

# An Achievable DoF Region for the Two-User Non-Coherent MIMO Broadcast Channel with Statistical CSI

Ngo Khac-Hoang, Sheng Yang, Maxime Guillaud

► **To cite this version:**

Ngo Khac-Hoang, Sheng Yang, Maxime Guillaud. An Achievable DoF Region for the Two-User Non-Coherent MIMO Broadcast Channel with Statistical CSI. 2017 IEEE Information Theory Workshop (ITW 2017), Nov 2017, Kaohsiung, Taiwan. pp.604-608, 10.1109/itw.2017.8277972 . hal-01567036

**HAL Id: hal-01567036**

**<https://hal-centralesupelec.archives-ouvertes.fr/hal-01567036>**

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.

# An Achievable DoF Region for the Two-User Non-Coherent MIMO Broadcast Channel with Statistical CSI

Khac-Hoang Ngo<sup>\*†</sup>, Sheng Yang<sup>†</sup>, Maxime Guillaud<sup>\*</sup>

<sup>\*</sup>Mathematical and Algorithmic Sciences Lab, France Research Center, Huawei Technologies France SASU, 92100 Boulogne-Billancourt, France

<sup>†</sup>LSS, CentraleSupélec, 91190 Gif-sur-Yvette, France

Email: {ngo.khac.hoang, maxime.guillaud}@huawei.com, sheng.yang@centralesupelec.fr

**Abstract**—In this paper, we study the two-user non-coherent multiple-input multiple-output broadcast channels with spatially correlated Rayleigh block fading. We propose a scheme to exploit the statistical channel state information, namely, the knowledge of the covariance matrix, and derive the corresponding achievable degrees of freedom region. The main idea of the proposed scheme is based on rate-splitting, additive superposition coding, and channel training. Our result shows that statistical CSI can play an important role in enhancing the degrees of freedom of a non-coherent broadcast channel.

## I. INTRODUCTION

Since the pioneering works [1], [2] in the late 90s, the multiple-input multiple-output (MIMO) technology has become an attractive and efficient solution to exploit the extra spatial degrees of freedom (DoF). Under the ideal assumption that the channel matrix is well conditioned and known to either end of the channel, it can be shown that the capacity of a point-to-point MIMO channel scales linearly with the number of antennas as  $C \sim \min\{M, N\} \log \text{SNR}$  when the signal-to-noise ratio (SNR) is large, where  $M$  and  $N$  are the numbers of transmit and receive antennas, respectively. The degrees of freedom, defined as the pre-log of the capacity at high SNR, is  $\min\{M, N\}$  in this case.

In practice, however, the channel matrix is *a priori* unknown and varies over time. Communication without *a priori* channel state information (CSI), also known as *non-coherent* communication, has been investigated in a large body of work (see, e.g., [3], [4], [5], [6], [7] and the references therein). In particular, it was shown in [4] that when the channel varies fast enough, the channel capacity scales as  $\log \log \text{SNR} + \chi(\mathbf{H}) + o(1)$  where  $\chi(\mathbf{H})$  is the so-called fading number of the channel. For non-coherent *block fading* MIMO Rayleigh channels with coherence time  $T$ , it was shown [3], [5], [7] that the DoF is  $M^*(1 - M^*/T)$  where  $M^* := \min\{M, N, \lfloor \frac{T}{2} \rfloor\}$ . Remarkably, the optimal DoF can be achieved either by carefully designed space-time modulations, or by training-based strategies [8].

Non-coherent multi-user communications, such as the broadcast channels (BC) and the multiple access channels (MAC), have been studied in the block fading case (see, e.g., [9] and the references therein). For the MAC, although the exact DoF

region is still unknown, it has been shown that the optimal sum DoF can be achieved with a training-based scheme [9]. For the BC, the exact DoF region is known with *isotropic* Rayleigh fading, since it is a special case of stochastically degraded broadcast channel. It has been shown that time division multiple access (TDMA) is DoF optimal in this case [9].

In this work, we are interested in the broadcast channels with spatially correlated (i.e., non-isotropic) fading. Note that such channels are not degraded in general and the capacity region is unknown. Our setting is motivated by the fact that spatial correlation is actually present in most communication systems where the transmit antennas are co-located. Furthermore, it is a reasonable assumption that the covariance matrices can be tracked and known perfectly since the time-variation of the spatial correlation is usually slow. The main purpose of this work is to understand how the statistical knowledge can be exploited to improve the DoF of a non-coherent MIMO BC. To that end, we study the two-user case and propose a transmission scheme based on rate-splitting, additive superposition coding, and channel training. Specifically, the private signals are linearly precoded and superimposed over the common signal in such a way that the private signal does not cause any interference to the unintended receiver whereas the common signal can be decoded by both users. Then, at each receiver both the private and common signals can be decoded as in a non-coherent MAC channel with a training-based scheme. As the main result of this paper, an achievable DoF region of this BC is derived. We show that substantial DoF gain can be obtained over the case without statistical CSI for which TDMA is optimal. To the best of our knowledge, this work is the first one to investigate the non-coherent BC with spatial correlation, although the point-to-point case has been previously studied in [6].

The remainder of the paper is organized as follows. The system model and assumptions are presented in Section II and the main result on achievable DoF region of the BC is stated in Section III. In Section IV, we describe the proposed scheme and derive the DoF region. Finally, we conclude the paper in Section V.

Throughout the paper, we use the following notational conventions. 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.,  $\mathbf{M}$ , for matrices. Deterministic quantities are denoted with italic letters, e.g., a scalar  $x$ , a vector  $\mathbf{v}$ , and a matrix  $\mathbf{M}$ . The transpose and conjugated transpose of  $\mathbf{M}$  are  $\mathbf{M}^\top$  and  $\mathbf{M}^\text{H}$ , respectively.

