Resource Allocation for Wireless Fading Relay Channels: Max-Min Solution

Yingbin Liang, *Member, IEEE,* Venugopal V. Veeravalli, *Fellow, IEEE,* H. Vincent Poor, *Fellow, IEEE*

arXiv:0707.0969v1 [cs.IT] 6 Jul 2007

***Abstract*— Resource allocation is investigated for fading relay channels under separate power constraints at the source and relay nodes. As a basic information-theoretic model for fading relay channels, the parallel relay channel is first studied, which consists of multiple independent three-terminal relay channels as subchannels. Lower and upper bounds on the capacity are derived, and are shown to match, and thus establish the capacity for the parallel relay channel with degraded subchannels. This capacity theorem is further demonstrated via the Gaussian parallel relay channel with degraded subchannels, for which the synchronized and asynchronized capacities are obtained. The capacity achieving power allocation at the source and relay nodes among the subchannels is partially characterized for the synchronized case and fully characterized for the asynchronized case. The fading relay channel is then studied, which is based on the three-terminal relay channel with each communication link being corrupted by a multiplicative fading gain coefficient as well as an additive Gaussian noise term. For each link, the fading state information is assumed to be known at both the transmitter and the receiver. The source and relay nodes are allowed to allocate their power adaptively according to the instantaneous channel state information. The source and relay nodes are assumed to be subject to separate power constraints. For both the full-duplex and half-duplex cases, power allocations that maximize the achievable rates are obtained. In the half-duplex case, the power allocation needs to be jointly optimized with the channel resource (time and bandwidth) allocation between the two orthogonal channels over which the relay node transmits and receives. Capacities are established for fading relay channels that satisfy certain conditions.**

***Index Terms*— Capacity, max-min, parallel relay channels, resource allocation, wireless relay channels.**

I. Introduction

The three-terminal relay channel was introduced by van der Meulen [1] and was initially studied primarily in the context of multiuser information theory [1], [2], [3]. In recent years, relaying has emerged as a powerful technique to improve the reliability and throughput of wireless networks. An understanding of wireless relay channels has thus become an important area of research. Wireless relay channels and networks have been addressed from various aspects, including information-theoretic capacity [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], diversity [22], [23], [24], [25], outage performance [26], [27], and cooperative coding [28], [29], [30]. Central to the study of wireless relay channels is the problem of resource allocation. For example, the source and relay nodes can dynamically allocate their transmit powers to achieve a better rate if the fading state information is available. Resource allocation for relay channels and networks has been studied by several recent papers, including [9], [31], [32], [33], [34], [26]. Common to all of these studies is the assumption that the source and relay nodes are subject to a total power constraint.

In this paper, we study wireless fading relay channels, where we assume that the source and relay nodes are subject to separate power constraints instead of a total power constraint. This assumption is more practical for wireless networks, because the source and relay nodes are usually geographically separated, and are hence supported by separate power supplies. Under this assumption, the resource allocation problem falls under a class of *max-min* problems. We connect such *max-min* problems to the *minimax* two hypothesis testing problem (see, e.g., [35, II.C]), and apply a similar technique to find optimal (in the *max-min* sense) resource allocation strategies for fading relay channels.

We first study the parallel relay channel, which consists of multiple independent relay channels and serves as a basic information-theoretic model for fading relay channels. We derive a lower bound on the capacity based on the partial decode-and-forward scheme as well as a cut-set upper bound. We show that the two bounds match and establish the capacity for the parallel relay channel with degraded subchannels. This generalizes the capacity result in [36, Th. 12] to multiple subchannels. We also demonstrate that the parallel relay channel is not a simple combination of subchannels in that the capacity of the parallel relay channel can be larger than the sum of the capacities of subchannels, as was also remarked in [36, Sec. VII].

We then study the Gaussian parallel relay channel with degraded subchannels. There are two types of capacity that can be defined for this channel. The first is the synchronized capacity, where the source and relay inputs are allowed to be correlated. To achieve the capacity, the source and relay nodes need to choose an optimal correlation parameter for each subchannel, and further to choose an optimal power allocation across the subchannels under separate power constraints. We characterize the optimal solutions for the cases where the optimization is convex, and provide equations that need to

The material in this paper was presented in part at the Asilomar Conference on Signals, Systems and Computers, Pacific Grove, California, Nov. 2004.

This research was supported by the National Science Foundation under CAREER/PECASE Grant CCR-00-49089 and Grants ANI-03-38807 and CNS-06-25637, and by a Vodafone Foundation Graduate Fellowship.

Yingbin Liang is with the Department of Electrical Engineering, Princeton University, Engineering Quadrangle, Olden Street, Princeton, NJ 08544; e-mail: yingbinl@princeton.edu; Venugopal V. Veeravalli is with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, 106 CSL, 1308 West Main Street, Urbana, IL 61801; e-mail: vvv@uiuc.edu; H. Vincent Poor is with the Department of Electrical Engineering, Princeton University, Engineering Quadrangle, Olden Street, Princeton, NJ 08544; e-mail: poor@princeton.edu

be solved numerically for cases where the optimization is nonconvex. We also study the asynchronized capacity, where the source and relay inputs are required to be independent. This capacity is easier to achieve in practice due to the simpler transceiver design for the source and relay nodes. We fully characterize the capacity-achieving power allocation at the source and relay nodes in closed form.

We then move on to study the fading relay channel, which is based on the classical relay channel with each transmission link being corrupted by a multiplicative stationary and ergodic fading process as well as an additive white Gaussian noise process. The fading relay channel is a special case of the parallel relay channel, with each subchannel corresponding to one fading state realization. We assume that both the transmitter and the receiver know the channel state information, so that the source and relay nodes can allocate their transmission powers adaptively according to the instantaneous fading state information. We consider the resource allocation problem for two fading relay models: full-duplex and half-duplex.

The fading full-duplex relay channel has been studied in [9], where lower and upper bounds on the capacity were derived, along with the resource allocation that optimizes these bounds, under a total power constraint for the source and relay nodes. In this paper, we assume separate power constraints for the source and relay nodes and study the power allocation that optimizes the capacity bounds. We focus on the more practical asynchronized case. We obtain the power allocation that maximizes an achievable rate, and show that the optimal power allocation may be *two-level water-filling*, *orthogonal division water-filling*, or *iterative water-filling* depending on the channel statistics and the power constraints. We also establish the asynchronized capacity for channels that satisfy a certain condition.

We further study a fading half-duplex relay channel model, where the source node transmits to the relay and destination nodes in one channel, and the relay node transmits to the destination node in an orthogonal channel. We introduce a parameter θ to represent the channel resource (time and bandwidth) allocation between the two orthogonal channels. We study three scenarios. In Scenario I, where the two orthogonal channels share the channel resource equally, i.e., $\theta = 1/2$, we show that the optimal power allocation falls into three cases depending on the ranges of power constraints at the source and relay nodes. The optimal power allocation for the relay node is always water-filling, but the power allocation for the source node is not water-filling in general. In scenario II, the channel resource allocation parameter θ needs to be same for all channel states but can be jointly optimized with the power allocation. In Scenario III, which is the most general scenario, θ can change with channel realizations and is jointly optimized with power allocation. For both Scenarios II and III, we derive the jointly optimal θ and power allocation that maximize the achievable rate. Furthermore, we show that the lower bound achieves the cut-set upper bound if the channel statistics and power constraint satisfy a certain condition. We hence establish the capacity for these channels over all possible power and channel resource allocations.

The paper is organized as follows. In Section II, the parallel relay channel is introduced and studied. In Section III, the optimal resource allocation that achieves the capacity for the Gaussian parallel relay channel with degraded subchannels is studied. In Section IV, resource allocation for the fading full-duplex relay channel is presented. In Section V, resource allocation for the fading half-duplex relay channel is studied, where the three scenarios described above are considered. Finally in Section VI, we give concluding remarks.

II. Parallel Relay Channels

In this section, we study the parallel relay channel, which serves as a basic information-theoretic model for the fading relay channels that are considered in Sections IV and V. The parallel relay channel also models the relay channel where the source and relay nodes can transmit over multiple frequency bands with each subchannel corresponding to the channel over one frequency band. It is shown in this section that in contrast to the parallel point-to-point channel, the parallel relay channel is not a simple combination of independent subchannels.

***Definition** 1:* A parallel relay channel with K subchannels (see Fig. 1) consists of K finite source input alphabets $\mathcal{X}_1, \dots, \mathcal{X}_K$, K finite relay input alphabets $\mathcal{X}_{R1}, \dots, \mathcal{X}_{RK}$, K finite destination output alphabets $\mathcal{Y}_1, \dots, \mathcal{Y}_K$ and K finite relay output alphabets $\mathcal{Y}_{R1}, \dots, \mathcal{Y}_{RK}$. The transition probability distribution is given by

$$\prod_{k=1}^{K} p_k(y_k, y_{Rk} | x_k, x_{Rk}) \tag{1}$$

where $x_k \in \mathcal{X}_k$, $x_{Rk} \in \mathcal{X}_{Rk}$, $y_k \in \mathcal{Y}_k$, and $y_{Rk} \in \mathcal{Y}_{Rk}$ for $k = 1, \dots, K$.

A $\left(2^{nR}, n\right)$ code consists of the following:

- One message set $\mathcal{W} = \{1, 2, \dots, 2^{nR}\}$ with the message W uniformly distributed over $\mathcal{W}$;
- One encoder at the source node that maps each message $w \in \mathcal{W}$ to a codeword

$$(x_{11}, \dots, x_{1n}, \dots, x_{K1}, \dots, x_{Kn});$$

- A set of relay functions $\{f_i\}_{i=1}^n$ such that for $1 \le i \le n$:

$$\begin{aligned}&(x_{R1i}, \dots, x_{RKi})\\ &\quad = f_i(y_{R11}, \dots, y_{R1[i-1]}, \dots, y_{RK1}, \dots, y_{RK[i-1]});\end{aligned}$$

- One decoder at the destination node that maps a received sequence $(y_{11}, \dots, y_{1n}, \dots,\ y_{K1}, \dots, y_{Kn})$ to a message $\hat{w} \in \mathcal{W}$.

Note that the relay node is allowed to jointly encode and decode across the K parallel subchannels.

A rate R is *achievable* if there exists a sequence of $\left(2^{nR}, n\right)$ codes with the average probability of error at the destination node going to zero as n goes to infinity.

The following theorem provides lower and upper bounds on the capacity of the parallel relay channel.

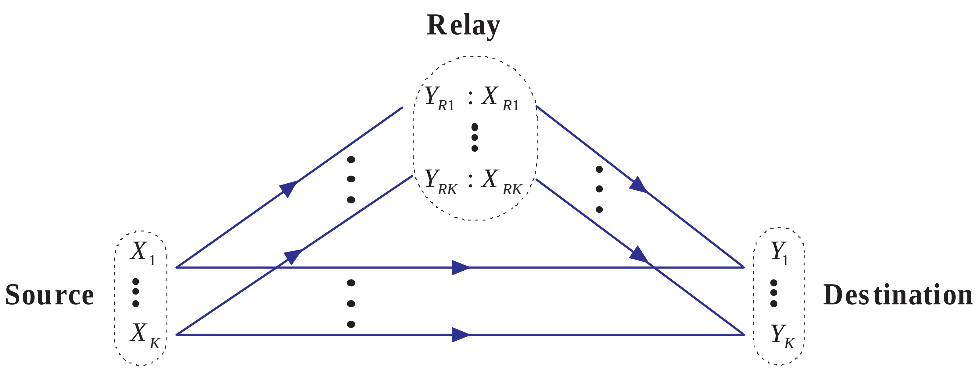

Fig. 1. Parallel relay channel

***Theorem** 1:* For the parallel relay channel, a lower bound on the capacity is given by

$$\begin{aligned} C_{\texttt{low}} = \max\min\Bigg\{ &\sum_{k=1}^{K} I(X_k, X_{Rk}; Y_k), \\ &\sum_{k=1}^{K} I(Q_k; Y_{Rk}|X_{Rk}) + I(X_k; Y_k|Q_k, X_{Rk}) \Bigg\} \end{aligned} \tag{2}$$

where Q_k for $k = 1, \ldots, K$ are auxiliary random variables. The maximum in (2) is over the joint distribution

$$\prod_{k=1}^{K} p_k(q_k, x_{Rk}, x_k) p_k(y_k, y_{Rk}|x_k, x_{Rk}).$$

An upper bound on the capacity is given by

$$\begin{aligned} C_{\texttt{up}} = \max\min\Bigg\{ &\sum_{k=1}^{K} I(X_k, X_{Rk}; Y_k), \\ &\sum_{k=1}^{K} I(X_k; Y_k, Y_{Rk}|X_{Rk}) \Bigg\} \end{aligned} \tag{3}$$

where the maximum in (3) is over the joint distribution

$$\prod_{k=1}^{K} p_k(x_{Rk}, x_k) p_k(y_k, y_{Rk}|x_k, x_{Rk}).$$

***Remark** 1:* The lower bound (2) generalizes the rate given in [37, Theorem 1] based on the decode-and-forward scheme.

Proof: To derive the lower bound (2), we use the following achievable rate for the relay channel based on the partial decode-and-forward scheme given in [3]:

$$\begin{aligned} R < &\max\min \\ &\Big\{ I(X_R, X; Y),\ I(Q; Y_R|X_R) + I(X; Y|Q, X_R) \Big\} \end{aligned} \tag{4}$$

We set $Q = (Q_1, \ldots, Q_K)$, $X = (X_1, \ldots, X_K)$, $X_R = (X_{R1}, \ldots, X_{RK})$, $Y = (Y_1, \ldots, Y_K)$, and $Y_R = (Y_{R1}, \ldots, Y_{RK})$ in the above achievable rate. We further choose (Q_1, X_1, X_{R1}), ..., (Q_K, X_K, X_{RK}) to be independent, and then obtain the lower bound (2).

The upper bound (3) is based on the cut-set bound [2, Theorem 4] and the independency of the K parallel subchannels. ∎

***Remark** 2:* In the achievable scheme, the relay node first decodes information sent by the source node over each subchannel. The relay node then reassigns total decoded information to each subchannel to forward to the destination node. Hence information that was sent to the relay node over one subchannel may be forwarded to the destination node over other subchannels, as long as the total rate at which the relay node can forward information to the destination node over all subchannels is larger than the total rate at which the relay node can decode information from the source node.

The lower and upper bounds in Theorem 1 do not match in general. We next study a class of parallel relay channels with degraded subchannels. For this channel, the lower and upper bounds match, and we hence establish the capacity. Moreover, this capacity provides an achievable rate for the case where the subchannels are either stochastically degraded or reversely degraded (e.g., fading relay channels).

***Definition** 2:* Consider the parallel relay channel with degraded subchannels. Assume each subchannel is either degraded or reversely degraded, i.e., each subchannel satisfies either

$$\begin{aligned} &p_k(y_k, y_{Rk}|x_k, x_{Rk}) \\ &\quad = p_k(y_{Rk}|x_k, x_{Rk}) p_k(y_k|y_{Rk}, x_{Rk}), \end{aligned} \tag{5}$$

or

$$\begin{aligned} &p_k(y_k, y_{Rk}|x_k, x_{Rk}) \\ &\quad = p_k(y_k|x_k, x_{Rk}) p_k(y_{Rk}|y_k, x_{Rk}). \end{aligned} \tag{6}$$

We note that the parallel relay channel with degraded subchannels has been studied in [36, Sec. VII] for the two-subchannel case. We now generalize the result in [36, Sec. VII] to channels with multiple subchannels. In fact, our main focus is on the Gaussian case considered in this section and Section III.

We define the set A to contain the indices of the subchannels that satisfy (5), i.e., those subchannels where the source-to-relay channel is stronger than the source-to-destination channel. Then the set A^c contains the indices of the subchannels that satisfy (6), i.e., those subchannels where the source-to-relay channel is weaker than the source-to-destination channel. Note that in general the parallel relay channel with degraded subchannels is neither a degraded relay channel nor a reversely degraded channel. For this channel, the lower and upper bounds given in Theorem 1 match and establish the following capacity theorem.

***Theorem** 2:* For the parallel relay channel with degraded subchannels, the capacity is given by

$$C = \max \min \left\{ \sum_{k=1}^{K} I(X_k, X_{Rk}; Y_k), \right. \\ \left. \sum_{k \in A} I(X_k; Y_{Rk}|X_{Rk}) + \sum_{k \in A^c} I(X_k; Y_k|X_{Rk}) \right\}. \tag{7}$$

where the maximum is over the joint distribution

$$\prod_{k=1}^{K} p_k(x_{Rk}, x_k) p_k(y_k, y_{Rk}|x_k, x_{Rk}).$$

***Remark** 3:* Theorem 2 generalizes the capacity of the parallel relay channel with unmatched degraded subchannels in [36, Theorem 12] to channels with multiple subchannels.

Proof: The achievability follows from $C_{\texttt{low}}$ in (2) by setting $Q_k = X_k$ for $k \in A$ and setting $Q_k = \phi$ for $k \in A^c$. The converse follows from $C_{\texttt{up}}$ in (3) by applying the degradedness conditions (5) and (6). ■

Note that the partial decode-and-forward scheme achieves the capacity of the parallel relay channel with degraded subchannels. From the selection of Q_k in the above achievability proof, it can be seen that the relay node decodes all the information sent over the degraded subchannels, i.e., $Q_k = X_k$ for $k \in A$, and decodes no information sent over the reversely degraded subchannels, i.e., $Q_k = \phi$ for $k \in A^c$. Hence for the subchannels with $k \in A^c$, the link from the source node to the relay node can be eliminated without changing the capacity of the channel.

However, the relay node still plays an important role in the reversely degraded subchannels by forwarding information that it has decoded in other degraded subchannels to the destination node. This is different from the role of the relay node in a single reversely degraded channel, where it does not forward information at all. Furthermore, we see that in the parallel relay channel, information may be transmitted from the source node to the relay node in one subchannel, and be forwarded to the destination node over other subchannels, as we have commented in Remark 2. More importantly, in contrast to the parallel point-to-point channel, the capacity of the parallel relay channel with degraded subchannels in Theorem 2 can be larger than the following sum of the capacities of the subchannels

$$\max \; \min \left\{ \sum_{k \in A} I(X_k, X_{Rk}; Y_k), \; \sum_{k \in A} I(X_k; Y_{Rk}|X_{Rk}) \right\} \\ + \sum_{k \in A^c} I(X_k; Y_k|X_{Rk}). \tag{8}$$

This demonstrates that the parallel relay channel is not a simple combination of independent subchannels. This fact has also been pointed out in [36, Remark 15] for two-subchannel case.

