The Degrees of Freedom Region of Temporally-Correlated MIMO Networks with Delayed CSIT

Xinping Yi, *Student Member, IEEE*, Sheng Yang, *Member, IEEE*,
David Gesbert, *Fellow, IEEE*, Mari Kobayashi, *Member, IEEE*

***Abstract*—We consider the temporally-correlated Multiple-Input Multiple-Output (MIMO) broadcast channels (BC) and interference channels (IC) where the transmitter(s) has/have (i) *delayed* channel state information (CSI) obtained from a latency-prone feedback channel as well as (ii) imperfect *current* CSIT, obtained, e.g., from prediction on the basis of these past channel samples based on the temporal correlation. The degrees of freedom (DoF) regions for the two-user broadcast and interference MIMO networks with general antenna configuration under such conditions are fully characterized, as a function of the prediction quality indicator. Specifically, a simple unified framework is proposed, allowing to attain optimal DoF region for the general antenna configurations and current CSIT qualities. Such a framework builds upon *block-Markov encoding* with *interference quantization*, optimally combining the use of both outdated and instantaneous CSIT. A striking feature of our work is that, by varying the power allocation, *every* point in the DoF region can be achieved with one single scheme. As a result, instead of checking the achievability of every corner point of the outer bound region, as typically done in the literature, we propose a new systematic way to prove the achievability.**

***Index Terms*—MIMO, Broadcast Channels, Interference Channels, Degrees of Freedom, Delayed CSIT**

I. Introduction

While the capacity region of the Multiple-Input Multiple-Output (MIMO) broadcast channel (BC) was established in [1], the characterization of the capacity of Gaussian interference channel (IC) has been a long-standing open problem, even for the two-user single-antenna case. Recent progress sheds light on this problem from various perspectives, among which the authors in [2] characterized the degrees of freedom (DoF) region, specializing to the large signal-to-noise-ratio (SNR) regime, for the two-user MIMO IC. The number of DoF represents the slope with which the rate increases with the logarithm of SNR. Note that when taking additional system limitations into account such as imperfect hardware, finite modulation levels, and cost of channel training in a time-varying environment, the sum rate inevitably saturates in the very large SNR limit [3]. However, the DoF can be shown to be meaningful within a reasonable interval of practical SNRs for properly designed systems. Furthermore, it provides us with a first-order approximation from which novel transmission schemes and insights emerge. In most works, the DoF analysis for multiuser channels involves the full knowledge of channel state information (CSI) at both the transmitter and receiver sides. In practice, however, the acquisition of perfect CSI at the transmitters is difficult, if not impossible, especially for fast fading channels. The CSIT obtained via feedback suffers from delays, which renders the available CSIT feedback possibly fully obsolete (i.e., uncorrelated with the current true channel) under the fast fading channel and, seemingly non-exploitable in view of designing the spatial precoding.

Recently, this common accepted viewpoint in such scenario (referred to as "delayed CSIT") was challenged by an interesting information theoretic work [4], in which a novel scheme (termed here as "MAT alignment") was proposed for the MISO BC to demonstrate that even the completely outdated channel feedback is still useful. The precoders are designed achieving strictly better DoF than what is obtained without any CSIT. The essential ingredient for the proposed scheme in [4] lies in the use of a multi-slot protocol initiating with the transmission of unprecoded information symbols to the user terminals, followed by the *analog* forwarding of the interferences created in the previous time slots. Most recently, generalizations under the similar principle to the MIMO BC [5], MIMO IC [6] settings, the MIMO BC with secrecy constraints [7], among others, were also addressed, where the DoF regions are fully characterized with arbitrary antenna configurations, again establishing DoF strictly beyond the ones obtained without CSIT [8]–[10] but below the ones with perfect CSIT [1], [2]. Note that other recent interesting lines of work combining instantaneous and delayed forms of feedback were reported in [11], [12].

Albeit inspiring and fascinating from a conceptual point of view, these works made an assumption that the channel is independent and identically distributed (i.i.d.) across time, where the delayed CSIT bears no correlation with the current channel realization. Hence, these results pessimistically consider that no estimate for the *current* channel realization is producible to the transmitter. Owing to the finite Doppler spread behavior of fading channels, it is however the case in many real life situations that the past channel realizations can provide *some*

Manuscript received October 10, 2012; revised June 05, 2013; accepted September 30, 2013. This work has been performed in the framework of the European research projects HARP, SHARING and HIATUS (FET-Open grant number: 265578), as well as the French ANR project FIREFLIES (ANR-10-INTB-0302). A part of preliminary results of this work has been presented at IEEE Int. Symp. Inf. Theory 2013, Istanbul, Turkey.

X. Yi and D. Gesbert are with the Mobile Communications Dept., EURECOM, 06560 Sophia Antipolis, France (email: {xinping.yi, david.gesbert}@eurecom.fr).

S. Yang and M. Kobayashi are with the Telecommunications Dept., SUPELEC, 91190 Gif-sur-Yvette, France (e-mail: {sheng.yang, mari.kobayashi}@supelec.fr).

information about the current one. Therefore a scenario where the transmitter is endowed with delayed CSI in addition to some (albeit imperfect) estimate of the current channel is of practical relevance. Together with the delayed CSIT, the benefit of such imperfect current CSIT was first exploited in [13] for the MISO BC whereby a novel transmission scheme was proposed which improves over pure MAT alignment in constructing precoders based on delayed *and* current CSIT estimate. The full characterization of the optimal DoF for this hybrid CSIT was later reported in [14], [15] for MISO BC under this setting. The key idea behind the schemes (termed hereafter as "α-MAT alignment") in [13]–[15] lies in the modification of the MAT alignment such that i) the initial time slot involves transmission of *precoded* symbols, which enables to reduce the power of mutual interferences and efficiently compress them; ii) the subsequent slots perform a digital transmission of quantized residual interferences together with new private symbols. Most recently, this philosophy was extended to the MIMO networks (BC/IC) but only with symmetric antenna configurations [16], as well as the K-user MISO case [17]. The generalization to the MISO BC with different qualities of imperfect current CSIT was also studied in [18]. Remarkably, the authors of [18] showed that, in order to balance the asymmetry of the CSIT quality, an infinite number of time slots are required. As such, they extended the number of phases of the α-MAT alignment [14] to infinity and varied the length of each phase.

Unfortunately, extending the previous results to the MIMO case with arbitrary antenna configurations is not a trivial step, even with the symmetric current CSIT quality assumption. The main challenges are two-fold: (a) the extra spatial dimension at the receiver side introduces a non-trivial tradeoff between the useful signal and the mutual interference, and (b) the asymmetry of receive antenna configurations results in the discrepancy of common-message-decoding capability at different receivers. In particular, the total number of streams that can be delivered as common messages to both receivers is inevitably limited by the weaker one (i.e., with fewer antennas). Such a constraint prevents the system from achieving the optimal DoF of the symmetric case by simply extending the previous schemes developed in [16].

To counter these new challenges posed by the asymmetry of antenna configurations, we develop a new strategy that balances the discrepancy of common-message-decoding capability at two receivers. This allows us to fully characterize the DoF region of both MIMO BC and MIMO IC, achieved by a unified and simple scheme built upon *block-Markov encoding*. This encoding concept was first introduced in [19] for relay channels and then became a standard tool for communication problems involving interaction between nodes, such as feedback (e.g., [20], [21]) or user cooperation (e.g., [22]). It turns out that our problem with both delayed and instantaneous CSIT, closely related to [20], can also be solved with this scheme. As it will become clear later, in each block, the transmitter superimposes the common information about the interferences created in the past block (due to the imperfect instantaneous CSIT) on the new private information (thus creating new interferences). At the receiver side, backward decoding is employed, i.e., the decoding of each block relies on the common side information from the decoding of future blocks. Due to the repetitive nature in each block, the proposed scheme can be uniquely characterized with the parameters such as the power allocation and rate splitting of the superposition. Surprisingly enough, our block-Markov scheme can also include the asymmetry of current CSIT with a simple parameter change, and thus somehow balance the global asymmetry, i.e., antenna asymmetry and CSIT asymmetry, in the system.

Overall, our results allow to bridge between previously reported CSIT scenarios such as the pure delayed CSIT of [4], [5] and the pure instantaneous CSIT scenarios [1], [2] for the MIMO setting. We tackle both the BC and IC configurations as we point out the tight connection between the DoF achieving transmission strategies in both settings. More specifically, we obtain the following key results:

- We establish outer bounds on the DoF region for the two-user temporally-correlated MIMO BC and IC with perfect delayed and imperfect current CSIT, as a function of the current CSIT quality *exponent*. By introducing a virtual received signal for the IC, we nicely link the outer bound to that of the BC, arriving at the similar outer bound results for both cases. In addition to the genie-aided bounding techniques and the application of the extremal inequality in [14], we develop a set of upper and lower bounds of ergodic capacity for MIMO channels, which is essential for the MIMO case but not extendible from MISO.
- We propose a unified framework relying on block-Markov encoding uniquely parameterized by the rate splitting and power allocation, by which the optimal DoF regions confined by the outer bounds are achievable with perfect delayed plus imperfect current CSIT. For any antenna and current CSIT settings, every point in the outer bound region can be achieved with one single scheme. For instance, the MIMO BC with $M = 3$, $N_1 = 2$ and $N_2 = 1$ achieves optimal sum DoF $\frac{15+4\alpha_1+2\alpha_2}{7}$ when $3\alpha_1 - 2\alpha_2 \leq 1$ and $\frac{7+2\alpha_2}{3}$ otherwise, where α_1 and α_2 are imperfect current CSIT qualities for both users' channels. This smoothly connects three special cases: the case with pure delayed CSIT [5] ($\alpha_1 = \alpha_2 = 0$), that with perfect current CSIT [1] ($\alpha_1 = \alpha_2 = 1$), and that with perfect CSIT at Receiver 1 and delayed CSIT at Receiver 2 [24] ($\alpha_1 = 1, \alpha_2 = 0$).
- We propose a new systematic way to prove the achievability. In the proposed framework, the achievability region is defined by the decodability conditions in terms of the rate splitting and power allocation. The achievability is proved by mapping the outer bound region into a set of proper rate and power allocation and showing that this set lies within the decodability region. This contrasts with most existing proofs in the literature where the achievability of each corner point is checked.

It is worth noting that our results cover the previously reported particular cases: the perfect CSIT setting [1], [2] (i.e., current CSIT of perfect quality), the pure delayed CSIT setting [6] (i.e., current CSIT of zero quality), the partial/hybrid CSIT MIMO BC/IC case [24]–[26] (with perfect CSIT at one receiver and delayed CSIT at the other one), and the special MISO case [13]–[15] with $N_1 = N_2 = 1$, symmetric MIMO case [16], as well as the MISO case with asymmetric current CSIT qualities [18]. In a parallel work [23], a similar scheme was independently revealed, also built on the block-Markov encoding, evolving from the multi-phase scheme initially proposed in [18]. While they focus on the MISO BC in a more general evolving CSIT setting, our work deals with a wider class of channel configurations (both MIMO BC and IC) with static CSIT.

The rest of the paper is organized as follows. We present the system model and assumptions in the coming section, followed by the main results on DoF region characterization for both MIMO BC and MIMO IC cases in Section III. Some illustrative examples of the achievability schemes are provided in Section IV, followed by the general formulation in Section V. In Section VI, we present the proofs of outer bounds. Finally, we conclude the paper in Section VII.

Notation: Matrices and vectors are represented as uppercase and lowercase letters, respectively. Matrix transport, Hermitian transport, inverse, rank, determinant and the Frobenius norm of a matrix are denoted by $\boldsymbol{A}^{\mathsf{T}}$, $\boldsymbol{A}^{\mathsf{H}}$, $\boldsymbol{A}^{-1}$, $\mathrm{rank}(\boldsymbol{A})$, $\det(\boldsymbol{A})$ and $\|\boldsymbol{A}\|_{\mathrm{F}}$, respectively. $\boldsymbol{A}_{[k_1:k_2]}$ represents the submatrix of $\boldsymbol{A}$ from k_1-th row to k_2-th row when $k_1 \le k_2$. $\boldsymbol{h}^{\perp}$ is the normalized orthogonal component of any non-zero vector $\boldsymbol{h}$. We use $\mathbf{I}_M$ to denote an $M \times M$ identity matrix where the dimension is omitted whenever confusion is not probable. The approximation $f(P) \sim g(P)$ is in the sense of $\lim_{P\to\infty} \frac{f(P)}{g(P)} = C$, where $C > 0$ is a constant that does not scale as P. Partial ordering of Hermitian matrices is denoted by $\succeq$ and $\preceq$, i.e., $\boldsymbol{A} \preceq \boldsymbol{B}$ means $\boldsymbol{B}-\boldsymbol{A}$ is positive semidefinite. Logarithms are in base 2. $(x)^+$ means $\max\{x, 0\}$, and $\mathbb{R}_+^n$ represents the set of n-tuples of non-negative real numbers. $f = O(g)$ follows the standard Landau notation, i.e., $\lim \frac{f}{g} \le C$ where the limit depends on the context. With some abuse of notation, we use $O_X(g)$ to denote any f such that $\mathbb{E}_X(f) = O(\mathbb{E}_X(g))$. Finally, the range or null spaces mentioned in this paper refer to the column spaces.

II. System Model

A. Two-user MIMO Broadcast Channel

For a two-user (M, N_1, N_2) MIMO broadcast channel (BC) with M antennas at the transmitter and N_i antennas at Receiver i, the discrete time signal model is given by

$$\boldsymbol{y}_i(t) = \boldsymbol{H}_i(t)\boldsymbol{x}(t) + \boldsymbol{z}_i(t) \tag{1}$$

for any time instant t, where $\boldsymbol{H}_i(t) \in \mathbb{C}^{N_i \times M}$ is the channel matrix for Receiver i ($i = 1, 2$); $\boldsymbol{z}_i(t) \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_{N_i})$ is the normalized additive white Gaussian noise (AWGN) vector at Receiver i and is independent of channel matrices and transmitted signals; the coded input signal $\boldsymbol{x}(t) \in \mathbb{C}^{M\times 1}$ is subject to the power constraint $\mathbb{E}\big(\|\boldsymbol{x}(t)\|^2\big) \le P$, $\forall t$.

B. Two-user MIMO Interference Channel

For a two-user (M_1, M_2, N_1, N_2) MIMO interference channel (IC) with M_i antennas at Transmitter i and N_j antennas at Receiver j, for $i, j = 1, 2$, the discrete time signal model is given by

$$\boldsymbol{y}_i(t) = \boldsymbol{H}_{i1}(t)\boldsymbol{x}_1(t) + \boldsymbol{H}_{i2}(t)\boldsymbol{x}_2(t) + \boldsymbol{z}_i(t) \tag{2}$$

for any time instant t, where $\boldsymbol{H}_{ji}(t) \in \mathbb{C}^{N_j \times M_i}$ $(i, j = 1, 2)$ is the channel matrix between Transmitter i and Receiver j; the coded input signal $\boldsymbol{x}_i(t) \in \mathbb{C}^{M_i\times 1}$ is subject to the power constraint $\mathbb{E}\big(\|\boldsymbol{x}_i(t)\|^2\big) \le P$ for $i = 1, 2$, $\forall t$.

In the rest of this paper, we refer to MIMO BC/IC as MIMO networks. For notational brevity, we define the ensemble of channel matrices, i.e., $\mathcal{H}(t) \triangleq \{\boldsymbol{H}_1(t), \boldsymbol{H}_2(t)\}$ (resp. $\mathcal{H}(t) \triangleq \{\boldsymbol{H}_{11}(t), \boldsymbol{H}_{21}(t), \boldsymbol{H}_{12}(t), \boldsymbol{H}_{22}(t)\}$), as the channel state for BC (resp. IC). We further define $\mathcal{H}^k \triangleq \{\mathcal{H}(t)\}_{t=1}^k$, and $\hat{\mathcal{H}}^k \triangleq \{\hat{\mathcal{H}}(t)\}_{t=1}^k$, where $k = 1, \cdots, n$.

C. Assumptions and Definitions

Assumption 1 (perfect delayed and imperfect current CSIT)**.** *At each time instant t, the transmitters know perfectly the delayed CSI $\mathcal{H}^{t-1}$, and obtain an imperfect estimate of the current CSI $\hat{\mathcal{H}}(t)$, which could, for instance, be produced by standard prediction based on past samples. The current CSIT estimate is modeled by*

$$\boldsymbol{H}_i(t) = \hat{\boldsymbol{H}}_i(t) + \tilde{\boldsymbol{H}}_i(t) \tag{3}$$

$$\boldsymbol{H}_{ij}(t) = \hat{\boldsymbol{H}}_{ij}(t) + \tilde{\boldsymbol{H}}_{ij}(t) \tag{4}$$

for BC and IC, respectively, where estimation error $\tilde{\boldsymbol{H}}_i(t)$ (resp. $\tilde{\boldsymbol{H}}_{ij}(t)$) and the estimate $\hat{\boldsymbol{H}}_i(t)$ (resp. $\hat{\boldsymbol{H}}_{ij}(t)$) are mutually independent, and each entry is assumed[1] *to be $\mathcal{N}_{\mathbb{C}}\left(0, \sigma_i^2\right)$ and $\mathcal{N}_{\mathbb{C}}\left(0, 1-\sigma_i^2\right)$. Further, we assume the following Markov chain*

$$(\mathcal{H}^{t-1}, \hat{\mathcal{H}}^{t-1}) \to \hat{\mathcal{H}}(t) \to \mathcal{H}(t), \quad \forall t, \tag{5}$$

which means $\mathcal{H}(t)$ is independent of $(\mathcal{H}^{t-1}, \hat{\mathcal{H}}^{t-1})$ conditional on $\hat{\mathcal{H}}(t)$. Furthermore, at the end of the transmission, i.e., at time instant n, the receivers know perfectly $\mathcal{H}^n$ and $\hat{\mathcal{H}}^n$.

It readily follows that, for any fat submatrix $\boldsymbol{H}$ of $\boldsymbol{H}_i$ or $\boldsymbol{H}_{ij}$, $\mathbb{E}(\log\det(\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}})) > -\infty$ and $\mathbb{E}(\log\det(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\mathsf{H}})) = O(1)$ when σ_i^2 goes to 0.

The assumption on the CSI at the receiver (CSIR) is in accordance with previous works with delayed CSIT, and does not add any limitation over the assumption made in [4]–[6]. We point out that only local CSIT/CSIR (the channel links with which the node is connected) is really helpful and leads to the same result. Nevertheless, we assume the CSIT/CSIR to be available in a global fashion for simplicity of presentation.

We are interested in characterizing the degrees of freedom (DoF) of the above system as functions of the quality of current CSIT, thus bridging between the two previously investigated extremes which are the perfect instantaneous CSIT and the fully outdated (non-instantaneous) CSIT cases. As it was established in previous works [13], [14], the imperfect current CSIT has beneficial value (in terms of improving the DoF) only if the CSIT estimation error decays at least exponentially with the

[1] We make the above assumption on the fading distribution to simplify the presentation, although the results can be applied to a broader class of distributions.

$$\begin{aligned}
d_1 &\le \min\{M, N_1\}, && \text{(6a)}\\
d_2 &\le \min\{M, N_2\}, && \text{(6b)}\\
d_1 + d_2 &\le \min\{M, N_1 + N_2\}, && \text{(6c)}
\end{aligned}$$

$$\frac{d_1}{\min\{M,N_1\}} + \frac{d_2}{\min\{M,N_1+N_2\}} \le 1 + \frac{\min\{M,N_1+N_2\} - \min\{M,N_1\}}{\min\{M,N_1+N_2\}}\alpha_1, \tag{6d}$$

$$\frac{d_1}{\min\{M,N_1+N_2\}} + \frac{d_2}{\min\{M,N_2\}} \le 1 + \frac{\min\{M,N_1+N_2\} - \min\{M,N_2\}}{\min\{M,N_1+N_2\}}\alpha_2, \tag{6e}$$

SNR or faster. Thus it is reasonable to study the regime by which the CSIT quality can be parameterized by an indicator $\alpha_i \ge 0$ such that:

$$\alpha_i \triangleq -\lim_{P\to\infty} \frac{\log \sigma_i^2}{\log P} \tag{7}$$

if the limit exists. This α_i indicates the quality of current CSIT corresponding to Receiver i at high SNR. While $\alpha_i = 0$ reflects the case with no current CSIT, $\alpha_i \to \infty$ corresponds to that with perfect instantaneous CSIT. As a matter of fact, when $\alpha_i \ge 1$, the quality of the imperfect current CSIT is sufficient to avoid the DoF loss, and ZF precoding with this imperfect CSIT is able to achieve the maximum DoF [27]. Therefore, we focus on the case $\alpha_i \in [0,1]$ henceforth. The connections between the above model and the linear prediction over existing time-correlated channel models with prescribed user mobility are highlighted in [13], [14]. According to the definition of the estimated current CSIT, we have $\mathbb{E}\big(|\boldsymbol{h}_k^{\mathsf{H}}(t)\hat{\boldsymbol{h}}_k^{\perp}(t)|^2\big) = \sigma_i^2 \sim P^{-\alpha_i}$, with $\boldsymbol{h}_k^{\mathsf{H}}$ representing any row of channel matrices $\boldsymbol{H}_i(t)$ (resp. $\boldsymbol{H}_{ij}(t)$), and $\hat{\boldsymbol{h}}_k^{\mathsf{H}}$ being its corresponding estimate.

