Service Outage Based Power and Rate Allocation for Parallel Fading Channels

Jianghong Luo, *Member, IEEE,* Roy Yates, *Member, IEEE,* and Predrag Spasojević, *Member, IEEE*

***Abstract*— The service outage based allocation problem explores variable rate transmission schemes and combines the concepts of ergodic capacity and outage capacity for fading channels. A service outage occurs when the transmission rate is below a given basic rate r_0. The allocation problem is to maximize the expected rate subject to the average power constraint and the constraint that the outage probability is less than ϵ. A general class of probabilistic power allocation schemes is considered for an M-parallel fading channel model. The optimum power allocation scheme is derived and shown to be deterministic except at channel states of a boundary set. The resulting service outage achievable rate ranges from $1-\epsilon$ of the outage capacity up to the ergodic capacity with increasing average power. Two near optimum schemes are also derived by exploiting the fact that the outage probability is usually small. The second near optimum scheme significantly reduces the computational complexity of the optimum solution; moreover, it has a simple structure for the implementation of transmission of mixed real-time and non real-time services.**

***Index Terms*— Service outage, power allocation, fading channel, adaptive transmission, ergodic capacity, outage capacity**

I. Introduction

Wireless communication channels vary with time due to multipath, mobility of users, and changes in the environment. For a time varying channel, dynamic allocation of resources such as power and rate can yield improved performance over fixed allocation strategies. Since system performance criteria are usually application specific, different classes of applications will result in different adaptive transmission schemes. In order to differentiate real-time service from non real-time service, three capacity measures have been defined in the literature: ergodic capacity [7], delay limited capacity [8], and capacity versus outage [4], [17]. A comprehensive survey of these concepts can be found in [2].

The ergodic capacity [7] determines the maximum achievable rate without a delay constraint. The corresponding optimum power allocation is the well known water-filling allocation [5], [6]. In a fast fading environment, the ergodic capacity can be achieved by using a constant-rate variable-power long code that experiences the ergodicity of the fading process [4]. But in the presence of slow fading and a decoding delay constraint, ergodic capacity is usually achieved by multiplexing variable-rate variable-power short codes [4]. Thus, in deep fades the code rate could be very low or even equal to zero, which could be detrimental to real-time applications. To address this issue, the notion of capacity versus outage was developed for constant-rate real-time applications for a block fading channel model [4], [17]. In the capacity versus outage problem, a constant target rate aims to be maintained throughout the fading process. An information outage occurs when the target rate is higher than the instantaneous mutual information [4]. The maximum target rate that can be achieved with an outage probability less than ϵ is called the outage capacity, and the zero outage capacity is called the delay-limited capacity in [4].

We observe that for some variable rate real time applications, neither the ergodic capacity nor the outage capacity is appropriate. For example, for applications with simultaneous voice and data transmissions, as soon as a basic rate r_0 for the voice service has been guaranteed, any excess rate can be used to transmit data in a best effort fashion. For some video or audio applications, the source rate can be adapted according to the fading channel conditions to provide multiple quality of service levels. Typically, a nonzero basic rate r_0 is required to achieve a minimum acceptable service quality. For these applications, maximizing the long term average rate while meeting a basic rate requirement for the instantaneous rate allocation is a desirable property. However, neither the ergodic capacity nor the outage capacity can achieve this goal, since the ergodic capacity offers no guarantee on the instantaneous rate while the outage capacity achieves a low long term average rate. Therefore, in this paper we combine the notion of ergodic capacity and outage capacity, and formulate the service outage based allocation problem, which maximizes the long term average rate subject to basic rate and average power constraints.

In a Rayleigh fading channel, infinite average power is needed to achieve any nonzero rate at all times. Hence, we impose the basic service rate requirement in a probabilistic way to obviate a need for infinite average power. The service is said to be in an outage when the instantaneous rate is smaller than the basic service rate r_0. A service outage constraint dictates that the probability of a service outage be less than ϵ, a parameter indicating the outage tolerance of the application. Unlike the information outage in the capacity versus outage problem [4], [15], the bits transmitted during the service outage may still be valuable in that they will be transmitted reliably and will contribute to the average rate. Related to service outage [13], [14], a minimum rate requirement was recently proposed in [10] for the fading broadcast channel.

The service outage based allocation problem was previously studied for the single fading channel in [13], [14]. In this paper, we generalize the results to the M-parallel flat fading channel. The M-parallel flat fading channel model can characterize a variety of systems, including an OFDM system with frequency selective fading and the multiple antenna signal model when the perfect channel state information is available at transmitter

and singular value decomposition is employed. In practical systems, the service outage based allocation problem is more relevant for the M-parallel channel model, since multimedia applications usually require large bandwidth or spatial dimension.

We will see in this work that solving the allocation problem in the M-parallel fading channel is quite difficult. In the single channel scenario [14], it is intuitive that the outage should happen for a set of worst fraction fading states. However, in the case of parallel channels, the channel state is specified by a vector and there is no absolute ordering of these state vectors. Unlike the single channel case in [13], [14], we show in Section V-B that choosing the outage set to be the set of channels that consume the most power to achieve the basic rate is suboptimal. Therefore, we cannot simply extend the approach in [13], [14] to the M-parallel fading channel.

In order to solve the allocation problem for the M-parallel fading channel, we formulate the problem in the general class of probabilistic power allocation schemes. Probabilistic power allocation schemes were previously studied in [4]. In this paper, the optimum power allocation scheme is derived using the generalized Karush-Kuhn-Tucker conditions in vector space [11]. The computation of the parameters in the optimum solution is complicated, which motivates us to find some simpler near optimum schemes. Two near optimum power allocation schemes are derived by exploiting the fact that the outage probability is usually a small value.

The remainder of this paper is organized as follows. In Section II, the channel model and the service outage based allocation problem are presented. The generalized Karush-Kuhn-Tucker conditions for functional optimization problems are reviewed in Section III. The optimum power allocation is derived in Section IV. Two near optimum power allocation policies are derived in Section V. Numerical results are given in Section VI and conclusions are drawn in Section VII. All proofs in this paper are provided in Appendix A.

II. System Model and Allocation Problem

In an M-parallel flat fading channel model, each fading block consists of M subchannels as follows

$$y_i = \sqrt{h_i} x_i + n_i \quad i = 1, 2, \ldots, M\,. \tag{1}$$

For a subchannel i, x_i is the channel input, y_i is the channel output, and h_i is the channel state. The noise components $n_1, \ldots, n_M$ are independent Gaussian random variables with normalized unit variance. It is assumed that the channel state vector $\mathbf{h} = (h_1, \ldots, h_M)$ stays the same within one fading block but may vary from block to block. For a typical slow fading environment, it is also assumed that block length $N \to \infty$ so that the information theoretic results can be applied. One codeword spans M subchannels in one fading block and perfect channel state information is available at both the transmitter and the receiver. The vector fading process is ergodic within the communication session.

Throughout this paper, we use the following notation:

- For a vector of channel states $\mathbf{h} = (h_1, \ldots, h_M)$, the power allocation vector is $\mathbf{p}(\mathbf{h}) = (p_1(\mathbf{h}), \ldots, p_M(\mathbf{h}))$. Here $p_i(\mathbf{h})$, the power allocated to subchannel i, depends on the current channel state vector $\mathbf{h}$.
- Given a vector $\mathbf{a}$ of length M, we denote its arithmetic mean by $\langle \mathbf{a} \rangle = M^{-1} \sum_{i=1}^{M} a_i$.
- The maximum mutual information of an M-parallel Gaussian channel $\mathbf{h}$ with power allocation $\mathbf{p}(\mathbf{h})$ is

$$r(\mathbf{h}, \mathbf{p}(\mathbf{h})) = \frac{1}{M} \sum_{i=1}^{M} \log(1 + h_i p_i(\mathbf{h})). \tag{2}$$

  To simplify the derivations, we use the natural logarithm and drop the usual factor $1/2$ in the Gaussian capacity expression. The rate unit is nats/subchannel. In this paper, the rate is averaged over parallel channels, since in practical systems the expected rate as well as the basic rate usually scales with the number of dimensions (bandwidth or number of antennas).
- For a scalar x, $[x]^+ = \max(x, 0)$. For a vector $\mathbf{x} = (x_1, \ldots, x_M)$, $[\mathbf{x}]^+ = ([x_1]^+, \ldots, [x_M]^+)$.
- The indicator function $1(x)$ is equal to 1 if x is true and is equal to 0 otherwise.
- For two vectors $\mathbf{a}$ and $\mathbf{b}$ of length M, we write $\mathbf{a} \geq \mathbf{b}$ if $a_i \geq b_i$ for all $i = 1, \ldots, M$.

It can be seen that with perfect channel state information at the transmitter and receiver, the maximum achievable rate of a given power vector $\mathbf{p}(\mathbf{h})$ at fading block $\mathbf{h}$ is given by (2). Thus, we only need to identify the optimum power allocation scheme.

Although we could formulate the allocation problem for M-parallel fading channels in the class of deterministic schemes, as we did for $M = 1$ fading channel in [14], the deterministic allocation problem turns out to be difficult to solve. Moreover, as shown in [4], the optimum allocation for the outage capacity is a probabilistic policy for discrete channel distribution, suggesting that deterministic schemes are likely to be suboptimal. Therefore, this paper formulates the allocation problem using the more general class of probabilistic schemes.

The probabilistic power allocation is a vector of random variables with a conditional pdf $f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})$. The physical interpretation of a probabilistic scheme is that the probability of using a particular power vector is the time-sharing factor for that power vector. In a probabilistic power allocation scheme, each realization of the power allocation is associated with a coding scheme. A service outage occurs when the code rate is less than the basic rate r_{o} specified by the application. Since multiple codes are employed in a probabilistic manner, at each channel state we can have the situation where some code rates are less than r_{o} while others are greater or equal to r_{o}. Thus, at each channel state, a service outage occurs with some probability. In order to simplify the derivation, we use $\mathbf{P}(\mathbf{h})$ to indicate a probabilistic power allocation scheme with conditional PDF $f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})$, while using $\mathbf{p}(\mathbf{h})$ to indicate a deterministic scheme. Due to the assumptions of ergodicity and perfect channel state information, the power allocation only depends on the current channel state vector. We use $F(\mathbf{h})$ to represent the cdf of channel state vector $\mathbf{h}$.

For a given probabilistic power allocation $\mathbf{P}(\mathbf{h})$, the average

rate, average power, and outage probability are given by

$$\begin{aligned}
&\mathrm{E}\{r(\mathbf{h},\mathbf{P}(\mathbf{h}))\} = \int\int r(\mathbf{h},\mathbf{p}) f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})\, d\mathbf{p}\, dF(\mathbf{h}) \\
&\mathrm{E}\{\langle \mathbf{P}(\mathbf{h})\rangle\} = \int\int \langle \mathbf{p}\rangle f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})\, d\mathbf{p}\, dF(\mathbf{h}) \\
&\Pr\{r(\mathbf{h},\mathbf{P}(\mathbf{h})) < r_{\mathrm{o}}\} \\
&\quad = \int\int 1(r(\mathbf{h},\mathbf{p}) < r_{\mathrm{o}}) f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})\, d\mathbf{p}\, dF(\mathbf{h}) . \qquad (3)
\end{aligned}$$

The service outage based allocation problem is to identify the optimum conditional PDF $f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})$ as follows:

$$\begin{aligned}
R^* = \max_{f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})} \ & \mathrm{E}\{r(\mathbf{h},\mathbf{P}(\mathbf{h}))\} && (4)\\
\text{subject to} \quad & \mathrm{E}\{\langle \mathbf{P}(\mathbf{h})\rangle\} \le p_{\mathrm{av}} && (4a)\\
& \Pr\{r(\mathbf{h},\mathbf{P}(\mathbf{h})) < r_{\mathrm{o}}\} \le \epsilon , && (4b)
\end{aligned}$$

where the conditional PDF $f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})$ is a set of functions for each $\mathbf{h}$ satisfying

$$\int f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h}) d\,\mathbf{p} = 1, \qquad f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h}) \ge 0 \quad \text{for all } \mathbf{h} . \qquad (5)$$

The resulting maximum average rate R^* is called the service outage achievable rate. This work can be extended to other rate expressions besides the Shannon capacity in (2), which may depend on decoding error probability, and the set of modulation and coding schemes in a practical system. Problem (4) may seem to be more complicated than the corresponding deterministic allocation problem, but in fact it will be easier to solve. In later sections, we will see that this problem can be simplified and solved using generalized Karush-Kuhn-Tucker conditions [11].

III. Functional Optimization

In this section, we briefly review the Karush-Kuhn-Tucker conditions for functional optimization, since in this work the optimization variables are functions instead of vectors in an Euclidean space. Readers are referred to texts [9], [11] for comprehensive results on optimization theory in a general vector space and [1] on optimization theory in an Euclidean space.

Specifically, we are interested in the following type of functional optimization problem in an Lebesgue L_p space with measure m:

$$\begin{aligned}
\min_{x(t)} \ & \int y_1(t,x(t)) dm(t) && (6)\\
\text{subject to} \quad & \int y_2(t,x(t)) dm(t) \le 0 && (6a)\\
& \int y_3(t,x(t)) dm(t) = 0 && (6b)\\
& y_4(t,x(t)) \le 0 && (6c)\\
& a \le x(t) \le b && (6d)
\end{aligned}$$

where functions $x(t)$ and $y_i(t,x(t))$ belong to the L_p space with measure m. The L_p space consists of those real-valued measurable functions x for which $\int |x(t)|^p dm(t)$ is finite [11], [16]. It is shown that the Lagrange multiplier associated with constraint (6a) and (6b) are scalers denoted as u and λ, while the Lagrange multiplier associated with constraint (6c) is a function $v(t) \in L_q$ where $1/p + 1/q = 1$ [3], [11]. Usually, no Lagrange multipliers are employed for simple constraints such as (6d), instead it is absorbed in the Karush-Kuhn-Tucker conditions as shown below. Let

$$\begin{aligned}
&l(x(t),u,\lambda,v(t)) = \\
&y_1(t,x(t)) + u y_2(t,x(t)) + \lambda y_3(t,x(t)) + v(t) y_4(t,x(t)) .
\end{aligned}$$

The Lagrangian of problem (6) is $L(x(t),u,\lambda,v(t)) = \int l(x(t),u,\lambda,v(t)) dm(t)$. The variation of $L(x(t),u,\lambda,v(t))$ with respect to $x(t)$ is equal to 0 iff the derivative of $l(x,u,\lambda,v(t))$ with respect to x at $x = x(t)$ is equal to zero. Thus, according to the generalized Karush-Kuhn-Tucker necessary conditions theorem [11], if the optimum solution $x^*(t)$ is a regular point (constraint qualification), it must satisfy the following conditions:

$$\left.\frac{dl(x,u,\lambda,v(t))}{dx}\right|_{x=x^*(t)} \begin{cases} = 0 & a < x^*(t) < b \\ \ge 0 & x^*(t) = a \\ \le 0 & x^*(t) = b \end{cases} \qquad (7)$$

$$u \int y_2(t,x^*(t)) dm(t) = 0 \qquad (8)$$

$$v(t) y_4(t,x^*(t)) = 0 \qquad (9)$$

$$u \ge 0, \quad v(t) \ge 0 \qquad (10)$$

In addition, $x^*(t)$ must also satisfy the constraints (6a), (6b), and (6c). Note that (7) incorporates the constraint (6d), and that (9) follows from $\int v(t) y_4(t,x^*(t)) dm(t) = 0$ due to the fact that $y_4(t,x(t)) \le 0$ for all t and $v(t) \ge 0$.

Furthermore, if $y_1(t,x(t))$, $y_2(t,x(t))$, and $y_4(t,x(t))$ are convex functionals with respect to $x(t)$, and $y_3(t,x(t))$ is a linear functional with respect to $x(t)$, conditions (7)-(10) and constraints (6a)-(6d) are sufficient for the global optimum solution of (6).

A similar approach can be applied to the more general case where $y_i(t,x(t))$ is replaced by $y_i(t,x_1(t),\ldots,x_n(t))$. In this case, we just replace the derivation with respect to $x(t)$ in (7) with the partial derivatives with respect to $x_j(t)$ for all $j = 1,\ldots,n$.

IV. The Optimum Service Outage Based Allocation

A. Allocations for an M-parallel Fading Channel

In this section, we introduce two deterministic power allocation schemes: the multi-dimensional water-filling allocation $\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0)$ and the basic-rate power allocation $\mathbf{p}_{\mathrm{r_0}}(\mathbf{h})$. These two allocations will be used to characterize the optimum solution in later sections.