We now consider a Gaussian example of the parallel relay channel with degraded subchannels. The channel input-output relationship at one time instant is as follows.

$$\text{For } k \in A, \qquad \begin{aligned} Y_{Rk} &= X_k + Z_{Rk} \\ Y_k &= X_k + \sqrt{\rho_{Rk}} X_{Rk} + Z_{Rk} + Z'_k, \end{aligned} \tag{9}$$

where Z_{Rk} and Z'_k are independent Gaussian random variables with variances σ_{Rk}^2 and $\sigma_k^2 - \sigma_{Rk}^2$, respectively. For $k \in A$, $\sigma_k^2 > \sigma_{Rk}^2$.

$$\text{For } k \in A^c, \qquad \begin{aligned} Y_{Rk} &= X_k + Z_k + Z'_{Rk} \\ Y_k &= X_k + \sqrt{\rho_{Rk}} X_{Rk} + Z_k, \end{aligned} \tag{10}$$

where Z_k and Z'_{Rk} are independent Gaussian random variables with variances σ_k^2 and $\sigma_{Rk}^2 - \sigma_k^2$, respectively. For $k \in A^c$, $\sigma_{Rk}^2 \geq \sigma_k^2$. In (9) and (10), ρ_{Rk} (assumed to be positive) indicates the ratio of the relay-to-destination SNR to the source-to-destination SNR for subchannel k. We assume that the source and relay input sequences are subject to the following average power constraints:

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \mathrm{E}\left[X_{ki}^2\right] \leq P \;, \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \mathrm{E}\left[X_{Rki}^2\right] \leq P_R \;. \tag{11}$$

where i is the time index.

It can be seen from (9) and (10) that the subchannels with $k \in A$ satisfy the degradedness condition (5) and the subchannels with $k \in A^c$ satisfy the degradedness condition (6). Hence the Gaussian channel defined in (9) and (10) is the parallel relay channel with degraded subchannels. The following capacity theorem is based on Theorem 2.

***Theorem** 3:* The capacity of the Gaussian parallel relay channel with degraded subchannels is given by

$$C = \max_{\substack{\sum_{k=1}^{K} P_k \leq P, \;\; \sum_{k=1}^{K} P_{Rk} \leq P_R, \\ 0 \leq \beta_k \leq 1, \;\; \text{for } k=1,\ldots,K}} \\ \min \left\{ \sum_{k=1}^{K} \mathcal{C}\left(\frac{P_k + \rho_{Rk} P_{Rk} + 2\sqrt{\bar{\beta}_k \rho_{Rk} P_k P_{Rk}}}{\sigma_k^2} \right), \right. \\ \left. \sum_{k \in A} \mathcal{C}\left(\frac{\beta_k P_k}{\sigma_{Rk}^2} \right) + \sum_{k \in A^c} \mathcal{C}\left(\frac{\beta_k P_k}{\sigma_k^2} \right) \right\}. \tag{12}$$

where $\bar{\beta}_k = 1 - \beta_k$, and the function $\mathcal{C}(x) := \frac{1}{2}\log(1+x)$. In (12), the parameter $\bar{\beta}_k$ indicates correlation between the source input and the relay input to subchannel k, and P_k and P_{Rk} indicate the source and relay powers that are allocated for transmission over subchannel k.

Proof: The achievability follows from Theorem 2 by choosing the following joint distribution:

$$\begin{aligned} & X_{Rk} \sim \mathcal{N}(0, P_{Rk}), \\ & X'_k \sim \mathcal{N}(0, \beta_k P_k), \qquad \text{with } X'_k \text{ independent of } X_{Rk}, \\ & X_k = \sqrt{\frac{\bar{\beta}_k P_k}{P_{Rk}}} X_{Rk} + X'_k \end{aligned} \tag{13}$$

The converse is similar to the steps in the converse proof in [2, Sec. IV], and is omitted. ■

Note that the capacity in Theorem 3 is sometimes referred to as the synchronized capacity, because the source and relay nodes are allowed to use correlated inputs to exploit coherent combining gain. This may not be practical for encoder design. It is hence interesting to study the asynchronized capacity, where the source and relay nodes are assumed to use independent inputs. The following asynchronized capacity is derived by setting $\beta_k = 1$ for $k = 1, \ldots, K$ in (12).

***Corollary** 1:* For the Gaussian parallel relay channel with degraded subchannels, the asynchronized capacity is given by

$$C = \max_{\substack{\sum_{k=1}^K P_k \le P, \\ \sum_{k=1}^K P_{Rk} \le P_R}} \min \left\{ \sum_{k=1}^K \mathcal{C}\left(\frac{P_k + \rho_{Rk} P_{Rk}}{\sigma_k^2} \right), \right. \tag{14}$$
$$\left. \sum_{k \in A} \mathcal{C}\left(\frac{P_k}{\sigma_{Rk}^2} \right) + \sum_{k \in A^c} \mathcal{C}\left(\frac{P_k}{\sigma_k^2} \right) \right\}.$$

To obtain the capacity in Theorem 3 and the asynchronized capacity in Corollary 1, we still need to solve the optimization problems in (12) and (14), i.e., to find the jointly optimal correlation parameters $\{\beta_k, \text{ for } k = 1, \ldots, K\}$ and power allocations $\{(P_k, P_{Rk}), \text{ for } k = 1, \ldots, K\}$ in (12), and to find the optimal power allocations $\{(P_k, P_{Rk}), \text{ for } k = 1, \ldots, K\}$ in (14). We study these optimization problems in the next section.

III. Optimal Resource Allocation for Gaussian Parallel Relay Channels with Degraded Subchannels

In this section, we study the optimization problems in (12) and (14), which are *max-min* optimization problems. We first introduce a general technique for solving this class of *max-min* optimization problems. We then demonstrate the application of this technique by finding the optimal solutions in (12) and (14). We obtain the analytic form of the jointly optimal correlation parameters $\{\beta_k, \text{ for } k = 1, \ldots, K\}$ and power allocation $\{(P_k, P_{Rk}), \text{ for } k = 1, \ldots, K\}$ that achieve the synchronized capacity for the cases where the optimization problem is convex. We also obtain a closed-form solution for the optimal $\{(P_k, P_{Rk}), \text{ for } k = 1, \ldots, K\}$ that achieve the asynchronized capacity. This optimal solution may have three different structures depending on the channel SNRs and power constraints. This optimal power allocation is directly related to the power allocation for the fading full-duplex relay channel presented in Section IV.

A. Technique to Solve a Class of Max-Min Problem

Consider the following max-min problem:

$$\max_{\underline{t} \in \mathcal{G}} \min \{R_1(\underline{t}),\ R_2(\underline{t})\} \tag{15}$$

where $\underline{t}$ is a real vector in a set $\mathcal{G}$, and $R_1(\underline{t})$ and $R_2(\underline{t})$ are real continuous functions of $\underline{t}$. An optimal $\underline{t}^*$ is referred to as a *max-min rule*.

We now introduce a technique to solve the max-min problem (15). We will also illustrate this technique with a geometric interpretation. This technique is similar to that used in finding the *minimax* detection rule in the two hypothesis testing problem (see, e.g., [35, Sec. II.C]).

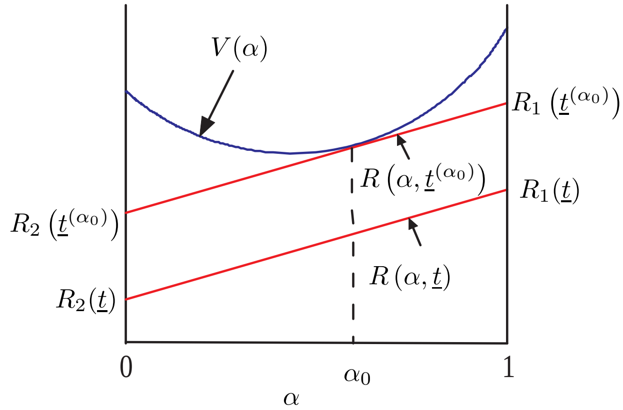

Fig. 2. Illustration of functions $V(\alpha)$ and $R(\alpha, \underline{t})$

Consider the following function:

$$R(\alpha, \underline{t}) := \alpha R_1(\underline{t}) + (1-\alpha) R_2(\underline{t}), \qquad 0 \le \alpha \le 1. \tag{16}$$

As a function of α, $R(\alpha, \underline{t})$ is a straight line from $R(0, \underline{t}) = R_2(\underline{t})$ to $R(1, \underline{t}) = R_1(\underline{t})$. Hence the maximization in (15) corresponds to maximizing the minimal of the two end points of the line $R(\alpha, \underline{t})$ over all possible $\underline{t} \in \mathcal{G}$.

We further define a function

$$V(\alpha) := \max_{\underline{t} \in \mathcal{G}} R(\alpha, \underline{t}) = R(\alpha, \underline{t}^{(\alpha)}), \tag{17}$$

where $\underline{t}^{(\alpha)}$ maximizes $R(\alpha, \underline{t})$ for fixed α. From the definitions of $V(\alpha)$ and $R(\alpha, \underline{t})$, it is easy to see the following two facts (see Fig. 2 for an illustration):

Fact 1: The function $V(\alpha)$ is continuous and convex for $\alpha \in [0, 1]$;

Fact 2: For any power allocation rule $\underline{t} \in \mathcal{G}$, $R(\alpha, \underline{t})$ as a function of α is completely below the convex curve $V(\alpha)$ or tangent to it.

A known general solution to the max-min optimization problem in (15) is summarized in the following proposition.

***Proposition** 1:* Suppose α^* is a solution to $V(\alpha^*) = \min_{\alpha \in [0,1]} V(\alpha)$. Then $\underline{t}^{(\alpha^*)}$ is a max-min rule, i.e., a solution to the max-min problem in (15). The relationship between $R_1(\underline{t}^{(\alpha^*)})$ and $R_2(\underline{t}^{(\alpha^*)})$ falls into the following three cases (see Fig. 3):

Case 1: If $\alpha^* = 0$, $R_1(\underline{t}^{(\alpha^*)}) \ge R_2(\underline{t}^{(\alpha^*)})$;

Case 2: If $\alpha^* = 1$, $R_1(\underline{t}^{(\alpha^*)}) \le R_2(\underline{t}^{(\alpha^*)})$;

Case 3: (Equalizer rule) If $0 < \alpha^* < 1$, $R_1(\underline{t}^{(\alpha^*)}) = R_2(\underline{t}^{(\alpha^*)})$.

This technique of finding the max-min solution is applied throughout the paper.

B. Optimal Resource Allocation for Gaussian Parallel Relay Channel: Synchronized Case

In this subsection, we apply Proposition 1 to find jointly optimal $\{\beta_k, \text{ for } k = 1, \ldots, K\}$ and $\{(P_k, P_{Rk}), \text{ for } k = 1, \ldots, K\}$ that solve the max-min problem in (12). This

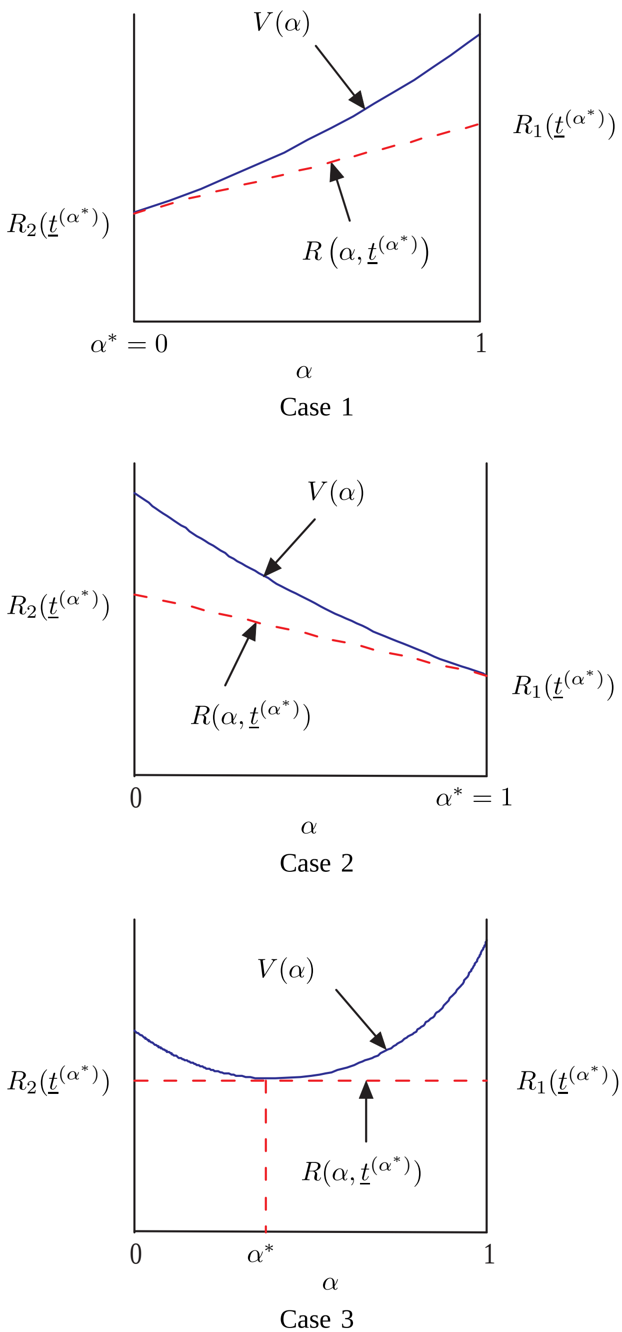

Fig. 3. Illustration of Cases 1, 2, and 3 in Proposition 1

optimal solution provides the optimal correlation between the source and relay inputs over each subchannel and the optimal source and relay power allocation among the K subchannels that achieve the synchronized capacity of the Gaussian parallel relay channel with degraded subchannels. We study the asynchronized case in the next subsection.

To simplify notation, we let

$$\underline{P} = (P_1, \dots, P_K), \qquad \underline{P}_R = (P_{R1}, \dots, P_{RK}), \qquad \underline{\beta} = (\beta_1, \dots, \beta_K) \tag{18}$$

and

$$\mathcal{G} = \Big\{ (\underline{P}, \underline{P}_R, \underline{\beta}) : \ \sum_{k=1}^{K} P_k \le P, \ \ \sum_{k=1}^{K} P_{Rk} \le P_R, \ \ 0 \le \beta_k \le 1, \ \text{ for } k = 1, \dots, K \Big\} \tag{19}$$

The max-min optimization problem in (12) can be written in the following compact form.

$$\boxed{\begin{aligned} &C = \max_{(\underline{P}, \underline{P}_R, \underline{\beta}) \in \mathcal{G}} \min\{\mathcal{R}_1(\underline{P}, \underline{P}_R, \underline{\beta}), \ \mathcal{R}_2(\underline{P}, \underline{P}_R, \underline{\beta})\} \\ &\text{where} \\ &\mathcal{R}_1(\underline{P}, \underline{P}_R, \underline{\beta}) \\ &\quad = \sum_{k=1}^{K} \mathcal{C}\left(\frac{P_k + \rho_{Rk} P_{Rk} + 2\sqrt{\beta_k \rho_{Rk} P_k P_{Rk}}}{\sigma_k^2} \right) \\ &\mathcal{R}_2(\underline{P}, \underline{P}_R, \underline{\beta}) = \sum_{k \in A} \mathcal{C}\left(\frac{\beta_k P_k}{\sigma_{Rk}^2} \right) + \sum_{k \in A^c} \mathcal{C}\left(\frac{\beta_k P_k}{\sigma_k^2} \right) \end{aligned}} \tag{20}$$

According to Proposition 1, the max-min rule that solves (20) may fall into the following three cases.

Case 1: $\alpha^* = 0$, and $(\underline{P}^{(0)}, \underline{P}_R^{(0)}, \underline{\beta}^{(0)})$ is a max-min rule, which needs to satisfy the condition

$$R_1(\underline{P}^{(0)}, \underline{P}_R^{(0)}, \underline{\beta}^{(0)}) \ge R_2(\underline{P}^{(0)}, \underline{P}_R^{(0)}, \underline{\beta}^{(0)}). \tag{21}$$

By definition, $(\underline{P}^{(0)}, \underline{P}_R^{(0)}, \underline{\beta}^{(0)})$ maximizes

$$R(0, \underline{P}, \underline{P}_R, \underline{\beta}) = R_2(\underline{P}, \underline{P}_R, \underline{\beta}). \tag{22}$$

It is readily seen that the following $\underline{\beta}^{(0)}$ is optimal:

$$\beta_k^{(0)} = \begin{cases} 1, & \text{if } P_k^{(0)} > 0; \\ \text{arbitrary} & \text{if } P_k^{(0)} = 0. \end{cases} \tag{23}$$

With $\underline{\beta}^{(0)}$ given in (23), $R_2(\underline{P}, \underline{P}_R, \underline{\beta})$ is a function of $\underline{P}$ only. Moreover, it is a convex function of $\underline{P}$. Then the Kuhn-Tucker condition (KKT condition) (see, e.g., [38, p. 314-315]) characterizes the necessary and sufficient condition that the optimal $\underline{P}^{(0)}$ needs to satisfy. The Lagrangian is given by

$$\mathcal{L} = \sum_{k \in A} \mathcal{C}\left(\frac{P_k}{\sigma_{Rk}^2} \right) + \sum_{k \in A^c} \mathcal{C}\left(\frac{P_k}{\sigma_k^2} \right) - \lambda \left(\sum_{k=1}^{K} P_k - P \right), \tag{24}$$

which implies the following KKT condition:

$$\begin{aligned} &\frac{\partial \mathcal{L}}{\partial P_k} = \frac{1}{2 \ln 2} \cdot \frac{1}{\sigma_{Rk}^2 + P_k} - \lambda \le 0, \\ &\qquad\qquad \text{with equality if } P_k > 0, \quad \text{if } k \in A; \\ &\frac{\partial \mathcal{L}}{\partial P_k} = \frac{1}{2 \ln 2} \cdot \frac{1}{\sigma_k^2 + P_k} - \lambda \le 0, \\ &\qquad\qquad \text{with equality if } P_k > 0, \quad \text{if } k \in A^c. \end{aligned} \tag{25}$$

Hence the optimal $P_k^{(0)}$ is given by

$$P_k^{(0)} = \begin{cases} \left(\dfrac{1}{2 \ln 2\lambda} - \sigma_{Rk}^2 \right)^+, & \text{if } k \in A; \\ \left(\dfrac{1}{2 \ln 2\lambda} - \sigma_k^2 \right)^+, & \text{if } k \in A^c \end{cases} \tag{26}$$

where λ is chosen to satisfy the power constraint $\sum_{k=1}^{K} P_k \leq P$. The function $(\cdot)^+$ is defined as

$$(x)^+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases} \tag{27}$$

For case 1 to happen, $(\underline{P}^{(0)}, \underline{P}_R^{(0)}, \underline{\beta}^{(0)})$ needs to satisfy the condition (21). To characterize the least power P_R needed for case 1 to happen, $\underline{P}_R^{(0)}$ needs to maximize $R_1(\underline{P}^{(0)}, \underline{P}_R, \underline{\beta}^{(0)})$ with $\underline{\beta}^{(0)}$ given in (23) and $\underline{P}^{(0)}$ given in (26), respectively. The optimal $\underline{P}_R^{(0)}$ can be obtained by the KKT condition via the following Lagrangian:

$$\mathcal{L} = \sum_{k=1}^{K} \mathcal{C}\left(\frac{P_k^{(0)} + \rho_{Rk} P_{Rk}}{\sigma_k^2}\right) - \mu\left(\sum_{k=1}^{K} P_{Rk} - P_R\right), \tag{28}$$

The KKT condition is given by

$$\frac{\partial \mathcal{L}}{\partial P_{Rk}} = \frac{1}{2\ln 2} \cdot \frac{\rho_{Rk}}{\sigma_k^2 + P_k^{(0)} + \rho_{Rk} P_{Rk}} - \mu \leq 0, \quad \text{with equality if } P_{Rk} > 0 \tag{29}$$

which implies

$$P_{Rk}^{(0)} = \left(\frac{1}{2\ln 2\mu} - \frac{P_k^{(0)}}{\rho_{Rk}} - \frac{\sigma_k^2}{\rho_{Rk}}\right)^+, \qquad \text{for } k = 1, \ldots, K \tag{30}$$

where μ is chosen to satisfy the power constraint $\sum_{k=1}^{K} P_{Rk} \leq P_R$.

