Outage Capacities and Optimal Power Allocation for Fading Multiple-Access Channels

Lifang Li, Nihar Jindal, *Member, IEEE*, and Andrea Goldsmith, *Fellow, IEEE*

***Abstract*—We derive the outage capacity region of an M-user fading multiple-access channel (MAC) under the assumption that both the transmitters and the receiver have perfect channel side information (CSI). The outage capacity region is implicitly obtained by deriving the outage probability region for a given rate vector. Given a required rate and average power constraint for each user, we find a successive decoding strategy and a power allocation policy that achieves points on the boundary of the outage probability region. We discuss the scenario where an outage must be declared simultaneously for all users (common outage) and when outages can be declared individually (individual outage) for each user.**

***Index Terms*—Capacity region, fading channels, multiple-access channels (MACs), optimal power allocation, outage probability.**

I. Introduction

Wireless communication channels vary over time due to user mobility. By applying optimal dynamic power and rate allocation strategies, the 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 are obtained in [1], [2], and [3], respectively.[1] These results have also been extended to the fading multiple-antenna multiple-access and broadcast channels in [4], [5]. The Shannon capacity implies no complexity or delay constraints, and is obtained by varying the transmit power and possibly the rate relative to the channel fading conditions such that the average rate is maximized. For delay-constrained applications, Shannon capacity is not a good performance measure, since the transmission delay depends on the channel variation. Thus, a better performance measure for such systems is the zero-outage capacity, defined as the maximum instantaneous mutual information rate that can be maintained under all fading conditions through optimal power control. Under the assumption that CSI is available at both the transmitter and the receiver, the zero-outage capacity regions and the corresponding optimal power allocation schemes are derived for the fading MAC and the fading broadcast channel in [6] and [7], respectively.[2]

Zero-outage capacity is somewhat pessimistic, however, since a constant rate must be maintained under any fading condition. By allowing some transmission outage during severe fading conditions, the maximum mutual information rate that can be kept constant during nonoutage increases. This motivates the investigation of outage channel capacity, defined as the maximum instantaneous information rate that can be maintained under any fading condition during nonoutage such that the allowed average transmission outage probability is satisfied.

Outage capacity is most relevant in a slow-fading environment, where the channel can be assumed to be constant over the duration of a codeword. If only the receiver has CSI in a point-to-point channel, then the transmitter always transmits at a constant rate and cannot use any form of power control. In this scenario, an outage occurs whenever the channel cannot support transmission at the designated constant rate, i.e., whenever the instantaneous mutual information is less than the rate of transmission. Thus, the outage probability is equal to the probability that the channel cannot support the given rate, which is roughly equal to the probability of a decoding error when a channel code designed for a given rate is used on a channel with instantaneous mutual information below this given rate [8]. Alternatively, the outage probability can be viewed as the fraction of time that an incorrect codeword is received.

If both the receiver and the transmitter have CSI, the transmitter can use power control to conserve power by not transmitting at all during designated outage periods and by varying the amount of transmit power during nonoutages such that the instantaneous mutual information is exactly equal to (instead of exceeding) the rate of transmission. Here, the outage probability is described as the probability of not transmitting/receiving a codeword at all, instead of as the probability of decoding error, as is necessarily the case without transmitter CSI. We can also view the outage probability as the fraction of time that no codeword is received, which is relevant to many practical scenarios (e.g., in cellular systems, mobile units outside the service range of a base station are said to be in outage).

For a MAC, the same interpretations for outage probability and capacity hold. With or without transmitter CSI, the outage probability of User i is approximately equal to the fraction of time that an incorrect or no codeword is received from User i.

Manuscript received September 29, 1999; revised December 20, 2004. This work was supported by the National Science Foundation Career Award NCR-9501452, a grant from Pacific Bell, and by ONR under Grants N00014-99-1-0578 and N00014-02-1-0003. The material in this paper was presented in part at the IEEE Wireless Communications and Networking Conference (WCNC'99), New Orleans, LA, September 1999 and at the IEEE International Symposium on Information Theory, Sorrento, Italy, June 2000

L. Li is with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109-8099 USA (e-mail: lifang@systems.caltech.edu).

N. Jindal is with the University of Minnesota, Minneapolis, MN 55455 USA (nihar@ece.umn.edu).

A. Goldsmith is with Stanford University, Stanford, CA 94305-9515 USA (andrea@ee.stanford.edu).

Communicated by S. Shamai, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2005.844065

[1]The Shannon capacity of a fading channel is called "throughput capacity" in [2], and "ergodic capacity" in [3].

[2]The zero-outage capacity is called "delay-limited capacity" in [6].

A motivating example is an uplink channel in which each transmitter wishes to send constant-rate video to the base station. Severe fading may preclude sufficiently high-rate communication from occurring at all times, but high-rate communication may be possible 95% of the time.

In [8], the minimum outage probability problem is solved for the single-user fading channel. For an M-user fading broadcast channel, under different assumptions about whether the transmission to all users is turned off simultaneously or individually, the optimal power allocation strategy that minimizes the common outage probability or achieves the boundary of the outage probability region of the M users under a total average power constraint of all users is derived in [7]. An alternative notion of capacity combining the ideas of outage and ergodic capacity, referred to as the service outage capacity and minimum-rate outage capacity, has also been recently considered [9]–[12].

In this paper, we derive the outage capacity region and the optimal power allocation policies for an M-user fading MAC under similar assumptions about whether the outage declaration from each user is simultaneous or individual. The outage capacity region is explicitly defined in terms of the achievable rates corresponding to the set of all power policies meeting the individual power constraints. Essentially, a user is in outage whenever his power is equal to zero, and is transmitting at the nonoutage rate at all other times. This simple definition of outage capacity gives a unified framework that allows us to easily treat the cases of common and individual outage. Though no explicit coding theorem and converse are given in this paper, operational meaning is given to the definition of the outage capacity region by relating the outage capacity to the zero-outage capacity, for which a rigorous coding theorem and converse exists. Intuitively, the outage capacity is the set of all rates achievable in all nonoutage fading states. Thus, the outage capacity can be related to the zero-outage capacity of the conditional distribution of the fading states, where the conditioning is on a nonoutage event. For the single-user channel, it is intuitively easy to see that the outage states should be the fading states with the smallest amplitude [8]. However, for multiuser channels, no such simple ordering of the joint fading states is possible, and the difficulty remains in determining what set of states should be set as outage states.

The zero-outage capacity of the fading MAC is derived in [6]. Specifically, it is shown in [6] that the zero-outage capacity region is implicitly obtained by determining, for each given rate vector $\boldsymbol{R} = (R_1, \ldots, R_M)$, the set of average transmit powers such that each user can support rate R_i under any fading condition. For the general case where the allowed outage probability of each user is larger than zero, we will show that the outage capacity region is implicitly obtained by determining, for each given rate vector $\boldsymbol{R}$, the set of all common outage probabilities or individual outage probability vectors such that each user can support rate R_i under any nonoutage fading condition while satisfying his given average power constraint. Given the allowed outage probability of each user and a rate vector $\boldsymbol{R}$, we also solve the dual problem of finding the average power region of the M users required to support $\boldsymbol{R}$ for the given outage probability vector.

For a given rate vector $\boldsymbol{R}$, in order to solve the optimization problem of minimizing the common outage probability or bounding the outage probability region for a given average power constraint on each of the M users, we use the Lagrangian method with multiple constraints. Since there is an independent average power constraint for each of the M users, M Lagrangian multipliers are needed. For each given Lagrangian multiplier vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_M)$, the optimization problem is readily solved by applying the techniques developed in [6] and [7]. Standard convex optimization algorithms can be used to find the appropriate Lagrangian multiplier vector for the M users such that the M average power constraints are satisfied simultaneously.

The remainder of this paper is organized as follows. In Section II, we present the fading MAC model. In Sections III and IV, we give the definitions and notations that will be used in the rest of the paper. In Section V, for the case of simultaneous outage declaration, we derive the minimum common outage probability for a given rate vector $\boldsymbol{R}$ and the corresponding optimal power allocation strategy. In addition, the average power region for supporting $\boldsymbol{R}$ with a given common outage probability is obtained. As for individual outage declaration, we derive in Section VI the outage probability region boundary for a given rate vector $\boldsymbol{R}$, the corresponding optimal power allocation strategy, and the required average power region for supporting $\boldsymbol{R}$ with the given outage probability constraint of each user. In Section VII, we discuss the relationship between outage capacities of the MAC and the broadcast channel. In Section VIII, we present the main difference in the solutions to the above problems when additional peak power constraints are imposed on the M users. Our conclusions are given in Section IX.

II. The Fading Multiple-Access Channel (MAC)

We consider a discrete-time M-user fading MAC model as discussed in [6]

$$Y(n) = \sum_{i=1}^{M} \sqrt{H_i(n)} X_i(n) + Z(n) \tag{1}$$

where $X_i(n)$ and $H_i(n)$ are the transmitted signal and the fading process of the ith user, respectively, and $Z(n)$ is Gaussian noise with variance σ^2. Let $\boldsymbol{h}(n) = [H_1(n), \ldots, H_M(n)]$ denote the joint fading process, and let $\overline{P}_i$ be the average power constraint of User i. We assume that the joint fading process of the M users is stationary and ergodic, and the stationary distribution has continuous density.[3] For a slowly time-varying MAC, let $\boldsymbol{h} = (h_1, \ldots, h_M)$ be the joint fading state at a particular time n, i.e., $\boldsymbol{h}(n) = \boldsymbol{h}$, and let $\mathcal{H}_{\text{all}}$ denote the set of all possible joint fading states. We assume that the M transmitters and the receiver know the current joint fading state $\boldsymbol{h}$. Therefore, each transmitter can vary its transmit power and codewords relative to the joint fading condition of the M channels, and the receiver can vary its decoding order of the M users.

[3]As in the single-user case [8] and the broadcast communication case [7], our analysis can be easily extended to discrete distributions.

Notation: In this paper, we use boldface letters to denote M-dimensional vector quantities. In addition, all operations and inequalities for vectors are defined element-wise, i.e., $\boldsymbol{P} \leq \boldsymbol{Q}$ implies $P_i \leq Q_i$ for $i = 1, \ldots, M$, and $\boldsymbol{PQ} = (P_1Q_1, \ldots, P_MQ_M)$. The only exception to this is the inner product, which is defined in the standard manner

$$\boldsymbol{P} \cdot \boldsymbol{Q} = \sum_{i=1}^{M} P_i Q_i.$$

Finally, all expected values of random variables and probabilities of random events are assumed to be calculated with respect to both the randomization within each joint fading state and the randomization across all joint fading states, unless otherwise noted.

III. Outage Capacity Regions

In this section, we define the outage capacity region of an M-user MAC, where each transmitter may suspend transmission over a subset of fading states under a given average power constraint and an average outage probability constraint. We consider common outage and individual outage separately. Both outage capacity regions are defined based on power allocation policies and the corresponding achievable rates.

We define a power allocation policy over all possible fading states as a mapping from each fading state $\boldsymbol{h}$ to a set of random transmit powers for the M users, $\forall\, \boldsymbol{h} \in \mathcal{H}_{\text{all}}$. For a fully randomized power allocation policy $\mathcal{P}^{\text{rand}}$, the set of M transmit powers varies within each fading state $\boldsymbol{h}$ as well as across all fading states. Specifically, let $\boldsymbol{P}^{\text{rand}}(\boldsymbol{h})$ denote the set of random transmit power allocation functions for the M users, and let $f_{\boldsymbol{P}}^{\text{rand}}(\boldsymbol{h})$ denote their joint probability density function (PDF)[4] in the fading state $\boldsymbol{h}$, $\forall\, \boldsymbol{h} \in \mathcal{H}_{\text{all}}$. Then $\mathcal{P}^{\text{rand}}$ can be expressed as follows:

$$\mathcal{P}^{\text{rand}} : \boldsymbol{h} \rightarrow \left\{\boldsymbol{P}^{\text{rand}}(\boldsymbol{h}),\ f_{\boldsymbol{P}}^{\text{rand}}(\boldsymbol{h})\right\}, \qquad \forall\, \boldsymbol{h} \in \mathcal{H}_{\text{all}}. \quad (2)$$

Note that for each different fading state $\boldsymbol{h}$, the vector of random transmit powers $\boldsymbol{P}^{\text{rand}}(\boldsymbol{h})$ can have a different sample space, and the joint PDF $f_{\boldsymbol{P}}^{\text{rand}}(\boldsymbol{h})$ can be a different distribution. Considerations of the set of fully randomized power allocation policies allow for complete generality when defining outage capacities of the fading MAC, and also mirrors the approach taken in [8] for single-user outage capacity.

In the slowly fading environment that we are concerned with, if a deterministic power vector $\boldsymbol{P} = (P_1, \ldots, P_M)$ is allocated to the M users for a given fading state $\boldsymbol{h}$, then the following rate vectors are achievable in this given state $\boldsymbol{h}$:

$$\mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}) = \left\{ \boldsymbol{R} : \sum_{j \in S} R_j \leq \frac{1}{2} \log\left(1 + \frac{1}{\sigma^2} \sum_{j \in S} h_j P_j\right), \ \forall\, S \subseteq \{1, \ldots, M\} \right\}. \quad (3)$$

[4]When we refer to probability functions in generic terms, we use the term PDF to mean both discrete and continuous probability functions [13].

Note that $\mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P})$ is actually the capacity region of the equivalent Gaussian MAC for the fading state $\boldsymbol{h}$. This capacity region is an M-dimensional polyhedron with $M!$ corner points. Maximum-likelihood decoding can be used to achieve any rate vector in the capacity region, but the more computationally efficient technique of successive decoding is sufficient to achieve the corner points. We will later see that we only need to operate at these corner points in order to achieve the outage capacity.

If a *random* power allocation vector $\boldsymbol{P}^{\text{rand}}(\boldsymbol{h})$ is employed in fading state $\boldsymbol{h}$, then each possible value of the power allocation vector corresponds to a different rate region given by the above equation. In this case, $\mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h}))$ refers to the corresponding achievable rate region that varies based on the joint PDF $f_{\boldsymbol{P}}^{\text{rand}}(\boldsymbol{h})$ of the M random transmit powers in fading state $\boldsymbol{h}$. For example, if the power policy equiprobably chooses between two different power vectors in a certain fading state, then the system could operate at a rate vector in the MAC capacity region corresponding to the first power vector 50% of the time, and could operate at a rate vector in the MAC capacity region corresponding to the second power vector the remaining 50% of the time.

Using the framework provided by the above definitions of a random power policy and the corresponding achievable rate regions, we are able to precisely define common and individual outage capacity regions.

A. Common Outage Capacity Region

The common outage capacity region is defined as follows.

Definition 3.1: A rate vector $\boldsymbol{R}$ is in the common outage capacity region $\mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, Pr)$ if and only if there exists a random power policy $\mathcal{P}^{\text{rand}}$ that meets the power constraint $E[\boldsymbol{P}^{\text{rand}}(\boldsymbol{h})] \leq \overline{\boldsymbol{P}}$ and allows for the rate vector to be achieved with a probability of at least $1 - Pr$:

$$Pr[\boldsymbol{R} \in \mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h}))] \geq 1 - Pr. \quad (4)$$

That is,

$$\mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, Pr) \triangleq \bigcup_{\mathcal{P}^{\text{rand}} : E[\boldsymbol{P}^{\text{rand}}(\boldsymbol{h})] \leq \overline{\boldsymbol{P}}} \{\boldsymbol{R} : Pr[\boldsymbol{R} \in \mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h}))] \geq 1 - Pr\}. \quad (5)$$

Therefore, the common outage capacity region consists of all rate vectors that can be maintained (with arbitrarily small probability of error) with a common outage probability no larger than $\boldsymbol{Pr}$ under the average power constraint $\overline{\boldsymbol{P}}$. Though this definition is given only in terms of random power policies, we later relate the outage capacity to the zero-outage capacity to give an operational meaning to our definition. Notice that, in this definition, we have allowed for a fully randomized power policy $\mathcal{P}^{\text{rand}}$. However, as we will show later, the random power policy $\mathcal{P}^{\text{rand}}$ need only have cardinality of two in each fading state. In particular, in each given state $\boldsymbol{h} \in \mathcal{H}_{\text{all}}$, let $\boldsymbol{P}^{\text{rand}}(\boldsymbol{h}) = \boldsymbol{P}(\boldsymbol{h})$ with probability $w(\boldsymbol{h})$, and let $\boldsymbol{P}^{\text{rand}}(\boldsymbol{h}) = \boldsymbol{0}$ with probability $1 - w(\boldsymbol{h})$, where $\boldsymbol{P}(\boldsymbol{h})$ is a vector of deterministic power allocation functions of $\boldsymbol{h}$, and $w(\boldsymbol{h})$ is a deterministic probability function of $\boldsymbol{h}$, $0 \leq w(\boldsymbol{h}) \leq 1$. We denote this simple power allocation policy with cardinality of two in each fading state as $\mathcal{P}$ (i.e., without the superscript "rand")

$$\mathcal{P}: \boldsymbol{h} \rightarrow \begin{cases} \boldsymbol{0}, & \text{with prob. } 1-w(\boldsymbol{h}) \\ \boldsymbol{P}(\boldsymbol{h}), & \text{with prob. } w(\boldsymbol{h}) \end{cases} \qquad \forall\, \boldsymbol{h} \in \mathcal{H}_{\text{all}}. \tag{6}$$

In this power allocation policy $\mathcal{P}$, since $w(\boldsymbol{h})$ is the probability that the M users will be transmitting with the allocated power vector $\boldsymbol{P}(\boldsymbol{h})$, and $1-w(\boldsymbol{h})$ is the probability that no power will be allocated to any user (i.e., an outage will be declared from all users), we call $w(\boldsymbol{h})$ the probability of transmission function, $\forall\, \boldsymbol{h} \in \mathcal{H}_{\text{all}}$. The following proposition shows that it is sufficient to consider random power policies of cardinality two in each fading state in the definition of the common outage capacity region.

Proposition 3.1:

$$\mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, Pr) = \bigcup_{\mathcal{P} \in \mathcal{F}} \bigcap_{\{\boldsymbol{h}: w(\boldsymbol{h})>0\}} \mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h})) \tag{7}$$

where $\mathcal{F}$ is the set of all power policies $\mathcal{P}$ as defined in (6) that satisfy the conditions

$$E[\boldsymbol{P}(\boldsymbol{h})w(\boldsymbol{h})] \leq \overline{\boldsymbol{P}} \tag{8}$$

$$E[w(\boldsymbol{h})] \geq 1 - Pr. \tag{9}$$

Proof: See Part A of the Appendix. □

Since all M users must transmit at their specified rates during nonoutage periods, if we consider only nonoutage fading states, it appears as if constant rates were being maintained at all times, which is similar to the zero-outage case. Therefore, the common outage capacity region can be given in terms of the zero-outage capacity region of the MAC by conditioning on the nonoutage fading states. Specifically, let $p(\boldsymbol{h})$ denote the true PDF of the fading state $\boldsymbol{h} \in \mathcal{H}_{\text{all}}$. For a given power policy $\mathcal{P} \in \mathcal{F}$, as defined in Proposition 3.1, let $\mathcal{H}_{\text{tran}}$ denote the set of nonoutage fading states (transmission states), i.e., $\mathcal{H}_{\text{tran}} = \{\boldsymbol{h} : w(\boldsymbol{h}) > 0\}$. We now define a new PDF $g(\boldsymbol{h})$ for $\boldsymbol{h} \in \mathcal{H}_{\text{all}}$ as follows:

$$g(\boldsymbol{h}) = \begin{cases} \frac{1}{E[w(\boldsymbol{h})]} p(\boldsymbol{h}) w(\boldsymbol{h}), & \boldsymbol{h} \in \mathcal{H}_{\text{tran}} \\ 0, & \boldsymbol{h} \notin \mathcal{H}_{\text{tran}}. \end{cases} \tag{10}$$

It is easily verified that

$$\int_{\boldsymbol{h} \in \mathcal{H}_{\text{all}}} g(\boldsymbol{h})\, d\boldsymbol{h} = \int_{\boldsymbol{h} \in \mathcal{H}_{\text{tran}}} g(\boldsymbol{h})\, d\boldsymbol{h} = 1.$$

Proposition 3.2: The common outage capacity region defined in Definition 3.1 can be written in terms of the zero-outage capacity region as

$$\mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, Pr) = \bigcup_{\{w(\boldsymbol{h}): E[w(\boldsymbol{h})] = 1 - Pr\}} \mathcal{C}_{\text{zero}}\left(\frac{\overline{\boldsymbol{P}}}{1-Pr}, \mathcal{H}_{\text{tran}}\right) \tag{11}$$

where $\mathcal{C}_{\text{zero}}\left(\overline{\boldsymbol{P}}/(1-Pr), \mathcal{H}_{\text{tran}}\right)$ refers to the zero-outage capacity region of the equivalent fading MAC for which the set of all possible fading states is $\mathcal{H}_{\text{tran}}$, and the PDF of $\boldsymbol{h} \in \mathcal{H}_{\text{tran}}$ is $g(\boldsymbol{h})$.

Proof: See Part B of the Appendix. □

This proposition relates the outage capacity region to the zero-outage (or delay-limited) capacity region, for which a rigorous coding theorem and converse exists. For point-to-point channels, the outage capacity can precisely be related to the ϵ-achievable rate [8, Proposition 2], [14]. Here, the notion of outage capacity is intended to be the analogous quantity for MACs.

