Capacity Scaling for MIMO Two-Way Relaying

Rahul Vaze and Robert W. Heath Jr.
The University of Texas at Austin
Department of Electrical and Computer Engineering
Wireless Networking and Communications Group
1 University Station C0803
Austin, TX 78712-0240
email: vaze@ece.utexas.edu, rheath@ece.utexas.edu

Abstract

This paper considers capacity scaling in a MIMO (multiple input multiple output) two-way relay channel where two nodes exchange data with each other through multiple relay nodes. The sources and relay nodes have multiple antennas and operate in half-duplex mode. The source nodes are assumed to have perfect receive channel state information for all the channels in the two-way relay network, while the relay nodes either know their own transmit and receive channel state information (the coherent case) or have no knowledge of the channel state (the non-coherent case). The main results in this paper are on capacity scaling of the two-way channel in independent and identically distributed Gaussian matrix channels for large numbers of relays. For both the coherent and non-coherent case, we characterize the capacity region of MIMO two-way relay channel, as the number of relays grow large. The coherent capacity region is shown to scale linearly with the number of source antennas and logarithmically with the number of relay nodes, while the non-coherent capacity is shown to scale linearly with the number of transmit antennas at the source nodes, and logarithmically with signal to noise ratio (SNR), exactly as in the point-to-point capacity for the independent and identically distributed MIMO Gaussian channel. A main conclusion is that MIMO two-way relay channels (both coherent and non-coherent) it is possible to asymptotically (in the number of relays) obtain unidirectional full-duplex performance while using only half-duplex nodes.

I. Introduction

Relay channels are perhaps the most basic building block for cooperative and multihop communication in wireless networks. In a relay channel, one or more nodes called relays without data of their own to transmit help a particular source destination pair communicate. The origins of the relay channel - as a three terminal communication channel - go back to Van der Meulen [1]. Despite the passage of time, the capacity of even the most basic relay channels are still unknown. Nonetheless, bounds derived in [1], [2] show that using a relay, it is possible to increase the reliable rate of data transfer between the source and the destination.

Recently, the relay channel and its variations has been the subject of renewed research, under the general guise of cooperative communication [4], [5], [6], [7] for wireless channels. In the cooperative communication setting, users cooperate by taking turns relaying each others data. Thanks to the spatial separation between users, cooperation between users provides a means to obtain and exploit spatial diversity without requiring multiple antennas at each user. Several different protocols have been proposed to exploit the cooperative diversity gain in a wireless network. For example, a cooperation protocol for a cellular system is introduced in [4], [5] where each user transmits a part of the message directly to the base-station and other part of the message is relayed to base-station via some other user who acts as a relay. Many other cooperation protocols have been proposed and studied in literature, namely: amplify and forward [6], [7], [13], decode and forward [15], [18], with half-duplex [14], and full duplex assumptions [16].

Motivated by MIMO capacity improvements in point-to-point channels [30], there is now interest in MIMO relay channels where the source, destination, and receiver may have multiple antennas [3], [10], [11]. The capacity of Gaussian MIMO relay channels is studied in [3], where upper and lower bounds on capacity of MIMO relay channel are derived for both fixed channels and Rayleigh fading channels. An improved lower bound for MIMO relay channel for Rayleigh fading channels is provided by [10], where message splitting and superposition coding are used at the transmitter to improve the bounds provided in [3]. In [3], [10] only full-duplex relays (can transmit and receive at the same time) are considered. Some capacity bounds for the more practical Gaussian MIMO relay channel with half-duplex relays, where the relays cannot transmit and receive at the same time, were developed in [11]. The results in [3], [10], [11] do not give the exact capacity expression, but the bounds derived therein indicate that with relays there is a potential capacity gain to be leveraged by using multiple antennas.

Although the aforementioned research shows that there are benefits to using relays in wireless networks, general capacity analysis of relay channels, cooperative communication, or more broadly multi-hop networking is largely an open problem. Recently there has been some success, however, with establishing network capacity scaling laws in the limit of large number of nodes, where network capacity is defined as the sum of the capacity of all the simultaneous transmissions that can be supported in the network. For a dense wireless network with K nodes, where any given node randomly wants to communicate to any other node via multi-hop and each of the intermediate nodes employs decode and forward strategy, Gupta and Kumar [12] showed that the network capacity scales as $\mathcal{O}(\sqrt{K})$ as $K \to \infty$. Restricting this general setup to a network where there is only one active source-destination pair and all other nodes assist this source-destination transmission, Gastpar and Vetterli [19] showed that the network capacity scales logarithmically in the number of users when the number of users grows large. This result holds when each node in the network has only 1 antenna. For the same setup as in [19], Bölcskei et. al. showed that there is a M fold increase in the capacity, if both the source and destination are equipped with M antennas.

Prior work on the capacity scaling of large networks [12], [19], [20], assumes only one-way communication. Most networks, however, are two way in nature: the destination terminal also has some data to send to source terminal. An enhancement of the relay channel that includes two-way communication is known by several names including the two-way relay channel [21], bidirectional relay channel [24], [23] or analog network coding [28]. A specific embodiment of this channel that assumes half-duplex relays and the absence of a direct path between source and destination was proposed in [21]. We use the term two-way relay channel or two-way relay protocol to refer to this specific case of half-duplex relays. An illustration is provided in Fig. 1. In the first time slot, both terminals T_1 and T_2 are scheduled to transmit simultaneously while the relay receives. In the next time slot, the relay is scheduled to transmit while terminals T_1 and T_2 receive. The key idea is that each terminal can cancel the interference (generated by its own transmission) from the signal it receives from the relays, to recover the transmission from the other terminal. The idea is reminiscent of work in network coding [27], though note that here the coding is done in the analog domain [28] rather than in digital domain [27].

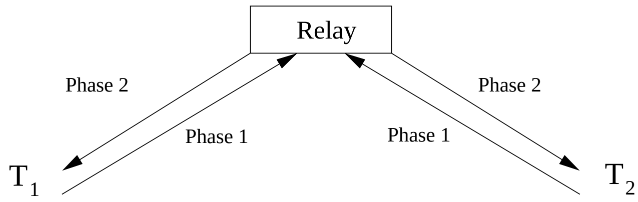


Fig. 1. Two way relaying Protocol

For the two-way relaying protocol, for the case of Rayleigh fading and Gaussian noise channels, an achievable sum rate (sum of the rates achievable on the $T_1 \to T_2$ and $T_2 \to T_1$ links) expression is derived in [21]. It is shown that by using two-way relaying, it is possible to remove the $\frac{1}{2}$ rate loss factor in spectral efficiency due to the half duplex assumption on the nodes. The optimal achievable rate region, however, is not known for this case. With multiple antennas at both T_1, T_2 and the relay, using decode and forward protocol at the relay, optimal achievable rates for Rayleigh fading and Gaussian noise two-way relay channel are derived in [23], which involve waterfilling at each node. Recently, for the case of restricted two-way relay channel (where the transmitters T_1 and T_2 encode their information independently) with half-duplex nodes, capacity region has been derived in [24], which is given by the minimum of the capacity of multiple access channel (when T_1 and T_2 both transmit to the relay) and the broadcast channel (when the relay transmits to both T_1 and T_2). For the full-duplex restricted two-way channel (all nodes can transmit and receive at the same time) achievable rate regions are derived in [22] for amplify and forward, decode and forward, and compress and forward relay strategies, which do not match with the best known upper bounds [25] and the capacity region for this case is unknown, in general. To the best of our knowledge, capacity region is unknown for general two-way channel (where T_1 and T_2 are allowed to jointly encode their information) for both half-duplex and full-duplex nodes.

In previous work on the two-way relay channel, only one relay node was considered. In this paper we consider a two-way channel with multiple relay nodes, with half duplex constraint on each node. We analyze the scaling behavior of the capacity of two-way relay channel as the number of relay nodes grow large. We assume that two terminals T_1 and T_2 want to communicate with each other via K relay nodes. None of the relays have any data of their own and only facilitate communication between T_1 and T_2. Both T_1 and T_2 are equipped with M antennas, while all the K relays have N antennas each (no restriction on N and M). We assume that there is no direct path between T_1 and T_2 and that the relays operate in half-duplex mode. Throughout this paper we assume that both T_1 and T_2 have perfect receive channel state information (CSI) for all the channels in the two-way relay network. This could be enabled through a combination of channel reciprocity and feedback, however, we do not explore the practicalities of this assumption in this work. We consider two different assumptions about CSI at the relays. In the *coherent* case, we assume that the relays have complete knowledge of their own transmit and receive channel states, while, in the *noncoherent* case, the relays have no knowledge of the channel state.

For the coherent case, our main result is summarized as follows. Using a *max flow min cut* argument, we derive an upper bound on the maximum achievable rates for $T_1 \rightarrow T_2$ and $T_2 \rightarrow T_1$ links which is given by

$$\begin{aligned} R_{12} &\leq \frac{M}{2}\log K + \mathcal{O}(1) \\ R_{21} &\leq \frac{M}{2}\log K + \mathcal{O}(1) \\ R_{12} + R_{21} &\leq M \log K + \mathcal{O}(1) \end{aligned}$$

as $K \rightarrow \infty$, where R_{12} and R_{21} are the rate of information transfer from $T_1 \rightarrow T_2$ and $T_2 \rightarrow T_1$, respectively, for any joint encoding between T_1, T_2 and all the relays. To derive a lower bound, we consider a constructive protocol where each relay performs maximal ratio combining and maximal ratio transmission on per data stream basis. We show that this protocol achieves the same capacity scaling as the upper bound on the coherent MIMO two-way relaying capacity up to a $\mathcal{O}(1)$ term, without any cooperation between T_1 and T_2. Hence we provide a complete characterization of the capacity scaling for the MIMO two-way relay channel under consideration. The lower bound provides some practical guidance in protocol design, indicating that it is sufficient to have the terminals encode independently and the relay nodes only need simple linear processing in the form of maximum ratio combining and maximal ratio transmission on a per stream basis, at least asymptotically.

In the coherent MIMO two-way relaying, all the relays have knowledge of their own transmit and receive CSI, requiring overhead due to additional training and higher complexity at the relay. Therefore we also provide a characterization of the capacity scaling for the non-coherent case when no CSI is available at any relay. We first derive an upper bound on the maximum rate of information transfer to and from T_1 and T_2 by using cutset bound and capacity results [30]. Next, using amplify and forwarding at relays, we lower bound the non-coherent two-way relaying capacity by providing a achievable rate region. We show that the achievable rate region provided by amplify and forward strategy meets the upper bound in the high SNR regime and therefore we characterize the high-SNR capacity of non-coherent MIMO two-way relaying.

Our results, surprising it might seem, show that with MIMO two-way relaying there is a improvement in the capacity scaling by a factor of 2, compared to one way relaying [20], for both the coherent and the non-coherent case. We show that with MIMO two-way relaying, both T_1 and T_2 can simultaneously communicate with each other at a rate which is equal to the maximum rate at which T_1 can communicate to T_2 if T_2 was silent. Therefore as $K \rightarrow \infty$, MIMO two-way relaying is shown to create two interference free parallel channels, one for $T_1 \rightarrow T_2$ and another for $T_2 \rightarrow T_1$, where on each channel a rate given by the maximum possible rate at which T_1 can communicate to T_2 link if T_2 was silent (one-way communication [20]) is achievable.

Notation: The following notation is used in this paper. The superscripts $^{T}, ^{H}$ represent the transpose and transpose conjugate. $\mathbf{M}$ denotes a matrix, $\mathbf{m}$ a vector and m_i the i^{th} element of $\mathbf{m}$. For a matrix $\mathbf{M} = [\mathbf{m}_1 \ \mathbf{m}_2 \ \dots \ \mathbf{m}_n]$ by v$ec(\mathbf{M})$ we mean $[\mathbf{m}_1^T \ \mathbf{m}_2^T \ \dots \ \mathbf{m}_n^T]^T$. $\mathcal{E}_x(f(x))$ denotes the expectation of function f with respect to x, $det(\mathbf{A})$ is the determinant of matrix $\mathbf{A}$. $||\cdot||$ denotes the usual Euclidean norm of a vector. $\mathbf{I}_m$ is a $m \times m$ identity matrix. $|\mathcal{X}|$ is the cardinality of set $\mathcal{X}$. We use the usual notation for $u(x) = \mathcal{O}(v(x))$ if $|\frac{u(x)}{v(x)}|$ remains bounded, as $x \rightarrow \infty$. $x \sim \mathcal{CN}(0, \sigma)$ means x is a circularly symmetric complex Gaussian random variable with zero mean and variance σ and $x|y \sim \mathcal{CN}(0, \sigma)$ means given y, x is a circularly symmetric complex Gaussian random variable with zero mean and variance σ.

The variance of a random variable a is denoted by var(a). $\mathcal{C}^{MN}$ denotes the set of $M \times N$ matrices with complex entries. $x \xrightarrow{w.p.1} y$ denotes that random variable x converges to y with probability 1. $I(x; y)$ denotes the mutual information between x and y, $H(x)$ the entropy of x and $h(x)$ the differential entropy of x [31].

Organization: The rest of the paper is organized as follows. In Section II, we describe the MIMO two-way relaying system model, the protocol under consideration and the key assump-

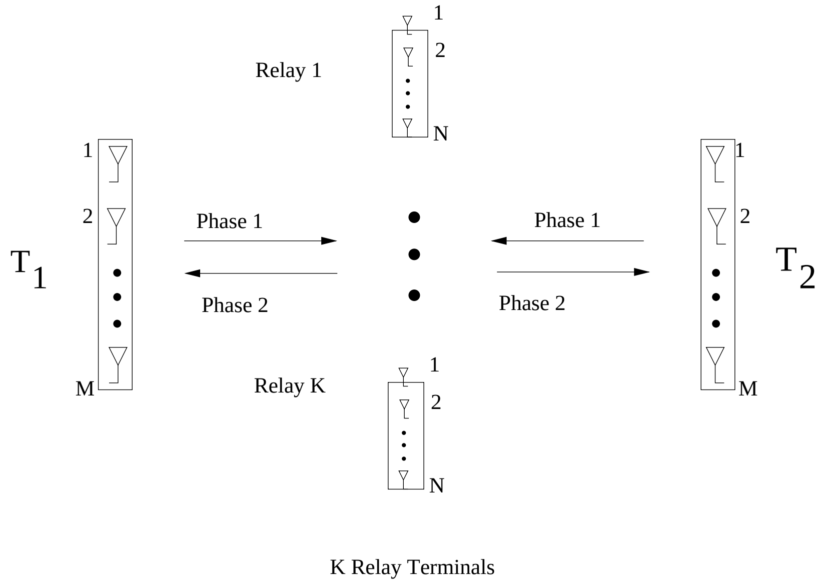


Fig. 2. Two way communication in two hops

tions. In Section III, we derive an upper bound on the asymptotic coherent MIMO two-way relaying capacity. In Section IV, by using a simple combining operation at the relays, we derive the asymptotic achievable rate region for coherent MIMO two-way relaying and show that it is possible to achieve the upper bound on the asymptotic coherent MIMO two-way relaying capacity up to an $\mathcal{O}(1)$ term. Section V summarizes and discusses the implication of the coherent MIMO two-way relaying capacity. For non-coherent MIMO two-way relaying, in Section VI-A we derive an upper bound on the achievable rate region. Section VI-B gives a result on asymptotic achievable rate region for non-coherent MIMO two-way relaying using amplify and forward strategy at relays. Final conclusions are made in Section VII.

II. MIMO Two-way Relaying System Model

In this section we describe the MIMO two-way relaying protocol under consideration, and then present the relevant signal and channel models.

