



A Constant-Gap Result on the Multi-Antenna Broadcast Channels with Linearly Precoded Rate Splitting

Sheng Yang, Zheng Li

► To cite this version:

Sheng Yang, Zheng Li. A Constant-Gap Result on the Multi-Antenna Broadcast Channels with Linearly Precoded Rate Splitting. 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2018), Jun 2018, Kalamata, Greece. 10.1109/SPAWC.2018.8446000 . hal-01912264

HAL Id: hal-01912264

<https://hal.science/hal-01912264>

Submitted on 10 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Constant-Gap Result on the Multi-Antenna Broadcast Channels with Linearly Precoded Rate Splitting

Sheng Yang[‡] and Zheng Li^{‡§}

[‡] L2S, CentraleSupélec-CNRS-Université Paris-Sud, 91192, Gif-sur-Yvette, France

[§]Orange Labs Networks, 92326, Châtillon, France

Email:{sheng.yang, zheng.li}@centralesupelec.fr

Abstract—In this paper, we consider a simple downlink channel with a multi-antenna base station and two single-antenna receivers. We assume that the channel is deterministic and known to all the nodes. Our contribution is two-fold. First, we show that linear precoding with private streams can have unbounded gap to the capacity of the channel. Second, we show that using rate-splitting with a simple power allocation one can achieve the sum capacity to within a constant gap for any channel realization.

I. INTRODUCTION

The capacity region of a broadcast channel (BC, also known as downlink channel) is still unknown in general [1]. In the particular case of degraded BC, in which the channel can be somehow ordered, the problem is solved and a simple superposition coding is capacity achieving [2]. Although the multi-antenna BC with Gaussian noise is not degraded, its capacity region has also been found [3], [4] and can be achieved with dirty paper coding (DPC) [5]. The implementation of DPC is however not trivial, due to its non-linear nature and the fact that it is sensitive to the channel state information at the transmitter side (CSIT) [6]. As such, linear precoding is used in most practical systems instead. Apart from the simplicity, it can be shown that linear precoding schemes such as zero-forcing achieve the maximum degrees of freedom (DoF) of the system [7], [8].

In this paper, we first show that *any* linear precoding scheme can be far from optimal, since the gap between the achievable sum rate of the best linear scheme and the sum capacity can be unbounded. Then, we show that by introducing rate-splitting and a common stream, one can achieve the sum capacity to within a constant number of bits for any channel realization. The precoding of both the common and private streams is still linear.

The idea of using rate-splitting to partially mitigate interference has been introduced in [9], [10] for the interference channels. Essentially, each transmitter splits the individual message into private and common parts, in which the common parts are decodable by (though not intended to) all receivers. Decodable interference can therefore be completely “removed”. It has been shown in [11] that such a scheme achieves the capacity region of the two-user interference channel to within one bit. For broadcast channels, rate splitting has been exploited in [12] in a broadcast channel with imperfect CSIT. Extensions have been made in later works [13], [14], [15]. In the current

work, our main contribution is to establish the optimality of linearly precoded rate splitting in the constant gap sense.

The rest of the paper is organized as follows. The problem is described formally in Section II. The achievable rate of linear precoding schemes is studied in Section III, whereas that of rate splitting is derived in Section IV. Finally, we conclude the paper with some remarks on possible extensions.

II. PROBLEM STATEMENT

Notation: In this paper, we use the following notational conventions. For random quantities, we use upper case non-italic letters, e.g., X , for scalars, upper case non-italic bold letters, e.g., \mathbf{V} , for vectors, and upper case letter with bold and sans serif fonts, e.g., \mathbf{M} , for matrices. Deterministic quantities are denoted in a rather conventional way with italic letters, e.g., a scalar x , a vector \mathbf{v} , and a matrix \mathbf{M} . Logarithms are in base 2. The Euclidean norm of a vector and a matrix is denoted by $\|\mathbf{v}\|$ and $\|\mathbf{M}\|$, respectively.

