

# Correspondence

## Gaussian Orthogonal Relay Channels: Optimal Resource Allocation and Capacity

Yingbin Liang, *Student Member, IEEE*, and  
Venugopal V. Veeravalli, *Senior Member, IEEE*

**Abstract**—A Gaussian orthogonal relay model is investigated, where the source transmits to the relay and destination in channel 1, and the relay transmits to the destination in channel 2, with channels 1 and 2 being orthogonalized in the time–frequency plane in order to satisfy practical constraints. The total available channel resource (time and bandwidth) is split into the two orthogonal channels, and the resource allocation to the two channels is considered to be a design parameter that needs to be optimized. The main focus of the analysis is on the case where the source-to-relay link is better than the source-to-destination link, which is the usual scenario encountered in practice. A lower bound on the capacity (achievable rate) is derived, and optimized over the parameter  $\theta$ , which represents the fraction of the resource assigned to channel 1. It is shown that the lower bound achieves the max-flow min-cut upper bound at the optimizing  $\theta$ , the common value thus being the capacity of the channel at the optimizing  $\theta$ . Furthermore, it is shown that when the relay-to-destination signal-to-noise ratio (SNR) is less than a certain threshold, the capacity at the optimizing  $\theta$  is also the maximum capacity of the channel over all possible resource allocation parameters  $\theta$ . Finally, the achievable rates for optimal and equal resource allocations are compared, and it is shown that optimizing the resource allocation yields significant performance gains.

**Index Terms**—Achievable rate, decode-and-forward relay, parallel relay channel.

### I. INTRODUCTION

The relay channel was introduced by van der Meulen [1], and was comprehensively studied by Cover and El Gamal [2]. In a commonly studied version of the relay channel, no restrictions are imposed on the transmitted and received signals at the relay node, and perfect “echo cancellation” of the transmitted signal at the receiver of the relay is implicitly assumed. However, such perfect echo cancellation may not be easy to implement in practice. A practical way to isolate the transmitted and received signals at the relay node is to orthogonalize these signals. Toward this end, recent papers on relay channels have considered frequency division [3], time division [4]–[7], code division [8], and general orthogonal division [9], [10] to orthogonalize transmitted and received signals at the relay node.

In this correspondence, we study a Gaussian orthogonal relay model, where the source transmits to the relay and destination in one orthogonal channel (channel 1), and the relay transmits to the destination in the other orthogonal channel (channel 2). Assuming a total available bandwidth of  $W$  hertz, the total channel resource of  $2W$  dimensions per second is split into the two channels, in order to orthogonalize the transmitted and received signals at the relay node. In contrast to previous studies of this model [3], [4], [7], we allow the resource allocation

Manuscript received August 17, 2004; revised January 26, 2005. This work was supported by the National Science Foundation under CAREER/PECASE award CCF-0049089 and by a Vodafone Foundation Graduate Fellowship. The material in this correspondence was presented at the 38th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, March 2004.

The authors are with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: yliang1@uiuc.edu; vvv@uiuc.edu).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2005.853305

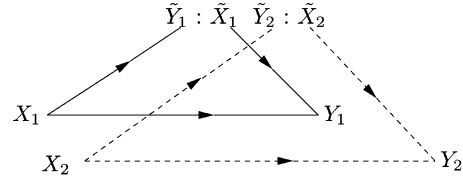


Fig. 1. Parallel relay channel.

represented by  $\theta$  (resource fraction allocated to channel 1) to be a design parameter that may be optimized. We focus on the case where the source-to-relay link is better than the source-to-destination link.

For this model, we derive a lower bound on the capacity (achievable rate) as a function of resource allocation parameter  $\theta$ , and characterize the value  $\theta^*$  that maximizes the achievable rate. We obtain the somewhat surprising result that the lower bound achieves the max-flow min-cut upper bound at  $\theta^*$ . We hence obtain the capacity of the channel when  $\theta$  is fixed at  $\theta^*$ . Furthermore, we show that when the relay-to-destination signal-to-noise (SNR) is less than a certain threshold, the capacity at  $\theta^*$  is also the maximum capacity of the channel over all possible resource allocation parameters  $\theta$ .

In the following sections, we first describe the channel model, and then present our main results.

### II. SYSTEM MODEL

In this section, we first present a two-dimensional discrete memoryless parallel relay channel whose capacity bounds are useful for our analysis. We then introduce the orthogonal relay model considered in this correspondence.

#### A. Parallel Relay Channel Model

Consider a two-dimensional discrete memoryless parallel relay channel, where the source, relay, and destination communicate on two parallel (independent) links as shown in Fig. 1. The two sets of links are indicated by solid and dashed lines, respectively, in the figure. We use  $(x_1, x_2)$  and  $(\tilde{x}_1, \tilde{x}_2)$  to denote the signals sent from the source and relay, respectively, and  $(y_1, y_2)$  and  $(\tilde{y}_1, \tilde{y}_2)$  to denote the signals received by the destination and relay, respectively. Under the assumption that the parallel links are independent we have the following channel transition probability:

$$p(y_1, y_2, \tilde{y}_1, \tilde{y}_2 | x_1, x_2, \tilde{x}_1, \tilde{x}_2) = p(y_1, \tilde{y}_1 | x_1, \tilde{x}_1) p(y_2, \tilde{y}_2 | x_2, \tilde{x}_2). \quad (1)$$

Note that the parallel relay channel can be seen as a special case of the classical relay channel with vector inputs and outputs. We can hence obtain the following bounds on the capacity of the parallel relay channel by applying the results in [2] and exploiting (1).

