EURECOM
Department of Mobile Communications
Campus Sophia Tech
Les Templiers
450 route des Chappes
B.P. 193
06410 Biot
FRANCE

Research Report RR-12-275

How Much CSIT Feedback is Necessary for the Multiuser MISO Broadcast Channels?

December 07 th, 2012
Last update February 04th, 2013

Jinyuan Chen, Sheng Yang, and Petros Elia

[1]EURECOM's research is partially supported by its industrial members: BMW Group, Cisco, Monaco Telecom, Orange, SAP, SFR, STEricsson, Swisscom, Symantec.

Tel : (+33) 4 93 00 81 00
Fax : (+33) 4 93 00 82 00
Email : {jinyuan.chen,petros.elia}@eurecom.fr, sheng.yang@supelec.fr

How Much CSIT Feedback is Necessary for the Multiuser MISO Broadcast Channels?

Jinyuan Chen, Sheng Yang, and Petros Elia

Abstract

This work considers the multiuser multiple-input single-output (MISO) broadcast channel (BC), where a transmitter with M antennas transmits information to K single-antenna users, and where - as expected - the quality and timeliness of channel state information at the transmitter (CSIT) is imperfect. Motivated by the fundamental question of how much feedback is necessary to achieve a certain performance, this work seeks to establish bounds on the tradeoff between degrees-of-freedom (DoF) performance and CSIT feedback quality. Specifically, this work provides a novel DoF region outer bound for the general K-user $M \times 1$ MISO BC with partial current CSIT, which naturally bridges the gap between the case of having no current CSIT (only delayed CSIT, or no CSIT) and the case with full CSIT. The work then characterizes the minimum CSIT feedback that is necessary for any point of the sum DoF, which is optimal for the case with $M \geq K$, and the case with $M = 2,\ K = 3$.

Index Terms

Broadcast channels, multiple-input single-output (MISO), multiuser, limited feedback, channel state information at the transmitter (CSIT), alternating CSIT, degrees-of-freedom (DoF).

Contents

List of Figures

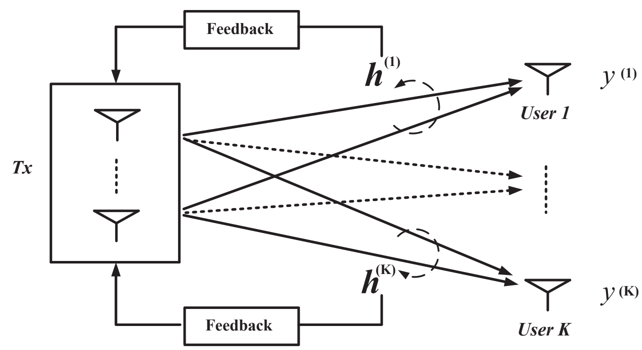

Figure 1: System model of K-user MISO BC with CSIT feedback.

1 Introduction

We consider the multiuser multiple-input single-output (MISO) broadcast channel (BC), where a transmitter with M antennas, transmits information to K single-antenna users. In this setting, the received signal at time t, is of the form

$$y_{k,t} = \boldsymbol{h}_{k,t}^{\mathsf{T}} \boldsymbol{x}_t + z_{k,t}, \quad k = 1, \cdots, K \tag{1}$$

where $\boldsymbol{h}_{k,t}$ denotes the $M \times 1$ channel vector for user k, $z_{k,t}$ denotes the unit power AWGN noise, and where $\boldsymbol{x}_t$ denotes the transmitted signal vector adhering to a power constraint $\mathbb{E}[||\boldsymbol{x}_t||^2] \leq P$, for P taking the role of the signal-to-noise ratio (snr). We here consider that the fading coefficients $\boldsymbol{h}_{k,t},\ k = 1, \cdots, K$, are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and unit variance, and are i.i.d. over time.

It is well known that the performance of the BC is greatly affected by the timeliness and quality of feedback; having full CSIT allows for the optimal $\min\{M, K\}$ sum degrees-of-freedom (DoF) (cf. [1])[1], while the absence of any CSIT reduces this to just 1 sum DoF (cf. [2, 3]). This gap has spurred a plethora of works that seek to analyze and optimize BC communications in the presence of delayed and imperfect feedback. One of the works that stands out is the work by Maddah-Ali and Tse [4] which recently revealed the benefits of employing delayed CSIT over the BC, even if this CSIT is completely obsolete. Several interesting generalizations followed, including the work in [5] which showed that in the BC setting with $K = M + 1$, combining delayed CSIT with perfect (current) CSIT (over the last $\frac{K-1}{K}$ fraction of communication period) allows for the optimal sum DoF M corresponding to full CSIT. A similar approach was exploited in [6] which revealed that, to achieve the maximum sum DoF $\min\{M, K\}$, each user has to symmetrically feed back perfect CSIT over a $\frac{\min\{M,K\}}{K}$ fraction of the communication time,

[1] We remind the reader that for an achievable rate tuple $(R_1, R_2, \cdots, R_K)$, where R_i is for user i, the corresponding DoF tuple $(d_1, d_2, \cdots, d_K)$ is given by $d_i = \lim_{P \to \infty} \frac{R_i}{\log P},\ i = 1, 2, \cdots, K$. The corresponding DoF region $\mathcal{D}$ is then the set of all achievable DoF tuples $(d_1, d_2, \cdots, d_K)$.

and that this fraction is optimal. Other interesting works in the context of utilizing delayed and current CSIT, can be found in [7–10] which explored the setting of combining perfect delayed CSIT with immediately available imperfect CSIT, the work in [11, 12] which additionally considered the effects of the quality of delayed CSIT, the work in [13] which considered alternating CSIT feedback, the work in [14] which considered delayed and progressively evolving (progressively improving) current CSIT, and the works in [15–21] and many other publications.

Our work here generalizes many of the above settings, and seeks to establish fundamental tradeoff between DoF performance and CSIT feedback quality, over the general K-user $M \times 1$ MISO BC.

1.1 CSIT quantification and feedback model

We proceed to describe the quality and timeliness measure of CSIT feedback, and how this measure relates to existing work. We here use $\hat{\boldsymbol{h}}_{k,t}$ to denote the current channel estimate (for channel $\boldsymbol{h}_{k,t}$) at the transmitter at timeslot t, and use

$$\tilde{\boldsymbol{h}}_{k,t} = \boldsymbol{h}_{k,t} - \hat{\boldsymbol{h}}_{k,t}$$

to denote the estimate error assumed to be mutually independent of $\hat{\boldsymbol{h}}_{k,t}$ and assumed to have i.i.d. Gaussian entries with power

$$\mathbb{E}\big[\|\tilde{\boldsymbol{h}}_{k,t}\|^2\big] \doteq P^{-\alpha_{k,t}},$$

for some CSI quality exponent $\alpha_{k,t} \in [0,1]$ describing the quality of this estimate. We note that $\alpha_{k,t} = 0$ implies very little current CSIT knowledge, and that $\alpha_{k,t} = 1$ implies perfect CSIT in terms of the DoF performance[2].

The approach extends over non-alternating CSIT settings in [4] and [7–10], as well as over an alternating CSIT setting (cf. [6, 13]) where CSIT knowledge alternates between perfect CSIT ($\alpha_{k,t} = 1$), and delayed or no CSIT ($\alpha_{k,t} = 0$).

In a setting where communication takes place over n such coherence periods ($t = 1, 2, \cdots, n$), this approach offers a natural measure of a per-user average feedback cost, in the form of

$$\bar{\alpha}_k \triangleq \frac{1}{n}\sum_{t=1}^{n} \alpha_{k,t}, \quad k = 1, 2, \cdots, K,$$

as well as a measure of current CSIT feedback cost

$$\mathsf{C}_{\mathrm{C}} \triangleq \sum_{k=1}^{K} \bar{\alpha}_k, \tag{2}$$

accumulated over all users.

[2]This can be readily derived, using for example the work in [22].

1.1.1 Alternating CSIT setting

In a setting where delayed CSIT is always available, the above model captures the alternating CSIT setting where the exponents are binary ($\alpha_{k,t} = 0, 1$), in which case

$$\bar{\alpha}_k = \delta_{\mathrm{P},k}$$

simply describes the fraction of time during which user k feeds back perfect CSIT, with

$$\mathsf{C}_{\mathrm{C}} = \mathsf{C}_{\mathrm{P}} \triangleq \sum_{k=1}^{K} \delta_{\mathrm{P},k}$$

describing the *total perfect CSIT feedback cost*.

1.1.2 Symmetric and asymmetric CSIT feedback

Motivated by the fact that different users might have different feedback capabilities due to the feedback channels with different capacities and different reliabilities, symmetric CSIT feedback ($\bar{\alpha}_1 = \cdots = \bar{\alpha}_K$) and asymmetric CSIT feedback ($\bar{\alpha}_k \neq \bar{\alpha}_{k'}\ \forall\, k \neq k^{'}$) are considered in this work.

1.2 Structure of the paper and Summary of Contributions

Section 2 provides the main results of this work:

- In Theorem 1 we first provide a novel outer bound on the DoF region, for the K-user $M \times 1$ MISO BC with partial current CSIT quantized with $\{\alpha_{k,t}\}_{k,t}$, which bridges the case with no current CSIT (only delayed CSIT, or no CSIT) and the case with full CSIT. This result manages to generalize the results by Maddah-Ali and Tse ($\alpha_{k,t} = 0,\ \forall t, k$), Yang et al. and Gou and Jafar ($K = 2,\ \alpha_{k,t} = \alpha,\ \forall t, k$), Maleki et al. ($K = 2,\ \alpha_{1,t} = 1, \alpha_{2,t} = 0,\ \forall t$), Chen and Elia ($K = 2,\ \alpha_{1,t} \neq \alpha_{2,t},\ \forall t$), Lee and Heath ($M = K + 1$, $\alpha_{k,t} \in \{0, 1\},\ \forall t, k$), and Tandon et al. ($\alpha_{k,t} \in \{0, 1\},\ \forall t, k$).

- From Theorem 1, we then provide the upper bound on the sum DoF, which is tight for the case with $M \geq K$ (cf. Theorem 2) and the case with $M = 2, K = 3$ (cf. Theorem 3, Corollary 3a).

- Furthermore, Theorem 4 characterizes the minimum total current CSIT feedback cost $\mathsf{C}_{\mathrm{P}}^{\star}$ to achieve the maximum sum DoF, where the total feedback cost $\mathsf{C}_{\mathrm{P}}^{\star}$ can be distributed among all the users with any (asymmetric and symmetric) combinations $\{\delta_{\mathrm{P},k}\}_k$.

- In addition, the work considers some other general settings of BC and provides the DoF inner bound as a function of the CSIT feedback cost.

The main converse proof, that is for Theorem 1, is shown in the Section 3 and appendix. Most of the achievability proofs are shown in the Section 4. Finally Section 5 concludes the paper.

1.3 Notation and conventions

Throughout this paper, we will consider communication over n coherence periods where, for clarity of notation, we will focus on the case where we employ a single channel use per such coherence period (unit coherence period). Furthermore, unless stated otherwise, we assume perfect delayed CSIT, as well as adhere to the common convention (see [4, 6, 8, 9, 13, 23]), and assume perfect and global knowledge of channel state information at the receivers.

In terms of notation, $(\bullet)^{\mathsf{T}}$, $(\bullet)^{\mathsf{H}}$, $\mathrm{tr}(\bullet)$ and $||\bullet||_F$ denote the transpose, conjugate transpose, trace and Frobenius norm of a matrix respectively, while $\mathrm{diag}(\bullet)$ denotes a diagonal matrix, $||\bullet||$ denotes the Euclidean norm, and $|\bullet|$ denotes either the magnitude of a scalar or the cardinality of a set. $o(\bullet)$ and $O(\bullet)$ come from the standard Landau notation, where $f(x) = o(g(x))$ implies $\lim_{x\to\infty} f(x)/g(x) = 0$. with $f(x) = O(g(x))$ implying that $\limsup_{x\to\infty} |f(x)/g(x)| < \infty$. We also use $\doteq$ to denote *exponential equality*, i.e., we write $f(P) \doteq P^B$ to denote $\lim\limits_{P\to\infty} \dfrac{\log f(P)}{\log P} = B$. Similarly $\dot{\geq}$ and $\dot{\leq}$ denote exponential inequalities. We use $\boldsymbol{A} \succeq \mathbf{0}$ to denote that $\boldsymbol{A}$ is positive semidefinite, and use $\boldsymbol{A} \preceq \boldsymbol{B}$ to mean that $\boldsymbol{B} - \boldsymbol{A} \succeq \mathbf{0}$. Logarithms are of base 2.

2 Main results

2.1 Outer bounds

We first present the DoF region outer bound for the general K-user $M \times 1$ MISO BC.

