The Optimal DoF Region for the Two-User Non-Coherent SIMO Multiple-Access Channel

Khac-Hoang Ngo, Sheng Yang, Maxime Guillaud

The Optimal DoF Region for the Two-User Non-Coherent SIMO Multiple-Access Channel

Khac-Hoang Ngo*†, Sheng Yang*, Maxime Guillaud†
*LSS, CentraleSupélec, 91190 Gif-sur-Yvette, France
†Mathematical and Algorithmic Sciences Lab, Paris Research Center, Huawei Technologies, 92100 Boulogne-Billancourt, France
Email: {ngo.khac.hoang, maxime.guillaud}@huawei.com, sheng.yang@centralesupelec.fr

***Abstract*—The optimal degree-of-freedom (DoF) region of the non-coherent multiple-access channels is still unknown in general. In this paper, we make some progress by deriving the *entire* optimal DoF region in the case of the two-user single-input multiple-output (SIMO) generic block fading channels. The achievability is based on a simple training-based scheme. The novelty of our result lies in the converse using a genie-aided bound and the duality upper bound. As a by-product, our result generalizes previous proofs for the single-user Rayleigh block fading channels.**

I. Introduction

The fundamental limit of communication over wireless fading channels depends on the availability of channel state information (CSI) at the transmitter/receiver. While the channel statistics are normally stable and can be assumed to be available, the assumption on *instantaneous* CSI varies with the context. When the instantaneous CSI is assumed to be *a priori* known, e.g., in fixed environments where it changes slowly and can be estimated accurately at negligible cost, at least at the receiver side, the communication is said to be *coherent*. On the other hand, if the instantaneous CSI is *a priori* unknown, e.g, when the estimation cost is not negligible, the communication is said to be *non-coherent*.

In a point-to-point multiple-input multiple-output (MIMO) channel with M transmit and N receive antennas, it is well known that the *coherent* capacity scales linearly with the number of antennas as $C \sim \min\{M,N\} \log \mathsf{SNR}$ at high signal-to-noise ratio (SNR) [1], [2]. In the *non-coherent* case with stationary fading, the capacity scales as $\log\log \mathsf{SNR} + \chi(\mathbf{H}) + o(1)$[1] [3], implying a DoF of 0. Nevertheless, if the channel remains constant during a certain amount of slots, say T slots, then the DoF becomes strictly positive as $M^*(1-\frac{M^*}{T})$ where $M^* := \min\left\{M, N, \lfloor \frac{T}{2} \rfloor\right\}$. This fading setup is commonly referred to as the block fading channel, and has been extensively investigated in the literature [4], [5], [6]. Remarkably, in the block fading case, the optimal DoF can be achieved either by well-designed space-time modulations [4], [5], [6], or by simple training-based strategies [7]. The converse in the aforementioned works was based on the Rayleigh fading assumption, using either a direct approximation at high SNR [5] or a duality upper bound with a carefully chosen auxiliary output distribution [6].

[1] $\chi(\mathbf{H})$ is called the fading number of the channel.

In multi-user MIMO channels, such as the broadcast channels (BC) and the multiple access channels (MAC), non-coherent communications have been studied in the block fading case. For the BC, the exact DoF region is known with *isotropic* Rayleigh fading (a special case of stochastically degraded BC) and can be achieved with time division multiple access (TDMA) [8]. Some achievable schemes have been proposed for the BC with spatially correlated fading [9], [10]. For the MAC, it has been shown that the optimal sum DoF can be achieved with a training-based scheme [8], but the optimal DoF region is still *unknown*.

In this work, we make some progress for the non-coherent single-input multiple-output (SIMO) MAC. Specifically, we derive the optimal DoF region in the case of two single-antenna transmitters (users) and a N-antenna receiver in block fading channel with coherence time T. When $N = 1$, the region is achieved with a simple time division multiplexing between two users. In this case, letting two users cooperate does not help exploit more degrees of freedom and it is optimal to activate only one user at a time to achieve $1 - \frac{1}{T}$ DoF for that user. When $N > 1$, a training-based scheme can achieve another DoF pair. We let two users send orthogonal pilots for channel estimation in the first 2 time slot, then send data simultaneously in the remaining $T-2$ time slots. In this way, each user can achieve $1 - \frac{2}{T}$ DoF.

The main technical contribution of this paper lies in the converse proof. Leveraging the duality upper bound [3], we carefully choose an output distribution with which we derive a tight outer bound on the DoF region. Unlike previous results such as [5], [6], we do not assume the Gaussianity of the channel coefficients, which makes our proof more general and our results stronger even in the single-user case.

The remainder of this paper is organized as follows. The system model and preliminaries are presented in Section II. In Section III, we provide the main result on the optimal DoF region of the two-user MAC, as well as the proof for the case $N = 1$ and the achievability for the case $N > 1$. We introduce the converse proof technique through a new proof for the single-user SIMO channel in Section IV, and use it to show the tight outer bound for the case $N > 1$ of the MAC in Section V. Finally, we conclude the paper in Section VI.

Notations: For random quantities, we use upper case non-italic letters: normal fonts, e.g., X, for scalars; bold fonts, e.g., $\mathbf{V}$, for vectors; and bold and sans serif fonts, e.g., M, for matrices. Deterministic quantities are denoted with italic letters, e.g., a scalar x, a vector $\boldsymbol{v}$, and a matrix $\boldsymbol{M}$. Throughout the paper, we adopt the column convention for vectors. The Euclidean norm of a vector and a matrix is denoted by $\|\boldsymbol{v}\|$ and $\|\boldsymbol{M}\|$, respectively. The trace, transpose and conjugated transpose of $\boldsymbol{M}$ is $\operatorname{tr}\{\boldsymbol{M}\}$, $\boldsymbol{M}^{\mathsf{T}}$ and $\boldsymbol{M}^{\mathsf{H}}$, respectively. $\mathsf{M}_{[i:j]}$ denotes the sub-matrix containing columns from i to j of a matrix M (thus $\mathsf{M}_{[i]}$ denotes column i). $\operatorname{diag}(x_1,\dots,x_N)$ denotes the diagonal matrix with diagonal entries $x_1,\dots,x_N$. $H(.)$, $h(.)$, and $D(.\|.)$ denote the entropy, differential entropy, and Kullback-Leibler divergence, respectively. Logarithms are in base 2. $(x)^+ = \max\{x,0\}$. ":=" means "is defined as". $\Gamma(x) = \int_0^\infty z^{x-1}e^{-z}dz$ is the Gamma function. Given two functions f and g, we write $f(x) = O(g(x))$ if there exists a constant $c > 0$ and some x_0 such that $f(x) \leq cg(x), \forall x \geq x_0$.

II. System Model and Preliminaries

We consider a single-input multiple-output (SIMO) multiple-access channel in which two single-antenna users send their signals to a receiver with N antennas. The channel between the users and the receiver is flat and block fading with equal and synchronous coherence interval of T symbol periods. That is, the channel vector $\mathbf{H}_k \in \mathbb{C}^{N\times 1}$, $k = 1, 2$, remains unchanged during each block of length T symbols and changes independently between blocks. The realizations of $\mathbf{H}_1$ and $\mathbf{H}_2$ are *unknown* to both the users and the receiver. The received signal during the coherence block b, $b = 1, 2, \dots$[2], is

$$\mathsf{Y}[b] = \mathbf{H}_1[b]\,\mathbf{X}_1^{\mathsf{T}}[b] + \mathbf{H}_2^{\mathsf{T}}[b]\,\mathbf{X}_2^{\mathsf{T}}[b] + \mathsf{Z}[b], \tag{1}$$

where $\mathbf{X}_1 \in \mathbb{C}^T$ and $\mathbf{X}_2 \in \mathbb{C}^T$ are the transmitted signals from user 1 and user 2, respectively, with the power constraint

$$\frac{1}{B}\sum_{b=1}^{B} \|\mathbf{X}_i[b]\|^2 \leq PT, \quad i = 1, 2, \tag{2}$$

where B is the number of the blocks spanned by a codeword. We assume that $\mathsf{Z} \in \mathbb{C}^{N\times T}$ is the additive white Gaussian noise with independent and identically distributed (i.i.d.) $\mathcal{CN}(0, 1)$ entries. The parameter P is the average power ratio between the transmitted signal and the noise, thus we refer to P as the SNR of the channel.

Since the channel is block memoryless[3], it is well known that a rate pair $(R_1(P), R_2(P))$ in bits per channel use is achievable at SNR P, i.e., lies within the capacity region $\mathcal{C}_{\text{Avg}}(P)$, for the MAC if and only if

$$R_1 + R_2 \leq \frac{1}{T} I(\mathbf{X}_1, \mathbf{X}_2; \mathsf{Y}), \tag{3}$$

$$R_1 \leq \frac{1}{T} I(\mathbf{X}_1; \mathsf{Y}|\mathbf{X}_2), \tag{4}$$

$$R_2 \leq \frac{1}{T} I(\mathbf{X}_2; \mathsf{Y}|\mathbf{X}_1), \tag{5}$$

for some input distribution subject to the average power constraint P (as the codeword length B goes to infinity) [11]. Then, we say that (d_1, d_2) is an achievable DoF pair with

$$d_k := \liminf_{P\to\infty} \frac{R_k(P)}{\log(P)}, \quad k = 1, 2. \tag{6}$$

The optimal DoF region $\mathcal{D}_{\text{Avg}}(P)$ is defined as the set of all achievable DoF pairs.

We assume that the channel vectors $\mathbf{H}_1$ and $\mathbf{H}_2$ are independent[4] and drawn from a generic distribution satisfying the following conditions:

$$h(\mathbf{H}_k) > -\infty, \quad \mathbb{E}\left[\|\mathbf{H}_k\|^2\right] < \infty, \quad k = 1, 2. \tag{7}$$

The following results, whose proofs are provided in Appendix A, are useful for our main analysis.

