

# Rate Bounds for MIMO Relay Channels

Caleb K. Lo and Robert W. Heath, Jr.

Wireless Networking and Communications Group  
Department of Electrical and Computer Engineering  
The University of Texas at Austin

1 University Station C0803

Austin, TX 78712-0240

Phone: (512) 232-2014

Fax: (512) 471-6512

Email: {clo, rheath}@ece.utexas.edu

## Abstract

This paper considers the multi-input multi-output (MIMO) relay channel where multiple antennas are employed by each terminal. Compared to single-input single-output (SISO) relay channels, MIMO relay channels introduce additional degrees of freedom, making the design and analysis of optimal cooperative strategies more complex. In this paper, a partial cooperation strategy that combines transmit-side message splitting and block-Markov encoding is presented. Lower bounds on capacity that improve on a previously proposed non-cooperative lower bound are derived for Gaussian MIMO relay channels<sup>1</sup>.

Keywords - Relay channels, MIMO systems, superposition coding, dirty-paper coding.

## 1 Introduction

Mesh networks that support multihop communication form an integral part of future-generation wireless communications [1–4]. Relay channels are the fundamental building blocks of multihop mesh networks. From [5], a discrete memoryless relay channel is defined by  $(\mathcal{X}_1 \times \mathcal{X}_2, p(y, y_1 | x_1, x_2), \mathcal{Y} \times \mathcal{Y}_1)$ . Here,  $\mathcal{X}_1$ ,  $\mathcal{X}_2$ ,  $\mathcal{Y}_1$  and  $\mathcal{Y}$  are finite sets corresponding to the transmitter, the relay and the receiver as shown in Fig. 1.

---

<sup>1</sup>These lower bounds are special cases of [5, Theorem 7], but are of interest as they are applied to the MIMO relay channel and provide intuition about the structure of good coding strategies for MIMO relaying.

Relay channels were introduced in [7] and upper bounds on their capacity were derived in [8]. Full-duplex relay channels were first analyzed from an information-theoretic perspective in [5], where inner and outer bounds were derived and exact capacity expressions were obtained for special cases such as the physically degraded and Gaussian degraded relay channels. The information-theoretic analysis in [5] relied on cooperation between the transmitter and the relay induced by block-Markov encoding.

Achievable rates in relay channels can be further improved via multi-input multi-output (MIMO) technology [9–12]. It has been shown that the capacity of a MIMO channel can scale linearly as the minimum of the number of transmit and receive antennas [13]. This encouraging result has led to research on multiuser MIMO channels such as Gaussian multiple access (MAC) [14–17] and broadcast (BC) [18–21] channels. Although discrete memoryless relay channels were analyzed decades ago, MIMO relay channels have only recently been studied [23]. In [23], a Gaussian relay channel with multiple antennas at each terminal is considered. Upper and lower bounds on capacity are shown for both deterministic and Rayleigh fading channels. The lower bounds for the case of fixed channels in [23] arise from a non-cooperative transmit strategy.

Higher achievable rates than those yielded by the non-cooperative approach in [23] can be obtained by observing that MIMO relay channels inherently contain more degrees of freedom than single-input single-output (SISO) relay channels, where each terminal employs only a single antenna. We assume that the relay performs partial decode-and-forward operations, where the relay decodes a portion of the transmitter's codeword, encodes the decoded message using its own codebook, and sends the encoded message to the receiver. In a MIMO relay channel, the channel eigenmodes can be exploited to optimize the cooperative role of the relay. Thus, coding strategies such as transmit-side message splitting [5, 6] can increase the achievable rate for MIMO relay channels.

We present transmission strategies that rely on message splitting to support varying levels of cooperation between the transmitter and the relay in a MIMO relay channel. In this policy, the transmitter has two messages and chooses its codeword as a function of both of them; the relay, though, only has to decode one of these messages. The key intuition behind the application of message splitting to MIMO relaying is as follows: in a Gaussian SISO relay channel, message splitting does not have any apparent benefits. This is based on the fact that in a Gaussian SISO relay channel, the received signals at the relay and the receiver can be statistically ordered. On the other hand, the received signals in a Gaussian MIMO relay channel cannot be statistically ordered. In particular, the channel eigenmodes determine the optimal level of cooperation between the transmitter and the relay in a MIMO relay channel, which is measured by how the transmitter chooses its codeword as a function of both messages.

We stress that our message splitting strategies are special cases of the partial cooperation approach in [5, Theorem 7]. Since a direct application of the general coding strategy in [5, Theorem 7] to Gaussian MIMO relay channels would require a computationally intensive optimization over several auxiliary random variables, we consider simplified coding approaches and obtain closed-form achievable rate expressions.

We propose lower bounds on the capacity of the MIMO relay channel by utilizing transmit-side message splitting. In particular, we consider both superposition coding and precoding at the transmitter. For the case of precoding in a Gaussian MIMO relay channel, dirty-paper coding [22] is employed at the transmitter. Our proposed lower bounds obtained via a combination of transmit-side message splitting and block-Markov encoding improve on the lower bounds from [23] that are obtained by a non-cooperative transmit strategy that does not employ block-Markov encoding. The non-cooperative approach in [23] is actually a special case of our proposed strategies. We also perform a simple numerical analysis that illustrates how the achievable rate from our precoding approaches depends on the exact channel state and not just on the channel norms.

This paper is organized as follows: In Section II we describe the system model. Section III reviews the upper and lower bounds on capacity from [23] for the Gaussian MIMO relay channel. In Section IV, we present our message splitting strategies for Gaussian MIMO relay channels along with their associated achievable rates. Numerical results are given in Section V. We conclude the paper in Section VI.

The appendix contains rigorous derivations of some of the rate bounds originally presented in [24].

We use boldface notation for matrices and vectors; uppercase notation is used for matrices while lowercase notation is used for vectors.  $\mathbb{E}$  represents mathematical expectation.  $\text{Re}(x)$  denotes the real part of a complex number  $x$ . For a matrix  $\mathbf{A}$ ,  $\mathbf{A}^\dagger$ ,  $\text{tr}(\mathbf{A})$  and  $\det(\mathbf{A})$  denote the transpose conjugate, trace, and determinant, respectively of  $\mathbf{A}$  while  $\mathbf{A} \succeq 0$  means that  $\mathbf{A}$  is positive semi-definite. SNR represents signal-to-noise ratio.  $\mathbf{I}_K$  denotes the  $K \times K$  identity matrix. We use  $\mathcal{CN}(\mathbf{b}, \mathbf{C})$  to represent the circularly symmetric complex Gaussian distribution with mean  $\mathbf{b}$  and covariance matrix  $\mathbf{C}$ . For a set  $\mathcal{R}$ ,  $\|\mathcal{R}\|$  denotes the cardinality of  $\mathcal{R}$ .

## 2 System Model

Consider the Gaussian MIMO full-duplex relay channel illustrated in Fig. 2. Let  $\mathbf{x}_1$  and  $\mathbf{x}_2$  be the  $M_t \times 1$  and  $M_r \times 1$  transmitted signals from the transmitter and the relay. Let  $\mathbf{y}$  and  $\mathbf{y}_1$  be the  $N_t \times 1$  and  $N_r \times 1$  received signals at the receiver and the relay. Define  $\mathbf{H}_1$ ,  $\mathbf{H}_2$ , and  $\mathbf{H}_3$  as  $N_r \times M_t$ ,  $N_t \times M_t$  and  $N_t \times M_r$  channel gain matrices. Define  $\mathbf{z}$  and  $\mathbf{z}_1$  as independent  $N_t \times 1$  and  $N_r \times 1$  circularly-symmetric complex Gaussian noise vectors with distributions  $\mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_t})$  and  $\mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_r})$ .

We assume that the transmitter is subject to a power constraint  $\mathbb{E}(\mathbf{x}_1^\dagger \mathbf{x}_1) \leq M_t$  and that the relay is also subject to a power constraint  $\mathbb{E}(\mathbf{x}_2^\dagger \mathbf{x}_2) \leq M_r$ . We also assume that the relay has two sets of antennas, with one set for the receiver and one for the transmitter, so it operates in a full-duplex mode. The relay also cancels out interference from its transmitter array at its receiver array. In addition, we assume that all channel matrices are fixed and known at all three terminals and that  $\mathbf{z}$  and  $\mathbf{z}_1$  are uncorrelated with  $\mathbf{x}_1$  and  $\mathbf{x}_2$ . We do not consider fading channels in this paper.

We define parameters related to the SNR at the receiver and at the relay as  $\gamma_1 = SNR_1/M_t$ ,  $\gamma_2 = SNR_2/M_t$ , and  $\gamma_3 = SNR_3/M_r$  where  $SNR_1$  and  $SNR_2$  are the expected SNR values for  $\mathbf{x}_1$  after fading at each receive antenna at the relay and the receiver, and  $SNR_3$  is the expected SNR for  $\mathbf{x}_2$  after fading at each receive antenna at the receiver [11].