The multi-dimensional water-filling allocation is the optimum allocation achieving the ergodic capacity in M-parallel fading channels as

$$\begin{aligned}
\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0) = \arg\max_{\mathbf{p}(\mathbf{h})} \ & \mathrm{E}\{r(\mathbf{h},\mathbf{p}(\mathbf{h}))\} && (11)\\
\text{subject to} \quad & \mathrm{E}\{\langle \mathbf{p}(\mathbf{h})\rangle\} \le p_{\mathrm{av}} && (11a)\\
& \mathbf{p}(\mathbf{h}) \ge 0 . && (11b)
\end{aligned}$$

We get $\mathbf{p}_{\rm wf}(\mathbf{h},h_0)=(p_{{\rm wf},1}(h_1,h_0),\ldots,p_{{\rm wf},M}(h_M,h_0))$, by applying the generalized Karush-Kuhn-Tucker condition in vector spaces [11], where

$$p_{{\rm wf},i}(h_i,h_0)=\left[\frac{1}{h_0}-\frac{1}{h_i}\right]^+ \quad i=1,\ldots,M\,, \tag{12}$$

and the water-filling cutoff h_0 is the solution to $\mathrm{E}\{\langle\mathbf{p}_{\rm wf}(\mathbf{h},h_0)\rangle\}=p_{\rm av}$.

The basic-rate power allocation is the power allocation that requires the minimum average power to maintain a basic rate at each channel state, as follows:

$$\mathbf{p}_{{\rm r}_0}(\mathbf{h})=\arg\min_{\mathbf{p}(\mathbf{h})}\langle\mathbf{p}(\mathbf{h})\rangle \tag{13}$$

$$\text{subject to}\quad r(\mathbf{h},\mathbf{p}(\mathbf{h}))=r_{\rm o} \tag{13a}$$

$$\mathbf{p}(\mathbf{h})\geq 0\,. \tag{13b}$$

The solution to the above problem is given by lemma 1 in [4], and is summarized below. The basic-rate allocation is $\mathbf{p}_{{\rm r}_0}(\mathbf{h})=(p_{{\rm r}_0,1}(\mathbf{h}),\ldots,p_{{\rm r}_0,M}(\mathbf{h}))$ with

$$p_{{\rm r}_0,i}(\mathbf{h})=\left[\lambda(\mathbf{h})-\frac{1}{h_i}\right]^+ \quad i=1,\ldots,M\,. \tag{14}$$

For a given $\mathbf{h}$, the basic-rate allocation also allocates power in the form of water-filling among subchannels, but the $\lambda(\mathbf{h})$ changes with $\mathbf{h}$ to ensure $r(\mathbf{h},\mathbf{p}_{{\rm r}_0}(\mathbf{h}))=r_{\rm o}$. Let $\pi(i)$ be the permutation of index i such that $h_{\pi(1)}\geq h_{\pi(2)}\geq\ldots\geq h_{\pi(M)}$. The $\lambda(\mathbf{h})$ is given by

$$\lambda(\mathbf{h})=\left(\frac{e^{Mr_{\rm o}}}{\prod_{l=1}^{\mu}h_{\pi(l)}}\right)^{\frac{1}{\mu}}\,, \tag{15}$$

where μ is the unique integer in $\{1,\ldots,M\}$ such that $\lambda(\mathbf{h})\geq h_{\pi(l)}^{-1}$ for $l\leq\mu$ and $\lambda(\mathbf{h})<h_{\pi(l)}^{-1}$ for $l>\mu$ [4]. Parameter μ indicates number of sub-channels with non zero power allocation at $\mathbf{h}$. When $M=1$ the basic-rate allocation

$\mathbf{p}_{{\rm r}_0}(\mathbf{h})$ becomes channel inversion, and when $M\to\infty$ it converges to the water-filling allocation $\mathbf{p}_{\rm wf}(\mathbf{h},h_0)$. An example of $\mathbf{p}_{{\rm r}_0}(\mathbf{h})$ for $M=2$ fading channel can be found in [4].

Based on the observation that for a given $\mathbf{h}$ both $\mathbf{p}_{\rm wf}(\mathbf{h},h_0)$ and $\mathbf{p}_{{\rm r}_0}(\mathbf{h})$ are in the form of 'water-filling' but with different water levels, we have the following proposition.

Proposition 1 *We have*

(a) $r(\mathbf{h},\mathbf{p}_{\rm wf}(\mathbf{h},h_0))\geq r_o \overset{\text{iff}}{\Longleftrightarrow} h_0^{-1}\geq\lambda(\mathbf{h})\overset{\text{iff}}{\Longleftrightarrow}\mathbf{p}_{\rm wf}(\mathbf{h},h_0)\geq\mathbf{p}_{{\rm r}_0}(\mathbf{h})$.

(b) *For any* $\mathbf{h}$*, either* $\mathbf{p}_{\rm wf}(\mathbf{h},h_0)\geq\mathbf{p}_{{\rm r}_0}(\mathbf{h})$ *or* $\mathbf{p}_{\rm wf}(\mathbf{h},h_0)\leq\mathbf{p}_{{\rm r}_0}(\mathbf{h})$ *holds.*

B. Feasibility and Outage Capacity

The feasibility of problem (4) is directly related to outage capacity in [4]. Let $C_\epsilon(p_{\rm av})$ be the outage capacity for a given $p_{\rm av}$. The $C_\epsilon(p_{\rm av})$ is the maximum instantaneous rate which can be transmitted with an outage probability ϵ. Thus, for a given $p_{\rm av}$ and ϵ, Problem (4) is feasible iff $r_{\rm o}\leq C_\epsilon(p_{\rm av})$. For convenience of subsequent derivations, the feasibility condition is expressed in the following equivalent form

$$p_{\rm av}\geq P_{\min}(r_{\rm o},\epsilon)\,, \tag{16}$$

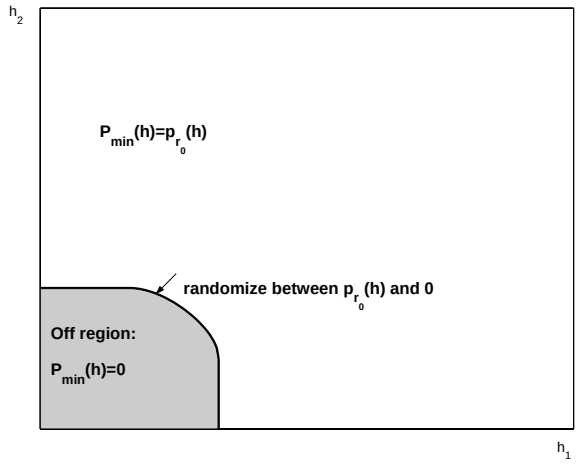

Fig. 1. The $\mathbf{P}_{\min}(\mathbf{h})$ in an $M=2$ parallel fading channel.

where the $P_{\min}(r_{\rm o},\epsilon)$ is the minimum average power needed to support $r_{\rm o}$ with an outage probability ϵ. When $p_{\rm av}=P_{\min}(r_{\rm o},\epsilon)$, we have $C_\epsilon(p_{\rm av})=r_{\rm o}$ and problem (4) shares the same optimum solution, denoted $\mathbf{P}_{\min}(\mathbf{h})$, with the outage capacity problem. For convenience of subsequent derivations, we rewrite Proposition 4 in [4] and express $\mathbf{P}_{\min}(\mathbf{h})$ as follows.

Definition 1 For any $\mathbf{h}$, let $X_w(\mathbf{h})$ be a Bernoulli $w(\mathbf{h})$ random variable: $X_w(\mathbf{h})=1$ with probability $w(\mathbf{h})$ and $X_w(\mathbf{h})=0$ with probability $1-w(\mathbf{h})$.

The minimum average power allocation is $\mathbf{P}_{\min}(\mathbf{h})=X_{w'}(\mathbf{h})\mathbf{p}_{{\rm r}_0}(\mathbf{h})$, where

$$w'(\mathbf{h})=\begin{cases}1 & \langle\mathbf{p}_{{\rm r}_0}(\mathbf{h})\rangle<s'\\ v' & \langle\mathbf{p}_{{\rm r}_0}(\mathbf{h})\rangle=s'\\ 0 & \langle\mathbf{p}_{{\rm r}_0}(\mathbf{h})\rangle>s'\end{cases}\,, \tag{17}$$

and the parameters s' and $0\leq v'\leq 1$ are solutions to $\mathrm{E}\{w'(\mathbf{h})\}=1-\epsilon$. That is

$$s'=\sup\{x:\Pr\{\langle\mathbf{p}_{{\rm r}_0}(\mathbf{h})\rangle<x\}<1-\epsilon\} \tag{18}$$

$$v'=\frac{1-\epsilon-\Pr\{\langle\mathbf{p}_{{\rm r}_0}(\mathbf{h})\rangle<s'\}}{\Pr\{\langle\mathbf{p}_{{\rm r}_0}(\mathbf{h})\rangle=s'\}}\,. \tag{19}$$

$\mathbf{P}_{\min}(\mathbf{h})$ is an on-off transmission policy. If the required sum power $\langle\mathbf{p}_{{\rm r}_0}(\mathbf{h})\rangle>s'$, transmission is turned off, while if $\langle\mathbf{p}_{{\rm r}_0}(\mathbf{h})\rangle\leq s'$, transmission is turned on and the power is allocated according to $\mathbf{p}_{{\rm r}_0}(\mathbf{h})$. $\mathbf{P}_{\min}(\mathbf{h})$ for an $M=2$ fading channel is plotted in Figure 1. The off region may or may not be a convex set depending on $r_{\rm o}$ and ϵ.

C. Derivation of the Optimum Allocation Scheme

In this section, we derive the optimum solution for the service outage based allocation problem (4). We first show that an optimum power allocation in (4) is a scheme which is randomized between two deterministic schemes.

Lemma 1 *There exists an optimum solution of problem (4) of the following form*

$$\mathbf{P}^*(\mathbf{h})=X_w(\mathbf{h})\mathbf{p}_{\rm a}(\mathbf{h})+(1-X_w(\mathbf{h}))\mathbf{p}_{\rm b}(\mathbf{h})\,, \tag{20}$$

where $r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h})) \geq r_o$ *for all* $\mathbf{h}$, $\mathrm{E}\{w(\mathbf{h})\} \geq 1-\epsilon$ *and* $\mathrm{E}\{\langle \mathbf{P}^*(\mathbf{h})\rangle\} = p_{\rm av}$.

Proof of Lemma 1 is based on the concavity of the rate function $r(\mathbf{h}, \mathbf{p})$.

By Lemma 1, we have $\mathbf{P}^*(\mathbf{h}) = \mathbf{p}_{\rm a}(\mathbf{h})$ with probability $w(\mathbf{h})$, and $\mathbf{P}^*(\mathbf{h}) = \mathbf{p}_{\rm b}(\mathbf{h})$ with probability $1-w(\mathbf{h})$. Moreover, the conditions $r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h})) \geq r_{\rm o}$ and $\mathrm{E}\{w(\mathbf{h})\} \geq 1-\epsilon$ ensure that the randomized scheme meets the service outage constraint. Thus, problem (4) can be simplified into a problem which requires identifying $\mathbf{p}_{\rm a}(\mathbf{h})$, $\mathbf{p}_{\rm b}(\mathbf{h})$, and $w(\mathbf{h})$ as follows:

$$\max_{\mathbf{p}_{\rm a}, \mathbf{p}_{\rm b}, w} \mathrm{E}\{w(\mathbf{h}) r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h})) + (1-w(\mathbf{h})) r(\mathbf{h}, \mathbf{p}_{\rm b}(\mathbf{h}))\} \tag{21}$$

$$\text{subject to} \quad \mathrm{E}\{w(\mathbf{h})\langle \mathbf{p}_{\rm a}(\mathbf{h})\rangle + (1-w(\mathbf{h}))\langle \mathbf{p}_{\rm b}(\mathbf{h})\rangle\} = p_{\rm av} \tag{21a}$$

$$\mathrm{E}\{w(\mathbf{h})\} \geq 1-\epsilon \tag{21b}$$

$$r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h})) \geq r_{\rm o} \tag{21c}$$

$$\mathbf{p}_{\rm a}(\mathbf{h}) \geq 0 \quad \mathbf{p}_{\rm b}(\mathbf{h}) \geq 0 \quad 0 \leq w(\mathbf{h}) \leq 1 \tag{21d}$$

In the following, we derive the optimum solution of problem (21) using the generalized Karush-Kuhn-Tucker conditions theorem described in Section III.

Let $\mathbf{p}_{\rm a}^*(\mathbf{h})$, $\mathbf{p}_{\rm b}^*(\mathbf{h})$, and $w^*(\mathbf{h})$ denote the optimum solution of (21). Let h_0^*, $s^* \geq 0$, and $u^*(\mathbf{h}) \geq 0$ denote the corresponding Lagrange multipliers for constraints (21a), (21b), and (21c), respectively. Define

$$\begin{aligned} &l(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h}), \mathbf{p}_{\rm b}(\mathbf{h}), w(\mathbf{h}), h_0, s, u(\mathbf{h})) \\ &= w(\mathbf{h})[r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h})) - h_0\langle \mathbf{p}_{\rm a}(\mathbf{h})\rangle] \\ &+(1-w(\mathbf{h}))[r(\mathbf{h}, \mathbf{p}_{\rm b}(\mathbf{h})) - h_0\langle \mathbf{p}_{\rm b}(\mathbf{h})\rangle] \\ &+s\, w(\mathbf{h}) + u(\mathbf{h}) r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h})) \end{aligned} \tag{22}$$

In following, for simplicity we use the notation $l(\cdots) \triangleq l(\mathbf{h}, \mathbf{p}_{\rm a}^*(\mathbf{h}), \mathbf{p}_{\rm b}^*(\mathbf{h}), w^*(\mathbf{h}), h_0^*, s^*, u^*(\mathbf{h}))$. According to the Karush-Kuhn-Tucker necessary conditions theorem, the optimum solution must satisfy the following conditions[1]:

$$\frac{\partial l(\cdots)}{\partial p_{a,i}^*(\mathbf{h})} \begin{cases} = 0 & p_{a,i}^*(\mathbf{h}) > 0 \\ \leq 0 & p_{a,i}^*(\mathbf{h}) = 0 \end{cases} \quad \text{for } i = 1, \ldots, M \tag{23}$$

$$\frac{\partial l(\cdots)}{\partial p_{b,i}^*(\mathbf{h})} \begin{cases} = 0 & p_{b,i}^*(\mathbf{h}) > 0 \\ \leq 0 & p_{b,i}^*(\mathbf{h}) = 0 \end{cases} \quad \text{for } i = 1, \ldots, M \tag{24}$$

$$\frac{\partial l(\cdots)}{\partial w^*(\mathbf{h})} \begin{cases} = 0 & 0 < w^*(\mathbf{h}) < 1 \\ \leq 0 & w^*(\mathbf{h}) = 0 \\ \geq 0 & w^*(\mathbf{h}) = 1 \end{cases} \tag{25}$$

$$u^*(\mathbf{h})[r(\mathbf{h}, \mathbf{p}_{\rm a}^*(\mathbf{h})) - r_{\rm o}] = 0, \qquad u^*(\mathbf{h}) \geq 0 \tag{26}$$

$$s^*[\mathrm{E}\{w^*(\mathbf{h})\} - (1-\epsilon)] = 0, \qquad s^* \geq 0 \tag{27}$$

$$\mathrm{E}\{w^*(\mathbf{h})\langle \mathbf{p}_{\rm a}^*(\mathbf{h})\rangle + (1-w^*(\mathbf{h}))\langle \mathbf{p}_{\rm b}^*(\mathbf{h})\rangle\} = p_{\rm av} \tag{28}$$

Moreover, the following lemma shows that any solution that satisfies the above conditions is an optimum solution.

Lemma 2 *The Karush-Kuhn-Tucker conditions (23)-(28) are sufficient conditions for the optimum solution of problem (21).*

[1]Notation $\frac{\partial l(\cdots)}{\partial p_{a,i}^*(\mathbf{h})}$ is the derivative over $p_{a,i}(\mathbf{h})$ evaluated at $p_{a,i}^*(\mathbf{h})$

The proof of Lemma 2 requires transforming of Problem (21) into a convex optimization problem. From Karush-Kuhn-Tucker conditions (23), (24), and (26), we have the following lemma.