Lemma 1. *Let $\boldsymbol{A} \in \mathbb{C}^{m\times t}$ have full column rank, $\mathsf{W} \in \mathbb{C}^{n\times m}$ be such that $h(\mathsf{W}) > -\infty$ and $\mathbb{E}\left[\|\mathsf{W}\|_{\text{F}}^2\right] < \infty$, then we have*

$$h(\mathsf{W}\boldsymbol{A}) = n \log\det(\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}) + c_0 \tag{8}$$

where c_0 is bounded by some constant that only depends on the statistics of W.

Lemma 2. *Let $\mathrm{X} \geq 0$ be some random variable such that $\mathbb{E}[\mathrm{X}] < \infty$ and $h(\mathrm{X}/\mathbb{E}[\mathrm{X}]) > -\infty$. Then, for any $\alpha < 1$,*

$$\mathbb{E}[\log(1 + \mathrm{X})] \geq \alpha \log(1 + \mathbb{E}[\mathrm{X}]) + c_0 \tag{9}$$

where $c_0 > -\infty$ is some constant that only depends on α.

From the above result, we observe that when $\mathbb{E}[\mathrm{X}] \to \infty$, $\frac{\mathbb{E}[\log(1+\mathrm{X})]}{\log(1+\mathbb{E}[\mathrm{X}])} \approx 1$ since we can let α be arbitrarily close to 1. The upper bound is simply from Jensen's inequality.

If the support of the input distribution is further bounded such that $\|\mathbf{X}_i\|^2 \leq P$, $i = 1, 2$, then we say that the input satisfies the peak power constraint P. In this case, the capacity region and DoF region are denoted $\mathcal{C}_{\text{Peak}}(P)$ and $\mathcal{D}_{\text{Peak}}(P)$, respectively. Since the peak power constraint implies the average power constraint, we have that

$$\mathcal{C}_{\text{Peak}}(P) \subseteq \mathcal{C}_{\text{Avg}}(P), \quad \mathcal{D}_{\text{Peak}}(P) \subseteq \mathcal{D}_{\text{Avg}}(P). \tag{10}$$

Lemma 3. *For any rate pair (R_1, R_2) achievable under the average power constraint P, for any $\beta > 1$, there exists (R'_1, R'_2) achievable under the peak power constraint P^β, such that*

$$R_k - R'_k = O(P^{1-\beta} \log P^\beta), \quad k = 1, 2, \tag{11}$$

In short,

$$\mathcal{C}_{\textit{Avg}}(P) \subseteq \mathcal{C}_{\textit{Peak}}(P^\beta) + O(P^{1-\beta} \log P^\beta), \quad \forall\, \beta > 1. \tag{12}$$

Since the pre-log of the gap $P^{1-\beta}\log P^\beta$ is vanishing at high SNR for any $\beta > 1$, we have the DoF region

$$\mathcal{D}_{\text{Avg}}(P) \subseteq \mathcal{D}_{\text{Peak}}(P^\beta) \subseteq \mathcal{D}_{\text{Avg}}(P^\beta), \quad \forall\, \beta > 1. \tag{13}$$

Letting β arbitrarily close to 1, we conclude that using the peak power constraint instead of the average power constraint does not change the optimal DoF region. We therefore consider

[2]Throughout, we omit the block index whenever confusion is unlikely.

[3]The results can be generalized to stationary fading as done in [3].

[4]Independence is not necessary but makes the analysis slightly simpler.

throughout the peak power constraint, which can simplify considerably the analysis.

Lemma 4. *Let $\mathbf{Y} \in \mathbb{C}^N$ be a vector-valued random variable with distribution $\mathcal{P}$. Consider another family of distributions $\mathcal{R}$ whose densities are given by*

$$r_{\mathbf{Y}}(\boldsymbol{y}) = \frac{\Gamma(N)|\det\boldsymbol{A}|^2}{\pi^N\beta^\alpha\Gamma(\alpha)}\|\boldsymbol{A}\boldsymbol{y}\|^{2(\alpha-N)}\exp\left(-\frac{\|\boldsymbol{A}\boldsymbol{y}\|^2}{\beta}\right), \quad (14)$$

for $\boldsymbol{y} \in \mathbb{C}^N$, where $\alpha, \beta > 0$, $\boldsymbol{A}$ is any nonsingular deterministic $N \times N$ complex matrix. When $\beta = \mathbb{E}_{\mathcal{P}}[\|\boldsymbol{A}\mathbf{Y}\|^2]$ and $\alpha = 1/\log(\beta) = 1/\log(\mathbb{E}_{\mathcal{P}}[\|\boldsymbol{A}\mathbf{Y}\|^2])$, denote this distribution as $\mathcal{R}(N, \boldsymbol{A})$. In this case,

$$\mathbb{E}_{\mathcal{P}}[-\log(r_{\mathbf{Y}}(\mathbf{Y}))] = -\log|\det\boldsymbol{A}|^2 + N\mathbb{E}_{\mathcal{P}}[\log\|\boldsymbol{A}\mathbf{Y}\|^2] + O(\log\log(\mathbb{E}\left[\|\boldsymbol{A}\mathbf{Y}\|^2\right])). \quad (15)$$

If we take $\mathbf{Y}$ as the channel output, as long as $\mathbb{E}\left[\|\boldsymbol{A}\mathbf{Y}\|^2\right] \leq P^{c_0}$ for any constant c_0 whose value only depends on the channel statistics, the term $O(\log\log(\mathbb{E}\left[\|\boldsymbol{A}\mathbf{Y}\|^2\right]))$ scales double-logarithmically with P. Therefore, in the DoF sense, it is enough to consider only the first two terms in (15).

III. Main Result

The main finding of this paper is the optimal DoF region of the MAC described above, as stated in Theorem 1.

Theorem 1. *For the non-coherent multiple-access channel with two single-antenna transmitters and a N-antenna receiver in flat and block fading with coherence time T, the optimal DoF region is characterized by*

$$d_1 + d_2 \leq 1 - \frac{1}{T}, \quad (16)$$

if $T \leq 2$ or $N = 1$, and

$$\frac{d_1}{T-2} + d_2 \leq 1 - \frac{1}{T}, \quad (17)$$

$$d_1 + \frac{d_2}{T-2} \leq 1 - \frac{1}{T}, \quad (18)$$

otherwise.

Remark III.1. *When $T \to \infty$, the optimal DoF region approaches the region in the coherent case: $d_1 + d_2 \leq 1$ if $N = 1$, and $\max\{d_1, d_2\} \leq 1$ if $N > 1$ (as shown in Figure 1).*

The case $T = 1$ (stationary fading) is trivial: zero DoF is achievable, even if two users cooperate [3]. If $T = 2$ or $N = 1$, the optimal DoF region is achieved with time division multiplexing between the users, noting that the active user can achieve $1 - \frac{1}{T}$ DoF by either a training-based scheme [7] or unitary space-time modulations [4], [5]. The tight outer bound follows by letting two users cooperate, then according to [5], [6], it is optimal to use $\min\{2, N, \lfloor\frac{T}{2}\rfloor\} = 1$ transmit antenna and achieve $1 - \frac{1}{T}$ DoF in total.

When $T \geq 3, N > 1$, the region is the convex hull of the origin and three points: $\left(1 - \frac{1}{T}, 0\right)$, $\left(0, 1 - \frac{1}{T}\right)$, and $\left(1 - \frac{2}{T}, 1 - \frac{2}{T}\right)$. The first two points are achieved by activating only one user. The third point is achieved with a training-based scheme: let two users send orthogonal pilots in the first two time slot for the receiver to learn their channel, then send data in the remaining $T - 2$ time slots. The region is then achieved with time sharing between these points. It remains to show the tight outer bound for this case $T \geq 3, N > 1$, but before that, let us introduce the proof technique by using it for a new proof of the tight DoF for the single-user SIMO channel in the next section.

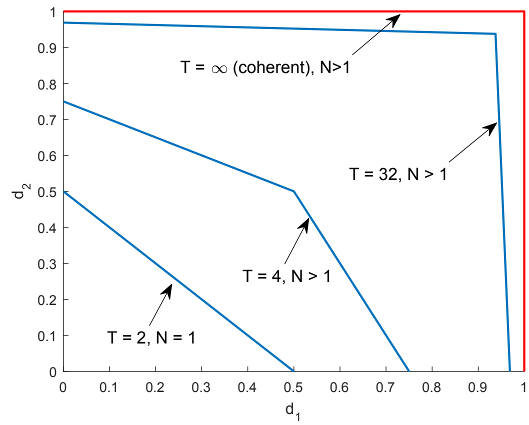

Fig. 1. The optimal DoF region of two-user SIMO MAC with N receive antennas in block fading with coherence time T.

IV. Single-User SIMO Channel Revisited

Consider the single-user (point-to-point) SIMO channel with block fading with coherence time T

$$\mathsf{Y} = \mathbf{H}\mathbf{X}^{\mathsf{T}} + \mathsf{Z}, \quad (19)$$

where we have the same assumptions as in the MAC channel. It was shown that the DoF of this channel is $1 - \frac{1}{T}$ and can be achieved with either a training-based scheme [7] or well-designed space time modulations [4], [5], [6]. For the converse of the high SNR capacity (which implies the converse of the DoF), while $h(\mathsf{Y}|\mathbf{X})$ can be calculated easily, the upper bound for $h(\mathsf{Y})$ is much more involved [5], [6]. In this section, we provide a simpler proof for the converse of the DoF using the duality approach as in [6] but with a simple choice of auxiliary output distribution.

First, let us define the random variable V as the index of the strongest input component, i.e.,[5]

$$\mathrm{V} := \arg\max_{i=1,2,\ldots,T}|\mathrm{X}_i|^2. \quad (20)$$

Thus, X_V denotes the entry in $\mathbf{X}$ with the largest magnitude. Let the genie give V to the receiver,[6] we have

$$\begin{aligned} I(\mathbf{X}; \mathsf{Y}) &\leq I(\mathbf{X}; \mathsf{Y}, \mathrm{V}) && (21)\\ &= I(\mathbf{X}; \mathsf{Y}|\mathrm{V}) + I(\mathbf{X}; \mathrm{V}) && (22)\\ &\leq h(\mathsf{Y}|\mathrm{V}) - h(\mathsf{Y}|\mathbf{X}, \mathrm{V}) + H(\mathrm{V}) && (23)\\ &\leq h(\mathsf{Y}|\mathrm{V}) - h(\mathsf{Y}|\mathbf{X}) + \log(T), && (24) \end{aligned}$$

[5]When there are more than one such components, we pick an arbitrary one.