With these definitions, the received signals at the relay and at the receiver are

$$\begin{aligned} \mathbf{y}_1 &= \sqrt{\gamma_1} \mathbf{H}_1 \mathbf{x}_1 + \mathbf{z}_1 \\ \mathbf{y} &= \sqrt{\gamma_2} \mathbf{H}_2 \mathbf{x}_1 + \sqrt{\gamma_3} \mathbf{H}_3 \mathbf{x}_2 + \mathbf{z}. \end{aligned} \quad (1)$$

### 3 Background

It was shown in [5, Sec. III] that the capacity  $C$  of a general relay channel is upper-bounded as

$$C \leq \max_{p(x_1, x_2)} \min\{I(X_1; Y, Y_1 | X_2), I(X_1, X_2; Y)\} \quad (2)$$

where the first term in the minimization is the rate from the transmitter to the relay and the receiver and the second term is the rate from the transmitter and the relay to the receiver.

Now let  $\mathbf{x}_1$  and  $\mathbf{x}_2$  be random vectors with mean zero and covariance matrices  $\boldsymbol{\Sigma}_{ij} = \mathbb{E}(\mathbf{x}_i \mathbf{x}_j^\dagger)$ . The authors of [23] established the following capacity upper bound and lower bound for the case where the channel gains are fixed and known at each terminal.

**Lemma 3.1.** [23, Sec. III] *An upper bound on the capacity of the Gaussian MIMO relay channel is given by*

$$C^G \leq C_{upper}^G = \max_{0 \leq \rho \leq 1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22}} \min(C_1^G, C_2^G) \quad (3)$$

where  $tr(\boldsymbol{\Sigma}_{11}) \leq M_t$ ,  $tr(\boldsymbol{\Sigma}_{22}) \leq M_r$  and

$$\begin{aligned} C_1^G &\triangleq \log \left[ \det \left( \mathbf{I}_{M_t} + (1 - \rho^2) \begin{bmatrix} \sqrt{\gamma_1} \mathbf{H}_1 \\ \sqrt{\gamma_2} \mathbf{H}_2 \end{bmatrix} \boldsymbol{\Sigma}_{11} \begin{bmatrix} \sqrt{\gamma_1} \mathbf{H}_1 \\ \sqrt{\gamma_2} \mathbf{H}_2 \end{bmatrix}^\dagger \right) \right] \\ C_2^G &\triangleq \inf_{a > 0} \log [\det(\mathbf{I}_{N_t} + (\gamma_2 + \frac{\rho^2}{a} \sqrt{\gamma_2 \gamma_3}) \mathbf{H}_2 \boldsymbol{\Sigma}_{11} \mathbf{H}_2^\dagger + (\gamma_3 + a \sqrt{\gamma_2 \gamma_3}) \mathbf{H}_3 \boldsymbol{\Sigma}_{22} \mathbf{H}_3^\dagger)]. \end{aligned} \quad (4)$$

**Lemma 3.2.** [23, Sec. III] *A lower bound on the capacity of the Gaussian MIMO relay channel is given by*

$$C^G \geq C_{lower}^G = \max(C_d^G, \min(C_3^G, C_4^G)), \quad (5)$$

where

$$\begin{aligned} C_d^G &\triangleq \max_{\Sigma_{11}} \log[\det(\mathbf{I}_{N_t} + \gamma_2 \mathbf{H}_2 \Sigma_{11} \mathbf{H}_2^\dagger)] \\ C_3^G &\triangleq \max_{\Sigma_{11}} \log[\det(\mathbf{I}_{N_r} + \gamma_1 \mathbf{H}_1 \Sigma_{11} \mathbf{H}_1^\dagger)] \\ C_4^G &\triangleq \max_{\Sigma_{22}} \log[\det(\mathbf{I}_{N_t} + \gamma_3 \mathbf{H}_3 \Sigma_{22} \mathbf{H}_3^\dagger (\mathbf{I}_{N_t} + \gamma_2 \mathbf{H}_2 \Sigma_{11}^* \mathbf{H}_2^\dagger)^{-1})] \end{aligned} \quad (6)$$

with

$$\Sigma_{11}^* \triangleq \arg \max_{\Sigma_{11} \succeq 0} \log[\det(\mathbf{I}_{N_r} + \gamma_1 \mathbf{H}_1 \Sigma_{11} \mathbf{H}_1^\dagger)]. \quad (7)$$

Our objective is to use transmit-side message splitting to improve upon the bound in Lemma 3.2. We outline this strategy in the next section.

## 4 Transmit-Side Message Splitting

Next we describe the transmission strategy that is employed in this paper. We divide the transmit message into two components, denoted by the random variables  $w_u$  and  $w_v$ .  $w_u$  is the message that is decoded by the relay and is thus cooperatively sent by the transmitter-relay pair to the receiver.  $w_v$ , however, is intended to be decoded only by the receiver, and thus is a source of “interference” at the relay that is known a-priori at the transmitter.

We consider two classes of transmission strategies with this setup. The first is superposition coding, where codebooks for  $w_u$  and  $w_v$  are determined separately and then simply superposed (added to one another) at the transmitter. The second strategy is to utilize precoding at the transmit end, where intuitively the transmitter attempts to mitigate the interference caused by  $w_v$  to the desirable signal corresponding to  $w_u$  at the relay. For both strategies, the transmitter and the relay cooperate in block-Markov encoding of  $w_u$ .

Note that the receiver must determine both  $w_u$  and  $w_v$  to decode the transmit message. Thus, if  $R_u$  denotes the rate for the codebook corresponding to  $w_u$  and  $R_v$  that for  $w_v$ , the net achievable rate for both superposition coding and precoding is  $R = R_u + R_v$ . Assuming the receiver successively decodes  $w_u$  and  $w_v$ , the order in which they are decoded impacts their rates. In this paper, we use both decoding orders and choose the order that maximizes the overall rate.

Let  $\mathbf{u}$  and  $\mathbf{v}$  be auxiliary variables representing the contribution of  $w_u$  and  $w_v$ , respectively to  $\mathbf{x}_1$ . Define

$\Sigma_u$ ,  $\Sigma_v$  and  $\Sigma_{x_2}$  to be the covariance matrices of  $\mathbf{u}$ ,  $\mathbf{v}$  and  $\mathbf{x}_2$  respectively. Also, define

$$\mathbf{A} = \begin{bmatrix} \Sigma_u & \mathbb{E}(\mathbf{u}\mathbf{x}_2^\dagger) \\ \mathbb{E}(\mathbf{x}_2\mathbf{u}^\dagger) & \Sigma_{x_2} \end{bmatrix}$$

and  $\mathbf{B} = [\sqrt{\gamma_2}\mathbf{H}_2 \quad \sqrt{\gamma_3}\mathbf{H}_3]$ . In this case,  $\mathbb{E}(\mathbf{u}\mathbf{u}^\dagger) = \Sigma_u$ . In addition, define  $\mathcal{X}_1$ ,  $\mathcal{X}_2$ ,  $\mathcal{U}$  and  $\mathcal{V}$  as the finite alphabets for  $\mathbf{x}_1$ ,  $\mathbf{x}_2$ ,  $\mathbf{u}$  and  $\mathbf{v}$ , respectively.

#### 4.1 Superposition Coding

Consider the system illustrated in Fig. 3. Assume that the receiver attempts to decode  $w_u$  before decoding  $w_v$ . Let  $R_{sc,u}$  be the achievable rate for this case. By applying the partial cooperation strategy of [5, Theorem 7] to this case, it can be proved that

$$R_{sc,u} = \sup_{p(x_1, x_2, u, v)} (R_{sc,u,1} + R_{sc,u,2}) \quad (8)$$

where

$$\begin{aligned} R_{sc,u,1} &= \min(I(U; Y_1 | X_2), I(U, X_2; Y)) \\ R_{sc,u,2} &= I(V; Y | U, X_2) \end{aligned} \quad (9)$$

and the supremum is taken over all joint distributions

$$p(x_1, x_2, u, v) = p(x_2)p(u|x_2)p(v)p(x_1|u, v)$$

on  $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{U} \times \mathcal{V}$ . For the Gaussian MIMO relay channel, we employ Gaussian codebooks for  $\mathbf{u}$  and  $\mathbf{v}$  at the transmitter. We prove in Appendix A.1 that

$$I(U; Y_1 | X_2) = \log \left( \frac{\det \left( \mathbf{I}_{N_r} + \gamma_1 \mathbf{H}_1 \left( \Sigma_u - \mathbb{E}(\mathbf{u}\mathbf{x}_2^\dagger) \Sigma_{x_2}^{-1} \mathbb{E}(\mathbf{x}_2\mathbf{u}^\dagger) + \Sigma_v \right) \mathbf{H}_1^\dagger \right)}{\det \left( \mathbf{I}_{N_r} + \gamma_1 \mathbf{H}_1 \Sigma_v \mathbf{H}_1^\dagger \right)} \right), \quad (10)$$

which is the maximum signaling rate for  $w_u$  over the transmitter-to-relay link. We also have

$$I(U, X_2; Y) = \log \left( \frac{\det \left( \mathbf{I}_{N_t} + \gamma_2 \mathbf{H}_2 \Sigma_v \mathbf{H}_2^\dagger + \mathbf{B}\mathbf{A}\mathbf{B}^\dagger \right)}{\det \left( \mathbf{I}_{N_t} + \gamma_2 \mathbf{H}_2 \Sigma_v \mathbf{H}_2^\dagger \right)} \right), \quad (11)$$

representing the maximum signaling rate for  $w_u$  over the effective multiple-access channel from the transmitter and relay to the receiver, and

$$I(V; Y | U, X_2) = \log(\det(\mathbf{I}_{N_t} + \gamma_2 \mathbf{H}_2 \Sigma_v \mathbf{H}_2^\dagger)), \quad (12)$$

which is the maximum signaling rate for  $w_v$  over the transmitter-to-receiver link.

