Degrees of Freedom Regions of Two-User MIMO Z and Full Interference Channels: The Benefit of Reconfigurable Antennas

Lei Ke, *Student Member, IEEE*, and

Zhengdao Wang (Contact Author), *Senior Member, IEEE*

Abstract

We study the degrees of freedom (DoF) regions of two-user multiple-input multiple-output (MIMO) Z and full interference channels in this paper. We assume that the receivers always have perfect channel state information. We first derive the DoF region of Z interference channel with channel state information at transmitter (CSIT). For full interference channel without CSIT, the DoF region has been fully characterized recently and it is shown that the previously known outer bound is not achievable. In this work, we investigate the no-CSIT case further by assuming that the transmitter has the ability of antenna mode switching. We obtain the DoF region as a function of the number of available antenna modes and reveal the incremental gain in DoF that each extra antenna mode can bring. It is shown that in certain cases the reconfigurable antennas can bring extra DoF gains. In these cases, the DoF region is maximized when the number of modes is at least equal to the number of receive antennas at the corresponding receiver, in which case the previously outer bound is achieved. In all cases, we propose systematic constructions of the beamforming and nulling matrices for achieving the DoF region. The constructions bear an interesting space-frequency interpretation.

Index Terms

Degrees of freedom region, interference channel, multiple-input multiple-output, reconfigurable antenna, antenna mode switching

L. Ke and Z. Wang are with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: kelei@iastate.edu; zhengdao@iastate.edu).

I. INTRODUCTION

Characterizing the capacity region of interference channel has been a long open problem. Many researchers investigated this important problem, and the capacity regions of certain interference channels are known when the interference is strong, e.g. [1]–[3]. However, when the interference is not strong, the capacity region is still unknown. Recent progress reveals the capacity region for two-user interference channel to within one bit [4], and after that the sum capacity for very weak interference channel is settled [5]–[7]. Recently, a deterministic channel model has been proposed and used to explore the capacity of Gaussian interference network [8]–[10] such that the gap to capacity region can be bounded up to a constant value.

When it comes to multiple-input multiple-output (MIMO) networks, the capacity regions of certain MIMO interference channels are known [11], [12]. Instead of trying to characterize the capacity region completely, the degrees of freedom (DoF) region characterizes how capacity scales with transmit power as the signal-to-noise ratio goes to infinity.

It is well-known that in certain cases, the absence of channel state information at transmitter (CSIT) will not affect the DoF for MIMO networks, e.g., in the multiple access channel [13]. In other cases, CSIT does play an important role. For example, using interference alignment scheme, it is shown that the total DoF of a K-user MIMO interference channel is $MK/2$, where M is the number of antennas of each user [14]. The key idea is to pack interferences from multiple sources so as to reduce the dimensionality of signal space spanned by interference.

The DoF region of two-user MIMO interference channel with CSIT has been obtained in [15], where it is shown that zero forcing is enough to achieve the DoF region. However, it is a different story in two-user MIMO X channel, where each transmitter has a message to every receiver. In [16] it is shown that interference alignment is the key to achieving the DoF region of MIMO X network. The DoF region of two-user MIMO broadcast channel and interference channel without CSIT are considered in [17], where there is an uneven trade-off between the two users. Except for a special case, the DoF region for the interference channel is known and achievable. Similar, but more general result of isotropic fading channel can be found in [18]. The DoF regions of the K-user MIMO broadcast, interference and cognitive radio channels are derived in [19] for some cases. However, the special case in [17] remains unsolved.

When only one of the two transmitter-receiver pairs is subject to interference, the interference

channel is termed as *Z interference channel* (ZIC). To avoid confusion, we will call the channel where both pairs are subject to interference the *full interference channel* (FIC). The capacity region of MIMO Gaussian ZIC is established in [20] under very strong interference and aligned strong interference assumptions. In [21], the authors considered the capacity region of a single antenna ZIC without CSIT using deterministic approach.

Recently, it is shown in [22] that if the channel is staggered block fading, we can explore the channel correlation structure to do interference alignment, where the upper bound in the converse can be achieved in some special cases. For example, it is shown that for two-user MIMO staggered block fading FIC with 1 and 3 antennas at transmitters, 2 and 4 antennas at their corresponding receivers and without CSIT, the DoF pair $(1, 1.5)$ can be achieved. The idea was further clarified in [23], where a blind interference alignment scheme is also proposed for K-user multiple-input single-output (MISO) broadcast channel to achieve DoF outer bound when CSIT is absent. Also recently, it is shown in [24] that the previous outer bound is not tight when the channels are independent and identical distributed (i.i.d.) over time and isotropic over spatial domain. So by now the DoF region of two-user MIMO FIC is completely known for both the case with CSIT and the no CSIT case (receiver-side CSI, or CSIR, is always assumed available), provided that the channel is i.i.d. over time and isotropic over spatial domain. However, when the channel is not i.i.d. over time such as in the "staggered" fading channels [22], the DoF could be larger.

In this paper, we consider the ZIC channel with CSIT, and both ZIC and FIC without CSIT but with reconfigurable antennas. Specifically, we obtain the DoF regions for the cases of:

1) ZIC with CSIT. We show that zero forcing is sufficient for achieving the DoF region in this case (Theorem 1).

2) ZIC and FIC when transmitter one has the number K of antennas modes at least equal to N_1 (Theorems 2 and 3). Increasing K beyond N_1 does not bring more gains in DoF.

3) ZIC and FIC when $M_1 \leq K < N_1$, in which case each additional antenna mode brings an incremental gain on the DoF region (Theorem 4).

We present joint beamforming and nulling schemes to achieve the DoF region in all cases. When reconfigurable antennas are used, our proposed schemes have an interesting space-frequency coding explanation.

The rest of the paper is organized as follows. We first present the system model in Section II. Known results on the DoF region of two-user MIMO FIC are also briefly reviewed. The DoF region of ZIC with CSIT is discussed in Section III. The DoF regions of ZIC and FIC without CSIT when there are enough antenna modes are investigated in Section IV. When there are not enough modes, the DoF region is given in Section V. Finally, Section VI concludes this paper.

Notation: boldface uppercase (lowercase) letters denote matrices (vectors). $\mathbb{R}, \mathbb{Z}, \mathbb{C}$ are the real, integer and complex numbers sets. $\mathcal{CN}(0,1)$ denotes a circularly symmetric complex Gaussian (CSCG) distribution with zero mean and unit variance. We use $\boldsymbol{A} \otimes \boldsymbol{B}$ to denote the Kronecker product between $\boldsymbol{A}$ and $\boldsymbol{B}$. $\boldsymbol{0}$ and $\boldsymbol{1}$ denote all one and all zero matrices (vectors), respectively. $\boldsymbol{A}^T$ and $\boldsymbol{A}^\dagger$ denote the transpose and Hermitian of $\boldsymbol{A}$, respectively. We also use notation like $\boldsymbol{A}_{m \times n}$ to emphasize that $\boldsymbol{A}$ is of size $m \times n$. We use $\boldsymbol{I}_m$ to denote a size $m \times m$ identity matrix and $\boldsymbol{1}_m$ to denote an all-one column vector with length m. Denote $\boldsymbol{g}_n(a) := [1, a, a^2, \ldots, a^{n-1}]^T$. A size $n \times m$ Vandermonde matrix based on a set of element $\{a_1, a_2, \ldots, a_m\}$ is defined as $\boldsymbol{V}_n(a_1, a_2, \ldots, a_m) = [\boldsymbol{g}_n(a_1), \boldsymbol{g}_n(a_2), \ldots, \boldsymbol{g}_n(a_m)]$. We use $\mathcal{I}(\boldsymbol{x}; \boldsymbol{y})$ to denote the mutual information between $\boldsymbol{x}$ and $\boldsymbol{y}$. The differential entropy of a continuous random variable $\boldsymbol{x}$ is denoted as $\mathcal{H}(\boldsymbol{x})$.

II. SYSTEM MODEL AND KNOWN RESULTS

A. Channel Model

Consider a MIMO interference channel with two transmitters and two receivers, the number of transmit (receive) antennas at the ith transmitter (receiver) is denoted as M_i (N_i), $i \in \{1, 2\}$. The system is termed as an (M_1, N_1, M_2, N_2) system, which can be described as

$$\boldsymbol{y}_1(t) = \boldsymbol{H}_{11}(t)\boldsymbol{x}_1(t) + \boldsymbol{H}_{12}(t)\boldsymbol{x}_2(t) + \boldsymbol{z}_1(t) \tag{1}$$

$$\boldsymbol{y}_2(t) = \boldsymbol{H}_{21}(t)\boldsymbol{x}_1(t) + \boldsymbol{H}_{22}(t)\boldsymbol{x}_2(t) + \boldsymbol{z}_2(t) \tag{2}$$

where t is the time index, $\boldsymbol{y}_i(t) \in \mathbb{C}^{N_i}$, $\boldsymbol{z}_i(t) \in \mathbb{C}^{N_i}$ are the received signal and additive noise of receiver i, respectively. The entries of $\boldsymbol{z}_i(t)$ are independent and identically $\mathcal{CN}(0,1)$ distributed in both time and space. The channel between the ith transmitter and the jth receiver is denoted as $\boldsymbol{H}_{ji}(t) \in \mathbb{C}^{N_j \times M_i}$. We assume the probability of $\boldsymbol{H}_{ji}(t)$ belonging to any subset of $\mathbb{C}^{N_j \times M_i}$ that has zero Lebesgue measure is zero. For the two-user MIMO ZIC, $\boldsymbol{H}_{21}(t) = 0$. $\boldsymbol{x}_i(t) \in \mathbb{C}^{M_i}$

is the input signal at transmitter i and $\boldsymbol{x}_1(t)$ is independent of $\boldsymbol{x}_2(t)$. The transmitted signals satisfy the following power constraint:

$$\mathrm{E}(||\boldsymbol{x}_i(t)||^2) \leq P \quad i = 1, 2. \tag{3}$$

Denote the capacity region of the two-user MIMO system as $C(P)$, which contains all the rate pairs (R_1, R_2) such that the corresponding probability of error can approach zero as coding length increases. The DoF region is defined as follows [17]

$$\mathcal{D} := \left\{ (d_1, d_2) \in \mathbb{R}_2^+ : \exists (R_1(P), R_2(P)) \in C(P), \text{ such that } d_i = \lim_{P \to \infty} \frac{R_i(P)}{\log(P)}, \quad i = 1, 2 \right\}.$$

B. Reconfigurable antennas

Assume the CSI at receiver (CSIR) is always available. We would like to study the DoF regions of MIMO FIC and ZIC with or without CSIT under an additional assumption that one transmitter is equipped with reconfigurable antennas. The reconfigurable antennas are different from the conventional antennas as they can be switched to different pre-determined modes so that the channel fluctuation can be introduced artificially. Similar to [23], we use reconfigurable antennas to explore multiplexing gain other than diversity gain.

We assume that only one transmitter is equipped with reconfigurable antennas. We define one antenna *mode* as one possible configuration of a single transmit antenna such that by switching to a different mode, the channel between this transmit antenna and all receive antennas is changed. Different antenna modes can be realized via spatially separated physical antennas, or the same physical antenna excited with different polarizations, and so on. The benefit of antenna mode switching lies in the fact that channel variation can be artificially created, without the need to increase the number of RF chains. We let K denote the total number of antenna modes available at the transmitter with reconfigurable antennas (usually transmitter one).

We make the following assumption of the channel in this paper: the channel is block fading with coherent length of L symbols. Within each coherent block, the channels between all the transmitter modes and the receive antennas remain constant. The channels between the K modes of the reconfigurable transmitter and both receivers are isotropic, in the sense of [24]. From block to block, the channel changes independently.

When $K > M$, the transmitter has the freedom to use different modes at different slots. For a given antenna mode usage pattern over the length of a whole coherent block, the effective

channel for the whole block is not isotropic fading and not i.i.d. over the time slots within the block.

One may view our model approximately as a transition from an effective channel where all the links have exactly the same coherent time as in [24] to an effective channel where the links do not have the same coherent time [22]. However, there are two important distinctions between antenna mode switching and variation of channel coherence time: i) Antenna switching can be initiated at will at the transmitter, whereas channel coherence structure is in general not controllable. ii) The resulting equivalent channel from antenna mode switching is not "staggered" [22], so methods therein do not apply here.

C. Known Results on FIC

We first present some known results on DoF region of MIMO full interference channel which will be useful for developing our results.

The total degrees of freedom of two-user MIMO full interference channel with CSIT is developed in [15, Theorem 2], which leads to the following DoF regions:

$$d_i \leq \min(M_i, N_i), \quad i = 1, 2; \tag{4}$$

$$d_1 + d_2 \leq \min(\max(N_1, M_2), \max(M_1, N_2), N_1 + N_2, M_1 + M_2). \tag{5}$$

An outer bound of degrees of freedom region of two-user MIMO full interference channel without CSIT is as follows [18, Theorem 1]:

$$d_i \leq \min(M_i, N_i), \quad i = 1, 2; \tag{6}$$

$$d_1 + \frac{\min(N_1, N_2, M_2)}{\min(N_2, M_2)} d_2 \leq \min(M_1 + M_2, N_1); \tag{7}$$

$$\frac{\min(N_1, N_2, M_1)}{\min(N_1, M_1)} d_1 + d_2 \leq \min(M_1 + M_2, N_2). \tag{8}$$

Note that the same result is also given in [17], though in a less compact form.

