Grassmannian Signalling Achieves Tight Bounds on the Ergodic High-SNR Capacity of the Noncoherent MIMO Full-Duplex Relay Channel

Ramy H. Gohary, *Senior Member, IEEE*, and Halim Yanikomeroglu, *Senior Member, IEEE*

Abstract—This paper considers the ergodic noncoherent capacity of a multiple-input multiple-output frequency-flat block Rayleigh fading full duplex relay channel at high signal-to-noise ratios (SNRs). It is shown that, for these SNRs, restricting the input distribution to be isotropic on a compact Grassmann manifold maximizes an upper bound on the cut-set bound. Furthermore, it is shown that, from a degrees of freedom point of view, no relaying is necessary and Grassmannian signalling at the source achieves the upper bound within an SNR-independent gap. When the source-relay channel is sufficiently stronger than the source-destination and relay-destination channels, it is shown that, with the number of relay transmit antennas appropriately chosen, a Grassmannian decode-and-forward scheme, which is devised herein, achieves the ergodic noncoherent capacity of the relay channel within an approximation gap that goes to zero as the SNR goes to infinity. Closed-form expressions for the optimal number of relay transmit antennas indicate that this number decreases monotonically with the source transmit power.

Index Terms—Noncoherent communication, decode-and-forward relaying, Grassmannian signalling, ergodic capacity.

I. INTRODUCTION

THE availability of a relaying node to assist the communication between a transmitter-receiver pair enables significant rate and performance gains to be achieved [1]. As such, future advances in the signalling techniques and transmission protocols of wireless data networks are expected to embrace terminal relaying schemes, wherein idle wireless nodes are efficiently exploited to perform various relaying functions. The effectiveness of incorporating relays in the design of wireless networks depends on the computational power of the relaying terminals and their location, mobility and versatility [1].

Relay operation modes have been traditionally classified into either half or full duplex operation. In half duplex operation, the relay transmission and reception channels are orthogonal, which facilitates the design of these relays and renders them easier to implement in practice. In contrast, in full duplex operation, the relays transmit and receive on the same physical channel. Despite the practical challenges that this operation mode might impose, it is generally more advantageous from a rate perspective than the half duplex one. The capacity of general relay channels is not known, and only an upper bound on it, known as the cut-set bound, is available [2]. This bound is known to be tight for most of the special cases in which the capacity is known. An instance in which this is not the case is provided in [3].

Another feature of future wireless networks is that their nodes will rely on multiple antennas for transmission and reception. Systems with this feature are referred to as multiple-input multiple-output (MIMO) ones. In comparison with their single antenna counterparts, MIMO systems possess additional degrees of design freedom that enable trading off higher communication rates with increased reliability, e.g., [4].

Extracting the gains offered by a MIMO system depends on the available channel state information (CSI). For instance, when CSI is available at the receivers, the MIMO channel is said to be coherent, whereas when no CSI is available at any of the communicating terminals, the channel is said to be non-coherent. To establish coherent operation, the transmitter typically sends pilot symbols, which are used by the receivers to learn the channel [5], [6]. Sending pilot symbols consumes valuable communication resources and can be quite inefficient when the channel undergoes frequent variations [6], [7]. In particular, point-to-point training-based coherent systems achieve the maximum number of communications degrees of freedom, but do not necessarily achieve an SNR-independent term in the achievable rate expression; a term that can be quite significant at moderate-to-high SNRs. The significance of the SNR-independent term increases with the number of receive antennas and decreases with the coherence interval [8], i.e., the time duration during which the channel remains constant. See, e.g., [9] and [10] for comparisons between the performance of coded non-coherent and training-based communication systems.

In contrast with training-based coherent systems, in non-coherent ones, no pilot symbols are used and the transmitted channel symbols are structured in a way that enables detection without using CSI; see [8] for high SNR communication scenarios and [11] for low SNR ones. Since non-coherent communication systems dispose of the channel learning phase, these systems will be more suited to future wireless networks, wherein the topology and the physical channels may undergo rapid variations.

Information-theoretic analysis of several coherent single-input single-output (SISO) relaying schemes was considered in [12] and [13]. This analysis is extended to coherent MIMO relay systems with fading channels in [14]. More recently, the capacity of non-coherent SISO relay channels was studied in [15], and a distributed signalling scheme that suits non-coherent signalling on MIMO relay channels was devised in [16]. The scheme in [16] is based on using the subspace occupied by the transmitted signal as the "information carrying object" [8]. Such a scheme enables decoding at the destination without any CSI. Because of its geometric properties, this type of signalling is referred to as Grassmannian signalling. Our analytical development will make use of this type of signalling and hence, a brief note on the Grassmann manifold, which underlies it, will be provided in the next section.

In this paper, we consider a non-coherent full duplex frequency-flat block Rayleigh fading MIMO relay system, in which the transmitter, the relay and the receiver have multiple antennas but no CSI is available. We invoke channel non-coherence to derive an expression for an efficiently computable bound on the cut-set bound. It is shown that, when the coherence time during which the channel remains constant, exceeds a certain threshold, the input distribution that maximizes this bound is the isotropic distribution on a compact Grassmann manifold [8]. Furthermore, it is shown that, from a degrees of freedom perspective, no relaying is necessary and Grassmannian signalling at the source achieves the upper bound within an SNR-independent gap. A similar observation pertaining to the non-coherent SISO case is reported in [15]. Next, we devise a Grassmannian decode-and-forward (DF) scheme, which, despite channel non-coherence, enables the source and the relay to cooperate coherently in the sense of [17, Chapter 16] to achieve higher communication rates than direct transmission. In fact, when the source-relay channel is sufficiently stronger than the source-destination and relay-destination channels, and the number of relay transmit antennas are appropriately chosen, the Grassmannian DF scheme achieves the ergodic non-coherent capacity of the relay channel within a gap that goes to zero as the SNR goes to infinity. To determine the number of optimal relay transmit antennas, we consider the gap between the rate achieved by Grassmannian DF and the upper bound. It is is shown that this gap is a convex function of the number of the relay transmit antennas. Using this observation, closed form expressions that yield the optimal number of these antennas are provided. These expressions indicate that the optimal number of relay transmit antennas is monotonically decreasing with the source transmit power. It is worth noting that the approaches adopted herein are extensible to half duplex relay systems. However, the signals and the number of transmit and receive antennas that attain the lower and upper bounds corresponding to these systems appear to be significantly different from their full duplex counterparts.

The paper is organized as follows. In Section II the system model and preliminaries necessary for subsequent mathematical developments are provided. In Section III an upper bound on the cut-set bound is derived and a description of the structure of the signal matrices that maximize this bound is provided. Section IV is concerned with the evaluation of the new bound at high SNRs. It is in this section that we
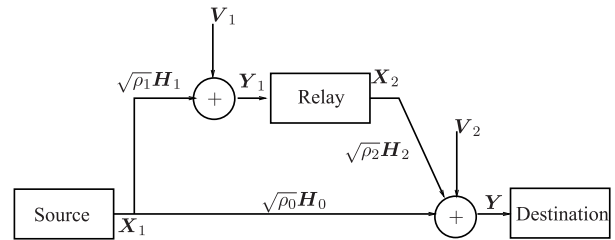

Fig. 1. A full-duplex MIMO relay channel.

show that, from a degrees of freedom perspective, no relaying is necessary and Grassmannian signalling at the source achieves the upper bound within an SNR-independent gap. In Section V a Grassmannian DF signalling scheme is devised. In Section VI we study the gap between the rate achieved by Grassmannian DF and the cut-set-based upper bound. The analysis of this gap enables us to determine conditions under which Grassmannian DF achieves capacity within an approximation error that goes to zero as the SNR goes to infinity. The analysis of the gap to the cut-set-based bound is used to expose the dependence of the number of relay transmit antennas on the source transmit power. Section VII provides numerical examples and Section VIII concludes the paper.

Notation: Standard notation will be used throughout: uppercase boldface letters will be used to denote random matrices and regular face letters to denote deterministic ones. The symbol $\oplus$ will be used to denote the direct sum operation [18], the symbol $\otimes$ will be used to denote the Kronecker product and the symbol vec$(\cdot)$ will be used to denote the operator that stacks the columns of the matrix argument on top of each other. The Stiefel and the Grassmann manifolds of $n \times m$ complex matrices, will be denoted by $\mathbb{V}_{n,m}$ and $\mathbb{G}_m(\mathbb{C}^n)$, respectively. The volume of these manifolds will be denoted by $|\mathbb{V}_{n,m}|$ and $|\mathbb{G}_m(\mathbb{C}^n)|$. The $q \times q$ identity matrix will be denoted by I_q. The standard $o(\cdot)$ and $O(1)$ notations are used throughout in conjunction with the condition that the SNR goes to infinity. In particular, $g_1(x) = o(f_1(x))$ is used to imply that $g_1(x)/f_1(x)$ tends to zero as x goes to infinity, and $g_2(x) = O(f_2(x))$ is used to imply that $g_2(x)/f_2(x)$ tends to a constant as x goes to infinity.

II. SYSTEM MODEL AND PRELIMINARIES

A. System Model

We consider the non-coherent MIMO communication system model shown in Fig. 1. In this model the physical channels are assumed to be frequency-flat block Rayleigh fading with a coherence time of T channel uses. Hence, for each block of T channel uses, each channel takes on a new statistically-independent realization and during each block it remains fixed. The dependence of T on the physical channel implies that its value can be assumed predetermined for a given communication scenario, e.g., packet-based or perfectly interleaved frequency-hopping terminal-relaying system [1], [8]. For tractability, it is assumed that the source-relay, source-destination, and relay-destination links have the same coherence time. This assumption is realistic because in practical communication scenarios, including current Long Term Evolution (LTE) and its "Advanced" version [19], the coherence time is determined by the frame structure, which is common for all terminals.

In the considered model, the source-destination and the source-relay channels are denoted by $\sqrt{\rho_0} H_0 \in \mathbb{C}^{M_1 \times N_2}$ and $\sqrt{\rho_1} H_1 \in \mathbb{C}^{M_1 \times N_1}$, respectively, and the relay-destination channel is denoted by $\sqrt{\rho_2} H_2 \in \mathbb{C}^{M_2 \times N_2}$, where $\sqrt{\rho_i}$, $i = 0, 1, 2$, are scaling parameters that represent the strengths of the MIMO channels relative to noise. The number of transmit antennas at the source and the relay is denoted by M_1 and M_2, respectively, and the number of receive antennas at the relay and the destination is denoted by N_1 and N_2, respectively. The additive noise at the relay and the destination is denoted by $V_1 \in \mathbb{C}^{T \times N_1}$ and $V_2 \in \mathbb{C}^{T \times N_2}$, respectively. The entries of H_i, $i = 0, 1, 2$ and V_i, $i = 1, 2$ are independent and identically distributed (i.i.d.) zero mean circularly-symmetric Gaussian random variables with unit variance. Using this notation, the received signals at the relay and the destination, Y_1 and Y, can be expressed as

$$Y_1 = \sqrt{\rho_1} X_1 H_1 + V_1, \tag{1}$$
$$Y = \sqrt{\rho_0} X_1 H_0 + \sqrt{\rho_2} X_2 H_2 + V_2, \tag{2}$$

where $X_1 \in \mathbb{C}^{T \times M_1}$ and $X_2 \in \mathbb{C}^{T \times M_2}$ are the transmitted signals of the source and the relay, respectively.

When the input data is encoded over a sufficiently large block, the law of large numbers can be used to express the average-power constraints at the source and the relay, respectively, as

$$\mathrm{E}\{\mathrm{Tr}(X_1^\dagger X_1)\} \leq P_s, \quad \text{and} \quad \mathrm{E}\{\mathrm{Tr}(X_2^\dagger X_2)\} \leq P_r. \tag{3}$$

B. Preliminaries

1) The Stiefel Manifold: The $p \times q$ Stiefel manifold, denoted by $\mathbb{V}_{p,q}$ when $p > q$ and by $\mathbb{V}_p$ when $p = q$, is the set of $p \times q$ matrices with orthonormal columns, i.e., $\mathbb{V}_{p,q} = \{Q \in \mathbb{C}^{p \times q} | Q^\dagger Q = I_q\}$. The volume of $\mathbb{V}_{p,q}$ is given by [8]

$$|\mathbb{V}_{p,q}| = \int_{Q \in \mathbb{V}_{p,q}} dQ = \prod_{i=p-q+1}^{p} \frac{2\pi^i}{(i-1)!}. \tag{4}$$

2) The Compact Grassmann Manifold: The compact Grassmann manifold is the set of equivalence classes of matrices with orthonormal columns that span the same subspace. In particular, denoting the equivalence class of $Q \in \mathbb{V}_{p,q}$ by $[Q]$, the compact Grassmann manifold can be expressed as $\mathbb{G}_q(\mathbb{C}^p) = \{[Q] | Q \in \mathbb{V}_{p,q}\}$, where $[Q] = \{\hat{Q} | \hat{Q} = QP, P \in \mathbb{V}_q\}$. The volume of $\mathbb{G}_q(\mathbb{C}^p)$ is given by

$$|\mathbb{G}_q(\mathbb{C}^p)| = \frac{|\mathbb{V}_{p,q}|}{|\mathbb{V}_q|} = \pi^{q(p-q)} \prod_{i=1}^{q} \frac{(i-1)!}{(i+p-q-1)!}. \tag{5}$$

3) Isotropically Distributed Matrices: A random matrix $Q \in \mathbb{C}^{p \times q}$, $p \geq q$, is said to be isotropically distributed if the probability distribution $p(Q) = p(UQ)$, for any fixed $U \in \mathbb{V}_p$. In other words, the distribution of Q is invariant under the left action of the group of unitary matrices.

4) The Cut-Set Bound: The cut-set bound is a generic outer bound on the set of all the rates that can be simultaneously achieved by a general network with multiple input and output terminals [20, Theorem 14.10.1]. Applying this bound to the considered full duplex single relay channel yields [2]

$$R \leq \frac{1}{T} \max_{p(X_1, X_2)} \min\{I(X_1, X_2; Y), I(X_1; Y, Y_1 | X_2)\}, \tag{6}$$

where R is the rate, and the division by T is because X_1, X_2, Y and Y_1 span T channel uses.