[6]This technique of giving the index of the strongest input component to the receiver was initially proposed in [12] for phase noise channel.

where the last inequality is because we have the Markov chain $\mathrm{V} \leftrightarrow \mathbf{X} \leftrightarrow \mathbf{Y}$ and $H(\mathrm{V}) \le \log(T)$. For a given $\mathbf{X}$, we can apply Lemma 1 with $\mathbf{W} = [\mathbf{H}\ \mathbf{Z}]$ and $\boldsymbol{A} = [\mathbf{X}\ \mathbf{I}_T]^{\mathsf{T}}$ to obtain

$$h(\mathbf{Y}|\mathbf{X}) = N\mathbb{E}\left[\log\det(\mathbf{I}_T + \mathbf{X}^*\mathbf{X}^{\mathsf{T}})\right] + O(1) \quad (25)$$
$$= N\mathbb{E}\left[\log(1 + \|\mathbf{X}\|^2)\right] + O(1). \quad (26)$$

To bound $h(\mathbf{Y}|\mathrm{V})$, we use the duality approach [3] as follows

$$\begin{aligned} h(\mathbf{Y}|\mathrm{V}) &= \mathbb{E}\left[-\log p(\mathbf{Y}|\mathrm{V})\right] \\ &= \mathbb{E}\left[-\log q(\mathbf{Y}|\mathrm{V})\right] - \mathbb{E}_{\mathrm{V}}[D(\mathcal{P}_{\mathbf{Y}|\mathrm{V}=v}\|\mathcal{Q})] \\ &\le \mathbb{E}\left[-\log q(\mathbf{Y}|\mathrm{V})\right], \end{aligned} \quad (27)$$

due to the non-negativity of the Kullback-Leibler divergence $D(\mathcal{P}_{\mathbf{Y}|\mathrm{V}=v}\|\mathcal{Q})$. Here, conditioned on V, the distribution $\mathcal{P}_{\mathbf{Y}|\mathrm{V}}$ with probability density function (pdf) $p(.)$ is imposed by the input, channel, and noise distributions, while $\mathcal{Q}$ is any distribution in $\mathbb{C}^{N\times T}$ with the pdf $q(.)$. Note that a proper choice of $\mathcal{Q}$ is the key to a tight upper bound. Our choice is inspired by a training-based scheme. Specifically, if we send a pilot symbol at time slot $v \in \{1,\dots,T\}$, then the output vector being the sum of $\mathbf{H}$ and $\mathbf{Z}_{[v]}$ should have comparable power in each direction since $\mathbf{H}$ is generic by assumption. Therefore, it is reasonable (in the DoF sense) to let $\mathbf{Y}_{[v]} \sim \mathcal{R}(N, \mathbf{I}_N)$, where the family of distributions $\mathcal{R}(N, \boldsymbol{A})$ is defined in Lemma 4. Now, $\mathbf{Y}_{[v]}$ should provide a rough estimate of the direction of the channel vector $\mathbf{H}$. Based on such an observation, it is also reasonable to assume that, given $\mathbf{Y}_{[v]}$, all other $\mathbf{Y}_{[i]}$, $i \ne v$, are mutually independent and follow

$$\mathbf{Y}_{[i]} \sim \mathcal{R}\left(N, \left(\mathbf{I}_N + \mathbf{Y}_{[v]}\mathbf{Y}_{[v]}^{\mathsf{H}}\right)^{-\frac{1}{2}}\right), \quad \forall i \ne v. \quad (28)$$

We thus obtain a "guess" of the auxiliary joint distribution $\mathcal{Q}_{\mathbf{Y}|\mathrm{V}=v}$.

Proposition 1. *With the above choice of auxiliary output distribution, it follows that*

$$\mathbb{E}\left[-\log q(\mathbf{Y}|\mathrm{V})\right] \le (N+T-1)\mathbb{E}\left[\log(1+|\mathrm{X}_{\mathrm{V}}|^2)\right] + N\mathbb{E}\left[\sum_{i=1, i\ne \mathrm{V}}^{T} \log\left(1 + \frac{|\mathrm{X}_i|^2}{1+|\mathrm{X}_{\mathrm{V}}|^2}\right)\right] + O(\log\log P). \quad (29)$$

Proof. See Appendix B. □

Plugging the bounds into (24), we obtain

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}) &\le (T-1)\mathbb{E}\left[\log(1+|\mathrm{X}_{\mathrm{V}}|^2)\right] + N\mathbb{E}\left[\log\frac{1+|\mathrm{X}_{\mathrm{V}}|^2}{1+\|\mathbf{X}\|^2}\right] \\ &\quad + N\mathbb{E}\left[\sum_{i=1, i\ne \mathrm{V}}^{T}\log\left(1+\frac{|\mathrm{X}_i|^2}{1+|\mathrm{X}_{\mathrm{V}}|^2}\right)\right] \\ &\quad + O(\log\log P) \\ &\le (T-1)\log(1+\mathbb{E}\left[|\mathrm{X}_{\mathrm{V}}|^2\right]) + O(\log\log P) \quad (30) \\ &\le (T-1)\log^+(P) + O(\log\log P), \quad (31) \end{aligned}$$

where we used the fact that $|\mathrm{X}_i|^2 \le |\mathrm{X}_{\mathrm{V}}|^2 \le \|\mathbf{X}\|^2$, $\forall i \ne \mathrm{V}$. Thus, the DoF is upper bounded by $\frac{T-1}{T}$, which is tight.

V. Two-User SIMO MAC

Let us get back to the MAC in this section and show that, when $T \ge 3, N > 1$, any achievable DoF pair (d_1, d_2) must satisfy (17) and (18).

A. *The $T \ge N+1 > 2$ case*

Let us consider the more straightforward case with $T \ge N+1 > 2$. We first bound R_1 and R_2 using similar techniques as for the single-user case, and then give the tight outer bound for the DoF region in the following steps.

Step 1: Output Rotation and Genie-Aided Bound: Given $\mathbf{X}_2$, the channel with respect to (w.r.t.) input $\mathbf{X}_1$ has equivalent noise $\mathbf{H}_2\mathbf{X}_2^{\mathsf{T}} + \mathbf{Z}$. Consider the following eigen-value decomposition

$$\mathbf{X}_2^*\mathbf{X}_2^{\mathsf{T}} = \mathbf{U}\ \mathrm{diag}(0,\dots,0,\|\mathbf{X}_2\|^2)\ \mathbf{U}^{\mathsf{H}}, \quad (32)$$

for some $T\times T$ unitary matrix $\mathbf{U}$. We consider the rotated output $\tilde{\mathbf{Y}} = \mathbf{Y}\mathbf{U} = \mathbf{H}_1\tilde{\mathbf{X}}_1^{\mathsf{T}} + \tilde{\mathbf{Z}}$, where $\tilde{\mathbf{X}}_1^{\mathsf{T}} = \mathbf{X}_1^{\mathsf{T}}\mathbf{U} = [\tilde{\mathrm{X}}_{11}\ \tilde{\mathrm{X}}_{12}\ \dots\ \tilde{\mathrm{X}}_{1T}]$ and $\tilde{\mathbf{Z}} = (\mathbf{H}_2\mathbf{X}_2^{\mathsf{T}} + \mathbf{Z})\mathbf{U}$. Note that given $\mathbf{X}_2$, the first $T-1$ columns of the noise $\tilde{\mathbf{Z}}$ are i.i.d. Gaussian whereas the last column is stronger as the sum of $\mathbf{H}_2\|\mathbf{X}_2\|$ and a Gaussian noise vector. Thus, we have

$$TR_1 \le I(\mathbf{X}_1;\mathbf{Y}|\mathbf{X}_2) = I(\tilde{\mathbf{X}}_1;\tilde{\mathbf{Y}}|\mathbf{X}_2). \quad (33)$$

Let us define the random variable V as the index of the strongest among the first $T-1$ elements of $\tilde{\mathbf{X}}_1$, namely,

$$\mathrm{V} = \arg\max_{i=1,2,\dots,T-1} |\tilde{\mathrm{X}}_{1i}|^2. \quad (34)$$

Similarly as in (24) with the genie-aided bound,

$$I(\tilde{\mathbf{X}}_1;\tilde{\mathbf{Y}}|\mathbf{X}_2) \le h(\tilde{\mathbf{Y}}|\mathbf{X}_2,\mathrm{V}) - h(\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}_1,\mathbf{X}_2) + \log(T-1). \quad (35)$$

Step 2: Bounding $h(\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}_1,\mathbf{X}_2)$ and $h(\tilde{\mathbf{Y}}|\mathbf{X}_2,\mathrm{V})$: Given $\tilde{\mathbf{X}}_1$ and $\mathbf{X}_2$, we can apply Lemma 1 with $\mathbf{W} = [\mathbf{H}_1\ \mathbf{H}_2\ \mathbf{Z}]$ and $\boldsymbol{A} = [\mathbf{X}_1\ \mathbf{X}_2\ \mathbf{I}_T]^{\mathsf{T}}\mathbf{U}$ to obtain

$$\begin{aligned} h(\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}_1,\mathbf{X}_2) &= N\mathbb{E}\left[\log\det(\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A})\right] + O(1) \\ &= N\mathbb{E}\left[\log\left((1+\|\mathbf{X}_2\|^2)\left(1+\sum_{i=1}^{T-1}|\tilde{\mathrm{X}}_{1i}|^2\right) + |\tilde{\mathrm{X}}_{1T}|^2\right)\right] \\ &\quad + O(1), \end{aligned} \quad (36)$$

where the last equality is obtained by applying $\tilde{\mathbf{X}}_1^{\mathsf{T}} = \mathbf{X}_1^{\mathsf{T}}\mathbf{U}$.

