# Grassmannian Signalling Achieves the Ergodic High SNR Capacity of the Non-Coherent MIMO Relay Channel within an SNR-Independent Gap

Ramy H. Gohary and Halim Yanikomeroglu

***Abstract*—This paper considers the ergodic non-coherent capacity of a multiple-input multiple-output frequency-flat block Rayleigh fading relay channel. It is shown that for this channel restricting the input distribution to be isotropic on a compact Grassmann manifold maximizes an upper bound on the cut-set bound at high signal-to-noise ratios (SNRs). Furthermore, Grassmannian signalling at the source achieves this bound within an SNR-independent gap. For moderate-to-high SNRs, a Grassmannian decode-and-forward (DF) relaying scheme is devised, and the optimal signalling dimensionality that minimizes the gap to the upper bound is obtained.**

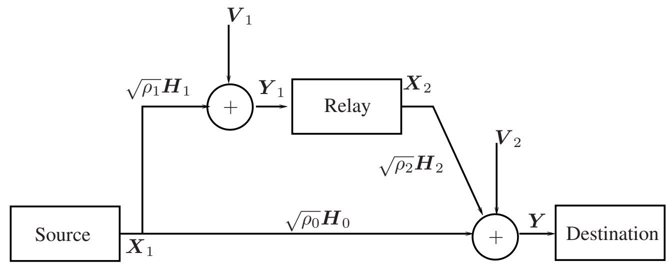


Fig. 1. A full-duplex MIMO relay channel.

## I. Introduction

One of the classifications of multiple-input multiple-output (MIMO) communication systems 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, upon reception, are used by the receivers to learn the channel [1], [2]. Sending pilot symbols consumes valuable communication resources and can be quite inefficient when the channel undergoes frequent variations [2]. In contrast with coherent systems, in non-coherent ones, no pilot symbols are used and the transmitted channel symbols are structured in such a way that enables detection without CSI being available; see e.g., [3] for high SNR communication scenarios and [4] for low SNR ones. Since non-coherent 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.

In this paper, we consider a non-coherent full-duplex MIMO relay system in which the transmitter, the relay and the receiver have multiple antennas but no CSI is available. We invoke the channel non-coherence in deriving an expression for an upper bound on the cut-set bound. It is shown that when the coherence time during which the channel remains essentially constant exceeds a certain threshold, the input distribution that maximizes this bound is the isotropic distribution on a compact Grassmann manifold [3]. Furthermore, Grassmannian signalling at the source achieves this bound within an SNR-independent gap. This implies that the gap between the rate obtained by the Grassmannian scheme and the ergodic non-coherent capacity of the MIMO channel becomes negligible as the SNR increases. For moderate-to-high SNRs, the insight drawn from the cut-set bound is used to devise a Grassmannian decode-and-forward (DF) relaying scheme, and to obtain the number of relay transmit antennas that minimizes the gap to the upper bound.

*Notation:* Random and deterministic matrices will be denoted by uppercase boldface and regular face, respectively. 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, and the direct sum operation will be denoted by $\oplus$.

## II. System Model and Preliminaries

### A. System Model

We consider the non-coherent MIMO communication system model shown in Figure 1. In this model the physical channels are assumed to be frequency-flat block Rayleigh fading with coherence time $T$, that is, each channel remains essentially fixed during $T$ channels uses and then takes on an independent realization. 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 source-destination and the source-relay channels are denoted by $\sqrt{\rho_0}\boldsymbol{H}_0 \in \mathbb{C}^{M_1 \times N_2}$ and $\sqrt{\rho_1}\boldsymbol{H}_1 \in \mathbb{C}^{M_1 \times N_1}$, respectively, and the relay-destination channel is denoted by $\sqrt{\rho_2}\boldsymbol{H}_2 \in \mathbb{C}^{M_2 \times N_2}$, where $\sqrt{\rho_i}$, $i = 0, 1, 2$, are scaling parameters that represent the relative strength of the MIMO channels with respect to noise. The additive noise at the relay and the destination is denoted by $\boldsymbol{V}_1 \in \mathbb{C}^{T \times N_1}$ and $\boldsymbol{V}_2 \in \mathbb{C}^{T \times N_2}$, respectively. The entries of $\boldsymbol{H}_i$, $i = 0, 1, 2$ and $\boldsymbol{V}_i$, $i = 1, 2$ are independent identically distributed Gaussian random variables with zero mean and unit variance. The received signals at the relay and the destination, $\boldsymbol{Y}_1$ and $\boldsymbol{Y}$, are given by

$$\boldsymbol{Y}_1 = \sqrt{\rho_1}\boldsymbol{X}_1\boldsymbol{H}_1 + \boldsymbol{V}_1, \quad (1)$$

$$\boldsymbol{Y} = \sqrt{\rho_0}\boldsymbol{X}_1\boldsymbol{H}_0 + \sqrt{\rho_2}\boldsymbol{X}_2\boldsymbol{H}_2 + \boldsymbol{V}_2, \quad (2)$$

where $\boldsymbol{X}_1 \in \mathbb{C}^{T\times M_1}$ and $\boldsymbol{X}_2 \in \mathbb{C}^{T\times M_2}$ are the signals transmitted by the source and the relay, respectively.

The signals of the source and the relay are subject to the following independent average power constraints:

$$\mathrm{E}\{\mathrm{Tr}(\boldsymbol{X}_1^\dagger \boldsymbol{X}_1)\} \le P_s, \qquad \mathrm{E}\{\mathrm{Tr}(\boldsymbol{X}_2^\dagger \boldsymbol{X}_2)\} \le P_r. \quad (3)$$

### 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 [3] $|\mathbb{V}_{p,q}| = \prod_{i=p-q+1}^{p} \frac{2\pi^i}{(i-1)!}$.