B. Individual Outage Capacity Region

For a given power policy $\mathcal{P}^{\text{rand}}$, as noted earlier, the term $\mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h}))$ refers to the achievable rate region that varies based on the joint PDF $f_{\boldsymbol{P}}^{\text{rand}}(\boldsymbol{h})$ of the M random transmit powers in fading state $\boldsymbol{h}$. If we assume that one possible value of $\boldsymbol{P}^{\text{rand}}(\boldsymbol{h})$ is $\boldsymbol{P}$, then the corresponding achievable rate region is $\mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P})$ as defined in (3). In practice, for the given power vector $\boldsymbol{P}$ and fading state $\boldsymbol{h}$, only *one* rate vector in $\mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P})$ can be chosen for transmission at any specific time. In this case, we let $\boldsymbol{R}(\boldsymbol{h}, \boldsymbol{P})$ denote the vector of rate allocation functions for the M users under the given power vector $\boldsymbol{P}$. That is, given the power allocation vector $\boldsymbol{P}$, the rate vector chosen for transmission is $\boldsymbol{R}(\boldsymbol{h}, \boldsymbol{P})$, and $\boldsymbol{R}(\boldsymbol{h}, \boldsymbol{P}) \in \mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P})$. Similarly, for the random power allocation vector $\boldsymbol{P}^{\text{rand}}(\boldsymbol{h})$, we let $\boldsymbol{R}\left(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h})\right)$ denote the corresponding vector of rate allocation functions for the M users, and

$$\boldsymbol{R}\left(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h})\right) \in \mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h}))$$

with probability one. Since $\boldsymbol{P}^{\text{rand}}(\boldsymbol{h})$ varies within each fading state $\boldsymbol{h}$ as well as across all fading states, the rate vector $\boldsymbol{R}\left(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h})\right)$ chosen for transmission from the varying rate region $\mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h}))$ will vary accordingly. That is, $\boldsymbol{R}\left(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h})\right)$ also varies within each fading state $\boldsymbol{h}$ and across all fading states. Then, obviously, for a given rate vector $\boldsymbol{R}$ and a given rate allocation function vector $\boldsymbol{R}\left(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h})\right)$, in each fading state $\boldsymbol{h}$ ($\boldsymbol{h} \in \mathcal{H}_{\text{all}}$) corresponding to the power policy $\mathcal{P}^{\text{rand}}$, the average probability of transmission for User i with a rate no smaller than R_i is

$$Pr\left[R_i(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h})) \geq R_i\right], \qquad i = 1, \ldots, M.$$

Therefore, the individual outage capacity region can be defined as follows.

Definition 3.2: A rate vector $\boldsymbol{R}$ is in the individual outage capacity region $\mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, \boldsymbol{Pr})$ if and only if there exist a power policy $\mathcal{P}^{\text{rand}}$ and a corresponding vector of rate allocation functions $\boldsymbol{R}\left(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h})\right)$ for each fading state $\boldsymbol{h}$ ($\boldsymbol{h} \in \mathcal{H}_{\text{all}}$) such that

$$E[\boldsymbol{P}^{\text{rand}}(\boldsymbol{h})] \leq \overline{\boldsymbol{P}} \tag{12}$$

$$Pr\left[R_i(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h})) \geq R_i\right] \geq 1 - Pr_i, \qquad \forall\, i = 1, \ldots, M. \tag{13}$$

That is, we have (14) at the top of the following page. In words, the individual outage capacity region consists of all rate vectors that can be maintained (with arbitrarily small probability of error) with an outage probability vector (i.e., where transmission from different users need not simultaneously be turned on or off) no larger than $\boldsymbol{Pr}$ under the average power constraint $\overline{\boldsymbol{P}}$.

This definition allows for a fully randomized power policy. We earlier showed that it is sufficient to consider random power polices of cardinality two (in each fading state) to define

$$\mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, \boldsymbol{Pr}) \triangleq \bigcup_{\mathcal{P}^{\text{rand}}: E[\boldsymbol{P}^{\text{rand}}(\boldsymbol{h})] \leq \overline{\boldsymbol{P}}} \left\{ \boldsymbol{R} : Pr\left[R_i(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h})) \geq R_i \right] \geq 1 - Pr_i, \quad \forall\, i = 1, \ldots, M \right\}. \tag{14}$$

common outage capacity. For individual outage capacity, we show that it is sufficient to consider random power policies of cardinality 2^M in each fading state. The 2^M term is equal to the number of different subsets of users that can simultaneously transmit. We will represent each of these 2^M possible combinations of users as an M-dimensional vector

$$\boldsymbol{\Psi}(k) = [\psi(k,1), \ldots, \psi(k,M)]$$

which equals the binary expansion of k (the index of the subset S_k), $0 \leq k \leq 2^M - 1$. For each vector $[\psi(k,1), \ldots, \psi(k,M)]$, if $\psi(k,i) = 1$, then User i is said to be transmitting (i.e., $i \in S_k$); otherwise, User i is not (i.e., $i \notin S_k$).

For a given fading state $\boldsymbol{h} \in \mathcal{H}_{\text{all}}$ and a given subset index k $(k = 0, \ldots, 2^M - 1)$, let $\boldsymbol{P}(\boldsymbol{h}, k)$ be a vector of deterministic power allocation functions for the M users, with $P_i(\boldsymbol{h}, k) = 0$ if $\psi(k,i) = 0$, $i = 1, \ldots, M$, i.e., $\boldsymbol{\Psi}(k)\boldsymbol{P}(\boldsymbol{h},k) = \boldsymbol{P}(\boldsymbol{h},k)$, and let $w(\boldsymbol{h},k)$ be a deterministic probability of transmission function for the kth subset of users satisfying

$$\sum_{k=0}^{2^M-1} w(\boldsymbol{h},k) = 1$$

in each fading state. Obviously, $\boldsymbol{P}(\boldsymbol{h},0) = \boldsymbol{0}$. Now in each fading state $\boldsymbol{h} \in \mathcal{H}_{\text{all}}$, let $\boldsymbol{P}^{\text{rand}}(\boldsymbol{h}) = \boldsymbol{P}(\boldsymbol{h},k)$ with probability $w(\boldsymbol{h},k)$, $k = 0, \ldots, 2^M - 1$. Since $\sum_{k=0}^{2^M-1} w(\boldsymbol{h},k) = 1$, this simple power allocation policy has cardinality 2^M in each fading state and we denote it as $\mathcal{P}^I$ (we use the superscript "I" to distinguish it from the power policy $\mathcal{P}$ with cardinality two in the common outage case)

$$\mathcal{P}^I : \boldsymbol{h} \rightarrow \{\boldsymbol{P}(\boldsymbol{h},k), w(\boldsymbol{h},k)\}_{k=0}^{2^M-1}, \qquad \forall\, \boldsymbol{h} \in \mathcal{H}_{\text{all}}. \tag{15}$$

More specifically

$$\mathcal{P}^I : \boldsymbol{h} \rightarrow \begin{cases} \boldsymbol{0}, & \text{with prob. } w(\boldsymbol{h},0) \\ \boldsymbol{P}(\boldsymbol{h},1), & \text{with prob. } w(\boldsymbol{h},1) \\ \vdots & \\ \boldsymbol{P}(\boldsymbol{h},2^M-1), & \text{with prob. } w(\boldsymbol{h},2^M-1), \end{cases} \quad \forall\, \boldsymbol{h} \in \mathcal{H}_{\text{all}}. \tag{16}$$

Notice that in power policy $\mathcal{P}^I$, the power allocation function vector $\boldsymbol{P}(\boldsymbol{h},k)$ and the probability of transmission function $w(\boldsymbol{h},k)$ are only deterministic functions of $\boldsymbol{h}$ and k. The following proposition shows that it is sufficient to consider random power policies of cardinality 2^M in each fading state in the definition of the individual outage capacity region.

Proposition 3.3:

$$\mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, \boldsymbol{Pr}) = \bigcup_{\mathcal{P}^I \in \mathcal{F}^I} \left\{ \boldsymbol{R} : \boldsymbol{\psi}(k)\boldsymbol{R} \in \bigcap_{\{\boldsymbol{h}: w(\boldsymbol{h},k) > 0\}} \mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h},k)), \forall\, K = 1, \ldots, 2^M - 1 \right\} \tag{17}$$

where $\mathcal{F}^I$ is the set of all power policies $\mathcal{P}^I$ as defined in (16) that satisfy the conditions

$$E\left[\sum_{k=0}^{2^M-1} \boldsymbol{P}(\boldsymbol{h},k)w(\boldsymbol{h},k)\right] \leq \overline{\boldsymbol{P}} \tag{18}$$

$$E\left[\sum_{k=1}^{2^M-1} w(\boldsymbol{h},k)\psi(k,i)\right] \geq 1 - Pr_i, \qquad \forall\, i = 1, \ldots, M. \tag{19}$$

Proof: See Part C of the Appendix. □

Similar to the common outage case, we will now relate the individual outage capacity region to the zero-outage capacity region. For a given power policy $\mathcal{P}^I \in \mathcal{F}^I$ as defined in Proposition 3.3, let $\mathcal{H}_{\text{tran}}(k)$ denote the set of nonoutage fading states (transmission states) for the kth subset of users, i.e.,

$$\mathcal{H}_{\text{tran}}(k) = \{\boldsymbol{h} : w(\boldsymbol{h},k) > 0\}, \qquad k = 1, \ldots, 2^M - 1$$

and define a new PDF $g(\boldsymbol{h},k)$ for $\boldsymbol{h} \in \mathcal{H}_{\text{all}}$ as follows:

$$g(\boldsymbol{h},k) = \begin{cases} \frac{1}{E[w(\boldsymbol{h},k)]} p(\boldsymbol{h}) w(\boldsymbol{h},k), & \boldsymbol{h} \in \mathcal{H}_{\text{tran}}(k) \\ 0, & \boldsymbol{h} \notin \mathcal{H}_{\text{tran}}(k) \end{cases} \tag{20}$$

where $p(\boldsymbol{h})$ is the true PDF of the fading state $\boldsymbol{h} \in \mathcal{H}_{\text{all}}$. It is easily verified that

$$\int_{\boldsymbol{h} \in \mathcal{H}_{\text{all}}} g(\boldsymbol{h},k)\, d\boldsymbol{h} = \int_{\boldsymbol{h} \in \mathcal{H}_{\text{tran}}(k)} g(\boldsymbol{h},k)\, d\boldsymbol{h} = 1, \qquad \forall\, k = 1, \ldots, 2^M - 1.$$

Proposition 3.4: The individual outage capacity region defined in Definition 3.2 can be written in terms of the zero-outage capacity region as

$$\mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, \boldsymbol{Pr}) = \bigcup_{\mathcal{P}^I \in \mathcal{F}^I} \left\{ \boldsymbol{R} : \boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\text{zero}}(\boldsymbol{A}(k), \mathcal{H}_{\text{tran}}(k)), \forall\, k = 1, \ldots, 2^M - 1 \right\} \tag{21}$$

$$= \bigcup_{\mathcal{P}^I \in \mathcal{F}^I} \bigcap_{k=1}^{2^M-1} \overline{\mathcal{C}}_{\text{zero}}(\boldsymbol{A}(k), \mathcal{H}_{\text{tran}}(k)) \tag{22}$$

where

$$\boldsymbol{A}(k) = E[\boldsymbol{P}(\boldsymbol{h},k)w(\boldsymbol{h},k)]/E[w(\boldsymbol{h},k)]$$

and $\mathcal{C}_{\text{zero}}(\boldsymbol{A}(k), \mathcal{H}_{\text{tran}}(k))$ refers to the zero-outage capacity region (with power constraint $\boldsymbol{A}(k)$) of the equivalent fading MAC for which the set of all possible fading states is $\mathcal{H}_{\text{tran}}(k)$,[5] and the PDF of $\boldsymbol{h} \in \mathcal{H}_{\text{tran}}(k)$ is $g(\boldsymbol{h},k)$. In addition,

[5]Note that if $\mathcal{H}_{\text{tran}}(k) = \varnothing$, we define

$$\mathcal{C}_{\text{zero}}(\boldsymbol{A}(k), \mathcal{H}_{\text{tran}}(k)) \equiv \{\boldsymbol{R} : R_i \geq 0, \forall\, i = 1, \ldots, M\}.$$

$\overline{\mathcal{C}}_{\rm zero}(\boldsymbol{A}(k), \mathcal{H}_{\rm tran}(k))$ refers to the *augmented* rate region of $\mathcal{C}_{\rm zero}(\boldsymbol{A}(k), \mathcal{H}_{\rm tran}(k))$, i.e.,

$$\overline{\mathcal{C}}_{\rm zero}(\boldsymbol{A}(k), \mathcal{H}_{\rm tran}(k)) = \{\boldsymbol{R} : \boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\rm zero}(\boldsymbol{A}(k), \mathcal{H}_{\rm tran}(k))\}.$$

Proof: See Part D of the Appendix. □

C. Boundary Characterization

We now define the notion of the "boundary" of these capacity regions.

Definition 3.3: The boundary surface of $\mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, Pr)$ (or $\mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, \boldsymbol{Pr})$) is the set of those rate vectors for which we cannot increase one component and remain in $\mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, Pr)$ (or $\mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, \boldsymbol{Pr})$) without decreasing another component.

With these definitions, we wish to find: a) the optimal power allocation strategy that achieves the boundary of the common outage capacity region $\mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, Pr)$; b) the optimal power allocation strategy that achieves the boundary of the individual outage capacity region $\mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, \boldsymbol{Pr})$. The regions $\mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, Pr)$ and $\mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, \boldsymbol{Pr})$ are easily determined given these optimal power allocation strategies.

In the next section (Section IV), we will show that finding the optimal power allocation policy that achieves the boundary of $\mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, Pr)$ is equivalent to deriving the power allocation policy that *minimizes* the common outage probability for a given rate vector $\boldsymbol{R}$ and power constraint vector $\overline{\boldsymbol{P}}$. In the individual outage case, there is a similar equivalence between the power allocation policy that achieves the boundary of $\mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, \boldsymbol{Pr})$ and the one that achieves the boundary of the outage probability region, which will be discussed in detail in Section IV.

D. Operational Meaning of Outage Capacity

As stated in the Introduction, outage capacity is most relevant in a slowly fading environment where the channel can be assumed to be constant for the duration of a codeword. In this situation, if allowing optimal power control, the outage probability of a user is the probability that no codeword is transmitted by that user. Furthermore, in the slow-fading environment, the decoding delay only depends on the code length employed and not on the time variation of the channel.

If, on the other hand, the channel is fast fading and cannot be assumed to be constant for the duration of a codeword, then outage capacity largely loses its operational meaning, though the mathematics may still go through. In the fast-fading scenario, decoding delay will depend on the time variation of the channel because outage periods may begin in the middle of the transmission of a codeword. In this case, the power control policy would force the transmitter to wait until the outage period is complete before finishing transmission of the codeword. In such an environment, it is probably more appropriate to consider the ergodic capacity, i.e., using very long codewords that utilize the ergodicity of the channel, or the zero-outage capacity, which has operational meaning in either fast- or slow-fading environments.

IV. Outage Probability Region

In this section, we consider the outage probability region, or the set of achievable outage probability scalars (common outage) or vectors (individual outage) for a given rate vector and power constraint vector. The outage capacity region is the set of all *rates* that are achievable while meeting an outage constraint and an average power constraint. The outage probability region, on the other hand, is the set of all *outage probabilities* that are achievable for a specified rate vector and power constraint vector.

The common outage probability set $\mathcal{O}_C(\overline{\boldsymbol{P}}, \boldsymbol{R})$ and the complementary common transmission (usage) probability set $\overline{\mathcal{O}}_C(\overline{\boldsymbol{P}}, \boldsymbol{R})$ are naturally defined in terms of the outage capacity regionas follows.

Definition 4.1: The outage probability Pr is in the common outage probability set $\mathcal{O}_C(\overline{\boldsymbol{P}}, \boldsymbol{R})$ if and only if the rate vector $\boldsymbol{R} \in \mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, Pr)$.

Definition 4.2: The usage probability $Pr^{\rm on}$ is in the common usage probability set $\overline{\mathcal{O}}_C(\overline{\boldsymbol{P}}, \boldsymbol{R})$ if and only if the rate vector $\boldsymbol{R} \in \mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, 1 - Pr^{\rm on})$.

Definition 4.3: The minimum common outage probability $Pr_{\rm min}(\overline{\boldsymbol{P}}, \boldsymbol{R})$ is the smallest probability in the set $\mathcal{O}_C(\overline{\boldsymbol{P}}, \boldsymbol{R})$.

Proposition 4.1: The common usage probability set is equivalently given by (23) at the bottom of the page.

Proof: The usage probability $Pr^{\rm on}$ is in the common usage probability set if and only if the rate vector $\boldsymbol{R} \in \mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, 1 - Pr^{\rm on})$, which by Proposition 3.1 is true if and only if there exists a deterministic power allocation function $\boldsymbol{P}(\boldsymbol{h})$ and a probability of transmission function $w(\boldsymbol{h})$ such that $E[w(\boldsymbol{h})\boldsymbol{P}(\boldsymbol{h})] \le \overline{\boldsymbol{P}}$, $\boldsymbol{R} \in \mathcal{C}_{\rm MAC}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h}))$, $\forall\ \boldsymbol{h} : w(\boldsymbol{h}) > 0$, and $E[w(\boldsymbol{h})] \ge Pr^{\rm on}$. Therefore, we have (24), also at the bottom of the page. However, notice that if there exists a power allocation function $\boldsymbol{P}(\boldsymbol{h})$ and a probability of transmission function $w(\boldsymbol{h})$ with $E[w(\boldsymbol{h})] = Pr^{\rm on}$, the function $w(\boldsymbol{h})$ can be reduced such that $E[w(\boldsymbol{h})] = \alpha$ for any $0 \le \alpha < Pr^{\rm on}$. Thus, the interval $[0, E[w(\boldsymbol{h})]]$ is not needed in the left-hand side of (24), and we have the result. □

$$\overline{\mathcal{O}}_C(\overline{\boldsymbol{P}}, \boldsymbol{R}) = \bigcup_{w(\boldsymbol{h}), \boldsymbol{P}(\boldsymbol{h}): E_{\boldsymbol{h}}[w(\boldsymbol{h})\boldsymbol{P}(\boldsymbol{h})] \le \overline{\boldsymbol{P}},\ \boldsymbol{R} \in \mathcal{C}_{\rm MAC}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h}))\ \forall\ \boldsymbol{h}: w(\boldsymbol{h}) > 0} E_{\boldsymbol{h}}[w(\boldsymbol{h})]. \tag{23}$$

$$\overline{\mathcal{O}}_C(\overline{\boldsymbol{P}}, \boldsymbol{R}) = \bigcup_{w(\boldsymbol{h}), \boldsymbol{P}(\boldsymbol{h}): E_{\boldsymbol{h}}[w(\boldsymbol{h})\boldsymbol{P}(\boldsymbol{h})] \le \overline{\boldsymbol{P}},\ \boldsymbol{R} \in \mathcal{C}_{\rm MAC}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h}))\ \forall\ \boldsymbol{h}: w(\boldsymbol{h}) > 0} [0, E_{\boldsymbol{h}}[w(\boldsymbol{h})]]. \tag{24}$$

For common outage, it is clear that the outage probability set is simply an interval of $[0,1]$. Given $0 \leq Pr \leq 1$, the outage capacity region $\mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, Pr)$ is implicitly obtained once the minimum common outage probability $Pr_{\min}(\overline{\boldsymbol{P}}, \boldsymbol{R})$ for a given rate vector $\boldsymbol{R}$ is calculated under the optimal power allocation. That is, for any rate vector $\boldsymbol{R}$, $\boldsymbol{R} \in \mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, Pr)$ if and only if $Pr_{\min}(\overline{\boldsymbol{P}}, \boldsymbol{R}) \leq Pr$.

The individual outage probability set $\mathcal{O}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$ and the complementary individual transmission (usage) probability set $\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$ are naturally defined in terms of the outage capacity region as follows.

Definition 4.4: The outage probability vector $\boldsymbol{Pr}$ is in the individual outage probability set $\mathcal{O}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$ if and only if the rate vector $\boldsymbol{R} \in \mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, \boldsymbol{Pr})$.

Definition 4.5: The usage probability vector $\boldsymbol{Pr}^{\textbf{on}}$ is in the individual usage probability set $\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$ if and only if the rate vector $\boldsymbol{R} \in \mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, \boldsymbol{1} - \boldsymbol{Pr}^{\text{on}})$.

