Capacity Bounds and Power Allocation for Wireless Relay Channels

Anders Høst-Madsen, *Senior Member, IEEE,* and Junshan Zhang, *Member, IEEE*

***Abstract*—We consider three-node wireless relay channels in a Rayleigh-fading environment. Assuming transmitter channel state information (CSI), we study upper bounds and lower bounds on the outage capacity and the ergodic capacity. Our studies take into account practical constraints on the transmission/reception duplexing at the relay node and on the synchronization between the source node and the relay node. We also explore power allocation. Compared to the direct transmission and traditional multihop protocols, our results reveal that optimum relay channel signaling can significantly outperform multihop protocols, and that power allocation has a significant impact on the performance.**

***Index Terms*—Channel capacity, cooperative diversity, ergodic capacity, power allocation, relay channel, wireless networks.**

I. Introduction

WIRELESS *ad hoc* networks consist of a number of terminals (nodes) communicating on a peer-to-peer basis, without the assistance of wired networks or centralized infrastructure. In such systems, the communications between nodes might take place through several intermediate nodes. In wireless communications, channel impairments that limit capacity include multipath fading, shadowing, and path loss.

Needless to say, high spectral efficiency is of vital importance in *ad hoc* wireless networks (see, e.g., [1]–[6]). One approach to increasing the capacity of wireless networks is to use *cooperative diversity*. Earlier works on cooperative communications can be found in [7] and [8]. Recently, cooperative diversity has been studied in [9]–[12] for cellular networks and in [13]–[17] for *ad hoc* networks. Roughly speaking, several terminals, each with one or more antennas, form a kind of "coalition" to cooperatively act as a large transmit or receive array. The channel can therefore exhibit some characteristics of the MIMO channel.

In this paper, we take steps to obtain a fundamental understanding of cooperative diversity. In particular, we investigate the three-node relay channel shown in Fig. 1. The desired transmission is from the source node (node 1) to the destination node (node 3), while the relay node (node 2) aids the communication by using its "capture" of the transmission between node 1 and node 3. The Gaussian relay channel was introduced by van der Meulen [18], [19] and used in connection with the Aloha system [20]. It was thoroughly analyzed by Cover and El Gamal in [21], where the capacity of the degraded relay channel was found, and upper and lower bounds for the general relay channel were presented. The work [22] used a simpler coding argument based on [7] to obtain the same achievable rate as in [21]. Very recently, partly spurred by the work on cooperative diversity, the relay channel has garnered much attention [23]–[36].

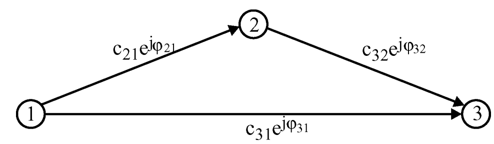

Fig. 1. The relay channel.

This paper focuses on wireless relay channels in a Rayleigh-fading environment. We take into account some practical constraints for wireless transceivers, such as synchronization and transmission duplexing. For convenience, we say that the relay node operates in the full duplex mode if the relay node can receive and transmit simultaneously on the same frequency channel. While there might exist some radio-frequency (RF) techniques making this possible [37], it is in general regarded unrealistic in practical systems, due to the dynamic range of incoming and outgoing signals and the bulk of ferroelectric components like circulators. We will therefore consider the cases where the relay node operates in a frequency-division (FD) manner or time-division (TD) manner.

We study upper bounds and lower bounds on the channel capacity. Specifically, we start with examining the capacity for the fixed channel gain case, building on which we explore the outage capacity. We then focus on the ergodic capacity.

The rest of the paper is organized as follows: we start with the channel model in Section II. In Section III, we present capacity bounds for the fixed channel gain case and outage capacity for the fading channel cases, and Section IV contains the main thrust on the ergodic capacity. Section V contains our conclusions. The proofs are relegated to Appendices A–E.

II. System Model

Consider the relay channel in Fig. 1. The transmitter at the source node sends a message $w \in \{1,\dots,M\}$ to the destination node, where the message is encoded into n symbols $x_1(w)[1],\dots,x(w)[n]$ and transmitted over the channel, under the power constraint $\frac{1}{n}\sum_{i=1}^{n} x_1[i]^2 \le P_1$. (The rate is $\log M/n$.) Let y_2 denote the received signal at the relay node. Building on *prior* received signals, the relay node then

Manuscript received September 20, 2003; revised February 1, 2005. The work of A. Høst-Madsen was supported in part by the National Science Foundation under Grant CCR-0329908. The work of J. Zhang was supported in part by the National Science Foundation under Grant ANI-0208135 and a grant from the Intel Research Council.

A. Høst-Madsen is with the Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, HI 96813 USA (e-mail: madsen@spectra.eng.hawaii.edu).

J. Zhang is with Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: junshan.zhang@asu.edu).

Communicated by G. Caire, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2005.847703

transmits a message x_2 that is intended to aid the transmission between source and destination, where $x_2[i]$ is encoded based on $\{y_2[i-1],\ldots,y_2[1])\}$, subject to the power constraint $\frac{1}{n}\sum_{i=1}^{n} x_2[i]^2 \le P_2$. Thus, the received signals at the relay node and the destination node are given by

$$\begin{aligned} y_2[i] &= c_{21}e^{j\varphi_{21}}x_1[i] + z_2[i] \\ y_3[i] &= c_{31}e^{j\varphi_{31}}x_1[i] + c_{32}e^{j\varphi_{32}}x_2[i] + z_3[i] \end{aligned} \quad (1)$$

where z_2 and z_3 are independent additive white Gaussian noise with unit variance (after normalization), and $\{c_{ik}e^{j\varphi_{ik}}\}$ are the channel gains with amplitude and phase separated, and modeled as independent (flat) fading processes.

A few more words on relaying techniques. As noted before, we say that the relay node operates in the full duplex mode if the relay can receive and transmit simultaneously on the same frequency channel. In contrast, if the relay node operates in a TD manner, then for a given time window D, the relay node is in the receive mode for a fraction of the time αD (we call this period the relay-receive period), and in the transmit mode for the rest $(1-\alpha)D$ (we call this period the relay-transmit period). Similarly, if the relay node operates in an FD manner, the bandwidth W is divided into a bandwidth of αW over which the relay node listens, and a bandwidth of $(1-\alpha)W$ over which the relay node transmits. The destination node listens over the whole bandwidth W. Clearly, from an information-theoretic point of view, the TD mode and the FD mode are equivalent for the fixed channel gain case. In fading channels, however, the TD mode has an advantage over the FD mode because α can be adjusted to the instantaneous channel conditions, whereas α is usually fixed in the FD mode. Thus motivated, we focus on the TD relay channels.

In the synchronized channel model, it is assumed that all nodes have complete channel state information (CSI), i.e., each node knows the instantaneous values (magnitudes and phases) of all c_{ij} as well as their statistics. It is furthermore assumed that all nodes are perfectly synchronized. It is relatively straightforward to obtain symbol (timing) synchronization between different nodes; however, carrier synchronization requires phase-locking-separated microwave oscillators, which is very challenging in practical systems [38]. In light of this observation, we also consider the asynchronous channel model in which there is a random phase offset $\theta[i]$ between the source and the relay, and the corresponding received signal at the destination is

$$y_3[i] = c_{31}e^{j\varphi_{31}}x_1[i] + e^{j\theta[i]}c_{32}e^{j\varphi_{32}}x_2[i] + z_3[i] \quad (2)$$

where $\theta[i]$ is random and ergodic, and is uniformly distributed in $[0, 2\pi)$ (a model where $\theta[i]$ is constant during the transmission is considered in [39]). We assume that only the destination knows (i.e., can estimate) $\theta[i]$. Note that the phase factor $\theta[i]$ can be incorporated into φ_{32}, that is, $\varphi'_{32} = \theta[i] + \varphi_{32}$ would have the same distribution as φ_{32}. Modeled in this way, the phase φ_{32} is unknown at the source and the relay. In summary, in the asynchronous channel model, the source and the relay know the amplitudes c_{ik}, but not necessarily the phases φ_{ik}.

We now outline the cases to be considered. We consider all four combinations of full duplex/TD and synchronized/asynchronous transceiver models. We first consider the case where the channel gains are fixed, and each node has a certain average power constraint (per frame). We then consider outage capacity, also with an average power constraint per node. We finally consider ergodic capacity with optimum power allocation over time and among the nodes; and the issue we investigate here is: what is the minimum network power (energy) needed to transmit a given amount (say one bit) of information at a given rate under a bandwidth constraint, and how much can be saved (by the network in total) by using a relay?

As a baseline for comparison, we will consider two strategies commonly applied in networks: direct transmission between source and destination and multihop transmission (routing). By the latter we mean that a packet is first transmitted to the relay; the relay decodes the packet, re-encodes it, and transmits it to the destination in the next time slot; that is, the destination only "uses" the transmission from the relay.

A few words about notation. The function log denotes the base 2 logarithm. We define the average channel gain on each link by

$$s_{ij} = E[c_{ij}^2]. \quad (3)$$

For notational convenience, we let C^+ denote an upper bound on the channel capacity, and C^- a lower bound. We will use R to denote the rate of a specific signaling scheme, i.e., an achievable rate, which also constitutes a lower bound. A quantity (function, parameter) x used in an upper bound is denoted x^+ and x^- in a lower bound. Expressions common to upper and lower bounds are written as $C^{\pm}$, which involves quantities associated either with an upper or a lower bound.

III. Preliminary: The Fixed Channel Gain Case

We start with the bounds on the capacity of the relay channel, assuming fixed channel gain coefficients c_{ij}. The results will also be used to find the outage capacity in a fading channel.

A. Full Duplex Relaying

The capacity of the general memoryless relay channel was studied in [21] (see also [40]), and the full duplex relay channel falls into this class. In particular, using the "max-flow-min-cut" theorem [40, Theorem 14.10.1] or [21, Theorem 4] yields the following upper bound:

$$C^+ = \max_{f(X_1,X_2)} \min\{I(X_1;Y_2,Y_3|X_2), I(X_1,X_2;Y_3)\} \quad (4)$$

$$= \max_{0\le\beta\le1} \min\left\{\frac{1}{2}\log\left(1+(1-\beta)P_1\left(c_{21}^2+c_{31}^2\right)\right),\right.$$
$$\left.\frac{1}{2}\log\left(1+c_{31}^2P_1+c_{32}^2P_2+2\sqrt{\beta c_{31}^2c_{32}^2P_1P_2}\right)\right\}. \quad (5)$$

If the relay node decodes and re-encodes the message, the following rate is achievable [21, Theorem 1]:

$$R_1 = \max_{f(X_1,X_2)} \min\{I(X_1;Y_2|X_2), I(X_1,X_2;Y_3)\} \quad (6)$$

$$= \max_{0\le\beta\le1} \min\left\{\frac{1}{2}\log(1+(1-\beta)c_{21}^2P_1),\right.$$
$$\left.\frac{1}{2}\log\left(1+c_{31}^2P_1+c_{32}^2P_2+2\sqrt{\beta c_{31}^2c_{32}^2P_1P_2}\right)\right\}. \quad (7)$$

Note that if $c_{21} < c_{31}$, the above achievable rate boils down to the rate of the direct transmission between source and destination. On the other hand, applying [21, Theorem 6][1] to the Gaussian relay channel gives a rate

$$R_2 = \frac{1}{2}\log\left(1 + c_{31}^2 P_1 + \frac{c_{21}^2 P_1}{1 + \frac{c_{31}^2 P_1 + c_{21}^2 P_1 + 1}{c_{32}^2 P_2}}\right). \quad (8)$$

Then, the following rate serves as a lower bound:

$$C^- = \max\{R_1, R_2\}. \quad (9)$$

Upper and lower bounds in the asynchronous case can be found by setting $\beta = 0$ (the proof of this is similar to the proofs of Propositions 1 and 2 later). Interestingly, for the asynchronous case, the upper bound (5) and lower bound (7) meet if c_{32} is sufficiently large, indicating that the exact capacity of the relay channel can be found, which was first observed in [36].

B. Time-Division Relaying

In this section we will generalize the results in the previous section to TD relaying, which was not included in [21]. Recall that in the TD mode, given a time window D, the relay is in the receive mode for a fraction of the time αD (the relay-receive period), and in the transmit mode for $(1-\alpha)D$ (the relay-transmit period). The source node can transmit in both the relay-receive and the relay-transmit periods. Suppose that the source node transmits with power $P_1^{(1)}$ during the relay-receive period, and with power $P_1^{(2)}$ during the relay-transmit period; the relay node transmits with power P_2. Denote the corresponding capacity as $C_R(\alpha, P_1^{(1)}, P_1^{(2)}, P_2)$. We then have the following upper bound on the capacity of the TD relay channel. This is a relatively straightforward application of [40, Theorem 14.10.1] or [21, Theorem 4] (the proof can be found in Appendix A).

Proposition 1 (Upper Bound): The capacity of the TD relay channel is upper-bounded by

$$C^+ = \max_{0\le\beta\le 1} \min\left\{C_1^+(\beta), C_2^+(\beta)\right\} \quad (10)$$

with

$$C_1^+(\beta) = \frac{\alpha}{2}\log\left(1 + (c_{31}^2 + c_{21}^2)P_1^{(1)}\right) + \frac{1-\alpha}{2}\log\left(1 + (1-\beta)c_{31}^2 P_1^{(2)}\right) \quad (11)$$

$$C_2^+(\beta) = \frac{\alpha}{2}\log\left(1 + c_{31}^2 P_1^{(1)}\right) + \frac{1-\alpha}{2}\log\Big(1 + c_{31}^2 P_1^{(2)} + c_{32}^2 P_2 + 2\sqrt{\beta c_{31}^2 P_1^{(2)} c_{32}^2 P_2}\Big). \quad (12)$$

Furthermore, an upper bound for the asynchronous relay channel is obtained by putting $\beta = 0$.

Consider the case where the relay can decode what it has received during the relay-receive period, re-encode it, and transmit during the relay-transmit period. We have the following proposition on the achievable rate (we call this rate the *decode-forward rate*).

Proposition 2 (Achievable Rate: Decode-Forward): The capacity of the TD relay channel is lower-bounded by

$$C^- = \max_{0\le\beta\le 1} \min\left\{C_1^-(\beta), C_2^-(\beta)\right\}$$

with

$$C_1^-(\beta) = \frac{\alpha}{2}\log\left(1 + c_{21}^2 P_1^{(1)}\right) + \frac{1-\alpha}{2}\log\left(1 + (1-\beta)c_{31}^2 P_1^{(2)}\right) \quad (13)$$

$$C_2^-(\beta) = \frac{\alpha}{2}\log\left(1 + c_{31}^2 P_1^{(1)}\right) + \frac{1-\alpha}{2}\log\Big(1 + c_{31}^2 P_1^{(2)} + c_{32}^2 P_2 + 2\sqrt{\beta c_{31}^2 P_1^{(2)} c_{32}^2 P_2}\Big). \quad (14)$$

The optimum value of β can be found in closed form (see [42]). Furthermore, an achievable rate for the asynchronous relay channel is obtained by putting $\beta = 0$ in (2).

The proof of Proposition 2, relegated to Appendix A, shows that, as opposed to the full duplex channel of [21], block-Markov coding is not needed for the TD relay channel.

Instead of using "decode-forward," an alternative approach is to let the relay use Wyner–Ziv lossy source coding [41] on the received signal. The compressed signal is then transmitted to the destination using an (error-free) channel encoding. The following proposition gives the rate for both the synchronized and the asynchronous cases (the proof is in Appendix A). The proposition can be viewed as a generalization of [21, Theorem 6].

Proposition 3 (Achievable Rate: Compress-Forward): The rate R is achievable if

$$R = \frac{\alpha}{2}\log\left(1 + c_{31}^2 P_1^{(1)} + \frac{c_{21}^2 P_1^{(1)}}{1+\sigma_w^2}\right) + \frac{1-\alpha}{2}\log\left(1 + c_{31}^2 P_1^{(2)}\right) \quad (15)$$

where σ_w^2 is the "compression noise" given by

$$\sigma_w^2 = \frac{c_{21}^2 P_1^{(1)} + c_{31}^2 P_1^{(1)} + 1}{\left(\left(1 + \frac{c_{32}^2 P_2}{1 + c_{31}^2 P_1^{(2)}}\right)^{(1-\alpha)/\alpha} - 1\right)(c_{31}^2 P_1^{(1)} + 1)}. \quad (16)$$

If $c_{21} < c_{31}$, the decode-forward approach would not work, and only the compress-forward approach can be used. The compress-forward method can be used for all channels and always gives a rate gain over the direct transmission, but as c_{21} becomes larger compared to c_{31}, the decode-forward rate would be eventually larger than the compress-forward rate (see Fig. 2). In general, a higher rate is obtained by taking the maximum of the rates of Propositions 2 and 3.

In the above, we have assumed that the parameters α, $P_1^{(1)}$, $P_1^{(2)}$, P_2 are fixed. We now consider that the source node and the relay node are subject to average power constraints P_1 and P_2. Since the relay only transmits during the relay-transmit period

[1][21, Theorem 6] only applies to the discrete alphabet relay channel. However, using [21, eqs. (76a) and (76b)] for Gaussian distributions yields (8). Alternatively, the method of proof we use for Proposition 3 in the Gaussian case can also be applied to the full duplex relay channel and will result in (8) in the Gaussian case.