*2) The Grassmann manifold:* The 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 $Q \in \mathbb{C}^{p\times q}$ by $[Q]$, the Grassmann manifold can be expressed as $\mathbb{G}_q(\mathbb{C}^p) = \{[Q] | Q \in \mathbb{V}_{p,q}\}$, where $[Q] = \{\hat{Q} | \hat{Q} = QP, P \in \mathbb{V}_q, PP^\dagger = P^\dagger P = I_q\}$. The volume of $\mathbb{G}_q(\mathbb{C}^p)$ is given by $|\mathbb{G}_q(\mathbb{C}^p)| = \frac{|\mathbb{V}_{p,q}|}{|\mathbb{V}_q|}$.

*3) Isotropically distributed matrices:* A random matrix $\boldsymbol{Q} \in \mathbb{C}^{p\times q}$, $p \ge q$, is said to be isotropically distributed if the probability distribution $p(\boldsymbol{Q}) = p(U\boldsymbol{Q})$, for any fixed $U \in \mathbb{V}_p$.

*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 [5, Theorem 14.10.1]. Applying this bound to the considered full-duplex single relay channel yields [6]

$$R \le \frac{1}{T} \max_{p(x_1,x_2)} \min\{I(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y}), I(\boldsymbol{X}_1; \boldsymbol{Y}, \boldsymbol{Y}_1 | \boldsymbol{X}_2)\}, \quad (4)$$

where $R$ is the source transmission rate, and the division by $T$ is because $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}$ and $\boldsymbol{Y}_1$ span $T$ channel uses.

## III. Maximization of the cut-set bound

To maximizes the right hand side of (4), we will begin by considering 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 Grassmannian structure, which implies that restricting the input signal to this structure incurs no loss of optimality.

### A. Maximizing $I(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y})$

In this section we will show that, at high SNR, Grassmannian signalling maximizes $I(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y})$. To do so, we begin by showing that $[\boldsymbol{X}_1, \boldsymbol{X}_2]$ that maximizes $I(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y})$ is isotropically distributed. We revisit the results in [7].

Conditioned on $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, the received signal matrix $\boldsymbol{Y}$ is zero mean Gaussian distributed with covariance

$$\mathrm{E}\{\boldsymbol{Y}\boldsymbol{Y}^\dagger | \boldsymbol{X}_1, \boldsymbol{X}_2\} = \rho_0 N_2 \boldsymbol{X}_1 \boldsymbol{X}_1^\dagger + \rho_2 N_2 \boldsymbol{X}_2 \boldsymbol{X}_2^\dagger + N_2 I_T.$$

Defining

$$\boldsymbol{X} = \begin{bmatrix} \sqrt{\rho_0}\boldsymbol{X}_1 & \sqrt{\rho_2}\boldsymbol{X}_2 \end{bmatrix}, \quad (5)$$

we can write

$$p(\boldsymbol{Y}|\boldsymbol{X}_1, \boldsymbol{X}_2) = \frac{\exp\Big(-\mathrm{Tr}\big(\boldsymbol{Y}^\dagger (N_2 \boldsymbol{X}\boldsymbol{X}^\dagger + N_2 I_T)^{-1} \boldsymbol{Y}\big)\Big)}{\pi^{TN_2} \det^{N_2}(N_2 \boldsymbol{X}\boldsymbol{X}^\dagger + N_2 I_T)}.$$

Hence, $p(\boldsymbol{Y}|\boldsymbol{X}_1, \boldsymbol{X}_2) = p(\boldsymbol{Y}|\boldsymbol{X})$, where $\boldsymbol{X}$ is given in (5). Since this distribution depends on the transmitted signals only through the $T \times T$ matrix $\boldsymbol{X}\boldsymbol{X}^\dagger$, it follows that increasing $M_1 + M_2$ beyond $T$ does not yield an additional rate gain. We will later obtain stricter constraints on $T$, but for now, it suffices to see that the matrix $\boldsymbol{X}$ is either "tall" or "square", but not "fat". In addition, let $\Psi \in \mathbb{C}^{(M_1+M_2)\times(M_1+M_2)}$ and $\Phi \in \mathbb{C}^{T\times T}$ be any two deterministic unitary square matrices, then in correspondence with properties 3 and 4 in [7], we have

$$p(\boldsymbol{Y}|\boldsymbol{X}_1, \boldsymbol{X}_2) = p(\boldsymbol{Y}|\boldsymbol{X}) = p(\boldsymbol{Y}|\boldsymbol{X}\Psi^\dagger) \quad (6)$$

$$p(\Phi\boldsymbol{Y}|\Phi\boldsymbol{X}_1, \Phi\boldsymbol{X}_2) = p(\Phi\boldsymbol{Y}|\Phi\boldsymbol{X}) = p(\boldsymbol{Y}|\boldsymbol{X}). \quad (7)$$

A direct consequence of these properties is given in Lemma 1 in [7], which asserts that if some $p_0(\boldsymbol{X})$ achieves the maximum of $I(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y})$, then this maximum is also achieved by $p_0(\Phi^\dagger \boldsymbol{X} \Psi)$, where $\Phi$ and $\Psi$ are arbitrary deterministic unitary square matrices with dimensions $T \times T$ and $(M_1 + M_2) \times (M_1 + M_2)$, respectively. This result together with Lemma 3 in [7], yield the counterpart of Theorem 2 therein for $I(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y})$. In particular, the signal matrix $\boldsymbol{X}$ in (8) that maximizes $I(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y})$ can be expressed as

$$\boldsymbol{X} = \boldsymbol{Q}_X \boldsymbol{D}, \quad (8)$$

where $\boldsymbol{Q}_X = [\boldsymbol{Q}_{X_1} \ \boldsymbol{Q}_{X_2}] \in \mathbb{V}_{T, M_1+M_2}$ is isotropically distributed and $\boldsymbol{D} = \boldsymbol{D}_1 \oplus \boldsymbol{D}_2$ is independent of $\boldsymbol{Q}_X$, where $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are diagonal matrices with non-negative entries.