Now assume that the receiver attempts to decode  $w_v$  before decoding  $w_u$ . Let  $R_{sc,v}$  be the achievable rate for this case. By applying the partial cooperation strategy of [5, Theorem 7] to this case, it can be proved that

$$R_{sc,v} = \sup_{p(x_1, x_2, u, v)} (R_{sc,v,1} + R_{sc,v,2}) \quad (13)$$

where

$$\begin{aligned} R_{sc,v,1} &= \min(I(U; Y_1|X_2), I(U, X_2; Y|V)) \\ R_{sc,v,2} &= I(V; Y) \end{aligned} \quad (14)$$

and the supremum is taken over all joint distributions

$$p(x_1, x_2, u, v) = p(x_2)p(u|x_2)p(v)p(x_1|u, v)$$

on  $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{U} \times \mathcal{V}$ . In this case our choice of Gaussian codebooks for  $\mathbf{u}$  and  $\mathbf{v}$  in a Gaussian MIMO relay channel yields

$$I(V; Y) = \log \left( \frac{\det(\mathbf{I}_{N_t} + \gamma_2 \mathbf{H}_2 \boldsymbol{\Sigma}_v \mathbf{H}_2^\dagger + \mathbf{B} \mathbf{A} \mathbf{B}^\dagger)}{\det(\mathbf{I}_{N_t} + \mathbf{B} \mathbf{A} \mathbf{B}^\dagger)} \right), \quad (15)$$

which is analogous to the rate in (12) and

$$I(U, X_2; Y|V) = \log(\det(\mathbf{I}_{N_t} + \mathbf{B} \mathbf{A} \mathbf{B}^\dagger)), \quad (16)$$

which is analogous to the rate in (11) while  $I(U; Y_1|X_2)$  is the same as in (10).

The objective is to choose the decoding order that yields a higher overall rate. We now state the following intuitively obvious result, which will not require a formal proof.

**Proposition 4.1.** *Let  $R_{sc}$  be the maximum signaling rate for the Gaussian MIMO relay channel where the transmitter employs superposition coding. Then*

$$R_{sc} = \max(R_{sc,u}, R_{sc,v}) \geq C_{lower}^G \quad (17)$$

where  $C_{lower}^G$  is given in Lemma 3.2.

By setting  $\mathbf{v} = \mathbf{x}_1$  and  $\mathbf{u} = 0$ , we can achieve  $C_d^G$ . Also, by setting  $\mathbf{u} = \mathbf{x}_1$ ,  $\mathbf{v} = 0$  and having the relay employ a codebook of the same cardinality as that of the codebook at the transmitter, we can achieve at least  $\min(C_3^G, C_4^G)$ .

## 4.2 Precoding

Instead of superposition coding, consider a strategy where the transmitter uses precoding to mitigate the interference caused by  $w_v$  to the desired signal corresponding to  $w_u$  at the relay. Assume that the receiver attempts to decode  $w_u$  before decoding  $w_v$ . Let  $R_{pre,u}$  be the achievable rate for this case. It is proved in Appendix A.2 that

$$R_{pre,u} = \sup_{p(x_1, x_2, u, v)} (R_{pre,u,1} + R_{pre,u,2}) \quad (18)$$

where

$$\begin{aligned} R_{pre,u,1} &= \min(I(U; Y_1|X_2) - I(U; V|X_2), I(U, X_2; Y)) \\ R_{pre,u,2} &= I(V; Y|U, X_2) \end{aligned} \quad (19)$$

and the supremum is taken over all joint distributions

$$p(x_1, x_2, u, v) = p(v)p(x_2)p(u|v, x_2)p(x_1|u, v)$$

on  $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{U} \times \mathcal{V}$ . Note from the form of the joint distributions that  $\mathbf{u}$  and  $\mathbf{v}$  are correlated, which differs from the case of superposition coding. The transmitter selects  $\mathbf{u}$  as a function of the known interference  $\mathbf{v}$  on the transmitter-to-relay channel  $\mathbf{H}_1$ .

For the Gaussian MIMO relay channel, we employ Gaussian codebooks for  $\mathbf{u}$  and  $\mathbf{v}$ . In particular, we choose  $\mathbf{u} = \mathbf{G}\mathbf{v} + \mathbf{x}'_1$  and  $\mathbf{x}_1 = \mathbf{x}'_1 + \mathbf{v}$ , where  $\mathbf{x}'_1$  and  $\mathbf{v}$  are chosen to be independent. Thus we are employing dirty-paper coding at the transmitter and the objective is to choose  $\mathbf{G}$  to maximize  $I(U; Y_1|X_2) - I(U; V|X_2)$ . We define  $\Sigma_{x'_1|x_2}$  to be the covariance matrix of  $\mathbf{x}'_1$  given knowledge of  $\mathbf{x}_2$ . By following a procedure similar to that in [25, Appendix C], we have

$$I(U; Y_1|X_2) - I(U; V|X_2) = \log(\det(\mathbf{I}_{N_r} + \gamma_1 \mathbf{H}_1 \Sigma_{x'_1|x_2} \mathbf{H}_1^\dagger)), \quad (20)$$

which is analogous to the rate in (10);  $I(U, X_2; Y)$  and  $I(V; Y|U, X_2)$  are the same as in (11) and (12) respectively.

Now assume that the receiver attempts to decode  $w_v$  before decoding  $w_u$ . Let  $R_{pre,v}$  be the achievable rate for this case. It is proved in Appendix A.3 that

$$R_{pre,v} = \sup_{p(x_1, x_2, u, v)} (R_{pre,v,1} + R_{pre,v,2}) \quad (21)$$

where

$$\begin{aligned} R_{pre,v,1} &= \min(I(U; Y_1|X_2) - I(U; V|X_2), I(U, X_2; Y|V)) \\ R_{pre,v,2} &= I(V; Y) \end{aligned} \quad (22)$$

and the supremum is taken over all joint distributions

$$p(x_1, x_2, u, v) = p(x_2)p(u, v|x_2)p(x_1|u, v)$$

on  $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{U} \times \mathcal{V}$ . In this case our choice of dirty-paper coding at the transmitter in a Gaussian MIMO relay channel results in  $I(V; Y)$ ,  $I(U; Y_1|X_2) - I(U; V|X_2)$  and  $I(U, X_2; Y|V)$  being the same as in (15), (20), and (16) respectively.

The objective is to choose the decoding order that yields a higher overall rate. We now state and prove the following result.

**Proposition 4.2.** *Let  $R_{pre}$  be the maximum signaling rate for the Gaussian MIMO relay channel employing dirty-paper coding at the transmitter. Then*

$$R_{pre} = \max(R_{pre,u}, R_{pre,v}) \geq R_{sc} \quad (23)$$

where  $R_{sc}$  is given in Proposition 4.1.