For $h(\tilde{\mathbf{Y}}|\mathbf{X}_2,\mathrm{V})$, we use the duality upper bound as before

$$h(\tilde{\mathbf{Y}}|\mathbf{X}_2,\mathrm{V}) = \mathbb{E}\left[-\log p(\tilde{\mathbf{Y}}|\mathbf{X}_2,\mathrm{V})\right] \le \mathbb{E}\left[-\log q(\tilde{\mathbf{Y}}|\mathbf{X}_2,\mathrm{V})\right],$$

where the only difference from the single-user case is the presence of $\mathbf{X}_2$. We choose the auxiliary pdf $q(.)$ as follows. Given $\mathrm{V} = v$, $v \le T-1$, we let $\tilde{\mathbf{Y}}_{[v]} \sim \mathcal{R}(N, \mathbf{I}_N)$, and given $\tilde{\mathbf{Y}}_{[v]}$, the other $\tilde{\mathbf{Y}}_{[i]}$'s are independent and follow

$$\tilde{\mathbf{Y}}_{[i]} \sim \mathcal{R}\left(N, \left(\mathbf{I}_N + \tilde{\mathbf{Y}}_{[v]}\tilde{\mathbf{Y}}_{[v]}^{\mathsf{H}}\right)^{-\frac{1}{2}}\right), \quad i \notin \{v, T\}, \quad (37)$$

$$\tilde{\mathbf{Y}}_{[T]} \sim \mathcal{R}\left(N, \left((1+\|\mathbf{X}_2\|^2)\mathbf{I}_N + \tilde{\mathbf{Y}}_{[v]}\tilde{\mathbf{Y}}_{[v]}^{\mathsf{H}}\right)^{-\frac{1}{2}}\right). \quad (38)$$

$$\mathbb{E}[-\log q(\tilde{\mathbf{Y}}|\mathbf{X}_2,\mathrm{V})] \le (N+T-2)\mathbb{E}\left[\log(1+|\tilde{\mathrm{X}}_{1\mathrm{V}}|^2)\right] + N\mathbb{E}\left[\sum_{i=1,i\neq V}^{T-1}\log\left(1+\frac{|\tilde{\mathrm{X}}_{1i}|^2}{1+|\tilde{\mathrm{X}}_{1\mathrm{V}}|^2}\right)\right]$$
$$+ N\mathbb{E}\left[\log(1+\|\mathbf{X}_2\|^2)\right] + \mathbb{E}\left[\log\left(1+\frac{|\tilde{\mathrm{X}}_{1\mathrm{V}}|^2}{1+\|\mathbf{X}_2\|^2}\right)\right] + N\mathbb{E}\left[\log\left(1+\frac{|\tilde{\mathrm{X}}_{1T}|^2}{1+\|\mathbf{X}_2\|^2+|\tilde{\mathrm{X}}_{1\mathrm{V}}|^2}\right)\right] + O(\log\log P). \quad (39)$$

$$f(\tilde{\mathbf{X}}_1,\mathbf{X}_2) := (N+T-2)\log\left(1+\max_{i=1,\dots,T-1}|\tilde{\mathrm{X}}_{1i}|^2\right) + \log\left(1+\frac{\max_{i=1,\dots,T-1}|\tilde{\mathrm{X}}_{1i}|^2}{1+\|\mathbf{X}_2\|^2}\right)$$
$$+ N\log\left(1+\frac{|\tilde{\mathrm{X}}_{1T}|^2}{1+\|\mathbf{X}_2\|^2+\max_{i=1,\dots,T-1}|\tilde{\mathrm{X}}_{1i}|^2}\right) - N\log\left(1+\sum_{i=1}^{T-1}|\tilde{\mathrm{X}}_{1i}|^2+\frac{|\tilde{\mathrm{X}}_{1T}|^2}{1+\|\mathbf{X}_2\|^2}\right). \quad (41)$$

Proposition 2. *With the above choice of auxiliary output distribution, we obtain the upper bound* (39) *for* $\mathbb{E}\left[-\log q(\tilde{\mathbf{Y}}|\mathbf{X}_2,\mathrm{V})\right]$*, and hence for* $h(\tilde{\mathbf{Y}}|\mathbf{X}_2,\mathrm{V})$.

Proof. See Appendix C. □

Step 3: Upper Bounds on R_1 and R_2: From (33), (35), (36) and (39), we have the bound for R_1

$$TR_1 \le \mathbb{E}\left[f(\tilde{\mathbf{X}}_1,\mathbf{X}_2)\right] + O(\log\log P), \quad (40)$$

where $f(\tilde{\mathbf{X}}_1,\mathbf{X}_2)$ is defined in (41). Following the exact same steps by swapping the users' role,

$$TR_2 \le \mathbb{E}\left[f(\tilde{\mathbf{X}}_2,\mathbf{X}_1)\right] + O(\log\log P), \quad (42)$$

where $\tilde{\mathbf{X}}_2 := \mathbf{X}_2\mathsf{U}_1$ with U_1 from the decomposition

$$\mathbf{X}_1^*\mathbf{X}_1^{\mathsf{T}} = \mathsf{U}_1 \operatorname{diag}(0,\dots,0,\|\mathbf{X}_1\|^2)\,\mathsf{U}_1^{\mathsf{H}}. \quad (43)$$

It follows that, for any $\lambda_1,\lambda_2\ge 0$, we have the following upper bound on the weighted sum rate

$$\begin{aligned}\lambda_1 R_1 + \lambda_2 R_2 &\le \frac{1}{T}\mathbb{E}\left[\lambda_1 f(\tilde{\mathbf{X}}_1,\mathbf{X}_2)+\lambda_2 f(\tilde{\mathbf{X}}_2,\mathbf{X}_1)\right]+O(\log\log P) \quad (44)\\ &\le \frac{1}{T}\sup_{\boldsymbol{x}_1,\boldsymbol{x}_2}[\lambda_1 f(\tilde{\boldsymbol{x}}_1,\boldsymbol{x}_2)+\lambda_2 f(\tilde{\boldsymbol{x}}_2,\boldsymbol{x}_1)]+O(\log\log P), \quad (45)\end{aligned}$$

where the supremum is over all $\boldsymbol{x}_1,\boldsymbol{x}_2$ subject to the peak power constraints $\|\boldsymbol{x}_1\|^2\le P$ and $\|\boldsymbol{x}_2\|^2\le P$.

Step 4: DoF upper bounds: Since we are only interested in the pre-log at high SNR, it is without loss of optimality to let $\|\boldsymbol{x}_1\|^2 = P^{\eta_1}, \|\boldsymbol{x}_2\|^2 = P^{\eta_2}$ for some $\eta_1,\eta_2\le 1$. In addition, we assume that

$$\max_{i=1,\dots,T-1}|\tilde{x}_{1i}|^2 = P^{\bar{\eta}_1}, \quad |\tilde{x}_{1T}|^2 = P^{\eta_{1T}}, \quad (46)$$
$$\max_{i=1,\dots,T-1}|\tilde{x}_{2i}|^2 = P^{\bar{\eta}_2}, \quad |\tilde{x}_{2T}|^2 = P^{\eta_{2T}}. \quad (47)$$

Hence, at high SNR, $\eta_1 = \max\{\bar{\eta}_1,\eta_{1T}\}$, $\eta_2 = \max\{\bar{\eta}_2,\eta_{2T}\}$. From (41) and (45), we have the weighted sum DoF bound

$$\begin{aligned}&\lambda_1 d_1 + \lambda_2 d_2\\ &\le \lambda_1\frac{N+T-2}{T}\bar{\eta}_1 + \lambda_1\frac{1}{T}(\bar{\eta}_1-\eta_2)^+ + \lambda_1\frac{N}{T}(\eta_{1T}-\max\{\bar{\eta}_1,\eta_2\})^+\\ &\quad - \lambda_1\frac{N}{T}\max\{\bar{\eta}_1,\eta_{1T}-\eta_2\}\\ &+\lambda_2\frac{N+T-2}{T}\bar{\eta}_2 + \lambda_2\frac{1}{T}(\bar{\eta}_2-\eta_1)^+ + \lambda_2\frac{N}{T}(\eta_{2T}-\max\{\bar{\eta}_2,\eta_1\})^+\\ &\quad - \lambda_2\frac{N}{T}\max\{\bar{\eta}_2,\eta_{2T}-\eta_1\}, \quad (48)\end{aligned}$$

subject to the constraints $\bar{\eta}_1,\eta_{1T}\le 1$ and $\bar{\eta}_2,\eta_{2T}\le 1$. Taking (λ_1,λ_2) as $\left(1,\frac{1}{T-2}\right)$ or $\left(\frac{1}{T-2},1\right)$, we can verify that, when $3\le N+1\le T$, (17) and (18) hold for all (d_1,d_2) satisfying (48). Thus the optimal DoF region is characterized.