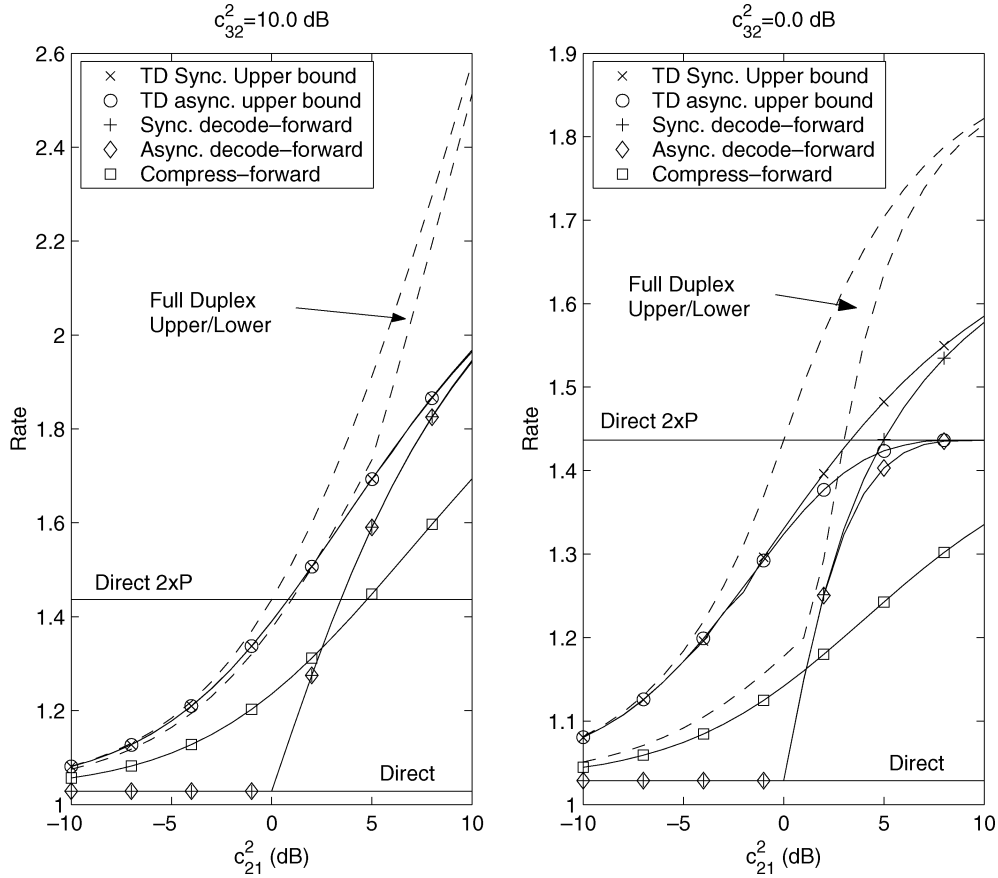

Fig. 2. Capacity bounds for TD relay channels with fixed channel coefficients, $P_1 = P_2 = 5$ dB and $c_{31} = 1$. The curve "Direct" is the rate for direct transmissions, and "Direct $2 \times P$" is the rate for direct transmissions when the source has twice the power.

of length $(1-\alpha)D$, it can use power $\frac{P_2}{1-\alpha}$ during the transmission period. Similarly, the source node transmits with power $\frac{\kappa P_1}{\alpha}$ during the relay-receive period and with power $\frac{(1-\kappa)P_1}{1-\alpha}$ during the relay-transmit period, where $0 \le \kappa \le 1$ so that the average power constraint is satisfied. The capacity is given by

$$C_{TD}(P_1, P_2) = \max_{0\le\alpha\le1, 0\le\kappa\le1} C_R\left(\alpha, \frac{\kappa P_1}{\alpha}, \frac{(1-\kappa)P_1}{1-\alpha}, \frac{P_2}{1-\alpha}\right). \quad (17)$$

While the optimum value of β in Propositions 2 and 1 can be found in closed form (see [42]), the optimization of α and κ needs to be done numerically. Upper and lower bounds to the capacity (17) can then be found using the propositions. Unfortunately, as opposed to the full duplex case, the upper and lower bounds never meet in nontrivial cases, even in the asynchronous case.

To illustrate the above results, Fig. 2 shows upper and lower bounds for different fixed values of c_{32}.

C. An Application to Outage Capacity

The bounds on capacity for fixed channel gains can be used to find bounds on outage capacity. The outage capacity of a fading channel is defined by [43]

$$C_o = \arg\max_C P(R(c_{ji}) \ge C) \ge p \quad (18)$$

where $R(c_{ji})$ is the rate for a specific realization of the fading process. Thus, to find the outage capacity, we first find the rate for specific values of c_{ji}, and the outage capacity can then be calculated as the $(1-p)$ percentile of the random variable $R(c_{ji})$. By using the upper and lower bounds from the previous section we can then find upper and lower bounds on the outage capacity.

Fig. 3 plots outage capacity versus the signal-to-noise ratio (SNR) in the direct link when all links experience independent Rayleigh fading. The curves are calculated based on an ensemble of $R(c_{ji})$ by Monte Carlo simulations for each value of the SNR. We also compare the results above with multihop transmission for which the achievable rate is

$$R = \max\left\{\frac{1}{2}\log\left(1+c_{31}^2 P_1\right), \min\left\{\frac{1}{4}\log\left(1+2c_{21}^2 P_1\right), \frac{1}{4}\log\left(1+2c_{32}^2 P_2\right)\right\}\right\}.$$

While the figure shows only two examples, they are typical for the many other examples we have considered. It should be noticed that upper and lower bounds are close, that the gain from synchronization is limited, and that relaying clearly outperforms the traditional multihop strategy, in particular at high SNR.

IV. Bounds on Ergodic Capacity

We now turn to studying the ergodic capacity for wireless relay channels. As stated in the Introduction, the problem here is what is the minimum network power (energy) needed to transmit a given amount (say one bit) of information at a certain

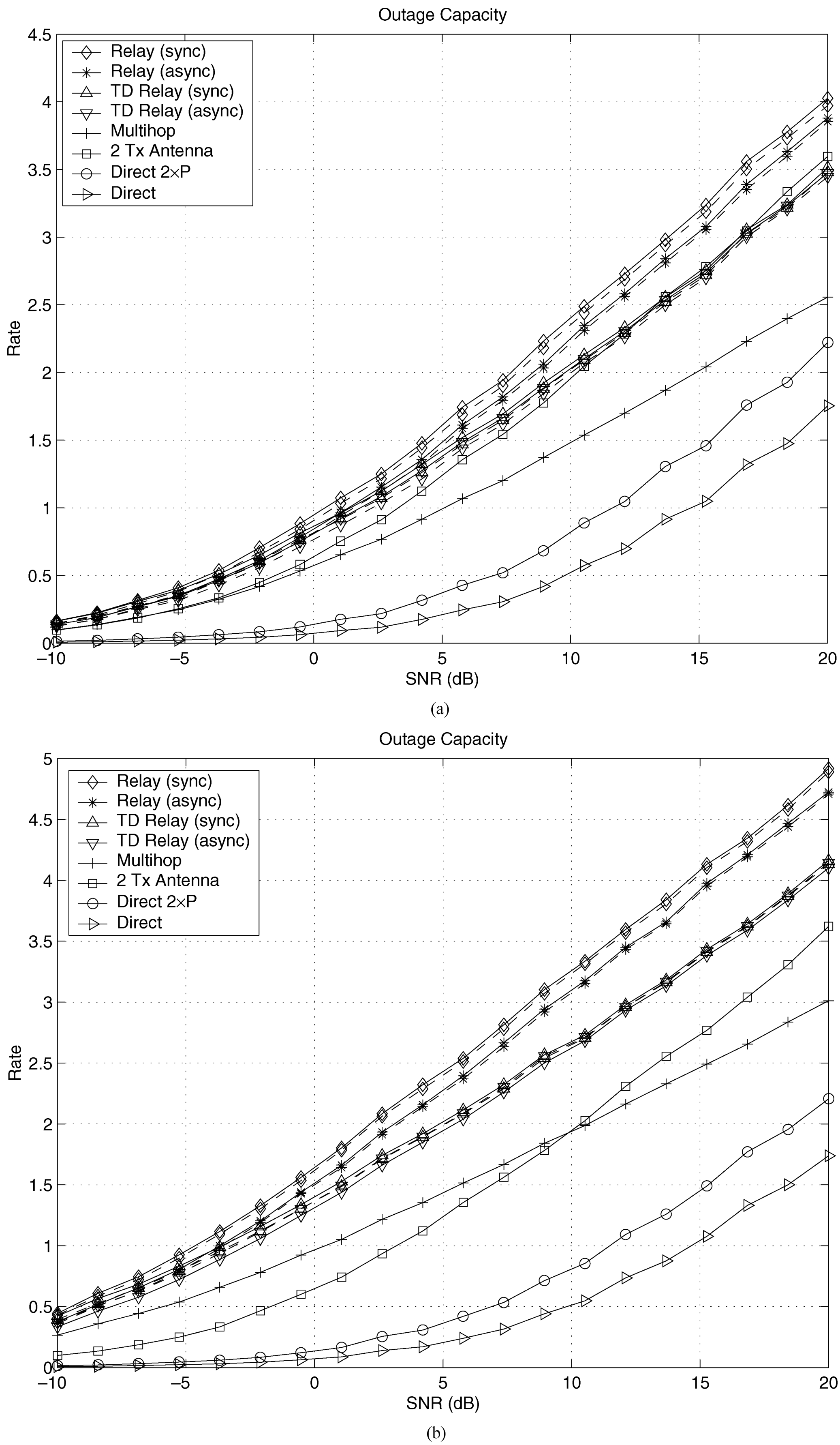

Fig. 3. Outage capacity for the relay channel with independent Rayleigh fading. Source and relay have equal power, and the x axis corresponds to the SNR for the direct link. The solid curve is for the upper bound, and the dashed curve is for the lower bound. Part (a) is for $s_{21} = s_{32} = 12$ dB and part (b) for $s_{21} = s_{32} = 20$ dB (s_{ij} is defined in (3)). "Direct $2 \times P$" is the rate for direct transmission when the source has twice the power. "2 Tx Antenna" is the rate for direct transmission when the source has two antennas.

rate under a bandwidth constraint. This is equivalent to finding the rate under a total power constraint, $P_1 + P_2 \leq P$.

A. Ergodic Capacity: The General SNR Case

First consider the full duplex case. Set $P_1 = P_1^{(1)} + P_1^{(2)}$, where $P_1^{(1)}$ is used for transmission to the relay, and $P_1^{(2)}$ for transmission to the destination (i.e., in earlier notation $P_1^{(1)} = (1-\beta)P_1$, $P_1^{(2)} = \beta P_1$). Power allocation dictates that $P_1^{(1)}$, $P_1^{(2)}$, and P_2 are functions of $\boldsymbol{c} = (c_{31}, c_{21}, c_{32})$. Using the upper bound (4), together with the fact that the optimum distribution for fixed $\boldsymbol{c}$ is Gaussian [44], [43], we have (a formal argument is given in the proof of Proposition 4)

$$C^+ = \max_{P_1^{(1)}, P_1^{(2)}, P_2} \min\left\{R_1^+(P_1^{(1)}), R_2^+(P_1^{(1)}, P_1^{(2)}, P_2)\right\} \tag{19}$$

with

$$R_1^+(P_1^{(1)}) = E\left[\frac{1}{2}\log\left(1 + P_1^{(1)}(\boldsymbol{c})\left(c_{21}^2 + c_{31}^2\right)\right)\right] \tag{20}$$

$$R_2^+(P_1^{(1)}, P_1^{(2)}, P_2) = E\left[\frac{1}{2}\log\left(1 + c_{31}^2\left(P_1^{(1)}(\boldsymbol{c}) + P_1^{(2)}(\boldsymbol{c})\right) + c_{32}^2 P_2(\boldsymbol{c}) + 2\sqrt{c_{31}^2 c_{32}^2 P_1^{(2)}(\boldsymbol{c}) P_2(\boldsymbol{c})}\right)\right]. \tag{21}$$

It remains to find the optimum functions $P_1^{(1)}(\boldsymbol{c})$, $P_1^{(2)}(\boldsymbol{c})$, and $P_2(\boldsymbol{c})$, and this is a power allocation problem. To this end, first fix $\boldsymbol{c}$. For fixed $P_j(\boldsymbol{c}) = P_1^{(2)}(\boldsymbol{c}) + P_2(\boldsymbol{c})$, the rate is maximized when we set

$$P_1^{(2)}(\boldsymbol{c}) = \frac{c_{31}^2}{c_{31}^2 + c_{32}^2} P_j(\boldsymbol{c})$$

and

$$P_2(\boldsymbol{c}) = \frac{c_{32}^2}{c_{31}^2 + c_{32}^2} P_j(\boldsymbol{c})$$

(a beamforming solution). We then get a simplified problem

$$C^+ = \max_{P_1^{(1)}, P_j} \min\left\{R_1^+(P_1^{(1)}), R_2^+(P_1^{(1)}, P_j)\right\} \tag{22}$$

with

$$R_1^+(P_1^{(1)}) = E\left[\frac{1}{2}\log\left(1 + P_1^{(1)}(\boldsymbol{c})\left(c_{21}^2 + c_{31}^2\right)\right)\right] \tag{23}$$

$$R_2^+(P_1^{(1)}, P_j) = E\left[\frac{1}{2}\log\left(1 + c_{31}^2 P_1^{(1)}(\boldsymbol{c}) + (c_{31}^2 + c_{32}^2) P_j(\boldsymbol{c})\right)\right]. \tag{24}$$

For the achievable rate, we similarly get a decode-forward solution (see Appendix B for details about the coding used)

$$C^- = \max_{P_1^{(1)}, P_j} \min\left\{R_1^-(P_1^{(1)}), R_2^-(P_1^{(1)}, P_j)\right\} \tag{25}$$

with

$$R_1^-(P_1^{(1)}) = E\left[\frac{1}{2}\log\left(1 + P_1^{(1)}(\boldsymbol{c}) \max\left\{c_{21}^2, c_{31}^2\right\}\right)\right] \tag{26}$$

$$R_2^-(P_1^{(1)}, P_j) = E\left[\frac{1}{2}\log\left(1 + c_{31}^2 P_1^{(1)}(\boldsymbol{c}) + (c_{31}^2 + c_{32}^2) P_j(\boldsymbol{c})\right)\right]. \tag{27}$$

TABLE I
POWER ALLOCATION ALGORITHM FOR THE FULL DUPLEX CASE. THE PROCEDURE CALCULATES $P_1^{(1)}$ AND P_j FOR SPECIFIC VALUES OF t AND μ

1. Set
$$P_1^{(1)} = 0, \qquad P_j = t\mu - \frac{1}{\tilde{c}(c_{31}^2, c_{32}^2)}$$
if
$$0 > \frac{\tilde{c}(c_{31}^2, c_{32}^2)(1-t)\mu}{c_{32}^2} - \frac{1}{c^{\pm}(c_{21}^2, c_{31}^2)}, \qquad 0 < P_j$$
exit;
2. Set
$$P_1^{(1)} = \frac{\tilde{c}(c_{31}^2, c_{32}^2)(1-t)\mu}{c_{32}^2} - \frac{1}{c^{\pm}(c_{21}^2, c_{31}^2)}, \qquad P_j = t\mu - \frac{1 + c_{31}^2 P_1^{(1)}}{\tilde{c}(c_{31}^2, c_{32}^2)}$$
if $P_1^{(1)} > 0$, $P_j > 0$ exit;
3. Set $P_j = 0$ and $P_1^{(1)}$ equal to the largest solution for x in
$$-\mu^{-1} c_{31}^2 c^{\pm}(c_{21}^2, c_{31}^2) x^2 + \left(c^{\pm}(c_{21}^2, c_{31}^2)(c_{31}^2 - \mu^{-1}) - \mu^{-1} c_{31}^2\right) x - \mu^{-1} + c_{31}^2 + (1-t) c^{\pm}(c_{21}^2, c_{31}^2) + t c_{31}^2 = 0$$
if $P_1^{(1)} > 0$ and
$$0 < t\mu - \frac{1}{\tilde{c}(c_{31}^2, c_{32}^2)} \quad \textbf{OR}$$
$$0 < (1-t)\log\left(1 + c^{\pm}(c_{21}^2, c_{31}^2) P_1^{(1)}\right) + t\log\left(1 + c_{31}^2 P_1^{(1)}\right) - (\mu \ln 2)^{-1} P_1^{(1)}$$
exit;
4. Set $P_1^{(1)} = P_j = 0$

For asynchronous signaling, we get a similar solution by replacing $c_{31}^2 + c_{32}^2$ by c_{32}^2 in (24) and (27). Since there is no beamforming gain, $P_1^{(1)} = P_1$, $P_1^{(2)} = 0$, and $P_j = P_2$ is the optimum. To unify the treatment of the different cases, we define

$$c^+(c_{21}^2, c_{31}^2) = c_{21}^2 + c_{31}^2 \tag{28}$$

$$c^-(c_{21}^2, c_{31}^2) = \max\{c_{21}^2, c_{31}^2\} \tag{29}$$

$$\tilde{c}(c_{31}^2, c_{32}^2) = \begin{cases} c_{31}^2 + c_{32}^2 & \text{synchronized} \\ c_{32}^2 & \text{asynchronous.} \end{cases} \tag{30}$$

The optimization problems for both upper and lower bounds can be solved by using a kind of "water-filling" techniques (see Appendix B), and the solutions are stated in the following proposition.