Note that (30) also follows directly from the standard water-filling solution if we further derive (28) in the following form:

$$\begin{aligned} \mathcal{L} = &\sum_{k=1}^{K} \mathcal{C}\left(\frac{P_k^{(0)}}{\sigma_k^2}\right) + \sum_{k=1}^{K} \mathcal{C}\left(\frac{\rho_{Rk} P_{Rk}}{P_k^{(0)} + \sigma_k^2}\right) \\ &- \mu\left(\sum_{k=1}^{K} P_{Rk} - P_R\right). \end{aligned} \tag{31}$$

With $\underline{P}^{(0)}$, $\underline{P}_R^{(0)}$, and $\underline{\beta}^{(0)}$ given in (26), (30), and (23), respectively, condition (21) becomes

$$\begin{aligned} &\sum_{k=1}^{K} \mathcal{C}\left(\frac{P_k^{(0)} + \rho_{Rk} P_{Rk}^{(0)}}{\sigma_k^2}\right) \\ &\qquad \geq \sum_{k \in A} \mathcal{C}\left(\frac{P_k^{(0)}}{\sigma_{Rk}^2}\right) + \sum_{k \in A^c} \mathcal{C}\left(\frac{P_k^{(0)}}{\sigma_k^2}\right) \end{aligned} \tag{32}$$

This condition is equivalent to the threshold condition $P_R \geq P_{R,u}(P)$. The threshold $P_{R,u}(P)$ is a function of the source power constraint P, and is determined by the value of P_R that results in equality in (32).

Therefore, if case 1 occurs, the optimal source power allocation $\underline{P}^{(0)}$ has a *water-filling* form, and the optimal relay power allocation $\underline{P}_R^{(0)}$ also has a *water-filling* form with $P_k^{(0)} + \sigma_k^2$ as the equivalent noise levels. The optimal correlation parameter $\underline{\beta}_k^{(0)} = 1$ for $P_k^{(0)} > 0$, which indicates that coherent combining is not needed for this case.

Case 2: $\alpha^* = 1$, and $(\underline{P}^{(1)}, \underline{P}_R^{(1)}, \underline{\beta}^{(1)})$ is a max-min rule, which needs to satisfy the condition

$$R_1(\underline{P}^{(1)}, \underline{P}_R^{(1)}, \underline{\beta}^{(1)}) \leq R_2(\underline{P}^{(1)}, \underline{P}_R^{(1)}, \underline{\beta}^{(1)}). \tag{33}$$

By definition, $(\underline{P}^{(1)}, \underline{P}_R^{(1)}, \underline{\beta}^{(1)})$ maximizes

$$R(1, \underline{P}, \underline{P}_R, \underline{\beta}) = R_1(\underline{P}, \underline{P}_R, \underline{\beta}). \tag{34}$$

We note that

$$\beta_k^{(1)} = \begin{cases} 0, & \text{if } P_k^{(1)} > 0, \text{ and } P_{Rk}^{(1)} > 0; \\ \text{arbitrary}, & \text{otherwise}. \end{cases} \tag{35}$$

It can be shown that $R_1(\underline{P}, \underline{P}_R, \underline{\beta})$ is a convex function of $(\underline{P}, \underline{P}_R)$ for $\underline{\beta}^{(1)}$ given in (35). To derive the optimal $(\underline{P}^{(1)}, \underline{P}_R^{(1)})$ that maximizes $R_1(\underline{P}, \underline{P}_R, \underline{\beta}^{(1)})$, the Lagrangian can be written as

$$\begin{aligned} \mathcal{L} = &\sum_{k=1}^{K} \mathcal{C}\left(\frac{P_k + \rho_{Rk} P_{Rk} + 2\sqrt{\rho_{Rk} P_k P_{Rk}}}{\sigma_k^2}\right) \\ &- \lambda\left(\sum_{k=1}^{K} P_k - P\right) - \mu\left(\sum_{k=1}^{K} P_{Rk} - P_R\right) \end{aligned} \tag{36}$$

The optimal $(\underline{P}^{(1)}, \underline{P}_R^{(1)})$ needs to satisfy the following KKT condition:

$$\begin{aligned} &\frac{\partial \mathcal{L}}{\partial P_k} = \frac{1}{2\ln 2} \cdot \frac{\sqrt{P_k} + \sqrt{\rho_{Rk} P_{Rk}}}{\sigma_k^2 + (\sqrt{P_k} + \sqrt{\rho_{Rk} P_{Rk}})^2} \leq \lambda\sqrt{P_k}, \\ &\qquad\qquad \text{with equality if } P_k > 0; \\ &\frac{\partial \mathcal{L}}{\partial P_{Rk}} = \frac{1}{2\ln 2} \cdot \frac{\sqrt{P_k} + \sqrt{\rho_{Rk} P_{Rk}}}{\sigma_k^2 + (\sqrt{P_k} + \sqrt{\rho_{Rk} P_{Rk}})^2} \leq \mu\sqrt{\frac{P_{Rk}}{\rho_{Rk}}}, \\ &\qquad\qquad \text{with equality if } P_{Rk} > 0. \end{aligned} \tag{37}$$

From (37), it is clear that $P_k^{(1)} = 0 \iff P_{Rk}^{(1)} = 0$. According to (35), we have $\beta_k^{(1)} P_k^{(1)} = 0$ for $k = 1, \ldots, K$, which implies $R_2(\underline{P}^{(1)}, \underline{P}_R^{(1)}, \underline{\beta}^{(1)}) = 0$. Hence condition (33) cannot be satisfied. Therefore, case 2 never happens.

Case 3: $0 < \alpha^* < 1$, and $(\underline{P}^{(\alpha^*)}, \underline{P}_R^{(\alpha^*)}, \underline{\beta}^{(\alpha^*)})$ is a max-min rule, where α^* is determined by the following condition

$$R_1(\underline{P}^{(\alpha^*)}, \underline{P}_R^{(\alpha^*)}, \underline{\beta}^{(\alpha^*)}) = R_2(\underline{P}^{(\alpha^*)}, \underline{P}_R^{(\alpha^*)}, \underline{\beta}^{(\alpha^*)}). \tag{38}$$

We need to derive $(\underline{P}^{(\alpha^*)}, \underline{P}_R^{(\alpha^*)}, \underline{\beta}^{(\alpha^*)})$ that maximizes

$$R(\alpha^*, \underline{P}, \underline{P}_R, \underline{\beta}) = \alpha^* R_1(\underline{P}, \underline{P}_R, \underline{\beta}) + (1 - \alpha^*) R_2(\underline{P}, \underline{P}_R, \underline{\beta}) \tag{39}$$

for a fixed α^*. This optimization problem is not convex. Now the KKT condition provides only a necessary condition that the optimal $(\underline{P}^{(\alpha^*)}, \underline{P}_R^{(\alpha^*)}, \underline{\beta}^{(\alpha^*)})$ needs to satisfy. One can still perform a brute force search over those $(\underline{P}, \underline{P}_R, \underline{\beta})$ that satisfy the KKT condition to find the optimal $(\underline{P}^{(\alpha^*)}, \underline{P}_R^{(\alpha^*)}, \underline{\beta}^{(\alpha^*)})$. However, it may be too complex to implement such an optimal solution that involves designing correlated source and relay inputs and also involves allocating the source and relay powers jointly with the correlation parameter for each subchannel. Hence it may not be worth searching for the jointly optimal solution $(\underline{P}^{(\alpha^*)}, \underline{P}_R^{(\alpha^*)}, \underline{\beta}^{(\alpha^*)})$, except in case 1, where using independent source and relay inputs is optimal and the optimal power allocation $(\underline{P}^{(\alpha^*)}, \underline{P}_R^{(\alpha^*)})$ is simpler. It is hence more interesting to study the asynchronized case, where it is assumed that the source and relay nodes use independent inputs.

C. Optimal Resource Allocation for Gaussian Parallel Relay Channel: Asynchronized Case

In this subsection, we solve the max-min problem in (14). This problem is simpler than the max-min problem in (12), because the optimization is over the power allocation $(\underline{P}, \underline{P}_R)$ only, and does not involve the correlation parameters $\underline{\beta}$. This also makes the optimal solution easy to implement in practice. In the following, we fully characterize the optimal power allocation, which may take three possible structures.

We let

$$\mathcal{G} = \left\{ (\underline{P}, \underline{P}_R) : \sum_{k=1}^{K} P_k \le P, \; \sum_{k=1}^{K} P_{Rk} \le P_R \right\}, \tag{40}$$

and rewrite the max-min optimization problem in (14) in the following manner:

$$\boxed{\begin{aligned} & C = \max_{(\underline{P}, \underline{P}_R) \in \mathcal{G}} \min\{\mathcal{R}_1(\underline{P}, \underline{P}_R),\, \mathcal{R}_2(\underline{P}, \underline{P}_R)\} \\ & \text{where} \\ & \mathcal{R}_1(\underline{P}, \underline{P}_R) = \sum_{k=1}^{K} \mathcal{C}\left(\frac{P_k + \rho_{Rk} P_{Rk}}{\sigma_k^2}\right) \\ & \mathcal{R}_2(\underline{P}, \underline{P}_R) = \sum_{k \in A} \mathcal{C}\left(\frac{P_k}{\sigma_{Rk}^2}\right) + \sum_{k \in A^c} \mathcal{C}\left(\frac{P_k}{\sigma_k^2}\right) \end{aligned}} \tag{41}$$

We apply Proposition 1 to solve (41), and consider the following three cases.

Case 1: $\alpha^* = 0$, and $(\underline{P}^{(0)}, \underline{P}_R^{(0)})$ is a max-min rule, which needs to satisfy the condition

$$R_1(\underline{P}^{(0)}, \underline{P}_R^{(0)}) \ge R_2(\underline{P}^{(0)}, \underline{P}_R^{(0)}). \tag{42}$$

The optimal $(\underline{P}^{(0)}, \underline{P}_R^{(0)})$ can be derived following the steps that are similar to those in case 1 of the synchronized case, and is given by

$$\begin{aligned} P_k^{(0)} &= \begin{cases} \left(\dfrac{1}{2\ln 2\lambda} - \sigma_{Rk}^2\right)^+, & \text{if } k \in A \\ \left(\dfrac{1}{2\ln 2\lambda} - \sigma_k^2\right)^+, & \text{if } k \in A^c \end{cases} \\ P_{Rk}^{(0)} &= \left(\frac{1}{2\ln 2\mu} - \frac{P_k^{(0)}}{\rho_{Rk}} - \frac{\sigma_k^2}{\rho_{Rk}}\right)^+, \quad \text{for } k = 1, \ldots, K \end{aligned} \tag{43}$$

where λ and μ are chosen to satisfy the power constraints $\sum_{k=1}^{K} P_k \le P$ and $\sum_{k=1}^{K} P_{Rk} \le P_R$.

We refer to the optimal $(P_k^{(0)}, P_{Rk}^{(0)})$ in (43) as *two-level water-filling* for the following reason. The optimal $\underline{P}^{(0)}$ is first obtained via *water-filling* with respect to the noise levels σ_{Rk}^2 and σ_k^2. The optimal $\underline{P}_R^{(0)}$ is then obtained via *water-filling* with $P_k^{(0)} + \sigma_k^2$ as equivalent noise levels, where $\underline{P}^{(0)}$ is treated as an additional noise level.

With $(P_k^{(0)}, P_{Rk}^{(0)})$ given in (43), condition (42) becomes

$$\begin{aligned} &\sum_{k=1}^{K} \mathcal{C}\left(\frac{P_k^{(0)} + \rho_{Rk} P_{Rk}^{(0)}}{\sigma_k^2}\right) \\ &\qquad \ge \sum_{k \in A} \mathcal{C}\left(\frac{P_k^{(0)}}{\sigma_{Rk}^2}\right) + \sum_{k \in A^c} \mathcal{C}\left(\frac{P_k^{(0)}}{\sigma_k^2}\right) \end{aligned} \tag{44}$$

This condition is equivalent to the threshold condition $P_R \ge P_{R,u}(P)$, where the threshold $P_{R,u}(P)$ is determined by the value of P_R that results in equality in (44). The threshold $P_{R,u}(P)$ is clearly a function of the source power constraint P.

Case 2: $\alpha^* = 1$, and $(\underline{P}^{(1)}, \underline{P}_R^{(1)})$ is a max-min rule, which needs to satisfy the condition

$$R_1(\underline{P}^{(1)}, \underline{P}_R^{(1)}) \le R_2(\underline{P}^{(1)}, \underline{P}_R^{(1)}). \tag{45}$$

By definition, $(\underline{P}^{(1)}, \underline{P}_R^{(1)})$ maximizes

$$R(1, \underline{P}, \underline{P}_R) = R_1(\underline{P}, \underline{P}_R). \tag{46}$$

We first note that $R_1(\underline{P}, \underline{P}_R)$ is a convex function of $(\underline{P}, \underline{P}_R)$. The Lagrangian can be written as

$$\begin{aligned} \mathcal{L} = &\sum_{k=1}^{K} \mathcal{C}\left(\frac{P_k + \rho_{Rk} P_{Rk}}{\sigma_k^2}\right) - \lambda\left(\sum_{k=1}^{K} P_k - P\right) \\ &- \mu\left(\sum_{k=1}^{K} P_{Rk} - P_R\right) \end{aligned} \tag{47}$$

According to the KKT condition, $(\underline{P}^{(1)}, \underline{P}_R^{(1)})$ needs to satisfy

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial P_k} &= \frac{1}{2\ln 2} \cdot \frac{1}{\sigma_k^2 + P_k + \rho_{Rk} P_{Rk}} \le \lambda, \\ &\qquad \text{with equality if } P_k > 0; \\ \frac{\partial \mathcal{L}}{\partial P_{Rk}} &= \frac{1}{2\ln 2} \cdot \frac{1}{\sigma_k^2 + P_k + \rho_{Rk} P_{Rk}} \le \frac{\mu}{\rho_{Rk}}, \\ &\qquad \text{with equality if } P_{Rk} > 0 \end{aligned} \tag{48}$$

which implies

$$\begin{aligned} &\text{If } \lambda < \frac{\mu}{\rho_{Rk}}, \quad P_k = \left(\frac{1}{2\ln 2\lambda} - \sigma_k^2\right)^+, \quad P_{Rk} = 0, \\ &\text{If } \lambda > \frac{\mu}{\rho_{Rk}}, \quad P_k = 0, \quad P_{Rk} = \left(\frac{1}{2\ln 2\mu} - \frac{\sigma_k^2}{\rho_{Rk}}\right)^+, \\ &\text{If } \lambda = \frac{\mu}{\rho_{Rk}}, \quad P_k + \rho_{Rk} P_{Rk} = \left(\frac{1}{2\ln 2\lambda} - \sigma_k^2\right)^+ \end{aligned} \tag{49}$$

where λ and μ are chosen to satisfy the power constraints. In general, $\lambda \ne \frac{\mu}{\rho_{Rk}}$. The equation (49) implies an *orthogonal division water-filling* power allocation, i.e., for each subchannel, either the source node or the relay node allocates a positive amount of power. This power allocation is similar to the optimal power allocation for fading multiple access channels [39].

For case 2 to happen, $(\underline{P}^{(1)}, \underline{P}_R^{(1)})$ needs to satisfy the condition (45), i.e.,

$$\begin{aligned} &\sum_{k=1}^{K} \mathcal{C}\left(\frac{P_k^{(1)} + \rho_{Rk} P_{Rk}^{(1)}}{\sigma_k^2}\right) \\ &\qquad \le \sum_{k \in A} \mathcal{C}\left(\frac{P_k^{(1)}}{\sigma_{Rk}^2}\right) + \sum_{k \in A^c} \mathcal{C}\left(\frac{P_k^{(1)}}{\sigma_k^2}\right) \end{aligned} \tag{50}$$

This condition essentially requires that the relay power P_R is small compared to the source power P.

Case 3: $0 < \alpha^* < 1$, and $(\underline{P}^{(\alpha^*)}, \underline{P}_R^{(\alpha^*)})$ is a max-min rule, where α^* is determined by the condition

$$R_1(\underline{P}^{(\alpha^*)}, \underline{P}_R^{(\alpha^*)}) = R_2(\underline{P}^{(\alpha^*)}, \underline{P}_R^{(\alpha^*)}). \quad (51)$$

We first derive $(\underline{P}^{(\alpha^*)}, \underline{P}_R^{(\alpha^*)})$ that maximizes

$$R(\alpha^*, \underline{P}, \underline{P}_R) = \alpha^* R_1(\underline{P}, \underline{P}_R) + (1-\alpha^*) R_2(\underline{P}, \underline{P}_R). \quad (52)$$

for a given α^*, and α^* will be determined later.