B. The $3\le T\le N$ case

When $T\le N$, the above choice of auxiliary output distribution is not sufficient for a tight DoF outer bound. To see this, let us take $(\lambda_1,\lambda_2)=\left(1,\frac{1}{T-2}\right)$, then if $\bar{\eta}_1+\eta_2\ge\eta_{1T}=1$ and $\eta_2=\bar{\eta}_1$, (48) becomes

$$d_1 + \frac{d_2}{T-2} \le \frac{T-1}{T}\bar{\eta}_1 + \frac{N}{T}(\eta_{1T}-\bar{\eta}_1), \quad (49)$$

which is loose since the right-hand side is larger than $1-\frac{1}{T}$ if $N\ge T$. Generally, the bound (48) can be loose when $\eta_{1T}>\max\{\bar{\eta}_1,\eta_2\}$ or $\eta_{2T}>\max\{\bar{\eta}_2,\eta_1\}$. To account for such scenarios, we ought to refine our choice of auxiliary output distribution for the duality upper bound. First, given $\mathbf{X}_2$, we define a pair of random variables (V,U) as

$$\mathrm{V} = \arg\max_{i=1,2,\dots,T}\frac{|\tilde{\mathrm{X}}_{1i}|^2}{\sigma_i^2}, \quad (50)$$

where $\sigma_i^2 = 1, \forall i<T$ and $\sigma_T^2 = 1+\|\mathbf{X}_2\|^2$, and

$$\mathrm{U} = \begin{cases}1, & \text{if } |\tilde{\mathrm{X}}_{1T}|^2 \ge \max\left\{\max_{i=1,\dots,T-1}|\tilde{\mathrm{X}}_{1i}|^2, 1+\|\mathbf{X}_2\|^2\right\},\\ 0, & \text{otherwise.}\end{cases} \quad (51)$$

$$\mathbb{E}\left[-\log q(\tilde{\mathbf{Y}}|\mathbf{X}_2, \mathrm{V}=v<T, \mathrm{U}=0)\right] \le (N+T-2)\mathbb{E}\left[\log(1+|\tilde{\mathrm{X}}_{1v}|^2)\right] + N\mathbb{E}\left[\log(1+\|\mathbf{X}_2\|^2)\right] + \mathbb{E}\left[\log\left(1+\frac{|\tilde{\mathrm{X}}_{1v}|^2}{1+\|\mathbf{X}_2\|^2}\right)\right] + O(\log\log P). \quad (54)$$

$$\mathbb{E}\left[-\log q(\tilde{\mathbf{Y}}|\mathbf{X}_2, \mathrm{V}=T)\right] \le N\mathbb{E}\left[\log(1+\|\mathbf{X}_2\|^2+|\tilde{\mathrm{X}}_{1T}|^2)\right] + (T-1)\mathbb{E}\left[\log\left(1+\frac{|\tilde{\mathrm{X}}_{1T}|^2}{1+\|\mathbf{X}_2\|^2}\right)\right] + O(\log\log P). \quad (55)$$

$$\mathbb{E}\left[-\log q(\tilde{\mathbf{Y}}|\mathbf{X}_2, \mathrm{V}=v<T, \mathrm{U}=1)\right] \le (N+T-2)\mathbb{E}\left[\log(1+|\tilde{\mathrm{X}}_{1v}|^2)\right] + N\mathbb{E}\left[\log\frac{1+\|\mathbf{X}_2\|^2+|\tilde{\mathrm{X}}_{1T}|^2}{1+\|\mathbf{X}_2\|^2+P}\right] + N\mathbb{E}\left[\log(1+\|\mathbf{X}_2\|^2)\right] + \mathbb{E}\left[\log\left(1+\frac{P}{1+\|\mathbf{X}_2\|^2}\right)\right] + O(\log\log P). \quad (58)$$

Thus, $\tilde{\mathrm{X}}_{1\mathrm{V}}$ is the input entry with the largest instantaneous SNR, and U determines a specific configuration of input entry powers in which the choice of auxiliary output distribution in the previous case possibly fails. Then similarly as for the case $T \ge N+1$, with output rotation, genie-aided bound, and duality upper bound, we have that

$$\begin{aligned} TR_1 &\le I(\tilde{\mathbf{X}}_1; \tilde{\mathbf{Y}}|\mathbf{X}_2) \\ &\le \mathbb{E}\left[-\log q(\tilde{\mathbf{Y}}|\mathbf{X}_2, \mathrm{V}, \mathrm{U})\right] - h(\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}_1, \mathbf{X}_2) + \log(2T), \end{aligned} \quad (52)$$

where $h(\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}_1, \mathbf{X}_2)$ was calculated in (36). For $\mathbb{E}\left[-\log q(\tilde{\mathbf{Y}}|\mathbf{X}_2, \mathrm{V}, \mathrm{U})\right]$, we choose the auxiliary pdf $q(\tilde{\mathbf{Y}}|\mathbf{X}_2, \mathrm{V}, \mathrm{U})$ as follows. Given $\mathrm{V}=v$ and $\mathrm{U}=u$, if $v=T$ or $\{v<T, u=0\}$, we let $\tilde{\mathbf{Y}}_{[v]} \sim \mathcal{R}(N, \mathbf{I}_N)$ and conditioned on $\tilde{\mathbf{Y}}_{[v]}$, the other $\tilde{\mathbf{Y}}_{[i]}$'s are independent and follow

$$\tilde{\mathbf{Y}}_{[i]} \sim \mathcal{R}\left(N, \left(\sigma_i^2 \mathbf{I}_N + \frac{\tilde{\mathbf{Y}}_{[v]}\tilde{\mathbf{Y}}_{[v]}^{\mathsf{H}}}{\sigma_v^2}\right)^{-\frac{1}{2}}\right), \quad i \neq v. \quad (53)$$

This choice is inspired by a training-based scheme in which the input symbol with strongest SNR is used as pilot. After some manipulations similar as for Propositions 1 and 2, we get the bounds (54) and (55). If $\{v<T, u=1\}$, we let $\tilde{\mathbf{Y}}_{[v]} \sim \mathcal{R}(N, \mathbf{I}_N)$ and given $\tilde{\mathbf{Y}}_{[v]}$, the other $\tilde{\mathbf{Y}}_{[i]}$'s are independent with

$$\tilde{\mathbf{Y}}_{[i]} \sim \mathcal{R}\left(N, \left(\mathbf{I}_N + \tilde{\mathbf{Y}}_{[v]}\tilde{\mathbf{Y}}_{[v]}^{\mathsf{H}}\right)^{-\frac{1}{2}}\right), \quad i \notin \{v, T\}, \quad (56)$$

$$\tilde{\mathbf{Y}}_{[T]} \sim \mathcal{R}\left(N, \left((1+\|\mathbf{X}_2\|^2)\mathbf{I}_N + \frac{P}{\|\tilde{\mathbf{Y}}_{[v]}\|^2}\tilde{\mathbf{Y}}_{[v]}\tilde{\mathbf{Y}}_{[v]}^{\mathsf{H}}\right)^{-\frac{1}{2}}\right), \quad (57)$$

where the only difference from (53) is the presence of the factor $\frac{P}{\|\tilde{\mathbf{Y}}_{[v]}\|^2}$. This factor is added to account for the fact that when $u=1$, $|\tilde{\mathrm{X}}_{1v}|^2 < |\tilde{\mathrm{X}}_{1T}|^2$, which can make the power of $\tilde{\mathbf{Y}}_{[v]}$ inferior to that of $\tilde{\mathbf{Y}}_{[T]}$. In this case, we have the bound (58).

These bounds and (36) give us the bound for R_1

$$TR_1 \le \mathbb{E}\left[g(\tilde{\mathbf{X}}_1, \mathbf{X}_2)\right] + O(\log\log P), \quad (59)$$

where $g(\tilde{\mathbf{X}}_1, \mathbf{X}_2)$ is defined in (60), and the similar bound for R_2

$$TR_2 \le \mathbb{E}\left[g(\tilde{\mathbf{X}}_2, \mathbf{X}_1)\right] + O(\log\log P). \quad (61)$$

The rest of the proof follows from a similar weighted sum bound for the rates and the DoFs as done in the previous case.

VI. Conclusion

In this work, we have proposed a new tight outer bound on the DoF region of the two-user non-coherent SIMO MAC with block fading. The outer bound region coincides with the inner bound region achieved by a simple training-based scheme. We expect to extend the results to the general MIMO case.

Appendix

A. Proof of mathematical preliminaries

1) Proof of Lemma 1: Consider the eigen-value decomposition $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}$, where $\boldsymbol{U} \in \mathbb{C}^{m\times m}$ and $\boldsymbol{V} \in \mathbb{C}^{t\times t}$ are unitary matrices, and $\boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}' \\ \mathbf{0}\end{bmatrix}$ with $\boldsymbol{\Sigma}' \in \mathbb{C}^{t\times t}$ a diagonal matrix containing the singular values of $\boldsymbol{A}$. Let $\mathbf{W}' = \mathbf{W}\boldsymbol{U}$, we have

$$\begin{aligned} h(\mathbf{W}\boldsymbol{A}) &= h(\mathbf{W}\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}) && (62)\\ &= h(\mathbf{W}'\boldsymbol{\Sigma}) && (63)\\ &= h(\mathbf{W}'_{[1:t]}\boldsymbol{\Sigma}') && (64)\\ &= h(\mathbf{W}'_{[1:t]}) + n\log|\det(\Sigma')|^2 && (65)\\ &= h(\mathbf{W}'_{[1:t]}) + n\log\det(\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}), && (66) \end{aligned}$$

where the second equality is because rotation does not change differential entropy; (66) follows from a change of variables. Next, it follows from $h(\mathbf{W}'_{[1:t]}) + h(\mathbf{W}'_{[t+1:m]}) \ge h(\mathbf{W}') = h(\mathbf{W}) > -\infty$ that

$$h(\mathbf{W}'_{[1:t]}) > -\infty - h(\mathbf{W}'_{[t+1:m]}) > -\infty, \quad (67)$$

where $h(\mathbf{W}'_{[t+1:m]}) < \infty$ since the average total power of $\mathbf{W}'_{[t+1:m]}$ is bounded by $\mathbb{E}\left[\|\mathbf{W}'\|_{\mathrm{F}}^2\right] = \mathbb{E}\left[\|\mathbf{W}\|_{\mathrm{F}}^2\right] < \infty$. We also have that $h(\mathbf{W}'_{[1:t]}) < \infty$. Therefore, $h(\mathbf{W}'_{[1:t]})$ is bounded by some constant that only depends on the statistics of $\mathbf{W}$. This concludes the proof.