Proposition 4.2: The independent usage probability set is equivalently given by

$$\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}, \boldsymbol{R}) = \bigcup_{w(\boldsymbol{h},k), \boldsymbol{P}(\boldsymbol{h},k), \boldsymbol{R}(\boldsymbol{h})} [w_1, \ldots, w_M] \tag{25}$$

where

$$w_i \triangleq \sum_{k=1}^{2^M-1} E\left[w(\boldsymbol{h},k)\psi(k,i)\right], \qquad \text{for } i = 1, \ldots, M$$

and the union is subject to the conditions

$$E\left[\sum_{k=1}^{2^M-1} w(\boldsymbol{h},k)\boldsymbol{P}(\boldsymbol{h},k)\right] \leq \overline{\boldsymbol{P}}$$
$$\boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h},k))$$

for any $w(\boldsymbol{h},k)$ satisfying $w(\boldsymbol{h},k) > 0$ and

$$\sum_{k=0}^{2^M-1} w(\boldsymbol{h},k) = 1, \qquad \forall\, \boldsymbol{h} \in \mathcal{H}_{\text{all}}.$$

Proof: The usage probability vector $\boldsymbol{Pr}^{\textbf{on}}$ is in the individual usage probability set if and only if the rate vector $\boldsymbol{R} \in \mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, \boldsymbol{1} - \boldsymbol{Pr}^{\textbf{on}})$, which is true if and only there exists a vector of deterministic power allocation functions $\boldsymbol{P}(\boldsymbol{h},k)$ and a probability of transmission function $w(\boldsymbol{h},k)$ such that

$$E\left[\sum_{k=1}^{2^M-1} w(\boldsymbol{h},k)\boldsymbol{P}(\boldsymbol{h},k)\right] \leq \overline{\boldsymbol{P}}$$
$$\boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h},k)), \qquad w(\boldsymbol{h},k) > 0$$

and

$$\sum_{k=0}^{2^M-1} w(\boldsymbol{h},k) = 1, \qquad \forall\, \boldsymbol{h} \in \mathcal{H}_{\text{all}}.$$

Using the idea from the proof of Proposition 4.1, we get the result. □

The individual outage probability region $\mathcal{O}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$ is a set in $[0,1]^M$ that contains the point $[1, \ldots, 1]$ and the individual

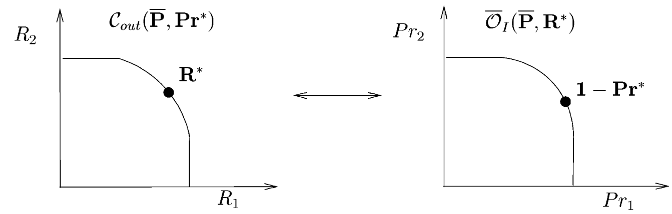

Fig. 1. Outage capacity region and usage probability region.

usage probability region $\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$ is a set in $[0,1]^M$ that contains the point $[0, \ldots, 0]$. We later show that the set $\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$ is in fact convex. With the above definitions,[6] it is easy to see the connection between the outage capacity and outage probability regions.

Given a probability vector $\boldsymbol{Pr}$, the outage capacity region $\mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, \boldsymbol{Pr})$ is implicitly obtained once the boundary of the outage (or usage) probability region $\mathcal{O}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$ (or $\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$) for a given rate vector $\boldsymbol{R}$ is derived through the optimal power allocation since, for any rate vector $\boldsymbol{R}$, $\boldsymbol{R} \in \mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, \boldsymbol{Pr})$ if and only if $\boldsymbol{1} - \boldsymbol{Pr} \in \overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$.[7]

An example of a two-user individual outage capacity region is plotted in Fig. 1. The corresponding usage probability region for a rate vector on the boundary of the outage capacity region is shown.

Having established that the outage capacity region can be found implicitly from the outage/usage probability region, we proceed by deriving the outage/usage probability region and the optimal power allocation policies. In Section V, we consider common outage and we derive the minimum common outage probability and the corresponding power allocation policy. In Section VI, we derive the usage probability region for individual outage, and also find the optimal power allocation policies.

V. Common Outage Capacity

In this section, we consider common outage, where outages are declared simultaneously for all users. We derive the minimum common outage probability $Pr_{\min}(\overline{\boldsymbol{P}}, \boldsymbol{R})$ in Section V-A and give the corresponding optimal power allocation strategy in Section V-B. This optimal strategy is given in terms of the optimal Lagrangian multipliers, and an algorithm to find these optimal multipliers is given in Section V-C. In Section V-D, we solve the dual problem of finding the average power region of the M users required to support $\boldsymbol{R}$ with a given common outage probability Pr, and finally we discuss the related notion of extreme points in Section V-E.

A. Minimum Common Outage Probability

For a given average power constraint vector $\overline{\boldsymbol{P}}$ and rate vector $\boldsymbol{R}$, from Definitions 4.1–4.3, it is obvious that deriving the min-

[6]Note that the definitions of $\mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, Pr)$, $\mathcal{C}_{\text{out}}(\overline{\boldsymbol{P}}, \boldsymbol{Pr})$, $Pr_{\min}(\overline{\boldsymbol{P}}, \boldsymbol{R})$, $\mathcal{O}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$, and $\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$ are similar to those for the broadcast channel in [7], where the power constraint is a total average power $\overline{P}$ instead of a vector $\overline{\boldsymbol{P}}$ for the M users.

[7]Since $\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$ is an M-dimensional region, it is not necessarily straightforward to determine if the region includes an arbitrary vector $\boldsymbol{1} - \boldsymbol{Pr}$. However, since the region $\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$ is convex (see Lemma 6.1), standard techniques can be used to answer this question efficiently.

imum common outage probability $Pr_{\min}(\overline{\boldsymbol{P}},\boldsymbol{R})$ is equivalent to deriving the maximum common usage probability in the set $\overline{\mathcal{O}}_C(\overline{\boldsymbol{P}},\boldsymbol{R})$.

That is, we need to solve the maximization problem

$$\max_{Pr^{\rm on}} Pr^{\rm on} \text{ subject to: } Pr^{\rm on}\in\overline{\mathcal{O}}_C(\overline{\boldsymbol{P}},\boldsymbol{R}). \tag{26}$$

For a given rate vector $\boldsymbol{R}$, define the set $\mathcal{Q}_C(\boldsymbol{R})$ as

$$\mathcal{Q}_C(\boldsymbol{R})=\{(Pr^{\rm on},\boldsymbol{P}) : Pr^{\rm on}\in\overline{\mathcal{O}}_C(\boldsymbol{P},\boldsymbol{R})\}. \tag{27}$$

Thus, we can rewrite the maximization in (26) as

$$\max_{(Pr^{\rm on},\boldsymbol{P})\in\mathcal{Q}_C(\boldsymbol{R})} Pr^{\rm on} \text{ subject to: } \boldsymbol{P}\le\overline{\boldsymbol{P}}. \tag{28}$$

We will require the following lemma to find the solution to (28).

Lemma 5.1: The set $\mathcal{Q}_C(\boldsymbol{R})$ is convex.

Proof: Convexity of the set is proven using a time-sharing argument. See Part E of the Appendix for details. □

Due to the convexity of the set $\mathcal{Q}_C(\boldsymbol{R})$, the pair $(Pr^{on*},\boldsymbol{P}^*)$ solves (28) if and only if there exists a Lagrangian multiplier vector $\boldsymbol{\lambda}\in\Re_+^M$ such that $(Pr^{on*},\boldsymbol{P}^*)$ is a solution to the maximization

$$\max_{(Pr^{\rm on},\boldsymbol{P})\in\mathcal{Q}_C(\boldsymbol{R})} [Pr^{\rm on}-\boldsymbol{\lambda}\cdot\boldsymbol{P}] \tag{29}$$

with $\boldsymbol{P}^*=\overline{\boldsymbol{P}}$. In convex optimization terms, this is akin to saying that the optimal solution must maximize the Lagrangian given the optimal Lagrange multipliers. Notice that there is no constraint on the power consumption in the maximization in (29).

By Proposition 4.1, a probability vector $Pr^{\rm on}$ is in $\overline{\mathcal{O}}_C(\boldsymbol{P},\boldsymbol{R})$ if and only if there exists a pair of functions $(w(\boldsymbol{h}),\boldsymbol{P}(\boldsymbol{h}))$ such that $Pr^{\rm on}=E_{\boldsymbol{h}}[w(\boldsymbol{h})]$, $\boldsymbol{P}=E_{\boldsymbol{h}}[w(\boldsymbol{h})P(\boldsymbol{h})]$, and $\boldsymbol{R}\in\mathcal{C}_{\rm MAC}(\boldsymbol{h},\boldsymbol{P}(\boldsymbol{h}))$ for all $\boldsymbol{h}$ such that $w(\boldsymbol{h})>0$. Therefore, we can equivalently perform the following maximization over the functions $w(\boldsymbol{h})$ and $\boldsymbol{P}(\boldsymbol{h})$:

$$\max_{w(\boldsymbol{h}),\boldsymbol{P}(\boldsymbol{h}):\boldsymbol{R}\in\mathcal{C}_{\rm MAC}(\boldsymbol{h},\boldsymbol{P}(\boldsymbol{h}))\ \forall\ \boldsymbol{h}:w(\boldsymbol{h})>0} E_{\boldsymbol{h}}[w(\boldsymbol{h})]-\boldsymbol{\lambda}\cdot E_{\boldsymbol{h}}[w(\boldsymbol{h})\boldsymbol{P}(\boldsymbol{h})]. \tag{30}$$

We will proceed to solve this maximization in two steps: we will first find a power allocation function $\boldsymbol{P}^*(\boldsymbol{h})$ that is optimal for any choice of $w(\boldsymbol{h})$, and then given such a power allocation function, we will maximize over the function $w(\boldsymbol{h})$.

In order to find the optimal power allocation function $\boldsymbol{P}^*(\boldsymbol{h})$, notice that if we fix $w(\boldsymbol{h})$, $\boldsymbol{P}^*(\boldsymbol{h})$ must satisfy the optimization in (30) over the variable $\boldsymbol{P}(\boldsymbol{h})$. That is, the optimal choice $\boldsymbol{P}^*(\boldsymbol{h})$ (for a given $w(\boldsymbol{h})$) must be the solution to

$$\min_{\boldsymbol{P}(\boldsymbol{h}):\boldsymbol{R}\in\mathcal{C}_{\rm MAC}(\boldsymbol{h},\boldsymbol{P}(\boldsymbol{h}))\ \forall\ \boldsymbol{h}:w(\boldsymbol{h})>0} \boldsymbol{\lambda}\cdot E_{\boldsymbol{h}}[w(\boldsymbol{h})P(\boldsymbol{h})] \tag{31}$$

which implies that $\boldsymbol{P}(\boldsymbol{h})$ must be optimal in *every* fading state $\boldsymbol{h}$ for which $w(\boldsymbol{h})>0$. Therefore, a power allocation function is optimal if and only if it is the solution to

$$\min_{\boldsymbol{P}(\boldsymbol{h})} \boldsymbol{\lambda}\cdot\boldsymbol{P}(\boldsymbol{h}) \text{ subject to: } \boldsymbol{R}\in\mathcal{C}_{\rm MAC}(\boldsymbol{h},\boldsymbol{P}(\boldsymbol{h})) \tag{32}$$

for every fading state $\boldsymbol{h}$ such that $w(\boldsymbol{h})>0$. This is identical to the problem posed when finding the zero-outage capacity in [6, eq. 4]. In [6], this problem is solved by exploiting the *polymatroid* structure of the MAC capacity region and by Lemma 3.3 of [2]. As the details of the solution are contained in [6], we will only state the results. The optimal power allocation function $\boldsymbol{P}^*(\boldsymbol{h})$ for all states $\boldsymbol{h}$ such that $w(\boldsymbol{h})>0$ is as shown in (33) at the bottom of the page, where the permutation $\pi(\cdot)$ satisfies

$$\frac{\lambda_{\pi(1)}}{h_{\pi(1)}}\ge\frac{\lambda_{\pi(2)}}{h_{\pi(2)}}\ge\cdots\ge\frac{\lambda_{\pi(M)}}{h_{\pi(M)}}. \tag{34}$$

The optimal solution is to allocate power to the M users in each fading state for which $w(\boldsymbol{h})>0$ so that the rate vector $\boldsymbol{R}$ can be achieved by performing successive decoding in the order specified by the permutation $\pi(\cdot)$. That is, the signal from User $\pi(M)$ is decoded first, treating all other users as noise. The codeword of User $\pi(M)$ is then subtracted off, and then User $\pi(M-1)$ is decoded, treating Users $\pi(M-2)$ through $\pi(1)$ as noise. The signal from User $\pi(1)$ is decoded last, with the signals from all other users being known and thus being subtracted from the total received signal. Note that the Lagrangian multiplier vector $\boldsymbol{\lambda}$ can be viewed as the *power price vector* of the M users, which we will use to refer to $\boldsymbol{\lambda}$ hereafter. From (34), we see that the decoding order in each fading state depends on both $\boldsymbol{\lambda}$ (users with larger power prices are decoded later) and the fading state $\boldsymbol{h}$ (users with smaller channel gains are decoded later). If no more than one component of $\boldsymbol{\lambda}$ is zero, the optimal decoding order in (34) is uniquely defined in each fading state. However, if two or more components of $\boldsymbol{\lambda}$ are equal to zero, then the decoding order is no longer uniquely defined for these users. We will discuss this case in the next subsection (Section V-B). Since the outage capacity can be stated in terms of the zero-outage capacity (Proposition 3.2), it should be intuitively clear why the optimal power allocation for the nonoutage fading states is identical to the optimal power allocation used to achieve zero-outage capacity.

Clearly, the power allocation function for fading states for which $w(\boldsymbol{h})=0$ is unimportant because these states do not affect the usage probability or the power constraint. Therefore, if we define $\boldsymbol{P}^*(\boldsymbol{h})$ by (33) for all fading states and not just fading states for which $w(\boldsymbol{h})>0$, then the power allocation function will be optimal for *any* choice of $w(\boldsymbol{h})$. Thus, if $\boldsymbol{P}^*(\boldsymbol{h})$ is defined by (33) for all fading states, $\boldsymbol{P}^*(\boldsymbol{h})$ is only a function of the Lagrangian multiplier vector $\boldsymbol{\lambda}$ (and not of $w(\boldsymbol{h})$) and is, therefore, optimal for any choice of $w(\boldsymbol{h})$.

$$P^*_{\pi(i)}(\boldsymbol{h})=\begin{cases}\frac{\sigma^2}{h_{\pi(1)}}\left[\exp\left(2R_{\pi(1)}\right)-1\right], & i=1\\ \frac{\sigma^2}{h_{\pi(i)}}\left[\exp\left(2\sum_{k=1}^{i}R_{\pi(k)}\right)-\exp\left(2\sum_{k=1}^{i-1}R_{\pi(k)}\right)\right], & 2\le i\le M\end{cases} \tag{33}$$

Having derived the optimal power allocation function $\boldsymbol{P}^*(\boldsymbol{h})$ for any $w(\boldsymbol{h})$, we can now perform the maximization of the usage probability (30) over only the function $w(\boldsymbol{h})$

$$\max_{w(\boldsymbol{h})} E_{\boldsymbol{h}}[w(\boldsymbol{h})] - \boldsymbol{\lambda} \cdot E_{\boldsymbol{h}}[w(\boldsymbol{h})\boldsymbol{P}^*(\boldsymbol{h})].$$

We can clearly simplify this as

$$\max_{w(\boldsymbol{h})} E_{\boldsymbol{h}}[w(\boldsymbol{h})(1 - \boldsymbol{\lambda} \cdot \boldsymbol{P}^*(\boldsymbol{h}))]. \tag{35}$$

Since $w(\boldsymbol{h})$ defines the probability of transmitting in each fading state, we must have $0 \leq w(\boldsymbol{h}) \leq 1$ for all fading states. In addition, since there are no other constraints on $w(\boldsymbol{h})$, it is clear that the optimal choice of $w(\boldsymbol{h})$ is

$$w^*(\boldsymbol{h}) = \begin{cases} 1, & \boldsymbol{\lambda} \cdot \boldsymbol{P}^*(\boldsymbol{h}) < 1 \\ 0, & \boldsymbol{\lambda} \cdot \boldsymbol{P}^*(\boldsymbol{h}) \geq 1. \end{cases} \tag{36}$$

Thus, $\boldsymbol{P}^*(\boldsymbol{h})$ and $w^*(\boldsymbol{h})$ (which are implicit functions of $\boldsymbol{\lambda}$) maximize (30).

B. Common Outage Transmission Policy

When the optimal Lagrangian multiplier vector $\boldsymbol{\lambda}^*$ is known (a simple algorithm to find $\boldsymbol{\lambda}^*$ is given in Section V-C), then the optimal transmission policy is known. For each fading state $\boldsymbol{h} \in \mathcal{H}_{\text{all}}$, the optimal transmission policy that minimizes the outage probability (given in terms of the vector of optimal Lagrangian multipliers $\boldsymbol{\lambda}^*$) is as follows.

1. In fading states that satisfy $\boldsymbol{\lambda}^* \cdot \boldsymbol{P}^*(\boldsymbol{h}) \geq 1$, an outage is declared and no users transmit, i.e., $w^*(\boldsymbol{h}) = 0$.
2. In fading states that satisfy $\boldsymbol{\lambda}^* \cdot \boldsymbol{P}^*(\boldsymbol{h}) < 1$, all M users transmit at their specified rates with probability one in that fading state, i.e., $w^*(\boldsymbol{h}) = 1$. Furthermore, successive decoding can be used with the decoding order described in (34).

There are a number of key properties to notice about the optimal transmission policy. First note that even though we allowed randomized power policies in each state, the optimal power policy is in fact *deterministic*, i.e., $w(\boldsymbol{h})$ is equal to either one or zero in each fading state.[8] This property is not so surprising if one notices that the optimization in (35) is a simple linear program for which we would expect the solution to lie on an extreme point of the space for the probability of transmission function $w(\boldsymbol{h})$: $0 \leq w(\boldsymbol{h}) \leq 1, \forall\, \boldsymbol{h} \in \mathcal{H}_{\text{all}}$.

Furthermore, the transmission policy can be viewed as a simple *threshold* policy, since simultaneous transmission by the M users is allowed if and only if the required minimum total weighted power $\boldsymbol{\lambda}^* \cdot \boldsymbol{P}^*(\boldsymbol{h})$ (where the weights are equal to the Lagrangian multipliers, or the power prices) for the M users to transmit their information at rate vector $\boldsymbol{R}$ in state $\boldsymbol{h}$ is less than 1. This is similar to the optimum transmission policies that minimize the outage probability for the single-user channel [8] and the common outage probability for the broadcast channel [7].

[8]It should be noted that the purely deterministic nature of the optimum power policy is only guaranteed for continuous fading distributions. For discrete fading distributions, a random power policy may be needed for states that satisfy $\boldsymbol{\lambda}^* \cdot \boldsymbol{P}^*(\boldsymbol{h}) = 1$. For continuous distributions, the set of such states has measure zero and thus need not be considered for our purposes.

Under this transmission policy, the resulting common outage probability is

$$\begin{aligned} Pr_{\min}(\boldsymbol{R}) &= 1 - E_{\boldsymbol{h}}[w^*(\boldsymbol{h})] \\ &= 1 - Pr[\boldsymbol{\lambda}^* \cdot \boldsymbol{P}^*(\boldsymbol{h}) < 1]. \end{aligned} \tag{37}$$

The average power used by each user is given by

$$\overline{P}_i = E_{\boldsymbol{h}}\left[w^*(\boldsymbol{h}) P_i^*(\boldsymbol{h})\right]. \tag{38}$$

The optimal Lagrangian multiplier vector $\boldsymbol{\lambda}^*$ guarantees that the power constraint of each user is satisfied. In fact, complementary slackness [15] guarantees that the power constraint is met with equality for every user i satisfying $\lambda_i^* > 0$.

However, if $\lambda_i^* = 0$ for two or more users, as noted earlier, the decoding order and subsequent power allocation policy in (34) and (33) is not uniquely defined. In this scenario, there can be *multiple* solutions to (30) given the optimal $\boldsymbol{\lambda}^*$, since there is no cost associated with allocating power to users with $\lambda_i = 0$. Therefore, additional power can be allocated to users with $\lambda_i = 0$ without affecting (30), which means that there are many different power allocation policies that achieve the maximum. However, by convex theory [15], we are guaranteed that at least one of them is a solution that satisfies the power constraints of all users (and not just those users with $\lambda_i^* > 0$), though it is not easy to find which solution that is.

When $\lambda_k^* = 0$, this indicates that User k is not a limiting factor in achieving the minimum common outage probability. In other words, the power constraint of User k is large enough such that User k can achieve rate R_k in all the nonoutage states even if he is decoded first (i.e., sees all other received power as interference). If multiple users have $\lambda_k^* = 0$, then a whole class of users is such that even if they are decoded before all other users (with some unknown decoding order within the class) in all fading states, they can still achieve their respective rates without exceeding their power constraints. The challenge then is to determine a decoding order for this set of users such that the corresponding power policy satisfies the power constraints. A simple way to find a decoding order that works in the case where two or more users have a zero Lagrangian multiplier is to lower the power constraint of one or more of these users until the Lagrangian multipliers are either strictly positive or zero for only one user.

C. Optimal Lagrangian Multipliers

In the previous subsections, we characterized the minimum common outage probability and the optimal transmission policy assuming knowledge of the optimum Lagrangian multiplier (power price) vector $\boldsymbol{\lambda}^*$. Therefore, given the power constraint vector $\overline{\boldsymbol{P}}$ and a target rate vector $\boldsymbol{R}$, an important question is how to obtain the optimal power price vector $\boldsymbol{\lambda}^*$ that corresponds to the minimum common outage probability. In this subsection, we will describe a standard convex optimization algorithm that provably converges to the optimum Lagrangian multipliers. This algorithm can also be used to find the optimal Lagrangian multipliers for individual outage (Section VI-C). We will use a convex optimization algorithm on the Lagrangian *dual* function. For a primer on dual functions and convex optimization, see [15].

