Degrees of Freedom of Time Correlated MISO Broadcast Channel with Delayed CSIT

Sheng Yang, *Member, IEEE,* Mari Kobayashi, *Member, IEEE,*

David Gesbert, *Fellow, IEEE,* Xinping Yi, *Student Member, IEEE*

Abstract

We consider the time correlated MISO broadcast channel where the transmitter has imperfect knowledge on the current channel state, in addition to delayed channel state information. By representing the quality of the current channel state as $P^{-\alpha}$ for the signal-to-noise ratio P and some constant $\alpha \geq 0$, we characterize the optimal degree of freedom region for this more general two-user MISO broadcast correlated channel. The essential ingredients of the proposed scheme lie in the quantization and multicasting of the overheard interferences, while broadcasting new private messages. Our proposed scheme smoothly bridges between the scheme recently proposed by Maddah-Ali and Tse with no current state information and a simple zero-forcing beamforming with perfect current state information.

I. INTRODUCTION

In most practical scenarios, perfect channel state information at transmitter (CSIT) may not be available due to the time-varying nature of wireless channels as well as the limited resource for channel estimation. However, many wireless applications must guarantee high-data rate and reliable communication in the presence of channel uncertainty. In this paper, we consider such scenario in the context of the two-user MISO broadcast channel, where the transmitter equipped with m antennas wishes to send two private messages to two receivers each with a single antenna. The discrete time signal model is given by

$$y_t = \boldsymbol{h}_t^{\mathsf{H}} \boldsymbol{x}_t + \varepsilon_t \tag{1a}$$

$$z_t = \boldsymbol{g}_t^{\mathsf{H}} \boldsymbol{x}_t + \omega_t \tag{1b}$$

for any time instant t, where $\boldsymbol{h}_t, \boldsymbol{g}_t \in \mathbb{C}^{m \times 1}$ are the channel vectors for user 1 and 2, respectively; $\varepsilon_t, \omega_t \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ are normalized additive white Gaussian noise (AWGN) at the respective receivers; the input signal $\boldsymbol{x}_t$ is subject to the power constraint $\mathbb{E}\left(\|\boldsymbol{x}_t\|^2\right) \leq P, \forall t$.

For the case of perfect CSIT, the optimal degrees of freedom of this channel is two and achieved by linear strategies such as zero-forcing (ZF) beamforming. When the transmitter suffers from constant inaccuracy of channel estimation, it has been shown in [1] that the degrees of freedom (DoF) per user is upper-bounded by $\frac{2}{3}$. It is also well known that the full multiplexing gain can be maintained under imperfect CSIT if the error in CSIT decreases as $O(P^{-1})$ or faster as P grows [2]. Moreover, for the case of the temporally correlated fading channel such that the transmitter can predict the current state with error decaying as $O(P^{-\alpha})$ for some constant $\alpha \in [0, 1]$, ZF can only achieve a fraction α of the optimal degrees of freedom [2]. This result somehow reveals the bottleneck of a family of precoding schemes relying only on instantaneous CSIT as the temporal correlation decreases ($\alpha \to 0$). Recently, a breakthrough has been made in order to overcome such problem. In [3], Maddah-Ali and Tse showed a surprising result that even completely outdated CSIT can be very useful in terms of degree of freedom, as long as it is accurate. For a system with $m \geq 2$ antennas and two users, the proposed scheme in [3], hereafter called MAT, achieves the multiplexing gain of $\frac{2}{3}$ per user, irrespectively of the temporal correlation. This work shifts the paradigm of broadcast precoding from space-only to space-time alignment. The role of perfect delayed CSIT can be re-interpreted as a feedback of the past signal/interference heard by the receivers. This side information enables the transmitter to perform "retrospective" alignment in the space *and* time domain, as demonstrated in different multiuser network systems (see e.g. [4]). Despite its DoF optimality, the MAT scheme is designed assuming the worst case scenario where the delayed channel feedback provides no information about the current one. This assumption is over pessimistic as most practical channels exhibit some form of temporal correlation. In fact, it readily follows that a selection strategy between ZF and MAT yields the degrees of freedom of $\max\{\alpha, \frac{2}{3}\}$ for $\alpha \in [0, 1]$. For either quasi-static fading channel ($\alpha \geq 1$) or very fast channels ($\alpha \to 0$), a selection approach is reasonable. However, for intermediate ranges of temporal correlation ($0 < \alpha < 1$), a fundamental question arises as to whether a better way of exploiting both delayed CSIT and current (imperfect) CSIT exists. Studying the achievable DoF under such CSIT assumption is of practical and theoretical interest.

The main contributions of this work are summarized as follows.

1) We propose a simple strategy (Scheme I) that combines the ZF precoding, based on the imperfect current state information, and the MAT alignment, based on the perfect past state information. The main role of current CSIT is to reduce, via spatial precoding, the overheard interference power in the original MAT alignment. The said power reduction then enables, via source compression/quantization, to save the resources related to the transmission of the overheard interference. Scheme I, initially reported in [6], achieves the DoF of

$$d_{\text{Scheme I}} = \frac{2 - \alpha}{3 - 2\alpha}, \quad \alpha \in [0, 1] \tag{2}$$

that is strictly better than both the MAT alignment and ZF precoding for $\alpha \in (0, 1)$. The key of this scheme is the quantization that is introduced into the MAT alignment to replace the analog transmission. Despite its

suboptimality as it will turn out, it is the indispensable part of the optimal scheme that we propose later.

2) We establish an outer bound on the DoF region of the two-user broadcast channel with perfect delayed and imperfect current state information. To that end, we use two powerful tools: the genie-aided model and the extremal inequality [8]. The outer bound is a source of inspiration of our optimal scheme.

3) Based on Scheme I and motivated by the outer bound, we propose an optimal scheme (Scheme II) that achieves the upper bound of the DoF

$$d_{\text{Scheme II}} = \frac{2+\alpha}{3}, \quad \alpha \in [0,1] \tag{3}$$

given by the converse. The enhancement is built on the observation that the second phase of Scheme I, i.e., multicast, does not exploit current CSI. This can be improved by sending two new private messages alongside the common message on the overheard interference. The gain of 2α degrees of freedom, i.e., from $2 - \alpha$ to $2 + \alpha$ is obtained at the price of 2α extra channel uses, i.e., from $3 - 2\alpha$ to 3. As confirmed by the converse, this is in fact the best tradeoff for the symmetric DoF. To achieve the other corner points of the region, we show that delayed CSIT is not necessary and the optimal strategy is to broadcast with common message that is useful for only one of the users.

4) As an extension to the main result, we derive the optimal DoF region of the same channel with common message. Another extension is the achievable DoF region when only imperfect delayed CSI is available (e.g., due to limited feedback rates). The case with general fading process (e.g., non-ergodic case in delay limited communications) is also discussed. Finally, in addition to the results on the optimal DoF region, we provide the exact achievable rate regions of the proposed schemes (cf. Appendix). At the time of submission, parallel independent work [12] was brought to our attention which also builds on the results of [6].

The paper is organized as follows. In Section II, after presenting the assumptions and some basic definitions of our model, we provide our main theorem on the optimal DoF region. The above contributions are then presented in order. Finally, we conclude the paper with some perspectives. Detailed proofs are deferred to the Appendix.

Throughout the paper, we will use the following notations. Matrix transpose, Hermitian transpose, inverse, and determinant are denoted by $\boldsymbol{A}^{\mathsf{T}}$, $\boldsymbol{A}^{\mathsf{H}}$, $\boldsymbol{A}^{-1}$, and $\det(\boldsymbol{A})$, respectively. $\boldsymbol{x}^{\perp}$ is any nonzero vector such that $\boldsymbol{x}^{\mathsf{H}}\boldsymbol{x}^{\perp} = 0$. Logarithm is in base 2.

II. SYSTEM MODEL AND MAIN RESULTS

For convenience, we provide the following definition on the channel states.

Definition 1 (channel states): The channel vectors $\boldsymbol{h}_t$ and $\boldsymbol{g}_t$ are called the states of the channel at instant t. For simplicity, we also define the state matrix $\boldsymbol{S}_t$ as $\boldsymbol{S}_t \triangleq \begin{bmatrix} \boldsymbol{h}_t^{\mathsf{H}} \\ \boldsymbol{g}_t^{\mathsf{H}} \end{bmatrix} \in \mathcal{S}$.

The assumptions on the fading process and the knowledge of the channel states are summarized as follows. Possible relaxations of the assumptions are discussed in Section VI.

Assumption 1 (mutually independent fading): At any given time instant t, the channel vectors for the two users $\boldsymbol{h}_t, \boldsymbol{g}_t$ are mutually independent and identically distributed (i.i.d.) with zero mean and covariance matrix $\mathbf{I}_m$. Moreover, we assume that $\text{rank}(\boldsymbol{S}_t) = 2$ with probability 1.

Assumption 2 (perfect delayed and imperfect current CSI): At each time instant t, the transmitter knows the delayed channel states up to instant $t-1$. In addition, the transmitter can somehow obtain an estimation $\hat{S}_t$ of the current channel state S_t, i.e., $\hat{h}_t$ and $\hat{g}_t$ are available to the transmitter with

$$h_t = \hat{h}_t + \tilde{h}_t \tag{4a}$$

$$g_t = \hat{g}_t + \tilde{g}_t \tag{4b}$$

where the estimate $\hat{h}_t$ (also $\hat{g}_t$) and estimation error $\tilde{h}_t$ (also $\tilde{g}_t$) are uncorrelated and both assumed to be zero mean with covariance $(1 - \sigma^2)\mathbf{I}_m$ and $\sigma^2\mathbf{I}_m$, respectively, with $\sigma^2 \leq 1$. The receivers knows perfectly $S_t \in \mathcal{S}$ and $\hat{S}_t \in \hat{\mathcal{S}}$ without delay.

For simplicity and tractability, we have the following assumption on the fading process.

Assumption 3 (Rayleigh fading): The processes $\left\{\hat{S}_t\right\}$ and $\left\{\tilde{S}_t\right\}$ are independent, stationary, and ergodic. For each t, the entries of $\hat{S}_t$ are i.i.d. $\mathcal{N}_\mathbb{C}\left(0, 1 - \sigma^2\right)$ distributed while the entries of $\tilde{S}_t$ are i.i.d. $\mathcal{N}_\mathbb{C}\left(0, \sigma^2\right)$ distributed. Moreover, we assume the following Markov chain

$$(\hat{S}^{t-1}, S^{t-1}) \leftrightarrow \hat{S}_t \leftrightarrow S_t, \quad \forall t. \tag{5}$$

Without loss of generality, we can introduce a parameter $\alpha_P \geq 0$ as the power exponent of the estimation error

$$\alpha_P \triangleq -\frac{\log(\sigma^2)}{\log P}. \tag{6}$$

The parameter α_P can be regarded as the quality of the current CSI in the high SNR regime. Note that $\alpha_P = 0$ corresponds to the case with no current CSIT at all while $\alpha_P \to \infty$ corresponds to the case with perfect current CSIT. In addition, we assume that $\lim_{P \to \infty} \alpha_P$ exists and define

$$\alpha \triangleq \lim_{P \to \infty} \alpha_P. \tag{7}$$

Hereafter, we use α instead of α_P, whenever no confusion is likely. Connections between the above model and practical time correlated models are highlighted in Section VI.

Definition 2 (achievable degrees of freedom): A code for the Gaussian MISO broadcast channel with delayed CSIT and imperfect current CSIT is

- A sequence of encoders at time t is given by $F_t : \mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{S}^{t-1} \times \hat{\mathcal{S}}^t \longmapsto \mathbb{C}^m$ where the message W_1 and W_2 are uniformly distributed over $\mathcal{W}_1$ and $\mathcal{W}_2$, respectively.
- A decoder for user k is given by the mapping $\hat{W}_k : \mathbb{C}^{1 \times n} \times \mathcal{S}^n \times \hat{\mathcal{S}}^n \longmapsto \mathcal{W}_k$, $k = 1, 2$.

The DoF pair (d_1, d_2) is said *achievable* if there exists a code that simultaneously satisfies the reliability condition

$$\limsup_{n \to \infty} \Pr\left\{W_k \neq \hat{W}_k\right\} = 0, \tag{8}$$

and

$$\lim_{P \to \infty} \liminf_{n \to \infty} \frac{\log_2 |\mathcal{W}_k(n, P)|}{n \log_2 P} \geq d_k, \quad k = 1, 2. \tag{9}$$

The union of all achievable DoF pairs is then called the optimal DoF region the broadcast channel.

The main result of this paper is stated in the following theorem.

Theorem 1: In the two-user MISO broadcast channel with delayed perfect CSIT and imperfect current CSIT, the optimal degrees of freedom region is characterized by

$$d_1 \leq 1 \tag{10a}$$

$$d_2 \leq 1 \tag{10b}$$

$$d_1 + 2d_2 \leq 2 + \alpha \tag{10c}$$

$$2d_1 + d_2 \leq 2 + \alpha \tag{10d}$$

The next three sections are devoted to proving the Theorem. We start with the achievability.

III. ACHIEVABILITY: A SIMPLE SCHEME

In this section, we describe a novel and simple scheme that is initially presented in [6]. Since this scheme builds on a variant of the MAT scheme, we briefly review the original MAT scheme and its variant.

A. MAT alignment revisited

In the two-user MISO case, the original MAT is a three-slot scheme, described by the following equations

$$\boldsymbol{x}_1 = \boldsymbol{u} \qquad \boldsymbol{x}_2 = \boldsymbol{v} \qquad \boldsymbol{x}_3 = [\boldsymbol{g}_1^{\mathsf{H}}\boldsymbol{u} + \boldsymbol{h}_2^{\mathsf{H}}\boldsymbol{v} \quad 0]^{\mathsf{T}} \tag{11a}$$

$$y_1 = \boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{u} \qquad y_2 = \boldsymbol{h}_2^{\mathsf{H}}\boldsymbol{v} \qquad y_3 = h_{31}(\boldsymbol{g}_1^{\mathsf{H}}\boldsymbol{u} + \boldsymbol{h}_2^{\mathsf{H}}\boldsymbol{v}) \tag{11b}$$

$$z_1 = \boldsymbol{g}_1^{\mathsf{H}}\boldsymbol{u} \qquad z_2 = \boldsymbol{g}_2^{\mathsf{H}}\boldsymbol{v} \qquad z_3 = g_{31}(\boldsymbol{g}_1^{\mathsf{H}}\boldsymbol{u} + \boldsymbol{h}_2^{\mathsf{H}}\boldsymbol{v}) \tag{11c}$$

where $\boldsymbol{x}_t \in \mathbb{C}^{m \times 1}, y_t, z_t \in \mathbb{C}$ are the transmitted signal, received signal at user 1, received signal at user 2, respectively, at time slot t; $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^{m \times 1}$ are useful signals to user 1 and user 2, respectively; for simplicity, we omit the noise in the received signals. The idea of the MAT scheme is to use the delayed CSIT to align the mutual interference into a reduced subspace with only one dimension ($\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{v}$ for user 1 and $\boldsymbol{g}_1^{\mathsf{H}}\boldsymbol{u}$ for user 2). And importantly, the reduction in interference is done without sacrificing the dimension of the useful signals. Specifically, a two-dimensional interference-free observation of $\boldsymbol{u}$ (resp. $\boldsymbol{v}$) is obtained at receiver 1 (resp. receiver 2).

