

The Generalized Degrees of Freedom Region of the MIMO Interference Channel

Sanjay Karmakar Mahesh K. Varanasi

Abstract

The generalized degrees of freedom (GDoF) region of the MIMO Gaussian interference channel (IC) is obtained for the general case of an arbitrary number of antennas at each node and where the signal-to-noise ratios (SNR) and interference-to-noise ratios (INR) vary with arbitrary exponents to a nominal SNR. The GDoF region reveals various insights through the joint dependence of optimal interference management techniques (at high SNR) on the SNR exponents that determine the relative strengths of direct-link SNRs and cross-link INRs and the numbers of antennas at the four terminals. For instance, it permits an in-depth look at the issue of rate-splitting and partial decoding and it reveals that, unlike in the scalar IC, treating interference as noise is not always GDoF-optimal even in the very weak interference regime. Moreover, while the DoF-optimal strategy that relies just on transmit/receive zero-forcing beamforming and time-sharing is not GDoF optimal (and thus has an unbounded gap to capacity), the precise characterization of the very strong interference regime – where single-user DoF performance can be achieved simultaneously for both users– depends on the relative numbers of antennas at the four terminals and thus deviates from what it is in the SISO case. For asymmetric numbers of antennas at the four nodes the shape of the symmetric GDoF curve can be a “distorted W” curve to the extent that for certain MIMO ICs it is a “V” curve.

This work was supported in part by NSF Grant CCF-0728955. The authors are with the Department of Electrical, Computer, and Energy Engineering, University of Colorado, Boulder, CO 80309-0425 USA (e-mail: sanjay.karmakar, varanasi@colorado.edu).

I. INTRODUCTION

In spite of research spanning over three decades, the capacity of the interference channel (IC) has been characterized only for some special cases (cf. [1], [2], [3], [4], [5], [6], [7], [8]) and remains an open problem to date for the general case. However, the understanding of the Gaussian interference channel has been significantly enriched through those works and through capacity approximations in recent years. In particular, the capacity region was characterized to within one bit for the single-antenna (SISO) IC [9] and to within a constant number of bits for the more general case of MIMO Gaussian ICs with an arbitrary number of antennas at each node in [10], [11]. Moreover, the degrees of freedom (DoF) region of the general MIMO IC was obtained in [12].

The constant gap capacity approximations of [9] and [11] provide performance guarantees to within a constant number of bits on SISO and MIMO ICs, respectively, by showing that simple Han-Kobayashi (HK) [13] coding schemes with Gaussian inputs and no time-sharing can achieve the capacity region to within a constant number of bits, with the constant being independent of the SNRs, INRs and the channel matrices. The capacity approximation through the DoF characterization of [12] shows that simple transmit beamforming and receive zero-forcing is sufficient to preserve the DoF optimality of the MIMO IC. While in the DoF characterization, it is assumed that the SNRs and interference-to-noise ratios (INRs) at each receiver scale in a similar way, *asymmetric* scaling of the SNRs and INRs on the *dB scale* has been shown to provide more insight about optimal interference management at high SNR as a function of that asymmetry. This phenomenon was captured succinctly through the so-called *generalized degrees of freedom* metric in the context of the SISO IC in [9]. This work obtains the GDoF region for the general MIMO IC, thereby generalizing the same result for the SISO IC obtained in [9] and the usual DoF region for general MIMO IC obtained in [12].

The GDoF region metric, as its name suggests, generalizes the notion of the conventional degrees of freedom (DoF) region metric by additionally emphasizing the signal level as a signaling dimension. It therefore characterizes the simultaneously accessible fractions of spatial and signal-level dimensions (per channel use) by the two users in the limit of high signal-to-noise ratio (SNR) while the ratios of the SNRs and INRs relative to a reference SNR, each expressed in the dB scale, are held constant, with each constant taken, in the most general case, to be arbitrary. The GDoF region was obtained for the SISO IC in [9] based on the constant gap to capacity result found therein. The symmetric GDoF, $d_{\text{sym}}(\alpha)$, which is the maximum common GDoF achievable by each of the two users, for the symmetric SISO IC with equal SNRs and equal INRs for the two users, i.e, with $\text{INR} = \text{SNR}^\alpha$ was evaluated in [9] to be the

well-known “W” curve. The W-curve clearly delineates the very weak, weak, moderate, strong and very strong interference regimes, depending on the value of α , pointing to the optimal (upto GDoF accuracy) interference management techniques as a function of the severity or mildness of the interference.

There have been several other recent works on characterizing the GDoF of various channels. For example, in [14], the symmetric GDoF of a class of symmetric MIMO ICs – for which the SNRs at each receiver are the same and the INRs at each receiver are also the same, with $\text{INR} = \text{SNR}^\alpha$ – and where both transmitters have M antennas and both receivers have N antennas, with the restriction $N \geq M$, was obtained and found to be a “W” curve also. In [15], the symmetric GDoF in the perfectly symmetric (with all direct links having identical gains and all cross links having identical gains) scalar K -user interference network was found (see also [16]). In [17], the symmetric GDoF was obtained for the $(N + 1)$ -user symmetric SIMO IC with N antennas at each receiver and with equal direct link SNRs and equal cross link INRs. The symmetric GDoF of a symmetric model of the scalar X-channel with real-valued channel coefficients was found in [18].

In this work, we obtain the GDoF region of the general MIMO IC with an arbitrary number of antennas at each node and in the most general case where the signal-to-noise ratios (SNR) and interference-to-noise ratios (INRs) vary with arbitrary exponents to a nominal SNR. This result is made possible by the recent constant gap to capacity characterization for the general MIMO IC in [10], [11]. The GDoF result of this paper thus generalizes the GDoF region of the SISO IC found in [9] to the MIMO IC. It also recovers the symmetric GDoF result of [14] for the class of symmetric MIMO ICs considered therein. Moreover, the single and unified constant-gap-to-capacity achievability scheme of [11] considered here, unlike that in [14], does not require the restriction on the numbers of antennas at the different nodes or on the values of SNR exponents and is GDoF optimal in the most general case. The main result of this work also recovers the conventional DoF region result obtained in [12] for the MIMO IC by setting all SNR exponents to unity. In addition to providing several insights that include whatever is common between certain symmetric (in numbers of antennas) MIMO ICs and SISO ICs and what is not, the GDoF result of this paper gives rise to new insights into optimal signaling strategies that make jointly optimal use of the available spatial and signal level dimensions.

The single, unified achievable scheme studied in depth here that is GDoF optimal (and indeed constant-gap-to-capacity optimal [11]) is a simple Han-Kobayashi coding scheme with mutually independent Gaussian input for the private and public messages of each user without time-sharing. The private and public message can be thought of consisting of several information streams. The private information streams are either directed along the null space of the corresponding cross-link channel matrix or

transmitted at power levels that ensure that they reach the unintended receiver below the noise floor. Such a scheme therefore jointly and optimally employs both *signal-level interference alignment* [19] as well as *transmit beamforming or signal-space interference alignment* [20] techniques. For a given DoF tuple in the GDoF region of the channel, we also explicitly specify the DoFs carried by the private and public messages of each user.

The rest of the paper is organized as follows. In Section II we describe the channel model and the GDoF optimal coding scheme. Section III contains the main result of this paper, namely, the GDoF region of the general MIMO IC. Specializations of this result to the SISO IC and to the DoF region of the MIMO IC are also given in Section III which recover the results of [9] and [12], respectively. Explicit specifications of the DoF-splitting between private and public sub-messages are obtained. The reciprocity property of the GDoF region (which denotes the invariability of GDoF with respect to direction of information flow) is described in IV as are specializations of the GDoF region to obtain the symmetric GDoF of the symmetric (M, N, M, N) MIMO IC, thereby recovering as a special case the result of [14] for $M \leq N$. Section V gives in-depth descriptions, through 2 specific examples of weak and mixed interference MIMO ICs, of how different DoF tuples in the GDoF region are achieved through the joint specification of the strategies at the transmitters and the receivers. In Section VI several novel insights revealed by the GDoF analysis are given. Section VII concludes the paper.

Notations: Let \mathbb{C} and \mathbb{R}^+ represent the field of complex numbers and the set of non-negative real numbers, respectively. An $n \times m$ matrix with entries coming from \mathbb{C} will be denoted by $A \in \mathbb{C}^{n \times m}$ and its entry in the i^{th} row and j^{th} column will be denoted by $[A]_{ij}$. We shall denote the transpose and the conjugate transpose of the matrix A by A^T and A^\dagger respectively. I_n represents the $n \times n$ identity matrix, $0_{m \times n}$ represents an all zero $m \times n$ matrix and $\mathbb{U}^{n \times n}$ represents the set of $n \times n$ unitary matrices. The k^{th} column (row) of the matrix A will be denoted by $A^{[k]}$ ($A^{(k)}$) whereas $A^{[k_1:k_2]}$ ($A^{(k_1:k_2)}$) will represent a matrix whose columns (rows) are same as the k_1^{th} to k_2^{th} columns (rows) of matrix A . If $x^{(k)} \in \mathbb{C}, \forall 1 \leq k \leq n$, then $\mathbf{x} \triangleq [x^{(1)}, x^{(2)}, \dots, x^{(n)}]^T$. $\{A, B, C, D\}$ will represent an ordered set of matrices. $I(x; y)$ and $I(x; y|z)$ will represent the mutual information and conditional mutual information of the arguments, respectively. $(x \wedge y)$, $(x \vee y)$ and $(x)^+$ represents the minimum and maximum between x and y and maximum between x and 0, respectively. We also use Landau notations for error terms in approximations. $o(1)$ denotes a term which goes to zero asymptotically and $\mathcal{O}(1)$ denotes a term which is bounded above by some constant. We say x is of the order of y if $\lim_{y \rightarrow \infty} \frac{x}{y} = 0$. All the logarithms in this paper are with base 2. We denote the distribution of a complex circularly symmetric Gaussian random vector with zero mean and covariance matrix Q , by $\mathcal{CN}(0, Q)$. Finally, the indicator function

$1(S)$ is defined as follows

$$1(S) = \begin{cases} 1, & \text{if } S \text{ is true;} \\ 0, & \text{if } S \text{ is false.} \end{cases}$$

■

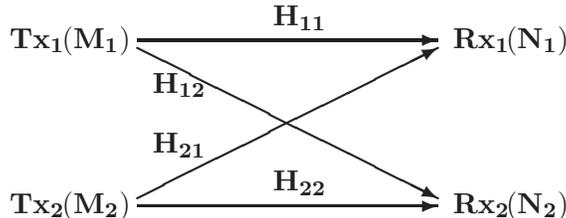


Fig. 1: The (M_1, N_1, M_2, N_2) MIMO IC.

II. CHANNEL MODEL AND PRELIMINARIES

In this section, we define the two-user MIMO IC, its capacity region and the generalized degrees of freedom region and state the upper and lower bounds on the capacity region that are within a constant gap of each other as obtained in [11] for the sake of completeness. We describe the achievable scheme of [11] and then give asymptotic (high SNR) approximations upto $O(1)$ of key quantities that arise in the bounds on capacity region. These approximations are used later to derive the main result of this paper on the GDoF region of the MIMO IC.

A. The MIMO IC

The 2-user MIMO IC, with M_i and N_i antennas at transmitter i (Tx_i) and receiver i (Rx_i), respectively, for $i = 1, 2$ as shown in Fig. 1 (hereafter referred to as (M_1, N_1, M_2, N_2) IC) is considered. We consider a time-invariant or fixed channel where the channel matrices, H_{ij} 's remain fixed for the entire duration of communication. We also assume that the entries of these matrices are drawn i.i.d. from a continuous and unitarily invariant [21] distribution, i.e., $UH_{ij}V$ is identically distributed to H_{ij} for any $U \in \mathbb{U}^{N_j \times N_j}$ and $V \in \mathbb{U}^{M_i \times M_i}$ which ensures that the channel matrices are full rank with probability one (w.p.1). This class of distributions will be denoted by \mathcal{P} in the rest of the paper. In the channel model, we also incorporate a real-valued path-loss or attenuation factor, denoted as η_{ij} , for the signal transmitted from

Tx_i to receiver Rx_j . At time t , Tx_i chooses a vector $X_{it} \in \mathbb{C}^{M_i \times 1}$ and sends $\sqrt{P_i}X_{it}$ over the channel, where we assume the following average input power constraint at Tx_i ,

$$\frac{1}{n} \sum_{t=1}^n \text{Tr}(Q_{it}) \leq 1, \quad (1)$$

for $i \in \{1, 2\}$, where $Q_{it} = \mathbb{E}(X_{it}X_{it}^\dagger)$ and Q_{it} 's can depend on the channel matrices. The received signals at time t can be written as

$$Y_{1t} = \sqrt{\rho_{11}}H_{11}X_{1t} + \sqrt{\rho_{21}}H_{21}X_{2t} + Z_{1t}, \quad (2)$$

$$Y_{2t} = \sqrt{\rho_{22}}H_{22}X_{2t} + \sqrt{\rho_{12}}H_{12}X_{1t} + Z_{2t}, \quad (3)$$

where $Z_{it} \in \mathbb{C}^{N_i \times 1}$ are i.i.d $\mathcal{CN}(\mathbf{0}, I_{N_i})$ across i and t , $\rho_{ii} = \eta_{ii}\sqrt{P_i}$ represents the signal-to-noise ratio (SNR) at receiver i and $\rho_{ij} = \eta_{ij}\sqrt{P_i}$ represents the interference-to-noise ratio (INR) at receiver j for $i \neq j \in \{1, 2\}$.

The performance on the MIMO IC should depend on the strength of the interference relative to the desired signal level on the dB scale with, for example, a better DoF performance expected when interference strength is much less or much higher than the signal strength as in the SISO IC [9]. This variation of performance due to relative difference in strengths of SNRs and INRs can not be captured (at high SNR) by a DoF analysis alone, i.e., if they differ say by only a constant. To characterize the DoF region under such a scenario we thus let the SNRs and INRs vary exponentially with respect to a nominal SNR, ρ , with different scaling factors as follows:

$$\lim_{\log(\rho) \rightarrow \infty} \frac{\log(\rho_{ij})}{\log(\rho)} = \alpha_{ij}, \text{ where } \alpha_{ij} \in \mathbb{R}^+ \text{ and } i, j \in \{1, 2\}. \quad (4)$$

Without loss of generality, we assume $\rho_{11} = \rho$ or $\alpha_{11} = 1$ throughout the rest of the paper. As mentioned earlier, this technique of varying different SNRs and INRs was first introduced in [9] to characterize the DoF region of the SISO 2-user IC and the corresponding DoF region was called the *generalized* DoF (GDoF) region.

B. Capacity region bounds to within a constant gap

In what follows, the MIMO IC with the channel matrices, SNRs and INRs as described above will be denoted by $\mathcal{IC}(\mathcal{H}, \bar{\rho})$, where $\mathcal{H} = \{H_{11}, H_{12}, H_{21}, H_{22}\}$ and $\bar{\rho} = [\rho_{11}, \rho_{12}, \rho_{21}, \rho_{22}]$ or equivalently as $\mathcal{IC}(\mathcal{H}, \bar{\alpha})$ where $\bar{\alpha} = [\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}]$. In what follows, $\bar{\rho}$ and $\bar{\alpha}$ will be used interchangeably to indicate the power levels of different links of the channel. The capacity region of $\mathcal{IC}(\mathcal{H}, \bar{\alpha})$ is defined in the usual way (cf. [11]) and will be denoted by $\mathcal{C}(\mathcal{H}, \bar{\alpha})$. Inner and outer bounds from [11] are stated next.