III. MAXIMIZATION OF THE CUT-SET BOUND—ARBITRARY SNR ANALYSIS

To maximize the right hand side of (6), we will consider the maximization of each expression in the minimization separately. It will then be seen that both expressions are maximized by input signals that possess the same structure, which implies that restricting the input signal to this structure incurs no loss of optimality, and that the order of the minimization and maximization operations in (6) can be swapped without affecting the tightness of the bound.

A. Maximizing $I(X_1, X_2; Y)$

In this section we will consider the maximization of $I(X_1, X_2; Y)$ at arbitrary SNRs. We begin by showing that the matrix $[X_1, X_2] \in \mathbb{C}^{T \times M_1 + M_2}$ that maximizes $I(X_1, X_2; Y)$ can be expressed as the product of an isotropically distributed unitary matrix and an independent diagonal matrix with non-negative entries. We revisit the results in [21].

We begin by deriving an expression for $p(Y|X_1, X_2)$. Since, conditioned on X_1 and X_2, Y is Gaussian and zero mean with covariance $\mathrm{E}\{YY^\dagger | X_1, X_2\} = \rho_0 N_2 X_1 X_1^\dagger + \rho_2 N_2 X_2 X_2^\dagger + N_2 I_T$,

$$p(Y|X_1, X_2) = p(Y|X)$$
$$= \frac{\exp\left(-\mathrm{Tr}\left(Y^\dagger (N_2 X X^\dagger + N_2 I_T)^{-1} Y\right)\right)}{\pi^{TN_2} \det^{N_2}(N_2 X X^\dagger + N_2 I_T)},$$
$$X = \left[\sqrt{\rho_0} X_1 \ \sqrt{\rho_2} X_2\right]. \tag{7}$$

Since $p(Y|X)$ depends on X only through the $T \times T$ matrix $XX^\dagger$, it follows from Theorem 1 in [21] that the maximum of $I(X_1, X_2; Y)$ is achieved by $M_1 + M_2 \leq T$. In other words, using a total number of source and relay transmit antennas greater than the coherence time is not more advantageous than using a total number of transmit antennas that is equal to the coherence time. We will later obtain stricter constraints on T when we consider high SNR operation, but for now, it suffices to see that the matrix X is either "tall" or "square", but not "fat".

To develop insight into the optimal distribution of X, let $\Psi \in \mathbb{V}_{M_1+M_2}$ and $\Phi \in \mathbb{V}_T$ be arbitrary deterministic matrices, then in correspondence with properties 3 and 4 in [21], we have

$$p(Y|X_1, X_2) = p(Y|X) = p(Y|X\Psi^\dagger), \tag{8}$$
$$p(\Phi Y|\Phi X_1, \Phi X_2) = p(\Phi Y|\Phi X) = p(Y|X). \tag{9}$$

In the current relay context, the properties in (8) and (9) say that the conditional distribution of Y is invariant under "mixing" the source and relay signals and under simultaneous rotation of the destination received signal and the source and relay transmitted signals by the same unitary matrix Φ. A direct consequence of these properties is given in Lemma 1 in [21], which, along with Lemma 3 in [21] yield that the signal matrix X in (7) that maximizes $I(X_1, X_2; Y)$ can be expressed as

$$X = Q_X D, \tag{10}$$

where $Q_X = [Q_{X_1} \ Q_{X_2}] \in \mathbb{V}_{T,M_1+M_2}$ is isotropically distributed and $D = D_1 \oplus D_2$ is independent of Q_X, and D_1 and D_2 are diagonal matrices with non-negative entries.

We note that the structure of X in (10) implies that coherent collaboration between the relay and the source [17, Chapter 16] is achieved by selecting the source codewords to be orthogonal to the codewords transmitted by the relay. This observation will be used in devising a DF scheme with an underlying Grassmannian signalling structure. In Section IV, we will invoke the methodology in [8] to obtain a closed form expression for $\max_{p(X_1,X_2)} I(X_1, X_2; Y)$ at high SNR. For now, we will focus on the second argument of the minimization in (6).

B. Maximization of an Upper Bound on $I(X_1; Y, Y_1|X_2)$

The second argument of the minimization in (6) is $I(X_1; Y, Y_1|X_2)$. Unfortunately, finding an input distribution that maximizes this expression appears to be difficult. This is because, with X_2 fixed, $\sqrt{\rho_2}X_2H_2+V_2$ resembles an additive Gaussian noise matrix with potentially high power and a non-isotropic distribution. To circumvent this difficulty, we derive an upper bound on this expression. Towards that end, we write $I(X_1; Y, Y_1|X_2) = h(X_1|X_2) - h(X_1|X_2, Y, Y_1)$. Since the channel H_2 is not known at either the source or the relay, it follows that X_1 and X_2 are independent of H_2, and

$$h(X_1|X_2) = h(X_1|X_2, H_2). \tag{11}$$

Since conditioning does not increase entropy, we have

$$h(X_1|X_2, Y, Y_1) \geq h(X_1|X_2, Y, Y_1, H_2) \tag{12}$$

Combining (11) and (12) yields

$$I(X_1; Y, Y_1|X_2) \leq I(X_1; Y, Y_1|X_2, H_2).$$

Our goal now is to determine the structure of the distribution that maximizes $I(X_1; Y, Y_1|X_2, H_2)$.

Using (1) and (2) it can be seen that

$$\begin{aligned} I(X_1; Y, &Y_1|X_2, H_2) \\ &\leq I(X_1[\sqrt{\rho_1}H_1 \ \sqrt{\rho_0}H_0] + [V_1 \ V_2]; X_1). \end{aligned} \tag{13}$$

This expression is equivalent to the mutual information of a channel with input X_1, channel matrix $[\sqrt{\rho_1}H_1 \ \sqrt{\rho_0}H_0]$, and additive noise $[V_1 \ V_2]$. Using the counterparts of the properties in (8) and (9) for this channel and applying Lemma 1 and Lemma 3 in [21], it can be readily concluded that the signal matrix, X_1, that maximizes $I(X_1[\sqrt{\rho_1}H_1 \ \sqrt{\rho_0}H_0] + [V_1 \ V_2]; X_1)$ can be written as

$$X_1 = Q_{X_1}D_1, \tag{14}$$

where $Q_{X_1} \in \mathbb{V}_{T,M_1}$ is isotropically distributed and D_1 is an independent diagonal matrix with non-negative entries; both Q_{X_1} and D_1 are independent of X_2, which renders (13) hold with equality.

Comparing (14) with (10), it can be seen that for the matrix X in (7) to be in the form in (10), with Q_X isotropically distributed and independent of D, the matrix X_1 must be in the form in (14), with Q_{X_1} isotropically distributed and independent of D_1; see Appendix A in [22]. Conversely, when X_1 assumes the form in (14) with Q_{X_1} isotropically distributed and independent of D_1, choosing

the matrix Q_{X_2} to be isotropically distributed in the null space of Q_{X_1} and independent of D_2 does not affect $I(X_1; Y, Y_1|X_2, H_2)$, and maximizes $I(X_1, X_2; Y)$. Hence, it can be concluded that choosing Q_X to be isotropically distributed and independent of D maximizes $I(X_1, X_2; Y)$ and $I(X_1; Y, Y_1|X_2, H_2)$. Therefore, the structure of X in (10) also maximizes $\frac{1}{T}\min\{I(X_1, X_2; Y), I(X_1; Y, Y_1|X_2, H_2)\}$. We will refer to this quantity as the cut-set-based bound to distinguish it from the cut-set bound in (6).

IV. THE CUT-SET-BASED BOUND AT HIGH SNR

In the previous section we showed that, at arbitrary SNRs, maximizing the cut-set-based bound,

$$\frac{1}{T}\min\{I(X_1, X_2; Y), I(X_1; Y, Y_1|X_2, H_2)\}, \tag{15}$$

can be achieved with a matrix X of the form in (10), where Q_X and D are independent. In this section we will use this form and the results on non-coherent point-to-point communications in [8] to evaluate this bound at high SNRs. Towards that end, we observe that the first term of the minimization in (15) corresponds to a multiple access channel with received signal Y and full collaboration between the transmitters of X_1 and X_2, whereas the second term of this minimization corresponds to a broadcast channel with transmitted signal X_1 and full collaboration between the receivers of Y and Y_1.

In contrast with the coherence time, T, which is usually predetermined, the number of transmit and receive antennas can be regarded as design parameters that ought to be properly chosen. Relationships between the number of transmit antennas, the number of receive antennas and T were investigated in [8] for a point-to-point communication scenario. In particular, it was shown therein that for efficient high SNR non-coherent communication, the number of transmit antennas must not exceed that of receive antennas, and, when the coherence time, T, exceeds a certain threshold that is a function of the number of transmit and receive antennas, relatively simple Grassmannian signalling suffices to achieve the asymptotic non-coherent high SNR ergodic capacity. Recent results on the ergodic non-coherent capacity of systems with T less than the threshold given in [8] have been obtained in [23]. It was shown therein that input signals that achieve the capacity of these systems have the form of the product of an isotropically distributed unitary matrix and a diagonal matrix with non-zero entries that are distributed as the square root of the eigenvalues of a Beta-distributed random matrix.

We will restrict attention to the case of T exceeding the threshold in [8] and apply the relevant results to the multiple access and broadcast cuts, corresponding to $I(X_1, X_2; Y)$ and $I(X_1; Y, Y_1|X_2, H_2)$ in (15), respectively. For the multiple access cut, $M_1 + M_2 \leq N_2$, and $T \geq N_2 + \min\{M_1 + M_2, N_2\}$. Notice that the condition on T is stronger than the one obtained in Section III-A for arbitrary SNRs [21].

Applying the results in [8] to the broadcast cut corresponding to $I(X_1; Y, Y_1|X_2, H_2)$ in (15) yields $M_1 \leq N_1+N_2$, and $T \geq N_1 + N_2 + \min\{M_1, N_1+N_2\}$. Combining the constraints for both cuts yields

$$M_1 + M_2 \leq N_2, \text{ and } T \geq M_1 + N_2 + \max\{M_2, N_1\}. \tag{16}$$

In the forthcoming analysis, we will assume that M_1, M_2, N_1, N_2 and T satisfy (16), and to avoid degenerate

setups, we will assume that P_s/P_r is fixed as $P_s, P_r \to \infty$. We will consider characterizations of achievable rates that are tight up to degrees of freedom (i.e., $O(1)$) and up to $o(1)$ approximations.

A. Computing a High SNR Upper Bound on $I(X_1, X_2; Y)$

To compute $I(X_1, X_2; Y)$ at high SNR, we compute $h(Y|X_1, X_2)$ and $h(Y)$. For $h(Y|X_1, X_2)$, we note that, conditioned on X_1 and X_2, the matrix Y is Gaussian distributed. Defining $H_{02} \triangleq [H_0^\dagger \; H_2^\dagger]^\dagger \in \mathbb{C}^{M_1+M_2 \times N_2}$, we write $E\{\text{vec}(Y)\,\text{vec}^\dagger(Y)|X\} = I_{N_2} \otimes (XX^\dagger + I_T)$, which implies that

$$h(Y|X_1, X_2) = N_2 T \log \pi e + N_2 \, E\{\log \det(D^2 + I_{M_1+M_2})\}. \tag{17}$$

Computing $h(Y)$ is significantly more difficult than computing $h(Y|X_1, X_2)$. The approach used in [8] for circumventing this difficulty is to compute $h(Y)$ in a different coordinate system, namely, the coordinate system corresponding to the singular value decomposition (svd) of Y. Using this change of coordinates, it was shown in [8] and [23] that the subspace spanned by the columns of Y can be decomposed into two subspace: one is spanned by the columns of X, and the other is spanned by the columns of Y, but lies in the null space of X. The uncertainty in the first subspace is dominated by the randomness in XH_{02}, whereas the uncertainty in the second subspace is due to noise only. To compute the entropy of Y in the svd coordinates, we use Lemma 12 in [23] and the approach in Appendix A herein with $M = M_1 + M_2$, $N = N_2$, $R = J_1 D$ and $H = H_{02}$ to write

$$\begin{aligned} h(Y) = \; & h(ZJ_1DH_{02}) \\ & + (T - M_1 - M_2) \, E\{\log \det(H_{02}H_{02}^\dagger)\} \\ & + (T - M_1 - M_2) \, E\{\log \det(J_1^2 D^2 + I_{M_1+M_2})\} \\ & + (N_2 - M_1 - M_2)(T - M_1 - M_2) \log \pi e \\ & + \log |\mathbb{G}_{M_1+M_2}(\mathbb{C}^T)| + o(1), \end{aligned} \tag{18}$$

where $J_1 \triangleq \sqrt{\rho_0} I_{M_1} \oplus \sqrt{\rho_2} I_{M_2}$ and $Z \in \mathbb{V}_{M_1+M_2}$ is isotropically distributed and independent of D and H_{02}. The $o(1)$ term in (18) follows from the asymptotic decoupling of the singular values of Y in Lemma 2 in Appendix A and the fact that at high SNRs $E\{\log \det(J_1^2 D^2)\} = E\{\log \det(J_1^2 D^2 + I_{M_1+M_2})\} + o(1)$. The latter approximation is only valid when D is full rank; a condition which will be shown to be satisfied by the optimal D for M_1, M_2, N_1, N_2 and T satisfying the inequalities in (16).

