Cooperative Multiple-Access Channels with Distributed State Information

Lorenzo Miretti, *Student Member, IEEE,* Mari Kobayashi, *Member, IEEE,*
David Gesbert, *Fellow, IEEE,* Paul de Kerret *Member, IEEE*

Abstract—This paper studies a memoryless state-dependent multiple access channel (MAC) where two transmitters wish to convey a message to a receiver under the assumption of *causal* and imperfect channel state information at transmitters (CSIT) and imperfect channel state information at receiver (CSIR). In order to emphasize the limitation of transmitter cooperation between physically distributed nodes, we focus on the so-called *distributed* CSIT assumption, i.e., where each transmitter has its individual channel knowledge, while the message can be assumed to be partially or entirely shared a priori between transmitters by exploiting some on-board memory. Under this setup, the first part of the paper characterizes the common message capacity of the channel at hand for arbitrary CSIT and CSIR structure. The optimal scheme builds on Shannon strategies, i.e., optimal codes are constructed by letting the channel inputs be a function of current CSIT only. For a special case when CSIT is a deterministic function of CSIR, the considered scheme also achieves the capacity region of a common message and two private messages. The second part addresses an important instance of the previous general result in a context of a cooperative multi-antenna Gaussian channel under i.i.d. fading operating in frequency-division duplex mode, such that CSIT is acquired via an explicit feedback of perfect CSIR. The capacity of the channel at hand is achieved by distributed linear precoding applied to Gaussian codes. Surprisingly, we demonstrate that it is suboptimal to send a number of data streams bounded by the number of transmit antennas as typically considered in a centralized CSIT setup. Finally, numerical examples are provided to evaluate the sum capacity of the binary MAC with binary states as well as the Gaussian MAC with i.i.d. fading.

Index Terms—Cooperative MAC, capacity, distributed CSIT, Shannon strategies, distributed precoding

I. INTRODUCTION

WIRELESS communication networks can substantially benefit from transmitter cooperation via joint processing among multiple transmitters (TX), as it enables to mitigate
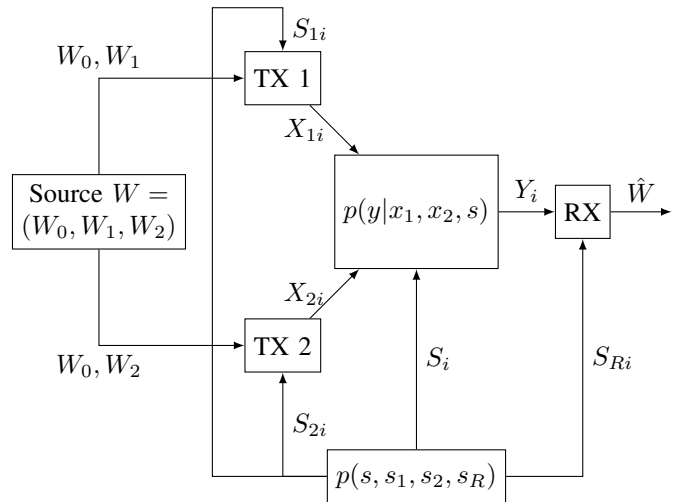
Fig. 1: Cooperative multiple-access channel with causal and distributed CSIT S_1, S_2 and imperfect CSIR S_R.

interference and enhance the network performance. Although the benefits of TX cooperation have been identified in terms of coverage, throughput scaling, spectral efficiency, and energy efficiency, most of the existing cooperative schemes and performance analysis build on the common assumption that perfect, or at least perfectly shared, channel state information at the transmitters (CSIT), referred to *centralized* CSIT, is available (see e.g. [1]–[3] and references therein). While such an assumption is convenient for analysis, it is however being challenged in a number of practical wireless scenarios. In fact, the acquisition of centralized CSIT always entails direct communication between transmitters or feedback from the receivers so that a given transmitter can collect the CSIT of other transmitters. This inevitably induces impairments and delays, which can be represented as a transmitter-specific distortion added to the channel state information. In order to capture such a limitation, we focus on the so-called *distributed* CSIT such that each transmitter has its own channel knowledge. On the other hand, we assume that messages are partially or entirely shared between transmitters prior to the actual data transmission. This assumption is justified for instance in cache-aided networks where parts of delay-tolerant web content (video files) can be pre-fetched at transmitters typically during off-peak hours [4], [5]. More generally, our setup can be thought as representing service situations where the CSIT time sensitivity is high in relation to that of the data contents.

By taking into account both practical CSIT limitation and inherent message-sharing opportunity, we study a state-dependent multiple access channel (MAC) illustrated in Fig. 1. Namely, two transmitters with respective state knowledge S_1, S_2 wish to cooperatively convey a message W through inputs X_1, X_2 to a single receiver (RX) with state knowledge S_R. We do not consider direct communication links between transmitters that enable further online interactions such as conferencing [6], [7]. Rather, we aim to design the transmission strategy for a predefined CSI distribution mechanism described by the joint distribution of (S, S_1, S_2, S_R) and a given message cooperation defined by the rate of (W_0, W_1, W_2). More precisely, the *a priori* cooperation among the TXs is modeled by the following two components:

- *State* cooperation, modeled by the distribution of the CSIT (S_1, S_2). Perfect state cooperation corresponds to a centralized CSIT configuration where the TXs share the same state information, i.e., $S_1 = S_2$ (note that this does not necessarily correspond to perfect CSIT).
- *Message* cooperation, modeled by splitting W into 3 sub-messages (W_0, W_1, W_2), where W_0 is the portion of W available at both TXs, and where W_k, $k = 1, 2$, is the portion available only at TX k. Perfect message cooperation corresponds to $W_0 = W$ or $W_1 = W_2 = \emptyset$.

Distributed CSIT gives rise to many interesting, yet challenging, problems because TXs must cooperate on the basis of uncertainties about each other's state information. There are roughly two classes of works. The first class focuses on signal processing methods [8] such as particular precoders optimization, and asymptotic ergodic rate analysis in the regime of high signal-to-noise ratio [9], [10] for cooperative multi-user networks with interference. The second class is based on the information theoretic models. These include the MAC with partial CSIT S_1, S_2 and full CSIR $S = S_R = (S_1, S_2)$ [11], the slow-fading Gaussian MAC with partial CSIT S_1, S_2 [11], the MAC with conferencing encoders under noncausal partial CSIT S_1, S_2 and full CSIR $S = S_R = (S_1, S_2)$ [7], the MAC with partial and strictly causal state information S_k at TX k and no CSIR [12], and the cooperative MAC with non-causal CSIT at one TX and strictly causal at the other [13]. Although useful system design and performance analysis are obtained from these two frameworks individually, they are disconnected each other in the sense that insights obtained from one class cannot be useful for another.

Motivated by such an observation, we wish to close the gap between these two approaches by designing a simple yet information theoretically optimal cooperative scheme under distributed channel state information. To the best of our knowledge, such a result was not reported before. In particular, we study the capacity of a common message and the capacity region of three messages over a memoryless state-dependent MAC with causal distributed CSIT. Before summarizing the main contributions of the current work in Section I-B, we first review the existing results on coding with causal CSIT under various network models.

A. Coding with causal CSIT

In [14], Shannon characterized the capacity of a memoryless state-dependent point-to-point channel with causal state knowledge at the transmitter $S = S_T \in \mathcal{S}$ and no CSIR $S_R = \emptyset$. The capacity of the channel at hand can be alternatively given by [15]

$$C = \max_{p(u),\, x=f(u,s)} I(U; Y),$$

where $U \in \mathcal{U}$ is an auxiliary random variable of finite cardinality and independent of S, and f is a deterministic function. Notice that U can be seen as an index for the family of functions $\mathcal{S} \to \mathcal{X}$, also known as *Shannon strategies*. This result has a practical impact to the design of modern wireless communication systems as it suggests that capacity is achieved by encoding the message through a function f depending only on the *current* CSIT.

In the following, we briefly summarize the existing results exploiting Shannon strategies. Shannon strategies were generalized to more general setups with imperfect CSIT S_T and imperfect CSIR S_R [16], and to particular cases of state process with memory in [16]. Shannon strategies were also extended to more general network models, including degraded broadcast channels [17], [18], degraded relay channels [18], [19], as well as multiple access channels [12], [13], [17], [18], [20]–[27]. By focusing on the MAC literature, the capacity region of the state-dependent MAC was studied by Das and Narayan in [20], where multi-letter formulas are given for very general channel and CSI models. Unfortunately the multi-letter expressions provide very little insights and cannot be easily computed. In contrast, single-letter expressions on an achievable rate region (R_1, R_2) have been derived for the case of a common state $S_1 = S_2 = S$ in [18]. When the two states (S_1, S_2) are independent, Shannon strategies are proved to achieve the sum capacity [21]. In practical wireless systems operating in frequency division duplexing (FDD) mode, it is typical to assume that the CSIT is a deterministic function of the CSIR as CSIT is acquired as an explicit feedback from the receiver. Under this condition, the full capacity region of the state dependent MAC has been characterized in [21] for independent states and then generalized to arbitrarily correlated states in [22]. For the case of degraded message sets as a special case of Fig. 1 when $W_2 = \emptyset$, the capacity region has been characterized for the case of one-sided CSIT $(S_2 = \emptyset)$ in [26, Section IV].

Interestingly, Shannon strategies are known to be suboptimal in general state-dependent MAC. In particular, Lapidoth and Steinberg demonstrated that Shannon strategies fail to achieve some rate pairs in the state-dependent MAC with no CSIR for the case of common state [24] and the case of independent states [25], causally or strictly causally available at the encoders. This is because block-Markov encoding can help the TXs to provide some CSIR at the RX by compressing and sending past CSIT cooperatively [24] or non-cooperatively [25]. Clearly, this strategy is not useful when the CSIT is already available at the RX, as in [22] and in parts of this work. The scheme proposed in [24], [25] have been further generalized in [12], where the TXs compress and send the

past codewords along with past channel states. The idea of sending the past codewords via block-Markov encoding has been proposed for the MAC with feedback [6], [28], [29], while the idea of sending the past state together with new messages was also considered in the simultaneous state and data communication (see e.g. [30] and references therein) and in the cooperative MAC with strictly causal CSIT [27].

B. Contributions and Paper Outline

This paper provides the following contributions:

1) We demonstrate that the common message capacity of the memoryless state-dependent MAC under distributed CSIT is achieved by Shannon strategies for any CSI distribution $p(s, s_1, s_2, s_R)$ in Theorem 1. The exact characterization of the common message capacity complements the existing results in [21], [24], [25], which consider only two private messages, the results in [26] which assumes $S_2 = \emptyset$ and $W_2 = \emptyset$, and it extends the single-user results in [14], [16] to two distributed TXs.

2) For the special case when the CSIT of each TX k is a deterministic function q_k of the CSIR, i.e., $S_k = q_k(S_R)$ for $k = 1, 2$, we prove that Shannon strategies achieve the full capacity region on three messages (W_0, W_1, W_2) in Theorem 2. This extends the existing result [22] to the case when a common message is present. The contribution of Theorem 2 lies in our converse proof based on a standard information inequality chain, which overcomes the technique used in [21] restricted to the case of independent states while significantly simplifying the approach in [22].