Proposition 4 (Ergodic Capacity: Full Duplex Relaying): Define

$$R_1^{\pm}(t,\mu) = E\left[\frac{1}{2}\log\left(1 + P_1^{(1)}(t,\mu,\boldsymbol{c})\left(c^{\pm}(c_{21}^2, c_{31}^2)\right)\right)\right] \tag{31}$$

$$R_2^{\pm}(t,\mu) = E\left[\frac{1}{2}\log\left(1 + c_{31}^2 P_1^{(1)}(t,\mu,\boldsymbol{c}) + \tilde{c}(c_{31}^2, c_{32}^2) P_j(t,\mu,\boldsymbol{c})\right)\right] \tag{32}$$

where the functions $P_1^{(1)}(t,\mu,\boldsymbol{c})$ and $P_j(t,\mu,\boldsymbol{c})$ are given in Table I. Let $(t^{\pm}, \mu^{\pm})$ be the solution to the equations

$$E\left[P_1^{(1)}(t,\mu,\boldsymbol{c}) + P_j(t,\mu,\boldsymbol{c})\right] = P \tag{33}$$

$$R_1^{\pm}(t,\mu) = R_2^{\pm}(t,\mu). \tag{34}$$

The capacity of the relay link C_r is then bounded by

$$R_1^{-}(t^-,\mu^-) \le C \le R_1^{+}(t^+,\mu^+). \tag{35}$$

In the TD relay channel, the relay cannot receive while it transmits. Accordingly, an upper bound on the capacity and a lower bound can therefore be obtained by the following modification of the full duplex problem:

$$C^{\pm} = \max_{P_1^{(1)},P_j} \min\left\{R_1^{\pm}(P_1^{(1)}), R_2^{\pm}(P_1^{(1)},P_j)\right\} \tag{36}$$

with

$$\begin{aligned} R_1^{\pm}(P_1^{(1)}) &= E\left[\frac{1}{2}\log\left(1+P_1^{(1)}(\boldsymbol{c})c^{\pm}\left(c_{21}^2,c_{31}^2\right)\right)\middle| P_j(\boldsymbol{c})=0\right] \\ &\quad + E\left[\frac{1}{2}\log\left(1+P_1^{(1)}(\boldsymbol{c})c_{31}^2\right)\middle| P_j(\boldsymbol{c})>0\right] \end{aligned} \tag{37}$$

$$R_2^{\pm}(P_1^{(1)},P_j) = E\left[\frac{1}{2}\log\left(1+c_{31}^2P_1^{(1)}(\boldsymbol{c})+(c_{31}^2+c_{32}^2)P_j(\boldsymbol{c})\right)\right]. \tag{38}$$

The details of the above optimization can be found in Appendix D. The solutions are stated as follows.

Proposition 5 (Ergodic Capacity: Time-Division Relaying): Define

$$\begin{aligned} R_1^{\pm}(t,\mu) &= E\left[\frac{1}{2}\log\left(1+P_1^{(1)}(t,\mu,\boldsymbol{c})\left(c^{\pm}(c_{21}^2,c_{31}^2)\right)\right)\middle| P_j(t,\mu,\boldsymbol{c})=0\right] \\ &\quad + E\left[\frac{1}{2}\log\left(1+P_1^{(1)}(t,\mu,\boldsymbol{c})c_{31}^2\right)\middle| P_j(t,\mu,\boldsymbol{c})>0\right] \end{aligned} \tag{39}$$

$$R_2^{\pm}(t,\mu) = E\left[\frac{1}{2}\log\left(1+c_{31}^2P_1^{(1)}(t,\mu,\boldsymbol{c})+\tilde{c}(c_{31}^2,c_{32}^2)P_j(t,\mu,\boldsymbol{c})\right)\right] \tag{40}$$

where $P_1^{(1)}(t,\mu,\boldsymbol{c})$, and $P_j(t,\mu,\boldsymbol{c})$ are given in Table II. Let $(t^{\pm},\mu^{\pm})$ be the solution to the equations

$$E\left[P_1^{(1)}(t,\mu,\boldsymbol{c})+P_j(t,\mu,\boldsymbol{c})\right] = P \tag{41}$$

$$R_1^{\pm}(t,\mu) = R_2^{\pm}(t,\mu). \tag{42}$$

The capacity of the relay link C_r is then bounded by

$$R_1^{-}(t^-,\mu^-) \le C \le R_1^{+}(t^+,\mu^+). \tag{43}$$

We also consider the asynchronous channel without power allocation; by this we mean that the network does not allocate power based on instantaneous channel state, while it still allocates power between source and relay based on channel statistics. It is then easily seen that the capacity bounds for full duplex are given by

$$C^{\pm} = \max_{\kappa\in[0,1]} \min\left\{E\left[\frac{1}{2}\log\left(1+\kappa P c^{\pm}\left(c_{21}^2,c_{31}^2\right)\right)\right], E\left[\frac{1}{2}\log\left(1+c_{31}^2\kappa P+c_{32}^2(1-\kappa)P\right)\right]\right\}. \tag{44}$$

TABLE II

POWER ALLOCATION ALGORITHM FOR THE TD CASE. THE PROCEDURE CALCULATES $P_1^{(1)}$ AND P_j FOR SPECIFIC VALUES OF t AND μ

1. Set
$$P_{1,1}^{(1)} = 0$$
$$P_{j,1} = t\mu - \frac{1}{\tilde{c}(c_{31}^2,c_{32}^2)}$$
if
$$0 > \frac{\tilde{c}(c_{31}^2,c_{32}^2)(1-t)\mu}{c_{32}^2} - \frac{1}{c_{31}^2}$$
$$0 < P_j$$
go to 3.
2. Set
$$P_{1,1}^{(1)} = \frac{\tilde{c}(c_{31}^2,c_{32}^2)(1-t)\mu}{c_{32}^2} - \frac{1}{c_{31}^2}$$
$$P_{j,1} = t\mu - \frac{1+c_{31}^2P_1^{(1)}}{\tilde{c}(c_{31}^2,c_{32}^2)}$$
3. Set $P_{j,2}=0$ and $P_{1,2}^{(1)}$ equal to the largest solution for x in
$$-\mu^{-1}c_{31}^2c^{\pm}(c_{21}^2,c_{31}^2)x^2 + \left(c^{\pm}(c_{21}^2,c_{31}^2)(c_{31}^2-\mu^{-1})-\mu^{-1}c_{31}^2\right)x - \mu^{-1}+c_{31}^2+(1-t)c^{\pm}(c_{21}^2,c_{31}^2)+tc_{31}^2 = 0$$
if $P_1^{(1)}>0$ and
$$0 < t\mu - \frac{1}{\tilde{c}(c_{31}^2,c_{32}^2)} \quad \textbf{OR}$$
$$0 < (1-t)\log\left(1+c^{\pm}(c_{21}^2,c_{31}^2)P_1^{(1)}\right) + t\log\left(1+c_{31}^2P_1^{(1)}\right) - (\mu\ln 2)^{-1}P_1^{(1)}$$
go to 5
4. Put $P_{1,2}^{(1)} = P_{j,2} = 0$
5. Calculate
$$R_{t,1} = (1-t)\log\left(1+P_{1,1}^{(1)}c_{31}^2\right) + t\log\left(1+c_{31}^2P_{1,1}^{(1)}+\tilde{c}(c_{31}^2,c_{32}^2)P_{j,1}\right) - (\mu\ln 2)^{-1}\left(P_{1,1}^{(1)}+P_{j,2}\right)$$
$$R_{t,2} = (1-t)\log\left(1+P_{1,2}^{(1)}c^{\pm}\left(c_{21}^2,c_{31}^2\right)\right) + t\log\left(1+c_{31}^2P_{1,2}^{(1)}\right) - (\mu\ln 2)^{-1}P_{1,2}^{(1)}$$
if $P_{1,1}^{(1)} \le 0$ or $R_{t,2} > R_{t,1}$ put $P_1^{(1)} = P_{1,2}^{(1)}, P_j = P_{j,2}$, otherwise $P_1^{(1)} = P_{1,1}^{(1)}, P_j = P_{j,1}$

For the TD case, we also assume that the network allocates the relay-receive and relay-transmit time intervals based on average channel conditions, and we then get the bounds

$$\begin{aligned} C^{\pm} = \max_{\alpha\in[0,1],\kappa\in[0,1]} \min &\left\{E\left[\alpha\frac{1}{2}\log\left(1+\kappa P c^{\pm}\left(c_{21}^2,c_{31}^2\right)\right) + (1-\alpha)\frac{1}{2}\log\left(1+\kappa P c_{31}^2\right)\right],\right. \\ &\left. E\left[\alpha\frac{1}{2}\log\left(1+\kappa P c_{31}^2\right) + (1-\alpha)\frac{1}{2}\log\left(1+c_{31}^2\kappa P+c_{32}^2\frac{1-\kappa}{1-\alpha}P\right)\right]\right\}. \end{aligned} \tag{45}$$

We now illustrate our findings on ergodic capacity via numerical results. We compare our results with simple multihop forwarding. For fair comparison, we allow the network to optimize power allocation and relay receive/transmit durations base on channel statistics. Accordingly, we have that

$$R = \max_{\kappa\in[0,1],\alpha\in[0,1]} \min\left\{E\left[\alpha\frac{1}{2}\log\left(1+c_{21}^2\kappa\frac{P}{\alpha}\right)\right],\right.$$

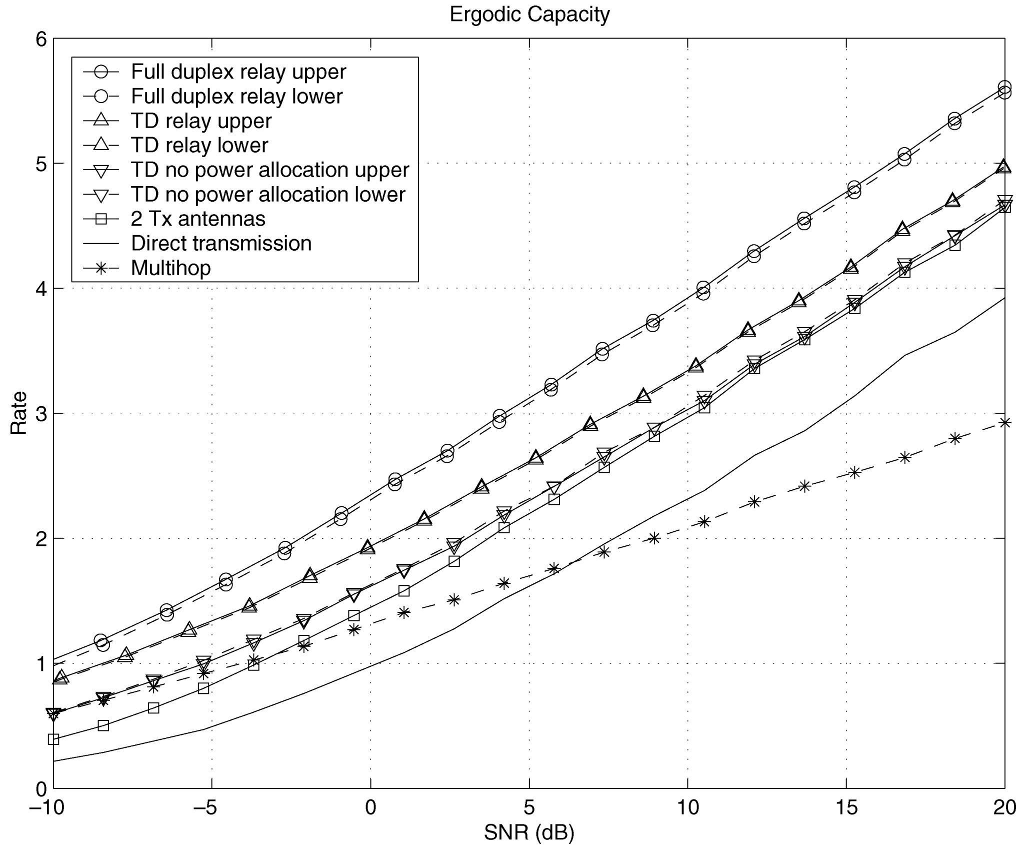

Fig. 4. Ergodic capacity of the relay channel in a Rayleigh fading channel with $s_{21} = s_{32} = 12$ dB relative to s_{31}, which is also the SNR. All curves are for synchronized transmission, except the curve for no power allocation.

$$E\left[(1-\alpha)\frac{1}{2}\log\left(1+c_{32}^2(1-\kappa)\frac{P}{1-\alpha}\right)\right]\Bigg\}. \quad (46)$$

Fig. 4 shows the different rates versus SNR for $s_{21} = s_{32} = 12$ dB relative to s_{31} (see (3); all channel gains are relative to s_{31}) and all paths are independently Rayleigh fading. The curves are generated using Monte Carlo simulation to generate an ensemble of values $\boldsymbol{c}$ which is then used to calculate upper and lower bounds.

We note that the upper and lower bounds are close in many cases of interest. Moreover, it should be noted that traditional multihop signaling is clearly suboptimum compared to the relay signaling, in the sense that the slope of the rate increase is smaller than that for the relay signaling, especially at high SNR. For SNR larger than 5 dB, all curves (except multihop) are approximately parallel lines. The offset between the lines (either in terms of rate or power) therefore can be used to characterize the gain of relaying at (reasonably) high SNR. It turns out that this offset can be calculated explicitly, and this is studied in the following section.

B. Ergodic Capacity in the High-SNR Regime

Clearly, the solutions for the bounds and the power allocation for general cases are complicated. It turns out, however, that in the high-SNR regime there exist simple closed-form solutions (cf. the high-SNR results for the multiple-input multiple-output (MIMO) channel in [45]). As mentioned in the previous section, in the high-SNR regime, the capacity of the relay channel, $C_{\text{relay}}(\text{SNR})$ satisfies

$$C_{\text{relay}}(\text{SNR}) \approx C_{\text{direct}}(\text{SNR}) + R_g$$

where $C_{\text{direct}}(\text{SNR})$ is the capacity for the direct link and R_g is a constant. To formalize this, we define the asymptotic relay gain by

$$R_g = \lim_{\text{SNR}\to\infty}(C_{\text{relay}}(\text{SNR}) - C_{\text{direct}}(\text{SNR})). \quad (47)$$

By SNR $\to \infty$, we mean that the noise power $\sigma^2 \to 0$ while the total power constraint is kept constant. Our goal is to find bounds for R_g. It turns out that at high SNR, the complicated power allocation rules in Tables I and II become much simpler.

Since only bounds on the capacity have been found, it is not clear that the limit (47) exists for the capacity. More formally, we say that $R_g^{\pm}$ are bounds for the relay gain if they satisfy

$$\limsup_{\text{SNR}\to\infty}(C_{\text{relay}}(\text{SNR}) - C_{\text{direct}}(\text{SNR})) \le R_g^+ \quad (48)$$

$$\liminf_{\text{SNR}\to\infty}(C_{\text{relay}}(\text{SNR}) - C_{\text{direct}}(\text{SNR})) \ge R_g^-. \quad (49)$$

Another way to look at the gain by relaying is in terms of the additional power needed for direct transmission to get the same rate as with a relay; we call this high-SNR limit the asymptotic power gain. It is easily seen that the relationship between the asymptotic relay gain and the asymptotic power gain (in decibels) is

$$P_g(\text{in dB}) = \frac{20}{\log 10} R_g \approx 6R_g. \tag{50}$$

We will now state the main results in the high-SNR regime. The asymptotic relay gain for full duplex is found in Appendix C and stated in the following proposition.

