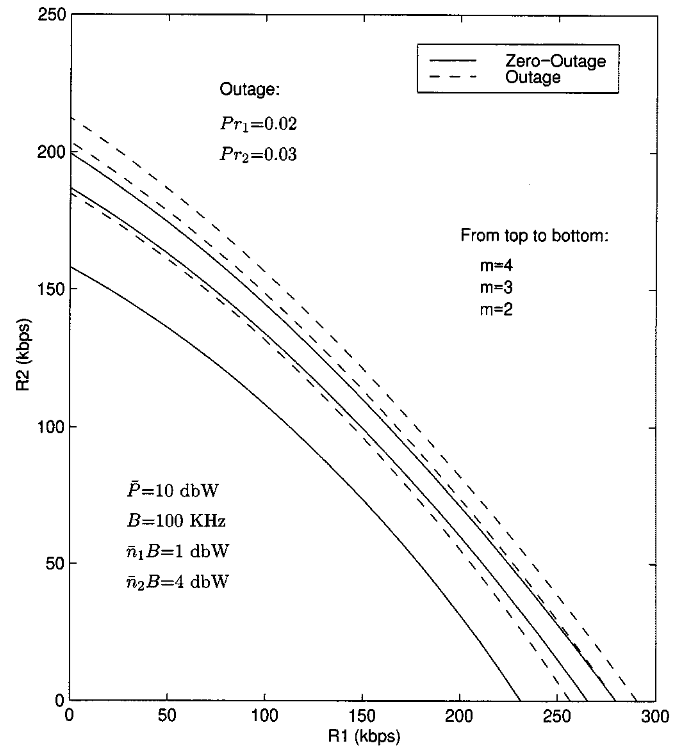

Fig. 12. Two-user capacity region for a given outage probability vector in Nakagami fading using CD: 3-dB SNR difference.

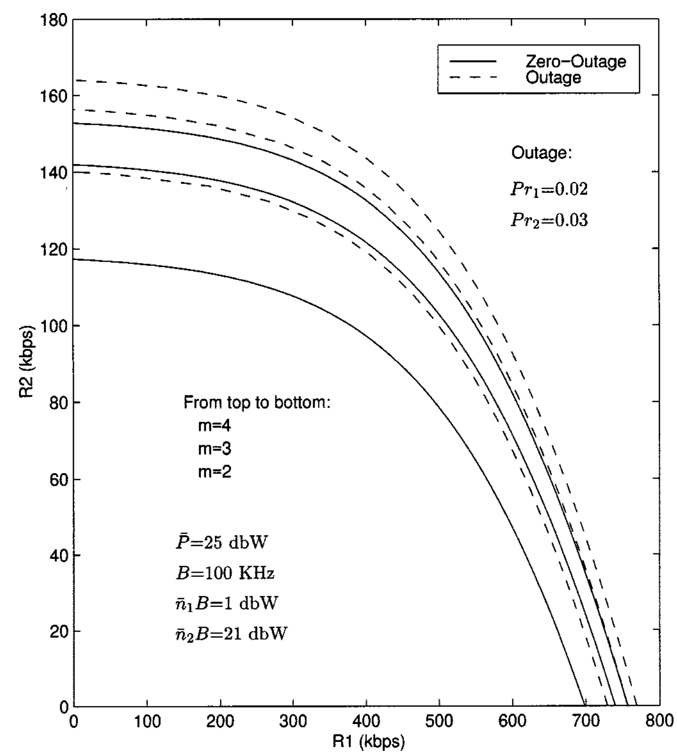

Fig. 13. Two-user capacity region for a given outage probability vector in Nakagami fading using CD: 20-dB SNR difference.

Fig. 14 shows the two-user outage probability regions with different total transmit power $\overline{P}$ for a given rate vector $(R_1, R_2) = (100, 130)$ kb/s. Note that the region below each curve is the outage probability region not achievable with the

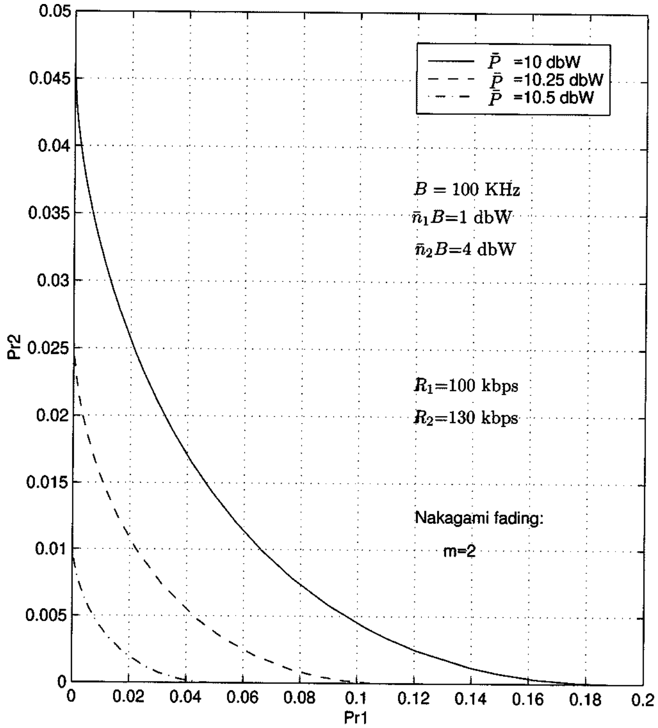

Fig. 14. Two-user outage probability region comparison for different average transmit power in Nakagami fading using CD: 3-dB SNR difference.

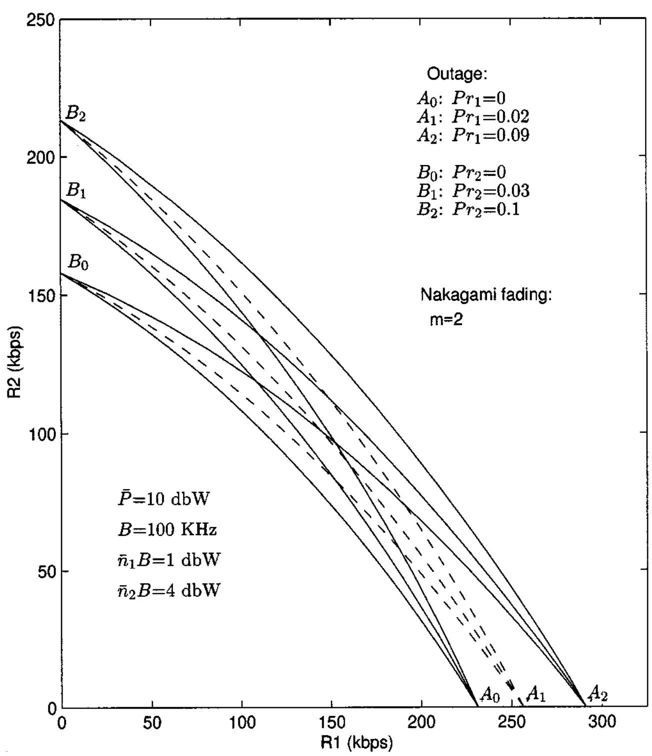

Fig. 15. Two-user capacity region comparison for different outage probability vectors in Nakagami fading using CD: 3-dB SNR difference.

corresponding transmit power $\overline{P}$. This nonachievable region shrinks quickly with the increase of transmit power $\overline{P}$ and disappears when $\overline{P} > 10.9$ dB since, when $\overline{P} \approx 10.9$ dB, according to (3) and (4), the rate vector $(R_1, R_2) = (100, 130)$ kb/s is on the boundary of the zero-outage capacity region. For a given transmit power $\overline{P}$, when the outage probability Pr_2 of User 2 decreases, there is a fast increase in the outage probability Pr_1 of User 1, since the average channel condition of User 2 is worse than that of User 1 and thus the total power required to support R_2 increases fast with the decrease of its outage probability. The intersections of the curves with the two axes in this figure denote the minimum outage probabilities for one user when there is no outage in the transmission for the other user.