*Proof.* We show that superposition coding is a special case of our precoding strategy. Without loss of generality, assume that  $w_u$  is decoded first at the receiver. Recall that

$$R_{pre,u} = \sup_{p(v)p(x_2)p(u|v,x_2)p(x_1|u,v)} (\min(I(U; Y_1|X_2) - I(U; V|X_2), I(U, X_2; Y)) + I(V; Y|U, X_2)). \quad (24)$$

By considering the case where  $\mathbf{u}$  and  $\mathbf{v}$  are independent random variables, we find that  $p(u|v, x_2) = p(u|x_2)$  and  $I(U; V|X_2) = 0$ . Thus, (24) reduces to

$$R_{sc,u} = \sup_{p(x_2)p(u|x_2)p(v)p(x_1|u,v)} (\min(I(U; Y_1|X_2), I(U, X_2; Y)) + I(V; Y|U, X_2)). \quad (25)$$

It immediately follows that  $R_{pre} \geq R_{sc}$ . □

## 5 Numerical Results

We employ a simple example to demonstrate how transmit-side message splitting outperforms the bounds in [23, Sec. III]. We choose  $\mathbf{H}_2 = [1 \ 0]$  and  $\mathbf{H}_3 = 1$ . We also choose  $\mathbf{H}_1 = [x \ y]$ , where  $x, y \in \mathbb{R}$ , and constrain  $\|\mathbf{H}_1\| = 10$ . By considering  $\mathbf{H}_1$  and  $\mathbf{H}_2$  as two-dimensional vectors, we can define an ‘‘angle’’  $\Theta(\mathbf{H}_1, \mathbf{H}_2)$  between them. We vary  $\Theta(\mathbf{H}_1, \mathbf{H}_2)$  over the range  $[0, \pi]$ , where  $\Theta(\mathbf{H}_1, \mathbf{H}_2)$  is expressed in radians. As  $\Theta(\mathbf{H}_1, \mathbf{H}_2) \rightarrow \pi/2$ , the gain between the second transmit antenna and the relay’s antenna, or  $y$ , increases.

Note that the norm constraint on  $\mathbf{H}_1$  causes the gain between the first transmit antenna and the relay's antenna, or  $x$ , to decrease.

We consider a system topology where the transmitter, the relay, and the receiver are equidistant; this is modeled by setting  $\gamma_1 = \gamma_2 = \gamma_3$  in (1). We observed that the lower bound from [23, Sec. III] is 1 bits/s/Hz for all values of  $\Theta(\mathbf{H}_1, \mathbf{H}_2)$ ; this results from our fixing  $\mathbf{H}_2$  at  $[1 \ 0]$ .

We need to solve the optimization problems (8), (13), (18) and (21) to obtain the achieved rates for our message splitting strategies. We employ numerical direct search methods such as the Nelder-Mead algorithm, differential evolution and simulated annealing to solve (8), (13), (18) and (21).

Fig. 4 shows the rates that are achieved by our message splitting strategies along with the upper and lower bounds from [23, Sec. III]. We see that the upper bound decreases as  $\Theta(\mathbf{H}_1, \mathbf{H}_2) \rightarrow \pi/2$  radians. Also, as  $\Theta(\mathbf{H}_1, \mathbf{H}_2) \rightarrow \pi/2$ , the transmitter uses more power on its second transmit antenna to exploit the rate benefits on the transmitter-to-relay link. This strategy, though, results in a loss of rate on the direct link since  $\mathbf{H}_2$  is fixed at  $[1 \ 0]$ . This leads to a monotonic decrease in the upper bound as  $\Theta(\mathbf{H}_1, \mathbf{H}_2) \rightarrow \pi/2$ .

We see that the achievable rates via superposition coding and dirty-paper coding always outperform the lower bound of 1 bits/s/Hz. Also, we see that the achievable rate from dirty-paper coding is never less than the achievable rate from superposition coding.

## 6 Conclusion

We derived new lower bounds on capacity for MIMO relay channels via transmit-side message splitting. Our proposed bounds improve upon the lower bounds that were introduced in [23]. In particular, our results show the benefits of employing the relay's assistance via superposition coding and precoding at the transmitter. Our results suggest that transmit-side message splitting should be an integral part of communication over MIMO relay channels, especially when the transmitter-to-relay link is strong relative to the transmitter-to-receiver and/or relay-to-receiver channels.

## 7 Acknowledgments

The authors would like to thank Sriram Vishwanath for numerous fruitful discussions that significantly improved the quality of the manuscript.

## A Proofs Of Rate Bounds

### A.1 Establishing (10)

We have  $I(U; Y_1|X_2) = h(Y_1|X_2) - h(Y_1|X_2, U)$ . Since the transmitter employs superposition coding, we have

$$\begin{aligned} \mathbf{y}_1 &= \sqrt{\gamma_1} \mathbf{H}_1 \mathbf{x}_1 + \mathbf{z}_1 \\ &= \sqrt{\gamma_1} \mathbf{H}_1 (\mathbf{u} + \mathbf{v}) + \mathbf{z}_1 \end{aligned} \quad (26)$$

and since  $\mathbf{u}$  and  $\mathbf{v}$  are independent given  $\mathbf{x}_2$ , and  $\mathbf{v}$  and  $\mathbf{x}_2$  are independent, we have

$$\begin{aligned} h(Y_1|X_2) &= h(\sqrt{\gamma_1} \mathbf{H}_1 (U + V) + Z_1|X_2) \\ &= \log((2\pi e)^{N_r} \det(\gamma_1 \mathbf{H}_1 (\boldsymbol{\Sigma}_{u|x_2} + \boldsymbol{\Sigma}_v) \mathbf{H}_1^\dagger + \mathbf{I}_{N_r})) \end{aligned} \quad (27)$$

and since  $\mathbf{z}_1$  is independent of  $\mathbf{u}$ ,  $\mathbf{v}$  and  $\mathbf{x}_2$  we have

$$\begin{aligned} h(Y_1|X_2, U) &= h(\sqrt{\gamma_1} \mathbf{H}_1 (U + V) + Z_1|X_2, U) \\ &= h(\sqrt{\gamma_1} \mathbf{H}_1 V + Z_1|X_2, U) \\ &= h(\sqrt{\gamma_1} \mathbf{H}_1 V + Z_1) \\ &= \log((2\pi e)^{N_r} \det(\gamma_1 \mathbf{H}_1 \boldsymbol{\Sigma}_v \mathbf{H}_1^\dagger + \mathbf{I}_{N_r})). \end{aligned} \quad (28)$$

Now we note that  $\log((2\pi e)^{M_t} \det(\boldsymbol{\Sigma}_{u|x_2})) = h(U|X_2) = h(U, X_2) - h(X_2) = \log((2\pi e)^{M_t+M_r} \det(A)) - \log((2\pi e)^{M_r} \det(\boldsymbol{\Sigma}_{x_2}))$  where

$$A = \begin{bmatrix} \boldsymbol{\Sigma}_u & \mathbb{E}(\mathbf{u}\mathbf{x}_2^\dagger) \\ \mathbb{E}(\mathbf{x}_2\mathbf{u}^\dagger) & \boldsymbol{\Sigma}_{x_2} \end{bmatrix}$$

so

$$\det(A) = \det(\boldsymbol{\Sigma}_{x_2}) \cdot \det(\boldsymbol{\Sigma}_u - \mathbb{E}(\mathbf{u}\mathbf{x}_2^\dagger) \boldsymbol{\Sigma}_{x_2}^{-1} \mathbb{E}(\mathbf{x}_2\mathbf{u}^\dagger))$$

and

$$\boldsymbol{\Sigma}_{u|x_2} = \boldsymbol{\Sigma}_u - \mathbb{E}(\mathbf{u}\mathbf{x}_2^\dagger) \boldsymbol{\Sigma}_{x_2}^{-1} \mathbb{E}(\mathbf{x}_2\mathbf{u}^\dagger).$$

Thus we have

$$h(Y_1|X_2) = \log((2\pi e)^{N_r} \det(\gamma_1 \mathbf{H}_1 (\boldsymbol{\Sigma}_u - \mathbb{E}(\mathbf{u}\mathbf{x}_2^\dagger) \boldsymbol{\Sigma}_{x_2}^{-1} \mathbb{E}(\mathbf{x}_2\mathbf{u}^\dagger) + \boldsymbol{\Sigma}_v) \mathbf{H}_1^\dagger + \mathbf{I}_{N_r})) \quad (29)$$

and finally we obtain

$$I(U; Y_1|X_2) = \log \left( \frac{\det \left( \mathbf{I}_{N_r} + \gamma_1 \mathbf{H}_1 \left( \boldsymbol{\Sigma}_u - \mathbb{E}(\mathbf{u}\mathbf{x}_2^\dagger) \boldsymbol{\Sigma}_{x_2}^{-1} \mathbb{E}(\mathbf{x}_2\mathbf{u}^\dagger) + \boldsymbol{\Sigma}_v \right) \mathbf{H}_1^\dagger \right)}{\det \left( \mathbf{I}_{N_r} + \gamma_1 \mathbf{H}_1 \boldsymbol{\Sigma}_v \mathbf{H}_1^\dagger \right)} \right). \quad (30)$$

## A.2 Achievability Proof of (18)

This proof relies on the concept of backward decoding, which was introduced in [26].

### A.2.1 Block Markov Encoding and Backward Decoding

Consider  $B+1$  blocks of transmission, each consisting of  $n$  symbols. A sequence of  $B$  messages,  $w_i = (w_{u,i}, w_{v,i}) \in \mathcal{W}$ ,  $i = 1, 2, \dots, B$ , each selected independently and uniformly over  $\mathcal{W}$  is to be sent over the channel in  $n(B+1)$  transmissions.