### B. Maximization of an upper bound on $I(\boldsymbol{X}_1; \boldsymbol{Y}, \boldsymbol{Y}_1 | \boldsymbol{X}_2)$

The second argument of the minimization in (4) is given by $I(\boldsymbol{X}_1; \boldsymbol{Y}, \boldsymbol{Y}_1 | \boldsymbol{X}_2)$. Finding an input distribution that maximizes this expression appears to be difficult. This is because, with $\boldsymbol{X}_2$ fixed, the matrix $\sqrt{\rho_2}\boldsymbol{X}_2\boldsymbol{H}_2 + \boldsymbol{V}_2$ resembles additive Gaussian noise with potentially high power and a non-isotropic distribution. Hence, we circumvent this difficulty by deriving a bound on this expression. To find such a bound, we invoke the non-coherence of the channel to show that

$$I(\boldsymbol{X}_1; \boldsymbol{Y}, \boldsymbol{Y}_1 | \boldsymbol{X}_2) \le I(\boldsymbol{X}_1; \boldsymbol{Y}, \boldsymbol{Y}_1 | \boldsymbol{X}_2, \boldsymbol{H}_2). \quad (9)$$

This inequality is proved in [8].

Our goal now is to determine the structure of the distribution that maximizes $I(\boldsymbol{X}_1; \boldsymbol{Y}, \boldsymbol{Y}_1 | \boldsymbol{X}_2, \boldsymbol{H}_2)$.

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

$$I(\boldsymbol{X}_1; \boldsymbol{Y}, \boldsymbol{Y}_1 | \boldsymbol{X}_2, \boldsymbol{H}_2) = I(\boldsymbol{X}_1[\sqrt{\rho_1}\boldsymbol{H}_1 \ \sqrt{\rho_0}\boldsymbol{H}_0] + [\boldsymbol{V}_1 \ \boldsymbol{V}_2]; \boldsymbol{X}_1). \quad (10)$$

This expression is equivalent to the mutual information of a channel with input $\boldsymbol{X}_1$, channel matrix $[\sqrt{\rho_1}\boldsymbol{H}_1 \ \sqrt{\rho_0}\boldsymbol{H}_0]$, and additive noise $[\boldsymbol{V}_1 \ \boldsymbol{V}_2]$. Applying the techniques in [7] to this channel and using the counterparts of properties in (6) and (7), it can be readily concluded that the signal matrix, $\boldsymbol{X}_1$, that maximizes can be written in the form

$$\boldsymbol{X}_1 = \boldsymbol{Q}_{X_1} \boldsymbol{D}_1, \quad (11)$$

where $\boldsymbol{Q}_{X_1} \in \mathbb{V}_{T,M_1}$ is isotropically distributed and $\boldsymbol{D}_1$ is an independent diagonal matrix with non-negative entries.

Comparing (11) with (8), it can be seen that for the matrix $\boldsymbol{X}$ in (5) to be in the form in (8), with $\boldsymbol{Q}_X$ isotropically distributed and independent of $\boldsymbol{D}$, the matrix $\boldsymbol{X}_1$ must be in the form in (11), with $\boldsymbol{Q}_{X_1}$ isotropically distributed and independent of $\boldsymbol{D}_1$; see Appendix A in [9]. Conversely, when $\boldsymbol{X}_1$ assumes the form in (11) with $\boldsymbol{Q}_{X_1}$ isotropically distributed and independent of $\boldsymbol{D}_1$, choosing the matrix $\boldsymbol{Q}_{X_2}$ to be isotropically distributed in the null space of $\boldsymbol{Q}_{X_1}$ and independent of $\boldsymbol{D}_2$ does not affect $I(\boldsymbol{X}_1;\boldsymbol{Y},\boldsymbol{Y}_1|\boldsymbol{X}_2,\boldsymbol{H}_2)$, and maximizes $I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y})$. Hence, it can be concluded that choosing $\boldsymbol{Q}_X$ to be isotropically distributed and independent of $\boldsymbol{D}$ maximizes both $I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y})$ and $I(\boldsymbol{X}_1;\boldsymbol{Y},\boldsymbol{Y}_1|\boldsymbol{X}_2,\boldsymbol{H}_2)$. Therefore, the structure of $\boldsymbol{X}$ in (8) also maximizes $\min\{I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y}), I(\boldsymbol{X}_1;\boldsymbol{Y},\boldsymbol{Y}_1|\boldsymbol{X}_2,\boldsymbol{H}_2)\}$. We will refer to this quantity as the cut-set-based bound to distinguish it from the cut-set bound given in (4).