Proposition 6: High-SNR Relay Gain: Full Duplex Synchronized Case: In the high-SNR regime (i.e., $\sigma^2 \to 0$) the optimum values of $P_1^{(1)}$ and P_j for both upper and lower bounds in the synchronized case converge toward

$$P_1^{(1)}(\boldsymbol{c}) = \begin{cases} \frac{c_{31}^2 + c_{32}^2}{c_{32}^2}(1-t)P, & \frac{c_{31}^2}{c_{32}^2} < \kappa \\ P, & \text{otherwise} \end{cases}$$

$$P_j(\boldsymbol{c}) = \begin{cases} \left(t - \frac{c_{31}^2}{c_{32}^2}(1-t)\right)P, & \frac{c_{31}^2}{c_{32}^2} < \kappa \\ 0, & \text{otherwise} \end{cases} \tag{51}$$

$$\kappa = \frac{t}{1-t}. \tag{52}$$

In a Rayleigh-fading channel, κ is given by (for $s_{31} \neq s_{32}$)

$$\kappa^- = \frac{s_{21}}{s_{32}} \tag{53}$$

for the achievable rate, and

$$\kappa^+ = \frac{2^{l(s_{21}, s_{31})} - s_{31}}{s_{32}} \tag{54}$$

for the upper bound. Here

$$l(a,b) = \begin{cases} \frac{a \log a - b \log b}{a - b}, & a \neq b \\ \log a + (\ln 2)^{-1}, & a = b. \end{cases} \tag{55}$$

In addition, the asymptotic relay gain is bounded by

$$R_g^{\pm} = \frac{1}{2} \frac{s_{32}}{s_{31} - s_{32}} \log\left(\frac{1 + \kappa^{\pm}}{1 + \kappa^{\pm}\frac{s_{32}}{s_{31}}}\right), \qquad s_{31} \neq s_{32}. \tag{56}$$

Note that power allocation rule (51) does not depend on the characteristic of the fading channel (i.e., it is not specific to Rayleigh fading), and it is the same for upper and lower bounds. (The above proposition is for the case $s_{31} \neq s_{32}$; results for $s_{31} = s_{32}$ can be found by taking the limit $s_{31} \to s_{32}$.)

For asynchronous signalling, we have the following.

Proposition 7: High-SNR Relay Gain: Full Duplex Asynchronous Case: In the high-SNR regime (i.e., $\sigma^2 \to 0$), the optimum values of P_1 and P_2 for both upper and lower bounds in the asynchronous case converge toward

$$P_1(\boldsymbol{c}) = \begin{cases} \frac{c_{32}^2}{c_{32}^2 - c_{31}^2}(1-t)P, & \frac{c_{31}^2}{c_{32}^2} < t \\ P, & \text{otherwise} \end{cases}$$

$$P_2(\boldsymbol{c}) = \begin{cases} \left(t - \frac{c_{31}^2}{c_{32}^2 - c_{31}^2}(1-t)\right)P, & \frac{c_{31}^2}{c_{32}^2} < t \\ ,\ \ 0 & \text{otherwise.} \end{cases} \tag{57}$$

In Rayleigh fading, the value of t is given by

$$t^{\pm} = \frac{2^{x^{\pm}} - s_{31}}{2^{x^{\pm}} + s_{32}} \tag{58}$$

where

$$x^- = \frac{(s_{31} + s_{32})\log(s_{21} + s_{31}) - s_{31}\log(s_{31})}{s_{32}} \tag{59}$$

for the achievable rate, and

$$x^+ = \frac{(s_{31} + s_{32}) l(s_{21}, s_{31}) - s_{31}\log(s_{31})}{s_{32}} \tag{60}$$

for the upper bound. The asymptotic rate gain is bounded by

$$R_g^{\pm} = \frac{1}{2}\log\left(t^{\pm}\frac{s_{32}}{s_{31}} + 1\right). \tag{61}$$

The preceding results give upper and lower bounds on the capacity (not the exact capacity). However, it is possible to bound the gap between the upper and lower bounds for both the synchronized and asynchronous cases (Propositions 6 and 7).

Corollary 1: The rate difference and the power difference between the upper and lower bounds in the Rayleigh-fading case satisfy

$$\text{rate: } R_g^+ - R_g^- \leq \frac{\ln 2 - 1}{2 \ln 2} \approx 0.221 \tag{62}$$

$$\text{power: } P_g^+ - P_g^- \leq \frac{10(\ln 2 - 1)}{\log 10 \ln 2} \approx 1.33 \text{ dB} \tag{63}$$

and the maximum of the differences occurs when $(s_{21}, s_{32}) = (s_{31}, \infty)$.

For most values of s_{ij}, $P_g^+ - P_g^-$ is smaller than 1.33 dB. A closer examination of the difference reveals that it is only in a small neighborhood of the ridge $s_{21} = s_{31}$ that the difference can be significant.

The achievable rates in Propositions 6 and 7 are based on using decode-forward. As seen for the fixed-channel case in Section III, a compress-forward scheme, based on Wyner–Ziv rate-distortion theory, can yield a higher rate than the decode-forward scheme in certain regions. We now consider a compress-forward scheme for ergodic capacity. As shown by Propositions 6 and 7, finding ergodic capacity is much simplified in the high-SNR regime, and we consider the compress-forward scheme only in this regime. The power control rule (57) combined with Gaussian Wyner–Ziv compression is applied for each given $\boldsymbol{c}$. The details of the approach can be found in Appendix E. We have the following proposition.

Proposition 8: High-SNR Relay Gain: Full Duplex Compress-Forward: Define the function $\Sigma(t)$ as the solution to the equation

$$\frac{\Sigma(t)s_{21}}{\Sigma(t)s_{21} - (1+\Sigma(t))s_{31}} \log\left(\frac{\Sigma(t)s_{21}}{(1+\Sigma(t))s_{31}}\right) + \log(1+\Sigma(t)) = \frac{s_{32}}{s_{31}+s_{32}} \log\left(\frac{1 + \frac{s_{32}}{s_{31}}t}{1-t}\right). \tag{64}$$

For Rayleigh fading, the following rate gain is achievable using compress-forward in the high-SNR regime (i.e., $\sigma^2 \to 0$):

$$R_g = \frac{1}{2}\max_{t \in [0,1]} \left\{ \frac{\Sigma(t)s_{21}}{\Sigma(t)s_{21} - (1+\Sigma(t))s_{31}} \log\left(\frac{\Sigma(t)s_{21}}{(1+\Sigma(t))s_{31}}\right) + \frac{s_{32}\log(1-t) + s_{31}\log\left(1 + t\frac{s_{32}}{s_{31}}\right)}{s_{31}+s_{32}} \right\}. \tag{65}$$

Numerical evaluation of this compress-forward rate shows that it can give a gain of around 0.5 dB over decode-forward in the region $s_{32} > 5$ dB, $s_{21} < 0$ dB, but otherwise is inferior to decode-forward.

For the asynchronous channel without power allocation, bounds for the asymptotic relay gain in Rayleigh fading can be easily found to be

$$R_g^+ = \frac{1}{2}\max_{\kappa\in[0,1]}\min\{l(\kappa s_{21},\kappa s_{31}), l(\kappa s_{31},(1-\kappa)s_{32})\} - \frac{1}{2}\log s_{31} \quad (66)$$

$$R_g^- = \frac{1}{2}\max_{\kappa\in[0,1]}\min\{\log((s_{21}+s_{31})\kappa), l(\kappa s_{31},(1-\kappa)s_{32})\} - \frac{1}{2}\log s_{31} \quad (67)$$

where the maximization over κ must be done numerically.

We now turn to the TD case. As in the full duplex case, we get a simple power allocation rule for the synchronized case for high SNR

$$P_1^{(1)}(\boldsymbol{c}) = \begin{cases} \frac{c_{31}^2+c_{32}^2}{c_{32}^2}(1-t)P, & s(t,\boldsymbol{c})>0 \text{ and } \frac{c_{31}^2}{c_{32}^2} < \frac{t}{1-t} \\ P, & \text{otherwise} \end{cases}$$

$$P_j(\boldsymbol{c}) = \begin{cases} \left(t - \frac{c_{31}^2}{c_{32}^2}(1-t)\right)P, & s(t,\boldsymbol{c})>0 \text{ and } \frac{c_{31}^2}{c_{32}^2} < \frac{t}{1-t} \\ 0, & \text{otherwise} \end{cases} \quad (68)$$

$$s(t,\boldsymbol{c}) = (1-t)\log\left(\frac{c_{31}^2(c_{31}^2+c_{32}^2)}{c_{32}^2 c^{\pm}(c_{21}^2,c_{31}^2)}(1-t)\right) + t\log\left(\frac{c_{31}^2+c_{32}^2}{c_{31}^2}t\right). \quad (69)$$

Unfortunately, because of the complicated condition for joint transmission (69) we cannot evaluate the Rayleigh integrals analytically, and we have to find t as well as the asymptotic relay gain R_g numerically, see Appendix D.

For the asynchronous TD case we get the asymptotic power allocation rule

$$P_1(\boldsymbol{c}) = \begin{cases} \frac{c_{32}^2}{c_{32}^2-c_{31}^2}(1-t)P, & s(t,\boldsymbol{c})>0 \text{ and } \frac{c_{31}^2}{c_{32}^2} < t \\ (P,0), & \text{otherwise} \end{cases}$$

$$P_2(\boldsymbol{c}) = \begin{cases} \left(t - \frac{c_{31}^2}{c_{32}^2-c_{31}^2}(1-t)\right)P, & s(t,\boldsymbol{c})>0 \text{ and } \frac{c_{31}^2}{c_{32}^2} < t \\ (P,0), & \text{otherwise} \end{cases} \quad (70)$$

$$s(t,\boldsymbol{c}) = (1-t)\log\left(\frac{c_{31}^2 c_{32}^2}{(c_{32}^2-c_{31}^2)c^{\pm}(c_{21}^2,c_{31}^2)}(1-t)\right) + t\log\left(\frac{c_{32}^2}{c_{31}^2}t\right). \quad (71)$$

Since we do not have closed-form expressions for the bounds in the TD case, we cannot exactly bound the difference between upper and lower bounds as in Corollary 1. However, extensive numerical evaluation of the bounds show that the difference also seems to be bounded, and that

$$P_g^+ - P_g^- \leq 1\text{ dB}. \quad (72)$$

For the TD case without power allocation, we can find the following bounds for the asymptotic relay gain in Rayleigh fading using the results from Propositions 1-2 and taking limits:

$$R_g^+ = \frac{1}{2}\max_{\kappa\in[0,1],\alpha\in[0,1]}\min\left\{\alpha l(\kappa s_{21},\kappa s_{31}) + (1-\alpha)\log(\kappa s_{31}),\ \alpha\log(\kappa s_{31}) + (1-\alpha)l\left(\kappa s_{31},\frac{1-\kappa}{1-\alpha}s_{32}\right)\right\} - \frac{1}{2}\log s_{31} \quad (73)$$

$$R_g^- = \frac{1}{2}\max_{\kappa\in[0,1],\alpha\in[0,1]}\min\left\{\alpha\log(\kappa s_{21}+\kappa s_{31}) + (1-\alpha)\log(\kappa s_{31}),\ \alpha\log(\kappa s_{31}) + (1-\alpha)l\left(\kappa s_{31},\frac{1-\kappa}{1-\alpha}s_{32}\right)\right\} - \frac{1}{2}\log s_{31}. \quad (74)$$

C. Discussions on the High-SNR Case

The asymptotic relay gain characterizes the gain from relaying at reasonably high SNR. Furthermore, since the difference between upper and lower bounds is bounded (see Corollary 1), we can use the achievable rates to characterize the gain of relaying. Figs. 5 and 6 compare the asymptotic relay gain for a number of different signaling schemes over Rayleigh fading. The comparison is done for values of s_{21} and s_{32} between − 10 dB and 20 dB. Fig. 5(a) shows the relay gain for full duplex synchronized relaying. Figs. 5(b) and 6 show the loss by different constraints on the signaling.

V. Conclusion

In this paper, we have studied upper and lower bounds on the ergodic capacity and the outage capacity, for a variety of wireless relay channel models. In many cases of interest, the gap between the upper bounds and lower bounds is small compared to the relaying gain (as shown by both analytical and numerical examples), which sheds insights on the channel capacity. A number of conclusions can be drawn based on our results.

- Compared to direct transmissions without a relay, relay channel signaling yields performance gain, for both ergodic capacity and outage capacity.
- Optimal relaying outperforms traditional multihop protocols.
- Power allocation can yield a significant gain in wireless relay channels, in particular when the relay operates in the TD mode.
- Since transmitter CSI makes power allocation possible, transmitter CSI leads to higher rates, even at high SNR. This is in contrast to the case of the point-to-point single-antenna channel [44], [46], but is in line with the point-to-point MIMO channel [47], [48].

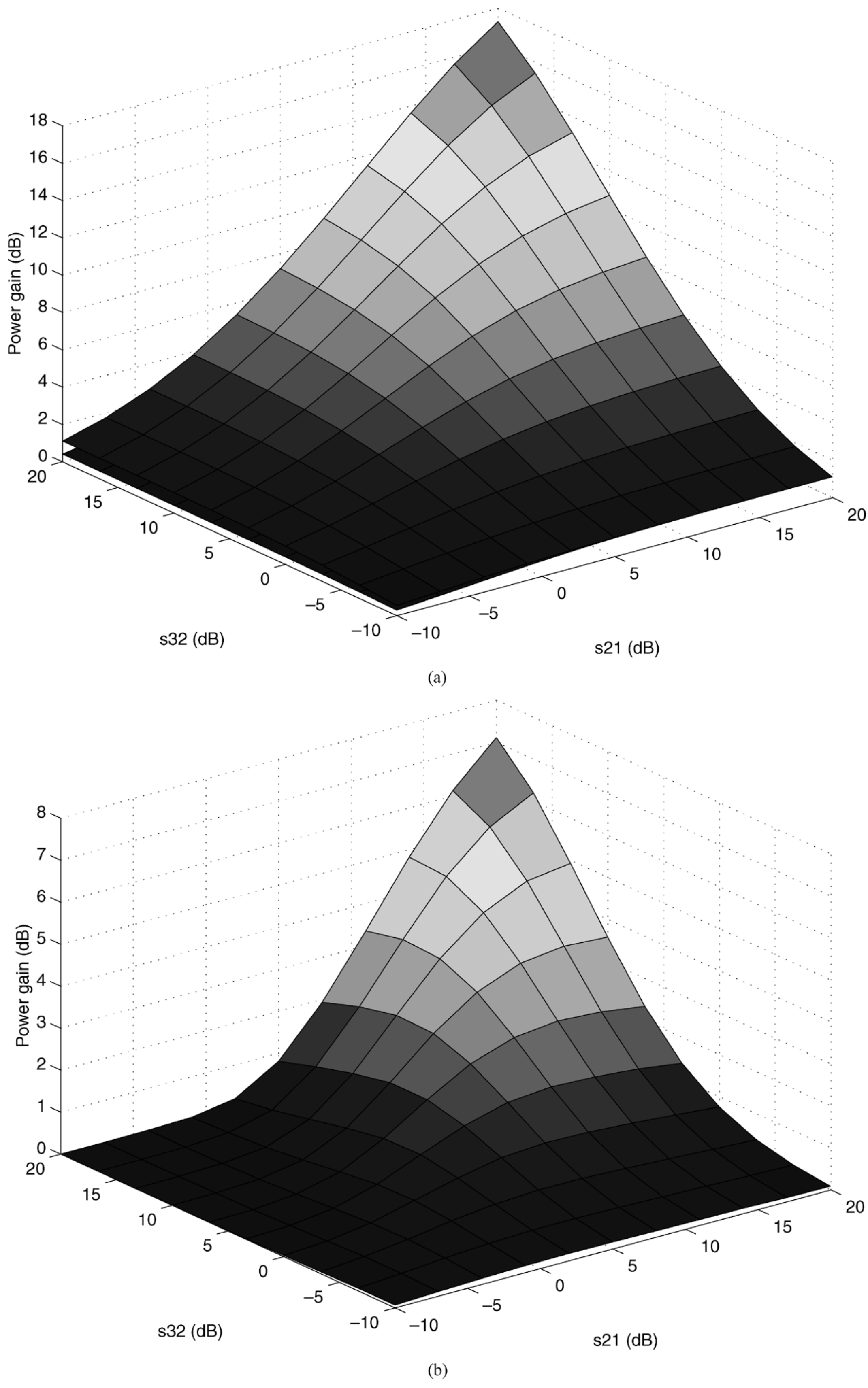

Fig. 5. (a) High-SNR relay gain for full duplex synchronized signaling. (b) Difference between full duplex and TD for synchronized signaling.