The Lagrangian can be written as

$$\begin{aligned}\mathcal{L} =& \alpha^* \sum_{k=1}^{K} \mathcal{C}\left(\frac{P_k + \rho_{Rk} P_{Rk}}{\sigma_k^2}\right) + (1-\alpha^*) \sum_{k \in A} \mathcal{C}\left(\frac{P_k}{\sigma_{Rk}^2}\right) \\ & + (1-\alpha^*) \sum_{k \in A^c} \mathcal{C}\left(\frac{P_k}{\sigma_k^2}\right) - \lambda \left(\sum_{k=1}^{K} P_k - P\right) \\ & - \mu \left(\sum_{k=1}^{K} P_{Rk} - P_R\right)\end{aligned} \quad (53)$$

which implies the following KKT condition:

$$\text{For } k \in A, \quad \frac{\partial \mathcal{L}}{\partial P_k} = \frac{\alpha^*}{2\ln 2} \cdot \frac{1}{\sigma_k^2 + P_k + \rho_{Rk} P_{Rk}} + \frac{1-\alpha^*}{2\ln 2} \cdot \frac{1}{\sigma_{Rk}^2 + P_k} \le \lambda, \quad \text{with equality if } P_k > 0; \quad (54)$$

$$\text{For } k \in A^c, \quad \frac{\partial \mathcal{L}}{\partial P_k} = \frac{\alpha^*}{2\ln 2} \cdot \frac{1}{\sigma_k^2 + P_k + \rho_{Rk} P_{Rk}} + \frac{1-\alpha^*}{2\ln 2} \cdot \frac{1}{\sigma_k^2 + P_k} \le \lambda, \quad \text{with equality if } P_k > 0; \quad (55)$$

$$\text{For } k = 1, \ldots, K, \quad \frac{\partial \mathcal{L}}{\partial P_{Rk}} = \frac{\alpha^*}{2\ln 2} \cdot \frac{\rho_{Rk}}{\sigma_k^2 + P_k + \rho_{Rk} P_{Rk}} \le \mu, \quad \text{with equality if } P_{Rk} > 0. \quad (56)$$

The optimal $(\underline{P}^{(\alpha^*)}, \underline{P}_R^{(\alpha^*)})$ can be solved by an iterative algorithm. For a given $\underline{P}_R$, the value of $\underline{P}$ can be obtained by solving (54) and (55), and its components have the following form:

$$P_k = \begin{cases} \text{positive root } x \text{ of (58) if it exists, otherwise } 0, & \text{if } k \in A; \\ \text{positive root } x \text{ of (59) if it exists, otherwise } 0, & \text{if } k \in A^c \end{cases} \quad (57)$$

where the roots are determined by the following equations:

$$\frac{\alpha^*}{2\ln 2} \cdot \frac{1}{x + \rho_{Rk} P_{Rk} + \sigma_k^2} + \frac{1-\alpha^*}{2\ln 2} \cdot \frac{1}{x + \sigma_{Rk}^2} = \lambda \quad (58)$$

$$\frac{\alpha^*}{2\ln 2} \cdot \frac{1}{x + \rho_{Rk} P_{Rk} + \sigma_k^2} + \frac{1-\alpha^*}{2\ln 2} \cdot \frac{1}{x + \sigma_k^2} = \lambda \quad (59)$$

where λ is chosen to satisfy the power constraint $\sum_{k=1}^{K} P_k \le P$. For a given $\underline{P}$, the value of $\underline{P}_R$ can be obtained by using (56), and its components have the following form:

$$P_{Rk} = \left(\frac{\alpha^*}{2\ln 2\mu} - \frac{P_k}{\rho_{Rk}} - \frac{\sigma_k^2}{\rho_{Rk}}\right)^+, \quad \text{for } k = 1, \ldots, K \quad (60)$$

where μ is chosen to satisfy the power constraint $\sum_{k=1}^{K} P_{Rk} \le P_R$.

If we iteratively obtain $\underline{P}$ and $\underline{P}_R$ according to (57) and (60) with an initial $\underline{P}_R$, we show in the following that $(\underline{P}, \underline{P}_R)$ converges to an optimal $(\underline{P}^{(\alpha^*)}, \underline{P}_R^{(\alpha^*)})$. We refer to this optimal power allocation as the *iterative water-filling* power allocation. We finally need to search over $0 \le \alpha \le 1$ to find α^* that satisfies the equalizer condition (51).

Proof of Convergenece: We show that $(\underline{P}, \underline{P}_R)$ obtained iteratively according to (57) and (60) converges to an optimal $(\underline{P}^{(\alpha^*)}, \underline{P}_R^{(\alpha^*)})$. We first note that after each iteration the objective function (52) either increases or remains the same. We also note that the objective function is bounded from the above because of the power constraints at the source and relay nodes. Hence the objective function must converge. It is easy to check that for a given $\underline{P}$, the objective function is a strictly concave function of $\underline{P}_R$, and (60) yields the unique optimal $\underline{P}_R$. It is also true that for a fixed $\underline{P}_R$, (57) yields the unique optimal $\underline{P}$. Hence as the objective function converges, $(\underline{P}, \underline{P}_R)$ must converge. Moreover, $(\underline{P}, \underline{P}_R)$ converges to the solution of the KKT conditions, which are sufficient for $(\underline{P}, \underline{P}_R)$ to be optimal because the objective function is concave over $(\underline{P}, \underline{P}_R) \in \mathcal{G}$. ∎

We now summarize the optimal power allocation that solves (41) in the following theorem.

***Theorem** 4:* The optimal solution to (41), i.e., the optimal power allocation that achieves the asynchronized capacity (14) falls into the following three cases:

Case 1: The optimal $(\underline{P}, \underline{P}_R)$ takes the *two-level water-filling* form and is given by (43). This case happens if $P_R > P_{R,u}(P)$ where the threshold $P_{R,u}(P)$ is determined by equality of (44).

Case 2: The optimal $(\underline{P}, \underline{P}_R)$ takes the *orthogonal division water-filling* form and is given by (49). This case happens if condition (50) is satisfied.

Case 3: The optimal $(\underline{P}, \underline{P}_R)$ takes the *iterative water-filling* form and is obtained iteratively by (57) and (60).

IV. Fading Full-Duplex Relay Channels

In this section, we study the three-terminal relay channel [1], [2] in the context of wireless networks, where nodes communicate over time-varying wireless channels. We are interested in how the source and relay nodes should dynamically change their power with wireless channel variation to achieve optimal performance. Such wireless relay channels can be modelled by the fading full-duplex relay model, where each transmission link of a three-terminal relay channel [1], [2] is corrupted by a multiplicative fading gain coefficient in addition to an additive white Gaussian noise term (see Fig. 4). The fading relay channel is referred to as the *full-duplex* channel because the relay node is allowed to transmit and receive at the same time and in the same frequency band.

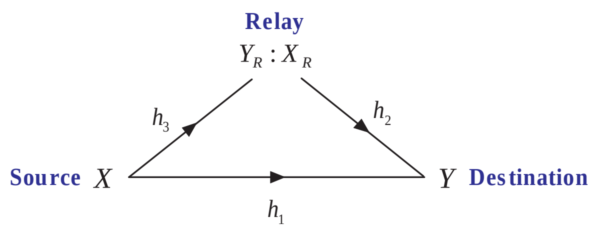

Fig. 4. Fading full-duplex relay channel

The channel input-output relationship at each symbol time can be written as

$$\begin{aligned} Y &= \sqrt{\rho_1}\, h_1 X + \sqrt{\rho_2}\, h_2 X_R + Z, \\ Y_R &= \sqrt{\rho_3}\, h_3 X + Z_R, \end{aligned} \tag{61}$$

where h_1, h_2, and h_3 are fading gain coefficients corresponding to the three transmission links, respectively, and are assumed to be independent complex proper random variables (not necessarily Gaussian) with variances normalized to 1. We further assume that the fading processes $\{h_{1i}\}$, $\{h_{2i}\}$, and $\{h_{3i}\}$ are stationary and ergodic over time, where i is the time index. In (61), the additive noise terms Z and Z_R are independent proper complex Gaussian random variables with variances also normalized to 1. The parameters ρ_1, ρ_2, and ρ_3 represent the link gain to noise ratios of the corresponding transmission links. The input symbol sequences $\{X_i\}$ and $\{X_{Ri}\}$ are subject to separate average power constraints P and P_R, respectively, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}|X_i|^2 \le P, \qquad \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}|X_{Ri}|^2 \le P_R. \tag{62}$$

***Remark** 4:* The fading relay channel is a special case of the parallel relay channel with each subchannel corresponding to one fading state realization. In particular, for a given fading state the fading relay channel is a Gaussian relay channel by (61). However, since this Gaussian channel is not physically degraded, the fading relay channel is not a Gaussian parallel relay channel with degraded subchannels that is considered in Sections II and III, where physically degradedness is assumed for each subchannel.

We assume that the transmitter and the receiver know the channel state information instantly. Hence the source and relay nodes can allocate their transmitted signal powers according to the channel state information to achieve the best performance. Our goal is to study the optimal power allocation at the source and relay nodes. As in Section III-C, we are interested in the asynchronized case for the fading full-duplex relay channel, where the source and relay nodes are required to use independent inputs. The main reason is because this simplifies the transmitter design, and is more practical in distributed networks, where nodes need to construct their codebooks independently.

For notational convenience, we collect the fading coefficients h_1, h_2 and h_3 in a vector $\underline{h} := (h_1, h_2, h_3)$. We define a set $A := \{\underline{h} : \rho_3|h_3|^2 > \rho_1|h_1|^2\}$, which contains all the fading states $\underline{h}$ with the source-to-relay link being better than the source-to-destination link. The complement of the set A is $A^c := \{\underline{h} : \rho_3|h_3|^2 \le \rho_1|h_1|^2\}$. We define a set $\mathcal{G}$ that contains all power allocation functions that satisfy the power constraints, i.e.,

$$\mathcal{G} = \{(P(\underline{h}), P_R(\underline{h})) : \mathrm{E}[P(\underline{h})] \le P, \;\; \mathrm{E}[P_R(\underline{h})] \le P_R\}. \tag{63}$$

The following lower and upper bounds on the asynchronized capacity of the fading full-duplex relay channel were given in [9].

***Lemma** 1:* ([9]) For the fading full-duplex relay channel, lower and upper bounds on the asynchronized capacity are given by

$$\begin{aligned} C_{\texttt{low}} = &\max_{(P(\underline{h}), P_R(\underline{h}))\in\mathcal{G}} \\ &\min\Big\{ 2\mathrm{E}\left[\mathcal{C}\Big(P(\underline{h})\rho_1|h_1|^2 + P_R(\underline{h})\rho_2|h_2|^2\Big)\right], \\ &\qquad 2\mathrm{E}_A\left[\mathcal{C}\Big(P(\underline{h})\rho_3|h_3|^2\Big)\right] + 2\mathrm{E}_{A^c}\left[\mathcal{C}\Big(P(\underline{h})\rho_1|h_1|^2\Big)\right] \Big\} \end{aligned} \tag{64}$$

$$\begin{aligned} C_{\texttt{up}} = &\max_{(P(\underline{h}), P_R(\underline{h}))\in\mathcal{G}} \\ &\min\Big\{ 2\mathrm{E}\left[\mathcal{C}\Big(P(\underline{h})\rho_1|h_1|^2 + P_R(\underline{h})\rho_2|h_2|^2\Big)\right], \\ &\qquad 2\mathrm{E}\left[\mathcal{C}\Big(P(\underline{h})(\rho_1|h_1|^2 + \rho_3|h_3|^2)\Big)\right] \Big\} \end{aligned} \tag{65}$$

Note that the rates in the lower bound of Lemma 1 are the same as the achievable rates in Corollary 1.

The optimal power allocation that maximizes the lower bound (64) and the upper bound (65) were obtained in [9] under a sum power constraint, i.e., the source and relay nodes are subject to a total power constraint. In this paper, we assume that the source and relay nodes are subject to the separate power constraints as given in (62) and (63), and derive the optimal power allocations that maximize the bounds (64) and (65), respectively. We also characterize the conditions where the lower and upper bounds match and determine the capacity of the channel.

Using the same technique as in Section III-C, we characterize the optimal power allocation that maximizes the lower bound (64) of the fading relay channel. This optimal power allocation takes the same three structures as those given in Section III-C, and is summarized in the following for the sake of completeness.

Optimal power allocation that maximizes the lower bound (64)**:**

Case 1 (two-level water-filling): If $P_R \ge P_{R,u}(P)$, the optimal $(P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h}))$ is given by

$$P^{(0)}(\underline{h}) = \begin{cases} \left(\frac{1}{\lambda \ln 2} - \frac{1}{\rho_3 |h_3|^2}\right)^+, & \text{if } \underline{h} \in A, \\ \left(\frac{1}{\lambda \ln 2} - \frac{1}{\rho_1 |h_1|^2}\right)^+, & \text{if } \underline{h} \in A^c \end{cases} \tag{66}$$

where λ is chosen to satisfy the power constraint $\mathrm{E}[P(\underline{h})] = P$.

$$P_R^{(0)}(\underline{h}) = \left(\frac{1}{\mu \ln 2} - \frac{1 + \rho_1 |h_1|^2 P^{(0)}(\underline{h})}{\rho_2 |h_2|^2}\right)^+ \tag{67}$$

where μ is chosen to satisfy the power constraint $\mathrm{E}[P_R(\underline{h})] = P_R$.
The threshold $P_{R,u}(P)$ as a function of the source power P can be solved using the following equation

$$\begin{aligned} &\mathrm{E}\left[\mathcal{C}\Big(P^{(0)}(\underline{h})\rho_1|h_1|^2 + P_R^{(0)}(\underline{h})\rho_2|h_2|^2\Big)\right] \\ &= \mathrm{E}_A\left[\mathcal{C}\Big(P^{(0)}(\underline{h})\rho_3|h_3|^2\Big)\right] \\ &\quad + \mathrm{E}_{A^c}\left[\mathcal{C}\Big(P^{(0)}(\underline{h})\rho_1|h_1|^2\Big)\right] \end{aligned} \tag{68}$$

Case 2 (orthogonal division water-filling): The optimal $(P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h}))$ is given by

$$\begin{aligned} &\text{If } \frac{\lambda}{\rho_1|h_1|^2} \le \frac{\mu}{\rho_2|h_2|^2}, \\ &\quad P^{(1)}(\underline{h}) = \left(\frac{1}{\lambda \ln 2} - \frac{1}{\rho_1|h_1|^2}\right)^+, \qquad P_R^{(1)}(\underline{h}) = 0; \\ &\text{If } \frac{\lambda}{\rho_1|h_1|^2} > \frac{\mu}{\rho_2|h_2|^2}, \\ &\quad P^{(1)}(\underline{h}) = 0, \qquad P_R^{(1)}(\underline{h}) = \left(\frac{1}{\mu \ln 2} - \frac{1}{\rho_2|h_2|^2}\right)^+ \end{aligned} \tag{69}$$

where λ and μ are chosen to satisfy the power constraints $\mathrm{E}[P(\underline{h})] = P$ and $\mathrm{E}[P_R(\underline{h})] = P_R$.
Case 2 happens if the following condition is satisfied:

$$\begin{aligned} &\mathrm{E}\left[\mathcal{C}\Big(P^{(1)}(\underline{h})\rho_1|h_1|^2 + P_R^{(1)}(\underline{h})\rho_2|h_2|^2\Big)\right], \\ &\le \mathrm{E}_A\left[\mathcal{C}\Big(P^{(1)}(\underline{h})\rho_3|h_3|^2\Big)\right] \\ &\quad + \mathrm{E}_{A^c}\left[\mathcal{C}\Big(P^{(1)}(\underline{h})\rho_1|h_1|^2\Big)\right] \end{aligned} \tag{70}$$

Case 3 (iterative water-filling): The optimal $(P^{(\alpha^*)}, P_R^{(\alpha^*)})$ can be obtained by the following iterative algorithm. For a given $P_R(\underline{h})$, the value of $P(\underline{h})$ is given by

$$P(\underline{h}) = \begin{cases} \text{positive root } x \text{ of (72) if it exists, otherwise } 0, \\ \qquad\qquad \text{if } \underline{h} \in A; \\ \text{positive root } x \text{ of (73) if it exists, otherwise } 0, \\ \qquad\qquad \text{if } \underline{h} \in A^c \end{cases} \tag{71}$$

where the roots are determined by the following equations:

$$\begin{aligned} &\frac{\alpha^*}{\ln 2} \cdot \frac{\rho_1|h_1|^2}{\rho_1|h_1|^2 x + P_R(\underline{h})\rho_2|h_2|^2 + 1} \\ &\quad + \frac{1-\alpha^*}{\ln 2} \cdot \frac{\rho_3|h_3|^2}{1 + \rho_3|h_3|^2 x} = \lambda \end{aligned} \tag{72}$$

$$\begin{aligned} &\frac{\alpha^*}{\ln 2} \cdot \frac{\rho_1|h_1|^2}{\rho_1|h_1|^2 x + P_R(\underline{h})\rho_2|h_2|^2 + 1} \\ &\quad + \frac{1-\alpha^*}{\ln 2} \cdot \frac{\rho_1|h_1|^2}{1 + \rho_1|h_1|^2 x} = \lambda \end{aligned} \tag{73}$$

where λ is chosen to satisfy the power constraint $\mathrm{E}[P(\underline{h})] = P$. For a given $P(\underline{h})$, the value of $P_R(\underline{h})$ is given by

$$P_R(\underline{h}) = \left(\frac{1}{\mu \ln 2} - \frac{1 + \rho_1|h_1|^2 P(\underline{h})}{\rho_2|h_2|^2}\right)^+ \tag{74}$$

where μ is chosen to satisfy the power constraint $\mathrm{E}[P_R(\underline{h})] = P_R$.
The power allocation $(P(\underline{h}), P_R(\underline{h}))$ obtained iteratively from (71) and (74) with an initial $P_R(\underline{h})$ converges to an optimal $(P^{(\alpha^*)}(\underline{h}), P_R^{(\alpha^*)}(\underline{h}))$. The parameter α^* is determined by the following equalizer condition:

$$\begin{aligned} &\mathrm{E}\left[\mathcal{C}\Big(P(\underline{h})\rho_1|h_1|^2 + P_R(\underline{h})\rho_2|h_2|^2\Big)\right] \\ &= \mathrm{E}_A\left[\mathcal{C}\Big(P(\underline{h})\rho_3|h_3|^2\Big)\right] \\ &\quad + \mathrm{E}_{A^c}\left[\mathcal{C}\Big(P(\underline{h})\rho_1|h_1|^2\Big)\right] \end{aligned} \tag{75}$$

The optimal power allocation for the upper bound (65) can be derived in a similar fashion but it is omitted here since this optimization does not have an operation meaning. In general, the upper and lower bounds do not match. In the following theorem, we characterize the condition where the two bounds match and establish the asynchronized capacity.

***Theorem** 5:* For the fading full-duplex relay channel, if the channel statistics and the power constraints at the source and relay nodes satisfy the condition (70), then the asynchronized capacity is given by

$$C = 2\mathrm{E}\left[\mathcal{C}\Big(P^{(1)}(\underline{h})\rho_1|h_1|^2 + P_R^{(1)}(\underline{h})\rho_2|h_2|^2\Big)\right] \tag{76}$$

where the capacity achieving power allocation $(P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h}))$ takes the orthogonal (time) division water-filling form given in (69).

Proof: The lower bound (64) and the upper bound (65) have one term in common inside the "min" in their expression. If condition (70) is satisfied, case 2 happens when solving the max-min problem for the lower bound (64). In this case, the common term of the bounds is optimized by the power allocation in (69) and determines both bounds that result in $C_{\mathtt{up}} = C_{\mathtt{low}}$. This common value is thus the asynchronized capacity. ■

The condition given in Theorem 5 essentially requires that the relay power P_R be small compared to the source power P. In this case, the optimal scheme is to maximize the rate at which the source and relay nodes can transmit to the destination node. The optimal scheme is to let the source and relay nodes have a time division access of the channel. For a given channel state realization, the node with a better channel to the destination node is allowed to transmit. This is similar to the optimal power allocation scheme for the fading multiple access channel studied in [39].

V. Fading Half-Duplex Relay Channels

In this section, we study a fading half-duplex relay channel model, where the source node transmits to the relay and destination nodes in one channel (channel 1), and the relay node transmits to the destination node in an orthogonal channel (channel 2). We introduce a parameter θ to represent the channel resource (time and bandwidth) allocation between the two orthogonal channels. We draw this fading half-duplex relay channel model in Fig. 5 with the solid and dashed lines indicating the transmission links of channels 1 and 2, respectively.

