Degrees of Freedom of MISO Broadcast Channel with Perfect Delayed and Imperfect Current CSIT

Sheng Yang, Mari Kobayashi, David Gesbert, Xinping Yi

***Abstract*— We consider the two-user MISO broadcast channel where the transmitter has imperfect knowledge on the current channel state, in addition to delayed channel state information. The degree of freedom region is completely characterized. The optimal 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. The essential ingredients of this scheme lie in the quantization and multicasting of the overheard interferences, while broadcasting new private messages.**

I. Introduction

In this paper, we consider 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

$$y_t = \boldsymbol{h}_t^{\mathsf{H}} \boldsymbol{x}_t + \varepsilon_t$$
$$z_t = \boldsymbol{g}_t^{\mathsf{H}} \boldsymbol{x}_t + \omega_t$$

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 $\mathbb{E}\left(\|\boldsymbol{x}_t\|^2\right) \le P$, $\forall\, t$.

For the case of perfect CSIT, the optimal degrees of freedom (DoF) 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 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 $m \ge 2$ 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. 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 ($\alpha \ge 1$) 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. In [4], 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. This power reduction then enables, via source compression/quantization, to save the resources related to the transmission of the overheard interference. This scheme achieves the symmetric 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 digitized transmission of the overheard interference, which replaces the analog one initially considered in the MAT alignment. Despite its suboptimality as it will turn out, it is the indispensable basis for the new optimal scheme.

The main contributions of this work are summarized as follows. We establish an outer bound on the DoF region of the two-user broadcast channel with perfect delayed and imperfect current state information. The outer bound turns out to be a source of inspiration of our optimal scheme. 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

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

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.

Due to the page limits, detailed proofs are omitted and available in the full paper [9]. At the time of submission of the full paper, parallel independent work [10] was brought to our attention which also builds on the results of [4].

II. System Model and Main Results

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 [9].

Assumption 1 (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 $\hat{\boldsymbol{S}}_t$ of the current channel state $\boldsymbol{S}_t$, i.e., $\hat{\boldsymbol{h}}_t$ and $\hat{\boldsymbol{g}}_t$ are available to the transmitter with

$$\boldsymbol{h}_t = \hat{\boldsymbol{h}}_t + \tilde{\boldsymbol{h}}_t, \quad \boldsymbol{g}_t = \hat{\boldsymbol{g}}_t + \tilde{\boldsymbol{g}}_t$$

where the estimate $\hat{\boldsymbol{h}}_t$ (also $\hat{\boldsymbol{g}}_t$) and estimation error $\tilde{\boldsymbol{h}}_t$ (also $\tilde{\boldsymbol{g}}_t$) 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 $\sigma^2 \le 1$. The receivers knows $\boldsymbol{S}_t \in \mathcal{S}$ and $\hat{\boldsymbol{S}}_t \in \hat{\mathcal{S}}$ without delay.

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

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

Without loss of generality, we can introduce a parameter $\alpha_P \ge 0$ as the power exponent of the estimation error[1]

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

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 $\alpha \triangleq \lim_{P\to\infty}\alpha_P$. Hereafter, we use α instead of α_P, whenever confusion is unlikely. The main result of this paper is stated below.

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

$$d_1 \le 1 \tag{3a}$$

$$d_2 \le 1 \tag{3b}$$

$$d_1 + 2d_2 \le 2 + \alpha \tag{3c}$$

$$2d_1 + d_2 \le 2 + \alpha. \tag{3d}$$

From the Theorem, we observe that the region collapses to the MAT region [3] when the quality of current CSIT is poor ($\alpha \to 0$), whereas it grows smoothly towards the DoF region with perfect CSIT with increasing α. The next three sections are devoted to proving the Theorem. We start with the achievability.

[1]Throughout the paper, logarithms are in base 2.

III. Achievability: A Simple Scheme

A. MAT alignment revisited

In the two-user MISO case, we consider the following variant of the MAT alignment presented in [3].

$$\begin{aligned} &\boldsymbol{x}_1 = \boldsymbol{u}+\boldsymbol{v} && \boldsymbol{x}_2 = [\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{v} \;\; 0]^{\mathsf{T}} && \boldsymbol{x}_3 = [\boldsymbol{g}_1^{\mathsf{H}}\boldsymbol{u} \;\; 0]^{\mathsf{T}} \\ &y_1 = \boldsymbol{h}_1^{\mathsf{H}}(\boldsymbol{u}+\boldsymbol{v}) && y_2 = h_{21}\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{v} && y_3 = h_{31}\boldsymbol{g}_1^{\mathsf{H}}\boldsymbol{u} \\ &z_1 = \boldsymbol{g}_1^{\mathsf{H}}(\boldsymbol{u}+\boldsymbol{v}) && z_2 = g_{21}\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{v} && z_3 = g_{31}\boldsymbol{g}_1^{\mathsf{H}}\boldsymbol{u} \end{aligned}$$