A. System Model

Consider a wireless network where there are two terminals T_1 and T_2 who want to exchange information via K relays, as shown in Fig. 2. The K relays do not have any data of their

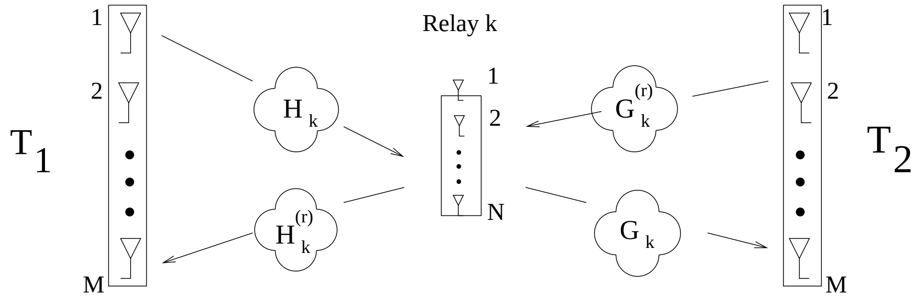


Fig. 3. Channel Model

own and only help T_1 and T_2 communicate. The K relays are assumed located randomly and independently in a area of fixed size, to ensure the signal-to-noise ratio (SNR) on the channels between each relay and T_1 and T_2 are independent. We also assume that there is no direct path between T_1 and T_2 and that they can communicate only through the K relays. This is a realistic assumption when relaying is used for coverage improvement in cellular systems, since at the cell edge the signal to noise ratio is extremely low for the direct path. In ad-hoc networks, it can be the case that two terminals want to communicate, but are out of each other's transmission range.

We assume that both the terminals T_1 and T_2 have M antennas and all the K relays have N antennas each, where $N \geq 1$ and is independent of M. We further assume that both the terminals and all the relays can operate only in half-duplex mode (cannot transmit and receive at the same time). The communication protocol is summarized from [21] as follows: In any given time slot, for the first α fraction of time, called the *transmit phase*, both T_1 and T_2 are scheduled to transmit and all the relays receive a superposition of the signals transmitted from T_1 and T_2. In the rest $(1-\alpha)$ fraction of the time slot, called the *receive phase*, all the relays are scheduled to transmit simultaneously and both the terminals receive. From here on, in this paper, we will refer to this protocol as *MIMO two-way relaying*.

B. Channel and Signal Model

To analyze the MIMO two-way relay channel, we assume that all the channels are frequency flat, slow fading channels and independently varying across integer multiples of time slots. As shown in Fig. 3, let the forward channel between T_1 and the k^{th} relay be $\mathbf{H}_k = [\mathbf{h}_{1k}\ \mathbf{h}_{2k}\ \ldots\ \mathbf{h}_{Mk}]$

and the backward channel between k^{th} relay and T_1 be $\mathbf{H}_k^{(r)} = [\mathbf{h}_{k1}^{(r)} \ \mathbf{h}_{k2}^{(r)} \ \ldots \ \mathbf{h}_{kM}^{(r)}]$. Similarly let the forward channel between k^{th} relay and T_2 be $\mathbf{G}_k = [\mathbf{g}_{k1} \ \mathbf{g}_{k2} \ \ldots \ \mathbf{g}_{kM}]$ and the backward channel between T_2 and the k^{th} relay be $\mathbf{G}_k^{(r)} = [\mathbf{g}_{1k}^{(r)} \ \mathbf{g}_{2k}^{(r)} \ \ldots \ \mathbf{g}_{Mk}^{(r)}]$. We assume that $\mathbf{H}_k, \mathbf{G}_k^{(r)} \in \mathbb{C}^{N\times M}, \mathbf{H}_k^{(r)}, \mathbf{G}_k \in \mathbb{C}^{M\times N}$ with independent and identically distributed (i.i.d.) $\mathcal{CN}(0,1)$ entries.

In the transmit phase, the $N \times 1$ received signal at the k^{th} relay is given by

$$\mathbf{r}_k = \sqrt{\frac{PE_k}{M}}\mathbf{H}_k\mathbf{x} + \sqrt{\frac{PF_k}{M}}\mathbf{G}_k^{(r)}\mathbf{u} + \mathbf{n}_k \tag{1}$$

where $\mathbf{x}$ and $\mathbf{u}$ are the $M \times 1$ signals transmitted from T_1 and T_2 to be decoded at T_2 and T_1 respectively, with $\mathcal{E}\{\mathbf{x}^H\mathbf{x}\} = \mathcal{E}\{\mathbf{u}^H\mathbf{u}\} = M$, P is the power transmitted by T_1 and T_2 and E_k and F_k are the path loss and shadowing parameters from T_1 and T_2 to the k^{th} relay, respectively. The noise $\mathbf{n}_k$ is the $N \times 1$ spatio-temporal white complex Gaussian noise independent across relays with $\mathcal{E}(\mathbf{n}_k\mathbf{n}_k^H) = \sigma^2\mathbf{I}_N$. Each relay processes its incoming signal to transmit a $N \times 1$ signal $\sqrt{\gamma_k}\mathbf{t}_k$ (with $\mathcal{E}\{\mathbf{t}_k^H\mathbf{t}_k\} = 1$) in the receive phase. We assume a power constraint of P at both T_1 and T_2 and a sum power constraint of $P_R(\sum_{k=1}^K \gamma_k = P_R)$ across all the relays. The $M \times 1$ received signal $\mathbf{v}$ and $\mathbf{y}$ at terminal T_1 and T_2 respectively in the receive phase, are given by

$$\mathbf{v} = \sum_{k=1}^K \sqrt{\gamma_k Q_k}\mathbf{H}_k^{(r)}\mathbf{t}_k + \mathbf{w} \tag{2}$$

$$\mathbf{y} = \sum_{k=1}^K \sqrt{\gamma_k P_k}\mathbf{G}_k\mathbf{t}_k + \mathbf{z} \tag{3}$$

where γ_k is the power transmitted by the k^{th} relay, Q_k, P_k are the path loss and shadowing parameters from the k^{th} relay to T_1 and T_2, respectively, while $\mathbf{w}$ and $\mathbf{z}$ are $M \times 1$ spatio-temporal white complex Gaussian noise vectors with $\mathcal{E}(\mathbf{w}\mathbf{w}^H) = \mathcal{E}(\mathbf{z}\mathbf{z}^H) = \sigma^2\mathbf{I}_M$.

Throughout this paper we assume that both T_1 and T_2 perfectly know $\{\mathbf{H}_k, \mathbf{H}_k^{(r)}, \mathbf{G}_k, \mathbf{G}_k^{(r)}\} \ \forall \, k$, $k = 1, 2, , \ldots K$ in the receive mode. To be precise, in the receive phase (i.e. when T_1 and T_2 receive signal from all the relays), T_1 and T_2 both know $\{\mathbf{H}_k, \mathbf{G}_k\}$ realization for the last transmit phase (i.e. when T_1 and T_2 transmit signal to all the relays), and $\{\mathbf{H}_k^{(r)}, \mathbf{G}_k^{(r)}\} \ \forall \, k, \ k = 1, 2, , \ldots K$ realization for the receive phase. We also assume that no transmit CSI is available at T_1 and T_2, i.e. in the transmit phase T_1 and T_2 have no information about what the realization of $\mathbf{H}_k$ and $\mathbf{G}_k$ is going to be when it transmits its signal to all the relays in the transmit phase, respectively. Capacity analysis for the case when transmit CSI is known at T_1 and T_2 is an interesting problem on its own but we do not consider it here, however, we show in later section, that the capacity of the coherent MIMO two-way relaying when transmit CSI is available to T_1 and T_2, can only differ by a $\mathcal{O}(1)$ term with compared to the no transmit CSI case, as $K \to \infty$.

In this paper we consider two different assumptions about the channel state information (CSI) at the relays. In the *coherent* case, we assume that all the relays have CSI in the transmit as well as the receive phases. In particular, in the receive phase, the k^{th} relay knows the realization of $\mathbf{H}_k$ ($\mathbf{G}_k^{(r)}$) for the receive phase and in the transmit phase it knows what the realization of $\mathbf{G}_k$ ($\mathbf{H}_k^{(r)}$) is going to be when it transmits the signal to T_1 and T_2 in the receive phase, which could be achieved through channel reciprocity or feedback. In the *non-coherent* case we assume that no CSI is available at any relay.

The path loss and shadowing effect parameters E_k and $P_k \; \forall \; k$ for the link between $T_1 \to T_2$, are assumed to be independent random variables, strictly positive, bounded and remain constant over the entire time period of interest. The randomness comes from the fact that the relays locations are chosen randomly, the strict positivity comes from the fact that the communication is happening over a fixed area, while the bounded assumption comes from the fact that none of the relays are too close to either T_1 or T_2. These are reasonable assumptions since in a random network with K nodes uniformly distributed on a fixed two-dimensional area, the minimum distance between any two nodes in the network is larger than $\frac{1}{K^{1+\delta}}$ with high probability, for any $\delta > 0$ [29]. Following the same argument, F_k and $Q_k \; \forall \; k$ are also assumed to be independent, strictly positive and bounded random variables. The results in this paper apply to independent, positive and bounded E_k, P_k, F_k, Q_k.

III. Upper Bound on the Coherent MIMO Two-Way Relaying Capacity

In this section we derive an upper bound on the capacity of coherent MIMO two-way relaying. The upper bound on the coherent MIMO two-way relaying is given by the following Theorem.

Theorem 1: If the number of relays $K \to \infty$, the capacity of coherent MIMO two-way relaying is upper bounded by

$$R_{12} \underset{w.p.1}{\leq} \frac{M}{2} \log K + \mathcal{O}(1)$$

$$R_{21} \underset{w.p.1}{\leq} \frac{M}{2} \log K + \mathcal{O}(1)$$

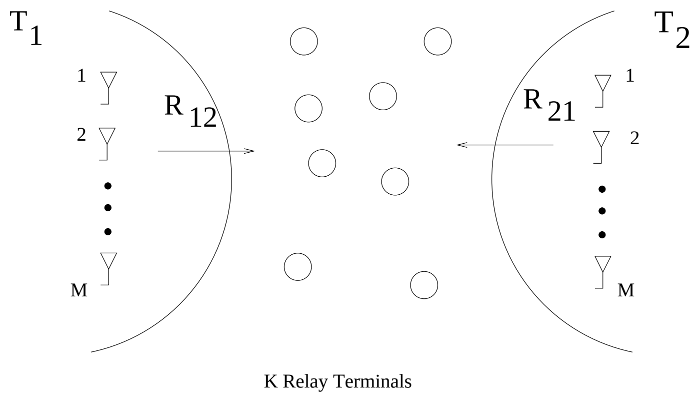


Fig. 4. Cut across both the sources: Broadcast Mode

and

$$R_{12} + R_{21} \underset{w.p.1}{\leq} M \log K + \mathcal{O}(1)$$

where R_{12} and R_{21} are the rates at which T_1 can communicate to T_2 and vice versa.

Outline of the Proof: Before presenting the formal proof, we sketch the proof to summarize the logical flow. For upper bounding the rate of information transfer from T_1 and T_2, we start by first separating T_1 and then T_2 from the network and apply the cut set bound [31] to upper bound the rate of information transfer between $T_1 \rightarrow T_2$ and $T_2 \rightarrow T_2$, respectively. Using cutset bound, we first show that the maximum rate at which information can be sent from $T_1 \rightarrow T_2$ ($T_2 \rightarrow T_1$) reliably, is upper bounded by the maximum rate of information transfer between T_1 (T_2) and $r_1, r_2, \ldots, r_K$ (broadcast cut) and also by the maximum rate of information transfer between $r_1, r_2, \ldots, r_K$ and T_2 (T_1) (multiple access cut), Fig. 4 and 5. In this work, we only consider Gaussian channels, therefore we use capacity results from [30] to upper bound the maximum rate through the broadcast cut for the case when CSI is only available at the receiver (all relays) and all the relays collaborate to decode the information. Similarly, for the multiple access cut as shown in Fig. 5, we upper bound the maximum rate at which all the $r_1, r_2, \ldots r_K$ can communicate to T_2 (T_1) by using capacity results from [30], when CSI is known both at the transmitter (all relays) and the receiver (T_1, T_2) and all the relays collaborate to transmit the information. The formal proof is as follows.

Proof: **Broadcast cut** - To prove the upper bound we make use of the cutset bound (Section 14.10 [31]). Separating the terminal T_1 from the rest of the network and applying the cutset bound on the broadcast cut as shown in Fig. 4,

$$R_{12} \leq \mathcal{E}_{\{\mathbf{H}_k,\mathbf{G}_k\}_{k=1}^K} \left\{ I(\mathbf{x};\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K,\mathbf{y}|\mathbf{t}_1,\mathbf{t}_2,\ldots\mathbf{t}_K,\mathbf{u}) \right\}. \tag{4}$$

Again applying the cutset bound while separating the terminal T_2,

$$R_{21} \leq \mathcal{E}_{\{\mathbf{H}_k^{(r)},\mathbf{G}_k^{(r)}\}_{k=1}^K} \left\{ I(\mathbf{u};\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K,\mathbf{v}|\mathbf{t}_1,\mathbf{t}_2,\ldots\mathbf{t}_K,\mathbf{x}) \right\} \tag{5}$$

for some joint distribution $p(\mathbf{x},\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_K,\mathbf{u})$, where R_{12} and R_{21} are the maximum rates at which T_1 can communicate to T_2 and T_2 can communicate to T_1 respectively, reliably. By definition of mutual information [31]

$$\begin{aligned} I(A;B,C|D) &= H(A|D) - H(A|B,C,D) \\ &= H(A|D) - H(A|C,D) + H(A|C,D) - H(A|B,C,D) \\ &= I(A;C|D) + I(A;B|C,D) \end{aligned}$$

for any A,B,C,D. and it follows that

$$\begin{aligned} I(\mathbf{x};\mathbf{r_1},\mathbf{r}_2,\ldots,\mathbf{r}_K,\mathbf{y}|\mathbf{t}_1,\mathbf{t}_2,\ldots\mathbf{t}_K,\mathbf{u}) &= I(\mathbf{x};\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K|\mathbf{t}_1,\mathbf{t}_2,\ldots\mathbf{t}_K,\mathbf{u}) \\ &\quad + I(\mathbf{x};\mathbf{y}|\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K,\ \mathbf{t}_1,\mathbf{t}_2,\ldots\mathbf{t}_K,\mathbf{u}). \end{aligned}$$

By expanding the mutual information in terms of entropy,

$$\begin{aligned} I(\mathbf{x};\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K|\mathbf{t}_1,\mathbf{t}_2,\ldots\mathbf{t}_K,\mathbf{u}) &= H(\mathbf{x}|\mathbf{t}_1,\mathbf{t}_2,\ldots\mathbf{t}_K,\mathbf{u}). \\ &\quad - H(\mathbf{x}|\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K,\mathbf{t}_1,\mathbf{t}_2,\ldots\mathbf{t}_K,\mathbf{u}) \end{aligned}$$

Since conditioning can only reduce entropy [31],

$$\begin{aligned} I(\mathbf{x};\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K|\mathbf{t}_1,\mathbf{t}_2,\ldots\mathbf{t}_K,\mathbf{u}) &\leq H(\mathbf{x}|\mathbf{u}) \\ &\quad - H(\mathbf{x}|\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K,\mathbf{t}_1,\mathbf{t}_2,\ldots\mathbf{t}_K,\mathbf{u}). \end{aligned}$$

Note that $\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_K$ is a function of $\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K$, which implies

$$\begin{aligned} I(\mathbf{x};\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K|\mathbf{t}_1,\mathbf{t}_2,\ldots\mathbf{t}_K,\mathbf{u}) &\leq H(\mathbf{x}|\mathbf{u}) \\ &\quad - H(\mathbf{x}|\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K,\mathbf{u}) \end{aligned}$$

and hence

$$I(\mathbf{x};\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K|\mathbf{t}_1,\mathbf{t}_2,\ldots\mathbf{t}_K,\mathbf{u}) \leq I(\mathbf{x};\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K|\mathbf{u}).$$