The senders use a triply-indexed set of codewords:

$$\begin{aligned} \mathcal{C} = \{ & (v^n(w_v), u^n(k, w_{2u}), x_2^n(w_{2u})) : w_v \in \{\phi, 1, 2, \dots, 2^{nR_v}\}, \\ & k \in \{1, 2, \dots, 2^{n(I(U; Y_1 | X_2) - \delta(\epsilon))}\}, w_{2u} \in \{\phi, 1, 2, \dots, 2^{nR_u}\}\}. \end{aligned} \quad (31)$$

$w_{2u}$  is sent cooperatively by both senders in block  $i$  to help the receiver decode the previous message  $w_{u,i-1}$ . See Table 1 for details.

Backward decoding is employed at the receiver to decode  $w_{u,i}$  and  $w_{v,i}$ . Thus, after block  $B+1$ ,  $\mathbf{y}(B+1)$  is used to decode  $w_{u,B}$  and  $w_{v,B}$ . Then,  $\mathbf{y}(B)$  and  $w_{u,B}$  are used to decode  $w_{u,B-1}$  and  $w_{v,B-1}$ . Next,  $\mathbf{y}(B-1)$  and  $w_{u,B-1}$  are used to decode  $w_{u,B-2}$  and  $w_{v,B-2}$ . The process continues until  $\mathbf{y}(2)$  and  $w_{u,2}$  are used to decode  $w_{u,1}$  and  $w_{v,1}$ .

### A.2.2 Generation of Random Code

Fix  $p(v)p(x_2)p(u|v, x_2)p(x_1|u, v)$ . Generate at random  $2^{nR_v}$  i.i.d.  $v^n$  sequences according to  $\sim \prod_{i=1}^n p(v_i)$ , and index them as  $v^n(w_v)$ ,  $w_v \in \{1, 2, \dots, 2^{nR_v}\}$ . Generate at random  $2^{nR_u}$  i.i.d.  $x_2^n$  sequences according to  $\sim \prod_{i=1}^n p(x_{2i})$ , and index them as  $x_2^n(w_{2u})$ ,  $w_{2u} \in \{1, 2, \dots, 2^{nR_u}\}$ . For each  $x_2^n(w_{2u})$ , generate  $2^{n(I(U; Y_1 | X_2) - \delta(\epsilon))}$  conditionally independent  $u^n \in T_\epsilon^{(n)}(u)$  sequences according to  $p(u|v, x_2)$ , and partition them into  $2^{nR_u}$  equal-sized bins for each  $x_2^n(w_{2u})$ . This defines the random codebook  $\mathcal{C} = \{(v^n(w_v), u^n(k, w_{2u}), x_2^n(w_{2u}))\}$ . Finally, generate the codeword  $x_1^n$  via  $p(x_1^n|u^n, v^n)$ .

The bin partitioning of the  $u^n$  sequences implicitly defines a function  $\mathcal{F}$  where  $\mathcal{F} : u^n(k, w_{2u}) \rightarrow w_u$ . Here,  $k \in \{1, 2, \dots, 2^{n(I(U; Y_1 | X_2) - \delta(\epsilon))}\}$ ,  $w_{2u} \in \{\phi, 1, 2, \dots, 2^{nR_u}\}$ , and  $w_u \in \{\phi, 1, 2, \dots, 2^{nR_u}\}$ . For example,  $\mathcal{F}(u^n(1, w_{2u})) = \mathcal{F}(u^n(2, w_{2u})) = \dots = \mathcal{F}(u^n(2^{n(I(U; Y_1 | X_2) - R_u - \delta(\epsilon))}, w_{2u})) = w_u = 1$ . We see that  $\mathcal{F}$  maps sequences  $u^n(k, w_{2u})$  into their corresponding bin (and hence, message) indices  $w_u$ . Since there is a one-to-one mapping between a sequence  $u^n(k, w_{2u})$  and its bin  $w_u$ , we can also write  $\mathcal{F}(u^n(k, w_{2u}))$  as  $\mathcal{F}(k, w_{2u})$ .

### A.2.3 Encoding

Let  $w_{u,i} \in \{1, 2, \dots, 2^{nR_u}\}$  and  $w_{v,i} \in \{1, 2, \dots, 2^{nR_v}\}$  comprise the new message to be sent in block  $i$ . Then, select any  $u^n(k, w_{u,i-1})$  in bin  $w_{u,i}$  such that  $(u^n(k, w_{u,i-1}), v^n(w_{v,i})) \in T_\epsilon^{(n)}$ . Use the selected  $u^n$  along with  $v^n(w_{v,i})$  to generate  $x_1^n$  via  $p(x_1^n | u^n, v^n)$  and transmit this  $x_1^n$ .

Here,  $P(x_1^n \in T_\epsilon^{(n)}(x_1^n | u^n(k, w_{u,i-1}), v^n(w_{v,i}))) > 1 - \epsilon$ .

Assuming that the relay estimated  $\hat{w}_{u,i-1}$  for  $w_{u,i-1}$  in block  $i - 1$ , then the relay sends  $x_2^n(\hat{w}_{u,i-1})$  in block  $i$ .

### A.2.4 Decoding and Error Analysis

Note that for  $w_{u,i}$  and  $w_{v,i}$ , we perform backward decoding at the receiver, though we perform block-by-block decoding at the relay. The following analysis is for the case where the receiver attempts to decode  $w_u$  before decoding  $w_v$ .

Here, we proceed through three decoding steps. We again employ the concept of strong typicality. Define the following error events:

- $E_{0i}$  as the event that  $(u^n(k_i, w_{2u,i}), x_2^n(w_{2u,i}), y_1^n(i), y^n(i)) \notin T_\epsilon^{(n)}$ , where  $y_1^n(i)$  and  $y^n(i)$  are the observations by the relay and receiver, respectively in block  $i$ .
- $E_{mi}$  as the event that there is an error in block  $i$  at decoding step  $m$  for  $m = 1, 2, 3$ .

Thus, the overall probability of error  $P_e^{(n)} = P(\bigcup_{m=0}^3 E_{mi}) \leq \sum_{m=0}^3 P(E_{mi})$ . We first note that for  $n$  sufficiently large,  $P(E_{0i}) < \epsilon$  by the asymptotic equipartition property (AEP). Now we bound  $P(E_{mi})$  for  $m = 1, 2, 3$  as follows.

Decoding step 1: The relay observes  $y_1^n(i)$  and looks for the unique  $u^n(k, w_{u,i-1})$  such that  $(u^n(k, w_{u,i-1}), y_1^n(i), x_2^n(w_{u,i-1})) \in T_\epsilon^{(n)}$ . If this unique  $u^n(k, w_{u,i-1})$  can be found, then the relay declares  $\mathcal{F}(u^n(k, w_{u,i-1}))$  as  $\hat{w}_{u,i}$ . Here,  $E_{1i}$  is the event where  $(u^n, v^n(w_{v,i}), x_2^n(w_{u,i-1})) \notin T_\epsilon^{(n)} \forall u^n$  such that  $\mathcal{F}(u^n) = w_{u,i}$ ; we have

$$\begin{aligned}
 P(E_{1i}) &= P((\mathbf{u}, v^n(w_{v,i}), \mathbf{x}_2(w_{u,i-1})) \notin T_\epsilon^{(n)}) \\
 &= \sum_{v^n} p(v^n) P((\mathbf{u}, v^n, \mathbf{x}_2(w_{u,i-1})) \notin T_\epsilon^{(n)}) \\
 &= \sum_{v^n} p(v^n) (1 - P((\mathbf{u}, v^n, \mathbf{x}_2(w_{u,i-1})) \in T_\epsilon^{(n)}))^{2^{n(I(U; Y_1 | X_2) - R_u - \delta(\epsilon))}} \\
 &\leq e^{-2^{n(I(U; Y_1 | X_2) - R_u - \delta(\epsilon))}} 2^{-n(I(U; V | X_2) + \delta_1(\epsilon))}
 \end{aligned} \tag{32}$$

which is arbitrarily small for  $n$  sufficiently large if

$$R_u < I(U; Y_1 | X_2) - I(U; V | X_2) - \delta_1(\epsilon) - \delta(\epsilon). \quad (33)$$

We note that (32) follows from the following two facts:

- i)  $2^{-n(I(U; V | X_2) + \delta_1(\epsilon))} \leq P((\mathbf{u}, v^n, \mathbf{x}_2(w_{u,i-1})) \in T_\epsilon^{(n)})$  and
- ii)  $(1-x)^k \leq e^{-kx}$  for  $0 \leq x \leq 1$  and  $k \geq 1$ .