Lemma 3 *The optimum* $\mathbf{p}_{\rm a}^*(\mathbf{h})$ *and* $\mathbf{p}_{\rm b}^*(\mathbf{h})$ *are*

$$\mathbf{p}_{\rm a}^*(\mathbf{h}) = \begin{cases} \mathbf{p}_{\rm wf}(\mathbf{h}, h_0^*) & r(\mathbf{h}, \mathbf{p}_{\rm wf}(\mathbf{h}, h_0^*)) \geq r_o \\ \mathbf{p}_{\rm r_0}(\mathbf{h}) & \text{otherwise} \end{cases} \tag{29}$$

$$\mathbf{p}_{\rm b}^*(\mathbf{h}) = \mathbf{p}_{\rm wf}(\mathbf{h}, h_0^*)\,. \tag{30}$$

Proposition 1(b) implies that $\mathbf{p}_{\rm a}^*(\mathbf{h})$ has an equivalent expression as

$$\mathbf{p}_{\rm a}^*(\mathbf{h}) = \mathbf{p}_{\rm wf}(\mathbf{h}, h_0^*) + [\mathbf{p}_{\rm r_0}(\mathbf{h}) - \mathbf{p}_{\rm wf}(\mathbf{h}, h_0)]^+\,. \tag{31}$$

We define the second term as the *supplemental power allocation*, that is

$$\mathbf{p}_{\rm s}(\mathbf{h}, h_0) = [\mathbf{p}_{\rm r_0}(\mathbf{h}) - \mathbf{p}_{\rm wf}(\mathbf{h}, h_0)]^+\,. \tag{32}$$

The supplemental power allocation provides the additional power needed for the water-filling allocation to meet the basic rate requirement. The rate achieved by $\mathbf{p}_{\rm a}^*(\mathbf{h})$ can be expressed as

$$r(\mathbf{h}, \mathbf{p}_{\rm a}^*(\mathbf{h})) = r(\mathbf{h}, \mathbf{p}_{\rm wf}(\mathbf{h}, h_0^*)) + r_{\rm s}(\mathbf{h}, h_0^*) \tag{33}$$

with $r_{\rm s}(\mathbf{h}, h_0) = [r_{\rm o} - r(\mathbf{h}, \mathbf{p}_{\rm wf}(\mathbf{h}, h_0))]^+$ being the additional rate allocation needed for water-filling allocation to meet the basic rate requirement.

Combining Lemma 1, Lemma 3 and expression (31), the optimum power allocation $\mathbf{P}^*(\mathbf{h})$ is

$$\mathbf{P}^*(\mathbf{h}) = \mathbf{p}_{\rm wf}(\mathbf{h}, h_0^*) + X_{w^*}(\mathbf{h}) \mathbf{p}_{\rm s}(\mathbf{h}, h_0^*)\,. \tag{34}$$

In the following, we determine $w^*(\mathbf{h})$. Employing (30), (31), and (33) in $l(\cdots)$, we have

$$l(\cdots) = w^*(\mathbf{h})[s^* - g(\mathbf{h}, h_0^*)] + r(\mathbf{h}, \mathbf{p}_{\rm wf}(\mathbf{h}, h_0^*)) - h_0^*\langle \mathbf{p}_{\rm wf}(\mathbf{h}, h_0^*)\rangle \tag{35}$$

with

$$g(\mathbf{h}, h_0) = h_0\langle \mathbf{p}_{\rm s}(\mathbf{h}, h_0)\rangle - r_{\rm s}(\mathbf{h}, h_0). \tag{36}$$

The first term in $g(\mathbf{h}, h_0)$ is the power expense of allocating supplemental power, the second term is the corresponding rate return, and h_0 is the Lagrange multiplier that connects the power with the rate. Thus, function $g(\mathbf{h}, h_0)$ provides a measure for the cost of allocating the supplemental power, and is called the *supplemental cost function*.

Lemma 4 *Properties of* $g(\mathbf{h}, h_0)$ *are as follows:*

(a) *If* $\mathbf{h}' \geq \mathbf{h}$, *then* $g(\mathbf{h}', h_0) \leq g(\mathbf{h}, h_0)$.

(b) *If* $\mathbf{p}_{\rm s}(\mathbf{h}, h_0) > 0$, *then* $g(\mathbf{h}, h_0) > 0$. *If* $\mathbf{p}_{\rm s}(\mathbf{h}, h_0) = 0$, *then* $g(\mathbf{h}, h_0) = 0$.

Lemma 4(a) shows that a higher cost is associated with a poorer channel state vector. Based on Lemma 4(b) and Proposition 1, we have the following equivalent statements.

$$g(\mathbf{h}, h_0) = 0 \Longleftrightarrow \mathbf{p}_{\rm s}(\mathbf{h}, h_0) = 0 \Longleftrightarrow r(\mathbf{h}, \mathbf{p}_{\rm wf}(\mathbf{h}, h_0)) \geq r_{\rm o}. \tag{37}$$

Taking the derivative of $l(\cdots)$ over $w^*(\mathbf{h})$, we have

$$\frac{\partial l(\cdots)}{\partial w^*(\mathbf{h})} = s^* - g(\mathbf{h}, h_0^*)\,. \tag{38}$$

From condition (25) and equality (38), we obtain

$$w^*(\mathbf{h}) = \begin{cases} 1 & g(\mathbf{h}, h_0^*) < s^* \\ v^*(\mathbf{h}) & g(\mathbf{h}, h_0^*) = s^* \\ 0 & g(\mathbf{h}, h_0^*) > s^* \end{cases}, \tag{39}$$

where $0 \le v^*(\mathbf{h}) \le 1$ needs to be determined. As we can see, the cost function $g(\mathbf{h}, h_0^*)$ determines the value of $w^*(\mathbf{h})$ and indicates where the supplemental power should be allocated.

Condition (27) implies the following two situations:

- when $s^* > 0$, we must have $\mathrm{E}\{w^*(\mathbf{h})\} = 1 - \epsilon$.
- when $s^* = 0$, we must have $\mathrm{E}\{w^*(\mathbf{h})\} \ge 1 - \epsilon$. $s^* = 0$ implies that either $g(\mathbf{h}, h_0^*) = 0$ or $w^*(\mathbf{h}) = 0$. Consequently, from Lemma 4(b) we have that $X_{w^*}(\mathbf{h})\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0^*) = 0$. Therefore, in this case $\mathbf{P}^*(\mathbf{h}) = \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*)$. Since no supplemental power is allocated, $0 \le v^*(\mathbf{h}) \le 1$ can be any function that meets $\mathrm{E}\{w^*(\mathbf{h})\} \ge 1 - \epsilon$. In order to simplify the presentation and without loss of generality, we choose $v^*(\mathbf{h})$ so that $\mathrm{E}\{w^*(\mathbf{h})\} = 1 - \epsilon$.[2]

The following theorem combines above results.

Theorem 1 *If problem (4) is feasible, an optimum power allocation is*

$$\mathbf{P}^*(\mathbf{h}) = \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*) + X_{w^*}(\mathbf{h})\left[\mathbf{p}_{\mathrm{r_o}}(\mathbf{h}) - \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*)\right]^+, \tag{40}$$

where

$$w^*(\mathbf{h}) = \begin{cases} 1 & g(\mathbf{h}, h_0^*) < s^* \\ v^*(\mathbf{h}) & g(\mathbf{h}, h_0^*) = s^* \\ 0 & g(\mathbf{h}, h_0^*) > s^* \end{cases}. \tag{41}$$

and h_0^*, s^*, and $0 \le v^*(\mathbf{h}) \le 1$ are solutions to

$$\mathrm{E}\{\langle \mathbf{P}^*(\mathbf{h})\rangle\} = p_{\mathrm{av}}, \quad \mathrm{E}\{w^*(\mathbf{h})\} = 1 - \epsilon.$$

The optimum power allocation can be viewed as a two layer allocation: the first layer is the water-filling allocation, and the second layer is the supplemental allocation. The supplemental allocation provides the additional power and rate for the water-filling allocation to meet the basic rate requirement. If the channel states are so poor that the cost $g(\mathbf{h}, h_0^*)$ is above a threshold, the supplemental allocation is turned off and a service outage is declared. $g(\mathbf{h}, h_0^*)$ divides the channel space into a service set $g(\mathbf{h}, h_0^*) < s^*$ with rates $r \ge r_{\mathrm{o}}$, a boundary set $g(\mathbf{h}, h_0^*) = s^*$ with a probabilistic policy, and an outage set $g(\mathbf{h}, h_0^*) > s^*$ with rates $r < r_{\mathrm{o}}$. The service set can be further divided into a basic-rate set with rate $r = r_{\mathrm{o}}$, and an enhanced-rate set with rate $r > r_{\mathrm{o}}$.

[2]In this case, the outage probability is less than or equal to $1 - \mathrm{E}\{w^*(\mathbf{h})\} = \epsilon$.

D. Properties of the Optimum Solution

In this section, we study the properties of the optimum solution. By examining $\mathbf{P}^*(\mathbf{h})$ in Theorem 1 further, it can be seen that the optimum solution is a combination of basic-rate allocation and water-filling allocation in the non-boundary channel state set, and is randomized between these two at the boundary set. The optimum solution for $M = 1$ fading channel can be found in [14]. The optimum solution for $M = 2$ fading channels is depicted in Fig 2. The optimum solution can be classified into four types as a function of an increasing p_{av} for any given $(r_{\mathrm{o}}, \epsilon)$ as follows.

- $\mathbf{P}^*(\mathbf{h})$ is Type **I** when $p_{\mathrm{av}} = P_{\min}(r_{\mathrm{o}}, \epsilon)$. In this case, we have $\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*) = 0$ and $w^*(\mathbf{h}) = w'(\mathbf{h})$ from (17). The optimum solution is the same allocation as the outage capacity, that is $\mathbf{P}^*(\mathbf{h}) = \mathbf{P}_{\min}(\mathbf{h}) = X_{w'}(\mathbf{h})\mathbf{p}_{\mathrm{r_o}}(\mathbf{h})$.
- $\mathbf{P}^*(\mathbf{h})$ is Type **II** when $\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*) = 0$ in the outage set. In this case, we have $w^*(\mathbf{h}) = w'(\mathbf{h})$ and the outage set is the same as for the outage capacity allocation. In this case, the cost function $g(\mathbf{h}, h_0^*)$ in the outage set reduces to $h_0^*\langle \mathbf{p}_{\mathrm{r_o}}(\mathbf{h})\rangle - r_{\mathrm{o}}$. Therefore, the optimum outage set, defined as $g(\mathbf{h}, h_0^*) > s^*$, can be rewritten as $\langle \mathbf{p}_{\mathrm{r_o}}(\mathbf{h})\rangle > s'$. Type II solution includes no transmission in the outage set, a probabilistic scheme in the boundary set, basic-rate allocation in the basic-rate set, and water-filling allocation with rate $r > r_{\mathrm{o}}$ in the enhanced-rate set.
- $\mathbf{P}^*(\mathbf{h})$ is Type **III** in the most general case. All other types can be considered as special cases of Type III. It includes water-filling allocation with rate $r < r_{\mathrm{o}}$ in the outage set, a probabilistic scheme in the boundary set, basic-rate allocation in the basic-rate set, and water-filling allocation with rate $r > r_{\mathrm{o}}$ in the enhanced-rate set.
- $\mathbf{P}^*(\mathbf{h})$ is Type **IV** when $\Pr\{r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*)) \ge r_{\mathrm{o}}\} \ge 1 - \epsilon$ holds. In this case, we have $s^* = 0$ and $X_{w^*}(\mathbf{h})\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0^*) = 0$. Thus, the optimum solution is $\mathbf{P}^*(\mathbf{h}) = \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*)$.

With increasing p_{av}, $\mathbf{P}^*(\mathbf{h})$ gradually changes from Type I solution $\mathbf{P}_{\min}(\mathbf{h})$, the optimum solution for the outage capacity, to Type IV solution $\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*)$, the optimum solution for the ergodic capacity. The service outage achievable rate gradually changes from $r_{\mathrm{o}}(1 - \epsilon)$ to the ergodic capacity. The outage probability is equal to ϵ for type I-III solution, and is less than ϵ for type IV solution.

The optimum solution is probabilistic at the boundary set only when $s^* > 0$. For a continuous channel distribution, the boundary set has a probability measure zero when $s^* > 0$. Therefore, the optimum solution is deterministic for the continuous channel distribution. As stated before, the deterministic allocation problem for M-parallel fading channel is hard to solve directly. However, by considering the probabilistic allocation problem (4), we in fact obtain the optimum solution for the corresponding deterministic allocation problem for continuous channel distributions.

E. Computation of the Optimum Parameters

In this section, we study the algorithm that determines the parameters of the optimum scheme $\mathbf{P}^*(\mathbf{h})$. The optimum parameters h_0^*, s^*, and $v^*(\mathbf{h})$ are the solutions of the average

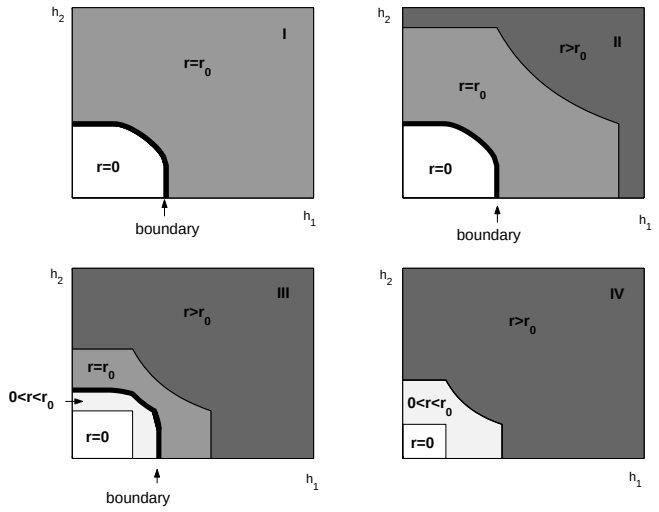

Fig. 2. Optimum solution types I-IV in $M = 2$ parallel fading channels. The optimum solution is probabilistic only at the boundary set.

power constraint and the outage probability constraint (necessary condition) as follows:

$$\mathrm{E}\{\langle \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)\rangle + \langle \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0)\rangle [1(g(\mathbf{h}, h_0) < s) + v(\mathbf{h}) 1(g(\mathbf{h}, h_0) = s)]\} = p_{\mathrm{av}}, \quad (42)$$

$$\Pr\{g(\mathbf{h}, h_0) < s)\} + \mathrm{E}\{v(\mathbf{h}) 1(g(\mathbf{h}, h_0) = s)\} = 1 - \epsilon \quad (43)$$

$$0 \le v(\mathbf{h}) \le 1\,. \quad (44)$$

A solution of (42)-(44) must exist when Problem (4) is feasible. Moreover, Lemma 2 shows that any solution of these equations is the optimum parameter set (sufficiency). In the following, we discuss algorithms to solve (42)-(44) for continuous channel distributions and discrete channel distributions respectively.

In the case of a continuous channel distribution function $F(\mathbf{h})$, $\{g(\mathbf{h}, h_0) = s\}$ is a set of probability measure zero for $s > 0$, and thus (42)-(44) can be reduced to

$$\mathrm{E}\{\langle \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)\rangle + 1(g(\mathbf{h}, h_0) \le s)\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0)\} = p_{\mathrm{av}}, \quad (45)$$

$$\Pr\{g(\mathbf{h}, h_0) \le s\} = 1 - \epsilon\,. \quad (46)$$

The left sides of (45) and (46) are continuous functions of h_0 and s, and a variety of well known root finding algorithms can be used [18].

The case of discrete channel distribution functions is more complicated, since we have to determine the value of $v(\mathbf{h})$ for the boundary set. For given h_0 and s, (42)-(44) form a linear programming problem on $v(\mathbf{h})$. Standard linear programming approaches, such as employing artificial variables and simplex method [12], can be used to determine whether there exists a feasible solution $v(\mathbf{h})$. If there exists a solution $v(\mathbf{h})$, the corresponding h_0 and s are the solutions we try to find. Therefore, the iterative algorithm is to search h_0 and s until a feasible solution $v(\mathbf{h})$ is found. The two dimensional search for h_0 and s can be carried in a sequential manner as shown below. For a given h_0, since $0 \le v(\mathbf{h}) \le 1$, (43) implies that

$$\Pr\{g(\mathbf{h}, h_0) < s)\} \le 1 - \epsilon \le \Pr\{g(\mathbf{h}, h_0) \le s\}\,. \quad (47)$$

Thus, for a given h_0, s can be expressed as

$$s(h_0) = \sup\{x : \Pr\{g(\mathbf{h}, h_0) < x\} < 1 - \epsilon\}\,. \quad (48)$$

The linear programming approach for solving $v(\mathbf{h})$ is a numerical method, and usually requires a lot of computation. Moreover, it gives no insight on the structure of the solution. Therefore, in Appendix B we have also derived the exact feasibility condition and a closed form solution of $v(\mathbf{h})$ for any (h_0, s) by exploiting the structure of (42)-(44).