Fig. 15 shows the two-user capacity regions with several different outage probability vectors and a total transmit power $\overline{P} = 10$ dB. In this figure, the points A_0, A_1, and A_2 are the single-user capacities of User 1 when the allowed outage probabilities are $\mathrm{Pr}_1 = 0$, $\mathrm{Pr}_1 = 0.02$, and $\mathrm{Pr}_1 = 0.09$, respectively; the points B_0, B_1, and B_2 are the single-user capacities of User 2 when the allowed outage probabilities are $\mathrm{Pr}_2 = 0$, $\mathrm{Pr}_2 = 0.03$, and $\mathrm{Pr}_2 = 0.1$, respectively. Let $Pa_0 = 0$, $Pa_1 = 0.02$, $Pa_2 = 0.09$, $Pb_0 = 0$, $Pb_1 = 0.03$, and $Pb_2 = 0.1$. The curves between points A_i and B_j are the boundaries of the capacity regions when the allowed outage probability vectors are $(\mathrm{Pr}_1, \mathrm{Pr}_2) = (Pa_i, Pb_j)$, $i, j = 0, 1, 2$. Note that when one of the outage probabilities Pr_1 and Pr_2 is zero, regardless of the time-varying channel state, the information of the corresponding user is always transmitted and the optimal power policy discussed in Section IV-C will be used for the other user to achieve the demonstrated capacity region in this figure under the constraint of its given outage probability.

Finally, in Figs. 16 and 17, where the SNR differences between the two users are 3 and 20 dB, respectively, the capacity regions using the optimal CD power policy with a common outage probability $\mathrm{Pr} = 0.1$ as discussed in Section IV-A and the optimal CD power policy with an outage probability vector $(\mathrm{Pr}_1, \mathrm{Pr}_2) = (0.1, 0.1)$ as discussed in Section IV-B are compared in Nakagami fading $(m = 2)$. Since the outage probability for each user is 0.1 using either of the two power policies, from the figures it is clear that by allowing a separate outage declaration for each user and using the corresponding optimal CD power policy, a larger capacity region can be achieved than by simply turning on or off the transmission for both users simultaneously based on the optimal power policy discussed in Section IV-A. However, the optimal power policy for the common outage declaration case is much less complex than that for the independent outage declaration case.

Also shown in Figs. 16 and 17 are the ergodic capacity regions of the fading broadcast channels as discussed in Part I of this paper. We see that these regions are much larger than the zero-outage capacity regions. Since the ergodic capacity and the zero-outage capacity correspond to the maximum average throughput and the maximum constant throughput of a fading channel, respectively, the comparison between these two different capacity regions demonstrates the throughput loss of a broadcast system transmitting at constant rates in any fading condition instead of transmitting at variable rates adapted to the fading channel states.

Note that by definition, each outage capacity region shown in the figures of this paper represents a set of constant rate vectors

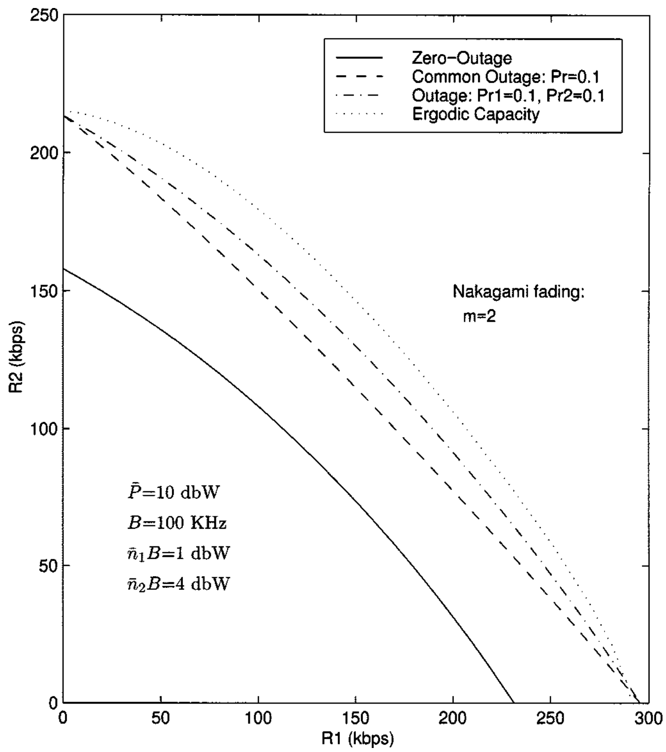

Fig. 16. Two-user capacity regions with a common outage probability and with an outage probability vector in Nakagami fading using CD: 3 dB SNR difference.

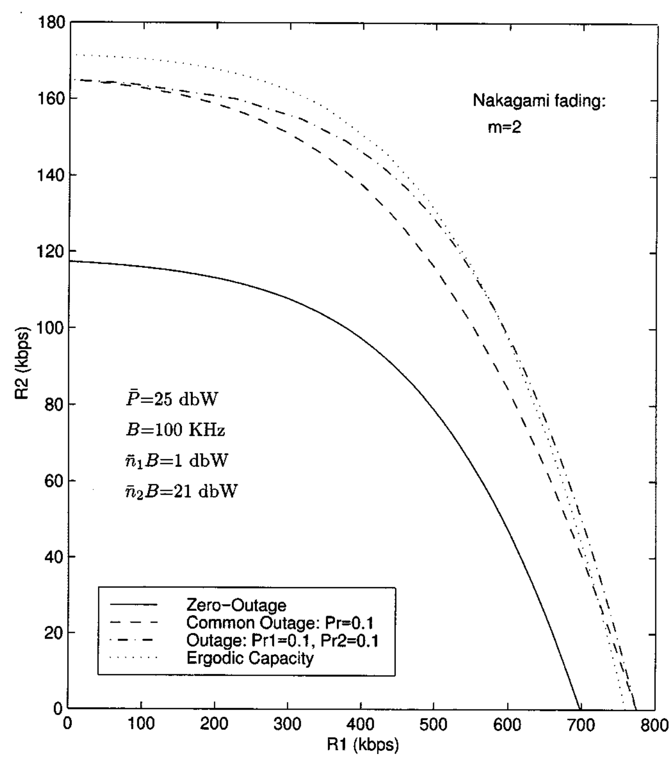

Fig. 17. Two-user capacity regions with a common outage probability and with an outage probability vector in Nakagami fading using CD: 20-dB SNR difference.

that can be maintained with the given outage probability vector $\mathbf{Pr}$ or the given common outage probability Pr satisfied. If we normalize the outage capacity of each user i $(i = 1, \ldots, M)$ with $1 - \mathrm{Pr}_i$ in the case of independent outage declaration and with $1 - \mathrm{Pr}$ in the case of common outage declaration, then the normalized outage capacity region will be an average throughput region. This average throughput region does not always include the zero-outage capacity region (maximum constant throughput region) as can be verified for the cases shown in Figs. 6, 9, 10, and 13. This is because the normalized outage capacity region varies with the given outage probability vector $\mathbf{Pr}$ or the given common outage probability Pr.

VII. Conclusion

We have obtained both the zero-outage capacity region and the minimum-outage capacity region of fading broadcast channels for TD, FD, and CD with and without successive decoding, assuming that perfect CSI is available at both the transmitter and the receivers. It is shown that optimal CD has the largest zero-outage capacity region, as expected. Moreover, we show that the capacity region can be greatly expanded by allowing some outage probability for each user. For a given rate vector, we have derived the optimal power policy that minimizes the common outage probability when transmission to all users is turned off simultaneously. When an outage can be declared for each user individually, we have also derived a general power allocation strategy to achieve boundaries of the outage probability regions under different spectrum-sharing techniques. We observe that these regions can increase dramatically with an increase in the total transmit power. Therefore, by applying the optimal dynamic power allocation strategies derived herein, tradeoffs between the maximum constant transmission rate, the outage probability for each user, and the total transmit power may be evaluated for the design of a broadcast communication system in a fading environment.

Appendix

A. Proof of Theorem 1

Achievability of the Capacity Region: We prove the achievability of the capacity region $\mathcal{C}_{\text{zero}}(\overline{P})$ in (1) by proving the achievability of $\boldsymbol{R} \in \bigcap_{\boldsymbol{n}\in\mathcal{N}} \mathcal{C}_{\text{CD}}(\boldsymbol{n}, \mathcal{P})$ in (2) for each given power allocation policy $\mathcal{P} \in \mathcal{F}$.

$\forall \mathcal{P} \in \mathcal{F}$, for $j = 1, 2, \ldots, M$, since $P_j(\boldsymbol{n})$ denotes the transmit power for User j in fading state $\boldsymbol{n}$

$$\forall \boldsymbol{R} \in \bigcap_{\boldsymbol{n}\in\mathcal{N}} \mathcal{C}_{\text{CD}}(\boldsymbol{n}, \mathcal{P}), \qquad \boldsymbol{R} = (R_1, R_2, \ldots, R_M)$$

we need to prove that for every $\epsilon > 0$, there exists a sequence of $((2^{R_1T}, 2^{R_2T}, \ldots, 2^{R_MT}), T)$ codes and a coding and decoding scheme with probability of error $P_e^{(T)} < \epsilon$ for every fading process with stationary distribution $\mathcal{Q}$, i.e., a coding delay T which is independent of the correlation structure of the fading. We prove in the following that this is true for the two-user case. The result can be easily generalized to the M-user case $(M > 2)$. Note that with the availability of CSI at both the transmitter and the receivers, the codewords can be chosen based on the realization of the fading process.