Given perfect channel knowledge at terminal T_2,

$$I(\mathbf{x};\mathbf{y}|\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K,\ \mathbf{t}_1,\mathbf{t}_2,\ldots\mathbf{t}_K,\mathbf{u}) = I(\mathbf{x},\mathbf{z})$$

where $\mathbf{z}$ is the AWGN noise. Since $\mathbf{x}$ and $\mathbf{z}$ are independent, $I(\mathbf{x},\mathbf{z}) = 0$, and therefore

$$I(\mathbf{x};\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K,\mathbf{y}|\mathbf{t}_1,\mathbf{t}_2,\ldots\mathbf{t}_K,\mathbf{u}) \leq I(\mathbf{x};\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K|\mathbf{u}).$$

Hence from (4),

$$R_{12} \leq \mathcal{E}_{\{\mathbf{H}_k,\mathbf{G}_k\}_{k=1}^K}\left\{I(\mathbf{x};\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K|\mathbf{u})\right\}.$$

Similarly we get the corresponding result for the $T_2 \to T_1$ link, by interchanging the roles of $\mathbf{x}$ and $\mathbf{u}$,

$$R_{21} \leq \mathcal{E}_{\{H_k^{(r)},\mathbf{G}_k^{(r)}\}_{k=1}^K}\left\{I(\mathbf{u};\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K|\mathbf{x})\right\}.$$

Therefore it is clear that R_{12}, R_{21} is upper bounded by the maximum information flow through the broadcast cut Fig. 4. Since we assume that the sources T_1 and T_2 transmit only for α fraction of the time in each time slot,

$$R_{12} \leq \mathcal{E}_{\{\mathbf{H}_k,\mathbf{G}_k\}_{k=1}^K}\left\{\alpha I(\mathbf{x};\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K|\mathbf{u})\right\}. \tag{6}$$

$$R_{21} \leq \mathcal{E}_{\{H_k^{(r)},\mathbf{G}_k^{(r)}\}_{k=1}^K}\left\{\alpha I(\mathbf{u};\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K|\mathbf{x})\right\}. \tag{7}$$

Expanding the mutual information in terms of differential entropy,

$$I(\mathbf{x};\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K,\mathbf{u}) = h(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K|\mathbf{u}) - h(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K|\mathbf{x},\mathbf{u}).$$

From (1),

$$\mathbf{r}_k = \sqrt{\frac{PE_k}{M}}\mathbf{H}_k\mathbf{x} + \sqrt{\frac{PF_k}{M}}\mathbf{G}_k^{(r)}\mathbf{u} + \mathbf{n}_k.$$

Now given F_k and $\mathbf{G}_k^{(r)}$ (which we assume is known at each of the relay)

$$h(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K|\mathbf{u}) = h\left(\sqrt{\frac{PE_1}{M}}\mathbf{H}_1\mathbf{x}+\mathbf{n}_1, \sqrt{\frac{PE_2}{M}}\mathbf{H}_2\mathbf{x}+\mathbf{n}_2,\ldots,\sqrt{\frac{PE_K}{M}}\mathbf{H}_K\mathbf{x}+\mathbf{n}_K|\mathbf{u}\right).$$

Since conditioning can only decrease entropy,

$$h(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_K|\mathbf{u}) \leq h\left(\sqrt{\frac{PE_1}{M}}\mathbf{H}_1\mathbf{x}+\mathbf{n}_1, \sqrt{\frac{PE_2}{M}}\mathbf{H}_2\mathbf{x}+\mathbf{n}_2,\ldots,\sqrt{\frac{PE_K}{M}}\mathbf{H}_K\mathbf{x}+\mathbf{n}_K\right).$$

With perfect knowledge of E_k, F_k and $\mathbf{H}_k, \mathbf{G}_k^{(r)}$ at each relay

$$h(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_K | \mathbf{x}, \mathbf{u}) = h(\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_K)$$

it follows that

$$\begin{aligned} I(\mathbf{x}; \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_K | \mathbf{u}) \;\; &\leq \;\; h\left(\sqrt{\frac{PE_1}{M}}\mathbf{H}_1\mathbf{x} + \mathbf{n}_1, \sqrt{\frac{PE_2}{M}}\mathbf{H}_2\mathbf{x} + \mathbf{n}_2, \ldots, \sqrt{\frac{PE_K}{M}}\mathbf{H}_K\mathbf{x} + \mathbf{n}_K\right) \\ & \quad - h(\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_K). \end{aligned}$$

Using results from [30] when CSI is only known at the receiver, this expression is upper bounded by

$$I(\mathbf{x}; \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_K | \mathbf{u}) \leq \log\det\left(\mathbf{I}_M + \frac{1}{\sigma^2}\sum_{k=1}^{K}\frac{PE_k}{M}\mathbf{H}_k^H\mathbf{H}_k\right)$$

and the maximum is achieved when $\mathbf{x}$ is circularly symmetric complex Gaussian with $\mathcal{E}(\mathbf{x}\mathbf{x}^H) = \mathbf{I}_M$. Interchanging the roles of $\mathbf{x}$ and $\mathbf{u}$ and replacing E_k with F_k and $\mathbf{H}_k$ with $\mathbf{G}_k$,

$$I(\mathbf{u}; \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_K | \mathbf{x}) \leq \log\det\left(\mathbf{I}_M + \frac{1}{\sigma^2}\sum_{k=1}^{K}\frac{PF_k}{M}\mathbf{G}_k^{(r)H}\mathbf{G}_k^{(r)}\right).$$

Therefore from (6) and (7),

$$R_{12} \leq \mathcal{E}_{\{\mathbf{H}_k, \mathbf{G}_k\}_{k=1}^K}\left\{\alpha\log\det\left(\mathbf{I}_M + \frac{1}{\sigma^2}\sum_{k=1}^{K}\frac{PE_k}{M}\mathbf{H}_k^H\mathbf{H}_k\right)\right\}.$$

$$R_{21} \leq \mathcal{E}_{\{\mathbf{H}_k^{(r)}, \mathbf{G}_k^{(r)}\}_{k=1}^K}\left\{\alpha\log\det\left(\mathbf{I}_M + \frac{1}{\sigma^2}\sum_{k=1}^{K}\frac{PF_k}{M}\mathbf{G}_k^{(r)H}\mathbf{G}_k^{(r)}\right)\right\}.$$

Since $\log\det(.)$ is a concave function and $\mathcal{E}\{\mathbf{H}_k^H\mathbf{H}_k\} = \mathcal{E}\{\mathbf{G}_k^{(r)H}\mathbf{G}_k^{(r)}\} = N\mathbf{I}_M$, applying Jensen's inequality [31]

$$R_{12} \leq \alpha M \log\left(1 + \frac{NP}{M\sigma^2}\sum_{k=1}^{K}E_k\right) \tag{8}$$

and

$$R_{21} \leq \alpha M \log\left(1 + \frac{NP}{M\sigma^2}\sum_{k=1}^{K}F_k\right). \tag{9}$$

Now consider an approximation of this upper bound in the limit $K \to \infty$. Recall that both E_k and F_k are assumed to be bounded $\forall k$, this implies var(E_k) and var(F_k) are also bounded $\forall k$ and hence

$$\sum_{k=1}^{\infty}\frac{\text{var}(E_k)}{k^2} \leq \infty$$

and

$$\sum_{k=1}^{\infty}\frac{\text{var}(F_k)}{k^2}\leq\infty.$$

Since the above sum is bounded for both E_k and F_k, from (Theorem $1.8D$ [32])

$$\sum_{k=1}^{K}\frac{E_k}{K}-\sum_{k=1}^{K}\frac{\mathcal{E}(E_k)}{K}\xrightarrow{w.p.1}0,\quad \sum_{k=1}^{K}\frac{F_k}{K}-\sum_{k=1}^{\infty}\frac{\mathcal{E}(F_k)}{K}\xrightarrow{w.p.1}0.$$

Defining $\mu_1=\sum_{k=1}^{K}\frac{\mathcal{E}(E_k)}{K}$ and $\mu_2=\sum_{k=1}^{K}\frac{\mathcal{E}(F_k)}{K}$, and noting that since $\log(.)$ is a continuous function, using (Theorem 1.7 [32])

$$R_{12}\underset{w.p.1}{\leq}\alpha M\log\left(\frac{KNP\mu_1}{M\sigma^2}\right)$$

and

$$R_{21}\underset{w.p.1}{\leq}\alpha M\log\left(\frac{KNP\mu_2}{M\sigma^2}\right).$$

Since $\mathcal{E}(E_k),\mathcal{E}(F_k)$ are bounded $\forall k$ μ_1 and μ_2 are finite, M,N are finite integers and $\frac{P}{\sigma^2}$ is finite, as $K\rightarrow\infty$

$$R_{12}\underset{w.p.1}{\leq}\alpha M\log(K)+\mathcal{O}(1) \tag{10}$$

and

$$R_{21}\underset{w.p.1}{\leq}\alpha M\log(K)+\mathcal{O}(1). \tag{11}$$

Multiple access cut - Again by using the cutset bound, we bound the maximum rate of information transfer R_{12} (R_{21}) from $T_1\rightarrow T_2$ ($T_1\rightarrow T_2$) by the maximum rate of information transfer across the multiple access cut as shown in Fig. 5. Using cutset bound, R_{12} and R_{21} are bounded by

$$R_{12}\leq\mathcal{E}_{\{\mathbf{H}_k\mathbf{G}_k\}_{k=1}^{K}}I(\mathbf{x},\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_K;\mathbf{y}|\mathbf{u}) \tag{12}$$

$$R_{21}\leq\mathcal{E}_{\{\mathbf{H}_k^{(r)}\mathbf{G}_k^{(r)}\}_{k=1}^{K}}I(\mathbf{u},\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_K;\mathbf{v}|\mathbf{x}) \tag{13}$$

Now,

$$\begin{aligned}I(\mathbf{x},\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_K;\mathbf{y}|\mathbf{u}) &= I(\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_K;\mathbf{y}|\mathbf{u})\\ &\quad +I(\mathbf{x};\mathbf{y}|\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_K,\mathbf{u})\\ &= H(\mathbf{y}|\mathbf{u})-H(\mathbf{y}|\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_K,\mathbf{u})\\ &\quad +H(\mathbf{y}|\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_K,\mathbf{u})-H(\mathbf{y}|\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_K,\mathbf{x},\mathbf{u})\end{aligned}$$

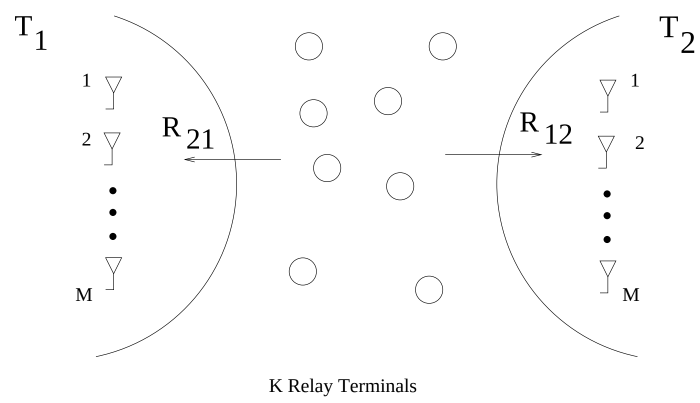


Fig. 5. Cut across both the sources: Multiple access Mode

Note that

$$H(\mathbf{y}|\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K, \mathbf{x}, \mathbf{u}) = H(\mathbf{y}|\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K, \mathbf{u}) = H(\mathbf{y}|\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K).$$

Therefore

$$I(\mathbf{x}, \mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K; \mathbf{y}|\mathbf{u}) = H(\mathbf{y}|\mathbf{u}) - H(\mathbf{y}|\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K)$$

Since conditioning can only reduce entropy,

$$I(\mathbf{x}, \mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K; \mathbf{y}|\mathbf{u}) \leq H(\mathbf{y}) - H(\mathbf{y}|\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K)$$

$$I(\mathbf{x}, \mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K; \mathbf{y}|\mathbf{u}) \leq I(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K, \mathbf{y})$$

Hence from (12),

$$R_{12} \leq \mathcal{E}_{\{\mathbf{H}_k \mathbf{G}_k\}_{k=1}^K} I(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K; \mathbf{y}) \quad (14)$$

Following similar steps we can also bound R_{21} as,

$$R_{21} \leq \mathcal{E}_{\{\mathbf{H}_k^{(r)} \mathbf{G}_k^{(r)}\}_{k=1}^K} I(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K; \mathbf{v}) \quad (15)$$

Clearly R_{12}, R_{21} are bounded by the maximum rate of information across the multiple access cut Fig. 5. Recall from (3) that the received signal $\mathbf{y}$ is given by

$$\mathbf{y} = \sum_{k=1}^{K} \sqrt{\gamma_k P_k} \mathbf{G}_k \mathbf{t}_k + \mathbf{z}$$

Note that

$$I(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K; \mathbf{y}) = I\left(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K; \frac{\mathbf{y}}{\sqrt{K}}\right).$$

Therefore dividing $\mathbf{y}$ by $\sqrt{K}$, we get

$$\frac{\mathbf{y}}{\sqrt{K}} = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} \sqrt{\gamma_k P_k} \mathbf{G}_k \mathbf{t}_k + \frac{\mathbf{z}}{\sqrt{K}}$$

This can also be written as

$$\frac{\mathbf{y}}{\sqrt{K}} = \underbrace{\frac{1}{\sqrt{K}} \left[\sqrt{P_1}\mathbf{G}_1 \; \sqrt{P_2}\mathbf{G}_2 \; \dots \; \sqrt{P_K}\mathbf{G}_K \right]}_{\Phi} [\sqrt{\gamma_1}\mathbf{t}_1 \sqrt{\gamma_2}\mathbf{t}_2 \dots \sqrt{\gamma_K}\mathbf{t}_K]^T + \frac{\mathbf{z}}{\sqrt{K}}.$$

Note that Φ is a $M \times NK$ matrix. Now assuming that all the relays know $\mathbf{G}_k, \; \forall k$ (allowing cooperation among all relays), with total power available across all relays bounded by P_R, we have from [30],

$$I\left(\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_K; \frac{\mathbf{y}}{\sqrt{K}}\right) \leq \sum_{l=1}^{\min\{NK,M\}} \max\left\{0, \log\left(\frac{K\lambda_l \nu}{\sigma^2}\right)\right\} \tag{16}$$

where $\lambda_l, l = 1, 2, \dots, \min\{NK, M\}$ are the eigen values of $\Phi\Phi^H$ matrix and ν is chosen such that

$$\sum_{l=1}^{\min\{NK,M\}} \max\{0, \nu - \frac{1}{\lambda_l}\} = P_R$$

By definition $\Phi\Phi^H = \frac{1}{K}\sum_{k=1}^{K} P_k G_k G_k^H$. Therefore as $K \to \infty$, from strong law of large numbers

$$\frac{1}{K}\sum_{k=1}^{K} P_k G_k G_k^H \xrightarrow{w.p.1} \frac{1}{K}\sum_{k=1}^{K} \mathcal{E}\left\{P_k G_k G_k^H\right\} = \frac{1}{K}\sum_{k=1}^{K} \mathcal{E}\left\{P_k\right\} \mathcal{E}\left\{G_k G_k^H\right\} = \frac{N}{K}\sum_{k=1}^{K} \mathcal{E}\left\{P_k\right\} \mathbf{I}_M$$

since $\mathcal{E}\left\{G_k G_k^H\right\} = N\mathbf{I}_M$. Note that P_k is bounded for all $k = 1, 2, \dots, K$ which implies $\frac{1}{K}\sum_{k=1}^{K} \mathcal{E}\left\{P_k\right\}$ is finite. Therefore defining $\frac{1}{K}\sum_{k=1}^{K} \mathcal{E}\{P_k\} = \rho$, it follows that

$$\lambda_i = N\rho \;\; \forall \; i = 1, 2, \dots M.$$

which implies

$$\nu = \left(\frac{P_R}{M} + \frac{1}{N\rho}\right)$$

and from capacity expression (16),

$$I\left(\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_K; \frac{\mathbf{y}}{\sqrt{K}}\right) \underset{w.p.1}{\leq} \sum_{l=1}^{M} \log\left(\frac{KN\rho}{\sigma^2}\left(\frac{P_R}{M} + \frac{1}{N\rho}\right)\right)$$