A. Channel model

We consider a two-user multiple-input single-output (MISO) broadcast channel (BC). For simplicity, we assume that the number of transmit antennas is $n_t = 2$. We assume that the channel matrix is drawn from the set of generic matrices and is known globally. Let us fix the channel realization to \mathbf{H} , and the channel output at time t is $\mathbf{Y}_k[t] = \mathbf{h}_k^T \mathbf{x}[t] + Z_k[t]$, $k = 1, 2$, or in a compact form

$$\begin{bmatrix} Y_1[t] \\ Y_2[t] \end{bmatrix} = \mathbf{H} \mathbf{x}[t] + \mathbf{Z}[t], \quad t = 1, \dots, n,$$

where $\mathbf{Z}[t] \sim \mathcal{CN}(0, \mathbf{I}_2)$ is the temporally i.i.d. additive white Gaussian noise (AWGN) in time. The input sequence is subject to the power constraint $\frac{1}{n} \sum_{t=1}^n \|\mathbf{x}[t]\|^2 \leq P$.

B. Sum capacity

The capacity region of this channel is well known and can be achieved with dirty paper coding. The capacity region can also be characterized with the dual multiple access channel (MAC) with sum power constraint. In particular, the sum capacity is

$$C_{\text{sum}} = \max_{\text{tr}(\mathbf{\Lambda}) \leq P} \log \det(\mathbf{I} + \mathbf{H}^H \mathbf{\Lambda} \mathbf{H}), \quad (1)$$

where $\mathbf{\Lambda}$ is diagonal. Since we have

$$\log \det(\mathbf{I} + P \mathbf{H}^H \mathbf{H}) - 2 \leq C_{\text{sum}} \leq \log \det(\mathbf{I} + P \mathbf{H}^H \mathbf{H}),$$

for our purpose, it is without loss of optimality to consider the quantity $\log \det(\mathbf{I} + P\mathbf{H}^H\mathbf{H})$ as the capacity as we are only interested in capacity to within a constant gap. Further,

$$\begin{aligned} \log \det(\mathbf{I} + P\mathbf{H}^H\mathbf{H}) \\ = \log(1 + P\|\mathbf{h}_1\|^2 + P\|\mathbf{h}_2\|^2 + P^2\det(\mathbf{H}\mathbf{H}^H)), \end{aligned}$$

which is within $\log 3$ bits/s/Hz to

$$\max \left\{ \log(1 + P\|\mathbf{h}_1\|^2), \log(1 + P\|\mathbf{h}_2\|^2), \log(1 + P^2\det(\mathbf{H}\mathbf{H}^H)) \right\}. \quad (2)$$

Note that the first two terms in (2) can be achieved with single-user transmission, by choosing the stronger user. Therefore, the only non-trivial case is when $\log(1 + P^2\det(\mathbf{H}\mathbf{H}^H))$ is the dominating term in (2).

III. LINEAR PRECODING

Now let us restrict ourselves to linear precoding schemes at the transmitter and treating interference as noise at the receivers. In particular, we let $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$ such that $\mathbb{E}[\mathbf{X}_1\mathbf{X}_1^H] = \mathbf{A}$ and $\mathbb{E}[\mathbf{X}_2\mathbf{X}_2^H] = \mathbf{B}$ with the following eigenvalue decompositions

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} u_A & v_A \\ \tilde{v}_A & \tilde{u}_A \end{bmatrix} \begin{bmatrix} \lambda_A & \\ & \mu_A \end{bmatrix} \begin{bmatrix} u_A^* & \tilde{v}_A^* \\ v_A^* & \tilde{u}_A^* \end{bmatrix}, \\ \mathbf{B} &= \begin{bmatrix} u_B & v_B \\ \tilde{v}_B & \tilde{u}_B \end{bmatrix} \begin{bmatrix} \lambda_B & \\ & \mu_B \end{bmatrix} \begin{bmatrix} u_B^* & \tilde{v}_B^* \\ v_B^* & \tilde{u}_B^* \end{bmatrix}, \end{aligned} \quad (3)$$

where $|\tilde{u}_A|^2 = |u_A|^2 = 1 - |\tilde{v}_A|^2 = 1 - |v_A|^2$ and $\lambda_A \geq \mu_A \geq 0$ without loss of generality; same convention is applied for \mathbf{B} .

While the general case shall be considered later, let us first assume that the channel matrix has the following triangular form

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ f & g \end{bmatrix}, \quad (4)$$

where the normalization can be done by scaling the transmit power; hence, $\mathbf{h}_1 = [1 \ 0]^T$ and $\mathbf{h}_2 = [f \ g]^T$. In this case, the sum rate (2) becomes

$$C_{\text{sum}} \approx \max \left\{ \log(1 + P), \log(1 + P|f|^2 + P|g|^2), \log(1 + P^2|g|^2) \right\}, \quad (5)$$

where we use “ \approx ” for constant-gap approximation.