$$g(\tilde{\mathbf{X}}_1,\mathbf{X}_2) := \begin{cases} (T-2)\log\left(1+\max\limits_{i=1,\dots,T-1}|\tilde{\mathrm{X}}_{1i}|^2\right)+\log\left(1+\frac{\max\limits_{i=1,\dots,T-1}|\tilde{\mathrm{X}}_{1i}|^2}{1+\|\mathbf{X}_2\|^2}\right), \\ \qquad \text{if } \frac{|\tilde{\mathrm{X}}_{1T}|^2}{1+\|\mathbf{X}_2\|^2} < \max\limits_{i=1,\dots,T-1}|\tilde{\mathrm{X}}_{1i}|^2 \text{ and } |\tilde{\mathrm{X}}_{1T}|^2 \le \max\left\{\max\limits_{i=1,\dots,T-1}|\tilde{\mathrm{X}}_{1i}|^2, 1+\|\mathbf{X}_2\|^2\right\}, \\ (T-2)\log(1+\max\limits_{i=1,\dots,T-1}|\tilde{\mathrm{X}}_{1i}|^2)+N\log\left(\frac{1+\|\mathbf{X}_2\|^2+|\tilde{\mathrm{X}}_{1T}|^2}{1+\|\mathbf{X}_2\|^2+P}\right)+\log\left(1+\frac{P}{1+\|\mathbf{X}_2\|^2}\right), \\ \qquad \text{if } \frac{|\tilde{\mathrm{X}}_{1T}|^2}{1+\|\mathbf{X}_2\|^2} < \max\limits_{i=1,\dots,T-1}|\tilde{\mathrm{X}}_{1i}|^2 \text{ and } |\tilde{\mathrm{X}}_{1T}|^2 > \max\left\{\max\limits_{i=1,\dots,T-1}|\tilde{\mathrm{X}}_{1i}|^2, 1+\|\mathbf{X}_2\|^2\right\}, \\ (T-1)\log\left(1+\frac{|\tilde{\mathrm{X}}_{1T}|^2}{1+\|\mathbf{X}_2\|^2}\right), \quad \text{if } \frac{|\tilde{\mathrm{X}}_{1T}|^2}{1+\|\mathbf{X}_2\|^2} > \max\limits_{i=1,\dots,T-1}|\tilde{\mathrm{X}}_{1i}|^2. \end{cases} \tag{60}$$

2) Proof of Lemma 2: Let $p(.)$ be the density of X. We introduce an auxiliary distribution with density $q(x) = \left(\frac{1}{\alpha}-1\right)(1+x)^{-1/\alpha}$, $x \ge 0$, with $\alpha < 1$. Then it follows that $h(\mathrm{X}) + \mathbb{E}[\log(q(x))] = -D(p\|q) \le 0$, which yields

$$\mathbb{E}[\log(1+\mathrm{X})] \ge \alpha h(\mathrm{X}) + \alpha\log\left(\frac{1}{\alpha}-1\right). \tag{68}$$

If $\mathbb{E}[\mathrm{X}] \le 1$, then (9) holds readily with $c_0 = -\alpha$.

If $\mathbb{E}[\mathrm{X}] > 1$, we have

$$\begin{aligned} h(\mathrm{X}) &= h\left(\mathbb{E}[\mathrm{X}]\frac{\mathrm{X}}{\mathbb{E}[\mathrm{X}]}\right) &(69)\\ &= \log(\mathbb{E}[\mathrm{X}]) + h\left(\frac{\mathrm{X}}{\mathbb{E}[\mathrm{X}]}\right) &(70)\\ &\ge \log(1+\mathbb{E}[\mathrm{X}]) - 1 + h\left(\frac{\mathrm{X}}{\mathbb{E}[\mathrm{X}]}\right), &(71)\end{aligned}$$

then applying (68), (9) holds with $c_0 = \alpha\log\left(\frac{1}{\alpha}-1\right) - \alpha + \alpha h\left(\frac{\mathrm{X}}{\mathbb{E}[\mathrm{X}]}\right) > -\infty$.

3) Proof of Lemma 3: We prove the lemma by construction. Consider a rate pair (R_1, R_2) achievable with some input pdf $p_{\|\mathbf{X}_1\|^2}(.)$ and $p_{\|\mathbf{X}_2\|^2}(.)$ satisfying the average power constraints P. Let us define a new input distribution with the truncated pdf as

$$p_{\underline{\mathbf{X}}_i}(x) = \begin{cases} \frac{p_{\|\mathbf{X}_i\|^2}(x)}{\Pr(\|\mathbf{X}_i\|^2 < P^\beta)}, & \text{if } x < P^\beta, \\ 0, & \text{if } x \ge P^\beta, \end{cases} \tag{72}$$

for $i = 1, 2$, with $\beta > 1$. For convenience, let us denote the inputs following $p_{\underline{\mathbf{X}}_1}(x)$ and $p_{\underline{\mathbf{X}}_2}(x)$ as $\underline{\mathbf{X}}_1$ and $\underline{\mathbf{X}}_2$, respectively. Then $\underline{\mathbf{X}}_1$ and $\underline{\mathbf{X}}_2$ satisfy the peak power constraint P^β. Similarly, we define $\bar{\mathbf{X}}_1$ and $\bar{\mathbf{X}}_2$ with pdf

$$p_{\bar{\mathbf{X}}_i}(x) = \begin{cases} \frac{p_{\|\mathbf{X}_i\|^2}(x)}{\Pr(\|\mathbf{X}_i\|^2 \ge P^\beta)}, & \text{if } x \ge P^\beta, \\ 0, & \text{if } x < P^\beta. \end{cases} \tag{73}$$

Clearly, $\mathbf{X}_i$ equals $\underline{\mathbf{X}}_i$ if $\|\mathbf{X}_i\|^2 < P^\beta$ and $\bar{\mathbf{X}}_i$ otherwise. We define the random variable V as

$$\mathrm{V} = \begin{cases} 0, & \text{if } \|\mathbf{X}_1\|^2 = \|\underline{\mathbf{X}}_1\|^2 \text{ and } \|\mathbf{X}_2\|^2 = \|\underline{\mathbf{X}}_2\|^2, \\ 1, & \text{otherwise.} \end{cases} \tag{74}$$

By Markov's inequality,

$$\begin{aligned} \Pr(\|\mathbf{X}_i\|^2 = \|\bar{\mathbf{X}}_i\|^2) &= \Pr(\|\mathbf{X}_i\|^2 \ge P^\beta) \\ &\le \frac{\mathbb{E}\left[\|\mathbf{X}_i\|^2\right]}{P^\beta} \le TP^{1-\beta}, \quad i = 1, 2, \end{aligned} \tag{75}$$

then

$$\begin{aligned} \Pr(\mathrm{V}=1) &= 1 - \Pr(\|\mathbf{X}_1\|^2 = \|\underline{\mathbf{X}}_1\|^2)\Pr(\|\mathbf{X}_2\|^2 = \|\underline{\mathbf{X}}_2\|^2) \\ &\le 1 - \left(1 - TP^{1-\beta}\right)^2 \le 2TP^{1-\beta}. \end{aligned} \tag{76}$$

Let the genie give V to the receiver, we have that

$$\begin{aligned} TR_1 &\le I(\mathbf{X}_1;\mathbf{Y}|\mathbf{X}_2) &(77)\\ &\le I(\mathbf{X}_1;\mathbf{Y},\mathrm{V}|\mathbf{X}_2) &(78)\\ &= I(\mathbf{X}_1;\mathbf{Y}|\mathbf{X}_2,\mathrm{V}) + I(\mathbf{X}_1;\mathrm{V}|\mathbf{X}_2) &(79)\\ &\le \Pr(\mathrm{V}=0)I(\mathbf{X}_1;\mathbf{Y}|\mathbf{X}_2,\mathrm{V}=0) \\ &\qquad + \Pr(\mathrm{V}=1)I(\mathbf{X}_1;\mathbf{Y}|\mathbf{X}_2,\mathrm{V}=1) + 1, &(80)\\ &\le I(\underline{\mathbf{X}}_1;\mathbf{Y}|\underline{\mathbf{X}}_2) + \Pr(\mathrm{V}=1)I(\mathbf{X}_1;\mathbf{Y}|\mathbf{X}_2,\mathrm{V}=1) + 1, &(81)\end{aligned}$$

where (80) is due to $I(\mathbf{X}_1;\mathrm{V}|\mathbf{X}_2) \le H(\mathrm{V}) \le 1$ bits. Next, since removing noise and giving CSI increase the rate,

$$\begin{aligned} &\Pr(\mathrm{V}=1)I(\mathbf{X}_1;\mathbf{Y}|\mathbf{X}_2,\mathrm{V}=1) \\ &\le \Pr(\mathrm{V}=1)I(\mathbf{X}_1;\mathbf{H}_1\mathbf{X}_1^{\mathsf{T}}+\mathbf{Z}|\mathbf{H}_1,\mathrm{V}=1) &(82)\\ &\le N\Pr(\mathrm{V}=1)\log(1+\mathbb{E}\left[\|\mathbf{X}_1\|^2|\mathrm{V}=1\right]) &(83)\\ &\le N\Pr(\mathrm{V}=1)\log\left(1+\frac{P}{\Pr(\|\mathbf{X}_1\|^2 \ge P^\beta)}\right) &(84)\\ &\le N\Pr(\mathrm{V}=1)\log(1+P) \\ &\qquad - N\Pr(\mathrm{V}=1)\log\Pr(\|\mathbf{X}_1\|^2 \ge P^\beta) &(85)\\ &= O(P^{1-\beta}\log P^\beta). &(86)\end{aligned}$$

where (84) is because

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{X}_1\|^2|\mathrm{V}=1\right] &\le \mathbb{E}\left[\|\bar{\mathbf{X}}_1\|^2\right] = \frac{\int_{P^\beta}^\infty x p_{\|\mathbf{X}_1\|^2}(x)dx}{\Pr(\|\mathbf{X}_1\|^2 \ge P^\beta)} \\ &\le \frac{\int_0^\infty x p_{\|\mathbf{X}_1\|^2}(x)dx}{\Pr(\|\mathbf{X}_1\|^2 \ge P^\beta)} \le \frac{P}{\Pr(\|\mathbf{X}_1\|^2 \ge P^\beta)}, \end{aligned} \tag{87}$$

and the last equality follows from (75) and (76). Plugging this into (81) yields

$$TR_1 \le I(\underline{\mathbf{X}}_1;\mathbf{Y}|\underline{\mathbf{X}}_2) + O(P^{1-\beta}\log P^\beta). \tag{88}$$

Following the same steps by swapping the users' role, we get the bound for R_2

$$TR_2 \le I(\underline{\mathbf{X}}_2;\mathbf{Y}|\underline{\mathbf{X}}_1) + O(P^{1-\beta}\log P^{\beta}). \tag{89}$$

Using similar techniques, we can also show that

$$T(R_1+R_2) \le I(\underline{\mathbf{X}}_1,\underline{\mathbf{X}}_2;\mathbf{Y}) + O(P^{1-\beta}\log P^{\beta}). \tag{90}$$

Therefore, there exists (R_1', R_2') satisfying

$$R_1'+R_2' \le \frac{1}{T} I(\underline{\mathbf{X}}_1,\underline{\mathbf{X}}_2;\mathbf{Y}), \tag{91}$$

$$R_1' \le \frac{1}{T} I(\underline{\mathbf{X}}_1;\mathbf{Y}|\underline{\mathbf{X}}_2), \tag{92}$$

$$R_2' \le \frac{1}{T} I(\underline{\mathbf{X}}_2;\mathbf{Y}|\underline{\mathbf{X}}_1), \tag{93}$$

i.e., achievable with the constructed inputs $\underline{\mathbf{X}}_1$ and $\underline{\mathbf{X}}_2$ satisfying the peak power constraint P^{β}, such that (11) holds. This concludes the proof.