## II. SYSTEM MODEL

We consider a MIMO broadcast channel consisting of a  $M$ -antenna base station (BS) and two users: user 1 with  $N_1$  antennas and user 2 with  $N_2$  antennas. The channel between the BS and user  $k$  is flat- and block-fading with equal-length and synchronous coherence interval of  $T$  slots. That is, the channel propagation matrix  $\mathbf{H}_k \in \mathbb{C}^{N_k \times M}$ ,  $k = 1, 2$ , remains constant during each block of length  $T$  symbols and changes to an independent value in the next block. Let the matrix  $\mathbf{X}[b] \in \mathbb{C}^{M \times T}$  be the transmitted signal from the  $M$  antennas during the coherence interval  $b$ . Then, the received signal matrix at user  $k$  during interval  $b$ ,  $b = 1, \dots, B$ , is

$$\mathbf{Y}_k[b] = \mathbf{H}_k[b]\mathbf{X}[b] + \sqrt{\frac{M}{\rho}}\mathbf{Z}_k[b], \quad k = 1, 2, \quad (1)$$

where  $\mathbf{Z}_k[b] \in \mathbb{C}^{N_k \times T}$  is the additive white Gaussian noise with independent and identically distributed (i.i.d.)  $\mathcal{CN}(0, 1)$  entries, independent across users and coherence intervals. The input is subject to the power constraint  $\frac{1}{B} \sum_{b=1}^B \|\mathbf{X}[b]\|^2 \leq MT$ , and thus  $\rho$  is ratio between the average transmit power per antenna and the noise power. Hereafter, we refer to  $\rho$  as the SNR of the channel.

**Correlation structure:** We model the channel matrix as

$$\mathbf{H}_k[b] = \mathbf{W}_k[b] \mathbf{R}_k^{\frac{1}{2}}, \quad k = 1, 2, \quad (2)$$

where  $\mathbf{W}_k[b] \in \mathbb{C}^{N_k \times r_k}$  has i.i.d.  $\mathcal{CN}(0, 1)$  entries and is unknown;  $\mathbf{R}_k^{\frac{1}{2}} \in \mathbb{C}^{r_k \times M}$  captures the spatial correlation at the transmitter side<sup>1</sup> and is assumed to be known and deterministic during the whole communication. We assume  $r_k \leq \lfloor T/2 \rfloor$ . For simplicity, we omit the index  $b$  when confusion is not likely.

We factorize the correlation matrix  $\mathbf{R}_k = \frac{1}{N_k} \mathbb{E}[\mathbf{H}_k^\text{H} \mathbf{H}_k]$  using eigendecomposition

$$\mathbf{R}_k = \mathbf{U}_k^\text{H} \mathbf{\Sigma}_k \mathbf{U}_k, \quad k = 1, 2, \quad (3)$$

where  $\mathbf{\Sigma}_k$  is a  $r_k \times r_k$  diagonal matrix whose elements are the non-zero eigenvalues of  $\mathbf{R}_k$ , and  $\mathbf{U}_k$  is a  $r_k \times M$  orthogonal matrix whose unit row vectors are the eigenvectors of  $\mathbf{R}_k$  corresponding to the non-zero eigenvalues. Then we can write  $\mathbf{R}_k^{\frac{1}{2}} = \mathbf{\Sigma}_k^{\frac{1}{2}} \mathbf{U}_k$ ,  $k = 1, 2$ . The column vectors of  $\mathbf{H}_k$  belong to a  $r_k$ -dimensional subspace  $\mathcal{V}_k$  of  $\mathbb{C}^M$  spanned by the rows of  $\mathbf{U}_k$ . The two channel subspaces  $\mathcal{V}_1$  and  $\mathcal{V}_2$  can be overlapping and the intersection  $\mathcal{V}_1 \cap \mathcal{V}_2$  is (if not empty) also a subspace of  $\mathbb{C}^M$ . We assume that  $\mathcal{V}_1 \cap \mathcal{V}_2$  is  $r_0$ -dimensional. The dimension of the union subspace  $\mathcal{V}_1 \cup \mathcal{V}_2$  is  $r_1 + r_2 - r_0 =: M' \leq M$ , which is also the rank of the aggregate channel matrix  $[\mathbf{H}_1^\text{H} \ \mathbf{H}_2^\text{H}]^\text{T}$ .

<sup>1</sup>The correlation at the receiver side is ignored since it can be removed at each receiver.

If the rate pair  $(R_1(\rho), R_2(\rho))$  is achievable at SNR  $\rho$ ,  $\forall \rho \geq 0$ , i.e., lies within the capacity region of the channel, then we say that  $(d_1, d_2)$  is an achievable DoF pair with

$$d_k := \liminf_{\rho \rightarrow \infty} \frac{R_k(\rho)}{\log_2(\rho)}, \quad k = 1, 2. \quad (4)$$

## III. MAIN RESULTS

The main finding of this paper is an achievable DoF region of the channel described above, as stated in Theorem 1.

**Theorem 1.** *Let us consider a two-user non-coherent MIMO broadcast channel with statistical CSI as described in Section II, and assume without loss of generality that  $r_2 \geq r_1$ . The following DoF pairs  $(d_1, d_2)$  can be achieved:*

$$D_1 = \left( r_1 \left( 1 - \frac{r_1}{T} \right), 0 \right), \quad (5)$$

$$D_2 = \left( 0, r_2 \left( 1 - \frac{r_2}{T} \right) \right), \quad (6)$$

$$D_3 = \left( (r_1 - r_0) \left( 1 - \frac{r_1 - r_0}{T} \right), (r_2 - r_0) \left( 1 - \frac{r_2 - r_0}{T} \right) \right), \quad (7)$$

$$D_4 = \left( (r_1 - r_0) \left( 1 - \frac{r_1}{T} \right) + r_0 \left( 1 - \frac{r_2}{T} \right), (r_2 - r_0) \left( 1 - \frac{r_2}{T} \right) \right), \quad (8)$$

$$D_5 = \left( (r_1 - r_0) \left( 1 - \frac{r_1}{T} \right), r_2 \left( 1 - \frac{r_2}{T} \right) \right). \quad (9)$$

The convex hull of these five points and the origin in the  $d_1 - d_2$  plane is an achievable region for this channel.

In Fig. 1, we illustrate the achievable DoF region in Theorem 1 for a fixed coherence time  $T = 25$  in two different cases:  $r_1 = r_2$  and  $r_1 < r_2$ . In general, the region enlarges when  $r_0$  descends from  $\min\{r_1, r_2\}$  to 0.