To obtain an expression for $I(X_1, X_2; Y)$, we subtract (17) from (18), which yields

$$\begin{aligned} I(X_1, X_2; Y) = \; & h(ZDJ_1H_{02}) \\ & + \big((N_2 - M_1 - M_2)(T - M_1 - M_2) - N_2 T\big) \log \pi e \\ & + (T - M_1 - M_2 - N_2) \, E\{\log \det(J_1^2 D^2 + I_{M_1+M_2})\} \\ & + (T - M_1 - M_2) \, E\{\log \det(H_{02}H_{02}^\dagger)\} \\ & + \log |\mathbb{G}_{M_1+M_2}(\mathbb{C}^T)| + o(1). \end{aligned} \tag{19}$$

Using (19), a high SNR bound on $I(X_1, X_2; Y)$ can be obtained by seeking a bound on $h(ZDJ_1H_{02})$. Towards that

end, we compute the covariance

$$\begin{aligned} E_{Z,D,H_{02}}\{\text{vec}(ZDJ_1H_{02}) \, \text{vec}^\dagger(ZJ_1DH_{02})\} \\ = I_{N_2} \otimes E_{Z,D}\{ZJ_1^2 D^2 Z^\dagger\}. \end{aligned} \tag{20}$$

To compute the last expectation we use the following result from [24] and [25].

Lemma 1: For any isotropically distributed unitary matrix $\Phi \in \mathbb{C}^{n \times n}$, $E\{[\Phi]_{ij}^* [\Phi]_{k\ell}\} = \frac{1}{n} \delta_{ik} \delta_{j\ell}$, where δ_{ik} is the Kronecker delta. $\square$

Applying this lemma to the right hand side of (20) yields $E_{Z,D}\{ZJ_1^2 D^2 Z^\dagger\} = \frac{\text{Tr}(J_1^2 E\{D^2\})}{M_1+M_2} I_{M_1+M_2}$. This, together with Theorem 9.6.5 in [20], imply that $h(ZDJ_1H_{02})$ is bounded by the entropy of an $(M_1 + M_2) \times N_2$ matrix with i.i.d. Gaussian entries with variance $\frac{\text{Tr}(J_1^2 E\{D^2\})}{M_1+M_2}$, i.e.,

$$\begin{aligned} h(ZJ_1DH_{02}) & \leq N_2(M_1+M_2) \log\left(\frac{\pi e \, \text{Tr}(J_1^2 E\{D^2\})}{M_1+M_2}\right) \\ & \leq N_2(M_1+M_2) \log\left(\frac{\pi e(\rho_0 P_s + \rho_2 P_r)}{M_1+M_2}\right), \end{aligned} \tag{21}$$

where the second inequality follows from invoking (10) and the power constraints in (3).

We now consider the bound on $I(X_1, X_2; Y)$ in (19). We note that, since T satisfies (16), the coefficient of the $\log \det(\cdot)$ function in (19) is non-negative. Now, the concavity of the $\log \det(\cdot)$ function and Jensen's inequality imply that $E\{\log \det(J_1^2 D^2 + I_{M_1+M_2})\} \leq \log \det(J_1^2 E\{D^2\} + I_{M_1+M_2})$, and equality holds if and only if D is deterministic. Hence, we conclude that maximizing the bound on $I(X_1, X_2; Y)$ requires the matrix D to be deterministic, i.e., $D = D$. We will later show that setting $D = D$ also maximizes the second term in the minimization in (15).

Substituting from (21) and setting $D = D$ in (19) yields

$$\begin{aligned} I(X_1, X_2; Y) \leq \; & N_2(M_1 + M_2) \log\left(\frac{\rho_0 P_s + \rho_2 P_r}{M_1 + M_2}\right) \\ & + (T - M_1 - M_2) \, E\{\log \det(H_{02}H_{02}^\dagger)\} \\ & + (T - M_1 - M_2 - N_2) \log \det(J_1^2 D^2 + I_{M_1+M_2}) \\ & + \log |\mathbb{G}_{M_1+M_2}(\mathbb{C}^T)| \\ & - (M_1 + M_2)(T - M_1 - M_2) \log \pi e + o(1). \end{aligned} \tag{22}$$

So far we have developed a high SNR upper bound on $I(X_1, X_2; Y)$. This bound is a function of D, which will be optimized in Section IV-C. To provide a high SNR upper bound on the cut-set-based bound in (15), in the next section we will obtain a high SNR upper bound on $I(X_1; Y_1, Y|X_2, H_2)$.

B. Computing a High SNR Bound on $I(X_1; Y_1, Y|X_2, H_2)$

To derive the desired bound, we note that the expressions in (2) can be written as

$$[Y_1 \; Y] = X_1 [H_1 \; H_0] J_2 + \sqrt{\rho_2} X_2 [0_{M_2 \times N_1} \; H_2] + [V_1 \; V_2],$$

where $J_2 = \sqrt{\rho_1} I_{N_1} \oplus \sqrt{\rho_0} I_{N_2}$. To derive an expression for $I(X_1; Y_1, Y|X_2, H_2)$, we compute $h(Y_1, Y|X_1, X_2, H_2)$ and $h(Y_1, Y|X_2, H_2)$.

For $h(Y_1, Y|X_1, X_2, H_2)$, we have that, conditioned on (X_1, X_2, H_2), the entries of the matrix $[Y_1 \ Y]$ are Gaussian with covariance

$$\mathrm{E}\{\mathrm{vec}([Y_1 \ Y]) \, \mathrm{vec}^\dagger([Y_1 \ Y])|X_1, X_2, H_2)\}$$
$$= J_2^2 \otimes X_1 X_1^\dagger + I_{T(N_1+N_2)}.$$

Hence,

$$h(Y_1, Y|X_1, X_2, H_2)$$
$$= \mathrm{E}\{\log \det(\pi e(J_2^2 \otimes X_1 X_1^\dagger + I_{T(N_1+N_2)}))\}$$
$$= T(N_1+N_2) \log \pi e + N_1 \mathrm{E}\{\log \det(\rho_1 D_1^2 + I_{M_1})$$
$$+ N_2 \log \det(\rho_0 D_1^2 + I_{M_1})\}. \tag{23}$$

Analogous to the technique used to obtain (18), we invoke the analysis in Appendix A herein with $M = M_1$, $N = N_1 + N_2$, $R = D_1$ and $H = [H_1 \ H_0]J_2$. Doing so yields

$$h(Y_1, Y|X_2, H_2)$$
$$\leq h(X_1 [H_1 \ H_0] J_2 + [V_1 \ V_2])$$
$$= h(W D_1 [H_1 \ H_0] J_2) + \log |\mathbb{G}_{M_1}(\mathbb{C}^T)|$$
$$+ (T - M_1) \mathrm{E}\{\log \det(D_1^2)\}$$
$$+ (T - M_1) \mathrm{E}\{\log \det(\rho_1 H_1 H_1^\dagger + \rho_0 H_0 H_0^\dagger)\}$$
$$+ (N_1 + N_2 - M_1)(T - M_1) \log \pi e + o(1), \tag{24}$$

where $W \in \mathbb{V}_{M_1}$ is isotropically distributed and independent of D_1 and $[H_1 \ H_0]$. Using Theorem 9.6.4 in [20], we have

$$h(W D_1 [H_1 \ H_0] J_2) = h(W D_1 [H_1 \ H_0])$$
$$+ M_1(N_1 \log \rho_1 + N_2 \log \rho_0). \tag{25}$$

Applying an approach analogous to the one used in deriving (21), it can be shown that the covariance of the matrix $W D_1 [H_1 \ H_0]$ can be expressed as

$$\mathrm{E}\{\mathrm{vec}(W D_1 [H_1 \ H_0]) \, \mathrm{vec}^\dagger(W D_1 [H_1 \ H_0])\}$$
$$= I_{N_1+N_2} \otimes \mathrm{E}\{W D_1^2 W^\dagger\}$$
$$= \frac{\mathrm{Tr}(\mathrm{E}\{D_1^2\})}{M_1} I_{M_1(N_1+N_2)},$$

which, using (3) and (10), implies that

$$h(W D_1 [H_1 \ H_0]) \leq M_1(N_1 + N_2) \log\left(\pi e \frac{P_s}{M_1}\right). \tag{26}$$

To bound the right hand side of (24), we use the fact that $\mathrm{E}\{\log \det(D_1^2)\} = \mathrm{E}\{\log \det(\rho_i D_1^2)\} - M_1 \log \rho_i$, $i = 0, 1$. Hence, for any $\lambda \in [0, 1]$, we can write

$$\mathrm{E}\{\log \det(D_1^2)\} = \lambda \mathrm{E}\{\log \det(\rho_1 D_1^2)\}$$
$$+ (1 - \lambda) \mathrm{E}\{\log \det(\rho_0 D_1^2)\}$$
$$- M_1(1 - \lambda) \log \rho_0 - \lambda M_1 \log \rho_1$$
$$\leq \lambda \mathrm{E}\{\log \det(\rho_1 D_1^2 + I_{M_1})\}$$
$$+ (1 - \lambda) \mathrm{E}\{\log \det(\rho_0 D_1^2 + I_{M_1})\}$$
$$- \lambda M_1 \log \rho_1 - M_1(1 - \lambda) \log \rho_0. \tag{27}$$

Combining (24) with (25), (26) and (27) yields

$$h(X_1 [H_1 \ H_0] + [V_1 \ V_2])$$
$$\leq \log |\mathbb{G}_{M_1}(\mathbb{C}^T)| + M_1(N_1 + N_2) \log\left(\pi e \frac{P_s}{M_1}\right)$$

$$+ \lambda(T - M_1) \mathrm{E}\{\log \det(\rho_1 D_1^2 + I_{M_1})\}$$
$$+ (1 - \lambda)(T - M_1) \mathrm{E}\{\log \det(\rho_0 D_1^2 + I_{M_1})\}$$
$$+ (T - M_1) \mathrm{E}\{\log \det(\rho_1 H_1 H_1^\dagger + \rho_0 H_0 H_0^\dagger)\}$$
$$+ (N_1 + N_2 - M_1)(T - M_1) \log \pi e$$
$$+ M_1(N_1 - \lambda(T - M_1)) \log \rho_1$$
$$+ M_1(N_2 - (1 - \lambda)(T - M_1)) \log \rho_0 + o(1). \tag{28}$$

Combining (28) with (23) yields

$$I(X_1; Y_1, Y|X_2, H_2)$$
$$\leq M_1(N_1 + N_2) \log\left(\frac{P_s}{M_1}\right) + \log |\mathbb{G}_{M_1}(\mathbb{C}^T)|$$
$$- M_1(T - M_1) \log \pi e$$
$$+ (\lambda(T - M_1) - N_1) \mathrm{E}\{\log \det(\rho_1 D_1^2 + I_{M_1})\}$$
$$+ ((1 - \lambda)(T - M_1) - N_2) \mathrm{E}\{\log \det(\rho_0 D_1^2 + I_{M_1})\}$$
$$+ (T - M_1) \mathrm{E}\{\log \det(\rho_1 H_1 H_1^\dagger + \rho_0 H_0 H_0^\dagger)\}$$
$$+ M_1(N_1 - \lambda(T - M_1)) \log \rho_1$$
$$+ M_1(N_2 - (1 - \lambda)(T - M_1)) \log \rho_0 + o(1). \tag{29}$$

This bound on $I(X_1; Y_1, Y|X_2, H_2)$ is valid for any $\lambda \in [0, 1]$, and its maximization can be facilitated by choosing λ to ensure that it is concave in D_1^2. This can be can be accomplished by setting λ to be equal to $\frac{\max\{M_2, N_1\}}{N_2 + \max\{M_2, N_1\}}$. This choice of λ, along with the constraints in (16), imply that $\lambda(T - M_1) - N_1 \geq \max\{M_2, N_1\} - N_1 \geq 0$ and $(1 - \lambda)(T - M_1) - N_2 \geq 0$, which ensures that the coefficients of the terms involving D_1^2 are nonnegative, and subsequently that the bound in (29) is concave in D_1^2.

Now, the concavity of $\log \det(\cdot)$ implies that $\mathrm{E}\{\log \det(\rho_i D_1^2 + I_{M_1})\} \leq \log \det(\rho_i \mathrm{E}\{D_1^2\} + I_{M_1})$, $i = 0, 1$. This inequality holds with equality if and only if $\mathrm{E}\{D_1^2\} = D_1^2$, i.e., D_1 is deterministic. Hence,

$$I(X_1; Y_1, Y|X_2, H_2)$$
$$\leq M_1(N_1 + N_2) \log\left(\frac{P_s}{M_1}\right) + \log |\mathbb{G}_{M_1}(\mathbb{C}^T)|$$
$$- M_1(T - M_1) \log \pi e$$
$$+ (\lambda(T - M_1) - N_1) \log \det(\rho_1 D_1^2 + I_{M_1})$$
$$+ ((1 - \lambda)(T - M_1) - N_2) \log \det(\rho_0 D_1^2 + I_{M_1})$$
$$+ (T - M_1) \mathrm{E}\{\log \det(\rho_1 H_1 H_1^\dagger + \rho_0 H_0 H_0^\dagger)\}$$
$$+ M_1(N_1 - \lambda(T - M_1)) \log \rho_1$$
$$+ M_1(N_2 - (1 - \lambda)(T - M_1)) \log \rho_0 + o(1). \tag{30}$$

Note that, because $I(X_1; Y_1, Y|X_2, H_2)$ pertains to the broadcast cut of the cut-set bound, it only depends on the transmit power of the source, P_s, but not on the transmit power of the relay, P_r.

C. Maximization and Tightness of Bounds

Now that we have obtained expressions for bounds on $I(X_1, X_2; Y)$ and $I(X_1; Y, Y_1|X_2, H_2)$, to obtain a bound on the rate that can be achieved by the non-coherent MIMO relay network, we need to maximize the minimum of the bounds on $I(X_1, X_2; Y)$ in (22) and $I(X_1; Y, Y_1|X_2, H_2)$ in (30). We have argued that, for Jensen's inequality used in deriving these bounds to hold with equality, the matrix D must be

deterministic, i.e., $\mathbf{D} = D$. In this case, the power constraints in (3) can be expressed as

$$\text{Tr}(D_1^2) \leq P_s, \quad \text{and} \quad \text{Tr}(D_2^2) \leq P_r. \qquad (31)$$

Subject to these constraints, the bound on $I(\underline{X}_1, \underline{X}_2; \underline{Y})$ in (22) is maximized by setting $D_1 = \sqrt{\frac{P_s}{M_1}} I_{M_1}$ and $D_2 = \sqrt{\frac{P_r}{M_2}} I_{M_2}$. Similarly, the bound on $I(\underline{X}_1; \underline{Y}, \underline{Y}_1 | \underline{X}_2, \underline{H}_2)$ in (30) is maximized by setting $D_1 = \sqrt{\frac{P_s}{M_1}} I_{M_1}$. Since both bounds are maximized by the same D, this choice maximizes their minimum.