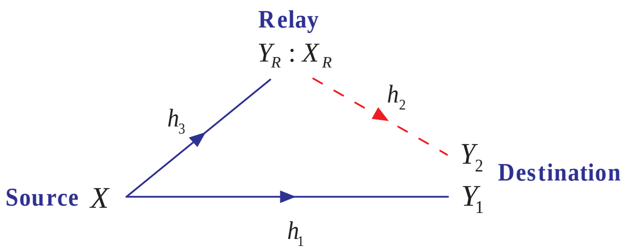

Fig. 5. Fading half-duplex relay model

The input-output relationship for the fading half-duplex relay channel is given by

$$\begin{aligned} Y_1 &= \sqrt{\rho_1}\, h_1 X + Z_1, \\ Y_2 &= \sqrt{\rho_2}\, h_2 X_R + Z_2, \\ Y_R &= \sqrt{\rho_3}\, h_3 X + Z_R, \end{aligned} \tag{77}$$

where h_1, h_2, and h_3 are fading gain coefficients that satisfy the same assumptions as for the fading full-duplex relay channel in Section IV. The additive noise terms Z_1, Z_2, and Z_R are independent proper complex Gaussian random variables with variances normalized to 1. The parameters ρ_1, ρ_2, and ρ_3 represent the link gain to noise ratios of the corresponding transmission links. The source and relay input sequences are subject to the same power constraints (62) as in the fading full-duplex relay channel.

As in the full-duplex case, the channel state information is assumed to be known at both the transmitter and the receiver. Hence the source and relay nodes can allocate their powers adaptively according to the instantaneous channel state information. The half-duplex channel has an additional channel resource allocation parameter θ that may also be optimized. Our goal is to find the jointly optimal θ and power allocation for the source and relay nodes that achieve the best rate. We also derive an upper bound on the capacity, which helps to establish capacity theorems for some special cases.

We study three scenarios. In Scenario I, we fix $\theta = \frac{1}{2}$, and only consider the maximization of the achievable rate over the power allocation at the source and relay nodes. In Scenario II, we restrict θ to be same for all channel states, and jointly optimize the achievable rate over this single parameter θ and power allocation. In Scenario III, which is the most general scenario, we further allow θ to change with channel state realizations, and optimize the achievable rate over all possible channel resource and power allocations.

A. Scenario I: Fixed $\theta = 1/2$

In this subsection, we study Scenario I, where the two orthogonal channels share the channel resource equally, i.e., the channel resource allocation parameter $\theta = 1/2$. We use this scenario to demonstrate the three basic structures of the optimal power allocation, which take simple forms. The optimal power allocation can be implemented in a distributed manner at the source and relay nodes, because each node needs to know only the channel state information of the links over which it transmits.

In the following, we first give an achievable rate for this channel, and then find an optimal power allocation that maximizes this achievable rate.

***Proposition** 2:* An achievable rate for the fading half-duplex relay channel Scenario I is given by

$$\begin{aligned} C_{\texttt{low}} = & \max_{(P(\underline{h}),P_R(\underline{h}))\in\mathcal{G}} \\ & \min\Bigg\{ \mathrm{E}\left[\mathcal{C}\left(2P(\underline{h})\rho_1|h_1|^2\right)+\mathcal{C}\left(2P_R(\underline{h})\rho_2|h_2|^2\right)\right], \\ & \qquad \mathrm{E}_A\left[\mathcal{C}\left(2P(\underline{h})\rho_3|h_3|^2\right)\right]+\mathrm{E}_{A^c}\left[\mathcal{C}\left(2P(\underline{h})\rho_1|h_1|^2\right)\right]\Bigg\} \end{aligned} \tag{78}$$

Proposition 2 follows easily by using steps that are similar to the achievability proof for Theorems 2 and 3 and by using the channel definition (77).

The optimal power allocation that maximizes $C_{\texttt{low}}$ in (78) can be derived by applying Proposition 1, and are given in the following three cases. The details of the proof are relegated to Appendix I.

Optimal power allocation that maximizes the lower bound (78)**:**

Case 1: If $P_R \geq P_{R,u}(P)$, the optimal $(P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h}))$ is given by

$$P^{(0)}(\underline{h}) = \begin{cases} \frac{1}{2}\left(\frac{1}{\lambda\ln 2}-\frac{1}{\rho_3|h_3|^2}\right)^+, & \text{if } \underline{h}\in A, \\ \frac{1}{2}\left(\frac{1}{\lambda\ln 2}-\frac{1}{\rho_1|h_1|^2}\right)^+, & \text{if } \underline{h}\in A^c \end{cases} \tag{79}$$

where λ is chosen to satisfy the power constraint $\mathrm{E}[P(\underline{h})] = P$.

$$P_R^{(0)}(\underline{h}) = \frac{1}{2}\left(\frac{1}{\mu\ln 2}-\frac{1}{\rho_2|h_2|^2}\right)^+ \tag{80}$$

where μ is chosen to satisfy the power constraint $\mathrm{E}[P_R(\underline{h})] = P_R$.
The threshold $P_{R,u}(P)$ as a function of the source power P can be solved using the following equation:

$$\begin{aligned} & \mathrm{E}\left[\mathcal{C}\left(2P_R^{(0)}(\underline{h})\rho_2|h_2|^2\right)\right] \\ & = \mathrm{E}_A\left[\mathcal{C}\left(2P^{(0)}(\underline{h})\rho_3|h_3|^2\right)-\mathcal{C}\left(2P^{(0)}(\underline{h})\rho_1|h_1|^2\right)\right]. \end{aligned} \tag{81}$$

Case 2: If $P_R \leq P_{R,l}(P)$, the optimal $(P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h}))$ is given by

$$P^{(1)}(\underline{h}) = \frac{1}{2}\left(\frac{1}{\lambda\ln 2}-\frac{1}{\rho_1|h_1|^2}\right)^+ \tag{82}$$

$$P_R^{(1)}(\underline{h}) = \frac{1}{2}\left(\frac{1}{\mu\ln 2}-\frac{1}{\rho_2|h_2|^2}\right)^+ \tag{83}$$

where λ and μ are chosen to satisfy the power constraints $\mathrm{E}[P(\underline{h})] = P$ and $\mathrm{E}[P_R(\underline{h})] = P_R$.

The threshold $P_{R,l}(P)$ can be solved using the following equation:

$$\begin{aligned}&\mathrm{E}\left[\mathcal{C}\left(2P_R^{(1)}(\underline{h})\rho_2|h_2|^2\right)\right]\\&=\mathrm{E}_A\left[\mathcal{C}\left(2P^{(1)}(\underline{h})\rho_3|h_3|^2\right)-\mathcal{C}\left(2P^{(1)}(\underline{h})\rho_1|h_1|^2\right)\right].\end{aligned} \tag{84}$$

Case 3: If $P_{R,l}(P)\leq P_R\leq P_{R,u}(P)$, the optimal $(P^{(\alpha^*)}(\underline{h}),P_R^{(\alpha^*)}(\underline{h}))$ is given by

$$P^{(\alpha^*)}(\underline{h})=\begin{cases}\text{positive root } x \text{ of (86) if it exists, otherwise } 0,\\ \qquad\qquad\qquad\qquad \text{if } \underline{h}\in A;\\ \frac{1}{2}\left(\frac{1}{\lambda\ln 2}-\frac{1}{\rho_1|h_1|^2}\right)^+, \qquad \text{if } \underline{h}\in A^c\end{cases} \tag{85}$$

where the root x is determined by the following equation

$$\frac{\alpha^*}{2\ln 2}\cdot\frac{1}{\frac{1}{2\rho_1|h_1|^2}+x}+\frac{1-\alpha^*}{2\ln 2}\cdot\frac{1}{\frac{1}{2\rho_3|h_3|^2}+x}-\lambda=0. \tag{86}$$

$$P_R^{(\alpha^*)}(\underline{h})=\frac{1}{2}\left(\frac{\alpha^*}{\mu\ln 2}-\frac{1}{\rho_2|h_2|^2}\right)^+ \tag{87}$$

The parameters λ and μ are chosen to satisfy the power constraints given in (63). The parameter α^* is determined by the following condition:

$$\begin{aligned}&\mathrm{E}\left[\mathcal{C}\left(2P^{(\alpha^*)}(\underline{h})\rho_1|h_1|^2\right)+\mathcal{C}\left(2P_R^{(\alpha^*)}(\underline{h})\rho_2|h_2|^2\right)\right]\\&=\mathrm{E}_A\left[\mathcal{C}\left(2P^{(\alpha^*)}(\underline{h})\rho_3|h_3|^2\right)\right]\\&\quad+\mathrm{E}_{A^c}\left[\mathcal{C}\left(2P^{(\alpha^*)}(\underline{h})\rho_1|h_1|^2\right)\right].\end{aligned} \tag{88}$$

It can be seen that in all cases the optimal power allocation $P_R(\underline{h})$ for the relay node depends only on the fading gain h_2 of the relay-to-destination link and it is always a water-filling solution. However, the optimal power allocation $P(\underline{h})$ for the source node in general depends on the fading gains h_1 and h_3 corresponding to two links (source-to-destination and source-to-relay), and it is not a water-filling solution in general. Only in cases where P_R is large or small compared to P, i.e., where $P_R\geq P_{R,u}(P)$ or $P_R\leq P_{R,l}(P)$, the optimal $P(\underline{h})$ depends only on the fading gain of one link and it reduces to a water-filling solution. This is intuitive because when P_R is small compared to P, we should make the multiple access transmission from the source and relay nodes to the destination node as strong as possible, and hence the power allocation at the source node should be based on the fading gain h_1 of the source-to-destination link. When P_R is large compared to P, we should transmit as much information as possible from the source node to the relay node, and hence the power allocation at the source node should be based on the fading gain h_3 of the source-to-relay link.

We now provide numerical results for a Rayleigh fading half-duplex relay channel. We assume that the fading coefficients h_1, h_2 and h_3 are independent, zero-mean, unit variance, proper complex Gaussian random variables (i.e., the amplitudes $|h_1|$, $|h_2|$ and $|h_3|$ have a Rayleigh distribution). We further assume $\rho_1=0.1$, $\rho_2=0.1$, and $\rho_3=1$. We assume the power constraint at the source node is $P=3$ dB. This corresponds to the practical environment where the relay node is close to the source node. In Fig. 6, we plot the achievable rates for Scenario I optimized over power allocation $(P(\underline{h}),P_R(\underline{h}))$. We also indicate the corresponding max-min optimization cases to achieve these rates. It can be seen that the achievable rate increases as the relay power increases in cases 2 and 3, and saturates when the relay power falls into case 1. This is because in case 1 the relay power is large enough to forward all the information decoded at the relay node to the destination node, and the achievable rate is limited by the capacity of the source-to-relay link.

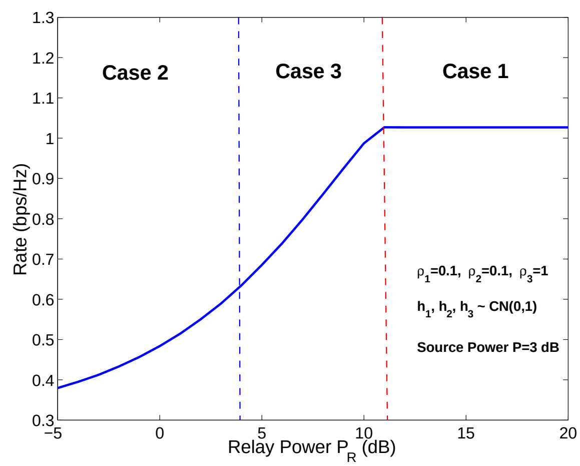

Fig. 6. Optimal achievable rates in Scenario I

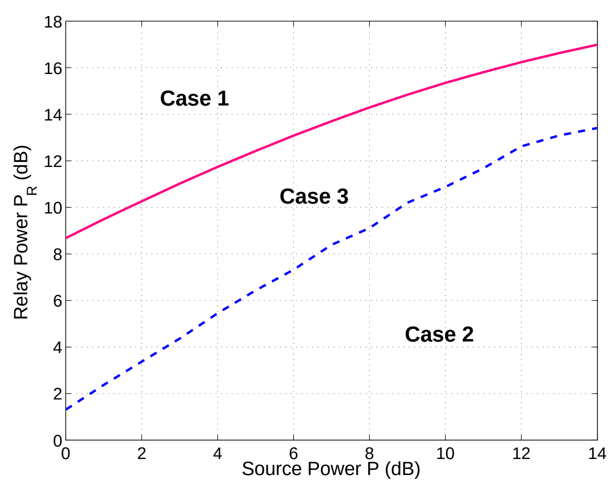

Fig. 7. Ranges of source and relay powers with corresponding max-min optimization cases in Scenario I

In Fig. 7, we plot the ranges of the source and relay powers with their corresponding max-min optimization cases. The solid line in the graph divides cases 1 and 3, and corresponds to the threshold function $P_{R,u}(P)$. The dashed line divides cases 2 and 3, and corresponds to the threshold function $P_{R,l}(P)$. It is clear from the graph that when the relay power is small compared to the source power, the optimal power allocation falls into case 2, and when the relay power is large compared to the source power, the optimal power allocation falls into case 1. Since the achievable rate (based on the decode-and-forward scheme) saturates in case 1, it is not useful to increase the relay power beyond the solid line in Fig. 7 if the decode-and-forward scheme is adopted. Hence the solid line $P_{R,u}$ defines the relay powers that provide the best decode-and-forward rates under Scenario I for the corresponding source powers.

B. Scenario II: Same θ for All Channel States

In Scenario I, θ is fixed at $1/2$; i.e., the channel resource of time and bandwidth is equally allocated for the two orthogonal channels. Such equal channel resource allocation may not be optimal, and therefore we consider Scenario II, where the channel resource allocation parameter θ needs to be optimized jointly with power allocation. We also assume that θ is the same for all channel states to make the system design simple. As in Scenario I, the optimal solution of Scenario II can also be implemented in a distributed manner at the source and relay nodes. This is because the optimal θ depends only on the channel statistics, not on the channel state realizations. The power allocation at each node depends only on the channel state of the links over which the node transmits.

We first give an achievable rate (lower bound on the capacity) and a cut-set upper bound on the capacity. We then study the joint channel resource and power allocations that optimize these bounds. We also characterize the condition when the two bounds match and establish the capacity.

***Proposition** 3:* An achievable rate for the fading half-duplex relay channel scenario II is given by

$$\begin{aligned} C_{\mathtt{low}} = & \max_{0\le\theta\le1,(P(\underline{h}),P_R(\underline{h}))\in\mathcal{G}} \\ & \min\left\{ \mathrm{E}\left[2\theta\mathcal{C}\left(\frac{P(\underline{h})\rho_1|h_1|^2}{\theta}\right)+2\bar{\theta}\mathcal{C}\left(\frac{P_R(\underline{h})\rho_2|h_2|^2}{\bar{\theta}}\right)\right], \right. \\ & \left. \mathrm{E}_A\left[2\theta\mathcal{C}\left(\frac{P(\underline{h})\rho_3|h_3|^2}{\theta}\right)\right]+\mathrm{E}_{A^c}\left[2\theta\mathcal{C}\left(\frac{P(\underline{h})\rho_1|h_1|^2}{\theta}\right)\right]\right\} \end{aligned} \tag{89}$$

where $\bar{\theta}=1-\theta$. An upper bound on the capacity is given by

$$\begin{aligned} C_{\mathtt{up}} = & \max_{0\le\theta\le1,(P(\underline{h}),P_R(\underline{h}))\in\mathcal{G}} \\ & \min\left\{ \mathrm{E}\left[2\theta\mathcal{C}\left(\frac{P(\underline{h})\rho_1|h_1|^2}{\theta}\right)+2\bar{\theta}\mathcal{C}\left(\frac{P_R(\underline{h})\rho_2|h_2|^2}{\bar{\theta}}\right)\right], \right. \\ & \left. \mathrm{E}\left[2\theta\mathcal{C}\left(\frac{P(\underline{h})(\rho_3|h_3|^2+\rho_1|h_1|^2)}{\theta}\right)\right]\right\}. \end{aligned} \tag{90}$$

We provide the optimal channel resource and power allocations $(\theta, P(\underline{h}), P_R(\underline{h}))$ that solve (89) in the following. The proof of optimality is relegated to Appendix II.

Optimal resource allocation that maximizes the lower bound (89):

Case 1: This case is included in case 3 with the parameter α being allowed to take the value of 0.

Case 2: The optimal $(\theta^{(1)}, P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h}))$ can be obtained by the following iterative algorithm. For a given θ, the power allocation $(P(\underline{h}), P_R(\underline{h}))$ are given by

$$P(\underline{h})=\theta\left(\frac{1}{\lambda\ln 2}-\frac{1}{\rho_1|h_1|^2}\right)^+ \tag{91}$$

$$P_R(\underline{h})=\bar{\theta}\left(\frac{1}{\mu\ln 2}-\frac{1}{\rho_2|h_2|^2}\right)^+, \tag{92}$$

where λ and μ are chosen to satisfy the power constraints. For a given $(P(\underline{h}), P_R(\underline{h}))$, the value of θ is given by the root of the following equation:

$$\begin{aligned} & 2\mathrm{E}\left[\mathcal{C}\left(\frac{P(\underline{h})\rho_1|h_1|^2}{\theta}\right)-\mathcal{C}\left(\frac{P_R(\underline{h})\rho_2|h_2|^2}{\bar{\theta}}\right)\right] \\ & \quad =\frac{1}{\ln 2}\mathrm{E}\left[\frac{P(\underline{h})\rho_1|h_1|^2}{\theta+P(\underline{h})\rho_1|h_1|^2}-\frac{P_R(\underline{h})\rho_2|h_2|^2}{\bar{\theta}+P_R(\underline{h})\rho_2|h_2|^2}\right]. \end{aligned} \tag{93}$$

The resource allocation $(\theta, P(\underline{h}), P_R(\underline{h}))$ obtained iteratively from (91),(92), and (93) converges to the optimal $(\theta^{(1)}, P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h}))$.
This case happens if the following condition is satisfied

$$\begin{aligned} & \bar{\theta}^{(1)}\mathrm{E}\left[\mathcal{C}\left(\frac{P_R^{(1)}(\underline{h})\rho_2|h_2|^2}{\bar{\theta}^{(1)}}\right)\right] \\ & \qquad \le \theta^{(1)}\mathrm{E}_A\left[\mathcal{C}\left(\frac{P^{(1)}(\underline{h})\rho_3|h_3|^2}{\theta^{(1)}}\right)\right. \\ & \qquad\qquad \left. -\mathcal{C}\left(\frac{P^{(1)}(\underline{h})\rho_1|h_1|^2}{\theta^{(1)}}\right)\right]. \end{aligned} \tag{94}$$

Case 3: The optimal $(\theta^{(\alpha^*)}, P^{(\alpha^*)}(\underline{h}), P_R^{(\alpha^*)}(\underline{h}))$ can be obtained by the following iterative algorithm. For a given θ, the power allocation $(P(\underline{h}), P_R(\underline{h}))$ is given by

$$P(\underline{h})=\begin{cases} \text{positive root } x \text{ of (96) if it exists, otherwise } 0, & \text{if } \underline{h}\in A, \\ \theta\left(\frac{1}{\lambda\ln 2}-\frac{1}{\rho_1|h_1|^2}\right)^+ & \text{if } \underline{h}\in A^c; \end{cases} \tag{95}$$

where the root x is determined by the following equation:

$$\frac{\alpha^*\theta}{\ln 2}\cdot\frac{1}{\frac{\theta}{\rho_1|h_1|^2}+x}+\frac{(1-\alpha^*)\theta}{\ln 2}\cdot\frac{1}{\frac{\theta}{\rho_3|h_3|^2}+x}-\lambda=0. \tag{96}$$

$$P_R(\underline{h})=\bar{\theta}\left(\frac{\alpha^*}{\mu\ln 2}-\frac{1}{\rho_2|h_2|^2}\right)^+. \tag{97}$$

The parameters λ and μ are chosen to satisfy the power constraints.
For a given $(P(\underline{h}), P_R(\underline{h}))$, the value of θ is the root of the following equation:

$$\begin{aligned} & 2\alpha^*\mathrm{E}_A\left[\mathcal{C}\left(\tfrac{P(\underline{h})\rho_1|h_1|^2}{\theta}\right)\right]-\tfrac{\alpha^*}{\ln 2}\mathrm{E}_A\left[\tfrac{P(\underline{h})\rho_1|h_1|^2}{\theta+P(\underline{h})\rho_1|h_1|^2}\right] \\ & -2\alpha^*\mathrm{E}\left[\mathcal{C}\left(\tfrac{P_R(\underline{h})\rho_2|h_2|^2}{\theta}\right)\right]+\tfrac{\alpha^*}{\ln 2}\mathrm{E}\left[\tfrac{P_R(\underline{h})\rho_2|h_2|^2}{\theta+P_R(\underline{h})\rho_2|h_2|^2}\right] \\ & +2\mathrm{E}_{A^c}\left[\mathcal{C}\left(\tfrac{P(\underline{h})\rho_1|h_1|^2}{\theta}\right)\right]-\tfrac{1}{\ln 2}\mathrm{E}_{A^c}\left[\tfrac{P(\underline{h})\rho_1|h_1|^2}{\theta+P(\underline{h})\rho_1|h_1|^2}\right] \\ & +2(1-\alpha^*)\mathrm{E}_A\left[\mathcal{C}\left(\tfrac{P(\underline{h})\rho_3|h_3|^2}{\theta}\right)\right] \\ & -\tfrac{1-\alpha^*}{\ln 2}\mathrm{E}_A\left[\tfrac{P(\underline{h})\rho_3|h_3|^2}{\theta+P(\underline{h})\rho_3|h_3|^2}\right] \\ & \quad =0 \end{aligned} \tag{98}$$

The resource allocation $(\theta, P(\underline{h}), P_R(\underline{h}))$ obtained iteratively from (95),(97), and (98) converges to the optimal $(\theta^{(\alpha^*)}, P^{(\alpha^*)}(\underline{h}), P_R^{(\alpha^*)}(\underline{h}))$.