Theorem 1 (DoF region outer bound) *The DoF region of the K-user $M \times 1$ MISO BC, is outer bounded as*

$$\sum_{k=1}^{K} \frac{d_{\pi(k)}}{\min\{k, M\}} \leq 1 + \sum_{k=1}^{K-1} \left(\frac{1}{\min\{k, M\}} - \frac{1}{\min\{K, M\}} \right) \bar{\alpha}_{\pi(k)} \tag{3}$$

$$d_k \leq 1, \quad k = 1, 2, \cdots, K \tag{4}$$

where π denotes a permutation of the ordered set $\{1, 2, \cdots, K\}$, and $\pi(k)$ denotes the k th element of set π.

Proof: The proof is shown in Section 3. □

Remark 1 *It is noted that the bound captures the results in [4] ($\alpha_{k,t} = 0, \ \forall t, k$), in [8,9] ($K = 2$, $\alpha_{k,t} = \alpha, \ \forall t, k$), in [23] ($M = K = 2$, $\alpha_{1,t} = 1, \alpha_{2,t} = 0, \ \forall\, t$), in [10] ($K = 2$, $\alpha_{1,t} \neq \alpha_{2,t}, \ \forall\, t$), in [6,13] ($\alpha_{k,t} \in \{0, 1\}, \ \forall t, k$).*

Summing up the K different bounds from the above, we directly have the following upper bound on the sum DoF $d_\Sigma \triangleq \sum_{k=1}^{K} d_k$, which is presented using the following notation

$$d_{\text{MAT}} \triangleq \frac{K}{1+\frac{1}{\min\{2,M\}}+\frac{1}{\min\{3,M\}}+\cdots+\frac{1}{\min\{K,M\}}} \tag{5}$$

$$\Gamma \triangleq \frac{M}{\sum_{i=1}^{K-M}\frac{1}{i}(\frac{M-1}{M})^{i-1}+(\frac{M-1}{M})^{K-M}(\sum_{i=K-M+1}^{K}\frac{1}{i})}. \tag{6}$$

Corollary 1a (Sum DoF outer bound) *For the K-user $M\times 1$ MISO BC, the sum DoF is outer bounded as*

$$d_\Sigma \le d_{MAT} + \left(1-\frac{d_{MAT}}{\min\{K,M\}}\right)\sum_{k=1}^{K}\bar{\alpha}_k. \tag{7}$$

The above then readily translates onto a lower bound on the minimum possible total current CSIT feedback cost $\mathsf{C}_{\text{C}}=\sum_{k=1}^{K}\bar{\alpha}_k$ needed to achieve the maximum sum DoF[3] $d_\Sigma=\min\{K,M\}$.

Corollary 1b (Bound on CSIT cost for maximum DoF) *The minimum C_C required to achieve the maximum sum DoF $\min\{K,M\}$ of the K-user $M\times 1$ MISO BC, is lower bounded as*

$$\mathsf{C}_C^\star \ge \min\{K,M\}. \tag{8}$$

Transitioning to the alternating CSIT setting where $\alpha_{k,t}\in\{0,1\}$, we have the following sum-DoF outer bound as a function of the perfect-CSIT duration $\bar{\alpha}_k=\delta_{\text{P},k}=\delta_{\text{P}},\ \forall k$. We note that the bound holds irrespective of whether, in the remaining fraction of the time $1-\delta_{\text{P}}$, the CSIT is delayed or non existent.

Corollary 1c (Outer bound, alternating CSIT) *For the K-user $M\times 1$ MISO BC, the sum DoF is outer bounded as*

$$d_\Sigma \le d_{MAT} + \left(K-\frac{K\,d_{MAT}}{\min\{K,M\}}\right)\min\left\{\delta_P,\frac{\min\{K,M\}}{K}\right\}. \tag{9}$$

2.2 Optimal cases of DoF characterizations

We now provide the optimal cases of DoF characterizations. The case with $M\ge K$ is first considered in the following.

[3]Naturally the result is limited to the case where $\min\{K,M\}>1$.

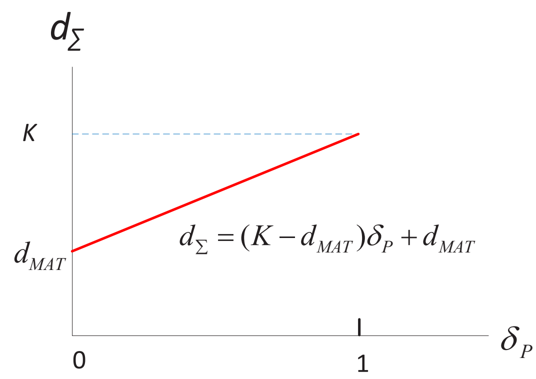

Figure 2: Optimal sum DoF d_Σ vs. δ_P for the MISO BC with $M \geq K$.

Theorem 2 (Optimal case, $M \geq K$) *For the K-user $M \times 1$ MISO BC with $M \geq K$, the optimal sum DoF is characterized as*

$$d_\Sigma = (K - d_{\mathit{MAT}}) \min\{\delta_\mathrm{P}, 1\} + d_{\mathit{MAT}}. \tag{10}$$

Proof: The converse and achievability proofs are derived from Corollary 1c and Proposition 2 (shown in the next subsection), respectively. □

Remark 2 *It is noted that, for the special case with $M = K = 2$, the above characterization captures the result in [13].*

Moving to the case where $M < K$, we have the following optimal sum DoF characterizations for the case with $M = 2,\ K = 3$. The first interest is placed on the minimum $\mathsf{C}^\star_\mathrm{P}(d_\Sigma)$ to achieve a sum DoF d_Σ, recalling that $\mathsf{C}^\star_\mathrm{P} = \sum_{k=1}^{K} \delta_{\mathrm{P},k}$ describes the total perfect CSIT feedback cost.

Theorem 3 (Optimal case, $M = 2, K = 3$) *For the three-user 2×1 MISO BC, the minimum total perfect CSIT feedback cost is given as*

$$\mathsf{C}^\star_\mathrm{P}(d_\Sigma) = (4d_\Sigma - 6)^+, \quad \forall\, d_\Sigma \in [0, 2] \tag{11}$$

where the total feedback cost $\mathsf{C}^\star_\mathrm{P}(d_\Sigma)$ can be distributed among all the users with some combinations $\{\delta_{\mathrm{P},k}\}_k$ such that $\delta_{\mathrm{P},k} \leq \mathsf{C}^\star_\mathrm{P}(d_\Sigma)/2$ for any k.

Proof: The converse proof is directly from Corollary 1a, while the achievability proof is shown in Section 4.2. □

Theorem 3 reveals the fundamental tradeoff between sum DoF and total perfect CSIT feedback cost (see Fig 3). The following examples are provided to offer some insights corresponding to Theorem 3.

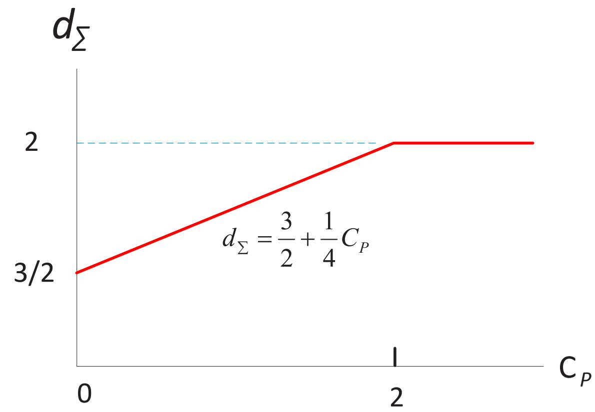

Figure 3: Optimal sum DoF (d_Σ) vs. total perfect CSIT feedback cost (C_P) for three-user 2×1 MISO BC.

Example 1 *For the target sum DoF $d_\Sigma = 3/2,\ 7/4,\ 2$, the minimum total perfect CSIT feedback cost is $\mathsf{C}_P^\star = 0,\ 1,\ 2$, respectively.*

Example 2 *The target $d_\Sigma = 7/4$ is achievable with asymmetric feedback $\boldsymbol{\delta}_P = [1/6 \quad 1/3 \quad 1/2]$, and symmetric feedback $\boldsymbol{\delta}_P = [1/3 \quad 1/3 \quad 1/3]$, and some other feedback such that $\mathsf{C}_P^\star(7/4) = 1$.*

Example 3 *The target $d_\Sigma = 2$ is achievable with asymmetric feedback $\boldsymbol{\delta}_P = [1/3 \quad 2/3 \quad 1]$, and symmetric feedback $\boldsymbol{\delta}_P = [2/3 \quad 2/3 \quad 2/3]$, and some other feedback such that $\mathsf{C}_P^\star(2) = 2$.*

Transitioning to the symmetric setting where $\delta_{\mathrm{P},k} = \delta_\mathrm{P}\ \forall\, k$, from Theorem 3 we have the fundamental tradeoff between optimal sum DoF and CSIT feedback cost δ_P.

Corollary 3a (Optimal case, $M = 2, K = 3$, δ_P) *For the three-user 2×1 MISO BC with symmetrically alternating CSIT feedback, the optimal sum DoF is given as*

$$d_\Sigma = \min\left\{\frac{3(2+\delta_P)}{4}, 2\right\}. \tag{12}$$

Now we address the questions of what is the minimum $\mathsf{C}_\mathrm{P}^\star$ to achieve the maximum sum DoF $\min\{M, K\}$ for the general BC, and how to distributed $\mathsf{C}_\mathrm{P}^\star$ among all the users, recalling again that $\mathsf{C}_\mathrm{P}^\star$ is the total perfect CSIT feedback cost.

Theorem 4 (Minimum cost for maximum DoF) *For the K-user $M \times 1$ MISO BC, the minimum total perfect CSIT feedback cost to achieve the maximum DoF is* *given by*

$$\mathsf{C}_P^\star(\min\{M,K\}) = \begin{cases} 0, & \textit{if} \quad \min\{M,K\}=1 \\ \min\{M,K\}, & \textit{if} \quad \min\{M,K\}>1 \end{cases}$$

where the total feedback cost $\mathsf{C}_P^\star$ can be distributed among all the users with any combinations $\{\delta_{P,k}\}_k$.

Proof: For the case with $\min\{M,K\}=1$, simple TDMA is optimal in terms of the DoF performance. For the case with $\min\{M,K\}>1$, the converse proof is directly derived from Corollary 1b, while the achievability proof is shown in Section 4.1. □

It is noted that Theorem 4 is a generalization of the result in [6] where only symmetric feedback was considered. The following examples are provided to offer some insights corresponding to Theorem 4.

Example 4 *For the case where $M=2,\ K=4$, the optimal 2 sum DoF performance is achievable, with asymmetric feedback $\boldsymbol{\delta}_P=[1/5 \quad 2/5 \quad 3/5 \quad 4/5]$, and symmetric feedback $\boldsymbol{\delta}_P=[1/2 \quad 1/2 \quad 1/2 \quad 1/2]$, and any other feedback such that $\mathsf{C}_P^\star=2$.*

Example 5 *For the case where $M=3,\ K=5$, the optimal 3 sum DoF performance is achievable, with asymmetric feedback $\boldsymbol{\delta}_P=[1/5 \quad 2/5 \quad 3/5 \quad 4/5 \quad 1]$, and symmetric feedback $\boldsymbol{\delta}_P=[3/5 \quad 3/5 \quad 3/5 \quad 3/5 \quad 3/5]$, and any other feedback such that $\mathsf{C}_P^\star=3$.*

The following corollary is derived from Theorem 4, where the case of having $\min\{M,K\}>1$ is considered.

Corollary 4a (Minimum cost for maximum DoF) *For the K-user $M\times 1$ MISO BC, where J users instantaneously feed back perfect (current) CSIT, with the other users feeding back delayed CSIT, then the minimum number J is $\min\{M,K\}$, in order to achieve the maximum sum DoF $\min\{M,K\}$.*

2.3 Inner bounds

In this subsection, we provide the following inner bounds on the sum DoF as a function of the CSIT cost, which are tight for many cases as stated.

Proposition 1 (Inner bound, $M=2, K\geq 3$) *For the $K(\geq 3)$-user 2×1 MISO BC, the sum DoF is bounded as*

$$d_\Sigma \geq \frac{3}{2} + \frac{K}{4}\min\{\delta_P, \frac{2}{K}\}. \tag{13}$$

Proof: The proof is shown in Section 4.3.□

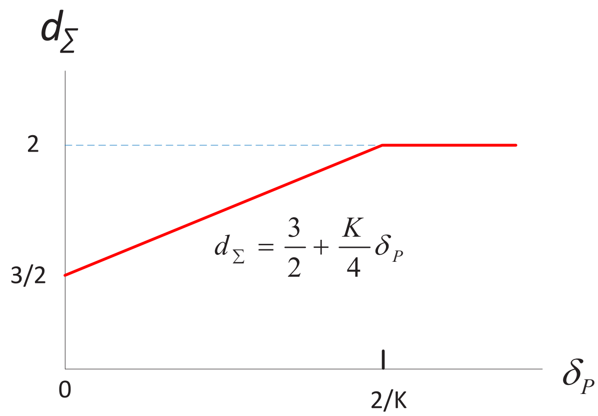

Figure 4: Achievable sum DoF d_Σ vs. δ_{P} for the $K(\geq 3)$-user 2×1 MISO BC.