Assuming Gaussian signaling for the linearly precoded signals \mathbf{X}_1 and \mathbf{X}_2 , we have

$$\begin{aligned} R_1 &= \log \left(1 + \frac{\|\mathbf{h}_1\|_{\mathbf{A}}^2}{1 + \|\mathbf{h}_1\|_{\mathbf{B}}^2} \right) = \log \left(1 + \frac{A_{11}}{1 + B_{11}} \right), \\ R_2 &= \log \left(1 + \frac{\|\mathbf{h}_2\|_{\mathbf{B}}^2}{1 + \|\mathbf{h}_2\|_{\mathbf{A}}^2} \right), \end{aligned} \quad (6)$$

where we use the notation $\|\mathbf{h}_k\|_{\mathbf{A}}^2 = \mathbf{h}_k^T \mathbf{A} \mathbf{h}_k$ and $\|\mathbf{h}_k\|_{\mathbf{B}}^2 = \mathbf{h}_k^T \mathbf{B} \mathbf{h}_k$. Note that we are only interested in the case with

$$\frac{A_{11}}{1 + B_{11}} \geq 1 \quad \text{and} \quad \frac{\|\mathbf{h}_2\|_{\mathbf{B}}^2}{1 + \|\mathbf{h}_2\|_{\mathbf{A}}^2} \geq 1, \quad (7)$$

for otherwise it is equivalent to the single user case to within a constant gap. In this case, the achievable sum rate with linear precoding can be written as

$$\begin{aligned} R_1 + R_2 &\approx \log \left(\frac{A_{11}}{1 + B_{11}} \right) + \log \left(\frac{\|\mathbf{h}_2\|_{\mathbf{B}}^2}{1 + \|\mathbf{h}_2\|_{\mathbf{A}}^2} \right) \\ &= \log \left(\frac{A_{11}}{1 + \|\mathbf{h}_2\|_{\mathbf{A}}^2} \right) + \log \left(\frac{\|\mathbf{h}_2\|_{\mathbf{B}}^2}{1 + B_{11}} \right). \end{aligned} \quad (8)$$

We can now maximize over \mathbf{A} and over \mathbf{B} separately. In fact, one can show the following lemma.

Lemma 1. For any \mathbf{A} and \mathbf{B} in (3), we can show that

$$\begin{aligned} \frac{A_{11}}{1 + \|\mathbf{h}_2\|_{\mathbf{A}}^2} &\leq 2 \min \left\{ \frac{2}{|f|^2} + 2 \frac{|g|^2}{|f|^2} \lambda_A, \lambda_A \right\}, \\ \frac{\|\mathbf{h}_2\|_{\mathbf{B}}^2}{1 + B_{11}} &\leq 2|f|^2 + 2|g|^2 \lambda_B. \end{aligned}$$

Proof. Let us first consider the part with \mathbf{A} .

$$\frac{A_{11}}{1 + \|\mathbf{h}_2\|_{\mathbf{A}}^2} = \frac{|u_A|^2 \lambda_A + |v_A|^2 \mu_A}{|f u_A + g \tilde{v}_A|^2 \lambda_A + |f v_A + g \tilde{u}_A|^2 \mu_A + 1}.$$

Since $|f u_A + g \tilde{v}_A|^2 + |f v_A + g \tilde{u}_A|^2 = |f|^2 + |g|^2$, we use $a \leq b$ to denote the ordered version of $|f u_A + g \tilde{v}_A|^2$ and $|f v_A + g \tilde{u}_A|^2$ and have $a \geq (|f u_A| - |g \tilde{v}_A|)^2$ and $2b \geq |f|^2 + |g|^2$. Using the fact that $\lambda_A \geq \mu_A$, we have the following upper bound