A rate pair (R_1, R_2) is said to be *achievable* for the two-user MIMO networks with perfect delayed and imperfect current CSIT if there exists a $\big(2^{nR_1}, 2^{nR_2}, n\big)$ code scheme with:

- two message sets $\mathcal{W}_1 \triangleq [1 : 2^{nR_1}]$ and $\mathcal{W}_2 \triangleq [1 : 2^{nR_2}]$, from which two independent messages W_1 and W_2 intended respectively to Receiver 1 and Receiver 2 are uniformly chosen;
- one encoding function for (each) transmitter:

$$\begin{aligned} &\text{BC:} \quad \boldsymbol{x}(t) = f_t\big(W_1, W_2, \mathcal{H}^{t-1}, \hat{\mathcal{H}}^t\big)\\ &\text{IC:} \quad \boldsymbol{x}_i(t) = f_{i,t}\big(W_i, \mathcal{H}^{t-1}, \hat{\mathcal{H}}^t\big),\ i = 1,2; \end{aligned} \tag{8}$$

- one decoding function at the corresponding receiver,

$$\hat{W}_j = g_j\big(\boldsymbol{Y}_j^n, \mathcal{H}^n, \hat{\mathcal{H}}^n\big),\ j = 1,2 \tag{9}$$

for Receiver j, where $\boldsymbol{Y}_j^n \triangleq \{\boldsymbol{y}_j(t)\}_{t=1}^n$,

such that the average decoding error probability $P_e^{(n)}$, defined as $P_e^{(n)} \triangleq \mathbb{P}\big((W_1, W_2) \neq (\hat{W}_1, \hat{W}_2)\big)$, vanishes as the code length n tends to infinity. The capacity region $\mathcal{C}$ is defined as the set of all achievable rate pairs. Accordingly, the DoF region can be defined as follows:

Definition 1 (degrees of freedom region)**.** *The degrees of freedom (DoF) region for the two-user MIMO network is defined as*

$$\mathcal{D} = \Bigg\{(d_1, d_2) \in \mathbb{R}_+^2 \,|\, \forall (w_1, w_2) \in \mathbb{R}_+^2, w_1 d_1 + w_2 d_2 \le \limsup_{P\to\infty}\left(\sup_{(R_1,R_2)\in\mathcal{C}} \frac{w_1 R_1 + w_2 R_2}{\log P}\right)\Bigg\}. \tag{10}$$

III. Main Results

According to the assumptions and definitions in the previous section, the main results of this paper are stated as the following two theorems:

Theorem 1. *For the two-user (M, N_1, N_2) MIMO BC with delayed and imperfect current CSIT, the optimal DoF region $\{(d_1, d_2)|(d_1, d_2) \in \mathbb{R}_+^2\}$ is characterized by eq-(6) on the top of this page, where $\alpha_i \in [0,1]$ $(i = 1,2)$ indicates the current CSIT quality exponent of Receiver i's channel.*

Proof: The proof of achievability will be presented in Section IV showing some insights with toy examples, and in Section V for the general formulation. The converse proof will be given in Section VI focusing on (6d) and (6e), because the first three bounds correspond to the upper bounds under perfect CSIT settings and thus hold trivially under delayed and imperfect current CSIT settings. ■

Remark 1. *This result yields a number of previous results as special cases: the delayed CSIT case [5] for $\alpha_1 = \alpha_2 = 0$, where the sum DoF bound* (6c) *is inactive; perfect CSIT case [1] for $\alpha_1 = \alpha_2 = 1$, where the weighted sum DoF bounds* (6d) *and* (6e) *are inactive; partial CSIT (i.e., perfect CSIT for one channel and delayed CSIT for the other one) case [24] for $\alpha_1 = 1, \alpha_2 = 0$, where only* (6b) *and* (6e) *are active; delayed CSIT in MISO BC for $N_1 = N_2 = 1$ [14], [15], [18].*

Before presenting the optimal DoF region for MIMO IC, we specify two conditions.

Definition 2 (Condition C_k)**.** *Given $k \in \{1,2\}$, the condition C_k holds, indicating the following inequalities*

$$M_k \ge N_j, \quad M_j < N_k, \quad M_1 + M_2 > N_1 + N_2 \tag{11}$$

are true, $\forall\ j \in \{1,2\},\ j \neq k$.

Remark 2. *This definition that points out the existence of the corresponding outer bound, is different from that in [6], in which the condition implies the activation of the outer bounds.*

Theorem 2. *For the two-user (M_1, M_2, N_1, N_2) MIMO IC with delayed and imperfect current CSIT, the optimal DoF region $\{(d_1, d_2)|(d_1, d_2) \in \mathbb{R}_+^2\}$ is characterized by eq-*(12)

$$
\begin{aligned}
d_1 &\le \min\{M_1, N_1\}, && \text{(12a)}\\
d_2 &\le \min\{M_2, N_2\}, && \text{(12b)}\\
d_1 + d_2 &\le \min\{M_1+M_2, N_1+N_2, \max\{M_1,N_2\}, \max\{M_2,N_1\}\}, && \text{(12c)}\\
\frac{d_1}{\min\{M_2,N_1\}} + \frac{d_2}{\min\{M_2,N_1+N_2\}} &\le \frac{\min\{N_1, M_1+M_2\}}{\min\{M_2,N_1\}} + \frac{\min\{M_2,N_1+N_2\} - \min\{M_2,N_1\}}{\min\{M_2,N_1+N_2\}}\alpha_1, && \text{(12d)}\\
\frac{d_1}{\min\{M_1,N_1+N_2\}} + \frac{d_2}{\min\{M_1,N_2\}} &\le \frac{\min\{N_2, M_1+M_2\}}{\min\{M_1,N_2\}} + \frac{\min\{M_1,N_1+N_2\} - \min\{M_1,N_2\}}{\min\{M_1,N_1+N_2\}}\alpha_2, && \text{(12e)}\\
d_1 + \frac{N_1+2N_2-M_2}{N_2} d_2 &\le N_1+N_2+(N_1-M_2)\alpha_2, \quad \text{if } C_1 \text{ holds} && \text{(12f)}\\
d_2 + \frac{N_2+2N_1-M_1}{N_1} d_1 &\le N_1+N_2+(N_2-M_1)\alpha_1, \quad \text{if } C_2 \text{ holds} && \text{(12g)}
\end{aligned}
$$

---

on the top of this page, where $\alpha_i \in [0,1]$ $(i=1,2)$ indicates the current CSIT quality exponent corresponds to Receiver i.

Proof: The general formulation of achievability will be presented in Section V, and the converse will be given in Section VI. For the converse, the first three inequalities correspond to the outer bounds for the case of perfect CSIT, which should also hold for our setting. Hence, it is sufficient to prove the last four bounds. Due to the symmetry property of the bounds (12d) and (12e), (12f) and (12g), it is sufficient to prove the bounds (12d) and (12f). ■

Remark 3. *Some previous reported results can be regarded as special cases of our results: delayed CSIT case [6] for $\alpha_1 = \alpha_2 = 0$; perfect CSIT case [2] for $\alpha_1 = \alpha_2 = 1$, where the weighted sum DoF bounds* (12d)-(12g) *are inactive; hybrid CSIT (i.e., perfect CSIT for one channel and delayed CSIT for the other one) case [26] for $\alpha_1 = 1, \alpha_2 = 0$, where the bounds* (12e) *and* (12f) *are active.*

IV. Achievability: Toy Examples

To introduce the main idea of our achievability scheme, we revisit MAT [4] and α-MAT alignment [13]–[15] for the case of MISO BC in Section IV. A, followed by an alternative way built on *block-Markov encoding and backward decoding* in Section IV. B, as well as some examples in Section IV. C and IV. D showing that block-Markov encoding allows us to balance the asymmetry both in current CSIT qualities and antenna configurations. Although MAT [4] and α-MAT alignment [13]–[15] appear to be conceptually different, these schemes boil down into a *single* block-Markov encoding scheme (of an infinite number of *constant-length* blocks). In fact, both schemes can be represented exactly in the same manner with different parameters.

A. MAT v.s. α-MAT Revisit

Let us take the simplest antenna configuration, i.e., $(2,1,1)$ BC, as an example. Recall that both MAT and α-MAT deliver symbol under the same structure. Specifically, in the first phase (Phase I), two independent messages w_1 and w_2 are encoded into two independent vectors $\boldsymbol{u}_1(w_1)$ and $\boldsymbol{u}_2(w_2)$ with different covariance matrices $\boldsymbol{Q}_1 \triangleq \mathbb{E}\left(\boldsymbol{u}_1\boldsymbol{u}_1^{\mathsf{H}}\right)$ and $\boldsymbol{Q}_2 \triangleq \mathbb{E}\left(\boldsymbol{u}_2\boldsymbol{u}_2^{\mathsf{H}}\right)$. The sum of these vectors are sent out, i.e.,

$$
\begin{aligned}
&\boldsymbol{x}[1] = \boldsymbol{u}_1 + \boldsymbol{u}_2,\\
&s.t. \begin{cases} \text{MAT:} & \boldsymbol{Q}_1 = \boldsymbol{Q}_2 = P\mathbf{I},\\ \alpha\text{-MAT:} & \begin{cases} \boldsymbol{Q}_1 = P_1\boldsymbol{\Phi}_{\hat{h}_2} + P_2\boldsymbol{\Phi}_{\hat{h}_2^\perp}\\ \boldsymbol{Q}_2 = P_1\boldsymbol{\Phi}_{\hat{h}_1} + P_2\boldsymbol{\Phi}_{\hat{h}_1^\perp}\end{cases}\end{cases}
\end{aligned} \tag{13}
$$

where $P_1 \sim P^{1-\alpha}$, $P_2 = P - P_1 \sim P$, $\forall\, \alpha \in [0,1]$, and $\boldsymbol{\Phi}_h \triangleq \frac{\boldsymbol{h}\boldsymbol{h}^{\mathsf{H}}}{\|\boldsymbol{h}\|^2}$. Each receiver experiences some interferences caused by the symbols dedicated to the other receiver

$$
\begin{cases} \eta_1 \triangleq \boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{u}_2\\ \eta_2 \triangleq \boldsymbol{h}_2^{\mathsf{H}}\boldsymbol{u}_1 \end{cases} s.t. \begin{cases} \text{MAT:} & \mathbb{E}\left(|\eta_i|^2\right) \sim P\\ \alpha\text{-MAT:} & \mathbb{E}\left(|\eta_i|^2\right) \sim P^{1-\alpha} \end{cases} \tag{14}
$$

Then, the task of the second phase is to multicast the interferences (η_1, η_2) to *both* receivers. The main difference between the MAT and α-MAT lies in the way in which the interferences are sent. While the analog version of η_k is sent in two slots with MAT, the digitized version is sent with α-MAT instead. Note that the covariance matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$, or equivalently, the spatial precoding and power allocation, of α-MAT are such that the mutual interferences (η_1, η_2) have a reduced power level $P^{1-\alpha}$. According to the rate-distortion theorem [28], each interference η_k, $k=1,2$, can be compressed with a source codebook of size $P^{1-\alpha}$ or $(1-\alpha)\log P$ bits into an index l_k, in such a way that the average distortion between η_k and the source codeword $\hat{\eta}_k(l_k)$ is comparable to the AWGN level [14]. Then, the index l_k is encoded with a channel codebook into a codeword $\boldsymbol{x}_c(l_k) \sim P\mathbf{I}_2$ and sent as the common message to both receivers. Thanks to the reduced range of l_k, there is still room to transmit private messages. The structure of the two slots in the second phase (Phase II) is

$$
\begin{cases} \text{MAT:} & \boldsymbol{x}[2] = \boldsymbol{v}_k\eta_k,\\ \alpha\text{-MAT:} & \boldsymbol{x}[2] = \boldsymbol{x}_c(l_k) + \boldsymbol{u}_{p1} + \boldsymbol{u}_{p2} \end{cases} \tag{15}
$$

where $k=1,2$, $\boldsymbol{v}_k$ is a randomly chosen vector; the covariances of the private signals $\boldsymbol{u}_{p1}$ and $\boldsymbol{u}_{p2}$ are respectively $\boldsymbol{Q}_{u_{p1}} = P^\alpha\boldsymbol{\Phi}_{\hat{h}_2^\perp}$ and $\boldsymbol{Q}_{u_{p2}} = P^\alpha\boldsymbol{\Phi}_{\hat{h}_1^\perp}$ in such a way that they are drown in the AWGN at the unintended receivers without creating noticeable interferences (at high SNR). At Receiver k, the common messages l_1 and l_2 are first decoded from the two slots in Phase II, by treating the private signal $\boldsymbol{u}_{p1}$ or $\boldsymbol{u}_{p2}$ as noise. The common messages are then used to 1) reconstruct

η_1 and η_2 that will be used with the received signal in Phase I to decode w_k and recover $2-\alpha$ DoF, and 2) to reconstruct $\boldsymbol{x}_c(l_k)$ and remove it from the received signals in Phase II so as to decode the private messages and recover 2α DoF (in two slots). In the end, $2-\alpha+2\alpha = 2+\alpha$ DoF per user is achievable in three slots, yielding an average DoF of $\frac{2+\alpha}{3}$ per user. The interested readers may refer to [14] for more details of α-MAT alignment.

B. An Alternative: Block-Markov Implementation

In fact, both MAT and α-MAT can be implemented in a block-Markov fashion, the concept of which is shown in Fig. 1 for $\alpha = 0$. The common message $\boldsymbol{x}_c(l_{b-1})$ comes from the previous block $b-1$, and $\boldsymbol{u}_k(w_{kb})$ is the new private message dedicated to Receiver k $(k=1,2)$. Essentially, we "squeeze" the Phase II of block $b-1$ and the Phase I of block b into one single block, with proper power and rate scaling.

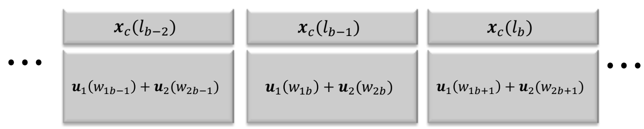


Fig. 1: Block-Markov Encoding.

The transmission consists of B blocks of length n. For simplicity of demonstration, we set $n=1$. In block b, the transmitter sends a mixture of two new private messages w_{1b} and w_{2b} together with one common message l_{b-1}, for $b = 1,\ldots,B$. As it will become clear, the message l_{b-1} is the compression index of the mutual interferences experienced by the receivers in the previous block $b-1$. By encoding w_{1b}, w_{2b}, and l_{b-1} into $\boldsymbol{u}_1(w_{1b})$, $\boldsymbol{u}_2(w_{2b})$, and $\boldsymbol{x}_c(l_{b-1})$, respectively, with independent channel codebooks, the transmitted signal is written as

$$\boldsymbol{x}[b] = \boldsymbol{x}_c(l_{b-1}) + \boldsymbol{u}_1(w_{1b}) + \boldsymbol{u}_2(w_{2b}), \quad b = 1,\ldots,B \tag{16}$$

where we set $l_0 = 1$ to initiate the transmission and $w_{1B} = w_{2B} = 1$ to end it. As before, the common message $\boldsymbol{x}_c(l_{b-1})$ is with power P, whereas the precoding in $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ is with a reduced power, parameterized by A, A', with $0 \le A, A' \le 1$, such that

$$\boldsymbol{Q}_1 = P^A \boldsymbol{\Phi}_{\hat{h}_2} + P^{A'} \boldsymbol{\Phi}_{\hat{h}_2^\perp}, \quad \boldsymbol{Q}_2 = P^A \boldsymbol{\Phi}_{\hat{h}_1} + P^{A'} \boldsymbol{\Phi}_{\hat{h}_1^\perp} \tag{17}$$

where $A \triangleq (A'-\alpha)^+$. The mutual interferences are defined similarly and their powers are now reduced

$$y_1[b] = \underbrace{\boldsymbol{h}_1^{\mathsf{H}} \boldsymbol{x}_c(l_{b-1})}_{P} + \underbrace{\boldsymbol{h}_1^{\mathsf{H}} \boldsymbol{u}_1(w_{1b})}_{P^{A'}} + \underbrace{\boldsymbol{h}_1^{\mathsf{H}} \boldsymbol{u}_2(w_{2b})}_{\eta_{1b} \sim P^A} \tag{18}$$

$$y_2[b] = \underbrace{\boldsymbol{h}_2^{\mathsf{H}} \boldsymbol{x}_c(l_{b-1})}_{P} + \underbrace{\boldsymbol{h}_2^{\mathsf{H}} \boldsymbol{u}_2(w_{2b})}_{P^{A'}} + \underbrace{\boldsymbol{h}_2^{\mathsf{H}} \boldsymbol{u}_1(w_{1b})}_{\eta_{2b} \sim P^A} \tag{19}$$

where we omit the block indices for the channel coefficients as well as the AWGN for brevity. At the end of block b, (η_{1b}, η_{2b}) are compressed with a codebook of size P^{2A} into an index $l_b \in \{1,\ldots,P^{2A}\}$. The distortion between (η_{1b}, η_{2b}) and $(\hat{\eta}_1(l_b), \hat{\eta}_2(l_b))$ is at the noise level.

At the end of B blocks, Receiver k would like to retrieve $w_{k1},\ldots,w_{k,B-1}$. Let us focus on Receiver 1, without loss of generality. In this particular case, l_{b-1} can be decoded at the end of block b, by treating the private signals as noise, i.e., with signal-to-interference-and-noise-ratio (SINR) level $P^{1-A'}$, for $b = 2,\ldots,B$. The correct decoding of l_{b-1} is guaranteed if the SINR can support the DoF of $2A$ for the common message, i.e.,

$$2A \le 1 - A'. \tag{20}$$

Given that this condition is satisfied, $l_0, l_1, \ldots, l_{B-1}$ are available to both receivers. Therefore, η_{1b}, η_{2b}, $b = 1,\ldots,B-1$, are known, up to the noise level. To decode w_{1b}, Receiver 1 uses η_{1b}, η_{2b}, and l_{b-1} to form the following 2×2 MIMO system

$$\begin{bmatrix} y_1[b] - \boldsymbol{h}_1^{\mathsf{H}} \boldsymbol{x}_c(l_{b-1}) - \eta_{1b} \\ \eta_{2b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}_1^{\mathsf{H}} \\ \boldsymbol{h}_2^{\mathsf{H}} \end{bmatrix} \boldsymbol{u}_1(w_{1b}) \tag{21}$$

where the equivalent channel matrix has rank 2 almost surely. This decoding strategy for the private message boils down to the backward decoding, where the mutual interferences (η_{1b}, η_{2b}) decoded in the future block are utilized in current block as side information. From the covariance matrix $\boldsymbol{Q}_1$ of $\boldsymbol{u}_1$ from (17), we deduce that the correct decoding of w_{1b} is guaranteed if the DoF d_{1b} of w_{1b} satisfies

$$d_{1b} \le A + A'. \tag{22}$$

Combining (20) and (22), it readily follows that the optimal A' should equalize (20), i.e., $A'^* = \frac{1+2\alpha}{3}$. Thus, we achieve $d_{1b} = \frac{2+\alpha}{3}$. Due to the symmetry, d_{2b} has the same value. Finally, we have

$$d_k = \frac{1}{B}\sum_{b=1}^{B-1} d_{kb} = \frac{B-1}{B}\frac{2+\alpha}{3}, \quad k = 1,2 \tag{23}$$

which goes to $\frac{2+\alpha}{3}$ when $B \to \infty$.

By now, we have shown that both MAT and α-MAT schemes can be interpreted under a common framework of block-Markov encoding with power allocation parameters (A, A') and that they only differ from the choice of these parameters. As we will show in the following subsections, the strength (or benefit) of the block-Markov encoding framework becomes evident in the asymmetric system setting, where the original α-MAT alignment fails to achieve the optimal DoF in general.

C. Asymmetry in Current CSIT Qualities

Let us consider again the MISO BC case but assume now that the CSIT qualities of two channels are different, i.e., $\alpha_1 \neq \alpha_2$, where α_k $(k=1,2)$ is for Receiver k. The signal model is in the exact same form as in (16) with a more general precoding, parameterized by A_k, A'_k, with $0 \le A_k, A'_k \le 1$, such that

$$\boldsymbol{Q}_1 = P^{A_1} \boldsymbol{\Phi}_{\hat{h}_2} + P^{A'_1} \boldsymbol{\Phi}_{\hat{h}_2^\perp}, \quad \boldsymbol{Q}_2 = P^{A_2} \boldsymbol{\Phi}_{\hat{h}_1} + P^{A'_2} \boldsymbol{\Phi}_{\hat{h}_1^\perp} \tag{24}$$

where $A_k \triangleq (A'_k - \alpha_j)^+$, $j \neq k \in \{1,2\}$. Following the same footsteps as in the symmetric case, it is readily shown that $\eta_{1b} \sim P^{A_2}$ and $\eta_{2b} \sim P^{A_1}$ and that (η_{1b}, η_{2b}) can be

TABLE I: Parameter Setting for the $(2,1,1)$ BC Case ($\alpha_1 \geq \alpha_2$)

Condition	A'_1	A'_2	Corner Point (d_1, d_2)
$2\alpha_1 - \alpha_2 \leq 1$	$A'_1 = \frac{1+\alpha_1+\alpha_2}{3}$	$A'_2 = \frac{1+\alpha_1+\alpha_2}{3}$	$\left(\frac{2+2\alpha_1-\alpha_2}{3}, \frac{2-\alpha_1+2\alpha_2}{3}\right)$
	$A'_1 = \frac{1+\alpha_2}{2}$	$A'_2 = \alpha_1$	$(1, \alpha_1)$
$2\alpha_1 - \alpha_2 > 1$	$A'_1 = \frac{1+\alpha_2}{2}$	$A'_2 = \frac{1+\alpha_2}{2}$	$(1, \frac{1+\alpha_2}{2})$
-	$A'_1 = \alpha_2$	$A'_2 = \frac{1+\alpha_1}{2}$	$(\alpha_2, 1)$

compressed up to the noise level with a source codebook of size $P^{A_1+A_2}$. The decoding at both receivers is the same as before. To decode the common message l_{b-1} by treating the private signals as noise, since the SINR is $P^{1-A'_1}$ at Receiver 1 and $P^{1-A'_2}$ at Receiver 2, the DoF of the common message should satisfy

$$A_1 + A_2 \leq \min\{1 - A'_1, 1 - A'_2\}. \tag{25}$$

Using the common messages l_b and l_{b-1} as side information, w_{1b} and w_{2b} can be decoded at the respective receivers if

$$d_{1b} \leq A_1 + A'_1 \quad \text{and} \quad d_{2b} \leq A_2 + A'_2. \tag{26}$$

By carefully selecting the parameters A'_1 and A'_2, all corner points of the DoF outer bound can be achieved, as shown in Table I on the top of this page where the condition is to make sure the corner points exist. Note that the corner point $(\alpha_2, 1)$ always exists as long as $\alpha_1 \geq \alpha_2$.