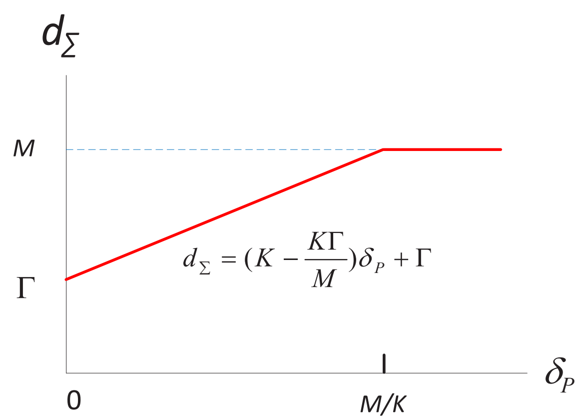

Figure 5: Achievable sum DoF d_Σ vs. δ_{P} for the MISO BC with $M < K$.

Proposition 2 (Inner bound, $M \geq K$ and $M < K$) *For the K-user $M \times 1$ MISO BC, the sum DoF for the case with $M \geq K$ is bounded as*

$$d_\Sigma \geq (K - d_{MAT}) \min\{\delta_P, 1\} + d_{MAT}, \tag{14}$$

while for the case with $M < K$, the sum DoF is bounded as

$$d_\Sigma \geq (K - \frac{K\Gamma}{M}) \min\{\delta_P, \frac{M}{K}\} + \Gamma. \tag{15}$$

Proof: The proof is shown in Section 4.4.□

Finally, we consider a case of BC with delayed CSIT feedback only, where $\delta_{\mathrm{P}} = 0$. In this case, we use $\delta_{\mathrm{D},k}$ to denote the fraction of time during which CSIT fed back from user k is delayed, and focus on the case with $\delta_{\mathrm{D},k} = \delta_{\mathrm{D}},\ \forall k$.

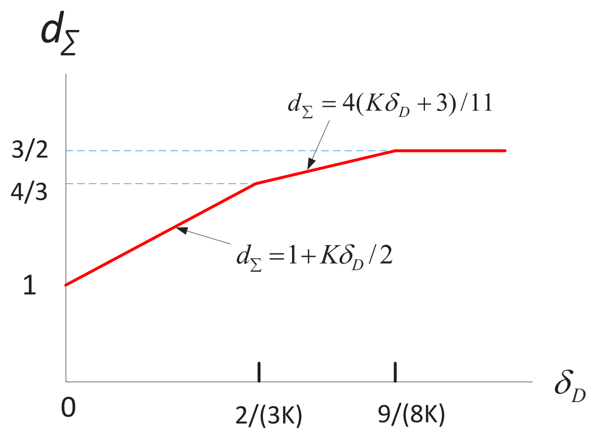

Figure 6: Achievable sum DoF d_Σ vs. δ_D for the MISO BC with $K \geq 3, M = 2$, where $\delta_\mathrm{P} = 0$.

Proposition 3 (Inner bound on DoF with delayed CSIT) *For the $K(\geq 3)$-user (2×1) MISO BC, and for the case of $\delta_\mathrm{P} = 0$, the sum DoF is bounded as*

$$d_\Sigma \geq \min\left\{1 + \frac{K}{2}\delta_\mathrm{D},\ \frac{12}{11} + \frac{4K}{11}\delta_\mathrm{D},\ \frac{3}{2}\right\}. \tag{16}$$

Proof: The proof is shown in Section 4.5.□

Remark 3 *For the K-user MISO BC with current and delayed CSIT feedback, by increasing the number of users, the same DoF performance can be achievable with decreasing feedback cost per user. For example, for the K-user MISO BC with $M = 2$, by increasing K we can achieve any fixed DoF within the range of $(1, 2]$, with decreasing $\delta_\mathrm{P} \leq \frac{2}{K}$, and $\delta_\mathrm{D} \leq \frac{9}{8K}$, both of which approach to 0 as K is large.*

3 Converse proof of Theorem 1

In this proof we will use Proposition 4 shown in the appendix, and use the following notations:

$$\begin{aligned}
\boldsymbol{S}_t &\triangleq \left[\boldsymbol{h}_{1,t}\ \cdots\ \boldsymbol{h}_{K,t}\right]^\mathsf{T} \\
\hat{\boldsymbol{S}}_t &\triangleq \left[\hat{\boldsymbol{h}}_{1,t}\ \cdots\ \hat{\boldsymbol{h}}_{K,t}\right]^\mathsf{T} \\
\Omega^n &\triangleq \{\boldsymbol{S}_t, \hat{\boldsymbol{S}}_t\}_{t=1}^n \\
y_k^n &\triangleq \{y_{k,t}\}_{t=1}^n.
\end{aligned}$$

Giving the observations and messages of users $1, \ldots, k-1$ to user k, we establish the following genie-aided upper bounds on the achievable rates

$$nR_1 \leq I(W_1; y_1^n \,|\, \Omega^n) + n\epsilon \tag{17}$$
$$nR_2 \leq I(W_2; y_1^n, y_2^n \,|\, W_1, \Omega^n) + n\epsilon \tag{18}$$
$$\vdots$$
$$nR_K \leq I(W_K; y_1^n, y_2^n, \ldots, y_K^n \,|\, W_1, \ldots, W_{K-1}, \Omega^n) + n\epsilon \tag{19}$$

where we apply Fano's inequality and some basic chain rules of mutual information using the fact that messages from different users are independent. Alternatively, we have

$$nR_1 \leq h(y_1^n \,|\, \Omega^n) - h(y_1^n \,|\, W_1, \Omega^n) + n\epsilon \tag{20}$$
$$nR_2 \leq h(y_1^n, y_2^n \,|\, W_1, \Omega^n) - h(y_1^n, y_2^n \,|\, W_1, W_2, \Omega^n) + n\epsilon \tag{21}$$
$$\vdots$$
$$\begin{aligned} nR_K \leq\ & h(y_1^n, \ldots, y_K^n \,|\, W_1, \ldots, W_{K-1}, \Omega^n) - h(y_1^n, \ldots, y_K^n \,|\, W_1, \ldots, W_K, \Omega^n) \\ & + n\epsilon. \end{aligned} \tag{22}$$

Therefore, it follows that

$$\begin{aligned}
&\sum_{k=1}^{K} \frac{n}{k'}(R_k - \epsilon) \\
&\leq \sum_{k=1}^{K-1}\left(\frac{1}{(k+1)'}h(y_1^n, \ldots, y_{k+1}^n | W_1, \ldots, W_k, \Omega^n) - \frac{1}{k'}h(y_1^n, \ldots, y_k^n \,|\, W_1, \ldots, W_k, \Omega^n)\right) \\
&\qquad + h(y_1^n \,|\, \Omega^n) - \frac{1}{K'}h(y_1^n, \ldots, y_K^n \,|\, W_1, \ldots, W_K, \Omega^n) \qquad (23) \\
&\leq \sum_{k=1}^{K-1}\sum_{t=1}^{n}\left(\frac{1}{(k+1)'}h(y_{1,t}, \ldots, y_{k+1,t} \,|\, y_1^{t-1}, \ldots, y_k^{t-1}, W_1, \ldots, W_k, \Omega^n)\right. \\
&\quad \left. - \frac{1}{k'}h(y_{1,t}, \ldots, y_{k,t} \,|\, y_1^{t-1}, \ldots, y_k^{t-1}, W_1, \ldots, W_k, \Omega^n)\right) + n\log P + n\, o(\log P) \\
&\qquad (24) \\
&\leq \log P \sum_{k=1}^{K-1}\sum_{t=1}^{n} \frac{(k+1)' - k'}{k'(k+1)'} \sum_{i=1}^{k} \alpha_{i,t} + n\log P + n\, o(\log P) \qquad (25) \\
&= n\log P \sum_{k=1}^{K-1} \frac{(k+1)' - k'}{k'(k+1)'} \sum_{i=1}^{k} \bar{\alpha}_i + n\log P + n\, o(\log P) \qquad (26) \\
&= n\log P \sum_{k=1}^{K-1} \left(\frac{1}{k'} - \frac{1}{K'}\right)\bar{\alpha}_k + n\log P + n\, o(\log P) \qquad (27)
\end{aligned}$$

where we define

$$k' \triangleq \min\{k, M\}; \tag{28}$$

the inequality (24) is due to 1) the chain rule of differential entropy, 2) the fact that removing condition does not decrease differential entropy, 3) $h(y_{1,t} \,|\, \Omega^n) \le \log P + o(\log P)$, i.e., Gaussian distribution maximizes differential entropy under covariance constraint, and 4) $h(y_1^n, \ldots, y_K^n \,|\, W_1, \ldots, W_K, \Omega^n) = h(z_1^n, \ldots, z_K^n) > 0$; (25) is from Proposition 4 by setting $U = \{y_1^{t-1}, \ldots, y_k^{t-1}, W_1, \ldots, W_k, \Omega^n\} \setminus \{\boldsymbol{S}_t, \hat{\boldsymbol{S}}_t\}$, $H = \boldsymbol{S}_t$, and $\hat{H} = \hat{\boldsymbol{S}}_t$; the last equality is obtained after putting the summation over k inside the summation over i and some basic manipulations. Similarly, we can interchange the roles of the users and obtain the same genie-aided bounds. Finally, the single antenna constraint gives that $d_i \le 1,\ i = 1, \cdots, K$. With this, we complete the proof.

4 Details of achievability proofs

In this section, we provide the details of the achievability proofs. Specifically, the achievability proof of Theorem 4 is first described in Section 4.1, which can be applied in parts for the achievability proof of Theorem 4.2 shown in Section 4.2, with the proposition proofs shown in the rest of this section.

4.1 Achievability proof of Theorem 4

We will prove that, the optimal sum DoF $d_\Sigma = \min\{M, K\}$ is achievable with any CSIT feedback cost $\boldsymbol{\delta}_{\mathrm{P}} \triangleq [\delta_{\mathrm{P},1} \quad \delta_{\mathrm{P},2} \quad \cdots \quad \delta_{\mathrm{P},K}] \in \mathbb{R}^K$ such that $\mathsf{C}_{\mathrm{P}} = \sum_{k=1}^K \delta_{\mathrm{P},k} = \min\{M, K\}$. First of all, we note that there exists a minimum number n such that

$$\boldsymbol{\delta}'_{\mathrm{P}} \triangleq [\delta'_{\mathrm{P},1} \quad \delta'_{\mathrm{P},2} \ \cdots \ \delta'_{\mathrm{P},K}] \triangleq n\boldsymbol{\delta}_{\mathrm{P}} = [n\delta_{\mathrm{P},1} \quad n\delta_{\mathrm{P},2} \ \cdots \ n\delta_{\mathrm{P},K}] \in \mathbb{Z}^K$$

is an integer vector. The explicit communication with n channel uses is given as follows:

- Step 1: Initially set time index $t = 1$.
- Step 2: Permute user indices orderly into a set $\mathcal{U}$ such that $\delta'_{\mathrm{P},\mathcal{U}(1)} \le \delta'_{\mathrm{P},\mathcal{U}(2)} \le \cdots \le \delta'_{\mathrm{P},\mathcal{U}(K)}$, where $\mathcal{U}(k)$ denotes the k th element of the $\mathcal{U}$ set, and where $\mathcal{U}(k) \in \{1, 2, \cdots, K\}$.
- Step 3: Select $\min\{M, K\}$ users to communicate: users $\mathcal{U}(K - \min\{M, K\} + 1), \cdots, \mathcal{U}(K-1), \mathcal{U}(K)$.
- Step 4: Let selected users feed back perfect CSIT at time t, keeping the rest users silent.

Table 1: Summary of the scheme for achieving $d^*_{\sum} = 2$ with $\mathsf{C}^\star_{\text{P}} = 2$, where $M = 2,\ K = 3$, $\delta_{\text{P},1} = 1/3,\ \delta_{\text{P},2} = 2/3,\ \delta_{\text{P},3} = 1$.

time t	1	2	3
$\mathcal{U}$	$\{1,2,3\}$	$\{1,2,3\}$	$\{2,1,3\}$
$\{\delta'_{\text{P},\mathcal{U}(1)},\ \delta'_{\text{P},\mathcal{U}(2)},\ \delta'_{\text{P},\mathcal{U}(3)}\}$	$\{1,2,3\}$	$\{1,1,2\}$	$\{0,1,1\}$
Active users	user 2, 3	user 2, 3	user 1, 3
Perfect CSIT feedback	user 3: yes user 2: yes user 1: no	user 3: yes user 2: yes user 1: no	user 3: yes user 2: no user 1: yes
No. of transmitted symbols	2	2	2

- Step 5: The transmitter sends $\min\{M, K\}$ independent symbols to those selected users respectively, which can be done with simple zero-forcing.
- Step 6: Set $\delta'_{\text{P},\mathcal{U}(k)} = \delta'_{\text{P},\mathcal{U}(k)} - 1,\ k = K - \min\{M, K\} + 1, \cdots, K-1, K$.
- Step 7: Set $t = t + 1$. If renewed $t > n$ then terminate, else go back to step 2.