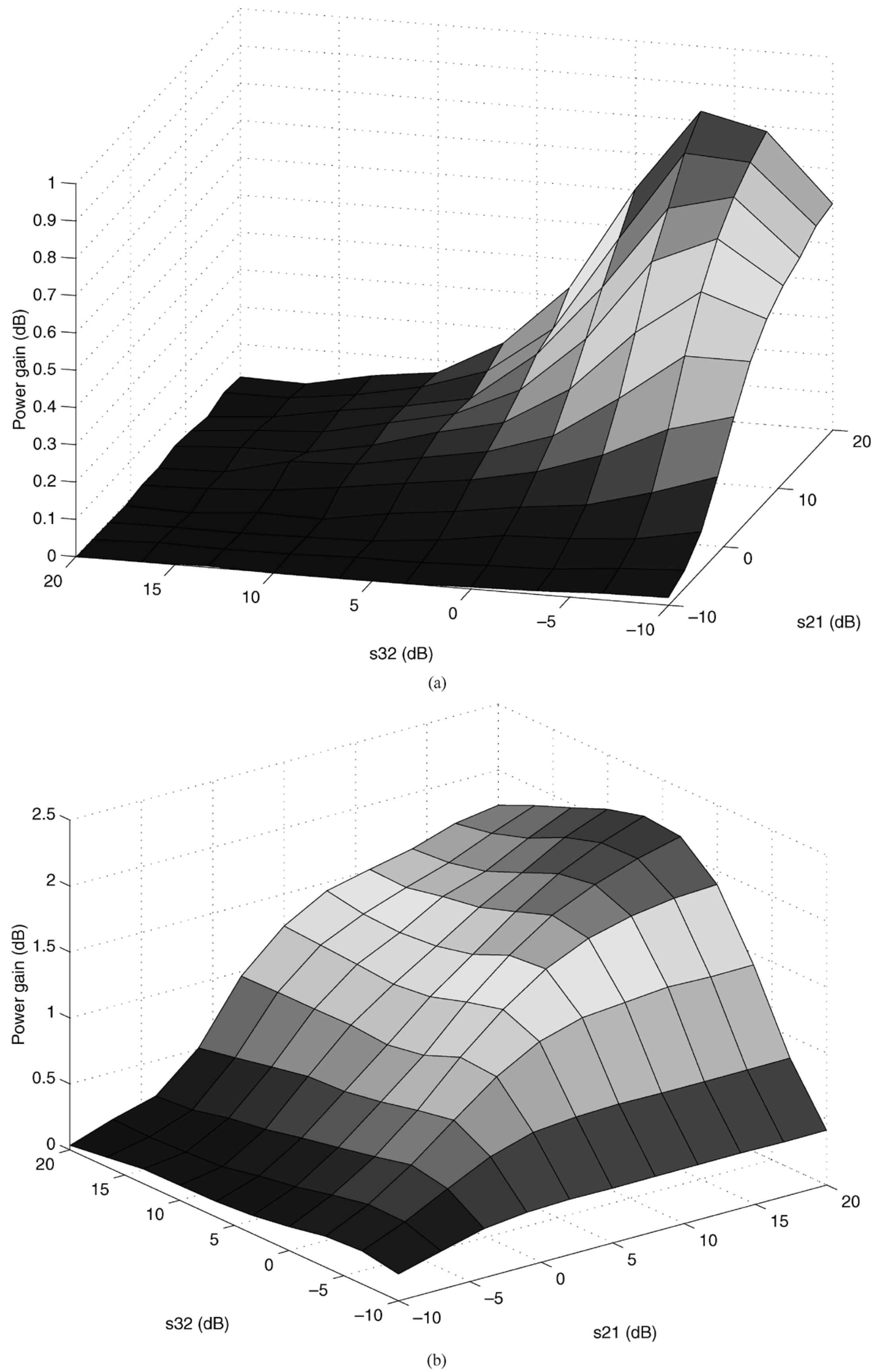

Fig. 6. High-SNR relay gain, achievable rate for TD. (a) Difference between synchronized and asynchronous signaling. (b) Difference between asynchronous signaling with power allocation and asynchronous signaling without power allocation.

APPENDIX A
PROOFS FOR THE FIXED CHANNEL GAIN CASE

Proof of Proposition 1

Let $X_{1,1}, X_{1,2}, \ldots, X_{1,n}$ be the transmitted sequence from node 1. The relay, node 2, receives the first $\lfloor \alpha n \rfloor$ transmitted signals as $Y_{2,1}, Y_{2,2}, \ldots, Y_{2,\lfloor \alpha n \rfloor}$, and after that transmits a sequence $X_{2,\lfloor \alpha n \rfloor+1}, X_{2,\lfloor \alpha n \rfloor+2}, \ldots, X_{2,n}$. The result in [21, Theorem 4] or [40, Theorem 14.10.1] gives the following nonsingle letter bound (i.e., [21, eq. (53)]) on the rate R:

$$nR \le \min \left\{ \sum_{k=1}^{n} I(X_{1,k}; Y_{2,k}, Y_{3,k} | X_{2,k}), \sum_{k=1}^{n} I(X_{1,k}, X_{2,k}; Y_{3,k}) \right\} + n\epsilon_n. \quad (75)$$

In the relay-receive period, the relay does not transmit so that $X_{2,k} = 0$. On the other hand, in the relay-transmit period, the relay does not receive, and therefore for $k > \lfloor \alpha n \rfloor$ I

$$(X_{1,k}; Y_{2,k}, Y_{3,k} | X_{2,k}) = I(X_{1,k}; Y_{3,k} | X_{2,k}).$$

We then get

$$\begin{aligned} nR \le \min \Bigg\{ & \sum_{k=1}^{\lfloor \alpha n \rfloor} I(X_{1,k}; Y_{2,k}, Y_{3,k} | X_{2,k} = 0) \\ & + \sum_{k=\lfloor \alpha n \rfloor+1}^{n} I(X_{1,k}; Y_{3,k} | X_{2,k}), \\ & \sum_{k=1}^{\lfloor \alpha n \rfloor} I(X_{1,k}; Y_{3,k} | X_{2,k} = 0) \\ & + \sum_{k=\lfloor \alpha n \rfloor+1}^{n} I(X_{1,k}, X_{2,k}; Y_{3,k}) \Bigg\} + n\epsilon_n. \quad (76) \end{aligned}$$

By letting $n \to \infty$ and using standard arguments as in [40] we get the following single-letter bound on the capacity:

$$\begin{aligned} C \le \max_{X_1^{(1)}, X_1^{(2)}, X_2} \min \Big\{ & \alpha I(X_1^{(1)}; Y_2^{(1)}, Y_3^{(1)} | X_2 = 0) \\ & + (1-\alpha) I(X_1^{(2)}; Y_3^{(2)} | X_2), \\ & \alpha I(X_1^{(1)}; Y_3^{(1)} | X_2 = 0) \\ & + (1-\alpha) I(X_1^{(2)}, X_2; Y_3^{(2)}) \Big\}. \end{aligned} \quad (77)$$

In the synchronized Gaussian case, the same argument as in [21] shows that this bound is maximized by letting $X_1^{(1)}$ be Gaussian with power $P_1^{(1)}$, X_2, and $X_1^{(2)}$ Gaussian with powers P_2 and $P_1^{(2)}$, respectively, and

$$E[X_1^{(2)} X_2^*] = \sqrt{\beta P_1^{(2)} P_2}.$$

Theorem 1 follows.

In the asynchronous case, the phases of the source and relay transmissions cannot be synchronized, i.e., the argument of the complex random variables $X_1^{(2)}$ and X_2 are independent. To derive an upper bound in this case, first notice that

$$\begin{aligned} I(X_1^{(2)}; Y_3^{(2)} | X_2) &= H(c_{31} e^{j\varphi_{31}} X_1^{(2)} + c_{32} e^{j\varphi_{32}} X_2 + Z_3 | X_2) \\ &\quad - H(Y_3^{(2)} | X_1^{(2)}, X_2) \quad (78) \\ &= H(c_{31} e^{j\varphi_{31}} X_1^{(2)} + Z_3 | X_2) - H(Z_3) \quad (79) \\ &\le H(c_{31} e^{j\varphi_{31}} X_1^{(2)} + Z_3) - H(Z_3) \quad (80) \end{aligned}$$

since conditioning reduces entropy. Inserting this in (77) we get a weaker upper bound

$$\begin{aligned} C \le \max_{X_1^{(1)}, X_1^{(2)}, X_2} \min \Big\{ & \alpha I(X_1^{(1)}; Y_2^{(1)}, Y_3^{(1)} | X_2 = 0) \\ & + (1-\alpha)(H(c_{31} e^{j\varphi_{31}} X_1^{(2)} + Z_3) - H(Z_3)), \\ & \alpha I(X_1^{(1)}; Y_3^{(1)} | X_2 = 0) \\ & + (1-\alpha) I(X_1^{(2)}, X_2; Y_3^{(2)}) \Big\}. \end{aligned} \quad (81)$$

Next, Lemma 1 below shows that $I(X_1^{(2)}, X_2; Y_3^{(2)})$ (and $H(c_{31} e^{j\varphi_{31}} X_1^{(2)} + Z_3)$) are maximized for $X_1^{(2)}, X_2$ independent Gaussian. This shows that (81) provides an upper bound in the asynchronous case by letting $X_1^{(2)}, X_2$ be independent Gaussian, which is equivalent to putting $\beta = 0$ in the expressions (11) and (12).

Lemma 1: Let X and Y be complex random variables with an arbitrary joint distribution satisfying $E[|X|^2] \le P_x$, $E[|Y|^2] \le P_y$, let Z be zero-mean, circular Gaussian independent of X and Y, and let θ be uniform over $[0, 2\pi]$, and independent of X, Y, and Z. Put $R = X + Ye^{j\theta} + Z$. Then the differential entropy $H(R)$ is maximized for X and Y independent Gaussian.

Proof: Notice that

$$\begin{aligned} E[|R|^2] &= E[|X|^2] + E[|Y|^2] + 2\Re\{E[XY^*]E[e^{j\theta}]\} \\ &\quad + E[|Z|^2] \quad (82) \\ &= E[|X|^2] + E[|Y|^2] + E[|Z|^2] \quad (83) \\ &\le P_x + P_y + E[|Z|^2]. \quad (84) \end{aligned}$$

But among all random variables with power bounded by $P_x + P_y + E[|Z|^2]$, the Gaussian random variable with power $P_x + P_y + E[|Z|^2]$ maximizes entropy [40], and that is achieved when X and Y are independent Gaussian. □

Proof of Proposition 2

There are two coding methods for the relay channel that give the same rate (and the capacity for the degraded relay channel [21]): the original list decoding method in [21] and backward decoding in [22] based on [7]. A simpler argument using parallel (Gaussian) channel arguments, used by several authors [8], [49], [50], gives the same rate as the preceding two methods. It can be shown that also in the TD case, all three coding methods give the same rate; we will provide an argument based on parallel (Gaussian) channels. It can be noticed that it is not necessary to use block-Markov coding [21], as opposed to the full duplex relay channel.

We split the message into two independent parts: A part w_d which is transmitted directly to the destination without the help of the relay at a rate R_d, and a part w_r which is transmitted through the relay to the destination at a rate R_r. The total rate is then $R = R_r + R_d$.

We define three codebooks: $\boldsymbol{X}_1^{(1)}(w_r)$ has αn elements independent and identically distributed (i.i.d.) according to a Gaussian distribution with power $P_1^{(1)}$, $\boldsymbol{X}_2(w_r)$ has $(1-\alpha)n$ elements with power $\beta P_1^{(2)}$, and $\boldsymbol{X}_1^{(2)}(w_d)$ has $(1-\alpha)n$ elements with power $(1-\beta)P_1^{(2)}$.

The transmission and decoding scheme is as follows. During the relay-receive period, the source transmits $\boldsymbol{X}_1^{(1)}(w_r)$. The relay can decode w_r if

$$R_r < \alpha\frac{1}{2}\log\left(1 + c_{21}^2 P_1^{(1)}\right). \tag{85}$$

During the relay-transmit period, the relay transmits

$$\sqrt{\frac{P_2}{\beta P_1^{(2)}}}\boldsymbol{X}_2(w_r).$$

The source transmits $\boldsymbol{X}_2(w_r) + \boldsymbol{X}_1^{(2)}(w_d)$.

The destination starts by decoding w_r from $(y_3^{(1)}, y_3^{(2)}$, treating $\boldsymbol{X}_1^{(2)}(w_d)$ as noise. Using the theory for parallel Gaussian channel [40], it can do so if

$$R_r \le \frac{\alpha}{2}\log\left(1 + c_{31}^2 P_1^{(1)}\right) + \frac{1-\alpha}{2}\log\left(1 + \frac{\left(\sqrt{\beta c_{31}^2 P_1^{(2)}} + \sqrt{c_{32}P_2}\right)^2}{1 + (1-\beta)c_{31}^2 P_1^{(2)}}\right). \tag{86}$$

It then subtracts $\boldsymbol{X}_2(w_r)$ from the received signal

$$\tilde{y}_3^{(2)} = y_3^{(2)} - \left(\sqrt{\beta c_{31}^2 P_1^{(2)}} + \sqrt{c_{32}P_2}\right)\boldsymbol{X}_2(w_r)$$

and decodes $\boldsymbol{X}_1^{(2)}(w_d)$ from $\tilde{y}_3^{(2)}$; it can do so if

$$R_d \le \frac{1-\alpha}{2}\log\left(1 + c_{31}^2(1-\beta)P_1^{(2)}\right). \tag{87}$$

Adding (87) and (86) gives (13); (85) and (86) gives (14).

Proof of Proposition 3

As in the proof of Proposition 2, the message is split into two parts: w_d transmitted directly and w_r transmitted through the relay, with a total rate $R = R_d + R_r$.

The encoding and transmission schemes are as follows. The message w_r is encoded using the codebook $\boldsymbol{X}_1^{(1)}(w_r)$ with αn elements, each following a Gaussian distribution with power $P_1^{(1)}$. The signal $\boldsymbol{X}_1^{(1)}(w_r)$ is transmitted during the relay-receive period. The resulting received signal at the relay $\boldsymbol{Y}_2$ is compressed to the index $s \in \{1, \dots, 2^{nR_0}\}$ (the mechanism is outlined below). During the relay-transmit period, the relay transmits $\boldsymbol{X}_2(s)$, with a Gaussian codebook with $(1-\alpha)n$ elements with power P_2. Simultaneously, the source transmits $\boldsymbol{X}_1^{(2)}(w_d)$ with $(1-\alpha)n$ with power $P_1^{(2)}$.

The destination starts by decoding s and w_d. During the relay-transmit period, the channel is a multiple-access channel (MAC), and it can therefore do so if[2]

$$R_0 \le \frac{1-\alpha}{2}\log\left(1 + \frac{c_{32}^2 P_2}{1 + c_{31}^2 P_1^{(2)}}\right) \tag{88}$$

$$R_d \le \frac{1-\alpha}{2}\log\left(1 + c_{31}^2 P_1^{(2)}\right). \tag{89}$$

It now uses s and $\boldsymbol{Y}_3^{(1)}$ to decode w_r.

The compression at the relay follows [51], [41], [40], particularly the rate distortion theory with side information at the decoder in [51] which extended [40, Theorem 14.9.1] to the continuous-alphabet case. The main idea is to use s and the side information $\boldsymbol{Y}_3^{(1)}$ to construct an estimate $\hat{\boldsymbol{Y}}_2$ at the destination. The proof of the "direct half" in [51] is done by quantizing the ranges of $\hat{Y}_2$ and $Y_3^{(1)}$. Let $\bar{\hat{Y}}_2$ and $\bar{Y}_3^{(1)}$ denote the corresponding quantized versions. For any $\epsilon > 0$, the quantization can be chosen so that

$$I(X; \bar{\hat{Y}}_2, \bar{Y}_3^{(1)}) \le I(X; \hat{Y}_2, Y_3^{(1)}) - \epsilon \tag{90}$$

and so that Lemma 5.3 in [51] is still valid (by refining partitions).

For a given quantization level, the decoder at the destination uses joint typicality to decode w_r from $\bar{\hat{\boldsymbol{Y}}}_2$ and $\bar{\boldsymbol{Y}}_3^{(1)}$. By standard arguments it can do so with small probability of error if

$$R_r/\alpha < I(X; \bar{\hat{Y}}_2, \bar{Y}_3^{(1)}) - 3\epsilon \tag{91}$$

$$\le I(X; \hat{Y}_2, Y_3^{(1)}) - 4\epsilon \tag{92}$$

according to [51, Fig. 4] we have

$$I(X; \hat{Y}_2, Y_3^{(1)}) = I(X; Y_2 + Z_W + kY_3^{(1)}, Y_3^{(1)}) \tag{93}$$

$$= \frac{1}{2}\log\left(1 + c_{31}^2 P_1 + \frac{c_{21}^2 P_1}{1 + \sigma_w^2}\right) \tag{94}$$

where k is a constant (that can be found from [51] but does not influence the final result), and Z_W is a Gaussian variable independent of Y_2 and $Y_3^{(1)}$ with variance σ_w^2. This variance is determined by the fact that the rate of the compressed signal $\hat{\boldsymbol{Y}}_2$ is constrained to the rate R_0 the relay has available, so according to [51, Theorem 2.2] and Section III

$$R_0/\alpha \ge I(Y_2; W) - I(Y_3; W) \tag{95}$$

$$= H(W|Y_3) - H(W|Y_2) \tag{96}$$

$$= \frac{1}{2}\log\left(c_{21}^2 P_1 + 1 + \sigma_w^2 - \frac{(c_{31}c_{21}P_1)^2}{c_{31}^2 P_1 + 1}\right) - \frac{1}{2}\log(\sigma_w^2) \tag{97}$$

$$\sigma_w^2 \ge \frac{c_{31}^2 P_1 + c_{21}^2 P_1 + 1}{\left(2^{2R_0/\alpha} - 1\right)\left(c_{31}^2 P_1 + 1\right)}. \tag{98}$$

Inserting (88) in (98) gives (16) and (89), (92), and (94) with $R = R_r + R_d$ gives (15).

[2] R_0 and R_d can be chosen anywhere on the "sum rate side" in the MAC pentagon (between C and B in [40, Fig. 14.14]) with the same resulting rate for the relay channel.