Since M, N, P_R, σ^2 and ρ are all finite, as $K \to \infty$,

$$I\left(\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_K; \frac{\mathbf{y}}{\sqrt{K}}\right) \underset{w.p.1}{\leq} M \log K + \mathcal{O}(1)$$

Moreover, since the relays transmit only for $(1-\alpha)$ fraction of time in any given slot,

$$R_{12} \underset{w.p.1}{\leq} (1-\alpha) I\left(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K; \frac{\mathbf{y}}{\sqrt{K}}\right) \leq (1-\alpha) M \log K + \mathcal{O}(1) \tag{17}$$

Similarly, we can get a bound for R_{21} by using (2),

$$R_{21} \underset{w.p.1}{\leq} (1-\alpha) I\left(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K; \frac{\mathbf{v}}{\sqrt{K}}\right) \leq (1-\alpha) M \log K + \mathcal{O}(1) \tag{18}$$

Combining (10), (11), (17) and (18)

$$R_{12} \underset{w.p.1}{\leq} \min\{\alpha, 1-\alpha\} M \log K + \mathcal{O}(1)$$

$$R_{21} \underset{w.p.1}{\leq} \min\{\alpha, 1-\alpha\} M \log K + \mathcal{O}(1)$$

Since $\alpha \in [0,1]$, $\min\{\alpha, 1-\alpha\} \leq \frac{1}{2}$, therefore

$$R_{12} \underset{w.p.1}{\leq} \frac{M}{2} \log K + \mathcal{O}(1)$$

$$R_{21} \underset{w.p.1}{\leq} \frac{M}{2} \log K + \mathcal{O}(1)$$

and trivially

$$R_{12} + R_{21} \underset{w.p.1}{\leq} M \log K + \mathcal{O}(1)$$

■

Discussion: In Theorem 1, we obtained upper bounds on R_{12} and R_{21} by using cutset bound on the broadcast cut (Fig. 4) and the multiple access cut (Fig. 5). For the broadcast cut, the upper bound corresponds to the case when the transmitter T_1 or T_2 has no CSI while all the relays collaborate to decode the message sent by T_1 or T_2 with perfect CSI, while the upper bound in the multiple access cut corresponds to the case when all the relays collaborate to transmit data to T_1 or T_2 using all their NK antennas with transmit CSI available at all relays. For Gaussian MIMO channel, capacity is known for both these scenarios [30] and hence serves as an upper bound for our MIMO two-way relaying protocol. An important observation to make is that the upper bound obtained in Theorem 1 is for any arbitrary α and not for any fixed α. Therefore, by using any particular scheme with a fixed α, one can do no better than the upper bound provided by Theorem 1.

To achieve this upper bound, consider two well known protocols that could be used at each of the relay, namely: decode and forward and amplify and forward. If a decode and forward protocol is used at each of the relay, then to decode T_1's message at any relay, T_2's message will be treated as interference and vice-versa. Therefore the achievable rate region in this case would be same as that of the achievable rate region for multiple access channel [31], which is strictly less that what is given by the upper bound derived in Theorem 1. Therefore one cannot hope to achieve the upper bound given by Theorem 1 by using decode and forward strategy at each of the relays, in case upper bound provided by Theorem 1 is tight.

If an amplify and forward protocol is used, then each relay amplifies the incoming signal, applies some linear processing using CSI, and sends it to T_1 and T_2 in the receive phase. Noting the fact that each terminal T_1 and T_2 knows what it transmits (i.e. T_1 knows $\mathbf{x}$ and T_2 knows $\mathbf{u}$), therefore with perfect receive CSI can cancel the self interference their own signals generate, employing an amplify and forward protocol at each relay looks to be a better option than decode and forward to achieve the upper bound.

A priori its not clear that even employing amplify and forward protocol at each relay will achieve the upper bound given by Theorem 1, but in the next section we are able to show that by using a simple amplify and forwarding protocol at each relay and without any cooperation between the relays or the terminals T_1 and T_2, it is possible to achieve the upper bound up to a $\mathcal{O}(1)$ factor, as $K \to \infty$.

IV. Lower Bound on Coherent MIMO Two-Way Relaying Capacity

In this section we provide a lower bound on the capacity of the coherent MIMO two-way relaying as $K \to \infty$. We show that the lower bound and upper bound (Theorem 1) are the same up to a $\mathcal{O}(1)$ term as $K \to \infty$.

A. Relaying Protocol and Capacity Analysis

We consider the following transmission strategy to obtain a lower bound on the coherent MIMO two-way relaying. Both T_1 and T_2 transmit data in spatial multiplexing mode i.e. both terminals transmit M independent data streams from their M antennas. We represent the M data streams going from T_1 to T_2 by $x_1, x_2, \dots x_M$ and the M data streams going from T_2 to T_1 by $u_1, u_2, \dots u_M$, where $\mathbf{x} = [x_1, x_2, \dots x_M]^T$ and $\mathbf{u} = [u_1, u_2, \dots u_M]^T$. Each data stream x_i and u_i $\forall\ i = 1, 2, \dots M$ is generated using random Gaussian codebook with $\mathcal{E}\{\mathbf{x}^H\mathbf{x}\} = \mathcal{E}\{\mathbf{x}^H\mathbf{x}\} = M$ (to meet the power constraint). Moreover, we fix $\alpha = \frac{1}{2}$ i.e. T_1 and T_2 transmit and receive for same amount of time. We partition the set of relays into $2M$ sets, with each set (of size $K/2M$) being responsible for one of the $2M$ data streams. We denote by $\mathcal{X}_l$ and $\mathcal{U}_l$, the set of relays assigned to data stream x_l and u_l, $l = 1, 2, \ldots, M$, respectively. Each relay belonging to the set $\mathcal{X}_l$ or $\mathcal{U}_l$ does MRC on the incoming signal and MRT while transmitting the signal, for data stream x_l and u_l, respectively. Finally we assume that at each terminal (T_1 or T_2), the i^{th} data stream is independently decoded by the i^{th} receive antenna and we do not consider any cooperation between the terminals T_1 and T_2.

The main idea behind this protocol is the following. Each relay in the network is assigned to one out of the total $2M$ data streams, for which it performs MRC while receiving the signal and MRT while transmitting the signal. Since there are K relays in the network, each data stream is served by $K/2M$ relays. As $K \to \infty$, more and more relays help each data stream in a coherent fashion and provides with an approximate SNR of $\log K$ for each data stream. This protocol is similar to one given by [19], [20]. With the help of this protocol we can prove the following theorem on the achievable rates for coherent MIMO two-way relaying.

Theorem 2: For coherent MIMO two-way relaying between T_1 and T_2, each equipped with M antennas, via K relays (each with N fixed number of antennas), the following set of rates are achievable

$$\begin{aligned} R_{12} &= \frac{M}{2}\log(K) + \mathcal{O}(1) \\ R_{21} &= \frac{M}{2}\log(K) + \mathcal{O}(1) \\ R_{12} + R_{21} &= M\log(K) + \mathcal{O}(1) \end{aligned}$$

as $K \to \infty$, with no cooperation required between T_1 and T_2.

Proof: Using relaying protocol as described above, from (1), the received signal at the k^{th} and the m^{th} relay, respectively, is given by

$$\mathbf{r}_k = \sqrt{\frac{PE_k}{M}}\sum_{j=1}^{M}\mathbf{h}_{jk}x_j + \sqrt{\frac{PF_k}{M}}\sum_{j=1}^{M}\mathbf{g}_{jk}^{(r)}u_j + \mathbf{n}_k$$

and

$$\mathbf{r}_m = \sqrt{\frac{PE_m}{M}}\sum_{j=1}^{M}\mathbf{h}_{jm}x_j + \sqrt{\frac{PF_m}{M}}\sum_{j=1}^{M}\mathbf{g}_{jm}^{(r)}u_j + \mathbf{n}_m.$$

Let us assume that the relay k is assigned to data stream x_i ($k \in \mathcal{X}_i$) and relay m is assigned to data stream u_i ($m \in \mathcal{U}_i$).

Employing MRC on the received signal for x_i and u_i at relay k and m, the maximal ratio combined signal at relay k and m respectively, is given by

$$p_k \triangleq \mathbf{h}_{ik}^H \mathbf{r}_k = \sqrt{\frac{PE_k}{M}} \left(||\mathbf{h}_{ik}||^2 x_i + \sum_{j=1, j \neq i}^{M} \mathbf{h}_{ik}^H \mathbf{h}_{jk} x_j \right) + \sqrt{\frac{PF_k}{M}} \sum_{j=1}^{M} \mathbf{h}_{ik}^H \mathbf{g}_{jk}^{(r)} x_j + \mathbf{h}_{ik}^H \mathbf{n}_k$$

and

$$q_m \triangleq \mathbf{g}_{im}^{(r)H} \mathbf{r}_m = \sqrt{\frac{PF_m}{M}} \left(||\mathbf{g}_{im}^{(r)}||^2 u_i + \sum_{j=1, j \neq i}^{M} \mathbf{g}_{im}^{(r)H} \mathbf{g}_{jm}^{(r)} u_j \right) + \sqrt{\frac{PE_m}{M}} \sum_{j=1}^{M} \mathbf{g}_{im}^{(r)H} \mathbf{h}_{jm} x_j + \mathbf{g}_{im}^{(r)H} \mathbf{n}_k.$$

Note that

$$\mathcal{E}(|p_k|^2) = \frac{PE_k}{M}(N^2 + (M-1)N) + \frac{PF_k}{M} MN + N\sigma^2$$

and

$$\mathcal{E}(|q_m|^2) = \frac{PF_m}{M}(N^2 + (M-1)N) + \frac{PE_m}{M} MN + N\sigma^2,$$

where the averaging is over random channel, signal and noise.

For transmission each relay employs MRT for a single data stream which it is serving for the forward channel. Therefore, the transmitted signals $\mathbf{t}_k$ and $\mathbf{t}_m$ by relay k and m respectively, are given by

$$\begin{aligned} \mathbf{t}_k &= \frac{p_k}{\sqrt{\frac{PE_k N}{M}(N+M-1) + PF_k N + N\sigma^2}} \frac{\mathbf{g}_{ki}^H}{||\mathbf{g}_{ki}||} \\ \mathbf{t}_m &= \frac{q_m}{\sqrt{\frac{PF_m N}{M}(N+M-1) + PE_m N + N\sigma^2}} \frac{\mathbf{h}_{mi}^{(r)H}}{||\mathbf{h}_{mi}^{(r)}||}. \end{aligned}$$

The scaling factor is required to ensure that $\mathcal{E}\{\mathbf{t}_k^H \mathbf{t}_k\} = 1, \ \forall \ k$.

Consider the T_1 to T_2 link. First, we compute the achievable rate on this link. From (3), the received signal at the i^{th} antenna of T_2 is given by

$$y_i = \sum_{k=1}^{K} \sqrt{\gamma_k P_k} \mathbf{g}_{ki} \mathbf{t}_k + z_i. \tag{19}$$

Substituting $\mathbf{t}_k$ in (19)

$$y_i = \sum_{k\in\mathcal{X}_i} \frac{\sqrt{\gamma_k P_k}||\mathbf{g}_{ki}||\left(\sqrt{\frac{PE_k}{M}}\left(||\mathbf{h}_{ik}||^2 x_i + \sum_{j=1,j\neq i}^{M}\mathbf{h}_{ik}^H\mathbf{h}_{jk}x_j\right) + \sqrt{\frac{PF_k}{M}}\sum_{j=1,j\neq i}^{M}\mathbf{h}_{ik}^H\mathbf{g}_{jk}^{(r)}u_j + \mathbf{h}_{ik}^H\mathbf{n}_k\right)}{\sqrt{\frac{PE_kN}{M}(N+M-1)+PF_kN+N\sigma^2}}$$

$$+ \sum_{m=1,m\neq i}^{M}\left(\sum_{k\in\mathcal{X}_m}\frac{\sqrt{\gamma_k P_k}\mathbf{g}_{ki}\mathbf{g}_{km}^H\sqrt{\frac{PE_k}{M}}\left(||\mathbf{h}_{mk}||^2 x_m + \sum_{j=1,j\neq m}^{M}\mathbf{h}_{mk}^H\mathbf{h}_{jk}x_j\right)}{||\mathbf{g}_{km}||\sqrt{\frac{PE_kN}{M}(N+M-1)+PF_kN+N\sigma^2}}\right)$$

$$+ \sum_{m=1,m\neq i}^{M}\left(\sum_{k\in\mathcal{X}_m}\frac{\sqrt{\gamma_k P_k}\mathbf{g}_{ki}\mathbf{g}_{km}^H\left(\sqrt{\frac{PF_k}{M}}\sum_{j=1}^{M}\mathbf{h}_{mk}\mathbf{g}_{jk}^{(r)}u_j + \mathbf{h}_{km}^H\mathbf{n}_k\right)}{||\mathbf{g}_{km}||\sqrt{\frac{PE_kN}{M}(N+M-1)+PF_kN+N\sigma^2}}\right)$$

$$+ \sum_{m=1}^{M}\left(\sum_{k\in\mathcal{U}_m}\frac{\sqrt{\gamma_k P_k}\mathbf{g}_{ki}\mathbf{h}_{km}^H\sqrt{\frac{PF_k}{M}}\left(||\mathbf{g}_{mk}^{(r)}||^2 u_m + \sum_{j=1,j\neq m}^{M}\mathbf{g}_{mk}^{(r)H}\mathbf{g}_{jk}^{(r)}u_j\right)}{||\mathbf{h}_{km}^{(r)}||\sqrt{\frac{PF_kN}{M}(N+M-1)+PE_kN+N\sigma^2}}\right)$$

$$+ \sum_{m=1}^{M}\left(\sum_{k\in\mathcal{U}_m}\frac{\sqrt{\gamma_k P_k}\mathbf{g}_{ki}\mathbf{h}_{km}^H\left(\sum_{j=1}^{M}\sqrt{\frac{PE_k}{M}}g_{mk}^{(r)H}\mathbf{h}_{jk}x_j + \mathbf{g}_{mk}^H\mathbf{n}_k\right)}{||\mathbf{h}_{km}^{(r)}||\sqrt{\frac{PF_kN}{M}(N+M-1)+PE_kN+N\sigma^2}}\right) + z_i.$$

Observe that y_i consists of contribution from x_i (desired), interference $x_l,\ l\neq i$, $u_l, l = 1,2,\dots M$, forwarded noise from the relays and the receiver noise. Separating the signal, interference and the noise components, rewriting y_i

$$y_i = \mathbf{h}_i^{sig}x_i + \sum_{j=1,j\neq i}^{M}\mathbf{h}_{i,j}^{int}x_j + \sum_{j=1}^{M}\mathbf{h}_j^{selfInt}u_j + N_i.$$