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. 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., $\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{v}$ and $\boldsymbol{g}_1^{\mathsf{H}}\boldsymbol{u}$. 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

B. Integrating the imperfect current CSI

Based on the above variant of the MAT scheme, the following two-stage scheme, called Scheme I, was proposed in [4]. As it is an important building block of the new scheme, we briefly review the key ingredients in the follow. Since only $\boldsymbol{h}_1$ and $\boldsymbol{g}_1$ are involved below, we drop the time indices for convenience.

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 $\boldsymbol{Q}_{\boldsymbol{u}} \triangleq \mathbb{E}(\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}})$ and $\boldsymbol{Q}_{\boldsymbol{v}} \triangleq \mathbb{E}(\boldsymbol{v}\boldsymbol{v}^{\mathsf{H}})$ that may depend on the estimates of the current channel. The power constraint is respected by choosing $\boldsymbol{Q}_{\boldsymbol{u}}$ and $\boldsymbol{Q}_{\boldsymbol{v}}$ such that $\mathrm{tr}(\boldsymbol{Q}_{\boldsymbol{u}}) + \mathrm{tr}(\boldsymbol{Q}_{\boldsymbol{v}}) \le P$.

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 $(\boldsymbol{h}^{\mathsf{H}}\boldsymbol{v}, \boldsymbol{g}^{\mathsf{H}}\boldsymbol{u})$ required by both receivers. However, unlike the original MAT scheme (4) where these symbols are transmitted in an analog fashion, we quantize them and then transmit the digital version. 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 $\boldsymbol{\eta} \triangleq (\eta_1, \eta_2)$. Note that for a given channel realization, the average power of η_1 and η_2 are

$$\sigma_{\eta_1}^2 \triangleq \boldsymbol{h}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{v}}\boldsymbol{h} \quad \text{and} \quad \sigma_{\eta_2}^2 \triangleq \boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{g}.$$

Assume that an R_{η_k}-bits quantizer is used for η_k, $k=1,2$.

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

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: Each user first tries to recover $(\hat{\eta}_1, \hat{\eta}_2)$. If this step is done successfully, receiver 1 has

$$\begin{aligned} y &= \boldsymbol{h}^{\mathsf{H}}\boldsymbol{u} + \eta_1 + \varepsilon \\ \hat{\eta}_1 &= \eta_1 - \Delta_1 \\ \hat{\eta}_2 &= \eta_2 - \Delta_2 = \boldsymbol{g}^{\mathsf{H}}\boldsymbol{u} - \Delta_2 \end{aligned}$$

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

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

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

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

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

C. Achievable degrees of freedom

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 symmetric rate of Scheme I is given by $\frac{R_{\mathsf{mimo}}}{1+R_{\boldsymbol{\eta}}/R_{\mathsf{mc}}}$. In the rest of the section, we show the following rates

$$\begin{aligned} R_{\mathsf{mc}} &= \log P + O(1) && \text{(5a)} \\ R_{\boldsymbol{\eta}} &= 2(1-\alpha)\log P + O(1) \\ R_{\mathsf{mimo}} &= (2-\alpha)\log P + O(1) \end{aligned}$$

which yields the 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 a pre-log loss of α in total. Since we need to show the achievability for any $m \geq 2$, it is enough to consider the case $m=2$. We fix the parameters of Scheme I as follows:

- We send two streams per user in two orthogonal directions:
$$\boldsymbol{Q}_{\boldsymbol{u}} = P_1 \boldsymbol{\Psi}_{\hat{\boldsymbol{g}}^{\perp}} + P_2 \boldsymbol{\Psi}_{\hat{\boldsymbol{g}}}, \quad \boldsymbol{Q}_{\boldsymbol{v}} = P_1 \boldsymbol{\Psi}_{\hat{\boldsymbol{h}}^{\perp}} + P_2 \boldsymbol{\Psi}_{\hat{\boldsymbol{h}}} \quad (6)$$
where $\boldsymbol{\Psi}_{\hat{\boldsymbol{g}}} \triangleq \frac{\hat{\boldsymbol{g}}\hat{\boldsymbol{g}}^{\mathsf{H}}}{\|\hat{\boldsymbol{g}}\|^2}$ and $\boldsymbol{\Psi}_{\hat{\boldsymbol{g}}^{\perp}}$, $\boldsymbol{\Psi}_{\hat{\boldsymbol{h}}}$, and $\boldsymbol{\Psi}_{\hat{\boldsymbol{h}}^{\perp}}$ are similarly defined with $\boldsymbol{x}^{\perp}$ being any nonzero vector such that $\boldsymbol{x}^{\mathsf{H}}\boldsymbol{x}^{\perp} = 0$.
- 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 $P_1 = P - P_2 \sim P$.
- The distortions D_1 and D_2 are set to the noise level, i.e., $D_1 = D_2 = 1 \sim P^0$.

First, (5a) is achievable by using any DoF optimal single user code. Second, we can upper-bound the quantization rate $R_{\boldsymbol{\eta}}$ as

$$\begin{aligned} R_{\boldsymbol{\eta}} &\leq \mathbb{E}\left(\log\left(\frac{\boldsymbol{h}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{v}}\boldsymbol{h}}{D_1}\right)\right) + \mathbb{E}\left(\log\left(\frac{\boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{g}}{D_2}\right)\right) \\ &\leq \log\left(\mathbb{E}\left(\boldsymbol{h}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{v}}\boldsymbol{h}\right)\right) + \log\left(\mathbb{E}\left(\boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{g}\right)\right) \\ &= 2\log\left(\mathbb{E}\left(\boldsymbol{g}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{g}\right)\right) && (7) \\ &= 2\log\left(\mathbb{E}\left(\tilde{\boldsymbol{g}}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{u}}\tilde{\boldsymbol{g}}\right) + \mathbb{E}\left(\hat{\boldsymbol{g}}^{\mathsf{H}}\boldsymbol{Q}_{\boldsymbol{u}}\hat{\boldsymbol{g}}\right)\right) && (8) \\ &\leq 2\log\left(P\sigma^2 + P_2\mathbb{E}\left(\hat{\boldsymbol{g}}^{\mathsf{H}}\boldsymbol{\Psi}_{\hat{\boldsymbol{g}}}\hat{\boldsymbol{g}}\right)\right) \\ &\leq 2\log\left(P\sigma^2 + 2P_2(1-\sigma^2)\right) \\ &= 2(1-\alpha)\log P + O(1) \end{aligned}$$

where the first inequality is from the rate-distortion theorem and by saying that Gaussian source is the hardest to compress [5]; the second inequality is from the concavity of the log function; (7) is from the symmetry between the channels and between the strategies; (8) is from (6). Finally, we lower-bound the MIMO rate R_{mimo} of user 1 as

$$\begin{aligned} R_{\mathsf{mimo}} &= \mathbb{E}\left(I(U;\tilde{Y} \,|\, S = \boldsymbol{S})\right) \\ &= \mathbb{E}\left(I(\boldsymbol{S}U;\tilde{Y})\right) \\ &= \mathbb{E}\left(h(\boldsymbol{S}U) - h(\boldsymbol{S}U \,|\, \tilde{Y})\right) \\ &= \mathbb{E}\left(h(\boldsymbol{S}U) - h(E+\Delta_1, -\Delta_2 \,|\, \tilde{Y})\right) \\ &\geq \mathbb{E}\left(h(\boldsymbol{S}U) - h(E+\Delta_1, -\Delta_2)\right) && (9) \\ &\geq \mathbb{E}\left(\log\det\left(\boldsymbol{S}\boldsymbol{Q}_{\boldsymbol{u}}\boldsymbol{S}^{\mathsf{H}}\right) - \log(1+D_1) - \log(D_2)\right) \\ &= \log(P_1 P_2) + O(1) \\ &= (2-\alpha)\log(P) + O(1) \end{aligned}$$

where (9) holds since conditioning reduces differential entropies; the last inequality follows because $\boldsymbol{u}$ is Gaussian, and 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: 1) Genie-aided model to construct a degraded broadcast channel, as in [3]; 2) Extremal inequality to bound the weighted difference of differential entropies [6]; 3) Isotropic property of the channel uncertainty to upper-bound the pre-log factor.