D. Asymmetry in Antenna Configurations

We use the $(4,3,2)$ MIMO BC case to show that the block-Markov encoding can achieve the optimal performance in asymmetric antenna settings. Recall that, in the previous subsections, the backward decoding is performed to decode the private messages, and that the common messages can be decoded *block by block*. In this case, however, we also need backward decoding to decode the common messages as well.

The same transmission signal model (16) is used here, with the following precoding, parameterized by A_k and A'_k, $k = 1,2$, $0 \leq A_k \leq A'_k \leq 1$:

$$\boldsymbol{Q}_1 = P^{A_1}\boldsymbol{\Phi}_{\hat{H}_2} + P^{A'_1}\boldsymbol{\Phi}_{\hat{H}_2^{\perp}}, \quad \boldsymbol{Q}_2 = P^{A_2}\boldsymbol{\Phi}_{\hat{H}_1} + P^{A'_2}\boldsymbol{\Phi}_{\hat{H}_1^{\perp}} \tag{27}$$

where A_k, $k \neq j \in \{1,2\}$, is defined as

$$A_k \triangleq \begin{cases} (A'_k - \alpha_j)^+, & d_k \leq 4 - N_j\alpha_j, \\ \frac{d_k - (4-N_j)}{N_j}, & d_k > 4 - N_j\alpha_j \end{cases} \tag{28}$$

with $d_k \in \mathbb{R}_+$ being the achievable DoF associated with Receiver k. It is readily verified that $A'_k - \alpha_j \leq A_k \leq A'_k$ is always true, such that the created interference at intended Receiver j is of power level A_k, and the desired signal at Receiver k is of level A'_k.

We recall that the common message $\boldsymbol{x}_c(l_{b-1})$ is transmitted with power P and that the ranks of $\boldsymbol{\Phi}_{\hat{H}_2}$, $\boldsymbol{\Phi}_{\hat{H}_2^{\perp}}$, $\boldsymbol{\Phi}_{\hat{H}_1}$, and $\boldsymbol{\Phi}_{\hat{H}_1^{\perp}}$ are respectively 2, 2, 3, and 1, almost surely. The received signals are now vectors given by

$$\boldsymbol{y}_1[b] = \underbrace{\boldsymbol{H}_1\boldsymbol{x}_c(l_{b-1})}_{P\mathbf{I}_3} + \boldsymbol{H}_1\boldsymbol{u}_1(w_{1b}) + \underbrace{\boldsymbol{H}_1\boldsymbol{u}_2(w_{2b})}_{\boldsymbol{\eta}_{1b}\sim P^{A_2}\mathbf{I}_3}, \tag{29}$$

$$\boldsymbol{y}_2[b] = \underbrace{\boldsymbol{H}_2\boldsymbol{x}_c(l_{b-1})}_{P\mathbf{I}_2} + \boldsymbol{H}_2\boldsymbol{u}_2(w_{2b}) + \underbrace{\boldsymbol{H}_2\boldsymbol{u}_1(w_{1b})}_{\boldsymbol{\eta}_{2b}\sim P^{A_1}\mathbf{I}_2}. \tag{30}$$

Following the same footsteps as in the single receive antenna case, it is readily shown that $(\boldsymbol{\eta}_{1b}, \boldsymbol{\eta}_{2b})$ can be compressed up to the noise level with a source codebook of size $P^{2A_1+3A_2}$. For convenience, let us define

$$d_\eta \triangleq 2A_1 + 3A_2. \tag{31}$$

Unlike the MISO case where the common messages can be decoded independently in each block without loss of optimality, backward decoding is required to *jointly* decode the common and private messages in the general MIMO case, in order to achieve the optimal DoF. As we will see later on, the common rate can be improved with backward decoding in general. The decoding starts at block B. Since w_{1B} and w_{2B} are both known, the private signals can be removed from the received signals $\boldsymbol{y}_1[B]$ and $\boldsymbol{y}_2[B]$. The common message l_{B-1} can be decoded at both receivers if $d_\eta \leq 2$. At block b, for $b = B-1, \ldots, 2$, assuming l_b is known perfectly from the decoding of block $b+1$, $\boldsymbol{\eta}_{1b}$ and $\boldsymbol{\eta}_{2b}$ can be reconstructed up to the noise level. The following MIMO system can be obtained

$$\begin{bmatrix} \boldsymbol{y}_1[b] - \boldsymbol{\eta}_{1b} \\ \boldsymbol{\eta}_{2b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_1 \\ 0 \end{bmatrix} \boldsymbol{x}_c(l_{b-1}) + \begin{bmatrix} \boldsymbol{H}_1 \\ \boldsymbol{H}_2 \end{bmatrix} \boldsymbol{u}_1(w_{1b}). \tag{32}$$

Note that this is a multiple-access channel (MAC) from which l_{b-1} and w_{1b} can be correctly decoded if the rate pair lies within the following region

$$d_\eta \leq 3 \tag{33}$$

$$d_{1b} \leq 2A_1 + 2A'_1 \tag{34}$$

$$d_\eta + d_{1b} \leq 3 + 2A_1, \tag{35}$$

whose general proof is provided in Appendix A. Let us set d_{1b} to equalize (34). Then, (33) and (35) imply $d_\eta \leq 3 - 2A'_1$. Similar analysis on Receiver 2 will lead to $d_\eta \leq 2 - A'_2$, by setting $d_{2b} = A'_2 + 3A_2$. Therefore, from (31), we obtain the following constraint

$$2A_1 + 3A_2 \leq \min\{3 - 2A'_1, 2 - A'_2\} \tag{36}$$

to achieve any (d_{1b}, d_{2b}) such that

$$d_{1b} \leq 2A_1 + 2A'_1 \quad \text{and} \quad d_{2b} \leq A'_2 + 3A_2. \tag{37}$$

By letting $B \to \infty$, $d_1 = 2A_1 + 2A'_1$ and $d_2 = A'_2 + 3A_2$ can be achieved for any $A'_1, A'_2 \leq 1$ given the definition of (A_1, A_2) in (28), as long as (36) is satisfied. We can show that, by properly choosing (A'_1, A'_2), all the corner points given by the outer bound can be achieved. For example, by setting $\alpha_1 = \alpha_2 = \alpha$, the values (A'_1, A'_2) that achieve the corner points are illustrated in Table II on the top of the next page. Note that $(\frac{12}{5}, \frac{4}{5} + \alpha)$ exists only when $\alpha \leq \frac{4}{5}$, whereas $(3\alpha, 4 - 3\alpha)$ and $(4 - 2\alpha, 2\alpha)$ exist only when $\alpha > \frac{4}{5}$.

TABLE II: Parameter Setting for the $(4,3,2)$ BC Case with $\alpha_1=\alpha_2=\alpha$.

Corner Point (d_1,d_2)	Cond.	(A'_1,A'_2)	(A_1,A_2)	d_η
$(3,\alpha)$	$\alpha\le\frac{1}{2}$	$(\frac{3+2\alpha}{4},\alpha)$	$(\frac{3-2\alpha}{4},0)$	$\frac{3-2\alpha}{2}$
	$\alpha>\frac{1}{2}$	$(1,\alpha)$	$(\frac{1}{2},0)$	1
$(2\alpha,2)$	$\alpha\le\frac{2}{3}$	$(\alpha,\frac{2+3\alpha}{4})$	$(0,\frac{2-\alpha}{4})$	$\frac{6-3\alpha}{4}$
	$\alpha>\frac{2}{3}$	$(\alpha,1)$	$(0,\frac{1}{3})$	1
$(\frac{12}{5},\frac{4}{5}+\alpha)$	$\alpha\le\frac{4}{5}$	$(\frac{3}{5}+\frac{1}{2}\alpha,\frac{1}{5}+\alpha)$	$(\frac{3}{5}-\frac{1}{2}\alpha,\frac{1}{5})$	$\frac{9}{5}-\alpha$
$(3\alpha,4-3\alpha)$	$\alpha>\frac{4}{5}$	$(1,1)$	$(\frac{3\alpha-2}{2},1-\alpha)$	1
$(4-2\alpha,2\alpha)$	$\alpha>\frac{4}{5}$	$(1,1)$	$(1-\alpha,\frac{2\alpha-1}{3})$	1

V. Achievability: the General Formulation

As aforementioned, the key ingredients of the achievability scheme consist of:

- *block-Markov encoding with a constant block length*: the fresh messages in the current block and the interferences created in the past blocks are encoded together with the proper rate splitting and power scaling;
- *spatial precoding with imperfect current CSIT*: with proper power allocation over the range and null spaces of the inaccurate current channel, the interference power at unintended receiver can be reduced as compared to that without any CSIT;
- *interference quantization*: instead of forwarding the overheard interference directly in an analog way as done in pure delayed CSIT scenario, the reduced-power interferences are compressed first with a reduced number of bits, and forwarded in a digital fashion with lower rate;
- *backward decoding*: the messages are decoded from the last block to the first one, where in each block the messages are decoded with the aid of side information provided by the blocks in the future.

In the following, the general achievability scheme will be described in detail for BC and IC respectively.

A. Broadcast Channels

First of all, we notice that the region (6) given in Theorem 1 does not depend on M (resp. N_k) when $M>N_1+N_2$ (resp. $N_k>M$). Therefore, it is sufficient to prove the achievability for the case $M\le N_1+N_2$ and $N_k\le M$. And the achievability for the other cases can be inferred by simply switching off the additional transmit/receive antennas. Thus, it yields

$$\begin{aligned} M &= \min\{M, N_1+N_2\}, \\ N_k &= \min\{M, N_k\}, \quad k=1,2. \end{aligned} \tag{38}$$

<u>Block-Markov encoding</u>

The block-Markov encoding has the same structure as before, namely,

$$\boldsymbol{x}[b]=\boldsymbol{x}_c(l_{b-1})+\boldsymbol{u}_1(w_{1b})+\boldsymbol{u}_2(w_{2b}), \quad b=1,\dots,B \tag{39}$$

where we recall that we set $l_0=1$ to initiate the transmission and $w_{1B}=w_{2B}=1$ to end it.

<u>Spatial precoding</u>

Both $\boldsymbol{u}_1,\boldsymbol{u}_2\in\mathbb{C}^{M\times 1}$ are precoded signals of M streams, such that

$$\boldsymbol{Q}_1=P^{A_1}\boldsymbol{\Phi}_{\hat{H}_2}+P^{A'_1}\boldsymbol{\Phi}_{\hat{H}_2^\perp}, \quad \boldsymbol{Q}_2=P^{A_2}\boldsymbol{\Phi}_{\hat{H}_1}+P^{A'_2}\boldsymbol{\Phi}_{\hat{H}_1^\perp} \tag{40}$$

where the rank of $\boldsymbol{\Phi}_{\hat{H}_k}$ is N_k whereas the rank of $\boldsymbol{\Phi}_{\hat{H}_k^\perp}$ is $M-N_k$, $k=1,2$. In other words, for Receiver k, N_j streams are sent in the subspace of the unintended Receiver j with power level A_k and the other $M-N_j$ streams are sent in the null space of Receiver j with power level A'_k, where (A_k,A'_k) satisfies

$$0\le A_k\le A'_k\le 1 \quad \text{and} \quad A_k\ge A'_k-\alpha_j \tag{41}$$

for $j\neq k\in\{1,2\}$. Note that the above condition guarantees that the interference at Receiver j has power level A_k and the desired signal at Receiver k is of power level A'_k.

<u>Interference quantization</u>

Recall that the common message $\boldsymbol{x}_c(l_{b-1})$ is sent with power P. The received signals in block b are given by

$$\boldsymbol{y}_1[b]=\underbrace{\boldsymbol{H}_1\boldsymbol{x}_c(l_{b-1})}_{P\mathbf{I}_{N_1}}+\boldsymbol{H}_1\boldsymbol{u}_1(w_{1b})+\underbrace{\boldsymbol{H}_1\boldsymbol{u}_2(w_{2b})}_{\boldsymbol{\eta}_{1b}\sim P^{A_2}\mathbf{I}_{N_1}}, \tag{42}$$

$$\boldsymbol{y}_2[b]=\underbrace{\boldsymbol{H}_2\boldsymbol{x}_c(l_{b-1})}_{P\mathbf{I}_{N_2}}+\boldsymbol{H}_2\boldsymbol{u}_2(w_{2b})+\underbrace{\boldsymbol{H}_2\boldsymbol{u}_1(w_{1b})}_{\boldsymbol{\eta}_{2b}\sim P^{A_1}\mathbf{I}_{N_2}}. \tag{43}$$

It is readily shown that $(\boldsymbol{\eta}_{1b},\boldsymbol{\eta}_{2b})$ can be compressed up to the noise level with a source codebook of size $P^{N_2A_1+N_1A_2}$ into an index l_b. For convenience, let us define

$$d_{\eta_1}\triangleq N_1A_2,\ d_{\eta_2}\triangleq N_2A_1,\ \text{and}\ d_\eta\triangleq d_{\eta_1}+d_{\eta_2}. \tag{44}$$

<u>Backward decoding</u>

The decoding starts at block B. Since w_{1B} and w_{2B} are both known, the private signals can be removed from the received signals $\boldsymbol{y}_1[B]$ and $\boldsymbol{y}_2[B]$. The common message l_{B-1} can be decoded at both receivers if $d_\eta\le\min\{N_1,N_2\}$. At block b, assuming l_b is known perfectly from the decoding of block $b+1$, $\boldsymbol{\eta}_{1b}$ and $\boldsymbol{\eta}_{2b}$ can be reconstructed up to the noise level, for $b=B-1,\dots,2$. The following MIMO system can be obtained at Receiver k, $k=1,2$

$$\begin{bmatrix}\boldsymbol{y}_k[b]-\boldsymbol{\eta}_{kb}\\ \boldsymbol{\eta}_{jb}\end{bmatrix}=\begin{bmatrix}\boldsymbol{H}_k\\ 0\end{bmatrix}\boldsymbol{x}_c(l_{b-1})+\begin{bmatrix}\boldsymbol{H}_k\\ \boldsymbol{H}_j\end{bmatrix}\boldsymbol{u}_k(w_{kb}) \tag{45}$$

for $j \neq k \in \{1,2\}$. Since the common message l_{b-1} and the private message w_{kb} are both desired by Receiver k, this system corresponds to a multiple-access channel (MAC). As formally proved in Appendix A, Receiver k can decode correctly both messages if the following conditions are satisfied.

$$d_\eta \leq N_k \tag{46}$$
$$d_{kb} \leq N_j A_k + (M - N_j)A'_k \tag{47}$$
$$d_\eta + d_{kb} \leq N_k + N_j A_k. \tag{48}$$

Let us choose d_{kb} to be equal to the right hand side of (47) for $k = 1,2$ and $b = 1,..,B-1$. Then, the equality in (47) together with (44), (46), (48) implies, when letting $B \to \infty$, the following lemma.

Lemma 1 (decodability condition for BC)**.** *Let us define*

$$\mathcal{A}_{\rm BC} \triangleq \big\{(A_1, A'_1, A_2, A'_2) \mid A_k, A'_k \in [0,1], \\ A'_k - \alpha_j \leq A_k \leq A'_k, \forall\, k \neq j \in \{1,2\}\big\} \tag{49}$$
$$\mathcal{D}_{\rm BC} \triangleq \{(d_1,d_2) \mid d_k \in [0, N_k], \quad \forall\, k \in \{1,2\}\} \tag{50}$$

and

$$f_{A\text{-}d} \;: \mathcal{A}_{\rm BC} \to \mathcal{D}_{\rm BC} \tag{51}$$
$$(A_k, A'_k) \mapsto d_k \triangleq N_j A_k + (M - N_j)A'_k, \forall k \neq j \in \{1,2\}. \tag{52}$$

Then $(d_1,d_2) = f_{A\text{-}d}(\boldsymbol{A})$, for some $\boldsymbol{A} \in \mathcal{A}_{\rm BC}$, is achievable with the proposed scheme, if

$$d_{\eta_1} + d_1 \leq N_1, \tag{53}$$
$$d_{\eta_2} + d_2 \leq N_2. \tag{54}$$

where we recall $d_{\eta_1} \triangleq N_1 A_2$ and $d_{\eta_2} \triangleq N_2 A_1$.

Remark 4. *In the above lemma, d_{η_k} can be interpreted as the degrees of freedom occupied by the interference at Receiver k. Therefore,* (53) *and* (54) *are clearly outer bounds for any transmission strategies, i.e., the sum of the dimension of the useful signal and the dimension of the interference signal at the receiver side cannot exceed the total dimension of the signal space. These bounds are in general not tight except for special cases such as the "strong interference" regime where interference can be decoded completely and removed or the "weak interference" regime where the interference can be treated as noise while the useful signal power dominates the received power. Remarkably, the proposed scheme achieves these outer bounds. This is due to two of the main ingredients of our scheme, namely, the block-Markov encoding and interference quantization. The block-Markov encoding places the digitized interference in the "upper level" of the signal space (with full power) and thus "pushes" the channel into the "strong interference" regime in which the digitized interference can be decoded thanks to the structure brought by the interference quantization.*

Definition 3 (achievable region for BC)**.** *Let us define*

$$\mathcal{I}_A^{\rm BC} \triangleq \left\{ (A_1, A'_1, A_2, A'_2) \in \mathcal{A}_{\rm BC} \;\middle|\; \begin{array}{l} (d_1,d_2) = f_{A\text{-}d}(A_1,A'_1,A_2,A'_2), \\ \frac{d_k}{N_k} \leq 1 - A_j, \quad k \neq j \in \{1,2\} \end{array} \right\} \tag{55}$$

and the achievable DoF region of the proposed scheme

$$\mathcal{I}_d^{\rm BC} \triangleq f_{A\text{-}d}(\mathcal{I}_A^{\rm BC}) \triangleq \left\{ (d_1,d_2) \;\middle|\; \begin{array}{l} (d_1,d_2) = f_{A\text{-}d}(A_1,A'_1,A_2,A'_2), \\ (A_1,A'_1,A_2,A'_2) \in \mathcal{A}_{\rm BC}, \\ \frac{d_k}{N_k} \leq 1 - A_j, \quad k \neq j \in \{1,2\} \end{array} \right\}. \tag{56}$$

<u>*Achievability analysis*</u>

In the following, we would like to show that any pair (d_1,d_2) in the outer bound region defined by (6), hereafter referred to as $\mathcal{O}_d^{\rm BC}$, can be achieved by the proposed strategy. Therefore, it is sufficient to show that $\mathcal{O}_d^{\rm BC} \subseteq \mathcal{I}_d^{\rm BC}$. The main idea is as follows. If there exists a function

$$f_{d\text{-}A} \;:\; \mathcal{O}_d^{\rm BC} \to \mathcal{A}_{\rm BC} \tag{57}$$

such that

$$(d_1,d_2) = f_{A\text{-}d}(f_{d\text{-}A}(d_1,d_2)), \quad \text{and} \tag{58}$$
$$f_{d\text{-}A}(d_1,d_2) \in \mathcal{I}_A^{\rm BC}, \tag{59}$$

then for every $(d_1,d_2) \in \mathcal{O}_d^{\rm BC}$ we can use the power allocation $(A_1,A'_1,A_2,A'_2) = f_{d\text{-}A}(d_1,d_2)$ on the proposed scheme to achieve it, i.e.,

$$\mathcal{O}_d^{\rm BC} = f_{A\text{-}d}(f_{d\text{-}A}(\mathcal{O}_d^{\rm BC})) \subseteq f_{A\text{-}d}(\mathcal{I}_A^{\rm BC}) = \mathcal{I}_d^{\rm BC} \tag{60}$$

from which the achievability is proved. Now, we define formally the power allocation function.

Definition 4 (power allocation for BC)**.** *Let us define $f_{d\text{-}A}$: $\mathcal{O}_d^{\rm BC} \to \mathcal{A}_{\rm BC}$:*

$$(d_1,d_2) \mapsto (A_1,A'_1) \triangleq f_1(d_1),\; (A_2,A'_2) \triangleq f_2(d_2) \tag{61}$$

where f_k, $j \neq k \in \{1,2\}$, is specified as below.

- *When $M = N_j$: $A'_k = A_k = \frac{d_k}{M}$;*
- *When $M > N_j$ and $d_k < M - N_j\alpha_j$: $A_k = (A'_k - \alpha_j)^+$, and thus*

$$A'_k = \begin{cases} \frac{d_k}{M-N_j}, & \text{if } d_k < (M-N_j)\alpha_j; \\ \frac{d_k + N_j\alpha_j}{M}, & \text{otherwise}; \end{cases} \tag{62}$$

- *When $M > N_j$ and $d_k \geq M - N_j\alpha_j$: $A'_k = 1$, and thus $A_k = \frac{d_k - (M-N_j)}{N_j}$.*

It is readily shown that, for any $(d_1,d_2) \in \mathcal{O}_d^{\rm BC}$, the resulting power allocation always lies in $\mathcal{A}_{\rm BC}$ as defined in (49) and that (58) is always satisfied. It remains to show that (59) holds as well, i.e., the decodability condition in (55) is satisfied. To that end, for any $(d_1,d_2) \in \mathcal{O}_d^{\rm BC}$, we first define $(A_1,A'_1,A_2,A'_2) \triangleq f_{d\text{-}A}(d_1,d_2)$ which implies $d_j = N_k A_j + (M-N_k)A'_j$, $j \neq k \in \{1,2\}$. Applying this equality on the constraints in the outer bound $\mathcal{O}_d^{\rm BC}$ in (6), we have

$$\frac{d_k}{N_k} \leq \frac{M - (M-N_k)A'_j}{N_k} - A_j, \tag{63}$$
$$\frac{d_k}{N_k} \leq 1 - \left[\frac{(M-N_k)(A'_j - \alpha_k) + N_k A_j}{M}\right]^+ \tag{64}$$

for $k \neq j \in \{1,2\}$, where the first one is from the sum rate constraint (6c) whereas the second one is from the rest of the constraints in (6). The final step is to show that either of (63) and (64) implies the last constraint in (55):

- When $M = N_k$, (64) is identical to the last constraint in (55);
- When $M > N_k$ and $d_j \geq M - N_k\alpha_k$, we have $A'_j = 1$ according to the mapping $f_{d\text{-}A}$ defined in Definition 4. Hence, (63) is identical to the last constraint in (55);
- When $M > N_k$ and $d_j < M - N_k\alpha_k$, we have $A_j = (A'_j - \alpha_k)^+$ according to Definition 4. Hence,

$$\left[\frac{(M-N_k)(A'_j-\alpha_k)+N_kA_j}{M}\right]^+ \geq A_j \tag{65}$$

  with which (64) implies the last constraint in (55).