Let

$$\Phi_1 \triangleq \{\boldsymbol{n}: n_1 < n_2, \boldsymbol{n} \in \mathcal{N}\}$$

$$\Phi_2 \triangleq \{\boldsymbol{n}: n_1 > n_2, \boldsymbol{n} \in \mathcal{N}\}.$$

Recall that we assume $\Pr\{n_i = n_j\} = 0, \forall i \neq j$. Thus,

$$\mathcal{N} = \Phi_1 \cup \Phi_2.$$

When the channel fading states are in Φ_1, let

$$\boldsymbol{N}^{(T)} = (\boldsymbol{N}(1), \boldsymbol{N}(2), \ldots, \boldsymbol{N}(T))$$

denote the realization of the slowly time-varying fading process. Let $M_1 = 2^{TR_1}$ and $M_2 = 2^{TR_2}$. Generate M_2 independent codewords $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{M_2}$ of length T according to the normal distribution $N(0, 1)$, scaled by

$$\sqrt{P_2(\boldsymbol{N}(n))}, \qquad n = 1, 2, \ldots, T.$$

For each codeword $\boldsymbol{u}_i = (u_{i,1}, u_{i,2}, \ldots, u_{i,T})$ ($\boldsymbol{u}_i$ is called a cluster center, $1 \le i \le M_2$), generate M_1 independent codewords $\boldsymbol{x}_{1,i}, \boldsymbol{x}_{2,i}, \ldots, \boldsymbol{x}_{M_1,i}$ of length T according to the conditional normal distribution $N(u_{i,n}, 1)$, scaled by

$$\sqrt{P_1(\boldsymbol{N}(n))}, \qquad n = 1, 2, \ldots, T.$$

Assuming that the source for User 1 produces integer m $(1 \le m \le M_1)$ and the source for User 2 produces integer i $(1 \le i \le M_2)$, the encoder maps the pair (m, i) into a codeword

$$\boldsymbol{x}_{m,i} = (x_{m,i,1}, x_{m,i,2}, \ldots, x_{m,i,T})$$

which is then transmitted. Let $\boldsymbol{y}$ and $\boldsymbol{z}$ be the received sequences for User 1 and User 2, respectively. We use the decoding rule in [12]. That is, the decoder of User 2 decodes that i for which $p(\boldsymbol{z}|\boldsymbol{u}_i)$ is maximized (a decoding failure occurs when there is a tie for the maximum). Let $P_{e,2}^{(T)}$ be the probability of decoding error for User 2. The decoder for User 1 first decodes the cluster center $\boldsymbol{u}_i$ in the same way as the decoder of User 2 does, and then uses its estimate of i to choose the m for which $p(\boldsymbol{y}|\boldsymbol{x}_{m,i})$ is maximized. For User 1, let $P_{e,12}^{(T)}$ be the probability of decoding error for index i and let $P_{e,11}^{(T)}$ be the probability of decoding error for index m. Thus, by denoting $P_{e,1}^{(T)}$ as the probability of decoding error for User 1, we have

$$P_{e,1}^{(T)} \le P_{e,11}^{(T)} + P_{e,12}^{(T)}. \tag{67}$$

Based on the above encoding and decoding rules, the probability of decoding error for User 2 is bounded by [12][10]

$$P_{e,2}^{(T)} \le \exp(\rho T R_2/(2B)) \times \sum_{\boldsymbol{n} \in \Phi_1} \left\{ f(\boldsymbol{n}) \sum_{\boldsymbol{z}} \left(\sum_{\boldsymbol{u}} Q_1(\boldsymbol{u}|\boldsymbol{n}) [p(\boldsymbol{z}|\boldsymbol{u}, \boldsymbol{n})]^{\frac{1}{1+\rho}} \right)^{1+\rho} \right\} \tag{68}$$

for any $\rho > 0$, where $f(\boldsymbol{n})$ denotes the pdf of $\boldsymbol{n}$, $Q_1(\boldsymbol{u}|\boldsymbol{n})$ is the conditional pdf of $\boldsymbol{u}$, conditional on the fading being $\boldsymbol{n}$, and $p(\boldsymbol{z}|\boldsymbol{u}, \boldsymbol{n})$ is the conditional pdf of the received sequence $\boldsymbol{z}$, conditional on the codeword being $\boldsymbol{u}$ and the fading being $\boldsymbol{n}$. Since

$$Q_1(\boldsymbol{u}|\boldsymbol{n}) = \prod_{n=1}^{T} Q_1(u_n|\boldsymbol{n}), \tag{69}$$

$$p(\boldsymbol{z}|\boldsymbol{u}, \boldsymbol{n}) = \prod_{n=1}^{T} p(z_n|u_n, \boldsymbol{n}) = \prod_{n=1}^{T} \left[\sum_{x} Q_2(x|u_n, \boldsymbol{n}) p(z_n|x, \boldsymbol{n}) \right] \tag{70}$$

where $\forall\, 1 \le n \le T$

$$Q_1(u_n|\boldsymbol{n}) \sim N(0, P_2(\boldsymbol{n}))$$
$$Q_2(x|u_n, \boldsymbol{n}) \sim N(u_n, P_1(\boldsymbol{n}))$$

and

$$p(z_n|x, \boldsymbol{n}) \sim N(x, n_2 B).$$

Therefore, in (70)

$$p(z_n|u_n, \boldsymbol{n}) \sim N(u_n, n_2 B + P_1(\boldsymbol{n})).$$

Substituting (69) and (70) into (68), it is easy to verify that

$$P_{e,2}^{(T)} \le \sum_{\boldsymbol{n} \in \Phi_1} f(\boldsymbol{n}) \cdot \exp\left\{ -\rho \left[-T \frac{R_2}{2B} + \frac{1}{2} \sum_{n=1}^{T} \log\left(1 + \frac{P_2(\boldsymbol{n})}{(1+\rho)(n_2 B + P_1(\boldsymbol{n}))}\right) \right] \right\}.$$

By assumption, $\exists \delta_2 > 0$ such that

$$R_2 \le B \log\left(1 + \frac{P_2(\boldsymbol{n})}{n_2 B + P_1(\boldsymbol{n})}\right) - B\delta_2 \quad \forall \boldsymbol{n} \in \Phi_1. \tag{71}$$

Thus,

$$P_{e,2}^{(T)} \le \exp\left\{ -\frac{\rho T}{2} [\delta_2 - \log(1+\rho)] \right\} \tag{72}$$

since $\sum_{\boldsymbol{n} \in \Phi_1} f(\boldsymbol{n}) \le 1$.

Similarly, we can show that

$$P_{e,12}^{(T)} \le \exp(\rho T R_2/(2B)) \times \sum_{\boldsymbol{n} \in \Phi_1} \left\{ f(\boldsymbol{n}) \sum_{\boldsymbol{y}} \left(\sum_{\boldsymbol{u}} Q_1(\boldsymbol{u}|\boldsymbol{n}) [p(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{n})]^{\frac{1}{1+\rho}} \right)^{1+\rho} \right\} \tag{73}$$

for any $\rho > 0$, where $Q_1(\boldsymbol{u}|\boldsymbol{n})$ is given by (69) and $p(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{n})$ is the conditional probability density function of the received sequence $\boldsymbol{y}$, conditional on the codeword being $\boldsymbol{u}$ and the fading being $\boldsymbol{n}$. Since

$$p(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{n}) = \prod_{n=1}^{T} p(y_n|u_n, \boldsymbol{n}) = \prod_{n=1}^{T} \left[\sum_{x} Q_2(x|u_n, \boldsymbol{n}) p(y_n|x, \boldsymbol{n}) \right] \tag{74}$$

where $\forall\, 1 \le n \le T$, $Q_2(x|u_n, \boldsymbol{n}) \sim N(u_n, P_1(\boldsymbol{n}))$ as given above, and $p(y_n|x, \boldsymbol{n}) \sim N(x, n_1 B)$. Therefore, in (74), $p(y_n|u_n, \boldsymbol{n}) \sim N(u_n, n_1 B + P_1(\boldsymbol{n}))$. Substituting (69) and (74) into (73), it is easily shown that

$$P_{e,12}^{(T)} \le \sum_{\boldsymbol{n} \in \Phi_1} f(\boldsymbol{n}) \cdot \exp\left\{ -\rho \left[-T \frac{R_2}{2B} + \frac{1}{2} \sum_{n=1}^{T} \log\left(1 + \frac{P_2(\boldsymbol{n})}{(1+\rho)(n_1 B + P_1(\boldsymbol{n}))}\right) \right] \right\}.$$

[10]Note that the unit for R_1 and R_2 in this paper is "bits per second," while the unit in [12] is "bits per sample." This is why R_2 in (68) is divided by $2B$, the number of samples per second for the band-limited channel.