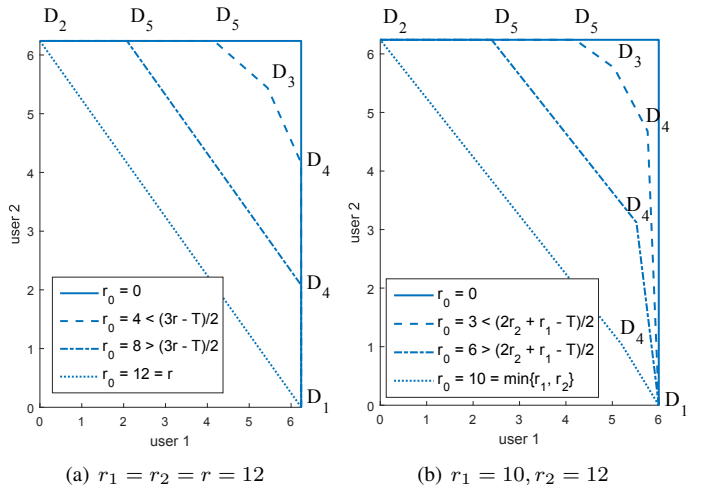


Fig. 1. The achievable DoF region of two-user non-coherent broadcast channel with spatial correlation for identical coherence time  $T = 25$  and  $r_k \leq \min\{M, N_k, \lfloor T/2 \rfloor\}$ . When  $r_0$  is below  $\frac{2r_2 + r_1 - T}{2}$ ,  $D_3$  becomes a corner point (see Remark 2).

For further insights, the following remarks are in order.

**Remark 1.**  $D_1$  and  $D_2$  contain the maximum individual DoF one user can achieve when the other user is deactivated. The BS exploits all the  $r_k$  dimensions of channel subspace  $\mathcal{V}_k$  to transmit signal to the active user  $k$ . Without knowing the

channel correlation structure, the best strategy one can do is time sharing between  $D_1$  and  $D_2$ , which is optimal when the channel subspaces  $\mathcal{V}_1$  and  $\mathcal{V}_2$  are completely overlapping.

**Remark 2.**  $D_3$  is achieved by transmitting a private message to each user in the non-overlapping part of the channel subspaces, and leave the intersection  $\mathcal{V}_1 \cap \mathcal{V}_2$  unoccupied to avoid interference. The number of (interference-free) effective signal directions to user  $k$  is  $\dim(\mathcal{V}_k) - \dim(\mathcal{V}_1 \cap \mathcal{V}_2) = r_k - r_0$ .  $D_3$  is a corner point of the region only when  $T + 2r_0 < r_1 + 2r_2$ , i.e., the subspace overlap is small enough.

**Remark 3.**  $D_4$  and  $D_5$  are achieved by transmitting a private message to each user in the non-overlapping part of the channel subspaces, and a common message to both users in the intersection simultaneously. When the channel is symmetric, i.e.,  $r_2 = r_1 =: r$ , either user can remain at maximum individual DoF while the DoF of the other user is increased up to a certain threshold. The achievable DoF is symmetric and defined by two axes and either

$$d_1 + d_2 \leq M' \left(1 - \frac{r}{T}\right), \quad (10)$$

if  $T + 2r_0 \geq 3r$ , or

$$\min \left\{ \frac{d_1}{T - M'} - \frac{d_2}{r - r_0}, \frac{d_2}{T - M'} - \frac{d_1}{r - r_0} \right\} \leq \left(1 - \frac{r}{T}\right) \left(\frac{r}{T - M'} - 1\right), \quad (11)$$

otherwise, where  $M' = 2r - r_0$  is the total number of effective signal directions for communication. When the channel is asymmetric, i.e.,  $r_2 > r_1$ , only the user with more effective signal directions (user 2) can remain at maximum individual DoF when we increase the DoF of the other user.

Note that for notational simplicity, we assumed that  $r_k \leq \lfloor T/2 \rfloor$ . When  $r_k > \lfloor T/2 \rfloor$ , one should send signal in only  $\lfloor T/2 \rfloor$  effective signal directions to maximize the individual achievable DoF of user  $k$ , as suggested in [5]. Two extreme cases, in which the achievable region given in Theorem 1 becomes optimal and coincides with some known results in the literature, are as follows.

1) *Best scenario: Non-overlapping subspaces:* In the most favorable case, the channel subspaces  $\mathcal{V}_1$  and  $\mathcal{V}_2$  are disjoint, i.e.,  $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$ ,  $r_0 = 0$  and  $r_1 + r_2 \leq M$ . Every row of  $\mathbf{U}_1$  is linearly independent of any row of  $\mathbf{U}_2$ .  $D_3$ ,  $D_4$ , and  $D_5$  coincide and equal  $(r_1(1 - \frac{r_1}{T}), r_2(1 - \frac{r_2}{T}))$ . The achievable region is simply a rectangle defined by the two axes and

$$d_1 \leq r_1 \left(1 - \frac{r_1}{T}\right), \quad d_2 \leq r_2 \left(1 - \frac{r_2}{T}\right). \quad (12)$$

Both users can simultaneously obtain the full DoF of a non-coherent point-to-point channel with  $r_k$  equivalent transmit antennas and  $N_k (\geq r_k)$  receive antennas. To achieve this, we simply transmit the signal to one user in its channel subspace, which is contained in the null space of the other user's channel.

2) *Worst scenario: Coinciding subspaces:* In this case, two channel subspaces  $\mathcal{V}_1$  and  $\mathcal{V}_2$  are completely overlapping, i.e.,  $r_1 = r_2 = r_0 = M'$ . Every row of  $\mathbf{U}_1$  is a linear combination of the rows of  $\mathbf{U}_2$ , and vice versa.  $D_4$  coincides with  $D_1$ ,  $D_5$  coincides with  $D_2$ , and  $D_3$  collapses to the origin. The region is simply a triangle defined by the two axes and

$$d_1 + d_2 \leq M' \left(1 - \frac{M'}{T}\right). \quad (13)$$

This is achieved with TDMA. Note that although TDMA does not require any statistical CSI, the knowledge of  $\mathbf{R}_k$  is still useful for shaping the input signal. TDMA was shown in [9] to be DoF-optimal for i.i.d., i.e. *isotropic*, Rayleigh fading channels. When the channels are spatially correlated but span the same subspace, TDMA can still be shown to be DoF-optimal (we omit the proof due to the lack of space). Thus, in this case, the achievable DoF region in Theorem 1 is tight.

#### IV. PROPOSED SCHEME

In this section, we propose a scheme to achieve the DoF pairs in Theorem 1. This scheme is based on rate splitting, additive superposition, and channel training.