**Theorem 1:** For the discrete memoryless parallel relay channel with transition probability defined in (1), the following lower and upper bounds on the capacity hold:

$$\begin{aligned} C_{\text{low}} &= \sup_{p(x_1, \tilde{x}_1) p(x_2, \tilde{x}_2)} \min \{ I(X_1, \tilde{X}_1; Y_1) + I(X_2, \tilde{X}_2; Y_2), \\ &\quad I(X_1; \tilde{Y}_1 | \tilde{X}_1) + I(X_2; \tilde{Y}_2 | \tilde{X}_2) \}, \\ C_{\text{up}} &= \sup_{p(x_1, \tilde{x}_1) p(x_2, \tilde{x}_2)} \min \{ I(X_1, \tilde{X}_1; Y_1) + I(X_2, \tilde{X}_2; Y_2), \\ &\quad I(\tilde{X}_1; \tilde{Y}_1, Y_1 | \tilde{X}_1) + I(X_2; \tilde{Y}_2, Y_2 | \tilde{X}_2) \} \end{aligned} \quad (2)$$

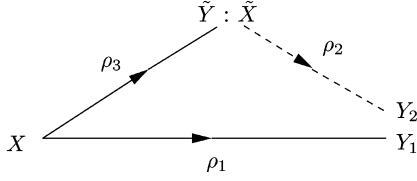


Fig. 2. Orthogonal relay channel model.

where the optimizing input distribution is such that  $(X_1, \tilde{X}_1)$  is independent of  $(X_2, \tilde{X}_2)$ .

A sketch of the proof of the above theorem is given in Appendix I.

### B. Gaussian Orthogonal Relay Channel Model

Orthogonal relay channels are special cases of the parallel relay channel, which satisfy the additional constraint that the relay should transmit and receive in orthogonal channels (channels 1 and 2). The reason for imposing this constraint is that in practice it is difficult to design the relay so that it can transmit and receive at the same time in the same frequency band. Assuming at total available bandwidth of  $W$  hertz, the total channel resource of  $2W$  dimensions per second is split into channels 1 and 2. Physically this can be realized by splitting either the total transmission time or the total bandwidth, or splitting the dimensions in the joint time–frequency space. We let the parameter  $\theta$  denote the fraction of dimensions allocated to channel 1, and hence  $\bar{\theta} := 1 - \theta$  is the fraction of dimensions allocated to channel 2. We refer to  $\theta$  as the channel resource allocation parameter. The total channel resource needs to be split in such a way that maximizes the achievable rate of the system.

We consider an orthogonal relay model, where the source transmits to the relay and destination in channel 1, and the relay transmits to the destination in channel 2. Although we may allow the source to transmit in channel 2 without violating the orthogonal constraint on the relay node, in our model we only permit the source to transmit in channel 1 in order to keep the source transmitter simple. It is easy to see that this model is a special case of the parallel relay channel, with  $\tilde{X}_1 = 0$ ,  $X_2 = 0$ , and  $\tilde{Y}_2 = 0$ . We then use the notation  $X := X_1$ ,  $\tilde{X} := \tilde{X}_2$ , and  $\tilde{Y} := \tilde{Y}_1$  and draw the orthogonal relay channel model in Fig. 2 with the solid and dashed lines indicating the links of channels 1 and 2, respectively. For this model, we consider the Gaussian case, where all the transmission links are corrupted by additive white Gaussian noises. Hence, the channel is memoryless across transmission symbols.

The relationships between input and output symbols in channels 1 and 2 can be written as

$$\begin{aligned} Y_1 &= X + Z_1 \\ Y_2 &= \tilde{X} + Z_2 \\ \tilde{Y} &= X + \tilde{Z} \end{aligned} \quad (3)$$

where  $Z_1$ ,  $Z_2$ , and  $\tilde{Z}$  are zero-mean independent Gaussian random variables with variances  $N_0/2$ ,  $N_0/2$  and  $\tilde{N}_0/2$ , respectively. The input symbol sequences  $\{X_n\}$  and  $\{\tilde{X}_n\}$  are subject to the average power constraints  $P$  and  $\tilde{P}$ , respectively. Note that channel 1 is assigned  $2W\theta$  dimensions per second, and hence we may transmit  $2W\theta$  symbols per second on this channel. Similarly, we may transmit  $2W\bar{\theta}$  symbols per second on channel 2. For notational convenience, we introduce three normalized SNR parameters,  $\rho_1 := (P/N_0W)$ ,  $\rho_2 := (\tilde{P}/N_0W)$ , and  $\rho_3 := (P/\tilde{N}_0W)$ , corresponding to the three transmission links of the channel (see Fig. 2). Throughout our discussion, we are interested in the case where  $\rho_3 > \rho_1$ , i.e., the source-to-relay link is better than the source-to-destination link.

Our goal is to find the resource allocation parameter  $\theta$  that maximizes the achievable rate (lower bound on the capacity), and to study the capacity at such an optimal resource allocation.

### III. MAIN RESULTS FOR GAUSSIAN ORTHOGONAL RELAY MODEL

The following theorem summarizes the bounds on the capacity for the Gaussian orthogonal relay model.

*Theorem 2:* For the Gaussian orthogonal relay model given in (3), a lower bound on the capacity is given by

$$C_{1ow} = \max_{0 \leq \theta \leq 1} \min \left\{ \theta C \left( \frac{\rho_1}{\theta} \right) + \bar{\theta} C \left( \frac{\rho_2}{\bar{\theta}} \right), \theta C \left( \frac{\rho_3}{\theta} \right) \right\} \text{ b/s/Hz} \quad (4)$$

where the function  $C(x) := (1/2) \log(1 + x)$ .