$$\begin{aligned} \frac{A_{11}}{1 + \|\mathbf{h}_2\|_{\mathbf{A}}^2} &\leq \frac{|u_A|^2 \lambda_A + |v_A|^2 \mu_A}{a \lambda_A + b \mu_A + 1} \\ &\leq \frac{|u_A|^2 \lambda_A}{a \lambda_A + 1} + \frac{|v_A|^2 \mu_A}{b \mu_A + 1} \\ &\leq \frac{|u_A|^2 \lambda_A}{(|f u_A| - |g \tilde{v}_A|)^2 \lambda_A + 1} + \frac{\mu_A}{\frac{|f|^2 + |g|^2}{2} \mu_A + 1} \\ &\leq \max_{u_A: |f u_A| \geq |g \tilde{v}_A|} \frac{|u_A|^2 \lambda_A}{(|f u_A| - |g \tilde{v}_A|)^2 \lambda_A + 1} \\ &\quad + \min \left\{ \left(\frac{|f|^2 + |g|^2}{2} \right)^{-1}, \mu_A \right\} \end{aligned} \quad (9)$$

$$\begin{aligned} &\leq \max_{u_A: |f u_A| \geq |g \tilde{v}_A|} \min \left\{ \frac{|u_A|^2}{(|f u_A| - |g \tilde{v}_A|)^2}, |u_A|^2 \lambda_A \right\} \\ &\quad + \min \left\{ \left(\frac{|f|^2 + |g|^2}{2} \right)^{-1}, \mu_A \right\} \end{aligned} \quad (10)$$

$$\begin{aligned} &\leq \min \left\{ \left(\frac{1}{\sqrt{\lambda_A}} + |g| \right)^2 \frac{\lambda_A}{|f|^2}, \lambda_A \right\} \\ &\quad + \min \left\{ \left(\frac{|f|^2 + |g|^2}{2} \right)^{-1}, \mu_A \right\} \end{aligned} \quad (11)$$

$$\begin{aligned} &\leq \min \left\{ \frac{2}{|f|^2} + 2 \frac{|g|^2}{|f|^2} \lambda_A, \lambda_A \right\} \\ &\quad + \min \left\{ \left(\frac{|f|^2 + |g|^2}{2} \right)^{-1}, \mu_A \right\} \\ &\leq 2 \min \left\{ \frac{2}{|f|^2} + 2 \frac{|g|^2}{|f|^2} \lambda_A, \lambda_A \right\}, \end{aligned} \quad (12)$$

where (9) is from the fact that the objective function is increasing with $|fu_A|$ when $|fu_A| \leq |g\tilde{v}_A|$; since in (10), $\frac{|u_A|^2}{(|fu_A| - |g\tilde{v}_A|)^2}$ is decreasing with $|u_A|$ and $|u_A|^2\lambda_A$ is increasing with $|u_A|$, the max-min is attained when both terms are equalized or when $|u_A| = 1$; (11) is indeed an upper bound of (10). The second part can be shown as follows.

$$\begin{aligned}
& \frac{\|\mathbf{h}_2\|_B^2}{1 + B_{11}} \\
&= \frac{|fu_B + g\tilde{v}_B|^2\lambda_B + |fv_B + g\tilde{u}_B|^2\mu_B}{|u_B|^2\lambda_B + |v_B|^2\mu_B + 1} \\
&\leq 2 \frac{|f|^2(|u_B|^2\lambda_B + |v_B|^2\mu_B) + |g|^2(|\tilde{v}_B|^2\lambda_B + |\tilde{u}_B|^2\mu_B)}{|u_B|^2\lambda_B + |v_B|^2\mu_B + 1} \\
&\leq 2|f|^2 \frac{|u_B|^2\lambda_B + |v_B|^2\mu_B}{|u_B|^2\lambda_B + |v_B|^2\mu_B + 1} \\
&\quad + 2|g|^2 \frac{\lambda_B}{|u_B|^2\lambda_B + |v_B|^2\mu_B + 1} \\
&\leq 2|f|^2 + 2|g|^2\lambda_B.
\end{aligned} \tag{13}$$

□

We are now ready to provide our first main result.

Proposition 1. *The gap between the achievable rate by linear precoding (LP) and the capacity is not bounded.*

Proof. We prove the statement with a specific example. We consider high SNR P and let the channel coefficients scale with P as $f = P^{\alpha_f}$ and $g = P^{\alpha_g}$ for some $\alpha_f, \alpha_g \in \mathbb{R}$. It follows that the achievable sum rate also scales with P as $d_{\text{LP}}(\alpha_f, \alpha_g) \log P + O(1)$. Similarly, the sum capacity scales as $d_{\text{DPC}}(\alpha_f, \alpha_g) \log P + O(1)$. In the literature, the pre-log factor is known as the generalized degree of freedom (GDoF). If we can show that $d_{\text{LP}}(\alpha_f, \alpha_g) < d_{\text{DPC}}(\alpha_f, \alpha_g)$ for some (α_f, α_g) , then the rate gap is necessarily unbounded when $P \rightarrow \infty$.