Interestingly, the alignment can be done in a different manner.

$$\boldsymbol{x}_1 = \boldsymbol{u} + \boldsymbol{v} \qquad \boldsymbol{x}_2 = [\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{v} \quad 0]^{\mathsf{T}} \qquad \boldsymbol{x}_3 = [\boldsymbol{g}_1^{\mathsf{H}}\boldsymbol{u} \quad 0]^{\mathsf{T}} \tag{12a}$$

$$y_1 = \boldsymbol{h}_1^{\mathsf{H}}(\boldsymbol{u} + \boldsymbol{v}) \qquad y_2 = h_{21}\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{v} \qquad y_3 = h_{31}\boldsymbol{g}_1^{\mathsf{H}}\boldsymbol{u} \tag{12b}$$

$$z_1 = \boldsymbol{g}_1^{\mathsf{H}}(\boldsymbol{u} + \boldsymbol{v}) \qquad z_2 = g_{21}\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{v} \qquad z_3 = g_{31}\boldsymbol{g}_1^{\mathsf{H}}\boldsymbol{u} \tag{12c}$$

In the first slot, the transmitter sends the private signals to both users by simply superposing them. In the second slot, the transmitter sends the interference overheard by receiver 1 in the first slot. The role of this stage is two-fold: *resolving interference for user 1 and reinforcing signal for user 2*. In the third slot, the transmitter sends the

interference overheard by user 2 to help both users the other way around. In summary, this variant of the MAT consists two phases: i) broadcasting the private signals, and ii) multicasting the overheard interference, i.e., $\boldsymbol{h}_1^{\text{H}}\boldsymbol{v}$ and $\boldsymbol{g}_1^{\text{H}}\boldsymbol{u}$. At the end of three slots, the observations at the receivers are given by

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{h}_1^{\text{H}} \\ 0 \\ h_{31}\boldsymbol{g}_1^{\text{H}} \end{bmatrix}}_{\text{rank}=2} \boldsymbol{u} + \underbrace{\begin{bmatrix} \boldsymbol{h}_1^{\text{H}} \\ h_{21}\boldsymbol{h}_1^{\text{H}} \\ 0 \end{bmatrix}}_{\text{rank}=1} \boldsymbol{v}, \tag{13a}$$

and

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{g}_1^{\text{H}} \\ g_{21}\boldsymbol{h}_1^{\text{H}} \\ 0 \end{bmatrix}}_{\text{rank}=2} \boldsymbol{v} + \underbrace{\begin{bmatrix} \boldsymbol{g}_1^{\text{H}} \\ 0 \\ g_{31}\boldsymbol{g}_1^{\text{H}} \end{bmatrix}}_{\text{rank}=1} \boldsymbol{u}. \tag{13b}$$

For each user, the useful signal lies in a two-dimensional subspace while the interference is aligned in a one-dimensional subspace. It readily follows that the variant enables each user to achieve two degrees of freedom in the three-dimensional time space as for the original MAT. Although two schemes are equivalent from the point of the space-time alignment, they differ conceptually in the way how the "order-two" symbols are delivered. More precisely, the variant spends two slots to deliver two symbols $\boldsymbol{h}_1^{\text{H}}\boldsymbol{v}, \boldsymbol{g}_1^{\text{H}}\boldsymbol{u}$ while the original MAT spends a single slot to deliver one symbols $\boldsymbol{h}_2^{\text{H}}\boldsymbol{v} + \boldsymbol{g}_1^{\text{H}}\boldsymbol{u}$. As seen shortly, multicasting two "order-two" symbols separately is crucial since it allows to shorten the time for multicasting when partial knowledge on the current state is available.

B. Integrating the imperfect current CSI

Based on the above variant of the MAT scheme, we propose the following two-stage scheme, called Scheme I, that integrates the estimates of the current CSI. Since only $\boldsymbol{h}_1$ and $\boldsymbol{g}_1$ are involved below, we drop the indices for convenience, whenever it is possible.

Phase 1 - Precoding and broadcasting the private signals: As in the above MAT variant, we first superpose the two private signals as $\boldsymbol{x} = \boldsymbol{u} + \boldsymbol{v}$, except that $\boldsymbol{u}$ and $\boldsymbol{v}$ are precoded beforehand. The precoding is specified by the covariance matrices $\boldsymbol{Q_u} \triangleq \mathbb{E}(\boldsymbol{u}\boldsymbol{u}^{\text{H}})$ and $\boldsymbol{Q_v} \triangleq \mathbb{E}(\boldsymbol{v}\boldsymbol{v}^{\text{H}})$ that may depend on the estimates of the current channel. The power constraint is respected by choosing $\boldsymbol{Q_u}$ and $\boldsymbol{Q_v}$ such that

$$\text{tr}(\boldsymbol{Q_u}) + \text{tr}(\boldsymbol{Q_v}) \leq P. \tag{14}$$

Phase 2 - Quantizing and multicasting the overheard interference : As the second phase of the MAT variant, the objective of this phase is to convey the overheard interferences $(\boldsymbol{h}^{\text{H}}\boldsymbol{v}, \boldsymbol{g}^{\text{H}}\boldsymbol{u})$ required by both receivers. However, unlike the original MAT scheme (12) where these symbols are transmitted in an analog fashion, we quantize them and then transmit the digital version. The rationale behind this choice is as follows. With (imperfect) CSI on the current channel, the transmitter can use the precoding to align the signals and allocate the transmit power in such a

way that the overheard interferences have a reduced power, without sacrificing *too much* received signal power.[1] As a result, we should be able to save the time resource for multicasting, which in turn improves the degree of freedom. The benefit can be significant when the current CSI is nearly perfect. In this case, the analog transmission is not longer suitable, due to the mismatch between the source (interference) power and available transmit power. Therefore, a good alternative is to quantize the interferences and to transmit the encoded symbols. The number of quantization bits depends naturally on the interference power that is related to the quality of the state information. For convenience, we define $\eta_1 \triangleq \boldsymbol{h}^{\mathsf{H}}\boldsymbol{v}$, $\eta_2 \triangleq \boldsymbol{g}^{\mathsf{H}}\boldsymbol{u}$, and $\boldsymbol{\eta} \triangleq (\eta_1, \eta_2)$.

The first step is the quantization. We quantize η_1 and η_2 separately. Note that for a given channel realization, the average power of η_1 and η_2 are

$$\sigma_{\eta_1}^2 \triangleq \boldsymbol{h}^{\mathsf{H}}\boldsymbol{Q}_v\boldsymbol{h} \quad \text{and} \quad \sigma_{\eta_2}^2 \triangleq \boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q}_u\boldsymbol{g}. \tag{15}$$

Let us assume that an R_{η_k}-bits quantizer is used for η_k, $k = 1, 2$. Hence, we have

$$\eta_k = \hat{\eta}_k + \Delta_k \tag{16}$$

where $\hat{\eta}_k$ and Δ_k are the quantized value and the quantization noise with average distortion $\mathbb{E}\left(|\Delta_k|^2\right) = D_k$, $k = 1, 2$, respectively. The index corresponding to $\hat{\boldsymbol{\eta}} \triangleq (\hat{\eta}_1, \hat{\eta}_2)$, represented in $R_{\boldsymbol{\eta}} \triangleq R_{\eta_1} + R_{\eta_2}$ bits, is then multicast to both users.

Decoding: At the receivers' side, each user first tries to recover $(\hat{\eta}_1, \hat{\eta}_2)$. If this step is done successfully, then receiver 1 has

$$y = \boldsymbol{h}^{\mathsf{H}}\boldsymbol{u} + \eta_1 + \varepsilon \tag{17}$$

$$\hat{\eta}_1 = \eta_1 - \Delta_1 \tag{18}$$

$$\hat{\eta}_2 = \eta_2 - \Delta_2 = \boldsymbol{g}^{\mathsf{H}}\boldsymbol{u} - \Delta_2 \tag{19}$$

from which an equivalent $m \times 2$ MIMO channel is obtained

$$\tilde{\boldsymbol{y}} \triangleq \begin{bmatrix} y - \hat{\eta}_1 \\ \hat{\eta}_2 \end{bmatrix} = \boldsymbol{S}\boldsymbol{u} + \begin{bmatrix} \varepsilon + \Delta_1 \\ -\Delta_2 \end{bmatrix} \tag{20}$$

where the noise $\boldsymbol{b} \triangleq [\varepsilon + \Delta_1 \;\; -\Delta_2]^{\mathsf{T}}$ depends on the input signals in general. Similarly, if receiver 2 can recover $(\hat{\eta}_1, \hat{\eta}_2)$ correctly, then the following term is available

$$\tilde{\boldsymbol{z}} \triangleq \begin{bmatrix} \hat{\eta}_1 \\ z - \hat{\eta}_2 \end{bmatrix} = \boldsymbol{S}\boldsymbol{v} + \begin{bmatrix} -\Delta_1 \\ \omega + \Delta_2 \end{bmatrix}. \tag{21}$$

In order to finally recover the message, each user performs conventional MIMO decoding of the above equivalent channel.

---

[1]With no CSIT on the current channel, the only way to reduce the interference power is to reduce the transmit power, therefore the received signal power.

C. Achievable degrees of freedom

In the following, we derive the achievable symmetric degrees of freedom of Scheme I. The exact achievable rate region from which the DoF can be proved in a more rigorous way will be provided in Appendix A.

Let R_{mimo}, $R_{\boldsymbol{\eta}}$, and R_{mc} be the average MIMO rate for each user, the quantization rate for $\boldsymbol{\eta}$, and the multicast rate of the channel, respectively. It is obvious that the average rate of Scheme I is given by

$$\bar{R}_{\text{I}} = \frac{R_{\text{mimo}}}{1 + R_{\boldsymbol{\eta}}/R_{\text{mc}}}. \tag{22}$$

In the rest of the section, we would like to show that the following rates are achievable

$$R_{\text{mc}} = \log P + O(1) \tag{23a}$$

$$R_{\boldsymbol{\eta}} = 2(1 - \alpha) \log P + O(1) \tag{23b}$$

$$R_{\text{mimo}} = (2 - \alpha) \log P + O(1) \tag{23c}$$

from which we have the symmetric DoF $\frac{2-\alpha}{3-2\alpha}$ given by (2). The interpretation of the achievable DoF is the following. By properly designing the precoding covariance matrices as well as the quantization, one can shorten the transmission duration by 2α channel uses at the price of losing a pre-log factor α in total rate. Since we need to show the achievability for any $m \geq 2$, it is enough to consider the case with $m = 2$. And we fix the parameters of Scheme I as follows:

- For each user, we send two streams in two orthogonal directions: one aligned with the estimated channel of the unintended user while the other one perpendicular to it, i.e.,

$$\boldsymbol{Q_u} = P_1 \boldsymbol{\Psi}_{\hat{g}^\perp} + P_2 \boldsymbol{\Psi}_{\hat{g}} \tag{24a}$$

$$\boldsymbol{Q_v} = P_1 \boldsymbol{\Psi}_{\hat{h}^\perp} + P_2 \boldsymbol{\Psi}_{\hat{h}} \tag{24b}$$

where

$$\boldsymbol{\Psi}_{\hat{g}} \triangleq \frac{\hat{g}\hat{g}^{\text{H}}}{\|\hat{g}\|^2} \tag{25}$$

and $\boldsymbol{\Psi}_{\hat{g}^\perp}$, $\boldsymbol{\Psi}_{\hat{h}}$, and $\boldsymbol{\Psi}_{\hat{h}^\perp}$ are similarly defined.

- The transmitted power in the direction of estimated channel is such that $P_2 \sim P^{1-\alpha}$ while the transmitted power in the orthogonal direction is $P_1 = P - P_2 \sim P$ for any $\alpha < 1$.

- The distortions D_1 and D_2 are set to the noise level, i.e., $D_1 = D_2 = 1 \sim P^0$.

First, (23a) is achievable by using any single user code. Second, we can upper-bound the quantization rate $R_{\boldsymbol{\eta}}$ as

$$R_{\boldsymbol{\eta}} \leq \mathbb{E}\left(\log\left(\frac{\boldsymbol{h}^{\mathsf{H}}\boldsymbol{Q_v}\boldsymbol{h}}{D_1}\right)\right) + \mathbb{E}\left(\log\left(\frac{\boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q_u}\boldsymbol{g}}{D_2}\right)\right) \tag{26}$$

$$\leq \log\left(\mathbb{E}\left(\boldsymbol{h}^{\mathsf{H}}\boldsymbol{Q_v}\boldsymbol{h}\right)\right) + \log\left(\mathbb{E}\left(\boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q_u}\boldsymbol{g}\right)\right) \tag{27}$$

$$= 2\log\left(\mathbb{E}\left(\boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q_u}\boldsymbol{g}\right)\right) \tag{28}$$

$$= 2\log\left(\mathbb{E}\left(\tilde{\boldsymbol{g}}^{\mathsf{H}}\boldsymbol{Q_u}\tilde{\boldsymbol{g}}\right) + \mathbb{E}\left(\hat{\boldsymbol{g}}^{\mathsf{H}}\boldsymbol{Q_u}\hat{\boldsymbol{g}}\right)\right) \tag{29}$$

$$\leq 2\log\left(P\sigma^2 + P_2\mathbb{E}\left(\hat{\boldsymbol{g}}^{\mathsf{H}}\boldsymbol{\Psi}_{\hat{\boldsymbol{g}}}\hat{\boldsymbol{g}}\right)\right) \tag{30}$$

$$\leq 2\log\left(P\sigma^2 + 2P_2(1-\sigma^2)\right) \tag{31}$$

$$= 2(1-\alpha)\log P + O(1) \tag{32}$$

where the first inequality is from the rate-distortion theorem and by saying that Gaussian source is the hardest to compress [7]; the second inequality is from the concavity of the log function; (28) is from the symmetry between the channels and between the strategies; (29) is from (24a). Finally, we lower-bound the MIMO rate R_{mimo} of user 1 as

$$R_{\text{mimo}} = \mathbb{E}\left(I(U;\tilde{Y} \mid S = \boldsymbol{S})\right) \tag{33}$$

$$= \mathbb{E}\left(I(\boldsymbol{S}U;\tilde{Y})\right) \tag{34}$$

$$= \mathbb{E}\left(h(\boldsymbol{S}U) - h(\boldsymbol{S}U \mid \tilde{Y})\right) \tag{35}$$

$$= \mathbb{E}\left(h(\boldsymbol{S}U) - h(E + \Delta_1, -\Delta_2 \mid \tilde{Y})\right) \tag{36}$$

$$\geq \mathbb{E}\left(h(\boldsymbol{S}U) - h(E + \Delta_1, -\Delta_2)\right) \tag{37}$$

$$\geq \mathbb{E}\left(\log\det\left(\boldsymbol{S}\boldsymbol{Q_u}\boldsymbol{S}^{\mathsf{H}}\right) - \log(1+D_1) - \log(D_2)\right) \tag{38}$$

$$= \mathbb{E}\left(\log\det\left(\boldsymbol{Q_u}\right)\right) + \mathbb{E}\left(\log\det\left(\boldsymbol{S}\boldsymbol{S}^{\mathsf{H}}\right) - \log(1+D_1) - \log(D_2)\right) \tag{39}$$

$$= \log(P_1 P_2) + O(1) \tag{40}$$

$$= (2-\alpha)\log(P) + O(1) \tag{41}$$

where (37) holds since conditioning reduces differential entropies; (38) follows because $\boldsymbol{u}$ is Gaussian, then by noticing that $E + \Delta_1$ and Δ_2 are independent with the corresponding differential entropy maximized by Gaussian distribution.