An important observation is that, T_2 knows $u_i,\ i = 1,2,\dots M$ and therefore with knowledge of $\mathbf{h}_{ki}, \mathbf{h}_{ki}^{(r)}, \mathbf{g}_{ki}$ and $\mathbf{g}_{ik}^{(r)}$ $\forall k$, it can cancel the $\sum_{j=1}^{M}\mathbf{h}_j^{selfInt}u_j$ term from the received signal at i^{th} receive antenna. Removing the self interference term from y_i

$$y_i' = \mathbf{h}_i^{sig}x_i + \sum_{j=1,j\neq i}^{M}\mathbf{h}_{i,j}^{int}x_j + N_i$$

where the signal contribution is given by

$$\mathbf{h}_i^{sig} = \sum_{k\in\mathcal{X}_i}\sqrt{\gamma_k}a_{ki} + \sum_{m=1,m\neq i}^{M}\sum_{k\in\mathcal{X}_m}\sqrt{\gamma_k}b_{k,i,m} + \sum_{m=1}^{M}\sum_{k\in\mathcal{U}_m}\sqrt{\gamma_k}c_{k,i,m}$$

where

$$a_{ki} = \frac{\sqrt{PP_k}\left(\sqrt{\frac{E_k}{M}}||\mathbf{g}_{ki}||||\mathbf{h}_{ik}||^2\right)}{\sqrt{\frac{PE_kN}{M}(N+M-1)+PF_kN+N\sigma^2}}$$

$$b_{k,i,m} = \frac{\sqrt{PP_k}\left(\sqrt{\frac{E_k}{M}}\mathbf{g}_{ki}\mathbf{g}_{km}^H\mathbf{h}_{mk}^H\mathbf{h}_{ik}\right)}{||\mathbf{g}_{km}||\sqrt{\frac{PE_kN}{M}(N+M-1)+PF_kN+N\sigma^2}}$$

$$c_{k,i,m} = \frac{\sqrt{PP_k}\left(\sqrt{\frac{E_k}{M}}\mathbf{g}_{ki}\mathbf{h}_{km}^{(r)H}\mathbf{g}_{mk}^{(r)}\mathbf{h}_{ik}\right)}{||\mathbf{h}_{km}^{(r)}||\sqrt{\frac{PF_kN}{M}(N+M-1)+PE_kN+N\sigma^2}}$$

the interference contribution is given by

$$\mathbf{h}_{ij}^{int} = \sum_{k\in\mathcal{X}_i}\sqrt{\gamma_k}d_{kij} + \sum_{k\in\mathcal{X}_j}\sqrt{\gamma_k}e_{kij} + \sum_{m=1,m\neq j}^{M}\sum_{k\in\mathcal{X}_m}\sqrt{\gamma_k}f_{k,i,j,m} + \sum_{m=1}^{M}\sum_{k\in\mathcal{U}_m}\sqrt{\gamma_k}l_{k,i,j,m}$$

where

$$d_{kij} = \frac{\sqrt{PP_k}\left(\sqrt{\frac{E_k}{M}}||g_{ki}||\mathbf{h}_{ik}^H\mathbf{h}_{jk}\right)}{\sqrt{\frac{PE_kN}{M}(N+M-1)+PF_kN+N\sigma^2}}$$

$$e_{k,i,j} = \frac{\sqrt{PP_k}\left(\sqrt{\frac{E_k}{M}}\mathbf{g}_{ki}\mathbf{g}_{kj}^H||\mathbf{h}_{mk}||^2\right)}{||\mathbf{g}_{kj}||\sqrt{\frac{PE_kN}{M}(N+M-1)+\frac{PF_k}{M}M+N\sigma^2}}$$

$$f_{k,i,j,m} = \frac{\sqrt{PP_k}\left(\sqrt{\frac{E_k}{M}}\mathbf{g}_{km}^H\mathbf{h}_{mk}^H\mathbf{h}_{jk}^{(r)}\mathbf{h}_{ik}\right)}{||\mathbf{g}_{km}^{(r)}||\sqrt{\frac{PE_kN}{M}(N+M-1)+PF_kN+N\sigma^2}}$$

$$l_{k,i,j,m} = \frac{\sqrt{PP_k}\left(\sqrt{\frac{E_k}{M}}\mathbf{g}_{ki}^H\mathbf{h}_{km}^{(r)}\mathbf{g}_{mk}^H\mathbf{h}_{jk}\right)}{||\mathbf{h}_{km}^{(r)}||\sqrt{\frac{PF_kN}{M}(N+M-1)+PE_kN+N\sigma^2}}$$

and the noise contribution is given by

$$N_i = \sum_{k\in\mathcal{X}_i}\sqrt{\gamma_k}n_{ki}^{(1)} + \sum_{m=1,m\neq i}^{M}\left(\sum_{k\in\mathcal{X}_i}\sqrt{\gamma_k}n_{kim}^{(2)}\right) + \sum_{m=1}^{M}\left(\sum_{k\in\mathcal{U}_m}\sqrt{\gamma_k}n_{kim}^{(3)}\right) + z_i$$

where

$$n_{ki}^{(1)}|\{E_k,P_k,F_k,\mathbf{H}_k,\mathbf{G}_k\}_{k=1}^K \sim \mathcal{CN}(0,\alpha_{k,i}\sigma^2)$$

$$n_{ki}^{(2)}|\{E_k,P_k,F_k,\mathbf{H}_k,\mathbf{G}_k\}_{k=1}^K \sim \mathcal{CN}(0,\beta_{k,i,m}\sigma^2)$$

$$n_{ki}^{(3)}|\{E_k,P_k,F_k,\mathbf{H}_k,\mathbf{G}_k,\mathbf{H}_k^{(r)},\mathbf{G}_k^{(r)}\}_{k=1}^K \sim \mathcal{CN}(0,\theta_{k,i,m}\sigma^2)$$

with

$$\alpha_{k,i} = \frac{P_k||\mathbf{g}_{ki}||^2||\mathbf{h}_{ik}||^2}{\left(\frac{PE_kN}{M}(N+M-1) + PF_kN + N\sigma^2\right)}$$

$$\beta_{k,i,m} = \frac{P_k|\mathbf{g}_{ki}\mathbf{g}_{km}|^2||\mathbf{h}_{km}||^2}{||\mathbf{g}_{km}||^2\left(\frac{PE_kN}{M}(N+M-1) + PF_kN + N\sigma^2\right)}$$

$$\theta_{k,i,m} = \frac{P_k|\mathbf{g}_{ki}\mathbf{h}_{km}^{(r)}|^2||\mathbf{g}_{mk}^{(r)}||^2}{||\mathbf{h}_{km}^{(r)}||^2\left(\frac{PF_kN}{M}(N+M-1) + PE_kN + N\sigma^2\right)}.$$

Recall that each data stream $x_i,\ i = 1, 2, \ldots M$ is generated using a random Gaussian codebook. Assuming that T_2 knows $\{E_k, P_k, F_k, \mathbf{H}_k, \mathbf{G}_k, \mathbf{H}_k^{(r)}, \mathbf{G}_k^{(r)}\}_{k=1}^K$ perfectly, the noise contribution N_i and $\sum_{j=1,j\neq i}^M \mathbf{h}_{ij}^{int}x_j$ are circularly symmetric Gaussian and therefore noise plus interference contribution is circularly symmetric complex Gaussian and hence with independent decoding of $x_i,\ i = 1, 2, \ldots M$ at the i^{th} antenna of T_2, the achievable rate for the $T_1 \to T_2$ link is given by [30]

$$R_{12} = \sum_{i=1}^M \mathcal{E}_{\{\mathbf{H}_k,\mathbf{G}_k\}_{k=1}^K} I_i \tag{20}$$

where I_i is given by

$$I_i = \frac{1}{2}\log\left(1 + \frac{|\mathbf{h}_i^{sig}|^2}{\sum_{j=1,j\neq i}^M |\mathbf{h}_i^{int}|^2 + \sigma^2\left(1 + \Delta_{kim}\right)}\right) \tag{21}$$

and

$$\Delta_{kim} = \sum_{k\in\mathcal{X}_i}\gamma_k\alpha_{ki} + \sum_{m=1,m\neq i}^M\sum_{k\in\mathcal{X}_m}\gamma_k\beta_{kim} + \sum_{m=1}^M\sum_{k\in\mathcal{U}_m}\gamma_k\theta_{kim}.$$

By assumption, E_k and P_k are positive and bounded $\forall k$, which implies a_{ki}, b_{kim}, c_{kim}, d_{kij}, e_{kij}, f_{kijm}, l_{kijm} and $\alpha_{ki}, \beta_{kim}, \gamma_{kim}$ are bounded $\forall i,\ j,\ k,\ m$. Hence the conditions (Theorem

$1.8D$[32]) are satisfied and hence

$$
\begin{aligned}
\sum_{k\in\mathcal{X}_i}\frac{a_{ki}}{K/2M}-\sum_{k\in\mathcal{X}_i}\frac{\mathcal{E}\{a_{ki}\}}{K/2M} &\xrightarrow{w.p.\ 1} 0\\
\sum_{k\in\mathcal{X}_m}\frac{b_{kim}}{K/2M}-\sum_{k\in\mathcal{X}_m}\frac{\mathcal{E}\{b_{kim}\}}{K/2M} &\xrightarrow{w.p.\ 1} 0\\
\sum_{k\in\mathcal{U}_m}\frac{c_{kim}}{K/2M}-\sum_{k\in\mathcal{U}_m}\frac{\mathcal{E}\{c_{kim}\}}{K/2M} &\xrightarrow{w.p.\ 1} 0\\
\sum_{k\in\mathcal{X}_i}\frac{d_{kij}}{K/2M}-\sum_{k\in\mathcal{X}_i}\frac{\mathcal{E}\{d_{kij}\}}{K/2M} &\xrightarrow{w.p.\ 1} 0\\
\sum_{k\in\mathcal{X}_j}\frac{e_{kij}}{K/2M}-\sum_{k\in\mathcal{X}_j}\frac{\mathcal{E}\{e_{kij}\}}{K/2M} &\xrightarrow{w.p.\ 1} 0\\
\sum_{k\in\mathcal{X}_m}\frac{f_{kijm}}{K/2M}-\sum_{k\in\mathcal{X}_m}\frac{\mathcal{E}\{f_{kijm}\}}{K/2M} &\xrightarrow{w.p.\ 1} 0\\
\sum_{k\in\mathcal{U}_m}\frac{l_{kijm}}{K/2M}-\sum_{k\in\mathcal{U}_m}\frac{\mathcal{E}\{l_{kijm}\}}{K/2M} &\xrightarrow{w.p.\ 1} 0\\
\sum_{k\in\mathcal{X}_i}\frac{\alpha_{ki}}{K/2M}-\sum_{k\in\mathcal{X}_i}\frac{\mathcal{E}\{\alpha_{ki}\}}{K/2M} &\xrightarrow{w.p.\ 1} 0\\
\sum_{k\in\mathcal{X}_m}\frac{\beta_{kim}}{K/2M}-\sum_{k\in\mathcal{X}_m}\frac{\mathcal{E}\{\beta_{kim}\}}{K/2M} &\xrightarrow{w.p.\ 1} 0\\
\sum_{k\in\mathcal{U}_m}\frac{\theta_{kim}}{K/2M}-\sum_{k\in\mathcal{U}_m}\frac{\mathcal{E}\{\theta_{kim}\}}{K/2M} &\xrightarrow{w.p.\ 1} 0
\end{aligned}
\tag{22}
$$

as $K\to\infty$. Also $b_{kim}, c_{kim}, d_{kij}, e_{kij}, f_{kijm}, l_{kijm}$ involve product of two independent channel coefficients, each with mean 0, which implies $\mathcal{E}\{b_{kim}\}=\mathcal{E}\{c_{kim}\}=\mathcal{E}\{d_{kij}\}=\mathcal{E}\{e_{kij}\}=\mathcal{E}\{f_{kijm}\}=\mathcal{E}\{l_{kijm}\}=0\ \forall i,j,k,m$.

Choosing $\gamma_k=\frac{P_R}{K}$ and multiplying and dividing the argument of the logarithm in (21) by $K/(2M)^2$, using (22) and applying (Theorem 1.7[32]),

$$
I_i \xrightarrow{w.p.\ 1} \frac{1}{2}\log\left(1+\frac{KPP_R\delta_i^2}{2M\sigma^2\left(2M+\tilde{\Delta}_{kim}\right)}\right)
\tag{23}
$$

as $K\to\infty$, where

$$
\delta_i=\sum_{k\in\mathcal{X}_i}\frac{\mathcal{E}\{a_{ki}\}}{K/2M}
$$

$$\tilde{\Delta}_{kim} = \sum_{k\in\mathcal{X}_i} \frac{P_R\mathcal{E}\{\alpha_{ki}\}}{K/2M} + \sum_{m=1,m\neq i}^{M}\left(\sum_{k\in\mathcal{X}_m} \frac{P_R\mathcal{E}\{\beta_{kim}\}}{K/2M}\right) + \sum_{m=1}^{M}\left(\sum_{k\in\mathcal{U}_m} \frac{P_R\mathcal{E}\{\theta_{kim}\}}{K/2M}\right).$$

Since $P,\ P_R$, α_{ki}, β_{kim} and θ_{kim} are finite $\forall\ k, i, m$, for $K \to \infty$

$$R_{12} \underset{w.p.1}{=} \frac{M}{2}\log(K) + \mathcal{O}(1).$$

For the achievable rate analysis on the $T_2 \to T_1$ link, consider the received signal v_i at the i^{th} antenna of T_1. From (2)

$$v_i = \sum_{k=1}^{K} \sqrt{\frac{Q_k}{\gamma_k}} \mathbf{h}_{ki}^{(r)} \mathbf{t}_k + w_i.$$

Compared to the received signal y_i (19) at the i^{th} antenna of T_2, v_i has random variables Q_k in place of P_k and $\mathbf{h}_{ki}^{(r)}$ in place of $\mathbf{g}_{ki}$, however the distribution of Q_k and $\mathbf{h}_{ki}^{(r)}$ is identical to that of P_k and $\mathbf{g}_{ki}$, respectively. Note that the number of relays serving u_i is the same as the number of relays serving $x_i,\ i = 1, 2, \ldots M$ and T_1 knows $x_i\ \forall i$ and therefore with channel knowledge of $\{\mathbf{H}_k, \mathbf{H}_k^{(r)}, \mathbf{G}_k, \mathbf{G}_k^{(r)}\}\ \forall k$ can also cancel the self interference. Putting all these facts together, the analysis for the achievable rate on the $T_2 \to T_1$ link is exactly the same as that of $T_1 \to T_2$ link and for the sake of brevity, we do not include it here. Hence the achievable rate for the $T_2 \to T_1$ link R_{21}, is also given by

$$R_{21} \underset{w.p.1}{=} \frac{M}{2}\log(K) + \mathcal{O}(1)$$

as $K \to \infty$. Therefore the achievable sum rate for two way relaying is given by

$$R_{12} + R_{21} \underset{w.p.1}{=} M\log(K) + \mathcal{O}(1)$$

as $K \to \infty$. ■

Discussion: Theorem 2 shows that the achievable rate in each direction $T_1 \to T_2$ or $T_2 \to T_1$ for coherent MIMO two-way relaying with MRC and MRT at each relay is given by $\frac{M}{2}\log(K) + \mathcal{O}(1)$ as $K \to \infty$. More importantly, Theorem 2 also shows that both T_1 and T_2 can simultaneously transmit at rate $\frac{M}{2}\log(K) + \mathcal{O}(1)$, without affecting each other's data rate and without requiring any cooperation between themselves. To get a better understanding of this result, consider the achievable rate for a single data stream. In our protocol, each data stream is served by $K/2M$ relays, which do MRC in the receive side and MRT in the transmit side. Consider a particular data stream, in the transmit phase (when the relays receive) if all the

$K/2M$ relays which are serving that stream are allowed to collaborate, by combining all their MRC signals it is easy to see that this stream will get a array gain of $\log{(K/2M)}$ over noise (one can neglect interference terms, because as $K \to \infty$, their effect is negligible). In our protocol, all the relays transmit the signal with MRT and we employ independent decoding of i^{th} data stream at i^{th} receive antenna, therefore effectively, our protocol allows all the $K/2M$ relays to collaborate in a distributed manner via the receive antenna which decodes that particular stream and hence we get a approximate SNR of $\log K$ for each data stream, as $K \to \infty$. The $\frac{1}{2}$ factor in the achievable rate expression is due to half-duplex assumption on nodes $\alpha = \frac{1}{2}$. Combining all the M data stream transmissions from both T_1 and T_2, our protocol of MRC and MRT at each relay provides with M parallel channels from $T_1 \to T_2$ and $T_2 \to T_1$, with rate $\frac{1}{2}\log K$ achievable on each channel. Moreover, with perfect channel knowledge, T_1 and T_2 can cancel the self interference their own transmitted signals create, which enable T_1 and T_2 to simultaneously achieve the rate of $\frac{M}{2}\log{(K)} + \mathcal{O}(1)$, without requiring any cooperation between themselves.