By now, we have proved the achievability through the existence of a proper power allocation function such that (58) and (59) are satisfied for every pair (d_1, d_2) in the outer bound.

B. Interference Channels

The proposed scheme for MIMO IC is similar to that for BC, with the differences that (a) the interferences can only be reconstructed at the transmitter from which the symbols are sent, and (b) antenna configuration does matter at both transmitters and receivers. Further, as with the broadcast channel, we notice that the region (12) given in Theorem 2 does not depend on M_k (resp. N_k) when $M_k > N_1 + N_2$ (resp. $N_k > M_1 + M_2$), $k = 1,2$. Therefore, it is sufficient to prove the achievability for the case $M_k \leq N_1 + N_2$ and $N_k \leq M_1 + M_2$, $k = 1,2$, since the achievability for the other cases can be inferred by simply switching off the additional transmit/receive antennas. Thus, it yields

$$\begin{aligned} M_k &= \min\{M_k, N_1+N_2\}, \\ N_k &= \min\{N_k, M_1+M_2\}, \quad k=1,2. \end{aligned} \tag{66}$$

We also define for notational convenience

$$N'_1 \triangleq \min\{N_1, M_2\}, \quad N'_2 \triangleq \min\{N_2, M_1\}. \tag{67}$$

<u>*Block-Markov encoding*</u>

The block-Markov encoding is done independently at both transmitters

$$\boldsymbol{x}_1[b] = \boldsymbol{x}_{1c}(l_{1,b-1}) + \boldsymbol{u}_1(w_{1b}), \tag{68}$$
$$\boldsymbol{x}_2[b] = \boldsymbol{x}_{2c}(l_{2,b-1}) + \boldsymbol{u}_2(w_{2b}), \quad b = 1,\ldots,B \tag{69}$$

where we set $l_{1,0} = l_{2,0} = 1$ to initiate the transmission and $w_{1B} = w_{2B} = 1$ to end it.

<u>*Spatial precoding*</u>

The signal $\boldsymbol{u}_k \in \mathbb{C}^{M_k\times 1}$, $k = 1,2$, is precoded signal of M_k streams, such that

$$\boldsymbol{Q}_1 = P^{A_1}\mathbf{\Phi}_{\hat{H}_{21}} + P^{A'_1}\mathbf{\Phi}_{\hat{H}_{21}^{\perp 1}} + P^{A''_1}\mathbf{\Phi}_{\hat{H}_{21}^{\perp 2}}, \tag{70}$$
$$\boldsymbol{Q}_2 = P^{A_2}\mathbf{\Phi}_{\hat{H}_{12}} + P^{A'_2}\mathbf{\Phi}_{\hat{H}_{12}^{\perp 1}} + P^{A''_2}\mathbf{\Phi}_{\hat{H}_{12}^{\perp 2}} \tag{71}$$

where we use $\hat{\boldsymbol{H}}_{jk}^{\perp 1}$ (resp. $\hat{\boldsymbol{H}}_{jk}^{\perp 2}$) to denote any matrix spanning the $(M_k - N'_j - \xi_k)$-dimensional (resp. ξ_k-dimensional) subspace of the null space of $\hat{\boldsymbol{H}}_{jk}$ where ξ_k will be specified later on. Therefore, the rank of $\mathbf{\Phi}_{\hat{H}_{jk}}$ is N'_j whereas the rank of $\mathbf{\Phi}_{\hat{H}_{jk}^{\perp 1}}$ and $\mathbf{\Phi}_{\hat{H}_{jk}^{\perp 2}}$ are respectively $M_k - N'_j - \xi_k$ and ξ_k, $k = 1,2$. The power levels (A_k, A'_k, A''_k) satisfy

$$\begin{aligned} &A_k, A'_k, A''_k \in [0,1], \\ A_k \leq A'_k,\ &A''_k \leq A'_k, \quad \text{and} \quad A_k \geq A'_k - \alpha_j \end{aligned} \tag{72}$$

for $j \neq k \in \{1,2\}$. Note that the above condition guarantees that the interference at Receiver j has power level A_k and the desired signal at Receiver k at power level A'_k.

<u>*Interference quantization*</u>

Recall that the common messages $\boldsymbol{x}_{1c}(l_{1,b-1})$ and $\boldsymbol{x}_{2c}(l_{2,b-1})$ are sent with power P. The received signals in block b are given by

$$\begin{aligned} \boldsymbol{y}_1[b] = &\underbrace{\boldsymbol{H}_{11}\boldsymbol{x}_{1c}(l_{1,b-1}) + \boldsymbol{H}_{12}\boldsymbol{x}_{2c}(l_{2,b-1})}_{P\mathbf{I}_{N_1}} \\ &+ \boldsymbol{H}_{11}\boldsymbol{u}_1(w_{1b}) + \underbrace{\boldsymbol{H}_{12}\boldsymbol{u}_2(w_{2b})}_{\boldsymbol{\eta}_{1b}\sim P^{A_2}\mathbf{I}_{N'_1}}, \end{aligned} \tag{73}$$

$$\begin{aligned} \boldsymbol{y}_2[b] = &\underbrace{\boldsymbol{H}_{22}\boldsymbol{x}_{2c}(l_{2,b-1}) + \boldsymbol{H}_{21}\boldsymbol{x}_{1c}(l_{1,b-1})}_{P\mathbf{I}_{N_2}} \\ &+ \boldsymbol{H}_{22}\boldsymbol{u}_2(w_{2b}) + \underbrace{\boldsymbol{H}_{21}\boldsymbol{u}_1(w_{1b})}_{\boldsymbol{\eta}_{2b}\sim P^{A_1}\mathbf{I}_{N'_2}}. \end{aligned} \tag{74}$$

It is readily shown that $\boldsymbol{\eta}_{1b}$ and $\boldsymbol{\eta}_{2b}$ can be compressed *separately* up to the noise level with two independent source codebooks of size $P^{N'_1A_2}$ and $P^{N'_2A_1}$, into indices $l_{2,b}$ and $l_{1,b}$, respectively. For convenience, let us define

$$d_{\eta_1} \triangleq N'_1A_2,\ d_{\eta_2} \triangleq N'_2A_1, \text{ and } d_\eta \triangleq d_{\eta_1} + d_{\eta_2}. \tag{75}$$

<u>*Backward decoding*</u>

The decoding starts at block B. Since w_{1B} and w_{2B} are both known, the private signals can be removed from the received signals $\boldsymbol{y}_1[B]$ and $\boldsymbol{y}_2[B]$. The common messages $l_{1,B-1}$ and $l_{2,B-1}$ can be decoded at both receivers if

$$d_{\eta_k} \leq \min\{M_j, N_1, N_2\}, \tag{76}$$
$$d_{\eta_1} + d_{\eta_2} \leq \min\{N_1, N_2\}, \tag{77}$$

i.e., the common rate pair should lie within the intersection of MAC regions at both receivers for the common messages. At block b, assuming both $l_{1,b}$ and $l_{2,b}$ are known perfectly from the decoding of block $b+1$, $\boldsymbol{\eta}_{1b}$ and $\boldsymbol{\eta}_{2b}$ can be reconstructed up to the noise level, for $b = B-1,\ldots,2$. The following MIMO system can be obtained at Receiver k

$$\begin{aligned} \begin{bmatrix} \boldsymbol{y}_k[b] - \boldsymbol{\eta}_{kb} \\ \boldsymbol{\eta}_{jb} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_{kk} \\ 0 \end{bmatrix} \boldsymbol{x}_{kc}(l_{k,b-1}) + \begin{bmatrix} \boldsymbol{H}_{kj} \\ 0 \end{bmatrix} \boldsymbol{x}_{jc}(l_{j,b-1}) \\ + \begin{bmatrix} \boldsymbol{H}_{kk} \\ \boldsymbol{H}_{jk} \end{bmatrix} \boldsymbol{u}_k(w_{kb}) \end{aligned} \tag{78}$$

for $j \neq k \in \{1,2\}$. Note that this system corresponds to a multiple-access channel from which the three independent messages $l_{1,b-1}$, $l_{2,b-1}$, and w_{kb} are to be decoded. It will be shown in the Appendix A that the three messages can be correctly decoded if the DoF quadruple $(d_{\eta_1}, d_{\eta_2}, d_{1b}, d_{2b})$ lies within the following region

$$d_{kb} \le N'_j A_k + (M_k - N'_j - \xi_k)A'_k + \xi_k A''_k \tag{79}$$
$$d_{\eta_k} \le \min\{M_j, N_1, N_2\} \tag{80}$$
$$d_{\eta_1} + d_{\eta_2} \le \min\{N_1, N_2\} \tag{81}$$
$$d_{\eta_k} + d_{kb} \le N'_k + \min\left\{M_k - N'_j, N_k - N'_k\right\} A'_k + N'_j A_k \tag{82}$$
$$d_{\eta_j} + d_{kb} \le \min\{M_k, N_k\} + N'_j A_k \tag{83}$$
$$d_{\eta_1} + d_{\eta_2} + d_{kb} \le N_k + N'_j A_k. \tag{84}$$

Now, let us fix

$$d_{kb} \triangleq N'_j A_k + (M_k - N'_j - \xi_k)A'_k + \xi_k A''_k \tag{85}$$
$$d_{\eta_j} \triangleq N'_j A_k \tag{86}$$

from which we can reduce the region defined by (79)-(84). First, we remove (79) that is implied by (85). Second, (80) is not active as it is implied by (86) and (81). Third, (81) is implied by (84) and (85). Finally, from (86), (83) is equivalent to $d_{kb} \le \min\{M_k, N_k\}$ that is implied by (85). Therefore, by letting $B \to \infty$, we have the following counterpart of Lemma 1 for interference channels.

Lemma 2 (decodability condition for IC)**.** *Let us define*

$$\mathcal{A}_{\rm IC} \triangleq \left\{ (A_1, A'_1, A''_1, A_2, A'_2, A''_2) \;\middle|\; \begin{array}{l} A_k, A'_k, A''_k \in [0,1] \\ A'_k - \alpha_j \le A_k \le A'_k, \;\; A''_k \le A'_k, \\ \xi_k A''_k \le N'_k(1 - A_j), \quad k \ne j \in \{1,2\} \end{array} \right\} \tag{87}$$

$$\mathcal{D}_{\rm IC} \triangleq \left\{(d_1, d_2) \mid d_k \in [0, \min\{M_k, N_k\}], \quad \forall\, k \in \{1,2\}\right\} \tag{88}$$

and

$$f_{A\text{-}d} \;:\; \mathcal{A}_{\rm IC} \to \mathcal{D}_{\rm IC} \tag{89}$$
$$(A_k, A'_k, A''_k) \mapsto d_k \triangleq N'_j A_k + (M_k - N'_j - \xi_k)A'_k + \xi_k A''_k, \quad \forall k \ne j \in \{1,2\} \tag{90}$$

where

$$\xi_k \triangleq \begin{cases} (M_k - N'_j)^+ - (N_k - N'_k)^+, & \text{if } C_k \text{ holds} \\ 0, & \text{otherwise.} \end{cases} \tag{91}$$

Then $(d_1, d_2) = f_{A\text{-}d}(\boldsymbol{A})$, for some $\boldsymbol{A} \in \mathcal{A}_{\rm IC}$, is achievable with the proposed scheme, if

$$d_{\eta_1} + d_1 \le N_1, \tag{92}$$
$$d_{\eta_2} + d_2 \le N_2. \tag{93}$$

where we recall $d_{\eta_1} \triangleq N'_1 A_2$ and $d_{\eta_2} \triangleq N'_2 A_1$.

Proof: It has been shown that with (86) and (90), only (82) and (84) are active. With ξ_k defined in (91), we can verify that $M_k - N'_j - \xi_k = \min\left\{M_k - N'_j, N_k - N'_k\right\}$. Thus, from (86), (90), (91), and the last constraint in (87), it follows that (82) always holds. Finally, the only constraint that remains is (84) that can be equivalently written as (92) and (93). ∎

Definition 5 (achievable region for IC)**.** *Let us define*

$$\mathcal{I}_A^{\rm IC} \triangleq \left\{ (A_1, A'_1, A''_1, A_2, A'_2, A''_2) \in \mathcal{A}_{\rm IC} \;\middle|\; \begin{array}{l} (d_1, d_2) = f_{A\text{-}d}(A_1, A'_1, A''_1, A_2, A'_2, A''_2), \\ \frac{d_k}{N'_k} \le \frac{N_k}{N'_k} - A_j, \quad k \ne j \in \{1,2\} \end{array} \right\} \tag{94}$$

and the achievable DoF region of the proposed scheme

$$\mathcal{I}_d^{\rm IC} \triangleq f_{A\text{-}d}(\mathcal{I}_A^{\rm IC}) \triangleq \left\{ (d_1, d_2) \;\middle|\; \begin{array}{l} (d_1, d_2) = f_{A\text{-}d}(A_1, A'_1, A''_1, A_2, A'_2, A''_2), \\ (A_1, A'_1, A''_1, A_2, A'_2, A''_2) \in \mathcal{A}_{\rm IC}, \\ \frac{d_k}{N'_k} \le \frac{N_k}{N'_k} - A_j, \quad k \ne j \in \{1,2\} \end{array} \right\}. \tag{95}$$

<u>Achievability analysis</u>

The analysis is similar to the BC case, i.e., it is sufficient to find a function $f_{d\text{-}A} : \mathcal{O}_d^{\rm IC} \to \mathcal{A}_{\rm IC}$ where $\mathcal{O}_d^{\rm IC}$ denotes the outer bound region defined by (12), such that

$$(d_1, d_2) = f_{A\text{-}d}(f_{d\text{-}A}(d_1, d_2)), \quad \text{and} \tag{96}$$
$$f_{d\text{-}A}(d_1, d_2) \in \mathcal{I}_A^{\rm IC}. \tag{97}$$

Now, we define formally the power allocation function.

Definition 6 (power allocation for IC)**.** *Let us define γ_k, $k \ne j \in \{1,2\}$, as*

$$\gamma_k \triangleq \min\left\{1, \frac{M_j - d_j}{\xi_k}\right\}. \tag{98}$$

Then, we define $f_{d\text{-}A} \;:\; \mathcal{O}_d^{\rm IC} \to \mathcal{A}_{\rm IC}$:

$$(d_1, d_2) \mapsto \begin{array}{l} (A_1, A'_1, A''_1) \triangleq f_1(d_1, d_2), \\ (A_2, A'_2, A''_2) \triangleq f_2(d_1, d_2) \end{array} \tag{99}$$

where f_k, $k \ne j \in \{1,2\}$, such that (90) *is satisfied, and that*

- *when $M_k = N'_j$: $A'' = A'_k = A_k = \frac{d_k}{M_k}$;*
- *when $M_k > N'_j$, $d_k < (M_k - N'_j)\gamma_k + N'_j(\gamma_k - \alpha_j)^+$:*

$$A_k = (A'_k - \alpha_j)^+, \quad A'_k = A''_k < \gamma_k; \tag{100}$$

- *when $M_k > N'_j$, $d_k \ge (M_k - N'_j)\gamma_k + N'_j(\gamma_k - \alpha_j)^+$, and $\gamma_k < 1$:*

$$A_k = (A'_k - \alpha_j)^+, \quad A'_k > A''_k = \gamma_k; \tag{101}$$

- *when $M_k > N'_j$, $d_k \ge (M_k - N'_j)\gamma_k + N'_j(\gamma_k - \alpha_j)^+$, and $\gamma_k = 1$:*

$$A'_k = A''_k = 1. \tag{102}$$

First, one can verify, with some basic manipulations that, $f_{d\text{-}A}(\mathcal{O}_d^{\rm IC}) \subseteq \mathcal{A}_{\rm IC}$. Second, (96) is satisfied by construction. Finally, it remains to show that (97) holds as well, i.e., the decodability condition in (94) is satisfied. Since the region $\mathcal{O}_d^{\rm IC}$ depends on whether the condition C_k holds, we prove the achievability accordingly.

1) Neither C_1 nor C_2 holds ($\xi_1 = \xi_2 = 0$): For any $(d_1, d_2) \in \mathcal{O}_d^{\mathrm{IC}}$, we can define $(A_1, A_1', A_1'', A_2, A_2', A_2'') \triangleq f_{d\text{-}A}(d_1, d_2)$ which implies, in this case,

$$d_j = N_k' A_j + (M_j - N_k')A_j', \quad j \neq k \in \{1,2\}. \tag{103}$$

Applying this equality on the constraints in the outer bound $\mathcal{O}_d^{\mathrm{IC}}$ in (12), we have

$$\frac{d_k}{N_k'} \le \frac{\min\{\max\{M_1, N_2\}, \max\{M_2, N_1\}\}}{N_k'} - \frac{(M_j - N_k')A_j'}{N_k'} - A_j, \tag{104}$$

$$\frac{d_k}{N_k'} \le \frac{\min\{M_k, N_k\}}{N_k'} - \left[\frac{\min\{M_k, N_k\} - N_k}{N_k'} + \frac{(M_j - N_k')(A_j' - \alpha_k) + N_k' A_j}{M_j}\right]^+, \tag{105}$$

for $k \neq j \in \{1,2\}$, where the first one is from the sum rate constraint (12c) whereas the second one is from the rest of the constraints in (12). The final step is to show that either of (104) and (105) implies the last constraint in (94).

- When $M_j = N_k'$, (105) implies the last constraint in (94) because $\frac{\min\{M_k, N_k\} - N_k}{N_k} \le 0$;
- When $M_j > N_k'$ and $d_j \ge M_j - N_k'\alpha_k$, we have $A_j' = 1$ according to the mapping $f_{d\text{-}A}$ defined in Definition 6, since $\gamma_j = 1$. Hence, the right hand side (RHS) of (104) is not greater than $\frac{N_k}{N_k'} - A_j$, which implies the last constraint in (94);
- When $M_j > N_k'$ and $d_j < M_j - N_k'\alpha_k$, we have $A_j = (A_j' - \alpha_k)^+$ according to Definition 6 with $\gamma_j = 1$. Since $\frac{\min\{M_k, N_k\} - N_k}{N_k} \le 0$, we can show that

$$\left[\frac{\min\{M_k, N_k\} - N_k}{N_k} + \frac{(M_j - N_k')(A_j' - \alpha_k) + N_k' A_j}{M_j}\right]^+ \ge \frac{\min\{M_k, N_k\} - N_k}{N_k} + A_j \tag{106}$$

  with which (105) implies the last constraint in (94).

2) C_k holds ($\xi_k > 0$, $\xi_j = 0$): In this case, it is readily shown, from (90) and (91), that

$$d_k = N_j A_k + (N_k - M_j)A_k' + \xi_k A_k'', \tag{107}$$

$$d_j = M_j A_j. \tag{108}$$

Applying the mapping $d_j = M_j A_j$ on (12c) results in

$$\frac{d_k}{N_k'} \le \frac{\min\{M_k, N_k\}}{N_k'} - A_j \tag{109}$$

that always implies $\frac{d_k}{N_k'} \le \frac{N_k}{N_k'} - A_j$. Due to the asymmetry, we also need to prove that $\frac{d_j}{N_j} \le 1 - A_k$. Therefore, the final step is to show that it can be implied by at least one of the constraints in (12), together with (107) and (108).

- When $d_k < (M_k - N_j)\gamma_k + N_j(\gamma_k - \alpha_j)^+$, we have $A_k' = A_k'' < \gamma_k$ according to (100). Therefore, $d_k = N_j A_k + (M_k - N_j)A_k'$, plugging which into (12e), we obtain

$$\frac{d_j}{N_j} \le \frac{\min\{M_j, N_j\}}{N_j} - \left[\frac{\min\{M_j, N_j\} - N_j}{N_j} + \frac{(M_k - N_j)(A_k' - \alpha_j) + N_j A_k}{M_k}\right]^+ \tag{110}$$

$$\le \frac{\min\{M_j, N_j\}}{N_j} - \left[\frac{\min\{M_j, N_j\} - N_j}{N_j} + A_k\right] \tag{111}$$

  where the $[\cdot]^+$ in (110) is from the single user bound (12b); the last inequality is due to $A_k = (A_k' - \alpha_j)^+$ and $\frac{\min\{M_j, N_j\} - N_j}{N_j} \le 0$.
- When $d_k \ge (M_k - N_j)\gamma_k + N_j(\gamma_k - \alpha_j)^+$, we have $A_k' \ge A_k'' = \gamma_k$ according to (101) and (102).
  - If $\gamma_k < 1$, then $A_k'' = \gamma_k = \frac{M_j - d_j}{\xi_k}$ and $d_k = (N_k - M_j)A_k' + M_j - d_j + N_j A_k$. Plugging the latter into (12f), we obtain

$$\frac{d_j}{N_j} \le \frac{\min\{M_j, N_j\}}{N_j} - \left[\frac{\min\{M_j, N_j\} - N_j}{N_j} + \frac{(N_k - M_j)(A_k' - \alpha_j) + N_j A_k}{N_k + N_j - M_j}\right]^+ \tag{112}$$

$$\le \frac{\min\{M_j, N_j\}}{N_j} - \left[\frac{\min\{M_j, N_j\} - N_j}{N_j} + A_k\right] \tag{113}$$

    where the $[\cdot]^+$ in (112) is from the single user bound (12b); the last inequality is due to $A_k = (A_k' - \alpha_j)^+$ and $\frac{\min\{M_j, N_j\} - N_j}{N_j} \le 0$.
  - If $\gamma_k = 1$, then $A_k' = A_k'' = 1$ and $d_k = M_k - N_j + N_j A_k$. Plugging the latter into (12c), we obtain

$$\frac{d_j}{N_j} \le \frac{\min\{M_k, N_k\} - M_k + N_j - N_j A_k}{N_j} \tag{114}$$

$$\le 1 - A_k. \tag{115}$$

Thus, the last constraint in (94) is shown in all cases. By now, we have proved the achievability through the existence of a proper power allocation function such that (96) and (97) are satisfied for every pair (d_1, d_2) in the outer bound.