V. Near Optimum Allocation Schemes

In Section IV, we derived the optimum solution for problem (4). As shown in Section IV-E, the computation of the optimum parameters, especially the optimum value of $v(\mathbf{h})$ in the boundary set is quite complicated. This motivates us to find simpler near optimum solutions in this section. As shown later in this section, there exist many near optimum schemes as long as certain requirements are satisfied. In particular, we develop two near optimum schemes, each with a specific physical interpretation.

In this section, we consider power allocations with the same two-layer structure

$$\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h})) = \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) + X_w(\mathbf{h})\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0)\,. \quad (49)$$

as the optimum solution $\mathbf{P}^*(\mathbf{h})$. As we can see, any $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ that satisfies $\mathrm{E}\{\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\} = p_{\mathrm{av}}$ and $\mathrm{E}\{w(\mathbf{h})\} \ge 1 - \epsilon$ is a feasible scheme for problem (4). In this section, by choosing some particular $(h_0, w(\mathbf{h}))$, we obtain two near optimum schemes in the form of (49). These schemes are simpler to implement than the optimum $\mathbf{P}^*(\mathbf{h})$ policy.

We first develop bounds for the average rate achieved by the $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ as follows.

Lemma 5 *The average rate achieved by* $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ *with* $\mathrm{E}\{w(\mathbf{h})\} \ge 1 - \epsilon$ *is bounded as*

$$R_{\mathrm{u}}(h_0) - r_o\epsilon \le \mathrm{E}\{r(\mathbf{h}, \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h})))\} \le R_{\mathrm{u}}(h_0),$$

where $R_{\mathrm{u}}(h_0) = r_o + \mathrm{E}\{[r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)) - r_o]^+\}$.

The upper bound is achieved when we have zero outage, and the lower bound is achieved when the rate during the outage is equal to zero.

It can be seen that for small ϵ the average rate performance is determined mainly by the value of $R_{\mathrm{u}}(h_0)$. Since $r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0))$ is a decreasing function of h_0, $R_{\mathrm{u}}(h_0)$ is a decreasing function of h_0. Thus, in order to achieve a high average rate, the value of h_0 should be small.

For a given h_0, $w(\mathbf{h})$ that satisfies the average power and outage constraints is not unique. There could be an infinite number of choices of $w(\mathbf{h})$ for a given h_0 when $\mathbf{h}$ is a vector of continuous random variables. The bounds of Lemma 5 imply that the average rate performance is relatively insensitive to the value of $w(\mathbf{h})$ for a given h_0. When the outage probability is sufficiently small, there are many near optimum schemes with small h_0 and the exact shape of the outage set is not critical.

A. Near Optimum Power Allocation I

In this section, we develop the near optimum scheme I using the general structure $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ but with particular h_0 and $w(\mathbf{h})$, so that the outage occurs when the supplemental power is above a threshold. The near optimum scheme I is simpler than the optimum solution in that parameter $v(\mathbf{h})$ is constant over the boundary set.

Consider a policy $\hat{\mathbf{P}}(\mathbf{h}) = \mathbf{p}_{\rm wf}(\mathbf{h}, \hat{h}_0) + X_{\hat{w}}(\mathbf{h})\mathbf{p}_{\rm s}(\mathbf{h}, \hat{h}_0)$ with

$$\hat{w}(\mathbf{h}) = \begin{cases} 1 & \langle \mathbf{p}_{\rm s}(\mathbf{h}, \hat{h}_0)\rangle < \hat{s} \\ \hat{v}(\mathbf{h}) & \langle \mathbf{p}_{\rm s}(\mathbf{h}, \hat{h}_0)\rangle = \hat{s} \\ 0 & \langle \mathbf{p}_{\rm s}(\mathbf{h}, \hat{h}_0)\rangle > \hat{s} \end{cases}, \tag{50}$$

where $\hat{h}_0$, $\hat{s}$, and $0 \le \hat{v}(\mathbf{h}) \le 1$ are solutions to $\mathrm{E}\{\langle \hat{\mathbf{P}}(\mathbf{h})\rangle\} = p_{\rm av}$ and $\mathrm{E}\{\hat{w}(\mathbf{h})\} = 1 - \epsilon$. We will see that it is sufficient to choose $\hat{v}(\mathbf{h}) = \hat{v}$. For any solution $(\hat{h}_0, \hat{s}, \hat{v}(\mathbf{h}))$, it can be directly verified that $(\hat{h}_0, \hat{s}, \hat{v})$ with

$$\hat{v} = \frac{\mathrm{E}\{\hat{v}(\mathbf{h}) 1(\langle \mathbf{p}_{\rm s}(\mathbf{h}, \hat{h}_0)\rangle = \hat{s})\}}{\mathrm{E}\{1(\langle \mathbf{p}_{\rm s}(\mathbf{h}, \hat{h}_0)\rangle = \hat{s})\}}$$

is also a solution to $\mathrm{E}\{\langle \hat{\mathbf{P}}(\mathbf{h})\rangle\} = p_{\rm av}$ and $\mathrm{E}\{\hat{w}(\mathbf{h})\} = 1 - \epsilon$. This simplifies the computation of parameters relative to the computation of the optimum solution $\mathbf{P}^*(\mathbf{h})$. In this section, we show that $\hat{\mathbf{P}}(\mathbf{h})$ is a near optimum scheme for problem (4). Applying the generalized Karush-Kuhn-Tucker conditions, we have the following lemma characterizing $\hat{w}(\mathbf{h})$.

Lemma 6 *$\hat{w}(\mathbf{h})$ is the optimum solution of the following problem.*

$$\hat{w}(\mathbf{h}) = \arg \min_{0 \le w(\mathbf{h}) \le 1} \mathrm{E}\{w(\mathbf{h})\langle \mathbf{p}_{\rm s}(\mathbf{h}, \hat{h}_0)\rangle\} \tag{51}$$
$$\textit{subject to} \quad \mathrm{E}\{w(\mathbf{h})\} \ge 1 - \epsilon$$

The following lemma is a corollary of Lemma 6.

Lemma 7 *Policy $\hat{\mathbf{P}}(\mathbf{h})$ has the minimum water-filling parameter h_0 among all $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ that satisfy $\mathrm{E}\{\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\} = p_{\rm av}$ and $\mathrm{E}\{w(\mathbf{h})\} \ge 1 - \epsilon$.*

In Lemma 5, it is shown that the average rate of $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ is mainly determined by $R_{\rm u}(h_0)$, which is a decreasing function of h_0. This implies that $\hat{\mathbf{P}}(\mathbf{h})$ with the minimum h_0 should be a good scheme. Applying Lemma 5, we have

$$\begin{aligned} R^* - r_{\rm o}\epsilon &\overset{(a)}{\le} R_{\rm u}(h_0^*) - r_{\rm o}\epsilon \overset{(b)}{\le} R_{\rm u}(\hat{h}_0) - r_{\rm o}\epsilon \\ &\overset{(c)}{\le} \mathrm{E}\{r(\mathbf{h}, \hat{\mathbf{P}}(\mathbf{h}))\} \le R^*. \end{aligned} \tag{52}$$

Here R^* is the maximum average rate achieved by $\mathbf{P}^*(\mathbf{h})$. Inequalities (a) and (c) are direct results of the rate bounds in Lemma 5. Inequality (b) follows from the fact that $\hat{h}_0 \le h_0^*$ by Lemma 7. Therefore, $\hat{\mathbf{P}}(\mathbf{h})$ achieves a rate between $R^* - r_{\rm o}\epsilon$ and R^*, and is a near optimum solution for Problem (4) for small ϵ.

To determine $(\hat{h}_0, \hat{s}, \hat{v})$, we have to solve

$$\begin{aligned} &\mathrm{E}\{\langle \mathbf{p}_{\rm wf}(\mathbf{h}, h_0)\rangle + \langle \mathbf{p}_{\rm s}(\mathbf{h}, h_0)\rangle 1(\langle \mathbf{p}_{\rm s}(\mathbf{h}, h_0)\rangle < s)\} \\ &\quad + s\mathrm{E}\{v1(\langle \mathbf{p}_{\rm s}(\mathbf{h}, h_0)\rangle = s)\} = p_{\rm av}, \end{aligned} \tag{53}$$

$$\Pr\{\langle \mathbf{p}_{\rm s}(\mathbf{h}, h_0)\rangle < s)\} + \mathrm{E}\{v1(\langle \mathbf{p}_{\rm s}(\mathbf{h}, h_0)\rangle = s)\} = 1 - \epsilon \tag{54}$$

$$0 \le v \le 1\,. \tag{55}$$

We search $(\hat{h}_0, \hat{s}, \hat{v})$ in an iterative way. For a given h_0, we first examine (54) and (55), and obtain $v(h_0)$ and $s(h_0)$ as a function of h_0 as

$$s(h_0) = \sup\{x : \Pr\{\langle \mathbf{p}_{\rm s}(\mathbf{h}, h_0)\rangle < x\} < 1 - \epsilon\} \tag{56}$$

$$v(h_0) = \frac{1 - \epsilon - \Pr\{\langle \mathbf{p}_{\rm s}(\mathbf{h}, h_0)\rangle < s(h_0)\}}{\Pr\{\langle \mathbf{p}_{\rm s}(\mathbf{h}, h_0)\rangle = s(h_0)\}}\,. \tag{57}$$

Then we adjust the value of h_0 until $\mathrm{E}\{\langle \hat{\mathbf{P}}(\mathbf{h})\rangle\} = p_{\rm av}$. Each time we adjust h_0, we have to compute $s(h_0)$ and $v(h_0)$ according to (56) and (57).

B. Near Optimum Power Allocation II

In this section, we develop the near optimum scheme II using the general structure $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ but with particular h_0 and $w(\mathbf{h})$, so that the policy has the same outage set as the outage capacity derived in [4]. As shown later in this section, the computational complexity of the near optimum scheme II is significantly less than that of either the optimum solution or the near optimum scheme I. Moreover, we will see that it has an appealing structure for the implementation of mixed real-time and non real-time services.

Consider a policy $\mathbf{P}'(\mathbf{h}) = \mathbf{p}_{\rm wf}(\mathbf{h}, h_0') + X_{w'}(\mathbf{h})\mathbf{p}_{\rm s}(\mathbf{h}, h_0')$ with $w'(\mathbf{h})$ given by (17) and h_0' satisfying $\mathrm{E}\{\langle \mathbf{P}'(\mathbf{h})\rangle\} = p_{\rm av}$. Recall that

$$w'(\mathbf{h}) = \begin{cases} 1 & \langle \mathbf{p}_{\rm r_0}(\mathbf{h})\rangle < s' \\ v' & \langle \mathbf{p}_{\rm r_0}(\mathbf{h})\rangle = s' \\ 0 & \langle \mathbf{p}_{\rm r_0}(\mathbf{h})\rangle > s' \end{cases}. \tag{58}$$

Policy $\mathbf{P}'(\mathbf{h})$ allocates the supplemental power at channel states where $\langle \mathbf{p}_{\rm r_0}(\mathbf{h})\rangle$ is below a threshold. Since $w'(\mathbf{h})$ does not depend on h_0', the outage set of $\mathbf{P}'(\mathbf{h})$ is much simpler than $\mathbf{P}^*(\mathbf{h})$ and $\hat{\mathbf{P}}(\mathbf{h})$. Applying the equality

$$\mathbf{p}_{\rm wf}(\mathbf{h}, h_0) + \mathbf{p}_{\rm s}(\mathbf{h}, h_0) = \mathbf{p}_{\rm r_0}(\mathbf{h}) + [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0) - \mathbf{p}_{\rm r_0}(\mathbf{h})]^+,$$

$\mathbf{P}'(\mathbf{h})$ can be expressed equivalently as

$$\mathbf{P}'(\mathbf{h}) = \mathbf{P}_{\min}(\mathbf{h}) + [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0') - \mathbf{P}_{\min}(\mathbf{h})]^+\,. \tag{59}$$

Recall that $\mathbf{P}_{\min}(\mathbf{h}) = X_{w'}(\mathbf{h})\mathbf{p}_{\rm r_0}(\mathbf{h})$ achieves the minimum sufficient power to meet the outage constraint. The physical meaning of $\mathbf{P}'(\mathbf{h})$ is that we first assign $\mathbf{P}_{\min}(\mathbf{h})$ to meet the outage constraint with the minimum sufficient power, and then we allocate the remaining power in an optimum way to maximize the excess rate. The $[\mathbf{p}_{\rm wf}(\mathbf{h}, h_0') - \mathbf{P}_{\min}(\mathbf{h})]^+$ is in fact a 'water-filling' allocation when $\mathbf{P}_{\min}(\mathbf{h})$ is viewed as interference.

It is interesting to see that $\mathbf{P}'(\mathbf{h})$ is not the optimum solution, since choosing the outage set to minimize the power for the basic rate may not maximize the expected rate at

the same time. Nevertheless, we show that $\mathbf{P}'(\mathbf{h})$ is a near optimum scheme for Problem (4) as follows.

It is hard to show directly that $\mathbf{P}'(\mathbf{h})$ is a near optimum scheme. Our approach is to introduce, as an intermediate step, a second scheme

$$\mathbf{P}''(\mathbf{h}) = \mathbf{P}_{\min}(\mathbf{h}) + [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0'') - \mathbf{p}_{\rm r_o}(\mathbf{h})]^+ , \qquad (60)$$

where h_0'' is the solution to $\mathrm{E}\{\langle \mathbf{P}''(\mathbf{h})\rangle\} = p_{\rm av}$. The following lemmas on $\mathbf{P}''(\mathbf{h})$ allow us to show that $\mathbf{P}''(\mathbf{h})$ is a near optimum scheme for problem (4).

Lemma 8 *The average rate achieved by* $\mathbf{P}''(\mathbf{h})$ *satisfies* $\mathrm{E}\left\{r(\mathbf{h}, \mathbf{P}''(\mathbf{h}))\right\} \geq R_{\rm u}(h_0'') - r_o\epsilon$.

Lemma 9 *We have* $h_0'' \leq \hat{h}_0$, *where the* $\hat{h}_0$ *is the water-filling parameter in* $\hat{\mathbf{P}}(\mathbf{h})$.

Applying Lemma 8 and Lemma 9, we have

$$\begin{aligned} R^* &\overset{\rm (a)}{\geq} \mathrm{E}\left\{r(\mathbf{h}, \mathbf{P}''(\mathbf{h}))\right\} \overset{\rm (b)}{\geq} R_{\rm u}(h_0'') - r_{\rm o}\epsilon \\ &\overset{\rm (c)}{\geq} R_{\rm u}(h_0^*) - r_{\rm o}\epsilon \overset{\rm (d)}{\geq} R^* - r_{\rm o}\epsilon\,. \end{aligned} \qquad (61)$$

Inequality (a) holds since $\mathbf{P}''(\mathbf{h})$ is a feasible scheme for problem (4). Inequality (b) follows from Lemma 8. Inequality (c) holds since $R_{\rm u}(h_0)$ is a decreasing function of h_0 and $h_0'' < \hat{h}_0 \leq h_0^*$ by Lemma 7 and Lemma 9. Applying the rate bounds in Lemma 5 to the optimum allocation $\mathbf{P}^*(\mathbf{h})$, we have $R^* = \mathrm{E}\{r(\mathbf{h}, \mathbf{P}^*(\mathbf{h}))\} \leq R_{\rm u}(h_0^*)$ and thus get inequality (d). Therefore, $\mathbf{P}''(\mathbf{h})$ is a near optimum scheme for problem (4) for small ϵ.

As we can see, both $\mathbf{P}'(\mathbf{h})$ and $\mathbf{P}''(\mathbf{h})$ first allocate $\mathbf{P}_{\min}(\mathbf{h})$ to meet the outage constraint, but the $\mathbf{P}'(\mathbf{h})$ allocates the remaining power in an optimum way to maximize the additional rate.

Lemma 10 *Scheme* $\mathbf{P}'(\mathbf{h})$ *achieves a higher average rate than* $\mathbf{P}''(\mathbf{h})$.

Therefore, $\mathbf{P}'(\mathbf{h})$ is a near optimum scheme for problem (4) for small ϵ.

The computation of $\mathbf{P}'(\mathbf{h})$ is much simpler than $\mathbf{P}^*(\mathbf{h})$ and $\hat{\mathbf{P}}(\mathbf{h})$, since its parameters (s', v') do not depend on h_0'. In particular, (s', v') can be determined by solving the outage probability constraint (4b) alone. This solution is given by (18) and (19). It follows that h_0' can be determined by solving $\mathrm{E}\{\langle \mathbf{P}'(\mathbf{h})\rangle\} = p_{\rm av}$ using a line search technique. Therefore, in $\mathbf{P}'(\mathbf{h})$ the (s', v') are the same for all values of $p_{\rm av}$, while in $\mathbf{P}^*(\mathbf{h})$ and $\hat{\mathbf{P}}(\mathbf{h})$ we have to compute (s, v) for each value of $p_{\rm av}$.