An upper bound on the capacity as a function of  $\theta$  is given by

$$C_{up}(\theta) = \min \left\{ \theta C \left( \frac{\rho_1}{\theta} \right) + \bar{\theta} C \left( \frac{\rho_2}{\bar{\theta}} \right), \theta C \left( \frac{\rho_3 + \rho_1}{\theta} \right) \right\} \text{ b/s/Hz}. \quad (5)$$

*Remark 1:* The lower bound is maximized over the resource allocation parameter  $\theta$ . Throughout the correspondence, we use “optimal resource allocation” to refer to the value  $\theta^*$  that maximizes the lower bound (4) on the capacity.

*Proof:* First, the capacity bounds in Theorem 1 can be specialized for the orthogonal relay model, considering that the source transmits  $2W\theta$  symbols per second in channel 1, and the relay transmits  $2W\bar{\theta}$  symbols per second in channel 2

$$C_{1ow} = \sup_{p(x)p(\tilde{x})} \min \{ \theta I(X; Y_1) + \bar{\theta} I(\tilde{X}; Y_2), \theta I(X; \tilde{Y}) \} \text{ b/s/Hz} \quad (6)$$

$$C_{up} = \sup_{p(x)p(\tilde{x})} \min \{ \theta I(X; Y_1) + \bar{\theta} I(\tilde{X}; Y_2), \theta I(X; \tilde{Y}, Y_1) \} \text{ b/s/Hz}. \quad (7)$$

Note that the bounds given in (6) and (7) are normalized to be in bits per second per hertz. Moreover, the optimal input distribution has  $X$  and  $\tilde{X}$  being statistically independent. We apply (6) and (7) to the Gaussian orthogonal relay channel. For the lower bound, it is straightforward to see that the optimal input distributions for  $X$  and  $\tilde{X}$  are Gaussian. For the upper bound, the same Gaussian input  $X$  maximizes  $I(X; \tilde{Y}, Y_1)$ . This can be shown by using the maximum entropy theorem for random vectors with a given covariance [11, Theorem 9.6.5, p. 234]. The bounds (4) and (5) then follow by choosing the source input symbol  $X$  to be  $\mathcal{N}(0, P/(2W\theta))$ , and the relay input symbol  $\tilde{X}$  to be  $\mathcal{N}(0, \tilde{P}/(2W\bar{\theta}))$  independently from the source symbol. Note that the lower bound is further tightened by maximizing over the resource allocation parameter  $\theta$ .  $\square$

*Remark 2:* In the preceding proof, we directly applied the upper bound for the discrete memoryless channel to the Gaussian channel. This is only a heuristic argument. A rigorous proof for the upper bound follows the converse proof for the discrete memoryless channel, applying the average power constraint for each codeword. The derivation follows easily using the steps in [12, pp. 7-18 to 7-20] and is hence omitted in this correspondence.

For convenience, we use  $C_{1,1ow}(\theta)$  and  $C_{2,1ow}(\theta)$  to denote, respectively, the first and second terms over which the minimization is taken in the expression for the lower bound (4), and define

$$C_{1ow}(\theta) := \min \{ C_{1,1ow}(\theta), C_{2,1ow}(\theta) \}. \quad (8)$$

Since we are only interested in the case where the source-to-relay link is better than the source-to-destination link, we always assume that  $\rho_3 > \rho_1$ .

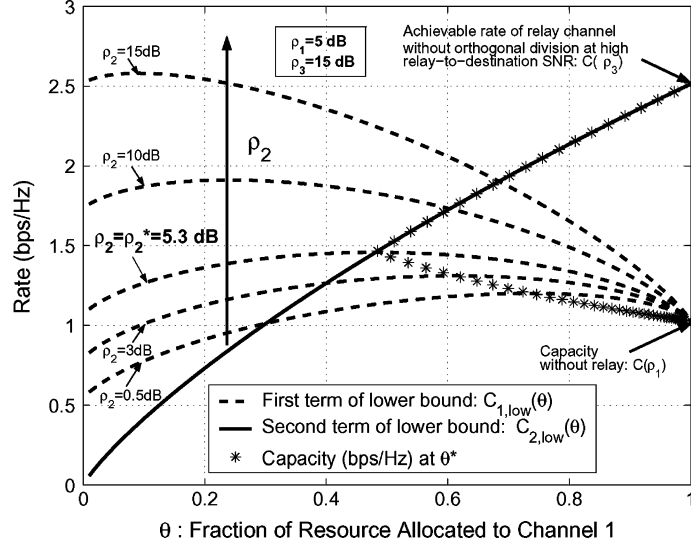


Fig. 3. Capacity curve for the Gaussian orthogonal relay model.

We now discuss how to maximize the lower bound over the resource allocation parameter  $\theta$ , for given  $\rho_1$  and  $\rho_3$ . We essentially need to tradeoff the two curves  $C_{1,low}(\theta)$  and  $C_{2,low}(\theta)$  as functions of  $\theta$  shown by the dashed and solid lines, respectively, in Fig. 3. Since

$$C_{1,low}(0) > C_{2,low}(0) \quad \text{and} \quad C_{1,low}(1) < C_{2,low}(1)$$

there exists at least one cross point  $\theta$  such that  $C_{1,low}(\theta) = C_{2,low}(\theta)$  (we use  $\theta^r$  to indicate the largest such value). Note that  $C_{2,low}(\theta)$  is a strictly increasing function of  $\theta$ . It can be easily shown that  $C_{1,low}(\theta)$  is a strictly concave function for  $\theta \in [0, 1]$  (see Fig. 3). Therefore,  $C_{1,low}(\theta)$  has a unique maximum over  $\theta \in [0, 1]$ . Let  $\theta^o$  denote the corresponding maximizing  $\theta$ .