The original problem of maximizing the common usage probability is (28)

$$1 - Pr_{\min}(\boldsymbol{R}, \overline{\boldsymbol{P}}) = \max_{(Pr^{\text{on}}, \boldsymbol{P}) \in \mathcal{Q}_C(\boldsymbol{R})} Pr^{\text{on}} \text{ subject to: } \boldsymbol{P} \leq \overline{\boldsymbol{P}}. \tag{39}$$

The Lagrangian of this maximization is

$$L(Pr^{\text{on}}, \boldsymbol{P}, \boldsymbol{\lambda}) = Pr^{\text{on}} - \boldsymbol{\lambda} \cdot (\boldsymbol{P} - \overline{\boldsymbol{P}}). \tag{40}$$

The dual function $g(\boldsymbol{\lambda})$ is found by taking the supremum over all $(Pr^{\text{on}}, \boldsymbol{P}) \in \mathcal{Q}_C(\boldsymbol{R})$ for each Lagrangian multiplier vector $\boldsymbol{\lambda}$:

$$g(\boldsymbol{\lambda}) = \sup_{(Pr^{\text{on}}, \boldsymbol{P}) \in \mathcal{Q}_C(\boldsymbol{R})} L(Pr^{\text{on}}, \boldsymbol{P}, \boldsymbol{\lambda}). \tag{41}$$

The dual function is a supremum of affine functions of $\boldsymbol{\lambda}$, and is therefore a convex function [15]. Due to the definition of $\overline{\mathcal{O}}_C(\boldsymbol{P}, \boldsymbol{R})$, we can equivalently write the dual function as

$$g(\boldsymbol{\lambda}) = \sup_{w(\boldsymbol{h}), \boldsymbol{P}(\boldsymbol{h})} L'(\boldsymbol{P}(\boldsymbol{h}), w(\boldsymbol{h}), \boldsymbol{\lambda}) \tag{42}$$

subject to the constraint

$$\boldsymbol{R} \in \mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h})) \ \forall\, \boldsymbol{h} : w(\boldsymbol{h}) > 0 \tag{43}$$

where $L'(\boldsymbol{P}(\boldsymbol{h}), w(\boldsymbol{h}), \boldsymbol{\lambda})$ is defined as

$$L'(\boldsymbol{P}(\boldsymbol{h}), w(\boldsymbol{h}), \boldsymbol{\lambda}) \triangleq E_{\boldsymbol{h}}[w(\boldsymbol{h})] - \boldsymbol{\lambda} \cdot (E_{\boldsymbol{h}}[w(\boldsymbol{h})\boldsymbol{P}(\boldsymbol{h})] - \overline{\boldsymbol{P}}). \tag{44}$$

For any fixed $\boldsymbol{\lambda}$, the optimum $(w(\boldsymbol{h}), P(\boldsymbol{h}))$ that achieves the supremum in (42) corresponds to the solution described in Section V-B (i.e., successive decoding with decoding order determined by the fading states and $\boldsymbol{\lambda}$ and a threshold policy based on the weighted-sum power required in each fading state).

For convex maximizations, the *minimum* of the dual function $g(\boldsymbol{\lambda})$ over all nonnegative Lagrangian multipliers is equal to the maximum of the original objective function. That is,

$$1 - Pr_{\min}(\boldsymbol{R}, \overline{\boldsymbol{P}}) = \min_{\boldsymbol{\lambda} \geq 0} g(\boldsymbol{\lambda}) = g(\boldsymbol{\lambda}^*) \tag{45}$$

where $\boldsymbol{\lambda}^*$ is the optimum Lagrangian multiplier vector. Furthermore, the optimal $(\boldsymbol{P}^*(\boldsymbol{h}), w^*(\boldsymbol{h}))$ achieves the supremum in (42) for $\boldsymbol{\lambda}^*$. Our goal is to find the Lagrangian multiplier vector (namely, $\boldsymbol{\lambda}^*$) that minimizes the dual function $g(\boldsymbol{\lambda})$. Since $g(\boldsymbol{\lambda})$ is a convex function, we can use standard convex optimization techniques to find the minimum value of the dual function and the optimal Lagrangian multiplier vector. Since it is not clear if the function $g(\boldsymbol{\lambda})$ is differentiable (though it is continuous), we use the ellipsoid algorithm, which can be slightly modified to work for nondifferentiable convex functions.

Now we briefly describe the ellipsoid algorithm as applied to our problem, and defer the details to Part F of the Appendix. The ellipsoid algorithm belongs to the family of cutting-plane methods [15, Ch. 12], the simplest of which is the one-dimensional bisection method. Of course, in our case the problem is M-dimensional, corresponding to the Lagrangian multiplier of each of the M users. In each iteration, a cutting-plane method eliminates a half-space from the feasible set (i.e., the set where the optimal solution can lie) by evaluating a gradient or a subgradient of the function to be minimized. This allows the feasible set to continually decrease in size, until it is small enough to satisfy convergence criteria. In the ellipsoid method, a minimum volume ellipsoid is formed around the feasible set and the function is then evaluated at the middle of this ellipsoid in order to generate a new cutting plane. This process is repeated indefinitely until the desired accuracy is reached.

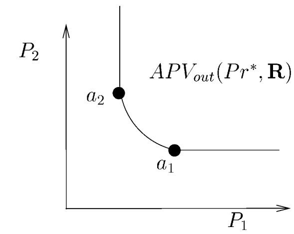

Fig. 2. Power region.

This method is applied to the problem at hand by first finding an ellipsoid in which $\boldsymbol{\lambda}^*$ must lie. The function $g(\boldsymbol{\lambda})$ is evaluated at some initial $\boldsymbol{\lambda_0}$ in this ellipsoid. Given that the functions $w^*(\boldsymbol{h}, \boldsymbol{\lambda_0})$ and $\boldsymbol{P}^*(\boldsymbol{h}, \boldsymbol{\lambda_0})$ maximize $L'(\boldsymbol{P}(\boldsymbol{h}), w(\boldsymbol{h}), \boldsymbol{\lambda_0})$, it can be shown that $g(\boldsymbol{\lambda}) \geq g(\boldsymbol{\lambda_0})$ for all $\boldsymbol{\lambda}$ satisfying

$$(E_{\boldsymbol{h}}[w^*(\boldsymbol{h}, \boldsymbol{\lambda_0})\boldsymbol{P}^*(\boldsymbol{h}, \boldsymbol{\lambda_0})] - \overline{\boldsymbol{P}})(\boldsymbol{\lambda} - \boldsymbol{\lambda_0})^T \leq 0.$$

This fact allows us to eliminate a halfspace of the space in the domain. A minimum volume ellipsoid covering the new feasible set (i.e., the original ellipsoid minus the eliminated halfspace) is then formed, and the process is repeated at the center of the new ellipsoid. It can be shown that the volume of the feasible ellipsoid converges to zero, and that the algorithm actually converges to the optimal $g(\boldsymbol{\lambda}^*)$. More details on this convergence can be found in Part F of the Appendix.

D. Average Power Region

In Sections V-A and V-B, given the power constraint vector $\overline{\boldsymbol{P}}$ and rate vector $\boldsymbol{R}$ of the M users, we derived the minimum common outage probability $Pr_{\min}(\overline{\boldsymbol{P}}, \boldsymbol{R})$ and the corresponding optimal power allocation policy. In this subsection, we find for a given rate vector $\boldsymbol{R}$ and common outage probability Pr^*, the required average power region $APV_{\text{out}}(Pr^*, \boldsymbol{R})$, defined as the set of all possible power constraint vectors that can support rate vector $\boldsymbol{R}$ with a common outage probability no larger than Pr^*. That is,

$$APV_{\text{out}}(Pr^*, \boldsymbol{R}) \triangleq \{\boldsymbol{P} : Pr^* \in \mathcal{O}_C(\boldsymbol{P}, \boldsymbol{R})\}. \tag{46}$$

The set $APV_{\text{out}}(Pr^*, \boldsymbol{R})$ is convex due to the convexity of the set $\mathcal{Q}_C(\boldsymbol{R})$, which is defined in (27). An example of a power region is shown in Fig. 2. Notice that the power region lies above (i.e., up and to the right of) the boundary. The points a_1 and a_2 are referred to as extreme points [6], while all points between these two extremes are referred to as regular points. In this subsection, we discuss only the characterization of regular points. Extreme points are discussed in Section V-E.

Due to the convexity of the average power region, the boundary of $APV_{\text{out}}(Pr^*, \boldsymbol{R})$ can be traced out by solving

$$\min_{\boldsymbol{P}} \boldsymbol{\lambda} \cdot \boldsymbol{P} \text{ subject to: } Pr^* \in \mathcal{O}_C(\boldsymbol{P}, \boldsymbol{R}) \tag{47}$$

for all power price vectors $\boldsymbol{\lambda} \in \Re_+^M$ such that $\sum_{i=1}^M \lambda_i = 1$. For a given power price vector $\boldsymbol{\lambda} \in \Re_+^M$, this minimization is equivalent to

$$\min_{(Pr^{\text{on}}, \boldsymbol{P}) \in \mathcal{Q}_C(\boldsymbol{R})} \boldsymbol{\lambda} \cdot \boldsymbol{P} \text{ subject to: } Pr^{\text{on}} \geq 1 - Pr^*. \tag{48}$$

An average power vector $\boldsymbol{P}$ solves (48) if and only if there exists a Lagrangian multiplier s^* such that $(1 - Pr^*, \boldsymbol{P})$ is a solution to the problem

$$\min_{(Pr^{\text{on}}, \boldsymbol{P}) \in \mathcal{Q}_C(\boldsymbol{R})} \{\boldsymbol{\lambda} \cdot \boldsymbol{P} - s^* \cdot Pr^{\text{on}}\}. \tag{49}$$

By the definition of the set $\mathcal{Q}_C(\boldsymbol{R})$, the following is an equivalent minimization:

$$\min_{w(\boldsymbol{h}), \boldsymbol{P}(\boldsymbol{h}): \boldsymbol{R} \in \mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h}))\ \forall\ \boldsymbol{h}: w(\boldsymbol{h}) > 0} \boldsymbol{\lambda} \cdot E_{\boldsymbol{h}}[w(\boldsymbol{h}) P(\boldsymbol{h})] - s^* E_{\boldsymbol{h}}[w(\boldsymbol{h})]. \tag{50}$$

Notice that this optimization problem is very similar to the one in (30) in Section V-A. By the same arguments used to solve (30), the optimum choice of $\boldsymbol{P}(\boldsymbol{h})$ is $\boldsymbol{P}^*(\boldsymbol{h})$ as described by (33) and (34). Clearly, the optimum choice of $w(\boldsymbol{h})$ is

$$w^*(\boldsymbol{h}) = \begin{cases} 1, & \boldsymbol{\lambda} \cdot \boldsymbol{P}^*(\boldsymbol{h}) < s^* \\ 0, & \boldsymbol{\lambda} \cdot \boldsymbol{P}^*(\boldsymbol{h}) > s^*. \end{cases} \tag{51}$$

The Lagrangian multiplier s^* should be chosen such that $E_{\boldsymbol{h}}[w(\boldsymbol{h})] = 1 - Pr^*$. Since s^* is a scalar, this can easily be done by the bisection method. The optimum transmission policy is identical to the policy derived to minimize the common outage probability, with the only exception being that the threshold level s^* is not necessarily 1 as it was in Section V-A. Again, notice the deterministic nature of the optimum transmission policy.

Given the derived $\boldsymbol{P}^*(\boldsymbol{h})$ and $w^*(\boldsymbol{h})$ for every state $\boldsymbol{h} \in \mathcal{H}_{\text{all}}$, the complete power allocation policy is known and the corresponding average power vector is $\overline{\boldsymbol{P}} = E_{\boldsymbol{h}}[w^*(\boldsymbol{h}) \boldsymbol{P}^*(\boldsymbol{h})]$, where both $w^*(\boldsymbol{h})$ and $\boldsymbol{P}^*(\boldsymbol{h})$ implicitly depend on $\boldsymbol{\lambda}$. By varying the power price vector $\boldsymbol{\lambda} \in \Re_+^M$, we can obtain different average power vectors $\overline{\boldsymbol{P}}$ that lie on the boundary surface of $APV_{\text{out}}(Pr^*, \boldsymbol{R})$.

E. Extreme Points

As discussed in [6, Sec. III] for the zero-outage capacity case, there are other average power vectors on the boundary surface of $APV_{\text{out}}(Pr^*, \boldsymbol{R})$ that cannot be parameterized by any $\boldsymbol{\lambda} \in \Re_+^M$. We refer to these points as extreme points. In Fig. 2, the points a_1 and a_2 are extreme points. At the point a_1, the power used by User 1 is the minimum power that User 1 requires to maintain the given R_1 with outage probability equal to Pr^* in the absence of User 2. In other words, a_1 corresponds to the single-user (i.e., in the absence of User 2) power region boundary of User 1. In order for User 1 to achieve his single-user bound, clearly User 1 must be decoded last in every fading state so that he experiences no interference from User 2. Thus, the point a_1 corresponds to giving User 1 absolute priority in the sense that User 1 is decoded last in every fading state. Similarly, a_2 corresponds to decoding User 2 last in every fading state.

For the two-user case, the extreme point a_1 can actually be characterized using the method described in Section V-D with $\lambda_1 > 0$ and $\lambda_2 = 0$, which ensures that User 1 is decoded last in every fading state according to the optimal decoding order described in (34). However, when there are more than two users, the required decoding order can no longer be characterized by a power price vector. Therefore, it is necessary to use a more general method [6] in which we give absolute decoding order priority to subsets of users. Users are partitioned into N subsets and, in all fading states, users in the Nth subset are decoded first, followed by users in the $(N-1)$th subset, and so on. Within each subset, the decoding order is determined by the power price vector and the fading state according to (34). Technical details of the characterization of extreme points are given in Part G of the Appendix.

VI. Individual Outage Capacity

In this section, we consider individual outage, where outages can be declared separately from each user. We characterize the boundary of the usage probability region in Section VI-A, and give the corresponding optimal power allocation strategy in Section VI-B. In Section VI-C, we describe an algorithm that finds the optimal Lagrangian multipliers. We discuss extreme points of the usage probability region in Section VI-D. Finally, we characterize the average power region in Section VI-E.

A. Outage Probability Region

In this subsection, we explicitly characterize the boundary of the individual outage probability region. A word on notation: For any given vector $\mathbf{X} = (X_1, \ldots, X_M)$ and any set $S \in \{S_k\}_{k=1}^{2^M-1}$, let $|S|$ denote the total number of users in the set S, and let $\mathbf{X}^{(S)}$ denote the subvector of $\mathbf{X}$ consisting of components corresponding to the $|S|$ users in the set S.

From Definitions 4.4 and 4.5, it is clear that, for a given average power constraint vector $\overline{\boldsymbol{P}}^*$ and rate vector $\boldsymbol{R}$, deriving the boundary of the outage probability region $\mathcal{O}_I(\overline{\boldsymbol{P}}^*, \boldsymbol{R})$ is equivalent to deriving the boundary of the usage probability region $\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}^*, \boldsymbol{R})$. Define

$$\mathcal{Q}_I(\boldsymbol{R}) = \{(\boldsymbol{Pr^{\text{on}}}, \overline{\boldsymbol{P}}) : \boldsymbol{Pr^{\text{on}}} \in \overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})\}. \tag{52}$$

We will require the following lemma to derive the boundary of $\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}^*, \boldsymbol{R})$ and the corresponding optimal power allocation policy that achieves this boundary.

Lemma 6.1: Both the usage probability region $\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$ and the set $\mathcal{Q}_I(\boldsymbol{R})$ are convex.

Proof: Convexity is proven using a time-sharing argument. See Part H of the Appendix for details. □

Due to the convexity of $\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}^*, \boldsymbol{R})$, an average usage probability vector will be on the boundary surface of $\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}^*, \boldsymbol{R})$ if and only if it is a solution to

$$\max_{\boldsymbol{Pr^{on}} \in \overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}^*, \boldsymbol{R})} \boldsymbol{\mu} \cdot \boldsymbol{Pr^{on}} \tag{53}$$

for some nonnegative vector $\boldsymbol{\mu} \in \Re^M$. Here μ_i can be viewed as the channel usage reward if the information from User i is transmitted,[9] $\forall\, i = 1, \ldots, M$. In this subsection, we focus on strictly positive vectors $\boldsymbol{\mu}$ for a parameterization of the *regular points* on the boundary surface of $\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}^*, \boldsymbol{R})$. The *extreme points* correspond to the case where some components of the vector $\boldsymbol{\mu}$ are equal to zero.[10] Although one can get arbitrarily close to an extreme point by letting some components of $\boldsymbol{\mu}$ go to zero, we show in Section VI-D how to explicitly obtain the extreme points based on the regular point method described here.

Since the set $\mathcal{Q}_I(\boldsymbol{R})$ is convex, for a given channel usage reward vector $\boldsymbol{\mu} \in \Re_+^M$, vector

$$\boldsymbol{Pr^{on}} = [Pr_1^{\rm on}, \ldots, .Pr_M^{\rm on}]$$

solves (53) if and only if there exists a Lagrangian multiplier vector $\boldsymbol{\lambda} \in \Re_+^M$ such that $(\boldsymbol{Pr^{on}}, \boldsymbol{P}^*)$ is a solution to the problem

$$\max_{(\boldsymbol{Pr^{on}}, \overline{\boldsymbol{P}}) \in \mathcal{Q}_I(\boldsymbol{R})} \left[\boldsymbol{\mu} \cdot \boldsymbol{Pr^{on}} - \boldsymbol{\lambda} \cdot \overline{\boldsymbol{P}}\right] \tag{54}$$

with $\boldsymbol{P}^* = \overline{\boldsymbol{P}}^*$.

Equivalently, this maximization can be written as (55) at the bottom of the page.

Similar to the procedure used to maximize the common usage probability in (30) of Section V-A, we will proceed to solve this maximization problem in two steps: we first find a power allocation function $\boldsymbol{P}^*(\boldsymbol{h}, k)$ that is optimal for any choice of $w(\boldsymbol{h}, k)$, and then given such a power allocation function, we maximize with respect to $w(\boldsymbol{h}, k)$.

In order to find the optimal power allocation function $\boldsymbol{P}^*(\boldsymbol{h}, k)$, notice that if we fix $w(\boldsymbol{h}, k)$, $\boldsymbol{P}^*(\boldsymbol{h}, k)$ must satisfy the optimization in (55) over the variable $\boldsymbol{P}(\boldsymbol{h}, k)$. That is, the optimal choice $\boldsymbol{P}^*(\boldsymbol{h}, k)$ (for a given $w(\boldsymbol{h}, k)$) must be the solution to

$$\min_{\boldsymbol{P}(\boldsymbol{h},k): \boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\rm MAC}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h},k))\ \forall\ \boldsymbol{h},k: w(\boldsymbol{h},k)>0} E_{\boldsymbol{h}}\left[\sum_{k=1}^{2^M-1} w(\boldsymbol{h}, k) \boldsymbol{\lambda} \cdot \boldsymbol{P}(\boldsymbol{h}, k)\right]. \tag{56}$$

Therefore, a power allocation function is optimal if and only if it is the solution to

$$\min_{\boldsymbol{P}(\boldsymbol{h},k)} \boldsymbol{\lambda} \cdot \boldsymbol{P}(\boldsymbol{h}, k) \text{ subject to: } \boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\rm MAC}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h}, k)) \tag{57}$$

for every $w(\boldsymbol{h}, k)$ such that $w(\boldsymbol{h}, k) > 0$. This is clearly identical to the minimization of weighted-sum power in (32). Thus, $\forall\, k = 1, \ldots, 2^M - 1$, the optimal power allocation function $\boldsymbol{P}^*(\boldsymbol{h}, k)$ for all states $\boldsymbol{h}$ such that $w(\boldsymbol{h}, k) > 0$ is shown in (58) at the bottom of the page, where the permutation $\pi_s(\cdot)$ for the $|S_k|$ users in the subset S_k satisfies

$$\frac{\lambda_{\pi_s(1)}}{h_{\pi_s(1)}} \geq \frac{\lambda_{\pi_s(2)}}{h_{\pi_s(2)}} \geq \cdots \geq \frac{\lambda_{\pi_s(|S_k|)}}{h_{\pi_s(|S_k|)}}. \tag{59}$$

Similar to the common outage case, the solution in (59) indicates that $\boldsymbol{\lambda}^{(S_k)}$ determines the decoding order of the users in the subset S_k in each fading state, and we refer to $\boldsymbol{\lambda}$ as the power price vector.

Clearly, the vector of power allocation functions for fading states for which $w(\boldsymbol{h}, k) = 0$ is unimportant, since these states do not affect the usage probability or the power constraint. Therefore, $\forall\ k = 1, \ldots, 2^M - 1$, if we define $\boldsymbol{P}^*(\boldsymbol{h}, k)$ by (58) for *every* fading state $\boldsymbol{h} \in \mathcal{H}_{\rm all}$, then the power allocation function will be optimal for any choice of $w(\boldsymbol{h}, k)$.

Having derived the optimal power allocation function $\boldsymbol{P}^*(\boldsymbol{h}, k)$ for any $w(\boldsymbol{h}, k)$, we can now perform the maximization of the usage probability (55) over only the function $w(\boldsymbol{h}, k)$. If we define the total reward for all users in the set S_k to transmit information as

$$\eta_k \triangleq \sum_{i=1}^{M} \mu_i \psi(k, i), \qquad 1 \leq k \leq 2^M - 1 \tag{60}$$

then we can rewrite the optimization in (55) as

$$\max_{w(\boldsymbol{h},k)} E_{\boldsymbol{h}}\left[\sum_{k=1}^{2^M-1} w(\boldsymbol{h}, k)\eta_k\right] - \sum_{i=1}^{M} \lambda_i E_{\boldsymbol{h}}\left[\sum_{k=1}^{2^M-1} w(\boldsymbol{h}, k) P_i^*(\boldsymbol{h}, k)\right]$$

[9]Note that μ_i can also be viewed as the channel outage penalty if an outage is declared from User i.