The structure of $\mathbf{P}'(\mathbf{h})$ in (59) suggests a simple implementation of transmission of mixed real-time and non real-time services. The $\mathbf{P}_{\min}(\mathbf{h})$ can be used to transmit the real-time service with the basic-rate requirement, and $[\mathbf{p}_{\rm wf}(\mathbf{h}, h_0') - \mathbf{P}_{\min}(\mathbf{h})]^+$ can be used to transmit the non real-time service. Two codebooks will be generated according to the corresponding power assignments to these two services and transmitted simultaneously using superposition coding, while successive decoding is employed at receiver.

C. Discussion

We derived three allocation schemes $\mathbf{P}^*(\mathbf{h})$, $\hat{\mathbf{P}}(\mathbf{h})$, and $\mathbf{P}'(\mathbf{h})$. All of these three schemes have a similar two-layer structure: $\mathbf{p}_{\rm wf}(\mathbf{h}, h_0) + X_w(\mathbf{h})\mathbf{p}_{\rm s}(\mathbf{h}, h_0)$ but with different h_0 and $w(\mathbf{h})$. The $w(\mathbf{h})$ determines where to allocate the supplemental power according to a metric. The metric is $g(\mathbf{h}, h_0^*)$ for $\mathbf{P}^*(\mathbf{h})$, $\langle \mathbf{p}_{\rm s}(\mathbf{h}, \hat{h}_0)\rangle$ for $\hat{\mathbf{P}}(\mathbf{h})$, and $\langle \mathbf{p}_{\rm r_o}(\mathbf{h})\rangle$ for $\mathbf{P}'(\mathbf{h})$. The service outage happens at the channel states where the metric is above a threshold. Since the metrics $g(\mathbf{h}, h_0^*)$, $\langle \mathbf{p}_{\rm s}(\mathbf{h}, \hat{h}_0)\rangle$, and $\langle \mathbf{p}_{\rm r_o}(\mathbf{h})\rangle$ are all non-increasing functions of $\mathbf{h}$, outage occurs at poor channel states for a good scheme, which is consistent with the intuition.

In previous sections, we have shown that $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ achieve a rate between $R^* - r_{\rm o}\epsilon$ and R^*. It can be directly verified that for sufficiently small $p_{\rm av}$ such that $\mathbf{p}_{\rm wf}(\mathbf{h}, h_0^*) = 0$ holds in the outage set, we have $\mathbf{P}^*(\mathbf{h}) = \hat{\mathbf{P}}(\mathbf{h}) = \mathbf{P}'(\mathbf{h})$. For sufficiently large average power, all three policies become water-filling allocation. Thus, for sufficiently small and large $p_{\rm av}$, both $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ are optimum. Moreover, as shown below, we have $\hat{\mathbf{P}}(\mathbf{h}) = \mathbf{P}'(\mathbf{h}) = \mathbf{P}^*(\mathbf{h})$ in an $M = 1$ fading channel for all parameters. When $M = 1$, since the metrics $g(h, h_0^*)$, $\langle \mathbf{p}_{\rm s}(h, \hat{h}_0)\rangle$, and $\langle \mathbf{p}_{\rm r_o}(h)\rangle$ are all non-increasing functions of h, all $w^*(h)$, $\hat{w}(h)$, and $w'(h)$ can be expressed in the same way as

$$w(h) = \begin{cases} 1 & h < h_b \\ v & h = h_b \\ 0 & h > h_b \end{cases} , \qquad (62)$$

Hence, v and h_b are the same for all three schemes since they are the solutions of $\mathrm{E}\{w(h)\} = 1 - \epsilon$. With the same $w(h)$, the average power constraints implies that $h_0^* = \hat{h}_0 = h_0'$. Thus, we have $\hat{\mathbf{P}}(\mathbf{h}) = \mathbf{P}'(\mathbf{h}) = \mathbf{P}^*(\mathbf{h})$ in an $M = 1$ fading channel.

VI. Numerical Results

In most of our numerical results, the two near optimum schemes achieve an average rate almost equal to the maximum rate R^* achieved by $\mathbf{P}^*(\mathbf{h})$, and the lower bound $R^* - r_{\rm o}\epsilon$ is loose, especially for large ϵ. To highlight the performance difference between these three schemes, we construct a particular two state model as follows: in an $M = 2$ channel, the channel state vector $\mathbf{h} = (h_1, h_2)$ is equal to $(0.1238, 0.1238)$ with probability $1/2$, and equal to $(0.1827, 0)$ with probability $1/2$. The average rate versus the average power performance for $\mathbf{P}^*(\mathbf{h})$, $\hat{\mathbf{P}}(\mathbf{h})$, and $\mathbf{P}'(\mathbf{h})$ in this model with fixed $r_{\rm o} = 0.36$ bits/symbol and $\epsilon = 1/2$ is given by Fig. 3. It can be seen that the optimum solution $\mathbf{P}^*(\mathbf{h})$ achieves a slightly higher average rate than the near optimum schemes $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$. For sufficiently small and high average power, both $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ are equal to $\mathbf{P}^*(\mathbf{h})$. In this example, $\hat{\mathbf{P}}(\mathbf{h})$ is slightly better than $\mathbf{P}'(\mathbf{h})$ in a range of parameters. As shown in Fig. 4 with $r_{\rm o} = 0.5$ bits/symbol, $\mathbf{P}'(\mathbf{h})$ is indistinguishable from $\mathbf{P}^*(\mathbf{h})$, and both are better than $\hat{\mathbf{P}}(\mathbf{h})$. The relative performance difference is still much less than $\epsilon = 0.5$.

As observed in Fig. 4, the average rate versus the average power performance for $\hat{\mathbf{P}}(\mathbf{h})$ is discontinuous near 6.1 dB and also near 7.2 dB. In fact, the performance of $\mathbf{P}^*(\mathbf{h})$ is

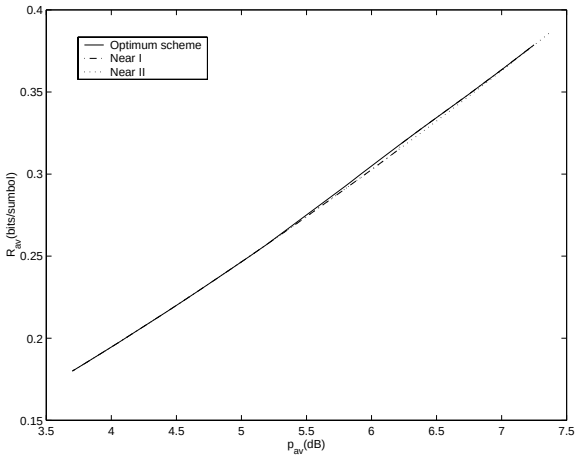

Fig. 3. The average rate performance of the optimum scheme $\mathbf{P}^*(\mathbf{h})$ versus two near optimum schemes $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ for a two state model with fixed $r_{\rm o} = 0.36$ bits/symbol and $\epsilon = 1/2$.

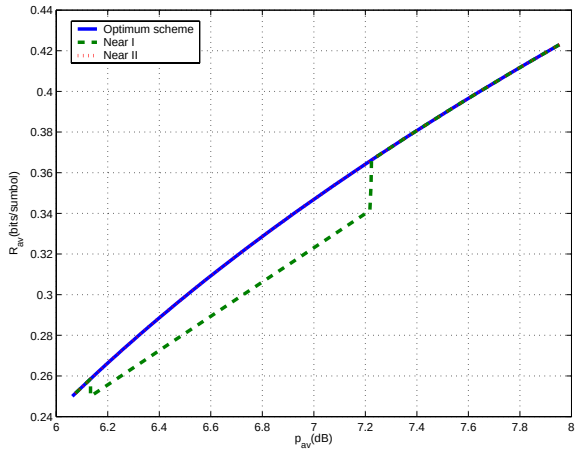

Fig. 4. The average rate performance of the optimum scheme $\mathbf{P}^*(\mathbf{h})$ versus two near optimum schemes $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ for a two state model with fixed $r_{\rm o} = 0.5$ bits/symbol and $\epsilon = 1/2$.

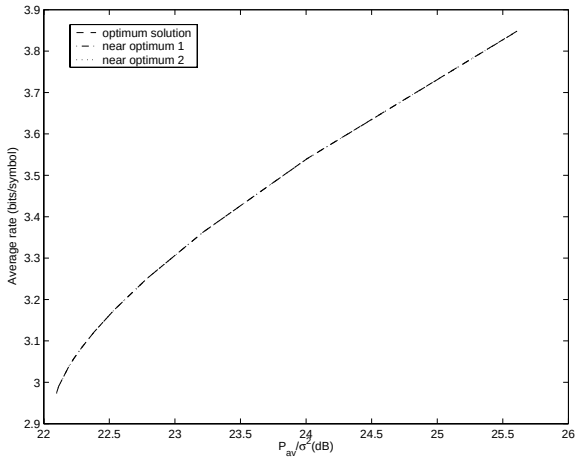

Fig. 5. The average rate performance of the optimum scheme $\mathbf{P}^*(\mathbf{h})$ versus two near optimum schemes $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ in Rayleigh fading channel with fixed $\epsilon = 0.01$ and $r_{\rm o} = 3$ bits/symbol.

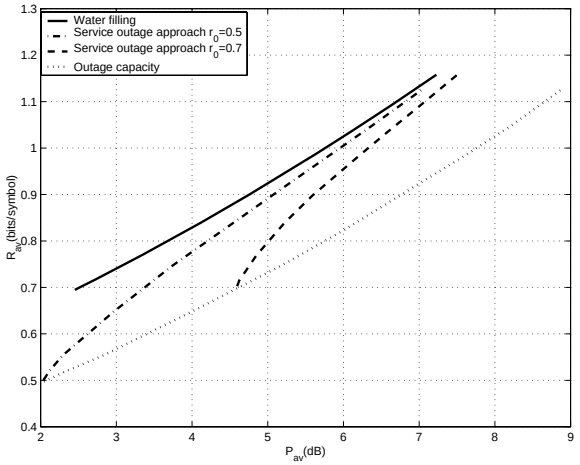

Fig. 6. Comparison of the service outage achievable rate with other capacity notions in $M = 2$ Rayleigh fading channel for a fixed $\epsilon = 0.01$.

always continuous, since $\mathbf{P}^*(\mathbf{h})$ is the optimum solution of the continuous optimization problem (4) and the complicated solution in the boundary of $\mathbf{P}^*(\mathbf{h})$ ensures the continuity of its performance. However, for the two near optimum schemes $\mathbf{P}'(\mathbf{h})$ and $\hat{\mathbf{P}}(\mathbf{h})$, we are not guaranteed a continuous solution when the channel state is discrete. The discontinuous points occur at the values of $p_{\rm av}$ for which the outage region changes.

We apply the results to the $M = 2$ parallel Rayleigh fading channel model. To simplify the computations, we assume that the sub-channels are iid with the joint PDF:

$$f(h_1, h_2) = \begin{cases} e^{-(h_1+h_2)} & h_1 \geq 0, h_2 \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{63}$$

In Fig. 5, the average rate versus the average power is plotted for $\mathbf{P}^*(\mathbf{h})$, $\hat{\mathbf{P}}(\mathbf{h})$, and $\mathbf{P}'(\mathbf{h})$ with fixed $\epsilon = 0.01$ and $r_{\rm o} = 3$ bits/symbol in Rayleigh fading channel. As we can see, the near optimum schemes are indistinguishable from the optimum solution $\mathbf{P}^*(\mathbf{h})$. In Fig 6, the service outage achievable rates with different $r_{\rm o}$ are plotted against the ergodic capacity and the outage capacity in $M = 2$ Rayleigh fading channel. As we can see, for the given outage probability $\epsilon = 0.01$, the outage capacity has a close to 2 dB loss in the average power compared to the ergodic capacity. A larger average power loss is expected when the outage probability is smaller. Between the outage capacity and the ergodic capacity, a number of service outage achievable rates with different r_0 exist. The service outage achievable rate ranges from $(1 - \epsilon)$ of the outage capacity up to the ergodic capacity. Starting from $r_0(1 - \epsilon)$, it approaches the ergodic capacity as the average power increases. The outage probability achieved by the water filling allocation with respect to different r_0 is also plotted against the service outage solution with a given $\epsilon = 0.01$ in Fig. 7. It can be observed that, for a range of $P_{\rm av}$, the service outage solution achieves a rate very close to the ergodic capacity, and, at the same time, significantly reduces the outage probability. Hence, the service outage approach strikes good balance between average rate and outage probability.

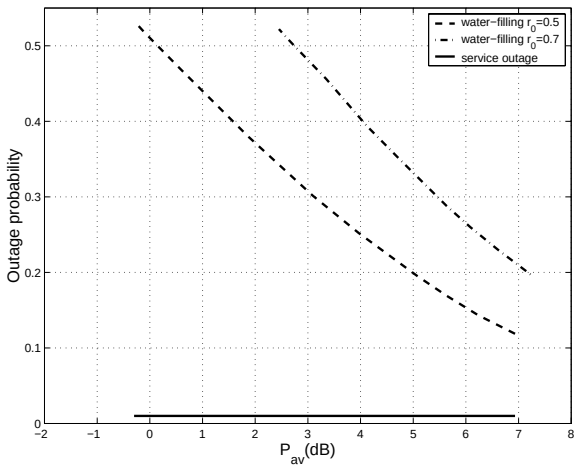

Fig. 7. The outage probability of water-filling allocation with respect to $r_{\rm o} = 0.5$ is compared to the service outage solution in $M = 2$ Rayleigh fading channel.

VII. Conclusion

The service outage based allocation problem is to maximize the expected rate subject to the average power constraint and the outage probability constraint in the class of probabilistic power allocation schemes. The feasibility condition of this allocation problem can be obtained from the capacity versus outage probability problem [4]. The optimum power allocation is derived for an M-parallel fading channels model. The result can be applied to both discrete and continuous fading distributions.

The optimum power allocation is shown to be a combination of the water-filling allocation and the basic-rate allocation, and is deterministic except at the boundary set. It can be viewed as a two-layer allocation: the first layer is the water-filling allocation, and the second layer the supplemental power allocation. The supplemental power is only allocated at channel states where the supplemental cost $g(\mathbf{h}, h_0^*)$ is below a threshold. With increasing average power, the optimum power allocation gradually changes from $\mathbf{P}_{\min}(\mathbf{h})$, which is the optimum solution for the outage capacity, to $\mathbf{p}_{\rm wf}(\mathbf{h}, h_0^*)$, which is the optimum solution for the ergodic capacity. The service outage based achievable rate R^* gradually changes from $r_{\rm o}(1-\epsilon)$ to the ergodic capacity. The service outage approach strikes a good balance between the outage probability and the average rate.

Two near optimum schemes are also derived: the $\hat{\mathbf{P}}(\mathbf{h})$ with the minimum h_0 and the $\mathbf{P}'(\mathbf{h})$ based on $\mathbf{P}_{\min}(\mathbf{h})$. Both $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ have similar structures as the optimum solution, but the supplemental power is allocated according to different metric functions. We have $\hat{\mathbf{P}}(\mathbf{h}) = \mathbf{P}'(\mathbf{h}) = \mathbf{P}^*(\mathbf{h})$ in an $M = 1$ fading channel and for a range of parameters in $M \geq 2$ fading channels. Otherwise, $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ achieve a rate between $R^* - r_{\rm o}\epsilon$ and R^*. The derivation of near optimum schemes shows that the exact shape of the outage set is not critical, a feasible scheme in the form of $\mathbf{p}_{\rm wf}(\mathbf{h}, h_0) + X_w(\mathbf{h})\mathbf{p}_{\rm s}(\mathbf{h}, h_0)$ achieves a high average rate as long as the corresponding h_0 is small. The near optimum scheme $\mathbf{P}'(\mathbf{h})$ has an immediate application on transmission of mixed real-time and non real-time services. Its computation is also significantly simpler than the optimum solution $\mathbf{P}^*(\mathbf{h})$ and the near optimum scheme $\hat{\mathbf{P}}(\mathbf{h})$.