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):

$$\begin{aligned} nR_1 &\leq H(W_1) \\ &= H(W_1 \,|\, S^n, \hat{S}^n) \\ &= I(W_1; Y^n, Z^n \,|\, S^n, \hat{S}^n) + n\epsilon_n \\ &\leq I(W_1; Y^n, Z^n, W_2 \,|\, S^n, \hat{S}^n) + n\epsilon_n \\ &= I(W_1; Y^n, Z^n \,|\, S^n, \hat{S}^n, W_2) + n\epsilon_n \end{aligned}$$

$$= \sum_{i=1}^{n} I(W_1; Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, S^n, \hat{S}^n, W_2) + n\epsilon_n$$
$$\leq \sum_{i=1}^{n} I(X_i; Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, S^n, \hat{S}^n, W_2) + n\epsilon_n$$
$$= \sum_{i=1}^{n} I(X_i; Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, S^i, \hat{S}^i, W_2) + n\epsilon_n$$
$$= \sum_{i=1}^{n} h(Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, S^i, \hat{S}^i, W_2) - h(Y_i, Z_i \,|\, X_i, Y^{i-1}, Z^{i-1}, S^i, \hat{S}^i, W_2) + n\epsilon_n$$
$$= \sum_{i=1}^{n} h(Y_i, Z_i \,|\, T_i, S_i) - h(E_i, \Omega_i) + n\epsilon_n$$
$$\leq \sum_{i=1}^{n} h(Y_i, Z_i \,|\, T_i, S_i) + n\epsilon_n \quad (10)$$

$$nR_2 \leq H(W_2)$$
$$\leq I(W_2; Z^n \,|\, S^n, \hat{S}^n) + n\epsilon_n$$
$$= \sum_{i=1}^{n} I(W_2; Z_i \,|\, Z^{i-1}, S^i, \hat{S}^i) + n\epsilon$$
$$\leq \sum_{i=1}^{n} h(Z_i \,|\, S_i) - h(Z_i \,|\, Y^{i-1}, Z^{i-1}, S^i, \hat{S}^i, W_2) + n\epsilon$$
$$= \sum_{i=1}^{n} h(Z_i \,|\, S_i) - h(Z_i \,|\, T_i, S_i) + n\epsilon \quad (11)$$

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 [8] (also see [5]), with the integration of the channel states. From (10) and (11), we have

$$n(R_1 + 2R_2) \leq \sum_{i=1}^{n} (h(Y_i, Z_i \,|\, T_i, S_i) - 2h(Z_i \,|\, T_i, S_i) + 2h(Z_i \,|\, S_i)) + 3n\epsilon_n. \quad (12)$$

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

$$\max_{P_{T_i} P_{X_i|T_i}} 2h(Z_i \,|\, S_i) \leq 2\,\mathbb{E}_{G_i}\left(\max_{P_{X_i|G_i = \boldsymbol{g}_i}} h(\boldsymbol{g}_i^{\mathsf{H}} X_i + E_i) \right)$$
$$\leq 2\,\mathbb{E}_{G_i}\left(\log(1 + P\|\boldsymbol{g}_i\|^2)\right)$$
$$\leq 2 \log P + O(1) \quad (13)$$

where we used the fact that Gaussian distribution maximizes differential entropy under the covariance constraint, that the logarithmic function is monotonically increasing, and $\mathsf{Cov}(X_i \,|\, \boldsymbol{g}_i) \preceq \mathsf{Cov}(X_i) \preceq P\mathbf{I}$. Then, another bound is

$$\max_{P_{T_i} P_{X_i|T_i}} (h(Y_i, Z_i \,|\, T_i, S_i) - 2h(Z_i \,|\, T_i, S_i))$$
$$\leq \mathbb{E}_{\hat{S}_i}\Big(\max_{\boldsymbol{K}: \boldsymbol{K} \succeq 0, \mathrm{tr}(\boldsymbol{K}) \leq P} \mathbb{E}_{S_i|\hat{S}_i} (\log(1 + \boldsymbol{h}_i^{\mathsf{H}} \boldsymbol{K} \boldsymbol{h}_i) - \log(1 + \boldsymbol{g}_i^{\mathsf{H}} \boldsymbol{K} \boldsymbol{g}_i))\Big) \quad (14)$$

which can be shown using essentially the extremal inequality [6], [7]. Details of the proof are given in [9].