Since $\forall \boldsymbol{n} \in \Phi_1$, $n_1 < n_2$, from (71) we obtain

$$R_2 \le B \log\left(1 + \frac{P_2(\boldsymbol{n})}{n_1 B + P_1(\boldsymbol{n})}\right) - B\delta_2 \qquad \forall \boldsymbol{n} \in \Phi_1.$$

Thus,

$$P_{e,12}^{(T)} \le \exp\left\{-\frac{\rho T}{2}\left[\delta_2 - \log(1+\rho)\right]\right\}. \tag{75}$$

Furthermore, for User 1, the probability of decoding error $P_{e,11}^{(T)}$ is bounded by [12]

$$P_{e,11}^{(T)} \le \exp(\rho T R_1/(2B)) \times \sum_{\boldsymbol{n}\in\Phi_1} \left\{ f(\boldsymbol{n}) \sum_{\boldsymbol{u}} Q_1(\boldsymbol{u}|\boldsymbol{n}) \times \sum_{\boldsymbol{y}} \left(\sum_{\boldsymbol{x}} Q_2(\boldsymbol{x}|\boldsymbol{u}, \boldsymbol{n}) [p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{n})]^{\frac{1}{1+\rho}} \right)^{1+\rho} \right\} \tag{76}$$

for any $\rho > 0$, where $Q_1(\boldsymbol{u}|\boldsymbol{n})$ is given in (69)

$$Q_2(\boldsymbol{x}|\boldsymbol{u}, \boldsymbol{n}) = \prod_{n=1}^{T} Q_2(x_n|u_n, \boldsymbol{n}) \tag{77}$$

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{n}) = \prod_{n=1}^{T} p(y_n|x_n, \boldsymbol{n}) \tag{78}$$

and $\forall\, 1 \le n \le T$,

$$Q_2(x_n|u_n, \boldsymbol{n}) \sim N(u_n, P_1(\boldsymbol{n}))$$
$$p(y_n|x_n, \boldsymbol{n}) \sim N(x_n, n_1 B).$$

Therefore, by substituting (69), (77), and (78) into (76), it is easy to verify that

$$P_{e,11}^{(T)} \le \sum_{\boldsymbol{n}\in\Phi_1} f(\boldsymbol{n}) \cdot \exp\left\{-\rho\left[-T\frac{R_1}{2B} + \frac{1}{2}\sum_{n=1}^{T}\log\left(1+\frac{P_1(\boldsymbol{n})}{(1+\rho)n_1 B}\right)\right]\right\}.$$

By assumption, $\exists\, \delta_1 > 0$ such that

$$R_1 \le B\log\left(1+\frac{P_1(\boldsymbol{n})}{n_1 B}\right) - B\delta_1, \qquad \forall \boldsymbol{n}\in\Phi_1.$$

Thus,

$$P_{e,11}^{(T)} \le \exp\left\{-\frac{\rho T}{2}\left[\delta_1 - \log(1+\rho)\right]\right\}. \tag{79}$$

Denoting $\delta = \min\{\delta_1, \delta_2\}$, then by (72) we obtain

$$P_{e,2}^{(T)} \le \exp\left\{-\frac{\rho T}{2}\left[\delta - \log(1+\rho)\right]\right\}$$

and by (67), (75), and (79) we obtain

$$P_{e,1}^{(T)} \le 2\cdot\exp\left\{-\frac{\rho T}{2}\left[\delta - \log(1+\rho)\right]\right\}.$$

Therefore, when the channel fading states are in Φ_1, the overall probability of decoding error $P_e^{(T)}(\Phi_1)$ for the two users is

$$P_e^{(T)}(\Phi_1) \le P_{e,1}^{(T)} + P_{e,2}^{(T)} \le 3\cdot\exp\left\{-\frac{\rho T}{2}\left[\delta - \log(1+\rho)\right]\right\}.$$

By taking ρ sufficiently small, we have $\delta - \log(1+\rho) > 0$ and it follows that the probability of error $P_e^{(T)}(\Phi_1)$ decreases exponentially with T, i.e., $\forall \epsilon > 0$, $\exists\, T_d(\Phi_1) > 0$, $\forall T > T_d(\Phi_1)$, $P_e^{(T)}(\Phi_1) < \epsilon$.

It can be similarly shown that when the channel-fading states are in Φ_2, there exists a sequence of $((2^{R_1 T}, 2^{R_2 T}, \ldots, 2^{R_M T}), T)$ codes and a coding and decoding scheme for which the probability of error $P_e^{(T)}(\Phi_2)$ decays exponentially with T, i.e., $\forall \epsilon > 0$, $\exists\, T_d(\Phi_2) > 0$, $\forall T > T_d(\Phi_2)$, $P_e^{(T)}(\Phi_2) < \epsilon$. Thus, $\forall \boldsymbol{n} \in \mathcal{N}$, there exists a sequence of $((2^{R_1 T}, 2^{R_2 T}, \ldots, 2^{R_M T}), T)$ codes and a coding and decoding scheme for which the probability of error $P_e^{(T)} \to 0$ as $T \to \infty$. Moreover, $P_e^{(T)}$ decreases in T at a rate independent of the correlation character of the fading, i.e., by denoting $T_d = \max\{T_d(\Phi_1), T_d(\Phi_2)\}$, $\forall \epsilon > 0$, $\forall T > T_d$, we have $P_e^{(T)} < \epsilon$ for every fading process with stationary distribution $\mathcal{Q}$. □

Converse: Suppose that rate vector $\boldsymbol{R}$ is achievable, i.e., $\boldsymbol{R} \in \mathcal{C}_{\text{zero}}(\overline{P})$. We need to prove that $\boldsymbol{R}$ cannot be outside of the region defined in (1). The proof is similar to that of the MAC capacity region [5].

Define $v_i = \frac{i}{m}$, $i = 0, 1, 2, \ldots, mI$. Since the time-varying noise density n_j of each user ranges from 0 to ∞, we say that a subchannel is in state S_i, $i = 0, 1, 2, \ldots, mI$ if $v_i \le n_j < v_{i+1}$, where $v_{mI+1} \triangleq \infty$. Therefore, there are $(mI+1)^M = N$ discrete joint channel states. We denote the kth $(0 \le k \le N-1)$ of these N states as $\boldsymbol{S}_k = [S_{\phi(k,1)}, S_{\phi(k,2)}, \ldots, S_{\phi(k,M)}]$, where $[\phi(k,1), \phi(k,2), \ldots, \phi(k,M)]$ is the base-$(mI+1)$ expansion of k, with $\phi(k,1)$ being the least important component. That is, $0 \le \phi(k,j) \le mI$ for all $1 \le j \le M$ and

$$k = \sum_{j=1}^{M} \phi(k,j)\cdot(mI+1)^{j-1}.$$

Note that a channel state $\boldsymbol{n} \in \boldsymbol{S}_k$ if and only if $n_j \in S_{\phi(k,j)}$, $\forall\, 1 \le j \le M$.

Consider a sequence of Markov processes defined on $\mathcal{N}$ by using a Markov chain which is composed of the above mentioned N channel states with transition probabilities $t(\boldsymbol{S}_j, \boldsymbol{S}_k)$. The process remains in a state $\boldsymbol{S}_j$ for an exponential time $\tau(\boldsymbol{S}_j) \triangleq$ Exponential$(\lambda(\boldsymbol{S}_j))$ and then selects a new state according to $t(\boldsymbol{S}_j, \boldsymbol{S}_k)$. By choosing the appropriate $\{\tau(\boldsymbol{S}_j)\}_{j=0}^{N-1}$ and transition probabilities, we assume that the Markov process has the required stationary distribution $\mathcal{Q}$ of the fading channel.