To that end, we let $\alpha_f > \alpha_g > \alpha_f - \frac{1}{2} \geq 0$. With this setting, (12) scales as $P^{1+2\alpha_g-2\alpha_f}$ and (13) scales as $P^{1+2\alpha_g}$. It follows that $d_{\text{LP}} \leq 2 + 4\alpha_g - 2\alpha_f$. From (5), we verify that $d_{\text{DPC}} = \max\{1, 1 + 2\alpha_f, 1 + 2\alpha_g, 2 + 2\alpha_g\} = 2 + 2\alpha_g$. Thus, we have shown that $d_{\text{DPC}} > d_{\text{LP}}$ for such (α_f, α_g) . □

Remark III.1. *When the channel matrix \mathbf{H} is bounded, the linear precoding schemes do achieve the optimal DoF as shown in [7], [8]. However, DoF is a coarse measure that only characterizes the prelog of the channel capacity when the channel gains are bounded. If one allows the channel gains to grow with the transmit power polynomially, then we obtain a finer measure from the prelog of the channel capacity, which is known as the generalized DoF (GDoF). The constant-gap optimality is stronger than GDoF and DoF since we allow arbitrary channel gains without any constraint. As revealed in the above example, although even ZF precoding can achieve the optimal DoF, no linear precoding schemes can be constant-gap optimal.*

Now, let us consider the general case with non-triangular channel matrix. We can transform the channel as follows. Assuming $\mathbf{h}_1 \neq 0$, we use the LQ decomposition

$$\mathbf{H} = \mathbf{L}\mathbf{Q} = \begin{bmatrix} \|\mathbf{h}_1\| & 0 \\ f\|\mathbf{h}_1\| & g\|\mathbf{h}_1\| \end{bmatrix} \mathbf{Q}, \tag{14}$$

where \mathbf{Q} is unitary and $(|f|^2 + |g|^2)\|\mathbf{h}_1\|^2 = \|\mathbf{h}_2\|^2$. Then, the received signal becomes

$$\mathbf{H}\mathbf{x} = \begin{bmatrix} 1 & 0 \\ f & g \end{bmatrix} \|\mathbf{h}_1\| \mathbf{Q}\mathbf{x} = \begin{bmatrix} 1 & 0 \\ f & g \end{bmatrix} \tilde{\mathbf{x}}, \tag{15}$$

where we define $\tilde{\mathbf{x}} := \|\mathbf{h}_1\| \mathbf{Q}\mathbf{x}$. Note that the original channel is equivalent to the new one from $\tilde{\mathbf{x}}$ to \mathbf{y} , since \mathbf{Q} is unitary and $\|\mathbf{h}_1\| \neq 0$ and both are known globally. The only difference is now the sum power constraint becomes $\tilde{P} = P\|\mathbf{h}_1\|^2$. Same arguments apply if we swap the roles of user 1 and user 2. We can obtain the following results.

Proposition 2. *The achievable rate of linear precoding is upper bounded, to within a constant gap, by*

$$\begin{aligned}
& \max \left\{ \log(1 + P\|\mathbf{h}_1\|^2), \log(1 + P\|\mathbf{h}_2\|^2), \right. \\
& \quad \left. \log(1 + \beta_\rho P^2 \det(\mathbf{H}\mathbf{H}^H)) \right\},
\end{aligned}$$

where $\beta_\rho := \min \left\{ \frac{1-\rho^2}{\rho^2}, 1 \right\}$ with $\rho := \left| \frac{\mathbf{h}_1^H \mathbf{h}_2}{\|\mathbf{h}_1\| \|\mathbf{h}_2\|} \right|$.