1) *An outer bound to the capacity region:*

Lemma 1 (Lemma 1 of [11]): For a given \mathcal{H} and $\bar{\alpha}$ the capacity region, $\mathcal{C}(\mathcal{H}, \bar{\alpha})$ of a 2-user MIMO Gaussian IC, with input power constraint (1), is contained within the set of rate tuples $\mathcal{R}^u(\mathcal{H}, \bar{\alpha})$, i.e.,

$$\mathcal{C}(\mathcal{H}, \bar{\alpha}) \subseteq \mathcal{R}^u(\mathcal{H}, \bar{\alpha}),$$

where $\mathcal{R}^u(\mathcal{H}, \bar{\alpha})$ represents the set of rate pairs (R_1, R_2) , satisfying the following constraints:

$$R_1 \leq \log \det \left(I_{N_1} + \rho_{11} H_{11} H_{11}^\dagger \right); \quad (5)$$

$$R_2 \leq \log \det \left(I_{N_2} + \rho_{22} H_{22} H_{22}^\dagger \right); \quad (6)$$

$$R_1 + R_2 \leq \log \det \left(I_{N_2} + \rho_{12} H_{12} H_{12}^\dagger + \rho_{22} H_{22} H_{22}^\dagger \right) + \log \det \left(I_{N_1} + \rho_{11} H_{11} K_1 H_{11}^\dagger \right); \quad (7)$$

$$R_1 + R_2 \leq \log \det \left(I_{N_1} + \rho_{21} H_{21} H_{21}^\dagger + \rho_{11} H_{11} H_{11}^\dagger \right) + \log \det \left(I_{N_2} + \rho_{22} H_{22} K_2 H_{22}^\dagger \right); \quad (8)$$

$$\begin{aligned} R_1 + R_2 \leq & \log \det \left(I_{N_1} + \rho_{21} H_{21} H_{21}^\dagger + \rho_{11} H_{11} K_1 H_{11}^\dagger \right) \\ & + \log \det \left(I_{N_2} + \rho_{12} H_{12} H_{12}^\dagger + \rho_{22} H_{22} K_2 H_{22}^\dagger \right); \end{aligned} \quad (9)$$

$$\begin{aligned} 2R_1 + R_2 \leq & \log \det \left(I_{N_1} + \rho_{21} H_{21} H_{21}^\dagger + \rho_{11} H_{11} H_{11}^\dagger \right) + \log \det \left(I_{N_1} + \rho_{11} H_{11} K_1 H_{11}^\dagger \right) + \\ & \log \det \left(I_{N_2} + \rho_{12} H_{12} H_{12}^\dagger + \rho_{22} H_{22} K_2 H_{22}^\dagger \right); \end{aligned} \quad (10)$$

$$\begin{aligned} R_1 + 2R_2 \leq & \log \det \left(I_{N_2} + \rho_{12} H_{12} H_{12}^\dagger + \rho_{22} H_{22} H_{22}^\dagger \right) + \log \det \left(I_{N_2} + \rho_{22} H_{22} K_2 H_{22}^\dagger \right) + \\ & \log \det \left(I_{N_1} + \rho_{21} H_{21} H_{21}^\dagger + \rho_{11} H_{11} K_1 H_{11}^\dagger \right), \end{aligned} \quad (11)$$

where $K_i = M_i K_{iu}$ and K_{iu} is as specified in equation (13) and $\rho_{ij} = \rho^{\alpha_{ij}}$ for $i, j \in \{1, 2\}$.

2) *An inner bound to the capacity region:* A simple HK coding scheme, denoted as $\mathcal{HK}(K_{1u}, K_{1w}, K_{2u}, K_{2w})$ is described in detail in the next subsection. Here we state an inner bound to the rate region achievable using this scheme found in [11] which is hence also an inner bound to the capacity region.

Lemma 2: Let us denote the achievable rate region of $\mathcal{HK}(K_{1u}, K_{1w}, K_{2u}, K_{2w})$ on $\mathcal{IC}(\mathcal{H}, \bar{\alpha})$ by

$\mathcal{R}_a(\mathcal{H}, \bar{\alpha})$, then $\mathcal{R}_a(\mathcal{H}, \bar{\alpha})$ is the set of rate tuples $(R_1, R_2) \in \mathbb{R}^{+2}$ satisfying the following conditions:

$$\begin{aligned}
R_1 &\leq \log \det \left(I_{N_1} + \rho_{11} H_{11} H_{11}^\dagger \right) - n_1; \\
R_2 &\leq \log \det \left(I_{N_2} + \rho_{22} H_{22} H_{22}^\dagger \right) - n_2; \\
R_1 + R_2 &\leq \log \det \left(I_{N_2} + \rho_{12} H_{12} H_{12}^\dagger + \rho_{22} H_{22} H_{22}^\dagger \right) \\
&\quad + \log \det \left(I_{N_1} + \rho_{11} H_{11} K_1 H_{11}^\dagger \right) - (n_1 + n_2); \\
R_1 + R_2 &\leq \log \det \left(I_{N_1} + \rho_{21} H_{21} H_{21}^\dagger + \rho_{11} H_{11} H_{11}^\dagger \right) \\
&\quad + \log \det \left(I_{N_2} + \rho_{22} H_{22} K_2 H_{22}^\dagger \right) - (n_1 + n_2); \\
R_1 + R_2 &\leq \log \det \left(I_{N_1} + \rho_{21} H_{21} H_{21}^\dagger + \rho_{11} H_{11} K_1 H_{11}^\dagger \right) \\
&\quad + \log \det \left(I_{N_2} + \rho_{12} H_{12} H_{12}^\dagger + \rho_{22} H_{22} K_2 H_{22}^\dagger \right) - (n_1 + n_2); \\
2R_1 + R_2 &\leq \log \det \left(I_{N_1} + \rho_{21} H_{21} H_{21}^\dagger + \rho_{11} H_{11} H_{11}^\dagger \right) + \log \det \left(I_{N_1} + \rho_{11} H_{11} K_1 H_{11}^\dagger \right) + \\
&\quad \log \det \left(I_{N_2} + \rho_{12} H_{12} H_{12}^\dagger + \rho_{22} H_{22} K_2 H_{22}^\dagger \right) - (2n_1 + n_2); \\
R_1 + 2R_2 &\leq \log \det \left(I_{N_2} + \rho_{12} H_{12} H_{12}^\dagger + \rho_{22} H_{22} H_{22}^\dagger \right) + \log \det \left(I_{N_2} + \rho_{22} H_{22} K_2 H_{22}^\dagger \right) + \\
&\quad \log \det \left(I_{N_1} + \rho_{21} H_{21} H_{21}^\dagger + \rho_{11} H_{11} K_1 H_{11}^\dagger \right) - (n_1 + 2n_2),
\end{aligned}$$

where n_i for $i = 1, 2$ are constants functions of the number of antennas only, K_{iw} and K_{iu} are given by equation (13) and $\rho_{ij} = \rho^{\alpha_{ij}}$ for $i, j \in \{1, 2\}$.

C. The simple HK coding scheme

We describe next the coding scheme whose rate region contains the rate region of Lemma 2 which is in turn within a constant gap to the capacity region. Let each of the users divide its message into two sub-messages (called the private and public messages hereafter) and use superposition coding to encode the two sub-messages with mutually independent random zero-mean Gaussian codewords so that we have

$$\begin{aligned}
X_1 &= U_1 + W_1, \\
X_2 &= U_2 + W_2,
\end{aligned} \tag{12}$$

where U_i and W_i represent the codewords of the private and public messages of user i , respectively¹. Moreover, the covariance matrices of the public and private messages are taken for each $i \in \{1, 2\}$ to be

$$\begin{aligned} K_{iu}(\mathcal{H}) &\triangleq \mathbb{E}(U_i U_i^\dagger) = \frac{1}{M_i} \left(I_{M_i} + \rho_{ij} H_{ij}^\dagger H_{ij} \right)^{-1}, \\ K_{iw}(\mathcal{H}) &\triangleq \mathbb{E}(W_i W_i^\dagger) = \left(\frac{I_{M_i}}{M_i} - K_{iu} \right). \end{aligned} \quad (13)$$

In what follows, we shall refer to such a coding scheme as the $\mathcal{HK}(\{K_{1u}, K_{1w}, K_{2u}, K_{2w}\})$ scheme, where we drop the dependence of K_{iu} and K_{iw} on the channel matrices for notational convenience. Let the singular value decomposition (SVD) of the channel matrix H_{ij} be given by $H_{ij} = V_{ij} \Sigma_{ij} U_{ij}^\dagger$, where $V_{ij} \in \mathbb{U}^{N_j \times N_j}$ and $U_{ij} \in \mathbb{U}^{M_i \times M_i}$ are unitary matrices and $\Sigma_{ij} \in \mathbb{C}^{N_j \times M_i}$ is a rectangular matrix containing the singular values along its diagonal. Using the SVD of the matrix H_{ij} , the covariance matrices for U_i and W_i of equation (13) can alternatively be written as

$$K_{iu} = U_{ij} \begin{bmatrix} \frac{1}{M_i} (I_{\min\{M_i, N_j\}} + \rho^{\alpha_{ij}} \Lambda_{ij})^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{M_i} I_{(M_i - N_j)^+} \end{bmatrix} U_{ij}^\dagger = U_{ij} D_{ij} U_{ij}^\dagger, \quad (14)$$

where Λ_{ij} is a diagonal matrix containing the non-zero eigenvalues of $H_{ij}^\dagger H_{ij}$ and denoting the quantity $M_i(1 + \rho^{\alpha_{ij}} \lambda_{ij}^{(k)}) = r_{ik}$ where $\lambda_{ij}^{(k)}$ is the k^{th} non-zero eigenvalues of $H_{ij}^\dagger H_{ij}$ for $1 \leq k \leq m_{ij} = \min\{M_i, N_j\}$ we have

$$[D_{ij}]_{kk} = \begin{cases} r_{ik}^{-1}, & \text{for } 1 \leq k \leq m_{ij}; \\ \frac{1}{M_i}, & \text{for } m_{ij} + 1 \leq k \leq M_i. \end{cases} \quad (15)$$

Similarly, we have

$$K_{iw} = U_{ij} \left(\frac{1}{M_i} I_{M_i} - D_{ij} \right) U_{ij}^\dagger = U_{ij}^{[1:m_{ij}]} \tilde{D}_{ij} (U_{ij}^{[1:m_{ij}]})^\dagger, \quad (16)$$

where $[\tilde{D}_{ij}]_{kk} = (\frac{1}{M_i} - \frac{1}{r_{ik}})$ for $1 \leq k \leq m_{ij}$. Now, it is well known that a Gaussian vector V with covariance matrix K can be expressed as $V = Ax$, where x is a Gaussian vector with identity as covariance matrix if $AA^\dagger = K$. Using this result along with equations (14) and (16) we can write

$$\begin{aligned} U_i &= U_{ij} \sqrt{D_{ij}} \mathbf{x}_{ip} = \sum_{l=1}^{M_i} \sqrt{[D_{ij}]_{ll}} x_{ip}^{(l)} U_{ij}^{[l]}, \\ W_i &= U_{ij} \sqrt{\tilde{D}_{ij}} \mathbf{x}_{ic} = \sum_{k=1}^{m_{ij}} \sqrt{[\tilde{D}_{ij}]_{kk}} x_{ic}^{(k)} U_{ij}^{[k]}, \end{aligned} \quad (17)$$

¹With a slight abuse of notation, we shall use the same symbol to represent a message and its corresponding codeword.

where $\mathbf{x}_{ic} = [x_{ic}^{(1)}, \dots, x_{ic}^{(m_{ij})}]^T \sim \mathcal{CN}(0, I_{m_{ij}})$ and $\mathbf{x}_{ip} = [x_{ip}^{(1)}, \dots, x_{ip}^{(M_i)}]^T \sim \mathcal{CN}(0, I_{M_i})$ are mutually independent normal Gaussian vectors. Substituting this in the expression for X_i we see that, the transmit signal at Tx_i can be written as

$$X_i = \sum_{k=1}^{m_{ij}} \sqrt{[\tilde{D}_{ij}]_{kk}} x_{ic}^{(k)} U_{ij}^{[k]} + \sum_{l=1}^{m_{ij}} \sqrt{[D_{ij}]_{ll}} x_{ip}^{(l)} U_{ij}^{[l]} + \sum_{m=1+M_i}^{M_i} \frac{1}{\sqrt{M_i}} x_{ip}^{(m)} U_{ij}^{[m]}. \quad (18)$$

In the above equation, $x_{ic}^{(l)}$ for $1 \leq l \leq m_{ij}$ and $x_{ip}^{(k)}$ for $1 \leq k \leq M_i$ represent the l^{th} and k^{th} stream of the public and private information along directions $U_{ij}^{[l]}$ and $U_{ij}^{[k]}$, respectively, for user i .

Remark 1: Note that each of the terms in the second sum of the right hand side of (18) have power proportional to ρ_{ij}^{-1} . Hence, all the streams encoded through $x_{ip}^{(k)}$ for $1 \leq k \leq m_{ij}$ after passing through the cross channel with strength ρ_{ij} reach Rx_j at the noise floor. This technique can be considered as a form of *interference alignment at the signal level*.

On the other hand, in the SVD of the matrix H_{ij} , the last $(M_i - N_j)^+$ columns of Σ_{ij} are all zeros and hence $H_{ij}U_{ij}^{[k]} = 0$ for $N_j < k \leq M_i$. In other words, each of the $U_{ij}^{[k]}$'s for $m_{ij} < k \leq M_i$ lie in the null space of the matrix H_{ij} . Therefore, any stream sent along one of these directions reaches Rx_j in a subspace which is perpendicular to the subspace in which the useful signals of Rx_2 lie. That is, each user can be said to align the interference to the undesired user in a particular subspace, which is a simple form of *signal space interference alignment*. This explains why we call the streams carried by $x_{ip}^{(k)}$, $1 \leq k \leq M_i$, *private streams*.

Thus the specific choice of the covariance matrices K_{iu} in $\mathcal{HK}(\{K_{1u}, K_{1w}, K_{2u}, K_{2w}\})$ for $i = 1, 2$ amounts to employing a technique to jointly utilize both types of interference alignments described above.

D. Generalized Degrees of Freedom Region

Definition 1: The GDoF region, $\mathcal{D}_o(\bar{M}, \bar{\alpha})$, of $\mathcal{IC}(\mathcal{H}, \bar{\alpha})$ is defined as

$$\mathcal{D}_o(\bar{M}, \bar{\alpha}) = \left\{ (d_1, d_2) : d_i = \lim_{\rho_{ii} \rightarrow \infty} \frac{R_i}{\log(\rho_{ii})}, i \in \{1, 2\} \text{ such that } (R_1, R_2) \in \mathcal{C}(\mathcal{H}, \bar{\alpha}) \right\}. \quad (19)$$

Since the capacity region of a MIMO IC is not known and a constant number of bits is insignificant in the GDoF analysis, to derive the GDoF region we shall use the constant-gap-to-capacity result found by the authors in [11]. In particular, since a constant number of bits is insignificant in the GDoF analysis, the $\mathcal{C}(\mathcal{H}, \bar{\alpha})$ in the definition of the GDoF region can be replaced by either $\mathcal{R}^u(\mathcal{H}, \bar{\alpha})$ or $\mathcal{R}_a(\mathcal{H}, \bar{\alpha})$ to compute the GDoF region of the MIMO IC. We state this fact as a lemma for easy further reference.

Lemma 3: The GDoF region of the MIMO IC is given as

$$\mathcal{D}_o(\bar{M}, \bar{\alpha}) = \left\{ (d_1, d_2) : d_i = \lim_{\rho_{ii} \rightarrow \infty} \frac{R_i}{\log(\rho_{ii})}, i \in \{1, 2\} \text{ and } (R_1, R_2) \in \mathcal{R}^u(\mathcal{H}, \bar{\alpha}) \right\}, \quad (20)$$

where $\mathcal{R}^u(\mathcal{H}, \bar{\alpha}) = \mathcal{R}^u(\mathcal{H}, \bar{\rho})$ is given by Lemma 1.

Proof: From Lemma 1 and 2 we have the following

$$\mathcal{R}_a(\mathcal{H}, \bar{\alpha}) \subseteq \mathcal{C}(\mathcal{H}, \bar{\alpha}) \subseteq \mathcal{R}^u(\mathcal{H}, \bar{\alpha}). \quad (21)$$

To obtain the desired result we use in the definition of the GDoF region of (19), the above set inclusions along with the fact that n_i 's in Lemma 2 are independent of ρ and \mathcal{H} . ■

E. Asymptotic Approximations

In the derivation of the GDoF region of the 2-user MIMO IC quantities like the sum rate upper bound on 2- and 3-user MIMO multiple-access channels (MACs) will appear frequently. Thus, in the following two lemmas, we provide asymptotic approximations up to $O(1)$ of such quantities for different $\bar{\alpha}$ and number of antennas.