Finally, the parameter α^* is determined by the condition

$$\begin{aligned}
&\mathrm{E}\left[\theta^{(\alpha^*)}\mathcal{C}\left(\frac{P^{(\alpha^*)}(\underline{h})\rho_1|h_1|^2}{\theta^{(\alpha^*)}}\right)\right. \\
&\qquad\left. +\,\bar{\theta}^{(\alpha^*)}\mathcal{C}\left(\frac{P_R^{(\alpha^*)}(\underline{h})\rho_2|h_2|^2}{\bar{\theta}^{(\alpha^*)}}\right)\right], \\
&\qquad\qquad = \mathrm{E}_A\left[\theta^{(\alpha^*)}\mathcal{C}\left(\frac{P^{(\alpha^*)}(\underline{h})\rho_3|h_3|^2}{\theta^{(\alpha^*)}}\right)\right] \\
&\qquad\qquad\quad + \mathrm{E}_{A^c}\left[\theta^{(\alpha^*)}\mathcal{C}\left(\frac{P^{(\alpha^*)}(\underline{h})\rho_1|h_1|^2}{\theta^{(\alpha^*)}}\right)\right].
\end{aligned} \tag{99}$$

The optimization for the upper bound (90) can be performed in a similar manner, and is not presented in this paper. In general, the lower bound (89) and the upper bound (90) do not match. In the following theorem, we characterize the condition under which the two bounds match and hence yield the capacity of this channel.

***Theorem** 6:* For the fading half-duplex relay channel Scenario II, if the channel statistics and the power constraints satisfy the condition (94), then the capacity is given by

$$\begin{aligned}
C = \mathrm{E}&\left[2\theta^{(1)}\mathcal{C}\left(\frac{P^{(1)}(\underline{h})\rho_1|h_1|^2}{\theta^{(1)}}\right)\right. \\
&\left. +\,2\bar{\theta}^{(1)}\mathcal{C}\left(\frac{P_R^{(1)}(\underline{h})\rho_2|h_2|^2}{\bar{\theta}}\right)\right]
\end{aligned} \tag{100}$$

where the capacity achieving resource allocation $(\theta^{(1)}, P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h}))$ can be obtained iteratively from (91), (92) and (93).

The proof of Theorem 6 is similar to the proof of Theorem 5, and is hence omitted.

***Remark** 5:* The capacity in Theorem 6 refers to the largest rate under Scenario II that can be achieved over all possible channel resource allocation parameters θ and over all possible power allocation rules $(P(\underline{h}), P_R(\underline{h}))$.

The condition given in Theorem 6 tends to be satisfied either when the relay power P_R is small compared to the source power P, or when the relay is much closer to the source than to the destination.

In Fig. 8, we plot the lower and upper bounds on the capacity of Scenario II for the same Rayleigh fading relay channel as in Fig. 6. Both bounds are optimized over $(\theta, P(\underline{h}), P_R(\underline{h}))$. It can be seen from Fig. 8 that when the relay power is less than a threshold (4 dB), the two bounds match and determine the capacity of Scenario II. This demonstrates our capacity result in Theorem 6 and the condition when the lower and upper bounds match. Fig. 8 also shows that the gap between the lower and upper bounds is small even when the relay power is large.

In Fig. 9, we plot the ranges of the source and relay powers with their corresponding max-min optimization cases. The dashed line in the graph divides cases 2 and 3. Similar to Fig. 7, the optimal power allocation falls into case 2 when the relay power is small compared to the source power. However, we see that Fig. 9 deviates from Fig. 7 in that case 1 (where the achievable rate saturates) is missing in Scenario II. This

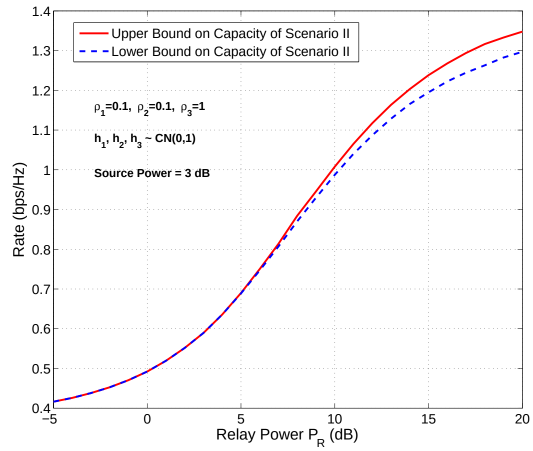

Fig. 8. Lower and upper bounds on capacity of Scenario II

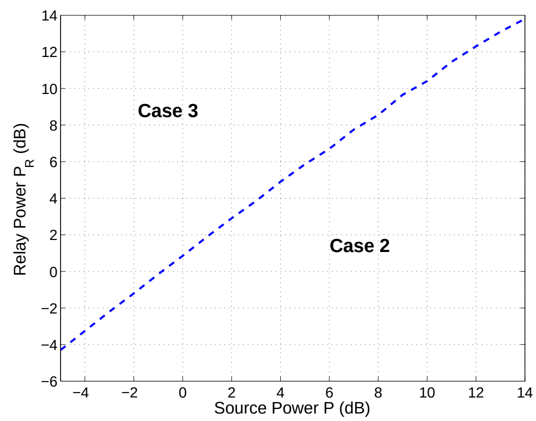

Fig. 9. Ranges of source and relay powers with corresponding max-min optimization cases in Scenario II

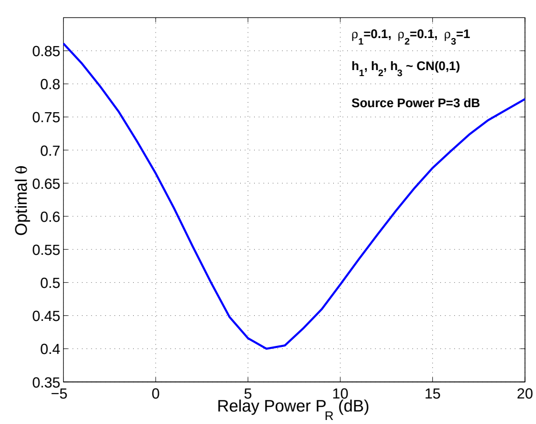

Fig. 10. Optimal θ as a function of relay power in Scenario II

explains why the achievable rate under Scenario II continues to increase beyond the point where the rate under Scenario I saturates (see Fig. 11 in Section V-C).

In Fig. 10, we plot the optimal value of θ as a function of the relay power, and observe that it is not a monotonic function. When the relay power is small, as the relay power increases, the optimal θ decreases so that more of the channel resource is assigned to the relay-to-destination link to make more use of the relay node. When the relay power is large, as the relay power increases, the optimal θ increases. This is because the relay power is now large enough to forward all the information decoded at the relay node to the destination node even with a small amount of the channel resource, and hence more of the channel resource is needed for the source node to transmit more information to the relay node. This behavior of the optimal θ is similar to that of the Gaussian half-duplex relay channel studied in [37].

C. Scenario III: θ Changes with Channel States

In Scenario II, the parameter θ is required to be the same for all channel states, and only the power allocations are dynamically adjusted according to the instantaneous channel state. In this subsection, we study Scenario III, where θ is also allowed to change with the channel state realizations, and $\theta(\underline{h})$ is optimized jointly with power allocation $(P(\underline{h}), P_R(\underline{h}))$. However, for the source and relay nodes to decide $\theta(\underline{h})$ for each channel state, each node needs to know the channel realizations on all transmission links. This makes the system design more complex, and not as practical as Scenario II. We include the analysis of the resource allocation for this scenario mainly for the sake of completeness.

***Proposition** 4:* An achievable rate for the fading half-duplex relay channel Scenario III is given by:

$$
\begin{aligned}
C_{\texttt{low}} = & \max_{\substack{0\le\theta(\underline{h})\le 1,\\ (P(\underline{h}),P_R(\underline{h}))\in\mathcal{G}}} \min \\
& \Bigg\{ \mathrm{E}\left[2\theta(\underline{h})\mathcal{C}\left(\tfrac{P(\underline{h})\rho_1|h_1|^2}{\theta(\underline{h})}\right)\right] + \mathrm{E}\left[2\bar{\theta}(\underline{h})\mathcal{C}\left(\tfrac{P_R(\underline{h})\rho_2|h_2|^2}{\bar\theta(\underline{h})}\right)\right], \\
& \mathrm{E}_A\left[2\theta(\underline{h})\mathcal{C}\left(\tfrac{P(\underline{h})\rho_3|h_3|^2}{\theta(\underline{h})}\right)\right] + \mathrm{E}_{A^c}\left[2\theta(\underline{h})\mathcal{C}\left(\tfrac{P(\underline{h})\rho_1|h_1|^2}{\theta(\underline{h})}\right)\right] \Bigg\}
\end{aligned} \tag{101}
$$

An upper bound on the capacity is given by

$$
\begin{aligned}
C_{\texttt{up}} = & \max_{\substack{0\le\theta(\underline{h})\le 1,\\ (P(\underline{h}),P_R(\underline{h}))\in\mathcal{G}}} \min \\
& \Bigg\{ \mathrm{E}\left[2\theta(\underline{h})\mathcal{C}\left(\tfrac{P(\underline{h})\rho_1|h_1|^2}{\theta(\underline{h})}\right)\right] + \mathrm{E}\left[2\bar{\theta}(\underline{h})\mathcal{C}\left(\tfrac{P_R(\underline{h})\rho_2|h_2|^2}{\bar\theta(\underline{h})}\right)\right], \\
& \mathrm{E}\left[2\theta(\underline{h})\mathcal{C}\left(\tfrac{P(\underline{h})(\rho_3|h_3|^2+\rho_1|h_1|^2)}{\theta(\underline{h})}\right)\right] \Bigg\}.
\end{aligned} \tag{102}
$$

The optimal resource allocation $(\theta(\underline{h}), P(\underline{h}), P_R(\underline{h}))$ that achieves the maximum of the lower bound (101) is given in the following. The proof of optimality is relegated to Appendix III.

Optimal resource allocation that maximizes the lower bound (101):

Case 1: The optimal resource allocation $(\theta^{(0)}(\underline{h}), P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h}))$ is given by

$$
P^{(0)}(\underline{h}) = \begin{cases} \left(\dfrac{1}{\lambda\ln 2} - \dfrac{1}{\rho_3|h_3|^2}\right)^+, & \text{if } \underline{h}\in A, \\ \left(\dfrac{1}{\lambda\ln 2} - \dfrac{1}{\rho_1|h_1|^2}\right)^+, & \text{if } \underline{h}\in A^c; \end{cases} \tag{103}
$$

$$
P_R^{(0)}(\underline{h}) = \begin{cases} \left(\dfrac{1}{\mu\ln 2} - \dfrac{1}{\rho_2|h_2|^2}\right)^+, & \text{if } P^{(0)}(\underline{h}) = 0; \\ 0, & \text{if } P^{(0)}(\underline{h}) > 0, \end{cases} \tag{104}
$$

where λ and μ are chosen to satisfy the power constraint given in (63).

$$
\theta^{(0)}(\underline{h}) = \begin{cases} 1, & \text{if } P^{(0)}(\underline{h}) > 0; \\ 0, & \text{if } P^{(0)}(\underline{h}) = 0. \end{cases} \tag{105}
$$

For case 1 to happen, $(\theta^{(0)}(\underline{h}), P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h}))$ needs to satisfy the following condition:

$$
\begin{aligned}
&\mathrm{E}\left[\mathcal{C}\left(P_R^{(0)}(\underline{h})\rho_2|h_2|^2\right)\right] \\
&\quad \ge \mathrm{E}_A\left[\mathcal{C}\left(P^{(0)}(\underline{h})\rho_3|h_3|^2\right) - \mathcal{C}\left(P^{(0)}(\underline{h})\rho_1|h_1|^2\right)\right].
\end{aligned} \tag{106}
$$

Case 2: The optimal $(\theta^{(1)}(\underline{h}), P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h}))$ can be determined by the following iterative algorithm. For a given $\theta(\underline{h})$, the power allocation $(P(\underline{h}), P_R(\underline{h}))$ is given by

$$
P(\underline{h}) = \theta(\underline{h})\left(\frac{1}{\lambda\ln 2} - \frac{1}{\rho_1|h_1|^2}\right)^+ \tag{107}
$$

$$
P_R(\underline{h}) = \bar\theta(\underline{h})\left(\frac{1}{\mu\ln 2} - \frac{1}{\rho_2|h_2|^2}\right)^+ \tag{108}
$$

where λ and μ are chosen to satisfy the power constraints given in (63). For a given $(P(\underline{h}), P_R(\underline{h}))$, the channel resource allocation $\theta(\underline{h})$ is the root of the following equation:

$$
\begin{aligned}
&2\mathcal{C}\left(\frac{P(\underline{h})\rho_1|h_1|^2}{\theta(\underline{h})}\right) - 2\mathcal{C}\left(\frac{P_R(\underline{h})\rho_2|h_2|^2}{\bar\theta(\underline{h})}\right) \\
&- \frac{1}{\ln 2}\frac{P(\underline{h})\rho_1|h_1|^2}{\theta(\underline{h}) + P(\underline{h})\rho_1|h_1|^2} + \frac{1}{\ln 2}\frac{P_R(\underline{h})\rho_2|h_2|^2}{\bar\theta(\underline{h}) + P_R(\underline{h})\rho_2|h_2|^2} \\
&\quad = 0.
\end{aligned} \tag{109}
$$

For case 2 to happen, $(\theta^{(1)}(\underline{h}), P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h}))$ needs to satisfy the following condition:

$$
\begin{aligned}
&\mathrm{E}\left[\bar\theta^{(1)}(\underline{h})\mathcal{C}\left(\frac{P_R^{(1)}(\underline{h})\rho_2|h_2|^2}{\bar\theta^{(1)}(\underline{h})}\right)\right] \\
&\quad \le \mathrm{E}_A\Bigg[\theta^{(1)}(\underline{h})\mathcal{C}\left(\frac{P^{(1)}(\underline{h})\rho_3|h_3|^2}{\theta^{(1)}(\underline{h})}\right) \\
&\qquad - \theta^{(1)}(\underline{h})\mathcal{C}\left(\frac{P^{(1)}(\underline{h})\rho_1|h_1|^2}{\theta^{(1)}(\underline{h})}\right)\Bigg]
\end{aligned} \tag{110}
$$

Case 3: The optimal $(\theta^{(\alpha^*)}(\underline{h}), P^{(\alpha^*)}(\underline{h}), P_R^{(\alpha^*)}(\underline{h}))$ can be obtained by the following iterative algorithm. For a given $\theta(\underline{h})$, the power allocation $(P(\underline{h}), P_R(\underline{h}))$ is given by

$$P(\underline{h}) = \begin{cases} \text{positive root } x \text{ of (112) if it exists, otherwise } 0, \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{if } \underline{h} \in A, \\ \theta(\underline{h}) \left(\frac{1}{\lambda \ln 2} - \frac{1}{\rho_1 |h_1|^2} \right)^+, \qquad \text{if } \underline{h} \in A^c; \end{cases} \tag{111}$$

where the root x is determined by the following equation:

$$\frac{\alpha^* \theta(\underline{h})}{\ln 2} \frac{1}{\frac{\theta(\underline{h})}{\rho_1 |h_1|^2} + x} + \frac{(1-\alpha^*)\theta(\underline{h})}{\ln 2} \frac{1}{\frac{\theta(\underline{h})}{\rho_3 |h_3|^2} + x} - \lambda = 0. \tag{112}$$

$$P_R(\underline{h}) = \bar{\theta}(\underline{h}) \left(\frac{\alpha^*}{\mu \ln 2} - \frac{1}{\rho_2 |h_2|^2} \right)^+ \tag{113}$$

where the parameters λ and μ are chosen to satisfy the power constraints (63).
For a given $(P(\underline{h}), P_R(\underline{h}))$, the channel resource allocation $\theta(\underline{h})$ is determined by