The optimal  $\theta^*$  that maximizes the lower bound  $C_{low}(\theta)$  falls into the following two cases. If

$$C_{1,low}(\theta^o) < C_{2,low}(\theta^o)$$

(equivalently,  $\theta^o > \theta^r$ ), then  $\theta^o$  maximizes the lower bound  $C_{low}(\theta)$ . If

$$C_{1,low}(\theta^o) \geq C_{2,low}(\theta^o)$$

(equivalently,  $\theta^o \leq \theta^r$ ), then  $\theta^r$  maximizes the lower bound  $C_{low}(\theta)$ .

As shown in Fig. 3, each value of  $\rho_2$  corresponds to a dashed curve  $C_{1,low}(\theta)$ , and this dashed curve jointly with the solid curve  $C_{2,low}(\theta)$  determines an optimal lower bound that is one point on the “star” line, and the corresponding value of  $\theta$  is the optimal  $\theta^*$ . Also, note that for fixed  $\rho_1$  and  $\rho_3$ ,  $\theta^o$  decreases as  $\rho_2$  increases, and  $C_{1,low}(\theta^o)$  increases as  $\rho_2$  increases. Hence, there exists a  $\rho_2^*$  such that  $\theta^o = \theta^r$ , and the two conditions  $\rho_2^* \leq \rho_2$  and  $\theta^o \leq \theta^r$  are equivalent, i.e.,  $\rho_2^* \leq \rho_2 \Leftrightarrow \theta^o \leq \theta^r$ . Hence, those two cases are also equivalently described by  $\rho_2 < \rho_2^*$  and  $\rho_2 \geq \rho_2^*$ , respectively. Note that  $\rho_2^*$  is completely determined by the given  $\rho_1$  and  $\rho_3$ .

We summarize the above analysis of the optimal resource allocation parameter  $\theta^*$  in the following theorem.

**Theorem 3:** Assume  $\rho_3 > \rho_1$ . Then, for given  $\rho_1$  and  $\rho_3$ , the optimal resource allocation parameter  $\theta^*$  that maximizes the lower bound falls into the following two cases.

Case 1: If  $0 < \rho_2 < \rho_2^*$ , i.e.,  $\theta^o > \theta^r$ , then the optimal resource allocation is achieved by  $\theta^* = \theta^o$ , and the corresponding lower bound is

$$C_{low} = C_{1,low}(\theta^o).$$

If  $\rho_2 = 0$ , then  $\theta^* = 1$ , and the lower bound becomes  $C(\rho_1)$ , the capacity of the channel without the relay node.

Case 2: If  $\rho_2^* \leq \rho_2$ , i.e.,  $\theta^o \leq \theta^r$ , then the optimal resource allocation is achieved by  $\theta^* = \theta^r$ , and the corresponding lower bound is

$$C_{low} = C_{1,low}(\theta^r) = C_{2,low}(\theta^r).$$

As  $\rho_2 \rightarrow \infty$ ,  $\theta^* \rightarrow 1$ , and the lower bound approaches  $C(\rho_3)$ , the achievable rate of the relay channel without orthogonal division at high relay-to-destination SNR.

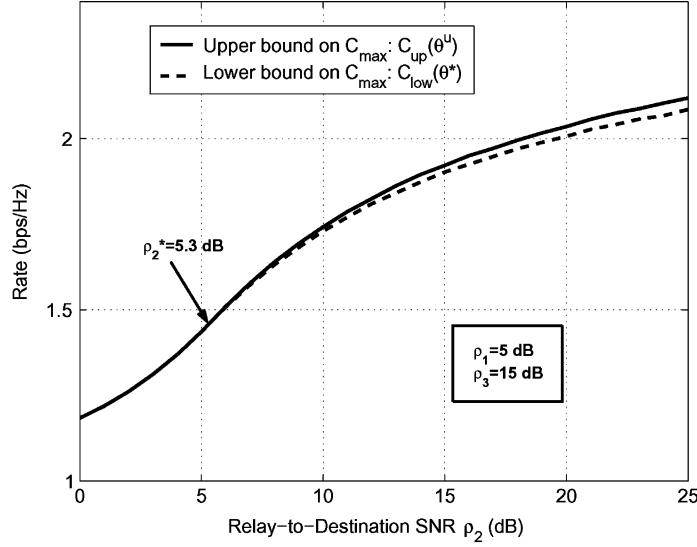
From Theorem 3, it is clear that the Gaussian orthogonal relay channel can achieve a rate that is at least equal to the capacity of the direct link from the source to destination. The rate increases as  $\rho_2$  increases, and approaches the achievable rate of the relay channel without orthogonal division as  $\rho_2$  goes to infinity.

Comparing the lower bound  $C_{low}(\theta^*)$  (using optimal resource allocation for  $C_{low}(\theta)$ ) with the upper bound evaluated at  $\theta^*$ ,  $C_{up}(\theta^*)$ , in the two cases of Theorem 3, we immediately conclude that the lower bound is tight when the optimal resource allocation is used.

**Theorem 4:** For given parameters  $\rho_1, \rho_2$ , and  $\rho_3$  with  $\rho_3 > \rho_1$ , the lower bound  $C_{low}(\theta^*)$  is the capacity of the Gaussian orthogonal relay channel with the resource allocation parameter being fixed at  $\theta^*$ .