Lemma 4: Let $H_1 \in \mathbb{C}^{u \times u_1}$ and $H_2 \in \mathbb{C}^{u \times u_2}$ are two full rank (w.p.1) channel matrices such that $H = [H_1 \ H_2]$ is also full rank w.p.1. Then for asymptotic ρ

$$\log \det \left(I_u + \rho^a H_1 H_1^\dagger + \rho^b H_2 H_2^\dagger \right) = f(u, (a, u_1), (b, u_2)) \log(\rho) + \mathcal{O}(1), \quad (22)$$

where for any $u \in \mathbb{R}^+$ and $(a_i, u_i) \in \mathbb{R}^2$ for $i \in \{1, 2\}$,

$$f(u, (a_1, u_1), (a_2, u_2)) \triangleq \begin{cases} \min\{u, u_1\}a_1^+ + \min\{(u - u_1)^+, u_2\}a_2^+, & \text{if } a_1 \geq a_2; \\ \min\{u, u_2\}a_2^+ + \min\{(u - u_2)^+, u_1\}a_1^+, & \text{if } a_1 < a_2. \end{cases} \quad (23)$$

Proof: This result was proved in [14] when $(u_1 + u_2) \geq u$. The proof for the case when $(u_1 + u_2) < u$ is given in Appendix A. ■

Remark 2: If $H_1 \in \mathbb{C}^{u \times u_1}$ and $H_2 \in \mathbb{C}^{u \times u_2}$ are mutually independent and $H_1, H_2 \in \mathcal{P}$ then $[H_1 \ H_2] \in \mathcal{P}$ and therefore is a full rank matrix w.p.1. That is, if H_1 and H_2 represent the two incoming channel matrices at any of the receivers in the MIMO IC then Lemma 4 holds.

Lemma 5: Let $H_i \in \mathbb{C}^{u \times u_i}$ for $i = 1, 2, 3$ be three channel matrices with statistics described at the beginning of this section, then for asymptotic ρ

$$\log \det \left(I_u + \sum_{i=1}^3 \rho^{a_i} H_i H_i^\dagger \right) = g(u, (a_1, u_1), (a_2, u_2), (a_3, u_3)) \log(\rho) + \mathcal{O}(1), \quad (24)$$

where for any $u \in \mathbb{R}^+$ and $(a_i, u_i) \in \mathbb{R}^2$ for $i \in \{1, 2, 3\}$,

$$g(u, (a_1, u_1), (a_2, u_2), (a_3, u_3)) \triangleq \min\{u, u_{i_1}\}a_{i_1}^+ + \min\{(u - u_{i_1})^+, u_{i_2}\}a_{i_2}^+ + \min\{(u - u_{i_1} - u_{i_2})^+, u_{i_3}\}a_{i_3}^+, \quad (25)$$

for $i_1, i_2, i_3 \in \{1, 2, 3\}$ such that $a_{i_1} \geq a_{i_2} \geq a_{i_3}$.

Proof: The proof is given in Appendix B. ■

Remark 3: Suppose $a_1 \geq \max\{a_2, a_3\}$ in Lemma 5, then $g(\cdot)$ can also be written as

$$g(u, (a_1, u_1), (a_2, u_2), (a_3, u_3)) = \min\{u, u_1\}a_1^+ + f((u - u_1)^+, (a_2, u_2), (a_3, u_3)).$$

For example, $g(10, (.5, 3), (1, 4), (1.2, 2)) = 2(1.2) + f(8, (.5, 3), (1, 4)) = 2.4 + 4(1) + 3(.5) = 7.9$.

Remark 4: $g(\cdot)$ in Lemma 5 represents the sum DoFs achievable on a 3-user MIMO multiple-access channel (MAC) with u antennas at the receiver, u_i antennas at the i^{th} transmitter, where the SNR of the i^{th} user is ρ^{a_i} for $i \in \{1, 2, 3\}$. Similarly, $f(\cdot)$ in Lemma 4 can be interpreted as the sum GDoF achievable on a 2-user MIMO MAC.

III. THE GDOF REGION OF THE MIMO IC

Using the explicit expression for the upper bounds to the capacity region of the MIMO IC from Lemma 1 and using it in Lemma 3 we get the main result of this paper.

Theorem 1: The GDoF region of $\mathcal{IC}(\bar{M}, \bar{\alpha})$ is the set of DoF tuples (d_1, d_2) , denoted by $\mathcal{D}_o(\bar{M}, \bar{\alpha})$, where $d_i \in \mathbb{R}^+$ for $i = 1, 2$ satisfy the following conditions:

$$\begin{aligned} d_1 &\leq \min\{M_1, N_1\}; \\ d_2 &\leq \min\{M_2, N_2\}; \\ (d_1 + \alpha_{22}d_2) &\leq f(N_2, (\alpha_{12}, M_1), (\alpha_{22}, M_2)) + f(N_1, (\beta_{12}, m_{12}), (\alpha_{11}, (M_1 - N_2)^+)); \\ (d_1 + \alpha_{22}d_2) &\leq f(N_1, (\alpha_{21}, M_2), (\alpha_{11}, M_1)) + f(N_2, (\beta_{21}, m_{21}), (\alpha_{22}, (M_2 - N_1)^+)); \\ (d_1 + \alpha_{22}d_2) &\leq g(N_1, (\alpha_{21}, M_2), (\beta_{12}, m_{12}), (1, (M_1 - N_2)^+)) + \\ &\quad g(N_2, (\alpha_{12}, M_1), (\beta_{21}, m_{21}), (\alpha_{22}, (M_2 - N_1)^+)); \\ (2d_1 + \alpha_{22}d_2) &\leq f(N_1, (\alpha_{21}, M_2), (\alpha_{11}, M_1)) + f(N_1, (\beta_{12}, m_{12}), (\alpha_{11}, (M_1 - N_2)^+)) + \\ &\quad g(N_2, (\alpha_{12}, M_1), (\beta_{21}, m_{21}), (\alpha_{22}, (M_2 - N_1)^+)); \\ (d_1 + 2\alpha_{22}d_2) &\leq f(M_2, (\alpha_{21}, N_1), (\alpha_{22}, N_2)) + f(N_2, (\beta_{21}, m_{21}), (\alpha_{22}, (M_2 - N_1)^+)) + \\ &\quad g(N_1, (\alpha_{21}, M_2), (\beta_{12}, m_{12}), (1, (M_1 - N_2)^+)), \end{aligned}$$

where $\beta_{ij} = (\alpha_{ii} - \alpha_{ij})^+$, functions $f(\cdot, \cdot, \cdot)$ and $g(\cdot, \dots, \cdot)$ are as defined in equation (23) and (25), respectively, for $i \neq j \in \{1, 2\}$ and $m_{ij} \triangleq \min\{M_i, N_j\}$ as defined before.

Proof of Theorem 1(Outline): From Lemma 3 we see that the GDoF region of the 2-user MIMO IC is simply the scaled version of the rate region $\mathcal{R}^u(\mathcal{H}, \bar{\alpha})$. Thus to evaluate the GDoF region we simply

need to scale all the terms in each of the equations of $\mathcal{R}^u(\mathcal{H}, \bar{\alpha})$. For example, consider the third bound in equation (7)

$$R_1 + R_2 \leq I_{b3}.$$

Dividing both sides by $\log(\rho_{11})$ and taking the limit we get

$$\begin{aligned} \lim_{\rho_{11} \rightarrow \infty} \frac{R_1 + R_2}{\log(\rho_{11})} &\leq \lim_{\rho_{11} \rightarrow \infty} \frac{I_{b3}}{\log(\rho_{11})}; \\ \Rightarrow \lim_{\rho \rightarrow \infty} \left\{ \frac{R_1}{\log(\rho)} + \frac{\alpha_{22} R_2}{\alpha_{22} \log(\rho)} \right\} &\leq \lim_{\rho \rightarrow \infty} \frac{I_{b3}}{\log(\rho)}; \\ \Rightarrow \lim_{\rho \rightarrow \infty} \frac{R_1}{\log(\rho)} + \lim_{\rho_{22} \rightarrow \infty} \frac{\alpha_{22} R_2}{\log(\rho_{22})} &\leq \lim_{\rho \rightarrow \infty} \frac{I_{b3}}{\log(\rho)}; \\ \Rightarrow d_1 + \alpha_{22} d_2 &\leq \lim_{\rho \rightarrow \infty} \frac{I_{b3}}{\log(\rho)}. \end{aligned}$$

Following the same steps for other bounds we get

$$\mathcal{D}_o(\bar{M}, \bar{\alpha}) = \left\{ (d_1, d_2) : d_1 \leq \lim_{\rho \rightarrow \infty} \frac{I_{b1}}{\log(\rho)}; \right. \quad (26)$$

$$d_2 \leq \lim_{\rho \rightarrow \infty} \frac{I_{b2}}{\log(\rho)}; \quad (27)$$

$$d_1 + \alpha_{22} d_2 \leq \lim_{\rho \rightarrow \infty} \frac{I_{b3}}{\log(\rho)}; \quad (28)$$

$$d_1 + \alpha_{22} d_2 \leq \lim_{\rho \rightarrow \infty} \frac{I_{b4}}{\log(\rho)}; \quad (29)$$

$$d_1 + \alpha_{22} d_2 \leq \lim_{\rho \rightarrow \infty} \frac{I_{b5}}{\log(\rho)}; \quad (30)$$

$$2d_1 + \alpha_{22} d_2 \leq \lim_{\rho \rightarrow \infty} \frac{I_{b6}}{\log(\rho)}; \quad (31)$$

$$d_1 + 2\alpha_{22} d_2 \leq \lim_{\rho \rightarrow \infty} \frac{I_{b7}}{\log(\rho)}; \left. \right\} \quad (32)$$

To prove the theorem we have to evaluate the right hand side limits, which can be done by finding the asymptotic approximations of the different I_{bi} s of Lemma 1 using Lemmas 4 and 5. The detailed proof is given in Appendix C. ■

The above theorem is specialized next to the SISO IC (by putting $M_1 = M_2 = N_1 = N_2 = 1$) in the following corollary, yielding its GDoF region.

Corollary 1: The GDoF region of the SISO IC is given as

$$\mathcal{D}_{\text{SISO}} = \left\{ (d_1, d_2) : d_1 \leq 1; \right. \quad (33a)$$

$$\left. d_2 \leq 1; \right. \quad (33b)$$

$$(d_1 + \alpha_{22}d_2) \leq \max\{\alpha_{22}, \alpha_{12}\} + (1 - \alpha_{12})^+; \quad (33c)$$

$$(d_1 + \alpha_{22}d_2) \leq \max\{1, \alpha_{21}\} + (\alpha_{22} - \alpha_{21})^+; \quad (33d)$$

$$(d_1 + \alpha_{22}d_2) \leq \max\{\alpha_{21}, (1 - \alpha_{12})^+\} + \max\{\alpha_{12}, (\alpha_{22} - \alpha_{21})^+\}; \quad (33e)$$

$$(2d_1 + \alpha_{22}d_2) \leq \max\{1, \alpha_{21}\} + (1 - \alpha_{12})^+ \max\{\alpha_{12}, (\alpha_{22} - \alpha_{21})^+\}; \quad (33f)$$

$$(d_1 + 2\alpha_{22}d_2) \leq \max\{\alpha_{22}, \alpha_{12}\} + \max\{\alpha_{21}, (1 - \alpha_{12})^+\} + (\alpha_{22} - \alpha_{21})^+ \} \quad (33g)$$

Remark 5: The region of Corollary 1 provides a single unified formula for the GDoF region of the SISO IC for all interference regimes. It can be specialized to obtain GDoF regions for different interference regimes given in Section V of [9] such as weak interference, mixed interference, strong interference, etc., separate formulas for each of which are given therein. For instance, in the *weak* interference regime defined by $\alpha_{12} \leq 1$ and $\alpha_{21} \leq \alpha_{22}$, we have

$$\max\{\alpha_{22}, \alpha_{12}\} + (1 - \alpha_{12})^+ = \max\{\alpha_{22}, \alpha_{12}\} + (1 - \alpha_{12}) = 1 + (\alpha_{22} - \alpha_{12})^+;$$

$$\max\{1, \alpha_{21}\} + (\alpha_{22} - \alpha_{21})^+ = \max\{1, \alpha_{21}\} + (\alpha_{22} - \alpha_{21}) = \alpha_{22} + (1 - \alpha_{21})^+.$$

Substituting these identities in equation (33) we recover equation (78) of [9] which represents the GDoF region of the SISO IC in the weak interference regime (note that α_{22}, α_{21} and α_{12} are denoted as α_1, α_2 and α_3 in [9]). Similarly for instance, equations (82) and (84) of [9] can be recovered for the mixed and strong interference channels, respectively, by simplifying the result of Corollary 1 according to the defining conditions on the α 's for those regimes.

Remark 6: The conventional DoF region of the MIMO IC obtained in [12] can also be recovered from Theorem 1 by putting $\alpha_{ij} = 1$, for $1 \leq i, j \leq 2$ in Theorem 1 and simplifying the different bounds.

Consequently, we get

$$\mathcal{D}_{\text{DoF}} = \left\{ (d_1, d_2) : d_1 \leq \min\{M_1, N_1\}; \right. \quad (34a)$$

$$\left. d_2 \leq \min\{M_1, N_1\}; \right. \quad (34b)$$

$$(d_1 + d_2) \leq (N_2 \wedge (M_1 + M_2)) + N_1 \wedge (M_1 - N_2)^+; \quad (34c)$$

$$(d_1 + d_2) \leq (N_1 \wedge (M_1 + M_2)) + N_2 \wedge (M_2 - N_1)^+; \quad (34d)$$

$$(d_1 + d_2) \leq (N_1 \wedge M_2) + ((N_1 - M_2)^+ \wedge (M_1 - N_2)^+) + \\ (N_2 \wedge M_1) + ((M_1 - N_2)^+ \wedge (M_2 - N_1)^+); \quad (34e)$$

$$(2d_1 + d_2) \leq (N_1 \wedge (M_1 + M_2)) + N_1 \wedge (M_1 - N_2)^+ + \\ (N_2 \wedge M_1) + ((M_1 - N_2)^+ \wedge (M_2 - N_1)^+); \quad (34f)$$

$$(d_1 + 2d_2) \leq (N_2 \wedge (M_1 + M_2)) + N_2 \wedge (M_2 - N_1)^+ + \\ (N_1 \wedge M_2) + ((N_1 - M_2)^+ \wedge (M_1 - N_2)^+) \left. \right\} \quad (34g)$$

Corollary 2 (The main result of [12]): The DoF region of the 2-user MIMO IC is given as

$$\mathcal{D}_{\text{DoF}} = \left\{ (d_1, d_2) : d_1 \leq \min\{M_1, N_1\}; \right. \\ \left. d_2 \leq \min\{M_2, N_2\}; \right. \\ \left. (d_1 + d_2) \leq \min\{(M_1 + M_2), (N_1 + N_2), \max(M_1, N_2), \max(M_1, N_2)\} \right\}.$$

Proof: We obtain this result starting from equation (34). Let us consider the sum bound of equation (34c),

$$(d_1 + d_2) \leq (N_2 \wedge (M_1 + M_2)) + N_1 \wedge (M_1 - N_2)^+, \\ = \{N_2 \wedge (M_1 + M_2)\}1(N_2 \geq M_1) + \{N_2 + (N_1 \wedge (M_1 - N_2))\}1(N_2 < M_1), \\ = \{N_2 \wedge (M_1 + M_2)\}1(N_2 \geq M_1) + \{(N_2 + N_1) \wedge M_1\}1(N_2 < M_1), \\ = \min\{(M_1 + M_2), (N_1 + N_2), \max(N_2, M_1)\}.$$

Similarly, simplifying the bound in equation (34d) it can be shown that

$$(d_1 + d_2) \leq \min\{(M_1 + M_2), (N_1 + N_2), \max(N_1, M_2)\}.$$

Moreover, in Appendix D it will be shown that the bounds in equations (34e)-(34g) are looser than those in equation (34a)-(34d). Finally, combining the simplified forms of the 3-rd and 4-th bounds above, the claim is proved. ■

Remark 7: The GDoF region of the 2-user MIMO multiple-access channel (MAC) with an arbitrary number of antennas at the three terminals can also be found as a by-product of the analysis of the MIMO IC. This is detailed in Appendix E.