It is shown in [17] that the outer bound given in (6)–(8) can be achieved by zero forcing or time sharing except for the case $M_1 < N_1 < \min(M_2, N_2)$, for which it was not known how to achieve

$$(d_1, d_2) = \left(M_1, \frac{\min(M_2, N_2)(N_1 - M_1)}{N_1} \right) \tag{9}$$

in general. The cases when $N_1 > N_2$ can be converted by switching the user indices. It is shown in [24] that when the channel is isotropic fading and i.i.d. over time, the outer bound given in (6)–(8) is not tight: if $N_1 \leq N_2$, the DoF region of FIC without CSIT can be given as follows:

$$d_i \leq \min(M_i, N_i), \quad i = 1, 2; \tag{10}$$

$$d_1 + \frac{\min(N_1, M_2) - \alpha}{\min(N_2, M_2) - \alpha}(d_2 - \alpha) \leq \min(M_1, N_1). \tag{11}$$

where $\alpha = \min(M_1 + M_2, N_1) - \min(M_1, N_1)$. In other words, (9) is not achievable when $M_1 < N_1 < \min(M_2, N_2)$, as (11) is reduced to

$$d_1 + \frac{M_1}{\min(M_2, N_2) - (N_1 - M_1)}d_2 \leq M_1 + \frac{M_1(N_1 - M_1)}{\min(M_2, N_2) - (N_1 - M_1)} \tag{12}$$

and the DoF pair $(d_1, d_2) = (M_1, N_1 - M_1)$ is on the boundary of the DoF region.

III. TWO-USER MIMO ZIC WITH CSIT

In this section, we prove the following theorem.

Theorem 1 (ZIC with CSIT): The DoF region of a two-user MIMO Z interference channel with CSIT is described by

$$d_i \leq \min(M_i, N_i), \quad i = 1, 2; \tag{13}$$

$$d_1 + d_2 \leq \min(\max(N_1, M_2), N_1 + N_2, M_1 + M_2). \tag{14}$$

Proof: We split the proof into the achievability and converse parts, as the following two lemmas. The theorem can be proved by showing the regions given by Lemma 1 and Lemma 2 are the same for all the cases. ∎

Lemma 1 (Achievability part of Theorem 1): The following region of two-user MIMO ZIC with CSIT is achievable:

$$d_i \leq \min(M_i, N_i), \quad i = 1, 2; \tag{15}$$

$$d_1 + d_2 \leq \min(N_1, M_1 + \min(N_2, M_2))1(M_2 < N_1)$$
$$+ \min(M_2, N_2 + \min(N_1, M_1))1(M_2 \geq N_1) \tag{16}$$

where $1(\cdot)$ is indicator function.

Proof: If $M_2 \geq N_1$ and assume transmitter 1 sends d_1 streams, transmitter 2 can send at most $M_2 - N_1$ streams along the null space of $\boldsymbol{H}_{12}$ without interfering receiver 1. Transmitter

2 can also send at most $N_1 - d_1$ streams along the row space of $\boldsymbol{H}_{12}$. Therefore user 2 can decode $\min((M_2 - N_1) + (N_1 - d_1), N_2)$ streams without interfering receiver 1. If $N_1 \geq M_2$ and assume transmitter 2 sends d_2 streams which interfere receiver 1, transmitter 1 can send $\min(N_1 - d_2, M_1)$ decodable streams to receiver 1. Combining these two cases, we have the achievable DoF region shown in this lemma. ∎

Lemma 2 (Conversepart of Theorem 1): The region given by (13) and (14) is a valid outer bound for the two-user MIMO ZIC with CSIT.

Proof: It is obvious that adding antennas at the receiver will not shrink the DoF region. Hence, we can add M_1 antennas to receiver 2 resulting an $(M_1, N_1, M_2, M_1 + N_2)$ MIMO FIC, and (14) follows from Corollary 1 in [15]. The outer bound of such a MIMO FIC is a valid outer bound of an (M_1, N_1, M_2, N_2) MIMO ZIC. Combining the trivial upper bound on point-to-point system, we have this lemma. ∎

Based on Lemma 1, zero forcing at receiver is sufficient to achieve the DoF region of ZIC when CSIT is available. The antenna mode switching ability is not needed in this case. However, we shall see later that such an ability is important for the case when CSIT is absent.

IV. Two-User MIMO ZIC and FIC without CSIT When Number of Modes $K \geq N_1$

In this section, we describe the DoF region of two-user ZIC and FIC without CSIT but with transmitter side reconfigurable antennas. We deal with the case that K, the number of antenna modes is at least equal to the N_1. The case $K < N_1$ will be dealt with in Section V.

Based on the antenna number configuration, the achievability scheme of ZIC and FIC without CSIT can be divided into two cases. In the first case, no reconfigurable antenna is needed to achieve an DoF outer bound — reconfigurable antennas are not helpful (Section IV-B). In the second case, the outer bound can be achieved with enough transmit side antenna modes (Section IV-C): reconfigurable antennas enlarges the DoF region. Our main results in this section are the following two theorems.

Theorem 2 (ZIC with Enough Reconfigurable Antenna Modes): The DoF region of two-user MIMO Z interference channel without CSIT is described by the following inequalities

$$d_i \leq \min(M_i, N_i), \quad i = 1, 2; \tag{17}$$

$$d_1 + \frac{\min(N_1, N_2, M_2)}{\min(N_2, M_2)} d_2 \leq \min(M_1 + M_2, N_1). \tag{18}$$

if either one of the following is true:

C1) $M_1 < N_1 < \min(M_2, N_2)$ and transmitter one can switch among N_1 antenna modes, or

C2) (M_1, N_1, M_2, N_2) do not satisfy the above condition. ∎

The DoF region in Theorem 2 is shown in Fig. 1.

Theorem 3 (FIC with Enough Reconfigurable Antenna Modes): The DoF region of two-user MIMO full interference channel without CSIT is described by the inequalities (6)–(8) if any one of the following is true:

C1) $M_1 < N_1 < \min(M_2, N_2)$ and transmitter one can switch among N_1 antenna modes, or

C2) $M_2 < N_2 < \min(M_1, N_1)$ and transmitter two can switch among N_2 antenna modes, or

C3) (M_1, N_1, M_2, N_2) are not one of the two above cases. ∎

A. Converse part

We first prove the converse part of the two theorems.

Lemma 3 (Converse part of Theorem 3): The outer bound of DoF region of two-user MIMO full interference channel given in (6)–(8) is still valid when either or both transmitters are using antenna mode switching.

Proof: The outer bound (7) has been derived based on the assumption that the rows of H_{12} and those of H_{22} are statistically equivalent [17], [18]. Similarly, the outer bound (8) has been derived based on the assumption that the rows of H_{11} and those of H_{21} are statistically equivalent. These assumptions are not affected by antenna mode switching at either or both transmitters. Hence, the DoF outer bound is still valid. ∎

Lemma 4 (Converse part of Theorem 2): The outer bound of degrees of freedom region of two-user MIMO Z interference channel without CSIT can be given as when transmitter one has the antenna mode switching ability

$$d_i \leq \min(M_i, N_i), \quad i = 1, 2; \tag{19}$$

$$d_1 + \frac{\min(N_1, N_2, M_2)}{\min(N_2, M_2)} d_2 \leq \min(M_1 + M_2, N_1). \tag{20}$$

Proof: This is the direct result of [18, Theorem 1] as in (6)–(8), by noticing that there is no interference from transmitter 1 to receiver 2 hence (8) is not longer needed. The antenna switching at transmitter one does not affect the upper bound, for the same reason stated in Lemma 3. ∎
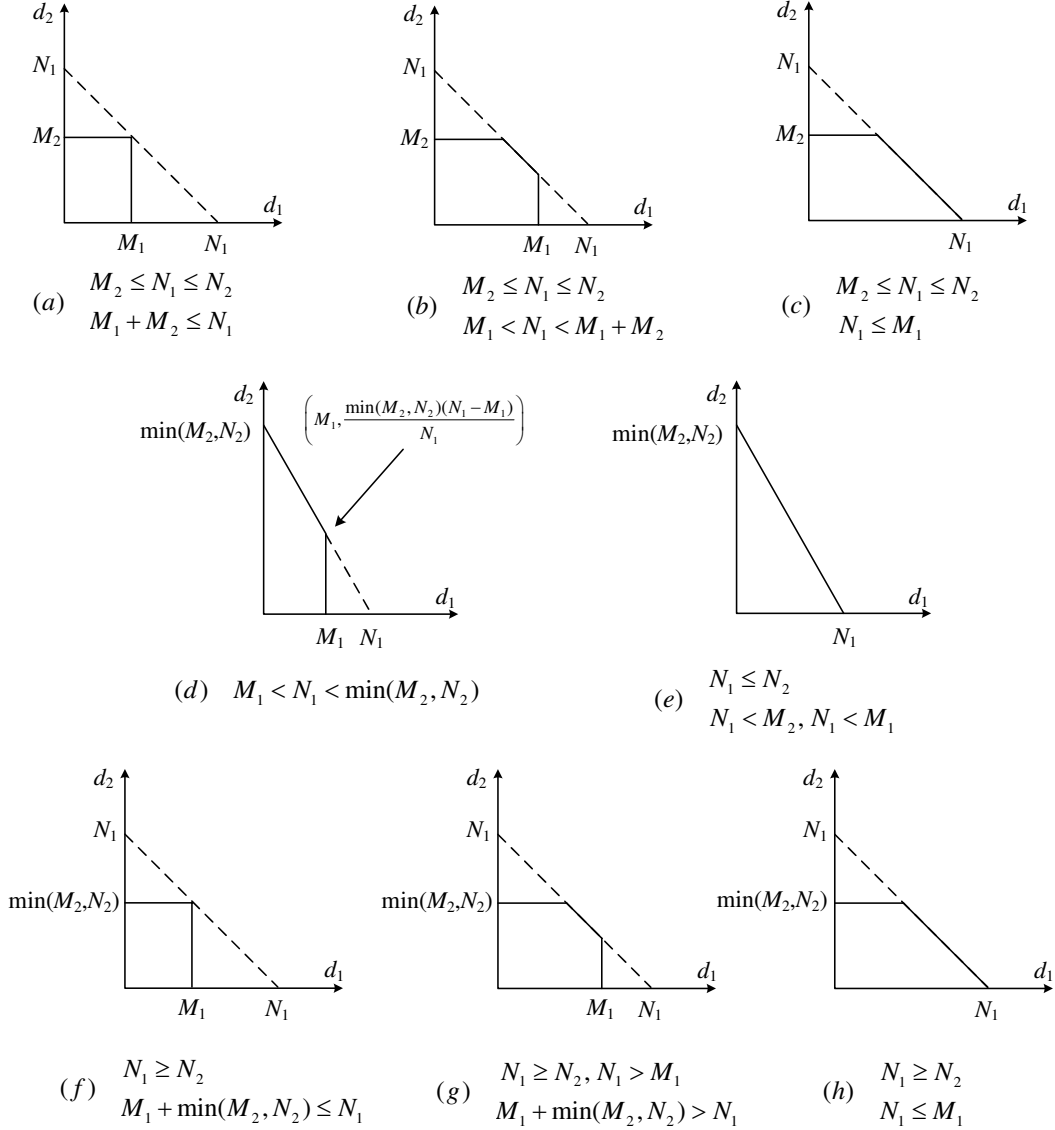
Fig. 1. DoF region of two-user MIMO ZIC without CSIT when number of antenna modes $K \geq N_1$. Figures (a)–(e) are for the case $N_1 \leq N_2$; Figures (f)–(h) are for the case $N_1 \geq N_2$.

B. Achievability: when antenna mode switching is not needed

In this section, we prove the achievability part for Case C2) of Theorem 2 and Case C3) of Theorem 3. Achievability for the remaining cases are left to Section IV-C.

Lemma 5: For the two-user MIMO Z interference channel without CSIT, when $N_1 \geq N_2$, (20) is achievable by zero forcing.

Proof: When $N_1 \geq N_2$, the corresponding outer regions are shown in Fig. 1 (f)–(h). Noticing that (20) is reduced to $d_1 + d_2 \leq \min(M_1 + M_2, N_1)$, zero forcing is sufficient to achieve the outer bound. ∎

Lemma 6: When CSIT is absent, the DoF outer region given by Lemma 4 of a two-user MIMO (M_1, N_1, M_2, N_2) ZIC is the same as that of an $(M_1, N_1, \min(M_2, N_2), \min(M_2, N_2))$ ZIC.

Proof: We give the proof case by case. It is trivial that when $M_2 \leq N_2$ reducing the number of antennas at receiver 2 to M_2 will not shrink the DoF region. When $M_2 > N_2$, we can further consider two sub-cases: $N_2 \geq N_1$ and $N_2 < N_1$.

1) When $M_2 > N_2 \geq N_1$, corresponding to Fig. 1 (d) and (e), the DoF bound (20) becomes $\frac{d_1}{N_1} + \frac{d_2}{N_2} \leq 1$. Hence the DoF outer region is the same as an (M_1, N_1, N_2, N_2) ZIC.

2) When $M_2 > N_2$ and $N_2 < N_1$, the DoF bound (20) becomes $d_1 + d_2 \leq \min(M_1 + M_2, N_1)$. Hence, if $M_1 \geq N_1 - N_2$, which implies $M_1 + \min(M_2, N_2) \geq N_1$, the DoF outer region is a pentagon or a tetragon; see Fig. 1 (g) and (h). Otherwise, it is a square, see Fig. 1 (f). One can show that the region is the same as that of an (M_1, N_1, N_2, N_2) ZIC.