In the above communication with n channel uses, the algorithm guarantees that user i is selected by $\delta'_{\text{P},k} = n\delta_{\text{P},k}$ times totally, and that $\min\{M, K\}$ different users are selected in each channel use. As a result, the optimal sum DoF $d_\Sigma = \min\{M, K\}$ is achievable.

Now we consider an example with $M = 2,\ K = 3$, and $\boldsymbol{\delta}_{\text{P}} = [1/3 \quad 2/3 \quad 1]$, and show that the optimal sum DoF $d_\Sigma = 2$ is achievable with the following communication:

- Let $n = 3$. Initially $\delta'_{\text{P},1} = n\delta_{\text{P},1} = 1$, $\delta'_{\text{P},2} = n\delta_{\text{P},2} = 2$, $\delta'_{\text{P},3} = n\delta_{\text{P},3} = 3$.
- For $t = 1$, we have $\mathcal{U} = \{1, 2, 3\}$, and $\delta'_{\text{P},\mathcal{U}(1)} = 1$, $\delta'_{\text{P},\mathcal{U}(2)} = 2$, $\delta'_{\text{P},\mathcal{U}(3)} = 3$. Users 3 and 2 are selected to communicate.
- For $t = 2$, we update the parameters as $\mathcal{U} = \{1, 2, 3\}$, and $\delta'_{\text{P},\mathcal{U}(1)} = 1$, $\delta'_{\text{P},\mathcal{U}(2)} = 1$, $\delta'_{\text{P},\mathcal{U}(3)} = 2$. At this time, again user 3 and user 2 are selected to communicate.
- For $t = 3$, we update the parameters as $\mathcal{U} = \{2, 1, 3\}$, and $\delta'_{\text{P},\mathcal{U}(1)} = 0$, $\delta'_{\text{P},\mathcal{U}(2)} = 1$, $\delta'_{\text{P},\mathcal{U}(3)} = 1$. At this time, user 3 and user 1 are selected to communicate. After that the communication terminates.

In the above communication with three channel uses, the transmitter sends two symbols in each channel use, which allows for the optimal sum DoF $d_\Sigma = 2$ (see Table 1).

4.2 Achievability proof of Theorem 3

We proceed to show that, any sum DoF $d_\Sigma \in [3/2, 2]$ is achievable with the feedback

$$\delta_{\mathrm{P},k} \le \frac{\mathsf{C}_\mathrm{P}}{2},\ k = 1,2,3, \quad \text{such that} \quad \mathsf{C}_\mathrm{P} = \sum_{k=1}^{3} \delta_{\mathrm{P},k} = 4d_\Sigma - 6.$$

First of all, we note that there exists a minimum number n such that

$$[2n\delta_{\mathrm{P},1}/\mathsf{C}_\mathrm{P} \quad 2n\delta_{\mathrm{P},2}/\mathsf{C}_\mathrm{P} \quad n2\delta_{\mathrm{P},3}/\mathsf{C}_\mathrm{P}] \in \mathbb{Z}^3, \quad \text{and} \quad 2n/\mathsf{C}_\mathrm{P} \in \mathbb{Z}.$$

The scheme has two blocks, with the first block consisting of n channel uses, and the second block consisting of

$$n^{'} = 2n/\mathsf{C}_\mathrm{P} - n$$

channel uses. In the first block, we use the algorithm shown in the Section 4.1 to achieve the full sum DoF in those n channel uses, during which user k feeds back perfect CSIT in $2n\delta_{\mathrm{P},3}/\mathsf{C}_\mathrm{P}$ channel uses, for $k = 1, 2, 3$. In the second block, we use the Maddah-Ali and Tse scheme in [4] to achieve 3/2 sum DoF in those $n^{'}$ channel uses, during which each user feeds back delayed CSIT only.

The communication with n channel uses for the first block is given as follows:

- Step 1: Let $\delta'_{\mathrm{P},k} = 2n\delta_{\mathrm{P},k}/\mathsf{C}_\mathrm{P}$ for all k. Initially, set $t = 1$.
- The steps 2, 3, 4, 5, 6 are the same as those in the algorithm shown in Section 4.1, for $M = 2, K = 3$.
- Step 7: Set $t = t + 1$. If renewed $t > n$ then terminate, else go back to step 2.

In the above communication with n channel uses, the algorithm guarantees that user k, $k = 1, 2, 3$, is selected by $\delta'_{\mathrm{P},k} = 2n\delta_{\mathrm{P},k}/\mathsf{C}_\mathrm{P}$ times. We note that $\delta'_{\mathrm{P},k} \le n$ under the constraint $\delta_{\mathrm{P},k} \le \mathsf{C}_\mathrm{P}/2$ for any k, and that $\sum_{k=1}^{K} \delta'_{\mathrm{P},k} = 2n$, to suggest that in each timeslot two different users are selected, which allows for the optimal 2 sum DoF in this block.

As stated, in the second block, we use the MAT scheme to achieve the 3/2 sum DoF in those $n^{'}$ channel uses, during which each user feeds back delayed CSIT only. As a result, in the total $n + n^{'}$ channel uses communication, user $k = 1, 2, 3$ feeds back perfect CSIT in $2n\delta_{\mathrm{P},k}/(\mathsf{C}_\mathrm{P}(n+n^{'})) = \delta_{\mathrm{P},k}$ fraction of communication period, with achievable sum DoF given as

$$d_\Sigma = \frac{2n}{(n+n^{'})} + \frac{3n^{'}}{2(n+n^{'})} = \frac{3}{2} + \frac{1}{4}\mathsf{C}_\mathrm{P}.$$

We note that the achievability scheme applies to the case of having some $\delta_{\mathrm{P},1}, \delta_{\mathrm{P},2}, \delta_{\mathrm{P},3} \le \mathsf{C}_\mathrm{P}/2$ such that $\mathsf{C}_\mathrm{P} = 4d_\Sigma - 6$, and allows to achieve any sum DoF $d_\Sigma \in [3/2, 2]$. Apparently, $\mathsf{C}_\mathrm{P} = 0$ allows for any sum DoF $d_\Sigma \in [0, 3/2]$, which completes the proof.

4.3 Proof of Proposition 1

The achievability scheme is based on time sharing between two strategies of CSIT feedback, i.e., delayed CSIT feedback with $\delta'_{\mathrm{P}} = 0$ and alternating CSIT feedback with $\delta''_{\mathrm{P}} = \frac{2}{K}$, where the first strategy achieves $d'_{\Sigma} = 3/2$ by applying Maddah-Ali and Tse (MAT) scheme (see in [4]), with the second strategy achieving $d''_{\Sigma} = 2$ by using alternating CSIT feedback manner (see in [6]).

Let $\Delta \in [0, 1]$ (res. $1 - \Delta$) be the fraction of time during which the first (res. second) CSIT feedback strategy is used in the communication. As a result, the final feedback cost (per user) is given as

$$\delta_{\mathrm{P}} = \delta'_{\mathrm{P}}\Delta + \delta''_{\mathrm{P}}(1 - \Delta), \tag{29}$$

implying that

$$\Delta = \frac{\delta''_{\mathrm{P}} - \delta_{\mathrm{P}}}{\delta''_{\mathrm{P}} - \delta'_{\mathrm{P}}}, \tag{30}$$

with final sum DoF given as

$$\begin{aligned} d_{\Sigma} &= d'_{\Sigma}\Delta + d''_{\Sigma}(1 - \Delta) \\ &= d''_{\Sigma} + \Delta(d'_{\Sigma} - d''_{\Sigma}) \\ &= d''_{\Sigma} + (d'_{\Sigma} - d''_{\Sigma})\frac{\delta''_{\mathrm{P}} - \delta_{\mathrm{P}}}{\delta''_{\mathrm{P}} - \delta'_{\mathrm{P}}} \\ &= \frac{3}{2} + \frac{K}{4}\delta_{\mathrm{P}} \end{aligned} \tag{31}$$

which completes the proof.

4.4 Proof of Proposition 2

For the case with $M \geq K$, the proposed scheme is based on time sharing between delayed CSIT feedback with $\delta'_{\mathrm{P}} = 0$ and full CSIT feedback with $\delta''_{\mathrm{P}} = 1$, where the first feedback strategy achieves $d^{'}_{\sum} = d_{\mathrm{MAT}}$ by applying MAT scheme, with the second one achieving $d^{''}_{\sum} = K$. As a result, following the steps in (29), (30), (31), the final sum DoF is calculated as

$$\begin{aligned} d_{\Sigma} &= d''_{\Sigma} + (d'_{\Sigma} - d''_{\Sigma})\frac{\delta''_{\mathrm{P}} - \delta_{\mathrm{P}}}{\delta''_{\mathrm{P}} - \delta'_{\mathrm{P}}} \\ &= (K - d_{\mathrm{MAT}})\delta_{\mathrm{P}} + d_{\mathrm{MAT}} \end{aligned}$$

where $\delta_{\mathrm{P}} \in [0, 1]$ is the final feedback cost (per user) for this case.

Similar approach is exploited for the case with $M < K$. In this case, we apply time sharing between delayed CSIT feedback with $\delta'_{\mathrm{P}} = 0$ and alternating CSIT feedback with $\delta''_{\mathrm{P}} = M/K$. In this case, the first feedback strategy achieves $d'_{\Sigma} = \Gamma$ by applying MAT scheme, with the second strategy achieving $d''_{\Sigma} = M$

Table 2: Summary of the achievability scheme for achieving $d_\Sigma = \frac{4}{3}$ with $\delta_\mathrm{D} = \frac{2}{3K}$.

block index	1	2	3	$\cdots$	K
No. of channel uses	3	3	3	$\cdots$	3
Active users	user 1, 2	user 2, 3	user 3, 4	$\cdots$	user K, 1
Delayed CSIT feedback fraction in a block	user 1: $1/3$ user 2: $1/3$ the rest: 0	user 2: $1/3$ user 3: $1/3$ the rest: 0	user 3: $1/3$ user 4: $1/3$ the rest: 0	$\cdots$	user K: $1/3$ user 1: $1/3$ the rest: 0
Sum DoF in a block	$4/3$	$4/3$	$4/3$	$\cdots$	$4/3$

by using alternating CSIT feedback manner. As a result, for $\delta_\mathrm{P} \in [0, \frac{M}{K}]$ being the final feedback cost for this case, the final sum DoF is calculated as

$$\begin{aligned} d_\Sigma &= d''_\Sigma + (d'_\Sigma - d''_\Sigma)\frac{\delta''_\mathrm{P} - \delta_\mathrm{P}}{\delta''_\mathrm{P} - \delta'_\mathrm{P}} \\ &= (K - \frac{K\Gamma}{M})\delta_\mathrm{P} + \Gamma \end{aligned}$$

which completes the proof.

4.5 Proof of Proposition 3

As shown in the Fig 6, the sum DoF performance has three regions:

$$d_\Sigma = \begin{cases} 1 + \frac{K}{2}\delta_\mathrm{D}, & \delta_\mathrm{D} \in [0, \frac{2}{3K}] \\ \frac{12}{11} + \frac{4K}{11}\delta_\mathrm{D}, & \delta_\mathrm{D} \in [\frac{2}{3K}, \frac{9}{8K}] \\ 3/2, & \delta_\mathrm{D} \in [\frac{9}{8K}, 1]. \end{cases}$$

In the following, we will prove that the sum DoF $d_\Sigma = 1,\ \frac{4}{3},\ \frac{3}{2}$ are achievable with $\delta_\mathrm{D} = 0,\ \frac{2}{3K},\ \frac{9}{8K}$, respectively. At the end, the whole DoF performance declared can be achievable by time sharing between those performance points.

The proposed scheme achieving $d_\Sigma = \frac{4}{3}$ with $\delta_\mathrm{D} = \frac{2}{3K}$, is a modified version of the MAT scheme in [4]. The new scheme has K blocks, with each block consisting of three channel uses. In each block, four independent symbols are sent to two orderly selected users, which can be done with MAT scheme with each of two chosen user feeding back delayed CSIT in one channel use. As a result, $d_\Sigma = \frac{4}{3}$ is achievable with $\delta_\mathrm{D} = \frac{2}{3K}$, using the fact that each of K users needs to feed back delayed CSIT twice only in the whole communication (see Table 2).