Remark 8 (Interpretation of the different bounds): We know that the GDoF optimal coding scheme divides each user's message into two sub-messages. Let the DoFs of the private and the public messages of user i be denoted by d_{ip} and d_{ic} , respectively. Note that H_{12} has a $(M_1 - N_2)^+$ -dimensional null space along which Tx_1 can send private information to its desired receiver at an SNR of $\rho^{\alpha_{11}}$. Along the remaining m_{12} dimensions Tx_1 can send private information only at a power level of $\rho^{-\alpha_{12}}$ which reaches Rx_1 at a power level of $\rho^{(\alpha_{11} - \alpha_{12})^+}$. Thus with respect to the private information of Tx_1 , Rx_1 is a MAC with 2 virtual transmitters having SNRs $\rho^{\alpha_{11}}$ and $\rho^{(\alpha_{11} - \alpha_{12})^+}$ and $(M_1 - N_2)^+$ and m_{12} transmit antennas, respectively. Hence, from Lemma 4, we have

$$d_{1p} \leq f(N_1, ((\alpha_{11} - \alpha_{12})^+, m_{12}), (\alpha_{11}, (M_1 - N_2)^+)).$$

On the other hand, since d_{1c} is decoded at Rx_2 , Rx_2 is a MAC receiver with respect to W_1 (having an SNR of $\rho^{\alpha_{12}}$ and M_1 transmit antennas) and X_2 (having an SNR of $\rho^{\alpha_{22}}$) and from Lemma 4 (recall Remark 4) we have

$$(d_{1c} + \alpha_{22}d_2) \leq f(N_2, (\alpha_{12}, M_1), (\alpha_{22}, M_2)).$$

Combining the above two equations we get the 3^{rd} bound of the GDoF region. The 4^{th} bound can be similarly interpreted just by interchanging the roles of Rx_1 and Rx_2 . As explained above, the two parts of the private message of Tx_1 can be thought of as two virtual users to the MAC receiver Rx_1 ; in addition to them, Tx_2 can send a maximum of $m_{21}\alpha_{21}$ public DoFs to Rx_1 through W_2 , which can be interpreted as the 3^{rd} virtual user (with SNR $\rho^{\alpha_{21}}$ and m_{21} transmit antennas) to the MAC receiver at Rx_1 , and therefore, Lemma 5 provides the following sum DoF upper bound

$$(d_{1p} + \alpha_{22}d_{2c}) \leq g(N_1, (\alpha_{21}, M_2), (\beta_{12}, m_{12}), (1, (M_1 - N_2)^+)).$$

A similar consideration regarding the DoFs decodable at Rx_2 gives

$$(d_{1c} + \alpha_{22}d_{2p}) \leq g(N_2, (\alpha_{12}, M_1), (\beta_{21}, m_{21}), (\alpha_{22}, (M_2 - N_1)^+)).$$

Combining the last two equations we get the 5^{th} bound of Theorem 1. The other two bounds of the theorem can be similarly interpreted.

Remark 9 (DoF-Split): The distributions and power levels of the codewords that encode each user's private and public messages are specified in Section II-C; however, to completely specify the encoding

scheme we further need to specify the DoFs carried by the private and public messages of each user, which are denoted by d_{ip} and d_{ic} , respectively, for $i = 1, 2$. The lemma below completes the specification of the GDoF optimal coding scheme by providing a set of 4-tuples, $\mathcal{G}(\bar{M}, \bar{\alpha}) = \{(d_{1c}, d_{1p}, d_{2c}, d_{2p})\}$, which is achievable on the 2-user MIMO IC by the GDoF optimal coding scheme of Section II-C. The $\mathcal{G}(\bar{M}, \bar{\alpha})$ region has the property that, for any $(d_1, d_2) \in \mathcal{D}_o(\bar{M}, \bar{\alpha})$ (which is specified in Theorem 1), there exists an $(d_{1c}, d_{1p}, d_{2c}, d_{2p}) \in \mathcal{G}(\bar{M}, \bar{\alpha})$ such that $(d_{ip} + d_{ic}) = d_i$ for $i = 1, 2$.

Lemma 6: The DoF pair $(d_1, d_2) \in \mathcal{D}_o(\bar{M}, \bar{\alpha})$ only if there exists a 4-tuple $(d_{1c}, d_{1p}, d_{2c}, d_{2p}) \in \mathcal{G}(\bar{M}, \bar{\alpha})$ such that $d_i = (d_{ic} + d_{ip})$ for $i = 1, 2$, where $\mathcal{G}(\bar{M}, \bar{\alpha}) = \mathcal{G}_1(\bar{M}, \bar{\alpha}) \cap \mathcal{G}_2(\bar{M}, \bar{\alpha})$ with $\mathcal{G}_1(\bar{M}, \bar{\alpha})$ defined below (and with $\mathcal{G}_2(\bar{M}, \bar{\alpha})$ obtained by interchanging the indexes 1 and 2 in the expression for $\mathcal{G}_1(\bar{M}, \bar{\alpha})$),

$$\mathcal{G}_1(\bar{M}, \bar{\alpha}) = \left\{ (d_{1p}, d_{1c}, d_{2c}) : \alpha_{11}d_{1p} \leq f(N_1, (\beta_{12}, m_{12}), (\alpha_{11}, (M_1 - N_2)^+)) \alpha_{11}; \right. \quad (35a)$$

$$\left. \alpha_{11}d_{1c} \leq \min\{N_1, M_1, N_2\} \alpha_{11}; \right. \quad (35b)$$

$$\left. \alpha_{22}d_{2c} \leq \min\{N_1, M_2\} \alpha_{21}; \right. \quad (35c)$$

$$\left. \alpha_{11}(d_{1p} + d_{1c}) \leq \min\{M_1, N_1\} \alpha_{11}; \right. \quad (35d)$$

$$\left. (\alpha_{11}d_{1p} + \alpha_{22}d_{2c}) \leq g(N_1, (\alpha_{21}, M_2), (\beta_{12}, m_{12}), (1, (M_1 - N_2)^+)); \right. \quad (35e)$$

$$\left. (\alpha_{11}d_{1c} + \alpha_{22}d_{2c}) \leq f(N_1, (\alpha_{21}, M_2), (\alpha_{11}, m_{12})); \right. \quad (35f)$$

$$\left. (\alpha_{11}d_{1p} + \alpha_{11}d_{1c} + \alpha_{22}d_{2c}) \leq f(N_1, (\alpha_{21}, M_2), (\alpha_{11}, M_1)); \right\} \quad (35g)$$

with $\beta_{12} = (\alpha_{11} - \alpha_{12})^+$ and functions $f(\cdot, \cdot, \cdot)$ and $g(\cdot, \cdot, \cdot, \cdot)$ are as defined in equation (23) and (25), respectively.

Proof: The proof is given in Appendix F. ■

IV. THE SYMMETRIC GDOF REGION OF THE (M, N, M, N) MIMO IC

Suppose the roles of the transmitters and receivers of the MIMO IC $\mathcal{IC}(\mathcal{H}, \bar{\alpha})$ are interchanged. In the notations defined in Section II, this resulting IC (hereafter referred to as the ‘‘reciprocal’’ channel) can be denoted by $\mathcal{IC}(\mathcal{H}^r, \bar{\alpha}^r)$, where $\mathcal{H}^r = \{H_{11}^T, H_{21}^T, H_{12}^T, H_{22}^T\}$ and $\bar{\alpha}^r = [\alpha_{11}, \alpha_{21}, \alpha_{12}, \alpha_{22}]$. Clearly, $\mathcal{D}_o(\bar{M}^r, \bar{\alpha}^r)$ denotes the GDoF region of the reciprocal channel where $\bar{M}^r = (N_1, M_1, N_2, M_2)$.

Corollary 3 (Reciprocity of the GDoF region): The GDoF region of the MIMO IC is same as that of its *reciprocal* channel i.e.,

$$\mathcal{D}_o(\bar{M}, \bar{\alpha}) = \mathcal{D}_o(\bar{M}^r, \bar{\alpha}^r).$$

Proof: It was proved in [11] that the capacity region of a 2-user MIMO IC and its *reciprocal* channel are within a constant (independent of ρ) number of bits to each other (see Lemma 6 of [11]). The corollary is easily proved by using this result in the definition of the GDoF region of the IC in equation (19), which states that the GDoF regions of two channels with capacity regions differing by only a constant number of bits are the same. ■

In other words, the GDoF region of the channel does not change if the roles of the transmitters and the receivers are interchanged. Note that this is a more general result than the reciprocity of the conventional DoF region proved in [12]. In what follows, we define the symmetric GDoF metric.

Definition 2 (Symmetric GDoF): Let $\mathcal{C}_s(\alpha) = \sup(R_1 + R_2)$ with $(R_1, R_2) \in \mathcal{C}(\mathcal{H}, \bar{\alpha})$ where $\bar{\alpha} = [1, \alpha, \alpha, 1]$ and $\sup \mathcal{A}$ represents the supremum of the set of elements in \mathcal{A} . Then the *symmetric GDoF* of the channel, denoted by d_s , is defined as

$$d_s \triangleq \lim_{\rho \rightarrow \infty} \frac{\mathcal{C}_s(\alpha)}{2 \log(\rho)}.$$

It is clear from Definition 1 and the above equation that

$$d_s = \frac{\sup_{\mathcal{D}_o(\bar{M}, \bar{\alpha})} (d_1 + d_2)}{2}. \quad (36)$$

The authors in [14] found the symmetric GDoF of the MIMO IC with an equal number of antennas at the two transmitters and an equal number of antennas at the receiver and with more receive than transmit antennas, i.e., for an (M, N, M, N) IC with $M \leq N$ and $\bar{\alpha} = [1, \alpha, \alpha, 1]$. They found that

$$d_s \leq \min\{M, \hat{D}(\alpha)\} \quad (37)$$

where

$$\hat{D}(\alpha) = \begin{cases} M - (2M - N)\alpha, & 0 \leq \alpha < \frac{1}{2}; \\ (N - M) + (2M - N)\alpha, & \frac{1}{2} \leq \alpha \leq \frac{2}{3}; \\ M - \frac{\alpha}{2}(2M - N), & \frac{2}{3} \leq \alpha \leq 1; \\ \frac{N}{2} + \frac{M}{2}(\alpha - 1), & 1 \leq \alpha. \end{cases} \quad (38)$$

It can be verified that the above result of [14] can be recovered by putting $N_1 = N_2 = N$, $M_1 = M_2 = M$ and $\bar{\alpha} = [1, \alpha, \alpha, 1]$, in Theorem 1 of this paper (the details are left to the reader). The result of [14] is however only valid for $M \leq N$ and does not extend to the $M > N$ case. Specializing Theorem 1 for $N_1 = N_2 = N$, $M_1 = M_2 = M > N$ and $\bar{\alpha} = [1, \alpha, \alpha, 1]$ we get the following.

Corollary 4: The symmetric DoF ($d_s = d_1 = d_2$) of a (M, N, M, N) IC with $M > N$ and $\bar{\alpha} = [1, \alpha, \alpha, 1]$ is given by

$$d_s \leq \min\{N, D(\alpha)\}$$

where $D(\alpha)$ is given as

$$D(\alpha) = \begin{cases} N - (2N - M)\alpha, & 0 \leq \alpha < \frac{1}{2}; \\ (M - N) + (2N - M)\alpha, & \frac{1}{2} \leq \alpha \leq \frac{2}{3}; \\ N - \frac{\alpha}{2}(2N - M), & \frac{2}{3} \leq \alpha \leq 1; \\ \frac{M}{2} + \frac{N}{2}(\alpha - 1), & 1 \leq \alpha. \end{cases} \quad (39)$$

Remark 10: The above formula is the same as the one given by (38) with M and N interchanged. This is in accordance with the reciprocity result in Corollary 3. In other words, d_s of the (M, N, M, N) IC with $M > N$ for a given $\frac{M}{N} = r$ is the same as that of the GDoF of a MIMO IC with $M \leq N$ and $\frac{M}{N} = \frac{1}{r}$.

Remark 11: It must be noted that the achievable schemes on the two channels are entirely different. While for $M \leq N$ the coding scheme need not depend on the channel matrices at the transmitters (see the achievability scheme of [14]), for $M > N$ the covariance matrices are necessarily functions of the channel matrices. Hence, a naive extension of the scheme of [14] to the case of $M > N$ is not GDoF optimal. In fact, such a scheme wouldn't even be DoF optimal because while on a MIMO IC with $M \leq N$, receive zero-forcing is sufficient to achieve the DoF region, for $M > N$, knowledge of channel state information at the transmitters (CSIT) is necessary to achieve DoF-optimal performance [22], [23], [24] such as through transmit zero-forcing beamforming [12].

Remark 12 (A scheme that ignores CSIT): The GDoF optimal coding scheme of [14] does not utilize any CSIT. The approach in [14] was to divide the range of α into the five regimes delineated in the SISO IC case in [9] and employ the main idea of the achievable schemes that are known to be GDoF-optimal in the SISO case (e.g., treat interference as noise in the very weak interference regime; set the power level of private messages so they arrive at the noise level at the unintended receiver in the moderate and weak interference regimes following the prescription of [9], send only common messages in the strong and very strong interference regimes). Hence, this coding scheme effectively only employs signal level interference alignment without any form of transmit beamforming. While on an (M, N, M, N) IC with $M \leq N$ beamforming is not necessary because neither of the cross-links have a null space, it is so when $M > N$. Therefore, the coding scheme of [14] when applied naively to the MIMO IC with $M > N$ cannot achieve the fundamental GDoF region of the channel (e.g., see Fig. 2(a) and Fig. 2(b)).

Remark 13 (Treating interference as noise (TIN):) On a 2-user MIMO IC with $\alpha \leq \frac{1}{2}$, $M_1 =$

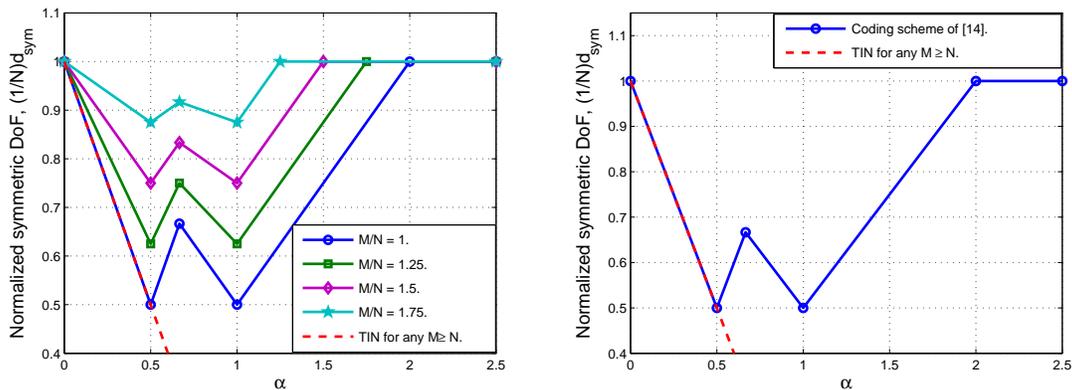
$M_2 = M$ and $N_1 = N_2 = N$, treating interference as noise (TIN) gives the following achievable region

$$\left\{ (R_1, R_2) : R_1 \leq \log \det \left(I_M + \rho H_{11}^\dagger \left(I_N + \rho^\alpha H_{21} H_{21}^\dagger \right)^{-1} H_{11} \right), \right. \\ \left. R_2 \leq \log \det \left(I_M + \rho H_{22}^\dagger \left(I_N + \rho^\alpha H_{12} H_{12}^\dagger \right)^{-1} H_{22} \right) \right\}. \quad (40)$$

The corresponding GDoF region is given as

$$\left\{ (d_1, d_2) : d_1 \leq f(N, (\alpha, M), (1, M)) - \min\{M, N\}\alpha = N(1 - \alpha), \right. \\ \left. d_2 \leq f(N, (\alpha, M), (1, M)) - \min\{M, N\}\alpha = N(1 - \alpha) \right\}.$$

Fig. 2(a) and Fig. 2(b) show the GDoF achievable by the GDoF optimal coding scheme of this paper in comparison with the coding scheme used in [14] and the TIN scheme (whose GDoF is denoted by the dashed red line). Comparing the GDoF curves of the three schemes, it is clear that both TIN and the No-CSIT coding scheme of [14] fail to achieve the fundamental GDoF of the (M, N, M, N) IC when $M > N$.