3) By specializing the model of Theorem 2, we establish the common message capacity of the multiple-input multiple-output (MIMO) Gaussian fading channel operating in frequency-division-duplexing (FDD) mode in Theorem 3. We demonstrate that distributed linear precoding over Gaussian codewords based on Shannon strategies is optimal. The difference with respect to centralized CSIT is that each TX shall choose its precoding vector as a function of its channel knowledge, rather than global CSIT. As a key ingredient for the achievability proof, we allow the number of data streams conveying W_0 to grow large up to a given upper bound that depends on the CSIT cardinalities. Furthermore, as a non-trivial extension, Theorem 4 characterizes the entire capacity region of the aforementioned channel. The converse proof exploits the underlying channel structure and functional dependencies, while achievability builds on superposition encoding and the distributed precoding technique developed for Theorem 3.

4) By letting the number of precoded data streams be the maximum dimension, in Proposition 2 and Proposition 3 we prove that the optimal distributed precoding design, belonging to the well-known class of difficult problems called *team decision* problems [8], can be cast into a convex form. Moreover, in Proposition 4 and Proposition 5 we provide a more in-depth analysis on the optimal number of common data streams. Surprisingly, we prove

that the common wisdom of limiting the number of precoded data streams by the number of transmit antennas is strictly suboptimal under a distributed CSIT setup. This is in sharp contrast to the case of centralized CSIT.

The paper is organized as follows. In Section II, we provide the formal system model and the main results for general MAC with distributed CSIT. Section III presents the results for the specific cooperative MIMO MAC at hand. The insights given by the above sections are then further illustrated via numerical examples in Section IV. For readability purposes, most of the proofs are moved to appendices.

II. COOPERATIVE MULTIPLE-ACCESS CHANNELS WITH CAUSAL AND DISTRIBUTED CSIT

This section first provides the general channel model and the basic definitions adopted throughout this work. Then, we present general results on the cooperative multiple access channel (MAC) with causal and distributed CSIT illustrated in Fig. 1. Hereafter, we follow the standard notation of [15].

A. System Model and Problem Statement

a) Channel Model: Consider the state-dependent MAC in Fig. 1, with a common message W_0, two private messages W_1, W_2, inputs $X_1 \in \mathcal{X}_1$, $X_2 \in \mathcal{X}_2$, output $Y \in \mathcal{Y}$, state $S \in \mathcal{S}$, memoryless channel law $p(y|x_1, x_2, s)$, *distributed* CSIT $(S_1, S_2) \in \mathcal{S}_1 \times \mathcal{S}_2$, and imperfect CSIR $S_R \in \mathcal{S}_R$. The sequence of tuples $\{(S_i, S_{1i}, S_{2i}, S_{Ri})\}_{i=1}^n$ is assumed to follow a generic memoryless law $p(s, s_1, s_2, s_R)$. An n-sequence of inputs, output and states is then governed by

$$p(y^n|x_1^n, x_2^n, s^n) = \prod_{i=1}^n p(y_i|x_{1i}, x_{2i}, s_i),$$
$$p(s^n, s_1^n, s_2^n, s_R^n) = \prod_{i=1}^n p(s_i, s_{1i}, s_{2i}, s_{Ri}).$$

We assume that three messages W_0, W_1, W_2 are independently and uniformly distributed over the sets $\mathcal{W}_j \triangleq \{1, \ldots, 2^{\lceil nR_j \rceil}\}$, $j = 0, 1, 2$, where $R_j \geq 0$ is the rate of the message W_j. All alphabets are assumed to be finite, unless otherwise stated.

b) Encoding and Decoding: A block code $(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n)$ of length n with causal CSIT is defined by a set of encoding functions

$$\phi_{ki} : \mathcal{W}_0 \times \mathcal{W}_k \times \mathcal{S}_k^i \to \mathcal{X}_k, \quad k = 1, 2, \quad i = 1, \ldots, n,$$

yielding the transmitted symbols $x_{ki} = \phi_{ki}(w_0, w_k, s_k^i)$, as well as a decoding function

$$\psi : \mathcal{Y}^n \times \mathcal{S}_R^n \to \mathcal{W}_0 \times \mathcal{W}_1 \times \mathcal{W}_2,$$

yielding the decoded messages $(\hat{w}_0, \hat{w}_1, \hat{w}_2) = \psi(y^n, s_R^n)$. Each encoder $k = 1, 2$ is subject to an average input cost constraint

$$\mathcal{E}[b^n(X_k^n)] \leq P_k, \quad b^n(x_k^n) \triangleq \frac{1}{n} \sum_{i=1}^n b_k(x_{ki}),$$

where $b_k : \mathcal{X}_k \to \mathbb{R}_+$ is a single-letter cost function upper-bounded by $b_{k,\max}$. A rate-cost tuple $(R_0, R_1, R_2, P_1, P_2)$ is

said to be achievable if, for the considered channel, there exists a family of block codes of length n defined as before such that the average probability of error satisfies $P_e^{(n)} \triangleq \Pr((\hat{W}_0, \hat{W}_1, \hat{W}_2) \neq (W_0, W_1, W_2)) \to 0$ as $n \to \infty$.

c) Figure of Merit: For a given cost pair (P_1, P_2), the closure of the set of all achievable rates (R_0, R_1, R_2) is the capacity-cost region $\mathscr{C}(P_1, P_2)$ of the considered channel. In this work, we are mostly interested in two operating points in $\mathscr{C}(P_1, P_2)$. Namely, the common message capacity, defined by $C_0(P_1, P_2) \triangleq \max\{R_0 \in \mathscr{C}(P_1, P_2)\}$ and the sum capacity $C_{\text{sum}}(P_1, P_2) \triangleq \max\{R_{\text{sum}} : (R_0, R_1, R_2) \in \mathscr{C}(P_1, P_2)\}$, where $R_{\text{sum}} \triangleq R_0 + R_1 + R_2$ denotes the rate of the aggregate message $W \triangleq (W_0, W_1, W_2)$ in Fig. 1.

B. General Results

As a non-trivial extension of [26, Corollary 3] with the one-sided CSIT ($S_2 = \emptyset$) and of [14], [16] for the centralized CSIT case $S_1 = S_2$ to a general CSI structure, we provide the main result of this section.

Theorem 1. *The common message capacity of the channel in Fig. 1 is given by*

$$C_0(P_1, P_2) = \max_{\substack{p(u) \\ x_k = f_k(u, s_k) \\ \boldsymbol{\mathcal{E}}[b_k(X_k)] \leq P_k \, \forall k}} I(U; Y|S_R), \qquad (1)$$

where $U \in \mathcal{U}$ is an auxiliary random variable of finite cardinality, independent of (S, S_1, S_2, S_R), and where f_k is a deterministic functions $\mathcal{U} \times \mathcal{S}_k \to \mathcal{X}_k$ for $k = 1, 2$.

Proof: The proof is given in Appendix A. ∎

Remark 1. *By replacing W_0 by the triple (W_0, W_1, W_2) in the converse proof of Theorem 1, it immediately follows that $C_{\text{sum}}(P_1, P_2) \equiv C_0(P_1, P_2)$.*

The main finding of Theorem 1 is that the common message capacity (or equivalently, the sum capacity) can be achieved by Shannon strategies, i.e., by coding over the current CSIT S_{1i}, S_{2i} only while neglecting the past CSIT sequences. In fact, the converse proof also shows that providing the strictly causal sequence (S_1^{i-1}, S_2^{i-1}) to both TXs does not increase the common message capacity.

It is also worth emphasizing the difference with respect to the centralized CSIT where both TXs share $S_1 = S_2 \triangleq S_T$. In such case, by omitting for simplicity the input cost constraints, we recover in fact the classical result of [14], [16]

$$C_0 = \max_{\substack{p(u) \\ (x_1, x_2) = f(u, s_T)}} I(U; Y|S_R).$$

Although Shannon strategies are optimal in both both distributed and centralized CSIT cases, the distributed CSIT assumption imposes the design of two different functions f_1, f_2 depending on the local CSIT only each, rather than a single f as in the (virtually) centralized case.

In order to prove the achievability part of Theorem 1, we obtain an achievable region for the MAC with a common message and two private messages as a byproduct. Specifically,

we obtain the following result by combining Slepian-Wolf coding [31] for the state-less MAC with common message and Shannon strategies [14].

Lemma 1. *For the channel in Fig. 1, $\mathscr{C}(P_1, P_2)$ includes the convex hull of all rate triples (R_0, R_1, R_2) such that*

$$R_1 \leq I(U_1; Y|U_2, U_0, S_R),$$
$$R_2 \leq I(U_2; Y|U_1, U_0, S_R),$$
$$R_1 + R_2 \leq I(U_1, U_2; Y|U_0, S_R),$$
$$R_0 + R_1 + R_2 \leq I(U_1, U_2; Y|S_R),$$

for some auxiliary variables $(U_0, U_1, U_2) \in \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2$ of finite cardinality, independent of the CSI (S, S_1, S_2, S_R), with pmf factorizing as $p(u_0)p(u_1|u_0)p(u_2|u_0)$, and for some deterministic functions $f_k : \mathcal{U}_k \times \mathcal{S}_k \to \mathcal{X}_k$, $x_k = f_k(u_k, s_k)$, satisfying $\boldsymbol{\mathcal{E}}[b_k(X_k)] \leq P_k$ for $k = 1, 2$.

It is well known that Shannon strategies, i.e., the scheme of Lemma 1, fail to achieve $\mathscr{C}(P_1, P_2)$ for general $p(s, s_1, s_2, s_R)$, as observed for a special case of a common CSIT $S_1 = S_2 = S$ and no CSIR $S_R = \emptyset$ in [24]. This is because block-Markov encoding enables two TXs to compress past state information and send it as a common message to provide possibly useful CSIR to the RX. Nevertheless, Theorem 1 shows that such a scheme based on block-Markov encoding is not necessary for achieving the sum capacity of the considered setup. Namely, provided that R_0 is large enough (in the worst case, equal to $C_{\text{sum}}(P_1, P_2)$), the sum capacity is indeed achievable by the scheme in Lemma 1.

In the following we focus on the particular case where each CSIT is a deterministic function of CSIR. This assumption is highly relevant to frequency-division duplex (FDD) systems, where each transmitter acquires channel knowledge via an explicit quantized feedback from the receiver. As a straight-forward extension of [22, Theorem 4] and [21, Theorem 5] restricted to two private messages[1], we characterize the capacity region for three messages as follows.

