Rate Bounds for MIMO Relay Channels Using Message Splitting

Caleb K. Lo, Sriram Vishwanath 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) 471-1190

Fax: (512) 471-6512

Email: {clo, sriram, 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 that allow for partial cooperation between the transmitter and the relay. A partial cooperation strategy that combines transmit-side message splitting and block-Markov encoding is presented. New lower capacity bounds that improve on a previously proposed non-cooperative lower bound are derived for both discrete memoryless relay channels and Gaussian relay channels.

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.

Relay channels were introduced in [6] and upper bounds on their capacity were derived in [7]. 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 [8–11]. 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 [12]. This encouraging result has led to research on multiuser MIMO channels such as Gaussian multiple access (MAC) [13–16] and broadcast (BC) [17–20] channels. Although discrete memoryless relay channels were analyzed decades ago, MIMO relay channels have only recently been studied [22]. In [22], a Gaussian relay channel with multiple antennas at each terminal is considered. Upper and lower capacity bounds are shown for both deterministic and Rayleigh fading channels. The lower bounds for the case of fixed channels in [22] arise from a non-cooperative transmit strategy.

Higher achievable rates than those yielded by the non-cooperative approach in [22] 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 decode-and-forward operations, where the relay decodes the source message, encodes it 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. For a SISO relay channel these degrees of freedom are not present. Thus, coding strategies such as message splitting that do not increase the achievable rate for SISO relay channels 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. For SISO relays as considered in [5], the relay either decodes all of the transmitter's codeword or offers no assistance to the transmitter. In contrast, the 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, can be optimized by exploiting the channel eigenmodes.

We propose new lower capacity bounds for 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 relay channel, dirty-paper coding [21] is employed at the transmitter. Our proposed lower bounds obtained via transmit-side message splitting improve on the lower bounds from [22], which turn out to be special cases 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 capacity bounds from [22] for the Gaussian MIMO relay channel. In Section IV, we present our message splitting strategies for both discrete memoryless and Gaussian 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 the key rate bounds that were originally presented in [23].

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 [10].

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

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

3 Background

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

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

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 $\mathbf{\Sigma}_{ij} = \mathbb{E}(\mathbf{x}_i\mathbf{x}_j^\dagger)$. The authors of [22] 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. [22, Sec. III] *An upper bound on the capacity of the Gaussian MIMO relay channel is given by*

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

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

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

Lemma 3.2. [22, Sec. III] *A lower bound on the capacity of the Gaussian MIMO relay channel is given by*

$$C^G \geq C^G_{lower} = \max(C^G_d, \min(C^G_3, C^G_4)), \tag{5}$$

where

$$\begin{aligned} C^G_d &\triangleq \max_{\boldsymbol{\Sigma}_{11}} \log[\det(\boldsymbol{I}_{N_t} + \gamma_2 \boldsymbol{H}_2 \boldsymbol{\Sigma}_{11} \boldsymbol{H}_2^\dagger)] \\ C^G_3 &\triangleq \max_{\boldsymbol{\Sigma}_{11}} \log[\det(\boldsymbol{I}_{N_r} + \gamma_1 \boldsymbol{H}_1 \boldsymbol{\Sigma}_{11} \boldsymbol{H}_1^\dagger)] \\ C^G_4 &\triangleq \max_{\boldsymbol{\Sigma}_{22}} \log[\det(\boldsymbol{I}_{N_t} + \gamma_3 \boldsymbol{H}_3 \boldsymbol{\Sigma}_{22} \boldsymbol{H}_3^\dagger (\boldsymbol{I}_{N_t} + \gamma_2 \boldsymbol{H}_2 \boldsymbol{\Sigma}^*_{11} \boldsymbol{H}_2^\dagger)^{-1})] \end{aligned} \tag{6}$$

with

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

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 $\mathbf{\Sigma}_u$, $\mathbf{\Sigma}_v$ and $\mathbf{\Sigma}_{x_2}$ to be the covariance matrices of $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{x}_2$ respectively. Also, define

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

and $\mathbf{B} = [\, \sqrt{\gamma_2}\mathbf{H}_2 \quad \sqrt{\gamma_3}\mathbf{H}_3 \,]$. In this case, $\mathbb{E}(\mathbf{u}\mathbf{u}^\dagger) = \mathbf{\Sigma}_u$. 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. It is proved in Appendix A.3 that

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

where

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

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

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

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

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

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

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

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. It is proved in Appendix A.4 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

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

which is analogous to the rate in (12) and

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

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 and prove the following result.

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

$$R_{sc} = \max(R_{sc,u}, R_{sc,v}) \geq C^G_{lower} \tag{17}$$

where C^G_{lower} is given in Lemma 3.2.

Proof. Recall from Lemma 3.2 that

$$C^G \geq C^G_{lower} = \max(C^G_d, \min(C^G_3, C^G_4)), \tag{18}$$

where

$$\begin{aligned} C^G_d &\triangleq \max_{\mathbf{\Sigma}_{11}} \log[\det(\mathbf{I}_{N_t} + \gamma_2 \mathbf{H}_2 \mathbf{\Sigma}_{11} \mathbf{H}_2^\dagger)] \\ C^G_3 &\triangleq \max_{\mathbf{\Sigma}_{11}} \log[\det(\mathbf{I}_{N_r} + \gamma_1 \mathbf{H}_1 \mathbf{\Sigma}_{11} \mathbf{H}_1^\dagger)] \\ C^G_4 &\triangleq \max_{\mathbf{\Sigma}_{22}} \log[\det(\mathbf{I}_{N_t} + \gamma_3 \mathbf{H}_3 \mathbf{\Sigma}_{22} \mathbf{H}_3^\dagger (\mathbf{I}_{N_t} + \gamma_2 \mathbf{H}_2 \mathbf{\Sigma}^*_{11} \mathbf{H}_2^\dagger)^{-1})] \end{aligned} \tag{19}$$

and

$$\mathbf{\Sigma}^*_{11} \triangleq \arg \max_{\mathbf{\Sigma}_{11} \succeq 0} \log[\det(\mathbf{I}_{N_r} + \gamma_1 \mathbf{H}_1 \mathbf{\Sigma}_{11} \mathbf{H}_1^\dagger)]. \tag{20}$$

We show that C^G_d, C^G_3, and C^G_4 arise from special cases of our superposition coding strategy.

First, we set $\mathbf{v} = \mathbf{x}_1$ and $\mathbf{u} = 0$; we obtain the following expression

$$\begin{aligned} R_{sc,u} &= R_{sc,v} \\ &= \sup_{p(x_1)} I(X_1; Y) \\ &= C^G_d. \end{aligned} \tag{21}$$

Next, we set $\mathbf{u} = \mathbf{x}_1$ and $\mathbf{v} = 0$. Also, instead of block-Markov encoding, the relay employs the following encoding approach: it uses its estimate of $\mathbf{u}$ to choose a codeword from its own codebook, which has the same cardinality as the transmitter's codebook. Thus, we have a one-to-one mapping between the elements of the codebooks for the transmitter and the relay. We obtain the following expression

$$\begin{aligned} R_{sc,u} &= R_{sc,v} \\ &= \sup_{p(x_1,x_2)} \min(I(X_1;Y_1|X_2), I(X_1,X_2;Y)) \\ &\geq \min(C_3^G, C_4^G). \end{aligned} \tag{22}$$

It immediately follows that $R_{sc} \geq C^G_{lower}$. □

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.5 that

$$R_{pre,u} = \sup_{p(x_1,x_2,u,v)} (R_{pre,u,1} + R_{pre,u,2}) \tag{23}$$

where

$$\begin{aligned} R_{pre,u,1} &= \min(I(U;Y_1|X_2) - I(U;V|X_2), I(U,X_2;Y)) \\ R_{pre,u,2} &= I(V;Y|U,X_2) \end{aligned} \tag{24}$$

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 $\mathbf{\Sigma}_{x_1'|x_2}$ to be the covariance matrix of $\mathbf{x}_1'$ given knowledge of $\mathbf{x}_2$ and prove in Appendix A.2 that

$$I(U;Y_1|X_2) - I(U;V|X_2) = \log(\det(\mathbf{I}_{N_r} + \gamma_1 \mathbf{H}_1 \mathbf{\Sigma}_{x_1'|x_2} \mathbf{H}_1^\dagger)), \tag{25}$$

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.6 that

$$R_{pre,v} = \sup_{p(x_1,x_2,u,v)} (R_{pre,v,1} + R_{pre,v,2}) \tag{26}$$

where

$$\begin{aligned} R_{pre,v,1} &= \min(I(U;Y_1|X_2) - I(U;V|X_2), I(U,X_2;Y|V)) \\ R_{pre,v,2} &= I(V;Y) \end{aligned} \tag{27}$$

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), (25), 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*

$$R_{pre} = \max(R_{pre,u}, R_{pre,v}) \geq R_{sc} \tag{28}$$

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

$$R_{pre,u} = \sup_{p(v)p(x_2)p(u|v,x_2)p(x_1|u,v)} (\min(I(U;Y_1|X_2) - I(U;V|X_2), I(U,X_2;Y)) + I(V;Y|U,X_2)). \quad (29)$$

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, (29) reduces to

$$R_{sc,u} = \sup_{p(x_2)p(u|x_2)p(v)p(x_1|u,v)} (\min(I(U;Y_1|X_2), I(U,X_2;Y)) + I(V;Y|U,X_2)). \quad (30)$$

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

4.3 Example Calculation

We investigate the performance of our precoding methods for a particular channel configuration. Consider a MIMO relay channel where the transmitter has two antennas, while the relay and receiver each have one antenna and $\mathbf{H}_1 = [5\sqrt{2}\ 5\sqrt{2}]$, $\mathbf{H}_2 = [1\ 0]$, and $\mathbf{H}_3 = 1$. Also, we set $\gamma_1 = \gamma_2 = \gamma_3 = 1$ to model a scenario where the relay is equidistant from the transmitter and the receiver.