Proof. First, let us apply (14) and (15), and we obtain the triangular form (4) with a new power constraint $\tilde{P} = P\|\mathbf{h}_1\|^2$. It can be verified that

$$|f| = \frac{\|\mathbf{h}_2\|}{\|\mathbf{h}_1\|} \rho, \quad |g| = \frac{\|\mathbf{h}_2\|}{\|\mathbf{h}_1\|} \sqrt{1 - \rho^2},$$

where $\rho := \left| \frac{\mathbf{h}_1^H \mathbf{h}_2}{\|\mathbf{h}_1\| \|\mathbf{h}_2\|} \right|$. Let us define \mathcal{S} as the set of \mathbf{A} and \mathbf{B} such that (7) is satisfied, and $\bar{\mathcal{S}}$ the complementary set. Then, applying Lemma 1 on (8), we have

$$\begin{aligned}
& \max_{(\mathbf{A}, \mathbf{B}) \in \mathcal{S}} R_1 + R_2 \\
& \lesssim \log \left(4 \min \left\{ 2 + 2\tilde{P}(|g|^2 + \frac{|g|^2}{|f|^2}) + 2\tilde{P}^2 \frac{|g|^4}{|f|^2}, \right. \right. \\
& \quad \left. \left. \tilde{P}|f|^2 + \tilde{P}^2|g|^2 \right\} \right) \\
& = \log \left(4 \min \left\{ 2 + 2P(1 - \rho^2) \left(\|\mathbf{h}_2\|^2 + \frac{\|\mathbf{h}_1\|^2}{\rho^2} \right) \right. \right. \\
& \quad \left. \left. + 2P^2 \frac{1 - \rho^2}{\rho^2} \det(\mathbf{H}\mathbf{H}^H), P\rho^2\|\mathbf{h}_2\|^2 + P^2 \det(\mathbf{H}\mathbf{H}^H) \right\} \right) \\
& \leq \log \left(4 \left\{ 2 + 2P(\|\mathbf{h}_2\|^2 + \|\mathbf{h}_1\|^2) + 2P^2\beta_\rho \det(\mathbf{H}\mathbf{H}^H) \right\} \right) \\
& \approx \max \left\{ \log(1 + P\|\mathbf{h}_1\|^2), \log(1 + P\|\mathbf{h}_2\|^2), \right. \\
& \quad \left. \log(1 + \beta_\rho P^2 \det(\mathbf{H}\mathbf{H}^H)) \right\},
\end{aligned} \tag{16}$$

where in the first inequality we used $\lambda_A, \lambda_B \leq \tilde{P} := P\|\mathbf{h}_1\|^2$; (16) can be verified by looking at both cases $\frac{1-\rho^2}{\rho^2} \geq 1$. For the set of \mathbf{A} and \mathbf{B} that do not satisfy (7), we have, from (6)

$$\max_{(\mathbf{A}, \mathbf{B}) \in \bar{\mathcal{S}}} R_1 + R_2$$

$$\leq \max \{1 + \log(1 + P\|\mathbf{h}_1\|^2), 1 + \log(1 + P\|\mathbf{h}_2\|^2)\}$$

$$\approx \max \{\log(1 + P\|\mathbf{h}_1\|^2), \log(1 + P\|\mathbf{h}_2\|^2)\}. \quad (18)$$

The proof is complete by combining both (17) and (18). \square

IV. RATE SPLITTING

Instead of sending only private messages, the transmitter can send both common and private messages and let the users “share” the common message. In this way, the potential interference in the previous case can be converted to common message that is decoded by both users. Intuitively, we somehow let each user decode the interference and remove it to improve the achievable rate. In the following, we show that a simple precoding scheme combined with rate splitting can achieve the capacity to within a constant gap.

Specifically, we let $\mathbf{X} = \mathbf{X}_0 + \mathbf{X}_1 + \mathbf{X}_2$ such that $\mathbb{E}[\mathbf{X}_k \mathbf{X}_k^H] = \mathbf{Q}_k$, $k = 0, 1, 2$, with

$$\mathbf{Q}_1 = P_1 \left(\mathbf{I} - \frac{\mathbf{h}_2^* \mathbf{h}_2^T}{\|\mathbf{h}_2\|^2} \right), \quad \mathbf{Q}_2 = P_2 \left(\mathbf{I} - \frac{\mathbf{h}_1^* \mathbf{h}_1^T}{\|\mathbf{h}_1\|^2} \right). \quad (19)$$