Theorem 2. *By assuming that $S_1 = q_1(S_R)$ and $S_2 = q_2(S_R)$, where q_1, q_2 are two deterministic functions, the capacity region $\mathscr{C}(P_1, P_2)$ of the channel in Fig. 1 is given by the convex hull of all rate triples satisfying*

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2, U, S_R), \\ R_2 &\leq I(X_2; Y|X_1, U, S_R), \\ R_1 + R_2 &\leq I(X_1, X_2; Y|U, S_R), \\ R_0 + R_1 + R_2 &\leq I(X_1, X_2; Y|S_R), \end{aligned} \qquad (2)$$

for some pmf $p(x_1|s_1, u)p(x_2|s_2, u)p(u)$, where $U \in \mathcal{U}$ is an auxiliary variable of finite cardinality and independent of (S, S_1, S_2, S_R), satisfying $\boldsymbol{\mathcal{E}}[b_k(X_k)] \leq P_k$ for $k = 1, 2$.

Proof: The proof is given in Appendix B. ∎

Our main contribution lies in the converse proof, which solves the issue highlighted in [21] through an appropriate identification of the auxiliary variable U, thus allowing to greatly simplify the non-traditional, yet innovative, converse

---

[1] [22, Theorem 4] generalized the case of independent states (S_1, S_2) to arbitrarily correlated states (S_1, S_2).

proof given by [22]. Note that Theorem 2 refers to a setup where the RX is fully informed about (S_1, S_2), hence there is no need for the TXs to convey (S_1, S_2) through block-Markov schemes as in [24], [25]. In contrast to the general CSI setup in Lemma 1, the private messages W_1, W_2 can be directly encoded into the input alphabets $\mathcal{X}_1, \mathcal{X}_2$ as observed already in [21, Theorem 5]. In light of Theorem 2 we highlight the following expression for the common message capacity, which will be used to prove the main result of the second part of this paper.

Remark 2. *Under the assumption $S_1 = q_1(S_R)$ and $S_2 = q_2(S_R)$, the expression in* (1) *is equivalently given by*

$$C_0(P_1, P_2) = \max_{\substack{p(u) \\ x_k = f_k(u, s_k) \\ \boldsymbol{\mathcal{E}}[b_k(X_k)] \leq P_k \, \forall k}} I(X_1, X_2; Y | S_R), \quad (3)$$

where $U \in \mathcal{U}$ is an auxiliary random variable of finite cardinality, independent of (S, S_1, S_2, S_R), and where f_k is a deterministic functions $\mathcal{U} \times \mathcal{S}_k \to \mathcal{X}_k$ for $k = 1, 2$.

We conclude the first part of this paper by providing the following outer bound.

Proposition 1. *Under the Markov chain $S_1 \to S_R \to S_2$, $\mathscr{C}(P_1, P_2)$ is included in the convex hull of all rate triples satisfying*

$$R_1 + R_2 \leq I(U_1, U_2; Y | U_0, S_R),$$
$$R_0 + R_1 + R_2 \leq I(U_1, U_2; Y | S_R),$$

for some auxiliary variables $(U_0, U_1, U_2) \in \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2$ of finite cardinality, independent of the CSI (S, S_1, S_2, S_R), with pmf factorizing as $p(u_0)p(u_1|u_0)p(u_2|u_0)$, and for some deterministic functions $f_k : \mathcal{U}_k \times \mathcal{S}_k \to \mathcal{X}_k$, $x_k = f_k(u_k, s_k)$, satisfying $\boldsymbol{\mathcal{E}}[b_k(X_k)] \leq P_k$ for $k = 1, 2$.

Proof. The proof is given in Appendix C. ∎

Propostion 1 shows that, for (S_1, S_2) conditionally independent given S_R, Shannon strategies may be sub-optimal only in terms of individual rates (indeed, [25] proves that higher individual rates are achievable for independent (S_1, S_2) and $S_R = \emptyset$). This extends [21, Theorem 4], which considered independent (S_1, S_2) and no common message. The bound on $R_1 + R_2$ was already reported in [23] and references therein by using the same technique as [22]. Similarly to Theorem 2, our contribution lies in a simpler converse proof.

III. FDD Cooperative MIMO Channel with Fading

In this section, we specialize the channel in Fig. 1 to a practical 2×2 cooperative Gaussian MIMO channel with fading operating in FDD mode, illustrated in Fig. 2. The goal of this section is to particularize the general results of Section II and derive operational rules for encoding in the Gaussian MIMO setting. We point out that the results can be generalized to more general antenna configurations and number of TXs as discussed in Section III-E.
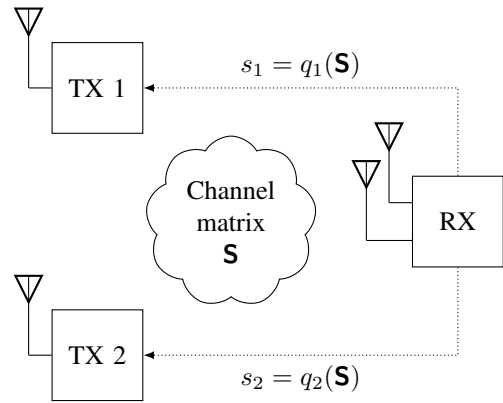

Fig. 2: Illustration of a cooperative MIMO channel with fading state matrix S known at the RX and distributed CSIT s_1, s_2 obtained from asymmetric feedback links.

A. Channel Model and Notation

For this second part of the article, we extend our notations to better highlight multi-dimensional quantities. In particular, we denote deterministic matrices, column vectors and scalars by X, $\boldsymbol{x}$, and x respectively. We denote by $\mathsf{X}_{i,j}$ the entry of X in the i-th column and j-th row. Random matrices, column vectors and scalars are denoted instead by $\mathbb{X}$, $\boldsymbol{X}$, and X respectively. The operators $(\cdot)^{\mathsf{H}}$ and $\|\cdot\|_{\mathrm{F}}$ denote respectively the Hermitian transpose and the Frobenius norm. We denote by $\boldsymbol{e}_i \in \{0, 1\}^d$ a standard column selector, i.e., with the i-th element set to 1 and all the other elements set to 0. The identity matrix of dimension n is denoted by I_n, or simply by I when the dimension is clear from the context. Finally, $\mathbb{C}$, $\mathbb{R}_+$, and $\mathbb{S}_+^n$ denote respectively the set of complex numbers, non-negative real numbers, and Hermitian positive-semidefinite matrices of dimension n.

We consider a classical Gaussian MIMO channel law $p(\boldsymbol{y}|x_1, x_2, \mathsf{S})$ and let the RX signal $\boldsymbol{Y} \in \mathbb{C}^{2 \times 1}$ for a given channel use be given by:

$$\boldsymbol{Y} = \mathbb{S}\boldsymbol{X} + \boldsymbol{Z} = \mathbb{S}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \boldsymbol{Z},$$

where the state $\mathbb{S} \in \mathbb{C}^{2 \times 2}$ is a matrix of random fading coefficients, $X_k \in \mathbb{C}$ is the signal transmitted by TX k, subject to an average power constraint $\boldsymbol{\mathcal{E}}[|X_k|^2] \leq P_k$, and where $\boldsymbol{Z} \sim \mathcal{CN}(\boldsymbol{0}, \mathsf{I}_2)$ is independent of $\mathbb{S}$. Furthermore, we focus on CSI distributions $p(\mathsf{S}, s_1, s_2, \mathsf{S}_R)$ where the RX has perfect CSIR $\mathbb{S}_R = \mathbb{S}$, and where the CSIT is a quantized version of the CSIR, i.e.,

$$S_k = q_k(\mathbb{S}), \quad k = 1, 2$$

$$q_k : \mathbb{C}^{2 \times 2} \to \mathcal{S}_k \triangleq \{1, \ldots, |\mathcal{S}_k|\}, \quad |\mathcal{S}_k| < \infty.$$

If $q_1 \neq q_2$, for example in the case of different feedback rates, we clearly fall into a distributed CSIT configuration. Note that, consistently with the definitions in Section II-A with S replaced by $\mathbb{S}$, we consider coding over a large number of i.i.d. fading realizations. By using classical wireless terminology, we recall that this corresponds to the so-called *fast-fading* regime.

B. Optimality of Distributed Linear Precoding with an Unconventional Number of Data Streams

In what follows, we establish the capacity region of the considered distributed setting and show that distributed linear precoding over Gaussian codewords is optimal. As we will see, the main novelty lies in an unconventional joint encoding technique for the common message W_0. For the sake of clarity, we present this technique by focusing on the common message capacity first.

Theorem 3. *The common message capacity of the channel in Section III-A is given by* $C_0(P_1, P_2) =$

$$\max_{\substack{\boldsymbol{g}_k(S_k) \in \mathbb{C}^{d'} \\ \boldsymbol{\mathcal{E}}\left[\|\boldsymbol{g}_k(S_k)\|^2\right] \leq P_k \, \forall k}} \boldsymbol{\mathcal{E}}\left[\log \det\left(\mathbf{I} + \mathbb{S}\boldsymbol{\Sigma}(S_1, S_2)\mathbb{S}^{\mathsf{H}}\right)\right], \quad (4)$$

where

$$\boldsymbol{\Sigma}(S_1, S_2) \triangleq \begin{bmatrix} \boldsymbol{g}_1^{\mathsf{H}}(S_1) \\ \boldsymbol{g}_2^{\mathsf{H}}(S_2) \end{bmatrix} \begin{bmatrix} \boldsymbol{g}_1(S_1) & \boldsymbol{g}_2(S_2) \end{bmatrix},$$

and where

$$d' \leq d \triangleq |\mathcal{S}_1| + |\mathcal{S}_2|.$$

Furthermore $C_0(P_1, P_2)$ *can be achieved by letting*

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{g}_1^{\mathsf{H}}(S_1) \\ \boldsymbol{g}_2^{\mathsf{H}}(S_2) \end{bmatrix} \boldsymbol{U}, \quad \boldsymbol{U} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{d'}), \quad (5)$$

where $\boldsymbol{U}$ *is the encoded common message.*

Proof: We first apply the well-known *maximum differential entropy lemma* [15, p. 21] to the mutual information in (3) using the conditional input covariance $\boldsymbol{\Sigma}(S_1, S_2) = \boldsymbol{\mathcal{E}}[\boldsymbol{X}\boldsymbol{X}^{\mathsf{H}}|S_1, S_2]$, obtaining the upper bound

$$R_0 \leq \boldsymbol{\mathcal{E}}\left[\log \det\left(\mathbf{I} + \mathbb{S}\boldsymbol{\Sigma}(S_1, S_2)\mathbb{S}^{\mathsf{H}}\right)\right].$$

For the achievability part, it suffices to show that every feasible $\boldsymbol{\Sigma}(S_1, S_2)$ can be obtained via distributed linear precoders of dimension $d = |\mathcal{S}_1| + |\mathcal{S}_2|$. To this end, we first define the random vectors

$$\tilde{\boldsymbol{X}}_k \triangleq \begin{bmatrix} f_k(U, 1) \\ \vdots \\ f_k(U, |\mathcal{S}_k|) \end{bmatrix}, \quad k = 1, 2,$$

collecting the random inputs X_k conditioned on each of the $|\mathcal{S}_k|$ realizations of S_k, and the covariance matrix

$$\mathbf{Q} \triangleq \boldsymbol{\mathcal{E}}\left[\begin{bmatrix} \tilde{\boldsymbol{X}}_1 \\ \tilde{\boldsymbol{X}}_2 \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{X}}_1^{\mathsf{H}} & \tilde{\boldsymbol{X}}_2^{\mathsf{H}} \end{bmatrix}\right] \in \mathbb{S}_+^{d \times d}.$$