Appendix

A. Lemma 1

Our approach is to show that for an arbitrary feasible probabilistic power allocation scheme $\mathbf{P}(\mathbf{h})$ with a conditional pdf $f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})$, we can always construct another feasible scheme $\mathbf{P}'(\mathbf{h})$ which is randomized between two deterministic schemes $\mathbf{p}_{\rm a}(\mathbf{h})$ and $\mathbf{p}_{\rm b}(\mathbf{h})$ with $r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h})) \geq r_{\rm o}$ and the sharing factor $w(\mathbf{h})$ satisfying $\mathrm{E}\{w(\mathbf{h})\} \geq 1-\epsilon$. It can be shown that $\mathbf{P}'(\mathbf{h})$ achieves a higher average rate than $\mathbf{P}(\mathbf{h})$. This implies that there exists an optimum scheme which is randomized between two deterministic schemes, and one of them has a rate higher or equal to $r_{\rm o}$ and $\mathrm{E}\{w(\mathbf{h})\} \geq 1-\epsilon$.

The feasibility of $\mathbf{P}(\mathbf{h})$ implies that $\mathrm{E}\{\langle \mathbf{P}(\mathbf{h})\rangle\} \leq p_{\rm av}$ and $\Pr\{r(\mathbf{h}, \mathbf{P}(\mathbf{h})) < r_{\rm o}\} \leq \epsilon$. Deterministic schemes $\mathbf{p}_{\rm a}(\mathbf{h})$ and $\mathbf{p}_{\rm b}(\mathbf{h})$, and the weighting function $w(\mathbf{h})$ are constructed as follows:

$$\begin{aligned} w(\mathbf{h}) &= \Pr\{r(\mathbf{h}, \mathbf{P}(\mathbf{h})) \geq r_{\rm o}|\mathbf{h}\} \\ \mathbf{p}_{\rm a}(\mathbf{h}) &= \mathrm{E}\{\mathbf{P}(\mathbf{h})|r(\mathbf{h}, \mathbf{p}(\mathbf{h})) \geq r_{\rm o}, \mathbf{h}\} \\ \mathbf{p}_{\rm b}(\mathbf{h}) &= \mathrm{E}\{\mathbf{P}(\mathbf{h})|r(\mathbf{h}, \mathbf{p}(\mathbf{h})) < r_{\rm o}, \mathbf{h}\}\,. \end{aligned} \tag{64}$$

Clearly, $1 - w(\mathbf{h})$ is the outage probability of $\mathbf{P}(\mathbf{h})$ for a given $\mathbf{h}$. Since $\mathbf{P}(\mathbf{h})$ meets the outage probability constraint, we must have $\mathrm{E}\{1-w(\mathbf{h})\} \leq \epsilon$. The $\mathbf{p}_{\rm a}(\mathbf{h})$ is the conditional average of $\mathbf{P}(\mathbf{h})$ whose rate is larger than or equal to $r_{\rm o}$, while the $\mathbf{p}_{\rm b}(\mathbf{h})$ is the conditional average of $\mathbf{P}(\mathbf{h})$ whose rate is smaller than $r_{\rm o}$. Since $r(\mathbf{h}, \mathbf{p})$ is concave on $\mathbf{p}$ for a given $\mathbf{h}$, applying Jensen's inequality we have

$$\begin{aligned} r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h})) &= r(\mathbf{h}, \mathrm{E}\{\mathbf{P}(\mathbf{h})|r(\mathbf{h}, \mathbf{p}(\mathbf{h})) \\ &\quad \geq r_{\rm o}, \mathbf{h}\}) \geq \mathrm{E}\{r(\mathbf{h}, \mathbf{P}(\mathbf{h}))|r(\mathbf{h}, \mathbf{p}(\mathbf{h})) \geq r_{\rm o}, \mathbf{h}\} \end{aligned} \tag{65}$$

$$\begin{aligned} r(\mathbf{h}, \mathbf{p}_{\rm b}(\mathbf{h})) &= r(\mathbf{h}, \mathrm{E}\{\mathbf{P}(\mathbf{h})|r(\mathbf{h}, \mathbf{p}(\mathbf{h})) < r_{\rm o}, \mathbf{h}\}) \\ &\quad \geq \mathrm{E}\{r(\mathbf{h}, \mathbf{P}(\mathbf{h}))|r(\mathbf{h}, \mathbf{p}(\mathbf{h})) < r_{\rm o}, \mathbf{h}\}\,. \end{aligned} \tag{66}$$

Consider a new probabilistic scheme $\mathbf{P}'(\mathbf{h})$ such that $\mathbf{P}'(\mathbf{h}) = \mathbf{p}_{\rm a}(\mathbf{h})$ with probability $w(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h}) = \mathbf{p}_{\rm b}(\mathbf{h})$ with probability $1 - w(\mathbf{h})$. The average power of $\mathbf{P}'(\mathbf{h})$ is

$$\begin{aligned} \mathrm{E}\{\langle \mathbf{P}'(\mathbf{h})\rangle\} &= \mathrm{E}\{w(\mathbf{h})\langle \mathbf{p}_{\rm a}(\mathbf{h})\rangle + (1-w(\mathbf{h}))\langle \mathbf{p}_{\rm b}(\mathbf{h})\rangle\} \\ &= \mathrm{E}\{\langle \mathbf{P}(\mathbf{h})\rangle\} \leq p_{\rm av}\,. \end{aligned} \tag{67}$$

Since $r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h})) \geq r_{\rm o}$ by (65) and $\Pr\{\mathbf{P}'(\mathbf{h}) = \mathbf{p}_{\rm a}(\mathbf{h})\} = w(\mathbf{h})$, we have

$$\Pr\{r(\mathbf{h}, \mathbf{P}'(\mathbf{h})) \geq r_{\rm o}|\mathbf{h}\} \geq w(\mathbf{h})\,.$$

Thus, the outage probability of $\mathbf{P}'(\mathbf{h})$ satisfies

$$\begin{aligned} \Pr\{r(\mathbf{h}, \mathbf{P}'(\mathbf{h})) < r_{\rm o}\} &= 1 - \mathrm{E}\{\Pr\{r(\mathbf{h}, \mathbf{P}'(\mathbf{h})) \geq r_{\rm o}|\mathbf{h}\}\} \\ &\leq 1 - \mathrm{E}\{w(\mathbf{h})\} \leq \epsilon\,. \end{aligned} \tag{68}$$

From (67) and (68), $\mathbf{P}'(\mathbf{h})$ is also a feasible scheme for problem (4). Inequalities (65) and (66) imply that $\mathbf{P}'(\mathbf{h})$

achieves an average rate higher than or equal to $\mathbf{P}(\mathbf{h})$, that is

$$\begin{aligned}&\mathrm{E}\{r(\mathbf{h},\mathbf{P}'(\mathbf{h}))\}\\&=\mathrm{E}\{w(\mathbf{h})r(\mathbf{h},\mathbf{p}_{\mathrm{a}}(\mathbf{h}))+(1-w(\mathbf{h}))r(\mathbf{h},\mathbf{p}_{\mathrm{b}}(\mathbf{h}))\}\\&\geq\mathrm{E}\{r(\mathbf{h},\mathbf{P}(\mathbf{h}))\}\,.\end{aligned}\tag{69}$$

Thus, from any arbitrary feasible power allocation we can always construct a better feasible power allocation which is randomized between two deterministic power allocations. This implies that there must exist an optimum power allocation which is randomized between two deterministic power allocations. Furthermore, it is required that $r(\mathbf{h},\mathbf{p}_{\mathrm{a}}(\mathbf{h}))\geq r_{\mathrm{o}}$ and $\mathrm{E}\{w(\mathbf{h})\}\geq 1-\epsilon$. Also it is easy to see that $\mathrm{E}\{\langle\mathbf{P}(\mathbf{h})\rangle\}=p_{\mathrm{av}}$ should hold for the optimum solution; otherwise, a higher average rate can be achieved by increasing the power.

B. Proof of Lemma 2

In order to prove Lemma 2, we need the following propositions.

Proposition 2 *If $f(y)$ is a concave function over y, then function $l(x,y)=xf(\frac{y}{x})$ is a concave function over non negative (x,y).*

Proof: Applying the fact that $f(y)$ is a concave function and $l(x,y)=xf(\frac{y}{x})$, we have

$$\begin{aligned}&l(\lambda x_1+(1-\lambda)x_2,\lambda y_1+(1-\lambda)y_2)\\&=(\lambda x_1+(1-\lambda)x_2)f\left(\frac{\lambda y_1+(1-\lambda)y_2}{\lambda x_1+(1-\lambda)x_2}\right)\\&=(\lambda x_1+(1-\lambda)x_2)\\&\quad f\left(\frac{\lambda x_1}{\lambda x_1+(1-\lambda)x_2}\frac{y_1}{x_1}+\frac{(1-\lambda)x_2}{\lambda x_1+(1-\lambda)x_2}\frac{y_2}{x_2}\right)\\&\overset{(a)}{\geq}(\lambda x_1+(1-\lambda)x_2)\\&\left[\frac{\lambda x_1}{\lambda x_1+(1-\lambda)x_2}f\left(\frac{y_1}{x_1}\right)+\frac{(1-\lambda)x_2}{\lambda x_1+(1-\lambda)x_2}f\left(\frac{y_2}{x_2}\right)\right]\\&=\lambda x_1 f\left(\frac{y_1}{x_1}\right)+(1-\lambda)x_2 f\left(\frac{y_2}{x_2}\right)\\&=\lambda l(x_1,y_1)+(1-\lambda)l(x_2,y_2)\,.\end{aligned}\tag{70}$$

Note that non-negativity of (x,y) is used in (a). □

In the following proposition, we use $\nabla f(\mathbf{x})$ to indicate the gradient of $f(\mathbf{x})$.

Proposition 3 *Let $\mathbf{x}$ and $\mathbf{y}$ be two vectors with equal lengths. Let $g(\mathbf{y})=f(\mathbf{X}(\mathbf{y}))$ where $\mathbf{x}=\mathbf{X}(\mathbf{y})$ is a one to one transformation between $\mathbf{x}$ and $\mathbf{y}$. If $\hat{\mathbf{x}}$ is a solution to $\nabla f(\mathbf{x})=0$, then $\hat{\mathbf{y}}=\mathbf{X}^{-1}(\hat{\mathbf{x}})$ is also a solution to $\nabla g(\mathbf{y})=0$.*

Now we return to the proof of Lemma 2, and show that problem (21) can be transformed into a convex optimization problem. Define $\mathbf{q}_a(\mathbf{h})=w(\mathbf{h})\mathbf{p}_{\mathrm{a}}(\mathbf{h})$ and $\mathbf{q}_b(\mathbf{h})=(1-w(\mathbf{h}))\mathbf{p}_{\mathrm{b}}(\mathbf{h})$. Problem (21) can be transformed as follows:

$$\begin{aligned}\max_{\mathbf{q}_a(\mathbf{h}),\mathbf{q}_b(\mathbf{h}),w(\mathbf{h})}\ &\mathrm{E}\Bigg\{w(\mathbf{h})\,r\left(\mathbf{h},\frac{\mathbf{q}_a(\mathbf{h})}{w(\mathbf{h})}\right)\\&\quad+(1-w(\mathbf{h}))\,r\left(\mathbf{h},\frac{\mathbf{q}_b(\mathbf{h})}{1-w(\mathbf{h})}\right)\Bigg\}\end{aligned}\tag{71}$$

$$\text{subject to}\quad \mathrm{E}\{\langle\mathbf{q}_a(\mathbf{h})\rangle+\langle\mathbf{q}_b(\mathbf{h})\rangle\}=p_{\mathrm{av}}\tag{71a}$$

$$\mathrm{E}\{w(\mathbf{h})\}\geq 1-\epsilon\tag{71b}$$

$$w(\mathbf{h})\,r\left(\mathbf{h},\frac{\mathbf{q}_a(\mathbf{h})}{w(\mathbf{h})}\right)-w(\mathbf{h})r_{\mathrm{o}}\geq 0\tag{71c}$$

$$\mathbf{q}_a(\mathbf{h})\geq 0\quad \mathbf{q}_b(\mathbf{h})\geq 0\quad 0\leq w(\mathbf{h})\leq 1\tag{71d}$$

Denoting $f_i(z)=\log(1+h_i z)$, it follows from Proposition 2 that

$$w(\mathbf{h})r(\mathbf{h},w^{-1}(\mathbf{h})\mathbf{q}_a(\mathbf{h}))=\frac{1}{M}\sum_{i=1}^{M}w(\mathbf{h})f_i\left(\frac{q_{\mathrm{a},i}(\mathbf{h})}{w(\mathbf{h})}\right)\tag{72}$$

is the sum of concave functions. Thus, $w(\mathbf{h})r(\mathbf{h},w^{-1}(\mathbf{h})\mathbf{q}_a(\mathbf{h}))$ is a concave function over $(w(\mathbf{h}),\mathbf{q}_a(\mathbf{h}))$. Similarly $(1-w(\mathbf{h}))r(\mathbf{h},(1-w(\mathbf{h}))^{-1}\mathbf{q}_b(\mathbf{h}))$ is a concave function over $(w(\mathbf{h}),\mathbf{q}_b(\mathbf{h}))$. Thus, the objective function is concave over $(w(\mathbf{h}),\mathbf{q}_a(\mathbf{h}),\mathbf{q}_b(\mathbf{h}))$. It can be seen that the equality constraint (71a) is a linear function over $(\mathbf{q}_a(\mathbf{h}),\mathbf{q}_b(\mathbf{h}))$, the constraint (71b) is a linear function over $w(\mathbf{h})$. Since the left side of constraint (71c) is a concave function over $(w(\mathbf{h}),\mathbf{q}_a(\mathbf{h}))$, constraint (71c) is a convex set. Thus, the constraints specify a convex feasible set. Therefore, according to the Karush-Kuhn-Tucker sufficient conditions theorem [9], the Karush-Kuhn-Tucker conditions are sufficient conditions for the transformed problem (71). Let $(\hat{\mathbf{p}_{\mathrm{a}}}(\mathbf{h}),\hat{\mathbf{p}_{\mathrm{b}}}(\mathbf{h}),\hat{w}(\mathbf{h}))$ be a solution of the Karush-Kuhn-Tucker conditions of the original problem (21). According to Proposition 3, it is easy to see that the corresponding transformed variable $(\hat{\mathbf{q}_a}(\mathbf{h}),\hat{\mathbf{q}_b}(\mathbf{h}),\hat{w}(\mathbf{h}))$ satisfies the Karush-Kuhn-Tucker conditions of the transformed problem (71). Therefore, $(\hat{\mathbf{p}_{\mathrm{a}}}(\mathbf{h}),\hat{\mathbf{p}_{\mathrm{b}}}(\mathbf{h}),\hat{w}(\mathbf{h}))$ is the optimum solution of the original problem (21), and, thus, the Karush-Kuhn-Tucker conditions of problem (21) are also sufficient.

C. Proof of Lemma 3

Condition (23) yields

$$p^*_{a,i}(\mathbf{h})=\left(\frac{1+u^*(\mathbf{h})/w^*(\mathbf{h})}{h_0^*}-\frac{1}{h_{\mathrm{i}}}\right)^+,\quad i=1,\ldots,M\,.\tag{73}$$

Condition (26) implies that:

1) When $u^*(\mathbf{h})=0$ we have $r(\mathbf{h},\mathbf{p}^*_{\mathrm{a}}(\mathbf{h}))\geq r_{\mathrm{o}}$. Moreover, when $u^*(\mathbf{h})=0$, (73) implies $\mathbf{p}^*_{\mathrm{a}}(\mathbf{h})=\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0^*)$. Thus, in this case, we have $r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0^*))\geq r_{\mathrm{o}}$.
2) When $u^*(\mathbf{h})>0$ we have $r(\mathbf{h},\mathbf{p}^*_{\mathrm{a}}(\mathbf{h}))=r_{\mathrm{o}}$. Expression (73) and $r(\mathbf{h},\mathbf{p}^*_{\mathrm{a}}(\mathbf{h}))=r_{\mathrm{o}}$ imply that $\mathbf{p}^*_{\mathrm{a}}(\mathbf{h})=\mathbf{p}_{r_{\mathrm{o}}}(\mathbf{h})$ with $(1+u^*(\mathbf{h})/w^*(\mathbf{h}))/h_0^*=\lambda(\mathbf{h})$. Since $u^*(\mathbf{h})/w^*(\mathbf{h})>0$, we have $\lambda(\mathbf{h})>1/h_0^*$. According to Proposition 1(a), we have $\lambda(\mathbf{h})>1/h_0^*$ iff $r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0^*))<r_{\mathrm{o}}$. Therefore, when $r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0^*))<r_{\mathrm{o}}$, we have $\mathbf{p}^*_{\mathrm{a}}(\mathbf{h})=\mathbf{p}_{r_{\mathrm{o}}}(\mathbf{h})$ with a rate equal to r_{o}.