(a) The symmetric GDoF of the channel.

(b) Symmetric GDoF achievable by the coding scheme of [14].

Fig. 2: Symmetric GDoF of the (M, N, M, N) IC, with $M \geq N$.

V. HOW DOES $\mathcal{HK}(\{K_{1u}, K_{1w}, K_{2u}, K_{2w}\})$ ACHIEVE GDOF OPTIMALITY?

In what follows, we shall explain how the $\mathcal{HK}(\{K_{1u}, K_{1w}, K_{2u}, K_{2w}\})$ scheme achieves the fundamental GDoF region of the MIMO IC through some examples. The encoding scheme is completely specified in Subsection II-C and Lemma 6. As explained in Subsection II-C, in the GDoF optimal coding

scheme the private and public messages of each user are essentially a weighted sum of several independent streams of information, each stream directed along a beam which is dependent on the channel matrix of the cross link emerging from the corresponding transmitter. The direction of these beams and their weights are chosen in such a manner (e.g., see equations (14)-(18)) that the effective covariance matrix of the overall codeword corresponding to each of the message is as given by equation (13)². As for decoding, it is clear that with respect to U_i , W_i and W_j , Rx_i sees a MAC channel for $i \neq j \in \{1, 2\}$ and for any $(d_{1p}, d_{1c}, d_{2p}, d_{2c})$ -tuple belonging to the achievable region (see Lemma 6 in Appendix F) Rx_i can decode U_i , W_i and W_j with probability of error going to zero. Therefore, any decoding scheme which is capacity optimal on a MAC will be GDoF optimal for the MIMO IC if each receiver tries to decode the 2 public messages and its own private message while treating the other private message as noise.

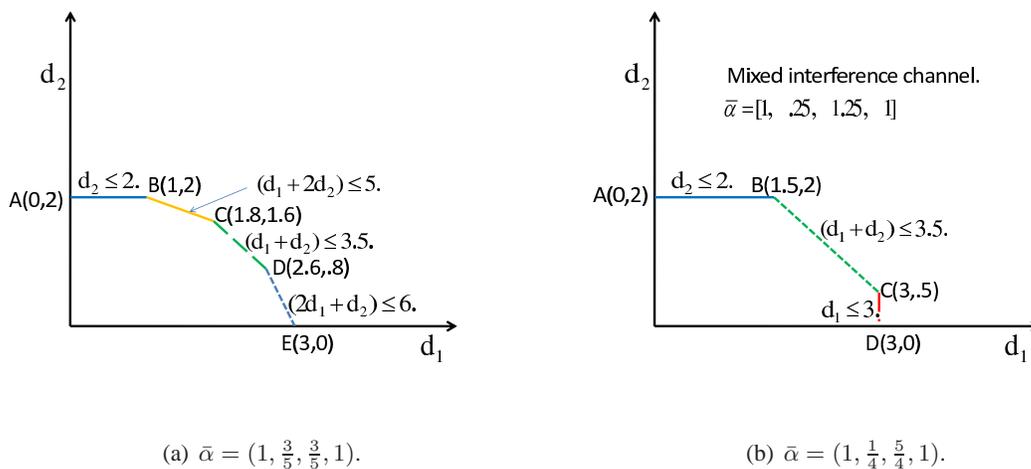


Fig. 3: GDoF region of the $(3, 3, 2, 2)$ IC.

Example 1 (A MIMO IC with weak interference): Figure 3(a) depicts the GDoF region of a $(3, 3, 2, 2)$ MIMO IC with $\bar{\alpha} = [1, \frac{3}{5}, \frac{3}{5}, 1]$. Clearly, it is sufficient to illustrate the achievability of the vertices of the GDoF region since any point on the line joining any two vertices can be achieved via time-sharing. The time sharing argument, however, is just a matter of convenience, and is not necessary to achieve a point in the GDoF region. In fact, we illustrate the achievability of a non-corner point in Remark 14.

²Instead of sending independent streams of information (which is without loss of GDoF optimality), if coding is also done across different streams, it is possible to achieve a larger error exponents.

Note that points A or E can be achieved simply by turning off Tx_1 or Tx_2 , respectively. To analyze the achievability of the other corner points we need to know the DoFs carried by the private and public messages of each user. For the $(3, 3, 2, 2)$ IC with $\bar{\alpha} = [1, \frac{3}{5}, \frac{3}{5}, 1]$, Lemma 6 gives

$$\begin{aligned}
d_{1p} &\leq 1.8; & d_{2p} &\leq .8; \\
d_{1c} &\leq 2; & d_{2c} &\leq 2; \\
d_{2c} &\leq 1.2; & d_{1c} &\leq 1.2; \\
(d_{1p} + d_{1c}) &\leq 3; & \text{and} & (d_{2p} + d_{2c}) \leq 2; \\
(d_{1p} + d_{2c}) &\leq 2.2; & (d_{2p} + d_{1c}) &\leq 1.2; \\
(d_{1c} + d_{2c}) &\leq 2.6; & (d_{2c} + d_{1c}) &\leq 2; \\
(d_{1p} + d_{1c} + d_{2c}) &\leq 3; & (d_{2p} + d_{2c} + d_{1c}) &\leq 2;
\end{aligned} \tag{41}$$

Achievability of point B: From the set of bounds in equation (41) we see the only choice for the different DoFs for the public and private messages of the two users are given as $d_{1p} = 1$, $d_{1c} = 0$, $d_{2p} = .8$ and $d_{2c} = 1.2$. Since the first user needs to send only private information having DoF 1, it is best to send it in the direction of the null space of H_{12} , i.e.,

$$X_1 = \frac{1}{\sqrt{3}} x_{1p}^{(3)} U_{12}^{[3]}. \tag{42}$$

On the other hand, the structure of the codeword for the second user is also clear from equation (18),

$$X_2 = \sum_{k=1}^2 \frac{\sqrt{\rho_{21} \lambda_{21}^{(k)}}}{\sqrt{2(1 + \rho_{21} \lambda_{21}^{(k)})}} x_{2c}^{(k)} U_{21}^{[k]} + \sum_{l=1}^2 \frac{1}{\sqrt{2(1 + \rho_{21} \lambda_{21}^{(l)})}} x_{2p}^{(l)} U_{21}^{[l]}, \tag{43}$$

where $x_{2c}^{(k)}$ and $x_{2p}^{(k)}$ carries .6 and .4 DoFs, respectively for both $k = 1, 2$.

Decoding: Rx_1 first projects the received signal on the 2 dimensional space which is perpendicular to $H_{11}U_{12}^{[3]}$ to remove the effect of $x_{1p}^{(3)}$ by zero forcing. In the resulting 2 dimensional signal space, only contribution from W_2 is present, carrying a DoF of 1.2. This can be decoded because the link from Tx_2 to Rx_1 is a 2×2 point-to-point MIMO channel with effective SNR of ρ^6 . Once decoded, Rx_1 removes its effect from the original received signal (the received signal before zero-forcing) and then it gets a interference-free channel from Tx_1 to itself. It can hence decode U_1 . On the other hand, Rx_2 does not face any interference³ from Tx_1 so that it can decode W_2 while treating U_2 as noise. This is possible because treating U_2 as noise only raises the noise floor to ρ^4 while the received signal power of W_2 is at

³The interference that reach below noise floor is irrelevant in the GDoF computation.

ρ which implies it can decode .6 DoFs from each receive dimension. Next, subtracting the contribution of W_2 from the received signal, Rx_2 can decode U_2 .

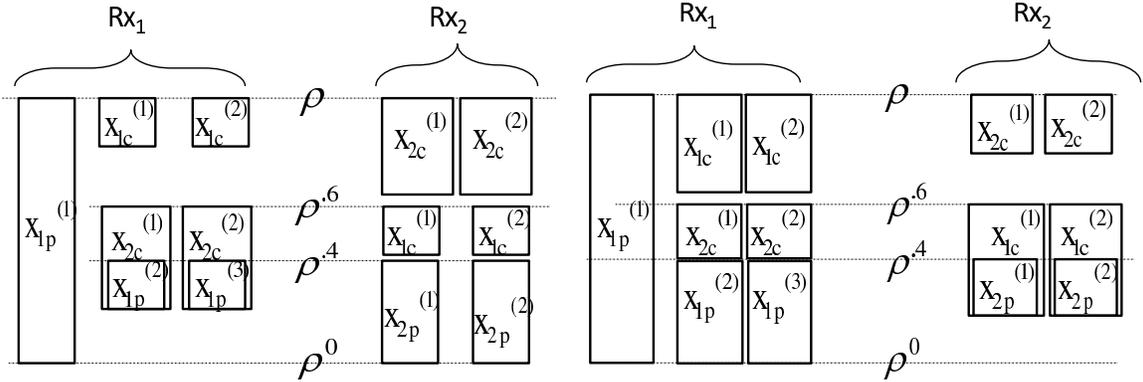
Achievability of point C: Since Rx_2 can support only 2 DoFs at point C, we have $d_{1c} \leq .4$. Combining this with equation (41) we get $d_{1c} = .4$, $d_{1p} = 1.4$, $d_{2p} = .8$ and $d_{2c} = .8$. For this choice of the different rates, the transmit signals at Tx_1 and Tx_2 are given by

$$X_1 = \sum_{k=1}^2 \frac{\sqrt{\rho_{12}\lambda_{12}^{(k)}}}{\sqrt{3(1 + \rho_{12}\lambda_{12}^{(k)})}} x_{1c}^{(k)} U_{12}^{[k]} + \sum_{l=1}^2 \frac{1}{\sqrt{3(1 + \rho_{12}\lambda_{12}^{(l)})}} x_{1p}^{(l)} U_{12}^{[l]} + \frac{1}{\sqrt{3}} x_{1p}^{(3)} U_{12}^{[3]}, \quad (44)$$

where $x_{1c}^{(k)}$ and $x_{1p}^{(k)}$ carries .2 DoFs, for both $k = 1, 2$ and $x_{1p}^{(3)}$ carries 1 DoF and

$$X_2 = \sum_{k=1}^2 \frac{\sqrt{\rho_{21}\lambda_{21}^{(k)}}}{\sqrt{2(1 + \rho_{21}\lambda_{21}^{(k)})}} x_{2c}^{(k)} U_{21}^{[k]} + \sum_{l=1}^2 \frac{1}{\sqrt{2(1 + \rho_{21}\lambda_{21}^{(l)})}} x_{2p}^{(l)} U_{21}^{[l]}, \quad (45)$$

where $x_{2c}^{(k)}$ and $x_{2p}^{(k)}$ carries .4 DoFs each, for both $k = 1, 2$. The different signals at both the receivers are depicted in Fig. 4(a), where each stream is represented by a box the top level of which marks its signal strength and the vertical height is proportional to the DoFs carried by it. Note that, $x_{1p}^{(1)}$ though transmitted at a power level of 1, does not appear at Rx_2 since it is transmitted along the null space of the channel from Tx_1 to Rx_2 .



(a) Receive signal spaces at DoF pair (1.8, 1.6).

(b) Receive signal spaces at DoF pair (2.6, .8).

Fig. 4: GDoF region of a (3, 3, 2, 2) MIMO IC and its explicit achievable scheme.

Decoding: The decoding procedure at Rx_1 is exactly the same as in the previous case. Rx_2 on the other hand, can decode W_2 , W_1 and U_2 , respectively, in that order through successive interference cancellation, i.e., it first decodes W_2 , treating W_1 and U_2 (both of which are received below ρ^6) as noise. Subtracting

the contribution of W_2 , it next decodes W_1 treating U_2 as noise. Finally, subtracting the contribution of W_1 it decodes U_2 . It should be noted that during the decoding of each of these messages the noise floor is actually at the power level of the messages being treated as noise.

Achievability of point D: Again from (41) we get $d_{1p} = 1.8$, $d_{1c} = .8$, $d_{2p} = .4$ and $d_{2c} = .4$, for which X_i for $1 \leq i \leq 2$ can be written as

$$X_1 = \sum_{k=1}^2 \frac{\sqrt{\rho_{12}\lambda_{12}^{(k)}}}{\sqrt{3(1 + \rho_{12}\lambda_{12}^{(k)})}} x_{1c}^{(k)} U_{12}^{[k]} + \sum_{l=1}^2 \frac{1}{\sqrt{3(1 + \rho_{12}\lambda_{12}^{(l)})}} x_{1p}^{(l)} U_{12}^{[l]} + \frac{1}{\sqrt{3}} x_{1p}^{(3)} U_{12}^{[3]}, \quad (46)$$

where $x_{1c}^{(k)}$ and $x_{1p}^{(k)}$ carries .4 DoFs each, for both $k = 1, 2$ and $x_{1p}^{(3)}$ carries 1 DoF and

$$X_2 = \sum_{k=1}^2 \frac{\sqrt{\rho_{21}\lambda_{21}^{(k)}}}{\sqrt{2(1 + \rho_{21}\lambda_{21}^{(k)})}} x_{2c}^{(k)} U_{21}^{[k]} + \sum_{l=1}^2 \frac{1}{\sqrt{2(1 + \rho_{21}\lambda_{21}^{(l)})}} x_{2p}^{(l)} U_{21}^{[l]}, \quad (47)$$

where $x_{2c}^{(k)}$ and $x_{2p}^{(k)}$ carries .2 DoFs each, for both $k = 1, 2$. The different received signals at both the receivers are depicted in Fig. 4(b). It is clear from Fig. 4(b) that a MAC receiver can decode all the messages.

Example 2 (A MIMO IC with mixed interference): Fig. 3(b) depicts the GDoF region of the $(3, 3, 2, 2)$ IC with *mixed* interference, i.e., $\alpha_{12} = .25 < 1$ and $\alpha_{21} = 1.25 > 1$. Points A and D of this channel are achievable simply by making Tx_1 and Tx_2 , respectively, silent. Now for the MIMO IC of Fig. 3(b), from Lemma 6, we get

$$\begin{aligned} d_{1p} &\leq 2.5; & d_{2p} &\leq 0; \\ d_{1c} &\leq 2; & d_{2c} &\leq 2; \\ d_{2c} &\leq 2.5; & d_{1c} &\leq .5; \\ (d_{1p} + d_{1c}) &\leq 3; & \text{and} & (d_{2p} + d_{2c}) &\leq 2; \\ (d_{1p} + d_{2c}) &\leq 3.5; & (d_{2p} + d_{1c}) &\leq .5; \\ (d_{1c} + d_{2c}) &\leq 3.5; & (d_{2c} + d_{1c}) &\leq 2; \\ (d_{1p} + d_{1c} + d_{2c}) &\leq 3.5; & (d_{2p} + d_{2c} + d_{1c}) &\leq 2; \end{aligned} \quad (48)$$

Achievability of point B: From equation (48) we get $d_1 = d_{1p} = 1.5$ and $d_2 = d_{2c} = 2$, which imply $d_{1c} = d_{2p} = 0$. With this choice of the GDoFs carried by the private and public messages we have

$$X_1 = \sum_{k=1}^2 \frac{1}{\sqrt{3(1 + \rho_{12}\lambda_{12}^{(k)})}} x_{1p}^{(k)} U_{12}^{[k]} + \frac{1}{\sqrt{3}} x_{1p}^{(3)} U_{12}^{[3]}, \quad (49)$$

where $x_{1p}^{(k)}$ carry .25, .25 and 1 DoFs, for $k = 1, 2$ and 3, respectively. On the other hand,

$$X_2 = \sum_{k=1}^2 \frac{1}{\sqrt{2}} x_{2c}^{(k)} U_{21}^{[k]}, \quad (50)$$

where $x_{2c}^{(k)}$ carries 1 DoFs for both $k = 1, 2$.