4.3.1 Upper Bound

We substitute these values for the channels into the expressions in Theorem 3.1. Here, we find that the upper bound $C^G_{upper} \approx 2.28793$ bits/s/Hz.

4.3.2 Lower Bound

We again substitute these values for the channels into the expressions in Theorem 3.2. We find that the lower bound $C^G_{lower} = 1$ bits/s/Hz.

4.3.3 Our Achievable Rate

We define

$$\mathbf{\Sigma}_u = \begin{bmatrix} a & b \\ b^* & d \end{bmatrix}, \mathbf{\Sigma}_v = \begin{bmatrix} g & q \\ q^* & r \end{bmatrix}, \mathbf{\Sigma}_{x_2} = k \text{ and } \mathbb{E}(\mathbf{u}\mathbf{x}_2^\dagger) = \begin{bmatrix} x \\ y \end{bmatrix}.$$

For superposition coding, if the receiver attempts to decode w_u before decoding w_v,

$$I(U;Y_1|X_2) = \log\left(1 + \frac{50(a + 2Re(b) + d - |x|^2 - 2Re(xy^*) - |y|^2)}{1 + 50(g + 2Re(q) + r)}\right), \tag{31}$$

$$I(U,X_2;Y) = \log\left(1 + \frac{a + 2Re(x) + k}{1 + g}\right), \tag{32}$$

and

$$I(V;Y|U,X_2) = \log(1+g). \tag{33}$$

We choose $\mathbf{\Sigma}_u$, $\mathbf{\Sigma}_v$, and $\mathbf{\Sigma}_{x_2}$ to maximize $R_{sc,u,1}$ and $R_{sc,u,2}$ subject to $\text{tr}(\mathbf{\Sigma}_u + \mathbf{\Sigma}_v) \leq 1$ and $0 \leq k \leq 1$ along with $\mathbf{\Sigma}_u + \mathbf{\Sigma}_v \succeq 0$, $|x| \leq 1$ and $|y| \leq 1$. We find that the optimal values are

$$\mathbf{\Sigma}_u = \begin{bmatrix} 0.504 & 0.0768 \\ 0.0768 & 0.16 \end{bmatrix}, \mathbf{\Sigma}_v = \begin{bmatrix} 0.108 & 0.0768 \\ 0.0768 & 0.141 \end{bmatrix},$$
$$\Sigma_{x_2} = [1], \text{ and } \mathbb{E}(\mathbf{u}\mathbf{x}_2^\dagger) = [\,0.146 \quad 0.0768\,].$$

Thus $R_{sc,u} = R_{sc,u,1} + R_{sc,u,2} \approx 1.5383$ bits/s/Hz.

If the receiver attempts to decode w_v before decoding w_u, $I(U;Y_1|X_2)$ is the same as in (31),

$$I(U,X_2;Y|V) = \log(1 + a + 2Re(x) + k), \tag{34}$$

and

$$I(V;Y) = \log\left(1 + \frac{g}{1 + a + 2Re(x) + k}\right). \tag{35}$$

We find that the optimal values are

$$\mathbf{\Sigma}_u = \begin{bmatrix} 0.523 & 0.0342 \\ 0.0342 & 0.16 \end{bmatrix}, \mathbf{\Sigma}_v = \begin{bmatrix} 0.116 & 0.0342 \\ 0.0342 & 0.182 \end{bmatrix},$$

$$\mathbf{\Sigma}_{x_2} = [1], \text{ and } \mathbb{E}(\mathbf{u}\mathbf{x}_2^\dagger) = [\,0.107 \quad 0.0342\,].$$

Thus $R_{sc,v} = R_{sc,v,1} + R_{sc,v,2} \approx 1.51192$ bits/s/Hz.

Comparing the achievable rate for both cases, we choose to decode w_u first and so $R_{sc} \approx 1.5383$ bits/s/Hz. We have used superposition coding to outperform the lower bound from [22, Sec. III].

For dirty-paper coding, we choose $\mathbf{\Sigma}_{x_1'|x_2}$, $\mathbf{\Sigma}_v$, and $\mathbf{\Sigma}_{x_2}$ to maximize $R_{pre,u,1}$ and $R_{pre,u,2}$. Since $\mathbf{u} = \mathbf{G}\mathbf{v} + \mathbf{x}_1'$, we have $\mathbf{\Sigma}_u = \mathbf{G}\mathbf{\Sigma}_v\mathbf{G}^\dagger + \mathbf{\Sigma}_{x_1'|x_2}$, where the optimal value of $\mathbf{G}$ is given in Appendix A.2.

We define

$$\mathbf{\Sigma}_{x_1'|x_2} = \begin{bmatrix} a & b \\ b^* & d \end{bmatrix}, \mathbf{\Sigma}_v = \begin{bmatrix} g & q \\ q^* & r \end{bmatrix}, \mathbf{\Sigma}_{x_2} = k \text{ and } \mathbb{E}(\mathbf{u}\mathbf{x}_2^\dagger) = \begin{bmatrix} x \\ y \end{bmatrix}.$$

If the receiver attempts to decode w_u before decoding w_v,

$$I(U;Y_1|X_2) - I(U;V|X_2) = \log(1 + 50(a + 2Re(b) + d)), \tag{36}$$

and $I(U, X_2; Y)$ and $I(V; Y|U, X_2)$ can be computed from (11) and (12), respectively. We find that the optimal values are

$$\mathbf{\Sigma}_{x_1'|x_2} = \begin{bmatrix} 0.5 & 6.81 \cdot 10^{-6} \\ 6.81 \cdot 10^{-6} & 0.103 \end{bmatrix}, \mathbf{\Sigma}_v = \begin{bmatrix} 0.264 & 0.173 \\ 0.173 & 0.124 \end{bmatrix},$$

$$\mathbf{\Sigma}_{x_2} = [0.988], \text{ and } \mathbb{E}(\mathbf{u}\mathbf{x}_2^\dagger) = [\,0.499 \quad 0\,].$$

Thus $R_{pre,u} = R_{pre,u,1} + R_{pre,u,2} \approx 2.07851$ bits/s/Hz.

If the receiver attempts to decode w_v before decoding w_u, $I(U;Y_1|X_2) - I(U;V|X_2)$ is the same as in (36) and $I(U, X_2; Y|V)$ and $I(V;Y)$ can be computed from (16) and (15), respectively. We find that the optimal values are the same as for the case where the receiver attempts to decode w_u before decoding w_v, and $R_{pre,v} = R_{pre,v,1} + R_{pre,v,2} \approx 2.07851$ bits/s/Hz.

Comparing the achievable rate for both cases, the receiver can choose either decoding order, and so $R_{pre} \approx 2.07851$ bits/s/Hz. We have used dirty-paper coding to outperform the lower bound from [22, Sec. III].

5 Numerical Results

We employ a simple example to demonstrate how transmit-side message splitting outperforms the bounds in [22, 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 $\Theta(\mathbf{H}_1, \mathbf{H}_2) \to \pi/2$, 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 [22, 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].

Fig. 4 shows the rates that are achieved by our message splitting strategies along with the upper and lower bounds from [22, Sec. III]. We see that the upper bound decreases as $\Theta(\mathbf{H}_1, \mathbf{H}_2) \to \pi/2$ radians. Also, as $\Theta(\mathbf{H}_1, \mathbf{H}_2) \to \pi/2$, 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 $\Theta(\mathbf{H}_1, \mathbf{H}_2) \to \pi/2$.

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 capacity bounds for MIMO relay channels via transmit-side message splitting. Our proposed bounds improve upon the lower bounds that were introduced in [22]. 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.

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

$$\begin{aligned} \mathbf{y}_1 &= \sqrt{\gamma_1}\mathbf{H}_1\mathbf{x}_1 + \mathbf{z}_1 \\ &= \sqrt{\gamma_1}\mathbf{H}_1(\mathbf{u}+\mathbf{v}) + \mathbf{z}_1 \end{aligned} \tag{37}$$

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

$$\begin{aligned} h(Y_1|X_2) &= h(\sqrt{\gamma_1}\mathbf{H}_1(U+V) + Z_1|X_2) \\ &= \log((2\pi e)^{N_r}\det(\gamma_1\mathbf{H}_1(\mathbf{\Sigma}_{u|x_2}+\mathbf{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})) \end{aligned} \tag{38}$$

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

$$\begin{aligned} h(Y_1|X_2,U) &= h(\sqrt{\gamma_1}\mathbf{H}_1(U+V) + Z_1|X_2,U) \\ &= h(\sqrt{\gamma_1}\mathbf{H}_1 V + Z_1|X_2,U) \\ &= h(\sqrt{\gamma_1}\mathbf{H}_1 V + Z_1) \\ &= \log((2\pi e)^{N_r}\det(\gamma_1\mathbf{H}_1\mathbf{\Sigma}_v\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})). \end{aligned} \tag{39}$$

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

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

so

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

and

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

Thus we have

$$h(Y_1|X_2) = \log((2\pi e)^{N_r} \det(\gamma_1 \mathbf{H}_1(\mathbf{\Sigma}_u - \mathbb{E}(\mathbf{u}\mathbf{x}_2^\dagger)\mathbf{\Sigma}_{x_2}^{-1}\mathbb{E}(\mathbf{x}_2\mathbf{u}^\dagger) + \mathbf{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})) \tag{40}$$

and finally we obtain

$$I(U;Y_1|X_2) = \log\left(\frac{\det\left(\mathbf{I}_{N_r} + \gamma_1\mathbf{H}_1\left(\mathbf{\Sigma}_u - \mathbb{E}(\mathbf{u}\mathbf{x}_2^\dagger)\mathbf{\Sigma}_{x_2}^{-1}\mathbb{E}(\mathbf{x}_2\mathbf{u}^\dagger) + \mathbf{\Sigma}_v\right)\mathbf{H}_1^\dagger\right)}{\det\left(\mathbf{I}_{N_r} + \gamma_1\mathbf{H}_1\mathbf{\Sigma}_v\mathbf{H}_1^\dagger\right)}\right). \tag{41}$$

A.2 Establishing (25)

Here we follow a procedure that is similar to the proof in Appendix C of [24]. Recall that 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. Our objective is to choose $\mathbf{G}$ to maximize $I(U;Y_1|X_2)$ - $I(U;V|X_2)$.