[10]In the average power region, extreme points correspond to the case where some components of $\boldsymbol{\lambda}$ are equal to zero. For both the usage probability region and the average power region, more general decoding orders are used to achieve the extreme points.

$$\max_{w(\boldsymbol{h},k), \boldsymbol{P}(\boldsymbol{h},k): \boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\rm MAC}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h},k))\ \forall\ \boldsymbol{h},k: w(\boldsymbol{h},k)>0} \sum_{i=1}^{M} \sum_{k=1}^{2^M-1} \mu_i E_{\boldsymbol{h}}[w(\boldsymbol{h}, k)\psi(k, i)] - \sum_{i=1}^{M} \lambda_i E_{\boldsymbol{h}}\left[\sum_{k=1}^{2^M-1} w(\boldsymbol{h}, k) P_i(\boldsymbol{h}, k)\right]. \tag{55}$$

$$\begin{aligned} P_j^*(\boldsymbol{h}, k) &= 0, \quad \forall\, j \notin S_k \\ P_{\pi_s(i)}^*(\boldsymbol{h}, k) &= \begin{cases} \frac{\sigma^2}{h_{\pi_s(1)}}\left[\exp(2R_{\pi_s(1)}) - 1\right], & \text{if } i = 1 \\ \frac{\sigma^2}{h_{\pi_s(i)}}\left[\exp(2\sum_{l=1}^{i} R_{\pi_s(l)}) - \exp(2\sum_{l=1}^{i-1} R_{\pi_s(l)})\right], & \forall\, 2 \leq i \leq |S_k| \end{cases} \end{aligned} \tag{58}$$

which can be simplified to

$$\max_{w(\boldsymbol{h},k)} E_{\boldsymbol{h}} \left[\sum_{k=1}^{2^M-1} w(\boldsymbol{h},k) \left(\eta_k - \sum_{i=1}^{M} \lambda_i P_i^*(\boldsymbol{h},k) \right) \right]. \tag{61}$$

Since $w(\boldsymbol{h},k)$ defines the probability of transmission for the subset S_k in each fading state, we must have $\sum_{k=0}^{2^M-1} w(\boldsymbol{h},k) = 1$ in each fading state and $0 \le w(\boldsymbol{h},k) \le 1, \forall\, \boldsymbol{h}, k$. Since both constraints apply to each fading state separately, $w^*(\boldsymbol{h},k)$ is optimal if and only if it is the solution to

$$\max_{w(\boldsymbol{h},k):0\le w(\boldsymbol{h},k)\le 1\ \forall\ k,\ \sum_{k=0}^{2^M-1} w(\boldsymbol{h},k)=1} \sum_{k=1}^{2^M-1} w(\boldsymbol{h},k) \cdot (\eta_k - \boldsymbol{\lambda}\cdot \boldsymbol{P}^*(\boldsymbol{h},k)) \tag{62}$$

in each fading state. It is then straightforward to show that the optimal $w^*(\boldsymbol{h},k)$ is defined in each fading state for $k = 1,\dots,2^M-1$ by

$$w^*(\boldsymbol{h},k) = \begin{cases} 1, & \text{if } k = k^* \text{ and } \eta_k - \boldsymbol{\lambda}\cdot \boldsymbol{P}^*(\boldsymbol{h},k) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{63}$$

where

$$k^* = \operatorname*{argmax}_{j=1,\dots,2^M-1} (\eta_j - \boldsymbol{\lambda}\cdot \boldsymbol{P}^*(\boldsymbol{h},j)).$$

That is, the subset with the largest reward $(\eta_k - \boldsymbol{\lambda}\cdot \boldsymbol{P}^*(\boldsymbol{h},k))$ in each fading state is chosen for transmission. If no subset has a strictly positive reward in a fading state, then a full outage (i.e., subset S_0 is active) is declared in that fading state. This is very similar to the optimum common outage transmission policy, except that, with individual outage, there is a different channel usage (transmission) reward (η_k) associated with each subset of users S_k. It is possible that two different values of k maximize the function $(\eta_k - \boldsymbol{\lambda}\cdot \boldsymbol{P}^*(\boldsymbol{h},k))$ in some fading states, but this occurs with probability zero since we are concerned only with continuous fading distributions. For discrete fading distributions, there is the possibility that more than one subset would be selected for some fading states. See [3, Sec. IV.B] for a discussion of the same problem with regard to the broadcast channel.

B. Individual Outage Transmission Policy

For each fading state $\boldsymbol{h} \in \mathcal{H}_{\text{all}}$, the optimal transmission policy that minimizes the outage probability (given in terms of the optimal Lagrangian multiplier vector $\boldsymbol{\lambda}^*$) is as follows.

1. If $w^*(\boldsymbol{h},0) = 1$, then an outage is declared from all users and no power is assigned to any user.
2. If $w^*(\boldsymbol{h},k) = 1$ for some $k \ne 0$, then all users in the subset S_k transmit at their respective rates using power $\boldsymbol{P}^*(\boldsymbol{h},k)$ as defined in (58). Furthermore, successive decoding can be used with the decoding order described in (59).

Notice that the optimal transmission policy is again *deterministic*, i.e., either no user transmits in a fading state or all users in some subset S_k transmit. The intuition behind this is the same as for the optimal common outage transmission policy, i.e., that the linear program in (62) is extremized at the boundary of the constraint set.

Unlike the optimal common outage transmission policy, this policy is not exactly a threshold policy, though it is in fact quite similar. In the case of individual outage, there are $2^M - 1$ rewards (corresponding to the $2^M - 1$ possible transmission subsets) to calculate in each fading state, i.e., $\eta_k - \boldsymbol{\lambda}\cdot \boldsymbol{P}^*(\boldsymbol{h},k)$. For common outage, there is only one possible set of transmitting users, i.e., all users, and the reward function is given by $1 - \boldsymbol{\lambda}\cdot \boldsymbol{P}^*(\boldsymbol{h})$. For common outage, if this single reward function is strictly positive, then it is worthwhile to transmit in that fading state. For individual outage, only the best subset, i.e., the subset with the largest reward, is considered. If the best subset has a positive reward, then that subset should transmit in that state.

Under this transmission policy, the resulting outage probability of each user i $(i = 1,\dots,M)$ is

$$Pr_i = 1 - E_{\boldsymbol{h}} \left[\sum_{k=1}^{2^M-1} w^*(\boldsymbol{h},k)\psi(k,i) \right]. \tag{64}$$

The average power used by each user i $(i = 1,\dots,M)$ is given by

$$\overline{P}_i = E_{\boldsymbol{h}} \left[\sum_{k=1}^{2^M-1} w^*(\boldsymbol{h},k) P_i^*(\boldsymbol{h},k) \right]. \tag{65}$$

The vector of optimal Lagrangian multipliers $\boldsymbol{\lambda}^*$ guarantees that the power constraint of each user is satisfied.

As discussed in Section V-B for the common outage case, if $\lambda_i^* = 0$ for two or more users, the decoding order and subsequent power allocation policy in (59) and (58) is not uniquely defined. In this scenario, there are *multiple* solutions to the dual problem given the optimal $\boldsymbol{\lambda}^*$. However, we are guaranteed that at least one of them is a solution that meets the power constraints of all users (not just those users with $\lambda_i^* > 0$). A simple way to circumvent this potential difficulty is to lower the power constraint of one or more of these users until the vector of optimum Lagrangian multipliers is either strictly positive or zero for only one user.

C. Optimal Lagrangian Multipliers

In the previous subsections, we characterized the outage/usage probability region and the optimal transmission policy assuming knowledge of the optimum Lagrangian multiplier (power price) vector $\boldsymbol{\lambda}^*$. Therefore, given a power constraint vector $\overline{\boldsymbol{P}}^*$, a target rate vector $\boldsymbol{R}$, and a transmission reward vector $\boldsymbol{\mu}$, an important question is how to obtain the optimal power price vector $\boldsymbol{\lambda}^*$ that corresponds to the boundary of the usage probability region. It is easy to see that the ellipsoid algorithm used to find the optimal power price vector in the common outage case (described in Section V-C and Part F of the Appendix) can also be used in the individual outage case. For the case of individual outage, the subgradient $\boldsymbol{s}$ at $\boldsymbol{\lambda}_0$ is again given by the difference between the average power constraint vector and the average transmit power vector

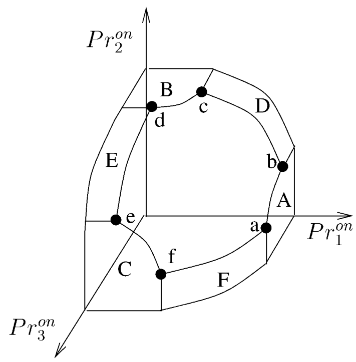

Fig. 3. Three-user usage probability region.

resulting from the optimal power policy corresponding to the price vector $\boldsymbol{\lambda_0}$. That is,

$$\boldsymbol{s} = \overline{\boldsymbol{P}}^* - E_{\boldsymbol{h}} \left[\sum_{k=1}^{2^M - 1} w^*(\boldsymbol{h}, k, \boldsymbol{\lambda_0}) \boldsymbol{P}^*(\boldsymbol{h}, k, \boldsymbol{\lambda_0}) \right] \qquad (66)$$

where $\boldsymbol{P}^*(\boldsymbol{h}, k, \boldsymbol{\lambda_0})$ and $w^*(\boldsymbol{h}, k, \boldsymbol{\lambda_0})$ are defined in (58) and (63), respectively.[11] With this choice of subgradient, the ellipsoid algorithm in Part F of the Appendix, including the procedure to find an initial polyhedron that $\boldsymbol{\lambda}^*$ must lie in, can be used without any modification.

D. Extreme Points

By varying the channel usage reward vector $\boldsymbol{\mu}$ and keeping $\boldsymbol{\mu}$ strictly positive, we can obtain all regular points on the boundary surface of the outage probability region $\mathcal{O}_I(\overline{\boldsymbol{P}}^*, \boldsymbol{R})$. However, there are points on the boundary surface of $\mathcal{O}_I(\overline{\boldsymbol{P}}^*, \boldsymbol{R})$ that cannot be explicitly characterized using this method. This corresponds to the case when $\boldsymbol{\mu}$ is not strictly positive. Alternatively, the extreme points correspond to boundary points that are also on the boundary surface of the usage probability region of some subset of users (i.e., on the boundary surface of the usage probability region of only Users 1 and 2 of a three-user system). The exact nature of these extreme points is best illustrated through an example.

In Fig. 3, the usage probability region for a three-user system is shown. The surface connecting points a,b,c,d,e, and f, is the boundary surface of the usage probability region. The regular points correspond to the interior of this surface, while the boundary of this surface composes the extreme points. If we only consider the fading distribution of the channel for User 1 (in the absence of Users 2 and 3), there is a certain maximum usage probability achievable with the given power constraint and given rate. This clearly is an upper bound to Pr_1^{on}. Face A in the figure is the area where this upper bound coincides with the usage probability region of the three-user system. In order for User 1 to achieve its single-user usage capacity, clearly User 1 must be decoded last in every fading state so that User 1 experiences no interference. Thus, User 1's outage states are chosen independent of the fading states of the other users. Given that User 1 is decoded last, Users 2 and 3 both treat User 1 as an additional source of background noise. Thus, face A is the usage probability region for Users 2 and 3 given that User 1 (which is treated as interference) is using his optimal strategy to maximize his usage probability. Notice that face A looks like a two-user usage probability region because, although User 1 is decoded last in every fading state, there is still a tradeoff between Users 2 and 3. At point a, in every fading state User 2 is decoded first, followed by User 3, followed by User 1. Similarly, at point b, User 3 is decoded first, followed by User 2, followed by User 1. At the points between a and b, the decoding order between Users 2 and 3 is determined by a power price vector and the fading state (i.e., according to (34)), or in other words, those points are similar to regular points for the two-user usage probability region. Similarly, face B corresponds to decoding User 2 last in every fading state and face C corresponds to decoding User 3 last in every fading state.

Face D corresponds to the area where Users 1 and 2 are always decoded after User 3. Thus, the rates achieved by Users 1 and 2 between points b and c correspond to the two-user usage probability region of Users 1 and 2, in the absence of User 3. Since User 3 is decoded first, he must treat Users 1 and 2 as noise. Similarly, face E corresponds to decoding User 1 first in every fading state and decoding Users 2 and 3 last. Finally, face F corresponds to decoding User 2 first in every fading state.

Thus, to characterize these extreme points, we must give absolute decoding order priority to groups of users, as we did to characterize extreme points of the average power region with common outage in Section V-E. Again, we let $\mathcal{L}$ represent a partition of $G = \{1, \ldots, M\}$ into $N \geq 1$ subsets and we successively decode users in the set G_N first, followed by the users in G_{N-1}, and so on. In Fig. 3, point a corresponds to $G_1 = \{1\}$, $G_2 = \{3\}$, and $G_3 = \{2\}$, and points on the line between a and b correspond to $G_1 = \{1\}$ and $G_2 = \{2, 3\}$. Given a partition, we can essentially treat each subset of users separately, with the caveat that some users see users in other subsets as interference. The technical details of this process are described in Part I of the Appendix.

E. Average Power Region

In Sections VI-A and VI-B, given the power constraint vector $\overline{\boldsymbol{P}}^*$ and rate vector $\boldsymbol{R}$ of the M users, we derived the outage probability region $\mathcal{O}_I(\overline{\boldsymbol{P}}^*, \boldsymbol{R})$ and the corresponding optimal power allocation policy. Now, for a given rate vector $\boldsymbol{R}$ and average outage probability vector $\boldsymbol{Pr}^*$, we consider the average power region $APV_{\text{out}}(\boldsymbol{Pr}^*, \boldsymbol{R})$, defined as the set of all possible average power vectors that can support rate vector $\boldsymbol{R}$ with the average outage probability of each user i no larger than Pr_i^*, $\forall\, 1 \leq i \leq M$. That is,

$$APV_{\text{out}}(\boldsymbol{Pr}^*, \boldsymbol{R}) \triangleq \{\boldsymbol{P} : \boldsymbol{Pr}^* \in \mathcal{O}_I(\boldsymbol{P}, \boldsymbol{R})\}. \qquad (67)$$

The average power region provides an alternative (i.e., an alternative to the individual outage/usage probability region) implicit characterization of the individual outage capacity region because a rate vector $\boldsymbol{R}$ is in $C_{\text{out}}(\overline{\boldsymbol{P}}, \boldsymbol{Pr})$ if and only if $\overline{\boldsymbol{P}}$ is in $APV_{\text{out}}(\boldsymbol{Pr}, \boldsymbol{R})$.

[11]Note that, for simplicity, we did not include $\boldsymbol{\lambda}$ as a parameter in the functions $\boldsymbol{P}^*(\boldsymbol{h}, k)$ and $w^*(\boldsymbol{h}, k)$ given by (58) and (63), respectively, though both of them depend on the value of $\boldsymbol{\lambda}$.

By convexity of the set $\mathcal{Q}_I(\boldsymbol{R})$ defined in (52), it is clear from (54) that an average power vector will be on the boundary surface of $APV_{\text{out}}(\boldsymbol{Pr}^*, \boldsymbol{R})$ if and only if it is a solution to the minimization problem

$$\min_{\boldsymbol{P}} \boldsymbol{\lambda} \cdot \boldsymbol{P} \text{ subject to: } \mathbf{1} - \boldsymbol{Pr}^* \in \overline{\mathcal{O}}_I(\boldsymbol{P}, \boldsymbol{R}) \qquad (68)$$

for some vector $\boldsymbol{\lambda} \in \Re_+^M$. Therefore, for each given power price vector $\boldsymbol{\lambda} \in \Re_+^M$, an average power vector $\boldsymbol{P}$ solves (68) if and only if there exists a Lagrangian multiplier vector $\boldsymbol{\mu}$ such that $(\boldsymbol{Pr}^{\mathbf{on}}, \boldsymbol{P})$ is a solution to the problem

$$\min_{(\boldsymbol{Pr}^{\mathbf{on}}, \overline{\boldsymbol{P}}) \in \mathcal{Q}_I(\boldsymbol{R})} \left\{ \boldsymbol{\lambda} \cdot \overline{\boldsymbol{P}} - \boldsymbol{\mu} \cdot \boldsymbol{Pr}^{\mathbf{on}} \right\} \qquad (69)$$

and $\boldsymbol{Pr}^{\mathbf{on}} \geq \mathbf{1} - \boldsymbol{Pr}^*$. This optimization problem is similar to the one in (54). However, in (54), the average transmit power vector $\overline{\boldsymbol{P}}^*$ and the channel usage reward vector $\boldsymbol{\mu}$ are given, but the appropriate power price vector $\boldsymbol{\lambda}$ must be found such that the average power constraint $\overline{P}_i^*$ of each user i $(1 \leq i \leq M)$ is satisfied. The resulting average outage probability vector will lie on the boundary surface of $\mathcal{O}_I(\overline{\boldsymbol{P}}^*, \boldsymbol{R})$ for the given $\boldsymbol{\mu}$. In (69), we have the opposite scenario: the power price vector $\boldsymbol{\lambda}$ and the average outage probability vector $\boldsymbol{Pr}^*$ are given, but the usage reward vector $\boldsymbol{\mu}$ must be found such that the average outage probability constraint Pr_i^* of each user i $(1 \leq i \leq M)$ is satisfied. Then, the resulting average power vector will lie on the boundary surface of $APV_{\text{out}}(\boldsymbol{Pr}^*, \boldsymbol{R})$ for the given $\boldsymbol{\lambda}$.

Given the optimal Lagrangian multiplier vector $\boldsymbol{\mu}$, by the equivalence of (54) and (69), the optimal choice of $\boldsymbol{P}(\boldsymbol{h}, k)$ and $w(\boldsymbol{h}, k)$ are given by (58) and (63), respectively. The optimal Lagrangian multiplier vector $\boldsymbol{\mu}$ guarantees that the outage probability constraints are met.

The optimal Lagrangian multiplier vector $\boldsymbol{\mu}$ can also be found using the ellipsoid method described in Section V-C and Part F of the Appendix. In this case, the subgradient is given by

$$\boldsymbol{s} = \boldsymbol{Pr}^* - E_{\boldsymbol{h}} \left[\sum_{k=1}^{2^M - 1} w^*(\boldsymbol{h}, k, \boldsymbol{\lambda_0}) \boldsymbol{\Psi}(k) \right].$$

Again, the vector of optimal Lagrangian multipliers is such that the outage constraints are met with equality (or possibly inequality if the Lagrangian multiplier of some user is zero).

As in the case with a given common outage probability, there are also extreme points on the boundary surface of $APV_{\text{out}}(\boldsymbol{Pr}^*, \boldsymbol{R})$ that cannot be parameterized by any $\boldsymbol{\lambda} \in \Re_+^M$. These extreme points are explicitly characterized for the average power region in Part J of the Appendix.

VII. Multiple-Access and Broadcast Channel Duality

The recently established duality between the Gaussian MAC and broadcast channel [16] applies very cleanly to outage capacity. The duality of these channels implies that the capacity region of the constant broadcast channel with channel gains $h_1, \ldots, h_M$, noise power of variance one at each receiver, and average power constraint $\overline{P}$, is exactly equal to the capacity region of the *dual* MAC, which has the same channel gains $h_1, \ldots, h_M$, noise power of variance one, and a *sum* power constraint of $\overline{P}$ across all M transmitters. In [16, Sec. V], this duality is extended to the outage capacity of the fading MAC and the broadcast channel. If we let $\mathcal{C}_{\text{out}}^{\text{MAC}}(\overline{\boldsymbol{P}}, Pr; \boldsymbol{H})$ denote the common outage capacity of the MAC with fading distribution $\boldsymbol{H}$, and let $\mathcal{C}_{\text{out}}^{\text{BC}}(\overline{P}, Pr; \boldsymbol{H})$ denote the common outage capacity of the broadcast channel (defined in [7]) with fading distribution $\boldsymbol{H}$, then the following relationships hold:

$$\mathcal{C}_{\text{out}}^{\text{BC}}(\overline{P}, Pr; \boldsymbol{H}) = \bigcup_{\{\overline{\boldsymbol{P}} : \mathbf{1} \cdot \overline{\boldsymbol{P}} = \overline{P}\}} \mathcal{C}_{\text{out}}^{\text{MAC}}(\overline{\boldsymbol{P}}, Pr; \boldsymbol{H}) \qquad (70)$$

$$\mathcal{C}_{\text{out}}^{\text{MAC}}(\overline{\boldsymbol{P}}, Pr; \boldsymbol{H}) = \bigcap_{\boldsymbol{\alpha} > \mathbf{0}} \mathcal{C}_{\text{out}}^{\text{BC}} \left(\mathbf{1} \cdot \frac{\overline{\boldsymbol{P}}}{\boldsymbol{\alpha}}, Pr; \boldsymbol{\alpha H} \right). \qquad (71)$$

These expressions are stated here for the common outage capacity only, but they hold for individual outage capacity (referred to as the independent outage capacity in [7]) as well. Duality shows that the outage capacity region of the broadcast channel is equal to the union of outage capacity regions of the MAC, and the outage capacity region of the MAC is equal to the intersection of outage capacity regions of the scaled broadcast channel. Additionally, the capacity-achieving power policies for the broadcast channel can be directly mapped via a simple transformation to the optimal power policies for the MAC, and *vice versa*. This mapping preserves the rates achieved in each fading state, as well as the sum power. See [16] for more details.