Fig. 3 plots the  $C_{1,low}(\theta)$  curves (dashed lines),  $C_{2,low}(\theta)$  curve (solid line), and the capacity curve (star line) at  $\rho_1 = 5$  dB and  $\rho_3 = 15$  dB. Each dashed line in the graph corresponds to one relay-to-destination SNR  $\rho_2$ , and the tradeoff between each dashed line and the solid line determines one optimal resource allocation parameter  $\theta^*$  and the capacity at the corresponding  $\theta^*$  and  $\rho_2$ . If  $\rho_2 < \rho_2^*$ , the optimal  $\theta^*$  and the capacity at  $\theta^*$  are determined by the peak of the dashed curve  $C_{1,low}(\theta)$ . The value  $\rho_2^*$  is the SNR where the corresponding  $C_{1,low}(\theta)$  achieves the peak value at its cross point with  $C_{2,low}(\theta)$ . If  $\rho_2 \geq \rho_2^*$ , the optimal  $\theta^*$  is determined by the cross point of  $C_{1,low}(\theta)$  and  $C_{2,low}(\theta)$ . And as  $\rho_2$  increases beyond  $\rho_2^*$ , the capacity at  $\theta^*$  (star curve) coincides with each corresponding point of  $C_{2,low}(\theta)$ , and finally approaches the achievable rate of the relay channel without orthogonal division as  $\rho_2$  goes to infinity.

The star line in the graph is the capacity curve for values of  $\rho_2$  in the range  $[0, \infty)$ . Each point on this curve corresponds to the capacity at a certain value of  $\rho_2$  with resource allocation  $\theta^*$ . It is clear from the graph that when  $\rho_2$  is small, as  $\rho_2$  increases we assign more of the resource to channel 2 ( $\theta^*$  decreases). This is because a higher  $\rho_2$  makes the relay more useful to the source, and hence, makes it more deserving of a


 Fig. 4. Bounds on  $C_{\max}$ .

larger portion of the resources. However, beyond the threshold  $\rho_2^*$ , as  $\rho_2$  increases, we assign a smaller portion of the resource to channel 2. This behavior is also as expected because at such high  $\rho_2$ , the relay-to-destination link is so strong that a small assignment of the resource is sufficient for the relay to help the source, and it is hence preferable to assign a larger portion of the resource to the source to allow for more information to be transferred to the relay.

We next consider the maximum rate that an orthogonal relay channel can support. We let  $C(\theta)$  denote the capacity of an orthogonal relay channel with resource allocation parameter  $\theta$ . Then

$$C_{\max} := \max_{0 \leq \theta \leq 1} C(\theta) \quad (9)$$

is the maximum rate this channel can reliably support. We refer to  $C_{\max}$  as the maximum capacity.

In the previous analysis, since  $\theta^*$  only maximizes the lower bound on the capacity  $C_{1ow}(\theta)$ , it does not necessarily maximize  $C(\theta)$ . Hence, the channel capacity  $C_{1ow}(\theta^*)$  at  $\theta^*$  may not equal  $C_{\max}$ . However, we show that if the relay-to-destination SNR  $\rho_2$  is smaller than a threshold  $\rho_2^*$ ,  $C_{1ow}(\theta^*)$  is indeed the maximum capacity of the channel.

**Theorem 5:** If  $\rho_2 < \rho_2^*$ ,  $\theta^*$  maximizes both the lower bound  $C_{1ow}(\theta)$  and the upper bound  $C_{up}(\theta)$ , and hence,  $C_{\max} = C_{1ow}(\theta^*)$ .

*Proof:* We only need to prove that, when  $\rho_2 < \rho_2^*$ , the  $\theta^*$  that maximizes the lower bound  $C_{1ow}(\theta)$  also maximizes the upper bound  $C_{up}(\theta)$ . Then  $C_{\max} = C_{1ow}(\theta^*)$  is an immediate consequence of this fact and Theorem 4.

In the proof, we will use the following two facts about  $\theta^*$ , which are clear in the reasoning for Theorem 3. For  $\rho_2 < \rho_2^*$ , the following facts hold:

Fact I:  $\theta^*$  maximizes  $C_{1,1ow}(\theta)$

Fact II:  $C_{1,1ow}(\theta^*) < C_{2,1ow}(\theta^*)$ .

We let  $C_{1,up}(\theta)$  and  $C_{2,up}(\theta)$  denote the two terms over which the minimization is taken in the expression for the upper bound (5). Then

$$C_{up}(\theta) = \min\{C_{1,up}(\theta), C_{2,up}(\theta)\}. \quad (10)$$

It is easy to see  $C_{1,up}(\theta) = C_{1,1ow}(\theta)$  from their expressions. Hence, from Fact I,  $\theta^*$  also maximizes  $C_{1,up}(\theta)$  for  $\rho_2 < \rho_2^*$ .

Comparing the expressions of  $C_{1,up}(\theta)$  and  $C_{2,up}(\theta)$  with the expressions of  $C_{1,1ow}(\theta)$  and  $C_{2,1ow}(\theta)$ , we have for  $\rho_2 < \rho_2^*$

$$C_{1,up}(\theta^*) = C_{1,1ow}(\theta^*) \stackrel{(a)}{<} C_{2,1ow}(\theta^*) < C_{2,up}(\theta^*) \quad (11)$$

where the inequality (a) is Fact II.