It is easy to see that $\mathbf{Q}$ contains all the elements of $\boldsymbol{\Sigma}(s_1, s_2)$, $\forall(s_1, s_2) \in \mathcal{S}_1 \times \mathcal{S}_2$. Note that, due to the power constraint $\sum_{s_k \in \mathcal{S}_k} \boldsymbol{\mathcal{E}}[|f_k(U, s_k)|^2]p(s_k) \leq P_k < \infty$, any feasible (f_1, f_2, p_U) must satisfy $\boldsymbol{\mathcal{E}}[|f_k(U, s_k)|^2] < \infty \, \forall s_k \in \mathcal{S}_k$, hence $\mathbf{Q}$ has finite entries. Since $\mathbf{Q}$ is Hermitian positive semi-definite, there exists a square matrix $\mathbf{F} \in \mathbb{C}^{d \times d}$ such that $\mathbf{F}^{\mathsf{H}}\mathbf{F} = \mathbf{Q}$. We denote its column vectors by

$$\mathbf{F} \triangleq \begin{bmatrix} \boldsymbol{g}_1(1) & \dots & \boldsymbol{g}_1(|\mathcal{S}_1|) & \boldsymbol{g}_2(1) & \dots & \boldsymbol{g}_2(|\mathcal{S}_2|) \end{bmatrix}. \quad (6)$$

Finally, simple calculations show that the scheme in (5) with distributed linear precoders designed using the above procedure, i.e., selected from (6), preserves the desired $\boldsymbol{\Sigma}(S_1, S_2)$, and attains the maximum entropy upper bound. $\blacksquare$

The main result of Theorem 3 is that distributed linear precoding [8] of shared Gaussian codewords achieves the performance limits of the considered cooperative MIMO setting. However, as a sufficient condition to prove achievability, Theorem 3 considers the transmission of possibly $d = |\mathcal{S}_1| + |\mathcal{S}_2|$ independent data streams. This unconventional design choice appears to be in sharp contrast with the centralized CSIT configuration (i.e., $S_1 = S_2 \triangleq S_T$), where the capacity of the 2×2 MIMO channel is achieved by encoding $d' \leq 2$ streams in the presence of perfect message cooperation. In this latter case, by considering the per-antenna power constraint, the capacity takes the well-known expression given for example by [32]

$$C_0(P_1, P_2) = \max_{\substack{\boldsymbol{\Sigma}(S_T) \in \mathbb{S}_+^2, \, \forall k \\ \boldsymbol{\mathcal{E}}[\Sigma_{k,k}(S_T)] \leq P_k}} \boldsymbol{\mathcal{E}}\left[\log \det\left(\mathbf{I} + \mathbb{S}\boldsymbol{\Sigma}(S_T)\mathbb{S}^{\mathsf{H}}\right)\right], \quad (7)$$

where $\boldsymbol{\Sigma}(s_T) \triangleq \boldsymbol{\mathcal{E}}[\boldsymbol{X}\boldsymbol{X}^{\mathsf{H}}|S_T]$ is the conditional input covariance. Clearly, the capacity in (7) can be achieved by taking the matrix square-root $\mathbf{G}(S_T) \triangleq \boldsymbol{\Sigma}^{\frac{1}{2}}(S_T) \in \mathbb{C}^{2 \times 2}$ and by letting

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \mathbf{G}^{\mathsf{H}}(S_T)\boldsymbol{U}, \quad \boldsymbol{U} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_2),$$

or, in other words, by precoding $d' = 2$ data streams only. Such an approach cannot be used for general distributed settings, as it generally leads to unfeasible linear precoders violating the functional dependencies $x_k = f_k(\boldsymbol{u}, s_k)$.

The proof of Theorem 3 addresses this issue by *increasing the dimensionality* d' of the linear precoders up to $d \gg 2$, i.e., beyond conventional design choices. We now show that this unconventional linear precoding technique can be applied to extend Theorem 3 to the full capacity region.

Theorem 4. *The capacity region* $\mathscr{C}(P_1, P_2)$ *of the channel in Section III-A is given by the union of all rate triples* (R_0, R_1, R_2) *such that*

$$R_1 \leq \boldsymbol{\mathcal{E}}\left[\log\left(1 + \gamma_1(S_1)\|\mathbb{S}\boldsymbol{e}_1\|^2\right)\right]$$
$$R_2 \leq \boldsymbol{\mathcal{E}}\left[\log\left(1 + \gamma_2(S_2)\|\mathbb{S}\boldsymbol{e}_2\|^2\right)\right]$$
$$R_1 + R_2 \leq \boldsymbol{\mathcal{E}}\left[\log \det\left(\mathbf{I} + \mathbb{S}\mathrm{diag}(\gamma_1(S_1), \gamma_2(S_2))\mathbb{S}^{\mathsf{H}}\right)\right]$$
$$R_0 + R_1 + R_2 \leq \boldsymbol{\mathcal{E}}\left[\log \det\left(\mathbf{I} + \mathbb{S}\boldsymbol{\Sigma}(S_1, S_2)\mathbb{S}^{\mathsf{H}}\right)\right]$$

where

$$\boldsymbol{\Sigma}(S_1, S_2) = \begin{bmatrix} \boldsymbol{g}_1^{\mathsf{H}}(S_1) \\ \boldsymbol{g}_2^{\mathsf{H}}(S_2) \end{bmatrix} \begin{bmatrix} \boldsymbol{g}_1(S_1) \, \boldsymbol{g}_2(S_2) \end{bmatrix} + \begin{bmatrix} \gamma_1(S_1) & 0 \\ 0 & \gamma_2(S_2) \end{bmatrix},$$

for some $\boldsymbol{g}_k(S_k) \in \mathbb{C}^{d'}$ *and* $\gamma_k(S_k) \in \mathbb{R}_+$ *such that*

$$d' \leq d \triangleq |\mathcal{S}_1| + |\mathcal{S}_2|$$

and $\boldsymbol{\mathcal{E}}\left[\|\boldsymbol{g}_k(S_k)\|^2\right] + \boldsymbol{\mathcal{E}}[\gamma_k(S_k)] \leq P_k$ *for* $k = 1, 2$. *Furthermore any point in* $\mathscr{C}(P_1, P_2)$ *can be achieved by letting*

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{g}_1^{\mathsf{H}}(S_1) & \sqrt{\gamma_1(S_1)} & 0 \\ \boldsymbol{g}_2^{\mathsf{H}}(S_2) & 0 & \sqrt{\gamma_2(S_2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{U} \\ V_1 \\ V_2 \end{bmatrix}, \quad (8)$$

where $(U, V_1, V_2) \sim \mathcal{CN}(0, I_{d'+2})$ *are the encoded common and private messages, respectively.*

Proof. The proof is given in Appendix D. ∎

Theorem 4 shows that superposition of jointly and independently encoded Gaussian codes achieves the capacity region. However, while encoding of the private messages W_1, W_2 follows traditional approaches (one power-controlled stream per TX), the joint encoding of W_0 may require a larger number of precoded data streams (beyond two streams in our case) as already seen for Theorem 3. A more detailed analysis of the role played by d' in terms of optimal distributed precoding is provided in the following sections.

C. Convex Reformulation for Optimal Distributed Precoding Design

The distributed precoding design problem (4) belongs to the class of *static team decision* problems [8], [33], for which no efficient solutions are known in general. However, in contrast to the traditional precoding design with $d' \leq N_T$, where $N_T = 2$ is the total number of TX antennas, by letting $d' \leq d$ we are able to recast the optimal precoding design problem (4) into an equivalent *convex* problem.

Proposition 2. *Problem* (4) *is equivalent to the following convex problem*

$$\underset{Q \in \mathbb{S}_+^d}{maximize} \quad \mathcal{E}\left[\log\det\left(I + \mathbb{S}_{eq} Q \mathbb{S}_{eq}^H\right)\right] \quad subject\ to$$

$$\sum_{i=1}^{|\mathcal{S}_1|} Q_{i,i} \Pr(S_1 = i) \leq P_1, \qquad (9)$$

$$\sum_{j=|\mathcal{S}_1|+1}^{|\mathcal{S}_1|+|\mathcal{S}_2|} Q_{j,j} \Pr(S_2 = j - |\mathcal{S}_1|) \leq P_2,$$

where we defined $\mathbb{S}_{eq} \triangleq \mathbb{S} E^H(S_1, S_2) \in \mathbb{C}^{2 \times d}$, *and*

$$E(S_1, S_2) \triangleq \begin{bmatrix} e_{S_1} & 0 \\ 0 & e_{S_2} \end{bmatrix} \in \mathbb{C}^{d \times 2},$$

where $e_{S_1} \in \{0,1\}^{|\mathcal{S}_1|}$ *and* $e_{S_2} \in \{0,1\}^{|\mathcal{S}_2|}$ *are standard column selectors.*

Proof: The proof follows by simply rewriting (4) in light of the technique used for the proof of Theorem 3. Specifically, we let $d' = d$ and use $\mathbb{S}\Sigma(S_1, S_2)\mathbb{S}^H = \mathbb{S}E^H(S_1, S_2)F^H F E(S_1, S_2)\mathbb{S}^H = \mathbb{S}_{eq} Q \mathbb{S}_{eq}^H$, where F is given by (6) and $Q = F^H F$. The power constraints correspond to linear constraints on the diagonal elements of Q. ∎

Problem (9) corresponds to the capacity of a virtual $d \times 2$ MIMO channel with state $\mathbb{S}_{eq}$, perfect CSIR, no CSIT, and (fixed) transmit covariance Q. The capacity achieving distributed precoders for the original channel can be then designed from the optimal $Q^\star$ in Problem (9) as follows:

$$\begin{bmatrix} g_1^\star(S_1) & g_2^\star(S_2) \end{bmatrix} = (Q^\star)^{\frac{1}{2}} E(S_1, S_2) \in \mathbb{C}^{d \times 2}. \qquad (10)$$

An important remark here is that if the constraint $d' \leq d$ of Problem (4) is replaced by $d' < d$, the technique of Proposition 2 does not lead to a convex reformulation. This is because

the $d \times d$ matrix F is replaced by a $d' \times d$ matrix F', hence introducing a non-convex constraint $\mathrm{rank}(Q) \leq d' < d$ to Problem (9). However, note that if the optimal $Q^\star$ for the unconstrained problem has rank $r < d$, then we can reduce with no loss of optimality the dimensionality of $g_k^\star(S_k)$ in (10) down to $d' = r$.