VIII. Auxiliary Constraints on Transmit Power

In Sections V and VI, we considered only average transmit power constraints for the M users. In practice, sometimes we have to consider the peak transmit power constraint of each user as well. That is, in addition to the average transmit power constraint vector $\overline{\boldsymbol{P}}$ of the M users, for each fading state $\boldsymbol{h}$, the transmit power vector of the M users must be no larger than $\hat{\mathbf{P}} = (\hat{P}_1, \ldots, \hat{P}_M)$. Under these auxiliary constraints, given a rate vector $\boldsymbol{R}$, the problem of deriving the minimum average common outage probability of all users or deriving the average outage probability region of the M users can be similarly solved as shown in Sections V and VI, except that now we have to solve the minimization problems (32) and (57) subject to the additional peak power constraint vector $\hat{\mathbf{P}}$, i.e., the additional constraint for problem (32) is

$$\mathbf{P}(\boldsymbol{h}) \leq \hat{\mathbf{P}}, \qquad \forall\, \boldsymbol{h} : w(\boldsymbol{h}) > 0$$

and the additional constraint for problem (57) is

$$\mathbf{P}(\boldsymbol{h}, k) \leq \hat{\mathbf{P}}, \qquad \forall\, \boldsymbol{h}, k : w(\boldsymbol{h}, k) > 0. \qquad (72)$$

Note that since $P_i(\boldsymbol{h}, k) = 0$ for any user $i \notin S_k$, the peak power constraint in (72) is equivalent to $P_i(\boldsymbol{h}, k) \leq \hat{P}_i, \forall\, i \in S_k$. The solution to problems (32) and (57) under the additional peak power constraint vector $\hat{\mathbf{P}}$ is given in [6, Sec. V]. That is, for a given power price vector $\boldsymbol{\lambda}$, assuming that $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_M$, the transmit power $P_i(\boldsymbol{h})$ $(1 \leq i \leq M)$ or $P_i(\boldsymbol{h}, k)$ $(i \in S_k)$ of each user i is now obtained through a greedy algorithm. Specifically, by denoting $x_i^{(k)}$ as the value of $P_i(\boldsymbol{h})$ in the kth step of the following algorithm, the solution $P_i(\boldsymbol{h})$ for each user i $(1 \leq i \leq M)$ is obtained after M steps.

- *Initialization:* Set $x_i^{(0)} = \hat{P}_i$ for all i $(1 \leq i \leq M)$. If $\boldsymbol{R} \notin \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \hat{\boldsymbol{P}})$ then stop. Otherwise, set $k = 1$.
- *Step* k: Let $\pi^{(k)}(\cdot)$ be a permutation on
$$\{1, \ldots, k-1, k+1, \ldots, M\}$$
(note that $\pi^{(k)}(k)$ does not exist) such that
$$\begin{aligned}\frac{x_{\pi^{(k)}(1)}h_{\pi^{(k)}(1)}}{R_{\pi^{(k)}(1)}} \leq \cdots \leq \frac{x_{\pi^{(k)}(k-1)}h_{\pi^{(k)}(k-1)}}{R_{\pi^{(k)}(k-1)}} \\ \leq \frac{x_{\pi^{(k)}(k+1)}h_{\pi^{(k)}(k+1)}}{R_{\pi^{(k)}(k+1)}} \\ \leq \cdots \leq \frac{x_{\pi^{(k)}(M)}h_{\pi^{(k)}(M)}}{R_{\pi^{(k)}(M)}}.\end{aligned}$$
Then set
$$x_i^{(k)} = \begin{cases} x_i^{(k-1)}, & \text{if } i \neq k \\ \frac{1}{h_i} \cdot \max\limits_{j \neq k}[f(S_j \cup \{k\}) - \sum\limits_{l \in S_j} x_l], & \text{if } i = k \end{cases}$$
where $f(\cdot)$ is defined as
$$f(F) = \sigma^2 \left[\exp(2\sum_{i \in F} R_i)\right] - 1$$
for any $F \subseteq \{1, \ldots, M\}$ and
$$S_j \equiv \{\pi^{(k)}(1), \ldots, \pi^{(k)}(j)\}$$
with $\pi^{(k)}(k)$ nonexistent. Go to Step $k+1$.
- Stop after M steps. △

Similarly, $P_i(\boldsymbol{h}, k)$ $(\forall\, i \in S_k)$ can be obtained by applying the above algorithm to the $|S_k|$ users in the set S_k instead of to the M users. Note that the solution $P_i^*(\boldsymbol{h})$ $(1 \leq i \leq M)$ in (33) and the solution $P_i^*(\boldsymbol{h}, k)$ $(\forall\, i \in S_k)$ in (58) that can be achieved by successive decoding are obtained through an algorithm that is actually a special case of the above greedy algorithm [6]. However, when peak power constraints are imposed on the M users, the solution $P_i(\boldsymbol{h})$ or $P_i(\boldsymbol{h}, k)$ obtained through this greedy algorithm cannot always be achieved by successive decoding in general.

IX. Conclusion

We have obtained the outage capacity region of fading MACs under the assumption that perfect CSI is available both at the transmitters and at the receiver. The capacity region is obtained implicitly by deriving the minimum common outage probability and the individual outage probability region for a given rate vector. Given the average power constraint of each user, we have derived the power allocation policy that minimizes the common outage probability for a given rate vector when transmission to all users is turned off simultaneously. When an outage can be declared for each user individually, we have derived a power allocation strategy to achieve the outage probability region boundary for the given rate vector.

In both cases, the optimal power allocation policies have been shown to be purely deterministic functions of the fading state, and standard convex optimization algorithms have been used for obtaining the optimal power control parameters. These optimal power allocation policies show that, similar to the zero-outage scenario, successive decoding is optimal and in each fading state, the decoding order is determined by the power prices of the users and their fading gains. By applying these optimal power allocation strategies, we have also obtained the average power regions that can support a rate vector with a given common outage probability or a given outage probability vector for the M users. When there are additional peak power constraints, the optimal power allocation for the M users in each fading state can be obtained through a greedy algorithm, though in general it cannot be achieved by successive decoding.

Appendix

A. *Proof of Proposition 3.1*

We prove (7) by showing that for every fully random power policy $\mathcal{P}^{\mathrm{rand}}$ satisfying the power constraint vector $\overline{\boldsymbol{P}}$, there exists an equivalent power policy $\mathcal{P}$ with cardinality two in each fading state that achieves the same rate vector $\boldsymbol{R}$ using the same or less power. Given an arbitrary random power policy $\mathcal{P}^{\mathrm{rand}}$, let $\boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h})$ denote the vector of random power allocation functions in each fading state $\boldsymbol{h} \in \mathcal{H}_{\mathrm{all}}$. We construct a new power policy $\mathcal{P}$ of the form in (6) by defining a deterministic power allocation function $\boldsymbol{P}(\boldsymbol{h})$ and a deterministic probability of transmission function $w(\boldsymbol{h})$ in *each* fading state as follows:

$$\boldsymbol{P}(\boldsymbol{h}) = E[\boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h}) | \boldsymbol{R} \in \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h}))] \tag{73}$$
$$w(\boldsymbol{h}) = Pr[\boldsymbol{R} \in \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h}))] \tag{74}$$

where the probability and expectation are computed over only the randomization in each fading state (i.e., over $f_{\boldsymbol{P}}^{\mathrm{rand}}(\boldsymbol{h})$), and not over the set of all fading states $\mathcal{H}_{\mathrm{all}}$. Notice that in each fading state $\boldsymbol{h}$

$$\begin{aligned} E[\boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h})] &= E[\boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h}) | \boldsymbol{R} \in \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h}))] \\ &\quad \cdot Pr[\boldsymbol{R} \in \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h}))] \\ &\quad + E[\boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h}) | \boldsymbol{R} \notin \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h}))] \\ &\quad \cdot Pr[\boldsymbol{R} \notin \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h}))] \\ &\geq \boldsymbol{P}(\boldsymbol{h}) w(\boldsymbol{h}) \end{aligned}$$

where, again, the probabilities and expectations are computed over only the randomization in each fading state (i.e., over $f_{\boldsymbol{P}}^{\mathrm{rand}}(\boldsymbol{h})$) and not over the set $\mathcal{H}_{\mathrm{all}}$. Therefore,

$$E[\boldsymbol{P}(\boldsymbol{h}) w(\boldsymbol{h})] \leq E[\boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h})].$$

Next, notice that if $\boldsymbol{Q}'$ and $\boldsymbol{Q}''$ are arbitrary power vectors satisfying $\boldsymbol{R} \in \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{Q}')$ and $\boldsymbol{R} \in \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{Q}'')$, then due to the concavity of the log function in the definition of $\mathcal{C}_{\rm MAC}(\boldsymbol{h},\boldsymbol{P})$ in (3), we have

$$\boldsymbol{R} \in \mathcal{C}_{\rm MAC}(\boldsymbol{h}, \alpha\boldsymbol{Q}' + (1-\alpha)\boldsymbol{Q}'')$$

for any $\alpha \in [0,1]$. Since $\boldsymbol{P}(\boldsymbol{h})$ is defined as the conditional expected value of $\boldsymbol{P}^{\rm rand}(\boldsymbol{h})$ on the set of power vectors for which $\boldsymbol{R}$ is admissible, the same argument yields $\boldsymbol{R} \in \mathcal{C}_{\rm MAC}(\boldsymbol{h},\boldsymbol{P}(\boldsymbol{h}))$ for any fading state with $w(\boldsymbol{h}) > 0$. In addition, it is clear that

$$E[w(\boldsymbol{h})] = Pr[\boldsymbol{R} \in \mathcal{C}_{\rm MAC}(\boldsymbol{h}, \boldsymbol{P}^{\rm rand}(\boldsymbol{h}))].$$

Therefore, the power policy $\mathcal{P}$, which is of cardinality two in each fading state, uses no more power than the power policy $\mathcal{P}^{\rm rand}$ and achieves $\boldsymbol{R}$ with the same probability.

B. Proof of Proposition 3.2

1. We first prove that, for any given rate vector $\boldsymbol{R}$, if

$$\boldsymbol{R} \in \mathcal{C}_{\rm zero}\left(\frac{\overline{\boldsymbol{P}}}{1-Pr}, \mathcal{H}_{\rm tran}\right)$$

and $E[w(\boldsymbol{h})] = 1 - Pr$, then $\boldsymbol{R} \in \mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, Pr)$.

By the definition of zero-outage capacity in [6, Theorem 2.3], a rate vector $\boldsymbol{R}$ is in $\mathcal{C}_{\rm zero}(\frac{\overline{\boldsymbol{P}}}{1-Pr}, \mathcal{H}_{\rm tran})$ if and only if there exists a *deterministic* power allocation function $\boldsymbol{Q}(\boldsymbol{h})$ ($\boldsymbol{h} \in \mathcal{H}_{\rm tran}$) that meets the power constraint

$$\int_{\boldsymbol{h}\in\mathcal{H}_{\rm tran}} \boldsymbol{Q}(\boldsymbol{h})g(\boldsymbol{h})\, d\boldsymbol{h} \le \overline{\boldsymbol{P}}/(1-Pr)$$

and satisfies $\boldsymbol{R} \in \mathcal{C}_{\rm MAC}(\boldsymbol{h}, \boldsymbol{Q}(\boldsymbol{h}))$, $\forall\, \boldsymbol{h} \in \mathcal{H}_{\rm tran}$. We can easily map from any such power allocation policy to one of the form $\mathcal{P}$ defined in (6) by letting $\boldsymbol{P}(\boldsymbol{h}) = \boldsymbol{Q}(\boldsymbol{h})$ ($\forall\, \boldsymbol{h} \in \mathcal{H}_{\rm tran}$) and using the same $w(\boldsymbol{h})$. Since $w(\boldsymbol{h}) = 0$ for any $\boldsymbol{h} \notin \mathcal{H}_{\rm tran}$ and $E[w(\boldsymbol{h})] = 1 - Pr$, given the true PDF $p(\boldsymbol{h})$ of the fading state $\boldsymbol{h}$ ($\boldsymbol{h} \in \mathcal{H}_{\rm all}$) and the new PDF $g(\boldsymbol{h})$ as defined in (10), we have

$$\int_{\boldsymbol{h}\in\mathcal{H}_{\rm all}} \boldsymbol{P}(\boldsymbol{h})w(\boldsymbol{h})p(\boldsymbol{h})\, d\boldsymbol{h} = \int_{\boldsymbol{h}\in\mathcal{H}_{\rm tran}} \boldsymbol{Q}(\boldsymbol{h})g(\boldsymbol{h})E[w(\boldsymbol{h})]\, d\boldsymbol{h} \le \overline{\boldsymbol{P}} \quad (75)$$

which implies that the power constraint of (8) in Proposition 3.1 is satisfied. In addition, $\boldsymbol{R} \in \mathcal{C}_{\rm MAC}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h}))$ for all $\boldsymbol{h} \in \mathcal{H}_{\rm tran}$. Therefore, according to Proposition 3.1, it is obvious that $\boldsymbol{R} \in \mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, Pr)$.

2. Next, we prove that, for any given rate vector $\boldsymbol{R}$, if $\boldsymbol{R} \in \mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, Pr)$ then there exists a power policy $\mathcal{P}$ such that $E[w(\boldsymbol{h})] = 1 - Pr$ and

$$\boldsymbol{R} \in \mathcal{C}_{\rm zero}\left(\frac{\overline{\boldsymbol{P}}}{1-Pr}, \mathcal{H}_{\rm tran}\right).$$

From Proposition 3.1, it is clear that if $\boldsymbol{R} \in \mathcal{C}_{\rm out}(\overline{\boldsymbol{P}}, Pr)$, then there exists a power policy $\mathcal{P}$ such that $E[w(\boldsymbol{h})] \ge 1 - Pr$ and

$$E[\boldsymbol{P}(\boldsymbol{h})w(\boldsymbol{h})] \le \overline{\boldsymbol{P}}.$$

If $E[w(\boldsymbol{h})] = 1 - Pr$, we let $\boldsymbol{Q}(\boldsymbol{h}) = \boldsymbol{P}(\boldsymbol{h})$, $\forall\, \boldsymbol{h} \in \mathcal{H}_{\rm tran}$, and it can be similarly shown as above that

$$\int_{\boldsymbol{h}\in\mathcal{H}_{\rm tran}} \boldsymbol{Q}(\boldsymbol{h})g(\boldsymbol{h})\, d\boldsymbol{h} \le \overline{\boldsymbol{P}}/(1-Pr).$$

Therefore, since $\boldsymbol{R} \in \mathcal{C}_{\rm MAC}(\boldsymbol{h}, \boldsymbol{Q}(\boldsymbol{h}))$, $\forall\, \boldsymbol{h} \in \{\boldsymbol{h} : w(\boldsymbol{h}) > 0\}$, we have

$$\boldsymbol{R} \in \mathcal{C}_{\rm zero}\left(\frac{\overline{\boldsymbol{P}}}{1-Pr}, \mathcal{H}_{\rm tran}\right).$$

In the case where $E[w(\boldsymbol{h})] > 1 - Pr$, we construct a new power policy by keeping the power allocation function $\boldsymbol{P}(\boldsymbol{h})$ unchanged while scaling down $w(\boldsymbol{h})$ in each fading state $\boldsymbol{h} \in \mathcal{H}_{\rm tran}$ with a positive factor α ($0 < \alpha < 1$) such that $E[\alpha w(\boldsymbol{h})] = 1 - Pr$. Under this new power policy, obviously $\mathcal{H}_{\rm tran}$ remains the same, and the power constraint is still satisfied and, in addition, it still holds that

$$\boldsymbol{R} \in \mathcal{C}_{\rm MAC}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h})), \qquad \forall\, \boldsymbol{h} \in \{\boldsymbol{h} : w(\boldsymbol{h}) > 0\}.$$

Therefore, given this new power policy, it can be similarly shown as above that

$$\boldsymbol{R} \in \mathcal{C}_{\rm zero}\left(\frac{\overline{\boldsymbol{P}}}{1-Pr}, \mathcal{H}_{\rm tran}\right).$$

Note that now the new PDF of $\boldsymbol{h} \in \mathcal{H}_{\rm tran}$ is $g(\boldsymbol{h}) = \alpha w(\boldsymbol{h})p(\boldsymbol{h})/(1-Pr)$.

Combining the proofs in the above two items, we conclude that Proposition 3.2 is true.

C. Proof of Proposition 3.3

We prove (17) by showing that for every fully random power policy $\mathcal{P}^{\rm rand}$ satisfying the power constraint vector $\overline{\boldsymbol{P}}$ and the outage probability constraint vector $\boldsymbol{Pr}$, there exists an equivalent power policy $\mathcal{P}^{I}$ with cardinality 2^M in each fading state that achieves the same rate vector $\boldsymbol{R}$ using the same or less power. Given such a random power policy $\mathcal{P}^{\rm rand}$, let $\boldsymbol{P}^{\rm rand}(\boldsymbol{h})$ denote the vector of random power allocation functions in each fading state $\boldsymbol{h} \in \mathcal{H}_{\rm all}$, and let $\boldsymbol{R}\left(\boldsymbol{h}, \boldsymbol{P}^{\rm rand}(\boldsymbol{h})\right)$ denote the corresponding rate allocation function vector that satisfies the outage probability constraints of the M users. In addition, let $\mathcal{A}_{\boldsymbol{h},k}$ denote the corresponding *event* that only users in the kth subset S_k ($k = 0, \ldots, 2^M - 1$) are transmitting at their designated rates (or higher rates) in fading state $\boldsymbol{h}$, and no other users are. That is,

$$\mathcal{A}_{\boldsymbol{h},k} = \Big\{ \boldsymbol{R}(\boldsymbol{h}, \boldsymbol{P}^{\rm rand}(\boldsymbol{h})) \ge \boldsymbol{\Psi}(k)\boldsymbol{R} \text{ and } \boldsymbol{R}(\boldsymbol{h}, \boldsymbol{P}^{\rm rand}(\boldsymbol{h})) < \boldsymbol{\Psi}(j)\boldsymbol{R},\ \forall\, j > k \text{ such that } S_j \supset S_k \Big\}.$$

We construct a new power policy $\mathcal{P}^{I}$ of the form in (16) by defining a deterministic power allocation function $\boldsymbol{P}(\boldsymbol{h}, k)$ and a deterministic probability of transmission function $w(\boldsymbol{h}, k)$ for the kth subset of users in *each* fading state as follows:

$$\boldsymbol{P}(\boldsymbol{h}, k) = E[\boldsymbol{\Psi}(k)\boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h})|\mathcal{A}_{\boldsymbol{h},k}] \tag{76}$$

$$w(\boldsymbol{h}, k) = Pr[\mathcal{A}_{\boldsymbol{h},k}] \tag{77}$$

where the expectation and probability are computed over only the randomization in each fading state, and not over the set of all fading states $\mathcal{H}_{\mathrm{all}}$. It is obvious that $\sum_{k=0}^{2^M-1} w(\boldsymbol{h}, k) = 1$, $P_i(\boldsymbol{h}, k) = 0$ for any $i \notin S_k$, and $\boldsymbol{P}(\boldsymbol{h}, 0) = \boldsymbol{0}$ since $\boldsymbol{\Psi}(0) = \boldsymbol{0}$. Notice that in each fading state $\boldsymbol{h}$

$$\begin{aligned} E[\boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h})] &= \sum_{k=0}^{2^M-1} E[\boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h})|\mathcal{A}_{\boldsymbol{h},k}] \cdot Pr[\mathcal{A}_{\boldsymbol{h},k}] \\ &\geq \sum_{k=0}^{2^M-1} \boldsymbol{P}(\boldsymbol{h}, k) w(\boldsymbol{h}, k) \end{aligned}$$

where, again, the probabilities and expectations are computed over only the randomization in each fading state and not over the set $\mathcal{H}_{\mathrm{all}}$. Therefore,

$$E\left[\sum_{k=0}^{2^M-1} \boldsymbol{P}(\boldsymbol{h}, k) w(\boldsymbol{h}, k)\right] \leq E[\boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h})].$$

Furthermore, by the same argument used as in the proof of Proposition 3.1, it is straightforward to show that $\boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h}, k))$ for any fading state with $w(\boldsymbol{h}, k) > 0$, $k = 1, \ldots, 2^M - 1$. Since $\boldsymbol{\Psi}(k)\boldsymbol{P}(\boldsymbol{h}, k) = \boldsymbol{P}(\boldsymbol{h}, k)$, we have $\boldsymbol{\Psi}(j)\boldsymbol{R} \notin \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h}, k))$, $\forall\, j > k$ such that $S_j \supset S_k$. Therefore, for any fading state with $w(\boldsymbol{h}, k) > 0$, $k = 1, \ldots, 2^M - 1$, there exists a corresponding rate allocation function vector $\boldsymbol{R}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h}, k)) \in \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h}, k))$ satisfying

$$\boldsymbol{R}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h}, k)) \geq \boldsymbol{\Psi}(k)\boldsymbol{R}$$

and

$$\boldsymbol{R}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h}, k)) < \boldsymbol{\Psi}(j)\boldsymbol{R}$$

$\forall\, j > k$ such that $S_j \supset S_k$. The resulting average probability of transmission for each user i with a rate no smaller than R_i is

$$\begin{aligned} E\left[\sum_{k=1}^{2^M-1} w(\boldsymbol{h}, k)\psi(k, i)\right] &= \sum_{k=1}^{2^M-1} Pr[\mathcal{A}_{\boldsymbol{h},k}]\psi(k, i) \\ &= Pr\left[R_i(\boldsymbol{h}, \boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h})) \geq R_i\right] \\ &\geq 1 - Pr_i, \qquad \forall\, i = 1, \ldots, M. \end{aligned}$$

In conclusion, the power policy $\mathcal{P}^I$, which is of cardinality 2^M in each fading state, uses no more power than the power policy $\mathcal{P}^{\mathrm{rand}}$ and achieves $\boldsymbol{R}$ with the same set of probabilities.