VI. CONVERSE

To obtain the outer bounds, the following ingredients are essential:

- *Genie-aided* bounding techniques by providing side information of one receiver to the other one [5], [6];
- *Extremal inequality* to bound the weighted difference of conditional differential entropies [29], [30];
- *Ergodic capacity upper and lower bounds* for MIMO channels with channel uncertainty.

In the following, we first present the proof of outer bound (6d) for MIMO BC and (12d) for MIMO IC, referred to in this section as L_4. It should be noticed that both bounds share the same structure. Then, we give the proof of bound (12f) for the MIMO IC case, referred to in this section as L_6, when the condition C_1 holds.

A. Proof of Bound L_4

We first provide the outer bounds by employing the genie-aided techniques for BC and IC, respectively, reaching a similar formulation of the rate bounds. With the help of extremal inequalities, the weighted sum rates are further bounded. Finally, the bounds in terms of (α_1, α_2) are obtained by deriving novel ergodic capacity bounds for MIMO channels with channel uncertainty.

To obtain the outer bounds, we adopt a genie-aided upper bounding technique reminisced in [5], [6], by providing Receiver 2 the side information of Receiver 1's message W_1 as well as received signal $\boldsymbol{Y}_1^n$. For notational brevity, we define a virtual received signal as

$$\bar{\boldsymbol{y}}_i(t) \triangleq \begin{cases} \boldsymbol{H}_i(t)\boldsymbol{x}(t) + \boldsymbol{z}_i(t) & \text{for BC} \\ \boldsymbol{H}_{i2}(t)\boldsymbol{x}_2(t) + \boldsymbol{z}_i(t) & \text{for IC} \end{cases} \tag{116}$$

and we also define $\boldsymbol{X}^n \triangleq \{\boldsymbol{x}(t)\}_{t=1}^n$, $\boldsymbol{X}_i^n \triangleq \{\boldsymbol{x}_i(t)\}_{t=1}^n$, $\boldsymbol{Y}_i^k \triangleq \{\boldsymbol{y}_i(t)\}_{t=1}^k$, and $\bar{\boldsymbol{Y}}_i^k \triangleq \{\bar{\boldsymbol{y}}_i(t)\}_{t=1}^k$. Denote also $n\epsilon_n \triangleq 1 + nRP_e^{(n)}$ where ϵ_n tends to zero as $n \to \infty$ by the assumption that $\lim_{n\to\infty} P_e^{(n)} = 0$.

1) Broadcast Channel: The genie-aided model is a degraded BC $\boldsymbol{X}^n \to (\boldsymbol{Y}_1^n, \boldsymbol{Y}_2^n) \to \boldsymbol{Y}_1^n$, and therefore we bound the achievable rates by applying Fano's inequality as

$$\begin{aligned}
&n(R_1 - \epsilon_n)\\
&\le I(W_1; \boldsymbol{Y}_1^n | \mathcal{H}^n, \hat{\mathcal{H}}^n) &(117)\\
&= \sum_{t=1}^n I(W_1; \boldsymbol{y}_1(t) | \mathcal{H}^n, \hat{\mathcal{H}}^n, \boldsymbol{Y}_1^{t-1}) &(118)\\
&= \sum_{t=1}^n \Big(h(\boldsymbol{y}_1(t) | \mathcal{H}^n, \hat{\mathcal{H}}^n, \boldsymbol{Y}_1^{t-1}) \\
&\qquad - h(\boldsymbol{y}_1(t) | \mathcal{H}^n, \hat{\mathcal{H}}^n, \boldsymbol{Y}_1^{t-1}, W_1) \Big) &(119)\\
&\le \sum_{t=1}^n \left(h(\boldsymbol{y}_1(t) | \mathcal{H}(t)) - h(\boldsymbol{y}_1(t) | \mathcal{U}(t), \mathcal{H}(t)) \right) &(120)\\
&\le n N_1' \log P - \sum_{t=1}^n h(\bar{\boldsymbol{y}}_1(t) | \mathcal{U}(t), \mathcal{H}(t)) + n \cdot O(1) &(121)\\
&n(R_2 - \epsilon_n)\\
&\le I(W_2; \boldsymbol{Y}_1^n, \boldsymbol{Y}_2^n, W_1 | \mathcal{H}^n, \hat{\mathcal{H}}^n) &(122)\\
&= I(W_2; \boldsymbol{Y}_1^n, \boldsymbol{Y}_2^n | W_1, \mathcal{H}^n, \hat{\mathcal{H}}^n) &(123)\\
&= \sum_{t=1}^n I(W_2; \boldsymbol{y}_1(t), \boldsymbol{y}_2(t) | \mathcal{H}^n, \hat{\mathcal{H}}^n, \boldsymbol{Y}_1^{t-1}, \boldsymbol{Y}_2^{t-1}, W_1) &(124)\\
&\le \sum_{t=1}^n I(\boldsymbol{x}(t); \boldsymbol{y}_1(t), \boldsymbol{y}_2(t) | \mathcal{H}^n, \hat{\mathcal{H}}^n, \boldsymbol{Y}_1^{t-1}, \boldsymbol{Y}_2^{t-1}, W_1) &(125)\\
&= \sum_{t=1}^n \Big(h(\boldsymbol{y}_1(t), \boldsymbol{y}_2(t) | \mathcal{H}^n, \hat{\mathcal{H}}^n, \boldsymbol{Y}_1^{t-1}, \boldsymbol{Y}_2^{t-1}, W_1) &(126)\\
&\qquad - h(\boldsymbol{y}_1(t), \boldsymbol{y}_2(t) | \boldsymbol{x}(t), \mathcal{H}^n, \hat{\mathcal{H}}^n, \boldsymbol{Y}_1^{t-1}, \boldsymbol{Y}_2^{t-1}, W_1) \Big) &(127)\\
&\le \sum_{t=1}^n h(\boldsymbol{y}_1(t), \boldsymbol{y}_2(t) | \mathcal{H}^n, \hat{\mathcal{H}}^n, \boldsymbol{Y}_1^{t-1}, \boldsymbol{Y}_2^{t-1}, W_1) &(128)\\
&= \sum_{t=1}^n h(\bar{\boldsymbol{y}}_1(t), \bar{\boldsymbol{y}}_2(t) | \mathcal{U}(t), \mathcal{H}(t)) &(129)
\end{aligned}$$

where $\mathcal{U}(t) \triangleq \left\{ \bar{\boldsymbol{Y}}_1^{t-1}, \bar{\boldsymbol{Y}}_2^{t-1}, \mathcal{H}^{t-1}, \hat{\mathcal{H}}^t, W_1 \right\}$ for BC and $N_1' \triangleq \min\{M, N_1\}$; (120) is from (116) and because (a) conditioning reduces differential entropy, and (b) $\{\bar{\boldsymbol{y}}_1(t), \bar{\boldsymbol{y}}_2(t)\}$ are independent of $\mathcal{H}_{t+1}^n$ and $\hat{\mathcal{H}}_{t+1}^n$, given the past states and channel outputs; (121) follows the fact that the rate of the point-to-point $M \times N_1$ MIMO channel (i.e., between the transmitter and Receiver 1) is bounded by $\min\{M, N_1\} \log P + O(1)$; (123) is due to the independence between W_1 and W_2; (125) follows date processing inequality; (128) is obtained by noticing (a) translation does not change differential entropy, (b) Gaussian noise terms are independent from instant to instant, and are also independent of the channel matrices and the transmitted signals, and (c) the differential entropy of Gaussian noise with normalized variance is non-negative and finite.

2) Interference Channel: Given the message and corresponding channel states, part of the received signal is deterministic and therefore removable without mutual information loss. Hence, similarly to the BC case, we formulate a degraded channel model, i.e., $\boldsymbol{X}_2^n \to \left(\bar{\boldsymbol{Y}}_1^n, \bar{\boldsymbol{Y}}_2^n\right) \to \bar{\boldsymbol{Y}}_1^n$. By applying Fano's inequality, the achievable rate of Receiver 1 and Receiver 2 can be bounded as

$$\begin{aligned}
&n(R_1 - \epsilon_n)\\
&\le I(W_1; \boldsymbol{Y}_1^n | \mathcal{H}^n, \hat{\mathcal{H}}^n) &(130)\\
&= I(W_1, W_2; \boldsymbol{Y}_1^n | \mathcal{H}^n, \hat{\mathcal{H}}^n) - I(W_2; \boldsymbol{Y}_1^n | W_1, \mathcal{H}^n, \hat{\mathcal{H}}^n) &(131)\\
&\le n\tilde{N}_1 \log P - I(W_2; \boldsymbol{Y}_1^n | W_1, \mathcal{H}^n, \hat{\mathcal{H}}^n) + n \cdot O(1) &(132)\\
&= n\tilde{N}_1 \log P - h(\boldsymbol{Y}_1^n | W_1, \mathcal{H}^n, \hat{\mathcal{H}}^n) \\
&\qquad + h(\boldsymbol{Y}_1^n | W_1, W_2, \mathcal{H}^n, \hat{\mathcal{H}}^n) + n \cdot O(1) &(133)\\
&= n\tilde{N}_1 \log P - h(\boldsymbol{Y}_1^n | W_1, \mathcal{H}^n, \hat{\mathcal{H}}^n) + n \cdot O(1) &(134)\\
&= n\tilde{N}_1 \log P - h(\bar{\boldsymbol{Y}}_1^n | \mathcal{H}^n, \hat{\mathcal{H}}^n) + n \cdot O(1) &(135)\\
&\le n\tilde{N}_1 \log P - \sum_{t=1}^n h(\bar{\boldsymbol{y}}_1(t) | \mathcal{H}^n, \hat{\mathcal{H}}^n, \bar{\boldsymbol{Y}}_1^{t-1}, \bar{\boldsymbol{Y}}_2^{t-1}) \\
&\qquad + n \cdot O(1) &(136)\\
&= n\tilde{N}_1 \log P - \sum_{t=1}^n h(\bar{\boldsymbol{y}}_1(t) | \mathcal{U}(t), \mathcal{H}(t)) + n \cdot O(1) &(137)\\
&n(R_2 - \epsilon_n)\\
&\le I(W_2; \boldsymbol{Y}_1^n, \boldsymbol{Y}_2^n, W_1 | \mathcal{H}^n, \hat{\mathcal{H}}^n) &(138)\\
&= I(W_2; \boldsymbol{Y}_1^n, \boldsymbol{Y}_2^n | W_1, \mathcal{H}^n, \hat{\mathcal{H}}^n) &(139)\\
&= I(W_2; \bar{\boldsymbol{Y}}_1^n, \bar{\boldsymbol{Y}}_2^n | \mathcal{H}^n, \hat{\mathcal{H}}^n) &(140)\\
&= \sum_{t=1}^n I(W_2; \bar{\boldsymbol{y}}_1(t), \bar{\boldsymbol{y}}_2(t) | \mathcal{H}^n, \hat{\mathcal{H}}^n, \bar{\boldsymbol{Y}}_1^{t-1}, \bar{\boldsymbol{Y}}_2^{t-1}) &(141)\\
&\le \sum_{t=1}^n I(\boldsymbol{x}_2(t); \bar{\boldsymbol{y}}_1(t), \bar{\boldsymbol{y}}_2(t) | \mathcal{H}^n, \hat{\mathcal{H}}^n, \bar{\boldsymbol{Y}}_1^{t-1}, \bar{\boldsymbol{Y}}_2^{t-1}) &(142)\\
&= \sum_{t=1}^n \Big(h(\bar{\boldsymbol{y}}_1(t), \bar{\boldsymbol{y}}_2(t) | \mathcal{H}^n, \hat{\mathcal{H}}^n, \bar{\boldsymbol{Y}}_1^{t-1}, \bar{\boldsymbol{Y}}_2^{t-1}) &(143)
\end{aligned}$$

$$- h(\bar{\boldsymbol{y}}_1(t), \bar{\boldsymbol{y}}_2(t)|\boldsymbol{x}_2(t), \mathcal{H}^n, \hat{\mathcal{H}}^n, \bar{\boldsymbol{Y}}_1^{t-1}, \bar{\boldsymbol{Y}}_2^{t-1})\Big) \quad (144)$$

$$\leq \sum_{t=1}^{n} h(\bar{\boldsymbol{y}}_1(t), \bar{\boldsymbol{y}}_2(t)|\mathcal{H}^n, \hat{\mathcal{H}}^n, \bar{\boldsymbol{Y}}_1^{t-1}, \bar{\boldsymbol{Y}}_2^{t-1}) \quad (145)$$

$$= \sum_{t=1}^{n} h(\bar{\boldsymbol{y}}_1(t), \bar{\boldsymbol{y}}_2(t)|\mathcal{U}(t), \mathcal{H}(t)) \quad (146)$$

where we define $\mathcal{U}(t) \triangleq \left\{\bar{\boldsymbol{Y}}_1^{t-1}, \bar{\boldsymbol{Y}}_2^{t-1}, \mathcal{H}^{t-1}, \hat{\mathcal{H}}^t\right\}$ for IC and $\tilde{N}_1 \triangleq \min\{M_1+M_2, N_1\}$; (132) follows the fact that the mutual information at hand is upper bounded by the rate of the $(M_1+M_2)\times N_1$ point-to-point MIMO channel created by letting the two transmitters cooperate perfectly, given by $\min\{M_1+M_2,N_1\}\log P + O(1)$; (134) is due to the fact that (a) transmitted signal $\boldsymbol{X}_i^n$ is a deterministic function of messages W_i, $\mathcal{H}^n$, and $\hat{\mathcal{H}}^{n-1}$ as specified in (8) for $i=1,2$, (b) translation does change differential entropy, and (c) the differential entropy of Gaussian noise with normalized variance is non-negative and finite; (135) and (140) are obtained because translation preserves differential entropy; (136) is because conditioning reduces differential entropy; (145) is because (a) translation does not change differential entropy, (b) Gaussian noise terms are independent from instant to instant, and are also independent of the channel matrices and the transmitted signals, and (c) the differential entropy of Gaussian noise with normalized variance is non-negative and finite; (146) is obtained due to the independence $\{\bar{\boldsymbol{y}}_1(t), \bar{\boldsymbol{y}}_2(t)\}$ of $\mathcal{H}_{t+1}^n$ and $\hat{\mathcal{H}}_{t+1}^n$, given the past state and channel outputs.

It is worth noting that BC and IC share the common structure of the achievable rate bounds, and therefore can be further bounded in a similar way. To avoid redundancy, we give the proof for IC, which can be straightforwardly extended to BC.

Define

$$\boldsymbol{S}(t) \triangleq \begin{bmatrix}\boldsymbol{H}_{12}(t)\\ \boldsymbol{H}_{22}(t)\end{bmatrix}, \quad \hat{\boldsymbol{S}}(t) \triangleq \begin{bmatrix}\hat{\boldsymbol{H}}_{12}(t)\\ \hat{\boldsymbol{H}}_{22}(t)\end{bmatrix}, \quad \boldsymbol{K}(t) \triangleq \mathbb{E}\{\boldsymbol{x}_2(t)\boldsymbol{x}_2^{\mathsf{H}}(t)\,|\,\mathcal{U}(t)\}. \quad (147)$$

Let $p=\min\{M_2, N_1+N_2\}$ and $q=\min\{M_2,N_1\}$. By following the footsteps in [14], we have

$$\frac{1}{p}h(\bar{\boldsymbol{y}}_1(t), \bar{\boldsymbol{y}}_2(t)|\mathcal{U}(t), \mathcal{H}(t)) - \frac{1}{q}h(\bar{\boldsymbol{y}}_1(t)|\mathcal{U}(t), \mathcal{H}(t)) \quad (148)$$

$$\leq \mathbb{E}_{\hat{\boldsymbol{S}}(t)} \max_{\substack{\boldsymbol{K}\succeq 0,\\ \mathrm{tr}(\boldsymbol{K})\leq P}} \mathbb{E}_{\boldsymbol{S}(t)|\hat{\boldsymbol{S}}(t)} \Big(\frac{1}{p}\log\det(\mathbf{I}+\boldsymbol{S}(t)\boldsymbol{K}(t)\boldsymbol{S}^{\mathsf{H}}(t)) - \frac{1}{q}\log\det(\mathbf{I}+\boldsymbol{H}_{12}(t)\boldsymbol{K}(t)\boldsymbol{H}_{12}^{\mathsf{H}}(t))\Big) \quad (149)$$

$$\leq -\frac{\min\{M_2,N_1+N_2\}-\min\{M_2,N_1\}}{\min\{M_2,N_1+N_2\}}\log\sigma_1^2 + O(1) \quad (150)$$

where (149) is obtained by applying extremal inequality [29], [30] for degraded outputs; the last inequality is obtained from the following lemma:

Lemma 3. *For two random matrices $\boldsymbol{S}=\hat{\boldsymbol{S}}+\tilde{\boldsymbol{S}}\in\mathbb{C}^{L\times M}$ and $\boldsymbol{H}=\hat{\boldsymbol{H}}+\tilde{\boldsymbol{H}}\in\mathbb{C}^{N\times M}$ with $L\geq N$, $\tilde{\boldsymbol{S}}$, $\tilde{\boldsymbol{H}}$ are respectively independent of $\hat{\boldsymbol{S}}$, $\hat{\boldsymbol{H}}$, and the entries of $\tilde{\boldsymbol{H}}$ are i.i.d. $\mathcal{N}_{\mathbb{C}}(0,\sigma^2)$. Then, given any $\boldsymbol{K}\succeq 0$ with eigenvalues $\lambda_1\geq\cdots\geq\lambda_M\geq 0$, we have*

$$\frac{1}{\min\{M,L\}}\mathbb{E}_{\tilde{\boldsymbol{S}}}\log\det(\mathbf{I}+\boldsymbol{S}\boldsymbol{K}\boldsymbol{S}^{\mathsf{H}}) - \frac{1}{\min\{M,N\}}\mathbb{E}_{\tilde{\boldsymbol{H}}}\log\det(\mathbf{I}+\boldsymbol{H}\boldsymbol{K}\boldsymbol{H}^{\mathsf{H}}) \leq -\frac{\min\{M,L\}-\min\{M,N\}}{\min\{M,L\}}\log(\sigma^2) + O_{\hat{S}}(1) + O_{\hat{H}}(1) \quad (151)$$

as σ^2 goes to 0.

Proof: See Appendix B. ■

Remark:

- This lemma can be regarded as the weighted difference of the ergodic capacity for two MIMO channels with uncertainty, where $\tilde{\boldsymbol{S}}$ and $\tilde{\boldsymbol{H}}$ are channel uncertainties. It can also be interpreted as the ergodic capacity difference of two Ricean MIMO channels with line-of-sight components $\hat{\boldsymbol{S}}$, $\hat{\boldsymbol{H}}$, and fading components $\tilde{\boldsymbol{S}}$, $\tilde{\boldsymbol{H}}$.
- This lemma also shows the change of the ergodic capacity per dimension as the dimensionality decreases. In other words, as the channel dimension decreases, the difference of the ergodic capacity per dimension is bounded by the dimension difference and the channel uncertainty.

According to the Markov chain $\boldsymbol{X}_2^n \to \left(\bar{\boldsymbol{Y}}_1^n, \bar{\boldsymbol{Y}}_2^n\right) \to \bar{\boldsymbol{Y}}_1^n$, we upper-bound the weighted sum rate as

$$n\left(\frac{R_1}{\min\{M_2,N_1\}} + \frac{R_2}{\min\{M_2,N_1+N_2\}} - \epsilon_n\right) \quad (152)$$

$$\leq n\cdot\frac{\min\{M_1+M_2,N_1\}}{\min\{M_2,N_1\}}\log P + \sum_{t=1}^{n}\Bigg(\frac{1}{\min\{M_2,N_1+N_2\}}h(\bar{\boldsymbol{y}}_1(t), \bar{\boldsymbol{y}}_2(t)|\mathcal{U}(t), \mathcal{H}(t)) \quad (153)$$

$$- \frac{1}{\min\{M_2,N_1\}}h(\bar{\boldsymbol{y}}_1(t)|\mathcal{U}(t), \mathcal{H}(t))\Bigg) + n\cdot O(1) \quad (154)$$

$$\leq n\frac{\min\{M_1+M_2,N_1\}}{\min\{M_2,N_1\}}\log P + n\cdot O(1) + n\frac{\min\{M_2,N_1+N_2\}-\min\{M_2,N_1\}}{\min\{M_2,N_1+N_2\}}\alpha_1\log P \quad (155)$$

and another outer bound can be similarly obtained by exchanging the roles of Receiver 1 and Receiver 2. Accordingly, the corresponding outer bound L_4 of the DoF region is obtained by the definition.