Hence, the lemma holds. ∎

We also have the following lemma regarding the relationship between DoF regions of ZIC and FIC.

Lemma 7: When $N_1 \leq N_2$, the MIMO ZIC and FIC have the same DoF regions. Any encoding scheme that is DoF optimal for one channel is also DoF optimal for the other.

Proof: Any point in the FIC is also trivially achievable in the ZIC because user 2's channel is interference free. Conversely, any point achievable in the ZIC region, is also achievable in FIC. This is based on the fact that the channels are statistically equivalent at both receivers. If receiver 1 can decode user 1's message, then receiver 2, having at least as many antennas, must also be able to decode the same message. Receiver 2 can then subtract the decoded message, which renders the resulting channel the same as in the ZIC. ∎

Due to Lemma 7, we can translate all achievability schemes from FIC to ZIC and vice versa when $N_1 \leq N_2$. Therefore the achievability schemes in [17] for FIC when $N_1 \leq N_2$ and $M_1 \geq N_1$ can be used for ZIC. Therefore, the achievability part for Case C2) of Theorem 2 is complete.

For the FIC, the achievability for the case $N_1 \leq N_2$, except when $M_1 < N_1 < \min(M_2, N_2)$, is shown in [17]. When $N_1 \geq N_2$, we can swap the indices of the two users, so that except for the Cases C1) and C2) the achievability scheme is known for FIC.

C. Achievability: with antenna mode switching when $K \geq M_1 N_1$

In this subsection, we prove a weaker version of the achievability for Case C1) of Theorem 2 and Cases C1) and C2) of Theorem 3. Namely, we assume that the number of antenna modes available is $K \geq M_1 N_1$. The scheme is simpler in this case, and the achievability scheme for the case $K = N_1$ will be built upon this case.

Based on Lemma 6 and Lemma 7, we only consider the two-user MIMO ZIC with $M_1 < N_1 < M_2 = N_2$ to prove the Cases C1) for both theorems. Case C2) of Theorem 3 is the Case of C1) with user indices swapped. Therefore, we want to show that the following DoF pair is achievable for ZIC with $K_1 = M_1 N_1$ modes:

$$(d_1, d_2) = \left(M_1, \frac{M_2(N_1 - M_1)}{N_1} \right). \tag{21}$$

We first notice that this point cannot be achieved by zero forcing over one time instant. This is because using zero forcing if transmitter 1 sends M_1 streams, transmitter 2 can only send $N_1 - M_1$ streams without interfering receiver 1. If transmitter 2 sends more streams, the desired signal and interference are not separable at receiver 1 as transmitter 2 does not know channel state information so it cannot send streams along the null space of $\boldsymbol{H}_{12}$. A simple example is the $(1, 2, 3, 3)$ case, where the outer bound gives us $(d_1, d_2) = (1, 1.5)$, which is not achievable via zero forcing over one time slot. We make the assumption that the channel $\boldsymbol{H}_{12}$ stays the same for at least N_1 time slots. It is sufficient to show that $(M_1 N_1, M_2(N_1 - M_1))$ streams can be achieved in N_1 time slots.

We first develop the beamforming and nulling design by assuming that there are $N_1 M_1$ antenna modes available at transmitter 1 such that it can use different antenna modes in different slots to create channel variation. We will further show that the resultant beamforming and nulling design still work even if there are only N_1 modes available.

Here and after, we use tilde notation to indicate the time expansion signals, where the number of slots of time expansion signals shall be clear within the context. The time expansion channel

between transmitter 1 and receiver 1 in N_1 time slots is

$$\tilde{\boldsymbol{H}}_{11} = \begin{bmatrix} \boldsymbol{H}_{11}(1) & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_{11}(2) & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{H}_{11}(N_1) \end{bmatrix}_{N_1^2 \times N_1 M_1}$$

and the channel between transmitter 2 and receiver 1 is

$$\tilde{\boldsymbol{H}}_{12} = \boldsymbol{I}_{N_1} \otimes \boldsymbol{H}_{12}(1) \tag{22}$$

as transmitter 2 does not create channel variation. We will use precoding at transmitter 2 only and nulling at receiver 1 only. Let $\tilde{\boldsymbol{P}}$ be the transmit beamforming matrix at transmitter 2 and $\tilde{\boldsymbol{Q}}$ be the nulling matrix at receiver 1. We propose to use the following structures for them

$$\tilde{\boldsymbol{P}}_{M_2 N_1 \times M_2(N_1 - M_1)} = \boldsymbol{P}_{N_1 \times (N_1 - M_1)} \otimes \boldsymbol{I}_{M_2} \tag{23}$$

$$\tilde{\boldsymbol{Q}}_{M_1 N_1 \times N_1^2} = \boldsymbol{Q}_{M_1 \times N_1} \otimes \boldsymbol{I}_{N_1}. \tag{24}$$

The received signal at receiver 1 can be written as

$$\tilde{\boldsymbol{y}}_1 = \tilde{\boldsymbol{H}}_{11} \tilde{\boldsymbol{x}}_1 + \tilde{\boldsymbol{H}}_{12} \tilde{\boldsymbol{P}} \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{z}}_1 \tag{25}$$

where $\tilde{\boldsymbol{x}}_1$ is a length $M_1 N_1$ vector, and $\tilde{\boldsymbol{x}}_2$ is a length $M_2(N_1 - M_1)$ vector. After applying nulling matrix $\tilde{\boldsymbol{Q}}$, we have

$$\tilde{\boldsymbol{Q}} \tilde{\boldsymbol{y}}_1 = \underbrace{\tilde{\boldsymbol{Q}} \tilde{\boldsymbol{H}}_{11}}_{\tilde{\boldsymbol{A}}} \tilde{\boldsymbol{x}}_1 + \underbrace{\tilde{\boldsymbol{Q}} \tilde{\boldsymbol{H}}_{12} \tilde{\boldsymbol{P}}}_{\tilde{\boldsymbol{B}}} \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{Q}} \tilde{\boldsymbol{z}}_1. \tag{26}$$

To achieve the degrees of freedom $(M_1 N_1, M_2(N_1 - M_1))$ for both users, it is sufficient to design our $\tilde{\boldsymbol{P}}$ and $\tilde{\boldsymbol{Q}}$ to satisfy the following conditions simultaneously

1) $\text{rank}(\tilde{\boldsymbol{A}}) = M_1 N_1$,
2) $\text{rank}(\tilde{\boldsymbol{P}}) = M_2(N_1 - M_1)$,
3) $\tilde{\boldsymbol{B}} = \boldsymbol{0}$.

The second condition can be easily satisfied. Because $\text{rank}(\tilde{\boldsymbol{P}}) = \text{rank}(\boldsymbol{P})\text{rank}(\boldsymbol{I}_{M_2})$, we only need to design $\boldsymbol{P}$ such that $\text{rank}(\boldsymbol{P}) = N_1 - M_1$. As to the third condition, notice that

$$\tilde{\boldsymbol{B}} = (\boldsymbol{Q} \otimes \boldsymbol{I}_{N_1})(\boldsymbol{I}_{N_1} \otimes \boldsymbol{H}_{12}(1))(\boldsymbol{P} \otimes \boldsymbol{I}_{M_2})$$

$$= (\boldsymbol{Q} \boldsymbol{I}_{N_1} \boldsymbol{P}) \otimes (\boldsymbol{I}_{N_1} \boldsymbol{H}_{12}(1) \boldsymbol{I}_{M_2})$$

$$= (\boldsymbol{Q} \boldsymbol{P}) \otimes \boldsymbol{H}_{12}(1).$$

It is therefore sufficient (and also necessary) to have $\boldsymbol{QP} = \boldsymbol{0}$. Then the key is to find a $\boldsymbol{Q}$ such that the equivalent channel of user 1 after nulling

$$\tilde{\boldsymbol{A}} = (\boldsymbol{Q} \otimes \boldsymbol{I}_{N_1})\tilde{\boldsymbol{H}}_{11} \tag{27}$$

has full rank $M_1 N_1$ with probability 1. The matrix $\tilde{\boldsymbol{A}}$ is of size $M_1 N_1 \times M_1 N_1$ and has the following structure

$$\tilde{\boldsymbol{A}} = \begin{bmatrix} q_{11}\boldsymbol{H}_{11}(1) & q_{12}\boldsymbol{H}_{11}(2) & \cdots & q_{1N_1}\boldsymbol{H}_{11}(N_1) \\ q_{21}\boldsymbol{H}_{11}(1) & q_{12}\boldsymbol{H}_{11}(2) & \cdots & q_{2N_1}\boldsymbol{H}_{11}(N_1) \\ \vdots & \vdots & \ddots & \vdots \\ q_{M_1 1}\boldsymbol{H}_{11}(1) & q_{M_1 2}\boldsymbol{H}_{11}(2) & \cdots & q_{M_1 N_1}\boldsymbol{H}_{11}(N_1) \end{bmatrix}.$$

To show that $\tilde{\boldsymbol{A}}$ has full rank, we need the following lemma, which is known before, and a proof of it can be found in e.g., [25].

Lemma 8: [25, Lemma 2] Consider an analytic function $h(\boldsymbol{x})$ of several variables $\boldsymbol{x} = [x_1, \ldots, x_n]^T \in \mathbb{C}^n$. If h is nontrivial in the sense that there exists $\boldsymbol{x}_0 \in \mathbb{C}^n$ such that $h(\boldsymbol{x}_0) \neq 0$, then the zero set of $f(\boldsymbol{x})$ $Z := \{\boldsymbol{x} \in \mathbb{C}^n | h(\boldsymbol{x}) = 0\}$ is of measure (Lebesgue measure in $\mathbb{C}^n$) zero. ∎

Because the determinant of $\tilde{\boldsymbol{A}}$ is an analytic polynomial function of elements of $\boldsymbol{H}_{11}(t), t = 1, \ldots, N_1$, we only need to find a specific pair of $\boldsymbol{Q}$ and $\boldsymbol{H}_{11}(t), t = 1, \ldots, N_1$, such that $\tilde{\boldsymbol{A}}$ is full rank. We propose the following:

$$\boldsymbol{Q} = [\boldsymbol{V}_{N_1}(1, \omega_{N_1}, \ldots, \omega_{N_1}^{M_1-1})]^T, \tag{28}$$

where $\omega_{N_1} := \exp(-j2\pi/N_1)$.

Let $\omega := \exp(-j2\pi/N_1^2)$. Take the realizations of $\boldsymbol{H}_{11}(t), t = 1, \ldots, N_1$, as

$$\boldsymbol{H}_{11}(t) = \boldsymbol{V}_{N_1}(\omega^{t-1}, \omega^{N_1+t-1}, \ldots, \omega^{(M_1-1)N_1+t-1}). \tag{29}$$

It can be verified that for such choices of $\boldsymbol{Q}$ and $\boldsymbol{H}_{11}(t)$, $\tilde{\boldsymbol{A}}$ is a Vandermonde matrix:

$$\tilde{\boldsymbol{A}} = \boldsymbol{V}_{M_1 N_1}(1, \omega^{N_1}, \ldots, \omega^{(M_1-1)N_1}, \omega^1, \omega^{N_1+1}, \ldots, \omega^{(M_1-1)N_1+1},$$

$$\ldots, \omega^{N_1-1}, \omega^{2N_1-1}, \ldots, \omega^{M_1 N_1-1}),$$

hence of full rank. We also notice that $\tilde{A}$ is a leading principal minor of a permuted fast Fourier transform (FFT) matrix with size $N_1^2 \times N_1^2$. The permutation is as follows: Index the columns of an FFT matrix $0, 1, \ldots, N_1^2 - 1$, and then permute them in an order shown below:

$$(0, N_1, 2N_1, \ldots, (M_1 - 1)N_1), (1, N_1 + 1, 2N_1 + 1, \ldots, (M_1 - 1)N_1 + 1), \ldots$$

Based on Lemma 8, if we choose the nulling matrix using Q as specified in (28), $\tilde{A}$ has full rank almost surely. One choice of the corresponding P matrix with respect to (28) is the following

$$P = \mathcal{V}_{N_1}(\omega_{N_1}^{-M_1}, \omega_{N_1}^{-(M_1+1)}, \ldots, \omega_{N_1}^{-(N_1-1)}), \tag{30}$$

which is orthogonal to Q. This completes the achievability part under conditions in Case C1) of Theorem 2 and Cases C1) and C2) of Theorem 3, but with $K \geq M_1 N_1$.