IV. CONVERSE

In this section, we establish the converse proof of the main result. Before going into the details, we would like to point out the essential elements of the upcoming proof:

- Genie-aided model is used to construct a degraded broadcast channel, as in [3].
- Extremal inequality is applied to bound the weighted difference of differential entropies [8].
- Isotropic property of the estimation error on the current state is exploited.

First, let us first consider the genie-aided model where the genie provides z_t to user 1 at each time instant t. This is a degraded broadcast channel $X \leftrightarrow (Y, Z) \leftrightarrow Z$. Therefore, we have the following upper bounds on the rates (R_1, R_2):

$$nR_1 \leq H(W_1) \tag{42}$$

$$= H(W_1 \,|\, S^n, \hat{S}^n) \tag{43}$$

$$= I(W_1; Y^n, Z^n \,|\, S^n, \hat{S}^n) + n\epsilon_n \tag{44}$$

$$\leq I(W_1; Y^n, Z^n, W_2 \,|\, S^n, \hat{S}^n) + n\epsilon_n \tag{45}$$

$$= I(W_1; Y^n, Z^n \,|\, S^n, \hat{S}^n, W_2) + n\epsilon_n \tag{46}$$

$$= \sum_{i=1}^{n} I(W_1; Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, S^n, \hat{S}^n, W_2) + n\epsilon_n \tag{47}$$

$$\leq \sum_{i=1}^{n} I(X_i; Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, S^n, \hat{S}^n, W_2) + n\epsilon_n \tag{48}$$

$$= \sum_{i=1}^{n} I(X_i; Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, S^i, \hat{S}^i, W_2) + n\epsilon_n \tag{49}$$

$$= \sum_{i=1}^{n} h(Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, S^i, \hat{S}^i, W_2) \tag{50}$$

$$- h(Y_i, Z_i \,|\, X_i, Y^{i-1}, Z^{i-1}, S^i, \hat{S}^i, W_2) + n\epsilon_n \tag{51}$$

$$= \sum_{i=1}^{n} h(Y_i, Z_i \,|\, T_i, S_i) - h(E_i, \Omega_i) + n\epsilon_n \tag{52}$$

$$\leq \sum_{i=1}^{n} h(Y_i, Z_i \,|\, T_i, S_i) + n\epsilon_n \tag{53}$$

$$nR_2 \leq H(W_2) \tag{54}$$

$$\leq I(W_2; Z^n \,|\, S^n, \hat{S}^n) + n\epsilon_n \tag{55}$$

$$= \sum_{i=1}^{n} I(W_2; Z_i \,|\, Z^{i-1}, S^i, \hat{S}^i) + n\epsilon \tag{56}$$

$$= \sum_{i=1}^{n} h(Z_i \,|\, Z^{i-1}, S^i, \hat{S}^i) - h(Z_i \,|\, Z^{i-1}, S^i, \hat{S}^i, W_2) + n\epsilon \tag{57}$$

$$\leq \sum_{i=1}^{n} h(Z_i \,|\, S_i) - h(Z_i \,|\, Y^{i-1}, Z^{i-1}, S^i, \hat{S}^i, W_2) + n\epsilon \tag{58}$$

$$= \sum_{i=1}^{n} h(Z_i \,|\, S_i) - h(Z_i \,|\, T_i, S_i) + n\epsilon \tag{59}$$

where we defined $T_i \triangleq (Y^{i-1}, Z^{i-1}, S^{i-1}, \hat{S}^i, W_2)$; we also used the fact that the differential entropy of the AWGN $h(E_i, \Omega_i) \geq 0$. Note that the above chains of inequalities follow closely Gallager's proof for the degraded broadcast channel [10] (also see [7]), with the integration of the channel states. Obviously, we have the Markov chain $X_i \leftrightarrow T_i \leftrightarrow (S^{i-1}, \hat{S}^{i-1}) \leftrightarrow \hat{S}_i \leftrightarrow S_i$. In the following, we would like to obtain an upper bound on $R_1 + 2R_2$.

From (53) and (59), we have

$$n(R_1 + 2R_2) \leq \sum_{i=1}^{n} \left(h(Y_i, Z_i \,|\, T_i, S_i) - 2h(Z_i \,|\, T_i, S_i) + 2h(Z_i \,|\, S_i) \right) + 3n\epsilon_n. \tag{60}$$

Now, we can have an upper bound for each i,

$$h(Y_i, Z_i \,|\, T_i, S_i) - 2h(Z_i \,|\, T_i, S_i) + 2h(Z_i \,|\, S_i) \tag{61}$$

$$\leq \max_{P_{T_i} P_{X_i | T_i}} h(Y_i, Z_i \,|\, T_i, S_i) - 2h(Z_i \,|\, T_i, S_i) + 2h(Z_i \,|\, S_i) \tag{62}$$

$$\leq \max_{P_{T_i} P_{X_i | T_i}} 2h(Z_i \,|\, S_i) + \max_{P_{T_i} P_{X_i | T_i}} \left(h(Y_i, Z_i \,|\, T_i, S_i) - 2h(Z_i \,|\, T_i, S_i) \right). \tag{63}$$

The first maximization can be upper-bounded as:

$$\max_{P_{T_i} P_{X_i | T_i}} 2h(Z_i \,|\, S_i) \leq 2 \, \mathbb{E}_{G_i} \left(\max_{P_{X_i | G_i = \boldsymbol{g}_i}} h(\boldsymbol{g}_i^{\mathsf{H}} X_i + E_i) \right) \tag{64}$$

$$\leq 2 \, \mathbb{E}_{G_i} \left(\log(1 + P\|\boldsymbol{g}_i\|^2) \right) \tag{65}$$

$$\leq 2 \log P + O(1) \tag{66}$$

where we used the fact that Gaussian distribution maximizes differential entropy under the covariance constraint, that the logarithmic function is monotonically increasing, and $\mathsf{Cov}(X_i \,|\, \boldsymbol{g}_i) \preceq \mathsf{Cov}(X_i) \preceq P\mathbf{I}$. The second maximization in (63) can also be bounded, but in a slightly more involved way:

$$\max_{P_{T_i} P_{X_i | T_i}} \left(h(Y_i, Z_i \,|\, T_i, S_i) - 2h(Z_i \,|\, T_i, S_i) \right)$$

$$\leq \max_{P_{T_i}} \mathbb{E}_{T_i} \left(\max_{P_{X_i | T_i}} \left(h(Y_i, Z_i \,|\, T_i = T, S_i) - 2h(Z_i \,|\, T_i = T, S_i) \right) \right) \tag{67}$$

$$= \max_{P_{T_i}} \mathbb{E}_{T_i} \left(\max_{P_{X_i | T_i}} \mathbb{E}_{S_i | T_i} \left(h(Y_i, Z_i \,|\, T_i = T, S_i = \boldsymbol{S}_i) - 2h(Z_i \,|\, T_i = T, S_i = \boldsymbol{S}_i) \right) \right) \tag{68}$$

$$= \max_{P_{T_i}} \mathbb{E}_{T_i} \left(\max_{P_{X_i | T_i}} \mathbb{E}_{S_i | \hat{S}_i} \left(h(\boldsymbol{S}_i X_i + N_i \,|\, T_i = T) - 2h(\boldsymbol{g}_i^{\mathsf{H}} X_i + E_i \,|\, T_i = T) \right) \right) \tag{69}$$

$$= \max_{P_{T_i}} \mathbb{E}_{T_i} \left(\max_{\substack{\boldsymbol{C}: \boldsymbol{C} \succeq 0, \mathrm{tr}(\boldsymbol{C}) \leq P}} \max_{\substack{P_{X_i | T_i}: \\ \mathsf{Cov}(X_i | T_i) \preceq \boldsymbol{C}}} \mathbb{E}_{S_i | \hat{S}_i} \left(h(\boldsymbol{S}_i X_i + N_i \,|\, T_i = T) - 2h(\boldsymbol{g}_i^{\mathsf{H}} X_i + E_i \,|\, T_i = T) \right) \right) \tag{70}$$

$$= \max_{P_{T_i}} \mathbb{E}_{T_i} \left(\max_{\substack{\boldsymbol{C}: \boldsymbol{C} \succeq 0, \mathrm{tr}(\boldsymbol{C}) \leq P}} \mathbb{E}_{S_i | \hat{S}_i} \left(\log \det \left(I + \boldsymbol{S}_i \boldsymbol{K}_* \boldsymbol{S}_i^{\mathsf{H}} \right) - 2 \log(1 + \boldsymbol{g}_i^{\mathsf{H}} \boldsymbol{K}_* \boldsymbol{g}_i) \right) \right) \tag{71}$$

$$\leq \mathbb{E}_{\hat{S}_i} \left(\max_{\substack{\boldsymbol{K}: \boldsymbol{K} \succeq 0, \mathrm{tr}(\boldsymbol{K}) \leq P}} \mathbb{E}_{S_i | \hat{S}_i} \left(\log \det \left(I + \boldsymbol{S}_i \boldsymbol{K} \boldsymbol{S}_i^{\mathsf{H}} \right) - 2 \log(1 + \boldsymbol{g}_i^{\mathsf{H}} \boldsymbol{K} \boldsymbol{g}_i) \right) \right) \tag{72}$$

$$\leq \mathbb{E}_{\hat{S}_i} \left(\max_{\substack{\boldsymbol{K}: \boldsymbol{K} \succeq 0, \mathrm{tr}(\boldsymbol{K}) \leq P}} \mathbb{E}_{S_i | \hat{S}_i} \left(\log(1 + \boldsymbol{h}_i^{\mathsf{H}} \boldsymbol{K} \boldsymbol{h}_i) - \log(1 + \boldsymbol{g}_i^{\mathsf{H}} \boldsymbol{K} \boldsymbol{g}_i) \right) \right) \tag{73}$$

where in (67) we put one of the maximization into the expectation, which only increases the value; in (69), we define $N_i \triangleq [E_i \ \ \Omega_i]^{\mathsf{T}}$; (70) is obtained by splitting one maximization into two, one with the trace constraint and the other with the covariance constraint; (71) is from the fact that with covariance constraint, Gaussain distribution optimizes the weighted difference of the two differential entropies, given that 1) S_i is independent of X_i conditional on T_i, and that Y_i is a degraded version of (Y_i, Z_i); it is an application of the extremal inequality [8], [9]; note that $\boldsymbol{K}_* \preceq \boldsymbol{C}$

is defined as the optimal covariance for the inner maximization; (72) is obtained by noticing that any $\boldsymbol{K}$ such that $0 \preceq \boldsymbol{K} \preceq \boldsymbol{C}$ with $\mathrm{tr}\,(\boldsymbol{C}) \leq P$ belongs to the set $\{\boldsymbol{K} : \ \boldsymbol{K} \succeq 0, \mathrm{tr}\,(\boldsymbol{K}) \leq P\}$, and that the whole term only depends on $\hat{S}_i$; the last inequality is from the fact that $\det\,(I + \boldsymbol{A}) \leq (1 + a_{11})(1 + a_{22})$ for any $\boldsymbol{A} \triangleq [a_{ij}]_{i,j=1,2} \succeq 0$.

Lemma 1: For any given $\boldsymbol{K} \succeq 0$ with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_m \geq 0$, we have

$$\mathbb{E}_{S_i | \hat{S}_i} \left(\log(1 + \boldsymbol{h}_i^{\mathsf{H}} \boldsymbol{K} \boldsymbol{h}_i) \right) \leq \log(1 + \|\hat{\boldsymbol{h}}_i\|^2 \lambda_1) + O(1) \tag{74}$$

$$\mathbb{E}_{S_i | \hat{S}_i} \left(\log(1 + \boldsymbol{g}_i^{\mathsf{H}} \boldsymbol{K} \boldsymbol{g}_i) \right) \geq \log(1 + e^{-\gamma} \sigma^2 \lambda_1) + O(1) \tag{75}$$

where γ is Euler's constant.

Proof: See Appendix C. ∎

Without loss of generality, we consider $\sigma^2 > 0$ in the following. The case with $\sigma^2 = 0$ corresponds to the case of perfect CSI, in which the optimal DoF is already known. From Lemma 1, we have

$$\mathbb{E}_{S_i | \hat{S}_i} \left(\log(1 + \boldsymbol{h}_i^{\mathsf{H}} \boldsymbol{K} \boldsymbol{h}_i) - \log(1 + \boldsymbol{g}_i^{\mathsf{H}} \boldsymbol{K} \boldsymbol{g}_i) \right)$$

$$\leq \log \frac{1 + \|\hat{\boldsymbol{h}}_i\|^2 \lambda_1}{1 + e^{-\gamma} \sigma^2 \lambda_1} + O(1) \tag{76}$$

$$\leq \log \left(1 + \frac{\|\hat{\boldsymbol{h}}_i\|^2}{e^{-\gamma} \sigma^2} \right) + O(1) \tag{77}$$

$$\leq -\log(\sigma^2) + \log \left(e^{-\gamma} \sigma^2 + \|\hat{\boldsymbol{h}}_i\|^2 \right) + O(1) \tag{78}$$

where (77) is from the fact that $\log \frac{1+ax}{1+bx} \leq \log(1 + \frac{a}{b})$, $\forall\, a, x \geq 0$, $b > 0$. Note that this upper bound does not depend on $\boldsymbol{K}$. From (72) and (78), we have

$$\max_{P_{T_i} P_{X_i | T_i}} \left(h(Y_i, Z_i \,|\, T_i, S_i) - 2h(Z_i \,|\, T_i, S_i) \right)$$

$$\leq \alpha \log P + \mathbb{E}_{\hat{S}_i} \left(\log \left(e^{-\gamma} + \|\hat{\boldsymbol{h}}_i\|^2 \right) \right) + O(1) \tag{79}$$

$$= \alpha \log P + O(1). \tag{80}$$

From (60), (66), (80), and letting $n \to \infty$, we have

$$R_1 + 2R_2 \leq (2 + \alpha) \log P + O(1) \tag{81}$$

from which we obtain (10c) by dividing both sides of the above inequality by $\log P$ and let $P \to \infty$. Similarly, from (59), (66), and letting $n \to \infty$, we have

$$R_2 \leq \log P + O(1) \tag{82}$$

from which we obtain the single user bound (10b) by dividing both sides of the above inequality by $\log P$ and let $P \to \infty$. To obtain (10d) and (10a), we can use the genie-aided model in which receiver 2 is helped by the genie and has perfect knowledge of y_t. Due to the symmetry, the same reasoning as above can be applied by swapping the roles of receiver 1 and receiver 2.

V. ACHIEVABILITY: CLOSING THE GAP

A. Inspiration from the upper bound

Let us compare the achievable symmetric DoF of Scheme I with the upper bound:

$$\frac{2-\alpha}{3-2\alpha} \quad \text{versus} \quad \frac{2+\alpha}{3} = \frac{2-\alpha+2\alpha}{3-2\alpha+2\alpha}. \tag{83}$$

A natural question arises. *Can we convey 2α more symbols by extending the transmission by 2 channel uses, i.e., in total over three channel uses?* We recall that the time saving of 2α channel uses has been made possible by exploiting the current CSIT during the first phase (of broadcasting). The comparison above reveals that Scheme I can be possibly enhanced if we exploit the current CSI during the multicasting phase as well.