Decoding: Rx_1 first projects the received signal on the 2-dimensional space which is perpendicular to $H_{11}U_{12}^{[3]}$ to remove the effect of $x_{1p}^{(3)}$ by zero-forcing. In the resulting 2-dimensional signal space, contribution from U_1 and W_2 are present, together they are carrying 2.5 DoFs and can be decoded by Rx_1 , since as a MAC receiver it has a sum GDoF of 2.5. Once decoded, it removes the effect of these signals from the original received signal (the received signal before zero-forcing), and decodes $x_{1p}^{(3)}$. On the other, Rx_2 does not face any interference from Tx_1 and so it can decode W_2 .

Achievability of point C: Similarly, at point C we have $d_{1c} = .5$, $d_{1p} = 2.5$, $d_{2c} = .5$, $d_{2p} = 0$ and the transmit signals are

$$X_1 = \sum_{k=1}^2 \frac{\sqrt{\rho_{12}\lambda_{12}^{(k)}}}{\sqrt{3(1 + \rho_{12}\lambda_{12}^{(k)})}} x_{1c}^{(k)} U_{12}^{[k]} + \sum_{l=1}^2 \frac{1}{\sqrt{3(1 + \rho_{12}\lambda_{12}^{(l)})}} x_{1p}^{(l)} U_{12}^{[l]} + \frac{1}{\sqrt{3}} x_{1p}^{(3)} U_{12}^{[3]}, \quad (51)$$

where $x_{1c}^{(k)}$ and $x_{1p}^{(k)}$ carry .25 and .75 DoFs, respectively, for both $k = 1, 2$ and $x_{1p}^{(3)}$ carries 1 DoF. On the other hand,

$$X_2 = \sum_{k=1}^2 \frac{1}{\sqrt{2}} x_{2c}^{(k)} U_{21}^{[k]}, \quad (52)$$

where $x_{2c}^{(k)}$ carries .25 DoFs, for both $k = 1, 2$.

Decoding: At Rx_1 , first the contribution of $x_{1p}^{(3)}$ is zero-forced, and then W_2 , W_1 and the remaining part of U_1 are decoded successively, in that order. During the decoding of each of these messages the others are treated as noise. After decoding a message, it is subtracted from the original signal. At Rx_2 also W_2 , W_1 are decoded successively following a similar method.

Remark 14 (Achievability of a non-corner point): To illustrate the fact that all the points in between the corner points can be achieved without time-sharing, we consider the point (2.25, 1.25). From equation (48) we know that $d_{1p} = 1.75$, $d_{1c} = .5$ and $d_2 = d_{2c} = 1.25$ is achievable. The corresponding codewords are

$$X_1 = \sum_{k=1}^2 \frac{\sqrt{\rho_{12}\lambda_{12}^{(k)}}}{\sqrt{3(1 + \rho_{12}\lambda_{12}^{(k)})}} x_{1c}^{(k)} U_{12}^{[k]} + \sum_{l=1}^2 \frac{1}{\sqrt{3(1 + \rho_{12}\lambda_{12}^{(l)})}} x_{1p}^{(l)} U_{12}^{[l]} + \frac{1}{\sqrt{3}} x_{1p}^{(3)} U_{12}^{[3]}, \quad (53)$$

where $x_{1c}^{(k)}$ and $x_{1p}^{(k)}$ carry .25 and $3/8$ DoFs, respectively for both $k = 1, 2$ and $x_{1p}^{(3)}$ carries 1 DoF. On the other hand,

$$X_2 = \sum_{k=1}^2 \frac{1}{\sqrt{2}} x_{2c}^{(k)} U_{21}^{[k]}, \quad (54)$$

where $x_{2c}^{(k)}$ carries $5/8$ DoFs, for both $k = 1, 2$.

Decoding: Decoding is similar to methods described above. The details are skipped.

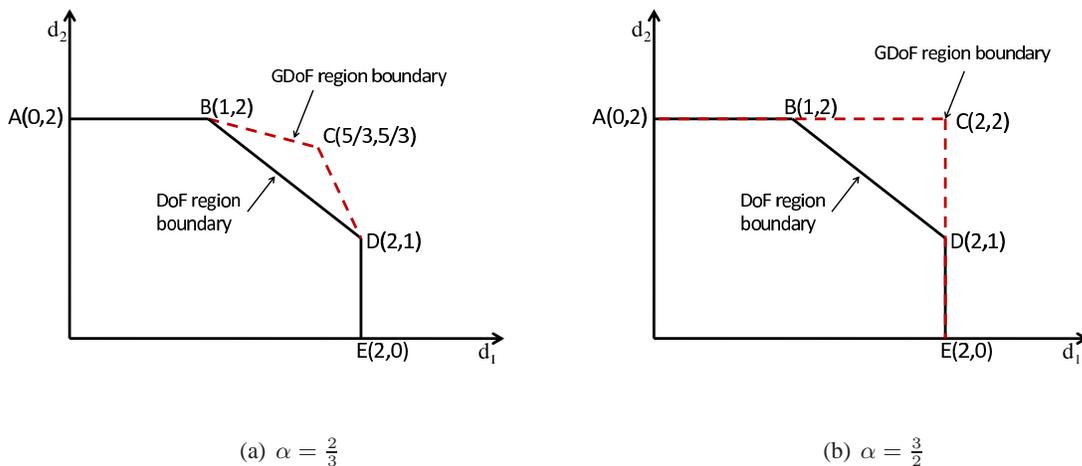


Fig. 5: GDoF region of the $(3, 2, 3, 2)$ MIMO IC with $\alpha_{11} = \alpha_{22} = 1$ and $\alpha_{12} = \alpha_{21} = \alpha$.

VI. FURTHER INSIGHTS

A. Only Tx/Rx ZF Beam-forming is not GDoF optimal

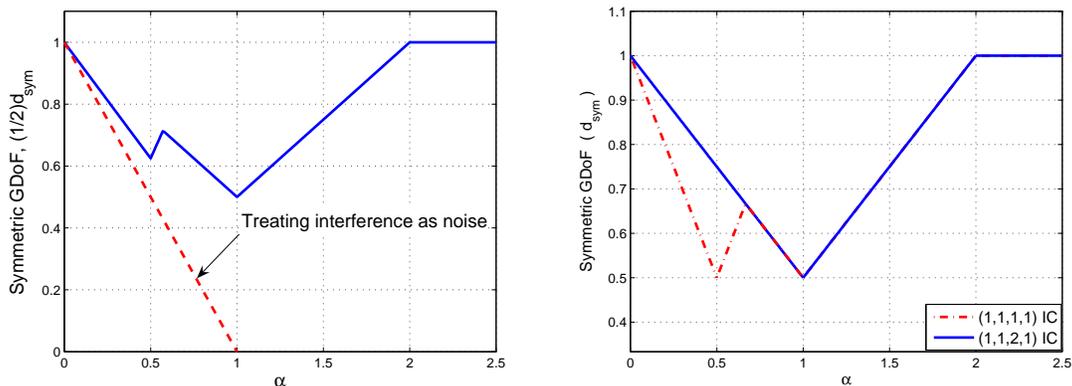
The fundamental GDoF gives a finer high SNR approximation than the DoF approximation and therefore reveals insights that are not revealed by the DoF analysis. Figure 5(a) illustrates this point by comparing the DoF and GDoF region of the $(3, 2, 3, 2)$ IC with $\bar{\alpha} = [1, \frac{2}{3}, \frac{2}{3}, 1]$. It is known from [12] that only transmit/receive zero-forcing beam-forming is sufficient to achieve any point in the DoF region of the channel. The DoF region achievable using this scheme is shown in Fig. 5 as against the fundamental GDoF region. It is easily seen that forgoing the opportunity to align signals in the signal-level dimension leads to a strictly GDoF suboptimal performance. In particular, this technique can not achieve any point in the triangular region BCD. However, the coding scheme of Section II which in addition to beamforming, also employs signal-level interference alignment, can achieve all the points in the region BCD.

B. On achieving single-user performance

It is well known that the achievability of single user DoFs on a MIMO MAC or BC depends on the number of antennas at the different nodes. Moreover, on a SISO IC, it depends on the interference level, α . On a MIMO IC it depends on *both*. From Corollary 4 we get that on a (M, N, M, N) IC with $M \geq N$, the single user GDoF is achieved (each user gets N DoFs) when

$$\alpha \geq \alpha^* = \left(3 - \frac{M}{N}\right).$$

In contrast to the case on a SISO IC, the value of α at which the single-user performance is achieved, denoted by α^* , decreases below 2 as M increases, giving a similar effect as in a MAC or BC (see Fig. 2(a)).



(a) Symmetric GDoF region of the $(3, 2, 2, 3)$ IC.

(b) Symmetric GDoF region of a $(1, 1, 2, 1)$ MIMO IC

Fig. 6: Sub-optimality of TIN and deviation of the GDoF boundary from the well known “W” shape.

C. Sub-optimality of treating interference as noise

Another fundamental difference of the MIMO IC from the SISO IC revealed by the GDoF analysis is this: in general, treating interference as noise (TIN) is *not* GDoF optimal on a MIMO IC even in the *very weak interference* regime, i.e., when $\alpha \leq \frac{1}{2}$. This is seen in Fig. 2(a) where the dotted line, which represents the symmetric GDoF achievable by TIN, is strictly sub-optimal with respect to the fundamental GDoF of the channel for $\alpha \leq \frac{1}{2}$ whenever $M/N > 1$. See also Fig. 6(a) which illustrates this point for the $(3, 2, 2, 3)$ MIMO IC.

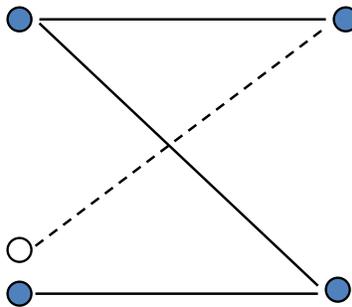


Fig. 7: Diagonalization of the cross links using ZF and BF.

D. Deviation from the “W” shape

Unlike in the SISO IC, the symmetric GDoF region of a MIMO IC in general need not maintain the “W” shape. The deviation in general is due to asymmetry in the numbers of antennas. For example, consider the $(1, 1, 2, 1)$ IC with $\alpha_{ii} = 1$ and $\alpha_{ij} = \alpha$, for $i \neq j \in \{1, 2\}$. The best achievable symmetric DoF ($d_{sym} = d_1 = d_2$) on this channel denoted by d is

$$d = \begin{cases} 1 - \frac{\alpha}{2}, & 0 \leq \alpha \leq 1; \\ \frac{\alpha}{2}, & 1 \leq \alpha \leq 2; \\ 1, & 2 \leq \alpha. \end{cases}$$

which is depicted in Figure 6(b). Diagonalizing the cross-link from Tx_2 to Rx_1 and then turning off the subchannel which interferes with Rx_1 gives the GDoF equivalent channel of Figure 7(a) which is a SISO “Z” IC. The symmetric GDoF region of this channel is indeed “V” shaped as found in [9]. Although a little more involved, the distorted “W” of Figure 6(a) for the $(3, 2, 2, 3)$ MIMO IC can be explained similarly.

VII. CONCLUSION

The GDoF analysis of this paper, unifies and generalizes the earlier results on GDoF of SISO IC [9], the DoF region [12] of MIMO IC and the symmetric GDoF [14] of MIMO IC through a single achievable scheme for all. The coding schemes in [12] and [14] are strictly suboptimal in the GDoF sense on a general 2-user MIMO IC in one case or other. The analysis here reveals various insights about the MIMO IC including the fact that in general, partially decoding the unintended user’s message is necessary to be GDoF optimal even in the so called very weak interference regime. The two types of signaling dimensions

available on a MIMO IC – namely, signal space and signal level – are jointly and optimally exploited in the GDoF optimal scheme.

APPENDIX A

PROOF OF LEMMA 4

We shall consider two different cases: 1) $(u_1 + u_2) \geq u$; and 2) $(u_1 + u_2) < u$. The first case was proved in Lemma 1 of [14] which gives

$$\log \det \left(I_u + \rho^a H_1 H_1^\dagger + \rho^b H_2 H_2^\dagger \right) = \min\{u, u_1\} a \log(\rho) + (u - u_1)^+ b \log(\rho) + \mathcal{O}(1). \quad (55)$$

However when $(u_1 + u_2) < u$ we have the following

$$\begin{aligned} \log \det \left(I_u + \rho^a H_1 H_1^\dagger + \rho^b H_2 H_2^\dagger \right) &= \log \det \left(I_u + [H_1 \ H_2] \begin{bmatrix} \rho^a I_{u_1} & 0 \\ 0 & \rho^a I_{u_2} \end{bmatrix} [H_1 \ H_2]^\dagger \right) \\ &= \log \det \left(I_{(u_1+u_2)} + H^\dagger H \begin{bmatrix} \rho^a I_{u_1} & 0 \\ 0 & \rho^a I_{u_2} \end{bmatrix} \right), \quad [\cdot: H = [H_1 \ H_2]] \\ &\stackrel{(a)}{=} \log \det \left(H^\dagger H \begin{bmatrix} \rho^a I_{u_1} & 0 \\ 0 & \rho^a I_{u_2} \end{bmatrix} \right) + o(1) \\ &= (u_1 a + u_2 b) \log(\rho) + \mathcal{O}(1), \end{aligned} \quad (56)$$

where step (a) follows from the fact that $H^\dagger H$ is full rank, since H is full rank by assumption. Finally combining equations (55) and (56) we have the desired result.

APPENDIX B

PROOF OF LEMMA 5

Without loss of generality, let us assume that $a_1 \geq \max\{a_2, a_3\}$. In the proof we shall use the following notations:

$$L_t = \log \det \left(I_u + \sum_{i=1}^3 \rho^{a_i} H_i H_i^\dagger \right), \quad \Lambda = \begin{bmatrix} \rho^{a_2} I_{u_2} & 0 \\ 0 & \rho^{a_3} I_{u_3} \end{bmatrix}, \quad \text{and } H_{23} = [H_2 \ H_3],$$

where the entries of H_{23} are iid that come from a continuous distribution and hence H_{23} is full rank w.p.1. Using the identity $\log \det(I + AB) = \log \det(I + BA)$ we get

$$\begin{aligned} L_t &= \log \det \left(I_u + \rho^{a_1} H_1 H_1^\dagger + H_{23} \Lambda H_{23}^\dagger \right), \\ &= \log \det \left(I_u + \rho^{a_1} H_1 H_1^\dagger \right) + \log \det \left(I_{u_2+u_3} + \Lambda H_{23}^\dagger \left(I_u + \rho^{a_1} H_1 H_1^\dagger \right)^{-1} H_{23} \right), \\ &= \min\{u, u_1\} a_1 \log(\rho) + \log \det \left(I_{u_2+u_3} + \Lambda H_{23}^\dagger \left(I_u + \rho^{a_1} H_1 H_1^\dagger \right)^{-1} H_{23} \right) + \mathcal{O}(1), \end{aligned} \quad (57)$$

Next, we approximate the second term on the right hand side of the last equation as

$$\begin{aligned} L_{t2} &= \log \det \left(I_{u_2+u_3} + \Lambda H_{23}^\dagger \left(I_u + \rho^{a_1} H_1 H_1^\dagger \right)^{-1} H_{23} \right) \\ &= \log \det \left(I_{u_2+u_3} + \Lambda H_{23}^\dagger U \begin{bmatrix} (I_{\min\{u,u_1\}} + \rho^{a_1} \Lambda_1)^{-1} & 0 \\ 0 & I_{(u-u_1)^+} \end{bmatrix} U^\dagger H_{23} \right), \end{aligned} \quad (58)$$

where in the last step we have used the eigen-decomposition of the matrix $H_1 H_1^\dagger$, where Λ_1 is a diagonal matrix containing the positive eigenvalues only and U is the unitary matrix containing all the eigen-vectors. Since both $H_2, H_3 \in \mathcal{P}$, $H \triangleq U^\dagger H_{23}$ is identically distributed as H_{23} and thus $H \in \mathcal{P}$. Suppose the rows of the matrix H are divided into two sub sets: $G_1^\dagger = H^{(1:\min\{u,u_1\})}$ and $G_2^\dagger = H^{(\min\{u,u_1\}+1:u)}$, then both $G_1^\dagger, G_2^\dagger \in \mathcal{P}$ and from the last equation we get

$$\begin{aligned} L_{t2} &= \log \det \left(I_{u_2+u_3} + \Lambda G_1 (I_{u_1} + \rho^{a_1} \Lambda_1)^{-1} G_1^\dagger + \Lambda G_2 G_2^\dagger \right), \\ &\stackrel{(c)}{=} \log \det \left(I_{u_2+u_3} + \Lambda G_2 G_2^\dagger \right) + \mathcal{O}(1), \\ &= \log \det \left(I_{(u-u_1)^+} + G_2^\dagger \Lambda G_2 \right) + \mathcal{O}(1), \end{aligned} \quad (59)$$

where step (c) follows from the fact that $a_1 \geq \max\{a_2, a_3\}$. Since $G_2^\dagger \in \mathcal{P}$, it is full rank and so are $(G_2^\dagger)^{[1:u_2]} = G_{21}$ and $(G_2^\dagger)^{[u_2+1:u_2+u_3]} = G_{22}$. Putting these into equation (59) we get

$$\begin{aligned} L_{t2} &= \log \det \left(I_{u-u_1} + \rho^{a_2} G_{21} G_{21}^\dagger + \rho^{a_3} G_{22} G_{22}^\dagger \right) + \mathcal{O}(1), \\ &= (\min\{(u-u_1)^+, u_2\} a_2 + \min\{(u-u_1-u_2)^+, u_3\} a_3) \log(\rho) + \mathcal{O}(1), \end{aligned}$$

where the last step follows from Lemma 4. Finally, substituting this into equation (57), we get the desired result.