We have $I(U;Y_1|X_2) = h(Y_1|X_2)$ - $h(U,Y_1|X_2) + h(U|X_2)$. Also, $I(U;V|X_2) = h(U|X_2)$ - $h(U|V,X_2)$. Thus, $R(\mathbf{G}) = I(U;Y_1|X_2)$ - $I(U;V|X_2) = h(Y_1|X_2)$ - $h(U,Y_1|X_2)$ + $h(U|V,X_2)$. Since $\mathbf{z}_1$ is independent of $\mathbf{x}_1'$ and $\mathbf{v}$, we have

$$\begin{aligned} h(Y_1|X_2) &= h(\sqrt{\gamma_1}\mathbf{H}_1 X_1 + Z_1|X_2) \\ &= \log((2\pi e)^{N_r}\det(\gamma_1\mathbf{H}_1(\mathbf{\Sigma}_{x_1'|x_2} + \mathbf{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})) \end{aligned} \tag{42}$$

along with

$$\begin{aligned} h(U|V,X_2) &= h(X_1'|X_2) \\ &= \log((2\pi e)^{M_t}\det(\mathbf{\Sigma}_{x_1'|x_2})) \end{aligned} \tag{43}$$

and

$$h(U,Y_1|X_2) = \log((2\pi e)^{N_r+M_t}\det(\mathbf{K})) \tag{44}$$

where

$$K = \begin{bmatrix} \gamma_1\mathbf{H}_1(\mathbf{\Sigma}_{x_1'|x_2} + \mathbf{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r} & \sqrt{\gamma_1}\mathbf{H}_1(\mathbf{\Sigma}_v\mathbf{G}^\dagger + \mathbf{\Sigma}_{x_1'|x_2}) \\ (\mathbf{G}\mathbf{\Sigma}_v + \mathbf{\Sigma}_{x_1'|x_2})\sqrt{\gamma_1}\mathbf{H}_1^\dagger & \mathbf{G}\mathbf{\Sigma}_v\mathbf{G}^\dagger + \mathbf{\Sigma}_{x_1'|x_2} \end{bmatrix}.$$

Using the Schur's complement formula for the determinant of $\mathbf{K}$, we find that

$$R(\mathbf{G}) = \log\left(\frac{\det(\mathbf{\Sigma}_{x_1'|x_2})}{B}\right). \tag{45}$$

Here we have

$$\begin{aligned} B &= \det(\mathbf{G}\mathbf{\Sigma}_v\mathbf{G}^\dagger + \mathbf{\Sigma}_{x_1'|x_2} - \gamma_1(\mathbf{G}\mathbf{\Sigma}_v + \mathbf{\Sigma}_{x_1'|x_2}) \\ &\quad \cdot\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1(\mathbf{\Sigma}_{x_1'|x_2} + \mathbf{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1(\mathbf{\Sigma}_v\mathbf{G}^\dagger + \mathbf{\Sigma}_{x_1'|x_2})). \end{aligned} \tag{46}$$

Now we expand B as $B = \det((\mathbf{G}a - b)(\mathbf{G}a - b)^\dagger + c)$ and find that

$$aa^\dagger = \mathbf{\Sigma}_v - \gamma_1\mathbf{\Sigma}_v\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1(\mathbf{\Sigma}_{x_1'|x_2} + \mathbf{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1\mathbf{\Sigma}_v$$

$$ba^\dagger = \gamma_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1(\mathbf{\Sigma}_{x_1'|x_2} + \mathbf{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1\mathbf{\Sigma}_v$$

$$bb^\dagger + c = \mathbf{\Sigma}_{x_1'|x_2} - \gamma_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1(\mathbf{\Sigma}_{x_1'|x_2} + \mathbf{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1\mathbf{\Sigma}_{x_1'|x_2}.$$

Thus, we find that the value of $\mathbf{G}$ that minimizes B is

$$\begin{aligned} \mathbf{G}_{min} &= (ba^\dagger)(aa^\dagger)^{-1} \\ &= \gamma_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1(\mathbf{\Sigma}_{x_1'|x_2} + \mathbf{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1\mathbf{\Sigma}_v \\ &\quad \cdot(\mathbf{\Sigma}_v - \gamma_1\mathbf{\Sigma}_v\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1(\mathbf{\Sigma}_{x_1'|x_2} + \mathbf{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1\mathbf{\Sigma}_v)^{-1} \\ &= \gamma_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1(\mathbf{\Sigma}_{x_1'|x_2} + \mathbf{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1 \\ &\quad \cdot(\mathbf{I}_{M_t} + \gamma_1\mathbf{\Sigma}_v\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1) \\ &= \gamma_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1. \end{aligned} \tag{47}$$

The minimum value of B is $\det(c)$, so we solve for c from $bb^\dagger + c$. We see that

$$\begin{aligned} bb^\dagger &= \mathbf{G}_{min}aa^\dagger(\mathbf{G}_{min})^\dagger \\ &= \gamma_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1 \\ &\quad \cdot(\mathbf{\Sigma}_v - \gamma_1\mathbf{\Sigma}_v\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1(\mathbf{\Sigma}_{x_1'|x_2} + \mathbf{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1\mathbf{\Sigma}_v) \\ &\quad \cdot(\gamma_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1)^\dagger \\ &= \gamma_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1(\mathbf{\Sigma}_v^{-1} + \gamma_1\mathbf{H}_1^\dagger \\ &\quad \cdot(\gamma_1\mathbf{H}_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1)^{-1}\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1\gamma_1\mathbf{\Sigma}_{x_1'|x_2} \\ &= \gamma_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1(\mathbf{\Sigma}_{x_1'|x_2} + \mathbf{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1} \\ &\quad \cdot\mathbf{H}_1\mathbf{\Sigma}_v\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1\mathbf{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1\gamma_1\mathbf{\Sigma}_{x_1'|x_2}. \end{aligned} \tag{48}$$

Now we find that

$$
\begin{aligned}
c &= \boldsymbol{\Sigma}_{x_1'|x_2} - \gamma_1 \boldsymbol{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1(\boldsymbol{\Sigma}_{x_1'|x_2}+\boldsymbol{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1\boldsymbol{\Sigma}_{x_1'|x_2} \\
&\quad -\gamma_1 \boldsymbol{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1(\boldsymbol{\Sigma}_{x_1'|x_2}+\boldsymbol{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1\boldsymbol{\Sigma}_v\mathbf{H}_1^\dagger \\
&\quad \cdot(\gamma_1\mathbf{H}_1\boldsymbol{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1\gamma_1\boldsymbol{\Sigma}_{x_1'|x_2} \\
&= \boldsymbol{\Sigma}_{x_1'|x_2} - \gamma_1 \boldsymbol{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1(\boldsymbol{\Sigma}_{x_1'|x_2}+\boldsymbol{\Sigma}_v)\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1 \\
&\quad \cdot(\mathbf{I}_{M_t} + \gamma_1\boldsymbol{\Sigma}_v\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1\boldsymbol{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1)\boldsymbol{\Sigma}_{x_1'|x_2} \\
&= \boldsymbol{\Sigma}_{x_1'|x_2} - \gamma_1 \boldsymbol{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger(\gamma_1\mathbf{H}_1\boldsymbol{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger + \mathbf{I}_{N_r})^{-1}\mathbf{H}_1\boldsymbol{\Sigma}_{x_1'|x_2}.
\end{aligned} \tag{49}
$$

Finally, we obtain

$$
R(\mathbf{G}_{min}) = \log\left(\frac{\det\left(\boldsymbol{\Sigma}_{x_1'|x_2}\right)}{\det\left(\boldsymbol{\Sigma}_{x_1'|x_2} - \gamma_1\boldsymbol{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger\left(\gamma_1\mathbf{H}_1\boldsymbol{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger + \mathbf{I}_{N_r}\right)^{-1}\mathbf{H}_1\boldsymbol{\Sigma}_{x_1'|x_2}\right)}\right) \tag{50}
$$

which results in

$$
R(\mathbf{G}_{min}) = \log(\det(\mathbf{I}_{N_r} + \gamma_1\mathbf{H}_1\boldsymbol{\Sigma}_{x_1'|x_2}\mathbf{H}_1^\dagger)). \tag{51}
$$

A.3 Achievability Proof of (8)

We want to send a codeword $\mathbf{x}$ from the transmitter. The relay observes $\mathbf{y}_1$, decodes it and sends an encoded version of the result as $\mathbf{x}_2$. Finally, the receiver observes $\mathbf{y}$ which is composed of both the transmitter's signal and the relay's signal. Let $\mathbf{u}$ and $\mathbf{v}$ be auxiliary variables that, taken together, determine $\mathbf{x}$. We perform transmit-side message splitting, where $\mathbf{v}$ is treated as a state variable for the transmit-to-relay channel; thus, the relay only decodes $\mathbf{u}$.

Define the set of messages at the transmitter to be $\mathcal{W} = \mathcal{W}_u \times \mathcal{W}_v$. Let $w_u \in \mathcal{W}_u$ represent the part of the message that the relay decodes, and let $w_v \in \mathcal{W}_v$ represent the part of the message that only the receiver decodes.

A.3.1 Block Markov Encoding

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

The senders use a triply-indexed set of codewords:

$$\begin{aligned} \mathcal{C} \; = \; & \{(v^n(w_v), u^n(w_u|a), x_2^n(a)) : w_v \in \{\phi, 1, 2, \ldots, 2^{nR_v}\}, w_u \in \{\phi, 1, 2, \ldots, 2^{nR_u}\}, \\ & a \in \{\phi, 1, 2, \ldots, 2^{nR_0}\}\}. \end{aligned} \tag{52}$$

$u^n(\phi|a)$ means that the codeword u^n only depends on a, and a_i 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.