For each $T = 1, 2, \ldots$, let $\boldsymbol{N}^{(T)}$ be a fading process starting with $\boldsymbol{N}^{(T)}(0) = \boldsymbol{N}(0)$, where $\boldsymbol{N}(0)$ is a random variable with the stationary distribution $\mathcal{Q}$. We assume all fading processes start with $\boldsymbol{N}^{(T)}(0) \equiv \boldsymbol{N}(0)$, $T = 1, 2, \ldots$. The initial sojourn time in state $\boldsymbol{N}(0)$ of fading process $\boldsymbol{N}^{(T)}$ is given by $\tau_T(\boldsymbol{N}(0))$, where $\tau_T(\boldsymbol{S}_j) \asymp$ Exponential$(r_T\lambda(\boldsymbol{S}_j))$, $j = 0, 1, \ldots, N-1$. The scaling constant r_T determines the fading speed for process $\boldsymbol{N}^{(T)}$. $\forall \delta > 0$ fixed, by selecting an appropriate decreasing sequence $\{r_T\}_{T=1}^{\infty}$ where $r_T \to 0$ as $T \to \infty$, we can have

$$\Pr(\forall T, \tau_T(\boldsymbol{S}_j) > T) > 1 - \delta, \qquad 0 \le j \le N-1. \tag{80}$$

Since $\boldsymbol{R} \in \mathcal{C}_{\text{zero}}(\overline{P})$, we can choose for each T a code of size $2^{TR_1} \cdot 2^{TR_2} \cdots 2^{TR_M}$ for which the probability of error $p(T)$ under fading process $\boldsymbol{N}^{(T)}$ goes to zero as $T \to \infty$. Let $\{\boldsymbol{X}^{(T)}(n)\}_{n=1}^{T}$ denote a random selection of codewords from the codebook for the M users and let $P_j^{(T)}(n)$ be the transmit power for User j $(1 \le j \le M)$. Note that $\{\boldsymbol{X}^{(T)}(n)\}_{n=1}^{T}$ can be chosen according to the fading process $\boldsymbol{N}^{(T)}$. For $0 \le k \le N-1$, let $\Omega(\boldsymbol{S}_k)$ be the subset of the sample space on which $\boldsymbol{N}(0) \in \boldsymbol{S}_k$ and $\forall T$, $\tau_T(\boldsymbol{S}_k) > T$. Let Q be an independent random variable uniformly distributed on $[0, T]$. For $1 \le j \le M$, define

$$\begin{aligned} V_j(\boldsymbol{S}_k, T) &= E_Q\left[P_j^{(T)}(Q) \middle| \Omega(\boldsymbol{S}_k)\right] \\ W_j(\boldsymbol{S}_k, T) &= E_Q\left[P_j^{(T)}(Q) \middle| [\boldsymbol{N}(0) \in \boldsymbol{S}_k] - \Omega(\boldsymbol{S}_k)\right] \\ Z_j(\boldsymbol{S}_k, T) &= V_j(\boldsymbol{S}_k, T) \Pr\{\forall T, \tau_T(\boldsymbol{S}_k) > T | \boldsymbol{N}(0) \in \boldsymbol{S}_k\} \\ &\quad + W_j(\boldsymbol{S}_k, T) \Pr\{\exists T: \tau_T(\boldsymbol{S}_k) \le T | \boldsymbol{N}(0) \in \boldsymbol{S}_k\}. \end{aligned}$$

Then the power constraint is that $\forall T$

$$\sum_{k=0}^{N-1} p(\boldsymbol{S}_k) \sum_{j=1}^{M} Z_j(\boldsymbol{S}_k, T) \le \overline{P}.$$

According to (80), we have

$$\sum_{k=0}^{N-1} p(\boldsymbol{S}_k) \sum_{j=1}^{M} V_j(\boldsymbol{S}_k, T) \le \frac{\overline{P}}{1-\delta}.$$

By the bounded convergence theorem, it is clear that there exists a convergent subsequence along which if taking the limit we obtain for $1 \le j \le M$ and $0 \le k \le N-1$

$$V_j(\boldsymbol{S}_k, T) \to V_j(\boldsymbol{S}_k), \qquad \text{as } T \to \infty.$$

Therefore

$$\sum_{k=0}^{N-1} p(\boldsymbol{S}_k) \sum_{j=1}^{M} V_j(\boldsymbol{S}_k) \le \frac{\overline{P}}{1-\delta}.$$

Now we define a new fading process $\boldsymbol{N}$ by

$$\boldsymbol{N}(n) = \sum_{k=0}^{N-1} 1[\boldsymbol{N}(0) \in \boldsymbol{S}_k] \boldsymbol{n}^l(\boldsymbol{S}_k)$$

where

$$\boldsymbol{n}^l(\boldsymbol{S}_k) = [n_1^l(\boldsymbol{S}_k), n_2^l(\boldsymbol{S}_k), \ldots, n_M^l(\boldsymbol{S}_k)]$$

and

$$n_j^l(\boldsymbol{S}_k) \triangleq v_{\phi(k,j)}, \qquad 1 \le j \le M.$$

Note that conditioned on $\boldsymbol{N}(0)$, the fading process is deterministic. $\forall 0 \le k \le N-1$, let $q(T|\Omega(\boldsymbol{S}_k))$ be the conditional probability of error for code $\boldsymbol{X}^{(T)}$ in the new fading channel. Then, obviously

$$p(T) \ge p(\Omega(\boldsymbol{S}_k)) q(T|\Omega(\boldsymbol{S}_k)).$$

By assumption, $p(T) \to 0$ as $T \to \infty$. Thus, $q(T|\Omega(\boldsymbol{S}_k)) \to 0$ as $T \to \infty$. But conditional on $\Omega(\boldsymbol{S}_k)$, we have a constant channel, and a sequence of codes satisfying the power constraint $\{V_j(\boldsymbol{S}_k)\}_{j=1}^{M}$. It follows that

$$\boldsymbol{R} \in \mathcal{C}_{\text{CD}}\left(\boldsymbol{n}^l(\boldsymbol{S}_k), \mathcal{V}\right), \qquad \forall 0 \le k \le N-1 \tag{81}$$

where $\mathcal{V}$ denotes the power allocation policy that assigns power $V_j(\boldsymbol{S}_k)$ to User j $(1 \le j \le M)$ in channel state $\boldsymbol{S}_k$, and $\mathcal{C}_{\text{CD}}(\cdot, \cdot)$ is as given in (2).

$\forall \mathcal{P} \in \mathcal{F}, \forall \boldsymbol{n} \in \mathcal{N}$, define

$$\overline{\mathcal{C}}_{\text{CD}}^{(N)}(\boldsymbol{n}, \mathcal{P}) = \mathcal{C}_{\text{CD}}(\boldsymbol{n}^l(\boldsymbol{S}_k), \mathcal{P}), \qquad \text{if } \boldsymbol{n} \in \boldsymbol{S}_k,\ 0 \le k \le N-1.$$

Denote $\mathcal{F}_{N,\delta}$ as the set of all power control policies that satisfy the power constraint $\frac{\overline{P}}{1-\delta}$ and are piecewise constant on each fading state $\boldsymbol{S}_k$, $0 \le k \le N-1$. Denote $\mathcal{P}_{N,\delta}$ as the power allocation policy that assigns power $P_j^{N,\delta}(\boldsymbol{n})$ to User j $(1 \le j \le M)$ in each fading state $\boldsymbol{n}$, where

$$P_j^{N,\delta}(\boldsymbol{n}) = \sum_{k=0}^{N-1} V_j(\boldsymbol{S}_k) 1[\boldsymbol{n} \in \boldsymbol{S}_k].$$

Thus, $\mathcal{P}_{N,\delta} \in \mathcal{F}_{N,\delta}$. Since $\forall \boldsymbol{n} \in \mathcal{N}$, there exists a fading state $\boldsymbol{S}_k$ $(0 \le k \le N-1)$ such that $\boldsymbol{n} \in \boldsymbol{S}_k$, from (81) and the definition of $\overline{\mathcal{C}}_{\text{CD}}^{(N)}(\cdot, \cdot)$ we know that, for any $\delta > 0$

$$\forall \boldsymbol{n} \in \mathcal{N}, \qquad \boldsymbol{R} \in \overline{\mathcal{C}}_{\text{CD}}^{(N)}(\boldsymbol{n}, \mathcal{P}_{N,\delta}).$$

It follows that

$$\boldsymbol{R} \in \bigcup_{\mathcal{P} \in \mathcal{F}_N} \bigcap_{\boldsymbol{n} \in \mathcal{N}} \overline{\mathcal{C}}_{\text{CD}}^{(N)}(\boldsymbol{n}, \mathcal{P})$$

where $\mathcal{F}_N = \mathcal{F}_{N,0}$. Thus,

$$\mathcal{C}_{\text{zero}}(\overline{P}) \subseteq \bigcup_{\mathcal{P} \in \mathcal{F}_N} \bigcap_{\boldsymbol{n} \in \mathcal{N}} \overline{\mathcal{C}}_{\text{CD}}^{(N)}(\boldsymbol{n}, \mathcal{P}). \tag{82}$$