Now, since for  $\rho_2 < \rho_2^*$ ,  $\theta^*$  maximizes  $C_{1,up}(\theta)$  and

$$C_{1,up}(\theta^*) < C_{2,up}(\theta^*)$$

we conclude that  $\theta^*$  maximizes the upper bound  $C_{up}(\theta)$ .  $\square$

Clearly, for fixed  $\rho_1$  and  $\rho_3$ ,  $C_{1ow} = C_{1ow}(\theta^*)$  always serves as a lower bound on  $C_{\max}$  for  $\rho_2 > 0$ . We now define

$$C_{up} := \max_{0 \leq \theta \leq 1} C_{up}(\theta). \quad (12)$$

It is easy to see that  $C_{up}(\theta)$  is a concave function of  $\theta$  for  $\theta \in [0, 1]$ , and is hence maximized at some value of  $\theta$ , denoted by  $\theta^u$ . Then the maximum value  $C_{up} = C_{up}(\theta^u)$  is an upper bound on  $C_{\max}$ . We draw these upper and lower bounds on  $C_{\max}$  in Fig. 4.

It is clear from Fig. 4 that when  $\rho_2 < \rho_2^*$ , the lower bound achieves the upper bound, and hence the common value is the maximum capacity  $C_{\max}$ . This verifies our result in Theorem 5. When  $\rho_2 > \rho_2^*$ , there is a gap between the lower and upper bounds on  $C_{\max}$ . Note that our lower bound  $C_{1ow}$  is derived based on the relay using the decode-and-forward scheme [2, Sec. II]. It is possible for other relaying schemes to yield tighter lower bounds for values  $\rho_2 > \rho_2^*$ . For example, the estimate-and-forward scheme [2, Theorem 6] can be shown to provide a tighter lower bound on  $C_{\max}$  for some values of  $\rho_2 > \rho_2^*$ . In particular, the estimate-and-forward scheme achieves the following rate  $C_{1ow}^{ef}$ :

$$C_{1ow}^{ef} = \max_{0 \leq \theta \leq 1} \theta C \left( \frac{\rho_1}{\theta} + \frac{\rho_3/\theta}{1+b} \right) \text{ b/s/Hz} \quad (13)$$

where the constant  $b$  is defined by

$$b := \frac{\frac{\rho_3}{\theta} + \frac{\rho_1}{\theta} + 1}{\left( \left( \frac{\rho_2}{1-\theta} + 1 \right)^{(1-\theta)/\theta} - 1 \right) \left( \frac{\rho_1}{\theta} + 1 \right)}. \quad (14)$$

The achievable rate  $C_{1ow}^{ef}$  satisfies the following limits:

$$\lim_{\rho_2 \rightarrow \infty} C_{1ow}^{ef} = \lim_{\rho_2 \rightarrow \infty} C_{up} = C(\rho_1 + \rho_3). \quad (15)$$

Equation (15) implies that  $C_{1ow}^{ef}$ , which serves as another lower bound on  $C_{\max}$ , converges to the upper bound on  $C_{\max}$  as  $\rho_2$  goes to infinity. The common limit then becomes the maximum capacity as  $\rho_2$  goes to infinity. Hence, the lower bound based on the estimate-and-forward scheme must outperform the lower bound based on the decode-and-forward scheme for large values of  $\rho_2$  in the range  $\rho_2 > \rho_2^*$ . A similar observation has been made in [13], [14] for various cases in relay networks without orthogonal division. However, our

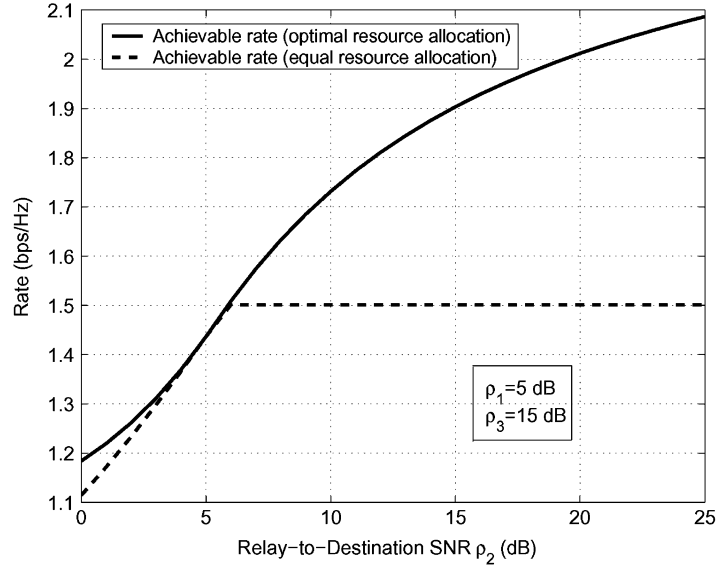


Fig. 5. Comparison of achievable rate with optimal resource allocation and achievable rate with equal resource allocation.

numerical results show that only when  $\rho_3$  is very close to  $\rho_1$ , the estimate-and-forward scheme outperforms the decode-and-forward scheme at moderate values of  $\rho_2$ . For most cases of  $\rho_3 > \rho_1$ , the convergence of  $C_{\text{low}}^{\text{ef}}$  is very slow, and the estimate-and-forward scheme outperforms the decode-and-forward scheme only at extremely large values of  $\rho_2$ . For SNRs of practical interest, the decode-and-forward scheme generally has a better achievable rate.

#### IV. COMPARISON WITH EQUAL RESOURCE ALLOCATION

Fig. 5 plots the achievable rates with the optimal resource allocation, and compares these rates with the rates achieved by equal resource allocation. It is clear from the graph that the performance is greatly enhanced by optimal resource allocation. The graph shows that there is a threshold value of  $\rho_2$  beyond which the rate achieved by equal resource allocation saturates. However, the achievable rate for optimal resource allocation continues to grow with increasing  $\rho_2$ .