B. Enhanced scheme

The key element of the new scheme is broadcasting with common message in the presence of imperfect CSI.

Lemma 2 (broadcast channel with common message and imperfect CSI): Let (R_0, R_1, R_2) be the rate of common message, private message for user 1, and private message for user 2, respectively. Furthermore, we let (d_0, d_1, d_2) be the corresponding DoF. Then, there exists a family of codes $\{\mathcal{X}_0(P), \mathcal{X}_{p1}(P), \mathcal{X}_{p2}(P)\}$, such that

$$d_0 = 1 - \alpha \tag{84}$$

$$d_1 = d_2 = \alpha \tag{85}$$

is achievable simultaneously.

Proof: A sketch of proof is as follows, with more details given in Appendix B. Let us consider a single channel use with a superposition scheme: $\boldsymbol{x} = \boldsymbol{x}_c + \boldsymbol{x}_{p1} + \boldsymbol{x}_{p2}$ with precoding such that $\mathbb{E}\left(\boldsymbol{x}_{p1}\boldsymbol{x}_{p1}^{\mathsf{H}}\right) = \frac{P_p}{2}\boldsymbol{\Psi}_{\hat{\boldsymbol{g}}^{\perp}}$ and $\mathbb{E}\left(\boldsymbol{x}_{p2}\boldsymbol{x}_{p2}^{\mathsf{H}}\right) = \frac{P_p}{2}\boldsymbol{\Psi}_{\hat{\boldsymbol{h}}^{\perp}}$. We set the power $P_p \sim P^{\alpha}$ such that the private signals are drowned by noise at the unintended receivers while remain the level P^{α} at the intended receivers. The power of the common signal is $P_c = \mathbb{E}\left(\|\boldsymbol{x}_c\|^2\right) \sim P$. The decoding is performed as follows. At each receiver, the common message is decoded first with the private signals treated as noise. The signal-to-interference-and-noise ratio (SINR) is approximately $P_c/P_p \sim P^{1-\alpha}$, from which the achievability of $d_0 = 1 - \alpha$ is shown. Then, each receiver proceeds with the decoding of their own private messages, after removing the decoded common message. The SINR for the private message being approximately P^{α}, $d_k = \alpha$ is thus achievable for user k, $k = 1, 2$. An important remark is that only current CSI is used to achieve this point. Later on, in Section VI-A, we will show that this is in fact the best tradeoff between private and common messages for the channel, even when delayed CSIT is available. ∎

It is now clear that we can trade α of common degrees of freedom for 2α private degrees of freedom. Therefore, Scheme I can be improved by modifying the second phase of the protocol. The new scheme, hereafter called Scheme II, is described as below.

1) The first phase of Scheme II is identical to the first phase of Scheme I: $\boldsymbol{x} = \boldsymbol{u} + \boldsymbol{v}$ with the same precoders.

2) As in Scheme I, the quantized version $\hat{\boldsymbol{\eta}} \triangleq (\hat{\eta}_1, \hat{\eta}_2)$ of the interferences η_1 and η_2 are coded in approximately $2(1-\alpha)\log P$ bits. However, instead of sending these bits in $2(1-\alpha)$ channel uses, they are sent in 2 channel
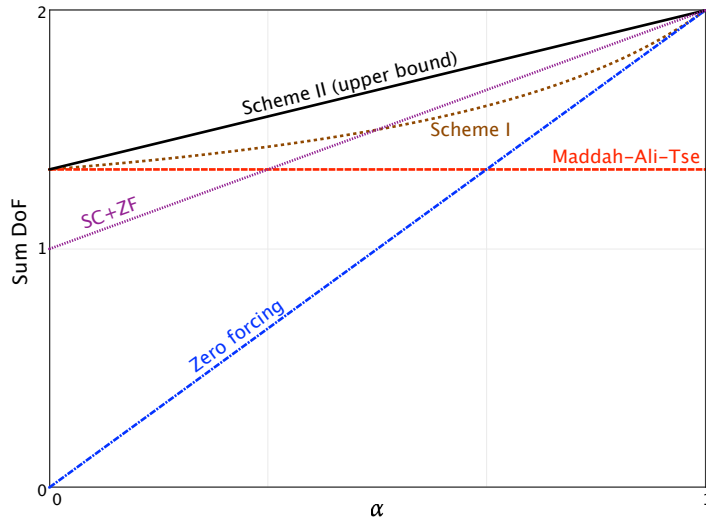
Fig. 1. Comparison of the achievable sum DoF between the proposed scheme and the zero-forcing and MAT alignment as a function of α.

uses with the code $\mathcal{C}_0$, as the common message for both users. Meanwhile, a new message of $\alpha \log P$ bits per channel use is sent to user k, as the private message of with codebook $\mathcal{X}_{p,k}$, $k = 1, 2$.

3) To decode, each receiver starts from the received signal at the second phase. First, according to Lemma 2, the private and common messages can be decoded reliably. Then, $\hat{\boldsymbol{\eta}}$ is restored in exactly the same manner as in Scheme I. Finally, the MAT part of $(2 - \alpha) \log P$ bits can also be recovered reliably.

Therefore, in three channel uses, $2\alpha + 2 - \alpha = 2 + \alpha$ DoF is achieved, yielding a symmetric DoF per channel use of $\frac{2+\alpha}{3}$.

Note that the region given by (10) is a polygon characterized by the vertices: $(0, 1)$, $(\alpha, 1)$, $(\frac{2+\alpha}{3}, \frac{2+\alpha}{3})$, $(1, \alpha)$, $(1, 0)$. Obviously, Scheme II achieves the symmetric point. From Lemma 2, we can see that by making the common message as the private message of one of the users, we achieve $(1, \alpha)$ and $(\alpha, 1)$. Therefore, by time sharing, the whole region is achievable. In Fig. 1, we compare the sum DoF of different schemes. The scheme "SC+ZF" (superposition coding and ZF precoding) is from the corner point $(1, \alpha)$ or $(\alpha, 1)$. Note that when α is close to 0, the estimation of current CSIT is bad and therefore useless. In this case, the optimal scheme is MAT [3], achieving DoF of $\frac{4}{3}$ for each user. On the other hand, when $\alpha \geq 1$, the estimation is good and the interference at the receivers due to the imperfect estimation is below the noise level and thus can be neglected as far as the DoF is concerned. In this case, ZF with the estimated current CSI is asymptotically optimal, achieving degrees of freedom 1 for each user. Our result (Scheme I and II) reveals that strictly larger DoF than $\max\{\frac{2}{3}, \alpha\}$ can be obtained by exploiting both the imperfect current CSIT and the perfect delayed CSIT in an intermediate regime $\alpha \in (0, 1)$.

In the Appendix, we provide the exact achievable rate region. Some examples of the achievable sum rates are plotted in Figure 2 and 3.[2] In Fig. A, we plot the sum rate performance of our proposed scheme II for different α.

[2]Note that parameters are fixed according the choices given in the Appendix without optimization.
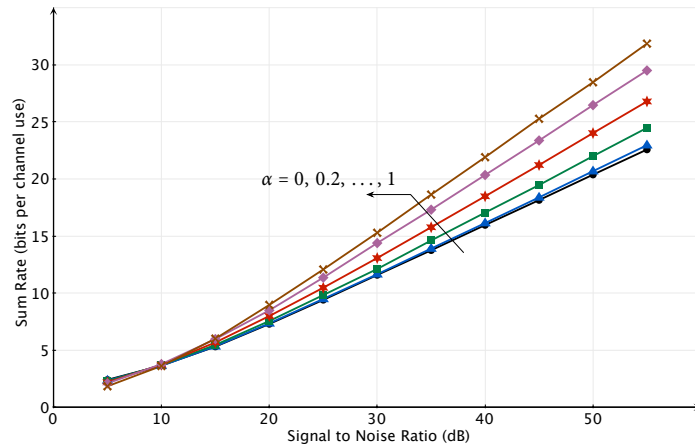
Fig. 2.   The achievable ergodic sum-rate of Scheme II for $\alpha = 0, 0.2, \ldots, 1$.
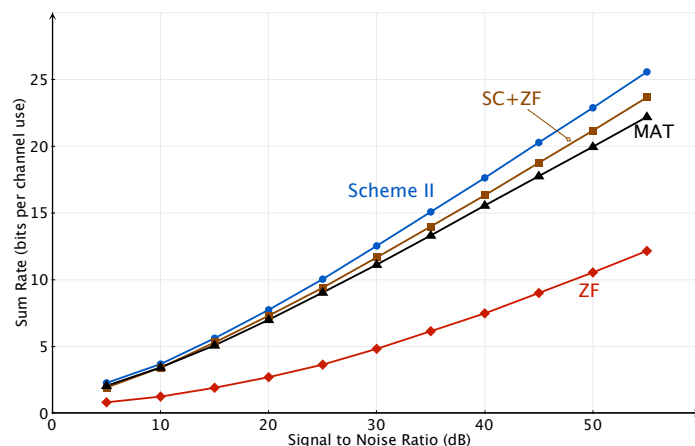

Fig. 3.   Comparison of the achievable ergodic sum-rate between the proposed scheme and the zero-forcing and MAT alignment. We set $\alpha = 0.5$.

We observe that as the quality of channel knowledge increases ($\alpha \to 0$) the sum rate improves significantly with the sharper slope promised by the DoF result. Notice that the performance with $\alpha = 0$ nearly corresponds to the sum rate achieved by MAT(cf. Fig. 3). In Figure 3, we compare different strategies: Scheme II, MAT, ZF, as well as "SC+ZF" in terms of the ergodic sum rate for $\alpha = 0.5$. For this quality of the current CSIT, ZF performs substantially worse than the others, achieving the pre-log of one. The sum rate with MAT, SC+ZF, scheme II increases with a slope of 4/3, 3/2, 5/3, respective, as expected from the DoF results.

VI. EXTENSIONS AND DISCUSSIONS

A. DoF with common message

The main result of this paper can be extended trivially to the case with common message.

Corollary 1: Let (d_0, d_1, d_2) be the degrees of freedom related to the common message, private message to user 1, and private message to user 2, respectively. Then, the optimal DoF region in terms of (d_0, d_1, d_2) is characterized by

$$d_0 + d_1 \leq 1 \tag{86a}$$

$$d_0 + d_2 \leq 1 \tag{86b}$$

$$2d_0 + d_1 + 2d_2 \leq 2 + \alpha \tag{86c}$$

$$2d_0 + 2d_1 + d_2 \leq 2 + \alpha \tag{86d}$$

Proof: The converse follows the same lines as in the case without common message, presented in Section IV. To obtain (86b) and (86c), we replace W_2 by $\tilde{W}_2 \triangleq (W_0, W_2)$ and R_2 by $\tilde{R}_2 \triangleq R_0 + R_2$ throughout Section IV and carry out exactly the same steps. To obtain the other two inequalities, we interchange the roles of user 1 and user 2. It follows straightforwardly due to the symmetry between the two users.

Note that the region is a polyhedron and completely characterized by the vertices in terms of (d_0, d_1, d_2):

- extreme points: $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$
- private points: $(0, 1, \alpha)$, $(0, \alpha, 1)$, $\left(0, \frac{2+\alpha}{3}, \frac{2+\alpha}{3}\right)$
- mixed point: $(1 - \alpha, \alpha, \alpha)$

which are all achievable with the proposed scheme. Therefore, the entire region is achievable by time sharing between the vertices. ∎

B. General fading processes

The main results are based on the fact that the channel gains are Rayleigh distributed and the fading processes are stationary and ergodic. However, a close examination on the achievability and the converse proofs reveals the results still hold in a more general setting. In the following, we discuss briefly some possible relaxations.

1) Non-Rayleigh fading: The Rayleigh distribution has been introduced for tractability of the achievable rates and the upper bound. This constraint can be relaxed for the DoF analysis. In order to obtain the same upper bound on the DoF, we need the following constraints:

- The estimation error $\tilde{S}$ should follow an isotropic (unitary invariant) distribution. Otherwise, it would be possible to exploit the statistical property of the direction of $\tilde{S}$ to precode the signal. The DoF can be thus be potentially increased.[3]
- $\mathbb{E}\left(\log(\|\boldsymbol{g}\|^2)\right)$, $\mathbb{E}\left(\log(\|\boldsymbol{h}\|^2)\right)$, $\mathbb{E}\left(\log(\|\hat{\boldsymbol{g}}\|^2)\right)$, and $\mathbb{E}\left(\log(\|\hat{\boldsymbol{h}}\|^2)\right)$ are finite, i.e., in $(-\infty, \infty)$.

On the other hand, in order to have the same achievability results, we need the second one of the above constraints as well as $\mathbb{E}\left(\log \det\left(\boldsymbol{S}\boldsymbol{S}^{\mathsf{H}}\right)\right) > -\infty$.

---

[3]This constraint may be further relaxed as follows. Let $\mathcal{V}_\delta$ be a set of $\mathbb{C}^{m \times m}$ unitary matrices and the Haar measure of this set is $\mu(\mathcal{V}_\delta) = \delta$. It can be shown that as long as the probability of $\tilde{S}$ lying within $\left\{\tilde{S}_0 \boldsymbol{V} : \boldsymbol{V} \in \mathcal{V}_\delta\right\}$ scales as δ when $\delta \to 0$, for any $\boldsymbol{S}_0$ and $\boldsymbol{V}_\delta$, the DoF results still hold. Detailed proofs are omitted.

2) Non-ergodic fading (delay-limited communications): The DoF results have been derived based on the ergodic rates. For non-ergodic fading processes, the DoF needs to be redefined. Here, we focus on the case with delay-limited communication, i.e., when n is finite. Note that the original DoF definition does not hold any more, since the original reliability condition (8) and (9) hinges on the fact that $n \to \infty$. However, it is possible to modify the definition by replacing the limit on n in (8) by a limit on $P \to \infty$ and removing the limit on n in (9). Following the definition of multiplexing gain in [5], we can redefine the achievable degrees of freedom in the delay-limited case as follows.

Definition 3 (achievable degrees of freedom in the non-ergodic case): For a family of codes $\{\mathfrak{X}(P)\}$ of length n and rate $R_k(P)$ bits per channel use for user k, $k = 1, 2$, we let $P_{e,k}(P)$ be the average probability of error for user k and let

$$r_k \triangleq \lim_{P \to \infty} \frac{R_k(P)}{\log P}, \quad k = 1, 2. \tag{87}$$

Then, the achievable degrees of freedom of $\mathfrak{X}$ is defined as

$$\mathsf{DoF}_k \triangleq \sup \left\{ r : \lim_{P \to \infty} -\frac{\log P_{e,k}(P)}{\log P} > 0 \right\}. \tag{88}$$

In other words, the DoF defined above is the maximum pre-log factor of the rate of a coding scheme for a reliable communication in the high SNR regime. Note that the code length n here is fixed.

The ergodicity is not necessary for the converse proof with the new reliability condition. This is because the upper bounds have been established in a symbol-wise manner. The only detail to take care of is the term $n\epsilon_n$ from Fano's inequality. Note that $n\epsilon_n = 1 + nP_eR$. In the original setting where we can have $n \to \infty$, P_e goes to 0 with n for any given rate R, which implies $\epsilon_n \to 0$ with n. In the new setting with finite n, P_e goes to 0 with P while R goes to ∞ with P. Fortunately, (87) and (88) ensure that the decaying rate of P_e (polynomial in P) is faster than the increasing rate of R (logarithmic in P). Therefore, we have $\epsilon_n = P_eR \to 0$ with P. As a result, every step in the converse remains the same as in the original setting.