A.3.2 Generation of Random Code

Fix $p(v)p(x_2)p(u|x_2)p(x_1|u,v)$. 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 $v^n(w_v)$, $w_v \in \{1, 2, \ldots, 2^{nR_v}\}$. Generate at random 2^{nR_0} i.i.d. x_2^n sequences according to $\sim \prod_{i=1}^n p(x_{2i})$, and index them as $x_2^n(a)$, $a \in \{1, 2, \ldots, 2^{nR_0}\}$. For each $x_2^n(a)$, generate 2^{nR_u} conditionally independent u^n sequences according to $\sim \prod_{i=1}^n p(u_i|x_{2i})$, and index them as $u^n(w_u|a)$ where $w_u \in \{1, 2, \ldots, 2^{nR_u}\}$. This defines the random codebook $\mathcal{C} = \{(v^n(w_v), u^n(w_u|a), x_2^n(a))\}$.

For each message $w_u \in \{1, 2, \ldots, 2^{nR_u}\}$ assign an index $a(w_u)$ at random from $\{1, 2, \ldots, 2^{nR_0}\}$. The set of messages with the same index a form a bin $A_a \subseteq \mathcal{W}_u$.

Finally, generate the codeword x_1^n via $p(x_1^n|u^n, v^n)$. In the special case of superposition coding with Gaussian signals,

$$x_1^n = v^n(w_v) + u^n(w_u|a). \tag{53}$$

A.3.3 Encoding

Let $w_{u,i} \in \{1, 2, \ldots, 2^{nR_u}\}$ and $w_{v,i} \in \{1, 2, \ldots, 2^{nR_v}\}$ comprise the new message to be sent in block i and assume that $w_{u,i-1} \in A_{a_i}$; the encoder sends x_1^n which is comprised of $v^n(w_{v,i})$ and $u^n(w_{u,i}|a_i)$.

Assuming that the relay estimated $\hat{w}_{u,i-1}$ for the previous index $w_{u,i-1}$, where $\hat{w}_{u,i-1} \in A_{\hat{a}_i}$, then the relay sends $x_2^n(\hat{a}_i)$ in block i. Here, $\hat{a}_i$ is the bin index for $w_{u,i-1}$; for example, $\hat{a}_2$ is the bin index for $w_{u,1}$.

A.3.4 Decoding and Error Analysis

Assume that at the end of block i - 1, the receiver has correctly estimated $(w_{v,1},w_{v,2},\ldots,w_{v,i-2})$, $(w_{u,1},w_{u,2},\ldots,w_{u,i-2})$, and $(a_2,\ldots,a_{i-1})$ while the relay has correctly estimated $(w_{u,1},w_{u,2},\ldots,w_{u,i-1})$ and consequently $(a_2,\ldots,a_i)$. The decoding procedures at the end of block i are as follows for the case where the receiver attempts to decode w_u before decoding w_v.

Our error analysis employs the concept of strong typicality. As defined in [26], $T_\epsilon^{(n)}(X)$ is the set of ϵ-strongly typical sequences with respect to some $p(x)$ on $\mathcal{X}$ if $\forall\ x^n \in \mathcal{X}^n$,

$$\forall a \in \mathcal{X} \text{ such that } p(a) > 0, \text{ we have } \left|\frac{1}{n}N(a|x^n) - p(a)\right| < \frac{\epsilon}{|\mathcal{X}|}$$

and

$$\forall a \in \mathcal{X} \text{ such that } p(a) = 0, \text{ we have } N(a|x^n) = 0$$

where $N(a|x^n)$ is the number of times that a appears in x^n.

We proceed through four decoding steps. Define the following error events:

- E_{0i} as the event that $(u^n(w_{u,i}|a_i),x_2^n(a_i),y_1^n(i),y^n(i)) \notin T_\epsilon^{(n)}$, 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 = 1,2,3,4.

Thus, the overall probability of error $P_e^{(n)} = P(\bigcup_{m=0}^4 E_{mi}) \leq \sum_{m=0}^4 P(E_{mi})$. 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 = 1,2,3,4 as follows.

Decoding step 1: Upon observing $y_1^n(i)$, the relay receiver declares that $\hat{w}_{u,i} = \hat{w}_u$ is sent if it is the unique index such that $(u^n(\hat{w}_u|a_i),x_2^n(a_i),y_1^n(i)) \in T_\epsilon^{(n)}$. Here, E_{1i} is the event that $\exists\ \hat{w}_u \neq w_{u,i}$ such that $(u^n(\hat{w}_u|a_i), x_2^n(a_i), y_1^n(i)) \in T_\epsilon^{(n)}$. Now, for $\hat{w}_u \neq w_{u,i}$,

$$
\begin{aligned}
P(E_{1i}|\hat{w}_u) &= P((\mathbf{u}(\hat{w}_u|a_i), \mathbf{x}_2(a_i), \mathbf{y}_1(i)) \in T_\epsilon^{(n)}) \\
&= \sum_{(u^n(\hat{w}_u|a_i), y_1^n(i)) \in T_\epsilon^{(n)}(U, Y_1|x_2^n), \hat{w}_u \neq w_{u,i}} p(u^n(\hat{w}_u|a_i), y_1^n(i)|x_2^n(a_i)) \\
&= \sum_{(u^n(\hat{w}_u|a_i), y_1^n(i)) \in T_\epsilon^{(n)}(U, Y_1|x_2^n), \hat{w}_u \neq w_{u,i}} p(u^n(\hat{w}_u|a_i)|x_2^n(a_i)) p(y_1^n(i)|x_2^n(a_i)) \quad (54) \\
&\leq \left|T_\epsilon^{(n)}(U, Y_1|x_2^n)\right| 2^{-n(H(U|X_2)-\delta(\epsilon))} 2^{-n(H(Y_1|X_2)-\delta_1(\epsilon))} \\
&\leq 2^{-n(H(U|X_2)+H(Y_1|X_2)-H(U,Y_1|X_2)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \\
&= 2^{-n(H(Y_1|X_2)-H(Y_1|X_2,U)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \\
&= 2^{-n(I(U;Y_1|X_2)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))}
\end{aligned}
$$

where (54) follows from the fact that $\mathbf{y}_1(i)$ and $\mathbf{u}(\hat{w}_u|a_i)$ are independent for $\hat{w}_u \neq w_{u,i}$. Also, we have $\delta(\epsilon) \to 0$, $\delta_1(\epsilon) \to 0$ and $\delta_2(\epsilon) \to 0$ as $\epsilon \to 0$. Thus, we have

$$
P(E_{1i}) \leq 2^{nR_u} \cdot 2^{-n(I(U;Y_1|X_2)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \quad (55)
$$

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

$$
R_u < I(U; Y_1|X_2). \quad (56)
$$

Decoding step 2: The receiver declares that $\tilde{a}_i = \tilde{a}$ was sent if it is the unique index such that $(x_2^n(\tilde{a}), y^n(i)) \in T_\epsilon^{(n)}$. Here, E_{2i} is the event that $\exists\ \tilde{a} \neq a_i$ such that $(x_2^n(\tilde{a}), y^n(i)) \in T_\epsilon^{(n)}$. Now, for $\tilde{a} \neq a_i$,

$$
\begin{aligned}
P(E_{2i}|\tilde{a}) &= P((\mathbf{x}_2(\tilde{a}), \mathbf{y}(i)) \in T_\epsilon^{(n)}) \\
&= \sum_{(x_2^n(\tilde{a}), y^n(i)) \in T_\epsilon^{(n)}(X_2, Y), \tilde{a} \neq a_i} p(x_2^n(\tilde{a}), y^n(i)) \\
&= \sum_{(x_2^n(\tilde{a}), y^n(i)) \in T_\epsilon^{(n)}(X_2, Y), \tilde{a} \neq a_i} p(x_2^n(\tilde{a}) p(y^n(i)) \quad (57) \\
&\leq \left|T_\epsilon^{(n)}(X_2, Y)\right| 2^{-n(H(X_2)-\delta(\epsilon))} 2^{-n(H(Y)-\delta_1(\epsilon))} \\
&\leq 2^{-n(H(X_2)+H(Y)-H(X_2,Y)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \\
&\leq 2^{-n(I(X_2;Y)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))}
\end{aligned}
$$

where (57) follows from the fact that $\mathbf{y}(i)$ and $\mathbf{x}_2(\tilde{a})$ are independent for $\tilde{a} \neq a_i$. Also, we have $\delta(\epsilon) \to 0$, $\delta_1(\epsilon) \to 0$ and $\delta_2(\epsilon) \to 0$ as $\epsilon \to 0$. Thus, we have

$$P(E_{2i}) \leq 2^{nR_0} \cdot 2^{-n(I(X_2;Y)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \tag{58}$$

and so $\tilde{a}_i = a_i$ with $P(E_{2i})$ arbitrarily small if n is sufficiently large and

$$R_0 < I(X_2;Y). \tag{59}$$

Decoding step 3: Assuming that a_i is decoded correctly at the receiver, it constructs the list $\mathcal{L}(y^n(i-1))$ of message w_u indices whose codewords are jointly typical with $y^n(i-1)$ in block i - 1. The receiver then declares that $\tilde{w}_{u,i-1}$ is sent in block i - 1 if it is the unique index in $A_{a_i} \cap \mathcal{L}(y^n(i-1))$. Here, $E_{3i} \triangleq E_{3i}^{'} \cup E_{3i}^{''}$ where

$$E_{3i}^{'} = \{w_{u,i-1} \notin A_{a_i} \cap \mathcal{L}(y^n(i-1))\} \tag{60}$$

and

$$E_{3i}^{''} = \{\exists \tilde{w} \in [1, 2^{nR_u}], \tilde{w} \neq w_{u,i-1}, \text{such that } \tilde{w} \in A_{a_i} \cap \mathcal{L}(y^n(i-1))\}. \tag{61}$$

We follow an approach similar to that in [5, pg. 576-577] to show that $\tilde{w}_{u,i-1} = w_{u,i-1}$ with $P(E_{3i})$ arbitrarily small if n is sufficiently large and

$$R_u < I(U;Y|X_2) + R_0. \tag{62}$$

Let the error event G_i for a decoding error for w_u in the i-th block be $G_i = \bigcup_{m=0}^{3} E_{mi}$. Define $\psi(w_u|y^n(i-1))$ to be 1 if $(u^n(w_u|a_{i-1}), x_2^n(a_{i-1}), y^n(i-1)) \in T_\epsilon^{(n)}$ and 0 otherwise. Then, $\|\mathcal{L}(y^n(i-1))\| = \sum_{w_u} \psi(w_u|y^n(i-1))$ and so

$$\mathbb{E}(\|\mathcal{L}(y^n(i-1))\||G_{i-1}^c) = \mathbb{E}(\psi(w_{u,i-1}|y^n(i-1))|G_{i-1}^c) + \sum_{w_u \neq w_{u,i-1}} \mathbb{E}(\psi(w_u|y^n(i-1))|G_{i-1}^c) \tag{63}$$

and for each $w_u \in [1, 2^{nR_u}]$, $w_u \neq w_{u,i-1}$, we have

$$\mathbb{E}(\psi(w_u|y^n(i-1))|G_{i-1}^c) \leq 2^{-n(I(U;Y|X_2)-\delta(\epsilon))} \tag{64}$$

where we have $\delta(\epsilon) \to 0$ as $\epsilon \to 0$.