Mirroring the previous section, the above result on common message capacity can be extended to the following weighted sum-rate maximization problem

$$\underset{(R_0, R_1, R_2) \in \mathscr{C}(P_1, P_2)}{maximize} \alpha_0 R_0 + \alpha_1 R_1 + \alpha_2 R_2, \qquad (11)$$

where $\alpha_k \geq 0$, $k = 0, 1, 2$ are non-negative weights identifying rate priorities. We recall that the above problem can be used to characterize the boundary of $\mathscr{C}(P_1, P_2)$, since the weights can be interpreted as coefficients of a supporting hyperplane to such boundary (see e.g. [15], [34] and references therein).

Proposition 3. *Problem* (11) *is equivalent to the following convex problem*

$$\underset{\substack{R_k \in \mathbb{R}_+, Q \in \mathbb{S}_+^d \\ \gamma_1(S_1), \gamma_2(S_2) \in \mathbb{R}_+}}{maximize} \alpha_0 R_1 + \alpha_1 R_1 + \alpha_2 R_2 \quad subject\ to$$

$$R_1 \leq \mathcal{E}\left[\log\left(1 + \gamma_1(S_1)\|\mathbb{S}e_1\|^2\right)\right],$$

$$R_2 \leq \mathcal{E}\left[\log\left(1 + \gamma_2(S_2)\|\mathbb{S}e_2\|^2\right)\right],$$

$$R_1 + R_2 \leq \mathcal{E}\left[\log\det\left(I + \mathbb{S}\mathrm{diag}(\gamma_1(S_1), \gamma_2(S_2))\mathbb{S}^H\right)\right],$$

$$R_0 + R_1 + R_2 \leq$$

$$\mathcal{E}\left[\log\det\left(I + \mathbb{S}_{eq} Q \mathbb{S}_{eq}^H + \mathbb{S}\mathrm{diag}(\gamma_1(S_1), \gamma_2(S_2))\mathbb{S}^H\right)\right],$$

$$\sum_{i=1}^{|\mathcal{S}_1|} (Q_{i,i} + \gamma_1(i)) \Pr(S_1 = i) \leq P_1,$$

$$\sum_{j=|\mathcal{S}_1|+1}^{|\mathcal{S}_1|+|\mathcal{S}_2|} (Q_{j,j} + \gamma_2(j - |\mathcal{S}_1|)) \Pr(S_2 = j - |\mathcal{S}_1|) \leq P_2,$$

$$(12)$$

where $\mathbb{S}_{eq}$ *is given as in Proposition 2.*

Proof: The proof follows by the same technique as in the proof of Proposition 2. The details are omitted. ∎

Problems (9) and (12) can be solved numerically via known convex optimization tools. A comprehensive discussion on the efficiency of various competing approaches is out of the scope of this work. Here, we point out two critical issues that should be taken into account in a practical system design. First, advanced stochastic optimization techniques may be required if the fading distribution $p(S)$ is continuous[2]. Second, classical second-order methods as interior-point methods for semi-definite optimization typically scale badly with the dimension d of Q. Hence, first-order methods may be more suitable whenever the cardinality of the CSIT alphabets $|\mathcal{S}_k|$ is large. As a result of the algorithmic complexity stemming out of the above considerations (which are still very active research topics), we envision that feasible implementations of the proposed distributed precoding design should operate in an *offline* fashion. Specifically, Problems (9) and (12) could be solved in a preliminary codebook design phase, while in

---

[2]Note that all the results presented in this section do not require $\mathcal{S}$ to be a discrete set, but only $|\mathcal{S}_k| < \infty$ for $k = 1, 2$.

the data transmission phase TX k simply selects the precoder from the pre-designed codebook based on the received CSIT index S_k.

D. Further Comments on the Optimal Number of Data Streams

In this section we further elaborate on the optimal number of data streams d' by addressing two important questions left open by Section III-B. Theorem 3 shows that using a number of precoded data streams $d' = d$ is a sufficient condition for achievability of the common message capacity. However, we know that for some CSIT configurations (e.g., for centralized CSIT, as already discussed), $d' > 2$ is not necessary. A first crucial question is whether there exists some distributed CSIT configuration for which such a condition is indeed necessary. In the next proposition we answer positively to this question.

Proposition 4. *For some $p(\mathbf{S}, s_2, s_2)$ and power constraints (P_1, P_2), restricting $d' \leq N_T$ in problem (4), where $N_T = 2$ is the total number of TX antennas, leads to strictly suboptimal rates.*

Proof: The proof is given in Appendix E, by showing the existence of a CSI distribution $p(\mathbf{S}, s_2, s_2)$ with binary CSIT $|\mathcal{S}_1| = |\mathcal{S}_2| = 2$ such that $d' \geq 3$ is necessary for achieving $C_0(P_1, P_2)$. ∎

A second natural question is whether the developed upper bound $d' \leq d = |\mathcal{S}_1| + |\mathcal{S}_2|$ is tight, for some $p(\mathbf{S}, s_1, s_2)$. In the following we answer negatively to this question, by showing that indeed we can consider a slightly tighter upper bound. However, we firstly remark that obtaining tighter bounds is not trivial and is in fact related to the well-known low-rank matrix completion problem [35]. Let us consider the matrix $\mathbf{Q} \in \$^d_+$ defined in the proof of Theorem 3, or equivalently in Proposition 2, and its partition into blocks

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_{12} \\ \mathbf{Q}_{12}^{\mathsf{H}} & \mathbf{Q}_2 \end{bmatrix}, \quad \mathbf{Q}_k \in \$^{|\mathcal{S}_k|}_+. \tag{13}$$

Informally, we recall that $\mathbf{Q}$ collects the elements of the conditional input covariances $\mathbf{\Sigma}(S_1, S_2)$ for all realizations of (S_1, S_2). By direct inspection of the capacity expression (4), or equivalently of the objective in Problem (9), we observe that the off-diagonal elements of the sub-matrices $\mathbf{Q}_k$ do not contribute to the achievable rate, since they do not correspond to any element of any realization of $\mathbf{\Sigma}(S_1, S_2)$. Hence, by letting $\tilde{\mathbf{Q}}$ be any optimal solution of (9), the solution $\mathbf{Q}^\star$ of the (non-convex) problem

$$\begin{aligned} \underset{\mathbf{Q} \in \$^d_+}{\text{minimize}} \quad & \text{rank}(\mathbf{Q}) \\ \text{subject to} \quad & \mathbf{Q}_{12} = \tilde{\mathbf{Q}}_{12} \\ & \mathbf{Q}_{i,i} = \tilde{\mathbf{Q}}_{i,i}, \quad i = 1, \ldots, d \end{aligned} \tag{14}$$

is also an optimal solution of (9), but where the off-diagonal elements of $\mathbf{Q}_k$ have been optimized such that the rank is minimized. Since we have seen that the rank $r \leq d$ of $\mathbf{Q}^\star$ corresponds to the dimension d' of optimal distributed precoders (see Section III-C), establishing a tighter upper-bound on d' can be cast into finding an upper-bound on the

solution of (14), which is an instance of a low-rank (semidefinite) matrix completion problem (see, e.g., [35]).

To the best of the authors knowledge, non-trivial upper-bounds to problems of the type (14) remain elusive. Nevertheless, in the following proposition we provide a simple result showing the existence of a tighter upper-bound than $d = |\mathcal{S}_1| + |\mathcal{S}_2|$.

Proposition 5. *The common message capacity $C_0(P_1, P_2)$ (resp. capacity region $\mathscr{C}(P_1, P_2)$) given by Theorem 3 (resp. Theorem 4) is also achievable by letting*

$$d' \leq |\mathcal{S}_1| + |\mathcal{S}_2| - 1.$$

Proof. The proof is given in Appendix F. ∎

The above result is by no means satisfactory, since the dimensionality reduction is marginal for large CSIT alphabets. Informally, the main limitation of the above bound is that the proof optimizes only one of the (coupled) variables in (14). However, note that the above bound is tight for the toy example considered in the proof of Proposition 4.

E. Extension to Arbitrary Users and Antenna Configurations

Theorem 3 can be readily extended to K TXs and arbitrary antenna configuration. By letting N_k and M be respectively the number of antennas at the k-th TX and at the RX, and by considering a fading state matrix $\mathbb{S} \in \mathbb{C}^{M \times (\sum_{k=1}^K N_k)}$ and distributed CSIT $(S_1, \ldots, S_K)$, $S_k = q_k(\mathbb{S}) \in \mathcal{S}_k$, $|\mathcal{S}_k| < \infty$, it can be shown that the common message capacity is given by $C_0(P_1, \ldots, P_K) =$

$$\max_{\substack{\mathbf{G}_k(S_k) \in \mathbb{C}^{d' \times N_k} \\ \mathcal{E}\left[\|\mathbf{G}_k(S_k)\|_{\mathsf{F}}^2\right] \leq P_k, \forall k}} \mathcal{E}\left[\log \det\left(\mathbf{I} + \mathbb{S}\mathbf{\Sigma}(S_1, \ldots, S_K)\mathbb{S}^{\mathsf{H}}\right)\right],$$
$$\tag{15}$$

where

$$\mathbf{\Sigma}(S_1, \ldots, S_K) = \begin{bmatrix} \mathbf{G}_1^{\mathsf{H}}(S_1) \\ \vdots \\ \mathbf{G}_K^{\mathsf{H}}(S_K) \end{bmatrix} \begin{bmatrix} \mathbf{G}_1(S_1) & \ldots & \mathbf{G}_K(S_K) \end{bmatrix},$$

and where $d' \leq d \triangleq \sum_{k=1}^K N_k |\mathcal{S}_k|$, and it is achievable by distributed linear precoding of d' i.i.d. Gaussian codewords.

The formal proof of the above statement follows from the same lines as for the proof of Theorem 3, by considering K functions $f_k : \mathcal{U} \times \mathcal{S}_k \to \mathbb{C}^{N_k \times 1}$. The detailed steps do not provide additional intuitions, hence they are omitted. The convex reformulation of Proposition 2 can be also similarly extended. Finally, as for Theorem 4, the full-capacity region for K private messages and a single common message can be achieved by superimposing standard MIMO MAC codes for $W_1, \ldots, W_K$ to the non-traditional distributed precoding technique achieving (15).

IV. NUMERICAL EXAMPLES

A. Channel with Additive Binary Inputs and State

We consider the following channel

$$Y = X_1 + X_2 + S,$$

with binary inputs and state, i.e., $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{S} = \{0, 1\}$, and where $\mathcal{Y} = \{0, 1, 2, 3\}$. We do not consider input cost constraints. We further assume $S \sim \text{Bernoulli}(q)$, no CSIR ($\mathcal{S}_R = \emptyset$), and distributed CSIT $p(s_1, s_2 | s) = p(s_1 | s) p(s_2 | s)$, where $p(s_k | s)$ is a binary symmetric channel with transition probability $\epsilon_k \in [0, 0.5]$. Under the above model, the common message capacity (which coincides with the sum capacity) is given by

$$C_0 = \max_{\substack{p(u) \\ x_k = f_k(u, s_k)}} I(U; Y).$$

A (non-scalable) method for optimally solving the above optimization problem is to adapt to the considered distributed setting the original idea of coding over the *alphabet of Shannon strategies* [14], [15], combined with classical results on the computation of the capacity of point-to-point channels [36]. More precisely, we proceed as follows:

1) We build the alphabet of *distributed Shannon strategies* by enumerating all the functions

$$t_u(s_1, s_2) = [t_{1,u}(s_1), t_{2,u}(s_2)], \quad t_{k,u} : \mathcal{S}_k \to \mathcal{X}_k,$$

where each function is indexed by U. There are $|\mathcal{U}| = |\mathcal{X}_1|^{|\mathcal{S}_1|} |\mathcal{X}_2|^{|\mathcal{S}_2|} = 16$ such functions.