##### A. Rate splitting with common and private messages

We would like to send the private signals  $\mathbf{X}_1$  to user 1,  $\mathbf{X}_2$  to user 2, and a common signal  $\mathbf{X}_0$  to both users. Each of these signals contains an information-carrying matrix pre-multiplied by a orthogonal precoding matrix. Let  $\mathbf{P}_k \in \mathbb{C}^{M \times (r_k - r_0)}$  be the precoding matrix in private signal  $\mathbf{X}_k$ , i.e.,

$$\mathbf{X}_k = \mathbf{P}_k \mathbf{S}_k, \quad k = 1, 2, \quad (14)$$

where  $\mathbf{S}_k \in \mathbb{C}^{(r_k - r_0) \times T}$  is the signal carrying the private message to user  $k$ . As we want the private signal to one user to be transparent to the other user, the precoders satisfy

$$\mathbf{U}_2 \mathbf{P}_1 = \mathbf{0}_{r_2 \times (r_1 - r_0)}, \quad \mathbf{U}_1 \mathbf{P}_2 = \mathbf{0}_{r_1 \times (r_2 - r_0)}, \quad (15)$$

$$\text{rank} \left( \mathbf{R}_k^{\frac{1}{2}} \mathbf{P}_k \right) = r_k - r_0, \quad k = 1, 2. \quad (16)$$

That is, the precoder of one user spans the null space of the channel of the other user. On the other hand, we precode the common signal as

$$\mathbf{X}_0 = \mathbf{P}_0 \mathbf{S}_0, \quad (17)$$

where  $\mathbf{S}_0 \in \mathbb{C}^{r_0 \times T}$  is the signal carrying the common message. Since  $\mathbf{X}_0$  is to be decoded by both users, the precoder  $\mathbf{U}_0^H$  is designed such that its span does not overlap with the null space of either user,<sup>2</sup> i.e.,  $\mathbf{P}_0$  is linearly independent of both  $\mathbf{P}_1$  and  $\mathbf{P}_2$ , and

$$\text{rank} \left( \mathbf{R}_1^{\frac{1}{2}} \mathbf{P}_0 \right) = \text{rank} \left( \mathbf{R}_2^{\frac{1}{2}} \mathbf{P}_0 \right) = r_0. \quad (18)$$

These signals are additively superposed and the  $M \times T$  transmitted signal matrix is simply

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_0. \quad (19)$$

<sup>2</sup>The columns of  $\mathbf{P}_0$  form an orthonormal basis for the intersection  $\mathcal{V}_1 \cap \mathcal{V}_2$ . Thus,  $\mathbf{P}_0$  can be calculated from the basis  $\mathbf{U}_1$  of  $\mathcal{V}_1$  and the basis  $\mathbf{U}_2$  of  $\mathcal{V}_2$ .

The received signals become

$$\mathbf{Y}_1 = \mathbf{W}_1 \Sigma_1^{\frac{1}{2}} \mathbf{U}_1 \mathbf{P}_1 \mathbf{S}_1 + \mathbf{W}_1 \Sigma_1^{\frac{1}{2}} \mathbf{U}_1 \mathbf{P}_0 \mathbf{S}_0 + \sqrt{\frac{M}{\rho}} \mathbf{Z}_1, \quad (20)$$

$$\mathbf{Y}_2 = \mathbf{W}_2 \Sigma_2^{\frac{1}{2}} \mathbf{U}_2 \mathbf{P}_2 \mathbf{S}_2 + \mathbf{W}_2 \Sigma_2^{\frac{1}{2}} \mathbf{U}_2 \mathbf{P}_0 \mathbf{S}_0 + \sqrt{\frac{M}{\rho}} \mathbf{Z}_2, \quad (21)$$

where the equivalent channels  $\mathbf{W}_1 \Sigma_1^{\frac{1}{2}} \mathbf{U}_1 \mathbf{P}_1 \in \mathbb{C}^{N_1 \times (r_1 - r_0)}$ ,  $\mathbf{W}_1 \Sigma_1^{\frac{1}{2}} \mathbf{U}_1 \mathbf{P}_0 \in \mathbb{C}^{N_1 \times r_0}$ ,  $\mathbf{W}_2 \Sigma_2^{\frac{1}{2}} \mathbf{U}_2 \mathbf{P}_2 \in \mathbb{C}^{N_2 \times (r_2 - r_0)}$ ,  $\mathbf{W}_2 \Sigma_2^{\frac{1}{2}} \mathbf{U}_2 \mathbf{P}_0 \in \mathbb{C}^{N_2 \times r_0}$  are unknown, correlated, and Gaussian distributed. Each column vector of these channels represents a (private or common) signal direction from the BS to users.

### B. DoF analysis

The received signal at each user is similar to a two non-coherent two-user MAC: (20) as the MAC 1 with  $(r_1 - r_0, r_0)$  equivalent transmit antennas and  $N_1$  receive antennas, (21) as the MAC 2 with  $(r_2 - r_0, r_0)$  equivalent transmit antennas and  $N_2$  receive antennas. Note that for each MAC, the two equivalent channel matrices are correlated and not independent of each other due to the terms  $\Sigma_k^{\frac{1}{2}}$ . In the case of i.i.d. Rayleigh block fading, a training-based scheme achieving sum DoF for the  $K$ -user non-coherent MAC was given in [9, Theorem 5]. This result, however, still holds in our setting.

**Proposition 1.** *For the MAC  $k$ , with training-based schemes, we can achieve the convex hull of three DoF pairs  $(d_k^p, d_k^c)$ :*

$$D_{k,1}^{\text{MAC}} = \left( (r_k - r_0) \left( 1 - \frac{r_k}{T} \right), r_0 \left( 1 - \frac{r_k}{T} \right) \right), \quad (22)$$

$$D_{k,2}^{\text{MAC}} = \left( (r_k - r_0) \left( 1 - \frac{r_k - r_0}{T} \right), 0 \right), \quad (23)$$

$$D_{k,3}^{\text{MAC}} = \left( 0, r_0 \left( 1 - \frac{r_0}{T} \right) \right). \quad (24)$$

*Proof.* See Appendix A.  $\square$

Specifically, to achieve  $D_{k,1}^{\text{MAC}}$ , the receiver learns  $r_0$  common signal directions and  $r_k - r_0$  private directions in the first  $r_k$  time slots, then receives data during the remaining  $T - r_k$  slots of the coherence interval.  $D_{k,2}^{\text{MAC}}$  and  $D_{k,3}^{\text{MAC}}$  are achieved by letting the receiver learn and receive data only in the private signal directions and common signal directions, respectively.