Then,

$$\begin{aligned} \mathbb{E}(\|\mathcal{L}(y^n(i-1))\| | G_{i-1}^c) &\leq 1 + (2^{nR_u} - 1)(2^{-n(I(U;Y|X_2)-\delta(\epsilon))}) \\ &\leq 1 + 2^{n(R_u - I(U;Y|X_2)+\delta(\epsilon))}. \end{aligned}$$

If G_{i-1}^c occurs, then $w_{u,i-1} \in \mathcal{L}(y^n(i-1))$. We also note that if E_{2i}^c occurs, then $\tilde{a}_i = a_i$ and so $w_{u,i-1} \in A_{\tilde{a}_i}$. It then follows that $\mathrm{P}(E_{3i}^{'} \cap E_{2i}^c \cap E_{0i}^c | G_{i-1}^c) = 0$. Let $E = E_{3i} \cap E_{2i}^c \cap E_{0i}^c$. We can now show that

$$\begin{aligned} P(E|G_{i-1}^c) &= P(E_{3i}^{''} \cap E_{2i}^c \cap E_{0i}^c | G_{i-1}^c) \\ &\leq P(\exists w_u \neq w_{u,i-1} \text{ where } w_u \in K_1 | G_{i-1}^c) \qquad (65) \\ &\leq \mathbb{E}\left(\sum_{w_u \neq w_{u,i-1}, w_u \in \mathcal{L}(y^n(i-1))} P(\{w_u \in A_{a_i}\}) | G_{i-1}^c \right) \\ &\leq \mathbb{E}(\|\mathcal{L}(y^n(i-1))\| 2^{-nR_0} | G_{i-1}^c) \\ &\leq 2^{-nR_0}(1 + 2^{n(R_u - I(U;Y|X_2)+\delta(\epsilon))}) \end{aligned}$$

where $K_1 = A_{a_i} \cap \mathcal{L}(y^n(i-1))$ in (65).

Therefore, if

$$R_u < I(U;Y|X_2) + R_0 - \delta(\epsilon) \qquad (66)$$

then $\mathrm{P}(E_{3i} \cap E_{2i}^c \cap E_{0i}^c | G_{i-1}^c) \leq \delta_1(\epsilon)/(4\mathrm{B})$ if n is sufficiently large. Here, we have $\delta_1(\epsilon) \to 0$ as $\epsilon \to 0$.

Thus, for $P(E_{3i})$ to be arbitrarily small when n is sufficiently large, we must have (using (59))

$$R_u < I(U;Y|X_2) + I(X_2;Y) = I(U, X_2; Y) \qquad (67)$$

and by combining (56) and (67) we obtain

$$R_u < \min(I(U;Y_1|X_2), I(U, X_2; Y)). \qquad (68)$$

Decoding step 4: The receiver estimates $\tilde{w}_{v,i-1} = \tilde{w}_v$ if it is the unique index such that $(v^n(\tilde{w}_v), y^n(i), u^n(\tilde{w}_{u,i-1}|a_i), x_2^n(a_i)) \in T_\epsilon^{(n)}$. Here, E_{4i} is the event that $\exists\, \tilde{w}_v \neq w_{v,i-1}$

such that $(v^n(\tilde{w}_v),y^n(i),u^n(\tilde{w}_{u,i-1}|a_i),x_2^n(a_i)) \in T_\epsilon^{(n)}$. Now, for $\tilde{w}_v \neq w_{v,i-1}$,

$$\begin{aligned}
P(E_{4i}|\tilde{w}_v) &= P((\mathbf{v}(\tilde{w}_v),\mathbf{y}(i),\mathbf{u}(\tilde{w}_{u,i-1}|a_i),\mathbf{x}_2(a_i)) \in T_\epsilon^{(n)}) \\
&= \sum_{(v^n(\tilde{w}_v),y^n(i)) \in T_\epsilon^{(n)}(V,Y|u^n,x_2^n),\tilde{w}_v \neq w_{v,i-1}} p(v^n(\tilde{w}_v),y^n(i)|u^n(\tilde{w}_{u,i-1}|a_i),x_2^n(a_i)) \\
&= \sum_{(v^n(\tilde{w}_v),y^n(i)) \in T_\epsilon^{(n)}(V,Y|u^n,x_2^n),\tilde{w}_v \neq w_{v,i-1}} p(v^n(\tilde{w}_v)|u^n(\tilde{w}_{u,i-1}|a_i),x_2^n(a_i)) \\
&\quad \cdot p(y^n(i)|u^n(\tilde{w}_{u,i-1}|a_i),x_2^n(a_i)) \qquad (69) \\
&\leq \left|T_\epsilon^{(n)}(V,Y|u^n,x_2^n)\right| 2^{-n(H(V|U,X_2)-\delta(\epsilon))} 2^{-n(H(Y|U,X_2)-\delta_1(\epsilon))} \\
&\leq 2^{-n(H(V|U,X_2)+H(Y|U,X_2)-H(V,Y|U,X_2)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \\
&\leq 2^{-n(I(V;Y|U,X_2)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))}
\end{aligned}$$

where (69) follows from the fact that $\mathbf{y}(i)$ and $\mathbf{v}(\tilde{w}_v)$ are independent for $\tilde{w}_v \neq w_{v,i-1}$. Also, we have $\delta(\epsilon) \to 0$, $\delta_1(\epsilon) \to 0$ and $\delta_2(\epsilon) \to 0$ as $\epsilon \to 0$. Thus, we have

$$P(E_{4i}) \leq 2^{nR_v} \cdot 2^{-n(I(V;Y|U,X_2)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \qquad (70)$$

and so $\tilde{w}_{v,i-1} = w_{v,i-1}$ with $P(E_{4i})$ arbitrarily small if n is sufficiently large and if

$$R_v < I(V;Y|U,X_2). \qquad (71)$$

A.4 Achievability Proof of (13)

Apply the code generation and encoding procedures from Section A.3.

A.4.1 Decoding and Error Analysis

Assume that at the end of block i - 1, the receiver has correctly estimated $(w_{v,1},w_{v,2},\dots,w_{v,i-1})$, $(w_{u,1},w_{u,2},\dots,w_{u,i-2})$, and $(a_2,\dots,a_{i-1})$ while the relay has correctly estimated $(w_{u,1},w_{u,2},\dots,w_{u,i-1})$ and consequently $(a_2,\dots,a_i)$. The decoding procedures at the end of block i are as follows for the case where the receiver attempts to decode w_v before decoding w_u.

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

- E_{0i} as the event that $(u^n(w_{u,i}|a_i), x_2^n(a_i), y_1^n(i), y^n(i)) \notin T_\epsilon^{(n)}$, 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 $m = 1,2,3,4$.

Thus, the overall probability of error $P_e^{(n)} = P(\bigcup_{m=0}^{4} E_{mi}) \leq \sum_{m=0}^{4} P(E_{mi})$. 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 $m = 1,2,3,4$ 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 $\tilde{w}_v \neq w_{v,i}$,

$$
\begin{aligned}
P(E_{1i}|\tilde{w}_v) &= P((\mathbf{v}(\tilde{w}_v), \mathbf{y}(i)) \in T_\epsilon^{(n)}) \\
&= \sum_{(v^n(\tilde{w}_v), y^n(i)) \in T_\epsilon^{(n)}(V,Y), \tilde{w}_v \neq w_{v,i}} p(v^n(\tilde{w}_v), y^n(i)) \\
&= \sum_{(v^n(\tilde{w}_v), y^n(i)) \in T_\epsilon^{(n)}(V,Y), \tilde{w}_v \neq w_{v,i}} p(v^n(\tilde{w}_v)) p(y^n(i)) \qquad (72) \\
&\leq \left| T_\epsilon^{(n)}(V,Y) \right| 2^{-n(H(V)-\delta(\epsilon))} 2^{-n(H(Y)-\delta_1(\epsilon))} \\
&\leq 2^{-n(H(V)+H(Y)-H(V,Y)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \\
&\leq 2^{-n(I(V;Y)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))}
\end{aligned}
$$

where (72) 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 $\delta(\epsilon) \to 0$, $\delta_1(\epsilon) \to 0$ and $\delta_2(\epsilon) \to 0$ as $\epsilon \to 0$. Thus, we have

$$
P(E_{1i}) \leq 2^{nR_v} \cdot 2^{-n(I(V;Y)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \qquad (73)
$$