Substituting $D_1 = \sqrt{\frac{P_s}{M_1}} I_{M_1}$ and $D_2 = \sqrt{\frac{P_r}{M_2}} I_{M_2}$ in (22) yields

$$I(\underline{X}_1, \underline{X}_2; \underline{Y}) \leq \phi_1 + o(1), \qquad (32)$$

$$\phi_1 \triangleq N_2(M_1 + M_2) \log\left(\frac{\rho_0 P_s + \rho_2 P_r}{M_1 + M_2}\right)$$

$$+ M_1(T - M_1 - M_2 - N_2) \log\left(\frac{\rho_0 P_s}{M_1}\right)$$

$$+ M_2(T - M_1 - M_2 - N_2) \log\left(\frac{\rho_2 P_r}{M_2}\right)$$

$$+ (T - M_1 - M_2) \mathrm{E}\{\log\det(\mathbf{H}_{02} \mathbf{H}_{02}^\dagger)\}$$

$$- (M_1 + M_2)(T - M_1 - M_2) \log \pi e$$

$$+ \log |\mathbb{G}_{M_1 + M_2}(\mathbb{C}^T)|. \qquad (33)$$

For arbitrary M_1 and M_2, the bound in (32) is tight from a degrees of freedom perspective, i.e., up to an error of order $O(1)$, but this bound is loose from an $o(1)$ perspective. This is because, for arbitrary M_1 and M_2, (21) holds with equality if and only if $\frac{\rho_0 P_s}{M_1} = \frac{\rho_2 P_r}{M_2}$. In Section VI we will show that this condition, not only yields the optimal M_2, but also ensures the $o(1)$ tightness of the bound in (32). To show the $O(1)$ tightness of the bound in (32), in Section IV-D we will derive an achievable lower bound on $I(\underline{X}_1, \underline{X}_2; \underline{Y})$ and will show that both the lower bound and the bound in (32) share the same number of degrees of freedom.

It is worth noting that, for arbitrary M_1 and M_2, other techniques, including the duality one in [26], might be used to yield tighter bounds on $I(\underline{X}_1, \underline{X}_2; \underline{Y})$.

For now, we continue with the derivation of an upper bound on $I(\underline{X}_1; \underline{Y}_1, \underline{Y} | \underline{X}_2, \underline{H}_2)$. Substituting $D_1 = \sqrt{\frac{P_s}{M_1}} I_{M_1}$ in (30) and using the fact that, for large P_s,

$$\log\left(\frac{P_s}{M_1} + 1\right) = \log\left(\frac{P_s}{M_1}\right) + o(1)$$

yield

$$I(\underline{X}_1; \underline{Y}_1, \underline{Y} | \underline{X}_2, \underline{H}_2) \leq \phi_2 + o(1), \qquad (34)$$

$$\phi_2 \triangleq M_1(T - M_1) \log\left(\frac{P_s}{M_1 \pi e}\right) + \log |\mathbb{G}_{M_1}(\mathbb{C}^T)|$$

$$+ (T - M_1) \mathrm{E}\{\log\det(\rho_1 \mathbf{H}_1 \mathbf{H}_1^\dagger + \rho_0 \mathbf{H}_0 \mathbf{H}_0^\dagger)\}. \qquad (35)$$

Unlike the bound in (32), this bound is $o(1)$ tight for any ρ_i, $i = 0, 1, 2$ and P_r as $P_s \to \infty$. This is because, in contrast with the matrix J_1 in (21), the matrix J_2 in (25) can be readily decoupled from the random matrices and these matrices do not depend on P_r.

D. A Lower Bound on $\max_{p(X_1, X_2)} I(\underline{X}_1, \underline{X}_2; \underline{Y})$

It was argued in the previous section that the bound that we obtained on $I(\underline{X}_1, \underline{X}_2; \underline{Y})$ is not $o(1)$ tight unless $\frac{\rho_0 P_s}{M_1} = \frac{\rho_2 P_r}{M_2}$. To investigate the number of degrees of freedom, we derive a lower bound on $I(\underline{X}_1, \underline{X}_2; \underline{Y})$ for arbitrary M_1 and M_2. We will show that the upper bound in (32) and the lower bound obtained in this section share the same number of degrees of freedom. This implies that the bound in (32) is $O(1)$ tight, that is, up to an SNR-independent term and an $o(1)$ approximation error. We will later show that the upper bound is maximized by M_2 satisfying $\frac{\rho_0 P_s}{M_1} = \frac{\rho_2 P_r}{M_2}$, and that this value of M_2 causes the gap between the upper and lower bounds to decay to zero as P_s goes to infinity.

By examining the derivation of the bound in (32), it can be seen that its looseness follows from that of the upper bound on $h(\mathbf{Z} J_1 \mathbf{D} \mathbf{H}_{02})$ in (21). Hence, to derive a lower bound on $\max_{p(X_1, X_2)} I(\underline{X}_1, \underline{X}_2; \underline{Y})$, we will derive a lower bound on $h(\mathbf{Z} J_1 \mathbf{D} \mathbf{H}_{02})$. Since conditioning does not increase entropy, we have

$$h(\mathbf{Z} J_1 \mathbf{D} \mathbf{H}_{02})$$

$$\geq h(\mathbf{Z} J_1 \mathbf{D} \mathbf{H}_{02} | \mathbf{Z} J_1 \mathbf{D})$$

$$= h\big((I_{N_2} \otimes \mathbf{Z} J_1 \mathbf{D}) \operatorname{vec}(\mathbf{H}_{02}) | \mathbf{Z} J_1 \mathbf{D}\big)$$

$$= h(\mathbf{H}_{02}) + 2N_2 \mathrm{E}\{\log |\det(\mathbf{Z} J_1 \mathbf{D})|\} \qquad (36)$$

$$= N_2(M_1 + M_2) \log \pi e + N_2 M_1 \log \rho_0$$

$$+ N_2 M_2 \log \rho_2 + N_2 \mathrm{E}\{\log\det \mathbf{D}^2\}, \qquad (37)$$

where (36) follows from [20, Theorem 9.6.3]. Using (37) in (19) yields,

$$\max_{p(X_1, X_2)} I(\underline{X}_1, \underline{X}_2; \underline{Y})$$

$$\geq \log |\mathbb{G}_{M_1 + M_2}(\mathbb{C}^T)|$$

$$+ (T - M_1 - M_2) \mathrm{E}\{\log\det(\mathbf{H}_{02} \mathbf{H}_{02}^\dagger)\}$$

$$+ (T - M_1 - M_2 - N_2) \mathrm{E}\{\log\det(J_1^2 \mathbf{D}^2 + I_{M_1 + M_2})\}$$

$$- (M_1 + M_2)(T - M_1 - M_2) \log \pi e + N_2 M_1 \log \rho_0$$

$$+ N_2 M_2 \log \rho_2 + N_2 \mathrm{E}\{\log\det \mathbf{D}^2\} + o(1). \qquad (38)$$

A lower bound on $\max_{p(X_1, X_2)} I(\underline{X}_1, \underline{X}_2; \underline{Y})$ can be obtained by setting $\mathbf{D}^2$ to have a particular distribution that satisfies the power constraint. One such setting is to let $\mathbf{D}^2$ be deterministic. Hence,

$$\max_{p(X_1, X_2)} I(\underline{X}_1, \underline{X}_2; \underline{Y}) \geq I(\underline{X}_1, \underline{X}_2; \underline{Y})\Big|_{\mathbf{D}^2 = \frac{P_s}{M_1} I_{M_1} \oplus \frac{P_r}{M_2} I_{M_2}}.$$

Using this in (38), yields

$$\max_{p(X_1, X_2)} I(\underline{X}_1, \underline{X}_2; \underline{Y}) \geq \phi_3 + o(1), \qquad (39)$$

$$\phi_3 \triangleq \log |\mathbb{G}_{M_1 + M_2}(\mathbb{C}^T)|$$

$$+ (T - M_1 - M_2) \mathrm{E}\{\log\det(\mathbf{H}_{02} \mathbf{H}_{02}^\dagger)\}$$

$$+ M_1(T - M_1 - M_2) \log\left(\frac{\rho_0 P_s}{M_1 \pi e}\right)$$

$$+ M_2(T - M_1 - M_2) \log\left(\frac{\rho_2 P_r}{M_2 \pi e}\right). \qquad (40)$$

E. A Degrees of Freedom View on the Cut-Set-Based Bound

The asymptotic $O(1)$ behaviour of a function $f(P_s)$ as $P_s \to \infty$ is captured by the number of its degrees of freedom,

χ_f, that is,

$$\chi_f = \lim_{P_s \to \infty} \frac{f(P_s)}{\log P_s}. \tag{41}$$

Hence, to study the asymptotic behaviour of the cut-set-based bound as $P_s, P_r \to \infty$, we assume that the ratio P_r/P_s is finite and we compute the number of degrees of freedom of ϕ_1 in (33), ϕ_2 in (34) and ϕ_3 in (39). Direct computation of these quantities yields

$$\chi_{\phi_1} = \chi_{\phi_3} = (M_1 + M_2)(T - M_1 - M_2), \tag{42}$$
$$\chi_{\phi_2} = M_1(T - M_1). \tag{43}$$

Notice that the inequalities in (16) ensure that $\chi_{\phi_i} > 0$, $i = 1, 2, 3$.

$$\begin{aligned}
\chi_{\phi_1} - \chi_{\phi_2} &\geq M_2(M_1 + N_2 + \max\{M_2, N_1\}) \\
&\qquad - M_2(2M_1 - M_2) \\
&\geq M_2(2M_1 + M_2 + \max\{M_2, N_1\}) \\
&\qquad - M_2(2M_1 - M_2) \\
&= M_2 \max\{M_2, N_1\} > 0.
\end{aligned}$$

In other words, at sufficiently high SNRs, the broadcast cut, corresponding to $I(X_1; Y, Y_1 | H_2, X_2)$, yields the constraining term in the cut-set-based bound in (15). Using the results in [8], it can be seen that χ_{ϕ_2} is equal to the number of communication degrees of freedom in a point-to-point MIMO system with coherence time T, number of transmit antennas M_1, and number of receive antennas $N_2 \geq M_1$. This implies that using Grassmannian signalling at the source, i.e., with X_1 isotropically distributed on $\mathbb{G}_{M_1}(\mathbb{C}^T)$, achieves χ_{ϕ_2}, that is, the number of communications degrees of freedom of the non-coherent relay channel. We therefore conclude that, from a degrees of freedom perspective, no relaying is necessary, and Grassmannian signalling at the source achieves the cut-set-based bound, and hence capacity, within an SNR-independent gap and an error term of order $o(1)$. A related result is reported in [15] in the context of "regular" fading channels wherein the fading process cannot be perfectly predicted from its infinite past. In particular, it was shown in [15, Corollaries 1 and 2] that, at asymptotically high SNRs, turning off the relay is optimal with respect to the so-called fading number [26] when the prediction error of the source-destination link is not larger than the prediction error of the relay-destination link.

In the next section we will devise a Grassmannian DF signalling scheme that will be shown to achieve the $o(1)$ asymptotic capacity in some cases. In doing so, we will be guided by the observation that the bound in (15) is maximized by restricting X in (10) to be in the form $Q_X\left(\sqrt{\frac{P_s}{M_1}} I_{M_1} \oplus \sqrt{\frac{P_r}{M_2}} I_{M_2}\right)$, where Q_X is isotropically distributed on $\mathbb{G}_{M_1+M_2}(\mathbb{C}^T)$.

V. DECODE-AND-FORWARD WITH GRASSMANNIAN CODEBOOKS

In this section we devise a Grassmannian DF scheme that is essentially an adaptation of the standard DF scheme in [2] to the non-coherent channel considered herein. Although the proposed scheme utilizes a random binning strategy analogous to the one used in [2, Theorem 5] for scalar Gaussian relay channels, the rate expressions achieved by this scheme can be also achieved with the block-Markov superposition encoding scheme proposed in [27, Chapter 9]. Using these rate expressions, we derive a lower bound on the rates that can be achieved by Grassmannian DF. We will show that, in contrast with the $O(1)$ bounds, which suggest that, from a degrees of freedom perspective, no relaying is necessary, we will show in this section that Grassmannian DF yields a rate gain, and in particular cases, achieves capacity within an $o(1)$ approximation error. In particular, we will show that when the source-relay channel is sufficiently stronger than the source-destination and relay-destination channels, and the number of relay transmit antennas are appropriately chosen, the Grassmannian DF scheme achieves the ergodic non-coherent capacity of the relay channel within an $o(1)$ approximation error. In addition, we will show that, apart from that error, the optimal number of relay transmit antennas in the proposed scheme decreases monotonically with the source transmit power.

A. Codebook Structure, Encoding and Decoding

The codebook structure proposed herein is motivated by the structure of the input distribution that maximizes the bounds derived on the cut-set-based bound in the previous section. The main features of this structure are: 1) The relay codewords are randomly generated according to the isotropic distribution on $\mathbb{G}_{M_2}(\mathbb{C}^T)$; and 2) The source codewords corresponding to a given relay codeword are M_1-dimensional subspaces that are randomly generated according to the isotropic distribution on the null space of that codeword. The set of these M_1-dimensional subspaces is isomorphic to $\mathbb{G}_{M_1}(\mathbb{C}^{T-M_2})$.

The source and relay codebooks are composed of length-n codewords, each containing a sequence of $\lceil n/T \rceil$ Grassmannian symbols, where the division by T follows from the fact that each Grassmannian symbol spans T channel uses.