Now combining (82) with the achievability result which indicates that

$$\bigcup_{\mathcal{P} \in \mathcal{F}_N} \bigcap_{\boldsymbol{n} \in \mathcal{N}} \mathcal{C}_{\text{CD}}(\boldsymbol{n}, \mathcal{P}) \subseteq \bigcup_{\mathcal{P} \in \mathcal{F}} \bigcap_{\boldsymbol{n} \in \mathcal{N}} \mathcal{C}_{\text{CD}}(\boldsymbol{n}, \mathcal{P}) \subseteq \mathcal{C}_{\text{zero}}(\overline{P})$$

we obtain

$$\bigcup_{\mathcal{P} \in \mathcal{F}_N} \bigcap_{\boldsymbol{n} \in \mathcal{N}} \mathcal{C}_{\text{CD}}(\boldsymbol{n}, \mathcal{P}) \subseteq \mathcal{C}_{\text{zero}}(\overline{P}) \subseteq \bigcup_{\mathcal{P} \in \mathcal{F}_N} \bigcap_{\boldsymbol{n} \in \mathcal{N}} \overline{\mathcal{C}}_{\text{CD}}^{(N)}(\boldsymbol{n}, \mathcal{P}).$$

Since the lower and upper bounds converge as $N \to \infty$, it is clear that

$$\mathcal{C}_{\text{zero}}(\overline{P}) = \bigcup_{\mathcal{P} \in \mathcal{F}} \bigcap_{\boldsymbol{n} \in \mathcal{N}} \mathcal{C}_{\text{CD}}(\boldsymbol{n}, \mathcal{P}). \qquad \square$$

B. Necessary and Sufficient Condition for Equations in (8) to Have Positive Solutions

We show that $\det A > 0$ is the necessary and sufficient condition for the M linear equations in (8) to have positive solutions for all $P_i^{\min}(\boldsymbol{n})$ $(1 \le i \le M)$ in every fading state $\boldsymbol{n}$.

Necessary Condition: Since the equations in (8) are equivalent to

$$A \cdot \boldsymbol{P}^{\min}(\boldsymbol{n}) = B\boldsymbol{n}$$

if (8) has a solution for each state $\boldsymbol{n}$, then $\det A \neq 0$ so that the inverse of A exists and the solution to $\boldsymbol{P}^{\min}(\boldsymbol{n})$ is as given in (10). Furthermore, if $\boldsymbol{P}^{\min}(\boldsymbol{n}) > 0$ [i.e., $\forall 1 \le i \le M$, $P_i^{\min}(\boldsymbol{n}) > 0$] for any given $\boldsymbol{n} > 0$, then every component of A^{-1} must be nonnegative.

Define the cofactor matrix

$$\operatorname{cof} A = \left[(-1)^{i+j} \det A_{ij}\right], \qquad i, j = 1, 2, \ldots, M$$

where A_{ij} is the $(M-1)\times(M-1)$ matrix formed by deleting the ith row and the jth column of A. Then

$$A^{-1} = \frac{1}{\det A} \cdot (\operatorname{cof} A)^T. \tag{83}$$

Given $i < j$ fixed, since each component of the $(j-1)$th row of A_{ij} is -1, by subtracting the $(j-1)$th row from all other rows of A_{ij}, it is easily seen that the expansion by the ith column of A_{ij} is

$$\det A_{ij} = (-1)^{i+j} \prod_{k\neq i, j} (a_{kk}+1).$$

Thus,

$$(-1)^{i+j} \det A_{ij} = \prod_{k\neq i, j} (a_{kk}+1) > 0 \tag{84}$$

since $\forall 1 \le k \le M$, $a_{kk} > 0$. It can be similarly shown that (84) holds for $i > j$ as well. Therefore, if every component of A^{-1} must be nonnegative, from (83) we know that $\det A > 0$.

Sufficient Condition: Denote the cofactor matrix of A as $\operatorname{cof} A = C = (c_{ij})$, $i, j = 1, 2, \ldots, M$, then as shown above

$$c_{ij} = (-1)^{i+j} \det A_{ij} > 0, \qquad \forall i \neq j. \tag{85}$$

Since $\forall 1 \le k \le M$, $a_{kk} > 0$, and the expansion by the kth column of A is

$$\det A = a_{kk} \cdot c_{kk} - \sum_{i=1, i\neq k}^{M} c_{ik}$$

if $\det A > 0$, then

$$\begin{aligned} c_{kk} &= \frac{1}{a_{kk}} \cdot \left(\det A + \sum_{i=1, i\neq k}^{M} c_{ik}\right) \\ &> 0, \qquad \forall 1 \le k \le M. \end{aligned} \tag{86}$$

Therefore, combining (85) and (86) we get

$$c_{ij} > 0 \qquad \forall 1 \le i, j \le M.$$

By (83) it is clear that every component of A^{-1} must be positive. Consequently, from the expression for $\boldsymbol{P}^{\min}(\boldsymbol{n})$ in (10) we conclude that $\boldsymbol{P}^{\min}(\boldsymbol{n}) > 0$ in every fading state $\boldsymbol{n}$ $(\boldsymbol{n} > 0)$. $\square$

C. Proof of Lemma 1

Since for $k > 0$, $i \ge 0$,

$$\sum_{n=0}^{i+k-1} \binom{i+k-1}{n} = 2^{i+k-1}$$

we have

$$\binom{i+k-1}{i} \le 2^{i+k-1}.$$

For $0 < p < 1/2$

$$\begin{aligned} \sum_{i=k+r+1}^{\infty} \binom{i+k-1}{i} p^i(1-p)^k &\le \sum_{i=k+r+1}^{\infty} 2^{i+k-1} p^i (1-p)^k \\ &= \frac{(2p)^{k+r+1}}{1-2p}\, 2^{k-1}(1-p)^k \\ &= \frac{(2p)^{r+1}}{2(1-2p)}\, [4p(1-p)]^k. \end{aligned}$$

Because $4p(1-p) < [p+(1-p)]^2 = 1$

$$\lim_{k\to\infty} [4p(1-p)]^k = 0.$$

Therefore, for fixed integer r

$$\lim_{k\to\infty} \sum_{i=k+r+1}^{\infty} \binom{i+k-1}{i} p^i(1-p)^k = 0. \tag{87}$$

For a given positive integer k, we know that [13]

$$\sum_{i=0}^{\infty} \binom{i+k-1}{i} p^i(1-p)^k = 1. \tag{88}$$

Thus, from (87) and (88) we obtain

$$\begin{aligned} &\lim_{k\to\infty} \sum_{i=0}^{k+r} \binom{i+k-1}{i} p^i(1-p)^k \\ &\quad = 1 - \lim_{k\to\infty} \sum_{i=k+r+1}^{\infty} \binom{i+k-1}{i} p^i(1-p)^k = 1. \quad \square \end{aligned}$$

D. Proof of Lemma 2

Since for $n = 1, 2, \ldots$

$$\int_x^{\infty} t^{n-1} e^{-t}\, dt = \begin{cases} e^{-x}(n-1)! \displaystyle\sum_{i=0}^{n-1} \frac{x^i}{i!}, & \text{if } x > 0 \\ (n-1)!, & \text{if } x = 0 \end{cases} \tag{89}$$

we have

$$\int_0^{\infty} \frac{p_i(\gamma_i)}{\gamma_i}\, d\gamma_i = \frac{m}{m-1}\, \overline{n}_i B, \qquad i = 1, 2$$

where $p_i(\gamma_i)$ is given in (20). Thus,

$$D_a(m) + D_d(m) = \frac{m}{m-1}\, \overline{n}_1 B \tag{90}$$

$$D_b(m) + D_c(m) = \frac{m}{m-1}\, \overline{n}_2 B. \tag{91}$$

In (20), let $v_i = m\overline{n}_i B \gamma_i$, then

$$p_i(v_i) = \frac{1}{(m-1)!}\, v_i^{m-1} e^{-v_i}, \qquad i = 1, 2. \tag{92}$$

Substituting (92) into (22) and (23) and using (89), we have

$$\begin{aligned} &D_a(m) \\ &\quad = \int_0^{\infty} \int_{\frac{\overline{n}_1 B}{\overline{n}_2 B} v_2}^{\infty} \frac{m\overline{n}_1 B}{v_1}\, p_1(v_1) p_2(v_2)\, dv_1\, dv_2 \\ &\quad = \int_0^{\infty} \left\{ \frac{m}{m-1}\, \overline{n}_1 B e^{-\frac{\overline{n}_1 B}{\overline{n}_2 B} v_2} \sum_{i=0}^{m-2} \frac{\left(\frac{\overline{n}_1 B}{\overline{n}_2 B} v_2\right)^i}{i!} \right\} \\ &\qquad \times p_2(v_2)\, dv_2 \end{aligned}$$