The following proposition relates the achievable DoF region of these two MACs and the region of the BC.

**Proposition 2.** *The BC can achieve the convex hull of all the points having the form*

$$f_k(D_{1,i_1}^{\text{MAC}}, D_{2,i_2}^{\text{MAC}}), \quad (25)$$

for  $k = 1, 2$  and  $i_1, i_2 \in \{1, 2, 3\}$ , where the mapping  $f$  is defined as

$$f_k: \mathbb{R}_+^2 \times \mathbb{R}_+^2 \mapsto \mathbb{R}_+^2 \quad (26)$$

$$(d_1^p, d_1^c), (d_2^p, d_2^c) \rightarrow \begin{cases} (d_1^p + \min\{d_1^c, d_2^c\}, d_2^p), & \text{if } k = 1, \\ (d_1^p, d_2^p + \min\{d_1^c, d_2^c\}), & \text{if } k = 2. \end{cases}$$

*Proof.* Each corner point of the MAC is achieved with a training-based scheme consisting of a training phase and a data transmission phase. By aligning these training and data transmission phases, we can combine the corner points of the two MACs to have achievable points of the BC. For example, to combine  $D_{1,1}^{\text{MAC}}$  and  $D_{2,1}^{\text{MAC}}$ , the alignment of the training and data transmission can be visualized from Fig. 2. In this figure, each bar represents a subspace in which the private or common signal lies on. First, it is clear that the (pilot or data) private signal transmissions in  $\mathcal{V}_1 \setminus (\mathcal{V}_1 \cap \mathcal{V}_2)$  and  $\mathcal{V}_2 \setminus (\mathcal{V}_1 \cap \mathcal{V}_2)$  do not interfere each other. Therefore, user  $k$  can achieve  $d_k = (r_k - r_0) \left( 1 - \frac{r_k}{T} \right)$  DoF for the private message independently. On the other hand, the common signal in  $\mathcal{V}_1 \cap \mathcal{V}_2$  interferes the private pilot signals of both users, so we must wait until the training of private signal directions is done for both users before starting the common data transmission. As a consequence, the duration for common data transmission is limited by the user with longer private training (user 2), and the BC can only achieve  $\min_k \{ r_0 \left( 1 - \frac{r_k}{T} \right) \} = r_0 \left( 1 - \frac{r_2}{T} \right)$  DoF for the common message.

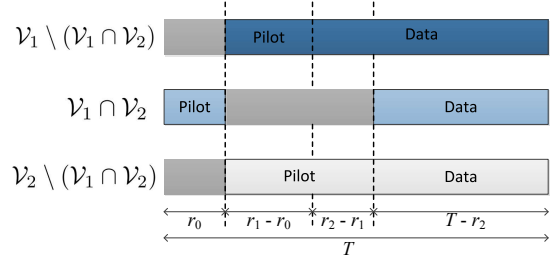


Fig. 2. The training model for the achievable scheme of the BC.

The combination of other corner points of the two MACs can be explained similarly. Therefore, for any achievable DoF pair  $(d_1^p, d_1^c)$  of the MAC 1 and  $(d_2^p, d_2^c)$  of the MAC 2, the BC can achieve  $d_1^p, d_2^p$  DoF for the private messages and  $\min\{d_1^c, d_2^c\}$  DoF for the common message, respectively. Next, the BC can dedicate the common message to either user to achieve two DoF pairs  $(d_1^p + \min\{d_1^c, d_2^c\}, d_2^p)$  and  $(d_1^p, d_2^p + \min\{d_1^c, d_2^c\})$ , which are the outputs of the mapping  $f$ .

We have shown that all the points having form (25) are achievable for the BC. The convex hull of these points can be achieved simply with time sharing.  $\square$

There are 18 points having the form (25), but some of them are dominated by the others. By eliminating the dominated points, we find five possible corner points of the BC, listed in Theorem 1, as follows:

$$D_1 = f_1(D_{1,1}^{\text{MAC}}, D_{2,3}^{\text{MAC}}) \quad (27)$$

$$D_2 = f_2(D_{1,3}^{\text{MAC}}, D_{2,1}^{\text{MAC}}) \quad (28)$$

$$D_3 = f_1(D_{1,2}^{\text{MAC}}, D_{2,2}^{\text{MAC}}) = f_2(D_{1,2}^{\text{MAC}}, D_{2,2}^{\text{MAC}}) \quad (29)$$

$$D_4 = f_1(D_{1,1}^{\text{MAC}}, D_{2,1}^{\text{MAC}}) \quad (30)$$

$$D_5 = f_2(D_{1,1}^{\text{MAC}}, D_{2,1}^{\text{MAC}}) \quad (31)$$

Specifically, these corner points can be achieved as follows.

- $D_1$  and  $D_2$  are individual extreme points and can be achieved by serving only one user as in the point-to-point case.
- $D_3$  can be achieved by transmitting only private signals  $\mathbf{X}_1$  and  $\mathbf{X}_2$  and letting  $\mathbf{X}_0 = \mathbf{0}_{M \times T}$ . (20) and (21) become two parallel non-coherent channels in Rayleigh fading with  $r_1 - r_0$  and  $r_2 - r_0$  equivalent transmit antennas, respectively. Then  $(r_k - r_0) \left(1 - \frac{r_k - r_0}{T}\right)$  DoF can be achieved for user  $k$  with a training-based scheme as follows:
  - During the first  $r_2 - r_0$  slots, user  $k$  learns its  $r_k - r_0$  private signal directions. Since user 1 finishes the learning first ( $r_1 \leq r_2$ ), the BS can start transmitting the private signal to user 1 using  $r_1 - r_0$  learned directions in  $r_2 - r_1$  time slots, when user 2 is still learning.
  - During the last  $T - (r_2 - r_0)$  slots, the BS transmits the private signal to user  $k$  in  $r_k - r_0$  learned directions.
- $D_4$  and  $D_5$  can be achieved by dedicating the common message in the training-based scheme illustrated in Fig. 2 to either user.

This method can be used to generalize the result to the  $K$ -user case. The main idea is to use rate splitting to transform the BC into  $K$  MACs, then construct the corner points of the BC from the corner points of the MACs by a generalization of the mapping (26).