V. MIMO Two-Way Relaying Capacity Results

Combining the results in Sections III and Section IV, we establish the following characterization of the capacity scaling in the MIMO two-way relay channel. We summarize this theorem and provide some remarks on its implications in this section.

Theorem 3: As the number of relays grow large ($K \to \infty$), neglecting the $\mathcal{O}(1)$ term, the capacity of coherent MIMO two-way relaying is given by the convex hull of

$$\begin{aligned} R_{12} &\underset{w.p.1}{=} \frac{M}{2}\log(K) \\ R_{21} &\underset{w.p.1}{=} \frac{M}{2}\log(K) \\ R_{12} + R_{21} &\underset{w.p.1}{=} M\log(K) \end{aligned}$$

where R_1 and R_2 are the rate of information transfer between $T_1 \to T_2$ and $T_2 \to T_1$, respectively.

Proof: Follows from Theorem 1 and 2. ■

Remark 1: (Transmit CSI at T_1 and T_2) As stated before, for the case of coherent MIMO two-way capacity scaling, we do not consider the case when transmit CSI is available at T_1 or T_2, but by noting the following facts it is easy to see that even if transmit CSI is available at

T_1 or T_2, the coherent MIMO two-way relaying capacity does not change up to a $\mathcal{O}(1)$ term. From the proof of Theorem 1, it is clear that if transmit CSI is made available to T_1 or T_2, the upper bound for the multiple access cut is unchanged while the upper bound on the broadcast cut can only become loose. Since R_{12} (R_{21}) in Theorem 1 is upper bounded by the minimum of the broadcast cut upper bound and the multiple access cut upper bound, it follows immediately that the upper bound on R_{12} (R_{21}) in Theorem 1 remains unchanged when CSI is available to T_1 or T_2. Moreover, in the last section, with the help of a coherent amplify and forward protocol (without transmit CSI at T_1 or T_2) we showed that the upper bound provided by Theorem 1 is achievable up to a $\mathcal{O}(1)$ term. Therefore, it is easy to see that the capacity of coherent MIMO two-way relaying is unchanged up to a $\mathcal{O}(1)$ term, even when transmit CSI is made available to T_1 or T_2. Since in practice it is always challenging to acquire CSI, we note here that having transmit CSI at T_1 or T_2 does not help in improving the capacity, as $K \to \infty$.

Remark 2: (Timing Capacity) For a unidirectional relay channel (when no data is sent from $T_2 \to T_1$), it is well known that by using a random α (time for which T_1 transmits) the achievable rate can increase by maximum of 1 bit (called the timing capacity) [26]. In our two-way relay channel setting, at any given time, both T_1 and T_2 either transmit to relays or receive from relays simultaneously and clearly there is no timing capacity gain available. To exploit timing capacity gain one can choose a strategy in which T_1 and T_2 transmit and receive in different time slots, however, this separation will result in capacity loss by a factor of $\frac{\log K}{2}$ in each direction due to half-duplex assumption on nodes. Clearly, the capacity loss incurred by such a strategy would be much more than the 1 bit which at maximum can be gained by choosing random transmission strategies. Hence we conclude that choosing different time slots for which T_1 and T_2 transmit and receive, and randomly choosing them cannot increase the capacity for two-way relaying.

Discussion: The capacity region of the coherent MIMO two-way relaying is illustrated in Fig. 6. Combining Theorems 1 and 2, the sum capacity of coherent MIMO two-way relaying is given by $M \log (K) + \mathcal{O}(1)$, as $K \to \infty$, which is exactly double of the capacity achievable in each direction $T_1 \to T_2$ or $T_2 \to T_1$ [20]. The $\mathcal{O}(1)$ term in the upper and lower bound can in general be different and hence we characterize the exact capacity up to a $\mathcal{O}(1)$ term. An important implication of Theorem 2 is that with enough relays, using MRC and MRT at each relay on a per data stream basis, is optimal in the sense of achieving the right capacity scaling.

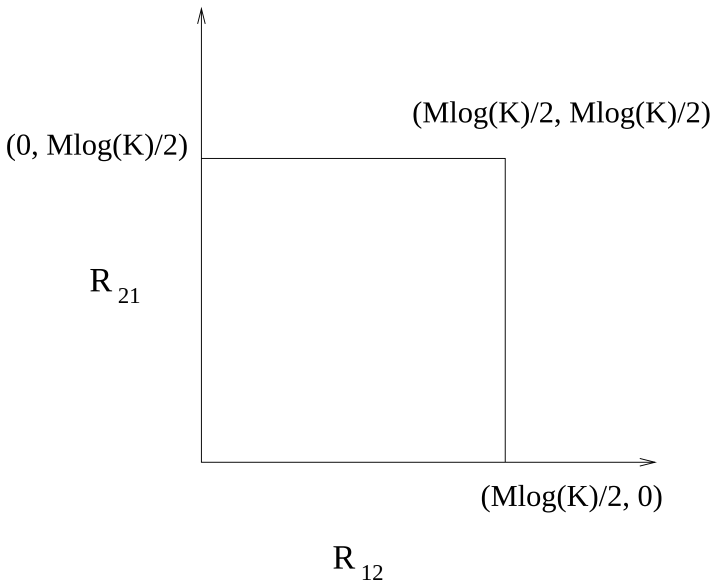


Fig. 6. Capacity of coherent MIMO two way relaying

Therefore, neglecting the $\mathcal{O}(1)$ term, what this result shows is that with two-way relaying, one can communicate at rate $\frac{M}{2}\log K$ from $T_2 \rightarrow T_1$ while simultaneously communicating at rate $\frac{M}{2}\log K$ from $T_1 \rightarrow T_2$. Since $\frac{M}{2}\log K$ is the maximum rate at which T_1 (T_2) can communicate with T_2 (T_1) even when there is T_2 is silent, two-way relaying provides with the best capacity region one can hope for, and hence is a optimal capacity achieving scheme.

Recall that we obtained the lower bound in Theorem 2 by fixing $\alpha = \frac{1}{2}$ i.e. T_1 and T_2 transmit and receive for equal amount of time. Since this lower bound is only a $\mathcal{O}(1)$ term away from the upper bound, it is clear that MIMO two-way relaying protocol which distributes equal amount of time for transmit and receive phase is optimal in achieving the right capacity scaling.

From Theorem 2, it is also clear that the upper bound on sum-capacity of MIMO two-way relaying is achievable up to a $\mathcal{O}(1)$ term without any cooperation between T_1 and T_2. This is significant since the upper bound is for some joint encoding between T_1 and T_2. This is made possible because with channel knowledge, both T_1 and T_2 are able to cancel off the self interference, their own signals generate.

Compared to the asymptotic results on the unidirectional MIMO relay channel [20], our results show that by MIMO two-way relaying one can remove the $\frac{1}{2}$ rate loss factor on the capacity, which comes from the half-duplex assumption on the terminals and relays. Therefore with MIMO two-way relaying one can achieve unidirectional full-duplex performance with half-duplex terminals.

As discussed above, coherent MIMO two-way relaying provides with capacity gains , but there is a strict requirement on CSI at both T_1, T_2 and all the relays. To perform MRC and MRT on a per stream basis for coherent MIMO two-way relaying, in the transmit and receive phase, the k^{th} relay needs to know the realization of $\mathbf{H}_k$ (or $\mathbf{G}_k^{(r)}$) of the transmit phase and the realization of $\mathbf{G}_k$ (or $\mathbf{H}_k^{(r)}$) of the receive phase, while to cancel the self-interference and to detect the incoming signal in the receive phase, both the terminals T_1 and T_2 need to know the realization of $\{\mathbf{H}_k, \mathbf{G}_k\}$ $\forall k$ for the transmit phase and $\{\mathbf{H}_k^{(r)}, \mathbf{G}_k^{(r)}\}$ $\forall k$ for the receive phase. In practice, this is a very strict and challenging requirement, but by sending training sequence and using standard channel estimation techniques, all the nodes can learn the required receive channel coefficients with good enough accuracy.

In particular, by sending training sequences from T_1 and T_2, k^{th} relay can learn $\mathbf{H}_k$ and $\mathbf{G}_k^{(r)}$ in the transmit phase (when T_1 and T_2 transmit signals to all the relays). Learning $\mathbf{G}_k$ and $\mathbf{H}_k^{(r)}$ realization at the k^{th} relay for the receive phase (when all the relays transmit and both T_1 and T_2 receive) is more challenging, however, in a time-division duplex system, by employing calibration at transmitter and receiver, the forward and backward channel can be assumed to be reciprocal in which case the realization of $\mathbf{H}_k^{(r)}$ and $\mathbf{G}_k$ for the receive phase is approximately equal to the realization of $\mathbf{H}_k^T$ and $\mathbf{G}_k^{(r)T}$ for the transmit phase. Instead if a frequency-division duplex (FDD) system is used, assuming block fading channel, $\mathbf{G}_k$ and $\mathbf{H}_k^{(r)}$ can be learnt at each relay for receive phase, by feeding back the information about $\mathbf{G}_k$ and $\mathbf{H}_k^{(r)}$ from T_1 and T_2 in the transmit phase, learnt in the last receive phase at T_1 and T_2.

To decode the incoming signal and to cancel the self interference, T_1 and T_2 needs to know the realization of $\mathbf{G}_k, \mathbf{H}_k$ of the transmit phase and the realization of $\mathbf{G}_k^{(r)}, \mathbf{H}_k^{(r)}$ of the receive phase. By sending training sequences from all the relays to T_1 and T_2, T_1 and T_2 can learn the realization of $\mathbf{H}_k^{(r)}, \mathbf{G}_k$ of the receive phase, respectively. Since at the start of receive phase each relay knows the realization of $\mathbf{G}_k, \mathbf{H}_k, \mathbf{G}_k^{(r)}, \mathbf{H}_k^{(r)}$, therefore if all the relays transmits the required channel coefficients to T_1 and T_2, both T_1 and T_2 can learn the required CSI in the receive phase.

VI. Non-Coherent MIMO Two-Way Relaying

In the last section we derived the capacity of MIMO two-way relaying when both T_1 and T_2 have receive CSI while all the relays have perfect transmit and receive CSI. We also discussed about different ways to acquire the required CSI in a practical system, but it turns out that acquiring accurate CSI in a real-time communication is a challenging problem (large overhead and complexity) and guaranteing perfect CSI is almost impossible in practice. Therefore in this section we study the capacity of MIMO two-way relaying protocol when no CSI is available at any of the relay and only receive CSI is available at T_1 and T_2. Furthermore, for this case we fix $\alpha = \frac{1}{2}$, i.e. T_1 and T_2 transmit and receive for equal amount of time (transmit phase is equal to receive phase). We call this setup as *non-coherent MIMO two-way relaying*.

For non-coherent MIMO two-way relaying, we first upper bound the possible rates from $T_1 \rightarrow T_2$ and $T_1 \rightarrow T_2$ using cut set bound for the multiple access cut and then using a simple amplify and forward strategy at each relay, we compute the achievable rates from $T_1 \rightarrow T_2$ and $T_1 \rightarrow T_2$ which are shown to match with the upper bound in the high signal to noise (SNR) regime, thereby characterizing high SNR capacity.

A. Upper Bound on Non-Coherent Two-Way Relaying Capacity

As proved in the section III, the rate of information transfer from $T_1 \rightarrow T_2$ ($T_1 \rightarrow T_2$) is upper bounded by the rate of information transfer between all-relays put together and $T_2(T_1)$ (multiple access cut). We evaluate this upper bound in the following Theorem, when CSI is not available at any of the relay.

Theorem 4: In high-SNR regime (large P_R), the rates R_{12} (R_{21}) from $T_1 \rightarrow T_2$ ($T_2 \rightarrow T_1$) for non-coherent MIMO two-way relaying are upper bounded by

$$\begin{aligned} R_{12} &\leq \frac{M}{2}\log\left(P_R\right) + \mathcal{O}(1) \\ R_{21} &\leq \frac{M}{2}\log\left(P_R\right) + \mathcal{O}(1) \\ R_{12} + R_{21} &\leq M\log\left(P_R\right) + \mathcal{O}(1). \end{aligned}$$

Proof: Using (14) and (15), we have

$$R_{12} \leq \mathcal{E}_{\{\mathbf{G}_k\}_{k=1}^{K}}\{I(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K; \mathbf{y})\}$$

and

$$R_{21} \leq \mathcal{E}_{\{\mathbf{H}_k^{(r)}\}_{k=1}^K}\{I(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K; \mathbf{v})\}.$$

which is to be maximized over joint distribution of $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K$ with no CSI at any relay. Recall from (3) that the received signal $\mathbf{y}$ is given by

$$\mathbf{y} = \sum_{k=1}^{K} \sqrt{\gamma_k P_k}\mathbf{G}_k \mathbf{t}_k + \mathbf{z}$$

Using MIMO capacity result from [30] for no transmit CSI case

$$I(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K; \mathbf{y}) \leq \log\det\left(I + \frac{\Sigma Q \Sigma^H}{\sigma^2}\right)$$

where

$$\Sigma = [\sqrt{P_1}G_1\sqrt{P_2}G_2 \ldots \sqrt{P_K}G_K]$$

and Q is the covariance matrix of

$$[\sqrt{\gamma_1}\mathbf{t}_1\sqrt{\gamma_2}\mathbf{t}_2 \ldots \sqrt{\gamma_K}\mathbf{t}_K]^T$$

with $[\mathbf{t}_1 \ \mathbf{t}_2 \ \ldots \ \mathbf{t}_K]^T$ circularly symmetric complex Gaussian and power constraint $tr(Q) \leq P_R$. As shown in [30], this expression's maximum is achieved when $Q = \frac{P_R}{NK}\mathbf{I}_{NK\times NK}$. Therefore

$$R_{12} \leq \mathcal{E}_{\{\mathbf{G}_k\}}\left\{\log\det\left(I + \frac{P_R}{NK\sigma^2}\sum_{k=1}^{K} P_k G_k G_k^H\right)\right\}$$

Using Jensen's inequality

$$R_{12} \leq \log\det\left(I + \frac{P_R}{K\sigma^2}\sum_{k=1}^{K} P_k\right).$$

Since $\frac{1}{K}\sum_{k=1}^{K} P_k$ and σ^2 are finite, for large P_R

$$R_{12} \leq M\log P_R + \mathcal{O}(1).$$

Since T_1 and T_2 transmit only for half the time ($\alpha = \frac{1}{2}$) in any given time slot

$$R_{12} \leq \frac{M}{2}\log P_R + \mathcal{O}(1)$$

Similarly, we can show

$$R_{21} \leq \frac{M}{2}\log P_R + \mathcal{O}(1)$$

and trivially

$$R_{12} + R_{21} \leq M\log P_R + \mathcal{O}(1)$$

■

B. Lower Bound on Non-coherent Two-way Relaying Capacity

In this subsection we compute achievable rates R_{12} and R_{21} for non-coherent MIMO two-way relaying for a simple amplify and forward strategy for each relay. The strategy is the following: with no CSI at any relay, each relay just normalizes the received signal to meet its power constraint and retransmits it in the receive phase. With receive CSI known at each destination T_1 (T_2), self interference generated by T_1 (T_2) can be removed from the received signal. After removing the self interference, the equivalent channel between $T_1 \rightarrow T_2$ ($T_2 \rightarrow T_1$) for non-coherent MIMO two-way relaying is given by $\sum_{k=1}^{K} \mathbf{H}_k \mathbf{G}_k$ ($\sum_{k=1}^{K} \mathbf{H}_k^{(r)} \mathbf{G}_k^{(r)}$). As $K \rightarrow \infty$ this channel is shown to behave as i.i.d. MIMO Gaussian channel and by using the results from [30], we lower bound the capacity of non-coherent MIMO two-way relaying. We show that with approximately same power used at $T_1(T_2)$ and all relays (i.e. $P \approx P_R$), the lower bound meets the upper bound in the high SNR regime (high P).