D. Achievability: with antenna mode switching when $K = N_1$

Assuming there are N_1 modes available at transmitter 1 and denote these channel vectors between receive antennas of user 1 and the ith mode as $h_i, 1 \leq i \leq N_1$ and let $\hat{H}_{N_1 \times N_1} = [h_1, h_2, \ldots, h_{N_1}]$. We choose the antenna modes to be switched cyclically:

$$H_{11}(1) = [h_1, h_2, \ldots, h_{M_1}], \tag{31}$$

$$H_{11}(2) = [h_2, h_3, \ldots, h_{M_1+1}], \tag{32}$$

$$\vdots$$

$$H_{11}(N_1) = [h_{N_1}, h_1, \ldots, h_{M_1-1}]. \tag{33}$$

We want to show that under this switching pattern, the equivalent channel $\tilde{A}$ in (27) between transmitter one and receiver one after nulling, is still full rank. To show this, indexing the columns of $\tilde{A}$ in (27) as $0, 1, \ldots, N_1 M_1 - 1$, we then permute and group the columns of $\tilde{A}$ in the following way:

$$(0, M_1, 2M_1, \ldots, (M_1 - 1)N_1), (1, M_1 + 1, 2M_1 + 1, \ldots, (M_1 - 1)N_1 + 1), \ldots$$

Denote the permutation result as $\tilde{A}'$ and it can be expressed as

$$\tilde{A}' = \begin{bmatrix} \hat{H} & \hat{H} & \cdots & \hat{H} \\ G\hat{H} & \omega_{N_1}^{-1}G\hat{H} & \cdots & \omega_{N_1}^{-M_1}G\hat{H} \\ \vdots & \vdots & \ddots & \vdots \\ G^{M_1}\hat{H} & (\omega_{N_1}^{-1}G)^{M_1}\hat{H} & \cdots & (\omega_{N_1}^{-M_1}G)^{M_1}\hat{H} \end{bmatrix},$$

where G is a size $N_1 \times N_1$ diagonal matrix and can be expressed as

$$G = \mathrm{diag}(1, \omega_{N_1}, \omega_{N_1}^2, \ldots, \omega_{N_1}^{N_1-1}). \tag{34}$$

Notice that $\tilde{A}' = R(I_{M_1} \otimes \hat{H})$, where

$$R = \begin{bmatrix} I_{N_1} & I_{N_1} & \cdots & I_{N_1} \\ G & \omega_{N_1}^{-1}G & \cdots & \omega_{N_1}^{-M_1}G \\ \vdots & \vdots & \ddots & \vdots \\ G^{M_1} & (\omega_{N_1}^{-1}G)^{M_1} & \cdots & (\omega_{N_1}^{-M_1}G)^{M_1} \end{bmatrix}.$$

Recall $\omega_{N_1} = \exp(-j2\pi/N_1)$. To show $\tilde{A}$ is full rank, it is necessary to show R is full rank as $I_{M_1} \otimes \hat{H}$ is full rank with probability 1. It can be verified that via row and column permutations R can be changed to a block diagonal matrix with the ith block being

$$\mathcal{V}_{M_1}(\omega_{N_1}^i, \omega_{N_1}^{i-1}, \cdots, \omega_{N_1}^{i-M_1+1}), \tag{35}$$

which is full rank due to Vandermonde structure. Hence R is full rank. It follows that A is full rank with probability 1. This completes the achievability part under conditions in Case C1) of Theorem 2 and Cases C1) and C2) of Theorem 3 for $K = N_1$.

E. Discussion

1) Frequency domain interpretation: We note that the matrix $[Q^\dagger, \ P]$ is an inverse FFT (IFFT) matrix in our construction (23), (24), (28) and (30). This observation yields an interesting frequency domain interpretation of our construction. The signal of user 2 is transmitted over frequencies corresponding to the last $N_1 - M_1$ columns of an IFFT matrix, whereas the first user's signal is transmitted on all frequencies. Due to the antenna mode switching at transmitter 1, the channel between transmitter 1 and receiver 1 is now time-varying and we manually introduce frequency spread. User 1's signal is spread from one frequency bin to all the frequencies while

user 2's signal remains in the last $N_1 - M_1$ frequency bins. Therefore the signal in the first M_1 bins is interference free, which can be used to decode user 1's message. The nulling matrix applied at receiver 1 has a projection explanation as well. Left multiplying the left and right hand sides of (26) with $\tilde{\boldsymbol{Q}}^\dagger$ yields

$$\tilde{\boldsymbol{Q}}^\dagger \tilde{\boldsymbol{Q}} \tilde{\boldsymbol{y}}_1 = \tilde{\boldsymbol{Q}}^\dagger \tilde{\boldsymbol{Q}} \tilde{\boldsymbol{H}}_{11} \tilde{\boldsymbol{x}}_1 + \tilde{\boldsymbol{Q}}^\dagger \boldsymbol{Q} \tilde{\boldsymbol{z}}_1$$
$$= ((\boldsymbol{Q}^\dagger \boldsymbol{Q}) \otimes \boldsymbol{I}_{N_1})(\tilde{\boldsymbol{H}}_{11} \tilde{\boldsymbol{x}}_1 + \tilde{\boldsymbol{z}}_1),$$

where $\boldsymbol{Q}^\dagger \boldsymbol{Q}$ is the frequency domain projection matrix. We can see that the signal of user 1 is projected from N_1 frequencies to the first M_1 frequencies.

2) The Loss of DoF due to lack of CSIT: In two-user MIMO Z interference channel without CSIT, losing CSIT will not shrink degrees of freedom region if $M_2 \leq N_1$ or $M_2 > N_1 \geq N_2 + M_1$. For all the other cases, the degrees of freedom region is strictly smaller when comparing with the CSIT case.

This observation can be verified case by case. Notice that it is already shown in [17, Theorem 2] that when $M_2 \leq N_1 \leq N_2$ absence of CSIT does not reduce DoF region in two-user MIMO FIC. Because MIMO FIC and ZIC has the same DoF region when $N_1 \leq N_2$. We only need to consider the sub cases when $N_1 > N_2$, corresponding to (f)–(h) in Fig. 1.

1) If $M_2 < N_1$ and $N_1 > N_2$, the total DoF of MIMO ZIC is upper bounded by N_1 due to (14), so the DoF region remains the same if CSIT is absent.

2) If $M_2 > N_1 > N_2$, the DoF region of MIMO ZIC without CSIT is a square only when $M_1 + N_2 \leq N_1$, same as that of ZIC with CSIT. Otherwise, the maximum total DoF of ZIC with CSIT is $\min(M_2, N_1 + N_2, \min(M_1, N_1) + N_2)$, strictly larger than N_1 which is the maximum total DoF when CSIT is absent, hence loss of CSIT reduces the DoF region.

3) Alternative construction when $N_1/M_1 = \beta \in \mathbb{Z}$: When $N_1/M_1 = \beta \in \mathbb{Z}$, instead of using the $\boldsymbol{Q}$ given in (28) we can use the following $\boldsymbol{Q}_{M_1 \times N_1} = \boldsymbol{I}_{M_1} \otimes \boldsymbol{1}_\beta^T$. We need to show that this $\boldsymbol{Q}$ matrix will lead to a full rank $\tilde{\boldsymbol{A}}$. This can be achieved by choosing $\tilde{\boldsymbol{H}}_{11}$ such that it can be

decomposed as $\tilde{\boldsymbol{H}}_{11} = \boldsymbol{I}_{M_1} \otimes \tilde{\boldsymbol{H}}'_{11}$, where

$$
\tilde{\boldsymbol{H}}'_{11} = \begin{bmatrix} \boldsymbol{H}_{11}(1) & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_{11}(2) & \dots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{H}_{11}(\beta) \end{bmatrix}_{N_1\beta \times N_1} .
$$

For this $\tilde{\boldsymbol{H}}_{11}$

$$
\tilde{\boldsymbol{A}} = (\boldsymbol{I}_{M_1} \otimes \boldsymbol{1}_\beta^T \otimes \boldsymbol{I}_{N_1})(\boldsymbol{I}_{M_1} \otimes \tilde{\boldsymbol{H}}'_{11})
$$
$$
= \boldsymbol{I}_{M_1} \otimes ((\boldsymbol{1}_\beta^T \otimes \boldsymbol{I}_{N_1})\tilde{\boldsymbol{H}}'_{11}),
$$

which has full rank. For this choice of $\boldsymbol{Q}$, we only use $\beta M_1 = N_1$ antenna modes in N_1 time slots.

Therefore, for the two-user MIMO ZIC and FIC when $M_1 < N_1 < \min(M_2, N_2)$ and $N_1/M_1 = \beta \in \mathbb{Z}$, β fold time expansion is enough to achieve the DoF region. We remark that this can be viewed as the generalization of the case we discussed in Section IV-C for $N_1 = \beta$ and $M_1 = 1$. In fact $\boldsymbol{1}_\beta^T$ is the nulling matrix $\boldsymbol{Q}$ given in (28) when $N_1 = \beta, M_1 = 1$.

4) Successive Decoding in ZIC: For the two-user MIMO FIC when $M_1 < N_1 < \min(M_2, N_2)$ and CSIT is absent, we need block decoding at both receivers in general, which introduces decoding delay. Successive interference cancellation decoder can be used at receiver 2 to reduce decoding delay. Taking the case $N_1/M_1 = \beta \in \mathbb{Z}$ as an example, we can use β fold time expansion and choose $\boldsymbol{Q} = \boldsymbol{1}_\beta^T$. The corresponding $\boldsymbol{P}$ matrix is not necessary to be the last $\beta - 1$ columns of an $\beta \times \beta$ FFT matrix. The following $\boldsymbol{P}$ matrix still satisfies the design constraint

$$
\boldsymbol{P}_{\beta \times (\beta-1)} = \begin{bmatrix} \boldsymbol{I}_{\beta-1} \\ \boldsymbol{1}_{\beta-1}^T \end{bmatrix}. \tag{36}
$$

Here, $\boldsymbol{P}$ has a nice structure. Every stream of user 2 can be decoded immediately as they are interference free. For other cases where M_1 cannot divide N_1, we can still find a $\boldsymbol{Q}$, $\boldsymbol{P}$ pair through numerical simulation such that the upper diagonal parts of $\boldsymbol{P}$ are all zeros and contain small number of nonzero entries. Such a beamforming matrix can guarantee the immediate decoding of user 2's signal the interference only comes from the streams already decoded .

V. Two-User MIMO ZIC and FIC without CSIT When Number of Modes $K < N_1$

In this section, we will present our result for the $K < N_1$ case. The main result of this section is the following theorem.

Theorem 4: When $M_1 < N_1 < \min(M_2, N_2)$ and the antennas of transmitter 1 can be switched among K antenna modes, where $K < N_1$, the DoF region of two-user MIMO ZIC and FIC without CSIT is given by the following inequalities

$$d_i \le \min(M_i, N_i), \quad i = 1, 2; \tag{37}$$

$$d_1 + \frac{K}{\min(M_2, N_2) - (N_1 - K)} d_2 \le M_1 + \frac{K(N_1 - M_1) + (\min(M_2, N_2) - N_1)(K - M_1)}{\min(M_2, N_2) - (N_1 - K)} \tag{38}$$

The DoF region of FIC for $M_2 < N_2 < \min(M_1, N_1)$ can be obtained by switching the two user indices. ∎

The method of proof is heavily based on that in [24], to which the reader is referred for several lemmas that will be used and their proofs. Some notation that is used in this section are the following. We use tilde notation to denote the time expanded signal *over L time slots* and $t \in [1, L]$ is the index of the slot within one block. In general, by default, for a vector $\boldsymbol{x}$, $\tilde{\boldsymbol{x}} = \text{vec}(\boldsymbol{x}(1), \boldsymbol{x}(2), \cdots, \boldsymbol{x}(L))$ and for a matrix $\boldsymbol{V}$, $\tilde{\boldsymbol{V}} = \text{diag}(\boldsymbol{V}(1), \boldsymbol{V}(2), \cdots, \boldsymbol{V}(L))$. In addition, for a time expanded vector $\tilde{\boldsymbol{x}}$, we use $\tilde{\boldsymbol{x}}^n$ or $\{\tilde{\boldsymbol{x}}\}^n$ to denote a sequence of n successive blocks of $\tilde{\boldsymbol{x}}$: $\tilde{\boldsymbol{x}}^n = \text{vec}(\boldsymbol{x}(1), \boldsymbol{x}(2), \ldots, \boldsymbol{x}(nL))$. Furthermore, $\boldsymbol{x}(t)^n$ is the sequence of $\boldsymbol{x}(t)$ which contains all the vector $\boldsymbol{x}$ of the tth slot of all n blocks: $\boldsymbol{x}(t)^n = \text{vec}(\boldsymbol{x}(t), \boldsymbol{x}(t+L), \ldots, \boldsymbol{x}(t+(n-1)L))$. Similar notation is defined for matrices as well. We use $\boldsymbol{H}$ denotes $(\boldsymbol{H}_{11}, \boldsymbol{H}_{12}, \boldsymbol{H}_{21}, \boldsymbol{H}_{22})$, hence $\tilde{\boldsymbol{H}}^n$ denotes all the channel matrices over n blocks. In addition, for a random vector $\boldsymbol{x}$, $\boldsymbol{x}_G$ is a corresponding CSCG vector that has the same covariance matrix as $\boldsymbol{x}$.

A. The Converse Part

We prove the converse part of Theorem 4 in the following. Recall that for $M_1 < N_1 < \min(M_2, N_2)$, the proof is equivalent for both FIC and ZIC. We will only show the proof for ZIC. To make the proof self-contained, we will go through some similar steps as in [24], but avoiding details.

The converse is developed based on blocking for every L slots. In each block, the channel $\boldsymbol{H}_{12}, \boldsymbol{H}_{22}$ stay the same with the decomposition $\boldsymbol{H}_{12} = \boldsymbol{W}_{12}\boldsymbol{\Lambda}_{12}\boldsymbol{V}_{12}^{\dagger}$ and $\boldsymbol{H}_{22} = \boldsymbol{W}_{22}\boldsymbol{\Lambda}_{22}\boldsymbol{V}_{22}^{\dagger}$, whereas $\boldsymbol{H}_{11}$ is time-varying among L slots due to antenna mode switching at transmitter 1. Transmitter 1 has K modes with $K < N_1$ and it can adopt arbitrary switching pattern. Let $\underline{\boldsymbol{H}}_{11}$ be an $N_1 \times N_1$ full rank random matrix such that $\underline{\boldsymbol{H}}_{11} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_{N_1}]$ and $\boldsymbol{h}_i, 1 \leq i \leq K$ is the random vector channel between the ith antenna mode and receive antennas of user 1. We introduce the fictitious vectors $\{\boldsymbol{h}_i, K + 1 \leq i \leq N_1\}$ to simplify the proof. We assume $\underline{\boldsymbol{H}}_{11}$ is isotropic fading and i.i.d. over blocks of length L each, where L naturally satisfy $L \geq \lceil K/M_1 \rceil$. We denote the decomposition of $\underline{\boldsymbol{H}}_{11}$ as $\underline{\tilde{\boldsymbol{W}}}_{11}\underline{\tilde{\boldsymbol{\Lambda}}}_{11}\underline{\tilde{\boldsymbol{V}}}_{11}^{\dagger}$.