D. Proof of Proposition 3.4

For a given power policy $\mathcal{P}^I : \boldsymbol{h} \rightarrow \{\boldsymbol{P}(\boldsymbol{h}, k), w(\boldsymbol{h}, k)\}_{k=0}^{2^M-1}$ satisfying $\mathcal{P}^I \in \mathcal{F}^I$ as defined in Proposition 3.3, in order to prove (21), we first show that, for any given rate vector $\boldsymbol{R}$, if

$$\boldsymbol{\Psi}(k)\boldsymbol{R} \in \bigcap_{\{\boldsymbol{h}: w(\boldsymbol{h},k)>0\}} \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{P}(\boldsymbol{h}, k)), \;\; \forall\, k = 1, \ldots, 2^M - 1$$

then $\boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\mathrm{zero}}(\boldsymbol{A}(k), \mathcal{H}_{\mathrm{tran}}(k))$, $\forall\, k = 1, \ldots, 2^M - 1$.

By the definition of zero-outage capacity in [6, Theorem 2.3], a rate vector $\boldsymbol{\Psi}(k)\boldsymbol{R}$ is in $\mathcal{C}_{\mathrm{zero}}(\boldsymbol{A}(k), \mathcal{H}_{\mathrm{tran}}(k))$ if and only if there exists a *deterministic* power allocation function $\boldsymbol{Q}(\boldsymbol{h})$ ($\boldsymbol{h} \in \mathcal{H}_{\mathrm{tran}}(k)$) that meets the power constraint

$$\int_{\boldsymbol{h} \in \mathcal{H}_{\mathrm{tran}}(k)} \boldsymbol{Q}(\boldsymbol{h}) g(\boldsymbol{h}, k)\, d\boldsymbol{h} \leq \boldsymbol{A}(k)$$

and satisfies

$$\boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{Q}(\boldsymbol{h})), \qquad \forall\, \boldsymbol{h} \in \mathcal{H}_{\mathrm{tran}}(k).$$

It is obvious that, by simply letting $\boldsymbol{Q}(\boldsymbol{h}) = \boldsymbol{P}(\boldsymbol{h}, k)$, $\forall\, \boldsymbol{h} \in \mathcal{H}_{\mathrm{tran}}(k)$, these conditions are all satisfied.

Next we show that, for any given rate vector $\boldsymbol{R}$, if

$$\boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\mathrm{zero}}(\boldsymbol{A}(k), \mathcal{H}_{\mathrm{tran}}(k)), \qquad \forall\, k = 1, \ldots, 2^M - 1$$

then there exists another power policy with the same set of probability functions $\{w(\boldsymbol{h}, k)\}_{k=0}^{2^M-1}$ and a different set of deterministic power allocation functions $\{\boldsymbol{B}(\boldsymbol{h}, k)\}_{k=0}^{2^M-1}$ that meets the power constraint $\overline{\boldsymbol{P}}$ and satisfies

$$\boldsymbol{\Psi}(k)\boldsymbol{R} \in \bigcap_{\{\boldsymbol{h}: w(\boldsymbol{h},k)>0\}} \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{B}(\boldsymbol{h}, k)), \qquad \forall\, k = 1, \ldots, 2^M - 1. \tag{78}$$

Specifically, since

$$\boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\mathrm{zero}}(\boldsymbol{A}(k), \mathcal{H}_{\mathrm{tran}}(k)), \qquad \forall\, k = 1, \ldots, 2^M - 1$$

there exists a *deterministic* power allocation function $\boldsymbol{Q}(\boldsymbol{h})$ ($\boldsymbol{h} \in \mathcal{H}_{\mathrm{tran}}(k)$) that satisfies

$$\int_{\boldsymbol{h} \in \mathcal{H}_{\mathrm{tran}}(k)} \boldsymbol{Q}(\boldsymbol{h}) g(\boldsymbol{h}, k)\, d\boldsymbol{h} \leq \boldsymbol{A}(k)$$

and $\boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{Q}(\boldsymbol{h}))$, $\forall\, \boldsymbol{h} \in \mathcal{H}_{\mathrm{tran}}(k)$. By letting $\boldsymbol{B}(\boldsymbol{h}, k) = \boldsymbol{Q}(\boldsymbol{h})$ ($\boldsymbol{h} \in \mathcal{H}_{\mathrm{tran}}(k)$), it is clear that (78) is satisfied. In addition

$$\begin{aligned} E[\boldsymbol{B}(\boldsymbol{h}, k) w(\boldsymbol{h}, k)] &= \int_{\boldsymbol{h} \in \mathcal{H}_{\mathrm{tran}}(k)} \boldsymbol{Q}(\boldsymbol{h}) w(\boldsymbol{h}, k) p(\boldsymbol{h})\, d\boldsymbol{h} \\ &\leq E[w(\boldsymbol{h}, k)]\boldsymbol{A}(k) \\ &= E[\boldsymbol{P}(\boldsymbol{h}, k) w(\boldsymbol{h}, k)]. \end{aligned}$$

Therefore,

$$\sum_{k=0}^{2^M-1} E[\boldsymbol{B}(\boldsymbol{h}, k) w(\boldsymbol{h}, k)] \leq \sum_{k=0}^{2^M-1} E[\boldsymbol{P}(\boldsymbol{h}, k) w(\boldsymbol{h}, k)] \leq \overline{\boldsymbol{P}}.$$

According to Proposition 3.3, we have $\boldsymbol{R} \in \mathcal{C}_{\mathrm{out}}(\overline{\boldsymbol{P}}, \boldsymbol{Pr})$ and (21) holds.

Finally, we show that (22) is the equivalent expression of (21). Since

$$\overline{\mathcal{C}}_{\mathrm{zero}}(\boldsymbol{A}(k), \mathcal{H}_{\mathrm{tran}}(k)) = \{\boldsymbol{R} : \boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\mathrm{zero}}(\boldsymbol{A}(k), \mathcal{H}_{\mathrm{tran}}(k))\}$$

we have

$$\begin{aligned} &\boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\mathrm{zero}}(\boldsymbol{A}(k), \mathcal{H}_{\mathrm{tran}}(k)), \quad \forall\, k = 1, \ldots, 2^M - 1 \\ &\quad \Longleftrightarrow \boldsymbol{R} \in \overline{\mathcal{C}}_{\mathrm{zero}}(\boldsymbol{A}(k), \mathcal{H}_{\mathrm{tran}}(k)), \quad \forall\, k = 1, \ldots, 2^M - 1 \\ &\quad \Longleftrightarrow \boldsymbol{R} \in \bigcap_{k=1}^{2^M-1} \overline{\mathcal{C}}_{\mathrm{zero}}(\boldsymbol{A}(k), \mathcal{H}_{\mathrm{tran}}(k)). \end{aligned}$$

Therefore, (21) and (22) are equivalent expressions.

E. Proof of Lemma 5.1

In the following, we prove that the set $\mathcal{Q}_C(\boldsymbol{R})$ defined in (27) is convex. For a given rate vector $\boldsymbol{R}$, consider two points $(Pr^{on(1)}, \overline{\boldsymbol{P}}^{(1)})$ and $(Pr^{on(2)}, \overline{\boldsymbol{P}}^{(2)})$ in $\mathcal{Q}_C(\boldsymbol{R})$. By Proposition 3.1, there must exist power allocation functions $\boldsymbol{P}^{(1)}(\boldsymbol{h})$ and $\boldsymbol{P}^{(2)}(\boldsymbol{h})$ and probability of transmission functions $w^{(1)}(\boldsymbol{h})$ and $w^{(2)}(\boldsymbol{h})$ such that

$$Pr^{on(i)} = E_{\boldsymbol{h}}\left[w^{(i)}(\boldsymbol{h})\right], \qquad i = 1, 2 \tag{79}$$

and

$$\overline{\boldsymbol{P}}^{(i)} = E_{\boldsymbol{h}}\left[w^{(i)}(\boldsymbol{h})\mathbf{P}^{(i)}(\boldsymbol{h})\right], \qquad i = 1, 2 \tag{80}$$

with $\boldsymbol{R} \in \mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{(1)}(\boldsymbol{h}))$ for all $\boldsymbol{h}$ satisfying $w^{(1)}(\boldsymbol{h}) > 0$ and $\boldsymbol{R} \in \mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{(2)}(\boldsymbol{h}))$ for all $\boldsymbol{h}$ satisfying $w^{(2)}(\boldsymbol{h}) > 0$.

We must show that

$$\left(\alpha Pr^{on(1)} + (1-\alpha)Pr^{on(2)}, \alpha\overline{\boldsymbol{P}}^{(1)} + (1-\alpha)\overline{\boldsymbol{P}}^{(2)}\right) \in \mathcal{Q}_C(\boldsymbol{R})$$

for any $\alpha \in [0, 1]$. In order to do so, we define a random power allocation function $\boldsymbol{P}^{\text{rand}}(\boldsymbol{h})$ ($\forall\, \boldsymbol{h} \in \mathcal{H}_{\text{all}}$) as shown in (81) at the bottom of the page. Then obviously the resulting average common usage probability is

$$\begin{aligned} Pr[\boldsymbol{R} \in \mathcal{C}_{\text{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{\text{rand}}(\boldsymbol{h}))] &= E_{\boldsymbol{h}}\left[\alpha w^{(1)}(\boldsymbol{h}) + (1-\alpha)w^{(2)}(\boldsymbol{h})\right] \\ &= \alpha Pr^{on(1)} + (1-\alpha)Pr^{on(2)} \end{aligned}$$

and the average transmit power is given by

$$\begin{aligned} E[\boldsymbol{P}^{\text{rand}}(\boldsymbol{h})] &= E\Big[\alpha w^{(1)}(\boldsymbol{h})\mathbf{P}^{(1)}(\boldsymbol{h}) \\ &\qquad + (1-\alpha)w^{(2)}(\boldsymbol{h})\mathbf{P}^{(2)}(\boldsymbol{h})\Big] \\ &= \alpha\overline{\boldsymbol{P}}^{(1)} + (1-\alpha)\overline{\boldsymbol{P}}^{(2)}. \end{aligned}$$

Therefore, we have

$$\left(\alpha Pr^{on(1)} + (1-\alpha)Pr^{on(2)}, \alpha\overline{\boldsymbol{P}}^{(1)} + (1-\alpha)\overline{\boldsymbol{P}}^{(2)}\right) \in \mathcal{Q}_C(\boldsymbol{R})$$

which implies that the set $\mathcal{Q}_C$ is convex.

F. Description of Ellipsoid Algorithm

For differential functions, the ellipsoid algorithm works by taking the gradient of the function $g(\boldsymbol{\lambda})$ and thereby eliminating portions of the domain that cannot contain the optimal solution. By the definition of convexity, any differential convex function $f(x)$ satisfies

$$f(x) \geq f(x_0) + \nabla f(x_0)(x - x_0)^T \tag{82}$$

where x and x_0 represent vectors here. Therefore, for any x satisfying $\nabla f(x_0)(x - x_0)^T \geq 0$ we have $f(x) \geq f(x_0)$. Given some point x_0 in the domain of f, by evaluating $\nabla f(x_0)$ we can eliminate a halfspace in our search for the minimizing x^*. The ellipsoid algorithm is based on the idea of continually eliminating a halfspace of the domain. If $f(x)$ is nondifferentiable, however, we cannot evaluate the gradient of the function. We can, however, eliminate a halfspace by finding a *subgradient* of the function f. The vector $\boldsymbol{s}$ is a subgradient of f at x if

$$f(y) \geq f(x) + \boldsymbol{s}(y - x)^T, \qquad \forall\, y. \tag{83}$$

Fortunately, continuous convex functions are subdifferentiable at all points in the domain of f. As long as we are able to find a subgradient at any point in the domain of our function, we are able to use the ellipsoid algorithm for nondifferentiable functions.

Since $g(\boldsymbol{\lambda})$ is the supremum of the Lagrangian $L(\boldsymbol{P}(\boldsymbol{h}), w(\boldsymbol{h}), \boldsymbol{\lambda})$, which is an affine function of $\boldsymbol{\lambda}$, we are easily able to find a subgradient for $g(\boldsymbol{\lambda})$. For any given $\boldsymbol{\lambda_0}$, let us denote the corresponding optimum probability of transmission function and the power allocation function by $w^*(\boldsymbol{h}, \boldsymbol{\lambda_0})$ and $\boldsymbol{P}^*(\boldsymbol{h}, \boldsymbol{\lambda_0})$, respectively. Here we write them as explicit functions of $\boldsymbol{\lambda_0}$ because they depend on the Lagrangian multipliers. Consider the derivative of $L'(\boldsymbol{P}(\boldsymbol{h}), w(\boldsymbol{h}), \boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda}$

$$\frac{\partial}{\partial\boldsymbol{\lambda}}L'(\boldsymbol{P}(\boldsymbol{h}), w(\boldsymbol{h}), \boldsymbol{\lambda}) = -(E_{\boldsymbol{h}}[w(\boldsymbol{h})\boldsymbol{P}(\boldsymbol{h})] - \overline{\boldsymbol{P}}). \tag{84}$$

Due to the affine nature of L', this implies that for all $\boldsymbol{\lambda}$ satisfying

$$(E_{\boldsymbol{h}}[w^*(\boldsymbol{h}, \boldsymbol{\lambda_0})\boldsymbol{P}^*(\boldsymbol{h}, \boldsymbol{\lambda_0})] - \overline{\boldsymbol{P}})(\boldsymbol{\lambda} - \boldsymbol{\lambda_0})^T \leq 0$$

we have

$$L'(\boldsymbol{P}^*(\boldsymbol{h}, \boldsymbol{\lambda_0}), w^*(\boldsymbol{h}, \boldsymbol{\lambda_0}), \boldsymbol{\lambda}) \geq L'(\boldsymbol{P}^*(\boldsymbol{h}, \boldsymbol{\lambda_0}), w^*(\boldsymbol{h}, \boldsymbol{\lambda_0}), \boldsymbol{\lambda_0}).$$

Since $g(\boldsymbol{\lambda})$ is defined as the supremum of the Lagrangian, this implies $g(\boldsymbol{\lambda}) \geq g(\boldsymbol{\lambda_0})$. Thus,

$$\boldsymbol{s} = \overline{\boldsymbol{P}} - E_{\boldsymbol{h}}[w^*(\boldsymbol{h}, \boldsymbol{\lambda_0})\boldsymbol{P}^*(\boldsymbol{h}, \boldsymbol{\lambda_0})]$$

is a subgradient of $g(\boldsymbol{\lambda})$ at $\boldsymbol{\lambda_0}$, which allows us to then eliminate a halfspace of the domain of $\boldsymbol{\lambda}$, further narrowing our search for the optimal Lagrangian multipliers. Notice that the subgradient is equal to the difference between the average power constraint vector and the average transmit power vector corresponding to the optimal power policy with the given power price vector.

$$\boldsymbol{P}^{\text{rand}}(\boldsymbol{h}) = \begin{cases} \boldsymbol{P}^{(1)}(\boldsymbol{h}) & \text{with prob. } \alpha w^{(1)}(\boldsymbol{h}) \\ \boldsymbol{P}^{(2)}(\boldsymbol{h}) & \text{with prob. } (1-\alpha)w^{(2)}(\boldsymbol{h}) \\ 0 & \text{with prob. } 1 - \alpha w^{(1)}(\boldsymbol{h}) - (1-\alpha)w^{(2)}(\boldsymbol{h}). \end{cases} \tag{81}$$

There are, in fact, a number of different algorithms to minimize nondifferentiable convex functions [15], [17]. Here we propose the ellipsoid algorithm to find the vector of optimal Lagrangian multipliers. The ellipsoid algorithm is not necessarily the fastest algorithm, but it is simple to use and describe. Readers interested in finding algorithms with better convergence properties can see [15], [17].

In the ellipsoid algorithm, we first find an ellipsoid in which the optimal Lagrangian multiplier vector $\boldsymbol{\lambda}^*$ *must* lie. We then take the subgradient at the center of this feasible ellipsoid to rule out a halfspace. We then consider the minimum volume ellipsoid covering the previous feasible ellipsoid intersected with the feasible halfspace (determined by the subgradient), and find a subgradient at the center of the new ellipsoid. If the center of the feasible ellipsoid is not nonnegative (i.e., some component of $\boldsymbol{\lambda}$ is negative), then it is easy to show that the vector $\boldsymbol{s}$ with $s_i = -1$ for all i such that $\lambda_i < 0$ and $s_i = 0$ for all other components is a subgradient.

Before describing the ellipsoid algorithm in detail, we first find a polyhedron in which $\boldsymbol{\lambda}^*$ must lie by individually bounding λ_i for each user i. We do this by finding a subgradient for $\boldsymbol{\lambda} = (0, \dots, c, \dots, 0)$ (i.e., $\lambda_i = c$ and $\lambda_j = 0, \forall\, j \neq i$) where c is some positive constant. Consider the optimal $w^*(\boldsymbol{h}, \boldsymbol{\lambda})$ and $\boldsymbol{P}^*(\boldsymbol{h}, \boldsymbol{\lambda})$ for this $\boldsymbol{\lambda}$. Due to the structure of $\boldsymbol{\lambda}$ and the optimum decoding order described in (34), User i is decoded *last* in all fading states. Therefore User i experiences no interference and

$$P_i^*(\boldsymbol{h}, \boldsymbol{\lambda}) = \frac{\sigma^2}{h_i}(e^{2R_i} - 1), \qquad \forall\, \boldsymbol{h} \in \mathcal{H}_{\text{all}}.$$

In addition, $w^*(\boldsymbol{h}, \boldsymbol{\lambda}) = 1$ only in states for which $cP_i^*(\boldsymbol{h}, \boldsymbol{\lambda}) < 1$. Therefore, if we let $c \to \infty$, we see that $w^*(\boldsymbol{h}, \boldsymbol{\lambda})$ will be zero in almost all fading states, and for sufficiently large c we have $E_{\boldsymbol{h}}[w^*(\boldsymbol{h}, \boldsymbol{\lambda})P_i^*(\boldsymbol{h}, \boldsymbol{\lambda})] < \overline{P}_i$. This implies that

$$\boldsymbol{s} = (0, \dots, \overline{P}_i - E_{\boldsymbol{h}}[w^*(\boldsymbol{h}, \boldsymbol{\lambda})P_i^*(\boldsymbol{h}, \boldsymbol{\lambda})], \dots, 0)$$

is a subgradient of $g(\boldsymbol{\lambda})$. Thus, $g(\boldsymbol{\lambda})$ can only increase if λ_i is increased beyond c, which implies $\lambda_i^* \leq c$. By repeating this procedure for all M users, we will find that $\mathbf{0} \leq \boldsymbol{\lambda}^* \leq \boldsymbol{c}$, where $\boldsymbol{c}$ is a positive vector.

Next we briefly describe the ellipsoid algorithm [15]. We define an ellipsoid $\mathcal{B}$ by

$$\mathcal{B}(x, A) = \{\boldsymbol{\lambda} : (\boldsymbol{\lambda} - x)A^{-1}(\boldsymbol{\lambda} - x)^T \leq 1\}$$

for an $M \times M$ positive-definite matrix A and an M-dimensional vector x (which is the center of the ellipse). Given the minimum volume ellipsoid covering the initial polyhedron $\mathbf{0} \leq \boldsymbol{\lambda} \leq \boldsymbol{c}$ and defining $\boldsymbol{\lambda}^{(0)}$ as the center of this ellipsoid, the algorithm is as follows:

1. Find a subgradient at $\boldsymbol{\lambda}^{(k)}$:
$$\boldsymbol{s} = \overline{\boldsymbol{P}} - E_{\boldsymbol{h}}[w^*(\boldsymbol{h}, \boldsymbol{\lambda}^{(k)})\boldsymbol{P}^*(\boldsymbol{h}, \boldsymbol{\lambda}^{(k)})]$$
where $\boldsymbol{P}^*(\boldsymbol{h}, \boldsymbol{\lambda}^{(k)})$ and $w^*(\boldsymbol{h}, \boldsymbol{\lambda}^{(k)})$ are given by (33) and (36), respectively.