## IV. The cut-set-based bound at high SNR

In the previous section we showed that the matrix $\boldsymbol{X}$ that maximizes the cut-set-based bound,

$$\min\{I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y}), I(\boldsymbol{X}_1;\boldsymbol{Y},\boldsymbol{Y}_1|\boldsymbol{X}_2,\boldsymbol{H}_2)\}, \tag{12}$$

has the form in (8). We will use this and the results in [3] to evaluate (12) at high SNR. The analysis in [3] indicates that efficient high SNR non-coherent communication can be achieved if the number of transmit antennas does not exceed that of receive antennas and the coherence time, $T$, exceeds a certain threshold. Applying these results to the multiple access channel corresponding to $I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y})$ and the broadcast channel corresponding to $I(\boldsymbol{X}_1;\boldsymbol{Y},\boldsymbol{Y}_1|\boldsymbol{X}_2,\boldsymbol{H}_2)$ in (12) yields the following constraints on $M_1, M_2, N_1, N_2$ and $T$, which will be assumed to hold throughout.

$$M_1+M_2 \le N_2, \qquad \text{and} \tag{13}$$

$$T \ge M_1+N_2+\max\{M_2,N_1\}. \tag{14}$$

### A. Computing a high SNR bound on $I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y})$

To compute $I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y})$ at high SNR, we compute $h(\boldsymbol{Y}|\boldsymbol{X}_1,\boldsymbol{X}_2)$ and $h(\boldsymbol{Y})$. Conditioned on $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, the matrix $\boldsymbol{Y}$ is Gaussian distributed. Defining $J_1 \triangleq \sqrt{\rho_0} I_{M_1} \oplus \sqrt{\rho_2} I_{M_2}$, it can be shown that

$$\begin{aligned} &h(\boldsymbol{Y}|\boldsymbol{X}_1,\boldsymbol{X}_2) \\ &\quad = N_2 T\log \pi e + N_2 \,\mathrm{E}\{\log\det(J_1^2\boldsymbol{D}^2 + I_{M_1+M_2})\}, \end{aligned} \tag{15}$$

Computing $h(\boldsymbol{Y})$ is significantly more difficult than computing $h(\boldsymbol{Y}|\boldsymbol{X}_1,\boldsymbol{X}_2)$. This difficulty is circumvented in [3] by noting that $h(\boldsymbol{Y}) = h(\boldsymbol{X}\boldsymbol{H}_{02}+\boldsymbol{V}_2) \approx h(\boldsymbol{X}\boldsymbol{H}_{02}+\boldsymbol{V}_2^{\perp})$, where $\boldsymbol{H}_{02} = \begin{bmatrix}\boldsymbol{H}_0^{\dagger} & \boldsymbol{H}_2^{\dagger}\end{bmatrix}^{\dagger}$ and $\boldsymbol{V}_2^{\perp} = \boldsymbol{Q}_X^{\perp}(\boldsymbol{Q}_X^{\perp})^{\dagger}\boldsymbol{V}_2$ is the noise component in the null space of $\boldsymbol{X}$, spanned by $\boldsymbol{Q}_X^{\perp}$. To proceed with the analysis, we will use the following lemma.

*Lemma 1:* For any isotropically distributed $\Xi \in \mathbb{C}^{\ell\times n}$, any deterministic unitary matrix, $\Phi \in \mathbb{V}_\ell$, and any arbitrarily distributed random matrix $\boldsymbol{G} \in \mathbb{C}^{n\times m}$ with $n \le m$ and $\boldsymbol{G}$ almost surely full row rank,

$$p(\Phi\Xi\boldsymbol{G}) = p(\Xi\boldsymbol{G}).$$

*Proof:* See Appendix A in [8]. ∎

This lemma says that the isotropic distribution of $\boldsymbol{X}$ induces an isotropic distribution on $\boldsymbol{X}\boldsymbol{H}_{02}$. Hence, invoking the change of variables in Lemma 6 and Appendix E in [3], yields

$$\begin{aligned} &h(\boldsymbol{X}\boldsymbol{H}_{02}+\boldsymbol{V}_2^{\perp}) \\ &\approx h(\boldsymbol{Z}\boldsymbol{D}J_1\boldsymbol{H}_{02}) + \log|\mathbb{G}_{M_1+M_2}(\mathbb{C}^T)| \\ &\quad + (T-M_1-M_2)\,\mathrm{E}\{\log\det(\boldsymbol{H}_{02}\boldsymbol{H}_{02}^{\dagger})\} \\ &\quad + (T-M_1-M_2)\,\mathrm{E}\{\log\det(J_1^2\boldsymbol{D}^2 + I_{M_1+M_2})\} \\ &\quad + (N_2-M_1-M_2)(T-M_1-M_2)\log\pi e, \end{aligned} \tag{16}$$

where $\boldsymbol{Z} \in \mathbb{C}^{M_1+M_2\times M_1+M_2}$ is an isotropically distributed unitary matrix. The approximation follows from the high SNR assumption, which implies that $\mathrm{E}\{\log\det(J_1^2\boldsymbol{D}^2)\} \approx \mathrm{E}\{\log\det(J_1^2\boldsymbol{D}^2 + I_{M_1+M_2})\}$. This approximation is only valid when $\boldsymbol{D}$ is full rank; a condition which we will show later to be satisfied by the optimal $\boldsymbol{D}$ for $M_1, M_2, N_1, N_2$ and $T$ satisfying (13) and (14).