$$= \frac{m}{m-1}\,\overline{n}_1 B \sum_{i=0}^{m-2} \frac{(i+m-1)!}{(m-1)!\,i!} \left(\frac{\overline{n}_1 B}{\overline{n}_1 B + \overline{n}_2 B}\right)^i \times \left(\frac{\overline{n}_2 B}{\overline{n}_1 B + \overline{n}_2 B}\right)^m \tag{93}$$

$$\begin{aligned} D_b(m) &= \int_0^\infty \int_{\frac{\overline{n}_1 B}{\overline{n}_2 B} v_2}^\infty \frac{m\overline{n}_2 B}{v_2}\, p_1(v_1) p_2(v_2)\, dv_1\, dv_2 \\ &= \int_0^\infty \left\{ e^{-\frac{\overline{n}_1 B}{\overline{n}_2 B} v_2} \sum_{i=0}^{m-1} \frac{\left(\frac{\overline{n}_1 B}{\overline{n}_2 B} v_2\right)^i}{i!} \right\} \frac{m\overline{n}_2 B}{v_2}\, p_2(v_2)\, dv_2 \\ &= \frac{m}{m-1}\,\overline{n}_2 B \sum_{i=0}^{m-1} \frac{(i+m-2)!}{(m-2)!\,i!} \left(\frac{\overline{n}_1 B}{\overline{n}_1 B + \overline{n}_2 B}\right)^i \times \left(\frac{\overline{n}_2 B}{\overline{n}_1 B + \overline{n}_2 B}\right)^{(m-1)}. \end{aligned} \tag{94}$$

When $\overline{n}_1 B < \overline{n}_2 B$, applying Lemma 1 to (93) and (94), we have

$$\lim_{m\to\infty} D_a(m) = \overline{n}_1 B$$
$$\lim_{m\to\infty} D_b(m) = \overline{n}_2 B.$$

Combining these limiting expressions with (90) and (91) we obtain

$$\lim_{m\to\infty} D_c(m) = 0$$
$$\lim_{m\to\infty} D_d(m) = 0.$$

E. Proof of Theorem 3

In the following, $\forall 1 \le m_0 \le N, \forall 1 \le j \le m_0$, for simplicity we denote $\lambda_j(\boldsymbol{n})$, $v_j(\boldsymbol{n})$, $z_j(\boldsymbol{n})$, and $u_j(\boldsymbol{n})$ as λ_j, v_j, z_j, and u_j, respectively.

For $m_0 = 1, 2, \ldots, N, \forall \boldsymbol{n} \in \Omega_{m_0}$, let

$$\boldsymbol{u^*} \triangleq [u_1^*, u_2^*, \ldots, u_{m_0}^*]$$
$$\boldsymbol{u} \triangleq [u_1, u_2, \ldots, u_{m_0}].$$

Since $\boldsymbol{u^*}$ satisfies

$$\begin{aligned} &\sum_{m_0=1}^N E_{\boldsymbol{n}\in\Omega_{m_0}}\left[\sum_{i=1}^{m_0} v_i u_i^*\right] \\ &= P(s^*) + \sum_{m_0=1}^N \sum_{i=1}^{m_0} [\tau^* v_j + (1-\tau^*) v_{j-1}] \Pr(\tilde{L}_j(m_0, s^*)) \\ &= \overline{P} \end{aligned}$$

$\forall \boldsymbol{u} \ne \boldsymbol{u^*}$, by denoting

$$A(m_0) \triangleq \left\{ \boldsymbol{n}\colon \boldsymbol{n} \in \Omega_{m_0},\, \frac{1}{s^*} > z_1 \right\}$$

we have

$$\begin{aligned} &\sum_{m_0=1}^N E_{\boldsymbol{n}\in\Omega_{m_0}}\left[\sum_{i=1}^{m_0} v_i u_i\right] - \overline{P} \\ &= \sum_{m_0=1}^N E_{\boldsymbol{n}\in\Omega_m}\left[\sum_{i=1}^{m_0} v_i u_i\right] - \sum_{m_0=1}^N E_{\boldsymbol{n}\in\Omega_{m_0}}\left[\sum_{i=1}^{m_0} v_i u_i^*\right] \\ &= \sum_{m_0=1}^N \int_{\boldsymbol{n}\in A(m_0)} \sum_{i=1}^{m_0} v_i u_i dF(\boldsymbol{n}) \\ &\quad + \sum_{m_0=1}^N \sum_{j=1}^{m_0} \int_{\boldsymbol{n}\in L_j(m_0, s^*)} \left(\sum_{i=1}^{m_0} v_i u_i - v_j\right) dF(\boldsymbol{n}) \\ &\quad + \sum_{m_0=1}^N \sum_{j=1}^{m_0} \left[\sum_{i=1}^{m_0} v_i u_i - v_j \tau^* - v_{j-1}(1-\tau^*)\right] \\ &\quad \times \Pr(\tilde{L}_j(m_0, s^*)). \end{aligned} \tag{95}$$

We now show from (60)–(63) that $\forall \boldsymbol{n} \in \Omega_{m_0}$

$$\frac{\lambda_i}{v_i} < \frac{\lambda_{i-1}}{v_{i-1}}, \qquad i = 2, 3, \ldots, m_0. \tag{96}$$

Since $z_1 > z_2$, i.e., $\frac{\lambda_1}{v_1} > \frac{\lambda_2-\lambda_1}{v_2-v_1}$, (96) is true for $i = 2$. Assuming that when $i = k$, (96) is true, or equivalently, $\frac{\lambda_k-\lambda_{k-1}}{v_k-v_{k-1}} < \frac{\lambda_k}{v_k}$, then because $z_{k+1} < z_k$, we have

$$\frac{\lambda_{k+1}-\lambda_k}{v_{k+1}-v_k} < \frac{\lambda_k-\lambda_{k-1}}{v_k-v_{k-1}} < \frac{\lambda_k}{v_k}.$$

That is,

$$\frac{\lambda_{k+1}}{v_{k+1}} < \frac{\lambda_k}{v_k}.$$

Thus, when $i = k+1$, (96) is also true. So (96) is true for all $i = 2, 3, \ldots, m_0$. Continuing the derivation in (95), we have

a) $\forall \boldsymbol{n} \in A(m_0)$, since $\frac{1}{s^*} > z_1 = \frac{\lambda_1}{v_1}$, from (96) we know that $\forall i = 1, 2, \ldots, m_0$, $\frac{1}{s^*} > \frac{\lambda_i}{v_i}$, i.e., $v_i > \lambda_i s^*$. Therefore, the first term in (95) will be

$$\begin{aligned} &\sum_{m_0=1}^N \int_{\boldsymbol{n}\in A(m_0)} \sum_{i=1}^{m_0} v_i u_i\, dF(\boldsymbol{n}) \\ &> \sum_{m_0=1}^N \int_{\boldsymbol{n}\in A(m_0)} \sum_{i=1}^{m_0} \lambda_i s^* u_i\, dF(\boldsymbol{n}) \\ &= s^* \sum_{m_0=1}^N \int_{\boldsymbol{n}\in A(m_0)} \sum_{i=1}^{m_0} \lambda_i u_i\, dF(\boldsymbol{n}). \end{aligned} \tag{97}$$

b) $\forall \boldsymbol{n} \in L_j(m_0, s^*)$, $j = 1, 2, \ldots, m$, since $z_j > \frac{1}{s^*} > z_{j+1}$, we have

$$v_{j+1} > v_j + (\lambda_{j+1} - \lambda_j)s^*$$
$$v_{j-1} > v_j + (\lambda_{j-1} - \lambda_j)s^*.$$