## V. CONCLUSION

For non-coherent MIMO broadcast channels with spatial correlation, knowing and adapting to the correlation structure can improve significantly the degrees of freedom, especially when the intersection of the users' channel subspaces is small. We show this by characterizing an achievable DoF region of this channel in the two-user case. The region is constructed from the achievable region of some multiple access channels transformed from the broadcast channel by rate splitting, and can be achieved with a training-based scheme. While training-based strategy with Gaussian signaling is DoF-optimal for point-to-point non-coherent MIMO channel, it is not clear if it is also optimal for broadcast channels with spatial correlation.

## REFERENCES

- [1] I. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, pp. 585–595, Nov./Dec. 1999.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] L. Zheng and D. N. 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, Sep. 2006.
- [6] S. A. Jafar and A. Goldsmith, "Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 990–997, May 2005.
- [7] 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.

- [8] 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.
- [9] 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.

## APPENDIX A

### SKETCH OF PROOF OF PROPOSITION 1

This proof applies for both MAC  $k$ ,  $k = 1, 2$ , so we omit the index  $k$  for brevity. The received signal of the MAC is  $\mathbf{Y} = \mathbf{H}\hat{\mathbf{P}}\hat{\mathbf{S}} + \sqrt{\frac{M}{\rho}}\mathbf{Z}$ , where  $\mathbf{H} = \mathbf{W}\mathbf{R}^{\frac{1}{2}}$ ,  $\hat{\mathbf{P}} = [\mathbf{P}_0 \mathbf{P}] \in \mathbb{C}^{M \times r}$ ,  $\hat{\mathbf{S}} = [\mathbf{S}_0^T \mathbf{S}^T]^T \in \mathbb{C}^{r \times T}$ .

We consider a training-based scheme with equal power for training and data transmission phases. During the training phase, a  $r \times T_\tau$  matrix  $\hat{\mathbf{S}}_\tau$  is transmitted, containing training symbols known to the receiver. The received signal is  $\mathbf{Y}_\tau = \mathbf{H}\hat{\mathbf{P}}\hat{\mathbf{S}}_\tau + \sqrt{\frac{M}{\rho}}\mathbf{Z}_\tau$ . We use the MMSE estimator  $\hat{\mathbf{H}} = \mathbf{Y}_\tau \left( \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H \mathbf{R} \hat{\mathbf{P}} \hat{\mathbf{S}}_\tau + \frac{M}{\rho} \mathbf{I}_{T_\tau} \right)^{-1} \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H \mathbf{R}$ . The estimation error  $\tilde{\mathbf{H}} = \mathbf{H} - \hat{\mathbf{H}}$  has zero mean and covariance  $\frac{1}{N} \mathbb{E}[\tilde{\mathbf{H}} \tilde{\mathbf{H}}^H] = \left( \mathbf{R}^{-1} + \frac{\rho}{M} \hat{\mathbf{P}} \hat{\mathbf{S}}_\tau^H \hat{\mathbf{S}}_\tau \hat{\mathbf{P}}^H \right)^{-1}$ . In the data transmission, a  $r \times (T - T_\tau)$  matrix  $\hat{\mathbf{S}}_d \sim \mathcal{CN}(0, \mathbf{I}_r)$  containing data symbols is transmitted. The received signal is  $\mathbf{Y}_d = \hat{\mathbf{H}}\hat{\mathbf{P}}\hat{\mathbf{S}}_d + \sqrt{\frac{M}{\rho}}\mathbf{Z}'_d$ , where  $\mathbf{Z}'_d = \sqrt{\frac{\rho}{M}}\tilde{\mathbf{H}}\hat{\mathbf{P}}\hat{\mathbf{S}}_d + \mathbf{Z}_d$  is the combined noise consisting of additive noise and channel estimation error. With MMSE estimator,  $\mathbf{Z}'_d$  and  $\hat{\mathbf{S}}_d$  are uncorrelated. From [8, Theorem 1], the worst uncorrelated additive noise has zero-mean Gaussian distribution. Furthermore, since the distribution of the channel  $\hat{\mathbf{H}}$  is left rotationally invariant, this worst case Gaussian noise has identity covariance matrix. Therefore, a lower bound on the capacity is obtained by replacing  $\mathbf{Z}'_d$  by i.i.d. Gaussian noise with the same variance  $\sigma_{\mathbf{Z}'_d}^2 = \frac{\rho}{MN} \text{tr} \left( \hat{\mathbf{P}}^H \mathbb{E}[\tilde{\mathbf{H}} \tilde{\mathbf{H}}^H] \hat{\mathbf{P}} \right) + 1$ :

$$C \geq C^l = \frac{T - T_\tau}{T} \mathbb{E} \log \det \left( \mathbf{I}_N + \frac{\rho}{M \sigma_{\mathbf{Z}'_d}^2} \hat{\mathbf{H}} \hat{\mathbf{P}} \hat{\mathbf{P}}^H \hat{\mathbf{H}}^H \right). \quad (32)$$

$C^l$  is maximized with  $T_\tau = r$  and  $\sigma_{\mathbf{Z}'_d}^2 = 1 + \frac{\rho}{M} \frac{\rho + \text{tr} \mathbf{R}^{-1}}{\text{tr}(\hat{\mathbf{P}}^H \hat{\mathbf{P}})^{1/2}} \text{tr} \left( (\hat{\mathbf{P}}^H \hat{\mathbf{P}})^{-1/2} \hat{\mathbf{P}} \hat{\mathbf{P}}^H \right)$ . Note that with the designed precoders, the matrix  $\frac{1}{M \sigma_{\mathbf{Z}'_d}^2} \hat{\mathbf{H}} \hat{\mathbf{P}} \hat{\mathbf{P}}^H \hat{\mathbf{H}}^H$  has rank  $r$  with probability 1. Let  $\mu$  be an arbitrary non-zero eigenvalue, then

$$C \geq C^l = \left(1 - \frac{r}{T}\right) r \mathbb{E} \log(1 + \rho \mu) \quad (33)$$

$$= r \left(1 - \frac{r}{T}\right) \log(\rho) + O(1), \quad \text{when } \rho \rightarrow \infty. \quad (34)$$

This means a total  $r \left(1 - \frac{r}{T}\right)$  DoF can be achieved for the two messages coded in  $\hat{\mathbf{S}}$ . Since  $\mathbf{S}$  and  $\mathbf{S}_0$  contribute  $r - r_0$  and  $r_0$  rows in  $\hat{\mathbf{S}}$ , respectively,  $(r - r_0) \left(1 - \frac{r}{T}\right)$  DoF can be achieved for the message in  $\mathbf{S}$  and  $r_0 \left(1 - \frac{r}{T}\right)$  DoF for  $\mathbf{S}_0$ . We have proved that the DoF pair  $D_1^{\text{MAC}} = \left( (r - r_0) \left(1 - \frac{r}{T}\right), r_0 \left(1 - \frac{r}{T}\right) \right)$  is achievable for the equivalent MAC. The other achievable pairs  $D_2^{\text{MAC}}$  and  $D_3^{\text{MAC}}$  can be shown similarly by modifying  $\hat{\mathbf{P}}$  and  $\hat{\mathbf{S}}$ .