To bound $h(\boldsymbol{Z}\boldsymbol{D}J_1\boldsymbol{H}_{02})$, we use a result from [10], [11] pertaining to isotropically distributed matrices to show that

$$h(\boldsymbol{Z}J_1\boldsymbol{D}\boldsymbol{H}_{02}) \le N_2(M_1+M_2)\log\frac{\pi e\,\mathrm{Tr}\big(J_1^2\,\mathrm{E}\{\boldsymbol{D}^2\}\big)}{M_1+M_2}. \tag{17}$$

This upper bound on $h(\boldsymbol{Z}\boldsymbol{D}J_1\boldsymbol{H}_{02})$ is maximized by choosing $\boldsymbol{D}$ to be deterministic; i.e., $\boldsymbol{D} = D$. Furthermore, the concavity of the $\log\det(\cdot)$ function and Jensen's inequality imply that $\mathrm{E}\{\log\det(J_1^2\boldsymbol{D}^2 + I_{M_1+M_2})\} \le \log\det\big(J_1^2\,\mathrm{E}\{\boldsymbol{D}^2\} + I_{M_1+M_2}\big)$, with equality if and only if $\boldsymbol{D}$ is deterministic. Since $T$ satisfies (14), it can be shown that the coefficient of the $\log\det(\cdot)$ function in the expression of $I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y})$ is non-negative. Hence, maximizing the bound on $I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y})$ requires the matrix $\boldsymbol{D}$ to be deterministic. We will later show that setting $\boldsymbol{D} = D$ also maximizes the second term in the minimization in (12).

Substituting from (17) and setting $\boldsymbol{D} = D$ yields

$$\begin{aligned} &I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y}) \le N_2(M_1+M_2)\log\Big(\frac{\mathrm{Tr}\big(J_1^2D^2\}\big)}{M_1+M_2}\Big) + \\ &\log|\mathbb{G}_{M_1+M_2}(\mathbb{C}^T)| + (T-M_1-M_2)\,\mathrm{E}\{\log\det(\boldsymbol{H}_{02}\boldsymbol{H}_{02}^{\dagger})\} \\ &\quad + (T-M_1-M_2-N_2)\log\det(J_1^2D^2+I_{M_1+M_2}) \\ &\quad - (M_1+M_2)(T-M_1-M_2)\log\pi e. \end{aligned} \tag{18}$$

### B. Computing a high SNR bound on $I(\boldsymbol{X}_1;\boldsymbol{Y}_1,\boldsymbol{Y}|\boldsymbol{X}_2,\boldsymbol{H}_2)$

To compute the desired expression, we note that the expressions in (1) and (2) can be compactly written as

$$[\boldsymbol{Y}_1\,\boldsymbol{Y}] = \boldsymbol{X}_1[\boldsymbol{H}_1\,\boldsymbol{H}_0]\,J_2 + \sqrt{\rho_2}\boldsymbol{X}_2[0\,\boldsymbol{H}_2] + [\boldsymbol{V}_1\,\boldsymbol{V}_2],$$

where $J_2 = \sqrt{\rho_1}I_{N_1} \oplus \sqrt{\rho_0}I_{N_2}$. Conditioned on $(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{H}_2)$, the entries of the matrix $(\boldsymbol{Y}_1,\boldsymbol{Y})$ are Gaussian and hence it can be shown that

$$\begin{aligned} &h(\boldsymbol{Y}_1,\boldsymbol{Y}|\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{H}_2) = N_1\,\mathrm{E}\{\log\det(\rho_1\boldsymbol{D}_1^2+I_{M_1}) \\ &\quad + N_2\log\det(\rho_0\boldsymbol{D}_1^2+I_{M_1})\} + T(N_1+N_2)\log\pi e. \end{aligned} \tag{19}$$

Now, the high SNR assumption implies that $h(\boldsymbol{Y}_1,\boldsymbol{Y}|\boldsymbol{X}_2,\boldsymbol{H}_2) \approx h\Big(\boldsymbol{X}_1[\boldsymbol{H}_1\ \boldsymbol{H}_0]\,J_2 + [\boldsymbol{V}_1\ \boldsymbol{V}_2]^{\perp}\Big)$, where $[\boldsymbol{V}_1\ \boldsymbol{V}_2]^{\perp} = \boldsymbol{Q}_{X_1}^{\perp}(\boldsymbol{Q}_{X_1}^{\perp})^{\dagger}\,[\boldsymbol{V}_1\ \boldsymbol{V}_2]$ is the component of the noise matrix $[\boldsymbol{V}_1\ \boldsymbol{V}_2]$ in the null space of $\boldsymbol{X}_1$. Using Lemma 1 and the change of variables lemma from [3] yields

$$\begin{aligned} h(\boldsymbol{X}_1\,[\boldsymbol{H}_1\ \boldsymbol{H}_0]\,J_2 + [\boldsymbol{V}_1\ \boldsymbol{V}_2]^{\perp}) &= h(\boldsymbol{W}\boldsymbol{D}_1\,[\boldsymbol{H}_1\ \boldsymbol{H}_0]\,J_2) \\ &+ \log|\mathbb{G}_{M_1}(\mathbb{C}^T)| + (T-M_1)\,\mathrm{E}\{\log\det(\boldsymbol{D}_1^2)\} \\ &+ (T-M_1)\,\mathrm{E}\{\log\det(\rho_1\boldsymbol{H}_1\boldsymbol{H}_1^{\dagger} + \rho_0\boldsymbol{H}_0\boldsymbol{H}_0^{\dagger})\} \\ &+ (N_1+N_2-M_1)(T-M_1)\log\pi e, \quad (20) \end{aligned}$$