Relay Codebook Structure: Let s_i denote the message to be sent by the relay during time-block i, where each time-block spans n channel uses. The codebook at the relay is composed of 2^{nR_0} codewords, each spanning n channel uses, and R_0 is the rate of the codebook. A codeword corresponding to the message s_i is denoted by $\{Q_{X_2}^{(j)}(s_i)\}_{j=1}^{\lceil n/T \rceil}$. For the transmitted codewords to possess the structure that maximizes the high SNR cut-set-based bound derived in Section IV, the symbols $Q_{X_2}^{(j)}(s_i)$ are randomly generated according to the isotropic distribution on $\mathbb{G}_{M_2}(\mathbb{C}^T)$, for $j \in \{1, \ldots, \lceil \frac{n}{T} \rceil\}$ and for every time-block i.

Source Codebook Structure: Let m_i denote the message to be sent by the source during time-block i. The codebook at the source is composed of a total of $2^{n(R_0+R)}$ codewords, where R is the transmission rate from source to destination. The source codewords are partitioned into 2^{nR_0} "clusters" of 2^{nR} codewords each. Each cluster corresponds to a relay codeword that is composed of a sequence of Grassmannian symbols, denoted by $\{Q_{X_1}^{(j)}(m_i|s_i)\}_{j=1}^{\lceil n/T \rceil}$. To enable coherent cooperation to take place between the source and the relay [17, Chapter 16], in each cluster of 2^{nR} source codewords corresponding to a given relay message s_i, the symbols $Q_{X_1}^{(j)}(m_i|s_i)$ are generated according to the isotropic distribution on the null space of $Q_{X_2}^{(j)}(s_i)$. The set of all 2^{nR} sequences corresponding to a specific relay message s_i, or equivalently to a specific relay sequence of Grassmannian symbols $\{Q_{X_2}^{(j)}(s_i)\}_{j=1}^{\lceil n/T \rceil}$, is partitioned into 2^{nR_0} equal par-

titions, each corresponding to one of the 2^{nR_0} possible relay messages. Since the number of partitions within a cluster is 2^{nR_0}, the total number of sequences within each partition is $2^{n(R-R_0)}$.

Encoding and Decoding: Suppose that in block $i-1$ the relay knows the source message m_{i-1}. The relay also knows its own message s_{i-1}. To determine s_i, the relay determines the partition in which the message m_{i-1} lies in the cluster indexed by s_{i-1}. In other words, the relay uses the pair (m_{i-1}, s_{i-1}) to determine the partition s_i, that being the one in which the sequence $\{Q_{X_1}^{(j)}(m_{i-1}|s_{i-1})\}_{j=1}^{\lceil n/T \rceil}$ lies. The relay chooses the index of this partition to be its message in the next time-block, i.e., time-block i. Since at time-block i, the source knows (m_{i-1}, s_{i-1}), it can also determine s_i. The source then transmits the sequence of Grassmannian symbols corresponding to its message, m_i, from the bin with index s_i, i.e., $\{Q_{X_1}^{(j)}(m_i|s_i)\}_{j=1}^{\lceil n/T \rceil}$. The label of the 2^{nR_0} partitions forms the basis of collaboration between the source and the relay [2].

For decoding, the relay uses maximum likelihood (ML) decoding, whereas the destination uses successive decoding to recover the source message. First, the destination uses ML decoding to recover the relay message containing information about the partition in which the source codeword lies. Then, it uses ML decoding to recover the source codeword lying in that partition.

The remainder of the encoding and decoding procedure follows that in [2]. However, we point out that the key difference in the proposed technique is that each partition of the source codewords corresponds to a sequence of $\lceil n/T \rceil$ Grassmannian symbols, which are used by the relay to transmit the bin indices, and the source codewords in each partition lie in the null space of these symbols.

B. Evaluation of the DF Rate

The achievable rate of the DF scheme is given by [2]

$$R_{\mathrm{DF}} = \frac{1}{T} \max_{p(X_1, X_2)} \min\{I(X_1, X_2; Y), I(X_1; Y_1|X_2)\}$$

$$\geq \frac{1}{T} \min\{I(X_1, X_2; Y), I(X_1; Y_1|X_2)\},$$

$$[X_1 \ X_2] \in \mathbb{F}(M_1, M_2, T), \quad (44)$$

where the set $\mathbb{F}(M_1, M_2, T)$ is defined as follows:

$$\mathbb{F}(M_1, M_2, T) \triangleq \Big\{ [X_1 \ X_2] = Q_X D \,\Big|\, Q_X \text{ isotropically}$$

$$\text{distributed on } \mathbb{G}_{M_1+M_2}(\mathbb{C}^T), D = \sqrt{\frac{P_s}{M_1}} I_{M_1} \oplus \sqrt{\frac{P_r}{M_2}} I_{M_2} \Big\}.$$

$$(45)$$

The inequality in (44) follows from restricting X_1 and X_2 to be the product of a random component that is generated according to the isotropic distribution on $\mathbb{G}_{M_1+M_2}(\mathbb{C}^T)$ and a deterministic diagonal component, given by D in $\mathbb{F}(M_1, M_2, T)$. Note that, by construction, $\mathbb{F}(M_1, M_2, T)$ ensures that the source codewords, specified by the first M_1 columns of Q_X lie in the null space of the relay codewords, which are specified by the remaining M_2 columns of Q_X.

The first term in the minimization in (44) is the same as the first term in the minimization in (6). Using the Grassmannian

structure described in this section and Section IV, we have

$$I(X_1, X_2; Y)\Big|_{[X_1 \ X_2] \in \mathbb{F}(M_1, M_2, T)} \geq \phi_3 + o(1),$$

where ϕ_3 is defined in (39).

To evaluate the second term in the minimization (44), we note that the encoding procedure implies that, conditioned on X_1, the random matrices X_2 and Y_1 are statistically independent. That is, X_1, X_2 and Y_1 form the Markov chain $X_2 \to X_1 \to Y_1$, which implies that $h(Y_1|X_1, X_2) = h(Y_1|X_1)$. To evaluate $h(Y_1|X_1)$, we note from (1) that, conditioned on X_1, the entries of the matrix Y_1 are zero mean Gaussian and its covariance is given by

$$\mathrm{E}\{\mathrm{vec}(Y_1) \, \mathrm{vec}^\dagger(Y_1)|X_1\} = \rho_1(I_{N_1} \otimes X_1 X_1^\dagger) + I_{TN_1}.$$

Hence, for any M_1, N_1 and T, we have

$$h(Y_1|X_1, X_2) = \log \det\Big(\pi e\big(\rho_1(I_{N_1} \otimes X_1 X_1^\dagger) + I_{TN_1}\big)\Big)$$

$$= TN_1 \log \pi e + N_1 M_1 \log\Big(\frac{\rho_1 P_s}{M_1} + 1\Big). \quad (46)$$

To compute $h(Y_1|X_2)$, we note that, by construction, Q_{X_1} lies in the null space of Q_{X_2}. Hence, for a given $Q_{X_2} = Q_{X_2}$, Q_{X_1} can be expressed as

$$Q_{X_1} = Q_{X_2}^\perp \boldsymbol{\Omega}, \quad (47)$$

where $Q_{X_2}^\perp \in \mathbb{V}_{T, T-M_2}$ is any predetermined representative of the null space of Q_{X_2}, and $\boldsymbol{\Omega} \in \mathbb{V}_{T-M_2, M_1}$. Now,

$$h(Y_1|X_2) = h\Big(\sqrt{\frac{\rho_1 P_s}{M_1}} Q_{X_1} H_1 + V_1|X_2\Big)$$

$$= \int_{\mathbb{G}_{T-M_2}(\mathbb{C}^T)} p_{Q_{X_2}^\perp}(Q_{X_2}^\perp)$$

$$\times h\Big(\sqrt{\frac{\rho_1 P_s}{M_1}} Q_{X_2}^\perp \boldsymbol{\Omega} H_1 + V_1\Big) dQ_{X_2}^\perp$$

$$= \frac{1}{|\mathbb{G}_{T-M_2}(\mathbb{C}^T)|}$$

$$\times \int_{\mathbb{G}_{T-M_2}(\mathbb{C}^T)} h\Big(\sqrt{\frac{\rho_1 P_s}{M_1}} Q_{X_2}^\perp \boldsymbol{\Omega} H_1 + V_1\Big) dQ_{X_2}^\perp.$$

$$(48)$$

To compute a lower bound on $h(Y_1|X_2)$, we will show that $h\Big(\sqrt{\frac{\rho_1 P_s}{M_1}} Q_{X_2}^\perp \boldsymbol{\Omega} H_1 + V_1\Big)$ does not depend on $Q_{X_2}^\perp$, which makes the evaluation of the integral in (48) trivial. To do so, we note that the matrix $\bar{Q}_X = [Q_{X_2}^\perp \ Q_{X_2}]$ is deterministic and $|\det \bar{Q}_X| = 1$. Since V_1 is isotropically distributed and statistically independent of $\boldsymbol{\Omega} H_1$, we have $V_1 \stackrel{d}{=} \bar{Q}_X^\dagger V_1 = \tilde{V}_1$, where $\stackrel{d}{=}$ denotes equality in distribution. Hence, using [20, Theorem 9.6.3], we can write

$$h\Big(\sqrt{\frac{\rho_1 P_s}{M_1}} Q_{X_2}^\perp \boldsymbol{\Omega} H_1 + V_1\Big)$$

$$= h\Big(\sqrt{\frac{\rho_1 P_s}{M_1}} \bar{Q}_X^\dagger Q_{X_2}^\perp \boldsymbol{\Omega} H_1 + \bar{Q}_X^\dagger V_1\Big)$$

$$= h\left(\begin{bmatrix} \sqrt{\frac{\rho_1 P_s}{M_1}} \boldsymbol{\Omega} H_1 + \tilde{V}_{11} \\ \tilde{V}_{12} \end{bmatrix}\right), \quad (49)$$

where $\tilde{V}_{11} \in \mathbb{C}^{T-M_2 \times N_1}$ and $\tilde{V}_{12} \in \mathbb{C}^{M_2 \times N_1}$ denote the matrices containing the upper $T - M_2$ and lower M_2 rows of $\tilde{V}_1$, respectively. Since the entries of $\tilde{V}_1$ are i.i.d. Gaussian-distributed with unit variance,

$$h\left(\sqrt{\frac{\rho_1 P_s}{M_1}} Q_{X_2}^{\perp} \boldsymbol{\Omega} \boldsymbol{H}_1 + \boldsymbol{V}_1\right)$$

$$= h\left(\sqrt{\frac{\rho_1 P_s}{M_1}} \boldsymbol{\Omega} \boldsymbol{H}_1 + \tilde{\boldsymbol{V}}_{11}\right) + h(\tilde{\boldsymbol{V}}_{12})$$

$$= h\left(\sqrt{\frac{\rho_1 P_s}{M_1}} \boldsymbol{\Omega} \boldsymbol{H}_1 + \tilde{\boldsymbol{V}}_{11}\right) + M_2 N_1 \log \pi e. \quad (50)$$

This expression does not depend on $Q_{X_2}^{\perp}$, and hence (48) yields

$$h(Y_1|X_2) = h\left(\sqrt{\frac{\rho_1 P_s}{M_1}} \boldsymbol{\Omega} \boldsymbol{H}_1 + \tilde{\boldsymbol{V}}_{11}\right) + M_2 N_1 \log \pi e. \quad (51)$$

To derive a lower bound on $I(X_1; Y_1|X_2)$ achieved by Grassmannian DF, we note that the dimension of the matrix $\boldsymbol{\Omega} \boldsymbol{H}_1$ in (51) is $T - M_2 \times N_1$ and we consider two distinct cases: $N_1 \leq M_1$ and $N_1 > M_1$. For the first case, the technique in [8] can be used directly to derive a lower bound on $h(Y_1|X_2)$, whereas for the second case, the technique in [23] appears to be more appropriate.

1) Case I—$N_1 < M_1$: To derive a lower bound on $I(X_1; Y_1|X_2)$ achieved by Grassmannian DF in this case, we invoke Theorem 12 in [8], which suggests that, up to an error of order $o(1)$, the maximum rate that can be achieved on a point-to-point link with $M_1 > N_1$ source transmit antennas is attained by Grassmannian signalling if the number of source transmit antennas is reduced to N_1, and the source transmit power is equally distributed among them. Motivated by this theorem, we derive a lower bound on $I(X_1; Y_1|X_2)$ achieved by Grassmannian DF by restricting the number of "active" antennas to N_1.