The achievability proof needs considerably more modifications for finite n for the following two reasons. First, the rate-distortion function has been used to establish an upper bound on the quantization rate, while it is valid only when $n \to \infty$. Second, the rate of the equivalent MIMO channel has been directly related to the mutual information, while the achievability is also based on the assumption $n \to \infty$. Same arguments also hold for the second stage of the proposed scheme. To circumvent these limitations, we propose a simple quantization scheme as well as a random coding scheme with finite length. The achievability is shown by analyzing the error probabilities. It can be shown that the achievable region of $(\mathsf{DoF}_1, \mathsf{DoF}_2)$ is the same as the non-ergodic case. Details are deferred to Appendix D.

C. Imperfect delayed CSI: Limited feedback

In most practical scenarios, delayed CSIT is obtained through feedback channel and the current state is then predicted based on the delayed CSIT. Due to various reasons, perfect delayed CSIT may not be available. For
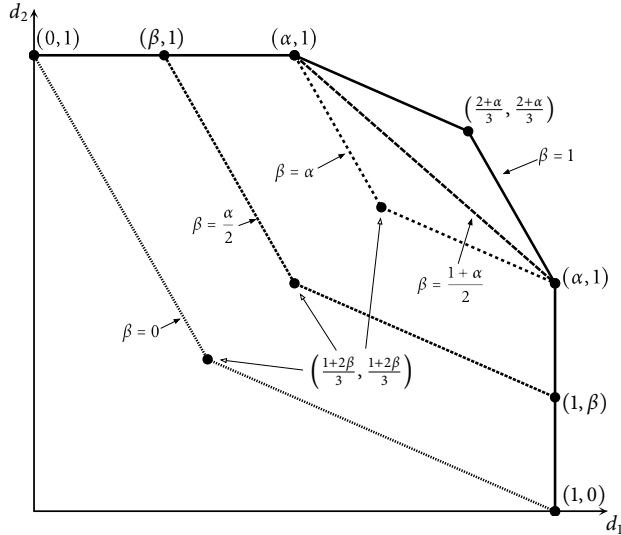
Fig. 4. Impact of imperfect delayed CSI on the DoF region. We fix $\alpha = 0.5$ and vary β from 1 to 0.

instance, the limited feedback rate may incur a distortion on the channel coefficients. In the following, we take a look at the impact of the imperfect delayed CSIT on the achievable degrees of freedom of the proposed scheme.

First, let us assume that the channel state $\boldsymbol{S}_{t-1}$ is quantized before being sent back to the transmitter (and the other receiver). The quantization model is

$$\boldsymbol{S}_{t-1} = \bar{\boldsymbol{S}}_{t-1} + \breve{\boldsymbol{S}}_{t-1} \tag{89}$$

where each entry of the quantization noise $\breve{\boldsymbol{S}}_{t-1}$ has the same variance σ_{FB}^2. We introduce a parameter β to characterize the precision of the quantization. As the definition of α, we define β as the power exponent of the quantization noise[4], i.e.,

$$\beta \triangleq -\frac{\log \sigma_{\text{FB}}^2}{\log P}. \tag{90}$$

Due to the lack of perfect delayed CSIT, instead of using $\boldsymbol{S}^{t-1}$ to predict $\boldsymbol{S}_t$ for the precoding and using $\boldsymbol{S}_{t-1}$ to perform the MAT alignment, the transmitter now predicts the quantized state $\bar{\boldsymbol{S}}_t$ with the past quantized state $\bar{\boldsymbol{S}}^{t-1}$ and uses $\bar{\boldsymbol{S}}_{t-1}$ for the alignment. Therefore, although the actual interference seen by the receivers is $(\boldsymbol{h}^{\mathsf{H}}\boldsymbol{v}, \boldsymbol{g}^{\mathsf{H}}\boldsymbol{u})$, the transmitter only has access to a noisy version of it $\boldsymbol{\eta} = (\bar{\boldsymbol{h}}^{\mathsf{H}}\boldsymbol{v}, \bar{\boldsymbol{g}}^{\mathsf{H}}\boldsymbol{u})$. Receiver 1 has

$$y = \boldsymbol{h}^{\mathsf{H}}\boldsymbol{u} + \boldsymbol{h}^{\mathsf{H}}\boldsymbol{v} + \varepsilon = \boldsymbol{h}^{\mathsf{H}}\boldsymbol{u} + \eta_1 + (\boldsymbol{h} - \bar{\boldsymbol{h}})^{\mathsf{H}}\boldsymbol{v} + \varepsilon \tag{91}$$

$$\hat{\eta}_1 = \eta_1 - \Delta_1 \tag{92}$$

$$\hat{\eta}_2 = \eta_2 - \Delta_2 = \bar{\boldsymbol{g}}^{\mathsf{H}}\boldsymbol{u} - \Delta_2. \tag{93}$$

[4]From the rate-distortion function, it is not difficult to relate β to the resource required for the CSI feedback, i.e., the feedback DoF.

The power of $\boldsymbol{\eta}$ is $\bar{\boldsymbol{h}}^{\mathsf{H}}\boldsymbol{Q_v}\bar{\boldsymbol{h}} + \bar{\boldsymbol{g}}^{\mathsf{H}}\boldsymbol{Q_u}\bar{\boldsymbol{g}}$ that depends on the "precision" of the prediction from $\bar{\boldsymbol{S}}^{t-1}$ to $\bar{\boldsymbol{S}}_t$. It can be shown[5] that the power exponent of this said prediction error is $\alpha' \triangleq \min\{\alpha, \beta\}$ where α is the power exponent of the prediction error when perfect delayed CSIT is present, i.e., predicting $\boldsymbol{S}_t$ from $\boldsymbol{S}^{t-1}$. Therefore, the achievable DoF of Scheme II would be $\frac{2+\alpha'}{3}$ without taking into account the "residual interference" $(\boldsymbol{h} - \bar{\boldsymbol{h}})^{\mathsf{H}}\boldsymbol{v}$ in (91). In fact, this interference costs a DoF loss of $1 - \beta$ over three slots, yielding the new DoF per user

$$d(\alpha, \beta) = \frac{2 + \alpha' - (1 - \beta)}{3} = \frac{1 + \min\{\alpha, \beta\} + \beta}{3}. \tag{94}$$

As in the case with perfect delayed CSIT, the DoF pairs $(1, \alpha')$ and $(\alpha', 1)$ are achievable without the MAT alignment. An example of the DoF region is shown in Fig. 4, where we fix the value α and vary β from 1 to 0. As shown in the figure, when $\beta = 1$, the DoF region is unchanged. When β is reduced to $\frac{1+\alpha}{2}$, the symmetric DoF point can be achieved by time sharing between the two corner points $(1, \alpha)$ and $(\alpha, 1)$. Delayed CSIT is not beneficial any more. As β continues to diminish to α, the symmetric DoF keeps dropping while the corner points remain still. At this point, using MAT alignment creates more interference than resolving it. When β goes below α, it becomes the dominating source of interference. The corner points become $(1, \beta)$ and $(\beta, 1)$. The above analysis reveals that even imperfect delayed CSI can beneficial, as long as the feedback accuracy β is larger than $\frac{1+\alpha}{2}$.

D. Bandwidth-limited Doppler process

The main result on the achievable DoF has been presented in terms of an artificial parameter α, denoting the speed of decay of the estimation error $\sigma^2 \sim P^{-\alpha}$ in the current CSIT. In this section, we provide an example showing the practical interpretation of this parameter. Focusing on receiver 1 due to symmetry, we describe the fading process, channel estimation, and feedback scheme as follows:

- The channel fading $\boldsymbol{h}_t$ follows a Doppler process with power spectral density $S_h(w)$. The channel coefficients are strictly band-limited to $[-F, F]$ with $F = \frac{v f_c T_f}{c} < \frac{1}{2}$ where v, f_c, T_f, c denotes the mobile speed in m/h, the carrier frequency in Hz, the slot duration in sec, the light speed in m/sec.

- The channel estimation is done at the receivers side with pilot-based downlink training. At slot t, receiver 1 estimates $\boldsymbol{h}_t$ based on a sequence of the noisy observations $\{\boldsymbol{s}_\tau = \sqrt{P}\boldsymbol{h}_\tau + \boldsymbol{\nu}_\tau\}$ up to t, where $\boldsymbol{\nu}_t$ is AWGN with zero mean unit covariance. The estimate is denoted by $\bar{\boldsymbol{h}}_t$ with

$$\boldsymbol{h}_t = \bar{\boldsymbol{h}}_t + \breve{\boldsymbol{h}}_t. \tag{95}$$

  Under this model, the estimation error vanishes as $\mathbb{E}\left(\|\breve{\boldsymbol{h}}_t\|^2\right) \sim P^{-1}$.

- At the end of slot t, the noisy observation $\boldsymbol{s}_t$ is sent to the transmitter and receiver 2 over a noise-free channel. At slot $t + 1$, based on the noisy observation $\{\boldsymbol{s}_\tau\}$ up to t, the transmitter and receiver 2 acquire the prediction

---

[5]Without going into the details, we can see that the following Markov chain holds $\bar{\boldsymbol{S}}^{t-1} \leftrightarrow \boldsymbol{S}^{t-1} \leftrightarrow \boldsymbol{S}_t \leftrightarrow \bar{\boldsymbol{S}}_t$. The prediction error from $\bar{\boldsymbol{S}}^{t-1}$ to $\bar{\boldsymbol{S}}_t$ is now the aggregation of two effects: the channel variation, characterized by $P^{-\alpha}$, and the quantization error due to limited feedback rate, characterized by $P^{-\beta}$. Hence, we have the power exponent of the aggregated error $\alpha' = \min\{\alpha, \beta\}$.

$\hat{h}_{t+1}$ of h_{t+1} and estimation $\tilde{h}_t$ of h_t. The corresponding prediction model is

$$h_t = \hat{h}_t + \tilde{h}_t. \tag{96}$$

From [2, Lemma 1], we have $\mathbb{E}\left(\|\tilde{h}_t\|^2\right) \sim P^{-(1-2F)}$.

In this channel with imperfect delayed CSIT, we can still apply the proposed scheme and analysis in exactly the same way as in the previous section with $\alpha = 1 - 2F$ and $\beta = 1$.

VII. Conclusions

A scheme achieving the optimal degrees of freedom region in a two-user MISO broadcast channel has been presented. The approach optimally exploits the combination of delayed channel feedback together with current imperfect CSIT. In practical scenarios, the current CSIT may be obtained from a prediction based on the delayed CSIT samples. When the quality of current CSIT is poor, the proposed scheme coincides with the previously reported MAT algorithm, whereas as the current CSIT prediction quality becomes ideal, the scheme relies on standard linear precoding. In between these extremal regimes, the proposed strategy advocates interference quantization followed by feedback.

Appendix

A. Achievable rate region of Scheme I

Let us recall that Scheme I consists of two phases: broadcast and multicast. In the following, let n_1 and n_2 denote the length of Phase 1 and Phase 2, in channel uses, respectively. The main ingredients are:

- Codebook generation:
  - Channel codebooks $\mathcal{X}_{\tilde{u}}$ of length n_1 and size $2^{n_1 R_{\mathsf{mimo},1}}$, $\mathcal{X}_{\tilde{v}}$ of length n_1 and size $2^{n_1 R_{\mathsf{mimo},2}}$. Entries of $\mathcal{X}_{\tilde{u}}$ and $\mathcal{X}_{\tilde{u}}$ are generated i.i.d. according to $\mathcal{N}_{\mathbb{C}}\left(0, \mathbf{\Lambda}_u\right)$ and $\mathcal{N}_{\mathbb{C}}\left(0, \mathbf{\Lambda}_v\right)$, respectively. $\mathbf{\Lambda}_u, \mathbf{\Lambda}_v \succeq 0$ are $m \times m$ matrices that can be assumed to be diagonal without loss of generality.
  - Channel codebook $\mathcal{X}_{\mathsf{mc}}$ of length n_2 and size $2^{n_2 R_{\mathsf{mc}}}$, the entries of which is generated i.i.d. according to $\mathcal{N}_{\mathbb{C}}\left(0, \mathbf{\Lambda}_{\mathsf{mc}}\right)$, where $\mathbf{\Lambda}_{\mathsf{mc}} \succeq 0$ is $m \times m$ diagonal matrix.
  - Source codebooks $\mathcal{C}_k$ of length n_1 and size $2^{n_1 R_{\eta_k}}$, $k = 1, 2$. Entries of $\mathcal{C}_1$ and $\mathcal{C}_2$ are generated i.i.d. according to $\mathcal{N}_{\mathbb{C}}\left(0, 1 - \tilde{D}_k\right)$, $\tilde{D}_k \leq 1$, $k = 1, 2$.

- Time-varying linear precoders that only depend on the estimate of the current state:

$$\mathbf{\Theta}_t, \mathbf{\Phi}_t, \mathbf{\Omega}_t : \hat{\mathcal{S}}_t \longmapsto \mathbb{C}^{m \times m}. \tag{97}$$

- Coding in Phase 1: The codewords $\{\tilde{u}_t\}_{t=1}^n$ and $\{\tilde{v}_t\}_{t=1}^n$ are selected from $\mathcal{X}_{\tilde{u}}$ and $\mathcal{X}_{\tilde{v}}$, according to $W_{\mathsf{mimo},1}, W_{\mathsf{mimo},2}$ respectively. The transmitted signal is $\{x_t\}_{t=1}^n$ with

$$x_t = u_t + v_t \tag{98}$$

$$= \mathbf{\Theta}_t \tilde{u}_t + \mathbf{\Phi}_t \tilde{v}_t. \tag{99}$$

- Quantization of the interferences $\boldsymbol{\eta}$: At the end of Phase 1, the transmitter knows $\{(\eta_{1,t}, \eta_{2,t})\}_{t=1}^{n_1}$ with $\eta_{1,t} \triangleq \boldsymbol{h}_t^{\mathsf{H}} \boldsymbol{v}_t \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_{\eta_{1,t}}^2)$ and $\eta_{2,t} \triangleq \boldsymbol{g}_t^{\mathsf{H}} \boldsymbol{u}_t \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_{\eta_{2,t}}^2)$, for a given channel realization $\{\boldsymbol{h}_t, \boldsymbol{g}_t\}_{t=1}^{n_1}$. The codebook $\mathcal{C}_k$, $k = 1, 2$, is used to quantize the normalized source $\left\{ \frac{\eta_{k,t}}{\sigma_{\eta_{k,t}}} \right\}_{t=1}^{n_1}$ that is i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1)$. The quantized output is represented in $n_1(R_{\eta_1} + R_{\eta_2})$ bits.