Furthermore, let $\boldsymbol{E}(t)$ of size $N_1 \times M_1$ denote the antenna mode selection matrix for time t. Let $\boldsymbol{e}_m, 1 \leq m \leq N_1$ be the mth column of $\boldsymbol{I}_{N_1}$. Let $i(t)$ denote the mode index selected by antenna i at time t. Then the ith column of $\boldsymbol{E}(t)$ is $\boldsymbol{e}_{i(t)}$. We have $\boldsymbol{H}_{11}(t) = \underline{\boldsymbol{H}}_{11}\boldsymbol{E}(t)$.

At receiver 1, from Fano's inequality, we have

$$nLR_1 - \delta_{nL} \leq \mathcal{I}(\tilde{\boldsymbol{y}}_1^n; \tilde{\boldsymbol{x}}_1^n | \tilde{\boldsymbol{H}}^n). \tag{39}$$

where $\delta_{nL} \to 0$ as $n \to \infty$. Denote

$$\tilde{\boldsymbol{r}} = \tilde{\boldsymbol{H}}_{11}\tilde{\boldsymbol{x}}_{1G} + \tilde{\boldsymbol{H}}_{12}^{\dagger}\tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{z}}_1 \tag{40}$$

$$\tilde{\boldsymbol{r}}_1 = \tilde{\boldsymbol{W}}_{12}^{\dagger}\tilde{\boldsymbol{H}}_{11}\tilde{\boldsymbol{x}}_{1G} + \tilde{\boldsymbol{V}}_{12}^{\dagger}\tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_1 \tag{41}$$

where $\tilde{\boldsymbol{n}}_1 = \tilde{\boldsymbol{W}}_{12}^{\dagger}\tilde{\boldsymbol{z}}_1$. Using [24, Theorem 3], which says that Gaussian input can reduce the mutual information by at most an $o(\log(P))$ quantity, and two uses of chain rule we have

$$nLR_1 - n\,o(\log(P))$$

$$\leq \mathcal{I}(\tilde{\boldsymbol{r}}^n; \boldsymbol{x}_{1G}^n | \tilde{\boldsymbol{H}}^n) \tag{42}$$

$$= \mathcal{I}\left(\tilde{\boldsymbol{r}}^n; \boldsymbol{x}_{1G}^n | \tilde{\boldsymbol{x}}_2^n, \tilde{\boldsymbol{H}}^n\right) + \mathcal{I}(\tilde{\boldsymbol{r}}^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n) - \mathcal{I}\left(\{\tilde{\boldsymbol{H}}_{12}^{\dagger}\tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{z}}_1\}^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right). \tag{43}$$

Using [24, Lemma 2], we have

$$\mathcal{I}\left(\{\tilde{\boldsymbol{H}}_{12}^{\dagger}\tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{z}}_1\}^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right) = \mathcal{I}\left(\{\tilde{\boldsymbol{W}}_{12}\tilde{\boldsymbol{\Lambda}}_{12}\tilde{\boldsymbol{V}}_{12}^{\dagger}\tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{z}}_1\}^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right) \tag{44}$$

$$= \mathcal{I}\left(\{\tilde{\boldsymbol{\Lambda}}_{12}\tilde{\boldsymbol{V}}_{12}^{\dagger}\tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_1\}^n; \tilde{\boldsymbol{x}}_2^n \tilde{\boldsymbol{H}}^n\right) \tag{45}$$

$$\geq \mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{12}^{\dagger}\tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_1\}^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right) - n\,o(\log(P)), \tag{46}$$

and

$$\mathcal{I}(\tilde{\boldsymbol{r}}^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n) = \mathcal{I}\left(\{\tilde{\boldsymbol{W}}_{12}^\dagger \tilde{\boldsymbol{H}}_{11} \tilde{\boldsymbol{x}}_{1G} + \tilde{\boldsymbol{\Lambda}}_{12} \tilde{\boldsymbol{V}}_{12}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_1\}^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right) \tag{47}$$

$$\leq \mathcal{I}\left(\tilde{\boldsymbol{r}}_1^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right) + n\, o(\log(P)). \tag{48}$$

Hence R_1 can be further bounded as

$$nLR_1 - n\, o(\log(P))$$
$$\leq \mathcal{I}\left(\tilde{\boldsymbol{r}}^n; \tilde{\boldsymbol{x}}_{1G}^n | \tilde{\boldsymbol{x}}_2^n, \tilde{\boldsymbol{H}}^n\right) + \mathcal{I}(\tilde{\boldsymbol{r}}_1^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n) - \mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{12}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_1\}^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right). \tag{49}$$

As to receiver 2, using Fano's inequality and [24, Lemma 2], we have

$$nLR_2 - \delta_{nL} \leq \mathcal{I}(\tilde{\boldsymbol{y}}_2^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n) \tag{50}$$

$$= \mathcal{I}\left(\{\tilde{\boldsymbol{W}}_{22} \tilde{\boldsymbol{\Lambda}}_{22} \tilde{\boldsymbol{V}}_{22}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{z}}_2\}^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right) \tag{51}$$

$$\leq \mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{22}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_2\}^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right) + n\, o(\log(P)), \tag{52}$$

where $\tilde{\boldsymbol{n}}_2 = \tilde{\boldsymbol{W}}_{22}^\dagger \tilde{\boldsymbol{z}}_2$. Hence

$$nLR_2 - n\, o(\log(P)) \leq \mathcal{I}\left(\tilde{\boldsymbol{r}}_1^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right) - \mathcal{I}\left(\tilde{\boldsymbol{r}}_1^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right) + \mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{22}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_2\}^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{r}}_1^n, \tilde{\boldsymbol{H}}^n\right). \tag{53}$$

Notice that by using Gaussian input, the following inequalities hold

$$\mathcal{I}\left(\tilde{\boldsymbol{r}}_1; \tilde{\boldsymbol{x}}_{1G}^n | \tilde{\boldsymbol{x}}_2^n, \tilde{\boldsymbol{H}}^n\right) \leq \mathrm{E} \log\left(\det(\boldsymbol{I}_{LN_1} + \frac{P}{M_1} \tilde{\boldsymbol{H}}_{11} \tilde{\boldsymbol{H}}_{11}^\dagger)\right) \tag{54}$$

$$= nLM_1 \log(P) + nLo(\log(P)), \tag{55}$$

$$\mathcal{I}\left(\tilde{\boldsymbol{r}}_1^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right) \leq n\mathrm{E} \log\left(\frac{\det(\boldsymbol{I}_{LN_1} + \frac{P}{M_2} \tilde{\boldsymbol{W}}_{12} \tilde{\boldsymbol{W}}_{12}^\dagger + \frac{P}{M_1} \tilde{\boldsymbol{H}}_{11} \tilde{\boldsymbol{H}}_{11}^\dagger)}{\det(\boldsymbol{I}_{LN_1} + \frac{P}{M_1} \tilde{\boldsymbol{H}}_{11} \tilde{\boldsymbol{H}}_{11}^\dagger)}\right) \tag{56}$$

$$= nL(N_1 - M_1) \log(P) + nLo(\log(P)). \tag{57}$$

Then let $n \to \infty$, multiply (53) with some positive scalar μ, add it with (49) and use (55), (57), we have the following inequality

$$nL[R_1 + uR_2 - o(\log(P))] \leq nLM_1 \log(P) + \mu nL(N_1 - M_1) \log(P) + \eta, \tag{58}$$

where μ is to be determined and

$$\eta = \mu\mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{22}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_2\}^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right) - \mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{12}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_1\}^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right) + (1-\mu)\mathcal{I}(\tilde{\boldsymbol{r}}_1^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n). \tag{59}$$

Divide (58) by $nL\log(P)$ and let $P \to \infty$, we have the following inequality on the DoF of two users

$$d_1 + \mu d_2 \leq M_1 + \mu(N_1 - M_1) + \lambda, \tag{60}$$

where

$$\lambda = \frac{1}{nL} \lim_{P \to \infty} \frac{\eta}{\log(P)}.$$

Recall that $\tilde{\boldsymbol{r}}_1 = \tilde{\boldsymbol{W}}_{12}^\dagger \tilde{\boldsymbol{H}}_{11} \tilde{\boldsymbol{x}}_{1G} + \tilde{\boldsymbol{V}}_{12}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_1$ and $\boldsymbol{H}_{11}(t) = \underline{\boldsymbol{H}}_{11} \boldsymbol{E}(t)$. We define

$$\tilde{\boldsymbol{r}}_2 = \underline{\tilde{\boldsymbol{\Lambda}}}_{11}^{-1} \underline{\tilde{\boldsymbol{W}}}_{11}^\dagger \tilde{\boldsymbol{W}}_{12} \tilde{\boldsymbol{V}}_{12}^\dagger \tilde{\boldsymbol{x}}_2 + \underline{\tilde{\boldsymbol{V}}}_{11}^\dagger \tilde{\boldsymbol{E}} \tilde{\boldsymbol{x}}_{1G} + \underline{\tilde{\boldsymbol{\Lambda}}}_{11}^{-1} \underline{\tilde{\boldsymbol{W}}}_{11}^\dagger \tilde{\boldsymbol{W}}_{12} \tilde{\boldsymbol{n}}_1 \tag{61}$$

$$\tilde{\boldsymbol{r}}_3 = \underline{\tilde{\boldsymbol{W}}}_{11}^\dagger \tilde{\boldsymbol{W}}_{12} \tilde{\boldsymbol{V}}_{12}^\dagger \tilde{\boldsymbol{x}}_2 + \underline{\tilde{\boldsymbol{V}}}_{11}^\dagger \tilde{\boldsymbol{E}} \tilde{\boldsymbol{x}}_{1G} + \underline{\tilde{\boldsymbol{\Lambda}}}_{11}^{-1} \underline{\tilde{\boldsymbol{W}}}_{11}^\dagger \tilde{\boldsymbol{W}}_{12} \tilde{\boldsymbol{n}}_1 \tag{62}$$

$$\tilde{\boldsymbol{r}}_4 = \underline{\tilde{\boldsymbol{V}}}_{11} \underline{\tilde{\boldsymbol{W}}}_{11}^\dagger \tilde{\boldsymbol{W}}_{12} \tilde{\boldsymbol{V}}_{12}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{E}} \tilde{\boldsymbol{x}}_{1G} + \underline{\tilde{\boldsymbol{V}}}_{11} \underline{\tilde{\boldsymbol{\Lambda}}}_{11}^{-1} \underline{\tilde{\boldsymbol{W}}}_{11}^\dagger \tilde{\boldsymbol{W}}_{12} \tilde{\boldsymbol{n}}_1 \tag{63}$$

$$\tilde{\boldsymbol{r}}_5 = \underline{\tilde{\boldsymbol{V}}}_{11} \underline{\tilde{\boldsymbol{W}}}_{11}^\dagger \tilde{\boldsymbol{W}}_{12} \tilde{\boldsymbol{V}}_{12}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{E}} \tilde{\boldsymbol{x}}_{1G} + \tilde{\boldsymbol{n}}_1 \tag{64}$$

$$\tilde{\boldsymbol{r}}_6 = \tilde{\boldsymbol{V}}_{12}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{E}} \tilde{\boldsymbol{x}}_{1G} + \tilde{\boldsymbol{n}}_1 \tag{65}$$

We have

$$\mathcal{I}(\tilde{\boldsymbol{r}}_1^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n) = \mathcal{I}(\tilde{\boldsymbol{r}}_2^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n) \tag{66}$$

$$= \mathcal{I}(\tilde{\boldsymbol{r}}_3^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n) + o(\log(P)) \tag{67}$$

$$= \mathcal{I}(\tilde{\boldsymbol{r}}_4^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n) + o(\log(P)) \tag{68}$$

$$= \mathcal{I}(\tilde{\boldsymbol{r}}_5^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n) + o(\log(P)) \tag{69}$$

$$= \mathcal{I}(\tilde{\boldsymbol{r}}_6^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n) + o(\log(P)), \tag{70}$$

where (67) due to [24, Lemma 2]; (66) and (68) hold as $\underline{\tilde{\boldsymbol{W}}}_{11} \underline{\tilde{\boldsymbol{\Lambda}}}_{11}$ and $\underline{\tilde{\boldsymbol{V}}}_{11}$ are full rank square matrices. (69) holds as changing noise variance will not change the DoF. (70) is true because $\underline{\tilde{\boldsymbol{V}}}_{11} \underline{\tilde{\boldsymbol{W}}}_{11}^\dagger \tilde{\boldsymbol{W}}_{12} \tilde{\boldsymbol{V}}_{12}^\dagger$ has the same distribution as $\tilde{\boldsymbol{V}}_{12}^\dagger$ and $\underline{\tilde{\boldsymbol{V}}}_{11} \underline{\tilde{\boldsymbol{W}}}_{11}^\dagger \tilde{\boldsymbol{W}}_{12}$ is independent of $\tilde{\boldsymbol{V}}_{12}^\dagger$. To find the DoF order of $\mathcal{I}(\tilde{\boldsymbol{r}}_6^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n)$, we first notice that for each slot t in one block, $\boldsymbol{V}_{12}^\dagger$ can be divided into three parts: $\boldsymbol{V}_{12,a}^\dagger(t)$, $\boldsymbol{V}_{12,b}^\dagger(t)$ and $\boldsymbol{V}_{12,c}^\dagger$.