With M_1 reduced to N_1, we will denote the square matrix corresponding to $\boldsymbol{H}_1$ in (51) by $\tilde{\boldsymbol{H}}_1$. Now, by the fact that conditioning does not increase entropy, we have

$$h\left(\sqrt{\frac{\rho_1 P_s}{N_1}} \boldsymbol{\Omega} \tilde{\boldsymbol{H}}_1 + \tilde{\boldsymbol{V}}_{11}\right) \geq h\left(\sqrt{\frac{\rho_1 P_s}{N_1}} \boldsymbol{\Omega} \tilde{\boldsymbol{H}}_1 + \tilde{\boldsymbol{V}}_{11} \middle| \tilde{\boldsymbol{V}}_{11}\right)$$

$$= h\left(\sqrt{\frac{\rho_1 P_s}{N_1}} \boldsymbol{\Omega} \tilde{\boldsymbol{H}}_1\right). \quad (52)$$

Using a technique similar to the one used in deriving Theorem 9 in [8], it can be readily verified that

$$h\left(\sqrt{\frac{\rho_1 P_s}{N_1}} \boldsymbol{\Omega} \tilde{\boldsymbol{H}}_1\right) = N_1(T - M_2) \log \frac{\rho_1 P_s}{N_1} + N_1^2 \log \pi e$$

$$+ (T - M_2 - N_1) \operatorname{E}\{\log \det(\tilde{\boldsymbol{H}}_1 \tilde{\boldsymbol{H}}_1^{\dagger})\} + \log |\mathbb{G}_{N_1}(\mathbb{C}^{T-M_2})|. \quad (53)$$

Combining (53), (52) and (50), with M_1 replaced with N_1 and $\boldsymbol{H}_1$ replaced with $\tilde{\boldsymbol{H}}_1$, and subtracting the expression of $h(Y_1|X_1, X_2)$ in (46) yields that Grassmannian DF results in

$$I(X_1; Y_1|X_2)\Big|_{[X_1 \, X_2] \in \mathbb{F}(N_1, M_2, T)} \geq \phi_4 + o(1),$$

where the error term follows from the fact that $\log\left(1 + \frac{\rho_1 P_s}{N_1}\right) = \log \frac{\rho_1 P_s}{N_1} + o(1)$, and

$$\phi_4 \triangleq N_1(T - M_2 - N_1) \log \frac{P_s}{N_1 \pi e} + \log |\mathbb{G}_{N_1}(\mathbb{C}^{T-M_2})|$$

$$+ (T - M_2 - N_1) \operatorname{E}\{\log \det(\rho_1 \tilde{\boldsymbol{H}}_1 \tilde{\boldsymbol{H}}_1^{\dagger})\}. \quad (54)$$

2) Case II—$N_1 \geq M_1$: In this case, the bound in (52) becomes trivial,[1] because the matrix $\frac{\rho_1 P_s}{N_1} \boldsymbol{\Omega} \boldsymbol{H}_1$ is singular and its differential entropy is $-\infty$. To avoid this pitfall, we will again make use of Lemma 12 in [23], which enables the vector of singular values of the matrix $\frac{\rho_1 P_s}{N_1} \boldsymbol{\Omega} \boldsymbol{H}_1 + \tilde{\boldsymbol{V}}_{11}$ to be decoupled into two vectors that become statistically independent asymptotically as $P_s \to \infty$. The first vector corresponds to the M_1 nonzero singular values of $\frac{\rho_1 P_s}{N_1} \boldsymbol{\Omega} \boldsymbol{H}_1$ and the second vector corresponds to the $N_1 - M_1$ singular values of an independent $(T - M_2 - M_1) \times (N_1 - M_1)$ random matrix with i.i.d. zero mean and unit variance Gaussian entries. The differential entropy of $\frac{\rho_1 P_s}{N_1} \boldsymbol{\Omega} \boldsymbol{H}_1$ is subsequently computed using the technique outlined in Appendix A herein with $M = M_1$, $N = N_1$, $R = I$, $H = H_1$, and with T replaced with $T - M_2$. Doing so yields

$$h\left(\sqrt{\frac{\rho_1 P_s}{M_1}} \boldsymbol{\Omega} \boldsymbol{H}_1 + \tilde{\boldsymbol{V}}_{11}\right)$$

$$= M_1(T - M_1 - M_2 + N_1) \log \frac{\rho_1 P_s}{M_1} + \log |\mathbb{G}_{M_1}(\mathbb{C}^{T-M_2})|$$

$$+ (M_1 N_1 + (T - M_1 - M_2)(N_1 - M_1)) \log \pi e$$

$$+ (T - M_1 - M_2) \operatorname{E}\{\log \det \boldsymbol{H}_1 \boldsymbol{H}_1^{\dagger}\} + o(1). \quad (55)$$

Using (55) with (50) and (46), it can be readily verified that Grassmannian DF yields

$$I(X_1; Y_1|X_2)\Big|_{[X_1 \, X_2] \in \mathbb{F}(M_1, M_2, T)} = \phi_5 + o(1), \quad (56)$$

where

$$\phi_5 \triangleq M_1(T - M_1 - M_2) \log \frac{P_s}{M_1 \pi e} + \log |\mathbb{G}_{M_1}(\mathbb{C}^{T-M_2})|$$

$$+ (T - M_1 - M_2) \operatorname{E}\{\log \det(\rho_1 \boldsymbol{H}_1 \boldsymbol{H}_1^{\dagger})\}. \quad (57)$$

Analogous to the upper bound on $I(X_1; Y_1, Y|X_2, H_2)$ in (30), from (54) and (57) we note that ϕ_4 and ϕ_5 depend on P_s only, but not on P_r. This is because ϕ_4 and ϕ_5 represent lower bounds on $I(X_1; Y_1|X_2)$, which describes the mutual information between the source transmitted signal and the relay received signal. In other words, $I(X_1; Y_1|X_2)$ is independent of the relay transmitted power.

C. The Degrees of Freedom of the Grassmannian DF Scheme

In Section IV-E we showed that the maximum number of degrees of freedom of the cut-set-based bound is χ_{ϕ_2}, and that this number is achieved with Grassmannian signalling at the source with no relaying. It is of interest to compare that with the number of degrees of freedom of Grassmannian DF relaying. Using (41), when $N_1 < M_1$, we have from (54) that the number of degrees of freedom of the Grassmannian

---

[1]This observation has been brought to our attention by an anonymous Reviewer.

DF scheme is $\chi_{\phi_4} = N_1(T - N_1 - M_2)$ and when $N_1 \geq M_1$, we have from (57) that the corresponding number is $\chi_{\phi_5} = M_1(T - M_1 - M_2)$. Comparing χ_{ϕ_4} and χ_{ϕ_5} it can be seen that $\chi_{\phi_5} - \chi_{\phi_4} = (M_1 - N_1)(T - M_1 - M_2 - N_1)$. Thus, for T and N_2 satisfying (16), $\chi_{\phi_5} - \chi_{\phi_4} > 0$, whenever $M_1 > N_1$. This implies that, from a degrees of freedom perspective, it is more beneficial for N_1 to be at least as great as M_1. We will later show that this also holds when $T \to \infty$. Hence, we now focus on the case of $M_1 \leq N_1$ and compare the degrees of freedom achieved by the cut-set-based bound and χ_{ϕ_5}. From (43) it can be seen that $\chi_{\phi_2} - \chi_{\phi_5} = M_1 M_2$. This implies that, when M_1, N_2 and T satisfy (16), choosing M_2 to be greater than zero and using Grassmannian DF incurs a strict loss from the degrees of freedom achieved by Grassmannian signalling at the source without relaying.

At this stage one may wonder whether, from the perspective of the $o(1)$ bounds derived earlier, Grassmannian DF relaying can be more advantageous than Grassmannian signalling at the source without relaying. To address this question, in the next section we will study the $o(1)$ gap between the cut-set-based bound and the rate achieved by Grassmannian DF relaying.

VI. GAP TO THE CUT-SET-BASED BOUND

In this section we consider the $o(1)$ gap between the lower and upper bounds characterized by ϕ_i, $i = 1, \ldots, 5$. We will identify cases in which the Grassmannian DF scheme achieves the ergodic non-coherent capacity within an approximation gap of order $o(1)$. We will also show that the number of relay transmit antennas that minimizes the $o(1)$ gap between the lower bound achievable by Grassmannian DF relaying and the cut-set-based bound is monotonically decreasing with the source transmit power.

We consider two possibilities. In the first possibility, up to an error of order $o(1)$, the lower bound on $I(X_1, X_2; Y)$, denoted by ϕ_3 in (39), is less than the lower bound on $I(X_1; Y_1|X_2)$, denoted by ϕ_4 when $N_1 < M_1$ in (54) and denoted by ϕ_5 when $N_1 \geq M_1$ in (57). In this case, the gap to the capacity of the considered relay channel is bounded by $\frac{1}{T}(\phi_1 - \phi_3)$. In the second possibility, up to an error of order $o(1)$, the lower bound on $I(X_1; Y_1|X_2)$ is less than the lower bound on $I(X_1, X_2; Y)$. In this case, the gap to the capacity of the considered relay channel is bounded by $\frac{1}{T}(\phi_2 - \phi_4)$ when $N_1 < M_1$ and by $\frac{1}{T}(\phi_2 - \phi_5)$ when $N_1 \geq M_1$. We will obtain expressions for these bounds and will show how the number of relay transmit antennas, M_2, can be chosen to maximize the Grassmannian DF achievable rate.

A. Case I: Lower Bound on $I(X_1, X_2; Y)$ Less Than Lower Bound on $I(X_1; Y_1|X_2)$

One of the situations in which the case that $I(X_1, X_2; Y) \leq I(X_1; Y_1|X_2)$ arises is when $\rho_1 \gg \max\{\rho_0, \rho_2\}$. Before considering the $o(1)$ high SNR upper bound on the gap between the rate achieved by Grassmannian DF signalling and the cut-set-based bound in this case, i.e., $\Delta_{13} \triangleq \frac{1}{T}(\phi_1 - \phi_3)$, we begin by considering the choice of M_1 and M_2 that would maximize ϕ_1 in (33), the upper bound on $I(X_1, X_2; Y)$. To maximize this bound, we let $K = M_1 + M_2$ and M_2 be the optimization variables instead of M_1 and M_2. With this change of variables, it can readily verified that ϕ_1 is continuous with a unique local

minimum in M_2. Hence, the optimal M_2 can be obtained by equating $\frac{\partial \phi_1}{\partial \hat{M}_2}$ to zero, where $\hat{M}_2$ is the extension of M_2 to the field of real numbers. Because of continuity and uniqueness of the local minimum, the equation $\frac{\partial \phi_1}{\partial \hat{M}_2} = 0$ has one solution. Subsequently, the optimal M_2 must be either $\lceil \hat{M}_2 \rceil$ or $\lfloor \hat{M}_2 \rfloor$. Computing $\frac{\partial \phi_1}{\partial \hat{M}_2}$ and equating it to zero yields

$$\frac{\rho_0 P_s}{K - \hat{M}_2} = \frac{\rho_2 P_r}{\hat{M}_2}. \tag{58}$$

It is worth noting from (42) that, for a given coherence time T, χ_{ϕ_1} depends only on $K = M_1 + M_2$, and is maximized when $K = \lfloor \frac{T}{2} \rfloor$ [8]; the $O(1)$ optimal K.

For simplicity we will assume that (58) yields an integer $\hat{M}_2 = M_2$. We now consider the $o(1)$ gap to the cut-set-based bound. Using (32), (33) and (39), this gap is given by

$$\Delta_{13} = \frac{N_2(M_1 + M_2)}{T}\left(-M_1 \log \frac{\rho_0 P_s}{M_1} - M_2 \log \frac{\rho_2 P_r}{M_2} + \log\left(\frac{M_1}{M_1 + M_2}\frac{\rho_0 P_s}{M_1} + \frac{M_2}{M_1 + M_2}\frac{\rho_2 P_r}{M_2}\right)\right). \tag{59}$$

The concavity of the $\log(\cdot)$ function and Jensen's inequality imply that $\Delta_{13} \geq 0$. Furthermore, $\Delta_{13} = 0$ if and only if (58) is satisfied.

Hence, when M_2 and $M_1 = K - M_2$ satisfy (58), the gap between the rate achieved by Grassmannian DF relaying and the cut-set-based bound is upper bounded by the $o(1)$ approximation errors. The condition in (58) implies that the optimal ratio of source to relay transmit antennas, $\frac{M_1}{M_2}$, is equal to the ratio of the effective source to relay transmit powers, $\frac{\rho_0 P_s}{\rho_2 P_r}$. In other words, the optimal ratio of source to relay transmit antennas depends only on the ratio $\frac{P_s}{P_r}$, rather than on the actual values of P_s and P_r. On the other hand, for a fixed number of source transmit antennas, M_1, and a fixed relay transmit power, P_r, the optimal number of relay transmit antennas, M_2, is monotonically decreasing in P_s. In the next section we will show that this is also true for the other case, which is considered therein.

B. Case II: Lower Bound on $I(X_1, X_2; Y)$ Greater Than Lower Bound on $I(X_1; Y_1|X_2)$

An upper bound on the $o(1)$ gap between the rate achieved by Grassmannian DF relaying and the cut-set-based bound is given by $\Delta_{24} \triangleq \frac{1}{T}(\phi_2 - \phi_4)$ when $N_1 < M_1$ and by $\Delta_{25} \triangleq \frac{1}{T}(\phi_2 - \phi_5)$, when $N_1 \geq M_1$. Before obtaining M_2 that minimizes this bound, we note from (35) that ϕ_2 does not depend on M_2. Hence, to minimize this bound, it suffices to maximize ϕ_4 and ϕ_5 in the respective cases.

1) $N_1 < M_1$: Substituting from (35) and (54) yields

$$\Delta_{24} = M_1\left(1 - \frac{M_1}{T}\right)\log \frac{P_s}{M_1 \pi e} + \frac{1}{T}\log \frac{|\mathbb{G}_{M_1}(\mathbb{C}^T)|}{|\mathbb{G}_{N_1}(\mathbb{C}^{T-M_2})|}$$
$$- N_1\left(1 - \frac{M_2 + N_1}{T}\right)\log \frac{P_s}{N_1 \pi e}$$
$$+ \left(1 - \frac{M_1}{T}\right)\text{E}\{\log \det(\rho_1 \boldsymbol{H}_1 \boldsymbol{H}_1^\dagger + \rho_0 \boldsymbol{H}_0 \boldsymbol{H}_0^\dagger)\}$$
$$- \left(1 - \frac{M_2 + N_1}{T}\right)\text{E}\{\log \det(\rho_1 \tilde{\boldsymbol{H}}_1 \tilde{\boldsymbol{H}}_1^\dagger)\}. \tag{60}$$

To determine M_2 that minimizes Δ_{24}, we extend M_2, as before, to the field of real numbers and denote it by $\hat{M}_2$. We then express the factorials in the expression of $|\mathbb{G}_{N_1}(\mathbb{C}^{T-M_2})|$ in (5) in terms of the gamma function, $\Gamma(\cdot)$. The resulting expression of Δ_{24} can be readily verified to be differentiable in $\hat{M}_2$. The optimal $\hat{M}_2$ must satisfy $\frac{\partial \Delta_{24}}{\partial \hat{M}_2} = 0$. We will show that this equation has a unique solution. This observation along with the continuity of Δ_{24} in $\hat{M}_2$ yield that the optimal M_2 must be one of the integers closest to the optimal $\hat{M}_2$, that is, M_2 that minimizes Δ_{24} is either $\lceil \hat{M}_2 \rceil$ or $\lfloor \hat{M}_2 \rfloor$, depending on which one yields a smaller Δ_{24}. The equation $\frac{\partial \Delta_{24}}{\partial \hat{M}_2} = 0$ can be readily expressed as

$$N_1 \log \frac{P_s}{N_1 e} + \mathrm{E}\{\log \det(\rho_1 \tilde{H}_1 \tilde{H}_1^\dagger)\}$$
$$- \frac{1}{\ln 2} \sum_{j=1}^{N_1} \Psi_0(j + T - \hat{M}_2 - N_1) = 0, \quad (61)$$

where $\Psi_0(\cdot)$ is the digamma function which is positive for all $j + T - \hat{M}_2 - M_1 \geq 2$; a condition ensured by (16) and the fact that $j \geq 1$. To show that (61) has a unique solution, we note that, for any $z \geq 0$, $\frac{d\Psi_0(z)}{dz} = \sum_{n=0}^{\infty} \frac{1}{(z+n)^2} > 0$, that is, $\Psi_0(z)$ is strictly monotonically increasing in z. This implies, not only that $\hat{M}_2$ that solves (61) is unique, but also that M_2 that minimizes Δ_{24} is monotonically decreasing with the source transmit power, P_s.