2) We set $x_k = f_k(u, s_k) = t_{k,u}(s_k)$ and compute the equivalent state-less point-to-point channel

$$p(y|u) = \sum_{\substack{x_1, x_2 \\ s, s_1, s_2}} p(y|x_1, x_2, s) p(x_1|u, s_1) p(x_2|u, s_2) p(s, s_1, s_2).$$

3) We run the Blahut-Arimoto algorithm for computing the capacity of the equivalent channel $p(y|u)$ [36].

Note that the above procedure is similar to the one outlined in [37] for centralized settings. Furthermore, it can be readily generalized to arbitrary CSIR by simply considering an augmented output $\tilde{Y} \triangleq (Y, S_R)$.

In Fig. 3 we plot the capacity C_0 versus the CSIT quality at TX 2, for various choices of CSIT quality at TX 1, and for $q = 0.5$. Note that $\epsilon_k = 0$ and $\epsilon_k = 0.5$ model respectively perfect and no CSIT at the k-th TX. Interestingly, the capacity of the system decreases with ϵ_2 down to a flat regime in which any further decrease in quality does not matter, and the turning point depends on ϵ_1. This can be interpreted as a regime in which the quality at one TX is so degraded that, although some CSIT is available, it does not allow for proper coordination with the better informed TX. Intuitively, it is important for the less informed TX to not act as unknown noise for the other TX. In fact, in the aforementioned regime it turns out that the optimal scheme at the less informed TX is to throw away completely its CSIT, making its behaviour not adaptive to the channel conditions but completely predictable by the more informed TX.

B. Cooperative AWGN MIMO with Rayleigh Fading and Quantized Feedback

In this section we simulate a practical cooperative MIMO channel with Rayleigh fading and with limited feedback rates. In particular, we let each element of $\mathbb{S}$ to be i.i.d. $\mathcal{CN}(0, 1)$, and
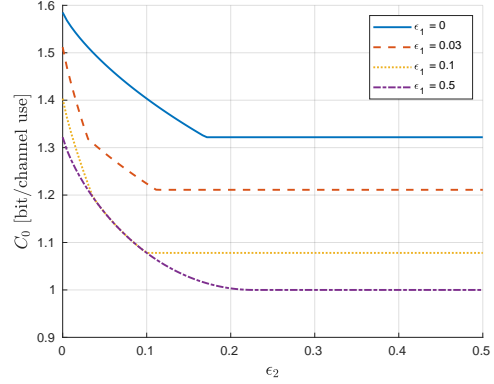

Fig. 3: Capacity vs. CSIT distortion ϵ_2 at TX 2, for various choices of CSIT distortion ϵ_1 at TX 1.

we set for simplicity $P_1 = P_2 \triangleq \text{SNR}$. The distributed CSIT configuration $p(\mathbf{S}, s_1, s_2)$ is given by two random quantizers with different rates β_1, β_2.

More precisely, let $\mathcal{S}_k = \{1, \ldots, 2^{\beta_k}\}$ be the index set of a codebook $\{\hat{\mathbf{S}}_{k,i}\}_{i=1}^{2^{\beta_k}}$ of randomly and independently generated codewords distributed as $p(\mathbf{S})$. We then let $q_k(\mathbf{S})$ to be a simple nearest neighbour vector quantizers in the Frobenius norm, i.e., $q_k(\mathbf{S}) = \arg\min_{i \in \mathcal{S}_k} \|\mathbf{S} - \hat{\mathbf{S}}_{k,i}\|_F$. This scenario corresponds to an error-free feedback link from the RX to the k-th TX with limited rate of β_k bits per channel realization. We set $\beta_1 = 4$ and $\beta_2 = 3$, which implies $d = |\mathcal{S}_1| + |\mathcal{S}_2| = 24$. We recall that the RX is assumed to have perfect CSIR.

We approximately solve Problem (9) through an off-the-shelf numerical solver for convex problems, by substituting $p(\mathbf{S}, s_1, s_2)$ with its empirical distribution $\hat{p}(\mathbf{S}, s_1, s_2) \triangleq \frac{1}{L} \sum_{l=1}^{L} \mathbb{1}[(\mathbf{S}, s_1, s_2) = (\mathbf{S}_l, q_1(\mathbf{S}_l), q_2(\mathbf{S}_l))]$ obtained from $L = 1000$ i.i.d. samples $\{\mathbf{S}_l\}_{l=1}^{L}$. This allows us to replace the expectation in (9) with a finite sum of L convex functions. The capacity obtained is exact for a channel with state distribution equal to the empirical distribution $\hat{p}(\mathbf{S}, s_1, s_2)$, and approximates the capacity for $p(\mathbf{S}, s_1, s_2)$ as L grows large. Furthermore, we repeat the above simulations by considering instead a single antenna at the RX, a setting denoted here as *cooperative MISO*.

In Fig. 4 we plot the common message capacity versus SNR of a given instance of the considered channel model. We also plot the common message capacity for perfect CSIT at both TXs

$$C_0^{(\text{perf. CSIT})} = \max_{\substack{\boldsymbol{\Sigma}(\mathbb{S}) \in \mathbb{S}_+^2 : \\ \mathcal{E}[\Sigma_{k,k}(\mathbb{S})] \le P_k \ \forall k}} \mathcal{E}\left[\log\det\left(\mathbf{I} + \mathbb{S}\boldsymbol{\Sigma}(\mathbb{S})\mathbb{S}^H\right)\right],$$

and for no CSIT

$$C_0^{(\text{no CSIT})} = \max_{\boldsymbol{\Sigma} \in \mathbb{S}_+^2 : \Sigma_{k,k} \le P_k \ \forall k} \mathcal{E}\left[\log\det\left(\mathbf{I} + \mathbb{S}\boldsymbol{\Sigma}\mathbb{S}^H\right)\right].$$

We recall that these CSIT configurations are equivalent to a centralized 2×2 MIMO system, hence we can simply use the classical MIMO results summarized e.g. in [32], adapted to a per-antenna power constraint. For a fair comparison, these capacities are computed over the same empirical marginal distribution $\hat{p}(\mathbf{S})$. As expected, for the MIMO case, the capacity
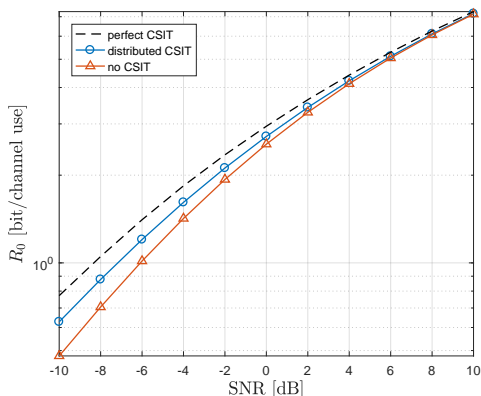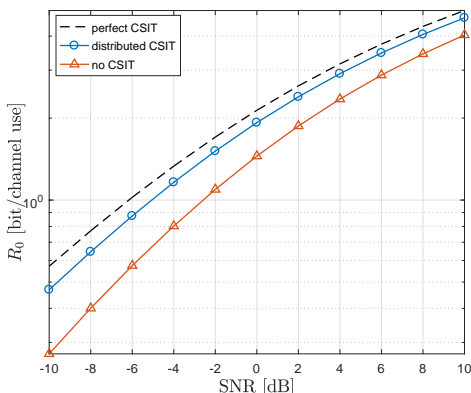
(a) 2×2 Cooperative MIMO

(b) 2×1 Cooperative MISO

Fig. 4: Common message capacity (4) vs SNR for (a) 2 RX antennas and (b) single RX antenna.

gain given by distributed CSIT w.r.t no CSIT follows the well-known *beamforming gain* trend of the perfect CSIT case, i.e., it vanishes in the high SNR regime. Similarly, for the MISO case, this gain converges to a constant power offset.

V. CONCLUSION

In this paper, we studied a two-user memoryless state-dependent multiple access channel under the assumption that causal and distributed CSIT is available and messages can be partly or entirely shared prior to the data transmission. We characterized the common message capacity of this channel and demonstrated that it is optimal to encode the message as a function of current CSIT only based on Shannon strategies. We provide an insightful example over an additive binary-input quaternary-output channel with binary states showing that, interestingly, in some cases there is a threshold in terms of CSIT quality below which one encoder shall not use its channel knowledge. For a special case when CSIT is a deterministic function of CSIR, the full capacity region of a common message and two private messages is also characterized. This last result is specialized to a practically relevant cooperative MIMO fading channel operating in FDD mode such that CSIT is acquired via an explicit feedback from the receiver. The cooperative MIMO example surprisingly

reveals that in a distributed CSIT setup the optimal number of data streams shall not be restricted to the minimum number of transmit antennas. This is in contrast to the classical MIMO design under the centralized CSIT assumption.

Interesting open problems include the evaluation of the minimum message cooperation (i.e., the minimum rate R_0) required such that the sum capacity is achievable via Shannon strategies, and the extension of the coding ideas derived for the cooperative MIMO case to systems with multiple receivers.