1) $\boldsymbol{V}_{12,a}^\dagger(t)$ is of size $M_1 \times M_2$ and consists of M_1 non-zero rows of $\boldsymbol{E}(t)\boldsymbol{V}_{12}^\dagger$.

2) $\boldsymbol{V}_{12,c}^\dagger$ is of size $(N_1 - K) \times M_2$ and is the same for all $1 \leq t \leq L$. It consists of $N_1 - K$ rows of $\boldsymbol{V}_{12}^\dagger$ that do not appear in any $\boldsymbol{V}_{12,a}(t)^\dagger, 1 \leq t \leq L$.

3) $\boldsymbol{V}_{12,b}^{\dagger}(t)$ is of size $(K - M_1) \times M_2$ and consists of $K - M_1$ rows of $\boldsymbol{V}_{12}^{\dagger}$ that neither in $\boldsymbol{E}(t)\boldsymbol{V}_{12,a}(t)^{\dagger}$ nor in $\boldsymbol{V}_{12,c}^{\dagger}$.

Example: Assume $N_1 = 5$, $M_1 = 2$, $L = 6$, $K = 4$ and $\boldsymbol{V}_{12} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{N_1}]$ where $\boldsymbol{v}_i$'s are $M_2 \times 1$ vectors. Assume $\boldsymbol{E}(t)$ is the following

$$\boldsymbol{E}(1) = [\boldsymbol{e}_1, \boldsymbol{e}_2], \quad \boldsymbol{E}(2) = [\boldsymbol{e}_1, \boldsymbol{e}_3], \quad \boldsymbol{E}(3) = [\boldsymbol{e}_1, \boldsymbol{e}_4],$$

$$\boldsymbol{E}(4) = [\boldsymbol{e}_1, \boldsymbol{e}_2], \quad \boldsymbol{E}(5) = [\boldsymbol{e}_2, \boldsymbol{e}_4], \quad \boldsymbol{E}(6) = [\boldsymbol{e}_2, \boldsymbol{e}_3].$$

We have

$$\boldsymbol{V}_{12,a}^{\dagger}(1) = [\boldsymbol{v}_1, \boldsymbol{v}_2]^{\dagger}, \quad \boldsymbol{V}_{12,a}^{\dagger}(2) = [\boldsymbol{v}_1, \boldsymbol{v}_3]^{\dagger}, \quad \boldsymbol{V}_{12,a}^{\dagger}(3) = [\boldsymbol{v}_1, \boldsymbol{v}_4]^{\dagger},$$

$$\boldsymbol{V}_{12,a}^{\dagger}(4) = [\boldsymbol{v}_1, \boldsymbol{v}_2]^{\dagger}, \quad \boldsymbol{V}_{12,a}^{\dagger}(5) = [\boldsymbol{v}_2, \boldsymbol{v}_4]^{\dagger}, \quad \boldsymbol{V}_{12,a}^{\dagger}(6) = [\boldsymbol{v}_2, \boldsymbol{v}_3]^{\dagger},$$

$$\boldsymbol{V}_{12,b}^{\dagger}(1) = [\boldsymbol{v}_3, \boldsymbol{v}_4]^{\dagger}, \quad \boldsymbol{V}_{12,b}^{\dagger}(2) = [\boldsymbol{v}_2, \boldsymbol{v}_4]^{\dagger}, \quad \boldsymbol{V}_{12,b}^{\dagger}(3) = [\boldsymbol{v}_2, \boldsymbol{v}_3]^{\dagger},$$

$$\boldsymbol{V}_{12,b}^{\dagger}(4) = [\boldsymbol{v}_3, \boldsymbol{v}_4]^{\dagger}, \quad \boldsymbol{V}_{12,b}^{\dagger}(5) = [\boldsymbol{v}_1, \boldsymbol{v}_3]^{\dagger}, \quad \boldsymbol{V}_{12,b}^{\dagger}(6) = [\boldsymbol{v}_1, \boldsymbol{v}_4]^{\dagger},$$

and $\boldsymbol{V}_{12,c}^{\dagger} = \boldsymbol{v}_5^{\dagger}$. Note that $\boldsymbol{V}_{12,c}^{\dagger}$ remains the same in one block of L slots. ∎

Suppose receiver 1 receives $\boldsymbol{r}_6$ as in (65) and wants to decode the message of $\boldsymbol{x}_2$ that goes through an equivalent channel $\boldsymbol{V}_{12}^{\dagger}$. Then $\boldsymbol{V}_{12,a}^{\dagger}(t)$ are the directions of interference from transmitter at time t, $\boldsymbol{V}_{12,b}^{\dagger}(t)$ are those directions that are temporarily interference-free at time t, and $\boldsymbol{V}_{12,c}^{\dagger}$ are the directions which are interference free for a whole block. The associated noises of the those directions are similarly defined as $\boldsymbol{n}_{2,a}(t), \boldsymbol{n}_{2,b}(t)$ and $\boldsymbol{n}_{2,c}(t)$.

To bound the DoF of $\mathcal{I}(\tilde{\boldsymbol{r}}_6^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n)$ of (71), we define

$$\boldsymbol{V}_{12,ab}^{\dagger}(t) = \begin{bmatrix} \boldsymbol{V}_{12,a}^{\dagger}(t) \\ \boldsymbol{V}_{12,b}^{\dagger}(t) \end{bmatrix}, \quad \boldsymbol{n}_{1,ab}(t) = \begin{bmatrix} \boldsymbol{n}_{1,a}(t) \\ \boldsymbol{n}_{1,b}(t) \end{bmatrix}, \tag{71}$$

$$\boldsymbol{V}_{12,bc}^{\dagger}(t) = \begin{bmatrix} \boldsymbol{V}_{12,b}^{\dagger}(t) \\ \boldsymbol{V}_{12,c}^{\dagger} \end{bmatrix}, \quad \boldsymbol{n}_{1,bc}(t) = \begin{bmatrix} \boldsymbol{n}_{1,b}(t) \\ \boldsymbol{n}_{1,c}(t) \end{bmatrix}, \tag{72}$$

and adopt the following notation for simplicity

$$\boldsymbol{y}_a(t) = \boldsymbol{V}_{12,a}^{\dagger}(t)\boldsymbol{x}_2(t) + \boldsymbol{x}_{1G}(t) + \boldsymbol{n}_{1,a}(t) \tag{73}$$

$$\boldsymbol{y}_b(t) = \boldsymbol{V}_{12,b}^{\dagger}(t)\boldsymbol{x}_2(t) + \boldsymbol{n}_{1,b}(t) \tag{74}$$

$$\boldsymbol{y}_c(t) = \boldsymbol{V}_{12,c}^{\dagger}\boldsymbol{x}_2(t) + \boldsymbol{n}_{1,c}(t) \tag{75}$$

$$\boldsymbol{y}_{bc}(t) = \boldsymbol{V}_{12,bc}^{\dagger}(t)\boldsymbol{x}_2(t) + \boldsymbol{n}_{1,bc}(t) \tag{76}$$

In addition, $\boldsymbol{y}_a(t)^n, \boldsymbol{y}_b(t)^n, \boldsymbol{y}_c(t)^n, \boldsymbol{y}_{bc}(t)^n$ are sequences of corresponding vectors of the tth slot over n blocks. The collection of $\boldsymbol{y}_a(1)^n, \boldsymbol{y}_a(2)^n, \ldots \boldsymbol{y}_a(t)^n$ is denoted as $\{\boldsymbol{y}_a^{(1:t)}\}^n$. We also define $\{\boldsymbol{y}_b^{(1:t)}\}^n$, $\{\boldsymbol{y}_c^{(1:t)}\}^n$ and $\{\boldsymbol{y}_{bc}^{(1:t)}\}^n$ similarly. Using the chain rule, we have

$$
\begin{aligned}
\mathcal{I}(\tilde{\boldsymbol{r}}_6^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n) = {} & \mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{12,c}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_{1,c}\}^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right) \\
& + \mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{12,b}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_{1,b}\}^n; \tilde{\boldsymbol{x}}_2^n | \{\tilde{\boldsymbol{V}}_{12,c}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_{1,c}\}^n, \tilde{\boldsymbol{H}}^n\right) \\
& + \mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{12,a}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{x}}_{1G} + \tilde{\boldsymbol{n}}_{1,a}\}^n; \tilde{\boldsymbol{x}}_2^n | \{\tilde{\boldsymbol{V}}_{12,bc}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_{1,bc}\}^n, \tilde{\boldsymbol{H}}^n\right)
\end{aligned} \tag{77}
$$

Now checking the second term in (77), we notice that

$$
\mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{12,b}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_{1,b}\}^n; \tilde{\boldsymbol{x}}_2^n | \{\tilde{\boldsymbol{V}}_{12,c}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_{1,c}\}^n, \tilde{\boldsymbol{H}}^n\right)
$$

$$
= \sum_{t=1}^{L} \mathcal{I}\left(\boldsymbol{y}_b(t)^n; \tilde{\boldsymbol{x}}_2^n | \{\boldsymbol{y}_b^{(1:t-1)}\}^n, \tilde{\boldsymbol{y}}_c^n, \tilde{\boldsymbol{H}}^n\right) \tag{78}
$$

$$
= \sum_{t=1}^{L} \mathcal{H}\left(\boldsymbol{y}_b(t)^n | \{\boldsymbol{y}_b^{(1:t-1)}\}^n, \tilde{\boldsymbol{y}}_c^n, \tilde{\boldsymbol{H}}^n\right) - \mathcal{H}\left(\boldsymbol{y}_b(t)^n | \tilde{\boldsymbol{x}}_2^n, \{\boldsymbol{y}_b^{(1:t-1)}\}^n, \tilde{\boldsymbol{y}}_c^n, \tilde{\boldsymbol{H}}^n\right) \tag{79}
$$

$$
= \sum_{t=1}^{L} \mathcal{H}\left(\boldsymbol{y}_b(t)^n | \{\boldsymbol{y}_b^{(1:t-1)}\}^n, \tilde{\boldsymbol{y}}_c^n, \tilde{\boldsymbol{H}}^n\right) - \mathcal{H}\left(\boldsymbol{y}_b(t)^n | \boldsymbol{x}_2(t)^n, \boldsymbol{y}_b(t)^n, \tilde{\boldsymbol{H}}(t)^n\right) \tag{80}
$$

$$
\leq \sum_{t=1}^{L} \mathcal{H}\left(\boldsymbol{y}_b(t)^n | \boldsymbol{y}_c(t)^n, \boldsymbol{H}(t)^n\right) - \mathcal{H}\left(\boldsymbol{y}_b(t)^n | \boldsymbol{x}_2(t)^n, \boldsymbol{y}_c(t)^n, \boldsymbol{H}(t)^n\right) \tag{81}
$$

$$
= \sum_{t=1}^{L} \mathcal{I}\left(\boldsymbol{y}_b(t)^n; \boldsymbol{x}_2(t)^n | \boldsymbol{y}_c(t)^n, \boldsymbol{H}(t)^n\right) \tag{82}
$$

$$
= \sum_{t=1}^{L} \left[\mathcal{I}\left(\boldsymbol{y}_b(t)^n, \boldsymbol{y}_c(t)^n; \boldsymbol{x}_2(t)^n | \boldsymbol{H}(t)^n\right) - \mathcal{I}\left(\boldsymbol{y}_c(t)^n; \boldsymbol{x}_2(t)^n | \boldsymbol{H}(t)^n\right) \right] \tag{83}
$$

$$
\leq \sum_{t=1}^{L} \left(\frac{N_1 - M_1}{N_1 - K} - 1\right) \mathcal{I}\left(\boldsymbol{y}_c(t)^n; \boldsymbol{x}_2(t)^n | \boldsymbol{H}(t)^n\right) \tag{84}
$$

$$
\leq nL(K - M_1) \log(P) + o(\log(P)) \tag{85}
$$

where:

- (78) and (83) follow by chain rule.
- (79) and (82) are expressing mutual information via entropy.
- (80) holds as the second term is the entropy of noise when conditioning on $\boldsymbol{x}_2(t)^n$.
- (81) is based on the fact that conditioning reduces entropy.

- (84) follows by [24, Lemma 3].
- (85) holds due to the fact that the DoF of an $(N_1 - K) \times M_2$ point-to-point MIMO channel is at most $\min(N_1 - K, M_2) = N_1 - K$.