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

$$
R_v < I(V;Y). \qquad (74)
$$

Decoding step 2: The analysis for this decoding step is similar to the analysis for decoding step 1 in Section A.3. Thus we have

$$R_u < I(U;Y_1|X_2). \tag{75}$$

Decoding step 3: The receiver declares that $\tilde{a}_i = \tilde{a}$ was sent if it is the unique index such that $(x_2^n(\tilde{a}),y^n(i),v^n(\tilde{w}_{v,i})) \in T_\epsilon^{(n)}$. Here, E_{3i} is the event that $\exists\ \tilde{a} \neq a_i$ such that $(x_2^n(\tilde{a}),y^n(i),v^n(\tilde{w}_{v,i})) \in T_\epsilon^{(n)}$. Now, for $\tilde{a} \neq a_i$,

$$\begin{aligned}
P(E_{3i}|\tilde{a}) &= P((\mathbf{x}_2(\tilde{a}), \mathbf{y}(i), \mathbf{v}(\tilde{w}_{v,i})) \in T_\epsilon^{(n)}) \\
&= \sum_{(x_2^n(\tilde{a}),y^n(i)) \in T_\epsilon^{(n)}(X_2,Y|v^n), \tilde{a}\neq a_i} p(x_2^n(\tilde{a}), y^n(i)|v^n(\tilde{w}_{v,i})) \\
&= \sum_{(x_2^n(\tilde{a}),y^n(i)) \in T_\epsilon^{(n)}(X_2,Y|v^n), \tilde{a}\neq a_i} p(x_2^n(\tilde{a})|v^n(\tilde{w}_{v,i}))p(y^n(i)|v^n(\tilde{w}_{v,i})) \qquad (76) \\
&\leq \left|T_\epsilon^{(n)}(X_2, Y|v^n)\right| 2^{-n(H(X_2|V)-\delta(\epsilon))} 2^{-n(H(Y|V)-\delta_1(\epsilon))} \\
&\leq 2^{-n(H(X_2|V)+H(Y|V)-H(X_2,Y|V)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \\
&\leq 2^{-n(I(X_2;Y|V)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))}
\end{aligned}$$

where (76) follows from the fact that $\mathbf{y}(i)$ and $\mathbf{x}_2(\tilde{a})$ are independent for $\tilde{a} \neq a_i$. Also, we have $\delta(\epsilon) \to 0$, $\delta_1(\epsilon) \to 0$ and $\delta_2(\epsilon) \to 0$ as $\epsilon \to 0$. Thus, we have

$$P(E_{3i}) \leq 2^{nR_0} \cdot 2^{-n(I(X_2;Y|V)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \tag{77}$$

and so $\tilde{a}_i = a_i$ with $P(E_{3i})$ arbitrarily small if n is sufficiently large and

$$R_0 < I(X_2;Y|V). \tag{78}$$

Decoding step 4: Assuming that a_i is decoded correctly at the receiver, it constructs the list $\mathcal{L}(y^n(i-1))$ of message w_u indices whose codewords are jointly typical with $y^n(i-1)$ in block i - 1. The receiver then declares that $\tilde{w}_{u,i-1}$ is sent in block i - 1 if it is the unique index in $A_{a_i} \cap \mathcal{L}(y^n(i-1))$. Here, $E_{4i} \triangleq E_{4i}^{'} \cup E_{4i}^{''}$ where

$$E_{4i}^{'} = \{w_{u,i-1} \notin A_{a_i} \cap \mathcal{L}(y^n(i-1))\} \tag{79}$$

and

$$E_{4i}^{''} = \{\exists \tilde{w} \in [1, 2^{nR_u}], \tilde{w} \neq w_{u,i-1}, \text{such that } \tilde{w} \in A_{a_i} \cap \mathcal{L}(y^n(i-1))\}. \quad (80)$$

Once again, we follow an approach similar to that in [5, pg. 576-577] to show that $\tilde{w}_{u,i-1} = w_{u,i-1}$ with $P(E_{4i})$ arbitrarily small if n is sufficiently large and

$$R_u < I(U;Y|X_2,V) + R_0. \quad (81)$$

Let the error event D_i for a decoding error for w_u in the i-th block as $D_i = E_{0i} \cup \bigcup_{m=2}^{4} E_{mi}$. Define $\psi(w_u|y^n(i-1))$ to be 1 if $(u^n(w_u|a_{i-1}),x_2^n(a_{i-1}),y^n(i-1),v^n(w_{v,i-1})) \in T_\epsilon^{(n)}$ and 0 otherwise. Then, $\|\mathcal{L}(y^n(i-1))\| = \sum_{w_u} \psi(w_u|y^n(i-1))$ and so

$$\mathbb{E}(\|\mathcal{L}(y^n(i-1))\||D_{i-1}^c) = \mathbb{E}(\psi(w_{u,i-1}|y^n(i-1))|D_{i-1}^c) + \sum_{w_u \neq w_{u,i-1}} \mathbb{E}(\psi(w_u|y^n(i-1))|D_{i-1}^c) \quad (82)$$

and for each $w_u \in [1,2^{nR_u}]$, $w_u \neq w_{u,i-1}$, we have

$$\mathbb{E}(\psi(w_u|y^n(i-1))|D_{i-1}^c) \leq 2^{-n(I(U;Y|X_2,V)-\delta(\epsilon))} \quad (83)$$

where we have $\delta(\epsilon) \to 0$ as $\epsilon \to 0$.

Then,

$$\begin{aligned} \mathbb{E}(\|\mathcal{L}(y^n(i-1))\||D_{i-1}^c) &\leq 1 + (2^{nR_u} - 1)(2^{-n(I(U;Y|X_2,V)-\delta(\epsilon))}) \\ &\leq 1 + 2^{n(R_u - I(U;Y|X_2,V)+\delta(\epsilon))}. \end{aligned}$$

If D_{i-1}^c occurs, then $w_{u,i-1} \in \mathcal{L}(y^n(i-1))$. We also note that if E_{3i}^c occurs, then $\tilde{a}_i = a_i$ and so $w_{u,i-1} \in A_{\tilde{a}_i}$. It then follows that $\mathrm{P}(E_{4i}^{'} \cap E_{3i}^c \cap E_{0i}^c|D_{i-1}^c) = 0$. Let $E = E_{4i} \cap E_{3i}^c \cap E_{0i}^c$. We can now show that

$$\begin{aligned} P(E|D_{i-1}^c) &= P(E_{4i}^{''} \cap E_{3i}^c \cap E_{0i}^c|D_{i-1}^c) \\ &\leq P(\exists w_u \neq w_{u,i-1} \text{ where } w_u \in K_3|D_{i-1}^c) \qquad (84) \\ &\leq \mathbb{E}\left(\sum_{w_u \neq w_{u,i-1}, w_u \in \mathcal{L}(y^n(i-1))} P(\{w_u \in A_{a_i}\})|D_{i-1}^c \right) \\ &\leq \mathbb{E}(\|\mathcal{L}(y^n(i-1))\| 2^{-nR_0}|D_{i-1}^c) \\ &\leq 2^{-nR_0}(1 + 2^{n(R_u - I(U;Y|X_2,V)+\delta(\epsilon))}) \end{aligned}$$

where $K_3 = A_{a_i} \cap \mathcal{L}(y^n(i-1))$ in (84).

Therefore, if

$$R_u < I(U;Y|X_2,V) + R_0 - \delta(\epsilon) \tag{85}$$

then $\mathrm{P}(E_{4i} \cap E_{3i}^c \cap E_{0i}^c | D_{i-1}^c) \leq \delta_1(\epsilon)/(4\mathrm{B})$ if n is sufficiently large. Here, we have $\delta_1(\epsilon) \to 0$ as $\epsilon \to 0$.

Thus, for $P(E_{4i})$ to be arbitrarily small when n is sufficiently large, we must have (using (78))

$$R_u < I(U;Y|X_2,V) + I(X_2;Y|V) = I(U,X_2;Y|V) \tag{86}$$

and by combining (75) and (86) we obtain

$$R_u < \min(I(U;Y_1|X_2), I(U,X_2;Y|V)). \tag{87}$$

A.5 Achievability Proof of (23)

This proof relies on the concept of backward decoding, which was introduced in [25].

A.5.1 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,. . . ,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:

$$\begin{aligned} \mathcal{C} \;=\; & \{(v^n(w_v), u^n(k, w_{2u}), x_2^n(w_{2u})) : w_v \in \{\phi, 1, 2, \ldots, 2^{nR_v}\}, \\ & k \in \{1, 2, \ldots, 2^{n(I(U;Y_1|X_2)-\delta(\epsilon))}\}, w_{2u} \in \{\phi, 1, 2, \ldots, 2^{nR_u}\}\}. \end{aligned} \tag{88}$$

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 2 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.5.2 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 $v^n(w_v)$, $w_v \in \{1, 2, \ldots, 2^{nR_v}\}$. 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 $x_2^n(w_{2u})$, $w_{2u} \in \{1, 2, \ldots, 2^{nR_u}\}$. 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 $\mathcal{F}$: $u^n(k, w_{2u}) \to w_u$. Here, $k \in \{1, 2, \ldots, 2^{n(I(U;Y_1|X_2)-\delta(\epsilon))}\}$, $w_{2u} \in \{\phi, 1, 2, \ldots, 2^{nR_u}\}$, and w_u $\in \{\phi, 1, 2, \ldots, 2^{nR_u}\}$. For example, $\mathcal{F}(u^n(1, w_{2u})) = \mathcal{F}(u^n(2, w_{2u})) = \ldots$ $= \mathcal{F}(u^n(2^{n(I(U;Y_1|X_2)-R_u-\delta(\epsilon))}, w_{2u})) = w_u = 1$. 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.5.3 Encoding

Let $w_{u,i} \in \{1, 2, \ldots, 2^{nR_u}\}$ and $w_{v,i} \in \{1, 2, \ldots, 2^{nR_v}\}$ 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, $\mathrm{P}(x_1^n \in T_\epsilon^{(n)}(x_1|u^n(k, w_{u,i-1}), v^n(w_{v,i}))) > 1$ - ϵ.