APPENDIX B
BOUNDS ON ERGODIC CAPACITY: THE FULL DUPLEX RELAYING CASE

For later use in calculating limits for SNR $\to\infty$, we will re-introduce the explicit noise power σ^2. Let $\{\boldsymbol{c}[i]\}_{i=1}^{\infty}$ be a specific realization of the fading process. The max-flow-min-cut bound gives

$$nR \le \sum_{i=1}^{n} I(X_1[i]; Y_2[i], Y_3[i] | X_2[i], \boldsymbol{c}[i]) + n\epsilon_n \tag{99}$$

$$\le \underbrace{\sum_{i=1}^{n}\frac{1}{2}\log\left(1+\frac{(c_{21}^2[i]+c_{31}^2[i])P_1^{(1)}[i]}{\sigma^2}\right)}_{R_1^n} + n\epsilon_n \tag{100}$$

and

$$nR \le \sum_{i=1}^{n} I(X_1[i], X_2[i]; Y_3[i] | \boldsymbol{c}[i]) + n\epsilon_n \tag{101}$$

$$\le \underbrace{\sum_{i=1}^{n}\frac{1}{2}\log\left(1+\frac{c_{31}^2[i]P_1^{(1)}[i]+(c_{31}^2[i]+c_{32}^2[i])P_j[i]}{\sigma^2}\right)}_{R_2^n} + n\epsilon_n. \tag{102}$$

The power constraint is $\frac{1}{n}\sum_{i=1}^{n}(P_1^{(1)}[i]+P_j[i]) \le nP$, together with $P_1^{(1)}[i]\ge 0$, $P_j[i]\ge 0$. Notice that we can always find an optimum solution with $R_1^n = R_2^n$. If we have a solution $(P_1^{(1)}[i], P_j[i])$ with $R_1^n > R_2^n$, put

$$(\tilde{P}_1^{(1)}[i], \tilde{P}_j[i]) = ((1-\varepsilon)P_1^{(1)}[i], P_j[i]+\varepsilon P_1^{(1)}[i])$$

and let the corresponding rates be $\tilde{R}_1^n$, $\tilde{R}_2^n$. Then $\tilde{R}_1^n$ is a decreasing function of ε, $\tilde{R}_1^n = 0$ for $\varepsilon = 1$, and $\tilde{R}_2^n$ an increasing function. By continuity we then have $\tilde{R}_1^n = \tilde{R}_2^n \ge R_2^n$ for some ε. Similarly for $R_1^n < R_2^n$.

Thus, we can state the optimization problem as the Lagrange problem of maximizing[3]

$$J = R_1^n + t(R_2^n - R_1^n) + \frac{\lambda}{2\ln 2}\sum_{i=1}^{n}(P_1^{(1)}[i]+P_j[i]) \tag{103}$$

$$= (1-t)R_1^n + tR_2^n + \frac{\lambda}{2\ln 2}\sum_{i=1}^{n}(P_1^{(1)}[i]+P_j[i]). \tag{104}$$

A few comments on this Lagrange optimization problem. Fix $\lambda < 0$, and suppose that for some t we have found an unconstrained maximum to J satisfying $R_1^n = R_2^n$. Then this is also a solution to the original constrained problem for *some* power constraint. In addition, the problem splits up in individual coordinate problems

$$J = \sum_{i=1}^{n} J[i] \tag{105}$$

$$J[i] = \frac{1-t}{2}\log\left(1+\frac{(c_{21}^2[i]+c_{31}^2[i])P_1^{(1)}[i]}{\sigma^2}\right) + \frac{t}{2}\log\left(1+\frac{c_{31}^2[i]P_1^{(1)}[i]+(c_{31}^2[i]+c_{32}^2[i])P_j[i]}{\sigma^2}\right) + \frac{\lambda}{2\ln 2}(P_1^{(1)}[i]+P_j[i]). \tag{106}$$

[3]The explicit Lagrange multipliers corresponding to the constraints $P_1^{(1)}[i] \ge 0$, $P_j[i] \ge 0$ have been omitted.

Thus, the problem reduces to the problem maximizing $J[i]$ for given t, λ.

Differentiating with respect to $P_1^{(1)}[i]$ and $P_j[i]$ and equating to zero, we get the equations

$$\frac{tc_{31}^2[i]}{\sigma^2+c_{31}^2[i]P_1^{(1)}[i]+(c_{31}^2[i]+c_{32}^2[i])P_j[i]} + \frac{(1-t)\left(c_{21}^2[i]+c_{31}^2[i]\right)}{\sigma^2+P_1^{(1)}[i]\left(c_{21}^2[i]+c_{31}^2[i]\right)} + \lambda = 0$$

$$\frac{t\left(c_{31}^2[i]+c_{32}^2[i]\right)}{\sigma^2+c_{31}^2[i]P_1^{(1)}[i]+(c_{31}^2[i]+c_{32}^2[i])P_j[i]} + \lambda = 0.$$

From the second equation we get

$$P_j[i] = -\frac{t}{\lambda} - \frac{\sigma^2+c_{31}^2[i]P_1^{(1)}[i]}{c_{31}^2[i]+c_{32}^2[i]}. \tag{107}$$

There are four possibilities, depending on whether $P_1^{(1)}[i]$ and $P_j[i]$ are zero, or strictly positive. First, consider the case $P_j[i] > 0$, $P_1^{(1)}[i] > 0$. Then

$$P_1^{(1)}[i] = -\frac{(1-t)(c_{31}^2[i]+c_{32}^2[i])}{c_{32}^2[i]\lambda} - \frac{\sigma^2}{c_{21}^2[i]+c_{31}^2[i]} > 0 \tag{108}$$

$$P_j[i] = -\frac{t}{\lambda} - \frac{\sigma^2+c_{31}^2[i]P_1^{(1)}[i]}{c_{31}^2[i]+c_{32}^2[i]} > 0. \tag{109}$$

Second, consider $P_j[i] > 0$, $P_1^{(1)}[i] = 0$. Then

$$P_j = -\frac{t}{\lambda} - \frac{\sigma^2}{c_{31}^2[i]+c_{32}^2[i]} > 0 \tag{110}$$

with the condition

$$-\frac{(1-t)(c_{31}^2[i]+c_{32}^2[i])}{c_{32}^2[i]\lambda} - \frac{\sigma^2}{c_{21}^2[i]+c_{31}^2[i]} \le 0. \tag{111}$$

Consider now the cases with $P_j[i] = 0$. $P_1^{(1)}[i]$ is then the solution to

$$\frac{tc_{31}^2[i]}{\sigma^2+c_{31}^2[i]P_1^{(1)}[i]} + \frac{(1-t)\left(c_{21}^2[i]+c_{31}^2[i]\right)}{\sigma^2+P_1^{(1)}[i]\left(c_{21}^2[i]+c_{31}^2[i]\right)} + \lambda = 0. \tag{112}$$

Or (with $\sigma^2 = 1$)

$$\lambda c_{31}^2[i](c_{21}^2[i]+c_{31}^2[i])P_1^{(1)}[i]^2 + (c_{31}^4[i]+c_{21}^2[i]c_{31}^2[i]+\lambda c_{21}^2[i] + 2\lambda c_{31}^2[i])P_1^{(1)}[i] + \lambda + c_{31}^2[i] + (1-t)c_{21}^2[i] = 0. \tag{113}$$

By looking at the coefficients of this second-order equation, it is seen that if

$$\frac{1}{c_{31}^2[i]+(1-t)c_{21}^2[i]} \le -\frac{1}{\lambda} \tag{114}$$

it has one negative and one positive root; otherwise, it has two negative roots. If (114) is satisfied, and (110) is *not* a solution, there are two solutions to the Lagrange problem: 1) $P_1^{(1)}[i] = 0$ and $P_j[i] = 0$ or $P_1^{(1)}[i]$ equal to the positive solution to (113) and $P_j[i] = 0$. The correct solution can be found by inserting each of the solutions in the expression (106) for $J[i]$ and choosing the one that maximizes $J[i]$.

Then, by letting $n \to \infty$ and using the ergodicity of the fading process, we arrive at the solution in Table I. For simplicity of notation we have used $\mu = -\frac{1}{\lambda}$ for the Lagrange parameter; μ can be interpreted as a water-filling level similar to ν in [40, Sec. 10.4].

For the achievable rate we consider decode-forward. If $c_{31} > c_{21}$, the source bypasses the relay. If $c_{31} < c_{21}$, the source transmits to the relay, which decodes the transmission, puts it into a queue, and transmits the output of the queue to the destination. Call the direct rate R_d, the rate through the relay R_r with total rate $R = R_d + R_r$. We then get the following rate:

$$R_d = P(c_{31} > c_{21}) E_{c_{31}>c_{21}} \left[\frac{1}{2}\log(1 + c_{31}^2 P_1(\boldsymbol{c}))\right] \quad (115)$$

$$\begin{aligned} R_r = \min & \Big\{ P(c_{31} < c_{21}) E_{c_{31}<c_{21}} \left[\frac{1}{2}\log(1 + c_{21}^2 P_1(\boldsymbol{c}))\right], \\ & P(c_{31} < c_{21}) E_{c_{31}<c_{21}} \left[\frac{1}{2}\log(1 + c_{31}^2 P_1(\boldsymbol{c}))\right] \\ & + E_{c_{31},c_{32}} \left[\frac{1}{2}\log\left(1 + \frac{(c_{31}^2 + c_{32}^2) P_j(\boldsymbol{c})}{1 + c_{31}^2 P_1(\boldsymbol{c})}\right)\right]\Big\} \end{aligned} \quad (116)$$

$$\begin{aligned} R = \min & \Big\{ E_{c_{31},c_{21}} \left[\frac{1}{2}\log(1 + \max\{c_{21}^2, c_{31}^2\} P_1(\boldsymbol{c}))\right], \\ & E_{c_{31},c_{32}}\left[\frac{1}{2}\log(1 + c_{31}^2 P_1(\boldsymbol{c}) + (c_{31}^2 + c_{32}^2) P_j(\boldsymbol{c}))\right]\Big\}. \end{aligned} \quad (117)$$

For achievable rate, any power allocation rule $(P_j(\boldsymbol{c}), P_1^{(1)}(\boldsymbol{c}))$ will give a valid rate as seen by the arguments in [44]. The results in Table I is obtained by replacing $c_{21}^2 + c_{31}^2$ with $\max\{c_{21}^2, c_{31}^2\}$ everywhere in the Lagrange optimization problem.

Finally, the asynchronous solution follows along the same lines by replacing $c_{31}^2 + c_{32}^2$ with c_{32}^2 everywhere.

Appendix C
Bounds on Asymptotic Rate Gains: The Full Duplex Relaying Case

Proof of Proposition 6

First we will prove the power allocation rule (51). Notice that for σ^2 suitably small, the condition (111) is not satisfied, and $P_j[i] > 0$, $P_1^{(1)}[i] = 0$ is not a solution. On the other hand, (108) is satisfied, and inserting (108) in (109) we get

$$\begin{aligned} P_j[i] = & -\frac{t}{\lambda} + \frac{(1-t) c_{31}^2[i]}{c_{32}^2[i]\lambda} \\ & - \left(1 - \frac{1}{c_{21}^2[i] + c_{31}^2[i]}\right) \frac{\sigma^2}{c_{31}^2[i] + c_{32}^2[i]}. \end{aligned} \quad (118)$$

We see that $P_j[i] > 0$ for arbitrary small σ^2 if and only if

$$\frac{c_{31}^2[i]}{c_{32}^2[i]} \leq \frac{t}{1-t} \triangleq \kappa. \quad (119)$$

For $\sigma^2 \to 0$, the largest solution of (112) converges toward $P_1^{(1)} = -\lambda^{-1}$, i.e., $P_1^{(1)}(\sigma^2) = -\lambda^{-1} + \epsilon(\sigma^2)$. The condition that this is a solution to the Lagrange problem is that $P_j[i]$ given by (107) is negative

$$-\frac{t}{\lambda} - -\frac{\sigma^2 + \epsilon(\sigma^2)}{c_{31}^2[i] + c_{32}^2[i]} + \frac{c_{31}^2[i]}{c_{31}^2[i] + c_{32}^2[i]}\frac{1}{\lambda} < 0. \quad (120)$$

For $\sigma^2 \to 0$ this is the opposite condition of (119).

Thus, we have proven that for a specific value of $\boldsymbol{c}$ and fixed λ, the Lagrange solution converges toward (again with $\mu = -\lambda^{-1}$)

$$\begin{aligned} & (P_1^\infty(\boldsymbol{c}), P_j^\infty(\boldsymbol{c})) \\ & \triangleq \lim_{\sigma^2 \to 0} (P_1^{(1)}(\boldsymbol{c}), P_j(\boldsymbol{c})) \\ & = \begin{cases} \left(\frac{c_{31}^2 + c_{32}^2}{c_{32}^2}(1-t)\mu, \left(t - \frac{c_{31}^2}{c_{32}^2}(1-t)\right)\mu\right), & \frac{c_{31}^2}{c_{32}^2} < \kappa \\ (\mu, 0), & \text{otherwise.} \end{cases} \end{aligned} \quad (121)$$

Furthermore, the solution (121) gives a constant total transmit power μ, and it follows (i.e., Lebegues dominated convergence) that $\lim_{\sigma^2 \to 0} E[P_1^{(1)} + P_j] = \mu$, and the solution (51) follows. The proof for the achievable rate is similar.

Let $f(c_{ij})$ denote the probability measure for c_{ij}^2 and $f(\boldsymbol{c})$ that for (c_{21}, c_{31}, c_{32}). If c_{ij}^2 has a probability density function $f'(c_{ij}^2)$, then $df(c_{ij}) = f'(c_{ij}^2) dc_{ij}^2$ in the following.

From continuity and the Lebesgue dominated convergence

$$\begin{aligned} & R_1^\infty(t) \\ & \triangleq \lim_{\sigma \to 0} \frac{1}{2} \int \log\left(1 + \frac{c(c_{21}^2, c_{31}^2)}{\sigma^2} P_1^{(1)}(\boldsymbol{c}, \sigma^2)\right) df(\boldsymbol{c}) \\ & \quad + \frac{1}{2}\log(\sigma^2) \\ & = \frac{1}{2}\int \log\left(c\left(c_{21}^2, c_{31}^2\right) P_1^\infty(\boldsymbol{c})\right) df(\boldsymbol{c}) \end{aligned} \quad (122)$$

$$\begin{aligned} & R_2^\infty(t) \\ & \triangleq \lim_{\sigma \to 0} \frac{1}{2} \int \\ & \quad \log\left(1 + \frac{c_{31}^2 P_1^{(1)}(\boldsymbol{c}, \sigma^2) + (c_{31}^2 + c_{32}^2) P_j(\boldsymbol{c}, \sigma^2)}{\sigma^2}\right) df(\boldsymbol{c}) \\ & \quad + \frac{1}{2}\log(\sigma^2) \\ & = \frac{1}{2}\int \log\left(c_{31}^2 P_1^\infty(\boldsymbol{c}) + (c_{31}^2 + c_{32}^2) P_j^\infty(\boldsymbol{c})\right) df(\boldsymbol{c}). \end{aligned} \quad (123)$$

We have to find t so that $R_1^\infty(t) = R_2^\infty(t)$. Define

$$A = \left\{(c_{31}, c_{32}) \,\middle|\, \frac{c_{31}^2}{c_{32}^2} < \frac{t}{1-t}\right\}.$$

Then

$$\begin{aligned} R_1^\infty(t) = & \int_A \int_R \log\left(c(c_{21}, c_{31}) \frac{c_{31}^2 + c_{32}^2}{c_{32}^2}(1-t)P\right) df(\boldsymbol{c}) \\ & + \int_{R^2 - A}\int_R \log\left(c(c_{21}, c_{31})P\right) df(\boldsymbol{c}) \end{aligned} \quad (124)$$

$$\begin{aligned} R_2^\infty(t) = & \int_A \log\left(\left(c_{31}^2 + c_{32}^2\right) tP\right) df(c_{31}) df(c_{32}) \\ & + \int_{R^2 - A} \log\left(c_{31}^2 P\right) df(c_{31}) df(c_{32}). \end{aligned} \quad (125)$$

The equation $R_1^\infty(t) = R_2^\infty(t)$ now reduces to

$$\begin{aligned}\log(t)\int_A df(c_{31})df(c_{32}) &+ \int_{R^2-A}\log\left(c_{31}^2\right)df(c_{31})df(c_{32}) \\ &= \int_{R^2}\log\left(c(c_{21},c_{31})\right)df(c_{31})df(c_{21}) \\ &\quad - \int_A \log\left(c_{32}^2\right)df(c_{31})df(c_{32}) \\ &\quad + \log(1-t)\int_A df(c_{31})df(c_{32}). \end{aligned} \tag{126}$$

Or

$$\begin{aligned}\log(\kappa)\int_A df(c_{31})df(c_{32}) &+ \int_{R^2-A}\log\left(c_{31}^2\right)df(c_{31})df(c_{32}) \\ &+ \int_A \log\left(c_{32}^2\right)df(c_{31})df(c_{32}) \\ &= \int_{R^2}\log\left(c(c_{21},c_{31})\right)df(c_{31})df(c_{21}). \end{aligned} \tag{127}$$

For Rayleigh fading, where c_{ij}^2 is exponentially distributed, all the integrals involved can be evaluated in closed form. For example

$$\begin{aligned}\int_A df(c_{31})df(c_{32}) &= \int_0^\infty \int_0^{\kappa c_{32}^2} s_{31}^{-1}\exp\left(-\frac{c_{31}^2}{s_{31}}\right)dc_{31}^2 s_{32}^{-1}\exp\left(-\frac{c_{32}^2}{s_{32}}\right)dc_{32}^2 \\ &= \int_0^\infty \left(1-\exp\left(-\frac{c_{32}^2\kappa}{s_{31}}\right)\right)s_{32}^{-1}\exp\left(-\frac{c_{32}^2}{s_{32}}\right)dc_{32}^2 \\ &= \frac{s_{32}\kappa}{s_{32}\kappa + s_{31}}. \end{aligned} \tag{128}$$

The other integrals needed are evaluated below. The evaluations and following simplifications have been done by a symbolic analysis program (Maple) aided by calculations by hand, and are rather straightforward but tedious

$$\begin{aligned}\ln 2\int_{R^2-A}\log\left(c_{31}^2\right)df(c_{31})df(c_{32}) &= \frac{s_{31}(\ln(s_{31})-\gamma)}{s_{32}\kappa+s_{31}} \\ &\quad + \frac{s_{32}\kappa(\ln(s_{32}\kappa+s_{31})-\ln(\kappa)-\ln(s_{32}))}{s_{32}\kappa+s_{31}} \end{aligned} \tag{129}$$

$$\begin{aligned}\ln 2\int_A \log\left(c_{32}^2\right)df(c_{31})df(c_{32}) &= \frac{\kappa s_{32}(\ln(s_{32})-\gamma)}{s_{32}\kappa+s_{31}} \\ &\quad + \frac{s_{31}(\ln(s_{32}\kappa+s_{31})-\ln(s_{31}))}{s_{32}\kappa+s_{31}} \end{aligned} \tag{130}$$

$$\begin{aligned}\ln 2\int_{R^2}\log\left(\max(c_{21}^2,c_{31}^2)\right)df(c_{31})df(c_{21}) &= \ln(s_{21}+s_{31})-\gamma \end{aligned} \tag{131}$$

$$\begin{aligned}\ln 2\int_{R^2}\log\left((c_{21}^2+c_{31}^2)\right)df(c_{31})df(c_{21}) &= \begin{cases}\frac{s_{21}\ln s_{21}-s_{31}\ln s_{31}}{s_{21}-s_{31}}-\gamma, & s_{21}\neq s_{31} \\ \ln s_{21}+1, & s_{21}=s_{31}\end{cases} \end{aligned} \tag{132}$$

where γ is Euler's constant. Inserting this in (127) we then get (54).