Thus,  $\hat{w}_{u,i} = w_{u,i}$  with  $P(E_{1i})$  arbitrarily small if  $n$  is sufficiently large and if

$$R_u < I(U; Y_1 | X_2) - I(U; V | X_2). \quad (34)$$

Decoding step 2: Backward decoding is employed to estimate  $w_{u,i}$  at the receiver. Assume that the receiver has estimated  $\tilde{w}_{u,i+1}$  for  $w_{u,i+1}$ . Now, the receiver looks for a unique  $w_{2u}$  such that  $(u^n(k_i, w_{2u}), x_2^n(w_{2u}), y^n(i)) \in T_\epsilon^{(n)}$ , where  $\mathcal{F}(k_i, w_{2u}) = \tilde{w}_{u,i+1}$ . It then declares  $\tilde{w}_u = w_{2u}$  if this unique  $w_{2u}$  exists. Here,  $E_{2i}$  is the event that  $\exists \tilde{w}_u \neq w_{u,i}$  such that  $(u^n(k_i, \tilde{w}_u), x_2^n(\tilde{w}_u), y^n(i)) \in T_\epsilon^{(n)}$ , where  $\mathcal{F}(k_i, \tilde{w}_u) = \tilde{w}_{u,i+1}$ . Now, for  $\tilde{w}_u \neq w_{u,i}$ ,

$$\begin{aligned} P(E_{2i} | \tilde{w}_u) &= P((\mathbf{u}, \mathbf{x}_2(\tilde{w}_u), \mathbf{y}(i)) \in T_\epsilon^{(n)}) \\ &= \sum_{(u^n(k_i, \tilde{w}_u), x_2^n(\tilde{w}_u), y^n(i)) \in T_\epsilon^{(n)}(U, X_2, Y), \tilde{w}_u \neq w_{u,i}, \mathcal{F}(k_i, \tilde{w}_u) = \tilde{w}_{u,i+1}} p(u^n(k_i, \tilde{w}_u), x_2^n(\tilde{w}_u), y^n(i)) \end{aligned} \quad (35)$$

$$\begin{aligned} &= \sum_{(u^n(k_i, \tilde{w}_u), x_2^n(\tilde{w}_u), y^n(i)) \in T_\epsilon^{(n)}(U, X_2, Y), \tilde{w}_u \neq w_{u,i}, \mathcal{F}(k_i, \tilde{w}_u) = \tilde{w}_{u,i+1}} p(u^n(k_i, \tilde{w}_u), x_2^n(\tilde{w}_u)) p(y^n(i)) \end{aligned} \quad (36)$$

$$\begin{aligned} &\leq \left| T_\epsilon^{(n)}(U, X_2, Y) \right| 2^{-n(H(U, X_2) - \delta(\epsilon))} 2^{-n(H(Y) - \delta_1(\epsilon))} \\ &\leq 2^{-n(H(U, X_2) + H(Y) - H(U, X_2, Y) - \delta(\epsilon) - \delta_1(\epsilon) - \delta_2(\epsilon))} \\ &= 2^{-n(I(U, X_2; Y) - \delta(\epsilon) - \delta_1(\epsilon) - \delta_2(\epsilon))} \end{aligned}$$

where (36) follows from the fact that  $\mathbf{y}(i)$  and  $(\mathbf{u}(k_i, \tilde{w}_u), \mathbf{x}_2(\tilde{w}_u))$  are independent for  $\tilde{w}_u \neq w_{u,i}$ . Also, we have  $\delta(\epsilon) \rightarrow 0$ ,  $\delta_1(\epsilon) \rightarrow 0$ , and  $\delta_2(\epsilon) \rightarrow 0$  as  $\epsilon \rightarrow 0$ . Thus, we have

$$P(E_{2i}) \leq 2^{nR_u} \cdot 2^{-n(I(U, X_2; Y) - \delta(\epsilon) - \delta_1(\epsilon) - \delta_2(\epsilon))} \quad (37)$$

and so  $\tilde{w}_u = w_{u,i}$  with  $P(E_{2i})$  arbitrarily small if  $n$  is sufficiently large and if

$$R_u < I(U, X_2; Y). \quad (38)$$

Now, we combine (34) and (38) to obtain

$$R_u < \min((I(U; Y_1|X_2) - I(U; V|X_2)), I(U, X_2; Y)). \quad (39)$$

Decoding step 3: Backward decoding is also employed to estimate  $w_{v,i}$  at the receiver. Assume that the receiver has estimated  $\tilde{w}_{u,i+1}$  for  $w_{u,i+1}$ . Recall that the receiver has estimated  $\tilde{w}_{u,i}$  for  $w_{u,i}$  in decoding step 2. Now, the receiver looks for a unique  $w_v$  such that  $(u^n(k_i, \tilde{w}_{u,i}), x_2^n(\tilde{w}_{u,i}), y^n(i), v^n(w_v)) \in T_\epsilon^{(n)}$ , where  $\mathcal{F}(k_i, \tilde{w}_{u,i}) = \tilde{w}_{u,i+1}$ . It then declares  $\tilde{w}_v = w_v$  if this unique  $w_v$  exists. Here,  $E_{3i}$  is the event that  $\exists \tilde{w}_v \neq w_{v,i}$  such that  $(u^n(k_i, \tilde{w}_{u,i}), x_2^n(\tilde{w}_{u,i}), y^n(i), v^n(\tilde{w}_v)) \in T_\epsilon^{(n)}$ , where  $\mathcal{F}(k_i, \tilde{w}_{u,i}) = \tilde{w}_{u,i+1}$ . Now, for  $\tilde{w}_v \neq w_{v,i}$ ,

$$\begin{aligned} P(E_{3i}|\tilde{w}_v) &= P((\mathbf{u}, \mathbf{x}_2(\tilde{w}_{u,i}), \mathbf{y}(i), \mathbf{v}(\tilde{w}_v)) \in T_\epsilon^{(n)}) \\ &= \sum_{(v^n(\tilde{w}_v), y^n(i)) \in T_\epsilon^{(n)}(V, Y|u^n, x_2^n), \tilde{w}_v \neq w_{v,i}, \mathcal{F}(k_i, \tilde{w}_{u,i}) = \tilde{w}_{u,i+1}} p(v^n(\tilde{w}_v), y^n(i)|u^n(k_i, \tilde{w}_{u,i}), x_2^n(\tilde{w}_{u,i})) \\ &= \sum_{(v^n(\tilde{w}_v), y^n(i)) \in T_\epsilon^{(n)}(V, Y|u^n, x_2^n), \tilde{w}_v \neq w_{v,i}, \mathcal{F}(k_i, \tilde{w}_{u,i}) = \tilde{w}_{u,i+1}} p(v^n(\tilde{w}_v)|u^n(k_i, \tilde{w}_{u,i}), x_2^n(\tilde{w}_{u,i})) p(y^n(i)|u^n(k_i, \tilde{w}_{u,i}), x_2^n(\tilde{w}_{u,i})) \\ &\leq \left| T_\epsilon^{(n)}(V, Y|u^n, x_2^n) \right| 2^{-n(H(V|U, X_2) - \delta(\epsilon))} 2^{-n(H(Y|U, X_2) - \delta_1(\epsilon))} \\ &\leq 2^{-n(H(V|U, X_2) + H(Y|U, X_2) - H(V, Y|U, X_2) - \delta(\epsilon) - \delta_1(\epsilon) - \delta_2(\epsilon))} \\ &= 2^{-n(I(V; Y|U, X_2) - \delta(\epsilon) - \delta_1(\epsilon) - \delta_2(\epsilon))} \end{aligned} \quad (40)$$

where (40) follows from the fact that  $\mathbf{y}(i)$  and  $\mathbf{v}(\tilde{w}_v)$  are independent for  $\tilde{w}_v \neq w_{v,i}$ . Also, we have  $\delta(\epsilon) \rightarrow 0$ ,  $\delta_1(\epsilon) \rightarrow 0$ , and  $\delta_2(\epsilon) \rightarrow 0$  as  $\epsilon \rightarrow 0$ . Thus, we have

$$P(E_{3i}) \leq 2^{nR_v} \cdot 2^{-n(I(V; Y|U, X_2) - \delta(\epsilon) - \delta_1(\epsilon) - \delta_2(\epsilon))} \quad (41)$$

and so  $\tilde{w}_v = w_{v,i}$  with  $P(E_{3i})$  arbitrarily small if  $n$  is sufficiently large and if

$$R_v < I(V; Y|U, X_2). \quad (42)$$

### A.3 Achievability Proof of (21)

This proof also relies on the concept of backward decoding. Apply the code generation and encoding procedures from Section A.2. Note that in this case, backward decoding is employed at the receiver to decode  $w_{u,i}$ , not both  $w_{u,i}$  and  $w_{v,i}$ . Thus, after block  $B+1$ ,  $\mathbf{y}(B+1)$  is used to decode  $w_{u,B}$ . Then,  $\mathbf{y}(B)$  and  $w_{u,B}$  are used to decode  $w_{u,B-1}$ . Next,  $\mathbf{y}(B-1)$  and  $w_{u,B-1}$  are used to decode  $w_{u,B-2}$ . The process continues

until  $\mathbf{y}(2)$  and  $w_{u,2}$  are used to decode  $w_{u,1}$ . The receiver can use block-by-block decoding to decode  $w_{v,i}$ ; thus,  $w_{v,i}$  can be decoded in block  $i$  after  $\mathbf{y}(i)$  is received, where  $i = 1, 2, \dots, B$ .