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.5.4 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

$$\begin{aligned} P(E_{1i}) &= P((\mathbf{u}, v^n(w_{v,i}), \mathbf{x}_2(w_{u,i-1})) \notin T_\epsilon^{(n)}) \\ &= \sum_{v^n} p(v^n) P((\mathbf{u}, v^n, \mathbf{x}_2(w_{u,i-1})) \notin T_\epsilon^{(n)}) \\ &= \sum_{v^n} p(v^n)(1 - P((\mathbf{u}, v^n, \mathbf{x}_2(w_{u,i-1})) \in T_\epsilon^{(n)}))^{2^{n(I(U;Y_1|X_2)-R_u-\delta(\epsilon))}} \\ &\leq e^{-2^{n(I(U;Y_1|X_2)-R_u-\delta(\epsilon))} 2^{-n(I(U;V|X_2)+\delta_1(\epsilon))}} \end{aligned} \tag{89}$$

which is arbitrarily small for n sufficiently large if

$$R_u < I(U;Y_1|X_2) - I(U;V|X_2) - \delta_1(\epsilon) - \delta(\epsilon). \tag{90}$$

We note that (89) follows from the following two facts:

i) $2^{-n(I(U;V|X_2)+\delta_1(\epsilon))} \leq \mathrm{P}((\mathbf{u}, v^n, \mathbf{x}_2(w_{u,i-1})) \in T_\epsilon^{(n)})$ 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

$$R_u < I(U;Y_1|X_2) - I(U;V|X_2). \tag{91}$$

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 $(\mathcal{F}(\tilde{w}_{u,i+1}, w_{2u}),x_2^n(w_{2u}),y^n(i)) \in T_\epsilon^{(n)}$. 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 $(\mathcal{F}(\tilde{w}_{u,i+1}, \tilde{w}_u),x_2^n(\tilde{w}_u),y^n(i)) \in T_\epsilon^{(n)}$. Now, for $\tilde{w}_u \neq w_{u,i}$,

$$\begin{aligned}
P(E_{2i}|\tilde{w}_u) &= P((\mathcal{F}(\tilde{w}_{u,i+1}, \tilde{w}_u), \mathbf{x}_2(\tilde{w}_u), \mathbf{y}(i)) \in T_\epsilon^{(n)}) \\
&= \sum_{(u^n(k_i,\tilde{w}_u),x_2^n(\tilde{w}_u),y^n(i))\in T_\epsilon^{(n)}(U,X_2,Y),\tilde{w}_u\neq w_{u,i},\mathcal{F}(k_i,\tilde{w}_u)=\tilde{w}_{u,i+1}} \\
&\quad p(u^n(k_i, \tilde{w}_u), x_2^n(\tilde{w}_u), y^n(i)) \qquad (92) \\
&= \sum_{(u^n(k_i,\tilde{w}_u),x_2^n(\tilde{w}_u),y^n(i))\in T_\epsilon^{(n)}(U,X_2,Y),\tilde{w}_u\neq w_{u,i},\mathcal{F}(k_i,\tilde{w}_u)=\tilde{w}_{u,i+1}} \\
&\quad p(u^n(k_i, \tilde{w}_u), x_2^n(\tilde{w}_u))p(y^n(i)) \qquad (93) \\
&\leq \left|T_\epsilon^{(n)}(U, X_2, Y)\right| 2^{-n(H(U,X_2)-\delta(\epsilon))}2^{-n(H(Y)-\delta_1(\epsilon))} \\
&\leq 2^{-n(H(U,X_2)+H(Y)-H(U,X_2,Y)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \\
&\leq 2^{-n(I(U,X_2;Y)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))}
\end{aligned}$$

where (93) 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 $\delta(\epsilon) \to 0$, $\delta_1(\epsilon) \to 0$, and $\delta_2(\epsilon) \to 0$ as $\epsilon \to 0$. Thus, we have

$$P(E_{2i}) \leq 2^{nR_u} \cdot 2^{-n(I(U,X_2;Y)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \tag{94}$$

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

$$R_u < I(U, X_2; Y). \tag{95}$$

Now, we combine (91) and (95) to obtain

$$R_u < \min((I(U;Y_1|X_2) - I(U;V|X_2)), I(U, X_2; Y)). \tag{96}$$

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 $(\mathcal{F}(\tilde{w}_{u,i+1},\tilde{w}_{u,i}),x_2^n(\tilde{w}_{u,i}),y^n(i),v^n(w_v)) \in T_\epsilon^{(n)}$. 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 $(\mathcal{F}(\tilde{w}_{u,i+1},\tilde{w}_{u,i}),x_2^n(\tilde{w}_{u,i}),y^n(i),v^n(\tilde{w}_v)) \in T_\epsilon^{(n)}$. Now, for $\tilde{w}_v \neq w_{v,i}$,

$$\begin{aligned}
P(E_{3i}|\tilde{w}_v) &= P((\mathcal{F}(\tilde{w}_{u,i+1},\tilde{w}_{u,i}),\mathbf{x}_2(\tilde{w}_{u,i}),\mathbf{y}(i),\mathbf{v}(\tilde{w}_v)) \in T_\epsilon^{(n)}) \\
&= \sum_{(v^n(\tilde{w}_v),y^n(i))\in T_\epsilon^{(n)}(V,Y|u^n,x_2^n),\tilde{w}_v\neq w_{v,i},\mathcal{F}(k_i,\tilde{w}_{u,i})=\tilde{w}_{u,i+1}} p(v^n(\tilde{w}_v),y^n(i)|u^n(k_i,\tilde{w}_{u,i}),x_2^n(\tilde{w}_{u,i})) \\
&= \sum_{(v^n(\tilde{w}_v),y^n(i))\in T_\epsilon^{(n)}(V,Y|u^n,x_2^n),\tilde{w}_v\neq w_{v,i},\mathcal{F}(k_i,\tilde{w}_{u,i})=\tilde{w}_{u,i+1}} p(v^n(\tilde{w}_v)|u^n(k_i,\tilde{w}_{u,i}),x_2^n(\tilde{w}_{u,i}))p(y^n(i)|u^n(k_i,\tilde{w}_{u,i}),x_2^n(\tilde{w}_{u,i})) \qquad (97) \\
&\leq \left|T_\epsilon^{(n)}(V,Y|u^n,x_2^n)\right| 2^{-n(H(V|U,X_2)-\delta(\epsilon))}2^{-n(H(Y|U,X_2)-\delta_1(\epsilon))} \\
&\leq 2^{-n(H(V|U,X_2)+H(Y|U,X_2)-H(V,Y|U,X_2)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \\
&\leq 2^{-n(I(V;Y|U,X_2)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))}
\end{aligned}$$

where (97) 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 $\delta(\epsilon) \to 0$, $\delta_1(\epsilon) \to 0$, and $\delta_2(\epsilon) \to 0$ as $\epsilon \to 0$. Thus, we have

$$P(E_{3i}) \leq 2^{nR_v} \cdot 2^{-n(I(V;Y|U,X_2)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \qquad (98)$$

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

$$R_v < I(V;Y|U,X_2). \qquad (99)$$

A.6 Achievability Proof of (26)

This proof also relies on the concept of backward decoding. Apply the code generation and encoding procedures from Section A.5. 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 $i = 1,2,\ldots,B$.

A.6.1 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 $\tilde{w}_v \neq w_{v,i}$,

$$\begin{aligned} P(E_{1i}|\tilde{w}_v) &= P((\mathbf{v}(\tilde{w}_v), \mathbf{y}(i)) \in T_\epsilon^{(n)}) \\ &= \sum_{(v^n(\tilde{w}_v), y^n(i)) \in T_\epsilon^{(n)}(V,Y), \tilde{w}_v \neq w_{v,i}} p(v^n(\tilde{w}_v), y^n(i)) \\ &= \sum_{(v^n(\tilde{w}_v), y^n(i)) \in T_\epsilon^{(n)}(V,Y), \tilde{w}_v \neq w_{v,i}} p(v^n(\tilde{w}_v)) p(y^n(i)) \qquad (100) \\ &\leq \left|T_\epsilon^{(n)}(V,Y)\right| 2^{-n(H(V)-\delta(\epsilon))} 2^{-n(H(Y)-\delta_1(\epsilon))} \end{aligned}$$

$$\leq \; 2^{-n(H(V)+H(Y)-H(V,Y)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))}$$
$$\leq \; 2^{-n(I(V;Y)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))}$$

where (100) 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 $\delta(\epsilon) \to 0$, $\delta_1(\epsilon) \to 0$, and $\delta_2(\epsilon) \to 0$ as $\epsilon \to 0$. Thus, we have

$$P(E_{1i}) \leq 2^{nR_v} \cdot 2^{-n(I(V;Y)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \tag{101}$$

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

$$R_v < I(V;Y). \tag{102}$$

Decoding step 2: The analysis for this decoding step is similar to the analysis for decoding step 1 in Section A.5. Thus we have

$$R_u < I(U;Y_1|X_2) - I(U;V|X_2). \tag{103}$$

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 $(\mathcal{F}(\tilde{w}_{u,i+1}, w_{2u}), x_2^n(w_{2u}), y^n(i), v^n(\tilde{w}_{v,i})) \in T_\epsilon^{(n)}$. 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 $(\mathcal{F}(\tilde{w}_{u,i+1}, \tilde{w}_u), x_2^n(\tilde{w}_u), y^n(i), v^n(\tilde{w}_{v,i})) \in T_\epsilon^{(n)}$. Now, for $\tilde{w}_u \neq w_{u,i}$,

$$\begin{aligned} P(E_{3i}|\tilde{w}_u) &= P((\mathcal{F}(\tilde{w}_{u,i+1}, \tilde{w}_u), \mathbf{x}_2(\tilde{w}_u), \mathbf{y}(i), \mathbf{v}(\tilde{w}_{v,i})) \in T_\epsilon^{(n)}) \\ &= \sum_{(u^n(k_i,\tilde{w}_u), x_2^n(\tilde{w}_u), y^n(i)) \in T_\epsilon^{(n)}(U,X_2,Y|v^n), \tilde{w}_u \neq w_{u,i}} p(u^n(k_i,\tilde{w}_u), x_2^n(\tilde{w}_u), y^n(i)|v^n(\tilde{w}_{v,i})) \\ &= \sum_{(u^n(k_i,\tilde{w}_u), x_2^n(\tilde{w}_u), y^n(i)) \in T_\epsilon^{(n)}(U,X_2,Y|v^n), \tilde{w}_u \neq w_{u,i}} p(u^n(k_i,\tilde{w}_u), x_2^n(\tilde{w}_u)|v^n(\tilde{w}_{v,i}) p(y^n(i)|v^n(\tilde{w}_{v,i}) \qquad (104) \\ &\leq \left|T_\epsilon^{(n)}(U,X_2,Y|v^n)\right| 2^{-n(H(U,X_2|V)-\delta(\epsilon))} 2^{-n(H(Y|V)-\delta_1(\epsilon))} \\ &\leq 2^{-n(H(U,X_2|V)+H(Y|V)-H(U,X_2,Y|V)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \\ &\leq 2^{-n(I(U,X_2;Y|V)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \end{aligned}$$

where (104) 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 $\delta(\epsilon) \to 0$, $\delta_1(\epsilon) \to 0$, and $\delta_2(\epsilon) \to 0$ as $\epsilon \to 0$. Thus, we have

$$P(E_{3i}) \leq 2^{nR_u} \cdot 2^{-n(I(U,X_2;Y|V)-\delta(\epsilon)-\delta_1(\epsilon)-\delta_2(\epsilon))} \tag{105}$$