- Coding in Phase 2: The quantized interference, denoted by W_{mc}, is coded into $\{\tilde{\boldsymbol{x}}_{\mathsf{mc},t}\}_{t=n_1+1}^{n_1+n_2}$ using codebook $\mathcal{X}_{\mathsf{mc}}$. It will be then precoded through the matrix $\{\boldsymbol{\Omega}_t\}$ and then multicast to both users. The transmitted signal is $\{\boldsymbol{x}_{\mathsf{mc},t} with\}_{t=n_1+1}^{n_1+n_2}$

$$\boldsymbol{x}_{\mathsf{mc},t} = \boldsymbol{\Omega}_t \tilde{\boldsymbol{x}}_{\mathsf{mc},t}. \tag{100}$$

For user k to recover its original message W_k correctly[6], when $n_1, n_2 \to \infty$, it is enough to

- recover the message W_{mc} at each receiver, which is possible if

$$n_1(R_{\eta_1} + R_{\eta_2}) \leq n_2 R_{\mathsf{mc}} \tag{101}$$

and

$$R_{\mathsf{mc}} < \min\{I(\tilde{X}_{\mathsf{mc}}; Y_{\mathsf{mc}} \mid S, \hat{S}), I(\tilde{X}_{\mathsf{mc}}; Z_{\mathsf{mc}} \mid S, \hat{S})\} \tag{102}$$

where Y_{mc} and Z_{mc} are the received signals corresponding to X_{mc} at receiver 1 and 2, respectively.

- reconstruct $\{\hat{\eta}_{k,t}\}_{t=1}^{n_1}$, $k = 1, 2$ with

$$\eta_{k,t} = \hat{\eta}_{k,t} + \Delta_{k,t}, \quad \Delta_{k,t} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_{\eta_{k,t}}^2 \tilde{D}_k), \tag{103}$$

which is possible if

$$R_{\eta_k} > \log\left(\frac{1}{\tilde{D}_k}\right), \quad k = 1, 2; \tag{104}$$

- then decode the message $W_{\mathsf{mimo},k}$, which is possible if

$$R_{\mathsf{mimo},1} < I(\tilde{U}; Y, \hat{\eta}_1, \hat{\eta}_2 \mid S, \hat{S}) \tag{105}$$

$$R_{\mathsf{mimo},2} < I(\tilde{V}; Z, \hat{\eta}_1, \hat{\eta}_2 \mid S, \hat{S}). \tag{106}$$

Putting all pieces together, we obtain the rate region of Scheme I in the following.

Proposition 1: Let us define the multicast rate R_{mc}, the compression rate R_{η_k}, and the MIMO rate as

$$R_{\mathsf{mc}} \triangleq \min\{\mathbb{E}\left(\log\left(1 + \boldsymbol{h}^{\mathsf{H}} \boldsymbol{Q}_{\mathsf{mc}} \boldsymbol{h}\right)\right), \mathbb{E}\left(\log\left(1 + \boldsymbol{g}^{\mathsf{H}} \boldsymbol{Q}_{\mathsf{mc}} \boldsymbol{g}\right)\right)\} \tag{107}$$

$$R_{\eta_k} \triangleq \log\frac{1}{\tilde{D}_k}, \quad k = 1, 2 \tag{108}$$

$$R_{\mathsf{mimo},1} \triangleq \mathbb{E}\left(\log\det\left(\mathbf{I} + \boldsymbol{D}_1 \boldsymbol{S} \boldsymbol{Q}_{\boldsymbol{u}} \boldsymbol{S}^{\mathsf{H}}\right)\right) \tag{109}$$

$$R_{\mathsf{mimo},2} \triangleq \mathbb{E}\left(\log\det\left(\mathbf{I} + \boldsymbol{D}_2 \boldsymbol{S} \boldsymbol{Q}_{\boldsymbol{v}} \boldsymbol{S}^{\mathsf{H}}\right)\right) \tag{110}$$

---

[6]Note that the assumption on the ergodicity and the Markov chain (5) makes the single-letter representation of the rates possible.

with

$$\boldsymbol{D}_1 \triangleq \mathrm{diag}\left(\frac{1}{1 + \boldsymbol{h}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{v}}\boldsymbol{h}\,\tilde{D}_1}, \ \frac{1 - \tilde{D}_2}{\boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{g}\,\tilde{D}_2}\right) \tag{111}$$

$$\boldsymbol{D}_2 \triangleq \mathrm{diag}\left(\frac{1 - \tilde{D}_1}{\boldsymbol{h}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{v}}\boldsymbol{h}\,\tilde{D}_1}, \ \frac{1}{1 + \boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{g}\,\tilde{D}_2}\right). \tag{112}$$

Then, the achievable rate region of Scheme I is the union of the rate pairs (R_1, R_2) such that

$$R_k = \frac{R_{\mathsf{mc}}\,R_{\mathsf{mimo},k}}{R_{\mathsf{mc}} + R_{\eta,1} + R_{\eta,2}} \tag{113}$$

over all policies $D(\hat{\boldsymbol{S}}) \triangleq \{\tilde{D}_1, \tilde{D}_2 : \ 0 \le \tilde{D}_k \le 1\}$ and

$$Q(\hat{\boldsymbol{S}}) \triangleq \{\boldsymbol{Q}_{\boldsymbol{u}}, \boldsymbol{Q}_{\boldsymbol{v}}, \boldsymbol{Q}_{\mathsf{mc}} \succeq 0 : \ \mathrm{tr}\,(\boldsymbol{Q}_{\boldsymbol{u}} + \boldsymbol{Q}_{\boldsymbol{v}}) \le P, \ \mathrm{tr}\,(\boldsymbol{Q}_{\mathsf{mc}}) \le P\} \tag{114}$$

that only depend on the estimate of the channels.

Proof: The average rate for user k is

$$R_k = \frac{n_1 R_{\mathsf{mimo},k}}{n_1 + n_2} \tag{115}$$

$$= \frac{R_{\mathsf{mimo},k}}{1 + n_2/n_1} \tag{116}$$

$$\le \frac{R_{\mathsf{mc}}\,R_{\mathsf{mimo},k}}{R_{\mathsf{mc}} + R_{\eta,1} + R_{\eta,2}} \tag{117}$$

where the last inequality is from (101). (107) can be obtained from (102), with $\boldsymbol{Q}_{\mathsf{mc}} = \boldsymbol{\Omega}\boldsymbol{\Lambda}_{\mathsf{mc}}\boldsymbol{\Omega}^{\mathsf{H}}$. To see (109), we write

$$I(\tilde{U}; Y, \hat{\eta}_1, \hat{\eta}_2 \mid S = \boldsymbol{S}, \hat{S} = \hat{\boldsymbol{S}}) = I(\tilde{U}; \hat{\eta}_1) + I(\tilde{U}; Y, \hat{\eta}_2 \mid \hat{\eta}_1) \tag{118}$$

$$= I(\tilde{U}; Y, \hat{\eta}_2 \mid \hat{\eta}_1) \tag{119}$$

$$= I(\tilde{U}; Y - \hat{\eta}_1, \hat{\eta}_2 \mid \hat{\eta}_1) \tag{120}$$

$$= I(\tilde{U}; \boldsymbol{h}^{\mathsf{H}}\tilde{U} + \Delta_1 + E, \hat{\eta}_2) \tag{121}$$

$$= I(\tilde{U}; \boldsymbol{h}^{\mathsf{H}}\tilde{U} + \Delta_1 + E, a_{\mathsf{mmse}}\,\boldsymbol{g}^{\mathsf{H}}\tilde{U} + E_{\mathsf{mmse}}) \tag{122}$$

$$= \log\det\left(\mathbf{I} + \mathrm{diag}\left((1 + \boldsymbol{h}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{v}}\boldsymbol{h}\,\tilde{D}_1)^{-1}, \frac{1 - \tilde{D}_2}{\boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{g}\,\tilde{D}_2}\right)\boldsymbol{S}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{S}^{\mathsf{H}}\right) \tag{123}$$

where (118) is from the chain rule of mutual information; (119) is from the fact that $\tilde{U}$ is independent of η_1; (121) is from the fact that $\hat{\eta}_1$ is independent of all the other terms. Since $\eta_2 = \hat{\eta}_2 + \Delta_2$ with $\hat{\eta}_2 \sim \mathcal{N}_{\mathbb{C}}\left(0, \boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{g}(1 - \tilde{D}_2)\right)$ and $\Delta_2 \sim \mathcal{N}_{\mathbb{C}}\left(0, \boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{g}\,\tilde{D}_2\right)$ being additive Gaussian noise, we can "estimate" $\hat{\eta}_2$ from η_2 and get the "backward channel" model

$$\hat{\eta}_2 = a_{\mathsf{mmse}}\,\eta_2 + e_{\mathsf{mmse}} \tag{124}$$

where $a_{\mathsf{mmse}} \triangleq 1 - \tilde{D}_2$ and the additive noise $e_{\mathsf{mmse}} \sim \mathcal{N}_{\mathbb{C}}\left(0, a_{\mathsf{mmse}}\boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{g}\,\tilde{D}_2\right)$ is independent of η_2. Thus, (123) follows as the mutual information of an equivalent Gaussian MIMO channel with Gaussian input, where

$\boldsymbol{Q_u} = \boldsymbol{\Theta}\boldsymbol{\Lambda}_u\boldsymbol{\Theta}^{\mathsf{H}}$ and $\boldsymbol{Q_v} = \boldsymbol{\Phi}\boldsymbol{\Lambda}_v\boldsymbol{\Phi}^{\mathsf{H}}$. Finally, (109) follows from (105) and (123). Due to the symmetry, (110) is straightforward. ∎

Note that the optimization in (113) is not trivial and is out of the scope of this paper. Instead of finding the exact rate, we focus on the symmetric degrees of freedom of the scheme with $m = 2$, by fixing the parameters Q and D as follows:

$$\boldsymbol{Q_u} = \frac{P_1}{2}\boldsymbol{\Psi}_{\hat{\boldsymbol{g}}^{\perp}} + \frac{P_2}{2}\boldsymbol{\Psi}_{\hat{\boldsymbol{g}}} \tag{125}$$

$$\boldsymbol{Q_v} = \frac{P_1}{2}\boldsymbol{\Psi}_{\hat{\boldsymbol{h}}^{\perp}} + \frac{P_2}{2}\boldsymbol{\Psi}_{\hat{\boldsymbol{h}}} \tag{126}$$

$$\boldsymbol{Q}_{\mathsf{mc}} = \frac{P}{2}\mathbf{I} \tag{127}$$

$$\tilde{D}_1 = \tilde{D}_2 = (2P)^{-1} \tag{128}$$

where we recall that $\boldsymbol{\Psi}_{\hat{\boldsymbol{g}}} \triangleq \frac{\hat{\boldsymbol{g}}\hat{\boldsymbol{g}}^{\mathsf{H}}}{\|\hat{\boldsymbol{g}}\|^2}$ and $\boldsymbol{\Psi}_{\hat{\boldsymbol{g}}^{\perp}}$, $\boldsymbol{\Psi}_{\hat{\boldsymbol{h}}}$, and $\boldsymbol{\Psi}_{\hat{\boldsymbol{h}}^{\perp}}$ are similarly defined. the power allocation (P_1, P_2) is specified by

$$P_2 = (1 - \hat{\alpha})\frac{P}{2}\hat{\sigma}^2 \tag{129a}$$

$$P_1 = P - P_2 \tag{129b}$$

with

$$\hat{\sigma}^2 \triangleq \max\left\{P^{-1}, \sigma^2\right\}, \quad \hat{\alpha} \triangleq -\frac{\log \hat{\sigma}^2}{\log P}. \tag{130}$$

The interpretation of the choices on the covariance matrices has already been given in Section III-C. For the power allocation (129) and the distortion (128):

- The scaling factor $(1 - \hat{\alpha})$ ensures that $P_1 = P$ and $P_2 = 0$ when the estimation error is small, i.e., $\sigma^2 \le P^{-1}$ while $P_1 = P_2 = \frac{P}{2}$ when the estimation error is high, i.e., $\sigma^2 = 1$.
- The distortion $\tilde{D}_1$ and $\tilde{D}_2$ are such that the error after reconstruction of $\boldsymbol{\eta}$ is at the same level as the channel noise.

It is readily shown that with the above choices, we obtain (23) and then the DoF (2).

B. Achievable rate region of Scheme II

The only difference between Scheme I and II is the private messages sent in second phase. That is, we need the following, in addition to the ingredients of Scheme I:

- Channel codebooks $\mathcal{X}_{p,1}$ of length n_2 and size $2^{n_2 R_{p,1}}$ and $\mathcal{X}_{p,2}$ of length n_2 and size $2^{n_2 R_{p,2}}$. Entries of $\mathcal{X}_{p,k}$, $k = 1, 2$, are generated i.i.d. according to $\mathcal{N}_{\mathbb{C}}(0, \boldsymbol{\Lambda}_{p,k})$, with $\boldsymbol{\Lambda}_{p,k} \succeq 0$ being $m \times m$ diagonal.
- Instead of $\mathcal{X}_{\mathsf{mc}}$, we use a channel codebook $\mathcal{X}_0$ of length n_2 and size $2^{n_2 R_0}$, the entries of which is generated i.i.d. according to $\mathcal{N}_{\mathbb{C}}(0, \boldsymbol{\Lambda}_0)$, with $\boldsymbol{\Lambda}_0 \succeq 0$ being $m \times m$ diagonal.
- Time-varying linear precoders that only depend on the estimate of the current state:

$$\boldsymbol{\Xi}_t, \boldsymbol{\Gamma}_t, \boldsymbol{\Omega}_t : \hat{\mathcal{S}}_t \longmapsto \mathbb{C}^{m \times m}.$$

- Coding in Phase 2: The quantized interferences, represented in $n_1(R_{\eta_1} + R_{\eta_2})$ bits and denoted by W_0, is coded in $\{\tilde{x}_{c,t}\}_{t=n_1+1}^{n_1+n_2} \in \mathcal{X}_0$, precoded, and then multicast to both users. Meanwhile, two private messages $W_{p,1}$ and $W_{p,2}$ for user 1 and 2 are coded in $\{\tilde{v}_{p,t}\}_{t=n_1+1}^{n_1+n_2} \in \mathcal{X}_{p,1}$ and $\{\tilde{u}_{p,t}\}_{t=n_1+1}^{n_1+n_2} \in \mathcal{X}_{p,2}$, respectively, precoded, and sent. The transmitted signal is $\{\boldsymbol{x}_{\mathsf{mix},t}\}_{t=n_1+1}^{n_1+n_2}$

$$\boldsymbol{x}_{\mathsf{mix},t} = \boldsymbol{\Omega}_t \tilde{\boldsymbol{x}}_{c,t} + \boldsymbol{\Xi}_t \tilde{\boldsymbol{u}}_{p,t} + \boldsymbol{\Gamma}_t \tilde{\boldsymbol{v}}_{p,t}. \tag{131}$$

Then, we can get the following result on the rate region.