If $\underline{h} \in A$,

$$\begin{aligned} &2\alpha^* \mathcal{C}\left(\frac{P(\underline{h})\rho_1 |h_1|^2}{\theta(\underline{h})} \right) - \frac{\alpha^*}{\ln 2} \frac{P(\underline{h})\rho_1 |h_1|^2}{\theta(\underline{h}) + P(\underline{h})\rho_1 |h_1|^2} \\ &- 2\alpha^* \mathcal{C}\left(\frac{P_R(\underline{h})\rho_2 |h_2|^2}{\bar{\theta}(\underline{h})} \right) + \frac{\alpha^*}{\ln 2} \frac{P_R(\underline{h})\rho_2 |h_2|^2}{\bar{\theta}(\underline{h}) + P_R(\underline{h})\rho_2 |h_2|^2} \\ &+ 2(1-\alpha*)\mathcal{C}\left(\frac{P(\underline{h})\rho_3 |h_3|^2}{\theta(\underline{h})} \right) \\ &- \frac{1-\alpha*}{\ln 2} \frac{P(\underline{h})\rho_3 |h_3|^2}{\theta(\underline{h}) + P(\underline{h})\rho_3 |h_3|^2} \\ &\quad = 0; \end{aligned} \tag{114}$$

If $\underline{h} \in A^c$,

$$\begin{aligned} &- 2\alpha^* \mathcal{C}\left(\frac{P_R(\underline{h})\rho_2 |h_2|^2}{\bar{\theta}(\underline{h})} \right) + \frac{\alpha^*}{\ln 2} \frac{P_R(\underline{h})\rho_2 |h_2|^2}{\bar{\theta}(\underline{h}) + P_R(\underline{h})\rho_2 |h_2|^2} \\ &+ 2\mathcal{C}\left(\frac{P(\underline{h})\rho_1 |h_1|^2}{\theta(\underline{h})} \right) - \frac{1}{\ln 2} \frac{P(\underline{h})\rho_1 |h_1|^2}{\theta(\underline{h}) + P(\underline{h})\rho_1 |h_1|^2} \\ &\quad = 0. \end{aligned}$$

The parameter α^* is determined by the condition

$$\begin{aligned} &\mathrm{E}\left[\theta^{(\alpha^*)}(\underline{h}) \mathcal{C}\left(\frac{P^{(\alpha^*)}(\underline{h})\rho_1 |h_1|^2}{\theta^{(\alpha^*)}(\underline{h})} \right) \right] \\ &+ \mathrm{E}\left[\bar{\theta}^{(\alpha^*)}(\underline{h}) \mathcal{C}\left(\frac{P_R^{(\alpha^*)}(\underline{h})\rho_2 |h_2|^2}{\bar{\theta}^{(\alpha^*)}(\underline{h})} \right) \right], \\ &\qquad = \mathrm{E}_A\left[\theta^{(\alpha^*)}(\underline{h}) \mathcal{C}\left(\frac{P^{(\alpha^*)}(\underline{h})\rho_3 |h_3|^2}{\theta^{(\alpha^*)}(\underline{h})} \right) \right] \\ &\qquad\quad + \mathrm{E}_{A^c}\left[\theta^{(\alpha^*)}(\underline{h}) \mathcal{C}\left(\frac{P^{(\alpha^*)}(\underline{h})\rho_1 |h_1|^2}{\theta^{(\alpha^*)}(\underline{h})} \right) \right]. \end{aligned} \tag{115}$$

The optimization for the upper bound (102) can be performed using steps that are similar to those for the lower bound. In general, the lower bound (101) and the upper bound (102) do not match. However, we show that if the channel statistics and the power constraints satisfy the following condition, the two bounds match and hence we obtain the capacity for this channel.

***Theorem** 7:* For the fading half-duplex relay channel Scenario III, if the channel statistics and the power constraints satisfy the condition (110), then the capacity is given by

$$\begin{aligned} C = &\mathrm{E}\left[2\theta^{(1)}(\underline{h}) \mathcal{C}\left(\frac{P^{(1)}(\underline{h})\rho_1 |h_1|^2}{\theta^{(1)}(\underline{h})} \right) \right] \\ &+ \mathrm{E}\left[2\bar{\theta}^{(1)}(\underline{h}) \mathcal{C}\left(\frac{P_R^{(1)}(\underline{h})\rho_2 |h_2|^2}{\bar{\theta}^{(1)}(\underline{h})} \right) \right] \end{aligned} \tag{116}$$

where the capacity achieving resource allocation $(\theta^{(1)}(\underline{h}), P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h}))$ can be obtained iteratively from (107), (108) and (109).

The proof of Theorem 7 is similar to that of Theorem 5, and is omitted.

***Remark** 6:* The capacity in Theorem 7 refers to the largest rate under Scenario III that can be achieved over all possible channel resource allocation $\theta(\underline{h})$ and power allocation $(P(\underline{h}), P_R(\underline{h}))$.

The condition given in Theorem 7 is similar to that in Theorem 6 for Scenario II, and these conditions tend to be satisfied either when the relay power P_R is small compared to the source power P, or when the relay node is much closer to the source node than to the destination node.

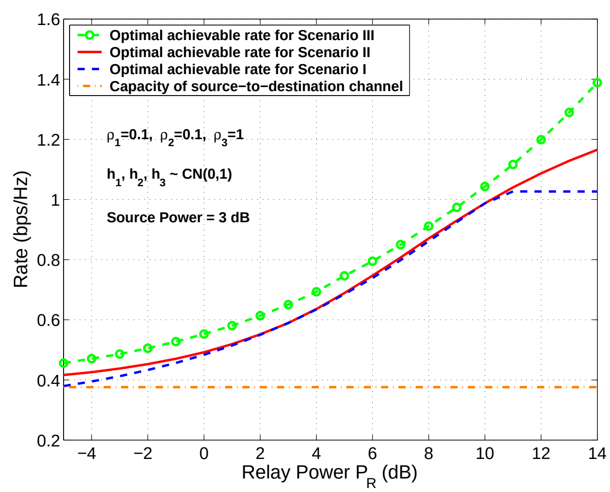

Fig. 11. Comparison of achievable rates with optimal resource allocations for Scenarios I, II, and III

In Fig. 11, we plot the achievable rates under Scenario III optimized over $(\theta(\underline{h}), P(\underline{h}), P_R(\underline{h}))$ for the same Rayleigh fading relay channel as in Fig. 6 and Fig. 8. We compare these rates with the achievable rates under Scenario I optimized over $(P(\underline{h}), P_R(\underline{h}))$ and Scenario II optimized over $(\theta, P(\underline{h}), P_R(\underline{h}))$, and with the capacity of the direct link from the source node to the destination node. It is clear from the graph that employing the relay node greatly improves the performance of the source-to-destination channel. Fig. 11 shows that the achievable rate under scenario II is larger than the achievable rate under Scenario I, particularly when the relay power is large and the achievable rate under scenario I saturates. This demonstrates that using a jointly optimal channel resource allocation parameter θ helps to improve the achievable rate. As we have commented for Fig. 9, Scenario

II does not have case 1, and hence the achievable rate under Scenario II continues to increase when the achievable rate under Scenario I saturates in case 1. Furthermore, Scenario III has larger achievable rates than Scenario II because $\theta(\underline{h})$ can be dynamically changed based on the instantaneous channel state information.

We note that finding an optimal resource allocation for Rayleigh fading relay channels is a high dimensional optimization problem, particularly in Scenario III where the optimization is jointly over $(\theta(\underline{h}), P(\underline{h}), P_R(\underline{h}))$. Although the problem is convex, the standard convex programming techniques may converge slowly. However, since we have obtained the analytical structures of the optimal solutions, our numerical algorithm converges extremely fast and takes only a few iterations.

VI. Concluding Remarks

We have studied capacity bounds for the parallel relay channel and its special case of the fading relay channel. We have established capacity theorems for several classes of channels including the parallel relay channel with degraded subchannels and its Gaussian case, the full-duplex relay channel that satisfies certain conditions in asynchronized case, and the half-duplex relay channel that satisfies certain conditions.

We have studied resource allocation for the Gaussian parallel relay channel with degraded subchannels and the fading relay channel under both full-duplex and half-duplex models. Our study of resource allocation is different from previous work on this topic in that we make the more practical assumption that the source and relay nodes are subject to separate power constraints rather than a total power constraint. We have shown that optimal resource allocation under this assumption may take three different forms depending on the channel statistics and values of the power constraints.

Finally, we note that the resource allocation problem we have considered falls under a class of *max-min* problems and we have provided a technique for solving such max-min problems. It is known that the achievable rates of relay channels when relay nodes use the decode-and-forward scheme are usually expressed by *max-min* forms. Our technique certainly applies to optimization problems arising in these contexts. In particular, our technique has been applied to study orthogonal relay broadcast channels in [40], and can be used to study more general classes of relay networks with fading links.

Appendix I
Proof of Resource Allocation that Maximizes $C_{\texttt{low}}$ (78) for Scenario I

We first let $R_1(P(\underline{h}), P_R(\underline{h}))$ and $R_2(P(\underline{h}), P_R(\underline{h}))$ denote the two terms over which the minimization in (78) is taken. We can then express (78) in the following compact form:

$$C_{\texttt{low}} = \max_{(P(\underline{h}),P_R(\underline{h}))\in\mathcal{G}} \min\left\{R_1(P(\underline{h}), P_R(\underline{h})),\ R_2(P(\underline{h}), P_R(\underline{h}))\right\}. \tag{117}$$

We apply Proposition 1 to derive the optimal power allocation rule, which falls into the following three cases.

Case 1: $\alpha^* = 0$, and $(P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h}))$ is an optimal power allocation, which needs to satisfy the condition

$$R_1(P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h})) \geq R_2(P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h})). \tag{118}$$

By definition, $(P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h}))$ maximizes

$$R(0, P(\underline{h}), P_R(\underline{h})) = R_2(P(\underline{h}), P_R(\underline{h})). \tag{119}$$

The optimal $P^{(0)}(\underline{h})$ given in (79) follows easily from the KKT condition. For case 1 to happen, $(P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h}))$ needs to satisfy the condition (118). It is clear that $R_2(P(\underline{h}), P_R(\underline{h}))$ depends only on $P(\underline{h})$. The term $R_1(P(\underline{h}), P_R(\underline{h}))$ depends on both $P(\underline{h})$ and $P_R(\underline{h})$. To characterize the most general condition for case 1 to happen, $P_R(\underline{h})$ needs to maximize $R_1(P(\underline{h}), P_R(\underline{h}))$. Such $P_R^{(0)}(\underline{h})$ can be obtained by the KKT condition and is given in (80). The condition (81) follows from equality of condition (118).

Case 2: $\alpha^* = 1$, and $(P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h}))$ is an optimal power allocation, which needs to satisfy the condition

$$R_1(P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h})) \leq R_2(P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h})). \tag{120}$$

The optimal $(P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h}))$ that maximizes

$$R(1, P(\underline{h}), P_R(\underline{h})) = R_1(P(\underline{h}), P_R(\underline{h})) \tag{121}$$

can be easily obtained by the KKT condition, and are given in (82) and (83). The condition (84) follows from equality of condition (120).

Case 3: $0 < \alpha^* < 1$, and $(P^{(\alpha^*)}(\underline{h}), P_R^{(\alpha^*)}(\underline{h}))$ is an optimal power allocation, where α^* is determined by the following condition

$$R_1(P^{(\alpha^*)}(\underline{h}), P_R^{(\alpha^*)}(\underline{h})) = R_2(P^{(\alpha^*)}(\underline{h}), P_R^{(\alpha^*)}(\underline{h})). \tag{122}$$

We need to derive $(P^{(\alpha^*)}(\underline{h}), P_R^{(\alpha^*)}(\underline{h}))$ that maximizes

$$\begin{aligned} &R(\alpha^*, P(\underline{h}), P_R(\underline{h})) \\ &\quad = \alpha^* R_1(P(\underline{h}), P_R(\underline{h})) + (1-\alpha^*) R_2(P(\underline{h}), P_R(\underline{h})). \end{aligned} \tag{123}$$

The Lagrangian can be written as

$$\begin{aligned} \mathcal{L} =& \alpha^* \mathrm{E}_A\left[\mathcal{C}\left(2P(\underline{h})\rho_1|h_1|^2\right)\right] + \alpha^* \mathrm{E}\left[\mathcal{C}\left(2P_R(\underline{h})\rho_2|h_2|^2\right)\right] \\ &+ \mathrm{E}_{A^c}\left[\mathcal{C}\left(2P(\underline{h})\rho_1|h_1|^2\right)\right] \\ &+ (1-\alpha^*)\mathrm{E}_A\left[\mathcal{C}\left(2P(\underline{h})\rho_3|h_3|^2\right)\right] \\ &- \lambda\Big(\mathrm{E}[P(\underline{h})] - P\Big) - \mu\Big(\mathrm{E}[P_R(\underline{h})] - P_R\Big) \end{aligned} \tag{124}$$

where λ and μ are Lagrange multipliers.

For $\underline{h} \in A$, the KKT condition is given by:

$$\begin{aligned} &\frac{\partial \mathcal{L}}{\partial P(\underline{h})} \\ &\quad = \frac{\alpha^*}{2\ln 2} \cdot \frac{1}{\frac{1}{2\rho_1|h_1|^2} + P(\underline{h})} + \frac{1-\alpha^*}{2\ln 2} \cdot \frac{1}{\frac{1}{2\rho_3|h_3|^2} + P(\underline{h})} \\ &\quad \leq \lambda, \qquad\qquad \text{with equality if } P(\underline{h}) > 0 \end{aligned} \tag{125}$$

It is easy to check that $P^{(\alpha^*)}(\underline{h})$ for $\underline{h} \in A$ given in (85) satisfies the preceding KKT condition. The $P^{(\alpha^*)}(\underline{h})$ for $\underline{h} \in A^c$ in (85) and $P_R^{(\alpha^*)}(\underline{h})$ in (87) also follow from the KKT condition.

APPENDIX II
PROOF OF RESOURCE ALLOCATION THAT MAXIMIZES $C_{\texttt{low}}$ (89) FOR SCENARIO II

We let $R_1(\theta, P(\underline{h}), P_R(\underline{h}))$ and $R_2(\theta, P(\underline{h}), P_R(\underline{h}))$ denote the two terms over which the minimization in (89) is taken. We can then express (89) in the following compact form:

$$C_{\texttt{low}} = \max_{0\le\theta\le 1, (P(\underline{h}),P_R(\underline{h}))\in\mathcal{G}} \min\left\{R_1(\theta, P(\underline{h}), P_R(\underline{h})),\ R_2(\theta, P(\underline{h}), P_R(\underline{h}))\right\}. \tag{126}$$

The max-min problem in (126) can be solved by using Proposition 1. The main step is to obtain $(\theta^{(\alpha)}, P^{(\alpha)}(\underline{h}), P_R^{(\alpha)}(\underline{h}))$ that maximizes

$$\begin{aligned} &R(\alpha, \theta, P(\underline{h}), P_R(\underline{h})) \\ &\quad := \alpha R_1(\theta, P(\underline{h}), P_R(\underline{h})) + (1-\alpha) R_2(\theta, P(\underline{h}), P_R(\underline{h})) \end{aligned} \tag{127}$$

for a given α. The following lemma states that maximizing the function $R(\alpha, \theta, P(\underline{h}), P_R(\underline{h}))$ over $(\theta, P(\underline{h}), P_R(\underline{h}))$ is a convex programming problem and hence can be solved by standard convex programming algorithms.

***Lemma** 2:* For a fixed α, $R(\alpha, \theta, P(\underline{h}), P_R(\underline{h}))$ is a concave function over $(\theta, P(\underline{h}), P_R(\underline{h}))$, where $0 \le \theta \le 1$ and $(P(\underline{h}), P_R(\underline{h})) \in \mathcal{G}$.

Lemma 2 can be verified by computing the Hessian of the function $R(\alpha, \theta, P(\underline{h}), P_R(\underline{h}))$ (for a fixed α) and showing that it is negative semidefinite.

We now apply Proposition 1 to study the max-min problem in (126) by considering the following three cases.

Case 1: $\alpha^* = 0$, and $(\theta^{(0)}, P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h}))$ is an optimal resource allocation, which needs to satisfy the condition

$$R_1(\theta^{(0)}, P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h})) \ge R_2(\theta^{(0)}, P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h})). \tag{128}$$

We first derive $(\theta^{(0)}, P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h}))$ that maximizes

$$R(0, \theta, P(\underline{h}), P_R(\underline{h})) = R_2(\theta, P(\underline{h}), P_R(\underline{h})). \tag{129}$$

It is easy to see that the optimal $\theta^{(0)} = 1$ from the expression of $R_2(\theta, P(\underline{h}), P_R(\underline{h}))$, and this results in

$$R_1(\theta^{(0)}, P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h})) \le R_2(\theta^{(0)}, P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h})). \tag{130}$$

Comparing (128) and (130), it is clear that only equality can be satisfied in (128). Hence this case can be included in the following case 3 with α^* being allowed to take the value of 0.

Case 2: $\alpha^* = 1$, and $(\theta^{(1)}, P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h}))$ is an optimal resource allocation, which needs to satisfy the condition

$$R_1(\theta^{(1)}, P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h})) \le R_2(\theta^{(1)}, P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h})). \tag{131}$$

We first derive $(\theta^{(1)}, P^{(1)}(\underline{h}), P_R^{(1)}(\underline{h}))$ that maximizes

$$R(1, \theta, P(\underline{h}), P_R(\underline{h})) = R_1(\theta, P(\underline{h}), P_R(\underline{h})). \tag{132}$$

The Lagrangian can be written as

$$\begin{aligned} \mathcal{L} = 2\theta \mathrm{E}\left[\mathcal{C}\left(\frac{P(\underline{h})\rho_1|h_1|^2}{\theta}\right)\right] + 2\bar{\theta}\mathrm{E}\left[\mathcal{C}\left(\frac{P_R(\underline{h})\rho_2|h_2|^2}{\bar{\theta}}\right)\right] \\ - \lambda\Big(\mathrm{E}[P(\underline{h})] - P\Big) - \mu\Big(\mathrm{E}[P_R(\underline{h})] - P_R\Big). \end{aligned} \tag{133}$$

It is easy to check that the KKT condition implies

$$P(\underline{h}) = \theta\left(\frac{1}{\lambda \ln 2} - \frac{1}{\rho_1|h_1|^2}\right)^+ \tag{134}$$

$$P_R(\underline{h}) = \bar{\theta}\left(\frac{1}{\mu \ln 2} - \frac{1}{\rho_2|h_2|^2}\right)^+ \tag{135}$$

The KKT condition also implies that the optimal $\theta^{(1)}$ needs to satisfy the following condition:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \theta} &= 2\mathrm{E}\left[\mathcal{C}\left(\frac{P(\underline{h})\rho_1|h_1|^2}{\theta}\right)\right] - 2\mathrm{E}\left[\mathcal{C}\left(\frac{P_R(\underline{h})\rho_2|h_2|^2}{\bar{\theta}}\right)\right] \\ &- \frac{1}{\ln 2}\mathrm{E}\left[\frac{P(\underline{h})\rho_1|h_1|^2}{\theta + P(\underline{h})\rho_1|h_1|^2}\right] + \frac{1}{\ln 2}\mathrm{E}\left[\frac{P_R(\underline{h})\rho_2|h_2|^2}{\bar{\theta} + P_R(\underline{h})\rho_2|h_2|^2}\right] \\ &\begin{cases} \le 0, & \text{if } \theta = 0; \quad \text{(does not happen)} \\ = 0, & \text{if } 0 < \theta < 1; \\ \ge 0, & \text{if } \theta = 1. \quad \text{(does not happen)} \end{cases} \end{aligned} \tag{136}$$

where the first and third cases do not happen because $\frac{\partial \mathcal{L}}{\partial \theta} \to \infty$ as $\theta \to 0$, and $\frac{\partial \mathcal{L}}{\partial \theta} \to -\infty$ as $\theta \to 1$. It can also be shown that $\frac{\partial \mathcal{L}}{\partial \theta}$ is monotonically decreasing for $0 \le \theta \le 1$. Hence $\frac{\partial \mathcal{L}}{\partial \theta}$ has at most one root for $0 \le \theta \le 1$.