where $\boldsymbol{W} \in \mathbb{C}^{M_1\times M_1}$ is an isotropically distributed unitary matrix independent of $\boldsymbol{D}_1$ and $[\boldsymbol{H}_1\ \boldsymbol{H}_0]$. Defining $\lambda = \frac{\max\{M_2,N_1\}}{N_2+\max\{M_2,N_1\}}$ and invoking (14) and the concavity of the $\log\det(\cdot)$ function yields

$$\begin{aligned} &\lambda(T-M_1)\,\mathrm{E}\{\log\det(\rho_1\boldsymbol{D}_1^2 + I_{M_1})\} \\ &+ (1-\lambda)(T-M_1)\,\mathrm{E}\{\log\det(\rho_0\boldsymbol{D}_1^2 + I_{M_1})\} \le \\ &\lambda(T-M_1)\log\det(\rho_1\,\mathrm{E}\{\boldsymbol{D}_1^2\} + I_{M_1}) \\ &+ (1-\lambda)(T-M_1)\log\det(\rho_0\,\mathrm{E}\{\boldsymbol{D}_1^2\} + I_{M_1}). \end{aligned}$$

The above inequality holds with equality if and only if $\mathrm{E}\{\boldsymbol{D}_1^2\} = D_1^2$, i.e., $\boldsymbol{D}_1$ is deterministic. Using the corollary of Theorem 9.6.4 in [5], and applying an approach similar to the one used in deriving (17), it can be shown that

$$\begin{aligned} I(\boldsymbol{X}_1;\boldsymbol{Y}_1,\boldsymbol{Y}|\boldsymbol{X}_2,\boldsymbol{H}_2) &\le M_1(N_1+N_2)\log\Big(\frac{\mathrm{Tr}(D_1^2)}{M_1}\Big) \\ &+ \big(\lambda(T-M_1)-N_1\big)\log\det(\rho_1 D_1^2 + I_{M_1}) \\ &+ \big((1-\lambda)(T-M_1)-N_2\big)\log\det(\rho_0 D_1^2 + I_{M_1}) \\ &+ (T-M_1)\,\mathrm{E}\{\log\det(\rho_1\boldsymbol{H}_1\boldsymbol{H}_1^{\dagger} + \rho_0\boldsymbol{H}_0\boldsymbol{H}_0^{\dagger})\} + \\ M_1\Big(\big(N_1-\lambda(T-M_1)\big)&\log\rho_1 + \big(N_2-(1-\lambda)(T-M_1)\big)\log\rho_0\Big) \\ &- M_1(T-M_1)\log\pi e + \log|\mathbb{G}_{M_1}(\mathbb{C}^T)|. \quad (21) \end{aligned}$$

### C. Maximization and tightness of bounds

For Jensen's inequality used in deriving the above bounds to hold, the matrix $\boldsymbol{D}$ must be deterministic, i.e., $\boldsymbol{D} = D$. The bound in (18) is maximized by $D_1 = \sqrt{\frac{P_s}{M_1}}I_{M_1}$ and $D_2 = \sqrt{\frac{P_r}{M_2}}I_{M_2}$, and the bound in (21) is maximized by $D_1 = \sqrt{\frac{P_s}{M_1}}I_{M_1}$. Since both bounds are maximized by the same choice of $D$, this choice maximizes their minimum. Substituting this choice in (18) yields

$$\begin{aligned} I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y}) &\le N_2(M_1+M_2)\log\Big(\frac{\rho_0P_s+\rho_2P_r}{M_1+M_2}\Big) \\ &+ M_1(T-M_1-M_2-N_2)\log\Big(\frac{\rho_0P_s}{M_1}+1\Big) \\ &+ M_2(T-M_1-M_2-N_2)\log\Big(\frac{\rho_2P_r}{M_2}+1\Big) \\ +(T-M_1-M_2)\,&\mathrm{E}\{\log\det(\boldsymbol{H}_{02}\boldsymbol{H}_{02}^{\dagger})\} + \log|\mathbb{G}_{M_1+M_2}(\mathbb{C}^T)| \\ &- (M_1+M_2)(T-M_1-M_2)\log\pi e \triangleq \phi_1. \quad (22) \end{aligned}$$

This bound is generally loose unless $P_s = P_r$ and $\rho_0 = \rho_2$. It is only in that special case that the bound in (22) approaches holding with equality as the SNR increases.

Substituting for $D_1 = \sqrt{\frac{P_s}{M_1}}I_{M_1}$ in (21) yields

$$\begin{aligned} I(\boldsymbol{X}_1;\boldsymbol{Y}_1,\boldsymbol{Y}|\boldsymbol{X}_2,\boldsymbol{H}_2) &\le M_1(N_1+N_2)\log\Big(\frac{P_s}{M_1}\Big) \\ &+ M_1\big(\lambda(T-M_1)-N_1\big)\log\Big(\frac{\rho_1P_s}{M_1}+1\Big) + \\ M_1\big((1-\lambda)(T-M_1)-N_2\big)&\log\Big(\frac{\rho_0P_s}{M_1}+1\Big) + \\ (T-M_1)\,\mathrm{E}\{\log\det(\rho_1\boldsymbol{H}_1\boldsymbol{H}_1^{\dagger}&+\rho_0\boldsymbol{H}_0\boldsymbol{H}_0^{\dagger})\} + \log|\mathbb{G}_{M_1}(\mathbb{C}^T)| \\ - M_1(T-M_1)\log\pi e &+ M_1\Big(N_1-\lambda(T-M_1)\Big)\log\rho_1 \\ &+ M_1\Big(N_2-(1-\lambda)(T-M_1)\Big)\log\rho_0 \triangleq \phi_2. \quad (23) \end{aligned}$$

Unlike the bound in (22), this bound can be shown to be asymptotically tight for any $\rho_i$, $i = 0, 1, 2$ and $P_r$.