Therefore,

$$\mathbf{p}_{\mathrm{a}}^{*}(\mathbf{h})=\begin{cases}\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0^{*}) & r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0))\geq r_{\mathrm{o}}\\ \mathbf{p}_{\mathrm{r_o}}(\mathbf{h}) & \text{otherwise}\end{cases}. \quad (74)$$

Lastly, condition (24) yields $\mathbf{p}_{\mathrm{b}}^{*}(\mathbf{h})=\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0^{*})$ directly.

D. Proof of Lemma 4

We need the following proposition to prove Lemma 4.

Proposition 4 *For $x\geq 0$, $t(x)=x-\log(1+x)$ is an increasing nonnegative function of x.*

Proof: When $x>0$, the first derivative $t'(x)=1-1/(1+x)>0$. Thus, $t(x)$ is increasing in x when $x\geq 0$. Since $t(0)=0$, $t(x)\geq 0$ for all $x\geq 0$. □

(a) To prove Lemma 4(a), we only need to show that function $g(\mathbf{h},h_0)$ is a non increasing function of h_i for all $i=1,2,\ldots,M$.
When $\mathbf{p}_{\mathrm{s}}(\mathbf{h},h_0)=0$, we have $r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0))\geq r_{\mathrm{o}}$ and $r_{\mathrm{s}}(\mathbf{h},h_0)=0$. Thus, we have $g(\mathbf{h},h_0)=0$ when $\mathbf{p}_{\mathrm{s}}(\mathbf{h},h_0)=0$.
When $\mathbf{p}_{\mathrm{s}}(\mathbf{h},h_0)>0$, we have $\mathbf{p}_{\mathrm{s}}(\mathbf{h},h_0)=\mathbf{p}_{\mathrm{r_o}}(\mathbf{h})-\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0)$ and $r_{\mathrm{s}}(\mathbf{h},h_0)=r_{\mathrm{o}}-r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0))$. Thus, in this case, we have

$$g(\mathbf{h},h_0)=h_0\langle\mathbf{p}_{\mathrm{r_o}}(\mathbf{h})\rangle-r_{\mathrm{o}}-q(\mathbf{h},h_0),$$

where $q(\mathbf{h},h_0)=h_0\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0)-r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0))$. We have

$$\frac{\partial q(\mathbf{h},h_0)}{\partial h_i}=\begin{cases}\frac{1}{Mh_i}\left(\frac{h_0}{h_i}-1\right)<0 & h_i>h_0\\ 0 & h_i\leq h_0\end{cases} \quad (75)$$

Partial derivative of $\langle\mathbf{p}_{\mathrm{r_o}}(\mathbf{h})\rangle$ can be computed according to the approach used in [4]. Without loss of generality, it is assumed that $h_1\geq\ldots\geq h_M$. Reference [4] shows that

$$\frac{\partial\langle\mathbf{p}_{\mathrm{r_o}}(\mathbf{h})\rangle}{\partial h_i}=\begin{cases}\frac{1}{Mh_i}\left(\frac{1}{h_i}-\lambda(\mathbf{h})\right) & i=1,\ldots,\mu\\ 0 & i=u+1,\ldots,M\end{cases}, \quad (76)$$

where μ is an integer employed in $\lambda(\mathbf{h})$. Parameter μ has a property such that $\lambda(\mathbf{h})\geq h_i^{-1}$ for $i\leq\mu$ and $\lambda(\mathbf{h})<h_i^{-1}$ for $i>\mu$ [4]. Thus, we have $(\partial/\partial h_i)\langle\mathbf{p}_{\mathrm{r_o}}(\mathbf{h})\rangle\leq 0$. From (75) and (76), it follows that

- when $h_i\leq h_0$, we have

$$(\partial/\partial h_i)g(\mathbf{h},h_0)=h_0(\partial/\partial h_i)\langle\mathbf{p}_{\mathrm{r_o}}(\mathbf{h})\rangle\leq 0.$$

- when $h_i>h_0$, we have

$$\begin{aligned}&(\partial/\partial h_i)g(\mathbf{h},h_0)\\&=h_0(\partial/\partial h_i)\langle\mathbf{p}_{\mathrm{r_o}}(\mathbf{h})\rangle-(\partial/\partial h_i)q(\mathbf{h},h_0).\end{aligned} \quad (77)$$

In the case of $\mathbf{p}_{\mathrm{s}}(\mathbf{h},h_0)>0$, by Proposition 1 we have $h_0^{-1}<\lambda(\mathbf{h})$, and thus $h_i^{-1}<h_0^{-1}<\lambda(\mathbf{h})$. By the definition of $\lambda(\mathbf{h})$, we have $\lambda(\mathbf{h})\geq h_i^{-1}$ iff $i\leq\mu$. Thus, in this case we must have $i<\mu$. Therefore, we have

$$(\partial/\partial h_i)g(\mathbf{h},h_0)=\frac{1}{Mh_i}(1-h_0\lambda(\mathbf{h}))\leq 0.$$

Thus, $g(\mathbf{h},h_0)$ is a non increasing function of h_i for all $i=1,2,\ldots,M$.

(b) Clearly, we have $g(\mathbf{h},h_0)=0$ when $\mathbf{p}_{\mathrm{s}}(\mathbf{h},h_0)=0$.
When $\mathbf{p}_{\mathrm{s}}(\mathbf{h},h_0)>0$, we want to show that $g(\mathbf{h},h_0)$ is strictly positive.
In this case, we have

$$\begin{aligned}&r_{\mathrm{s}}(\mathbf{h},h_0)\\&=r_{\mathrm{o}}-r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0))\\&=r(\mathbf{h},\mathbf{p}_{\mathrm{r_o}}(\mathbf{h},h_0))-r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0))\\&=1/M\sum_{i=1}^{M}\log\left(1+\frac{h_ip_{\mathrm{s,i}}(\mathbf{h},h_0)}{1+h_ip_{\mathrm{wf},i}(h_i,h_0)}\right)\end{aligned} \quad (78)$$

Let

$$g_i(\mathbf{h},h_0)=h_0p_{\mathrm{s,i}}(\mathbf{h},h_0)-\log\left(1+\frac{h_ip_{\mathrm{s,i}}(\mathbf{h},h_0)}{1+h_ip_{\mathrm{wf},i}(h_i,h_0)}\right),$$

then we have $g(\mathbf{h},h_0)=1/M\sum_{i=1}^{M}g_i(\mathbf{h},h_0)$.

1) When $p_{\mathrm{wf},i}(h_i,h_0)=0$, we have $h_i\leq h_0$ and $p_{\mathrm{s,i}}(\mathbf{h},h_0)=p_{\mathrm{r_o},i}(\mathbf{h})$. Then

$$\begin{aligned}g_i(\mathbf{h},h_0)&=h_0p_{\mathrm{r_o},i}(\mathbf{h})-\log(1+h_ip_{\mathrm{r_o},i}(\mathbf{h}))\\&=\frac{h_0}{h_i}h_ip_{\mathrm{r_o},i}(\mathbf{h})-\log(1+h_ip_{\mathrm{r_o},i}(\mathbf{h}))\\&\geq t(h_ip_{\mathrm{r_o},i}(\mathbf{h}))\,.\end{aligned} \quad (79)$$

2) When $p_{\mathrm{wf},i}(h_i,h_0)>0$, $1+h_ip_{\mathrm{wf},i}(h_i,h_0)=h_i/h_0$ holds, and thus

$$\begin{aligned}g_i(\mathbf{h},h_0)&=h_0p_{\mathrm{s,i}}(\mathbf{h},h_0)-\log(1+h_0p_{\mathrm{s,i}}(\mathbf{h},h_0))\\&=t(h_0p_{\mathrm{s,i}}(\mathbf{h},h_0))\,.\end{aligned} \quad (80)$$

Here, function $t(x)$ is an increasing nonnegative function of x when $x\geq 0$ by Proposition 4. Therefore, we have $g_i(\mathbf{h},h_0)>0$ for all i, and thus $g(\mathbf{h},h_0)>0$ when $\mathbf{p}_{\mathrm{s}}(\mathbf{h},h_0)>0$.

E. Proof of Lemma 5

The average rate achieved by $\mathbf{P}(\mathbf{h},h_0,w(\mathbf{h}))$ is

$$\begin{aligned}&\mathrm{E}\{r(\mathbf{h},\mathbf{P}(\mathbf{h},h_0,w(\mathbf{h})))\}\\&=\mathrm{E}\{r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0))+w(\mathbf{h})r_{\mathrm{s}}(\mathbf{h},h_0)\}\\&=\mathrm{E}\{r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0))+r_{\mathrm{s}}(\mathbf{h},h_0)\}\\&\quad-\mathrm{E}\{(1-w(\mathbf{h}))r_{\mathrm{s}}(\mathbf{h},h_0)\}\,.\end{aligned} \quad (81)$$

The first term in (81) is

$$\begin{aligned}&\mathrm{E}\{r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0))+r_{\mathrm{s}}(\mathbf{h},h_0)\}\\&=\mathrm{E}\{r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0))+[r_{\mathrm{o}}-r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0))]^{+}\}\\&\overset{(a)}{=}\mathrm{E}\{r_{\mathrm{o}}+[r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0))-r_{\mathrm{o}}]^{+}\}\\&=R_{\mathrm{u}}(h_0)\,.\end{aligned} \quad (82)$$

Equality (a) follows from $a+(b-a)^{+}=b+(a-b)^{+}$ for any a, b, and c.

The second term in (81) is bounded between 0 and $r_{\mathrm{o}}\epsilon$, since

$$0\leq r_{\mathrm{s}}(\mathbf{h},h_0)=[r_{\mathrm{o}}-r(\mathbf{h},\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0))]^{+}\leq r_{\mathrm{o}} \quad (83)$$

$$0\leq\mathrm{E}\{1-w(\mathbf{h})\}\leq\epsilon\,. \quad (84)$$

Thus,

$$R_{\rm u}(h_0) - r_{\rm o}\epsilon \le {\rm E}\{r(\mathbf{h}, \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h})))\} \le R_{\rm u}(h_0)\,. \quad (85)$$

F. Proof of Lemma 7

We need the following proposition to prove Lemma 7.

Proposition 5 *The average power* ${\rm E}\{\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\}$ *is decreasing in* h_0 *for a given* $w(\mathbf{h})$.

Proof: Proposition 5 The average power achieved by $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ can be expressed as follows:

$$\begin{aligned}
&{\rm E}\{\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\} \\
&= {\rm E}\{\langle \mathbf{p}_{\rm wf}(\mathbf{h}, h_0)\rangle + w(\mathbf{h})\langle \mathbf{p}_{\rm s}(\mathbf{h}, h_0)\rangle\} \\
&= {\rm E}\{(1 - w(\mathbf{h}))\langle \mathbf{p}_{\rm wf}(\mathbf{h}, h_0)\rangle \\
&\quad + w(\mathbf{h})(\langle \mathbf{p}_{\rm wf}(\mathbf{h}, h_0)\rangle + \langle \mathbf{p}_{\rm s}(\mathbf{h}, h_0)\rangle)\} \\
&= {\rm E}\{(1 - w(\mathbf{h}))\langle \mathbf{p}_{\rm wf}(\mathbf{h}, h_0)\rangle \\
&\quad + w(\mathbf{h})\left(\langle \mathbf{p}_{\rm r_o}(\mathbf{h})\rangle + \langle [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0) - \mathbf{p}_{\rm r_o}(\mathbf{h})]^+\rangle\right)\}\,. \quad (86)
\end{aligned}$$

Since $\langle \mathbf{p}_{\rm wf}(\mathbf{h}, h_0)\rangle$ is decreasing in h_0, the above expression implies that ${\rm E}\{\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\}$ is a decreasing function of h_0 for a given $w(\mathbf{h})$. □

Now we return to the proof of Lemma 7. For any $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ with ${\rm E}\{\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\} = P_{\rm av}$ and ${\rm E}\{w(\mathbf{h})\} \ge 1 - \epsilon$, consider a scheme $\mathbf{P}(\mathbf{h}, \hat{h}_0, w(\mathbf{h}))$. Then ${\rm E}\{\hat{w}(\mathbf{h})\langle \mathbf{p}_{\rm s}(\mathbf{h}, \hat{h}_0)\rangle\} \le {\rm E}\{w(\mathbf{h})\langle \mathbf{p}_{\rm s}(\mathbf{h}, \hat{h}_0)\rangle\}$ holds for any $w(\mathbf{h})$ that satisfies ${\rm E}\{w(\mathbf{h})\} \ge 1 - \epsilon$ according to Lemma 6. Then we have

$$\begin{aligned}
&{\rm E}\{\langle \mathbf{P}(\mathbf{h}, \hat{h}_0, w(\mathbf{h}))\rangle\} \\
&= {\rm E}\{\langle \mathbf{p}_{\rm wf}(\mathbf{h}, \hat{h}_0)\rangle + w(\mathbf{h})\langle \mathbf{p}_{\rm s}(\mathbf{h}, \hat{h}_0)\rangle\} \\
&\ge {\rm E}\{\langle \mathbf{p}_{\rm wf}(\mathbf{h}, \hat{h}_0)\rangle + \hat{w}(\mathbf{h})\langle \mathbf{p}_{\rm s}(\mathbf{h}, \hat{h}_0)\rangle\} \\
&= {\rm E}\{\langle \mathbf{P}(\mathbf{h}, \hat{h}_0, \hat{w}(\mathbf{h}))\rangle\} = p_{\rm av} \\
&= {\rm E}\{\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\} \quad (87)
\end{aligned}$$

Since ${\rm E}\{\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\}$ is a decreasing function of h_0 for a given $w(\mathbf{h})$ by Proposition 5, we have $\hat{h}_0 \le h_0$. Hence, $\hat{\mathbf{P}}(\mathbf{h})$ has the minimum water-filling parameter among all $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ that satisfies ${\rm E}\{\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\} = P_{\rm av}$ and ${\rm E}\{w(\mathbf{h})\} \ge 1 - \epsilon$.

G. Proof of Lemma 8

We show ${\rm E}\{r(\mathbf{h}, \mathbf{P}''(\mathbf{h}))\} \ge R_{\rm u}(h_0'') - r_{\rm o}\epsilon$ in this section.

For any $x_1 \ge 0$ and $x_2 \ge 0$, we have $\log(1 + x_1 + x_2) \le \log(1+x_1) + \log(1+x_2)$, and thus $\log(1+x_1) \ge \log(1+x_1+x_2) - \log(1+x_2)$. Therefore, for any two power allocations $\mathbf{p}_{\rm a}(\mathbf{h})$ and $\mathbf{p}_{\rm b}(\mathbf{h})$ we have

$$\begin{aligned}
&{\rm E}\{r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h}))\} \\
&\ge {\rm E}\{r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h}) + \mathbf{p}_{\rm b}(\mathbf{h}))\} - {\rm E}\{r(\mathbf{h}, \mathbf{p}_{\rm b}(\mathbf{h}))\}\,. \quad (88)
\end{aligned}$$

Let $\mathbf{p}_{\rm u}(\mathbf{h}, h_0) = \mathbf{p}_{\rm r_o}(\mathbf{h}) + [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0) - \mathbf{p}_{\rm r_o}(\mathbf{h})]^+$. Expanding $\mathbf{P}''(\mathbf{h})$ and applying $\mathbf{P}_{\rm min}(\mathbf{h}) = X_{w'}(\mathbf{h})\mathbf{p}_{\rm r_o}(\mathbf{h})$, we have

$$\begin{aligned}
\mathbf{P}''(\mathbf{h}) &= \mathbf{P}_{\rm min}(\mathbf{h}) + [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0'') - \mathbf{p}_{\rm r_o}(\mathbf{h})]^+ \\
&= \mathbf{P}_{\rm min}(\mathbf{h}) - \mathbf{p}_{\rm r_o}(\mathbf{h}) + \mathbf{p}_{\rm r_o}(\mathbf{h}) + [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0'') - \mathbf{p}_{\rm r_o}(\mathbf{h})]^+ \\
&= \mathbf{p}_{\rm u}(\mathbf{h}, h_0'') - (\mathbf{p}_{\rm r_o}(\mathbf{h}) - \mathbf{P}_{\rm min}(\mathbf{h})) \\
&= \mathbf{p}_{\rm u}(\mathbf{h}, h_0'') - X_{1-w'}(\mathbf{h})\mathbf{p}_{\rm r_o}(\mathbf{h})\,. \quad (89)
\end{aligned}$$

Applying Proposition 1 (b) and (c), it is easy to show that

$$r(\mathbf{h}, \mathbf{p}_{\rm u}(\mathbf{h}, h_0'')) = r_{\rm o} + [r(\mathbf{h}, \mathbf{p}_{\rm wf}(\mathbf{h}, h_0'')) - r_{\rm o}]^+\,. \quad (90)$$

Thus, ${\rm E}\{r(\mathbf{h}, \mathbf{p}_{\rm u}(\mathbf{h}, h_0''))\} = R_{\rm u}(h_0'')$.