Lemma 1 ([9]): For any given $\boldsymbol{K} \succeq 0$ with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_m \geq 0$, we have

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

where γ is Euler's constant.
Without loss of generality, we consider $\sigma^2 > 0$ in the following. From Lemma 1,

$$\mathbb{E}_{S_i|\hat{S}_i} (\log(1 + \boldsymbol{h}_i^{\mathsf{H}} \boldsymbol{K} \boldsymbol{h}_i) - \log(1 + \boldsymbol{g}_i^{\mathsf{H}} \boldsymbol{K} \boldsymbol{g}_i))$$
$$\leq \log \frac{1 + \|\hat{\boldsymbol{h}}_i\|^2 \lambda_1}{1 + e^{-\gamma} \sigma^2 \lambda_1} + O(1)$$
$$\leq \log \left(1 + \frac{\|\hat{\boldsymbol{h}}_i\|^2}{e^{-\gamma} \sigma^2}\right) + O(1) \quad (15)$$
$$\leq -\log(\sigma^2) + \log\left(e^{-\gamma}\sigma^2 + \|\hat{\boldsymbol{h}}_i\|^2\right) + O(1) \quad (16)$$

where (15) 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 (14) and (16) and by noticing that $\sigma^2 \leq 1$, we have

$$\max_{P_{T_i} P_{X_i|T_i}} (h(Y_i, Z_i \,|\, T_i, S_i) - 2h(Z_i \,|\, T_i, S_i))$$
$$\leq \alpha \log P + \mathbb{E}_{\hat{S}_i}\left(\log\left(e^{-\gamma} + \|\hat{\boldsymbol{h}}_i\|^2\right)\right) + O(1)$$
$$= \alpha \log P + O(1). \quad (17)$$

From (12), (13), (17), and letting $n \to \infty$, we have

$$R_1 + 2R_2 \leq (2 + \alpha) \log P + O(1)$$

from which we obtain (3c) by dividing both sides of the above inequality by $\log P$ and let $P \to \infty$. Similarly, from (11), (13), and letting $n \to \infty$, we have $R_2 \leq \log P + O(1)$, from which we obtain the single user bound (3b) by dividing both sides of the above inequality by $\log P$ and let $P \to \infty$. To obtain (3d) and (3a), we can use the genie-aided model in which receiver 2 is helped by the genie and has perfect knowledge of y_t. The converse proof is complete.

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:

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

A natural question arises. *Can we convey 2α more symbols by extending the transmission by 2α 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): 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 $\{\mathcal{X}_0, \mathcal{X}_{p,1}, \mathcal{X}_{p,2}\}_P$, such that the following is achievable simultaneously

$$d_0 = 1 - \alpha,\ d_1 = d_2 = \alpha.$$

Proof: A sketch of proof is as follows, with more details given in [9]. 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, $k = 1, 2$. ■

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 (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 is coded in approximately $2(1-\alpha)\log P$ bits. However, instead of being sent in a reduced duration of $2(1-\alpha)$ channel uses, these bits are sent in 2 channel 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 DoF per channel use of $\frac{2+\alpha}{3}$.

Note that the region given by (3) 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 DoF of different schemes.

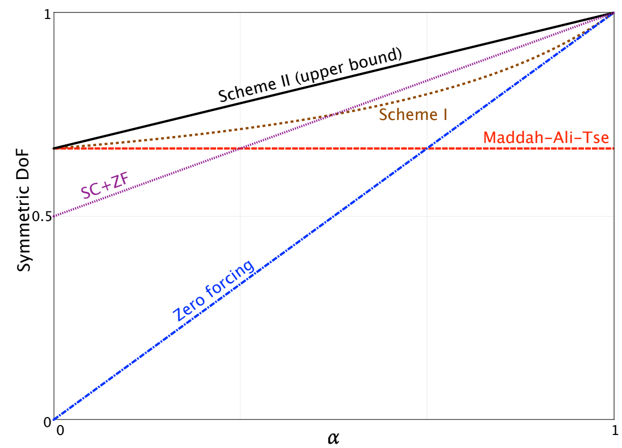

Fig. 1. Comparison of the achievable DoF between the proposed scheme and the zero-forcing and MAT alignment as a function of α.

The scheme "SC+ZF" (superposition coding and ZF precoding) is from equally time sharing between the corner points $(1,\alpha)$ and $(\alpha,1)$. 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 $\frac{2}{3}$ 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 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 $\alpha \in (0,1)$.

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] 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.

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

[6] 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.

[7] 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.

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

[9] S. Yang, M. Kobayashi, D. Gesbert, X. Yi, "Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT," submitted to *IEEE Trans. Inf. Theory* in March 2012, also available as *Arxiv preprint:1203.2550*.

[10] 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.