The third term in (77) can be bounded in a similar fashion. We have

$$\mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{12,a}^{\dagger}\tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{x}}_{1G} + \tilde{\boldsymbol{n}}_{1,a}\}^n; \tilde{\boldsymbol{x}}_2^n | \{\tilde{\boldsymbol{V}}_{12,bc}^{\dagger}\tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_{1,bc}\}^n, \tilde{\boldsymbol{H}}^n\right)$$

$$= \sum_{t=1}^{L} \mathcal{I}\left(\boldsymbol{y}_a(t)^n; \tilde{\boldsymbol{x}}_2^n | \{\boldsymbol{y}_a^{(1:t-1)}\}^n, \tilde{\boldsymbol{y}}_{bc}^n, \tilde{\boldsymbol{H}}^n\right) \tag{86}$$

$$= \sum_{t=1}^{L} \mathcal{H}\left(\boldsymbol{y}_a(t)^n | \{\boldsymbol{y}_a^{(1:t-1)}\}^n, \tilde{\boldsymbol{y}}_{bc}^n, \tilde{\boldsymbol{H}}^n\right) - \mathcal{H}\left(\boldsymbol{y}_a(t)^n | \tilde{\boldsymbol{x}}_2^n, \{\boldsymbol{y}_a^{(1:t-1)}\}^n, \tilde{\boldsymbol{y}}_{bc}^n, \tilde{\boldsymbol{H}}^n\right) \tag{87}$$

$$= \sum_{t=1}^{L} \mathcal{H}\left(\boldsymbol{y}_a(t)^n | \{\boldsymbol{y}_a^{(1:t-1)}\}^n, \tilde{\boldsymbol{y}}_{bc}^n, \tilde{\boldsymbol{H}}^n\right) - \mathcal{H}\left(\boldsymbol{y}_a(t)^n | \boldsymbol{x}_2(t)^n, \boldsymbol{y}_{bc}(t)^n, \tilde{\boldsymbol{H}}^n\right) \tag{88}$$

$$\leq \sum_{t=1}^{L} \mathcal{H}\left(\boldsymbol{y}_a(t)^n | \boldsymbol{y}_{bc}(t)^n, \boldsymbol{H}(t)^n\right) - \mathcal{H}\left(\boldsymbol{y}_a(t)^n | \boldsymbol{x}_2(t)^n, \boldsymbol{y}_{bc}(t)^n, \boldsymbol{H}(t)^n\right) \tag{89}$$

$$= \sum_{t=1}^{L} \mathcal{I}\left(\boldsymbol{y}_a(t)^n; \boldsymbol{x}_2(t)^n | \boldsymbol{y}_{bc}(t)^n, \boldsymbol{H}(t)^n\right) \tag{90}$$

$$\leq n \sum_{t=1}^{L} \mathcal{I}\left(\boldsymbol{y}_{aG}(t); \boldsymbol{x}_{2G}(t) | \boldsymbol{y}_{bcG}(t), \boldsymbol{H}(t)\right) \tag{91}$$

$$\leq nL \log\left(\det\left(\frac{P}{M_1}\boldsymbol{I}_{M_1} + \boldsymbol{I}_{M_1} + \frac{P}{M_2}\boldsymbol{I}_{M_1}\right)\det\left(\frac{P}{M_2}\boldsymbol{I}_{(N_1-M_1)} + \boldsymbol{I}_{(N_1-M_1)}\right)\right)$$

$$- nL \log\left(\det\left(\frac{P}{M_2}\boldsymbol{I}_{(N_1-M_1)} + \boldsymbol{I}_{(N_1-M_1)}\right)\det\left(\frac{P}{M_1}\boldsymbol{I}_{M_1} + \boldsymbol{I}_{M_1}\right)\right) \tag{92}$$

$$= o(\log(P)) \tag{93}$$

where:

- (86) follows by chain rule.
- (87) and (90) are expressing mutual information via entropy.
- (88) holds as the second term is the entropy of noise when conditioning on $\boldsymbol{x}_2(t)^n$.
- (89) is based on the fact that conditioning reduces entropy.
- (91) and (92) follows by [24, Lemma 3], where the covariance matrix of $\boldsymbol{x}_{1G}(t) + \boldsymbol{n}_{1,a}(t)$ and $\boldsymbol{n}_{1,bc}(t)$ are $\frac{P}{M_1}\boldsymbol{I}_{M_1} + \boldsymbol{I}_{M_1}$ and $\boldsymbol{I}_{(N_1-M_1)}$, respectively. In addition, the optimal input of $\boldsymbol{x}_2(t)$ is CSCG with covariance matrix $\frac{P}{M_2}\boldsymbol{I}_{(N_1-M_1)}$.

Substitute (85) and (93) in to (77), we have

$$\mathcal{I}(\tilde{\boldsymbol{r}}_1^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n) \leq \mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{12,c}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_{1,c}\}^n; \tilde{\boldsymbol{x}}_2^n | \tilde{\boldsymbol{H}}^n\right) + nL(K - M_1)\log(P) + o(\log(P)) \quad (94)$$

Now we go back to (59). Notice that if we choose $\boldsymbol{D} = [\boldsymbol{0}_{N_1 \times (\min(M_2, N_2) - N_1)}, \boldsymbol{I}_{N_1}]$, $(\boldsymbol{D}\boldsymbol{V}_{22}^\dagger, \boldsymbol{D}\boldsymbol{n}_2)$ has the same distribution as $(\boldsymbol{V}_{12}^\dagger, \boldsymbol{n}_1)$ as both $\boldsymbol{V}_{22}$ and $\boldsymbol{V}_{12}$ are uniformly distributed and $\boldsymbol{V}_{22}$ has no fewer columns than $\boldsymbol{V}_{12}$. (Please refer to [24, Sec. IV-C2] for more details). We have the following Markov chain:

$$\tilde{\boldsymbol{x}}_2 \text{ --- } \tilde{\boldsymbol{V}}_{22}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_2 \text{ --- } \tilde{\boldsymbol{V}}_{12}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_1 \text{ --- } \tilde{\boldsymbol{V}}_{12,c}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_{1,c}. \quad (95)$$

Denote $J = \min(M_2, N_2) - (N_1 - K)$. Let $\boldsymbol{V}_{22,a}$ contain the first J rows of $\boldsymbol{V}_{22}$, and $\boldsymbol{n}_{2,a}$ contain the first J elements of $\boldsymbol{n}_2$. We can bound η as

$$\begin{aligned}
\eta &\leq \mu \mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{22}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_2\}^n; \tilde{\boldsymbol{x}}_2^n | \{\tilde{\boldsymbol{V}}_{12,c}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_{1,c}\}^n, \tilde{\boldsymbol{H}}^n\right) \\
&\quad - \mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{12}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_1\}^n; \tilde{\boldsymbol{x}}_2^n | \{\tilde{\boldsymbol{V}}_{12,c}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_{1,c}\}^n, \tilde{\boldsymbol{H}}^n\right) \\
&\quad + (1 - \mu)nL(K - M_1)\log(P) + o(\log(P)) \quad (96) \\
&= \mu \mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{22,a}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_2\}^n; \tilde{\boldsymbol{x}}_2^n | \{\tilde{\boldsymbol{V}}_{12,c}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_{1,c}\}^n, \tilde{\boldsymbol{H}}^n\right) \\
&\quad - \mathcal{I}\left(\{\tilde{\boldsymbol{V}}_{12,ab}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_1\}^n; \tilde{\boldsymbol{x}}_2^n | \{\tilde{\boldsymbol{V}}_{12,c}^\dagger \tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{n}}_{1,c}\}^n, \tilde{\boldsymbol{H}}^n\right) \\
&\quad + (1 - \mu)nL(K - M_1)\log(P) + o(\log(P)) \quad (97)
\end{aligned}$$

Notice that the size of $\boldsymbol{V}_{12,ab}^\dagger$ is $K \times M_2$. Based on [24, Lemma 3], if we choose

$$\mu = \frac{K}{J} \quad (98)$$

the difference of the first two mutual information terms of (97) is at most in the order of $o(\log(P))$ and we have

$$\lambda \leq \left(1 - \frac{K}{J}\right)(K - M_1) = \frac{(\min(M_2, N_2) - N_1)(K - M_1)}{\min(M_2, N_2) - (N_1 - K)} \quad (99)$$

Recall that $d_1 + \mu d_2 \leq M_1 + \mu(N_1 - M_1) + \lambda$. We thus have the outer bound on the sum DoF as shown in (38) and the proof of the converse part of Theorem 4 is complete. ∎

B. Achievability

In order to show the achievability part of Theorem 4, we only need to construct an achievable scheme for the corner point of the DoF region. Without loss of generality, we assume that $M_2 = \min(M_2, N_2)$; otherwise, transmitter 2 can simply use N_2 transmit antennas. Since $K(N_1 - M_1) + (M_2 - N_1)(K - M_1) = M_2(K - M_1) + M_1(N_1 - K)$, it is sufficient to show that the following DoF pair

$$(d_1, d_2) = (KM_1, M_2(K - M_1) + M_1(N_1 - K)) \tag{100}$$

can be achieved over K slots with antenna mode switching at transmitter one among K modes. Similar to Section IV-D, we choose the mode switching pattern as follows:

$$\boldsymbol{E}(1) = [\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_{M_1}],$$

$$\boldsymbol{E}(2) = [\boldsymbol{e}_2, \boldsymbol{e}_3, \ldots, \boldsymbol{e}_{M_1+1}],$$

$$\vdots$$

$$\boldsymbol{E}(K) = [\boldsymbol{e}_K, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_{M_1-1}].$$

We propose to use a generalization of the joint nulling and beamforming design that is investigated in Section IV-C. Unlike the frequency nulling that has been used for $K = N_1$, this scheme requires that receiver 1 performs nulling in both frequency and spatial domains. We hereby use two superscripts F and S to indicate the matrices that associated with frequency processing and spatial processing.

The generalized joint nulling and beamforming has the following structure:

$$\tilde{\boldsymbol{Q}} = \boldsymbol{Q}^{\mathsf{F}}_{M_1 \times K} \otimes \boldsymbol{Q}^{\mathsf{S}}_{K \times N_1}, \tag{101}$$

$$\tilde{\boldsymbol{P}} = [\tilde{\boldsymbol{P}}_a, \tilde{\boldsymbol{P}}_b], \text{ where} \tag{102}$$

$$\tilde{\boldsymbol{P}}_a = [\boldsymbol{P}^{\mathsf{F}}_a]_{K \times (K - M_1)} \otimes [\boldsymbol{P}^{\mathsf{S}}_a]_{M_2 \times M_2}, \tag{103}$$

$$\tilde{\boldsymbol{P}}_b = [\boldsymbol{P}^{\mathsf{F}}_b]_{K \times M_1} \otimes [\boldsymbol{P}^{\mathsf{S}}_b]_{M_2 \times (N_1 - K)}. \tag{104}$$

The received signal at receiver 1 can be written as

$$\tilde{\boldsymbol{y}}_1 = \tilde{\boldsymbol{H}}_{11}\tilde{\boldsymbol{x}}_1 + \tilde{\boldsymbol{H}}_{12}[\tilde{\boldsymbol{P}}_a, \tilde{\boldsymbol{P}}_b]\tilde{\boldsymbol{x}}_2 + \tilde{\boldsymbol{z}}_1 \tag{105}$$

where $\tilde{\boldsymbol{x}}_1$ is a length $M_1 K$ vector, and $\tilde{\boldsymbol{x}}_2$ is a length $M_2(K - M_1) + M_1(N_1 - K)$ vector.
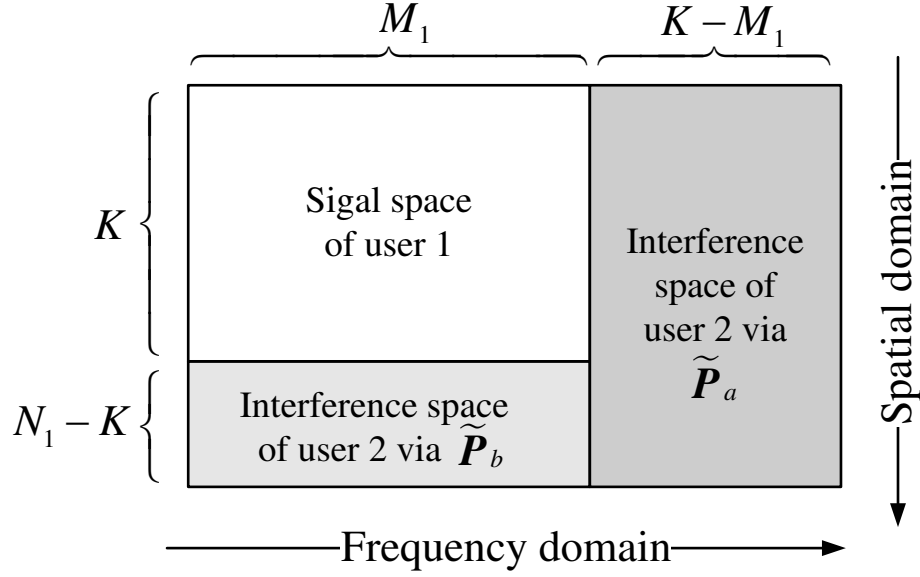
Fig. 2. Space-Frequency dimension allocation for the two users when $K < N_1$.

After applying nulling matrix $\tilde{Q}$, we have

$$\tilde{Q}\tilde{y}_1 = \underbrace{\tilde{Q}\tilde{H}_{11}}_{\tilde{A}}\tilde{x}_1 + \left[\underbrace{\tilde{Q}\tilde{H}_{12}\tilde{P}_a}_{\tilde{B}}, \underbrace{\tilde{Q}\tilde{H}_{12}\tilde{P}_b}_{\tilde{C}} \right]\tilde{x}_2 + \tilde{Q}\tilde{z}_1. \tag{106}$$

To achieve the degrees of freedom pair shown in (100) for both users, it is sufficient to design our $\tilde{P}$ and $\tilde{Q}$ to satisfy the following conditions simultaneously

1) $\operatorname{rank}(\tilde{A}) = M_1 K$,
2) $\operatorname{rank}([\tilde{P}_a, \tilde{P}_b]) = M_2(K - M_1) + M_1(N_1 - K)$,
3) $\tilde{B} = \mathbf{0}$,
4) $\tilde{C} = \mathbf{0}$.