In other words, the private streams \mathbf{X}_1 and \mathbf{X}_2 are precoded with zero-forcing. At the receiver side, each user decodes *jointly* its private stream and the common stream as in a two-user MAC channel. It readily follows that the rate triple (R_c, R_{p1}, R_{p2}) is achievable if

$$\begin{aligned} R_{p1} &\leq \log(1 + \mathbf{h}_1^T \mathbf{Q}_1 \mathbf{h}_1^*) \\ R_{p2} &\leq \log(1 + \mathbf{h}_2^T \mathbf{Q}_2 \mathbf{h}_2^*) \\ R_c + R_{p1} &\leq \log(1 + \mathbf{h}_1^T (\mathbf{Q}_1 + \mathbf{Q}_0) \mathbf{h}_1^*) \\ R_c + R_{p2} &\leq \log(1 + \mathbf{h}_2^T (\mathbf{Q}_2 + \mathbf{Q}_0) \mathbf{h}_2^*) \\ R_c &\leq \min \{ \log(1 + \mathbf{h}_1^T \mathbf{Q}_0 \mathbf{h}_1^*), \log(1 + \mathbf{h}_2^T \mathbf{Q}_0 \mathbf{h}_2^*) \}. \end{aligned}$$

Let $R_c = R_{c1} + R_{c2}$ and $R_1 = R_{p1} + R_{c1}$ and $R_2 = R_{p2} + R_{c2}$. We can use Fourier-Motzkin elimination [16] to obtain the following region on (R_1, R_2) .

$$R_1 \leq \min \{ \log(1 + \mathbf{h}_1^T \mathbf{Q}_1 \mathbf{h}_1^*) + \log(1 + \mathbf{h}_2^T \mathbf{Q}_0 \mathbf{h}_2^*), \log(1 + \mathbf{h}_1^T (\mathbf{Q}_1 + \mathbf{Q}_0) \mathbf{h}_1^*) \} \quad (20)$$

$$R_2 \leq \min \{ \log(1 + \mathbf{h}_2^T \mathbf{Q}_2 \mathbf{h}_2^*) + \log(1 + \mathbf{h}_1^T \mathbf{Q}_0 \mathbf{h}_1^*), \log(1 + \mathbf{h}_2^T (\mathbf{Q}_2 + \mathbf{Q}_0) \mathbf{h}_2^*) \} \quad (21)$$

$$\begin{aligned} R_1 + R_2 &\leq \min \{ \log(1 + \mathbf{h}_2^T \mathbf{Q}_2 \mathbf{h}_2^*) \\ &\quad + \log(1 + \mathbf{h}_1^T (\mathbf{Q}_1 + \mathbf{Q}_0) \mathbf{h}_1^*), \\ &\quad \log(1 + \mathbf{h}_1^T \mathbf{Q}_1 \mathbf{h}_1^*) + \log(1 + \mathbf{h}_2^T (\mathbf{Q}_2 + \mathbf{Q}_0) \mathbf{h}_2^*) \}. \end{aligned} \quad (22)$$

From the above region, it is not hard to verify that (22) provides the achievable sum rate $R_{\text{sum}}^{\text{RS}} := R_1 + R_2$ since it dominates the sum of (20) and (21). Let $P_1 = P_2 = \frac{P}{3}$ and $\mathbf{Q}_0 = \frac{P}{3} \mathbf{I}_2$, we have

$$\begin{aligned} R_{\text{sum}}^{\text{RS}} &= \min \left\{ \log(1 + \mathbf{h}_1^T \mathbf{Q}_1 \mathbf{h}_1^*) + \log \left(1 + \frac{P}{3} \|\mathbf{h}_2\|^2 + \mathbf{h}_2^T \mathbf{Q}_2 \mathbf{h}_2^* \right), \right. \\ &\quad \left. \log(1 + \mathbf{h}_2^T \mathbf{Q}_2 \mathbf{h}_2^*) + \log \left(1 + \frac{P}{3} \|\mathbf{h}_1\|^2 + \mathbf{h}_1^T \mathbf{Q}_1 \mathbf{h}_1^* \right) \right\}. \end{aligned} \quad (23)$$

Finally, we can prove the second main result of our paper.