Similarly, the proposed scheme achieving $d_\Sigma = \frac{3}{2}$ with $\delta_\mathrm{D} = \frac{9}{8K}$ has K blocks, with each block consisting of 8 channel uses. In each block, 3 out of K users are

Table 3: Summary of the achievability scheme for achieving $d_\Sigma = \frac{3}{2}$ with $\delta_{\mathrm{D}} = \frac{9}{8K}$.

block index	1	2	3	$\cdots$	K
No. of channel uses	8	8	8	$\cdots$	8
Active users	user 1, 2, 3	user 2, 3, 4	user 3, 4, 5	$\cdots$	user K, 1, 2
Delayed CSIT feedback fraction in a block	user 1: $3/8$ user 2: $3/8$ user 3: $3/8$ the rest: 0	user 2: $3/8$ user 3: $3/8$ user 4: $3/8$ the rest: 0	user 3: $3/8$ user 4: $3/8$ user 5: $3/8$ the rest: 0	$\cdots$	user K: $3/8$ user 1: $3/8$ user 2: $3/8$ the rest: 0
Sum DoF in a block	$3/2$	$3/2$	$3/2$	$\cdots$	$3/2$

selected to communicate. In this case, 12 independent symbols are sent to the chosen users during each block, which can be done with another MAT scheme with each of chosen users feeding back delayed CSIT in 3 channel uses. As a result, $d_\Sigma = \frac{3}{2}$ is achievable with $\delta_{\mathrm{D}} = \frac{9}{8K}$, using the fact that each of K users needs to feed back delayed CSIT 9 times only in the whole communication (see Table 3).

Finally, $d_\Sigma = 1$ is achievable without any CSIT. By now, we complete the proof.

5 Conclusions

This work considered the general multiuser MISO BC, and established inner and outer bounds on the tradeoff between DoF performance and CSIT feedback quality, which are optimal for many cases. Those bounds, as well as some analysis, were provided with the aim of giving insights on how much CSIT feedback to achieve a certain DoF performance.

6 Appendix

In this section, we will provide Proposition 4 used for the converse proof, as well as three lemmas to be used, together with corresponding proofs. For simplicity we drop the time index.

Proposition 4 *Let*

$$y_k = \boldsymbol{h}_k^{\mathsf{T}} \boldsymbol{x} + z_k, \tag{32}$$
$$\boldsymbol{y}_k \triangleq [y_1 \; y_2 \; \cdots \; y_k]^{\mathsf{T}} \tag{33}$$
$$\boldsymbol{z}_k \triangleq [z_1 \; z_2 \; \cdots \; z_k]^{\mathsf{T}} \tag{34}$$
$$\boldsymbol{H}_k \triangleq [\boldsymbol{h}_1 \; \boldsymbol{h}_2 \; \cdots \; \boldsymbol{h}_k]^{\mathsf{T}} \tag{35}$$
$$\boldsymbol{H} \triangleq [\boldsymbol{h}_1 \; \boldsymbol{h}_2 \; \cdots \; \boldsymbol{h}_K]^{\mathsf{T}} \tag{36}$$
$$\boldsymbol{H} = \hat{\boldsymbol{H}} + \tilde{\boldsymbol{H}} \tag{37}$$

where $\tilde{\boldsymbol{h}}_i \in \mathbb{C}^{M\times 1}$ has i.i.d. $\mathcal{N}_{\mathbb{C}}(0, \sigma_i^2)$ entries. Then, for any U such that $p_{X|U\hat{H}\tilde{H}} = p_{X|U\hat{H}}$ and $K \geq m \geq l$, we have

$$l'\, h(\boldsymbol{y}_m | U, \hat{H}, \tilde{H}) - m'\, h(\boldsymbol{y}_l | U, \hat{H}, \tilde{H}) \leq -(m' - l') \sum_{i=1}^{l} \log \sigma_i^2 + o(\log \mathsf{snr}) \tag{38}$$

where we define $l' \triangleq \min\{l, M\}$ and $m' \triangleq \min\{m, M\}$.

Lemma 1 [4] *Let $\boldsymbol{G} = \hat{\boldsymbol{G}} + \tilde{\boldsymbol{G}} \in \mathbb{C}^{m\times m}$ where $\tilde{\boldsymbol{G}}$ has i.i.d. $\mathcal{N}_c(0,1)$ entries, and $\tilde{\boldsymbol{G}}$ is independent of $\hat{\boldsymbol{G}}$. Then, we have*

$$\mathbb{E}_{\tilde{G}}\big[\log\det(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G})\big] = \sum_{i=1}^{\tau} \log\big(\lambda_i(\hat{\boldsymbol{G}}^{\mathsf{H}}\hat{\boldsymbol{G}})\big) + o(\log \mathsf{snr}) \tag{39}$$

where $\lambda_i(\hat{\boldsymbol{G}}^{\mathsf{H}}\hat{\boldsymbol{G}})$ denotes the i th largest eigenvalue of $\hat{\boldsymbol{G}}^{\mathsf{H}}\hat{\boldsymbol{G}}$; τ is the number of eigenvalues of $\hat{\boldsymbol{G}}^{\mathsf{H}}\hat{\boldsymbol{G}}$ that do not vanish with snr, i.e., $\lambda_i(\hat{\boldsymbol{G}}^{\mathsf{H}}\hat{\boldsymbol{G}}) = o(1)$ when snr is large, $\forall\, i > \tau$.

Lemma 2 *For $\boldsymbol{P} \in \mathbb{C}^{m\times m}$ a permutation matrix and $\boldsymbol{A} \in \mathbb{C}^{m\times m}$, let $\boldsymbol{AP} = \boldsymbol{QR}$ be the QR decomposition of the column permuted version of $\boldsymbol{A}$. Then, there exist at least one permutation matrix $\boldsymbol{P}$ such that*

$$r_{ii}^2 \geq \frac{1}{m-i+1} \lambda_i(\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}), \quad i = 1, \ldots, m \tag{40}$$

where as stated $\lambda_i(\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A})$ is the i th largest eigenvalue of $\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}$; r_{ii} is the i th diagonal elements of $\boldsymbol{R}$.

Lemma 3 *For any matrix $\boldsymbol{A} \in \mathbb{C}^{m\times m}$, there exists a column permuted version $\bar{\boldsymbol{A}}$, such that*

$$\det(\bar{\boldsymbol{A}}_{\mathcal{I}}^{\mathsf{H}} \bar{\boldsymbol{A}}_{\mathcal{I}}) \geq m^{-|\mathcal{I}|} \prod_{i\in\mathcal{I}} \lambda_i(\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}), \quad \forall \mathcal{I} \subseteq \{1, \ldots, m\} \tag{41}$$

where $\bar{\boldsymbol{A}}_{\mathcal{I}} = [A_{ji} : \; j \in \{1, \ldots, m\}, \, i \in \mathcal{I}] \in \mathbb{C}^{m\times|\mathcal{I}|}$ is the submatrix of $\boldsymbol{A}$ formed by the columns with indices in $\mathcal{I}$.

---

[4] We note that Lemma 1 is a slightly more general version of the result in [24, Lemma 6].

6.1 Proof of Lemma 1

Let us perform a singular value decomposition (SVD) on the matrix $\hat{\boldsymbol{G}}$, i.e., $\hat{\boldsymbol{G}} = \boldsymbol{U}\begin{bmatrix}\boldsymbol{D}_1 & \\ & \boldsymbol{D}_2\end{bmatrix}\boldsymbol{V}^{\mathsf{H}}$ where $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{C}^{m\times m}$ are unitary matrices and $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are $\tau'\times\tau'$ and $(m-\tau')\times(m-\tau')$ diagonal matrices of the singular values of $\hat{\boldsymbol{G}}$. Without loss of generality, we assume that the i th singluar value, $i = 1,\dots,m$, scales with snr as snr^{b_i}, when snr is large. Moreover, the singular values in $\boldsymbol{D}_1$ are such that $b_i > 0$ and those in $\boldsymbol{D}_2$ verify $b_i \le 0$. First, we have the following lower bound

$$
\begin{aligned}
&\mathbb{E}_{\tilde{G}}\big[\log\det(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G})\big] \\
&\quad= \mathbb{E}_{\boldsymbol{M}}\left[\log\det\left(\left(\begin{bmatrix}\boldsymbol{D}_1 & \\ & \boldsymbol{D}_2\end{bmatrix}+\boldsymbol{M}\right)^{\mathsf{H}}\left(\begin{bmatrix}\boldsymbol{D}_1 & \\ & \boldsymbol{D}_2\end{bmatrix}+\boldsymbol{M}\right)\right)\right] && (42)\\
&\quad\ge \mathbb{E}_{\boldsymbol{M}}\left[\log\det\left(\left(\begin{bmatrix}\boldsymbol{D}_1 & \\ & 0\end{bmatrix}+\boldsymbol{M}\right)^{\mathsf{H}}\left(\begin{bmatrix}\boldsymbol{D}_1 & \\ & 0\end{bmatrix}+\boldsymbol{M}\right)\right)\right] && (43)\\
&\quad= \mathbb{E}_{\boldsymbol{M}}\Big[\log\big|\det(\boldsymbol{D}_1+\boldsymbol{M}_{11})\det\big(\boldsymbol{M}_{22}-\underbrace{\boldsymbol{M}_{21}(\boldsymbol{D}_1+\boldsymbol{M}_{11})^{-1}}_{\boldsymbol{B}}\boldsymbol{M}_{12}\big)\big|^2\Big] && (44)\\
&\quad= \log\big|\det(\boldsymbol{D}_1)\big|^2 + \mathbb{E}_{\boldsymbol{M}_{11}}\Big[\log\big|\det\big(I+\boldsymbol{D}_1^{-1}\boldsymbol{M}_{11}\big)\big|^2\Big] \\
&\qquad+ \mathbb{E}_{\boldsymbol{B}}\mathbb{E}_{\tilde{\boldsymbol{M}}}\big[\log\det(\tilde{\boldsymbol{M}}^H(I+\boldsymbol{B}\boldsymbol{B}^H)\tilde{\boldsymbol{M}})\big] && (45)\\
&\quad\ge \log\big|\det(\boldsymbol{D}_1)\big|^2 + \mathbb{E}_{\boldsymbol{M}_{11}}\Big[\log\big|\det\big(I+\boldsymbol{D}_1^{-1}\boldsymbol{M}_{11}\big)\big|^2\Big] \\
&\qquad+ \underbrace{\mathbb{E}_{\tilde{\boldsymbol{M}}}\big[\log\det(\tilde{\boldsymbol{M}}^H\tilde{\boldsymbol{M}})\big]}_{(\ln 2)^{-1}\sum_{l=0}^{m-\tau'-1}\psi(m-\tau'-l)=O(1)} && (46)
\end{aligned}
$$

where we define $\boldsymbol{M} \triangleq \boldsymbol{U}^{\mathsf{H}}\tilde{\boldsymbol{G}}\boldsymbol{V} = \begin{bmatrix}\boldsymbol{M}_{11} & \boldsymbol{M}_{12}\\ \boldsymbol{M}_{21} & \boldsymbol{M}_{22}\end{bmatrix}$ with $\boldsymbol{M}_{11} \in \mathbb{C}^{\tau'\times\tau'}$, and remind that the entries of $\boldsymbol{M}$, thus of $\boldsymbol{M}_{ij}$, $i,j = 1,2$, are also i.i.d. $\mathcal{N}_c(0,1)$; (43) is from the fact that expectation of the log determinant of a non-central Wishart matrix is non-decreasing with in the "line-of-sight" component [25]; (44) is due to the identity $\det\left(\begin{bmatrix}\boldsymbol{N}_{11} & \boldsymbol{N}_{12}\\ \boldsymbol{N}_{21} & \boldsymbol{N}_{22}\end{bmatrix}\right) = \det(\boldsymbol{N}_{11})\det(\boldsymbol{N}_{22}-\boldsymbol{N}_{21}\boldsymbol{N}_{11}^{-1}\boldsymbol{N}_{12})$ whenever $\boldsymbol{N}_{11}$ is square and invertible; in (45), we notice that, given the matrix $\boldsymbol{B} \triangleq \boldsymbol{M}_{21}(\boldsymbol{D}_1+\boldsymbol{M}_{11})^{-1}$, the columns of $\boldsymbol{M}_{22}-\boldsymbol{B}\boldsymbol{M}_{12}$ are i.i.d. $\mathcal{N}_c(0, I+\boldsymbol{B}\boldsymbol{B}^H)$, from which $|\det(\boldsymbol{M}_{22}-\boldsymbol{B}\boldsymbol{M}_{12})|^2$ is equivalent in distribution to $\det(\tilde{\boldsymbol{M}}^H(I+\boldsymbol{B}\boldsymbol{B}^H)\tilde{\boldsymbol{M}})$ where $\tilde{\boldsymbol{M}} \in \mathbb{C}^{(m-\tau')\times(m-\tau')}$ has i.i.d. $\mathcal{N}_c(0,1)$ entries; the last inequality is from $\tilde{\boldsymbol{M}}^H(I+\boldsymbol{B}\boldsymbol{B}^H)\tilde{\boldsymbol{M}} \succeq \tilde{\boldsymbol{M}}^H\tilde{\boldsymbol{M}}$ and therefore $\det(\tilde{\boldsymbol{M}}^H(I+\boldsymbol{B}\boldsymbol{B}^H)\tilde{\boldsymbol{M}}) \ge \det(\tilde{\boldsymbol{M}}^H\tilde{\boldsymbol{M}})$, $\forall\,\boldsymbol{B}$; the closed-form term in the last inequality is due to [26] with $\psi(\cdot)$ being Euler's digamma function. In the following, we show that

$$\mathbb{E}\big[\log\big|\det\big(I+\boldsymbol{D}_1^{-1}\boldsymbol{M}_{11}\big)\big|^2\big] \ge O(1)$$

as well. To that end, we use the fact that the distribution of $\boldsymbol{M}_{11}$ is invariant to rotation, and so for $\boldsymbol{D}_1^{-1}\boldsymbol{M}_{11}$. Specifically, introducing $\theta \sim \text{Unif}(0, 2\pi]$ that is independent of the rest of the random variables, we have