2. Find the minimum volume ellipsoid covering $\mathcal{B}(\boldsymbol{\lambda}^{(k)}, A^{(k)}) \cap \{\boldsymbol{\lambda} : \boldsymbol{s}(\boldsymbol{\lambda} - \boldsymbol{\lambda}^{(k)})^T \leq 0\}$:
$$\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} - \frac{1}{M+1}\tilde{\boldsymbol{s}}\left(A^{(k)}\right)^T$$
$$A^{(k+1)} = \frac{M^2}{M^2 - 1}\left(A^{(k)} - \frac{2}{M+1}A^{(k)}\tilde{\boldsymbol{s}}^T\tilde{\boldsymbol{s}}A^{(k)}\right)$$
where
$$\tilde{\boldsymbol{s}} = \frac{\boldsymbol{s}}{\sqrt{\boldsymbol{s}A^{(k+1)}\boldsymbol{s}^T}}.$$

3. Increase k by 1 and return to Step 1.

This method can be repeated indefinitely until the desired accuracy is reached. At each step, the volume of the feasible ellipsoid is reduced by at least the factor $e^{-1/(2M)}$, which implies that the volume of the feasible ellipsoid goes to zero. Together with a Lipschitz condition on the objective function (which is implied by the convexity of $g(\boldsymbol{\lambda})$ [18]), convergence to the optimal $\boldsymbol{\lambda}^*$ can be shown. Details of the algorithm and convergence proofs are available in [15, Lecture 14].

G. Extreme Points of Average Power Region for Common Outage

In this section, we describe how to find the extreme points of $APV_{\text{out}}(Pr^*, \boldsymbol{R})$. As mentioned in Section V-E, we consider more general decoding orders by giving absolute decoding order priority (i.e., priority in *all* fading states) to certain groups of users. Specifically, let $\mathcal{L}$ denote a partition of $G \triangleq \{1, 2, \dots, M\}$ into N subsets $(G_1, \dots, G_N)$, which implies $G = \bigcup_{i=1}^N G_i$ and $G_i \cap G_j = \varnothing$ for $i \neq j$. In every fading state $\boldsymbol{h}$, users in the subset G_N are decoded first, followed by the users in G_{N-1}, up to the final subset G_1. The successive decoding order within each subset G_i is determined by $(\lambda_k)_{k \in G_i}$ as before (according to (34)). With this method, the users in set G_1 experience no interference from users in the other groups. However, the users in set G_2 are decoded before the users in G_1 and, thus, the users in G_1 provide interference to them. Similarly, the users in set G_3 see the users in G_1 and G_2 as interference. Interestingly, if powers are allocated according to (33), the total interference created by the users in any set G_i is independent of the decoding order used within that subset (i.e., $\sum_{j \in G_i} h_j P_j(\boldsymbol{h}) = \sigma^2 e^{2\sum_{j \in G_i} R_j}$ from (33)). Therefore, $(\lambda_k)_{k \in G_i}$ is used to determine the decoding order of users in G_i. If $\mathcal{L} = \{G\}$, then all the M users are of the same priority and their decoding order is then determined by the power price vector $\boldsymbol{\lambda}$ completely. Therefore, a decoding order determined by $(\boldsymbol{\lambda}, \mathcal{L})$ with $\mathcal{L} \neq \{G\}$ is different from any of the orders determined by $(\boldsymbol{\lambda}, \mathcal{L})$ with $\mathcal{L} = \{G\}$, and $(\boldsymbol{\lambda}, \mathcal{L})$ can be viewed as the *power allocation parameter pair* that describes all possible decoding orders. For every possible partition $\mathcal{L}$, a different surface is parameterized by the power price vector $\boldsymbol{\lambda}$. Thus, by considering every power price vector (with sum equal to one) for every partition, we can fully characterize the extreme points of $APV_{\text{out}}(Pr^*, \boldsymbol{R})$.

We now show how to find the extreme point corresponding to a given partition $\mathcal{L}$ and a given power price vector $\boldsymbol{\lambda}$. Note that the whole set of extreme points resulting from the partition $\mathcal{L}$ correspond to points which are on the boundary of the average power region (for the given common outage probability) of the users in subset G_1 *in the absence of all other users*. Given the partition $\mathcal{L}$ and the $\boldsymbol{\lambda}$, the problem is thus to minimize $\boldsymbol{\lambda}^{G_1} \cdot \boldsymbol{P}^{G_1}$, i.e., to minimize the weighted-sum power used by the users in G_1. This ensures that the average power used by the users in

G_1 will be on the boundary of the average power region of the users in G_1. Given that this has been accomplished, the secondary goal is to minimize $\boldsymbol{\lambda}^{G_2} \cdot \boldsymbol{P}^{G_2}$, followed by $\boldsymbol{\lambda}^{G_3} \cdot \boldsymbol{P}^{G_3}$, and so on. Since we are first concerned with the users in set G_1 and since the users of G_1 experience no interference from users of other subsets, we determine the outage states by only considering the users in G_1. If we did otherwise, the users in G_1 would not necessarily be able to achieve a point on the average power region of the users in G_1. Following the earlier arguments in this section, the transmit power is allocated to users in G_1 according to (85) at the bottom of the page, where $|G_1|$ denotes the number of users in set G_1 and the permutation $\pi_{G_1}(\cdot)$ of the $|G_1|$ users satisfies

$$\frac{\lambda_{\pi_{G_1}(1)}}{h_{\pi_{G_1}(1)}} \geq \frac{\lambda_{\pi_{G_1}(2)}}{h_{\pi_{G_1}(2)}} \geq \cdots \geq \frac{\lambda_{\pi_{G_1}(|G_1|)}}{h_{\pi_{G_1}(|G_1|)}}. \tag{86}$$

As before, the optimum choice of $w(\boldsymbol{h})$ is

$$w(\boldsymbol{h}) = \begin{cases} 1, & \boldsymbol{\lambda}^{(G_1)} \cdot \boldsymbol{P}^{(G_1)}(\boldsymbol{h}) < s^* \\ 0, & \boldsymbol{\lambda}^{(G_1)} \cdot \boldsymbol{P}^{(G_1)}(\boldsymbol{h}) > s^* \end{cases} \tag{87}$$

where the Lagrangian multiplier s^* is chosen such that $E_{\boldsymbol{h}}[w(\boldsymbol{h})] = 1 - Pr^*$. Since the outage states have been determined, all that is left to do is to allocate power to users in the lower priority groups. Since the users in group G_j ($\forall\, j = 2, \ldots, N$) will see users in groups $G_1, \ldots, G_{j-1}$ as interference, the optimum power allocation function for each user in G_j is as shown in (88) at the bottom of the page, where $|G_j|$ denotes the number of users in set G_j and the permutation $\pi_{G_j}(\cdot)$ of the $|G_j|$ users satisfies

$$\frac{\lambda_{\pi_{G_j}(1)}}{h_{\pi_{G_j}(1)}} \geq \frac{\lambda_{\pi_{G_j}(2)}}{h_{\pi_{G_j}(2)}} \geq \cdots \geq \frac{\lambda_{\pi_{G_j}(|G_j|)}}{h_{\pi_{G_j}(|G_j|)}}.$$

The corresponding average power vector is

$$\overline{\boldsymbol{P}} = E_{\boldsymbol{h}}[w(\boldsymbol{h})\boldsymbol{P}(\boldsymbol{h})] \tag{89}$$

where $\boldsymbol{P}(\boldsymbol{h})$ is defined in (85) and (88) and $w(\boldsymbol{h})$ is defined in (87).

H. Proof of Lemma 6.1

In the following, we prove that the set $\mathcal{Q}_I(\boldsymbol{R})$ defined in (27) is convex. This is a straightforward generalization of the proof in Part E of the Appendix. For a given rate vector $\boldsymbol{R}$, consider two points $(\boldsymbol{Pr}^{\boldsymbol{on}(1)}, \overline{\boldsymbol{P}}^{(1)})$ and $(\boldsymbol{Pr}^{\boldsymbol{on}(2)}, \overline{\boldsymbol{P}}^{(2)})$ in $\mathcal{Q}_I(\boldsymbol{R})$. By Proposition 3.3, there must exist power allocation functions $\boldsymbol{P}^{(1)}(\boldsymbol{h}, k)$ and $\boldsymbol{P}^{(2)}(\boldsymbol{h}, k)$ and probability of transmission functions $w^{(1)}(\boldsymbol{h}, k)$ and $w^{(2)}(\boldsymbol{h}, k)$ such that

$$\boldsymbol{Pr}^{\boldsymbol{on}(i)} = \sum_{k=1}^{2^M-1} E\left[w^{(i)}(\boldsymbol{h}, k)\boldsymbol{\Psi}(k)\right], \quad i = 1, 2 \tag{90}$$

$$\overline{\boldsymbol{P}}^{(i)} = E\left[\sum_{k=1}^{2^M-1} \boldsymbol{P}^{(i)}(\boldsymbol{h}, k) w^{(i)}(\boldsymbol{h}, k)\right], \quad i = 1, 2 \tag{91}$$

with $\boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{(1)}(\boldsymbol{h}, k))$ for all $\boldsymbol{h} : w^{(1)}(\boldsymbol{h}, k) > 0$ and $\boldsymbol{\Psi}(k)\boldsymbol{R} \in \mathcal{C}_{\mathrm{MAC}}(\boldsymbol{h}, \boldsymbol{P}^{(2)}(\boldsymbol{h}, k))$ for all $\boldsymbol{h} : w^{(2)}(\boldsymbol{h}, k) > 0$, $\forall\, k = 1, \ldots, 2^M - 1$.

We must show that

$$\left(\alpha\boldsymbol{Pr}^{\boldsymbol{on}(1)} + (1-\alpha)\boldsymbol{Pr}^{\boldsymbol{on}(2)}, \alpha\overline{\boldsymbol{P}}^{(1)} + (1-\alpha)\overline{\boldsymbol{P}}^{(2)}\right) \in \mathcal{Q}_I(\boldsymbol{R})$$

for any $\alpha \in [0, 1]$. In order to do so, we define a random power allocation function

$$\boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h}, k)\ (\forall\, \boldsymbol{h} \in \mathcal{H}_{\mathrm{all}}, \forall\, k = 1, \ldots, 2^M - 1)$$

as shown in (92) at the bottom of the page. Clearly, the resulting average transmission probability for each user i ($i = 1, \ldots, M$) is

$$\begin{aligned} &\alpha E_{\boldsymbol{h}}\left[\sum_{k=1}^{2^M-1} w^{(1)}(\boldsymbol{h}, k)\psi(k, i)\right] \\ &\quad + (1-\alpha)E_{\boldsymbol{h}}\left[\sum_{k=1}^{2^M-1} w^{(2)}(\boldsymbol{h}, k)\psi(k, i)\right] \\ &\qquad = \alpha Pr_i^{on(1)} + (1-\alpha)Pr_i^{on(2)} \end{aligned}$$

and the average transmit power vector is given by

$$\begin{aligned} E[\boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h})] &= E\left[\alpha \sum_{k=1}^{2^M-1} w^{(1)}(\boldsymbol{h}, k)\mathbf{P}^{(1)}(\boldsymbol{h}, k) \right. \\ &\qquad \left. + (1-\alpha) \sum_{k=1}^{2^M-1} w^{(2)}(\boldsymbol{h}, k)\mathbf{P}^{(2)}(\boldsymbol{h}, k)\right] \\ &= \alpha\overline{\boldsymbol{P}}^{(1)} + (1-\alpha)\overline{\boldsymbol{P}}^{(2)}. \end{aligned}$$

Therefore, we have

$$\left(\alpha\boldsymbol{Pr}^{\boldsymbol{on}(1)} + (1-\alpha)\boldsymbol{Pr}^{\boldsymbol{on}(2)}, \alpha\overline{\boldsymbol{P}}^{(1)} + (1-\alpha)\overline{\boldsymbol{P}}^{(2)}\right) \in \mathcal{Q}_I(\boldsymbol{R})$$

$$P_{\pi_{G_1}(i)}(\boldsymbol{h}) = \begin{cases} \frac{\sigma^2}{h_{\pi_{G_1}(1)}}\left[\exp(2R_{\pi_{G_1}(1)}) - 1\right], & \text{if } i = 1 \\ \frac{\sigma^2}{h_{\pi_{G_1}(i)}}\left[\exp\left(2\sum_{k=1}^{i} R_{\pi_{G_1}(k)}\right) - \exp\left(2\sum_{k=1}^{i-1} R_{\pi_{G_1}(k)}\right)\right], & \forall\, 2 \leq i \leq |G_1| \end{cases} \tag{85}$$

$$P_{\pi_{G_j}(i)}(\boldsymbol{h}) = \begin{cases} \frac{\sigma^2}{h_{\pi_{G_j}(1)}} \cdot \exp\left(2\sum_{k\in G_1,\ldots,G_{j-1}} R_k\right) \cdot \left[\exp(2R_{\pi_{G_j}(1)}) - 1\right], & \text{if } i = 1 \\ \frac{\sigma^2}{h_{\pi_{G_j}(i)}} \cdot \exp\left(2\sum_{k\in G_1,\ldots,G_{j-1}} R_k\right)\left[\exp\left(2\sum_{k=1}^{i} R_{\pi_{G_j}(k)}\right) - \exp\left(2\sum_{k=1}^{i-1} R_{\pi_{G_j}(k)}\right)\right], & \forall\, 2 \leq i \leq |G_j| \end{cases} \tag{88}$$

$$\boldsymbol{P}^{\mathrm{rand}}(\boldsymbol{h}) = \begin{cases} \boldsymbol{P}^{(1)}(\boldsymbol{h}, k) & \text{with prob. } \alpha w^{(1)}(\boldsymbol{h}, k) \\ \boldsymbol{P}^{(2)}(\boldsymbol{h}, k) & \text{with prob. } (1-\alpha) w^{(2)}(\boldsymbol{h}, k) \\ 0 & \text{with prob. } \alpha w^{(1)}(\boldsymbol{h}, 0) + (1-\alpha) w^{(2)}(\boldsymbol{h}, 0). \end{cases} \tag{92}$$

which implies that the set $\mathcal{Q}_I(\boldsymbol{R})$ is convex. Furthermore, this also implies that the individual usage probability region $\overline{\mathcal{O}}_I(\overline{\boldsymbol{P}}, \boldsymbol{R})$ is convex.

I. Extreme Points of Usage Probability Region for Individual Outage

In this section, we describe how to characterize the extreme points of the usage probability region for individual outage. As discussed in Section VI-D, we consider a partition of $G = \{1, \ldots, M\}$ and we successively decode users in set G_N first, followed by the users in G_{N-1}, and so on. Since users in G_i are given absolute priority over users in G_j for $i < j$ and users in G_1 are given the highest priority, for each given vector $\boldsymbol{\mu} \in \Re_+^M$, we can first derive the outage probability subvector $\boldsymbol{Pr}^{(G_1)}$ for users in set G_1 as if users in all other sets were nonexistent. This is necessary to ensure that the usage probabilities of the users in set G_1 are on the boundary surface of the usage probability region of the users in G_1. For $j \geq 2$, once the optimal power allocation for users in $\cup_{1 \leq i \leq j-1} G_i$ is determined, the outage probability subvector $\boldsymbol{Pr}^{(G_j)}$ for users in set G_j can be derived as if users in $\cup_{j+1 \leq i \leq N} G_i$ were nonexistent and the signals from users in $\cup_{1 \leq i \leq j-1} G_i$ were background noise. Note that subvector $\boldsymbol{Pr}^{(G_j)}$ $(1 \leq j \leq N)$ is a regular point on the boundary surface of the outage probability region $\mathcal{O}_I(\overline{\boldsymbol{P}}^{*(G_j)}, \boldsymbol{R}^{(G_j)})$ for the users in set G_j (treating signals from users in $\cup_{1 \leq i \leq j-1} G_i$ as background noise). Since we have already shown how to obtain a regular point on the boundary surface of region $\mathcal{O}_I(\overline{\boldsymbol{P}}^*, \boldsymbol{R})$ for an M-user system, the regular point $\boldsymbol{Pr}^{(G_j)}$ of region $\mathcal{O}_I(\overline{\boldsymbol{P}}^{*(G_j)}, \boldsymbol{R}^{(G_j)})$ for a $|G_j|$-user system can be similarly derived.

Therefore, for each given pair $(\boldsymbol{\mu}, \mathcal{L})$ with $\mathcal{L} \neq \{G\}$, we can obtain the corresponding boundary outage probability vector of region $\mathcal{O}_I(\overline{\boldsymbol{P}}^*, \boldsymbol{R})$. Such boundary vectors are the extreme points on the boundary surface of region $\mathcal{O}_I(\overline{\boldsymbol{P}}^*, \boldsymbol{R})$, and we can obtain all extreme points explicitly by varying $(\boldsymbol{\mu}, \mathcal{L})$ with $\boldsymbol{\mu} \in \Re_+^M$ and $\mathcal{L} \neq \{G\}$.

J. Extreme Points of Average Power Region for Individual Outage

In this section, we characterize the extreme points of the average power region with individual outage. As before, consider the case where subsets of users are given absolute priority over other subsets of users in every fading state. We let $\mathcal{L}$ represent a partition of $G = \{1, \ldots, M\}$ and we consider successively decoding users in set G_N first, followed by the users in G_{N-1}, and so on. Since users in G_i are given absolute priority over users in G_j for $i < j$ and users in G_1 are given the highest priority, for each given power price vector $\boldsymbol{\lambda} \in \Re_+^M$, we can first derive the average transmit power subvector $\overline{\boldsymbol{P}}^{(G_1)}$ for users in set G_1 as if users in all other sets were nonexistent. This ensures that the average power used by the users in G_1 will be on the boundary of the average power region of the users in G_1. For $j \geq 2$, once the optimal power allocation for users in $\cup_{1 \leq i \leq j-1} G_i$ is determined, the average transmit power subvector $\overline{\boldsymbol{P}}^{(G_j)}$ for users in set G_j can be derived as if users in $\cup_{j+1 \leq i \leq N} G_i$ were nonexistent and the signals from users in $\cup_{1 \leq i \leq j-1} G_i$ were background noise. Note that subvector $\overline{\boldsymbol{P}}^{(G_j)}$ $(1 \leq j \leq N)$ is a regular point on the boundary surface of the average power region $APV_{\text{out}}(\boldsymbol{Pr}^{*(G_j)}, \boldsymbol{R}^{(G_j)})$ for the users in set G_j (treating signals from users in $\cup_{1 \leq i \leq j-1} G_i$ as background noise). Since we have already shown how to obtain a regular point on the boundary surface of region $APV_{\text{out}}(\boldsymbol{Pr}^*, \boldsymbol{R})$ for an M-user system, the regular point $\overline{\boldsymbol{P}}^{(G_j)}$ of region $APV_{\text{out}}(\boldsymbol{Pr}^{*(G_j)}, \boldsymbol{R}^{(G_j)})$ for a $|G_j|$-user system can be similarly derived.

Therefore, for each given pair $(\boldsymbol{\lambda}, \mathcal{L})$ with $\mathcal{L} \neq \{G\}$, we can obtain the corresponding boundary average power vector. Such boundary vectors are the extreme points on the boundary surface of $APV_{\text{out}}(\boldsymbol{Pr}^*, \boldsymbol{R})$, and we can obtain all extreme points explicitly by varying $(\boldsymbol{\lambda}, \mathcal{L})$ with $\boldsymbol{\lambda} \in \Re_+^M$ and $\mathcal{L} \neq \{G\}$.

ACKNOWLEDGMENT

The authors wish to thank the Associate Editor, Shlomo Shamai, and the anonymous reviewers for their very helpful suggestions and comments, which greatly improved the quality of the manuscript.

REFERENCES

[1] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[2] D. N.C. Tse and S. V. Hanly, "Multiple-access fading channels: Part I: Polymatroidal structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.

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

[4] W. Yu, W. Rhee, and J. Cioffi, "Optimal power control in multiple access fading channels with multiple antennas," in *Proc. Int. Conf. Communications (ICC'01)*, vol. 2, Helsinki, Finland, Jun. 2001, pp. 575–579.

[5] W. Yu and W. Rhee, "Degrees of freedom in multi-user spatial multiplex systems with multiple antennas," *IEEE Trans. Commun.*, submitted for publication.

[6] S. V. Hanly and D. N. C. Tse, "Multiple-access fading channels: Part II: Delay-limited capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2816–2831, Nov. 1998.

[7] L. Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels: Part II: Outage capacity," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1103–1127, Mar. 2001.

[8] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1468–1489, Jul. 1999.

[9] J. Luo, L. Lin, R. Yates, and P. Spasojevic, "Service outage based power and rate allocation," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 323–330, Jan. 2003.

[10] J. Luo, R. Yates, and P. Spasojevic, "Service outage based power and rate allocation for parallel fading channels," in *Proc. IEEE Global Commuications. Conf. (GLOBECOM'03)*, vol. 2, San Francisco, CA, Dec. 2003, pp. 1003–1007.

[11] N. Jindal and A. Goldsmith, "Capacity and optimal power allocation for fading broadcast channels with minimum rates," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, pp. 2895–2909, Nov. 2003.

[12] C. Ng and A. Goldsmith, "Capacity of fading broadcast channels with rate constraints," in *Proc. 42nd Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2004.

[13] (2004, Oct.) *NIST/SEMATECH e-Handbook of Statistical Methods*. [Online] Available: http://www.itl.nist.gov/div898/handbook/eda/section3/eda361.htm

[14] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

[15] S. Boyd and L. Vandenberghe. (2001) Introduction to Convex Optimization With Engineering Applications. Course Reader. [Online]. Available: http://www.stanford.edu/class/ee364/

[16] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 768–783, May 2004.

[17] D. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.

[18] R. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.