4) Proof of Lemma 4: In this proof, all expectations are implicitly w.r.t. $\mathcal{P}$. A direct calculation from (14) yields

$$\mathbb{E}\left[-\log(r_{\mathbf{Y}}(\mathbf{Y}))\right] = -\log|\det\boldsymbol{A}|^2 + (N-\alpha)\mathbb{E}\left[\log\|\boldsymbol{A}\mathbf{Y}\|^2\right] + \frac{\mathbb{E}\left[\|\boldsymbol{A}\mathbf{Y}\|^2\right]}{\beta} + \log\Gamma(\alpha) + \log\beta^{\alpha} + \log\frac{\pi^N}{\Gamma(N)}. \tag{94}$$

When $\beta = \mathbb{E}\left[\|\boldsymbol{A}\mathbf{Y}\|^2\right]$ and $\alpha = \frac{1}{\log(\beta)} = \frac{1}{\log(\mathbb{E}[\|\boldsymbol{A}\mathbf{Y}\|^2])}$, this becomes

$$\begin{aligned}
&\mathbb{E}\left[-\log(r_{\mathbf{Y}}(\mathbf{Y}))\right] \\
&= -\log|\det\boldsymbol{A}|^2 + N\mathbb{E}\left[\log\|\boldsymbol{A}\mathbf{Y}\|^2\right] - \frac{\mathbb{E}\left[\log\|\boldsymbol{A}\mathbf{Y}\|^2\right]}{\log(\mathbb{E}\left[\|\boldsymbol{A}\mathbf{Y}\|^2\right])} \\
&\quad + \log\Gamma\left(\frac{1}{\log(\mathbb{E}\left[\|\boldsymbol{A}\mathbf{Y}\|^2\right])}\right) + \log\frac{e\pi^N}{\Gamma(N)}, && (95)\\
&= -\log|\det\boldsymbol{A}|^2 + N\mathbb{E}\left[\log\|\boldsymbol{A}\mathbf{Y}\|^2\right] \\
&\quad + O(\log\log(\mathbb{E}\left[\|\boldsymbol{A}\mathbf{Y}\|^2\right])), && (96)
\end{aligned}$$

where the last equality is because $0 < \frac{\mathbb{E}\left[\log\|\boldsymbol{A}\mathbf{Y}\|^2\right]}{\log(\mathbb{E}[\|\boldsymbol{A}\mathbf{Y}\|^2])} < 1$ and

$$\log\Gamma\left(\frac{1}{\log(\mathbb{E}\left[\|\boldsymbol{A}\mathbf{Y}\|^2\right])}\right) - \log\log(\mathbb{E}\left[\|\boldsymbol{A}\mathbf{Y}\|^2\right]) \to 0, \tag{97}$$

as $\mathbb{E}\left[\|\boldsymbol{A}\mathbf{Y}\|^2\right] \to \infty$ due to

$$\begin{aligned}
\lim_{x\to\infty}\log\Gamma\left(\frac{1}{x}\right) - \log x &= \lim_{x\to\infty}\log\left(\frac{1}{x}\Gamma\left(\frac{1}{x}\right)\right) \\
&= \lim_{x\to\infty}\log\left(\Gamma\left(1+\frac{1}{x}\right)\right) \\
&= \log(\Gamma(1)) \\
&= 0. && (98)
\end{aligned}$$

B. Proof of Proposition 1

Using Lemma 4, it follows that

$$\begin{aligned}
&\mathbb{E}\left[-\log q(\mathbf{Y}|\mathrm{V}=v)\right] \\
&= N\mathbb{E}\left[\log\|\mathbf{Y}_{[v]}\|^2\right] + \sum_{i=1,i\neq v}^{T}\mathbb{E}\Bigg[\log\det\left(\mathbf{I}_N + \mathbf{Y}_{[v]}\mathbf{Y}_{[v]}^{\mathsf{H}}\right) \\
&\quad + N\log\left\|\left(\mathbf{I}_N+\mathbf{Y}_{[v]}\mathbf{Y}_{[v]}^{\mathsf{H}}\right)^{-\frac{1}{2}}\mathbf{Y}_{[i]}\right\|^2\Bigg] + O(\log\log P) && (99)\\
&= N\mathbb{E}\left[\log\|\mathbf{Y}_{[v]}\|^2\right] + \sum_{i=1,i\neq v}^{T}\mathbb{E}\Bigg[\log\left(1+\|\mathbf{Y}_{[v]}\|^2\right) \\
&\quad + N\log\left(\|\mathbf{Y}_{[i]}\|^2 - \frac{|\mathbf{Y}_{[i]}^{\mathsf{H}}\mathbf{Y}_{[v]}|^2}{1+\|\mathbf{Y}_{[v]}\|^2}\right)\Bigg] + O(\log\log P) && (100)\\
&= (N+T-1)\mathbb{E}\left[\log(1+\|\mathbf{Y}_{[v]}\|^2)\right] \\
&\quad + N\sum_{i=1,i\neq v}^{T}\mathbb{E}\Big[\log(\|\mathbf{Y}_{[i]}\|^2 + \|\mathbf{Y}_{[i]}\|^2\|\mathbf{Y}_{[v]}\|^2 - |\mathbf{Y}_{[i]}^{\mathsf{H}}\mathbf{Y}_{[v]}|^2) \\
&\qquad - \log(1+\|\mathbf{Y}_{[v]}\|^2)\Big] + O(\log\log P), && (101)
\end{aligned}$$

where in the second equality, we used the identities $\det(\mathbf{I}+\boldsymbol{u}\boldsymbol{v}^{\mathsf{H}}) = 1+\boldsymbol{v}^{\mathsf{H}}\boldsymbol{u}$, $\det(c\boldsymbol{A}) = c^n\det(\boldsymbol{A})$ for $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, and $\|(\boldsymbol{A}+\boldsymbol{u}\boldsymbol{v}^{\mathsf{H}})^{-1/2}\boldsymbol{x}\|^2 = \boldsymbol{x}^{\mathsf{H}}(\boldsymbol{A}+\boldsymbol{u}\boldsymbol{v}^{\mathsf{H}})^{-1}\boldsymbol{x} = \boldsymbol{x}^{\mathsf{H}}\left(\boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathsf{H}}\boldsymbol{A}^{-1}}{1+\boldsymbol{v}^{\mathsf{H}}\boldsymbol{A}^{-1}\boldsymbol{u}}\right)\boldsymbol{x}$.

By expanding $\mathbf{Y}_{[1]},\dots,\mathbf{Y}_{[T]}$, we get that, given $\mathbf{X}$,

$$\mathbb{E}_{\mathbf{H},\mathbf{Z}}[\|\mathbf{Y}_{[i]}\|^2] = N(1+|\mathrm{X}_i|^2), \quad \forall i, \tag{102}$$

and

$$\begin{aligned}
&\mathbb{E}_{\mathbf{H},\mathbf{Z}}[\|\mathbf{Y}_{[i]}\|^2\|\mathbf{Y}_{[v]}\|^2 - |\mathbf{Y}_{[i]}^{\mathsf{H}}\mathbf{Y}_{[v]}|^2] \\
&\quad = (N^2-N)(1+|\mathrm{X}_v|^2+|\mathrm{X}_i|^2), \quad i\neq v. && (103)
\end{aligned}$$

Then, using Jensen's inequality and Lemma 2 (by letting α arbitrarily close to 1), we get that

$$\begin{aligned}
&\mathbb{E}\left[-\log q(\mathbf{Y}|\mathrm{V}=v)\right] \\
&\le (N+T-1)\mathbb{E}\left[\log(1+N+N|\mathrm{X}_v|^2)\right] \\
&\quad + N\sum_{i=1,i\neq \mathrm{v}}^{T}\mathbb{E}\left[\log\frac{N+N|\mathrm{X}_i|^2+(N^2-N)(1+|\mathrm{X}_v|^2+|\mathrm{X}_i|^2)}{1+N+N|\mathrm{X}_v|^2}\right] \\
&\quad + O(\log\log P) && (104)\\
&= (N+T-1)\mathbb{E}\left[\log(1+|\mathrm{X}_v|^2)\right] \\
&\quad + N\sum_{i=1,i\neq v}^{T}\mathbb{E}\left[\log\left(1+\frac{|\mathrm{X}_i|^2}{1+|\mathrm{X}_v|^2}\right)\right] + O(\log\log P). && (105)
\end{aligned}$$

Taking expectation over V, we obtain (29), which concludes the proof.