### D. A lower bound on $\max_{p(x_1,x_2)} I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y})$

To analyze the degrees of freedom, we derive a lower bound on $I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y})$. We will show that both the upper bound in (22) and the lower bound obtained below yield the same number of degrees of freedom, which implies that the bound in (22) is asymptotically tight, up to an SNR-independent term.

The looseness of the bound in (22) follows from the looseness of the upper bound on $h(\boldsymbol{Z}J_1\boldsymbol{D}\boldsymbol{H}_{02})$ in (17). Hence, to derive a lower bound on $\max_{p(x_1,x_2)} I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y})$, we will derive a lower bound on $h(\boldsymbol{Z}J_1\boldsymbol{D}\boldsymbol{H}_{02})$. Using the fact that conditioning reduces entropy, it can shown that

$$\begin{aligned} h(\boldsymbol{Z}J_1\boldsymbol{D}\boldsymbol{H}_{02}) &\ge N_2(M_1+M_2)\log\pi e + N_2M_1\log\rho_0 \\ &+ N_2M_2\log\rho_2 + 2N_2\,\mathrm{E}\{\log\det\boldsymbol{D}\}. \quad (24) \end{aligned}$$

Substituting from (24) yields the following bound:

$$\begin{aligned} \max_{p(x_1,x_2)} I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y}) &\ge \log|\mathbb{G}_{M_1+M_2}(\mathbb{C}^T)| \\ &+ (T-M_1-M_2)\,\mathrm{E}\{\log\det(\boldsymbol{H}_{02}\boldsymbol{H}_{02}^{\dagger})\} \\ +M_1(T-M_1-M_2-N_2)&\log\Big(\frac{\rho_0P_s}{M_1}+1\Big) + N_2M_1\log\Big(\frac{P_s\rho_0}{M_1}\Big) \\ +M_2(T-M_1-M_2-N_2)&\log\Big(\frac{\rho_2P_r}{M_2}+1\Big) + N_2M_2\log\Big(\frac{P_r\rho_2}{M_2}\Big) \\ &- (M_1+M_2)(T-M_1-M_2)\log\pi e \triangleq \phi_3. \quad (25) \end{aligned}$$

### E. A degrees of freedom view on the cut-set bound

Let $P_r/P_s$ be finite, and let $\chi_f = \lim_{P_s\to\infty}\frac{f(P_s)}{\log P_s}$ be the number of degrees of freedom of a function $f(P_s)$. Then,

$$\begin{aligned} \chi_{\phi_1} &= \chi_{\phi_3} = (M_1+M_2)(T-M_1-M_2), \\ \chi_{\phi_2} &= M_1(T-M_1), \end{aligned}$$

where $\phi_1$ is defined in (22), $\phi_2$ in (23), and $\phi_3$ in (25). Using the bounds on $N_2$ in (13) and $T$ in (14), we have

$$\chi_{\phi_1} - \chi_{\phi_2} = M_2\max\{M_2,N_1\} > 0.$$

In other words, at sufficiently high SNRs, the broadcast cut, corresponding to $I(\boldsymbol{X}_1;\boldsymbol{Y},\boldsymbol{Y}_1|\boldsymbol{H}_2,\boldsymbol{X}_2)$, is the constraining cut in (12). Using Grassmannian signalling at the source, i.e., with $\boldsymbol{X}_1$ isotropically distributed on $\mathbb{G}_{M_1}(\mathbb{C}^T)$, suffices for the maximum degrees of freedom of the cut-set bound to be achieved and hence, for (23) to hold with equality.

## V. Grassmannian Decode-and-Forward

While Grassmannian signalling at the source suffices for achieving the maximum degrees of freedom of the non-coherent relay channel, using Grassmannian DF relaying can be shown to be more beneficial at moderate-to-high SNRs.

### *A. Codebook structure, encoding and decoding*

The proposed codebook structure is motivated by the structure that maximizes the bounds on the cut-set-based bound in the previous section. The main features of this structure are: 1) The relay codewords are isotropically distributed on $\mathbb{G}_{M_2}(\mathbb{C}^T)$; and 2) Given a relay codeword, the source codewords corresponding to it are isotropically distributed in its $T \times M_1$ null space. This space can be shown to be isomorphic to $\mathbb{G}_{M_1}(\mathbb{C}^{T-M_2})$. We will now derive a lower bound on the rate that can be achieved by the Grassmannian DF scheme.