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{M}_{11}}\left[\log\left|\det\left(I + \boldsymbol{D}_1^{-1}\boldsymbol{M}_{11}\right)\right|^2\right] \\
&= \mathbb{E}_{\boldsymbol{M}_{11},\theta}\left[\log\left|\det\left(I + \boldsymbol{D}_1^{-1}\boldsymbol{M}_{11}e^{j\theta}\right)\right|^2\right] && (47)\\
&= \mathbb{E}_{\boldsymbol{M}_{11},\theta}\left[\log\left|\det\left(e^{-j\theta}I + \boldsymbol{D}_1^{-1}\boldsymbol{M}_{11}\right)\right|^2\right] && (48)\\
&= \sum_{i=1}^{\tau'} \mathbb{E}_J\mathbb{E}_\theta\left[\log|e^{-j\theta} + \underbrace{\lambda_i(\boldsymbol{D}^{-1}\boldsymbol{M}_{11})}_{J_i}|^2\right] && (49)\\
&= \sum_{i=1}^{\tau'} \mathbb{E}_J\mathbb{E}_\theta[\log(1 + |J_i|^2 + 2|J_i|\cos(\theta + \phi(J_i)))] && (50)\\
&= \sum_{i=1}^{\tau'} \mathbb{E}_J\mathbb{E}_\theta[\log(1 + |J_i|^2 + 2|J_i|\cos(\theta))] && (51)\\
&\geq \sum_{i=1}^{\tau'} \left[\mathbb{E}_J\left(\log(1 + |J_i|^2)\right) - 1\right] && (52)\\
&\geq -\tau' && (53)
\end{aligned}
$$

where the first equality is from the fact that $\boldsymbol{M}_{11}$ is equivalent to $\boldsymbol{M}_{11}e^{j\theta}$ as long as θ is independent of $\boldsymbol{M}_{11}$ and that $\boldsymbol{M}_{11}$ has independent circularly symmetric Gaussian entries; (49) is due to the characteristic polynomial of the matrix $-\boldsymbol{D}^{-1}\boldsymbol{M}_{11}$; in (50) we define $\phi(J_i)$ the argument of J_i that is independent of θ; (51) is from the fact that $\text{mod}(\theta + \phi)_{2\pi} \sim \text{Unif}(0, 2\pi]$ and is independent of ϕ, as long as $\theta \sim \text{Unif}(0, 2\pi]$ and is independent of ϕ, also known as the Crypto Lemma [27]; (52) is from the identity $\int_0^1 \log(a + b\cos(2\pi t))\,\mathrm{d}t = \log\frac{a+\sqrt{a^2-b^2}}{2} \geq \log(a) - 1, \forall\, a \geq b > 0$. Combining (46) and (53), we have the lower bound

$$
\mathbb{E}_{\tilde{G}}\left[\log\det\left(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G}\right)\right] \geq \log\left|\det\left(\boldsymbol{D}_1\right)\right|^2 + O(1) \tag{54}
$$

when snr is large. In fact, it has been shown that the $O(1)$ term here, sum of the $O(1)$ term in (46) and $-\tau'$ in (53), does not depend on snr at all.

The next step is to derive an upper bound on $\mathbb{E}\left[\log\det\left(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G}\right)\right]$. Following Jensen's inequality, we have

$$
\begin{aligned}
&\mathbb{E}_{\tilde{G}}\left[\log\det(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G})\right] \\
&\leq \log\det\left(\mathbb{E}_{\tilde{G}}[\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G}]\right) && (55)\\
&= \log\det\left(\begin{bmatrix} \boldsymbol{D}_1^2 & \\ & \boldsymbol{D}_2^2 \end{bmatrix} + \mathbb{E}[\boldsymbol{M}^{\mathsf{H}}\boldsymbol{M}]\right) && (56)\\
&= \log|\det\left(\boldsymbol{D}_1\right)|^2 + \underbrace{\log\det\left(I + m\boldsymbol{D}_1^{-2}\right)}_{o(1)} + \underbrace{\log\det\left(mI + \boldsymbol{D}_2^2\right)}_{o(\log\mathsf{snr})} && (57)
\end{aligned}
$$

$$= \log|\det(\boldsymbol{D}_1)|^2 + o(\log \mathsf{snr}) \tag{58}$$

Putting the lower and upper bounds together, we have

$$\mathbb{E}\left[\log\det(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G})\right] = \log|\det(\boldsymbol{D}_1)|^2 + o(\log \mathsf{snr}).$$

Finally, note that, since $\lambda_i(\hat{\boldsymbol{G}}^{\mathsf{H}}\hat{\boldsymbol{G}}) \doteq \mathsf{snr}^0$, $i = \tau' + 1, \ldots, \tau$, we have

$$\log\left|\det(\boldsymbol{D}_1)\right|^2 = \sum_{i=1}^{\tau'} \log\left(\lambda_i(\hat{\boldsymbol{G}}^{\mathsf{H}}\hat{\boldsymbol{G}})\right) \tag{59}$$

$$= \sum_{i=1}^{\tau} \log\left(\lambda_i(\hat{\boldsymbol{G}}^{\mathsf{H}}\hat{\boldsymbol{G}})\right) - \sum_{i=\tau'+1}^{\tau} \log\left(\lambda_i(\hat{\boldsymbol{G}}^{\mathsf{H}}\hat{\boldsymbol{G}})\right) \tag{60}$$

$$= \sum_{i=1}^{\tau} \log\left(\lambda_i(\hat{\boldsymbol{G}}^{\mathsf{H}}\hat{\boldsymbol{G}})\right) + o(\log \mathsf{snr}) \tag{61}$$

from which the proof is complete.

6.2 Proof of Lemma 2

The existence is proved by construction. Let $\boldsymbol{a}_j$, $j = 1, \ldots, m$, be the j th column of $\boldsymbol{A}$. We define j_1^* as the index of the column that has the largest Euclidean norm, i.e.,

$$j_1^* = \arg\max_{j=1,\ldots,m} \|\boldsymbol{a}_j\|. \tag{62}$$

Swapping the j_1^* and the first column, and denoting $\boldsymbol{A}_1 = \boldsymbol{A}$, we have

$$\boldsymbol{B}_1 \triangleq \boldsymbol{A}_1 \boldsymbol{T}_{1,j_1^*} \tag{63}$$

where $\boldsymbol{T}_{ij} \in \mathbb{C}^{m\times m}$ denotes the permutation matrix that swaps the i th and j th columns. Now, let $\boldsymbol{U}_1 \in \mathbb{C}^{m\times m}$ be any unitary matrix such that the first column is aligned with the first column of $\boldsymbol{B}_1$, i.e., equal to $\frac{\boldsymbol{a}_{j_1^*}}{\|\boldsymbol{a}_{j_1^*}\|}$. Then, we can construct a block-upper-triangular matrix $\boldsymbol{R}_1 = \boldsymbol{U}_1^{\mathsf{H}}\boldsymbol{B}_1 = \boldsymbol{U}_1^{\mathsf{H}}\boldsymbol{A}_1\boldsymbol{T}_{1,j_1^*}$ with the following form

$$\boldsymbol{R}_1 = \begin{bmatrix} r_{11} & * \\ \boldsymbol{0}_{(m-1)\times 1} & \boldsymbol{A}_2 \end{bmatrix} \tag{64}$$

where it is readily shown that

$$r_{11}^2 = \|\boldsymbol{a}_{j_1^*}\|^2 \tag{65}$$

$$\geq \frac{1}{m}||\boldsymbol{A}_1||_F^2 \tag{66}$$

$$\geq \frac{1}{m}\lambda_1(\boldsymbol{A}_1^{\mathsf{H}}\boldsymbol{A}_1). \tag{67}$$

Repeating the same procedure on $\boldsymbol{A}_2$, we will have $\boldsymbol{R}_2 = \boldsymbol{U}_2^{\mathsf{H}}\boldsymbol{B}_2 = \boldsymbol{U}_2^{\mathsf{H}}\boldsymbol{A}_2\boldsymbol{T}_{2,j_2^*}$ where all the involved matrices are similarly defined as above except for the reduced dimension $(m-1)\times(m-1)$ and

$$\boldsymbol{R}_2 = \begin{bmatrix} r_{22} & * \\ \boldsymbol{0}_{(m-2)\times 1} & \boldsymbol{A}_3 \end{bmatrix} \tag{68}$$

where it is readily shown that

$$r_{22}^2 \geq \frac{1}{m-1}\lambda_1(\boldsymbol{A}_2^{\mathsf{H}}\boldsymbol{A}_2) \tag{69}$$

$$\geq \frac{1}{m-1}\lambda_2(\boldsymbol{A}_1^{\mathsf{H}}\boldsymbol{A}_1). \tag{70}$$

Here, the last inequality is from the fact that, for any matrix $\boldsymbol{C}$ and a submatrix $\boldsymbol{C}_k$ by removing k rows or columns, we have [28, Corollary 3.1.3]

$$\lambda_i(\boldsymbol{C}_k^{\mathsf{H}}\boldsymbol{C}_k) \geq \lambda_{i+k}(\boldsymbol{C}^{\mathsf{H}}\boldsymbol{C}) \tag{71}$$

where we recall that λ_i is the i th largest eigenvalue. Let us continue the procedure on $\boldsymbol{A}_3$ and so on. At the end, we will have all the $\{\boldsymbol{U}_i\}$ and $\left\{\boldsymbol{T}_{i,j_i^*}\right\}$ such that

$$\underbrace{\begin{bmatrix} I_{m-1} & \\ & \boldsymbol{U}_m^{\mathsf{H}} \end{bmatrix} \cdots \begin{bmatrix} I_2 & \\ & \boldsymbol{U}_3^{\mathsf{H}} \end{bmatrix} \begin{bmatrix} 1 & \\ & \boldsymbol{U}_2^{\mathsf{H}} \end{bmatrix} \boldsymbol{U}_1^{\mathsf{H}}}_{\boldsymbol{Q}^{\mathsf{H}}} \boldsymbol{A} \underbrace{\boldsymbol{T}_{1,j_1^*} \begin{bmatrix} 1 & \\ & \boldsymbol{T}_{2,j_2^*} \end{bmatrix} \begin{bmatrix} I_2 & \\ & \boldsymbol{T}_{3,j_3^*} \end{bmatrix} \cdots \begin{bmatrix} I_{m-1} & \\ & \boldsymbol{T}_{m,j_m^*} \end{bmatrix}}_{\boldsymbol{P}}$$

$$= \underbrace{\begin{bmatrix} r_{11} & * & * & * \\ & r_{22} & * & * \\ & & \ddots & \vdots \\ & & & r_{mm} \end{bmatrix}}_{\boldsymbol{R}} \tag{72}$$

where it is obvious that $\boldsymbol{P}$ is a permutation matrix and $\boldsymbol{Q}$ is unitary. The proof is thus completed.

6.3 Proof of Lemma 3

Let $\bar{\boldsymbol{A}} \triangleq \boldsymbol{A}\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{R}$ with $\boldsymbol{P}$ a permutation matrix such that (40) holds. Then, we have

$$\det(\bar{\boldsymbol{A}}_{\mathcal{I}}^{\mathsf{H}}\bar{\boldsymbol{A}}_{\mathcal{I}}) = \det(\boldsymbol{R}_{\mathcal{I}}^{\mathsf{H}}\boldsymbol{Q}^{\mathsf{H}}\boldsymbol{Q}\boldsymbol{R}_{\mathcal{I}}) \tag{73}$$

$$= \det(\boldsymbol{R}_{\mathcal{I}}^{\mathsf{H}}\boldsymbol{R}_{\mathcal{I}}) \tag{74}$$

$$\geq \det(\boldsymbol{R}_{\mathcal{I}\mathcal{I}}^{\mathsf{H}}\boldsymbol{R}_{\mathcal{I}\mathcal{I}}) \tag{75}$$

$$= \prod_{i\in\mathcal{I}} r_{ii}^2 \tag{76}$$

$$\geq m^{-|\mathcal{I}|}\prod_{i\in\mathcal{I}}\lambda_i(\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}) \tag{77}$$

where the first inequality results from the Cauchy-Binet formula, and the last inequality is due to Lemma 2.

6.4 Proof of Proposition 4

The inequality (38) is trivial when $m \geq l \geq M$, i.e., $l' = m' = M$. From the chain rule $h(\boldsymbol{y}_m \,|\, U, \hat{H}, \tilde{H}) = h(\boldsymbol{y}_l \,|\, U, \hat{H}, \tilde{H}) + h(y_{l+1}, \ldots, y_m \,|\, \boldsymbol{y}_l, \hat{H}, \tilde{H}) = h(\boldsymbol{y}_l \,|\, U, \hat{H}, \tilde{H}) + o(\log \mathsf{snr})$, since with $l \geq M$, the observations $y_{l+1}, \ldots, y_m$ can be represented as a linear combination of $\boldsymbol{y}_l$, up to the noise error. In the following, we focus on the case $l \leq M$.