### A.3.1 Decoding and Error Analysis

Note that for  $w_{u,i}$ , we perform backward decoding at the receiver, though we still perform block-by-block decoding at the relay. We also perform block-by-block decoding at the receiver for  $w_{v,i}$ .

Once again, we proceed through three decoding steps and employ the concept of strong typicality. Define the following error events:

- $E_{0i}$  as the event that  $(u^n(k_i, w_{2u,i}), x_2^n(w_{2u,i}), y_1^n(i), y^n(i)) \notin T_\epsilon^{(n)}$ , where  $y_1^n(i)$  and  $y^n(i)$  are the observations by the relay and receiver, respectively in block  $i$ .
- $E_{mi}$  as the event that there is an error in block  $i$  at decoding step  $m$  for  $m = 1, 2, 3$ .

Thus, the overall probability of error  $P_e^{(n)} = P(\bigcup_{m=0}^3 E_{mi}) \leq \sum_{m=0}^3 P(E_{mi})$ . We first note that for  $n$  sufficiently large,  $P(E_{0i}) < \epsilon$  by the asymptotic equipartition property (AEP). Now we bound  $P(E_{mi})$  for  $m = 1, 2, 3$  as follows.

Decoding step 1: Upon observing  $y^n(i)$ , the receiver declares that  $\tilde{w}_{v,i} = \tilde{w}_v$  is sent if it is the unique index such that  $(v^n(\tilde{w}_v), y^n(i)) \in T_\epsilon^{(n)}$ . Here,  $E_{1i}$  is the event that  $\exists \tilde{w}_v \neq w_{v,i}$  such that  $(v^n(\tilde{w}_v), y^n(i)) \in T_\epsilon^{(n)}$ . Now, for  $\tilde{w}_v \neq w_{v,i}$ ,

$$\begin{aligned}
 P(E_{1i}|\tilde{w}_v) &= P((\mathbf{v}(\tilde{w}_v), \mathbf{y}(i)) \in T_\epsilon^{(n)}) \\
 &= \sum_{(v^n(\tilde{w}_v), y^n(i)) \in T_\epsilon^{(n)}(V, Y), \tilde{w}_v \neq w_{v,i}} p(v^n(\tilde{w}_v), y^n(i)) \\
 &= \sum_{(v^n(\tilde{w}_v), y^n(i)) \in T_\epsilon^{(n)}(V, Y), \tilde{w}_v \neq w_{v,i}} p(v^n(\tilde{w}_v))p(y^n(i)) \\
 &\leq |T_\epsilon^{(n)}(V, Y)| 2^{-n(H(V)-\delta(\epsilon))} 2^{-n(H(Y)-\delta_1(\epsilon))} \\
 &\leq 2^{-n(H(V)+H(Y)-H(V,Y)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \\
 &= 2^{-n(I(V;Y)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))}
 \end{aligned} \tag{43}$$

where (43) follows from the fact that  $\mathbf{y}(i)$  and  $\mathbf{v}(\tilde{w}_v)$  are independent for  $\tilde{w}_v \neq w_{v,i}$ . Also, we have  $\delta(\epsilon) \rightarrow 0$ ,  $\delta_1(\epsilon) \rightarrow 0$ , and  $\delta_2(\epsilon) \rightarrow 0$  as  $\epsilon \rightarrow 0$ . Thus, we have

$$P(E_{1i}) \leq 2^{nR_v} \cdot 2^{-n(I(V;Y)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \tag{44}$$

and so  $\tilde{w}_{v,i} = w_{v,i}$  with  $P(E_{1i})$  arbitrarily small if  $n$  is sufficiently large and if

$$R_v < I(V; Y). \quad (45)$$

Decoding step 2: The analysis for this decoding step is similar to the analysis for decoding step 1 in Section A.2. Thus we have

$$R_u < I(U; Y_1 | X_2) - I(U; V | X_2). \quad (46)$$

Decoding step 3: Backward decoding is employed to estimate  $w_{u,i}$  at the receiver. Assume that the receiver has estimated  $\tilde{w}_{u,i+1}$  for  $w_{u,i+1}$ . Recall that the receiver has estimated  $\tilde{w}_{v,i}$  for  $w_{v,i}$  in decoding step 1. Now, the receiver looks for a unique  $w_{2u}$  such that  $(u^n(k_i, w_{2u}), x_2^n(w_{2u}), y^n(i), v^n(\tilde{w}_{v,i})) \in T_\epsilon^{(n)}$ , where  $\mathcal{F}(k_i, w_{2u}) = \tilde{w}_{u,i+1}$ . It then declares  $\tilde{w}_u = w_{2u}$  if this unique  $w_{2u}$  exists. Here,  $E_{3i}$  is the event that  $\exists \tilde{w}_u \neq w_{u,i}$  such that  $(u^n(k_i, \tilde{w}_u), x_2^n(\tilde{w}_u), y^n(i), v^n(\tilde{w}_{v,i})) \in T_\epsilon^{(n)}$ , where  $\mathcal{F}(k_i, \tilde{w}_u) = \tilde{w}_{u,i+1}$ . Now, for  $\tilde{w}_u \neq w_{u,i}$ ,

$$\begin{aligned} P(E_{3i} | \tilde{w}_u) &= P((\mathbf{u}, \mathbf{x}_2(\tilde{w}_u), \mathbf{y}(i), \mathbf{v}(\tilde{w}_{v,i})) \in T_\epsilon^{(n)}) \\ &= \sum_{(u^n(k_i, \tilde{w}_u), x_2^n(\tilde{w}_u), y^n(i)) \in T_\epsilon^{(n)}(U, X_2, Y | v^n), \tilde{w}_u \neq w_{u,i}, \mathcal{F}(k_i, \tilde{w}_u) = \tilde{w}_{u,i+1}} p(u^n(k_i, \tilde{w}_u), x_2^n(\tilde{w}_u), y^n(i) | v^n(\tilde{w}_{v,i})) \\ &= \sum_{(u^n(k_i, \tilde{w}_u), x_2^n(\tilde{w}_u), y^n(i)) \in T_\epsilon^{(n)}(U, X_2, Y | v^n), \tilde{w}_u \neq w_{u,i}, \mathcal{F}(k_i, \tilde{w}_u) = \tilde{w}_{u,i+1}} p(u^n(k_i, \tilde{w}_u), x_2^n(\tilde{w}_u) | v^n(\tilde{w}_{v,i})) p(y^n(i) | v^n(\tilde{w}_{v,i})) \\ &\leq \left| T_\epsilon^{(n)}(U, X_2, Y | v^n) \right| 2^{-n(H(U, X_2 | V) - \delta(\epsilon))} 2^{-n(H(Y | V) - \delta_1(\epsilon))} \\ &\leq 2^{-n(H(U, X_2 | V) + H(Y | V) - H(U, X_2, Y | V) - \delta(\epsilon) - \delta_1(\epsilon) - \delta_2(\epsilon))} \\ &= 2^{-n(I(U, X_2; Y | V) - \delta(\epsilon) - \delta_1(\epsilon) - \delta_2(\epsilon))} \end{aligned} \quad (47)$$

where (47) follows from the fact that  $\mathbf{y}(i)$  and  $(\mathbf{u}(k_i, \tilde{w}_u), \mathbf{x}_2(\tilde{w}_u))$  are independent for  $\tilde{w}_u \neq w_{u,i}$ . Also, we have  $\delta(\epsilon) \rightarrow 0$ ,  $\delta_1(\epsilon) \rightarrow 0$ , and  $\delta_2(\epsilon) \rightarrow 0$  as  $\epsilon \rightarrow 0$ . Thus, we have

$$P(E_{3i}) \leq 2^{nR_u} \cdot 2^{-n(I(U, X_2; Y | V) - \delta(\epsilon) - \delta_1(\epsilon) - \delta_2(\epsilon))} \quad (48)$$

and so  $\tilde{w}_u = w_{u,i}$  with  $P(E_{3i})$  arbitrarily small if  $n$  is sufficiently large and if