For $i \ne j$, $1 \le i \le m_0$, denoting

$$x_i \triangleq \frac{\lambda_j - \lambda_i}{v_j - v_i}$$

it is easy to verify that $x_i > x_{i+1}$ by using (96). Since $x_{j-1} = z_j$, $x_{j+1} = z_{j+1}$, $\forall i = 1, 2, \ldots, j-1$, $\forall k = j+1, j+2, \ldots, m_0$, we have

$$x_i > \frac{1}{s^*} > x_k.$$

Therefore, $\forall i \ne j$, $1 \le i \le m_0$,

$$v_i > v_j + (\lambda_i - \lambda_j)s^*.$$

In (95), the second term will be

$$\sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\int_{\boldsymbol{n}\in L_j(m_0,\,s^*)}\left(\sum_{i=1}^{m_0} v_iu_i - v_j\right)\,dF(\boldsymbol{n})$$
$$> \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\int_{\boldsymbol{n}\in L_j(m_0,\,s^*)}\left(\sum_{i\neq j,\, i=1}^{m_0}[v_j+(\lambda_i-\lambda_j)s^*]u_i + v_ju_j - v_j\right)\,dF(\boldsymbol{n})$$
$$= \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\int_{\boldsymbol{n}\in L_j(m_0,\,s^*)}\left(\sum_{i\neq j,\, i=1}^{m_0}(\lambda_i-\lambda_j)s^*u_i\right)\,dF(\boldsymbol{n})$$
$$= \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\int_{\boldsymbol{n}\in L_j(m_0,\,s^*)}\left(\sum_{i\neq j,\, i=1}^{m_0}\lambda_is^*u_i - \lambda_js^*[1-u_j]\right)\,dF(\boldsymbol{n})$$
$$= s^*\sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\int_{\boldsymbol{n}\in L_j(m_0,\,s^*)}\left(\sum_{i\neq j,\, i=1}^{m}\lambda_iu_i - \lambda_j[1-u_j]\right)\,dF(\boldsymbol{n}). \tag{98}$$

c) $\forall\boldsymbol{n}\in\tilde{L}_j(m_0,\,s^*)$, $j=1,\,2,\,\ldots,\,m_0$, since $\frac{1}{s^*}=z_j=x_{j-1}$, we have

$$v_{j-1}=v_j+(\lambda_{j-1}-\lambda_j)s^*.$$

Because $\forall\, i=1,2,\ldots,j-2, \forall\, k=j+1,j+2,\ldots,m_0$,

$$x_i>\frac{1}{s^*}>x_k$$

$\forall\, i\neq j-1,\, j$ and $1\leq i\leq m_0$, we have

$$v_i>v_j+(\lambda_i-\lambda_j)s^*.$$

Therefore, in (95), the third term will be

$$\sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\left[\sum_{i=1}^{m_0}v_iu_i - v_j\tau^* - v_{j-1}(1-\tau^*)\right]\times\Pr(\tilde{L}_j(m_0,\,s^*))$$
$$> \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\left[\sum_{i\neq j,\, i=1}^{m_0}\{v_j+(\lambda_i-\lambda_j)s^*\}u_i + v_ju_j - v_j\tau^* - v_{j-1}(1-\tau^*)\right]\Pr(\tilde{L}_j(m_0,\,s^*))$$
$$= \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\left[\sum_{i\neq j,\, i=1}^{m_0}(\lambda_i-\lambda_j)s^*u_i + (v_j-v_{j-1})(1-\tau^*)\right]\times\Pr(\tilde{L}_j(m_0,\,s^*))$$
$$= \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\left[\left(\sum_{i\neq j,\, i=1}^{m_0}\lambda_is^*u_i\right) - \lambda_js^*(1-u_j) + (\lambda_j-\lambda_{j-1})(1-\tau^*)s^*\right]\Pr(\tilde{L}_j(m_0,\,s^*))$$
$$= \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\left\{\left(\sum_{i\neq j-1,\, i\neq j,\, i=1}^{m_0}\lambda_is^*u_i\right) + \lambda_js^*(u_j-\tau^*) + \lambda_{j-1}s^*[u_{j-1}-(1-\tau^*)]\right\}\times\Pr(\tilde{L}_j(m_0,\,s^*))$$
$$= s^*\sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\left\{\left(\sum_{i\neq j-1,\, i\neq j,\, i=1}^{m_0}\lambda_iu_i\right) + \lambda_j(u_j-\tau^*) + \lambda_{j-1}[u_{j-1}-(1-\tau^*)]\right\}\times\Pr(\tilde{L}_j(m_0,\,s^*)). \tag{99}$$

Therefore, substituting (97)–(99) into (95), we have

$$\sum_{m_0=1}^{N}E_{\boldsymbol{n}\in\Omega_{m_0}}\left[\sum_{i=1}^{m_0}v_iu_i\right]-\overline{P}$$
$$> s^*\sum_{m_0=1}^{N}\left\{\int_{\boldsymbol{n}\in A(m_0)}\sum_{i=1}^{m_0}\lambda_iu_i\,dF(\boldsymbol{n}) + \sum_{j=1}^{m_0}\int_{\boldsymbol{n}\in L_j(m_0,\,s^*)}\left(\sum_{i\neq j,\, i=1}^{m_0}\lambda_iu_i-\lambda_j[1-u_j]\right)\,dF(\boldsymbol{n}) + \sum_{j=1}^{m_0}\left[\left(\sum_{i\neq j-1,\, i\neq j,\, i=1}^{m_0}\lambda_iu_i\right)+\lambda_{j-1}[u_{j-1}-(1-\tau^*)] + \lambda_j(u_j-\tau^*)\right]\Pr(\tilde{L}_j(m_0,\,s^*))\right\}$$
$$= s^*\left\{\sum_{m_0=1}^{N}E_{\boldsymbol{n}\in\Omega_{m_0}}\left[\sum_{i=1}^{m_0}\lambda_iu_i\right] - \sum_{m_0=1}^{N}E_{\boldsymbol{n}\in\Omega_{m_0}}\left[\sum_{i=1}^{m_0}\lambda_iu_i^*\right]\right\}.$$

Thus, if

$$\sum_{m_0=1}^{N}E_{\boldsymbol{n}\in\Omega_{m_0}}\left[\sum_{i=1}^{m_0}\lambda_iu_i\right] > \sum_{m_0=1}^{N}E_{\boldsymbol{n}\in\Omega_{m_0}}\left[\sum_{i=1}^{m_0}\lambda_iu_i^*\right]$$

then

$$\sum_{m_0=1}^{N}E_{\boldsymbol{n}\in\Omega_{m_0}}\left[\sum_{i=1}^{m_0}v_iu_i\right] > \overline{P}$$

which means that $\boldsymbol{u}$ does not satisfy the total power constraint. Therefore, $\forall\, \boldsymbol{u} \neq \boldsymbol{u}^*$

$$\sum_{m_0=1}^{N} E_{\boldsymbol{n}\in\Omega_{m_0}}\left[\sum_{i=1}^{m_0}\lambda_i u_i\right] \leq \sum_{m_0=1}^{N} E_{\boldsymbol{n}\in\Omega_{m_0}}\left[\sum_{i=1}^{m_0}\lambda_i u_i^*\right]. \quad \square$$

ACKNOWLEDGMENT

The authors wish to thank the Editor and the anonymous reviewers for their very helpful suggestions and comments.

REFERENCES

[1] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.
[2] D. N. Tse and S. V. Hanly, "Multiple-access fading channels—Part I: Polymatroidal structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2796–2815, Nov. 1998.
[3] D. Hughes-Hartogs, "The capacity of a degraded spectral Gaussian broadcast channel," Ph.D. dissertation, Inform. Syst. Lab., Ctr. Syst. Res., Stanford Univ., Stanford, CA, July 1975.
[4] D. N. Tse, "Optimal power allocation over parallel Gaussian broadcast channels," *IEEE Trans. Inform. Theory*, submitted for publication.
[5] S. V. Hanly and D. N. Tse, "Multiple-access fading channels—Part II: Delay-limited capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2816–2831, Nov. 1998.
[6] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1468–1489, July 1999.
[7] S. H. Jamali and T. Le-Ngoc, *Coded-Modulation Techniques for Fading Channels*. Norwell, MA: Kluwer, 1994.
[8] A. J. Goldsmith, "The capacity of downlink fading channels with variable-rate and power," *IEEE Trans. Vehic. Technol.*, vol. 46, pp. 569–580, Aug. 1997.
[9] M. Nakagami, "The m-distribution—A general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*. Oxford, U.K.: Pergamon, 1960, pp. 3–36.
[10] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 197–207, Mar. 1973.
[11] ——, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 279–280, Mar. 1974.
[12] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Probl. Pered. Inform.*, vol. 10, no. 3, pp. 3–14, July–Sept. 1974.
[13] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. New York: Wiley, 1968, vol. 1.