APPENDIX B  
PROOF OF PROPOSITION 1 IN MORE DETAILS

This proof applies for both MAC  $k$ ,  $k = 1, 2$ , so we omit the user index  $k$  for brevity. The received signal of the MAC is

$$\mathbf{Y} = \mathbf{W}\mathbf{R}^{\frac{1}{2}}\mathbf{P}\mathbf{S} + \mathbf{W}\mathbf{R}^{\frac{1}{2}}\mathbf{P}_0\mathbf{S}_0 + \sqrt{\frac{M}{\rho}}\mathbf{Z} \quad (35)$$

$$= \hat{\mathbf{H}}\hat{\mathbf{P}}\hat{\mathbf{S}} + \sqrt{\frac{M}{\rho}}\mathbf{Z}, \quad (36)$$

where  $\hat{\mathbf{P}} = [\mathbf{P}_0 \ \mathbf{P}] \in \mathbb{C}^{M \times r}$ ,  $\hat{\mathbf{S}} = \begin{bmatrix} \mathbf{S}_0 \\ \mathbf{S} \end{bmatrix} \in \mathbb{C}^{r \times T}$ .

First, let us prove that we can achieve  $D_1^{\text{MAC}} = ((r - r_0)(1 - \frac{r}{T}), r_0(1 - \frac{r}{T}))$  using a training-based scheme. We assume equal power allocation for training and data transmission phases. During the training phase of  $T_\tau \geq r$  time slots, a  $r \times T_\tau$  matrix  $\hat{\mathbf{S}}_\tau$  is transmitted, containing training symbols known to the receiver. The received signal is

$$\mathbf{Y}_\tau = \mathbf{W}\mathbf{R}^{\frac{1}{2}}\hat{\mathbf{P}}\hat{\mathbf{S}}_\tau + \sqrt{\frac{M}{\rho}}\mathbf{Z}_\tau. \quad (37)$$

We use the MMSE estimator  $\hat{\mathbf{H}} = \mathbf{Y}_\tau \mathbf{A}$ , where  $\mathbf{A}$  is the minimizer of the mean square error

$$\begin{aligned} \frac{1}{N} \mathbb{E}[\|\mathbf{H} - \hat{\mathbf{H}}\|_F^2] &= \text{tr}(\mathbf{R}) - \text{tr}(\mathbf{R}\hat{\mathbf{P}}\hat{\mathbf{S}}_\tau \mathbf{A}) - \text{tr}(\mathbf{A}^H \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H \mathbf{R}) \\ &+ \text{tr} \left( \mathbf{A}^H \left( \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H \mathbf{R} \hat{\mathbf{P}} \hat{\mathbf{S}}_\tau + \frac{M}{\rho} \mathbf{I} \right) \mathbf{A} \right). \end{aligned} \quad (38)$$

Solving  $\frac{\partial}{\partial \mathbf{A}} \frac{1}{N} \mathbb{E}[\|\mathbf{H} - \hat{\mathbf{H}}\|_F^2] = 0$  yields the optimal  $\mathbf{A}_{\text{opt}} = \left( \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H \mathbf{R} \hat{\mathbf{P}} \hat{\mathbf{S}}_\tau + \frac{M}{\rho} \mathbf{I}_{T_\tau} \right)^{-1} \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H \mathbf{R}$ , hence

$$\hat{\mathbf{H}} = \mathbf{Y}_\tau \left( \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H \mathbf{R} \hat{\mathbf{P}} \hat{\mathbf{S}}_\tau + \frac{M}{\rho} \mathbf{I}_{T_\tau} \right)^{-1} \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H \mathbf{R}. \quad (39)$$

The estimation error  $\tilde{\mathbf{H}} = \mathbf{H} - \hat{\mathbf{H}}$  has zero mean and covariance

$$\frac{1}{N} \mathbb{E}[\tilde{\mathbf{H}} \tilde{\mathbf{H}}^H] = \mathbf{R} - \mathbf{R} \hat{\mathbf{P}} \hat{\mathbf{S}}_\tau \left( \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H \mathbf{R} \hat{\mathbf{P}} \hat{\mathbf{S}}_\tau + \frac{M}{\rho} \mathbf{I}_{T_\tau} \right)^{-1} \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H \mathbf{R} \quad (40)$$

$$= \left( \mathbf{R}^{-1} + \frac{\rho}{M} \hat{\mathbf{P}} \hat{\mathbf{S}}_\tau \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H \right)^{-1} \quad (41)$$

by the Woodbury identity.

In the data transmission, a  $r \times (T - T_\tau)$  matrix  $\hat{\mathbf{S}}_d$  containing data symbols is transmitted. The received signal is

$$\mathbf{Y}_d = \hat{\mathbf{H}} \hat{\mathbf{P}} \hat{\mathbf{S}}_d + \sqrt{\frac{M}{\rho}} \mathbf{Z}_d \quad (42)$$

$$= \hat{\mathbf{H}} \hat{\mathbf{P}} \hat{\mathbf{S}}_d + \sqrt{\frac{M}{\rho}} \mathbf{Z}'_d, \quad (43)$$

where  $\mathbf{Z}'_d = \sqrt{\frac{\rho}{M}} \tilde{\mathbf{H}} \hat{\mathbf{P}} \hat{\mathbf{S}}_d + \mathbf{Z}_d$  is the combined noise consisting of additive noise and channel estimation error. With MMSE estimator, this combined noise and data signal are uncorrelated.