### *B. Evaluation of the DF rate*

The achievable rate of the general DF scheme is given by [6]

$$\begin{aligned} R_{\text{DF}} &= \max_{p(x_1,x_2)} \min\{I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y}), I(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{X}_2)\} \\ &\geq \frac{1}{T} \min_{[\boldsymbol{X}_1\ \boldsymbol{X}_2]\in\mathbb{F}} \{I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y}), I(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{X}_2)\}, \quad (26) \end{aligned}$$

where the set $\mathbb{F}$ is defined as:

$$\mathbb{F} \triangleq \Big\{[\boldsymbol{X}_1\ \boldsymbol{X}_2] = \boldsymbol{Q}_X\boldsymbol{D}|\boldsymbol{Q}_X \text{ isotropically distributed on } \mathbb{G}_{M_1+M_2}(\mathbb{C}^T), \boldsymbol{D} = \sqrt{P_s/M_1}I_{M_1} \oplus \sqrt{P_r/M_2}I_{M_2}\Big\}. \quad (27)$$

The first term in the minimization in (26) is the same as the corresponding term in (4). Using the Grassmannian structure in Sections IV and V-A, and the definition of $\phi_3$ in (25) yields

$$I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y})\Big|_{[\boldsymbol{Q}_{X_1}\ \boldsymbol{Q}_{X_2}]\in\mathbb{G}_{M_1+M_2}(\mathbb{C}^T)} \geq \phi_3.$$

To evaluate the second term in the minimization (26), we note that the encoding procedure implies that $\boldsymbol{X}_2 \to \boldsymbol{X}_1 \to \boldsymbol{Y}_1$. Hence, we have

$$h(\boldsymbol{Y}_1|\boldsymbol{X}_1,\boldsymbol{X}_2) = TN_1\log\pi e + N_1M_1\log\Big(\frac{\rho_1P_s}{M_1}+1\Big).$$

To compute $h(\boldsymbol{Y}_1|\boldsymbol{X}_2)$, note that by construction, $\boldsymbol{Q}_{X_1}$ lies in the null space of $\boldsymbol{Q}_{X_2}$. Hence, $\boldsymbol{Q}_{X_1}$ can be expressed as

$$\boldsymbol{Q}_{X_1} = \boldsymbol{Q}_{X_2}^{\perp}\boldsymbol{\Omega}, \quad (28)$$

where $\boldsymbol{Q}_{X_2}^{\perp} \in \mathbb{V}_{T,T-M_2}$ is any representative of the null space of $\boldsymbol{Q}_{X_2}$, and $\boldsymbol{\Omega} \in \mathbb{V}_{T-M_2,M_1}$. Computing a lower bound on $h(\boldsymbol{Y}_1|\boldsymbol{X}_2)$ when $N_1 > M_1$ is more involved. Hence, herein, we restrict ourselves to the case of $N_1 = M_1$. The more general case will be considered later. Using the structure in (28) and the change of variable lemma in [3] we obtain

$$\begin{aligned} I(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{X}_2) &\geq M_1(T-M_1-M_2)\log\frac{\rho_1P_s}{M_1} \\ &\quad - M_1(T-M_1-M_2)\log\pi e + \log|\mathbb{G}_{M_1}(\mathbb{C}^{T-M_2})| \\ &\quad + (T-M_1-M_2)\,\mathrm{E}\{\log\det\boldsymbol{H}_1\boldsymbol{H}_1^{\dagger}\} \triangleq \phi_4. \quad (29) \end{aligned}$$

### *C. Optimizing $M_2$ for the Grassmannian DF scheme*

Comparing $\chi_{\phi_4}$ with $\chi_{\phi_2}$, it can be seen that $\chi_{\phi_2} - \chi_{\phi_4} = M_1M_2$. Hence, at asymptotically high SNRs, choosing $M_2 = 0$ is optimal. At moderate-to-high SNRs, either $I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y}) \leq I(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{X}_2)$, or if $I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y}) > I(\boldsymbol{X}_1;\boldsymbol{Y}_1|\boldsymbol{X}_2)$ the SNR-independent terms of $\phi_2$ and $\phi_4$ are not negligible. In the first case, the optimal $M_2$ satisfies $M_1 + M_2 = \lfloor T/2 \rfloor$, whereas in the second case, it can be shown that $\phi_2 - \phi_4$ can be expressed as the difference between two monotonically increasing functions. In that case, the optimal $M_2$ can be computed by rounding the solution of $\frac{\partial(\phi_2-\phi_4)}{\partial M_2} = 0$.

## VI. Conclusion

In this paper, it was shown that, in the high SNR regime, using isotropically distributed Grassmannian symbols at the source achieves an upper bound on the capacity of the non-coherent MIMO relay network withing an SNR-independent gap. In that case, it was also shown that the rate provided by the bound is constrained by the broadcast cut, rather than the multiple access one. For moderate-to-high SNR operation, a Grassmannian DF scheme was devised and its optimal signalling dimensionality was obtained.

## Acknowledgments

This work is supported, in part, by Research In Motion (RIM) and, in part, by the Collaborative Research and Development (CRD) grant by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

The authors would like to thank Dr. S.C. Bontu and Dr. J. Womack for their comments and support.

## References

[1] M. Brehler and M. K. Varanasi, "Training-codes for the noncoherent multi-antenna block-Rayleigh-fading channel," in *Proc. conf. information Sciences and Systems*, John Hopkins Univ., Mar. 2003.

[2] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, pp. 951–963, Apr. 2003.

[3] 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, pp. 359–383, Feb. 2002.

[4] 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, pp. 776–796, Feb. 2009.

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

[6] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 572–584, Sept. 1979.

[7] 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, pp. 139–157, Jan. 1999.

[8] R. H. Gohary and H. Yanikomeroglu, "Grassmannian decode-and-forward achieves the ergodic high SNR capacity of the non-coherent MIMO relay channel within a constant gap," 2012. Available at: http://www.sce.carleton.ca/~gohary.

[9] 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, pp. 1473–1484, June 2002.

[10] F. Hiai and D. Petz, "Asymptotic freeness almost everywhere for random matrices," *Acta. Sci. Math. (Szeged)*, vol. 66, pp. 809–834, 2000.

[11] 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.