#### V. CONCLUSION

We studied the capacity of a Gaussian orthogonal relay model, treating the resource allocation as a parameter that needs to be optimized. We found that at the resource allocation that maximizes the lower bound on the capacity (achievable rate), this lower bound achieves the max-flow min-cut upper bound on the capacity. This result demonstrates that the capacity can be obtained for an interesting case of the relay channel, i.e., the orthogonal relay model at the optimal resource allocation. Our results also underline the practical importance of resource allocation in orthogonal relay channels. Optimizing the resource allocation may be even more important for channels where the transmission loads for channels 1 and 2 are highly asymmetric, e.g., in the relay multiple-access channel studied in [15], [16] and in the relay broadcast channel studied in [14], [17], [18].

In this correspondence, our main focus has been on the case where  $\rho_3 > \rho_1$ , i.e., the source-to-relay link is better than the source-to-destination link. For the case where  $\rho_3 < \rho_1$ , the relay node can use the partial decode-and-forward scheme or the estimate-and-forward scheme to help the transmission. The analysis of resource allocation for the latter case would follow steps similar to those given in this correspondence. However, in practice, for  $\rho_3 < \rho_1$ , the relay node may not even be used, and hence, the study of this case is not of much interest.

In this correspondence, we have assumed that once the resource allocation parameter  $\theta$  is chosen it is fixed throughout transmission. A

recent work [19] studies the case where the resource allocation parameter is allowed to change during transmission.

#### APPENDIX I PROOF OF THEOREM 1

The following lower and upper bounds on the capacity are derived by a direct application of [2, Theorems 1 and 4], taking  $X = (X_1, X_2)$ ,  $\tilde{X} = (\tilde{X}_1, \tilde{X}_2)$ ,  $Y = (Y_1, Y_2)$ , and  $\tilde{Y} = (\tilde{Y}_1, \tilde{Y}_2)$

$$C_{\text{low}} = \sup_{p(x_1, x_2, \tilde{x}_1, \tilde{x}_2)} \min \{ I(X_1, X_2, \tilde{X}_1, \tilde{X}_2; Y_1, Y_2), \\ I(X_1, X_2; \tilde{Y}_1, \tilde{Y}_2 | \tilde{X}_1, \tilde{X}_2) \} \quad (16)$$

$$C_{\text{up}} = \sup_{p(x_1, x_2, \tilde{x}_1, \tilde{x}_2)} \min \{ I(X_1, X_2, \tilde{X}_1, \tilde{X}_2; Y_1, Y_2), \\ I(X_1, X_2; \tilde{Y}_1, \tilde{Y}_2, Y_1, Y_2 | \tilde{X}_1, \tilde{X}_2) \}. \quad (17)$$

The lower bound (16) is optimized over all joint distributions of  $X_1, X_2, \tilde{X}_1, \tilde{X}_2$ . We now prove that using independent  $(X_1, \tilde{X}_1)$  and  $(X_2, \tilde{X}_2)$  always increases the two mutual information terms in (16), and hence, the optimization of the lower bound only needs to be performed over such distributions.

For the first mutual information term in the lower bound expression (16), we have

$$I(X_1, X_2, \tilde{X}_1, \tilde{X}_2; Y_1, Y_2) \\ = H(Y_1, Y_2) - H(Y_1, Y_2 | X_1, X_2, \tilde{X}_1, \tilde{X}_2) \\ \leq H(Y_1) + H(Y_2) - H(Y_1 | X_1, \tilde{X}_1) - H(Y_2 | X_2, \tilde{X}_2) \\ = I(X_1, \tilde{X}_1; Y_1) + I(X_2, \tilde{X}_2; Y_2) \quad (18)$$

where the preceding inequality follows from the basic entropy inequality and the channel transition probability defined in (1). Note that this inequality becomes equality when  $(X_1, \tilde{X}_1)$  is independent of  $(X_2, \tilde{X}_2)$ .

Similarly, for the second mutual information term in the lower bound expression (16), we have

$$I(X_1, X_2; \tilde{Y}_1, \tilde{Y}_2 | \tilde{X}_1, \tilde{X}_2) \\ = H(\tilde{Y}_1, \tilde{Y}_2 | \tilde{X}_1, \tilde{X}_2) - H(\tilde{Y}_1, \tilde{Y}_2 | X_1, X_2, \tilde{X}_1, \tilde{X}_2) \\ \leq H(\tilde{Y}_1 | \tilde{X}_1) + H(\tilde{Y}_2 | \tilde{X}_2) \\ - H(\tilde{Y}_1 | X_1, \tilde{X}_1) - H(\tilde{Y}_2 | X_2, \tilde{X}_2) \\ = I(X_1; \tilde{Y}_1 | \tilde{X}_1) + I(X_2; \tilde{Y}_2 | \tilde{X}_2) \quad (19)$$

where the preceding inequality becomes an equality when  $(X_1, \tilde{X}_1)$  is independent of  $(X_2, \tilde{X}_2)$ .

It is clear from (18) and (19) that independent  $(X_1, \tilde{X}_1)$  and  $(X_2, \tilde{X}_2)$  yield a tighter lower bound. The desired lower bound then follows by combining (18) and (19) and optimizing over such independent input distributions. Similar arguments apply for the upper bound.

#### ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments.

#### REFERENCES

- [1] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Probab.*, vol. 3, pp. 120–154, 1971.
- [2] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.
- [3] A. E. Gamal, M. Mohseni, and S. Zahedi, "On reliable communication over additive white Gaussian noise relay channels," *IEEE Trans. Inf. Theory*, submitted for publication.
- [4] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
- [5] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2020–2040, Jun. 2005.
- [6] M. A. Khojastepour, A. Sabharwal, and B. Aazhang, "On the capacity of 'cheap' relay networks," in *Proc. Conf. Information Sciences and Systems (CISS)*, Baltimore, MD, Mar. 2003. Paper number 153.
- [7] R. U. Nabar, H. Bölcskei, and F. W. Kneubuhler, "Fading relay channels: Performance limits and space-time signal design," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1099–1109, Aug. 2004.
- [8] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—Part I and II," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1948, Nov. 2003.
- [9] A. A. El Gamal and S. Zahedi, "Capacity of a class of relay channels with orthogonal components," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1815–1817, May 2005.
- [10] M. A. Khojastepour and B. Aazhang, "'Cheap' relay channel: A unifying approach to time and frequency division relaying," in *Proc. Annu. Allerton Conf. Communication, Control and Computing*, Monticello, IL, Sep./Oct. 2004, pp. 1792–1801.
- [11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [12] A. A. El Gamal, *Lecture Notes for EE478 (Stanford University): Network Information Theory*. Stanford, CA: Stanford Univ., 2002.
- [13] M. Gastpar, G. Kramer, and P. Gupta, "The multiple-relay channel: Coding and antenna-clustering capacity," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Lausanne, Switzerland, Jun./Jul. 2002, p. 136.
- [14] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.
- [15] G. Kramer and A. J. van Wijngaarden, "On the white Gaussian multiple-access relay channel," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Sorrento, Italy, Jun. 2000, p. 40.
- [16] L. Sankaranarayanan, G. Kramer, and N. B. Mandayam, "Capacity theorems for the multiple-access relay channel," in *Proc. Annu. Allerton Conf. Communication, Control and Computing*, Monticello, IL, Sep./Oct. 2004, pp. 1782–1791.
- [17] Y. Liang and V. V. Veeravalli, "The impact of relaying on the capacity of broadcast channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Chicago, IL, Jun./Jul. 2004, p. 403.
- [18] —, "Cooperative relay broadcast channels," in *Proc. Wirellesscom, Symp. Information Theory*, Maui, HI, Jun. 2005.
- [19] G. Kramer, "Models and theory for relay channels with receiver constraints," in *Proc. Annu. Allerton Conf. Communication, Control and Computing*, Monticello, IL, Sep./Oct. 2004, pp. 1312–1321.

## A New Bound for the Zero-Error Capacity Region of the Two-User Binary Adder Channel

Mikael Mattas and Patric R. J. Östergård

**Abstract**—A new uniquely decodable (UD) code pair for the two-user binary adder channel (BAC) is presented. This code pair leads to an improved bound for the zero-error capacity region of such a channel. The highest known rate for a UD code pair for the two-user BAC is thereby improved to  $(\log_2 240)/6 \approx 1.3178$ . It is also demonstrated that the problem of finding UD code pairs for the closely related binary XOR channel is in one-to-one correspondence with a certain construction of binary one-error-correcting codes.

**Index Terms**—Binary adder channel, Shannon capacity, uniquely decodable (UD) code, zero-error capacity.

### I. INTRODUCTION

Multiuser channels have in a short time become ubiquitous due to mobile phones and other contemporary communication systems. In the last few years, this has renewed interest in studying some of the most basic multiuser channels. Real-world communication systems have to deal with issues like synchronization and error correction, which one can ignore in a theoretical study of ideal channels. Tolhuizen [14] recently studied one such channel with great success, the binary multiplying channel. Another channel, the binary adder channel, is considered in this correspondence.

With the two-user *binary adder channel* (BAC), two senders transmit binary symbols in a synchronized way, and the output of the channel is the ternary value obtained by *real* addition of the symbols (if addition is carried out modulo 2, we have the *binary XOR channel*). This channel, which is called the *binary erasure multiple access channel* in [5], is depicted in Fig. 1.

Since a received 1 cannot be unambiguously decoded, information should be transmitted in the form of codewords of a prescribed length  $n$ . Let  $C$  and  $D$  denote the block codes of the two users. Then decoding can be carried out unambiguously exactly when

$$\mathbf{c}_1 + \mathbf{d}_1 \neq \mathbf{c}_2 + \mathbf{d}_2$$

for all distinct  $\mathbf{c}_1, \mathbf{c}_2 \in C$  and distinct  $\mathbf{d}_1, \mathbf{d}_2 \in D$ . Alternatively, one may write

$$\mathbf{c}_1 - \mathbf{c}_2 \neq \mathbf{d}_2 - \mathbf{d}_1.$$

It should be emphasized that *real* addition and subtraction is carried out in the preceding formulas. The speed—the *rate* in bits per transmission—at which information can be transmitted from the respective user is  $R_1 = (\log_2 |C|)/n$  and  $R_2 = (\log_2 |D|)/n$ ; see [5] for an introduction to these and other basic concepts in information theory. The achievable rate pairs  $(R_1, R_2)$  form the capacity region; in the literature two types of capacity regions are considered.

Manuscript received December 7, 2004; revised May 20, 2005. This work was supported by the Academy of Finland under Grants 100500, 107493, and 202315.

M. Mattas was with the Department of Electrical and Communications Engineering, Helsinki University of Technology, P.O. Box 3000, 02015 TKK, Finland. He is now with TietoEnator Oyj, P.O. Box 171, 00381 Helsinki, Finland.

P. R. J. Östergård is with the Department of Electrical and Communications Engineering, Helsinki University of Technology, P.O. Box 3000, 02015 TKK, Finland.

Communicated by R. W. Yeung, Associate Editor for Shannon Theory. Digital Object Identifier 10.1109/TIT.2005.853309