2) $N_1 \geq M_1$: Analogous to the previous case, $\hat{M}_2$ that minimizes Δ_{25} solves

$$M_1 \log \frac{P_s}{M_1 e} + \mathrm{E}\{\log \det(\rho_1 H_1 H_1^\dagger)\}$$
$$- \frac{1}{\ln 2} \sum_{j=1}^{M_1} \Psi_0(j + T - \hat{M}_2 - M_1) = 0. \quad (62)$$

Hence, similar to the previous case, in this case too, M_2 that minimizes Δ_{25} decreases monotonically with P_s.

C. The Case of Infinite T

It is instructive to examine the effect of increasing T on the $o(1)$ gap between the rate achieved by Grassmannian DF signalling and the cut-set-based bound.

When the lower bound on $I(X_1, X_2; Y)$ is less than the lower bound on $I(X_1; Y_1|X_2)$, the asymptotic $o(1)$ gap is given by $\lim_{T \to \infty} \Delta_{13}$. From (59) it can be seen that, with all other parameters held constant this limit goes to zero, even when M_2 is not optimally chosen.

When the lower bound on $I(X_1, X_2; Y)$ is greater than the lower bound on $I(X_1; Y_1|X_2)$, and $N_1 < M_1$, the asymptotic gap is given by

$$\lim_{T \to \infty} \Delta_{24} = M_1 \log \frac{P_s}{M_1 \pi e} - N_1 \log \frac{P_s}{N_1 \pi e}$$
$$+ \mathrm{E}\{\log \det(\rho_0 H_0 H_0^\dagger + \rho_1 H_1 H_1^\dagger)\}$$
$$- \mathrm{E}\{\log \det(\rho_1 \tilde{H}_1 \tilde{H}_1^\dagger)\}$$
$$+ \lim_{T \to \infty} \frac{1}{T} \log \frac{|\mathbb{G}_{M_1}(\mathbb{C}^T)|}{|\mathbb{G}_{N_1}(\mathbb{C}^{T-M_2})|}.$$

Since the first two terms of this expression are the only ones that depend on P_s, it can be seen that the gap between

the rate of the Grassmannian DF scheme to the cut-set-based bound becomes unbounded as $P_s \to \infty$. Hence, it can be concluded that for large T and P_s the number of relay receive antennas must be at least equal to the number of source transmit antennas; i.e., $N_1 \geq M_1$.

For the case of $N_1 \geq M_1$, Δ_{25} provides an upper bound to the cut-set-based bound and

$$\lim_{T \to \infty} \Delta_{25} = \mathrm{E}\{\log \det(I_{M_1} + \frac{\rho_0}{\rho_1} H_0 H_0^\dagger (H_1^\dagger H_1)^{-1})\}, \quad (63)$$

where in writing (63), we have used (5) to conclude that $\lim_{T \to \infty} \frac{1}{T} \log \frac{|\mathbb{G}_{M_1}(\mathbb{C}^T)|}{|\mathbb{G}_{M_1}(\mathbb{C}^{T-M_2})|} = 0$.

D. A High SNR Comparison With the Coherent Case

The case of MIMO relaying over coherent frequency-flat block Rayleigh channels, in which the channels are known at the relay and the destination receivers, but not at the transmitters, was considered in [28]. It was shown in Theorem 4.1 therein that the capacity of the resulting relay channel is given by $C_c \leq \min\{C_{c_1}, C_{c_2}\}$, where

$$C_{c_1} = \mathrm{E}\left\{\log \det\left(I_{M_1} + \frac{\rho_1 P_s}{M_1} H_1 H_1^\dagger + \frac{\rho_0 P_s}{M_1} H_0 H_0^\dagger\right)\right\},$$
$$C_{c_2} = \mathrm{E}\left\{\log \det\left(I_{N_2} + \frac{\rho_0 P_s}{M_1} H_0^\dagger H_0 + \frac{\rho_2 P_r}{M_2} H_2^\dagger H_2\right)\right\}.$$

Using a classical result in [29] it can be shown that for $N_2 \geq M_1 + M_2$ and as $P_s, P_r \to \infty$,

$$C_{c_1} \to M_1 \log P_s + \mathrm{E}\left\{\log \det\left(\frac{\rho_1}{M_1} H_1 H_1^\dagger + \frac{\rho_0}{M_1} H_0 H_0^\dagger\right)\right\},$$
$$C_{c_2} \to (M_1 + M_2) \log P_s + \mathrm{E}\{\log \det(H_{02} H_{02}^\dagger)\}$$
$$+ M_1 \log \frac{\rho_0}{M_1} + M_2 \log \frac{\rho_2 \gamma}{M_2},$$

where $\gamma = P_r / P_s$. From these expressions, it can be seen that when P_s is sufficiently large, the capacity of the coherent channel, C_c, is bounded by C_{c_1}, that is, the number of degrees of freedom of the MIMO coherent capacity is bounded by M_1. This is the same as the number of degrees of freedom of a point-to-point MIMO system with M_1 transmit and $N_2 \geq M_1$ receive antennas. Comparing this with $\frac{\chi_{\phi_5}}{T}$, it can be seen that when $M_2 = 0$, the number of degrees of freedom of the MIMO non-coherent capacity is $M_1\left(1 - \frac{M_1}{T}\right)$, which implies that, from a degrees of freedom perspective, the non-coherence penalty in the MIMO relay channel is the same as the corresponding penalty in point-to-point MIMO systems, and decays to zero as T goes to infinity. This further implies that, in the coherent case at asymptotically high P_s, it is optimal, from a degrees of freedom perspective, for the relay to be silent.

VII. NUMERICAL EXAMPLES

In this section we provide numerical examples to compare the cut-set-based bound with the rates achievable by Grassmannian DF. The first example illustrates the monotonic decrease of the optimal number of relay transmit antennas in the Grassmannian DF scheme with the source transmit power, whereas the second example illustrates the variation of the gap between the cut-set-based bound and the rates achieved by the Grassmannian DF scheme with the source

transmit power. In both examples the channel coherence time is $T = 40$ channel uses, which, for the specifications of the Long Term Evolution standards [19] of symbol duration of 71.4 μs, 2 GHz carrier frequency, and Doppler shift $f_d = \frac{9}{16\pi T_s}$ [30, Chapter 5], where T_s is the coherence time in seconds, corresponds to a mobility of 33.8 km/hr.

In all forthcoming plots the source power, P_s, is assumed to be sufficiently high and the $o(1)$ terms are neglected.

Example 1: In this example we consider the case in which the total number of source and relay transmit antennas is set to be $K = M_1 + M_2 = 10$. The relay and the destination are assumed to have an equal number of receive antennas $N_1 = N_2 = 10$. These settings ensure that the conditions in (16) are satisfied. Furthermore, because $N_1 > M_1$, the rate expression corresponding to $I(X_1; Y_1 | X_2)$ for Grassmannian DF is given by ϕ_5 in (57). As mentioned in Section VI-B.2, having $N_1 < M_1$ results in a possibly unbounded SNR-dependent gap to the cut-set-based bound and is hence undesirable. The relay transmit power is set to be $P_r = 30$ dB, and the relative channel gains are set to be $\rho_0 = 2$, $\rho_1 = 3$ and $\rho_2 = 1$. We consider various values of P_s, the source transmit power, and the goal is to determine at each power the number of transmit antennas that should be allocated to the source and the relay to maximize the rate achieved by the Grassmannian DF scheme.

Fig. 2 depicts the upper and lower bounds on $I(X_1, X_2; Y)$, ϕ_1 in (33) and ϕ_3 in (40), respectively, and the upper and lower bounds on $I(X_1; Y_1 | X_2)$, ϕ_2 in (35) and ϕ_5 (57), respectively, for various values of relay transmit antennas, M_2, and source powers, P_s. The cut-set-based upper bound is given by $\min\{\phi_1, \phi_2\}$ and is plotted in the figure with thicker lines. The rates achieved by Grassmannian DF is given by $\min\{\phi_3, \phi_5\}$ and is also plotted with thicker lines.

From Fig. 2(a), it can be seen that, when $P_s = 15$ dB, the maximum rate achieved by Grassmannian DF is 6.915 bits per channel use (bpcu) and is attained by using $M_2 = 6$. Increasing P_s to 25, 35 and 45 dB, it can be seen from Fig. 2(b)–(d) that the maximum rates achieved by Grassmannian DF are, respectively, 23.35, 44.35 and 67.69 bpcu. The optimal numbers of relay transmit antennas, M_2, required to achieve these rates are respectively, 2, 1 and 0.

For comparison, Fig. 2 also shows the cut-set-based bound at various values of M_2 and P_s. From this figure, it can be seen that, at $P_s = 15$ dB, the gap to the cut-set-based bound at the optimal number of relay transmit antennas is approximately 4 bpcu and the bound is maximized by setting $M_2 = 4$. At $P_s = 25$ dB, this gap is reduced to about 2.27 bpcu and the bound is maximized by $M_2 = 4$. At $P_s = 35$ dB and $P_s = 45$ dB, this gap becomes negligible, and the values of M_2 that maximize the bound at these power settings are 1 and 0, respectively. The small gap at these settings is due to the fact that the solution of (58) is not integer for these power settings. □

Example 2: In this example we plot the cut-set-based upper bound and the lower bound achieved by Grassmannian DF with the source transmit power for two cases. In the first case, $\phi_3 < \phi_5$, and in the second case $\phi_5 < \phi_3$. For both cases, the number of source and relay transmit antennas is
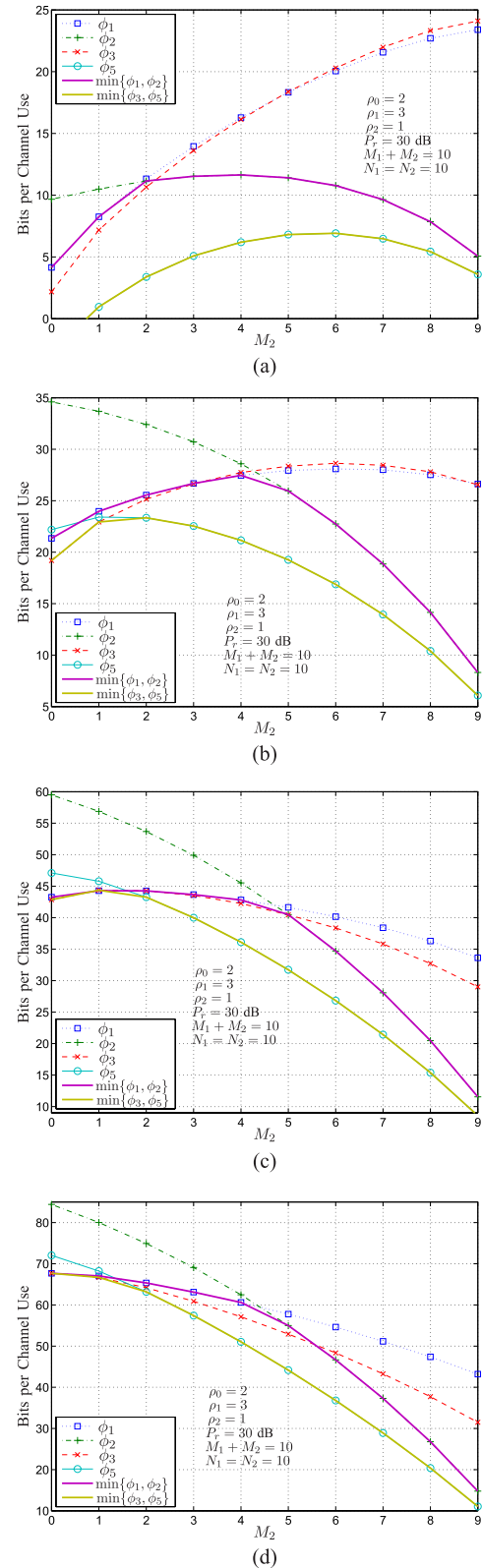
Fig. 2. The optimal number of relay transmit antennas, M_2, is monotonically decreasing with the source transmit power, P_s. A high-SNR comparison between the cut-set-based bound and the maximum rate achieved by Grassmannian DF. (a) $P_s = 15$ dB. (b) $P_s = 25$ dB. (c) $P_s = 35$ dB. (d) $P_s = 45$ dB.

$M_1 = 10$ and $M_2 = 2$, respectively. The number of relay and destination receive antennas is $N_1 = N_2 = 15$. The relative strengths of the source-destination and relay-destination links

Fig. 3. A comparison between the cut-set-based bound and the maximum rate achieved by Grassmannian DF: Case I (Section VI-A) and Case II (Section VI-B). (a) Case I: $\phi_3 \leq \phi_5$. (b) Case II: $\phi_5 \leq \phi_3$.