C. Proof of Proposition 2

Hence, we obtain from Lemma 4,

$$\begin{aligned}
&\mathbb{E}\left[-\log(q(\tilde{\mathbf{Y}}|\mathbf{X}_2, \mathrm{V}=v)\right] \\
&= N\mathbb{E}\left[\log\|\tilde{\mathbf{Y}}_{[v]}\|^2\right] + \sum_{i=1,i\neq v}^{T-1} \mathbb{E}\left[\log\det\left(\mathbf{I}_N + \tilde{\mathbf{Y}}_{[v]}\tilde{\mathbf{Y}}_{[v]}^{\mathsf{H}}\right)\right] \\
&\quad + N\sum_{i=1,i\neq v}^{T-1} \mathbb{E}\left[\log\left\|\left(\mathbf{I}_N + \tilde{\mathbf{Y}}_{[v]}\tilde{\mathbf{Y}}_{[v]}^{\mathsf{H}}\right)^{-\frac{1}{2}}\tilde{\mathbf{Y}}_{[i]}\right\|^2\right] \\
&\quad + \mathbb{E}\left[\log\det\left((1+\|\mathbf{X}_2\|^2)\mathbf{I}_N + \tilde{\mathbf{Y}}_{[v]}\tilde{\mathbf{Y}}_{[v]}^{\mathsf{H}}\right)\right] \\
&\quad + N\mathbb{E}\left[\log\left\|\left((1+\|\mathbf{X}_2\|^2)\mathbf{I}_N + \tilde{\mathbf{Y}}_{[v]}\tilde{\mathbf{Y}}_{[v]}^{\mathsf{H}}\right)^{-\frac{1}{2}}\tilde{\mathbf{Y}}_{[T]}\right\|^2\right] \\
&\quad + O(\log\log P) \qquad (106) \\
&\leq N\mathbb{E}\left[\log(1+|\tilde{\mathrm{X}}_{1v}|^2)\right] + \sum_{i=1,i\neq v}^{T} B_i + O(\log\log P), \qquad (107)
\end{aligned}$$

where

$$B_i := \mathbb{E}\left[\log\left(1+\|\tilde{\mathbf{Y}}_{[v]}\|^2\right) + N\log\left(\|\tilde{\mathbf{Y}}_{[i]}\|^2 - \frac{|\tilde{\mathbf{Y}}_{[i]}^{\mathsf{H}}\tilde{\mathbf{Y}}_{[v]}|^2}{1+\|\tilde{\mathbf{Y}}_{[v]}\|^2}\right)\right],$$

for $i \notin \{v, T\}$, and

$$\begin{aligned}
B_T := \mathbb{E}\Bigg[&\log\left((1+\|\mathbf{X}_2\|^2)^N\left(1+\frac{\|\tilde{\mathbf{Y}}_{[v]}\|^2}{1+\|\mathbf{X}_2\|^2}\right)\right) \\
&+ N\log\left(\frac{1}{1+\|\mathbf{X}_2\|^2}\left(\|\tilde{\mathbf{Y}}_{[T]}\|^2 - \frac{|\tilde{\mathbf{Y}}_{[T]}^{\mathsf{H}}\tilde{\mathbf{Y}}_{[v]}|^2}{1+\|\mathbf{X}_2\|^2+\|\tilde{\mathbf{Y}}_{[v]}\|^2}\right)\right)\Bigg].
\end{aligned}$$

By expanding $\tilde{\mathbf{Y}}_{[1]}, \ldots, \tilde{\mathbf{Y}}_{[T]}$, we get that, given $\mathbf{X}_1$ and $\mathbf{X}_2$,

$$\begin{aligned}
&\mathbb{E}_{\mathbf{H}_1,\mathbf{Z}}[\|\tilde{\mathbf{Y}}_{[v]}\|^2\|\tilde{\mathbf{Y}}_{[i]}\|^2 - |\tilde{\mathbf{Y}}_{[i]}^{\mathsf{H}}\tilde{\mathbf{Y}}_{[v]}|^2] \\
&= (N^2-N)\left(1+|\tilde{\mathrm{X}}_{1v}|^2+|\tilde{\mathrm{X}}_{1i}|^2\right), \quad i \notin \{v,T\}, \qquad (108)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}_{\mathbf{H}_1,\mathbf{Z}}[\|\tilde{\mathbf{Y}}_{[v]}\|^2\|\tilde{\mathbf{Y}}_{[T]}\|^2 - |\tilde{\mathbf{Y}}_{[T]}^{\mathsf{H}}\tilde{\mathbf{Y}}_{[v]}|^2] \\
&\quad = (N^2-N)\left((1+\|\mathbf{X}_2\|^2)(1+|\tilde{\mathrm{X}}_{1v}|^2)+|\tilde{\mathrm{X}}_{1T}|^2\right) \\
&\quad \leq (N^2-N)(1+\|\mathbf{X}_2\|^2)(1+|\tilde{\mathrm{X}}_{1v}|^2+|\tilde{\mathrm{X}}_{1T}|^2). \qquad (109)
\end{aligned}$$

Then, applying repeatedly Lemma 2 (by letting α arbitrarily close to 1), and Jensen's inequality,

$$\begin{aligned}
B_i &= \mathbb{E}\left[-(N-1)\log\left(1+\|\tilde{\mathbf{Y}}_{[v]}\|^2\right)\right] \\
&\quad + N\mathbb{E}\left[\log\left(\|\tilde{\mathbf{Y}}_{[i]}\|^2 + \|\tilde{\mathbf{Y}}_{[i]}\|^2\|\tilde{\mathbf{Y}}_{[v]}\|^2 - |\tilde{\mathbf{Y}}_{[i]}^{\mathsf{H}}\tilde{\mathbf{Y}}_{[v]}|^2\right)\right] \\
&\leq \mathbb{E}\left[-(N-1)\log(1+N+N|\tilde{\mathrm{X}}_{1v}|^2)\right] \\
&\quad + N\mathbb{E}\left[\log(N^2(1+|\tilde{\mathrm{X}}_{1i}|^2) + (N^2-N)|\tilde{\mathrm{X}}_{1v}|^2)\right] + O(1), \\
&= \mathbb{E}\left[\log(1+|\tilde{\mathrm{X}}_{1v}|^2)\right] + N\mathbb{E}\left[\log\left(1+\frac{|\tilde{\mathrm{X}}_{1i}|^2}{1+|\tilde{\mathrm{X}}_{1v}|^2}\right)\right] + O(1) \\
&\qquad (110)
\end{aligned}$$

for $i \notin \{v, T\}$, and

$$\begin{aligned}
B_T &= N\mathbb{E}\left[\log(1+\|\mathbf{X}_2\|^2)\right] + \mathbb{E}\left[\log\left(1+\frac{\|\tilde{\mathbf{Y}}_v\|^2}{1+\|\mathbf{X}_2\|^2}\right)\right] \\
&\quad + N\mathbb{E}\left[\log\left(\|\tilde{\mathbf{Y}}_{[T]}\|^2 + \frac{\|\tilde{\mathbf{Y}}_{[v]}\|^2\|\tilde{\mathbf{Y}}_{[T]}\|^2 - |\tilde{\mathbf{Y}}_{[T]}^{\mathsf{H}}\tilde{\mathbf{Y}}_{[v]}|^2}{1+\|\mathbf{X}_2\|^2}\right)\right] \\
&\quad - N\mathbb{E}\left[\log\left(1+\|\mathbf{X}_2\|^2+\|\tilde{\mathbf{Y}}_v\|^2\right)\right] \qquad (111) \\
&\leq N\mathbb{E}\left[\log(1+\|\mathbf{X}_2\|^2)\right] + \mathbb{E}\left[\log\left(1+\frac{N+N|\tilde{\mathrm{X}}_{1v}|^2}{1+\|\mathbf{X}_2\|^2}\right)\right] \\
&\quad + N\mathbb{E}\Big[\log\Big(N(1+\|\mathbf{X}_2\|^2) \\
&\qquad\qquad + (N^2-N)(1+|\tilde{\mathrm{X}}_{1v}|^2) + N^2|\tilde{\mathrm{X}}_{1T}|^2\Big)\Big] \\
&\quad - N\mathbb{E}\left[\log(1+\|\mathbf{X}_2\|^2+N+N|\tilde{\mathrm{X}}_{1v}|^2)\right] + O(1) \qquad (112) \\
&= N\mathbb{E}\left[\log(1+\|\mathbf{X}_2\|^2)\right] + \mathbb{E}\left[\log\left(1+\frac{|\tilde{\mathrm{X}}_{1v}|^2}{1+\|\mathbf{X}_2\|^2}\right)\right] \\
&\quad + N\mathbb{E}\left[\log\left(1+\frac{|\tilde{\mathrm{X}}_{1T}|^2}{1+\|\mathbf{X}_2\|^2+|\tilde{\mathrm{X}}_{1v}|^2}\right)\right] + O(1) \qquad (113)
\end{aligned}$$

Plugging (110) and (113) into (107) then taking expectation over V, we obtain (39), which concludes the proof.

References

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless personal communications*, vol. 6, no. 3, pp. 311–335, 1998.

[2] I. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, pp. 585–595, Nov./Dec. 1999.

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

[4] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.

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

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

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

[8] M. Fadel and A. Nosratinia, "Coherence disparity in broadcast and multiple access channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7383–7401, Dec. 2016.

[9] K. H. Ngo, S. Yang, and M. Guillaud, "An achievable DoF region for the two-user non-coherent MIMO broadcast channel with statistical CSI," in *2017 IEEE Information Theory Workshop (ITW)*, Nov. 2017, pp. 604–608.

[10] F. Zhang, M. Fadel, and A. Nosratinia, "Spatially correlated MIMO broadcast channel: Analysis of overlapping correlation eigenspaces," in *2017 IEEE International Symposium on Information Theory (ISIT)*, Jun. 2017, pp. 1097–1101.

[11] A. El Gamal and Y.-H. Kim, *Network Information Theory*. New York, NY, USA: Cambridge University Press, 2011.

[12] S. Yang and S. Shamai (Shitz), "On the multiplexing gain of discrete-time MIMO phase noise channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2394–2408, Apr. 2017.