From [8, Theorem 1], we know that the worst case uncorrelated additive noise has zero-mean Gaussian distribution. Furthermore, since the distribution of the channel  $\hat{\mathbf{H}}$  is left rotationally invariant, this worst case Gaussian noise has identity covariance matrix. Therefore, a lower bound on the capacity is obtained by replacing  $\mathbf{Z}'_d$  by i.i.d. Gaussian noise with the same power constraint

$$\sigma_{\mathbf{Z}'_d}^2 = \frac{1}{N(T - T_\tau)} \text{tr} \mathbb{E}[\mathbf{Z}'_d \mathbf{Z}'_d^H] \quad (44)$$

$$= \frac{\rho}{MN(T - T_\tau)} \text{tr} \left( \hat{\mathbf{P}}^H \mathbb{E}[\hat{\mathbf{H}} \hat{\mathbf{H}}^H] \hat{\mathbf{P}} \mathbb{E}[\hat{\mathbf{S}}_d \hat{\mathbf{S}}_d^H] \right) + 1 \quad (45)$$

$$= \frac{\rho}{MN} \text{tr} \left( \hat{\mathbf{P}}^H \mathbb{E}[\hat{\mathbf{H}} \hat{\mathbf{H}}^H] \hat{\mathbf{P}} \right) + 1, \quad (46)$$

with Gaussian signaling  $\hat{\mathbf{S}}_d \sim \mathcal{CN}(0, \mathbf{I}_r)$ . The capacity lower bound is

$$C \geq C^l = \frac{T - T_\tau}{T} \mathbb{E} \left[ \log \det \left( \mathbf{I}_N + \frac{\rho}{M \sigma_{\mathbf{Z}'_d}^2} \hat{\mathbf{H}} \hat{\mathbf{P}} \hat{\mathbf{P}}^H \hat{\mathbf{H}}^H \right) \right]. \quad (47)$$

To maximize  $C^l$ , we minimize  $\sigma_{\mathbf{Z}'_d}^2$  and  $T_\tau$ . Minimizing  $\sigma_{\mathbf{Z}'_d}^2$  is equivalent to

$$\min_{\text{tr}(\hat{\mathbf{P}} \hat{\mathbf{S}}_\tau \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H) = M} \text{tr} \left( \hat{\mathbf{P}}^H \left( \mathbf{R}^{-1} + \frac{\rho}{M} \hat{\mathbf{P}} \hat{\mathbf{S}}_\tau \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H \right)^{-1} \hat{\mathbf{P}} \right) \quad (48)$$

This can be solved using Lagrange multiplier method

$$\begin{aligned} L(\hat{\mathbf{S}}_\tau, \lambda) &= \text{tr} \left( \hat{\mathbf{P}}^H \left( \mathbf{R}^{-1} + \frac{\rho}{M} \hat{\mathbf{P}} \hat{\mathbf{S}}_\tau \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H \right)^{-1} \hat{\mathbf{P}} \right) \\ &+ \lambda (\text{tr}(\hat{\mathbf{P}} \hat{\mathbf{S}}_\tau \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H) - M) \end{aligned} \quad (49)$$

Solving  $\frac{\partial L(\hat{\mathbf{S}}_\tau, \lambda)}{\partial \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H} = 0$ , we obtain that the optimal training matrix should satisfy

$$\hat{\mathbf{P}} \hat{\mathbf{S}}_\tau \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H = \frac{M}{\rho \sqrt{\lambda}} (\hat{\mathbf{P}}^H \hat{\mathbf{P}})^{1/2} - \frac{M}{\rho} \mathbf{R}^{-1}. \quad (50)$$

Using the condition  $\text{tr}(\hat{\mathbf{P}} \hat{\mathbf{S}}_\tau \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H) = M$ , we find that

$$\hat{\mathbf{P}} \hat{\mathbf{S}}_\tau \hat{\mathbf{S}}_\tau^H \hat{\mathbf{P}}^H = \frac{M}{\rho} \frac{\rho + \text{tr} \mathbf{R}^{-1}}{\text{tr}(\hat{\mathbf{P}}^H \hat{\mathbf{P}})^{1/2}} (\hat{\mathbf{P}}^H \hat{\mathbf{P}})^{1/2} - \frac{M}{\rho} \mathbf{R}^{-1}. \quad (51)$$

Then

$$\sigma_{\mathbf{Z}'_d}^2 = 1 + \frac{\rho}{M} \frac{\rho + \text{tr} \mathbf{R}^{-1}}{\text{tr}(\hat{\mathbf{P}}^H \hat{\mathbf{P}})^{1/2}} \text{tr} \left( (\hat{\mathbf{P}}^H \hat{\mathbf{P}})^{-1/2} \hat{\mathbf{P}} \hat{\mathbf{P}}^H \right). \quad (52)$$

Now let us plug this worst noise variance to the capacity lower bound  $C^l$  given in (47). Note that with the designed precoder, the matrix  $\frac{1}{M \sigma_{\mathbf{Z}'_d}^2} \hat{\mathbf{H}} \hat{\mathbf{P}} \hat{\mathbf{P}}^H \hat{\mathbf{H}}^H$  has rank  $r$ , i.e.,  $r$  non-zero eigenvalues. Let  $\mu$  be an arbitrary eigenvalue among these, and notice that  $T_\tau \geq r$ , then

$$C \geq C^l = \left( 1 - \frac{r}{T} \right) r \mathbb{E} \log(1 + \rho \mu) \quad (53)$$

$$= r \left( 1 - \frac{r}{T} \right) \log(\rho) + O(1), \quad \text{when } \rho \rightarrow \infty. \quad (54)$$

This means a total  $r(1 - \frac{r}{T})$  DoF can be achieved for the two messages coded in  $\hat{\mathbf{S}}$ . Since  $\mathbf{S}$  and  $\mathbf{S}_0$  contribute  $r - r_0$  and  $r_0$

rows in  $\hat{\mathbf{S}}$ , respectively,  $(r - r_0) \left(1 - \frac{r}{T}\right)$  DoF can be achieved for the message in  $\mathbf{S}$  and  $r_0 \left(1 - \frac{r}{T}\right)$  DoF for  $\mathbf{S}_0$ .

We have proved that the DoF pair  $D_1^{\text{MAC}} = \left((r - r_0) \left(1 - \frac{r}{T}\right), r_0 \left(1 - \frac{r}{T}\right)\right)$  is achievable for the equivalent MAC. The other achievable pairs  $D_2^{\text{MAC}}$  and  $D_3^{\text{MAC}}$  can be shown similarly by modifying  $\hat{\mathbf{P}}$  and  $\hat{\mathbf{S}}$ .