We propose to use the following realizations:

$$Q^{\mathsf{F}} = [\mathcal{V}_K(1, \omega_K, \dots, \omega_K^{M_1-1})]^T, \tag{107}$$

$$P_a^{\mathsf{F}} = \mathcal{V}_K(\omega_K^{-M_1}, \omega_K^{-(M_1+1)}, \dots, \omega_K^{-(K-1)}). \tag{108}$$

$$P_b^{\mathsf{F}} = (Q^{\mathsf{F}})^{\dagger}, \tag{109}$$

$$P_a^{\mathsf{S}} = I_{M_2}, \tag{110}$$

$$P_b^{\mathsf{S}} = [I_{N_1-K}; 0], \tag{111}$$

$$Q^{\mathsf{S}} = \text{null}(H_{12}P_b^{\mathsf{S}})^T, \tag{112}$$

where (112) means that $Q^{\mathsf{S}} H_{12} P_b^{\mathsf{S}} = 0$. Here, we choose $((Q^{\mathsf{F}})^{\dagger}, P_a^{\mathsf{F}})$ to be a size $K \times K$ IFFT matrix, which offers the same frequency domain explanation as discussed in Section IV-E; see also Fig. 2. It is trivial to see $\tilde{B} = 0$. In other words, receiver 1 will simply ignore the signal in the last $K - M_1$ frequencies and only using the signal in the first K frequencies to decode his own message. Therefore, $\tilde{P}_a$ contains the interference directions from all the antennas of transmitter 2 but only in certain frequencies. Now, after applying the frequency nulling, there are $N_1 K$ dimensions remaining, which contain both user 1's message and the message of user 2 that is transmitted by $\tilde{P}_b$. Among all the $N_1 K$ dimensions, receiver 1 only requires $M_1 K$ dimensions to decode his own message, while leaving additional $K(N_1 - M_1)$ dimensions for user 2. Here we choose one possible way of decomposing the remaining dimensions. Transmitter 2 sends some messages in the first M_1 frequencies but only though $N_1 - K$ antennas, as shown in (111). Notice that

$$
\begin{aligned}
\tilde{C} &= \tilde{Q} \tilde{H}_{12} \tilde{P}_b \\
&= (Q^{\mathsf{F}} \otimes Q^{\mathsf{S}})(I_K \otimes H_{12})(P_b^{\mathsf{F}} \otimes P_b^{\mathsf{S}}) \tag{113} \\
&= (Q^{\mathsf{F}} P_b^{\mathsf{F}}) \otimes (Q^{\mathsf{S}} H_{12} P_b^{\mathsf{S}}) \tag{114}
\end{aligned}
$$

which means that the choice of Q^{S} as given in (112) is sufficient to set $\tilde{C} = 0$. It is clear that for the interference signal sent via $\tilde{P}_b$, receiver 1 only need to do spatial zero-forcing in our scheme, which can be seen from the fact $Q^{\mathsf{F}} P_b^{\mathsf{F}} = I_K$ due to (109).

To satisfy the second condition, notice that $\text{rank}(\tilde{P}_a) = M_2(K - M_1)$ and $\text{rank}(\tilde{P}_a) = M_1(N_1 - K)$, it is sufficient to show that $\tilde{P}_a \perp \tilde{P}_b$, which is obvious as

$$\tilde{P}_b^{\dagger} \tilde{P}_a = (Q^{\mathsf{F}} P_a^{\mathsf{F}}) \otimes ((P_b^{\mathsf{S}})^{\dagger} I_{M_2}) = 0 \tag{115}$$

because $Q^{\mathsf{F}} P_a^{\mathsf{F}} = 0$. This is not surprising as the signal of user 2 transmitted via $\tilde{P}_a$ and $\tilde{P}_b$ are orthogonal in frequency domain. The remaining part is to show the first condition holds, which is true because here $\tilde{A}$ has the same structure as $\tilde{A}'$ of (34) with N_1 replaced by K and h_i replaced by $Q_{K \times N_1}^{\mathsf{S}} h_i$.
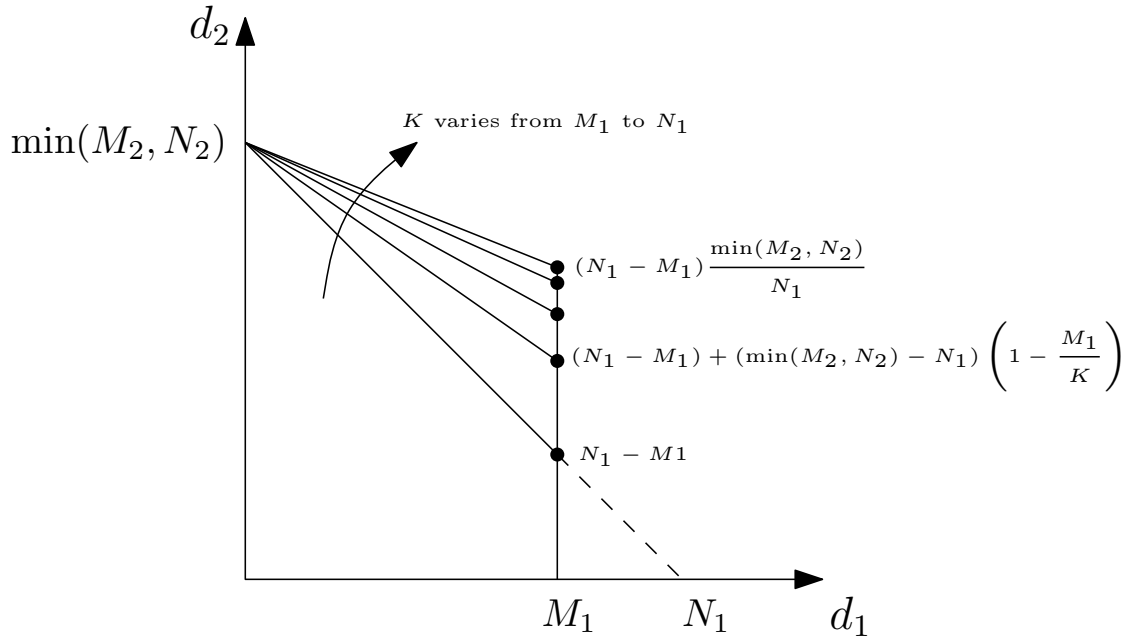
Fig. 3. The benefit of antenna mode switching on the DoF region, in the case of $M_1 < N_1 < \min(M_2, N_2)$.

C. Discussion

It is not surprising that when $K = N_1$, (38) implies

$$d_1 + \frac{N_1}{\min(M_2, N_2)} d_2 \leq N_1 \tag{116}$$

which is the same as (7) and that in [17, Theorem 3] when $M_1 < N_1 < \min(M_2, N_2)$. For the scheme that we discussed above, $\tilde{\boldsymbol{P}}_b$ disappears and it is the DoF achievable scheme that we developed in Section IV-C. In addition, when $K = M_1$, (38) becomes (12) and $\tilde{\boldsymbol{P}}_a$ disappears, the general scheme reduces to the DoF-optimal spatial zero-forcing as shown in [24]. Hence, for one extra mode at transmitter 1, we can further align $\min(M_2, N_2) - N_1$ streams of interference over K slots. The incremental gain per slot is reduced when K increases; see Fig. 3. Our result reveals the fundamental benefit that can be obtained from reconfigurable antenna modes when there is no CSIT and $M_1 < N_1 < \min(M_2, N_2)$. In addition, combining with the known results, we know that in order to achieve the DoF region of two-user FIC and ZIC, zero-forcing in frequency and spatial domains suffice regardless of the CSIT assumption.

VI. Conclusions

We derived the exact DoF region for the MIMO Z and full interference channels when perfect channel state information is available at receivers, including i) the Z interference channel with channel state information at the transmitter; ii) the Z and full interference channel without channel state information at the transmitter, but with reconfigurable antennas at the transmitters. For both FIC and ZIC, when the number of antenna modes K at the transmitter with the reconfigurable antennas is not less than the number of receive antennas at the corresponding receiver, the DoF region is maximized and no longer depends on the number of antenna modes. Otherwise, each additional antenna mode can bring extra gain in the DoF region when $M_1 < N_1 < \min(M_2, N_2)$ for both FIC and ZIC, and when $M_2 < N_2 < \min(M_1, N_1)$ for FIC. The incremental gain diminishes as K increases.

The achievability schemes we designed for the reconfigurable antenna cases rely on time expansion and joint beamforming and nulling over the time-expanded channel. Interestingly, they also bear a space-frequency coding interpretation. We completely characterized the DoF regions for both Z and full interference channels when transmitter antenna mode switching is allowed. Our result can specialize to previously known cases when there is no antenna mode switching by simply setting the number of antenna modes equal to the number of transmit antennas. Our work reveals how the channel variation introduced by the extra antenna mode switching brings benefits in the sense of the DoF region.

Acknowledgment

References

[1] A. Carleial, "A case where interference does not reduce capacity (Corresp.)," *IEEE Trans. Inform. Theory*, vol. 21, no. 5, pp. 569–570, May 1975.

[2] H. Sato, "The capacity of the Gaussian interference channel under strong interference (Corresp.)," *IEEE Trans. Inform. Theory*, vol. 27, no. 6, pp. 786–788, June 1981.

[3] A. El Gamal and M. Costa, "The capacity region of a class of deterministic interference channels (Corresp.)," *IEEE Trans. Inform. Theory*, vol. 28, no. 2, pp. 343–346, Feb. 1982.

[4] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[5] X. Shang, G. Kramer, and B. Chen, "Outer bound and noisy-interference sum-rate capacity for symmetric Gaussian interference channels," in *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on*, 2008, pp. 385–389.

[6] V. Sreekanth Annapureddy and V. Veeravalli, "Sum capacity of the Gaussian interference channel in the low interference regime," in *Information Theory and Applications Workshop, 2008*, 2008, pp. 422–427.

[7] A. Motahari and A. Khandani, "Capacity bounds for the Gaussian interference channel," in *Proc. IEEE Intl. Symp. on Info. Theory*, 2008, pp. 250–254.

[8] G. Bresler and D. Tse, "The two-user Gaussian interference channel: a deterministic view," *European Trans. on Telecomm.*, vol. 19, no. 4, pp. 333–354, Apr. 2008.

[9] G. Bresler, A. Parekh, and D. Tse, "The approximate capacity of the many-to-one and one-to-many gaussian interference channels," 2008. [Online]. Available: http://arxiv.org/abs/0809.3554v1

[10] S. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: a deterministic approach," 2009. [Online]. Available: http://arxiv.org/abs/0906.5394

[11] V. S. Annapureddy and V. V. Veeravalli, "Sum capacity of MIMO interference channels in the low interference regime," 2009. [Online]. Available: http://arxiv.org/abs/0909.2074

[12] X. Shang, B. Chen, G. Kramer, and H. V. Poor, "Capacity regions and sum-rate capacities of vector gaussian interference channels," 2009. [Online]. Available: http://arxiv.org/abs/0907.0472

[13] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, May 2003.

[14] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the K user interference channel," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[15] S. Jafar and M. Fakhereddin, "Degrees of freedom for the MIMO interference channel," *IEEE Trans. Inform. Theory*, vol. 53, no. 7, pp. 2637–2642, July 2007.

[16] S. Jafar and S. Shamai, "Degrees of freedom region of the MIMO X channel," *IEEE Trans. Inform. Theory*, vol. 54, no. 1, pp. 151–170, Jan. 2008.

[17] C. Huang, S. Jafar, S. Shamai, and S. Vishwanath, "On degrees of freedom region of MIMO networks without CSIT," 2009. [Online]. Available: http://arxiv.org/abs/0909.4017

[18] Y. Zhu and D. Guo, "Isotropic MIMO interference channels without CSIT: The loss of degrees of freedom," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, 2009, pp. 1338–1344.

[19] C. S. Vaze and M. K. Varanasi, "The degrees of freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT ," 2009. [Online]. Available: http://arxiv.org/abs/0909.5424

[20] X. Shang, B. Chen, G. Kramer, and H. V. Poor, "MIMO Z-interference channels: capacity under strong and noisy interference," 2009. [Online]. Available: http://arxiv.org/abs/0911.4530

[21] Y. Zhu and D. Guo, "Ergodic fading one-sided interference channels without state information at transmitters," 2009. [Online]. Available: http://arxiv.org/abs/0911.1082

[22] S. A. Jafar, "Exploiting channel correlations - simple interference alignment schemes with no CSIT," 2009. [Online]. Available: http://arxiv.org/abs/0910.0555

[23] C. Wang, T. Gou, and S. A. Jafar, "Aiming perfectly in the dark - blind interference alignment through staggered antenna switching," 2010. [Online]. Available: http://arxiv.org/abs/1002.2720

[24] Y. Zhu and D. Guo, "The Degrees of Freedom of MIMO Interference Channels without State Information at Transmitters," 2010. [Online]. Available: http://arxiv.org/abs/1008.5196

[25] T. Jiang, N. Sidiropoulos, and J. ten Berge, "Almost-sure identifiability of multidimensional harmonic retrieval," *Signal Processing, IEEE Transactions on*, vol. 49, no. 9, pp. 1849–1859, Sept. 2001.