B. Proof of Bound L_6

This bound is active when C_1 holds, i.e., $M_1\geq N_2$, $N_1 > M_2$, and $M_1+M_2 > N_1+N_2$. The proof follows the same lines of thought in [6]. Since $N_1>M_2$, we formulate a virtual received signal

$$\tilde{\boldsymbol{y}}_1(t) \triangleq \boldsymbol{U}\boldsymbol{y}_1(t) = \boldsymbol{U}\boldsymbol{H}_{11}(t)\boldsymbol{x}_1(t) + \boldsymbol{U}\boldsymbol{H}_{12}(t)\boldsymbol{x}_2(t) + \boldsymbol{U}\boldsymbol{z}_1(t) \quad (156)$$

where $\boldsymbol{U} \in \mathbb{C}^{N_1 \times N_1}$ is any unitary matrix such that the last $N_1 - M_2$ rows of $\boldsymbol{U}(t)\boldsymbol{H}_{12}(t)$ are with all zeros and is independent of the rest of random variables. Therefore, the last $N_1 - M_2$ outputs in $\tilde{\boldsymbol{y}}_1(t)$ are interference free, i.e., $\tilde{\boldsymbol{y}}_{1[M_2+1:N_1]}(t) \sim \boldsymbol{H}_{1[M_2+1:N_1]1}(t)\boldsymbol{x}_1(t)+\boldsymbol{z}_{1[M_2+1:N_1]}(t)$. For convenience, we also define

$$\tilde{\boldsymbol{y}}_2(t) \triangleq \boldsymbol{H}_{21}(t)\boldsymbol{x}_1(t) + \boldsymbol{z}_2(t). \tag{157}$$

Starting with Fano's inequality, the achievable rate can be bounded as

$$\begin{aligned}
&n(R_1 - \epsilon_n) \\
&\le I(W_1; \boldsymbol{Y}_1^n | \mathcal{H}^n, \hat{\mathcal{H}}^n) && (158)\\
&= I(W_1; \tilde{\boldsymbol{Y}}_1^n | \mathcal{H}^n, \hat{\mathcal{H}}^n) && (159)\\
&= I(W_1; \tilde{\boldsymbol{Y}}_{1[1:M_2]}^n | \mathcal{H}^n, \hat{\mathcal{H}}^n, \tilde{\boldsymbol{Y}}_{1[M_2+1:N_1]}^n) \\
&\qquad + I(W_1; \tilde{\boldsymbol{Y}}_{1[M_2+1:N_1]}^n | \mathcal{H}^n, \hat{\mathcal{H}}^n) && (160)\\
&\le n(M_2 - d_2)\log P + n \cdot O(1) \\
&\qquad + I(W_1; \tilde{\boldsymbol{Y}}_{1[M_2+1:N_1]}^n | \mathcal{H}^n, \hat{\mathcal{H}}^n) && (161)\\
&\le n(M_2 - d_2)\log P + n \cdot O(1) \\
&\qquad + I(W_1; \tilde{\boldsymbol{Y}}_{1[M_2+1:N_1]}^n, \tilde{\boldsymbol{Y}}_2^n | \mathcal{H}^n, \hat{\mathcal{H}}^n) && (162)\\
&= n(M_2 - d_2)\log P + n \cdot O(1) \\
&\qquad + \sum_{t=1}^n I\big(W_1; \tilde{\boldsymbol{y}}_{1[M_2+1:N_1]}(t), \tilde{\boldsymbol{y}}_2(t) | \mathcal{H}^n, \\
&\qquad\qquad \hat{\mathcal{H}}^n, \tilde{\boldsymbol{Y}}_{1[M_2+1:N_1]}^{t-1}, \tilde{\boldsymbol{Y}}_2^{t-1}\big) && (163)\\
&\le n(M_2 - d_2)\log P + n \cdot O(1) \\
&\qquad + \sum_{t=1}^n h(\tilde{\boldsymbol{y}}_{1[M_2+1:N_1]}(t), \tilde{\boldsymbol{y}}_2(t) | \mathcal{H}^n, \\
&\qquad\qquad \hat{\mathcal{H}}^n, \tilde{\boldsymbol{Y}}_{1[M_2+1:N_1]}^{t-1}, \tilde{\boldsymbol{Y}}_2^{t-1}) && (164)\\
&= n(M_2 - d_2)\log P + n \cdot O(1) \\
&\qquad + \sum_{t=1}^n h(\tilde{\boldsymbol{y}}_{1[M_2+1:N_1]}(t), \tilde{\boldsymbol{y}}_2(t) | \mathcal{U}(t), \mathcal{H}(t)) && (165)
\end{aligned}$$

$$\begin{aligned}
&n(R_2 - \epsilon_n) \\
&\le I(W_2; \boldsymbol{Y}_2^n | \mathcal{H}^n, \hat{\mathcal{H}}^n) && (166)\\
&= I(W_1, W_2; \boldsymbol{Y}_2^n | \mathcal{H}^n, \hat{\mathcal{H}}^n) - I(W_1; \boldsymbol{Y}_2^n | W_2, \mathcal{H}^n, \hat{\mathcal{H}}^n) && (167)\\
&\le nN_2 \log P - I(W_1; \tilde{\boldsymbol{Y}}_2^n | \mathcal{H}^n, \hat{\mathcal{H}}^n) + n \cdot O(1) && (168)\\
&\le nN_2 \log P - h(\tilde{\boldsymbol{Y}}_2^n | \mathcal{H}^n, \hat{\mathcal{H}}^n) \\
&\qquad + h(\tilde{\boldsymbol{Y}}_2^n | W_1, \mathcal{H}^n, \hat{\mathcal{H}}^n) + n \cdot O(1) && (169)\\
&= nN_2 \log P - h(\tilde{\boldsymbol{Y}}_2^n | \mathcal{H}^n, \hat{\mathcal{H}}^n) + n \cdot O(1) && (170)\\
&\le nN_2 \log P + n \cdot O(1) \\
&\qquad - \sum_{t=1}^n h(\tilde{\boldsymbol{y}}_2(t) | \mathcal{H}^n, \hat{\mathcal{H}}^n, \tilde{\boldsymbol{Y}}_{1[M_2+1:N_1]}^{t-1}, \tilde{\boldsymbol{Y}}_2^{t-1}) && (171)\\
&= nN_2 \log P - \sum_{t=1}^n h(\tilde{\boldsymbol{y}}_2(t) | \mathcal{U}(t), \mathcal{H}(t)) + n \cdot O(1) && (172)
\end{aligned}$$

where $\mathcal{U}(t) \triangleq \{\mathcal{H}^{t-1}, \hat{\mathcal{H}}^t, \tilde{\boldsymbol{Y}}_{1[M_2+1:N_1]}^{t-1}, \tilde{\boldsymbol{Y}}_2^{t-1}\}$ and $\tilde{\boldsymbol{Y}}_i^k \triangleq \{\tilde{\boldsymbol{y}}_i(t)\}_{t=1}^k$, $i = 1, 2$; (159) holds due to the fact that unitary transformation does not change the mutual information; (161) comes from Lemma 6 in [6], given by

$$\begin{aligned}
&I(W_1; \tilde{\boldsymbol{Y}}_{1[1:M_2]}^n | \mathcal{H}^n, \hat{\mathcal{H}}^n, \tilde{\boldsymbol{Y}}_{1[M_2+1:N_1]}^n) \\
&\qquad \le n(M_2 - d_2)\log P + n \cdot O(1)
\end{aligned} \tag{173}$$

where a similar proof can be straightforwardly obtained; (164) holds because (a) $\tilde{\boldsymbol{y}}_{1[M_2+1:N_1]}(t)$ and $\tilde{\boldsymbol{y}}_2(t)$ are deterministic functions of W_1, $\mathcal{H}^n$ and $\hat{\mathcal{H}}^n$, (b) translation does not change differential entropy, and (c) the differential entropy of Gaussian noise with normalized variance is non-negative and finite; (168) follows that the mutual information at hand is upper bounded by the capacity of an $(M_1 + M_2) \times N_2$ point-to-point MIMO channel, i.e., $N_2 \log P + O(1)$ since $M_1 + M_2 > N_2$ from the condition C_1; (170) holds because $\tilde{\boldsymbol{y}}_2(t)$ is a deterministic function of W_1, given channel states, and the differential entropy of the normalized Gaussian noise is finite; (171) is due to conditioning reduces the differential entropy; (165) and the last equality are due to that the received signals at instant t are independent of $\mathcal{H}_{t+1}^n$ and $\hat{\mathcal{H}}_{t+1}^n$, given the past states and channel outputs.

Next, we define

$$\boldsymbol{S}(t) \triangleq \begin{bmatrix} \boldsymbol{H}_{1[M_2+1:N_1]1}(t) \\ \boldsymbol{H}_{21}(t) \end{bmatrix} \in \mathbb{C}^{(N_1+N_2-M_2)\times M_1}. \tag{174}$$

Similarly to the proof for bound L_4, we obtain the weighted difference of two differential entropies by applying the extremal inequality and Lemma 3

$$\begin{aligned}
&\frac{1}{p} h(\tilde{\boldsymbol{y}}_{1[M_2+1:N_1]}(t), \tilde{\boldsymbol{y}}_2(t) | \mathcal{U}(t), \mathcal{H}(t)) - \frac{1}{q} h(\tilde{\boldsymbol{y}}_2(t) | \mathcal{U}(t), \mathcal{H}(t)) \\
&\qquad \le -\frac{N_1 - M_2}{N_1 + N_2 - M_2} \log \sigma_2^2 + O(1)
\end{aligned} \tag{175}$$

where we set $p = \min\{M_1, N_1 + N_2 - M_2\} = N_1 + N_2 - M_2$ and $q = \min\{M_1, N_2\} = N_2$.

Finally, we have

$$\begin{aligned}
&n\left(\frac{R_1}{N_1 + N_2 - M_2} + \frac{R_2}{N_2} - \epsilon_n\right) && (176)\\
&\le n\left(1 + \frac{M_2 - d_2}{N_1 + N_2 - M_2}\right)\log P \\
&\quad + \sum_{t=1}^n \left(\frac{1}{N_1 + N_2 - M_2} h(\tilde{\boldsymbol{y}}_{1[M_2+1:N_1]}(t), \tilde{\boldsymbol{y}}_2(t) | \mathcal{U}(t), \mathcal{H}(t))\right. && (177)\\
&\qquad\qquad \left. - \frac{1}{N_2} h(\tilde{\boldsymbol{y}}_2(t) | \mathcal{U}(t), \mathcal{H}(t))\right) + n \cdot O(1) && (178)\\
&\le n\left(1 + \frac{M_2 - d_2}{N_1 + N_2 - M_2}\right)\log P \\
&\qquad + n\frac{N_1 - M_2}{N_1 + N_2 - M_2}\alpha_2 \log P + n \cdot O(1) && (179)
\end{aligned}$$

which leads to

$$d_1 + \frac{N_1 + 2N_2 - M_2}{N_2} d_2 \le N_1 + N_2 + (N_1 - M_2)\alpha_2. \tag{180}$$

By exchanging the roles of Receiver 1 and Receiver 2, the outer bound (12g) can be obtained straightforwardly when the condition C_2 holds.

VII. Conclusion

In this work, we focus on the two-user MIMO broadcast and interference channels where the transmitter(s) has/have access to both delayed CSIT and an estimate of current CSIT. Specifically, the DoF region of MIMO networks (BC/IC) in this setting with general antenna configuration and general current CSIT qualities has been fully characterized, thanks to a simple yet unified framework employing interference quantization, block-Markov encoding and backward decoding techniques. Our DoF regions generalize a number of existing results under more specific CSIT settings, such as delayed CSIT [5], [6], perfect CSIT [1], [2], partial/hybrid/mixed CSIT [24]–[26]. The results further shed light on the benefits of the temporally correlated channel, when such correlation can be opportunistically taken into account for system designs.

Appendix

A. Achievable rate regions for the related MAC channels

1) Broadcast Channels: Let us focus on Receiver k, $k \neq j \in \{1,2\}$, without loss of generality. The channel in (45) is a MAC, rewritten as

$$\underbrace{\begin{bmatrix} \boldsymbol{y}_k[b] - \boldsymbol{\eta}_{kb} \\ \boldsymbol{\eta}_{jb} \end{bmatrix}}_{Y_k} = \underbrace{\begin{bmatrix} \boldsymbol{H}_k \\ 0 \end{bmatrix}}_{S_1} X_c + \underbrace{\begin{bmatrix} \boldsymbol{H}_k \\ \boldsymbol{H}_j \end{bmatrix}}_{S_2} X_k + Z_k \tag{181}$$

where $X_c \triangleq \boldsymbol{x}_c(l_{b-1})$ and $X_k \triangleq \boldsymbol{u}_k(w_{kb})$ are independent, with rate R_c and R_k, respectively; Z_k is the AWGN. It is well known [28] that a rate pair (R_c, R_k) is achievable in the channel if

$$R_c \leq I(X_c; Y_k \mid X_k, S) \tag{182}$$

$$R_k \leq I(X_k; Y_k \mid X_c, S) \tag{183}$$

$$R_c + R_k \leq I(X_c, X_k; Y_k \mid S) \tag{184}$$

for any input distribution $p_{X_c X_k} = p_{X_c} p_{X_k}$; $S \triangleq \{S_1, S_2\}$ denotes the state of the channel. Let $X_c \sim \mathcal{N}_{\mathbb{C}}(0, \boldsymbol{Q}_{\rm c})$ and $X_k \sim \mathcal{N}_{\mathbb{C}}(0, \boldsymbol{Q}_k)$ with $\boldsymbol{Q}_{\rm c} \triangleq P\mathbf{I}_M$ and $\boldsymbol{Q}_k \triangleq P^{A'_k} \boldsymbol{\Phi}_{\hat{H}_j^{\perp}} + P^{A_k} \boldsymbol{\Phi}_{\hat{H}_j}$. It readily follows that[2]

$$\begin{aligned} I(X_c; Y_k \mid X_k) &= \log\det(\mathbf{I} + P\boldsymbol{S}_1 \boldsymbol{S}_1^{\mathsf{H}}) \\ &= N_k \log P + O(1) \end{aligned} \tag{185}$$

$$\begin{aligned} I(X_k; Y_k \mid X_c) &= \log\det(\mathbf{I} + \boldsymbol{S}_2 \boldsymbol{Q}_k \boldsymbol{S}_2^{\mathsf{H}}) \\ &= ((M - N_j)A'_k + N_j A_k) \log P + O(1) \end{aligned} \tag{186}$$

since $\boldsymbol{S}_2 \in \mathbb{C}^{(N_1+N_2)\times M}$ has rank M almost surely, given the assumption $N_1 + N_2 \geq M$. For the sum rate constraint, we have

$$\begin{aligned} & I(X_c, X_k; Y_k) \\ &\quad = h(Y_k) - h(Z_k) && (187) \\ &\quad = h(\boldsymbol{H}_j X_k + Z_{k2}) + O(1) \\ &\qquad + h(\boldsymbol{H}_k(X_c + X_k) + Z_{k1} \mid \boldsymbol{H}_j X_k + Z_{k2}) && (188) \\ &\quad \geq h(\boldsymbol{H}_j X_k + Z_{k2}) + O(1) \\ &\qquad + h(\boldsymbol{H}_k(X_c + X_k) + Z_{k1} \mid \boldsymbol{H}_j X_k + Z_{k2}, X_k) && (189) \\ &\quad = h(\boldsymbol{H}_j X_k + Z_{k2}) + h(\boldsymbol{H}_k X_c + Z_{k1}) + O(1) && (190) \\ &\quad = N_j A_k \log P + N_k \log P + O(1) && (191) \end{aligned}$$

where we define Z_{k1} and Z_{k2} the first and second parts of the noise vector Z_k; the second equality is from the chain rule and the fact that the Gaussian noise Z_k is normalized; (189) is due to conditioning reduces differential entropy; (190) is from the independence between X_c and X_k and between the noises and the rest; the first term in (191) is essentially the differential entropy of the interference $\boldsymbol{\eta}_{jb}$. By relating the rate pair (R_c, R_k) to the DoF pair (d_η, d_{kb}), (46)-(48) is straightforward.

2) Interference Channels: In (78), each receiver sees a MAC with three independent messages. Let us focus on Receiver k, $k \neq j \in \{1,2\}$, without loss of generality. The channel in (78) is rewritten as

$$\underbrace{\begin{bmatrix} \boldsymbol{y}_k[b] - \boldsymbol{\eta}_{kb} \\ \boldsymbol{\eta}_{jb} \end{bmatrix}}_{Y_k} = \underbrace{\begin{bmatrix} \boldsymbol{H}_{kk} \\ 0 \end{bmatrix}}_{S_{k1}} X_{kc} + \underbrace{\begin{bmatrix} \boldsymbol{H}_{kj} \\ 0 \end{bmatrix}}_{S_{k2}} X_{jc} + \underbrace{\begin{bmatrix} \boldsymbol{H}_{kk} \\ \boldsymbol{H}_{jk} \end{bmatrix}}_{S_{k3}} X_k + Z_k \tag{192}$$

where $X_{kc} \triangleq \boldsymbol{x}_{kc}(l_{k,b-1})$, $X_{jc} \triangleq \boldsymbol{x}_{jc}(l_{j,b-1})$, and $X_k \triangleq \boldsymbol{u}_k(w_{kb})$, $k \neq j \in \{1,2\}$, are three independent signals, with rate R_{kc}, R_{jc}, and R_k, respectively; Z_k is the AWGN. It is well known [28] that a rate triple (R_{kc}, R_{jc}, R_k) is achievable in the channel if

$$R_{kc} \leq I(X_{kc}; Y_k \mid X_{jc}, X_k) \tag{193}$$

$$R_{jc} \leq I(X_{jc}; Y_k \mid X_{kc}, X_k) \tag{194}$$

$$R_k \leq I(X_k; Y_k \mid X_{kc}, X_{jc}) \tag{195}$$

$$R_{kc} + R_{jc} \leq I(X_{kc}, X_{jc}; Y_k \mid X_k) \tag{196}$$

$$R_{kc} + R_k \leq I(X_{kc}, X_k; Y_k \mid X_{jc}) \tag{197}$$

$$R_{jc} + R_k \leq I(X_{jc}, X_k; Y_k \mid X_{kc}) \tag{198}$$

$$R_{kc} + R_{jc} + R_k \leq I(X_{kc}, X_{jc}, X_k; Y_k) \tag{199}$$

for any $p_{X_{kc} X_{jc} X_k} = p_{X_{kc}} p_{X_{jc}} p_{X_k}$, where we omit the conditioning on the channel states S as in the BC case for brevity. Let $X_{kc} \sim \mathcal{N}_{\mathbb{C}}(0, \boldsymbol{Q}_{k\rm c})$ and $X_k \sim \mathcal{N}_{\mathbb{C}}(0, \boldsymbol{Q}_k)$ with $\boldsymbol{Q}_{k\rm c} \triangleq P\mathbf{I}_{M_k}$ and $\boldsymbol{Q}_k \triangleq P^{A_k} \boldsymbol{\Phi}_{\hat{H}_{jk}} + P^{A'_k} \boldsymbol{\Phi}_{\hat{H}_{jk}^{\perp 1}} + P^{A''_k} \boldsymbol{\Phi}_{\hat{H}_{jk}^{\perp 2}}$. It is readily shown that

$$\begin{aligned} I(X_{kc}; Y_k \mid X_{jc}, X_k) &= \log\det(\mathbf{I} + P\boldsymbol{S}_{k1} \boldsymbol{S}_{k1}^{\mathsf{H}}) \\ &= \min\{M_k, N_k\} \log P + O(1) \end{aligned} \tag{200}$$

$$\begin{aligned} I(X_{jc}; Y_k \mid X_{kc}, X_k) &= \log\det(\mathbf{I} + P\boldsymbol{S}_{k2} \boldsymbol{S}_{k2}^{\mathsf{H}}) \\ &= \min\{M_j, N_k\} \log P + O(1) \end{aligned} \tag{201}$$

$$\begin{aligned} I(X_k; Y_k \mid X_{kc}, X_{jc}) &= \log\det(\mathbf{I} + \boldsymbol{S}_{k3} \boldsymbol{Q}_k \boldsymbol{S}_{k3}^{\mathsf{H}}) \\ &= (N'_j A_k + (M_k - N'_j - \xi_k) A'_k \\ &\quad + \xi_k A''_k) \log P + O(1) \end{aligned} \tag{202}$$

$$\begin{aligned} I(X_{jc}, X_{kc}; Y_k \mid X_k) &= \log\det(\mathbf{I} + P\boldsymbol{S}_{k1} \boldsymbol{S}_{k1}^{\mathsf{H}} + P\boldsymbol{S}_{k2} \boldsymbol{S}_{k2}^{\mathsf{H}}) \\ &= N_k \log P + O(1) \end{aligned} \tag{203}$$

since $\boldsymbol{S}_{k3} \in \mathbb{C}^{(N_1+N_2)\times M_k}$ has rank M_k almost surely, given the assumption $N_1 + N_2 \geq M_k$. Following the same steps as

[2]Hereafter, we omit for notational brevity the expectation on the channel states S, whenever possible, which does not change the high SNR behavior in this case. We consider any realization $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ instead.

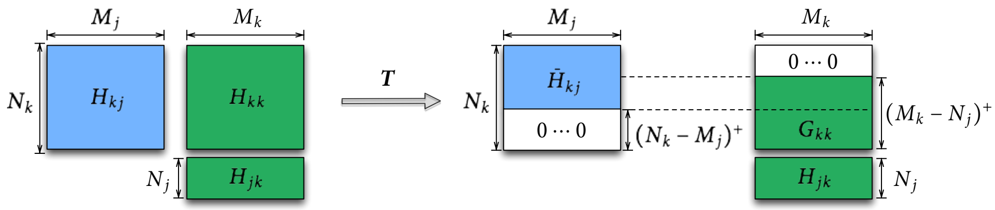


Fig. 2: Visualization of the interplay between X_{jc} and X_k.