We are now ready to prove the proposition. First of all, let us write

$$
\begin{aligned}
& h(\boldsymbol{y}_m|U, \hat{H}, \tilde{H}) - \mu\, h(\boldsymbol{y}_l|U, \hat{H}, \tilde{H}) \\
&= \mathbb{E}_{\hat{H}}\Big[\mathbb{E}_{\tilde{H}}[h(\boldsymbol{H}_m\boldsymbol{x} + \boldsymbol{z}_m \,|\, U, \hat{H} = \hat{\boldsymbol{H}}, \tilde{H} = \tilde{\boldsymbol{H}})] \\
&\quad - \mu\, \mathbb{E}_{\tilde{H}}[h(\boldsymbol{H}_l\boldsymbol{x} + \boldsymbol{z}_l \,|\, U, \hat{H} = \hat{\boldsymbol{H}}, \tilde{H} = \tilde{\boldsymbol{H}})]\Big]
\end{aligned} \tag{78}
$$

In the following, we focus on the term inside the expection over $\hat{H}$ in (78), i.e., for a given realization of $\hat{\boldsymbol{H}}$. Since $\boldsymbol{y}_l$ is a degraded version of $\boldsymbol{y}_m$, we can apply the results in [29, Corollary 4] and obtain the optimality of Gaussian input, i.e.,

$$
\begin{aligned}
&\max_{\substack{p_{X|U\hat{H}}:\\ \mathbb{E}[\mathrm{tr}(XX^{\mathsf{H}})] \leq \mathsf{snr}}} \mathbb{E}_{\tilde{H}}\big[h(\boldsymbol{y}_m|U, \hat{H} = \hat{\boldsymbol{H}}, \tilde{H} = \tilde{\boldsymbol{H}})\big] - \mu\, \mathbb{E}_{\tilde{H}}\big[h(\boldsymbol{y}_l|U, \hat{H} = \hat{\boldsymbol{H}}, \tilde{H} = \tilde{\boldsymbol{H}})\big] \\
&= \max_{\boldsymbol{\Psi} \succeq 0 : \mathrm{tr}(\boldsymbol{\Psi}) \leq \mathsf{snr}} \mathbb{E}_{\tilde{H}}\big[\log\det\left(I + \boldsymbol{H}_m \boldsymbol{\Psi} \boldsymbol{H}_m^{\mathsf{H}}\right)\big] - \mu\, \mathbb{E}_{\tilde{H}}\big[\log\det\left(I + \boldsymbol{H}_l \boldsymbol{\Psi} \boldsymbol{H}_l^{\mathsf{H}}\right)\big]
\end{aligned} \tag{79}
$$

for any $\mu \geq 1$. The next step is to upper bound the right hand side (RHS) of (79).

Next, let $\boldsymbol{\Psi} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathsf{H}}$ be the eigenvalue decomposition of the covariance matrix $\boldsymbol{\Psi}$ where $\boldsymbol{\Lambda}$ is a diagonal matrix and $\boldsymbol{V}$ is unitary. Note that it is without loss of generality to assume that all eigenvalues of $\boldsymbol{\Psi}$ are strictly positive, i.e., $\lambda_i(\boldsymbol{\Psi}) \geq c > 0$, $\forall i$, in the sense that

$$
\begin{aligned}
\log\det\left(I + \boldsymbol{H}\boldsymbol{\Psi}\boldsymbol{H}^{\mathsf{H}}\right) &\leq \log\det\left(I + \boldsymbol{H}(cI + \boldsymbol{\Psi})\boldsymbol{H}^{\mathsf{H}}\right) \\
&\leq \log\det\left(I + \boldsymbol{H}\boldsymbol{\Psi}\boldsymbol{H}^{\mathsf{H}}\right) + \log\det\left(I + c\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}}\right).
\end{aligned} \tag{80}
$$

In other words, a constant lift of the eigenvalues of $\boldsymbol{\Psi}$ does not have any impact on the high SNR behavior. This regularization will however simplify the analysis. The following is an upper bound for the first term in the RHS of (79).

$$
\begin{aligned}
& \mathbb{E}_{\tilde{H}}\Big[\log\det\big(I + \boldsymbol{H}_m\boldsymbol{\Psi}\boldsymbol{H}_m^{\mathsf{H}}\big)\Big] \\
&= \mathbb{E}_{\tilde{H}}\Big[\log\det\Big(I_M + \boldsymbol{\Psi}^{\frac{1}{2}}\boldsymbol{H}_m^{\mathsf{H}}\boldsymbol{H}_m\boldsymbol{\Psi}^{\frac{1}{2}}\Big)\Big] && (81) \\
&\leq \mathbb{E}_{\tilde{H}}\left[\log\det\left(I_M + \boldsymbol{\Psi}^{\frac{1}{2}}\boldsymbol{U}^H \begin{bmatrix} \|\boldsymbol{H}_m\|^2 I_{m'} & \\ & 0 \end{bmatrix} \boldsymbol{U}\boldsymbol{\Psi}^{\frac{1}{2}}\right)\right] && (82) \\
&= \mathbb{E}_{\tilde{H}}\Big[\log\det\big(I_{m'} + \|\boldsymbol{H}_m\|^2\,\tilde{\boldsymbol{\Psi}}\big)\Big] && (83) \\
&= \mathbb{E}_{\tilde{H}}\big[\log\det\big(\tilde{\boldsymbol{\Psi}}\big)\big] + \mathbb{E}_{\tilde{H}}\Big[\log\det\big(\tilde{\boldsymbol{\Psi}}^{-1} + \|\boldsymbol{H}_m\|^2 I\big)\Big] && (84)
\end{aligned}
$$

$$\leq \sum_{i=1}^{m'} \log \lambda_i(\boldsymbol{\Psi}) + \underbrace{\log\det\big((c^{-1} + m + \|\hat{\boldsymbol{H}}_m\|^2) I\big)}_{o(\log \mathsf{snr})} \tag{85}$$

$$\leq \log\det(\boldsymbol{\Lambda}) + o(\log \mathsf{snr}) \tag{86}$$

where $\boldsymbol{\Psi}^{\frac{1}{2}}$ is such that $\big(\boldsymbol{\Psi}^{\frac{1}{2}}\big)^2 = \boldsymbol{\Psi}$; (82) is due to fact that

$$\boldsymbol{H}_m^H \boldsymbol{H}_m \preceq \boldsymbol{U}^H \begin{bmatrix} \|\boldsymbol{H}_m\|^2 I_{m'} & \\ & 0 \end{bmatrix} \boldsymbol{U}$$

with $\boldsymbol{U}$ being the matrix of eigenvectors of $\boldsymbol{H}_m^H \boldsymbol{H}_m$ and $\|\boldsymbol{H}_m\|$ being the Frobenius norm of $\boldsymbol{H}_m$; in (83), we define $\tilde{\boldsymbol{\Psi}}$ as the $m' \times m'$ upper left block of $\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{U}^{\mathsf{H}}$; the first term in (85) is due to $\det(\tilde{\boldsymbol{\Psi}}) = \prod_{i=1}^{m'} \lambda_i(\tilde{\boldsymbol{\Psi}}) \leq \prod_{i=1}^{m'} \lambda_i(\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{U}^H) = \prod_{i=1}^{m'} \lambda_i(\boldsymbol{\Psi})$; the second term in (85) is from Jensen's inequality and using the fact that $\boldsymbol{\Psi}^{-1} \preceq c^{-1} I_M$ by assumption and that $\mathbb{E}_{\tilde{H}}(\boldsymbol{H}_m^H \boldsymbol{H}_m) = \sum_{k=1}^m \sigma_k^2 I_M + \hat{\boldsymbol{H}}_m^H \hat{\boldsymbol{H}}_m \preceq mI + \hat{\boldsymbol{H}}_m^H \hat{\boldsymbol{H}}_m$; the last inequality is from the assumption that every eigenvalue of $\boldsymbol{\Psi}$ is lower-bounded by some constant $c > 0$ independent of snr. Now, we need to lower bound the second expectation in the RHS of (79). To this end, let us write

$$\det(I_l + \boldsymbol{H}_l \boldsymbol{\Psi} \boldsymbol{H}_l^{\mathsf{H}}) = \det(I_l + \boldsymbol{H}_l \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}^{\mathsf{H}} \boldsymbol{H}_l^{\mathsf{H}}) \tag{87}$$

$$= \det(I_M + \boldsymbol{\Lambda} \boldsymbol{V}^{\mathsf{H}} \boldsymbol{H}_l^{\mathsf{H}} \boldsymbol{H}_l \boldsymbol{V}) \tag{88}$$

$$= \det\big(I_M + \boldsymbol{\Lambda} \boldsymbol{\Phi}^{\mathsf{H}} \boldsymbol{\Sigma}^2 \boldsymbol{\Phi}\big) \tag{89}$$

$$= 1 + \sum_{\substack{\mathcal{I} \subseteq \{1,\dots,M\} \\ \mathcal{I} \neq \emptyset}} \det(\boldsymbol{\Lambda}_{\mathcal{I}\mathcal{I}}) \det(\boldsymbol{\Phi}_{\mathcal{I}}^{\mathsf{H}} \boldsymbol{\Sigma}^2 \boldsymbol{\Phi}_{\mathcal{I}}) \tag{90}$$

$$\geq \det(\boldsymbol{\Sigma}^2) \sum_{j=1}^{M} \det(\boldsymbol{\Lambda}_{\mathcal{I}_j\mathcal{I}_j}) \det(\boldsymbol{\Phi}_{\mathcal{I}_j}^{\mathsf{H}} \boldsymbol{\Phi}_{\mathcal{I}_j}) \tag{91}$$

$$\geq M \det(\boldsymbol{\Sigma}^2) \left(\prod_{j=1}^{M} \Big(\det(\boldsymbol{\Lambda}_{\mathcal{I}_j\mathcal{I}_j}) \det(\boldsymbol{\Phi}_{\mathcal{I}_j}^{\mathsf{H}} \boldsymbol{\Phi}_{\mathcal{I}_j}) \Big) \right)^{\frac{1}{M}} \tag{92}$$

$$= M \det(\boldsymbol{\Sigma}^2) \det(\boldsymbol{\Lambda})^{\frac{l}{M}} \left(\prod_{j=1}^{M} \det(\boldsymbol{\Phi}_{\mathcal{I}_j}^{\mathsf{H}} \boldsymbol{\Phi}_{\mathcal{I}_j}) \right)^{\frac{1}{M}} \tag{93}$$

where (88) is an application of the identity $\det(I + \boldsymbol{A}\boldsymbol{B}) = \det(I + \boldsymbol{B}\boldsymbol{A})$; in (89), we define

$$\boldsymbol{\Sigma} \triangleq \mathrm{diag}(\sigma_1, \dots, \sigma_l), \; \boldsymbol{\Phi} \triangleq \boldsymbol{\Sigma}^{-1} \boldsymbol{H}_l \boldsymbol{V}, \; \text{and } \hat{\boldsymbol{\Phi}} \triangleq \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{H}}_l \boldsymbol{V};$$

in (90), we define $\boldsymbol{\Phi}_{\mathcal{I}} \triangleq [\Phi_{ji} : \; j = 1, \dots, l, \; i \in \mathcal{I}] \in \mathbb{C}^{l \times |\mathcal{I}|}$ as the submatrix of $\boldsymbol{\Phi}$ with columns indexed in $\mathcal{I}$ and $\boldsymbol{\Lambda}_{\mathcal{I}\mathcal{I}} = [\Lambda_{ji} : \; i, j \in \mathcal{I}] \in \mathbb{C}^{|\mathcal{I}| \times |\mathcal{I}|}$, with $\mathcal{I}$

denoting a nonempty set; the equality (90) is an application of the identity [30]

$$\det(I+\boldsymbol{A}) = 1 + \sum_{\substack{\mathcal{I}\subseteq\{1,\ldots,M\}\\ \mathcal{I}\neq\emptyset}} \det(\boldsymbol{A}_{\mathcal{I}\mathcal{I}})$$

for any $\boldsymbol{A}\in\mathbb{C}^{M\times M}$; in (91), we define $\mathcal{I}_1,\ldots,\mathcal{I}_M$ as the so-called sliding window of indices

$$\mathcal{I}_1 \triangleq \{1,2,\cdots,l\},\ \mathcal{I}_2 \triangleq \{2,3,\cdots,l,l+1\},\ \cdots,\ \mathcal{I}_M \triangleq \{M,1,2,\cdots,l-1\} \tag{94}$$

$$\text{i.e., } \mathcal{I}_j \triangleq \{\mathrm{mod}(j+i-1)_M + 1:\ i=0,1,\cdots,l-1\},\ j=1,2,\cdots,M \tag{95}$$

with $\mathrm{mod}(x)_M$ being the modulo operator; (92) is from the fact that arithmetic mean is not smaller than geometric mean; in (93), we use the fact that

$$\prod_{j=1}^{M} \det(\boldsymbol{\Lambda}_{\mathcal{I}_j\mathcal{I}_j}) = \det(\boldsymbol{\Lambda})^l.$$