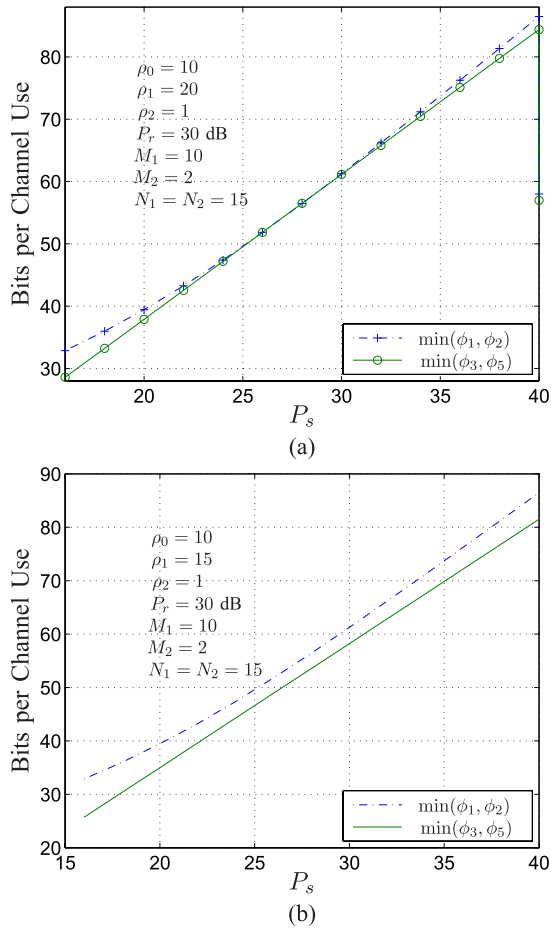are $\rho_0 = 10$ and $\rho_2 = 1$, respectively. The relative strengths of the source-relay link for the first and second cases are $\rho_1 = 20$ and $\rho_1 = 15$, respectively. From Fig. 3 it can be seen that, for the first case, Grassmannian DF achieves the cut-set-based bound when the value of P_s satisfies (58), i.e., $P_s = 27$ dB. In the second case, the gap between cut-set-based bound and the rate achieved by Grassmannian DF is minimum also at $P_s = 27$ dB, which is the value of P_s satisfying (62) with $\hat{M}_2 = 2$. This gap is about 2.85 bpcu. $\qquad\square$

VIII. CONCLUSION

In this paper we showed that isotropic Grassmannian signalling at the source without relaying suffices to achieve the degrees of freedom of the non-coherent full duplex MIMO relay channel. We derived $o(1)$ upper and lower bounds on the capacity of this channel and we devised a Grassmannian DF scheme that achieves the lower bound, and, under certain channel conditions and with a proper choice of the number of relay transmit antennas, also achieves the channel capacity, within an $o(1)$ gap. Finally, we obtained explicit expressions for computing the optimal number of relay transmit antennas for the Grassmannian DF scheme and we showed that this number is monotonically decreasing with the source transmit power.

APPENDIX A

PROOF OF (18) AND (24)

The proof of (18) and (24) follows from Lemma 12 in [23], which is stated here for completeness.

Lemma 2 ([23]): Let $\boldsymbol{X} = \boldsymbol{Q}_X \boldsymbol{R}$, where $\boldsymbol{Q}_X \in \mathbb{C}^{T \times M}$, $T > M$, is unitary and isotropically distributed. Let $\boldsymbol{R}$ be an $M \times M$ diagonal matrix satisfying $\text{E}\{\text{Tr}(\boldsymbol{R}^2)\} \leq P$, and let $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{H} + \boldsymbol{W}$, where the entries of $\boldsymbol{H} \in \mathbb{C}^{M \times N}$ and $\boldsymbol{W} \in \mathbb{C}^{T \times N}$, $N \geq M$, are i.i.d. Gaussian distributed with zero mean and unit variance. Let the singular values of $\boldsymbol{Y}$, $\{\sigma_i\}_{i=1}^N$ be ordered as $\sigma_1 > \cdots > \sigma_N$. Then, as $P \to \infty$, the joint probability density function (pdf) of the first M singular values, $\{\sigma_i\}_{i=1}^M$, converge pointwise to the joint pdf of the matrix $\boldsymbol{R}\boldsymbol{H}$ and the joint pdf of the remaining $N - M$ singular values, $\{\sigma_i\}_{i=M+1}^N$, converge pointwise to the joint pdf of the singular values of an independent $(T - M) \times (N - M)$ matrix $\boldsymbol{B}$ with i.i.d. entries that are Gaussian distributed with zero mean and unit variance. In particular, $h(\sigma_1, \ldots, \sigma_N) = h(\sigma_1, \ldots, \sigma_M) + h(\sigma_{M+1}, \ldots, \sigma_N) + o(1)$. $\qquad\square$

Invoking Lemma 2 in computing $h(\boldsymbol{Y})$ in the svd coordinates yields

$$h(\boldsymbol{Y}) = \log|\mathbb{V}_{T,N}| + \log|\mathbb{V}_N| + h(\sigma_1, \ldots, \sigma_M) \\ + h(\sigma_{M+1}, \ldots, \sigma_N) + \text{E}\{\log J_{T,N}(\sigma_1, \ldots, \sigma_N)\} + o(1), \tag{64}$$

where $J_{T,N}(\sigma_1, \ldots, \sigma_N) = \prod_{i<j}(\sigma_i^2 - \sigma_j^2)^2 \prod_{i=1}^N \sigma_i^{2(T-N)+1}$ is the Jacobian of the transformation from Cartesian to svd coordinates.

Let $\boldsymbol{Z} \in \mathbb{V}_M$ be isotropically distributed. Invoking Lemma 2 and computing $h(\boldsymbol{Z}\boldsymbol{R}\boldsymbol{H})$ and $h(\boldsymbol{B})$ in the svd coordinates yields

$$h(\boldsymbol{Z}\boldsymbol{R}\boldsymbol{H}) = \log|\mathbb{V}_N| + \log|\mathbb{V}_M| \\ + h(\sigma_1, \ldots, \sigma_M) + \text{E}\{\log J_{N,M}(\sigma_1, \ldots, \sigma_M)\}, \tag{65}$$

$$h(\boldsymbol{B}) = \log|\mathbb{V}_{T-M,N-M}| + \log|\mathbb{V}_{N-M}| + h(\sigma_{M+1}, \ldots, \sigma_N) \\ + \text{E}\{\log J_{T-N,N-M}(\sigma_{M+1}, \ldots, \sigma_N)\}. \tag{66}$$

The entries of $\boldsymbol{B}$ are i.i.d. Gaussian distributed with unit variance. Hence, $h(\boldsymbol{B}) = (T - M)(N - M)\log \pi e$.

Substituting from (65) and (66) into (64) and simplifying yields

$$h(\boldsymbol{Y}) = h(\boldsymbol{Z}\boldsymbol{R}\boldsymbol{H}) + (T - M)\,\text{E}\{\log \det(\boldsymbol{R}\boldsymbol{H}\boldsymbol{H}^\dagger\boldsymbol{R})\} \\ + (T - M)(N - M)\log \pi e + \log|\mathbb{G}_M(\mathbb{C}^T)| + o(1). \tag{67}$$

The proof of (18) and (24) follows directly from this result, with M, N, $\boldsymbol{R}$ and $\boldsymbol{H}$ properly chosen.

REFERENCES

[1] H. Yanikomeroglu, "Fixed and mobile relaying technologies for cellular networks," in *Proc. Workshop ASWN*, Jul. 2002, pp. 75–81.

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

[3] N. Fawaz and M. Médard, "A converse for the wideband relay channel with physically degraded broadcast," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Paraty, Brazil, Oct. 2011, pp. 425–429.

[4] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[5] M. Brehler and M. K. Varanasi, "Training-codes for the noncoherent multi-antenna block-Rayleigh-fading channel," in *Proc. Conf. Inf. Sci. Syst.*, Mar. 2003, pp. 1–6.

[6] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[7] R. H. Gohary and T. N. Davidson, "Non-coherent MIMO communication: Grassmannian constellations and efficient detection," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1176–1205, Mar. 2009.

[8] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[9] M. A. El-Azizy, R. H. Gohary, and T. N. Davidson, "A BICM-IDD scheme for non-coherent MIMO communication," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 541–546, Feb. 2009.

[10] G. W. Colman, R. H. Gohary, M. A. El-Azizy, T. N. Davidson, and T. J. Willink, "Quasi-Gray labelling for Grassmannian constellations," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 626–636, Feb. 2011.

[11] S. G. Srinivasan and M. K. Varanasi, "Optimal constellations for the low-SNR noncoherent MIMO block Rayleigh-fading channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 776–796, Feb. 2009.

[12] A. A. El Gamal and S. Zahedi, "Capacity of a class of relay channels with orthogonal components," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1815–1817, May 2005.

[13] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[14] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2020–2040, Jun. 2005.

[15] T. Koch and G. Kramer, "On noncoherent fading relay channels at high SNR," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2221–2241, Apr. 2013.

[16] Z. Utkovski, W. Li, and J. Lindner, "Distributed non-coherent Grassmann space-time codes for wireless relay networks," in *Proc. IEEE Int. Conf. Commun.*, Beijing, China, May 2008, pp. 1123–1128.

[17] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambrige Univ. Press, 2011.

[18] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1999.

[19] A. Ghosh, J. Zhang, J. G. Andrews, and R. Muhamed, *Fundamentals of LTE*. Upper Saddle River, NJ, USA: Prentice-Hall, 2010.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[21] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[22] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: The received signal density in closed form," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1473–1484, Jun. 2002.

[23] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 117–132, Feb. 2013.

[24] F. Hiai and D. Petz, "Asymptotic freeness almost everywhere for random matrices," *Acta Sci. Math. (Szeged)*, vol. 66, no. 2 pp. 809–834, 2000.

[25] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Found. Trend. Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.

[26] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[27] G. Kramer, "Topics in multi-user information theory," *Found. Trends Commun. Inf. Theory*, vol. 4, nos. 4–5, pp. 265–444, 2007.

[28] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 20–43, Jan. 2005.

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

[30] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2009.

Ramy H. Gohary (S'02–M'06–SM'13) is an adjunct professor and a senior research associate with the Department of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, Canada. He received the B.Eng. (Hons.) degree from Assiut University, Egypt in 1996, the M.Sc. degree from Cairo University, Egypt, in 2000, and the Ph.D. degree from McMaster University, Ontario, Canada in 2006, all in electronics and communications engineering. He received the Natural Sciences and Engineering Research Council visiting fellowship award in 2007. From 2006 to 2007, he was a Post-Doctoral fellow with McMaster University. From 2008 to 2010, he was a visiting scientist with the Terrestrial Wireless Systems Branch, Communications Research Centre, Canada. From 2010 to 2012, he was the project manager of the Carleton-BlackBerry (formerly Research In Motion (RIM)) research project, and from 2013 to present, he is a member of the Carleton-Huawei research team.

Dr. Gohary has co-authored in excess of 60 IEEE journal and conference papers. He has been involved in the organization of the IEEE Vehicular Technology Conference (VTC), Fall 2010, and the IEEE Signal Processing Advances for Wireless Communications (SPAWC), June 2014. He was also a technical program committee member for the 2013 and 2014 IEEE Wireless Communications and Netwoking Conference (WCNC).

Dr. Gohary's research interests include analysis and design of MIMO wireless communication systems, applications of optimization and geometry in signal processing and communications, information theoretic aspects of multiuser communication systems, and applications of iterative detection and decoding techniques in multiple antenna and multiuser systems.

Dr. Gohary is a registered Limited Engineering Licensee (LEL) in the province of Ontario, Canada.

Halim Yanikomeroglu (S'96–M'98–SM'12) was born in Giresun, Turkey, in 1968. He received the B.Sc. degree in Electrical and Electronics Engineering from the Middle East Technical University, Ankara, Turkey, in 1990, and the M.A.Sc. degree in Electrical Engineering (now ECE) and the Ph.D. degree in Electrical and Computer Engineering from the University of Toronto, Canada, in 1992 and 1998, respectively. He was with the R&D Group of Marconi Kominikasyon A.S., Ankara, Turkey, during 1993–1994.

Since 1998, Dr. Yanikomeroglu has been with the Department of Systems and Computer Engineering at Carleton University, Ottawa, where he is now a Full Professor. His research interests cover many aspects of wireless technologies with a special emphasis on cellular networks. Dr. Yanikomeroglu coauthored more than 50 IEEE journal papers, and has given a high number of tutorials and invited talks on wireless technologies in the leading international conferences. In recent years, Dr. Yanikomeroglu's research has been funded by Huawei, Blackberry, Samsung, Communications Research Centre of Canada (CRC), Telus, and Nortel. This collaborative research resulted in about 18 patents (granted and applied).

Dr. Yanikomeroglu has been involved in the organization of the IEEE Wireless Communications and Networking Conference (WCNC) from its inception, including serving as Steering Committee Member as well as the Technical Program Chair or Co- Chair of WCNC 2004, WCNC 2008, and WCNC 2014 to be held in Istanbul. Dr. Yanikomeroglu was the General Co-Chair of the IEEE Vehicular Technology Conference Fall 2010 held in Ottawa. Dr. Yanikomeroglu has served in the editorial boards of the IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and IEEE COMMUNICATIONS SURVEYS & TUTORIALS. He was the Chair of the IEEE's Technical Committee on Personal Communications (now called, Wireless Technical Committee). He is a Distinguished Lecturer for the IEEE Vehicular Technology Society.

Dr. Yanikomeroglu is a recipient of the Carleton University Faculty Graduate Mentoring Award in 2010, the Carleton University Graduate Students Association Excellence Award in Graduate Teaching in 2010, and the Carleton University Research Achievement Award in 2009. Dr. Yanikomeroglu spent the 2011–2012 academic year at TOBB University of Economics and Technology, Ankara, Turkey, as a Visiting Professor. He is a registered Professional Engineer in the province of Ontario, Canada.