APPENDIX C

PROOF OF THEOREM 1

1st and 2nd bound: We start with the first two constraints in equation (5) and (6),

$$I_1 \triangleq \log \det \left(I_{N_1} + \rho H_{11} H_{11}^\dagger \right) = \min\{M_1, N_1\} \log(\rho) + \mathcal{O}(1); \quad [\cdot : \text{Lemma 4 with } a = 1 \text{ and } b = 0]$$

$$I_2 \triangleq \log \det \left(I_{N_1} + \rho^{\alpha_{22}} H_{22} H_{22}^\dagger \right) = \min\{M_2, N_2\} \alpha_{22} \log(\rho) + \mathcal{O}(1). \quad [\cdot : \text{Lemma 4 with } a = \alpha_{22} \text{ and } b = 0]$$

Putting these into equations (26) and (27) we get

$$d_1 \leq \lim_{\rho \rightarrow \infty} \frac{I_1}{\log(\rho)} = \min\{M_1, N_1\}; \quad (60)$$

$$d_2 \leq \lim_{\rho \rightarrow \infty} \frac{I_2}{(\alpha_{22} \log(\rho))} = \min\{M_2, N_2\}. \quad (61)$$

3rd and 4th bound: Using Lemma 4 we get

$$\begin{aligned} I_{31} &\triangleq \log \det \left(I_{N_2} + \rho^{\alpha_{12}} H_{12} H_{12}^\dagger + \rho^{\alpha_{22}} H_{22} H_{22}^\dagger \right) \\ &= f(N_2, (\alpha_{12}, M_1), (\alpha_{22}, M_2)) \log(\rho) + \mathcal{O}(1), \end{aligned} \quad (62)$$

where $f(\cdot, \cdot, \cdot)$ is as defined in equation (23). The second term in equation (7) can again be approximated by using Lemma 4 twice (recall Remark 2) as follows.

$$\begin{aligned} I_{32} &= \log \det \left(I_{N_1} + \rho_{11} H_{11} \left(I_{M_1} + \rho_{12} H_{12}^\dagger H_{12} \right)^{-1} H_{11}^\dagger \right), \\ &= \log \det \left(I_{M_1} + \rho_{12} H_{12}^\dagger H_{12} + \rho_{11} H_{11}^\dagger H_{11} \right) - \log \det \left(I_{M_1} + \rho_{12} H_{12}^\dagger H_{12} \right), \\ &= f(M_1, (\alpha_{12}, N_2), (\alpha_{11}, N_1)) \log(\rho) - \min\{M_1, N_2\} \alpha_{12} \log(\rho) + \mathcal{O}(1), \end{aligned} \quad (63)$$

$$= f(N_1, ((1 - \alpha_{12})^+, m_{12}), (1, (M_1 - N_2)^+)) \log(\rho) + \mathcal{O}(1), \quad (64)$$

where the last step follows from the definition of $f(\cdot)$, i.e., equation (23). Next, putting equations (62), and (63) in equation (28) we get

$$\begin{aligned} d_1 + \alpha_{22} d_2 &\leq \lim_{\rho \rightarrow \infty} \frac{I_3}{\log(\rho)} = \lim_{\rho \rightarrow \infty} \frac{I_{31} + I_{32}}{\log(\rho)}; \\ &= f(N_2, (\alpha_{12}, M_1), (\alpha_{22}, M_2)) + f(N_1, ((1 - \alpha_{12})^+, m_{12}), (1, (M_1 - N_2)^+)). \end{aligned}$$

Similarly, approximating the terms in equation (8) by Lemma 4 and putting it in equation (29) we get

$$d_1 + \alpha_{22} d_2 \leq f(N_1, (\alpha_{21}, M_2), (\alpha_{11}, M_1)) + f(N_2, ((\alpha_{22} - \alpha_{21})^+, m_{21}), (\alpha_{22}, (M_2 - N_1)^+)).$$

5th bound: Note that neither of the terms in equation (9) are in a form on which we can apply Lemma 4 or 5. However, as we shall see next, these terms can be expressed in an alternative format on which Lemma 5 can be used. Let the eigenvalue decomposition of the matrix $H_{12}^\dagger H_{12}$ is given as

$$H_{12}^\dagger H_{12} = U \Lambda U^\dagger, \text{ where } \Lambda = \begin{bmatrix} \Lambda^+ & 0 \\ 0 & 0_{(M_1 - N_2)^+} \end{bmatrix}$$

and Λ^+ is a diagonal matrix containing only the positive eigenvalues. Using this decomposition we get

$$\begin{aligned} I_{51} &\triangleq \log \det \left(I_{N_1} + \rho^{\alpha_{21}} H_{21} H_{21}^\dagger + \rho H_{11} \left(I_{M_1} + \rho_{12} H_{12}^\dagger H_{12} \right)^{-1} H_{11}^\dagger \right) \\ &= \log \det \left(I_{N_1} + \rho^{\alpha_{21}} H_{21} H_{21}^\dagger + \rho H_{11} U \begin{bmatrix} (I_{m_{12}} + \rho^{\alpha_{12}} \Lambda^+)^{-1} & 0 \\ 0 & I_{(M_1 - N_2)^+} \end{bmatrix} U^\dagger H_{11}^\dagger \right), \\ &= \log \det \left(I_{N_1} + \rho^{\alpha_{21}} H_{21} H_{21}^\dagger + \rho \tilde{H} \begin{bmatrix} (I_{m_{12}} + \rho^{\alpha_{12}} \Lambda^+)^{-1} & 0 \\ 0 & I_{(M_1 - N_2)^+} \end{bmatrix} \tilde{H}^\dagger \right), \quad [\because \tilde{H} = H_{11} U] \end{aligned}$$

Note that \tilde{H} is identically distributed to H_{11} . Therefore, both $G_1 = \tilde{H}^{[1:m_{12}]}$ and $G_2 = \tilde{H}^{[(m_{12}+1):M_1]}$ have the same distribution as specified in section II, having all the properties of a typical channel matrix of the 2-user MIMO IC. Substituting this in the last equation we get

$$\begin{aligned} I_{51} &= \log \det \left(I_{N_1} + \rho^{\alpha_{21}} H_{21} H_{21}^\dagger + \rho G_1 (I_{m_{12}} + \rho^{\alpha_{12}} \Lambda^+)^{-1} G_1^\dagger + \rho G_2 G_2^\dagger \right) \\ &= \log \det \left(I_{N_1} + \rho^{\alpha_{21}} H_{21} H_{21}^\dagger + \rho^{(1-\alpha_{12})} G_1 G_1^\dagger + \rho G_2 G_2^\dagger \right) + o(1). \end{aligned}$$

Clearly, we can now apply Lemma 5 on equation (65),

$$I_{51} = g \left(N_1, (\alpha_{21}, M_2), ((1 - \alpha_{12})^+, m_{12}), (1, (M_1 - N_2)^+) \right) \log(\rho) + \mathcal{O}(1). \quad (65)$$

Applying similar technique for the other term in equation (9) we get

$$I_{52} = g \left(N_2, (\alpha_{12}, M_1), ((\alpha_{22} - \alpha_{21})^+, m_{21}), (\alpha_{22}, (M_2 - N_1)^+) \right) \log(\rho) + \mathcal{O}(1). \quad (66)$$

Finally, using this expressions for I_{51} and I_{52} in equation (30) we get the 5th bound for the GDoF region.

6th and 7th bound: Note that equations (10) and (11) involves terms whose approximations are already computed. Using those approximations we get the remaining 2 bounds of the GDoF region completing the proof.

APPENDIX D

PROOF OF COROLLARY 2

We have to prove that the bounds in equations (34e)-(34g) are looser than the others. To analyze the 5-th bound, we start from equation (34e),

$$\begin{aligned} (d_1 + d_2) &\leq (N_1 \wedge M_2) + ((N_1 - M_2)^+ \wedge (M_1 - N_2)^+) + (N_2 \wedge M_1) + ((M_1 - N_2)^+ \wedge (M_2 - N_1)^+); \\ &= \begin{cases} M_2 + N_2 + \{(N_1 - M_2) \wedge (M_1 - N_2)\}, & \text{if } N_1 > M_2 \text{ and } M_1 > N_2; \\ N_1 + N_2, & \text{if } N_1 \leq M_2 \text{ and } M_1 > N_2; \\ M_1 + M_2, & \text{if } N_1 > M_2 \text{ and } M_1 \leq N_2; \\ M_1 + N_1 + \{(N_2 - M_1) \wedge (M_2 - N_1)\}, & \text{if } N_1 \leq M_2 \text{ and } M_1 \leq N_2; \end{cases} \\ &= \begin{cases} \min\{(M_1 + M_2), (N_1 + N_2)\}, & \text{if } N_1 > M_2 \text{ and } M_1 > N_2; \\ N_1 + N_2, & \text{if } N_1 \leq M_2 \text{ and } M_1 > N_2; \\ M_1 + M_2, & \text{if } N_1 > M_2 \text{ and } M_1 \leq N_2; \\ \min\{(M_1 + M_2), (N_1 + N_2)\}, & \text{if } N_1 \leq M_2 \text{ and } M_1 \leq N_2; \end{cases} \end{aligned}$$

Clearly, this is looser than both the 3rd and the 4th bound. Consider next the 6th bound which, from equation (34f), is given as

$$\begin{aligned}
(2d_1 + d_2) &\leq (N_1 \wedge (M_1 + M_2)) + N_1 \wedge (M_1 - N_2)^+ + (N_2 \wedge M_1) + ((M_1 - N_2)^+ \wedge (M_2 - N_1)^+); \\
&= \begin{cases} (N_1 \wedge (M_1 + M_2)) + (N_1 \wedge (M_1 - N_2)) + N_2, & \text{if } M_1 \geq N_2; \\ (N_1 \wedge (M_1 + M_2)) + M_1 + ((N_2 - M_1) \wedge (M_2 - N_1)^+), & \text{if } M_1 < N_2; \end{cases} \\
&= \begin{cases} (N_1 \wedge (M_1 + M_2)) + (N_1 \wedge (M_1 - N_2)) + N_2, & \text{if } M_1 \geq N_2; \\ (N_1 \wedge (M_1 + M_2)) + \min\{N_2, (M_1 + M_2 - N_1)\}, & \text{if } M_1 < N_2 \text{ and } M_2 > N_1; \\ (N_1 \wedge (M_1 + M_2)) + M_1, & \text{if } M_1 < N_2 \text{ and } M_2 \leq N_1; \end{cases} \\
&= \begin{cases} (N_1 \wedge (M_1 + M_2)) + \min\{(N_1 + N_2), M_1\}, & \text{if } M_1 \geq N_2; \\ \min\{(M_1 + M_2), (N_1 + N_2)\}, & \text{if } M_1 < N_2 \text{ and } M_2 > N_1; \\ M_1 + (N_1 \wedge (M_1 + M_2)), & \text{if } M_1 < N_2 \text{ and } M_2 \leq N_1; \end{cases} \\
&= \begin{cases} (N_1 \wedge (M_1 + M_2)) + \min\{(N_1 + N_2), \max(N_2, M_1)\}, & \text{if } M_1 \geq N_2; \\ \min\{M_1 + \max(M_2, N_1), N_1 + \max(N_2, M_1)\}, & \text{if } M_1 < N_2 \text{ and } M_2 > N_1; \\ M_1 + \min\{\max(N_1, M_2), (M_1 + M_2)\}, & \text{if } M_1 < N_2 \text{ and } M_2 \leq N_1; \end{cases}
\end{aligned}$$

It is clear that this bound is looser than the sum of the 1st and 3rd or the sum of the 1st and the 4th bounds. The proof of the fact that the same is true for the 7th bound is identical and is hence skipped.

APPENDIX E

GDOF REGION OF THE 2-USER MIMO MAC

Both the set of lower and upper bounds to the capacity region of the 2-user MIMO IC contain terms that also appear in the capacity region of a 2-user MIMO MAC channel. Thus, as a by product we can obtain the GDoF region of the MIMO MAC channel.

Consider a MIMO MAC with two transmitters having M_1 and M_2 antennas, respectively, and with N receive antennas at the common receiver. The input-output relation for this channel can be written as

$$Y = \sqrt{\rho} H X_1 + \sqrt{\rho^\alpha} G X_2 + Z,$$

where $X_i \in \mathbb{C}^{M_i \times 1}$ is the transmitted signal from user i , where $Y \in \mathbb{C}^{N \times 1}$ is the received signal and $H \in \mathbb{C}^{N \times M_1}$ and $G \in \mathbb{C}^{N \times M_2}$ are the channel matrices from users 1 and 2 to the receiver, respectively, both of which are assumed to have full rank, and $Z \sim \mathcal{CN}(\mathbf{0}, I_N)$ is additive white Gaussian noise. Without loss of generality, we assume that the SNR of the second user is represented as ρ^α .

Let $\mathcal{C}_{MAC}(H, G)$ denote the capacity region of the 2-user MIMO MAC defined above. The GDoF region is defined as

$$\mathcal{D}_{MAC} = \left\{ (d_1, d_2) : d_i = \lim_{\rho \rightarrow \infty} \frac{R_i}{\log(\rho)}, i \in \{1, 2\} \text{ and } (R_1, R_2) \in \mathcal{C}_{MAC}(H, G) \right\}.$$

The result below gives the GDoF region of the 2-user MIMO MAC.

Corollary 5: The GDoF region of the 2-user MIMO MAC defined above is given as

$$\left\{ (d_1, d_2) : d_1 \leq \min\{M_1, N\}; \right. \\ d_2 \leq \min\{M_2, N\}\alpha; \\ \left. (d_1 + d_2) \leq f(N, (\alpha, M_2), (1, M_1)) \right\},$$

where $f(\cdot, \cdot, \cdot)$ is given by equation (23).