Without loss of generality, we assume that the M columns of $\boldsymbol{H}_l\boldsymbol{V}$ are ordered in such a way that 1) the first l columns are linearly independent, i.e., $\hat{\boldsymbol{\Phi}}_{\mathcal{I}_1}$ has full rank, and 2) $\boldsymbol{A} = \hat{\boldsymbol{\Phi}}_{\mathcal{I}_1}$ satisfies Lemma 3. Note that the former condition can almost always be satisfied since $\mathrm{rank}(\hat{\boldsymbol{\Phi}}) = l$ almost surely. Hence, we have

$$\begin{aligned}
\mathbb{E}_{\tilde{H}}\left[\log\det(\boldsymbol{\Phi}_{\mathcal{I}_j}^{\mathsf{H}}\boldsymbol{\Phi}_{\mathcal{I}_j})\right] &= \sum_{i=1}^{\mathrm{rank}(\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j})} \log\big(\lambda_i(\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j}^{\mathsf{H}}\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j})\big) + o(\log\mathsf{snr}) && (96)\\
&\geq \sum_{i=1}^{\mathrm{rank}(\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j\bigcap\mathcal{I}_1})} \log\big(\lambda_i(\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j}^{\mathsf{H}}\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j})\big) + o(\log\mathsf{snr}) && (97)\\
&\geq \sum_{i=1}^{\mathrm{rank}(\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j\bigcap\mathcal{I}_1})} \log\big(\lambda_i(\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j\bigcap\mathcal{I}_1}^{\mathsf{H}}\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j\bigcap\mathcal{I}_1})\big) + o(\log\mathsf{snr}) && (98)\\
&= \log\det(\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j\bigcap\mathcal{I}_1}^{\mathsf{H}}\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j\bigcap\mathcal{I}_1}) + o(\log\mathsf{snr}) && (99)\\
&\geq \log\prod_{i\in\mathcal{I}_j\bigcap\mathcal{I}_1} \lambda_i(\hat{\boldsymbol{\Phi}}^{\mathsf{H}}\hat{\boldsymbol{\Phi}}) + o(\log\mathsf{snr}) && (100)
\end{aligned}$$

where (96) is from Lemma 1 by noticing that $\boldsymbol{\Phi}_{\mathcal{I}_j} = \hat{\boldsymbol{\Phi}}_{\mathcal{I}_j} + \tilde{\boldsymbol{\Phi}}_{\mathcal{I}_j}$ with the entries of $\tilde{\boldsymbol{\Phi}}_{\mathcal{I}_j} \triangleq \boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{H}}_l\boldsymbol{V}$ being i.i.d. $\mathcal{N}_c(0,1)$; (97) is from the fact that $\mathrm{rank}(\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j}) \geq \mathrm{rank}(\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j\bigcap\mathcal{I}_1})$; (98) is due to $\lambda_i(\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j}^{\mathsf{H}}\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j}) \geq \lambda_i(\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j\bigcap\mathcal{I}_1}^{\mathsf{H}}\hat{\boldsymbol{\Phi}}_{\mathcal{I}_j\bigcap\mathcal{I}_1})$ where we recall that $\lambda_i(\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A})$ is defined as the i th largest eigenvalue of $\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}$; and the last inequality is due to Lemma 3. Summing over all j, we have

$$\sum_{j=1}^{M} \mathbb{E}_{\tilde{H}}\left[\log\det(\mathbf{\Phi}_{\mathcal{I}_j}^{\mathsf{H}}\mathbf{\Phi}_{\mathcal{I}_j})\right] \geq \log\left(\prod_{j=1}^{M}\prod_{i\in\mathcal{I}_j\bigcap\mathcal{I}_1}\lambda_i(\hat{\mathbf{\Phi}}^{\mathsf{H}}\hat{\mathbf{\Phi}})\right) + o(\log\mathsf{snr}) \tag{101}$$

$$= \log\left(\left(\prod_{i\in\mathcal{I}_1}\lambda_i(\hat{\mathbf{\Phi}}^{\mathsf{H}}\hat{\mathbf{\Phi}})\right)^{l}\right) + o(\log\mathsf{snr}) \tag{102}$$

$$\geq l\log\prod_{i\in\mathcal{I}_1}\lambda_i(\hat{\mathbf{\Phi}}_{\mathcal{I}_1}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathcal{I}_1}) + o(\log\mathsf{snr}) \tag{103}$$

$$= l\log\det\left(\hat{\mathbf{\Phi}}_{\mathcal{I}_1}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathcal{I}_1}\right) + o(\log\mathsf{snr}) \tag{104}$$

$$= -l\log\det\left(\mathbf{\Sigma}^2\right) + o(\log\mathsf{snr}) \tag{105}$$

where (103) is due to $\lambda_i(\hat{\mathbf{\Phi}}^{\mathsf{H}}\hat{\mathbf{\Phi}}) \geq \lambda_i(\hat{\mathbf{\Phi}}_{\mathcal{I}_1}^{\mathsf{H}}\hat{\mathbf{\Phi}}_{\mathcal{I}_1})$, $\forall\, i = 1,\dots,l$; the last equality is from the fact that $\hat{\mathbf{\Phi}}_{\mathcal{I}_1} = \mathbf{\Sigma}^{-1}\hat{\boldsymbol{H}}_l\boldsymbol{V}_{\mathcal{I}_1}$ and that $\hat{\boldsymbol{H}}_l\boldsymbol{V}_{\mathcal{I}_1}$ has full rank by construction. From (93) and (105), we obtain

$$\mathbb{E}_{\tilde{H}}\left[\log\det\left(I_l + \boldsymbol{H}_l\mathbf{\Psi}\boldsymbol{H}_l^{\mathsf{H}}\right)\right] \geq \frac{l}{M}\log\det(\mathbf{\Lambda}) + \frac{M-l}{M}\log\det(\mathbf{\Sigma}^2) + o(\log\mathsf{snr}) \tag{106}$$

and finally

$$\begin{aligned}&\mathbb{E}_{\tilde{H}}\left[\log\det\left(I_m + \boldsymbol{H}_m\mathbf{\Psi}\boldsymbol{H}_m^{\mathsf{H}}\right)\right] - \frac{M}{l}\mathbb{E}_{\tilde{H}}\left[\log\det\left(I_l + \boldsymbol{H}_l\mathbf{\Psi}\boldsymbol{H}_l^{\mathsf{H}}\right)\right]\\ &\leq -\frac{M-l}{l}\log\det(\mathbf{\Sigma}^2) + o(\log\mathsf{snr}).\end{aligned} \tag{107}$$

When $m < M$, the above bound (107) is not tight. However, we can show that, in this case, (107) still holds when we replace M with m. To see this, let us define $\mathbf{\Lambda}' \triangleq \mathrm{diag}(\lambda_1,\dots,\lambda_m)$. First, note that when $m < M$, (86) holds if we replace $\mathbf{\Lambda}$ with $\mathbf{\Lambda}'$ on the RHS. Then, the RHS of (87) becomes a lower bound if we replace $\mathbf{\Lambda}$ with $\mathbf{\Lambda}'$ and $\boldsymbol{V}$ with $\boldsymbol{V}' \in \mathbb{C}^{M\times m}$, the first m columns of $\boldsymbol{V}$. From then on, every step holds with M replaced by m. (107) thus follows with M replaced by m. By taking the expectation on both sides of (107) over $\hat{\boldsymbol{H}}$ and plugging it into (78), we complete the proof of (38).

References

[1] G. Caire and S. Shamai, “On the achievable throughput of a multiantenna Gaussian broadcast channel,” *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691 – 1706, Jul. 2003.

[2] S. Jafar and A. Goldsmith, “Isotropic fading vector broadcast channels: The scalar upper bound and loss in degrees of freedom,” *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 848 – 857, Mar. 2005.

[3] C. Huang, S. A. Jafar, S. Shamai, and S. Vishwanath, “On degrees of freedom region of MIMO networks without channel state information at transmitters,” *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 849–857, Feb. 2012.

[4] M. A. Maddah-Ali and D. N. C. Tse, “Completely stale transmitter channel state information is still very useful,” *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4418 – 4431, Jul. 2012.

[5] N. Lee and R. W. Heath Jr., “Not too delayed CSIT achieves the optimal degrees of freedom,” in *Proc. Allerton Conf. Communication, Control and Computing*, Oct. 2012, available on arXiv:1207.2211.

[6] R. Tandon, S. A. Jafar, and S. Shamai, “Minimum CSIT to achieve maximum degrees of freedom for the MISO BC,” Nov. 2012, available on arXiv:1211.4254v2.

[7] M. Kobayashi, S. Yang, D. Gesbert, and X. Yi, “On the degrees of freedom of time correlated MISO broadcast channel with delayed CSIT,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jul. 2012.

[8] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, “Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT,” *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 315–328, Jan. 2013.

[9] T. Gou and S. Jafar, “Optimal use of current and outdated channel state information: Degrees of freedom of the MISO BC with mixed CSIT,” *IEEE Communications Letters*, vol. 16, no. 7, pp. 1084 – 1087, Jul. 2012.

[10] J. Chen and P. Elia, “Degrees-of-freedom region of the MISO broadcast channel with general mixed-CSIT,” May 2012, available on arXiv:1205.3474v1.

[11] ——, “Can imperfect delayed CSIT be as useful as perfect delayed CSIT? DoF analysis and constructions for the BC,” in *Proc. Allerton Conf. Communication, Control and Computing*, Oct. 2012.

[12] ——, “Imperfect delayed CSIT can be as useful as perfect delayed CSIT: DoF and precoding schemes for BC,” Oct. 2012, submitted to *IEEE Trans. Inform. Theory*.

[13] R. Tandon, S. A. Jafar, S. Shamai, and H. V. Poor, “On the synergistic benefits of alternating CSIT for the MISO BC,” Aug. 2012, submitted to *IEEE Trans. Inform. Theory*, available on arXiv:1208.5071.

[14] J. Chen and P. Elia, “MISO broadcast channel with delayed and evolving CSIT,” Nov. 2012, submitted to *IEEE Trans. Inform. Theory*, available on arXiv:1211.1622.

[15] C. S. Vaze and M. K. Varanasi, “The degrees of freedom region of two-user and certain three-user MIMO broadcast channel with delayed CSI,” Dec. 2011, submitted to *IEEE Trans. Inf. Theory*, available on arXiv:1101.0306.

[16] A. Ghasemi, A. S. Motahari, and A. K. Khandani, “On the degrees of freedom of X channel with delayed CSIT,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jul. 2011.

[17] M. J. Abdoli, A. Ghasemi, and A. K. Khandani, “On the degrees of freedom of three-user MIMO broadcast channel with delayed CSIT,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jul. 2011.

[18] A. Ghasemi, A. S. Motahari, and A. K. Khandani, “Interference alignment for the MIMO interference channel with delayed local CSIT,” Feb. 2011, available on arXiv:1102.5673v1.

[19] J. Xu, J. G. Andrews, and S. A. Jafar, “Broadcast channels with delayed finite-rate feedback: Predict or observe?” *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1456 – 1467, Apr. 2012.

[20] Y. Lejosne, D. Slock, and Y. Yuan-Wu, “Degrees of freedom in the MISO BC with delayed-CSIT and finite coherence time: A simple optimal scheme,” in *Proc. IEEE Int. Conf. on Signal Processing, Communications and Control (ICSPCC)*, Aug. 2012.

[21] R. Tandon, M. A. Maddah-Ali, A. Tulino, H. V. Poor, and S. Shamai, “On fading broadcast channels with partial channel state information at the transmitter,” in *Proc. Int. Symp. on Wireless Communication Systems (ISWCS)*, Aug. 2012.

[22] 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, Jun. 2010.

[23] H. Maleki, S. Jafar, and S. Shamai, “Retrospective interference alignment over interference networks,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 3, pp. 228 – 240, Mar. 2012.

[24] X. Yi, S. Yang, D. Gesbert, and M. Kobayashi, “The degrees of freedom region of temporally-correlated MIMO networks with delayed CSIT,” Nov. 2012, submitted to *IEEE Trans. Inform. Theory*, available on arXiv:1211.3322.

[25] Y. H. Kim and A. Lapidoth, “On the log determinant of noncentral wishart matrices,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jul. 2003.

[26] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.

[27] G. D. Forney Jr., "On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener," in *Proc. Allerton Conf. Communication, Control and Computing*, Oct. 2003.

[28] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[29] H. Weingarten, T. Liu, S. Shamai, Y. Steinberg, and P. Viswanath, "The capacity region of the degraded multiple-input multiple-output compound broadcast channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5011 – 5023, Nov. 2009.

[30] A. C. Aitken, *Determinants and Matrices*, 8th ed. Edinburgh: Oliver and Boyd, 1954.