Proposition 3. *The selection between single-user transmission and the proposed rate splitting (RS) scheme achieves the sum-capacity (1) to within a constant gap.*

Proof. With single-user transmission, one can achieve the first two terms in (2). It is enough to show that the proposed rate splitting scheme achieves the third term in (2). Indeed, it is easy to verify that

$$\begin{aligned} &\log(1 + \mathbf{h}_1^T \mathbf{Q}_1 \mathbf{h}_1^*) + \log \left(1 + \frac{P}{3} \|\mathbf{h}_2\|^2 + \mathbf{h}_2^T \mathbf{Q}_2 \mathbf{h}_2^* \right) \\ &\geq \log \left(1 + \frac{P}{3} \mathbf{h}_1^T \mathbf{Q}_1 \mathbf{h}_1^* \|\mathbf{h}_2\|^2 \right) \\ &= \log \left(1 + \frac{P^2}{9} \det(\mathbf{H} \mathbf{H}^H) \right) \end{aligned} \quad (24)$$

$$\begin{aligned} &\log(1 + \mathbf{h}_2^T \mathbf{Q}_2 \mathbf{h}_2^*) + \log \left(1 + \frac{P}{3} \|\mathbf{h}_1\|^2 + \mathbf{h}_1^T \mathbf{Q}_1 \mathbf{h}_1^* \right) \\ &\geq \log \left(1 + \frac{P}{3} \mathbf{h}_2^T \mathbf{Q}_2 \mathbf{h}_2^* \|\mathbf{h}_1\|^2 \right) \\ &= \log \left(1 + \frac{P^2}{9} \det(\mathbf{H} \mathbf{H}^H) \right), \end{aligned} \quad (25)$$

where (24) and (25) are from the choices of \mathbf{Q}_1 and \mathbf{Q}_2 in (19). Therefore, plugging the above individual terms back to (23), we have

$$\begin{aligned} R_{\text{sum}}^{\text{RS}} &\geq \log \left(1 + \frac{P^2}{9} \det(\mathbf{H} \mathbf{H}^H) \right) \\ &\geq \log(1 + P^2 \det(\mathbf{H} \mathbf{H}^H)) - \log 9, \end{aligned}$$

which completes the proof. \square

V. CONCLUSION

We have investigated the two-user multiple-input single-output broadcast channel. We have shown that, using rate splitting combined with linear precoding, the sum capacity can be achieved to within a constant gap. Remarkably, using linear precoding alone can be far from optimal, with an unbounded gap to the capacity.

Although it is not considered in the current paper, we believe that the constant gap result holds for the whole capacity region. One should be able to prove the result using the dual MAC region, instead of the dual MAC sum rate (1). Another possible extension is for the case with multi-antenna receivers, for which we believe that the constant gap result continues to hold. Finally, it would be interesting to extend the result to the general multi-user MIMO case, which seems to be highly non-trivial with the increasing number of messages in the system.

REFERENCES

- [1] T. Cover, “Broadcast channels,” *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 2–14, 1972.
- [2] R. G. Gallager, “Capacity and coding for degraded broadcast channels,” *Problemy Peredachi Informatsii*, vol. 10, no. 3, pp. 3–14, 1974.
- [3] G. Caire and S. Shamai, “On the achievable throughput of a multiantenna Gaussian broadcast channel,” *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

- [4] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.
- [5] M. Costa, "Writing on dirty paper (corresp.)," *IEEE transactions on information theory*, vol. 29, no. 3, pp. 439–441, 1983.
- [6] S. Yang and J.-C. Belfiore, "The impact of channel estimation error on the DPC region of the two-user Gaussian broadcast channel," in *Proc. Allerton*, 2005.
- [7] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on selected areas in communications*, vol. 24, no. 3, pp. 528–541, 2006.
- [8] J. Lee and N. Jindal, "High SNR analysis for MIMO broadcast channels: Dirty paper coding versus linear precoding," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4787–4792, 2007.
- [9] A. Carleial, "Interference channels," *IEEE Transactions on Information Theory*, vol. 24, no. 1, pp. 60–70, 1978.
- [10] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.
- [11] R. H. Etkin, N. David, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Transactions on information theory*, vol. 54, no. 12, pp. 5534–5562, 2008.
- [12] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, "Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 315–328, 2013.
- [13] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, "A rate splitting strategy for massive MIMO with imperfect CSIT," *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 4611–4624, 2016.
- [14] A. G. Davoodi and S. A. Jafar, "GDoF of the MISO BC: Bridging the gap between finite precision CSIT and perfect CSIT," in *Information Theory (ISIT), 2016 IEEE International Symposium on*. IEEE, 2016, pp. 1297–1301.
- [15] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting multiple access for downlink communication systems: Bridging, generalizing and outperforming SDMA and NOMA," *CoRR*, vol. abs/1710.11018, 2017. [Online]. Available: <http://arxiv.org/abs/1710.11018>
- [16] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.