(187)-(191), we can obtain

$$
\begin{aligned}
&I(X_{kc}, X_k; Y_k \mid X_{jc}) \\
&\qquad \geq N'_j A_k \log P + \min\{M_k, N_k\} \log P + O(1). \qquad (204)
\end{aligned}
$$

It remains to bound the RHS of (198) and (199). First, using the chain rule, we have

$$
\begin{aligned}
&I(X_{jc}, X_k; Y_k \mid X_{kc}) \\
&\qquad = I(X_k; Y_k \mid X_{kc}) + I(X_{jc}; Y_k \mid X_k, X_{kc}) \qquad (205)
\end{aligned}
$$

where the scaling of the second term is already shown in (201). The first term can be interpreted as the rate of X_k by treating X_{jc} as noise in a two-user MAC with a channel matrix in the block upper triangular form $\left[\begin{smallmatrix} \boldsymbol{H}_{kj} & \boldsymbol{H}_{kk} \\ & \boldsymbol{H}_{jk} \end{smallmatrix}\right]$. As shown in Fig. 2, since $\boldsymbol{H}_{kj}$, $\boldsymbol{H}_{kk}$, and $\boldsymbol{H}_{jk}$ are mutually independent, there exists an invertible row transformation $\boldsymbol{T}$ that can convert the $(N_1+N_2)\times(M_1+M_2)$ matrix to the form on the right, almost surely. The interference created by X_{jc} is through the matrix $\bar{\boldsymbol{H}}_{kj}$, only affecting the overlapping part between X_{jc} and X_k, as shown in Fig. 2. Note that the dimension of the overlapping is $((M_k - N_j)^+ - (N_k - M_j)^+)^+$ that coincides with the definition of ξ_k in (91). From Fig. 2, the interference-free received signal for X_k is $\tilde{Y}_k = \left[\begin{smallmatrix} \boldsymbol{G}_{kk} \\ \boldsymbol{H}_{jk} \end{smallmatrix}\right] X_k + \tilde{Z}_k$. Thus,

$$
\begin{aligned}
I(X_k; Y_k \mid X_{kc}) &\geq I(X_k; \tilde{Y}_k) & (206)\\
&= \log\det\left(\mathbf{I} + \begin{bmatrix} \boldsymbol{G}_{kk} \\ \boldsymbol{H}_{jk} \end{bmatrix} \boldsymbol{Q}_k \begin{bmatrix} \boldsymbol{G}_{kk} \\ \boldsymbol{H}_{jk} \end{bmatrix}^{\mathsf{H}}\right) + O(1) & (207)\\
&\geq \log\det\left(\mathbf{I} + \begin{bmatrix} \tilde{\boldsymbol{G}}'_{kk} & \tilde{\boldsymbol{G}}_{kk} \\ \tilde{\boldsymbol{H}}'_{kk} & \tilde{\boldsymbol{H}}_{kk} \end{bmatrix} \begin{bmatrix} P^{A'_k}\mathbf{I}_{M_k - N'_j - \xi_k} & \\ & P^{A_k}\mathbf{I}_{N'_j} \end{bmatrix}\right. \\
&\qquad \left. \cdot \begin{bmatrix} \tilde{\boldsymbol{G}}'_{kk} & \tilde{\boldsymbol{G}}_{kk} \\ \tilde{\boldsymbol{H}}'_{kk} & \tilde{\boldsymbol{H}}_{kk} \end{bmatrix}^{\mathsf{H}}\right) + O(1) & (208)\\
&= ((M_k - N'_j - \xi_k)A'_k + N'_j A_k) \log P + O(1) & (209)
\end{aligned}
$$

where the $O(1)$ term in (207) is from the fact that the covariance of the noise $\tilde{Z}_k$ depends on $\boldsymbol{T}$ that does not scale with P; $\boldsymbol{G}_{kk}$ and $\boldsymbol{H}_{jk}$ remain independent. Next, let $\boldsymbol{Q}_k = \boldsymbol{U}_{jk} \operatorname{diag}(P^{A''_k}\mathbf{I}_{\xi_k}, P^{A'_k}\mathbf{I}_{M_k - N'_j - \xi_k}, P^{A_k}\mathbf{I}_{N'_j}) \boldsymbol{U}_{jk}^{\mathsf{H}}$ be the eigenvalue decomposition of $\boldsymbol{Q}_k$ and define the column partitions $[\tilde{\boldsymbol{G}}''_{kk}\ \tilde{\boldsymbol{G}}'_{kk}\ \tilde{\boldsymbol{G}}_{kk}] \triangleq \boldsymbol{G}_{kk}\boldsymbol{U}_{jk}$ and $[\tilde{\boldsymbol{H}}''_{jk}\ \tilde{\boldsymbol{H}}'_{jk}\ \tilde{\boldsymbol{H}}_{jk}] \triangleq \boldsymbol{H}_{jk}\boldsymbol{U}_{jk}$ where the number of columns of the submatrices is ξ_k, $M_k - N'_j - \xi_k$, and N'_j, respectively; inequality (208) is from the fact that removing one column block and the corresponding diagonal block of size ξ_k can only reduce the log-determinant; the last equality is from the fact that the square matrix $\left[\begin{smallmatrix} \tilde{\boldsymbol{G}}'_{kk} & \tilde{\boldsymbol{G}}_{kk} \\ \tilde{\boldsymbol{H}}'_{jk} & \tilde{\boldsymbol{H}}_{jk} \end{smallmatrix}\right]$ has full rank, almost surely, for the following reasons: 1) the matrices G and H are mutually independent since the column transform $\boldsymbol{U}_{jk}$ is unitary and independent of the G matrices; 2) the rows related to the matrices H are linearly independent, since it can be proved that $\operatorname{rank}(\tilde{\boldsymbol{H}}_{jk}) = \operatorname{rank}(\boldsymbol{H}_{jk}\boldsymbol{\Phi}_{\hat{H}_{jk}}\boldsymbol{H}_{jk}^{\mathsf{H}}) = \min\{M_k, N_j\}$, i.e., $\tilde{\boldsymbol{H}}_{jk}$ has full rank; 3) the rows related to the matrices G are linearly independent as well. Plugging (209) and (201) into (205), we have

$$
\begin{aligned}
&I(X_{jc}, X_k; Y_k \mid X_{kc}) \\
&\quad \geq (N'_k + (M_k - N'_j - \xi_k)A'_k + N'_j A_k) \log P + O(1). \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (210)
\end{aligned}
$$

Finally, for the sum rate constraint (199), we follow the same steps as (187)-(191), we can obtain

$$
\begin{aligned}
&I(X_{kc}, X_{jc}, X_k; Y_k) \\
&\qquad \geq N'_j A_k \log P + \min\{M_k + M_j, N_k\} \log P + O(1) \qquad (211)\\
&\qquad = (N_k + N'_j A_k) \log P + O(1) \qquad (212)
\end{aligned}
$$

By relating the rate pair (R_{kc}, R_{jc}, R_k) to the DoF pair $(d_{\eta_1}, d_{\eta_2}, d_{kb})$, (79)-(84) are straightforward.

B. *Proof of Lemma 3*

In order to prove Lemma 3, we provide the following preliminary results stated as Lemma 4-7.

Let $\boldsymbol{A} \in \mathbb{C}^{N\times M}$, $N \leq M$, be a full rank matrix and $\boldsymbol{A}' \in \mathbb{C}^{N\times M'}$, $M' \leq M$, be a submatrix of $\boldsymbol{A}$. We have the following lemmas that will be repeatedly used in the rest of the proof.

Lemma 4 (rank of submatrix)**.**

$$
\operatorname{rank}(\boldsymbol{A}') \geq \operatorname{rank}(\boldsymbol{A}) - (M - M'). \qquad (213)
$$

Lemma 5. *Let $\mathcal{I}_1, \ldots, \mathcal{I}_M$ be a cyclic sliding window of size N on the set of indices $\{1, \ldots, M\}$, i.e.,*

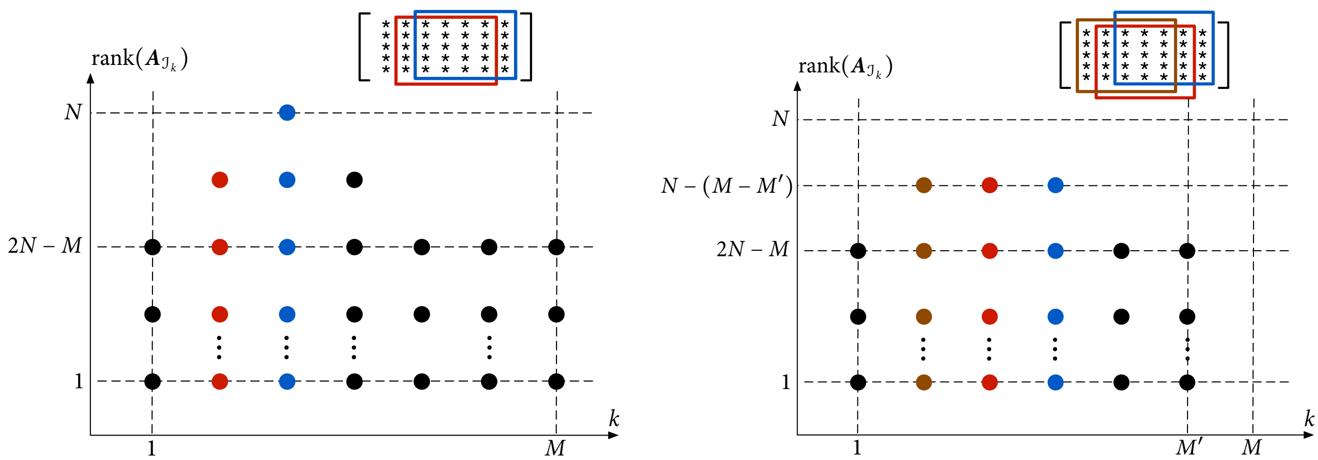
$$
\mathcal{I}_k \triangleq \{(k+i)_M + 1 \ : i \in [0, N-1]\}, \quad k = 1, \ldots, M. \qquad (214)
$$

If the columns of $\boldsymbol{A}$ are arranged such that $\operatorname{rank}(\boldsymbol{A}_{\mathcal{I}_k}) = N$ for some $k \in [1, M]$, then

$$
\sum_{k=1}^{M} \operatorname{rank}(\boldsymbol{A}_{\mathcal{I}_k}) \geq N^2 \qquad (215)
$$

where $\boldsymbol{A}_{\mathcal{I}_k}$ is the matrix composed of N columns of $\boldsymbol{A}$ defined by $\mathcal{I}_k$, i.e., $\boldsymbol{A}_{\mathcal{I}_k} \triangleq [A_{j,i}]_{j\in[1,N], i\in\mathcal{I}_k}$.

Proof: The sketch of the proofs for the above lemma is illustrated in Fig. 3a. Given that there exists k such that



(a) An example for Lemma 5

(b) An example for Lemma 6

Fig. 3: Illustrations of the worst-case ranks of the submatrices from a sliding window. For each k, the number of vertical dots represents the rank of the submatrix $\boldsymbol{A}_{\mathcal{I}_k}$. In particular, the number of red (resp. blue) dots is the rank of the submatrix selected by the red (resp. blue) window. The sum of the ranks can be found by counting the number of dots.

the submatrix selected by the window is full rank N (the blue window in Fig.3a), the rank of the submatrix selected by the window $\mathcal{I}_{k+1}$ or $\mathcal{I}_{k-1}$ (the red window in Fig.3b) is lower bounded by $N-1$. By applying the same argument, it is readily shown that the rank of the submatrix selected by the window $\mathcal{I}_{k+2}$ or $\mathcal{I}_{k-2}$ is lower bounded by $N-2$. This lower bound keeps decreasing when the window slides away from the blue one, until it hits another lower bound $N-(M-N)=2N-M$ given by Lemma 4. The submatrices within the sliding windows are of rank $2N-M$, which lasts $M-1-2(M-N)=2N-M-1$ times. With the help of Fig.3a, a lower bound on the sum of the ranks of all the submatrices visited by the sliding window, can be obtained by counting the dots in the figure, i.e.,

$$N+2\sum_{i=1}^{M-N}(N-i)+(2N-M)(2N-M+1)=N^2. \tag{216}$$

In fact, this can be found easily by "completing the triangle", the number of dots in which is N^2. ∎

Lemma 6. *$\boldsymbol{A}'\in\mathbb{C}^{N\times M'}$, $N\le M'\le M$, is a submatrix of $\boldsymbol{A}$. We define $\mathcal{I}'_1,\dots,\mathcal{I}'_{M'}$ as a cyclic sliding window of size N on the set of indices $\{1,\dots,M'\}$, i.e.,*

$$\mathcal{I}'_k\triangleq\{(k+i)_{M'}+1\ :i\in[0,N-1]\},\quad k=1,\dots,M'. \tag{217}$$

If the columns of $\boldsymbol{A}'$ are arranged such that the first $\operatorname{rank}(\boldsymbol{A}')$ columns of $\boldsymbol{A}'_{\mathcal{I}'_k}$ are linear independent for some $k\in[1,M]$, then we have

$$\sum_{k=1}^{M'}\operatorname{rank}(\boldsymbol{A}'_{\mathcal{I}'_k})\ge N(N-(M-M')) \tag{218}$$

where $\boldsymbol{A}'_{\mathcal{I}'_k}$ is the submatrix of $\boldsymbol{A}'$ with N columns defined by $\mathcal{I}'_k$, i.e., $\boldsymbol{A}'_{\mathcal{I}'_k}\triangleq[A'_{j,i}]_{j\in[1,N],i\in\mathcal{I}'_k}$.

Proof: The sketch of the proofs for the above lemma is illustrated in Fig.3b. Given that there exists k such that the submatrix selected by the window has rank $r=N-(M-M')$ given by Lemma 4 and that the first r columns are linearly independent (the blue window in Fig.3b), the rank of the submatrix selected by the windows $\mathcal{I}'_{k-1},\dots,\mathcal{I}'_{k-(N-r)}$ (the red and brown windows in Fig.3b) is lower bounded by $r-1$. This lower bound keeps decreasing when the window slides go away from these positions, until it hits another lower bound $N-(M-N)=2N-M$ given by Lemma 4. With the help of Fig.3b, a lower bound on the sum of the ranks of all the submatrices visited by the sliding window, can be obtained by counting the dots in the Figure. In fact, after some basic computations, it turns out that there are $N(N-(M-M'))$ dots. ∎

Lemma 7. *Given $\boldsymbol{H}=\hat{\boldsymbol{H}}+\tilde{\boldsymbol{H}}\in\mathbb{C}^{N\times M}$, $N\le M$, with the entries of $\tilde{\boldsymbol{H}}$ being i.i.d. $\mathcal{N}_{\mathbb{C}}\left(0,\sigma^2\right)$, $\sigma>0$, then*

$$\mathbb{E}_{\tilde{\boldsymbol{H}}}\log\det(\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}})\ge(N-\operatorname{rank}(\hat{\boldsymbol{H}}))\log\sigma^2+O_{\hat{H}}(1) \tag{219}$$

as σ^2 goes to 0.

Proof: According to [31, Lemma 1], for any $\boldsymbol{G}=\hat{\boldsymbol{G}}+\tilde{\boldsymbol{G}}\in\mathbb{C}^{N\times N}$ with the entries in $\tilde{\boldsymbol{G}}$ i.i.d. $\mathcal{N}_{\mathbb{C}}\left(0,1\right)$ independent of $\hat{\boldsymbol{G}}$, we have

$$\mathbb{E}_{\tilde{G}}\log\det(\boldsymbol{G}\boldsymbol{G}^{\mathsf{H}})\ge\sum_{i=1}^{\tau}\log(\lambda_i(\hat{\boldsymbol{G}}\hat{\boldsymbol{G}}^{\mathsf{H}}))+O(1) \tag{220}$$

where $\tau\le\operatorname{rank}(\hat{\boldsymbol{G}})$ is the number of eigenvalues of $\boldsymbol{G}$ that are larger than 1. From here, it follows that

$$\mathbb{E}_{\tilde{G}}\log\det(\boldsymbol{G}\boldsymbol{G}^{\mathsf{H}})\ge\sum_{i=1}^{\operatorname{rank}(\hat{\boldsymbol{G}})}\log(1+\lambda_i(\hat{\boldsymbol{G}}\hat{\boldsymbol{G}}^{\mathsf{H}}))+O(1) \tag{221}$$

since the remaining $\operatorname{rank}(\hat{\boldsymbol{G}})-\tau$ eigenvalues are smaller than 1 and do not contribute more than $O(1)$ to the expectation.

Therefore, for any $\sigma > 0$, we can apply the above inequality to $\sigma^{-1}\boldsymbol{H}$ and have

$$\begin{aligned}
&\mathbb{E}_{\tilde{\boldsymbol{H}}} \log\det((\sigma^{-1}\boldsymbol{H})(\sigma^{-1}\boldsymbol{H})^{\mathsf{H}}) \\
&\quad\geq \sum_{i=1}^{\mathrm{rank}(\hat{\boldsymbol{H}})} \log(\lambda_i(\sigma^{-2}\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\mathsf{H}})) + O(1) && (222)\\
&\quad= -\mathrm{rank}(\hat{\boldsymbol{H}})\log\sigma^2 + \sum_{i=1}^{\mathrm{rank}(\hat{\boldsymbol{H}})} \log(\lambda_i(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\mathsf{H}})) + O(1) && (223)\\
&\quad= -\mathrm{rank}(\hat{\boldsymbol{H}})\log\sigma^2 + O_{\hat{H}}(1) && (224)
\end{aligned}$$

where the last equality is from Assumption 1 that $\mathbb{E}_{\hat{H}}(\log\det(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\mathsf{H}})) > -\infty$. ■

In the following, we prove Lemma 3 case by case according to the value of M[3]. First, let us recall that $N \leq L$. Since the case with $M \leq N$ is trivial, we focus on the cases with $N < M < L$ and $M \geq L$.

1) Case A: $N < M < L$: Let us define M' as the number of eigenvalues of $\boldsymbol{K}$ that are not smaller than 1[4], and let $\boldsymbol{K} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathsf{H}}$ be the eigenvalue decomposition of $\boldsymbol{K}$. We first establish the following upper bound:

$$\begin{aligned}
\det(\mathbf{I} + \boldsymbol{S}\boldsymbol{K}\boldsymbol{S}^{\mathsf{H}}) &= \det(\mathbf{I} + \boldsymbol{\Lambda}\boldsymbol{V}^{\mathsf{H}}\boldsymbol{S}^{\mathsf{H}}\boldsymbol{S}\boldsymbol{V}^{\mathsf{H}}) && (225)\\
&\leq \det(\mathbf{I} + \lambda_{\max}(\boldsymbol{V}^{\mathsf{H}}\boldsymbol{S}^{\mathsf{H}}\boldsymbol{S}\boldsymbol{V})\boldsymbol{\Lambda}) && (226)\\
&\leq \det(\mathbf{I} + \|\boldsymbol{S}\|_{\mathsf{F}}^2\boldsymbol{\Lambda}) && (227)
\end{aligned}$$

where the last inequality is due to $\lambda_{\max}(\boldsymbol{V}^{\mathsf{H}}\boldsymbol{S}^{\mathsf{H}}\boldsymbol{S}\boldsymbol{V}) \leq \|\boldsymbol{S}\boldsymbol{V}\|_{\mathsf{F}}^2 = \|\boldsymbol{S}\|_{\mathsf{F}}^2$. Therefore, we have

$$\begin{aligned}
\mathbb{E}_{\tilde{\boldsymbol{S}}} \log\det(\mathbf{I} + \boldsymbol{S}\boldsymbol{K}\boldsymbol{S}^{\mathsf{H}}) &\leq \log\det(\mathbf{I} + \mathbb{E}_{\tilde{\boldsymbol{S}}}(\|\boldsymbol{S}\|_{\mathsf{F}}^2)\,\boldsymbol{\Lambda}) && (228)\\
&\leq \log\det(\mathbf{I} + \mathbb{E}_{\tilde{\boldsymbol{S}}}(\|\boldsymbol{S}\|_{\mathsf{F}}^2)\,\boldsymbol{\Lambda}') + O_{\hat{S}}(1) && (229)
\end{aligned}$$

where the first inequality is from (227) on which we apply Jensen's inequality; $\boldsymbol{\Lambda}'$ is a diagonal matrix composed of the M' largest eigenvalues of $\boldsymbol{K}$.

Next, let $\boldsymbol{\Phi} \triangleq \boldsymbol{H}\boldsymbol{V}$, $\boldsymbol{\Phi}' \triangleq \hat{\boldsymbol{H}}\boldsymbol{V}$. Without loss of generality, we assume that the columns of $\boldsymbol{\Phi}$ and $\boldsymbol{\Phi}'$ are arranged such that the conditions in Lemma 5 and Lemma 6 are satisfied (i.e., $\mathrm{rank}(\boldsymbol{\Phi}_{\mathcal{I}}) = N$, where $\mathcal{I}$ is the cyclic window with size N, and $\boldsymbol{\Phi}_{\mathcal{I}}$ is defined as in Lemma 5), respectively. This also implies that the eigenvalues in $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}'$ are arranged accordingly. In the following, given different values of M', we prove that

$$\begin{aligned}
&\mathbb{E}_{\tilde{\boldsymbol{H}}} \log\det(\mathbf{I} + \boldsymbol{H}\boldsymbol{K}\boldsymbol{H}^{\mathsf{H}}) \\
&\quad\geq \frac{N}{M}\log\det(\boldsymbol{\Lambda}') + \frac{N(M-N)}{M}\log\sigma^2 + O_{\hat{H}}(1). && (230)
\end{aligned}$$

<u>*Case $M' = M$:*</u> In this case, we have

$$\begin{aligned}
\det(\mathbf{I} + \boldsymbol{H}\boldsymbol{K}\boldsymbol{H}^{\mathsf{H}}) &= \det(\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathsf{H}}) && (231)\\
&= \sum_{\mathcal{I}\subseteq\{1,\dots,N\}} \det(\boldsymbol{\Lambda}_{\mathcal{I}})\det(\boldsymbol{\Phi}_{\mathcal{I}}^{\mathsf{H}}\boldsymbol{\Phi}_{\mathcal{I}}) && (232)\\
&\geq \sum_{k=1}^{M} \det(\boldsymbol{\Phi}_{\mathcal{I}_k}^{\mathsf{H}}\boldsymbol{\Phi}_{\mathcal{I}_k})\det(\boldsymbol{\Lambda}_{\mathcal{I}_k}) && (233)
\end{aligned}$$

where (232) is an application of the identity $\det(I + \boldsymbol{A}) = \sum_{\mathcal{I}\subseteq\{1,\dots,M\}}\det(\boldsymbol{A}_{\mathcal{I}\mathcal{I}})$ for any $\boldsymbol{A} \in \mathbb{C}^{M\times M}$ [32]; the lower bound is obtained by only considering a sliding window of size N for all the possible sub-determinant. Thus,

$$\begin{aligned}
&\log\det(\mathbf{I} + \boldsymbol{H}\boldsymbol{K}\boldsymbol{H}^{\mathsf{H}}) \\
&\quad\geq \log\left(\sum_{k=1}^{M} \det(\boldsymbol{\Phi}_{\mathcal{I}_k}^{\mathsf{H}}\boldsymbol{\Phi}_{\mathcal{I}_k})\det(\boldsymbol{\Lambda}_{\mathcal{I}_k})\right) && (234)\\
&\quad\geq \log\left(\frac{1}{M}\sum_{k=1}^{M} \det(\boldsymbol{\Phi}_{\mathcal{I}_k}^{\mathsf{H}}\boldsymbol{\Phi}_{\mathcal{I}_k})\det(\boldsymbol{\Lambda}_{\mathcal{I}_k})\right) && (235)\\
&\quad\geq \frac{1}{M}\log\left(\prod_{k=1}^{M} \det(\boldsymbol{\Phi}_{\mathcal{I}_k}^{\mathsf{H}}\boldsymbol{\Phi}_{\mathcal{I}_k})\det(\boldsymbol{\Lambda}_{\mathcal{I}_k})\right) && (236)\\
&\quad= \frac{1}{M}\left(N\log\det(\boldsymbol{\Lambda}) + \sum_{k=1}^{M}\log\det(\boldsymbol{\Phi}_{\mathcal{I}_k}^{\mathsf{H}}\boldsymbol{\Phi}_{\mathcal{I}_k})\right) && (237)
\end{aligned}$$

where (236) holds since arithmetic mean is not smaller than geometric mean; the last equality is from the sliding window property $\prod_{k=1}^{M}\det(\boldsymbol{\Lambda}_{\mathcal{I}_k}) = \det(\boldsymbol{\Lambda})^N$. Finally, we have

$$\begin{aligned}
&\mathbb{E}_{\tilde{\boldsymbol{H}}} \log\det(\mathbf{I} + \boldsymbol{H}\boldsymbol{K}\boldsymbol{H}^{\mathsf{H}}) \\
&\quad\geq \frac{1}{M}\left(N\log\det(\boldsymbol{\Lambda}) + \sum_{k=1}^{M}\mathbb{E}_{\tilde{\boldsymbol{H}}}\log\det(\boldsymbol{\Phi}_{\mathcal{I}_k}^{\mathsf{H}}\boldsymbol{\Phi}_{\mathcal{I}_k})\right) && (238)\\
&\quad\geq \frac{1}{M}\left(N\log\det(\boldsymbol{\Lambda}) + \log\sigma^2\sum_{k=1}^{M}(N - \mathrm{rank}(\hat{\boldsymbol{\Phi}}_{\mathcal{I}_k}))\right) + O_{\hat{H}}(1) && (239)\\
&\quad= \frac{1}{M}\left(N\log\det(\boldsymbol{\Lambda}) + \log\sigma^2\left(MN - \sum_{k=1}^{M}\mathrm{rank}(\hat{\boldsymbol{\Phi}}_{\mathcal{I}_k})\right)\right) + O_{\hat{H}}(1) && (240)\\
&\quad\geq \frac{N}{M}\left(\log\det(\boldsymbol{\Lambda}') + (M-N)\log\sigma^2\right) + O_{\hat{H}}(1) && (241)
\end{aligned}$$

where $\hat{\boldsymbol{\Phi}} \triangleq \hat{\boldsymbol{H}}\boldsymbol{V}$ and hence $\mathrm{rank}(\hat{\boldsymbol{\Phi}}) = \mathrm{rank}(\hat{\boldsymbol{H}})$; (239) is from Lemma 7; the last inequality is from Lemma 5 and that $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}'$ as $M = M'$.