The following theorem gives expression for achievable R_{12} and R_{21} pair, when each relay uses amplify and forwarding.

Theorem 5: In high-SNR regime, the achievable rate regions for non-coherent MIMO two-way relaying using amplify and forward protocol at each relay, is given by

$$\begin{aligned} R_{12} &= \frac{M}{2} \log\left(P_R\right) + \mathcal{O}(1) \\ R_{21} &= \frac{M}{2} \log\left(P_R\right) + \mathcal{O}(1) \\ R_{12} + R_{21} &= M \log\left(P_R\right) + \mathcal{O}(1) \end{aligned}$$

Proof: Recall from (1) that the received signal at each relay is given by

$$\mathbf{r}_k = \sqrt{\frac{PE_k}{M}} \mathbf{H}_k \mathbf{x} + \sqrt{\frac{PF_k}{M}} \mathbf{G}_k^{(r)} \mathbf{u} + \mathbf{n}_k \tag{24}$$

Therefore the average received signal plus noise power at each relay is given by $N(P(E_k + F_k) + \sigma^2)$. We assume that the k^{th} relay knows the average received signal plus noise power $N(P(E_k + F_k) + \sigma^2)$ and performs the normalization $\mathbf{t}_k = \left(\frac{1}{N(P(E_k+F_k)+\sigma^2))}\right)^{\frac{1}{2}} \mathbf{r}_k$ to ensure that $\mathcal{E}\{\mathbf{t}_k^H \mathbf{t}_k\} = 1$. With this normalization, from (2) and (3), the received signal at terminal T_1 and T_2 is given by $\mathbf{v}$ and $\mathbf{y}$, respectively, where

$$\mathbf{v} = \sum_{k=1}^{K} \sqrt{\frac{\gamma_k Q_k}{N(P(E_k + F_k) + \sigma^2)}} \mathbf{H}_k^{(r)} \mathbf{r}_k + \mathbf{w}.$$

$$\mathbf{y} = \sum_{k=1}^{K} \sqrt{\frac{\gamma_k P_k}{N(P(E_k+F_k)+\sigma^2)}} \mathbf{G}_k \mathbf{r}_k + \mathbf{z}.$$

substituting for $\mathbf{r}_k$ from (1) in the above equation

$$\begin{aligned} \mathbf{y} &= \sum_{k=1}^{K} \sqrt{\frac{\gamma_k P_k E_k}{NM(P(E_k+F_k)+\sigma^2)}} \mathbf{G}_k \mathbf{H}_k \mathbf{x} \\ &+ \sum_{k=1}^{K} \sqrt{\frac{\gamma_k P_k F_k}{NM(P(E_k+F_k)+\sigma^2)}} \mathbf{G}_k \mathbf{G}_k^{(r)} \mathbf{u} \\ &+ \sum_{k=1}^{K} \sqrt{\frac{\gamma_k P_k}{N(P(E_k+F_k)+\sigma^2)}} \mathbf{G}_k \mathbf{n}_k \\ &+ \mathbf{z}. \end{aligned}$$

Since T_2 knows $\mathbf{u}$ and has perfect CSI, it can cancel the self interference. Removing the self interference from $\mathbf{y}$ and dividing both sides by $\sqrt{K}$,

$$\begin{aligned} \mathbf{y}' &= \underbrace{\frac{1}{\sqrt{K}} \sum_{k=1}^{K} \sqrt{\frac{\gamma_k P_k E_k}{NM(P(E_k+F_k)+\sigma^2)}} \mathbf{G}_k \mathbf{H}_k \, \mathbf{x}}_{\mathbf{A}} \\ &+ \underbrace{\frac{1}{\sqrt{K}} \sum_{k=1}^{K} \sqrt{\frac{\gamma_k P_k}{N(P(E_k+F_k)+\sigma^2)}} \mathbf{G}_k \mathbf{n}_k + \frac{1}{\sqrt{K}} \mathbf{z}}_{\mathbf{b}}. \end{aligned}$$

Similarly T_1 knows $\mathbf{x}$ and also has perfect CSI, therefore it can also remove the self interference. Removing the self interference from $\mathbf{v}$ and dividing both sides by $\sqrt{K}$

$$\begin{aligned} \mathbf{v}' &= \underbrace{\frac{1}{\sqrt{K}} \sum_{k=1}^{K} \sqrt{\frac{\gamma_k Q_k F_k}{NM(P(E_k+F_k)+\sigma^2)}} \mathbf{H}_k^{(r)} \mathbf{G}_k^{(r)} \, \mathbf{u}}_{\mathbf{C}}. \\ &+ \underbrace{\frac{1}{\sqrt{K}} \sum_{k=1}^{K} \sqrt{\frac{\gamma_k Q_k}{N(P(E_k+F_k)+\sigma^2)}} \mathbf{H}_k^{(r)} \mathbf{n}_k + \frac{1}{\sqrt{K}} \mathbf{w}}_{\mathbf{d}}. \end{aligned}$$

As $K \to \infty$, it can be shown that (Theorem 3 [20])

$$\mathbf{A}_{i,j} \sim \mathcal{P}\left(0, \frac{1}{K} \sum_{k=1}^{K} \mathcal{E}\left\{\frac{\gamma_k P_k E_k}{M(P(E_k+F_k)+\sigma^2)}\right\}\right)$$

$$\mathbf{C}_{i,j} \sim \mathcal{P}\left(0, \frac{1}{K}\sum_{k=1}^{K}\mathcal{E}\left\{\frac{\gamma_k Q_k F_k}{M(P(E_k+F_k)+\sigma^2)}\right\}\right)$$
$$\mathbf{b}_i \sim \mathcal{P}\left(0, \frac{\sigma^2}{K}\sum_{k=1}^{K}\mathcal{E}\left\{\frac{\gamma_k P_k}{(P(E_k+F_k)+\sigma^2)}\right\}+1\right)$$
$$\mathbf{d}_i \sim \mathcal{P}\left(0, \frac{\sigma^2}{K}\sum_{k=1}^{K}\mathcal{E}\left\{\frac{\gamma_k Q_k}{(P(E_k+F_k)+\sigma^2)}\right\}+1\right)$$

and

$$R_{\mathbf{A}} \xrightarrow{w.p.1} \frac{1}{K}\sum_{k=1}^{K}\mathcal{E}\left\{\frac{\gamma_k P_k E_k}{M(P(E_k+F_k)+\sigma^2)}\right\}\mathbf{I}_{M^2}$$
$$R_{\mathbf{C}} \xrightarrow{w.p.1} \frac{1}{K}\sum_{k=1}^{K}\mathcal{E}\left\{\frac{\gamma_k Q_k F_k}{M(P(E_k+F_k)+\sigma^2)}\right\}\mathbf{I}_{M^2}$$

where $\mathbf{A}_{i,j}, \mathbf{C}_{i,j}$ denotes i^{th} row and j^{th} column entry of $\mathbf{A}$ and $\mathbf{C}$ respectively and $\mathbf{b}_i, \mathbf{d}_i$ denotes the i^{th} element of $\mathbf{b}$ and $\mathbf{d}$ respectively, $R_{\mathbf{A}} = \mathcal{E}_{\{\mathbf{H}_k,\mathbf{G}_k\}_{k=1}^{K}}\{\mathbf{a}\mathbf{a}^H\}$ where $\mathbf{a} = \mathrm{v}ec(\mathbf{A})$ and $R_{\mathbf{C}} = \mathcal{E}_{\{\mathbf{H}_k^{(r)},\mathbf{G}_k^{(r)}\}_{k=1}^{K}}\{\mathbf{c}\mathbf{c}^H\}$ where $\mathbf{c} = \mathrm{v}ec(\mathbf{C})$.

This shows that the channel matrices $\mathbf{A}, \mathbf{C}$ and the noise vectors $\mathbf{b}, \mathbf{d}$ are i.i.d. Gaussian, therefore using results from [30], the achievable rate R_{12} (R_{21}) of the $T_1 \rightarrow T_2$ ($T_2 \rightarrow T_1$) link for $\alpha = \frac{1}{2}$, is given by

$$R_{12} = \frac{1}{2}\mathcal{E}_{\mathbf{H}_w}\left\{\log\det\left(\mathbf{I}_M + \frac{\rho_1}{M}\mathbf{H}_w{\mathbf{H}_w}^H\right)\right\}$$

$$R_{21} = \frac{1}{2}\mathcal{E}_{\mathbf{H}_w}\left\{\log\det\left(\mathbf{I}_M + \frac{\rho_2}{M}\mathbf{H}_w{\mathbf{H}_w}^H\right)\right\}$$

where $\mathbf{H}_w$ is $M \times M$ matrix with each elements of $\mathbf{H}_w$ i.i.d. $\mathcal{CN}(0,1)$ and

$$\rho_1 = \frac{\frac{1}{K}\sum_{k=1}^{K}\mathcal{E}\left\{\frac{\gamma_k P_k E_k}{(P(E_k+F_k)+\sigma^2)}\right\}}{\frac{\sigma^2}{K}(\sum_{k=1}^{K}\mathcal{E}\left\{\frac{\gamma_k P_k}{(P(E_k+F_k)+\sigma^2)}\right\}+1)}$$
$$\rho_2 = \frac{\frac{1}{K}\sum_{k=1}^{K}\mathcal{E}\left\{\frac{\gamma_k Q_k F_k}{(P(E_k+F_k)+\sigma^2)}\right\}}{\frac{\sigma^2}{K}(\sum_{k=1}^{K}\mathcal{E}\left\{\frac{\gamma_k Q_k}{(P(E_k+F_k)+\sigma^2)}\right\}+1)}$$

Note that ρ_1 and ρ_2 are effective SNRs. Now letting E_k, P_k, F_k and Q_k to be independent and identically distributed, denoting $\eta_1 = \mathcal{E}\{Q_k F_k\} = \mathcal{E}\{E_k P_k\}\ \forall k$, $\eta_2 = \mathcal{E}\{Q_k\} = \mathcal{E}\{P_k\}\ \forall k$, $\eta = \mathcal{E}\{P(E_k+F_k)+\sigma^2\}\ \forall k$, we get

$$\rho_1 = \rho_2 = \frac{\frac{P\eta_1}{\eta}\sum_{k=1}^{K}\gamma_k}{\sigma^2(\frac{\eta_2}{\eta}\sum_{k=1}^{K}\gamma_k+1)}$$

Since the relay power is constrained by $\sum_{k=1}^{K} \gamma_k = P_R$

$$\rho_1 = \rho_2 = \frac{P P_R \eta_1}{\sigma^2 (P_R \eta_2 + \eta)}$$

Choosing $P \approx P_R$, we get

$$\rho_1 = \rho_2 \approx \frac{P_R}{\sigma^2}$$

since $E_k, F_k, P_k, Q_k \ \forall k$ are bounded. Therefore

$$R_{12} = R_{21} = \mathcal{E}_{\mathbf{H}_w} \left\{ \log\det \left(\mathbf{I}_M + \frac{P_R}{M\sigma^2} \mathbf{H}_w \mathbf{H}_w^H \right) \right\}.$$

In high SNR regime $P \approx P_R \gg 1$, from [30], it follows that

$$\begin{aligned} R_{12} &= \frac{M}{2} \log(P_R) + \mathcal{O}(1) \\ R_{21} &= \frac{M}{2} \log(P_R) + \mathcal{O}(1) \\ R_{12} + R_{21} &= M \log(P_R) + \mathcal{O}(1). \end{aligned}$$

where we have absorbed M and σ^2 in the $\mathcal{O}(1)$ term. ■

Discussion: In this section, we first obtained a upper bound on the capacity of non-coherent MIMO two-way relaying using multiple access cutset bound when CSI is only known at the receiver. Then with the help of a simple amplify and forward strategy we provide a lower bound which in the high SNR regime equals the upper bound and hence we get the capacity region of non-coherent MIMO two-way relaying. We find that, contrary to the coherent case, with non-coherent MIMO two-way relaying, as the number of relay nodes grow large the capacity expression is independent of the number of relays and no coherent combining gain (array gain) is available when there is no CSI at any relay. However, similar to the coherent case, it turns out that even in non-coherent case both T_1 and T_2 can simultaneously transmit at a rate which is equal to the maximum possible rate at which they could have transmitted when there is no data flowing in the opposite direction. Therefore we get two orthogonal channels one from $T_1 \to T_2$ and another from $T_1 \to T_2$ with rate $M \log P_R$ achievable on each link simultaneously, thereby removing the $\frac{1}{2}$ rate loss factor because of half-duplex nodes.

The lower bound provided by Theorem 5 shows that the achievable rate for non-coherent MIMO two-way relaying is same as the capacity of a point to point $M \times M$ i.i.d Gaussian channel with receive SNR P_R, with perfect CSI at receiver and no CSI at transmitter and where $\frac{1}{2}$ factor is due to the half-duplex requirement. This result is quite intuitive, since with absence of CSI at the relays, as $K \to \infty$ the equivalent channel between $T_1 \to T_2$ ($T_2 \to T_1$) converges to a $M \times M$ i.i.d Gaussian channel and therefore the results follow from [30].

Compared to (Theorem 3 [20]), this result shows that with non-coherent MIMO two-way relaying it is possible to remove the $\frac{1}{2}$ rate loss factor due to the half-duplex constraint and can achieve the same rate as promised by Theorem 3 [20] (for unidirectional communication), in each direction $T_1 \to T_2$ and $T_2 \to T_1$. This is again due to the fact that, with perfect CSI both T_1 and T_2 can cancel the self interference terms their own transmitted signals generate and hence the received signal at T_2 (T_1) when T_2 is also sending information is equivalent to the received signal at T_2 in [20], where there is no communication happening on $T_2 \to T_1$ link. Therefore there is a two-fold increase in achievable rate with non-coherent MIMO two-way relaying in comparison to [20].

VII. Conclusion

In this paper we developed capacity scaling laws for networks with MIMO two-way relaying under coherent and noncoherent assumptions. First we upper bounded the coherent MIMO two-way relaying capacity using the broadcast and multiple access cut-set bound. Then we gave an expression for the achievable rate region for the coherent MIMO two-way relaying, using coherent amplify and forwarding at each relay. The achievable rate region was shown to be a $\mathcal{O}(1)$ term away from the upper bound, as $K \to \infty$. Hence we characterized the coherent MIMO two-way relaying capacity up to a $\mathcal{O}(1)$ term. An interesting outcome of our analysis is that the achievable sum rate expression requires no cooperation between T_1 and T_2 while the upper bound is for some joint encoding between T_1 and T_2. This shows that the capacity can be achieved without any cooperation between T_1 and T_2, up to a $\mathcal{O}(1)$ term. We also showed that even if transmit CSI is available at T_1 and T_2 the capacity of coherent MIMO two-way relaying can only increase by an $\mathcal{O}(1)$ term as $K \to \infty$.