APPENDIX

A. Proof of Theorem 1

Converse: Let us define $U_i = (W_0, S_1^{i-1}, S_2^{i-1})$. We construct an upper-bound by assuming that past CSIT realizations (S_1^{i-1}, S_2^{i-1}) are available at both encoders. Hence, we assume that X_{1i} and X_{2i} are functions of $(W_0, S_1^i, S_2^{i-1}) = (U_i, S_{1i})$ and $(W_0, S_2^i, S_1^{i-1}) = (U_i, S_{2i})$ respectively. Note that U_i is independent of $(S_i, S_{1i}, S_{2i}, S_{Ri})$. Consider for brevity $\tilde{Y}_i = (Y_i, S_{Ri})$. We then have:

$$
\begin{aligned}
nR_0 &= H(W_0) \\
&= I(W_0; \tilde{Y}^n) + H(W_0|\tilde{Y}^n) \\
&\overset{(a)}{\le} I(W_0; \tilde{Y}^n) + n\epsilon_n \\
&= \sum_{i=1}^n I(W_0; \tilde{Y}_i|\tilde{Y}^{i-1}) + n\epsilon_n \\
&= \sum_{i=1}^n H(\tilde{Y}_i|\tilde{Y}^{i-1},) - H(\tilde{Y}_i|W_0, \tilde{Y}^{i-1}) + n\epsilon_n \\
&\le \sum_{i=1}^n H(\tilde{Y}_i|\tilde{Y}^{i-1}) - H(\tilde{Y}_i|U_i, \tilde{Y}^{i-1}) + n\epsilon_n \\
&\overset{(b)}{=} \sum_{i=1}^n H(\tilde{Y}_i|\tilde{Y}^{i-1}) - H(\tilde{Y}_i|U_i) + n\epsilon_n \\
&\le \sum_{i=1}^n H(\tilde{Y}_i) - H(\tilde{Y}_i|U_i) + n\epsilon_n \\
&= \sum_{i=1}^n I(U_i; \tilde{Y}_i) + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^n I(U_i; Y_i|S_{Ri}) + n\epsilon_n
\end{aligned}
\tag{16}
$$

where (a) follows from Fano's inequality ($\lim_{n\to\infty} \epsilon_n = 0$), (b) follows from the Markov chain $\tilde{Y}^{i-1} \to (W_0, S_1^{i-1}, S_2^{i-1}) \to \tilde{Y}_i$, and (c) is because S_{Ri} is independent of U_i. The code must also satisfy the input cost constraints

$$
P_1 \ge \mathcal{E}\left[\frac{1}{n}\sum_{i=1}^n b_1(X_{1i})\right], \quad P_2 \ge \mathcal{E}\left[\frac{1}{n}\sum_{i=1}^n b_2(X_{2i})\right].
\tag{17}
$$

We combine the bounds in (16) and (17) by means of a time-sharing variable Q uniformly distributed in $\{1, \ldots, n\}$ and independent of everything else, and by letting $U = (U_Q, Q)$, $X_1 = X_{1Q}$, $X_2 = X_{2Q}$ $Y = Y_Q$, $S = S_Q$, $S_1 = S_{1Q}$, $S_2 = S_{2Q}$, $S_R = S_{RQ}$. Note that the resulting distribution $p_{Y,X_1,X_2,S,S_1,S_2,S_R,U}$ factors as $p(y|x_1, x_2, s)\mathbb{1}[x_1 -$

$f_1(u, s_1)]\mathbb{1}[x_2 - f_2(u, s_2)]p(s, s_1, s_2, s_R)p(u)$, where $\mathbb{1}[\cdot]$ is an indicator function. With these identifications, we obtain

$$
\begin{aligned}
R_0 &\leq I(U; Y | S_R, Q) + \epsilon_n \\
&\leq H(Y | S_R) - H(Y | U, S_R, Q) + \epsilon_n \\
&\overset{(c)}{=} H(Y | S_R) - H(Y | U, S_R) + \epsilon_n \\
&= I(U; Y | S_R) + \epsilon_n,
\end{aligned}
$$

where (c) follows from the Markov chain $Q \to (U, S_R) \to Y$, and $P_1 \geq \mathcal{E}[b_1(X_1)]$, $P_2 \geq \mathcal{E}[b_2(X_2)]$. Hence, we finally have $R_0 \leq C_0(P_1, P_2) + \epsilon_n$. ∎

Achievability: Achievability follows from standard arguments, hence a formal proof is omitted. An informal yet intuitive proof can be obtained by the classical *physical device* argument of Shannon [14], [15]. More precisely, by fixing the functions $f_k(u_k, s_k)$, we can consider a new state-less and memoryless MAC with messages W_0, W_1, W_2, inputs U_1, U_2, and output (Y, S_R). For a given $f_k(u_k, s_k)$, the capacity region of this auxiliary channel is simply achievable by Slepian-Wolf coding for the MAC with common and independent messages [31], which gives the achievable region in Lemma 1.

The finite cardinality of U_1, U_2 follows directly by Shannon argument, which states that using Shannon strategies corresponds to coding over an augmented input alphabet of functions $\mathcal{S}_k \to \mathcal{X}_k$ of size $|\mathcal{X}_k|^{|\mathcal{S}_k|}$, indexed by U_k [15]. Hence we can consider $|\mathcal{U}_k| \leq |\mathcal{X}_k|^{|\mathcal{S}_k|}$. The finite cardinality of $\mathcal{U}_0$ follows by a simple application of the support lemma [15, Appendix C] applied to the Slepian-Wolf region of the auxiliary channel, which gives [15, p. 344]

$$
|\mathcal{U}_0| \leq \min\{|\mathcal{X}_1|^{|\mathcal{S}_1|}|\mathcal{X}_2|^{|\mathcal{S}_2|} + 2, |\mathcal{Y}||\mathcal{S}_R| + 3\}.
$$

Finally, the expression in Theorem 1 can be obtained by specializing the proof of Lemma 1 to the transmission of a common message only, i.e. by letting $R_1 = R_2 = 0$, and by identifying $U_0 = U_1 = U_2 \overset{\Delta}{=} U$. ∎

B. Proof of Theorem 2

Achievability: The proof follows the same lines as in [21]. Achievability builds on Lemma 1, where we rewrite the mutual information terms as follows. By focusing first on the sum-rate, we observe that

$$
\begin{aligned}
I(U_1, U_2; Y | S_R) &= \\
&= H(Y | S_R) - H(Y | U_1, U_2, S_R) \\
&\overset{(a)}{=} H(Y | S_R) - H(Y | U_1, U_2, S_1, S_2, S_R) \\
&\overset{(b)}{=} H(Y | S_R) - H(Y | X_1, X_2, U_1, U_2, S_1, S_2, S_R) \\
&\overset{(c)}{=} H(Y | S_R) - H(Y | X_1, X_2, S_R) \\
&= I(X_1, X_2; Y | S_R),
\end{aligned}
$$

where (a) comes from $(S_1, S_2) = (q_1(S_R), q_2(S_R))$, (b) is because (X_1, X_2) is a function of (S_1, S_2, U_1, U_2),

and (c) because of the Markov chain $(S_1, S_2, U_1, U_2) \to (X_1, X_2, S_R) \to Y$. Similarly, one can show

$$
\begin{aligned}
I(U_1; Y | S_R, U_2, U_0) &= H(Y | X_2, U_2, U_0, S_2, S_R) \\
&\quad - H(Y | X_1, X_2, U_1, U_2, U_0, S_1, S_2, S_R) \\
&\overset{(a)}{=} H(Y | X_2, U_0, S_R) - H(Y | X_1, X_2, U_0, S_R) \\
&= I(X_1; Y | X_2, U_0, S_R),
\end{aligned}
$$

where (a) follows from $(S_2, U_2) \to (U_0, X_2, S_R) \to Y$ and $(S_1, S_2, U_1, U_2) \to (U_0, X_1, X_2, S_R) \to Y$, and

$$
\begin{aligned}
I(U_2; Y | S_R, U_1, U_0) &= I(X_2; Y | X_1, U_0, S_R), \\
I(U_1, U_2; Y | S_R, U_0) &= I(X_1, X_2; Y | U_0, S_R).
\end{aligned}
$$

Finally, by the functional representation lemma [15, Appendix B] and since U_1 and U_2 do not appear in the bounds given by (2) and proven above, designing

$$
p(u_0)p(u_1|u_0)p(u_2|u_0)\mathbb{1}[x_1 - f_1(s_1, u_1)]\mathbb{1}[x_2 - f_2(s_2, u_2)],
$$

where $\mathbb{1}[\cdot]$ is an indicator function, is equivalent to designing $p(u)p(x_1|s_1, u)p(x_2|s_2, u)$ (replacing U_0 with U). ∎

Converse: Let us define $U_i = (W_0, S_1^{i-1}, S_2^{i-1})$. We construct an outer bound by assuming that past CSIT realizations (S_1^{i-1}, S_2^{i-1}) are available at both encoders. Hence, we assume that X_{1i} and X_{2i} are functions of $(W_0, W_1, S_1^i, S_2^{i-1}) = (W_1, U_i, S_{1i})$ and $(W_0, W_2, S_2^i, S_1^{i-1}) = (W_2, U_i, S_{2i})$ respectively. Note that U_i is independent of $(S_i, S_{1i}, S_{2i}, S_{Ri})$. We then bound

$$
\begin{aligned}
nR_1 &\overset{(a)}{\leq} I(W_1; Y^n, S_R^n | W_0, W_2) + n\epsilon_n \\
&\overset{(b)}{=} I(W_1; Y^n | W_0, W_2, S_R^n) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(W_1; Y_i | Y^{i-1}, W_0, W_2, S_R^n) + n\epsilon_n \\
&= \sum_{i=1}^{n} H(Y_i | W_0, W_2, Y^{i-1}, S_R^n) \\
&\quad - H(Y_i | W_0, W_1, W_2, Y^{i-1}, S_R^n) + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} H(Y_i | W_0, W_2, S_1^{i-1}, S_2^i, Y^{i-1}, S_R^n) \\
&\quad - H(Y_i | W_0, W_1, W_2, S_1^i, S_2^i, Y^{i-1}, S_R^n) + n\epsilon_n \\
&\overset{(d)}{=} \sum_{i=1}^{n} H(Y_i | W_2, X_{2i}, U_i, S_{2i}, Y^{i-1}, S_R^n) \\
&\quad - H(Y_i | X_{1i}, X_{2i}, U_i, S_{Ri}) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} H(Y_i | X_{2i}, U_i, S_{Ri}) \\
&\quad - H(Y_i | X_{1i}, X_{2i}, U_i, S_{Ri}) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(X_{1i}; Y_i | X_{2i}, U_i, S_{Ri}) + n\epsilon_n,
\end{aligned}
$$

where (a) follows from Fano's inequality ($\lim_{n \to \infty} \epsilon_n = 0$), (b) from the independence of W_1 and S_R^n, (c) from (S_{i1}, S_{2i}) being a function of S_{Ri}, and (d) from the Markov chain

$(W_0, W_1, W_2, S_1^i, S_2^i, Y^{i-1}, \{S_{Rj}\}_{j\neq i}) \to (X_{1i}, X_{2i}, S_{Ri}) \to Y_i$. Similarly, we have

$$nR_2 \leq \sum_{i=1}^{n} I(X_{2i}; Y_i | X_{1i}, U_i, S_{Ri}) + n\epsilon_n,$$

$$n(R_1 + R_2) \leq \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i | U_i, S_{Ri}) + n\epsilon_n,$$

$$nR_{\text{sum}} \leq \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i | S_{Ri}) + n\epsilon_n.$$

The code must also satisfy the input cost constraints

$$P_1 \geq \boldsymbol{\mathcal{E}} \left[\frac{1}{n} \sum_{i=1}^{n} b_1(X_{1i}) \right], \quad P_2 \geq \boldsymbol{\mathcal{E}} \left[\frac{1}{n} \sum_{i=1}^{n} b_2(X_{2i}) \right].$$