The iterative algorithm described in (91)-(93) converges to the solution of the KKT condition given in (134)-(136). Since the function $R_1(\theta, P(\underline{h}), P_R(\underline{h}))$ is concave, the solution of the KKT condition achieves the optimum. Condition (94) follows from condition (131).

Case 3: $0 < \alpha^* < 1$, and $(\theta^{(\alpha^*)}, P^{(\alpha^*)}(\underline{h}), P_R^{(\alpha^*)}(\underline{h}))$ is an optimal resource allocation, where α^* is determined by the following condition

$$\begin{aligned} &R_1(\theta^{(\alpha^*)}, P^{(\alpha^*)}(\underline{h}), P_R^{(\alpha^*)}(\underline{h})) \\ &\qquad = R_2(\theta^{(\alpha^*)}, P^{(\alpha^*)}(\underline{h}), P_R^{(\alpha^*)}(\underline{h})). \end{aligned} \tag{137}$$

We first derive $(\theta^{(\alpha^*)}, P^{(\alpha^*)}(\underline{h}), P_R^{(\alpha^*)}(\underline{h}))$ that maximizes

$$\begin{aligned} &R(\alpha^*, \theta, P(\underline{h}), P_R(\underline{h})) \\ &\quad = \alpha^* R_1(\theta, P(\underline{h}), P_R(\underline{h})) + (1-\alpha^*) R_2(\theta, P(\underline{h}), P_R(\underline{h})). \end{aligned} \tag{138}$$

for given α^*. The Lagrangian can be written as

$$\begin{aligned} \mathcal{L} =& 2\alpha^*\theta \mathrm{E}_A\left[\mathcal{C}\left(\frac{P(\underline{h})\rho_1|h_1|^2}{\theta}\right)\right] \\ &+ 2\alpha^*\bar{\theta}\mathrm{E}\left[\mathcal{C}\left(\frac{P_R(\underline{h})\rho_2|h_2|^2}{\bar{\theta}}\right)\right] \\ &+ 2\theta \mathrm{E}_{A^c}\left[\mathcal{C}\left(\frac{P(\underline{h})\rho_1|h_1|^2}{\theta}\right)\right] \\ &+ 2(1-\alpha^*)\theta \mathrm{E}_A\left[\mathcal{C}\left(\frac{P(\underline{h})\rho_3|h_3|^2}{\theta}\right)\right] \\ &- \lambda\Big(\mathrm{E}[P(\underline{h})] - P\Big) - \mu\Big(\mathrm{E}[P_R(\underline{h})] - P_R\Big). \end{aligned} \tag{139}$$

For a given θ, The optimal $(P(\underline{h}), P_R(\underline{h}))$ given in (95) and (97) follows from the KKT condition. The KKT condition also implies that the optimal θ for a given $(P(\underline{h}), P_R(\underline{h}))$ needs to satisfy the following condition

$$\begin{aligned}
&\frac{\partial \mathcal{L}}{\partial \theta} \\
&= 2\alpha^* \mathrm{E}_A \left[\mathcal{C}\left(\frac{P(\underline{h})\rho_1 |h_1|^2}{\theta} \right) \right] - \frac{\alpha^*}{\ln 2} \mathrm{E}_A \left[\frac{P(\underline{h})\rho_1 |h_1|^2}{\theta + P(\underline{h})\rho_1 |h_1|^2} \right] \\
&- 2\alpha^* \mathrm{E} \left[\mathcal{C}\left(\frac{P_R(\underline{h})\rho_2 |h_2|^2}{\bar{\theta}} \right) \right] + \frac{\alpha^*}{\ln 2} \mathrm{E} \left[\frac{P_R(\underline{h})\rho_2 |h_2|^2}{\bar{\theta} + P_R(\underline{h})\rho_2 |h_2|^2} \right] \\
&+ 2\mathrm{E}_{A^c} \left[\mathcal{C}\left(\frac{P(\underline{h})\rho_1 |h_1|^2}{\theta} \right) \right] - \frac{1}{\ln 2} \mathrm{E}_{A^c} \left[\frac{P(\underline{h})\rho_1 |h_1|^2}{\theta + P(\underline{h})\rho_1 |h_1|^2} \right] \\
&+ 2(1-\alpha^*) \mathrm{E}_A \left[\mathcal{C}\left(\frac{P(\underline{h})\rho_3 |h_3|^2}{\theta} \right) \right] \\
&- \frac{1-\alpha^*}{\ln 2} \mathrm{E}_A \left[\frac{P(\underline{h})\rho_3 |h_3|^2}{\theta + P(\underline{h})\rho_3 |h_3|^2} \right] \\
&\begin{cases} \le 0, & \text{if } \theta = 0; \quad \text{(does not happen)} \\ = 0, & \text{if } 0 < \theta < 1; \\ \ge 0, & \text{if } \theta = 1. \quad \text{(does not happen)} \end{cases}
\end{aligned} \tag{140}$$

where the first and third cases do not happen because $\frac{\partial \mathcal{L}}{\partial \theta} \to \infty$ as $\theta \to 0$, and $\frac{\partial \mathcal{L}}{\partial \theta} \to -\infty$ as $\theta \to 1$.

Therefore, the iterative algorithm described in (95)-(98) converges to the solution of the KKT condition. Since the function $R(\alpha, \theta, P(\underline{h}), P_R(\underline{h}))$ is concave for a given α, the solution of the KKT condition achieves the optimum. Condition (99) follows from condition (137).

Appendix III
Proof of Resource Allocation that Maximizes $C_{\texttt{low}}$ (101) for Scenario III

We let $R_1(\theta(\underline{h}), P(\underline{h}), P_R(\underline{h}))$ and $R_2(\theta(\underline{h}), P(\underline{h}), P_R(\underline{h}))$ denote the two terms over which the minimization in (101) is taken. We can then express (101) in the following compact form:

$$\begin{aligned}
C_{\texttt{low}} = &\max_{\substack{0 \le \theta(\underline{h}) \le 1, \, \forall \, \underline{h} \\ (P(\underline{h}), P_R(\underline{h})) \in \mathcal{G}}} \\
&\min \left\{ R_1(\theta(\underline{h}), P(\underline{h}), P_R(\underline{h})), \; R_2(\theta(\underline{h}), P(\underline{h}), P_R(\underline{h})) \right\}
\end{aligned} \tag{141}$$

As in Scenario II, one main step to solve the max-min problem in (141) is to obtain $(\theta^{(\alpha)}(\underline{h}), P^{(\alpha)}(\underline{h}), P_R^{(\alpha)}(\underline{h}))$ that maximizes

$$\begin{aligned}
R(\alpha, \theta(\underline{h}), P(\underline{h}), P_R(\underline{h})) := &\alpha R_1(\theta(\underline{h}), P(\underline{h}), P_R(\underline{h})) \\
&+ (1-\alpha) R_2(\theta(\underline{h}), P(\underline{h}), P_R(\underline{h}))
\end{aligned} \tag{142}$$

for a given α. The following lemma states that maximizing $R(\alpha, \theta(\underline{h}), P(\underline{h}), P_R(\underline{h}))$ over $(\theta(\underline{h}), P(\underline{h}), P_R(\underline{h}))$ is a convex programming and hence can be solved by standard convex programming algorithms.

***Lemma** 3:* For a fixed α, $R(\alpha, \theta(\underline{h}), P(\underline{h}), P_R(\underline{h}))$ is a concave function over $(\theta(\underline{h}), P(\underline{h}), P_R(\underline{h}))$, where $0 \le \theta(\underline{h}) \le 1$ for all $\underline{h}$ and $(P(\underline{h}), P_R(\underline{h})) \in \mathcal{G}$.

Lemma 3 can be verified by computing the Hessian of the function $R(\alpha, \theta(\underline{h}), P(\underline{h}), P_R(\underline{h}))$ (for a fixed α) and showing that it is negative semidefinite.

As for Scenarios I and II, we apply Proposition 1 to study the max-min problem in (141) by considering the following three cases.

Case 1: $\alpha^* = 0$, and $(\theta^{(0)}(\underline{h}), P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h}))$ is an optimal resource allocation, which needs to satisfy the condition

$$\begin{aligned}
&R_1(\theta^{(0)}(\underline{h}), P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h})) \\
&\qquad \ge R_2(\theta^{(0)}(\underline{h}), P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h})).
\end{aligned} \tag{143}$$

We first derive $(\theta^{(0)}(\underline{h}), P^{(0)}(\underline{h}), P_R^{(0)}(\underline{h}))$ that maximizes

$$R(0, \theta(\underline{h}), P(\underline{h}), P_R(\underline{h})) = R_2(\theta(\underline{h}), P(\underline{h}), P_R(\underline{h})). \tag{144}$$

It is clear that $\theta^{(0)}(\underline{h})$ given in (105) is optimal from the expression of $R_2(\theta(\underline{h}), P(\underline{h}), P_R(\underline{h}))$. The power allocation $P^{(0)}(\underline{h})$ given in (103) then easily follows from the KKT condition.

For case 1 to happen, condition (143) needs to be satisfied. To characterize the most general condition for case 1 to happen, for the given $\theta^{(0)}(\underline{h})$ and $P^{(0)}(\underline{h})$ $P_R(\underline{h})$ needs to maximize $R_1(\theta^{(0)}(\underline{h}), P^{(0)}(\underline{h}), P_R(\underline{h}))$, which has the following form:

$$\begin{aligned}
&R_1(\theta^{(0)}(\underline{h}), P^{(0)}(\underline{h}), P_R(\underline{h})) \\
&\qquad = 2\mathrm{E}_{\{\underline{h}: P^{(0)}(\underline{h}) = 0\}} \left[\mathcal{C}\left(P_R(\underline{h})\rho_2 |h_2|^2 \right) \right]
\end{aligned} \tag{145}$$

The optimal $P_R^{(0)}(\underline{h})$ given in (104) then follows from the KKT condition. Finally, condition (106) follows from condition (143).

The proofs for cases 2 and 3 are similar to those for Scenario II given in Appendix II, and are omitted.

Acknowledgment

The authors would like to thank Dr. Gerhard Kramer of Bell Laboratories, Alcatel-Lucent, for helpful discussions on the topic of parallel relay channels. The authors would also like to thank Dr. Jianwei Huang of Princeton University for helpful discussions on the topic of optimization.

References

[1] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[2] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 5, pp. 572–584, Sept. 1979.

[3] A. A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Trans. Inform. Theory*, vol. 28, no. 3, pp. 536–536, May 1982.

[4] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[5] ——, "User cooperation diversity-part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

[6] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Sorrento, Italy, June 2000, p. 22.

[7] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Trans. Inform. Theory*, vol. 50, no. 5, pp. 748–767, May 2004.

[8] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3037–3063, Sept. 2005.

[9] A. Host-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 2020–2040, June 2005.

[10] A. Host-Madsen, "Capacity bounds for cooperative diversity," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1522–1544, Apr. 2006.

[11] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 1, pp. 29–43, Jan. 2005.

[12] M. Katz and S. S. (Shitz), "Transmitting to colocated users in wireless ad hoc and sensor networks," *IEEE Trans. Inform. Theory*, vol. 51, no. 10, pp. 3540–3563, Oct. 2005.

[13] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 765–779, Mar. 2005.

[14] U. Mitra and A. Sabharwal, "On achievable rates of complexity constrained relay channels," in *Proc. Annu. Allerton Conf. Communication, Control and Computing*, Monticello, Illinois, Sept.-Oct. 2003, pp. 551–560.

[15] M. A. Khojastepour and B. Aazhang, "'Cheap' relay channel: A unifying approach to time and frequency division relaying," in *Proc. Annu. Allerton Conf. Communication, Control and Computing*, Monticello, Illinois, Sept.-Oct. 2004, pp. 1792–1801.

[16] R. U. Nabar, H. Bölcskei, and F. W. Kneubuhler, "Fading relay channels: Performance limits and space-time signal design," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1099–1109, Aug. 2004.

[17] L. Sankaranarayanan, G. Kramer, and N. B. Mandayam, "Capacity theorems for the multiple-access relay channel," in *Proc. Annu. Allerton Conf. Communication, Control and Computing*, Monticello, Illinois, Sept.-Oct. 2004, pp. 1782–1791.

[18] N. Jindal, U. Mitra, and A. Goldsmith, "Capacity of ad-hoc networks with node cooperation," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Chicago, June-July 2004, p. 271.

[19] Y. Liang and V. V. Veeravalli, "Cooperative relay broadcast channels," *IEEE Trans. Inform. Theory*, vol. 53, no. 3, pp. 900–928, Mar. 2007.

[20] R. Dabora and S. Servetto, "Broadcast channels with cooperating decoders," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5438–5454, Dec. 2006.

[21] A. A. El Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1545–1561, Apr. 2006.

[22] J. N. Laneman and G. W. Wornell, "Distributed space-time coded protocols for exploring cooperative diversity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[23] K. Azarian, H. E. Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4152– 4172, Dec. 2005.

[24] N. Pradad and M. K. Varanasi, "Diversity and multiplexing tradeoff bounds for cooperative diversity protocols," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Chicago, June-July 2004, p. 268.

[25] M. Yuksel and E. Erkip, "Diversity-multiplexing tradeoff in multiple-antenna relay systems," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Seattle, Washington, July 2006, pp. 1154–1158.

[26] D. Gunduz and E. Erkip, "Opportunistic cooperation by dynamic resource allocation," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1446–1454, Apr. 2007.

[27] A. S. Avestimehr and D. Tse, "Outage capacity of the fading relay channel in the low SNR regime," submitted to *IEEE Trans. Inform. Theory*, Feb. 2006.

[28] A. Stefanov and E. Erkip, "Cooperative coding for wireless networks," *IEEE Trans. Commun.*, vol. 52, no. 9, pp. 1470–1476, Sept. 2004.

[29] A. Murugan, K. Azarian, and H. E. Gamalr, "Cooperative lattice coding and decoding in half-duplex channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 268–279, Feb. 2007.

[30] X. Bao and J. Li, "An information theoretic analysis for adaptive-network-coded-cooperation (ANCC) in wireless relay networks," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Seattle, Washington, July 2006, pp. 2719–1723.

[31] Y. Yao, X. Cai, and G. B. Giannakis, "On energy efficiency and optimum resource allocation of relay transmissions in the low-power regime," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2917–2927, Nov. 2005.

[32] A. Reznik, S. Kulkarni, and S. Verdú, "Degraded Gaussian multirelay channel: Capacity and optimal power allocation," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3037–3046, Dec. 2004.

[33] I. Maric and R. Yates, "Forwarding strategies for Gaussian parallel-relay networks," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Chicago, June-July 2004, p. 269.

[34] I. Maric and R. D. Yates, "Bandwidth and power allocation for cooperative strategies in Gaussian relay networks," in *Proc. Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, California, Nov. 2004, pp. 1907–1911.

[35] H. V. Poor, *An Introduction to Signal Detection and Estimation,* 2nd ed. New York: Springer, 1994.

[36] Y. Liang and G. Kramer, "Rate regions for relay broadcast channels," *IEEE Trans. Inform. Theory*, vol. 53, no. 10, pp. ?–?, Oct. 2007.

[37] Y. Liang and V. V. Veeravalli, "Guassian orthogonal relay channels: Optimal resource allocation and capacity," *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3284–3289, Sept. 2005.

[38] D. G. Luenberger, *Linear and nonlinear programming, Second Edition*. Kluwer Academic Publishers, 2003.

[39] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE Int. Conf. on Communications (ICC)*, vol. 1, Seattle, Washington, June 1995, pp. 331–335.

[40] Y. Liang and V. V. Veeravalli, "Distributed optimal resource allocation for fading relay broadcast channels," in *IEEE Int. Workshop on Signal Processing Advances for Wireless Communications (SPAWC)*, New York, June 2005, pp. 573–577.

Yingbin Liang (S'01–M'05) received the Ph.D. degree in Electrical Engineering from the University of Illinois at Urbana-Champaign in 2005. Since September 2005, she has been working as a postdoctoral research associate at Princeton University. Her research interests include information theory, wireless communications, and wireless networks.

Dr. Liang was a Vodafone Fellow at the University of Illinois at Urbana-Champaign during 2003-2005, and received the Vodafone-U.S. Foundation Fellows Initiative Research Merit Award in 2005. She also received the M. E. Van Valkenburg Graduate Research Award from the ECE department, University of Illinois at Urbana-Champaign, in 2005.

Venugopal V. Veeravalli (S'86–M'92–SM'98–F'06) received the Ph.D. degree in 1992 from the University of Illinois at Urbana-Champaign, the M.S. degree in 1987 from Carnegie-Mellon University, Pittsburgh, PA, and the B. Tech. degree in 1985 from the Indian Institute of Technology, Bombay, (Silver Medal Honors), all in Electrical Engineering. He joined the University of Illinois at Urbana-Champaign in 2000, where he is currently a Professor in the department of Electrical and Computer Engineering, and a Research Professor in the Coordinated Science Laboratory. He served as a program director for communications research at the U.S. National Science Foundation in Arlington, VA from 2003 to 2005. He was an assistant professor at Cornell University, Ithaca, NY from 1996 to 2000.

His research interests include distributed sensor systems and networks, wireless communications, detection and estimation theory, and information theory. He is a Fellow of the IEEE and currently on the Board of Governors of the IEEE Information Theory Society. He was an Associate Editor for Detection and Estimation for the IEEE Transactions on Information Theory from 2000 to 2003, and an associate editor for the IEEE Transactions on Wireless Communications from 1999 to 2000. Among the awards he has received for research and teaching are the IEEE Browder J. Thompson Best Paper Award in 1996, the National Science Foundation CAREER Award in 1998, and the Presidential Early Career Award for Scientists and Engineers (PECASE) in 1999.

H. Vincent Poor (S'72–M'77–SM'82–F'87) received the Ph.D. degree in EECS from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is the Dean of Engineering and Applied Science, and the Michael Henry Strater University Professor of Electrical Engineering. Dr. Poor's research interests are in the areas of stochastic analysis, statistical signal processing and their applications in wireless networks and related fields. Among his publications in these areas is the recent book *MIMO Wireless Communications* (Cambridge University Press, 2007).

Dr. Poor is a member of the National Academy of Engineering, a Fellow of the American Academy of Arts and Sciences, and a former Guggenheim Fellow. He is also a Fellow of the Institute of Mathematical Statistics, the Optical Society of America, and other organizations. In 1990, he served as President of the IEEE Information Theory Society, and in 2004-07 as the Editor-in-Chief of these *Transactions*. Recent recognition of his work includes the 2005 IEEE Education Medal and the 2007 IEEE Marconi Prize Paper Award.