Proposition 2: Let us define the common and private rates as

$$R_0 \triangleq \min\left\{ \mathbb{E}\left(\log\left(1 + \frac{\boldsymbol{h}^{\mathsf{H}}\boldsymbol{Q}_0\boldsymbol{h}}{1 + \boldsymbol{h}^{\mathsf{H}}(\boldsymbol{Q}_{p1} + \boldsymbol{Q}_{p2})\boldsymbol{h}}\right)\right), \mathbb{E}\left(\log\left(1 + \frac{\boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q}_0\boldsymbol{g}}{1 + \boldsymbol{g}^{\mathsf{H}}(\boldsymbol{Q}_{p1} + \boldsymbol{Q}_{p2})\boldsymbol{g}}\right)\right)\right\} \tag{132}$$

$$R_{p,1} \triangleq \mathbb{E}\left(\log\left(1 + \frac{\boldsymbol{h}^{\mathsf{H}}\boldsymbol{Q}_{p1}\boldsymbol{h}}{1 + \boldsymbol{h}^{\mathsf{H}}\boldsymbol{Q}_{p2}\boldsymbol{h}}\right)\right), \quad R_{p,2} \triangleq \mathbb{E}\left(\log\left(1 + \frac{\boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q}_{p2}\boldsymbol{g}}{1 + \boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q}_{p1}\boldsymbol{g}}\right)\right) \tag{133}$$

and the compression rate R_{η_k} and the MIMO rate $R_{\mathsf{mimo},k}$ are defined by (108) and (109), respectively. Then, the achievable rate region of Scheme II is the union of the rate pairs (R_1, R_2) such that

$$R_k = \frac{R_0\, R_{\mathsf{mimo},k} + (R_{\eta,1} + R_{\eta,2})R_{p,k}}{R_0 + R_{\eta,1} + R_{\eta,2}} \tag{134}$$

over all policies $D(\hat{\boldsymbol{S}}) \triangleq \{\tilde{D}_1, \tilde{D}_2 : \ 0 \le \tilde{D}_k \le 1\}$ and

$$Q'(\hat{\boldsymbol{S}}) \triangleq \{\boldsymbol{Q}_{\boldsymbol{u}}, \boldsymbol{Q}_{\boldsymbol{v}}, \boldsymbol{Q}_0, \boldsymbol{Q}_{p,1}, \boldsymbol{Q}_{p,2} \succeq 0: \ \mathrm{tr}\,(\boldsymbol{Q}_{\boldsymbol{u}} + \boldsymbol{Q}_{\boldsymbol{v}}) \le P, \ \mathrm{tr}\,(\boldsymbol{Q}_0 + \boldsymbol{Q}_{p,1} + \boldsymbol{Q}_{p,2}) \le P\} \tag{135}$$

that only depend on the estimate of the channels.

Proof: The only differences from Scheme I are the multicast rate R_{mc} is now replaced by the common rate R_0, and that extra private rate is obtained. The common rate R_0 is obtained by treating the private signals as noise. Then, after removing the decoded common signal, the private message is obtained by treating the interference as noise. The covariance matrices are such that $\boldsymbol{Q}_0 = \boldsymbol{\Omega}\boldsymbol{\Lambda}_0\boldsymbol{\Omega}^{\mathsf{H}}$, $\boldsymbol{Q}_{p,1} = \boldsymbol{\Xi}\boldsymbol{\Lambda}_{p,1}\boldsymbol{\Xi}^{\mathsf{H}}$, $\boldsymbol{Q}_{p,2} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}_{p,2}\boldsymbol{\Gamma}^{\mathsf{H}}$. Further details are omitted. ∎

As with Scheme I, we focus on the symmetric case with $m = 2$ and fix the parameters Q and D as follows:

$$\boldsymbol{Q}_0 = \frac{P_c}{2}\mathbf{I} \tag{136}$$

$$\boldsymbol{Q}_{p1} = \frac{P_p}{2}\boldsymbol{\Psi}_{\hat{\boldsymbol{g}}^{\perp}} \tag{137}$$

$$\boldsymbol{Q}_{p2} = \frac{P_p}{2}\boldsymbol{\Psi}_{\hat{\boldsymbol{h}}^{\perp}} \tag{138}$$

$$\tilde{D}_1 = \tilde{D}_2 = (2P)^{-1} \tag{139}$$

where the power allocation (P_c, P_p) is specified by

$$P_p = \hat{\alpha}\hat{\sigma}^{-2} \tag{140a}$$

$$P_c = P - P_p \tag{140b}$$

where $\hat{\sigma}^2$ and $\hat{\alpha}$ are defined in (130). The interpretation of the above choices of $\boldsymbol{Q}_{p1}$ and $\boldsymbol{Q}_{p2}$ in the second phase is:

- The transmitted power of the private signals scales as $P_p \sim P^\alpha$, while the received power at the unintended receiver scale as P^0, i.e., the noise level. Therefore, the private signal does not incur any DoF loss for the unintended receiver.

- The scaling factor $\hat{\alpha}$ in (140a) is such that $P_p = P$ and $P_c = 0$ when the estimation error is small, i.e., $\sigma^2 \leq P^{-1}$ while leading to $P_c = 0$ and $P_p = P$ when the estimation error is high, i.e., $\sigma^2 = 1$.

It is readily shown that, with these choices, we have the high SNR approximation of the rates

$$R_0 = (1 - \alpha) \log P + O(1) \tag{141}$$

$$R_{p,k} = \alpha \log P + O(1), \quad k = 1, 2 \tag{142}$$

$$R_{\boldsymbol{\eta}} = 2(1 - \alpha) \log P + O(1) \tag{143}$$

$$R_{\text{mimo},k} = (2 - \alpha) \log P + O(1), \quad k = 1, 2 \tag{144}$$

from which we derive the DoF of Scheme II

$$d_{\text{Scheme II}} = \frac{2 + \alpha}{3}, \quad \alpha \in [0, 1]. \tag{145}$$

C. Proof of Lemma 1

First, we show (74) as follows.

$$\mathbb{E}_{S_i | \hat{S}_i} \left(\log(1 + \boldsymbol{h}_i^{\mathsf{H}} \boldsymbol{K} \boldsymbol{h}_i) \right) \leq \mathbb{E}_{S_i | \hat{S}_i} \left(\log(1 + \lambda_1 \|\boldsymbol{h}_i\|^2) \right) \tag{146}$$

$$\leq \log(1 + \lambda_1 \|\hat{\boldsymbol{h}}_i\|^2 + m\sigma^2 \lambda_1) \tag{147}$$

$$= \log(1 + \lambda_1 \|\hat{\boldsymbol{h}}_i\|^2) + \log \left(1 + \frac{m\sigma^2 \lambda_1}{1 + \|\hat{\boldsymbol{h}}_i\|^2 \lambda_1} \right) \tag{148}$$

$$\leq \log(1 + \lambda_1 \|\hat{\boldsymbol{h}}_i\|^2) + \log \left(1 + \frac{m\sigma^2}{\|\hat{\boldsymbol{h}}_i\|^2} \right) \tag{149}$$

where (147) is from the concavity of the log function. To derive (75), let us define $\hat{\boldsymbol{\psi}} \triangleq \boldsymbol{V}^{\mathsf{H}} \hat{\boldsymbol{g}}_i$ and $\tilde{\boldsymbol{\psi}} \triangleq \boldsymbol{V}^{\mathsf{H}} \tilde{\boldsymbol{g}}_i$ with $\boldsymbol{V}$ being the unitary matrix containing the eigenvectors of $\boldsymbol{K}$, i.e., $\boldsymbol{K} = \boldsymbol{V} \text{diag}(\lambda_1, \ldots, \lambda_m) \boldsymbol{V}^{\mathsf{H}}$. From isotropic assumption, $\tilde{\psi}_i$ are i.i.d. $\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$. We also need the following lemma.

Lemma 3: Let $X \sim \mathcal{N}_{\mathbb{C}}(\mu_X, \sigma_X^2)$ and define $\phi(\mu_X, \sigma_X^2) \triangleq \mathbb{E}(\log(|X|^2))$, then $\phi(\mu_X, \sigma_X^2)$ is monotonically increasing and concave in $|\mu_X|^2$ and we have

$$\phi(\mu_X, \sigma_X^2) \geq \phi(\mu_X, 0) \tag{150}$$

$$\phi(0, \sigma_X^2) = \log(e^{-\gamma} \sigma_X^2), \tag{151}$$

where γ is Euler's constant.

Proof: From [11, Lemma 10.1][7], we have

$$\phi \left(\frac{\mu_X}{\sigma_X}, 1 \right) = \log \left(\frac{|\mu_X|^2}{\sigma_X^2} \right) - \text{Ei} \left(-\frac{|\mu_X|^2}{\sigma_X^2} \right) \log e \tag{152}$$

---

[7]Note that unlike in [11], the logarithm in this paper is in base 2.

where $\mathrm{Ei}(x)$ is the exponential integral with $-\mathrm{Ei}(-x) = \int_x^\infty \frac{e^{-t}}{t}\,dt$ being positive and decreasing in x for $x > 0$. The concavity and monotonicity are directly shown in [11]. To show (150),

$$\phi(\mu_X, \sigma_X^2) = \log(\sigma_X^2) + \phi\left(\frac{\mu_X}{\sigma_X}, 1\right) \tag{153}$$

$$\geq \log(|\mu_X|^2) \tag{154}$$

$$= \phi(\mu_X, 0) \tag{155}$$

where the inequality is from (152). From (153) and using the fact that $\phi(0,1) = -\gamma \log e$, we obtain (151). ∎

Then, we have

$$\mathbb{E}_{S_i|\hat{S}_i}\left(\log(1 + \boldsymbol{g}_i^\mathsf{T}\boldsymbol{K}\boldsymbol{g}_i)\right) = \mathbb{E}_{\tilde{\boldsymbol{\psi}}}\left(\log\left(1 + \sum_{j=1}^m \lambda_j |\hat{\psi}_j + \tilde{\psi}_j|^2\right)\right) \tag{156}$$

$$\geq \mathbb{E}_{\tilde{\psi}}\left(\log(\lambda_1|\hat{\psi}_1 + \tilde{\psi}_1|^2)\right)^+ \tag{157}$$

$$\geq \left(\mathbb{E}_{\tilde{\psi}}\left(\log(\lambda_1|\hat{\psi}_1 + \tilde{\psi}|^2)\right)\right)^+ \tag{158}$$

$$= \left(\mathbb{E}_{\tilde{\psi}}\left(\log(\lambda_1|\tilde{\psi}|^2)\right)\right)^+ \tag{159}$$

$$= \left(\log(e^{-\gamma}\lambda_1\sigma^2)\right)^+ \tag{160}$$

$$= \log(1 + e^{-\gamma}\lambda_1\sigma^2) - 1 \tag{161}$$

where in (157), $(x)^+$ means $\max\{x, 0\}$; (158) is from the fact that moving the maximization outside of the expectation only reduces the value; in (159) we use the monotonic increasing property of $\mathbb{E}_{\tilde{\psi}}\left(\log(|x + \tilde{\psi}|^2)\right)$ from Lemma 3; (160) is from (151); in (161), we use the fact that $(\log(x))^+ \geq \log(1 + x) - 1$.

D. Proposed scheme for delay-limited communications

In the following, we set the finite length $n = 3$, which is only a particular case of $n \geq 3$. For simplicity, we only analyze the extension to Scheme I. The extension to Scheme II is straightforward. The main line of the proposed scheme is the same as in the ergodic case. The only difference is that instead of using a Gaussian codebook to quantize $\boldsymbol{\eta}$, a truncated uniform quantization with unit step and truncation value $\bar{\eta} = P^{\frac{1+\varsigma}{2}}\sigma$, for some $\varsigma > 0$, is used for both the real and imaginary parts of η_1 and η_2, i.e.,

$$\hat{\eta}_k = [\mathrm{trunc}(\mathrm{Re}\,(\eta_k))] + i\,[\mathrm{trunc}(\mathrm{Im}\,(\eta_k))] \tag{162}$$

where $\mathrm{trunc}(x) = x$ if $x \in [-\bar{\eta}, \bar{\eta}]$ and 0 otherwise; $[x]$ means $\lfloor x \rfloor + \frac{1}{2}$. Thus, the double indices of $(\hat{\eta}_1, \hat{\eta}_2)$, represented in

$$4\log(2\lceil\bar{\eta}\rceil) \approx 4 + 2(1 + \varsigma - \alpha_P)\log P \quad \text{bits}, \tag{163}$$

are sent with a multicast code.

We define the error event $\mathcal{E}$ as the event that one of the users cannot recover his message correctly. It can be shown that this event implies one of the following events:

- Quantization range error $\mathcal{E}_\Delta$: the amplitude of real or imaginary parts of interferences is out of $[-\bar{\eta}, \bar{\eta}]$;

- Multicast error $\mathcal{E}_{\mathsf{mc}}$: one of the users cannot recover the double indices of $(\hat{\eta}_1, \hat{\eta}_2)$ correctly;

- MIMO decoding error $\mathcal{E}_{\mathsf{mimo}}$: based on the received signal and the recovered indices, one of the users cannot recover his original message after performing a MIMO decoding of the equivalent channel (20) or (21).

That is, $\mathcal{E} \subseteq \mathcal{E}_\Delta \cup \mathcal{E}_{\mathsf{mc}} \cup \mathcal{E}_{\mathsf{mimo}}$. Therefore, we have

$$P_e \leq \mathbb{P}\left(\mathcal{E}_\Delta\right) + \mathbb{P}\left(\mathcal{E}_{\mathsf{mc}}\right) + \mathbb{P}\left(\mathcal{E}_{\mathsf{mimo}} \cap \bar{\mathcal{E}}_\Delta \cap \bar{\mathcal{E}}_{\mathsf{mc}}\right). \tag{164}$$

In the following, we examine the individual error events.

1) Quantization range error $\mathcal{E}_\Delta$: This event is the union of the four events: $\{\mathrm{Re}\,(\eta_1) > \bar{\eta}\}$, $\{\mathrm{Im}\,(\eta_1) > \bar{\eta}\}$, $\{\mathrm{Re}\,(\eta_2) > \bar{\eta}\}$, and $\{\mathrm{Im}\,(\eta_2) > \bar{\eta}\}$. This event implies that the quantization error is not bounded. Let us recall that $\eta_k \sim \mathcal{N}_\mathbb{C}\left(0, \sigma_{\eta_k}^2\right)$, i.e., $\mathrm{Re}\,(\eta_k), \mathrm{Im}\,(\eta_k) \sim \mathcal{N}(0, \frac{\sigma_{\eta_k}^2}{2})$, $k = 1, 2$, conditional on the channel state $\boldsymbol{S}$. We have

$$\mathbb{P}\left(\mathrm{Re}\,(\eta_1) > \bar{\eta}\right) = \mathbb{E}\left(Q\left(\frac{\bar{\eta}_1}{\sigma_{\eta_1}/\sqrt{2}}\right)\right) \tag{165}$$

$$\leq \mathbb{E}\left(\exp\left(-\frac{\bar{\eta}_1^2}{\sigma_{\eta_1}^2}\right)\right) \tag{166}$$

$$\leq \mathbb{E}\left(\exp\left(-\frac{\bar{\eta}_1^2}{A}\right)\right) \tag{167}$$

where the first equality comes from the Gaussian distribution conditional on the channel states; to obtain (166), we applied $Q(x) \leq e^{-x^2/2}$; the last inequality is from the fact that $\sigma_{\eta_1}^2 \leq A$ with $A \triangleq \|\tilde{\boldsymbol{h}}\|^2 P + \|\boldsymbol{h}\|^2 P_2$. Now, we can go further with the upper bound, by introducing $\epsilon > 0$,

$$\mathbb{P}\left(\mathrm{Re}\,(\eta_1) > \bar{\eta}\right)$$

$$\leq \mathbb{P}\left(A \leq \bar{\eta}^2 P^{-\epsilon}\right) e^{-P^\epsilon} + \mathbb{P}\left(A > \bar{\eta}^2 P^{-\epsilon}\right) \tag{168}$$

$$\leq e^{-P^\epsilon} + \mathbb{P}\left(A > \bar{\eta}^2 P^{-\epsilon}\right) \tag{169}$$

$$\leq e^{-P^\epsilon} + \mathbb{P}\left(\|\tilde{\boldsymbol{h}}\|^2 P > \frac{1}{2}\bar{\eta}^2 P^{-\epsilon}\right)$$

$$+ \mathbb{P}\left(\|\boldsymbol{h}\|^2 P_2 > \frac{1}{2}\bar{\eta}^2 P^{-\epsilon}\right) \tag{170}$$

$$= e^{-P^\epsilon} + \mathbb{P}\left(\frac{\|\tilde{\boldsymbol{h}}\|^2}{m\sigma^2} > \frac{1}{2m}P^{\zeta-\epsilon}\right)$$

$$+ \mathbb{P}\left(\frac{\|\boldsymbol{h}\|^2}{m} > \frac{1}{2m}P^{\zeta-\epsilon+1-\alpha_P-\xi_P}\right) \tag{171}$$

$$\leq e^{-P^\epsilon} + \frac{1}{4m^2}P^{-2(\zeta-\epsilon)} + \frac{1}{4m^2}P^{-2(\zeta-\epsilon+1-\alpha_P-\xi_P)} \tag{172}$$

where (170) is from the union bound; $\xi_P \triangleq \frac{\log P_2}{\log P}$; the last inequality is Chebyshev's inequality.