$$R_u < I(U, X_2; Y | V). \quad (49)$$

Now, we combine (46) and (49) to obtain

$$R_u < \min((I(U; Y_1 | X_2) - I(U; V | X_2)), I(U, X_2; Y | V)). \quad (50)$$

## References

- [1] R. Bruno, M. Conti and E. Gregori. Mesh networks: commodity multihop ad hoc networks. *IEEE Commun. Mag.*, 43(3):123–131, March 2005.
- [2] R. Pabst et al. Relay-based deployment concepts for wireless and mobile broadband radio. *IEEE Commun. Mag.*, 42(9):80–89, September 2004.
- [3] H. Viswanathan and S. Mukherjee. Throughput-range trade of wireless mesh backhaul networks. *IEEE J. Select. Areas Commun.*, 24(3):593–602, March 2006.
- [4] J. Jun and M.L. Sichitiu. The nominal capacity of wireless mesh networks. *IEEE Wireless Commun.*, 10(5):8–14, October 2003.
- [5] T. Cover and A. El Gamal. Capacity theorems for the relay channel. *IEEE Trans. Inform. Theory*, 25(5):572–584, September 1979.
- [6] A. El Gamal and M. Aref. The capacity of the semideterministic relay channel. *IEEE Trans. Inform. Theory*, 28(3):536, May 1982.
- [7] E.C. van der Meulen. Three-terminal communication channels. *Adv. Appl. Prob.*, 3:120–154, 1971.
- [8] H. Sato. Information transmission through a channel with relay. *Tech. Rep. B76-7*, The Aloha System, University of Hawaii, Honolulu, March 1976.
- [9] G.J. Foschini. Layered space-time architecture for wireless communication in fading environments when using multi-element antennas. *Bell Labs Tech. J.*, 1(2):41–59, Autumn, 1996.
- [10] J. Winters. On the capacity of radio communication systems with diversity in a Rayleigh fading environment. *IEEE J. Select. Areas Commun.*, 5(5):871–878, June 1987.
- [11] T. Marzetta and B. Hochwald. Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading. *IEEE Trans. Inform. Theory*, 45(1):139–157, January 1999.
- [12] D. Gesbert et al. From theory to practice: an overview of MIMO space-time coded wireless systems. *IEEE J. Select. Areas Commun.*, 21(3):281–302, April 2003.
- [13] E. Telatar. Capacity of multi-antenna Gaussian channels. *Eur. Trans. Telecommun. ETT*, 10(6):585–596, November 1999.

- [14] R.S. Cheng and S. Verdú. Gaussian multiaccess channels with ISI: capacity region and multiuser water-filling. *IEEE Trans. Inform. Theory*, 39(3):773–785, May 1993.
- [15] A.D. Wyner. Shannon-theoretic approach to a Gaussian cellular multiple-access channel. *IEEE Trans. Inform. Theory*, 40(6):1713–1727, November 1994.
- [16] P. Viswanath, D.N.C. Tse and V. Anantharam. Asymptotically optimal water-filling in vector multiple-access channels. *IEEE Trans. Inform. Theory*, 47(1):241–267, January 2001.
- [17] W. Yu, W. Rhee, S. Boyd, and J.M. Cioffi. Iterative water-filling for Gaussian vector multiple access channels. *IEEE Trans. Inform. Theory*, 50(1):145–152, January 2004.
- [18] G. Caire and S. Shamai. On the achievable throughput of a multiantenna Gaussian broadcast channel. *IEEE Trans. Inform. Theory*, 49(7):1691–1706, July 2003.
- [19] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz). The capacity region of the Gaussian MIMO broadcast channel. *IEEE Trans. Inform. Theory*, 52(9):3936–3964, September 2006.
- [20] N. Jindal, S. Vishwanath and A. Goldsmith. On the duality of Gaussian multiple-access and broadcast channels. *IEEE Trans. Inform. Theory*, 50(5):768–783, May 2004.
- [21] M. Mohseni and J.M. Cioffi. A proof of the converse for the capacity of Gaussian MIMO broadcast channels. In *Proc. of the IEEE Intl. Symp. on Inform. Theory*, 1:881–885, Seattle, WA, July 2006.
- [22] M. Costa. Writing on dirty paper. *IEEE Trans. Inform. Theory*, 29(3):439–441, May 1983.
- [23] B. Wang, J. Zhang, and A. Host-Madsen. On the capacity of MIMO relay channels. *IEEE Trans. Inform. Theory*, 51(1):29–43, January 2005.
- [24] C.K. Lo, S. Vishwanath and R.W. Heath, Jr. Rate bounds for MIMO relay channels using precoding. In *Proc. of the IEEE Global Telecommun. Conf.*, 3:1172–1176, St. Louis, MO, November 2005.
- [25] W. Yu. Competition and cooperation in multi-user communication environments. Ph.D. dissertation, Stanford University, Stanford, CA, June 2002.
- [26] F.M.J. Willems. Informationtheoretical results for the discrete memoryless multiple access channel. Ph.D. dissertation, Katholieke Universiteit Leuven, Leuven, Belgium, October 1982.
- [27] A. El Gamal. Network Information Theory. Course Notes for EE 478 (Stanford), 2002.

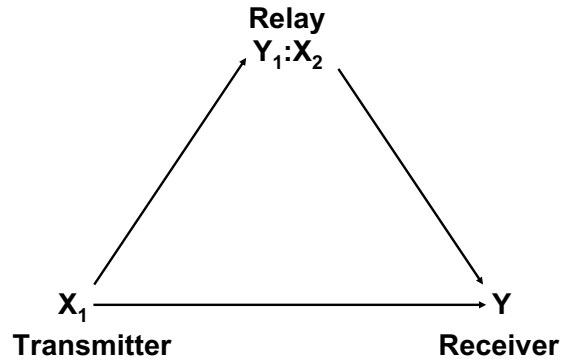


Figure 1: Discrete memoryless relay channel.

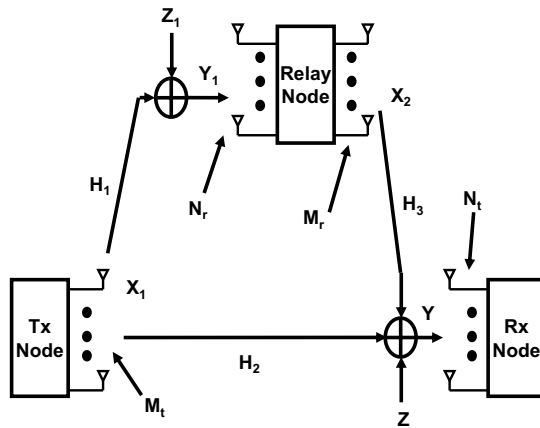


Figure 2: Gaussian MIMO Relay Channel.

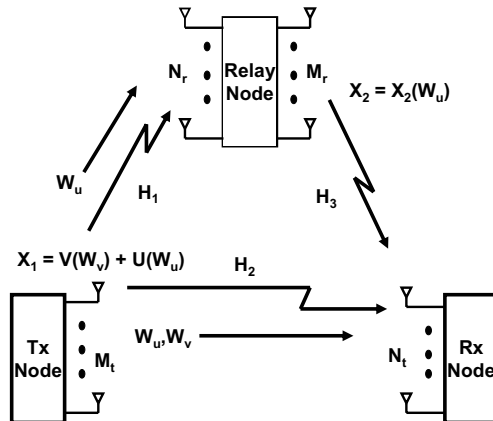


Figure 3: Gaussian MIMO relay channel with superposition coding.

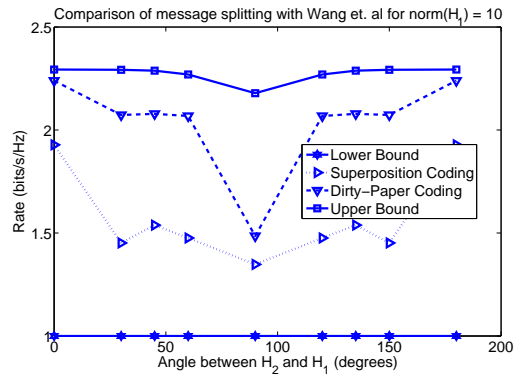


Figure 4: Achievable rate results for the case where the transmitter, the relay, and the receiver are all equidistant from each other, or  $\gamma_1 = \gamma_2 = \gamma_3$ .

Table 1: Block Markov encoding and backward decoding

Block	1	2	...	B	B+1
$\mathcal{V}$	$v^n(w_{v,1})$	$v^n(w_{v,2})$	...	$v^n(w_{v,B})$	$v^n(\phi')$
$\mathcal{U}$	$u^n(k_1, \phi)$	$u^n(k_2, w_{u,1})$	...	$u^n(k_B, w_{u,B-1})$	$u^n(k_{B+1}, w_{u,B})$
$X_2$	$x_2^n(\phi)$	$x_2^n(w_{u,1})$	...	$x_2^n(w_{u,B-1})$	$x_2^n(w_{u,B})$
$Y_1$	$\hat{w}_{u,1}$	$\hat{w}_{u,2}$	...	$\hat{w}_{u,B}$	$\phi''$