<u>*Case $M > M' \geq N$:*</u> For this case, we can first get

$$\begin{aligned}
\det(\mathbf{I} + \boldsymbol{H}\boldsymbol{K}\boldsymbol{H}^{\mathsf{H}}) &= \det(\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathsf{H}}) \\
&\geq \det(\mathbf{I} + \boldsymbol{\Phi}'\boldsymbol{\Lambda}'(\boldsymbol{\Phi}')^{\mathsf{H}}). && (242)
\end{aligned}$$

Following the same footsteps as in (238)-(240), we obtain

$$\begin{aligned}
&\mathbb{E}_{\tilde{\boldsymbol{H}}} \log\det(\mathbf{I} + \boldsymbol{H}\boldsymbol{K}\boldsymbol{H}^{\mathsf{H}}) \\
&\quad\geq \frac{1}{M'}\left(N\log\det(\boldsymbol{\Lambda}') + \log\sigma^2\left(M'N - \sum_{k=1}^{M'}\mathrm{rank}(\hat{\boldsymbol{\Phi}}'_{\mathcal{I}'_k})\right)\right) + O_{\hat{H}}(1) && (243)\\
&\quad\geq \frac{N}{M'}\left(\log\det(\boldsymbol{\Lambda}') + (M-N)\log\sigma^2\right) + O_{\hat{H}}(1) && (244)\\
&\quad\geq \frac{N}{M}\log\det(\boldsymbol{\Lambda}') + \frac{N(M-N)}{M}\log\sigma^2 + O_{\hat{H}}(1) && (245)
\end{aligned}$$

where the inequality (244) is from Lemma 6.

[3]The technique employed in this proof was first developed in our earlier version of this paper [16], and later applied and extended to tackle the K-user MISO case in [17], [31].

[4]Or any constant $c > 0$ that is independent of any parameter in the system. Note that M' can depend on $\hat{\boldsymbol{S}}$ and the SNR P.

Case $M' < N$: From (242) and given that $M' < N$, we have

$$
\begin{aligned}
&\mathbb{E}_{\tilde{H}} \log\det(\mathbf{I} + \boldsymbol{H}\boldsymbol{K}\boldsymbol{H}^{\mathsf{H}}) \\
&\ge \log\det(\boldsymbol{\Lambda}') + \log\sigma^2 \left(M' - \operatorname{rank}(\hat{\boldsymbol{\Phi}}')\right) + O_{\hat{H}}(1) && (246)\\
&\ge \log\det(\boldsymbol{\Lambda}') + \log\sigma^2 \left(M' - (N - (M - M'))\right) + O_{\hat{H}}(1) && (247)\\
&= \log\det(\boldsymbol{\Lambda}') + (M-N)\log\sigma^2 + O_{\hat{H}}(1) && (248)\\
&\ge \frac{N}{M}\log\det(\boldsymbol{\Lambda}') + \frac{N(M-N)}{M}\log\sigma^2 + O_{\hat{H}}(1) && (249)
\end{aligned}
$$

where (247) is from $\log\sigma^2 \le 0$ and $\operatorname{rank}(\hat{\boldsymbol{\Phi}}') \ge N - (M - M')$.

By now, (230) has been proved in all configurations of (M, N, M'). Combining (229) and (230), we have

$$
\begin{aligned}
&N\,\mathbb{E}_{\tilde{S}} \log\det(\mathbf{I} + \boldsymbol{S}\boldsymbol{K}\boldsymbol{S}^{\mathsf{H}}) - M\,\mathbb{E}_{\tilde{H}} \log\det(\mathbf{I} + \boldsymbol{H}\boldsymbol{K}\boldsymbol{H}^{\mathsf{H}}) \\
&\quad\le -N(M-N)\log\sigma^2 + O_{\hat{S}}(1) + O_{\hat{H}}(1) \\
&\qquad + N\log\det\big(\mathbb{E}_{\tilde{S}}(\|\boldsymbol{S}\|_{\mathrm{F}}^2)\,\mathbf{I} + (\boldsymbol{\Lambda}')^{-1}\big) && (250)\\
&\quad\le -N(M-N)\log\sigma^2 + O_{\hat{S}}(1) + O_{\hat{H}}(1) && (251)
\end{aligned}
$$

where the last inequality is from the fact that $\boldsymbol{\Lambda}' \succeq \mathbf{I}$ by construction and hence $\log\det\big(\mathbb{E}_{\tilde{S}}(\|\boldsymbol{S}\|_{\mathrm{F}}^2)\,\mathbf{I} + (\boldsymbol{\Lambda}')^{-1}\big) \le M'\log(1 + \mathbb{E}_{\tilde{S}}(\|\boldsymbol{S}\|_{\mathrm{F}}^2)) = O_{\hat{S}}(1)$. This completes the proof of (151) for the case $N < M < L$.

2) Case B: $M \ge L$: For the first term in (151), we bound it as follows

$$
\begin{aligned}
&\mathbb{E}_{\tilde{S}} \log\det(\mathbf{I} + \boldsymbol{S}\boldsymbol{K}\boldsymbol{S}^{\mathsf{H}}) \\
&= \mathbb{E}_{\tilde{S}} \log\det(\mathbf{I} + \boldsymbol{U}_{\boldsymbol{S}}\boldsymbol{\Sigma}_{\boldsymbol{S}}\boldsymbol{V}_{\boldsymbol{S}}^{\mathsf{H}}\boldsymbol{K}\boldsymbol{V}_{\boldsymbol{S}}\boldsymbol{\Sigma}_{\boldsymbol{S}}\boldsymbol{U}_{\boldsymbol{S}}^{\mathsf{H}}) && (252)\\
&= \mathbb{E}_{\tilde{S}} \log\det(\mathbf{I} + \boldsymbol{\Sigma}_{\boldsymbol{S}}^2\boldsymbol{V}_{\boldsymbol{S}}^{\mathsf{H}}\boldsymbol{K}\boldsymbol{V}_{\boldsymbol{S}}) && (253)\\
&\le \mathbb{E}_{\tilde{S}} \log\det(\mathbf{I} + \lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{S}}^2)\boldsymbol{V}_{\boldsymbol{S}}^{\mathsf{H}}\boldsymbol{K}\boldsymbol{V}_{\boldsymbol{S}}) && (254)\\
&= \sum_{i=1}^{L} \mathbb{E}_{\tilde{S}} \log(1 + \lambda_{\max}(\boldsymbol{S}\boldsymbol{S}^{\mathsf{H}})\lambda_i(\boldsymbol{V}_{\boldsymbol{S}}^{\mathsf{H}}\boldsymbol{K}\boldsymbol{V}_{\boldsymbol{S}})) && (255)\\
&\le \sum_{i=1}^{L} \mathbb{E}_{\tilde{S}} \log(1 + \lambda_{\max}(\boldsymbol{S}\boldsymbol{S}^{\mathsf{H}})\lambda_i) && (256)\\
&\le \sum_{i=1}^{L} \mathbb{E}_{\tilde{S}} \log(1 + \|\boldsymbol{S}\|_{\mathrm{F}}^2\lambda_i) && (257)\\
&\le \sum_{i=1}^{L} \log(1 + \mathbb{E}_{\tilde{S}}(\|\boldsymbol{S}\|_{\mathrm{F}}^2)\,\lambda_i) && (258)\\
&= \log\det(\mathbf{I} + \mathbb{E}_{\tilde{S}}(\|\boldsymbol{S}\|_{\mathrm{F}}^2)\,\boldsymbol{\Lambda}'') && (259)\\
&= \log\det(\mathbf{I} + \mathbb{E}_{\tilde{S}}(\|\boldsymbol{S}\|_{\mathrm{F}}^2)\,\boldsymbol{\Lambda}''') + O_{\hat{S}}(1) && (260)
\end{aligned}
$$

where in (252), $\boldsymbol{S} = \boldsymbol{U}_{\boldsymbol{S}}\boldsymbol{\Sigma}_{\boldsymbol{S}}\boldsymbol{V}_{\boldsymbol{S}}^{\mathsf{H}}$ with $\boldsymbol{\Sigma}_{\boldsymbol{S}} \in \mathbb{C}^{N\times N}$ and $\boldsymbol{V}_{\boldsymbol{S}} \in \mathbb{C}^{M\times L}$; (253) comes from the equality $\det(\mathbf{I} + \boldsymbol{A}\boldsymbol{B}) = \det(\mathbf{I} + \boldsymbol{B}\boldsymbol{A})$; (256) is due to Poincare Separation Theorem [32] that $\lambda_i(\boldsymbol{V}_{\boldsymbol{S}}^{\mathsf{H}}\boldsymbol{K}\boldsymbol{V}_{\boldsymbol{S}}) \le \lambda_i(\boldsymbol{K})$ for $i = 1, \cdots, N$; (257) is from the fact that $\lambda_{\max}(\boldsymbol{S}\boldsymbol{S}^{\mathsf{H}}) \le \|\boldsymbol{S}\|_{\mathrm{F}}^2$; (258) is obtained by applying Jensen's inequality; $\boldsymbol{\Lambda}'' \triangleq \operatorname{diag}(\lambda_1, \cdots, \lambda_L)$ and $\boldsymbol{\Lambda}''' \triangleq \operatorname{diag}(\lambda_1, \cdots, \lambda_{\min\{L, M'\}})$ with M' being the number of eigenvalues that are not smaller than 1, i.e., $\Lambda''' \succeq \mathbf{I}$.

For the second term in (151), we use the following lower bound

$$
\begin{aligned}
&\mathbb{E}_{\tilde{H}} \log\det(\mathbf{I} + \boldsymbol{H}\boldsymbol{K}\boldsymbol{H}^{\mathsf{H}}) \\
&\quad= \mathbb{E}_{\tilde{H}} \log\det(\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathsf{H}}) && (261)\\
&\quad\ge \mathbb{E}_{\tilde{H}} \log\det(\mathbf{I} + \boldsymbol{\Phi}'\boldsymbol{\Lambda}''\boldsymbol{\Phi}'^{\mathsf{H}}) && (262)\\
&\quad\ge \frac{N}{L}\log\det(\boldsymbol{\Lambda}''') + \frac{N(L-N)}{L}\log(\sigma^2) + O_{\hat{H}}(1) && (263)
\end{aligned}
$$

where $\boldsymbol{\Phi} \triangleq \boldsymbol{H}\boldsymbol{V} \in \mathbb{C}^{N\times M}$ with $\boldsymbol{V}$ being the unitary matrix containing the eigenvectors of $\boldsymbol{K}$, i.e., $\boldsymbol{K} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathsf{H}}$ with $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \cdots, \lambda_M)$; in (262), $\boldsymbol{\Phi}' = \boldsymbol{H}\boldsymbol{V}' \in \mathbb{C}^{N\times L}$ with $\boldsymbol{V}'$ being the first L columns of $\boldsymbol{V}$, and multiplying by the matrix $\boldsymbol{V}'$ does not change the distribution property, and $\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathsf{H}} \succeq \boldsymbol{\Phi}'\boldsymbol{\Lambda}''\boldsymbol{\Phi}'^{\mathsf{H}}$; the last inequality is obtained from (230) in the previous subsection.

Finally, it is readily shown that, following the same steps as in (250) and (251),

$$
\begin{aligned}
&\frac{1}{L}\mathbb{E}_{\tilde{S}} \log\det(\mathbf{I} + \boldsymbol{S}\boldsymbol{K}\boldsymbol{S}^{\mathsf{H}}) - \frac{1}{N}\mathbb{E}_{\tilde{H}} \log\det(\mathbf{I} + \boldsymbol{H}\boldsymbol{K}\boldsymbol{H}^{\mathsf{H}}) \\
&\qquad\le -\frac{L-N}{L}\log(\sigma^2) + O_{\hat{S}}(1) + O_{\hat{H}}(1). && (264)
\end{aligned}
$$

This completes the proof of (151) for the case $M \ge L$.

References

[1] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936 –3964, Sept. 2006.

[2] S. Jafar and M. Fakhereddin, "Degrees of freedom for the MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2637–2642, Jul. 2007.

[3] A. Lozano, R. W. Heath, and J. G. Andrews, "Fundamental limits of cooperation," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5213–5226, Sept. 2013.

[4] M. Maddah-Ali and D. 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] C. S. Vaze and M. Varanasi, "The degrees of freedom region of the two-user MIMO broadcast channel with delayed CSIT," *in Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, St. Petersburg, Russia, Jul. 2011.

[6] C. Vaze and M. Varanasi, "The degrees of freedom region and interference alignment for the MIMO interference channel with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4396–4417, Jul. 2012.

[7] S. Yang, M. Kobayashi, P. Piantanida, and S. Shamai, "Secrecy degrees of freedom of MIMO broadcast channels with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5244–5256, Sept. 2013.

[8] C. Huang, S. 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.

[9] Y. Zhu and D. Guo, "The degrees of freedom of isotropic MIMO interference channels without state information at the transmitters," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 341–352, Jan. 2012.

[10] C. Vaze and M. Varanasi, "The degree-of-freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5354–5374, Aug. 2012.

[11] R. Tandon, S. Jafar, S. Shamai, and H. Poor, "On the synergistic benefits of alternating CSIT for the MISO BC," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4106–4128, Jul. 2013.

[12] N. Lee and R. Heath Jr, "Not too delayed CSIT achieves the optimal degrees of freedom," *in Proc. 50th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, *arXiv preprint arXiv:1207.2211*, 2012.

[13] 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. Inf. Theory (ISIT)*, Boston, USA, Jul. 2012, pp. 2501–2505.

[14] 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.

[15] 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.

[16] X. Yi, S. Yang, D. Gesbert, and M. Kobayashi, "The degrees of freedom region of temporally-correlated MIMO networks with delayed CSIT," *arXiv preprint arXiv:1211.3322v1*, Oct. 2012.

[17] P. de Kerret, X. Yi, and D. Gesbert, "On the degrees of freedom of the K-user time correlated broadcast channel with delayed CSIT," *in Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, *arXiv preprint arXiv:1301.2138*, Jul. 2013.

[18] J. Chen and P. Elia, "Degrees-of-freedom region of the MISO broadcast channel with general mixed-CSIT," *in Proc. Inf. Theory Appl. Workshop*, La Jolla, CA, USA, Feb. 2013.

[19] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sept. 1979.

[20] L. Ozarow and S. Leung-Yan-Cheong, "An achievable region and outer bound for the Gaussian broadcast channel with feedback," *IEEE Trans. Inf. Theory*, vol. 30, no. 4, pp. 667–671, Jul. 1984.

[21] C. Suh and D. Tse, "Feedback capacity of the Gaussian interference channel to within 2 bits," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2667 –2685, May 2011.

[22] F. Willems and E. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 313–327, May 1985.

[23] J. Chen and P. Elia, "Toward the performance vs. feedback tradeoff for the two-user MISO broadcast channel," *accepted in IEEE Trans. Inf. Theory, arXiv preprint arXiv:1306.1751v4*, Sept. 2013.

[24] R. Tandon, M. Maddah-Ali, A. Tulino, H. Poor, and S. Shamai, "On fading broadcast channels with partial channel state information at the transmitter," *in Proc. IEEE Int. Symp. Wireless Communication Systems (ISWCS)*, 2012, pp. 1004–1008.

[25] 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, Jun. 2012.

[26] K. Mohanty, C. S. Vaze, and M. Varanasi, "The degrees of freedom region for the MIMO interference channel with hybrid CSIT," *arXiv preprint arXiv:1209.0047*, 2012.

[27] 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.

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

[29] 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.

[30] 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.

[31] J. Chen, S. Yang, and P. Elia, "On the fundamental feedback-vs-performance tradeoff over the MISO-BC with imperfect and delayed CSIT," *in Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013.

[32] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

Xinping Yi (S'12) received the B.E. degree from Huazhong University of Science and Technology and M.E. degree from University of Electronic Science and Technology, China, both in Electrical Engineering. Currently, He is pursuing the Ph.D. degree at Mobile Communication Department, EURECOM, Sophia Antipolis, France. From 2009 to 2011, he was a research engineer in Huawei Technologies, Shenzhen, China. His current research interests include multiuser information theory and interference management for wireless communications.

Sheng Yang (M'07) received the B.E. degree in electrical engineering from Jiaotong University, Shanghai, China, in 2001, and both the engineer degree and the M.Sc. degree in electrical engineering from École Nationale Supérieure des Télécommunications (ENST), Paris, France, in 2004, respectively. From 2004 to 2007, he worked as teaching and research assistant in the Communications & Electronics department in ENST. During the same period, he completed his Ph.D., graduating in 2007 from Université de Pierre et Marie Curie (Paris VI). From October 2007 to November 2008, he was with Motorola Research Center in Gif-sur-Yvette, France, as a senior staff research engineer. Since December 2008, he has joined the Telecommunications department at SUPELEC where he is currently an assistant professor. His research interests include cooperative diversity schemes, wireless networks information theory, and coding/decoding techniques for multi-antenna communication systems.

David Gesbert (F'11) is Professor and Head of the Mobile Communications Department, EURECOM, France, where he also heads the Communications Theory Group. He obtained the Ph.D. degree from Ecole Nationale Superieure des Telecommunications, France, in 1997. From 1997 to 1999 he has been with the Information Systems Laboratory, Stanford University. In 1999, he was a founding engineer of Iospan Wireless Inc., San Jose, Ca., a startup company pioneering MIMO-OFDM (now Intel). Between 2001 and 2003 he has been with the Department of Informatics, University of Oslo as an adjunct professor. D. Gesbert has published over 200 papers and several patents all in the area of signal processing, communications, and wireless networks.

D. Gesbert was a co-editor of several special issues on wireless networks and communications theory, for JSAC (2003, 2007, 2009), EURASIP Journal on Applied Signal Processing (2004, 2007), Wireless Communications Magazine (2006). He served on the IEEE Signal Processing for Communications Technical Committee, 2003-2008. He was an associate editor for IEEE Transactions on Wireless Communications and the EURASIP Journal on Wireless Communications and Networking. He authored or co-authored papers winning the 2012 SPS Signal Processing Magazine Best Paper Award, 2004 IEEE Best Tutorial Paper Award (Communications Society), 2005 Young Author Best Paper Award for Signal Proc. Society journals, and paper awards at conferences 2011 IEEE SPAWC, 2004 ACM MSWiM workshop. He co-authored the book Space time wireless communications: From parameter estimation to MIMO systems, Cambridge Press, 2006. In 2013, he is a General Chair for the IEEE Communications Theory Workshop, and a Program Chair for the Communications Theory Symposium of ICC2013.

Mari Kobayashi (M'06) received the B.E. degree in electrical engineering from Keio University, Yokohama, Japan, in 1999, the M.S. degree in Mobile Radio and the Ph.D. from Ecole Nationale Supérieure des Télécommunications, Paris, France, in 2000 and in 2005 respectively. From November 2005 to March 2007, she was a post-doc researcher at Centre Tecnològic de Telecomunicacions de Catalunya, Barcelona, Spain. Since May 2007, she is assistant professor in Supélec, Gif-sur-Yvette, France, where she is associate professor. She received Newcom++ Best paper award in 2009, the Communications Society/Information Theory Society Joint Paper Award in 2011. Her current research interests include MIMO communication systems, multiuser communication theory.