We finally calculate the asymptotic relay gain. First notice that in the high-SNR regime, power control for *direct transmission* is not needed (i.e., it does not give any gain). This follows directly from the results in [44]. We can therefore define the following rate for direct transmission:

$$\begin{aligned}R_d^\infty &\triangleq \lim_{\sigma\to 0}\frac{1}{2}\int_R \log\left(1+\frac{c_{31}^2P}{\sigma^2}\right)df(c_{31}) + \frac{1}{2}\log(\sigma^2) \\ &= \frac{1}{2}\int_R \log\left(c_{31}^2P\right)df(c_{31}). \end{aligned} \tag{133}$$

The rate gain relative to direct transmission for $\sigma\to 0$ can then be found as $R_2^\infty(t) - R_d^\infty$ (the term $\frac{1}{2}\log(\sigma^2)$ appears in both definitions, and therefore cancels out). We get

$$\begin{aligned}R_g &= R_2^\infty(t) - R_d^\infty \\ &= \int_A \frac{1}{2}\log\left(\left(c_{31}^2+c_{32}^2\right)tP\right)df(c_{31})df(c_{32}) \\ &\quad - \int_A \frac{1}{2}\log\left(c_{31}^2P\right)df(c_{31})df(c_{32}) \\ &= \int_A \frac{1}{2}\log\left(\left(1+\frac{c_{32}^2}{c_{31}^2}\right)\right)df(c_{31})df(c_{32}) \\ &\quad + \frac{1}{2}\log(t)\int_A df(c_{31})df(c_{32}) \end{aligned} \tag{134}$$

$$= \frac{1}{2}\frac{s_{32}}{s_{32}-s_{31}}\log\left(\frac{1+k}{1+k\frac{s_{32}}{s_{31}}}\right). \tag{135}$$

Proof of Proposition 7

For the asynchronous case, we can go through the same set of calculations, replacing P_j with P_2 and $c_{31}^2 + c_{32}^2$ with c_{32}^2. Notice that for $c_{32}^2 < c_{31}^2$ it does not pay to let the relay transmit: the destination knows at least as much about the message as the relay, and it has a better connection. By going through the Lagrange solution, taking limits, and excluding solutions with $c_{32}^2 < c_{31}^2$ and $P_2 > 0$, we get the asymptotic power allocation law

$$\begin{aligned}(P_1^\infty(\boldsymbol{c}), P_2^\infty(\boldsymbol{c})) &\triangleq \lim_{\sigma^2\to 0}(P_1^{(1)}(\boldsymbol{c}), P_2(\boldsymbol{c})) \\ &= \begin{cases}\left(\frac{c_{32}^2}{c_{32}^2-c_{31}^2}(1-t)\mu,\right. \\ \left.\left(t-\frac{c_{31}^2}{c_{32}^2-c_{31}^2}(1-t)\right)\mu\right), & \frac{c_{31}^2}{c_{32}^2}<t \\ (\mu,0), & \text{otherwise.}\end{cases} \end{aligned} \tag{136}$$

As for the synchronized case, the transmit power is constant, and we therefore arrive at the solution (57). We have to find t to solve $R_1^\infty(t) = R_2^\infty(t)$. Denote by $A = \{(c_{31},c_{32}) | \frac{c_{31}^2}{c_{32}^2} < t\}$. Then

$$R_1^\infty(t) = \int_A\int_R \log\left(c(c_{21},c_{31})\frac{c_{32}^2}{c_{32}^2-c_{31}^2}(1-t)P\right)df(\boldsymbol{c})$$

$$+\int_{R^2-A}\int_R \log\left(c(c_{21},c_{31})P\right)df(\boldsymbol{c}) \tag{137}$$

$$R_2^\infty(t)=\int_A \log\left(c_{32}^2tP\right)df(c_{31})df(c_{32}) + \int_{R^2-A}\log\left(c_{31}^2P\right)df(c_{31})df(c_{32}). \tag{138}$$

The equation $R_1^\infty(t)=R_2^\infty(t)$ reduces to

$$\begin{aligned}\log(t)\int_A df(c_{31})df(c_{32}) &+ \int_{R^2-A}\log\left(c_{31}^2\right)df(c_{31})df(c_{32})\\ &= \int_{R^2}\log\left(c(c_{21},c_{31})\right)df(c_{31})df(c_{21})\\ &-\int_A \log\left(c_{32}^2-c_{31}^2\right)df(c_{31})df(c_{32})\\ &+\log(1-t)\int_A df(c_{31})df(c_{32}).\end{aligned} \tag{139}$$

Or

$$\begin{aligned}\log\left(\frac{t}{1-t}\right)\int_A df(c_{31})df(c_{32}) &+\int_{R^2-A}\log\left(c_{31}^2\right)df(c_{31})df(c_{32})\\ &+\int_A\log\left(c_{32}^2-c_{31}^2\right)df(c_{31})df(c_{32})\\ &=\int_{R^2}\log\left(c(c_{21},c_{31})\right)df(c_{31})df(c_{21}).\end{aligned} \tag{140}$$

To solve this we need the following integral in addition to (128)–(132):

$$\ln 2\int_A \log\left(c_{32}^2-c_{31}^2\right)df(c_{31})df(c_{32}) = \int_0^\infty \ln(y)\int_0^{\frac{t}{1-t}y} f'_{c_{32}}(c_{31}^2+y)f'_{c_{31}}(c_{31}^2)dc_{31}^2dy. \tag{141}$$

The integral can be evaluated in closed form, but gives a complicated expression that we will not write down here. However, inserting this in (140) and reducing, we get the following left-hand side of (140):

$$\frac{s_{32}\ln\left(\frac{ts_{32}+s_{31}}{1-t}\right)+s_{31}\ln(s_{31})}{s_{31}+s_{32}}-\gamma \tag{142}$$

while the right-hand side is unchanged from the synchronized case. The rate gain relative to direct transmission for $\sigma\to 0$ can now be evaluated to

$$R_g=\int_A\frac{1}{2}\log\left(c_{32}^2tP\right)df(c_{31})df(c_{32}) - \int_A\frac{1}{2}\log\left(c_{31}^2P\right)df(c_{31})df(c_{32}) \tag{143}$$

$$=\frac{1}{2}\log\left(t\frac{s_{32}}{s_{31}}+1\right). \tag{144}$$

Proof of Corollary 1

We will briefly outline the proof of the corollary. For convenience, we normalize all s_{ij} by s_{31}, $\tilde{s}_{ij}=s_{ij}/s_{31}$. It is first proven that for each fixed value of $\tilde{s}_{21}$, $R_g^+-R_g^-$ is an increasing function of $\tilde{s}_{32}$ for both the synchronous and asynchronous case; therefore,

$$R_g^+-R_g^-\le\lim_{s_{32}\to\infty}R_g^+-R_g^-$$

(provided the limit exists). In both cases

$$\lim_{\tilde{s}_{32}\to\infty}R_g^+-R_g^-=\frac{1}{2}\log\left(\frac{2^{l(\tilde{s}_{21},1)}}{\tilde{s}_{21}+1}\right). \tag{145}$$

It is now easy to prove that the expression inside the log has its maximum at $\tilde{s}_{21}=1$. The expression (145) actually gives an upper bound on the difference $R_g^+-R_g^-$ for any fixed value of $\tilde{s}_{21}=s_{21}/s_{31}$, that can be used to show how close the bounds are away from the worst case.

APPENDIX D
BOUNDS ON ERGODIC CAPACITY: THE TIME-DIVISION RELAYING CASE

Let $\{\boldsymbol{c}[i]\}_{i=1}^\infty$ be a specific realization of the fading process. The max-flow-min-cut bound gives

$$nR\le\sum_{i=1}^n I(X_1[i];Y_2[i],Y_3[i]|X_2[i],\boldsymbol{c}[i])+n\epsilon_n \tag{146}$$

$$\le\underbrace{\sum_{i=1}^n f_1[i]\left(P_1^{(1)}[i],P_j[i]\right)}_{R_1^n}+n\epsilon_n \tag{147}$$

$$nR\le\sum_{i=1}^n I(X_1[i],X_2[i];Y_3[i]|\boldsymbol{c}[i])+n\epsilon_n \tag{148}$$

$$\le\underbrace{\sum_{i=1}^n\frac{1}{2}\log\left(1+\frac{c_{31}^2[i]P_1^{(1)}[i]+(c_{31}^2[i]+c_{32}^2[i])P_j[i]}{\sigma^2}\right)}_{R_2^n} + n\epsilon_n \tag{149}$$

$$f_1[i](P_1,P_j)=\begin{cases}\frac{1}{2}\log\left(1+\frac{(c_{21}^2[i]+c_{31}^2[i])P_1}{\sigma^2}\right), & P_j=0\\ \frac{1}{2}\log\left(1+\frac{c_{31}^2[i]P_1}{\sigma^2}\right), & P_j>0.\end{cases} \tag{150}$$

The power constraint is $\frac{1}{n}\sum_{i=1}^n(P_1^{(1)}[i]+P_j[i])\le nP$, together with $P_1^{(1)}[i]\ge 0$, $P_j[i]\ge 0$. We will argue that we can restrict our attention to solutions with $R_1^n=R_2^n$. First, if we have a solution with $R_1^n<R_2^n$, we can find a solution which is at least as good with $R_1^n=R_2^n$ by decreasing $P_j[i]$. Therefore, consider a sequence of solutions $\{R_1^n,R_2^n\}_{n=1}^\infty$ with $R_1^n\ge R_2^n$ and $\lim_{n\to\infty}\frac{1}{n}R_2^n=R$. R_1^n is a sum of terms where the relay receives ($P_j[i]=0$) and terms where the relay transmits ($P_j[i]>0$). By "switching off" the relay reception for some of the terms where the relay receives, we can decrease R_1^n to $\tilde{R}_1^n\le R_2^n$. Furthermore, for any $\varepsilon>0$ and for n large enough, we can do this so that $\frac{1}{n}(R_2^n-\tilde{R}_1^n)<\varepsilon$. By decreasing R_2^n as above, we then have a solution $\tilde{R}_1^n=\tilde{R}_2^n$ and $\frac{1}{n}(R_2^n-\tilde{R}_1^n)<\varepsilon$. We can, therefore, find a sequence of solutions $\{\tilde{R}_1^n,\tilde{R}_2^n\}_{n=1}^\infty$ with $\tilde{R}_1^n=\tilde{R}_2^n$ and $\lim_{n\to\infty}\frac{1}{n}\tilde{R}_2^n=R$.

As for full duplex, the problem reduces to maximization of $J[i]$, where

$$J = \sum_{i=1}^{n} J[i] \tag{151}$$

$$J[i] = (1-t) f_1[i]\left(P_1^{(1)}[i], P_j[i]\right) + \frac{t}{2}\log\left(1 + \frac{c_{31}^2[i]P_1^{(1)}[i] + (c_{31}^2[i] + c_{32}^2[i])P_j[i]}{\sigma^2}\right) + \frac{\lambda}{2\ln 2}(P_1^{(1)}[i] + P_j[i]). \tag{152}$$

The difference is that $J[i]$ is no longer continuous in the variable P_j, see (150). We therefore have to find both the maximum on the boundary $P_j[i]=0$ (this corresponds to the solution $P_{1,2}^{(1)}$, $P_{j,2}$ in Table II) and in the interior $P_j[i]>0$ (this corresponds to the solution $P_{1,1}^{(1)}$, $P_{j,1}$ in Table II). These two solutions should then be compared to see which one maximizes $J[i]$, which is done in Table II by comparing $R_{t,1}$ and $R_{t,2}$. Detailed calculations to find the solutions are along the same lines as for full duplex.

For the asymptotic solution when $\sigma \to 0$, we can argue as for full duplex that only two of the solutions are relevant for σ small enough, and that the solution with $P_j \neq 0$ (*joint transmission*) is selected if $\frac{c_{31}^2}{c_{32}^2} < \frac{t}{1-t}$, with the additional condition that the joint transmission solution is only used if it maximizes $J[i]$, which in the asymptotic limit reduces to the condition

$$(1-t)\log\left(\frac{c_{31}^2(c_{31}^2+c_{32}^2)}{c_{32}^2}(1-t)P\right) + t\log\left((c_{31}^2+c_{32}^2)tP\right) > (1-t)\log\left((c_{21}^2+c_{31}^2)P\right) + t\log\left(c_{31}^2 P\right) \tag{153}$$

or

$$(1-t)\log\left(\frac{c_{31}^2(c_{31}^2+c_{32}^2)}{c_{32}^2(c_{21}^2+c_{31}^2)}(1-t)\right) > t\log\left(\frac{c_{31}^2}{c_{31}^2+c_{32}^2}\frac{1}{t}\right). \tag{154}$$

We then get (68)–(69).

Define A as the region of (c_{31}, c_{21}, c_{32}) where joint transmission is used. We can then write the rates as

$$R_1^\infty(t) = \int_A \log\left(c_{31}^2 \frac{c_{31}^2 + c_{32}^2}{c_{32}^2}(1-t)P\right) df(\boldsymbol{c}) + \int_{R^2-A} \log\left(c(c_{21}, c_{31})P\right) df(\boldsymbol{c}) \tag{155}$$

$$R_2^\infty(t) = \int_A \log\left(\left(c_{31}^2 + c_{32}^2\right) tP\right) df(c_{31}) df(c_{32}) + \int_{R^2-A} \log\left(c_{31}^2 P\right) df(c_{31}) df(c_{32}). \tag{156}$$

The equation $R_1^\infty(t) = R_2^\infty(t)$ reduces to

$$\log\left(\frac{t}{1-t}\right)\int_A df(c_{31})df(c_{32}) + \int_{R^2-A}\log\left(c_{31}^2\right) df(c_{31})df(c_{32}) \tag{157}$$

$$= \int_{R^2-A}\log\left(c(c_{21},c_{31})\right) df(c_{31})df(c_{21}) + \int_A \log\left(\frac{c_{31}^2}{c_{32}^2}\right) df(c_{31})df(c_{32}). \tag{158}$$

It appears that this equation (and the integrals) must be solved numerically to find t because of the complicated shape of A. The rate gain can be found (numerically) from (134) with the new definition of A.