For coherent MIMO two-way relaying, there is a strict requirement that all the nodes need to know perfect CSI, which can be quite challenging and resource consuming. Therefore we also considered the case when only T_1 and T_2 have perfect receive CSI and none of the relays have any CSI, which is referred to as non-coherent MIMO two-way relaying. For this case we upper bounded the capacity region using only the multiple access cut-set bound and fixing $\alpha = \frac{1}{2}$ (i.e. $T_1,\ T_2$ and all the relays transmit for equal amount of time in each time slot). Then with the help of a simple amplify and forward protocol, we showed that in the high SNR regime by choosing similar transmit powers from at T_1 (T_2) and all relays together ($P \approx P_R$), we can achieve the upper bound upto a $\mathcal{O}(1)$ term, and hence characterize high SNR capacity of non-coherent MIMO two-way relaying.

Compared to [19], [20], our capacity scaling results for coherent and non-coherent MIMO two-way relaying shows that with MIMO two-way relaying there is a two-fold increase in the capacity than unidirectional communication with large number of relays. Hence, MIMO two-way relaying helps in improving the spectral efficiency and unidirectional full-duplex performance while using half-duplex terminals.

REFERENCES

[1] E.C. Van der Meulen, “Three terminal communication channels,” *Adv. Appl. Probab.*, vol. 3, pp. 120-154, 1971.

[2] T.M. Cover and A.A. El Gamal, “Capacity theorems for relay channels,” *IEEE Trans. on Information Theory*, vol. 25, no. 5, pp. 572-584, Sept 1979.

[3] B. Wang, J. Zhang and A. Host-Madsen, “On the capacity of MIMO relay channels, ” *IEEE Trans. on Information Theory*, vol. 51, no. 1, pp. 29- 43, Jan. 2005.

[4] A. Sendonaris, E. Erkip, B Aazhang, “User cooperation diversity. Part I. System description,” *IEEE Trans. on Communications*, vol. 51, no. 11, pp. 1927- 1938, Nov. 2003.

[5] A. Sendonaris, E. Erkip, B Aazhang, “User cooperation diversity. Part II. Implementation aspects and performance analysis,” *IEEE Trans. on Communications*, vol. 51, no. 11, pp. 1939- 1948, Nov. 2003.

[6] R.U. Nabar, H. Bölcskei, F.W. Kneubuhler, “Fading relay channels: performance limits and space-time signal design,” *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1099- 1109, Aug. 2004.

[7] J. Laneman, D.N.C. Tse and G.W. Wornell, “Cooperative diversity in wireless networks: Efficient protocols and outage behavior,” *IEEE Trans. on Information Theory*, vol. 50, no. 12, pp. 3062- 3080, Dec. 2004.

[8] S. M. Alamouti, “A simple transmit diversity technique for wireless communications ,” *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451-1458, Oct 1998.

[9] V. Tarokh, N.Seshadri and A. R. Calderbank, “Space-Time block codes for high data rate wireless communication:Performance criterion and code construction,” *IEEE Trans. on Information Theory*, vol. 44, no. 2, pp. 744-765, Mar 1998.

[10] C.K. Lo, S. Vishwanath and R.W. Heath, Jr., “ Rate bounds for MIMO relay channels using precoding,” in *Proc. IEEE GLOBECOM*, vol. 3, pp. 1172-1176, St. Louis, MO, Nov. 2005.

[11] A. Host-Madsen, J. Zhang, “Capacity bounds and power allocation for wireless relay channels,” *IEEE Trans. on Information Theory*, vol. 51, no. 6, pp. 2020- 2040, June 2005.

[12] P. Gupta, P.R. Kumar, “The capacity of wireless networks,” *IEEE Trans. on Information Theory*, vol. 46, no. 2, pp. 388-404, Mar. 2000.

[13] G. Kramer, M Gastpar and P. Gupta, “Cooperative Strategies and Capacity Theorems for Relay Networks,”*IEEE Trans. on Information Theory*, vol. 51, no. 9, pp. 3037-3063,Sept. 2005.

[14] S Vishwanath, S Jafar, S Sandhu, "Half-duplex relays: cooperative communication strategies and outer bounds," in *P*roc. International Conference on Wireless Networks, Communications and Mobile Computing, vol. 2, pp. 1455 - 1459, 13-16 June 2005.

[15] J.N. Laneman, G.W. Wornell, "Distributed Space-Time-Coded Protocols for Exploiting Cooperative Diversityin Wireless Networks,"*IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2415-2425, Oct 2003.

[16] A.H. Maden, "On the Capacity of Wireless Relaying," in *Proc. Vehicular Technology Conference, 2002. VTC 2002-Fall. 2002* IEEE 56th, vol.3, pp. 1333-1337, 2002.

[17] M. Yuksel and E. Erkip, "Cooperative Wireless Systems: A Diversity-Multiplexing Tradeoff Perspective," Submitted to IEEE Transactions on Information Theory Sep 2006, available online at *http://arxiv.org/abs/cs/0609122*.

[18] S. Wei, D.L. Goeckel, and M.C. Valenti, "Asynchronous cooperative diversity," "*IEEE Trans. on Wireless Communications*, vol. 5, no. 6, pp. 1547-1557, June 2006.

[19] M. Gastpar and M. Vetterli, "On the Capacity of Wireless Networks: The Relay Case," in *Proc. IEEE INFOCOM*, vol. 3, pp. 1577- 1586, New York June 2002.

[20] H. Bölcskei, R. U. Nabar, Ö. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. on Wireless Communications*, vol. 5, no. 6, pp. 1433-1444, June 2006.

[21] B. Rankov and A. Wittneben, "Spectral Efficient Signaling for Half-duplex Relay Channels," in *Proc. Asilomar Conference on Signals, Systems, and Computers 2005*, pp. 1066-1071, Pacific Grove, CA, Oct-Nov. 2005.

[22] B. Rankov and A. Wittneben, "Achievable Rate Regions for the Two-way Relay Channel," in *Proc. IEEE Int. Symposium on Information Theory (ISIT)*, pp. 1668-1672, Seattle, USA, July 2006.

[23] T.J Oechtering and H. Boche, "Optimal Tranmsit Strategies In Multi-Antenna Bidirectional Relaying," in *Proc. IEEE Intern. Conf. on Acoustics, Speech, and Signal Processing (ICASSP '07)*, vol. 3, pp. 145-148, Honolulu, Hawaii, USA, Apr. 2007.

[24] S. Joon Kim, P. Mitran and V. Tarokh, "Performance Bounds for Bi-Directional Coded Cooperation Protocols." available at http://arxiv.org/PS_cache/cs/pdf/0703/0703017v1.pdf.

[25] G. Kramer and S.A. Savari, "On networks of two-way channels," *Algebraic Coding Theory and Information Theory, DIMACS Workshop, Dec. 15-18, 2003*, Rutgers University, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 68, A. Ashikhmin and A. Barg, eds., pp. 133-143, available at *http://cm.bell-labs.com/who/gkr/*.

[26] G. Kramer, "Models and theory for relay channels with receive constraints," in *Proc. 42nd Annual Allerton Conf. on Commun. Control and Comp*, (Monticello, IL, USA), pp. 1312-1321, Sept. 29-Oct. 1, 2004.

[27] R. Ahlswede, N. Cai, S.-Y. R. Li and R. W. Yeung, "Network information flow," *IEEE Trans. on Information Theory*, vol. 46, no. 4, pp. 1204-1216, July 2000.

[28] S. Katti, R. Hariharan, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORs In The Air: Practical Wireless Network Coding," in *Proc. ACM SIGCOMM*, pp. 243-254, Pisa, Italy, Sept. 2006.

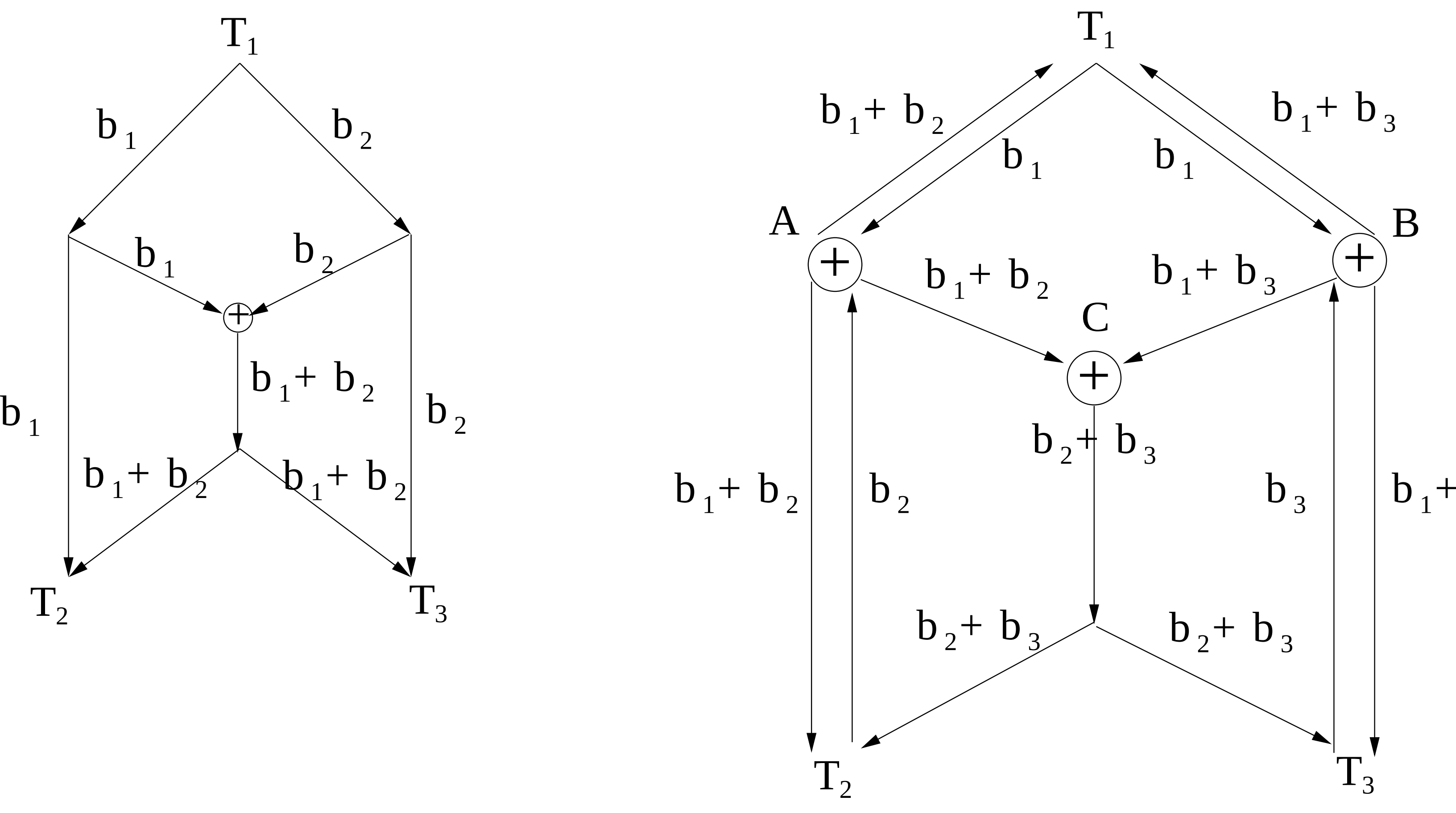
[29] A. Özgür, O. Leveque and D. Tse, "Hierarchical Cooperation Achieves Optimal Capacity Scaling in Ad Hoc Networks," in *Proc. Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sept. 2006.

[30] I.E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," *European Trans. on Telecommunications*, vol. 10, no. 6, pp. 585-595, Nov./Dec. 1999.

[31] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons 2004.

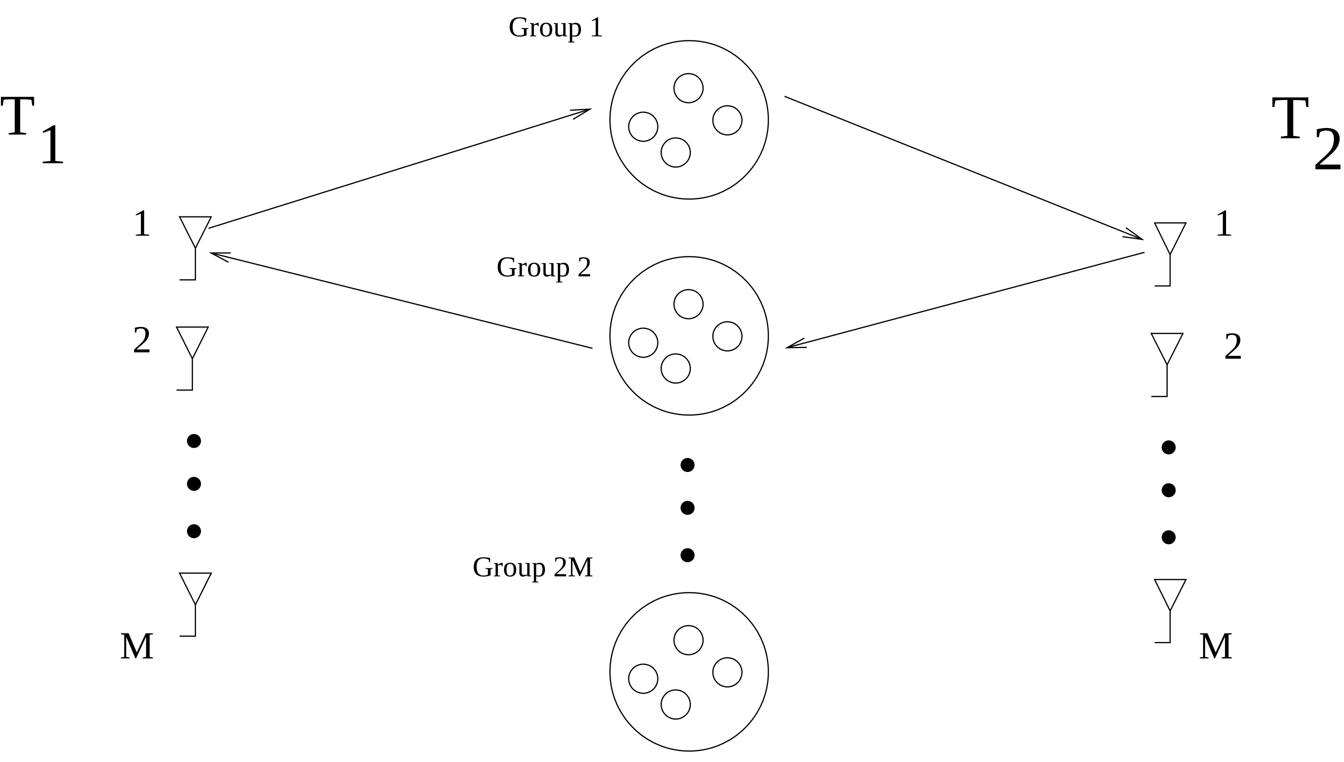
[32] R. J. Serfling, *Approximation theorems of mathematical statistics,* New York, NY, Wiley, 1980.

[33] L. Zheng and D.N.C. Tse, “Communication on the Grassmann manifold: a geometric approach to the -noncoherent multiple-antenna channel,” *IEEE Trans. on Information Theory*, vol. 48, no. 2, pp. 359-383, Feb. 2002.


T 1
b 1
b 2
b 1
b 2
b 1
b 1+ b 2
b 2
b 1+ b 2
b 1+ b 2
T 2
T 3
T 1
b 1+ b 2
b 1+ b 3
b 1
b 1
A
B
b 1+ b 2
b 1+ b 3
C
b 2+ b 3
b 1+ b 2
b 2
b 3
b 1+
b 2+ b 3
b 2+ b 3
T 2
T 3

(a)

(b)



K Relay Terminals