Proof: Following the analysis of the MIMO IC in [11], it can be easily shown that an achievable rate region of the MIMO MAC is given as

$$\mathcal{R}_A = \left\{ (R_1, R_2) : R_1 \leq \log \det \left(I_N + \rho H H^\dagger \right) - N \log(M_1); \right. \\ R_2 \leq \log \det \left(I_N + \rho^\alpha G G^\dagger \right) - N \log(M_2); \\ \left. R_1 + R_2 \leq \log \det \left(I_N + \rho H H^\dagger + \rho^\alpha G G^\dagger \right) - N \log(\max\{M_1, M_2\}) \right\},$$

and an upper bound is given as

$$\mathcal{R}^U = \left\{ (R_1, R_2) : R_1 \leq \log \det \left(I_N + \rho H H^\dagger \right); \right. \\ R_2 \leq \log \det \left(I_N + \rho^\alpha G G^\dagger \right); \\ \left. R_1 + R_2 \leq \log \det \left(I_N + \rho H H^\dagger + \rho^\alpha G G^\dagger \right) \right\}.$$

Note that the two regions differ only by constant (independent of SNR) number of bits. The desired result now follows by replacing \mathcal{C}_{MAC} in the definition of the GDoF region by \mathcal{R}^U or \mathcal{R}_A , since a constant number of bits are insignificant in the GDoF analysis. ■

Remark 15: The GDoF regions for the case when $N \geq (M_1 + M_2)$ is depicted in Fig. 8(a) and the case when $\max\{M_1, M_2\} < N < (M_1 + M_2)$ is depicted in Fig. 8(b), where $A = (M_1, (N - M_1)\alpha)$, $B = ((N - M_2)\alpha + M_1(1 - \alpha), M_2\alpha)$, $A' = (M_1, (N - M_1))$, $B' = ((N - M_2), M_2)$, $A'' = (M_1, (N - M_1) + M_2(\alpha - 1))$ and $B'' = ((N - M_2), M_2\alpha)$. Although, the GDoF analysis reveals the possibility of achieving a larger sum DoFs when one of the link's strength is exponentially larger than the other ($\alpha > 1$), it is not as interesting as the MIMO IC since the GDoF region of the MAC can be achieved

using independent Gaussian codes with scaled identity input covariances at each transmitter and joint decoding just as in a MAC with $\alpha = 1$. In other words, this DoF-optimal scheme is also GDoF-optimal.

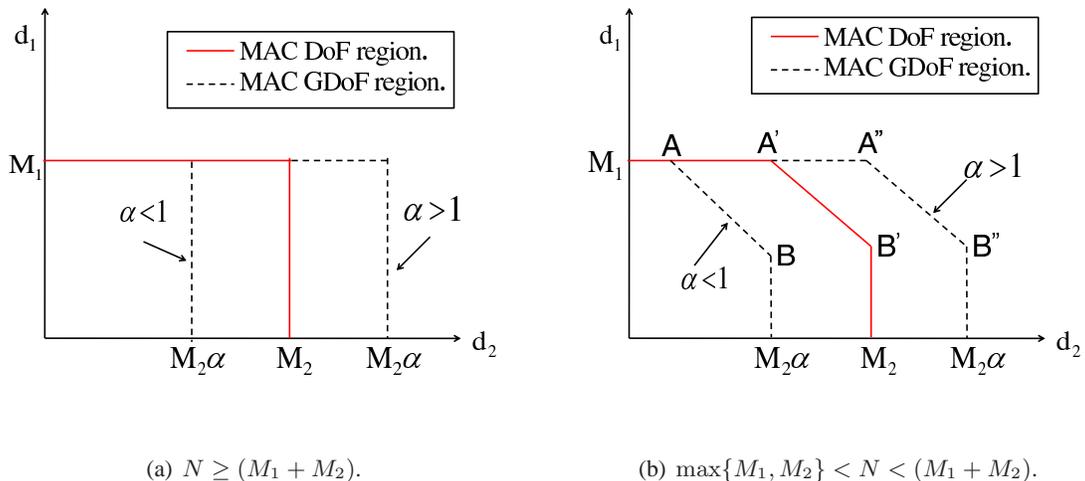


Fig. 8: The GDoF region of the MIMO MAC.

APPENDIX F

PROOF OF EQUIVALENT GDOF REGION

In the HK coding scheme [13], each user's message is divided into two parts called the private (U_i) and public (W_i) messages with rate S_i and T_i , respectively. It was proved in [13] that for any given probability distribution $P(\cdot)$ which factors as

$$P(Q, W_1, U_1, W_2, U_2, X_1, X_2) = P(Q)P(U_1|Q)P(W_1|Q)P(U_2|Q)P(W_2|Q) \\ P(X_1|U_1, U_2, Q)P(X_2|U_2, W_2, Q), \quad (67)$$

the rate region $\mathcal{R}_{\text{HK}}^o(P) = \mathcal{R}_{\text{HK}}^{(o,1)}(P) \cap \mathcal{R}_{\text{HK}}^{(o,2)}(P)$ is achievable where

$$\begin{aligned} \mathcal{R}_{\text{HK}}^{(o,i)}(P) = \left\{ (S_1, T_1, S_2, T_2) : \right. & S_i \leq I(U_i; Y_i | W_i, W_j, Q); \\ & T_i \leq I(W_i; Y_i | U_i, W_j, Q); \\ & T_j \leq I(W_j; Y_i | W_i, U_i, Q); \\ & (S_i + T_i) \leq I(U_i W_i; Y_i | W_j, Q); \\ & (S_i + T_j) \leq I(U_i W_j; Y_i | W_i, Q); \\ & (T_i + T_j) \leq I(W_i W_j; Y_i | U_i, Q); \\ & \left. (S_i + T_i + T_j) \leq I(U_i W_i W_j; Y_i | Q); \right\} \end{aligned}$$

for $i \neq j \in \{1, 2\}$. Let the 2-dimensional projection of the set $\mathcal{R}_{\text{HK}}^o(P)$ be denoted by $\Pi(\mathcal{R}_{\text{HK}}^o)$, which is defined as follows

$$\Pi(\mathcal{R}_{\text{HK}}^o(P)) = \{(0 \leq R_1 \leq (S_1 + T_1), 0 \leq R_2 \leq (S_2 + T_2)) : (S_1, T_1, S_2, T_2) \in \mathcal{R}_{\text{HK}}^o(P)\}.$$

Clearly, if $(R_1, R_2) \in \Pi(\mathcal{R}_{\text{HK}}^o(P))$ then there exists a 4-tuple $(S_1, T_1, S_2, T_2) \in \mathcal{R}_{\text{HK}}^o(P)$ such that $(S_i + T_i) = R_i$ for $i = 1, 2$, and vice versa. This is true for any distribution satisfying (67). Using the Fourier-Motzkin elimination method, a compact formula for the rate region $\Pi(\mathcal{R}_{\text{HK}}^o(P))$ was recently derived in Lemma 1 of [25], which when evaluated for the input distributions specified in Section II-C, results in an achievable rate region containing the rate region given in Lemma 2 (See Theorem 2 of [11]). Let us denote the rate region $\mathcal{R}_{\text{HK}}^o(P)$ by $\mathcal{R}_{\text{HK}}^G$, when P is same as the distributions specified in Section II-C. Using the technique in the proof of Lemma 2 (given in [11]) it then follows that $\mathcal{R}_{\text{HK}}^G = \mathcal{R}_{\text{HK}}^{G1} \cap \mathcal{R}_{\text{HK}}^{G2}$, where

$$\begin{aligned} \mathcal{R}_{\text{HK}}^{Gi} = \left\{ (S_i, T_i, T_j) : \right. & S_i \leq \log \det \left(I_{N_i} + \rho_{ii} H_{ii} K_{iu} H_{ii}^\dagger + \rho_{ji} H_{ji} K_{ju} H_{ji}^\dagger \right) - \tau_{ji}; \\ & T_i \leq \log \det \left(I_{N_i} + \rho_{ii} H_{ii} K_{iw} H_{ii}^\dagger + \rho_{ji} H_{ji} K_{ju} H_{ji}^\dagger \right) - \tau_{ji}; \\ & T_j \leq \log \det \left(I_{N_i} + \frac{\rho_{ji}}{M_j} H_{ji} H_{ji}^\dagger \right) - \tau_{ji}; \\ & (S_i + T_i) \leq \log \det \left(I_{N_i} + \frac{\rho_{ii}}{M_i} H_{ii} H_{ii}^\dagger + \rho_{ji} H_{ji} K_{ju} H_{ji}^\dagger \right) - \tau_{ji}; \\ & (S_i + T_j) \leq \log \det \left(I_{N_i} + \frac{\rho_{ji}}{M_j} H_{ji} H_{ji}^\dagger + \rho_{ii} H_{ii} K_{iu} H_{ii}^\dagger \right) - \tau_{ji}; \\ & (T_i + T_j) \leq \log \det \left(I_{N_i} + \frac{\rho_{ji}}{M_j} H_{ji} H_{ji}^\dagger + \rho_{ii} H_{ii} K_{iw} H_{ii}^\dagger \right) - \tau_{ji}; \\ & \left. (S_i + T_i + T_j) \leq \log \det \left(I_{N_i} + \frac{\rho_{ji}}{M_j} H_{ji} H_{ji}^\dagger + \frac{\rho_{ii}}{M_i} H_{ii} H_{ii}^\dagger \right) - \tau_{ji}; \right\} \end{aligned}$$

for $i \neq j \in \{1, 2\}$ and τ_{ji} 's for $1 \leq i \neq j \leq 2$ are constants independent of ρ or channel matrices. The GDoF region corresponding to the above achievable rate region can be defined as follows

$$\mathcal{G}_i(\bar{M}, \bar{\alpha}) = \left\{ (d_{1p}, d_{1c}, d_{2p}, d_{2c}) : d_{ip} = \lim_{\rho_{ii} \rightarrow \infty} \frac{S_i}{\log(\rho_{ii})}, d_{ic} = \lim_{\rho_{ii} \rightarrow \infty} \frac{T_i}{\log(\rho_{ii})} \text{ for } i = 1, 2, \right. \\ \left. \text{and } (S_1, T_1, S_2, T_2) \in \mathcal{R}_{\text{HK}}^{G_i} \right\}.$$

Using this definition, and following a similar approach as in Theorem 1, we get equation (35).

From the above analysis on one hand, we have the achievable rate region $\mathcal{R}_{\text{HK}}^G$ for the Gaussian IC, which is $\mathcal{R}_{\text{HK}}^o(P)$ evaluated for the distribution of Subsection II-C. on the other hand, we have $\mathcal{R}_a(\mathcal{H}, \bar{\alpha})$, which is a *subset* of the rate region obtained when $\Pi(\mathcal{R}_{\text{HK}}^o(P))$ is evaluated at the distribution in Subsection II-C. This two facts together imply that

$$\mathcal{R}_a(\mathcal{H}, \bar{\alpha}) \subseteq \Pi(\mathcal{R}_{\text{HK}}^G), \quad (68)$$

i.e., for any rate pair $(R_1, R_2) \in \mathcal{R}_a(\mathcal{H}, \bar{\alpha})$ there exists a 4-tuple $(S_1, T_1, S_2, T_2) \in \mathcal{R}_{\text{HK}}^G$ such that $(S_i + T_i) = R_i$ for $i = 1, 2$. In other words, $\mathcal{R}_a(\mathcal{H}, \bar{\alpha})$ is a subset of the 2-dimensional projection of the set $\mathcal{R}_{\text{HK}}^G$. Since $\mathcal{G}(\bar{M}, \bar{\alpha})$ and $\mathcal{D}_o(\bar{M}, \bar{\alpha})$ are the high SNR scaled versions of the rate regions $\mathcal{R}_{\text{HK}}^G$ and $\mathcal{R}_a(\mathcal{H}, \bar{\alpha})$, respectively the same is true for them. That is $\mathcal{D}_o(\bar{M}, \bar{\alpha})$ is a subset of the 2-dimensional projection of the set $\mathcal{G}(\bar{M}, \bar{\alpha})$ or for every $(d_1, d_2) \in \mathcal{D}_o(\bar{M}, \bar{\alpha})$, there exists a 4-tuple $(d_{1p}, d_{1c}, d_{2p}, d_{2c}) \in \mathcal{G}(\bar{M}, \bar{\alpha})$ such that $(d_{ip} + d_{ic}) = d_i$ for $i = 1, 2$.

REFERENCES

- [1] A. B. Carleial, "A case where interference does not reduce the capacity," *IEEE Trans. on Inform. Th.*, vol. 21, pp. 569–570, Sep, 1975.
- [2] R. Benzel, "The capacity region of a class of discrete additive degraded interference channels," *IEEE Trans. on Inform. Th.*, vol. 25, pp. 228–231, Mar, 1979.
- [3] H. Sato, "The capacity of Gaussian interference channel under strong interference," *IEEE Trans. on Inform. Th.*, vol. 27, pp. 786–788, Nov, 1981.
- [4] A. A. E. Gamal and M. H. M. Costa, "The capacity region of a class of deterministic interference channels," *IEEE Trans. on Inform. Th.*, vol. 28, pp. 343–346, Mar, 1982.
- [5] A. S. Motahari and A. K. Khandani, "Capacity bounds for the gaussian interference channel," *IEEE Trans. on Inform. Th.*, vol. 55, pp. 620–643, Feb, 2009.
- [6] X. Shang, G. Kramer, and B. Chen, "A new outer bound and the noisy-interference sum-rate capacity for Gaussian interference channels," *IEEE Trans. on Inform. Th.*, vol. 55, pp. 689–699, Feb, 2009.
- [7] V. S. Annapureddy and V. V. Veeravalli, "Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region," *IEEE Trans. on Inform. Th.*, vol. 55, pp. 3032–3050, Jun, 2009.
- [8] X. Shang, B. Chen, G. Kramer, and H. V. Poor, "Capacity regions and sum-rate capacities of vector gaussian interference channels," *IEEE Trans. on Inform. Th.*, vol. 56, pp. 5030–5044, Sep, 2010.

- [9] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. on Inform. Th.*, vol. 54, pp. 5534–5562, Dec, 2008.
- [10] E. Telatar and D. N. C. Tse, "Bounds on the capacity region of a class of interference channels," in *Proc. IEEE Int. Symp. on Inform. Th.*, Jun, 2007.
- [11] S. Karmakar and M. K. Varanasi, "Capacity of the MIMO interference channel to within a constant gap," Feb, 2011, preprint, available at <http://arxiv.org/abs/1102.0267>.
- [12] S. A. Jafar and M. J. Fakhreddin, "Degrees of freedom for the MIMO interference channel," *IEEE Trans. on Inform. Th.*, vol. 53, pp. 2637–2642, Jul, 2007.
- [13] T. S. Han and K. Kobayashi, "A new achievable region for the interference channel," *IEEE Trans. on Inform. Th.*, vol. 27, pp. 49–60, Jan, 1981.
- [14] P. A. Parker, D. W. Bliss, and V. Tarokh, "On the degrees-of-freedom of the MIMO interference channel," in *Proc. of Inform. Sciences and Systems*, Mar, 2008, pp. 62–67.
- [15] S. Jafar and S. Vishwanath, "Generalized degrees of freedom of the symmetric Gaussian K user interference channel," in *arXiv:cs.IT/0804.4489*, 2008.
- [16] B. Bandemer, A. E. Gamal, and G. Vasquez-Vilar, "On the sum capacity of the class of cyclically symmetric deterministic interference channels," in *Proc. IEEE Int. Symp. on Inform. Th., Seoul, South Korea*, Jun, 2009, pp. 2622–2626.
- [17] T. Gou and S. A. Jafar, "Capacity of a class of symmetric SIMO gaussian interference channels within $O(1)$," May, 2009, preprint, available at <http://arxiv.org/abs/0905.1745>.
- [18] C. Huang, V. R. Cadambe, and S. A. Jafar, "On the capacity and generalized degrees of freedom of the X channel," preprint, available at <http://arxiv.org/abs/0810.4741>.
- [19] G. Bresler, A. Parekh, and D. Tse, "The approximate capacity of the many-to-one and one-to-many gaussian interference channels," *IEEE Trans. on Inform. Th.*, vol. 56, pp. 4566–4592, Sep, 2010.
- [20] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the k-user interference channel," *IEEE Trans. on Inform. Th.*, vol. 54, pp. 3425–3441, Aug, 2008.
- [21] A. M. Tulino and S. Verdu, *Random Matrix Theory and Wireless Communications*. Now, 2004.
- [22] C. Huang, S. A. Jafar, S. S. (Shitz), and S. Vishwanath, "On degrees of freedom region of MIMO networks without CSIT," Sep. 2009, available Online: <http://arxiv.org/pdf/0909.4017>.
- [23] C. S. Vaze and M. K. Varanasi, "The degrees of freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT," Sept. 2009, available Online: <http://arxiv.org/abs/0909.5424>.
- [24] Y. Zhu and D. Guo, "The degrees of freedom of mimo interference channels without state information at transmitters," Aug, 2010, preprint, available at <http://arxiv.org/abs/1008.5196>.
- [25] H. Chong, M. Motani, H. Garg, and H. E. Gamal, "On the Han-Kobayashi region for the interference channel," *IEEE Trans. on Inform. Th.*, vol. 54, pp. 3188–3195, Jul, 2008.