Since ${\rm E}\{w'(\mathbf{h})\} = 1 - \epsilon$, we have

$${\rm E}\{r(\mathbf{h}, X_{1-w'}(\mathbf{h})\mathbf{p}_{\rm r_o}(\mathbf{h}))\} = {\rm E}\{1 - w'(\mathbf{h})\}r_{\rm o} = r_{\rm o}\epsilon\,. \quad (91)$$

Thus, applying (88) we have

$$\begin{aligned}
&{\rm E}\{r(\mathbf{h}, \mathbf{P}''(\mathbf{h}))\} \\
&\ge {\rm E}\{r(\mathbf{h}, \mathbf{p}_{\rm u}(\mathbf{h}, h_0''))\} - {\rm E}\{r(\mathbf{h}, X_{1-w'}(\mathbf{h})\mathbf{p}_{\rm r_o}(\mathbf{h}))\} \\
&= R_{\rm u}(h_0'') - r_{\rm o}\epsilon\,. \quad (92)
\end{aligned}$$

H. Proof of Lemma 9

To show $h_0'' \le \hat{h}_0$, we only need to show that $h_0'' \le h_0$ for any $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ that satisfies ${\rm E}\{\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\} = p_{\rm av}$ and ${\rm E}\{w(\mathbf{h})\} \ge 1 - \epsilon$.

We have

$$\begin{aligned}
&{\rm E}\{\langle \mathbf{P}''(\mathbf{h})\rangle\} \\
&= {\rm E}\{\langle \mathbf{P}_{\rm min}(\mathbf{h}) + [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0'') - \mathbf{p}_{\rm r_o}(\mathbf{h})]^+\rangle\} = p_{\rm av} \\
&= {\rm E}\{\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\} \\
&= {\rm E}\{\langle \mathbf{p}_{\rm wf}(\mathbf{h}, h_0) + w(\mathbf{h})\mathbf{p}_{\rm s}(\mathbf{h}, h_0)\rangle\} \\
&\overset{(a)}{\ge} {\rm E}\{\langle \mathbf{p}_{\rm wf}(\mathbf{h}, h_0) + [w(\mathbf{h})\mathbf{p}_{\rm r_o}(\mathbf{h}) - \mathbf{p}_{\rm wf}(\mathbf{h}, h_0)]^+\rangle\} \\
&\overset{(b)}{=} {\rm E}\{\langle w(\mathbf{h})\mathbf{p}_{\rm r_o}(\mathbf{h}) + [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0) - w(\mathbf{h})\mathbf{p}_{\rm r_o}(\mathbf{h})]^+\rangle\} \\
&\overset{(c)}{\ge} {\rm E}\{\langle w(\mathbf{h})\mathbf{p}_{\rm r_o}(\mathbf{h}) + [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0) - \mathbf{p}_{\rm r_o}(\mathbf{h})]^+\rangle\} \\
&\overset{(d)}{\ge} {\rm E}\{\langle \mathbf{P}_{\rm min}(\mathbf{h}) + [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0) - \mathbf{p}_{\rm r_o}(\mathbf{h})]^+\rangle\}\,. \quad (93)
\end{aligned}$$

Inequality (a) and (c) follows from $0 \le w(\mathbf{h}) \le 1$. Equality (b) follows from $a + (b - a)^+ = b + (a - b)^+$. Inequality (d) holds since $\mathbf{P}_{\rm min}(\mathbf{h})$ achieves the minimum power that needed to support $r_{\rm o}$ with probability $1 - \epsilon$, that is ${\rm E}\{\langle \mathbf{P}_{\rm min}(\mathbf{h})\rangle\} \le {\rm E}\{\langle w(\mathbf{h})\mathbf{p}_{\rm r_o}(\mathbf{h})\rangle\}$ for any $w(\mathbf{h})$ that satisfies ${\rm E}\{w(\mathbf{h})\} \ge 1 - \epsilon$.

Thus, it follows that

$${\rm E}\{\langle [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0'') - \mathbf{p}_{\rm r_o}(\mathbf{h})]^+\rangle\} \ge {\rm E}\{\langle [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0) - \mathbf{p}_{\rm r_o}(\mathbf{h})]^+\rangle\}\,. \quad (94)$$

Since $\langle \mathbf{p}_{\rm wf}(\mathbf{h}, h_0)\rangle$ is a decreasing function of h_0, we have $h_0'' \le h_0$, and thus $h_0'' \le \hat{h}_0$.

I. Proof of Lemma 10

In this section, we show that $\mathbf{P}'(\mathbf{h})$ achieves a higher average rate than $\mathbf{P}''(\mathbf{h})$.

For any nonnegative $\mathbf{p}_{\rm a}(\mathbf{h})$ and $\mathbf{p}_{\rm b}(\mathbf{h})$, let Π denote a set of probabilistic schemes with average power $p_{\rm av}$ such that $\mathbf{P}(\mathbf{h}) = \mathbf{p}_{\rm r_o}(\mathbf{h}) + \mathbf{p}_{\rm a}(\mathbf{h})$ with probability $w'(\mathbf{h})$ and $\mathbf{P}(\mathbf{h}) = \mathbf{p}_{\rm b}(\mathbf{h})$ with probability $1 - w'(\mathbf{h})$. Here $w'(\mathbf{h})$ is given by (17). It is easy to show that $\mathbf{P}'(\mathbf{h}) \in \Pi$ with $\mathbf{p}_{\rm a}(\mathbf{h}) = [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0') - \mathbf{p}_{\rm r_o}(\mathbf{h})]^+$ and $\mathbf{p}_{\rm b}(\mathbf{h}) = \mathbf{p}_{\rm wf}(\mathbf{h}, h_0')$, and $\mathbf{P}''(\mathbf{h}) \in \Pi$ with $\mathbf{p}_{\rm a}(\mathbf{h}) = \mathbf{p}_{\rm b}(\mathbf{h}) = [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0'') - \mathbf{p}_{\rm r_o}(\mathbf{h})]^+$.

Consider the following optimization problem

$$\max_{\mathbf{p}_{\rm a}(\mathbf{h}),\mathbf{p}_{\rm b}(\mathbf{h})} \mathrm{E}\{w'(\mathbf{h})r(\mathbf{h},\mathbf{p}_{\rm r_0}(\mathbf{h})+\mathbf{p}_{\rm a}(\mathbf{h})) + (1-w'(\mathbf{h}))r(\mathbf{h},\mathbf{p}_{\rm b}(\mathbf{h}))\} \quad (95)$$

$$\text{subject to } \mathrm{E}\{w'(\mathbf{h})\langle\mathbf{p}_{\rm r_0}(\mathbf{h})+\mathbf{p}_{\rm a}(\mathbf{h})\rangle + (1-w'(\mathbf{h}))\langle\mathbf{p}_{\rm b}(\mathbf{h})\rangle\} \le p_{\rm av} \quad (95a)$$

$$\mathbf{p}_{\rm a}(\mathbf{h}) \ge 0, \quad \mathbf{p}_{\rm b}(\mathbf{h}) \ge 0\,.$$

Applying the generalized Karush-Kuhn-Tucker conditions [11], the optimum solution of (95) is

$$\mathbf{p}^*_{\rm a}(\mathbf{h}) = [\mathbf{p}_{\rm wf}(\mathbf{h},h'_0) - \mathbf{p}_{\rm r_0}(\mathbf{h})]^+, \quad \mathbf{p}^*_{\rm b}(\mathbf{h}) = \mathbf{p}_{\rm wf}(\mathbf{h},h'_0)\,,$$

Thus, $\mathbf{P}'(\mathbf{h})$ is the optimum power allocation that maximizes the average rate in set II. Therefore, $\mathbf{P}'(\mathbf{h})$ achieves a higher average rate than $\mathbf{P}''(\mathbf{h})$.

As shown in section IV-E, to determine $(h^*_0, s^*, v^*(\mathbf{h}))$ for $\mathbf{P}^*(\mathbf{h})$, it requires solving a linear programming problem on $v(\mathbf{h})$ for given h_0 and s.

For a given h_0 and the corresponding $s(h_0)$ in (48), let

$$p = p_{\rm av} - \mathrm{E}\{\langle\mathbf{p}_{\rm wf}(\mathbf{h},h_0)\rangle - \langle\mathbf{p}_{\rm s}(\mathbf{h},h_0)\rangle \mathit{1}(g(\mathbf{h},h_0) < s(h_0))\}, \quad (96)$$

$$\delta = 1-\epsilon - \Pr\{g(\mathbf{h},h_0) < s(h_0)\}\,. \quad (97)$$

Define event B as $g(\mathbf{h},h_0) = s(h_0)$. Equations (42)-(44) become

$$\mathrm{E}\{v(\mathbf{h})\langle\mathbf{p}_{\rm s}(\mathbf{h},h_0)\rangle|B\} = p, \quad (98)$$

$$\mathrm{E}\{v(\mathbf{h})|B\} = \delta \quad (99)$$

$$0 \le v(\mathbf{h}) \le 1\,. \quad (100)$$

To determine whether (98)-(100) is feasible, we first solve the following two optimization problems.

$$p_{\min} = \min_{0\le v(\mathbf{h})\le 1} \mathrm{E}\{v(\mathbf{h})\langle\mathbf{p}_{\rm s}(\mathbf{h},h_0)\rangle|B\} \quad (101)$$

$$\text{subject to} \quad \mathrm{E}\{v(\mathbf{h})|B\} = \delta, \quad (101a)$$

and

$$p_{\max} = \max_{0\le v(\mathbf{h})\le 1} \mathrm{E}\{v(\mathbf{h})\langle\mathbf{p}_{\rm s}(\mathbf{h},h_0)\rangle|B\} \quad (102)$$

$$\text{subject to} \quad \mathrm{E}\{v(\mathbf{h})|B\} = \delta, \quad (102a)$$

Let $v_1(\mathbf{h})$ and $v_2(\mathbf{h})$ denote the corresponding optimum solution for $p_{\min}$ and $p_{\max}$ respectively.

We have the following lemma for the sub-problem.

Lemma 11 *For a given h_0, problem (98)-(100) is feasible iff $p_{\min} \le p \le p_{\max}$. When it is feasible, one solution is $v(\mathbf{h}) = \lambda v_1(\mathbf{h}) + (1-\lambda)v_2(\mathbf{h})$, where λ is the solution to $p = \lambda p_{\min} + (1-\lambda)p_{\max}$.*

The $v_1(\mathbf{h})$ can be obtained by solving problem (101) using the Karush-Kuhn-Tucker conditions. We have

$$v_1(\mathbf{h}) = \begin{cases} 1 & \mathbf{p}_{\rm s}(\mathbf{h},h_0) < t_1 \\ v_1 & \mathbf{p}_{\rm s}(\mathbf{h},h_0) = t_1 \\ 0 & \mathbf{p}_{\rm s}(\mathbf{h},h_0) > t_1 \end{cases}, \quad (103)$$

where parameters t_1 and $0 \le v_1 \le 1$ are solutions of (101a) as

$$t_1 = \sup\{x : \Pr\{B,\mathbf{p}_{\rm s}(\mathbf{h},h_0) < x\} < \delta\} \quad (104)$$

$$v_1 = \frac{\delta - \Pr\{B,\mathbf{p}_{\rm s}(\mathbf{h},h_0) < t_1\}}{\Pr\{B,\mathbf{p}_{\rm s}(\mathbf{h},h_0) = t_1\}}\,. \quad (105)$$

Similarly, solving problem (102), we have

$$v_2(\mathbf{h}) = \begin{cases} 1 & \mathbf{p}_{\rm s}(\mathbf{h},h_0) > t_2 \\ v_2 & \mathbf{p}_{\rm s}(\mathbf{h},h_0) = t_2 \\ 0 & \mathbf{p}_{\rm s}(\mathbf{h},h_0) < t_2 \end{cases}, \quad (106)$$

where t_2 and v_2 are

$$t_2 = \inf\{x : \Pr\{B,\mathbf{p}_{\rm s}(\mathbf{h},h_0) > x\} < \delta(h_0)\} \quad (107)$$

$$v_2 = \frac{\delta - \Pr\{B,\mathbf{p}_{\rm s}(\mathbf{h},h_0) > t_2\}}{\Pr\{B,\mathbf{p}_{\rm s}(\mathbf{h},h_0) = t_2\}}\,. \quad (108)$$

REFERENCES

[1] M. S. Bazaraa and C. M. Shetty. *Nonlinear programming theory and algorithms*. John Wiley and Sons, New York, 1979.

[2] E. Biglieri, J. Proakis, and S. Shamai. Fading channels: Information-theoretic and communications aspects. *IEEE Transactions on Information Theory*, 44(6):2619–2692, Oct. 1998.

[3] D. M. Burley. *Studies in optimization*. Wiley, New York, 1974.

[4] G. Caire, G. Taricco, and E. Biglieri. Optimum power control over fading channels. *IEEE Transactions on Information Theory*, 45(5):1468–1489, July 1999.

[5] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons, New York, 1991.

[6] R. Gallager. *Information Theory and Reliable Communication*. John Wiley and Sons, New York, 1968.

[7] A. J. Goldsmith and P. Varaiya. Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory*, 43(6):pp. 1986–1992, Nov. 1997.

[8] S. V. Hanly and D. N. C. Tse. Multiaccess fading channels: part II: Delay-limited capacities. *IEEE Transactions on Information Theory*, 44(7):2816–2831, Nov. 1997.

[9] A. D. Ioffe and V. M. Tihomirov. *Theory of extremal problems*. North-Holland, New York, 1979.

[10] N. Jindal and A. J. Goldsmith. Capacity and optimum power allocation for fading broadcast channels with minimum rates. In *IEEE Globecom, San Antonio, TX*, volume 2, pages 25–29, Nov., 2001.

[11] D. G. Luenberger. *Optimization by vector space methods*. John Wiley and Sons, New York, 1968.

[12] D. G. Luenberger. *Linear and nonlinear programming*. Addison-Wesley, Mass., 1984.

[13] J. Luo, L. Lin, R. Yates, and P. Spasojević. Service outage based power and rate allocation. In *Proceeding of Conference on Information Science and Systems, Johns Hopkins Univ., MD*, March 2001.

[14] J. Luo, R. Yates, and P. Spasojević. Service outage based capacity and power allocation in parallel fading channel. In *IEEE International Symposium on Information Theory*, page 108, June 2002.

[15] R. Negi, M. Charikar, and J. Cioffi. Minimum outage transmission over fading channel with delay constraint. In *Proc. of the IEEE International Communication Conference(ICC'99)*, pages 282–286, June 2000.

[16] G. O. Okikiolu. *Aspects of the theory of bounded integral operators in L^P space*. Academic Press, New York, 1971.

[17] L. H. Ozarow, S. Shamai, and A. D. Wyner. Information theoretic considerations for cellular mobile radio. *IEEE Transactions on Vehicular Technology*, 43(2):359–378, May 1994.

[18] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, New York, 1992.

Jianghong Luo received her B.S. and M.S. degrees in 1994 and 1997 from University of Science and Technology of China, and Ph.D. degree in 2004 from Rutgers University, all in Electrical Engineering. She was a research scientist in Lehigh University in 2004, and a reserach associate in Princeton University in 2005. Her research interests are in the general areas of communication theory, information theory, and signal processing.

Roy Yates received the B.S.E. degree in 1983 from Princeton University, and the S.M. and Ph.D. degrees in 1986 and 1990 from M.I.T., all in Electrical Engineering. Since 1990, he has been with the Wireless Information Networks Laboratory (WINLAB) and the ECE department at Rutgers University. Presently, he is an Associate Director of WINLAB and a Professor in the ECE Dept. He is a co-author (with David Goodman) of the text "Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers" published by John Wiley and Sons. He is a co-recipient of the 2003 IEEE Marconi Paper Prize Award in Wireless Communications. His research interests include power control, interference suppression and spectrum regulation for wireless systems.

Predrag Spasojević (M'00) received the Diploma of Engineering degree from the School of Electrical Engineering, University of Sarajevo, in 1990; and Master of Science and Doctor of Philosophy degrees in electrical engineering from Texas A&M University, College Station, Texas, in 1992 and 1999, respectively. From 2000 to 2001, he has been with WINLAB, Rutgers University, as Lucent Post-Doctoral Fellow, where he is currently an Assistant Professor in the Department of Electrical and Computer Engineering. His research interests are in the general areas of communication theory, information theory, and signal processing.