For the asynchronous case, we use joint transmission if

$$(1-t)\log\left(\frac{c_{31}^2 c_{32}^2}{c_{32}^2 - c_{31}^2}(1-t)P\right) + t\log\left(c_{32}^2 tP\right) > (1-t)\log\left(c(c_{21}, c_{31})P\right) + t\log\left(c_{31}^2 P\right) \tag{159}$$

or

$$(1-t)\log\left(c_{31}^2 \frac{c_{32}^2}{(c_{32}^2 - c_{31}^2)c(c_{21}, c_{31})}(1-t)\right) + t\log\left(\frac{c_{32}^2}{c_{31}^2}t\right) > 0. \tag{160}$$

The rates are now

$$R_1^\infty(t) = \int_A \int_R \log\left(c_{31}^2 \frac{c_{32}^2}{c_{32}^2 - c_{31}^2}(1-t)P\right) df(\boldsymbol{c}) + \int_{R^2-A}\int_R \log\left(c(c_{21}, c_{31})P\right) df(\boldsymbol{c}) \tag{161}$$

$$R_2^\infty(t) = \int_A \log\left(c_{32}^2 tP\right) df(c_{31})df(c_{32}) + \int_{R^2-A} \log\left(c_{31}^2 P\right) df(c_{31})df(c_{32}). \tag{162}$$

The equation $R_1^\infty(t) = R_2^\infty(t)$ reduces to

$$\log\left(\frac{t}{1-t}\right)\int_A df(c_{31})df(c_{32}) + \int_{R^2-A}\log\left(c_{31}^2\right) df(c_{31})df(c_{32}) = \int_{R^2-A}\log\left(c(c_{21},c_{31})\right) df(c_{31})df(c_{21}) + \int_A \log\left(\frac{c_{31}^2}{c_{32}^2 - c_{31}^2}\right) df(c_{31})df(c_{32}) \tag{163}$$

and the rate gain can be found from (143).

APPENDIX E
PROOF OF PROPOSITION 8

We use the asymptotic power allocation rule from the asynchronous case, i.e., $P_1(\boldsymbol{c})$ and $P_2(\boldsymbol{c})$ are functions of $\boldsymbol{c}$ given by (57). For now we let the value of t be open (i.e., some fixed unknown value); the value of t is found by optimizing over t the solution found in the following. Consider a fixed $\boldsymbol{c}$. The received signal at nodes 2 and 3 is then Gaussian, and we can use the same coding method as in Proposition 3, modified to full duplex. Thus, the received signal at node 2 is Wyner–Ziv compressed to a rate $R_s(\boldsymbol{c})$, and the power of the resulting "compression noise" given by (98), which for full duplex and with σ^2 explicitly reintroduced is

$$\sigma_w(\boldsymbol{c}) = \sigma^2 \frac{c_{31}^2 P_1(\boldsymbol{c}) + c_{21}^2 P_1(\boldsymbol{c}) + \sigma^2}{\left(2^{R_s(\boldsymbol{c})} - 1\right)\left(c_{31}^2 P_1(\boldsymbol{c}) + \sigma^2\right)}. \tag{164}$$

The problem is how to choose the function $R_s(\boldsymbol{c})$. We suggest to choose $R_s(\boldsymbol{c})$ so that σ_w becomes a constant independent of $\boldsymbol{c}$. Solving for $R_s(\boldsymbol{c})$ results in

$$R_s(\boldsymbol{c}) = \frac{1}{2}\log\left(1 + \frac{\sigma^2}{\sigma_w^2}\frac{c_{31}^2 P_1(\boldsymbol{c}) + c_{21}^2 P_1(\boldsymbol{c}) + \sigma^2}{c_{31}^2 P_1(\boldsymbol{c}) + \sigma^2}\right). \tag{165}$$

The compressed signal is transmitted on the channel between the relay and the destination. The capacity of this channel, with the power allocation rule (57) is given by

$$R_f = \frac{1}{2}\int_A \log\left(1+\frac{c_{32}^2 P_2(\boldsymbol{c})}{\sigma^2 + c_{31}^2 P_1(\boldsymbol{c})}\right) df(c_{32})df(c_{31}) \quad (166)$$

with

$$A = \left\{(c_{31},c_{32})\left|\frac{c_{31}^2}{c_{32}^2} < t\right.\right\}.$$

Now, as in the proof of Proposition 3, the rate of the compressed signal must match the capacity of the channel between relay and the destination, i.e., we must have $E_{\boldsymbol{c}}[R_s(\boldsymbol{c})] = R_f$. Since the power allocation rule is fixed (for specific t), this amounts to finding the value of the constant σ_w. We will restrict ourselves to solving this problem in the high-SNR regime $\sigma^2 \to 0$ for Rayleigh fading. First, the capacity of the forwarding channel between relay and destination is

$$\begin{aligned} R_f^\infty &= \lim_{\sigma^2\to 0}\frac{1}{2}\int_A \log\left(1+\frac{c_{32}^2 P_2}{\sigma^2 + c_{31}^2 P_1}\right) df(c_{32})df(c_{31}) \\ &= \frac{1}{2}\int_A \\ &\quad \log\left(1+\frac{c_{32}^2\left(t-\frac{c_{31}^2}{c_{32}^2-c_{31}^2}(1-t)\right)}{c_{31}^2\frac{c_{32}^2}{c_{32}^2-c_{31}^2}(1-t)}\right) df(c_{32})df(c_{31}) \\ &= \frac{1}{2}\int_A \log\left(\frac{t}{1-t}\left(\frac{c_{32}^2}{c_{31}^2}-1\right)\right) df(c_{32})df(c_{31}) \\ &= \frac{1}{2}\frac{s_{32}}{s_{31}+s_{32}}\log\left(\frac{1+\frac{s_{32}}{s_{31}}t}{1-t}\right). \end{aligned} \quad (167)$$

Clearly, σ_w is also a function of σ^2. As will be seen in the following, we get a valid solution for $\sigma^2 \to 0$ if we put $\sigma_w = \Sigma^{-1}\sigma^2$, where Σ is a constant independent of σ^2. $R_s(\boldsymbol{c})$ then becomes

$$\begin{aligned} R_s^\infty(\boldsymbol{c}) &= \frac{1}{2}\log\left(1+\Sigma\left(1+\frac{c_{21}^2}{c_{31}^2}\right)\right) \\ &= \frac{1}{2}\log\left(1+\frac{\Sigma}{1+\Sigma}\frac{c_{21}^2}{c_{31}^2}\right) + \frac{1}{2}\log(1+\Sigma) \end{aligned} \quad (168)$$

and further

$$\begin{aligned} E_{\boldsymbol{c}}[R_s^\infty(\boldsymbol{c})] &= \frac{1}{2}\int_{R^2}\log\left(1+\Sigma\left(1+\frac{c_{21}^2}{c_{31}^2}\right)\right) df(c_{21})df(c_{31}) \\ &= \frac{1}{2}\frac{\Sigma s_{21}}{\Sigma s_{21}-(1+\Sigma)s_{31}}\log\left(\frac{\Sigma s_{21}}{(1+\Sigma)s_{31}}\right) \\ &\quad + \frac{1}{2}\log(1+\Sigma). \end{aligned} \quad (169)$$

For given t we can then find Σ by equalizing R_f^∞ given by (167) and $E_{\boldsymbol{c}}[R_s^\infty(\boldsymbol{c})]$ given by (169).

We will finally calculate the asymptotic relay gain achieved. First, by using maximum ratio combining at the destination as in the proof of Proposition 3, we get the following rate:

$$R = \frac{1}{2}\int_{R^3}\log\left(1+\frac{c_{31}^2 P_1(\boldsymbol{c})}{\sigma^2}+\frac{c_{21}^2 P_1(\boldsymbol{c})}{\sigma^2+\sigma_w^2}\right) df(\boldsymbol{c}) \quad (170)$$

and

$$\begin{aligned} R^\infty &\triangleq \lim_{\sigma^2\to 0}\left(R+\frac{1}{2}\log(\sigma^2)\right) \\ &= \frac{1}{2}\int_{R^3}\log\left(c_{31}^2 P_1(\boldsymbol{c}) + \frac{\Sigma}{1+\Sigma}c_{21}^2 P_1(\boldsymbol{c})\right) df(\boldsymbol{c}). \end{aligned} \quad (171)$$

The asymptotic relay gain can now be found as $R^\infty - R_d^\infty$ with R_d^∞ given by (133)

$$\begin{aligned} 2R_g &= \int_{R^2-A}\log\left(\left(c_{31}^2+\frac{\Sigma}{1+\Sigma}c_{21}^2\right)P\right) df(c_{21})df(c_{31}) \\ &\quad + \int_A \log\left(c_{31}^2+\frac{\Sigma}{1+\Sigma}c_{21}^2\right) df(\boldsymbol{c}) \\ &\quad + \int_A \log\left(\frac{c_{32}^2}{c_{32}^2-c_{31}^2}(1-t)P\right) df(\boldsymbol{c}) \\ &\quad - \int_R \log\left(c_{31}^2 P\right) df(c_{31}) \\ &= \int_{R^2}\log\left(1+\frac{\Sigma}{1+\Sigma}\frac{c_{21}^2}{c_{31}^2}\right) df(c_{21})df(c_{31}) \\ &\quad + \int_A \log\left(\left(\frac{c_{32}^2}{c_{32}^2-c_{31}^2}(1-t)\right)\right) df(c_{21})df(c_{32}) \\ &= \frac{\Sigma s_{21}}{\Sigma s_{21}-(1+\Sigma)s_{31}}\log\left(\frac{\Sigma s_{21}}{(1+\Sigma)s_{31}}\right) \\ &\quad + \frac{s_{32}\log(1-t)+s_{31}\log\left(1+t\frac{s_{32}}{s_{31}}\right)}{s_{31}+s_{32}}. \end{aligned} \quad (172)$$

This should then be maximized with respect to t, bearing in mind that Σ is also a function of t.

ACKNOWLEDGMENT

We wish to thank the anonymous reviewers for their many helpful comments that improved the presentation of this paper.

REFERENCES

[1] A. Goldsmith and S. Wicker, "Design challenges for energy-constrained ad-hoc wireless networks," *IEEE Wireless Commun.*, vol. 9, no. 4, pp. 8–27, Aug. 2002.

[2] W. Stark, H. Wang, A. Worthen, S. Lafortune, and D. Teneketzis, "Low-energy wireless communication network design," *IEEE Wireless Commun.*, vol. 9, no. 4, pp. 60–72, Aug. 2002.

[3] A. Ephremides, "Energy concerns in wireless networks," *IEEE Wireless Commun.*, vol. 9, no. 4, pp. 48–59, Aug. 2002.

[4] R. Min, M. Bhardwaj, S. Cho, N. Ickes, E. Shih, A. Sinha, A. Wang, and A. Chandrakasan, "Energy-centric enabling technologies for wireless sensor networks," *IEEE Wireless Commun.*, vol. 9, no. 4, pp. 28–39, Aug. 2002.

[5] O. Koymen, V. Rodoplu, and T. Meng, "Throughput characteristics of a minimum energy wireless network," in *Proc. IEEE Int. Conf. Communications (ICC'01)*, vol. 8, Helsinki, Finland, Jun. 2001, pp. 2568–2572.

[6] V. Rodoplu and T. Meng, "Minimum energy mobile wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 8, pp. 1333–1344, Aug. 1999.

[7] F. Willems and E. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 313–327, May 1985.

[8] A. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 6, pp. 841–850, Nov. 1982.

[9] A. Sendonaris, E. Erkip, and B. Aazhang, "Increasing uplink capacity via user cooperation diversity," in *Proc. IEEE Int. Symp. Information Theory (ISIT'98)*, Cambridge, MA, Aug. 1998, p. 156.

[10] ——, "User cooperation diversity-Part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

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

[12] A. Stefanov and E. Erkip, "Cooperative information transmission in wireless networks," in *Proc. Asian-European ITW 2002*, Breisach, Germany, Jun. 2002.

[13] J. Laneman and G. Wornell, "Exploiting distributed spatial diversity in wireless networks," in *Proc. 38th Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2000.

[14] ——, "Energy-efficient antenna sharing and relaying for wireless networks," in *Proc. IEEE Wireless Communications and Networking Conf. (WCNC'00)*, Chicago, IL, Sep. 2000.

[15] J. N. Laneman, G. W. Wornell, and D. N. C. Tse, "Cooperative diversity in wireless networks, efficient protocols and outtage behavior," *IEEE Trans. Inf. Theory*, submitted for publication.

[16] J. Laneman and G. Wornell, "Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks," in *Proc. Global Telecommunications Conf. (GLOBECOM'02)*, Taipei, Taiwan, Nov. 2002.

[17] J. Laneman, E. Martinian, G. Wornell, J. Apostolopoulos, and J. Wee, "Comparing application- and physical-layer approaches to diversity on wireless channels," in *Proc. IEEE Int. Conf. Communications (ICC'03)*, Anchorage, AK, 2003.

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

[19] ——, "A survey of multi-way channels in information theory: 1961–1976," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 1, pp. 1–37, Jan. 1977.

[20] H. Sato, "Information transmission through a channel with relay," University of Hawaii, Honolulu, Tech. Rep. B76–7, 1976.

[21] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.

[22] C. Zeng, F. Kuhlmann, and A. Buzo, "Achievability proof of some multiuser channel coding theorems using backward decoding," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1160–1165, Nov. 1989.

[23] M. Hasna and M. Alouini, "Optimal power allocation for relayed transmissions over Rayleigh fading channels," in *Proc. 57th IEEE Semiannual Vehicular Technology Conf. (VTC'03-Spring)*, Apr. 2003.

[24] ——, "A performance study of dual-hop transmissions with fixed gain relays," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP '03)*, Hong Kong, Apr. 2003.

[25] L. Yang, M. Hasna, and M. Alouini, "Average outage duration of multihop communication systems with regenerative relays," in *Proc. 57th IEEE Semiannu. Vehicular Technology Conf. (VTC'03-Spring)*, Jeju Island, Korea, Apr. 2003.

[26] M. Hasna and M. Alouini, "Application of the harmonic mean statistics to the end-to-end performance of transmission systems with relays," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM'02)*, Taipei, Taiwan, Nov. 2002.

[27] ——, "Outage probability of multihop transmission over Nakagami fading channels," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 216–218, May 2003.

[28] ——, "Outage probability of multihop transmission over Nakagami fading channels," in *Proc. IEEE Int. Workshop on Advances in Wireless Communications (ISWC'02)*, Taiwan, 2002.

[29] M. Valenti and N. Correal, "Exploiting macrodiversity in dense multihop networks and relay channels," in *Proc. IEEE Vehicular Technology Conf. (VTC'02 Fall)*, Vancouver, BC, Oct. 2002.

[30] I. Maric and R. Yates, "Efficient multihop broadcast for wideband systems," in *Proc. 40th Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2002.

[31] A. Reznik, S. Kulkarni, and S. Verdú, "Capacity and optimal resource allocation in the degraded Gaussian relay channel with multiple relays," in *Proc. 40th Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2002.

[32] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *Proc. IEEE Int. Symp. Information Theory (ISIT'00)*, Sorrento, Italy, Jun. 2000.

[33] P. Vamroose and E. van der Meulen, "Uniquely decodable codes for deterministic relay channels," *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1203–1212, Jul. 1992.

[34] A. El Gamal and S. Zahedi, "Minimum energy communication over a relay channel," in *Proc. IEEE Int. Symp. Information Theory (ISIT'03)*, Yokohama, Japan, Jun./Jul. 2003.

[35] M. Gastpar and M. Vetterli, "On the asymptotic capacity of Gaussian relay networks," in *Proc. IEEE Int. Symp. Information Theory (ISIT'02)*, Lausanne, Switzerland, Jun./Jul. 2002, p. 195.

[36] M. Gastpar, G. Kramer, and P. Gupta, "The multiple-relay channel: Coding and antenna-clustering capacity," in *Proc. IEEE Int. Symp. Information Theory (ISIT'02)*, Lausanne, Switzerland, Jun. 2002, p. 137.

[37] D. M. Pozar, *Microwave Engineering*. New York: Wiley, 1998.

[38] W. Shiroma and M. P. D. Lisio, "Quasioptical circuits," in *Wiley Encyclopedia of Electrical and Electronics Engineering*. New York: Wiley, 1999, pp. 523–533.

[39] G. Kramer, P. Gupta, and M. Gastpar, "Information-theoretic multihopping for relay networks," in *Proc. 2004 Int. Zurich Seminar on Communications*, Zurich, Switzerland, 2004.

[40] T. Cover and J. Thomas, *Information Theory*. New York: Wiley, 1991.

[41] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[42] A. Høst-Madsen, "On the capacity of wireless relaying," in *Proc. IEEE Vehicular Technology Conf. (VTC'02 Fall)*, Vancouver, BC, Canada, 2002.

[43] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

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

[45] G. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, Autumn 1996.

[46] M. Alouini and A. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp. 1165–1181, Jul. 1999.

[47] S. K. Jayaweera and H. V. Poor, "Capacity of multiple-antenna systems with both receiver and transmitter channel state information," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2697–2709, Oct. 2003.

[48] J. Liu, J. Chen, M. P. Fossorier, and A. Høst-Madsen, "Capacity-approaching multiple coding for MIMO Rayleigh fading systems with transmit channel state information," *IEEE Trans. Commun.*, submitted for publication.

[49] M. C. Valenti and B. Zhao, "Capacity approaching distributed turbo codes for the relay channel," in *Proc. 57th IEEE Semiannual Vehicular Technology Conf. (VTC'03-Spring)*, Jeju Island, Korea, Apr. 2003.

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

[51] A. Wyner, "The rate-distortion function for source coding with side information at the decoder-ii: General sources," *Inf. Control*, vol. 38, pp. 60–80, 1978.