Note that, due to the symmetry, the probabilities for the four events, i.e., $\mathbb{P}\left(\mathrm{Re}\,(\eta_1) > \bar{\eta}\right)$, $\mathbb{P}\left(\mathrm{Im}\,(\eta_1) > \bar{\eta}\right)$, $\mathbb{P}\left(\mathrm{Re}\,(\eta_2) > \bar{\eta}\right)$, and $\mathbb{P}\left(\mathrm{Im}\,(\eta_2) > \bar{\eta}\right)$, have the same upper bound (172). Therefore, by the union bound, we

have $\mathbb{P}(\mathcal{E}_\Delta) \leq 4\,\mathbb{P}(\mathrm{Re}\,(\eta_1) > \bar{\eta})$. From (172), a sufficient condition for $\lim_{P\to\infty} \mathbb{P}(\mathcal{E}_\Delta) = 0$ is $\zeta > \epsilon > 0$ and $\lim_{P\to\infty} 1 - \alpha_P - \xi_P \geq 0$, i.e.,

$$\lim_{P\to\infty} \xi_P \leq 1 - \alpha, \tag{173}$$

meaning that the power P_2 should not scale faster than $P^{1-\alpha}$.

2) Multicast error $\mathcal{E}_{mc}$: First, we provide the following lemma.

Lemma 4: For the considered MISO broadcast channel with common message, $(\mathsf{DoF}_0, \mathsf{DoF}_1, \mathsf{DoF}_2) = (1 - \alpha, \alpha, \alpha)$ is achievable with a single-letter Gaussian code.

Proof: The coding scheme is the same superposition coding with Gaussian codebooks as in the ergodic case, except that the length is 1. As shown before, successive interference cancellation decoding is used at each receiver. By treating the private signals as noise, the SINR for the common signal scales as $P^{1-\alpha}$. After removing the common signal successfully, the useful private message is decoded with a SINR scaling as P^α. To show the DoF, it is enough to show that $\mathsf{DoF} = \kappa$ is achievable if the SINR scales as P^κ, $\kappa \geq 0$. Since the interferences and noise are all Gaussian, it can be easily shown with the union bound. ■

Note that the number of bits needed to describe the indices is approximately $4 + 2(1 + \zeta - \alpha_P) \log P$. From Lemma 4, we know that for any $\delta < 0$, a rate $(1 - \delta) \log P$ can be achieved reliably when $P \to \infty$. Therefore, as long as the number of channel uses

$$l \geq \frac{4}{(1-\delta)\log P} + \frac{2(1 + \zeta - \alpha_P)}{1 - \delta}, \tag{174}$$

we can guarantee that $\mathbb{P}\{\mathcal{E}_{mc}\} \to 0$ when $P \to \infty$.

3) MIMO decoding error $\mathcal{E}_{mimo}$: Let $\mathcal{E}_{mimo,k}$ be the MIMO decoding error at receiver k, $k = 1, 2$. It is obvious that $\mathbb{P}(\mathcal{E}_{mimo}) \leq \mathbb{P}(\mathcal{E}_{mimo,1}) + \mathbb{P}(\mathcal{E}_{mimo,2})$. Due to the symmetry, we can focus on $\mathcal{E}_{mimo,1}$. First, we introduce $\epsilon' > 0$ and define

$$\mathcal{O}_{\epsilon'} \triangleq \{ \boldsymbol{S} : \ \log\det(\mathbf{I} + \boldsymbol{S}\boldsymbol{Q_u}\boldsymbol{S}^{\mathsf{H}}) < R + \epsilon' \log P \}. \tag{175}$$

Therefore, the error probability can be upper-bounded by

$$\mathbb{P}\left(\mathcal{E}_{mimo,1} \cap \bar{\mathcal{E}}_\Delta \cap \bar{\mathcal{E}}_{mc}\right)$$

$$\leq \mathbb{P}\left(\mathcal{E}_{mimo,1} \cap \underbrace{\bar{\mathcal{E}}_\Delta \cap \bar{\mathcal{E}}_{mc} \cap \bar{\mathcal{O}}_{\epsilon'}}_{\mathcal{B}}\right) + \mathbb{P}(\mathcal{O}_{\epsilon'}) \tag{176}$$

$$= \mathbb{P}\left((W_1 \neq \hat{W}_1) \cap \mathcal{B}\right) + \mathbb{P}(\mathcal{O}_{\epsilon'}). \tag{177}$$

We can bound the two terms separately. For simplicity, we assume that minimum Euclidean distance decoding is used[8]. To that end, we look into the pair-wise error probability for a pair of different codewords $\boldsymbol{u}(0), \boldsymbol{u}(1) \in \mathcal{X}_1$,

---

[8]Since the noise is not Gaussian and depends on the signal in general, it does not correspond to maximum likelihood decoding.

denoted by $\mathbb{P}(\boldsymbol{u}(0) \to \boldsymbol{u}(1))$. For a given channel realization $\boldsymbol{S}$, we have

$$\mathbb{P}((\boldsymbol{u}(0) \to \boldsymbol{u}(1)) \cap \mathcal{B} \,|\, \boldsymbol{S})$$

$$\leq \mathbb{P}\left(\left(\frac{\|\boldsymbol{S}(\boldsymbol{u}(0) - \boldsymbol{u}(1))\|}{2} \leq \|\boldsymbol{b}\|\right) \cap \mathcal{B}\right) \tag{178}$$

$$\leq \mathbb{P}\left(\left(\|\boldsymbol{S}\boldsymbol{u}_d\|^2 \leq 2(|\varepsilon|^2 + |\Delta_1|^2 + |\Delta_2|^2)\right) \cap \mathcal{B}\right)$$

$$\leq \mathbb{P}\left(\left(\|\boldsymbol{S}\boldsymbol{u}_d\|^2 \leq 2|\varepsilon|^2 + 1\right) \cap \mathcal{B}\right) \tag{179}$$

$$\leq \mathbb{P}\left(\left(\|\boldsymbol{S}\boldsymbol{u}_d\|^2 \leq 2|\varepsilon|^2 + 1\right) \cap \bar{\mathcal{O}}_{\epsilon'}\right) \tag{180}$$

$$\leq \mathbb{P}\left(\|\boldsymbol{S}\boldsymbol{u}_d\|^2 \leq 2|\varepsilon|^2 + 1\right) \mathbf{1}_{\bar{\mathcal{O}}_{\epsilon'}}(\boldsymbol{S}) \tag{181}$$

where $\mathbf{1}_{\bar{\mathcal{O}}_{\epsilon'}}(\boldsymbol{S})$ is the indicator function that gives 1 if $\boldsymbol{S} \in \bar{\mathcal{O}}_{\epsilon'}$ and 0 otherwise; (179) is from the fact that the quantization error $|\Delta_k|$ is bounded by $\frac{1}{2}$, $k = 1, 2$. We can go further with the probability term

$$\mathbb{P}\left(\left(\|\boldsymbol{S}\boldsymbol{u}_d\|^2 \leq 2|\varepsilon|^2 + 1\right)\right) \tag{182}$$

$$\leq \mathbb{P}\left(\|\boldsymbol{S}\boldsymbol{u}_d\|^2 \leq 4|\varepsilon|^2\right) + \mathbb{P}\left(\|\boldsymbol{S}\boldsymbol{u}_d\|^2 \leq 2\right) \tag{183}$$

$$\leq \mathbb{E}\left(\exp\left(-\frac{1}{4}\|\boldsymbol{S}\boldsymbol{u}_d\|^2\right)\right) + \mathbb{P}\left(\|\boldsymbol{S}\boldsymbol{u}_d\|^2 \leq 2\right) \tag{184}$$

$$\leq \det\left(\mathbf{I} + \frac{1}{4}\boldsymbol{S}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{S}^{\mathsf{H}}\right)^{-1}$$

$$+ \mathbb{P}\left(\rho_1 \leq \frac{2}{\mu_1}\right) \mathbb{P}\left(\rho_2 \leq \frac{2}{\mu_2}\right) \tag{185}$$

$$\leq 16 \det\left(\mathbf{I} + \boldsymbol{S}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{S}^{\mathsf{H}}\right)^{-1} + (1 - e^{-\frac{2}{\mu_1}})(1 - e^{-\frac{2}{\mu_2}}) \tag{186}$$

$$\leq 16 \det\left(\mathbf{I} + \boldsymbol{S}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{S}^{\mathsf{H}}\right)^{-1} + \frac{16}{(2 + \mu_1)(2 + \mu_2)} \tag{187}$$

$$= 32 \det\left(\mathbf{I} + \boldsymbol{S}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{S}^{\mathsf{H}}\right)^{-1} \tag{188}$$

where $\boldsymbol{u}_d \triangleq (\boldsymbol{u}_0 - \boldsymbol{u}_1)/\sqrt{2} \sim \mathcal{N}_{\mathbb{C}}(0, \boldsymbol{Q}_{\boldsymbol{u}})$; $\|\boldsymbol{S}\boldsymbol{u}_d\|^2 \overset{d}{=} \mu_1 \rho_1 + \mu_2 \rho_2$ with $\rho_1, \rho_2 \sim \exp(1)$ and $\mu_1 \geq \mu_2$ being the two eigenvalues of $\boldsymbol{S}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{S}^{\mathsf{H}}$; (187) is obtained by applying $1 - \exp\left(-\frac{1}{x}\right) \leq \frac{2}{1+x}$, $\forall\, x \geq 0$. With the union bound on all possible codewords pairs, we finally obtain

$$\mathbb{P}\left((W_1 \neq \hat{W}_1) \cap \mathcal{B}\right)$$

$$\leq 2^R \, \mathbb{E}_{\boldsymbol{S}}\left(\mathbb{P}((\boldsymbol{u}(0) \to \boldsymbol{u}(1)) \cap \mathcal{B} \,|\, \boldsymbol{S})\right) \tag{189}$$

$$\leq 32\, P^r \, \mathbb{E}_{\boldsymbol{S}}\left(\det\left(\mathbf{I} + \boldsymbol{S}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{S}^{\mathsf{H}}\right)^{-1} \mathbf{1}_{\bar{\mathcal{O}}_{\epsilon'}}(\boldsymbol{S})\right) \tag{190}$$

$$\leq 32\, P^r \, \mathbb{E}_{\boldsymbol{S}}\left(P^{-(r+\epsilon')}\right) \tag{191}$$

$$= 32\, P^{-\epsilon'} \tag{192}$$

where we used the fact that $\det\left(\mathbf{I} + \boldsymbol{S}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{S}^{\mathsf{H}}\right)^{-1} \leq P^{-(r+\epsilon')}$ for any $\boldsymbol{S} \in \bar{\mathcal{O}}_{\epsilon'}$, from the definition (175).

In the following, we set $m = 2$ to simplify the calculation. For $m \geq 2$, the error probability can only be smaller. The probability $\mathbb{P}(\mathcal{O}_{\epsilon'})$ is upper-bounded as follows

$$\mathbb{P}(\mathcal{O}_{\epsilon'}) \leq \mathbb{P}\left(\det\left(\boldsymbol{S}\boldsymbol{Q_u}\boldsymbol{S}^{\mathsf{H}}\right) < P^{r+\epsilon'}\right) \tag{193}$$

$$= \mathbb{P}\left(\det\left(\boldsymbol{S}\boldsymbol{S}^{\mathsf{H}}\right)\det\left(\boldsymbol{Q_u}\right) < P^{r+\epsilon'}\right) \tag{194}$$

$$= \mathbb{P}\left(\det\left(\boldsymbol{S}\boldsymbol{S}^{\mathsf{H}}\right) < P^{r-1-\xi_P+\epsilon'}\right) \tag{195}$$

$$\leq \mathbb{P}\left(\lambda_{\min}(\boldsymbol{S}\boldsymbol{S}^{\mathsf{H}}) < P^{\frac{r-1-\xi_P+\epsilon'}{2}}\right). \tag{196}$$

As long as

$$r < 1 + \xi_P - \epsilon', \tag{197}$$

the probability in (196) scales as $P^{-\frac{(r+1+\xi_P-\epsilon')}{2}}$, according to the near-zero behavior of the minimum eigenvalue of the Wishart matrix $\boldsymbol{S}\boldsymbol{S}^{\mathsf{H}}$.

From (174) and (197), by letting $P \to \infty$, and making ζ, ϵ', ϵ'', and δ as close to 0 as possible, the proposed scheme can achieve an average DoF per user

$$\frac{r}{1+l} = \frac{2-\alpha}{3-2\alpha}. \tag{198}$$

Due to the symmetry, same proof applies to finding precisely the same DoF for user 2. ∎

REFERENCES

[1] A. Lapidoth, S. Shamai (Shitz), and M. A. Wigger, "On the capacity of fading MIMO broadcast channels with imperfect transmitter side-information," in Proc. of 43rd Allerton Conference, Sept. 28–30, 2005.

[2] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory* vol. 56, no. 6, pp. 2845–2866, June 2010.

[3] M. A. Maddah-Ali and D. N. C. Tse, "Completely stale transmitter channel state information is still very useful," *Arxiv preprint arXiv:1010.1499, Submitted to IEEE Trans. Inf. Theory*, 2011.

[4] S. A. Jafar, "Interference Alignment: A New Look at Signal Dimensions in a Communication Network," *Foundations and Trends in Communications and Information Theory*, vol. 7, no. 1, pp. 1–136, 2011.

[5] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[6] M. Kobayashi, S. Yang, D. Gesbert, X. Yi, "On the degrees of freedom of time correlated MISO broadcast channel with delayed CSIT," submitted to ISIT 2012, also available as *Arxiv preprint:1202.1909*, Feb. 2012.

[7] T. M. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[8] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information-theoretic problems," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1839–1851, May 2007.

[9] H. Weingarten, Y. Steinberg, S. Shamai (Shitz), "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.

[10] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Probl. Peredachi Inf.*, vol. 10, no. 3, pp. 3–14, 1974.

[11] A. Lapidoth and S. M. 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.

[12] T. Gou and S. A. Jafar, "Optimal use of current and outdated channel state information–Degrees of freedom of the MISO BC with mixed CSIT," *Arxiv preprint arXiv:1203.1301*, March 2012.