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

$$R_u < I(U, X_2; Y|V). \tag{106}$$

Now, we combine (103) and (106) to obtain

$$R_u < \min((I(U;Y_1|X_2) - I(U;V|X_2)), I(U, X_2; Y|V)). \tag{107}$$

References

[1] Bruno R, Conti M, Gregori E. Mesh networks: commodity multihop ad hoc networks. *IEEE Commun. Mag.*, **43(3)**:123–131, March 2005.

[2] Pabst R et al. Relay-based deployment concepts for wireless and mobile broadband radio. *IEEE Commun. Mag.*, **42(9)**:80–89, September 2004.

[3] Viswanathan H, Mukherjee S. Throughput-range trade of wireless mesh backhaul networks. *IEEE J. Select. Areas Commun.*, **24(3)**:593–602, March 2006.

[4] Jun J, Sichitiu ML. The nominal capacity of wireless mesh networks. *IEEE Wireless Commun.*, **10(5)**:8–14, October 2003.

[5] Cover T, El Gamal A. Capacity theorems for the relay channel. *IEEE Trans. Inform. Theory*, **25(5)**:572–584, September 1979.

[6] van der Meulen EC. Three-terminal communication channels. *Adv. Appl. Prob.*, **3**:120–154, 1971.

[7] Sato H. Information transmission through a channel with relay. *Tech. Rep. B76-7*, The Aloha System, University of Hawaii, Honolulu, March 1976.

[8] Foschini GJ. 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.

[9] Winters J. 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.

[10] Marzetta T, Hochwald B. Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading. *IEEE Trans. Inform. Theory*, **45(1)**:139–157, January 1999.

[11] Gesbert D 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.

[12] Telatar E. Capacity of multi-antenna Gaussian channels. *Eur. Trans. Telecommun. ETT*, **10(6)**:585–596, November 1999.

[13] Cheng RS, Verdú S. Gaussian multiaccess channels with ISI: capacity region and multiuser water-filling. *IEEE Trans. Inform. Theory*, **39(3)**:773–785, May 1993.

[14] Wyner AD. Shannon-theoretic approach to a Gaussian cellular multiple-access channel. *IEEE Trans. Inform. Theory*, **40(6)**:1713–1727, November 1994.

[15] Viswanath P, Tse DNC, Anantharam V. Asymptotically optimal water-filling in vector multiple-access channels. *IEEE Trans. Inform. Theory*, **47(1)**:241–267, January 2001.

[16] Yu W, Rhee W, Boyd S, Cioffi JM. Iterative water-filling for Gaussian vector multiple access channels. *IEEE Trans. Inform. Theory*, **50(1)**:145–152, January 2004.

[17] Caire G, Shamai S. On the achievable throughput of a multiantenna Gaussian broadcast channel. *IEEE Trans. Inform. Theory*, **49(7)**:1691–1706, July 2003.

[18] Weingarten H, Steinberg Y, Shamai S. The capacity region of the Gaussian MIMO broadcast channel. *IEEE Trans. Inform. Theory*, **52(9)**:3936–3964, September 2006.

[19] Jindal N, Vishwanath S, Goldsmith A. On the duality of Gaussian multiple-access and broadcast channels. *IEEE Trans. Inform. Theory*, **50(5)**:768–783, May 2004.

[20] Mohseni M, Cioffi JM. 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.

[21] Costa M. Writing on dirty paper. *IEEE Trans. Inform. Theory*, **29(3)**:439–441, May 1983.

[22] Wang B, Zhang J, Host-Madsen A. On the capacity of MIMO relay channels. *IEEE Trans. Inform. Theory*, **51(1)**:29–43, January 2005.

[23] Lo CK, Vishwanath S, Heath Jr RW. Rate bounds for MIMO relay channels using precoding. In *Proc. of the IEEE Global Telecommun. Conf.*, **3**:1172-1176, St. Louis, MO, November 2005.

[24] Yu W. Competition and cooperation in multi-user communication environments. Ph.D. dissertation, Stanford University, Stanford, CA, June 2002.

[25] Willems FMJ. Informationtheoretical results for the discrete memoryless multiple access channel. Ph.D. dissertation, Katholieke Universiteit Leuven, Leuven, Belgium, October 1982.

[26] El Gamal A. Network Information Theory. Course Notes for EE 478 (Stanford), 2002.

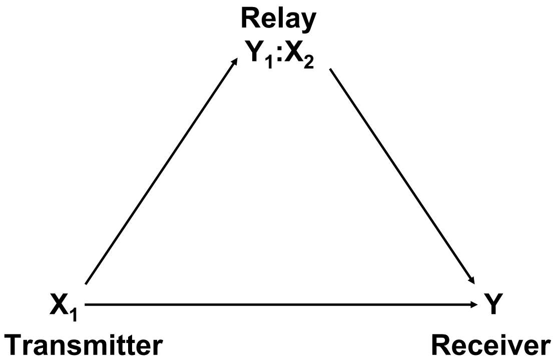

Figure 1: Discrete memoryless relay channel.

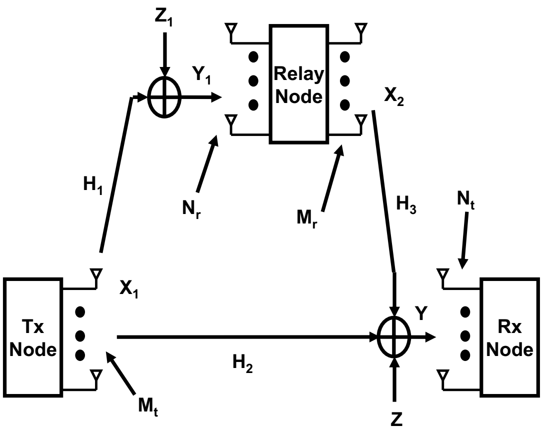

Figure 2: Gaussian MIMO Relay Channel.

Table 1: Block Markov encoding and decoding

Block	1	2	...	B-1	B
$\mathcal{V}$	$v^n(w_{v,1})$	$v^n(w_{v,2})$	...	$v^n(w_{v,B-1})$	$v^n(\phi)$
$\mathcal{U}$	$u^n(w_{u,1}\|\phi)$	$u^n(w_{u,2}\|a_2)$	...	$u^n(w_{u,B-1}\|a_{B-1})$	$u^n(\phi\|a_B)$
X_2	$x_2^n(\phi)$	$x_2^n(\hat{a}_2)$	...	$x_2^n(\hat{a}_{B-1})$	$x_2^n(\hat{a}_B)$
Y_1	$\hat{w}_{u,1}$,$\hat{a}_2$	$\hat{w}_{u,2}$,$\hat{a}_3$	...	$\hat{w}_{u,B-1}$,$\hat{a}_B$	ϕ
Y	ϕ	$\tilde{a}_2$,$\tilde{w}_{u,1}$,$\tilde{w}_{v,1}$	...	$\tilde{a}_{B-1}$,$\tilde{w}_{u,B-2}$,$\tilde{w}_{v,B-2}$	$\tilde{a}_B$,$\tilde{w}_{u,B-1}$,$\tilde{w}_{v,B-1}$

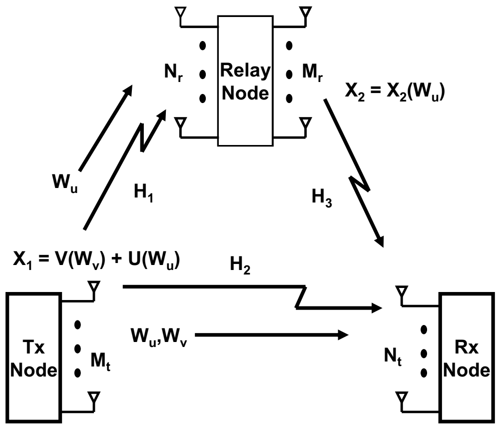

Figure 3: Gaussian MIMO relay channel with superposition coding.

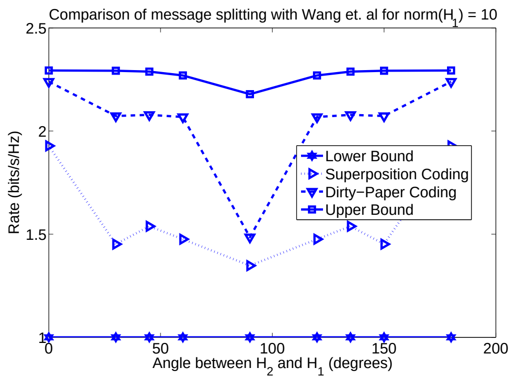

Figure 4: Achievable rate results for the case where the transmitter, the relay, and the receiver are all equidistant from each other, or $\gamma_1 = \gamma_2 = \gamma_3$.

Table 2: 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	$\hat{w}_{u,1}$	$\hat{w}_{u,2}$	...	$\hat{w}_{u,B}$	ϕ''