We combine all the bounds by means of a time-sharing variable Q uniformly distributed in $\{1, \dots, n\}$ and independent of everything else, and by letting $U = (U_Q, Q)$, $X_1 = X_{1Q}$, $X_2 = X_{2Q}$ $Y = Y_Q$, $S = S_Q$, $S_1 = S_{1Q}$, $S_2 = S_{2Q}$, $S_R = S_{RQ}$. Note that the resulting distribution $p_{Y,X_1,X_2,S,S_1,S_2,S_R,U}$ factors as

$$p(y|x_1, x_2, s)p(x_1|s_1, u)p(x_2|s_2, u)p(s, s_1, s_2, s_R)p(u)$$

as required. With these identifications, we readily obtain

$$R_1 \leq I(X_1; Y | X_2, U, S_R) + \epsilon_n$$
$$R_2 \leq I(X_2; Y | X_1, U, S_R) + \epsilon_n$$
$$R_1 + R_2 \leq I(X_1, X_2; Y | U, S_R) + \epsilon_n$$
$$R_{\text{sum}} \leq I(X_1, X_2; Y | S_R, Q) + \epsilon_n$$
$$\leq I(X_1, X_2; Y | S_R) + \epsilon_n,$$
$$P_1 \geq \boldsymbol{\mathcal{E}}[b_1(X_1)], P_2 \geq \boldsymbol{\mathcal{E}}[b_2(X_2)].$$

∎

C. Proof of Proposition 1

Let us define $U_{0i} := (W_0, S_R^{i-1})$, $U_{1i} := (W_1, S_1^{i-1}, U_{0i})$, and $U_{2i} := (W_2, S_2^{i-1}, U_{0i})$. Note that X_{1i} and X_{2i} are functions of (U_{1i}, S_{1i}) and (U_{2i}, S_{2i}) respectively, and, due to the Markov chain $S_{1i} \to S_{Ri} \to S_{2i}$, we also have $U_{1i} \to U_{0i} \to U_{2i}$ as required. By Fano's inequality ($\lim_{n\to\infty} \epsilon_n = 0$), and by following similar steps as in the previous sections, we obtain

$$n(R_1 + R_2)$$
$$\leq I(W_1, W_2; Y^n, S_R^n | W_0) + n\epsilon_n$$
$$= I(W_1, W_2; Y^n | W_0, S_R^n) + n\epsilon_n$$
$$= \sum_{i=1}^{n} I(W_1, W_2; Y_i | Y^{i-1}, W_0, S_R^n) + n\epsilon_n$$
$$\leq \sum_{i=1}^{n} I(W_1, W_2, S_1^{i-1}, S_2^{i-1}; Y_i | Y^{i-1}, W_0, S_R^n) + n\epsilon_n$$
$$= \sum_{i=1}^{n} I(U_{1i}, U_{2i}; Y_i | Y^{i-1}, U_{0i}, S_{Ri}^n) + n\epsilon_n$$
$$\leq \sum_{i=1}^{n} I(U_{1i}, U_{2i}; Y_i | U_{0i}, S_{Ri}) + n\epsilon_n,$$

where the last inequality comes from the memoryless property of the channel. Following similar lines one can prove

$$nR_{\text{sum}} \leq \sum_{i=1}^{n} I(U_{1i}, U_{2i}; Y_i | S_{Ri}) + n\epsilon_n,$$

which can be combined with the bound on $R_1 + R_2$ and the power constraints by means of the usual time-sharing step.

D. Proof of Theorem 4

We construct an outer bound $\mathscr{C}_o(P_1, P_2)$ by following similar steps as in [34], but starting from the single-letter formulation of $\mathscr{C}(P_1, P_2)$ given by Theorem 2 extended to continuous alphabets similarly to [15], [16], [32]. We consider the following applications of the *maximum differential entropy lemma* [15, p. 21], adapted to the complex field. From the rightmost bound in [15, Eq. 2.6], we have

$$h(\boldsymbol{Y}|\mathbb{S} = \mathbf{S}) \leq \log \left((\pi e)^2 \det \left(\boldsymbol{\mathcal{E}} \left[\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}} | \mathbb{S} = \mathbf{S} \right] \right) \right)$$
$$= \log \left((\pi e)^2 \det \left(\mathbf{S}\boldsymbol{\Sigma} \left(q_1(\mathbf{S}), q_2(\mathbf{S}) \right) \mathbf{S}^{\mathsf{H}} + \mathbf{I} \right) \right),$$

where $\boldsymbol{\Sigma}(S_1, S_2) = \boldsymbol{\mathcal{E}} \left[\boldsymbol{X}\boldsymbol{X}^{\mathsf{H}} | S_1, S_2 \right]$. Then, from the bound in [15, Eq. 2.7], we have

$$h(\boldsymbol{Y}|U, \mathbb{S} = \mathbf{S}) \leq \log \left((\pi e)^2 \det \left(\boldsymbol{\mathcal{E}} \left[(\boldsymbol{Y} - \boldsymbol{\mathcal{E}}[\boldsymbol{Y}|U, \mathbb{S} = \mathbf{S}]) \right. \right. \right.$$
$$\left. \left. \left. \times (\boldsymbol{Y} - \boldsymbol{\mathcal{E}}[\boldsymbol{Y}|U, \mathbb{S} = \mathbf{S}])^{\mathsf{H}} | \mathbb{S} = \mathbf{S} \right] \right) \right)$$
$$= \log \left((\pi e)^2 \det \left(\mathbf{S}\boldsymbol{\Gamma} \left(q_1(\mathbf{S}), q_2(\mathbf{S}) \right) \mathbf{S}^{\mathsf{H}} + \mathbf{I} \right) \right),$$

where $\boldsymbol{\Gamma}(S_1, S_2) =$

$$\boldsymbol{\Sigma}(S_1, S_2) - \boldsymbol{\mathcal{E}} \left[\boldsymbol{\mathcal{E}}[\boldsymbol{X}|U, S_1, S_2]\boldsymbol{\mathcal{E}}[\boldsymbol{X}^{\mathsf{H}}|U, S_1, S_2] | S_1, S_2 \right].$$

By the structure of the input distribution, we now observe that $\boldsymbol{\Gamma}(S_1, S_2) = \text{diag}(\gamma_1(S_1), \gamma_2(S_2))$ and that, $\forall (s_1, s_2) \in \mathcal{S}_1 \times \mathcal{S}_2$,

$$\boldsymbol{\Sigma}(s_1, s_2) = \begin{bmatrix} \gamma_1(s_1) & 0 \\ 0 & \gamma_2(s_2) \end{bmatrix}$$
$$+ \begin{bmatrix} \boldsymbol{\mathcal{E}}[|\mu_1(U, s_1)|^2] & \boldsymbol{\mathcal{E}}[\mu_1(U, s_1)\mu_2^\star(U, s_2)] \\ \boldsymbol{\mathcal{E}}[\mu_2(U, s_2)\mu_1^\star(U, s_1)] & \boldsymbol{\mathcal{E}}[|\mu_2(U, s_2)|^2] \end{bmatrix},$$
(18)

where we define the functions

$$\mu_k(U, S_k) \triangleq \boldsymbol{\mathcal{E}}[X_k | U, S_k]$$
$$\gamma_k(S_k) \triangleq \boldsymbol{\mathcal{E}} \left[|X_k|^2 | S_k \right] - \boldsymbol{\mathcal{E}} \left[|\mu_k(U, S_k)|^2 | S_k \right] \geq 0.$$

By following similar steps for $h(\boldsymbol{Y}|X_1, U, \mathbb{S} = \mathbf{S})$ and $h(\boldsymbol{Y}|X_2, U, \mathbb{S} = \mathbf{S})$, and by applying the resulting bounds to the mutual information terms in Theorem 2, we obtain

$$R_1 \leq \boldsymbol{\mathcal{E}} \left[\log \left(1 + \gamma_1(S_1) \|\mathbb{S}\boldsymbol{e}_1\|^2 \right) \right],$$
$$R_2 \leq \boldsymbol{\mathcal{E}} \left[\log \left(1 + \gamma_2(S_2) \|\mathbb{S}\boldsymbol{e}_2\|^2 \right) \right],$$
$$R_1 + R_2 \leq \boldsymbol{\mathcal{E}} \left[\log \det \left(\mathbf{I} + \mathbb{S}\text{diag}(\gamma_1(S_1), \gamma_2(S_2))\mathbb{S}^{\mathsf{H}} \right) \right],$$
$$R_0 + R_1 + R_2 \leq \boldsymbol{\mathcal{E}} \left[\log \det \left(\mathbf{I} + \mathbb{S}\boldsymbol{\Sigma}(S_1, S_2)\mathbb{S}^{\mathsf{H}} \right) \right].$$
(19)

The outer bound $\mathscr{C}_o(P_1, P_2)$ is then established by taking the convex hull of the union of all rate triples (R_0, R_1, R_2) satisfying (19) for some $p(x_1|s_1, u)p(x_2|s_2, u)p(u)$ such that $\boldsymbol{\mathcal{E}}[|X_k|^2] \leq P_k$.

Similarly to the proof of Theorem 3, it can be now shown that every $\boldsymbol{\Sigma}(S_1, S_2)$ as in (18) induced by any $p(x_1|s_1, u)p(x_2|s_2, u)p(u)$ can also be obtained by the scheme in (8), i.e., via superposition of linearly precoded Gaussian codes. The key point is showing that the second term in the RHS of (18) can be obtained via distributed linear precoders of dimension d. This follows by the same technique used in the proof of Theorem 3, by simply replacing the functions $f_k(u, s_k)$ with $\mu_k(u, s_k)$. Finally, since the inputs are conditionally Gaussian, standard arguments [15] show that (8) attains the bound $\mathscr{C}_o(P_1, P_2)$, without time sharing (i.e., we can omit the convex hull operation).

E. Proof of Proposition 4

The proof is split for the sake of clarity in the following three steps:

1) We fix a specific conditional input covariance matrix $\boldsymbol{\Sigma}^\star(S_1, S_2)$, and we show that it is achievable via distributed linear precoding if and only if $d' > 2$.
2) We construct a specific $p(\mathbf{S}, s_1, s_2)$ such that $\boldsymbol{\Sigma}^\star(S_1, S_2)$ is the unique optimal solution to Problem (4).
3) We combine the above steps to show that there exists a channel for which $d' \leq 2$ leads to strictly suboptimal rates.

Step 1: Consider binary D-CSIT alphabets, i.e., $\mathcal{S}_1 = \mathcal{S}_2 = \{0, 1\}$, and let $\boldsymbol{\Sigma}^\star(S_1, S_2)$ be given by

$$\begin{aligned}
\boldsymbol{\Sigma}^\star(0,0) &= \mathbf{I}, \quad \boldsymbol{\Sigma}^\star(1,0) = \mathbf{I}, \\
\boldsymbol{\Sigma}^\star(0,1) &= \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}^\star(1,1) = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}.
\end{aligned} \quad (20)$$

Define the set $\mathcal{G}(\tilde{d})$ of conditional input covariance matrices $\boldsymbol{\Sigma}(S_1, S_2)$ which are achievable via distributed linear precoders of maximal dimension $\tilde{d}$, i.e.,

$$\mathcal{G}(\tilde{d}) \triangleq \begin{cases} \boldsymbol{\Sigma}(S_1, S_2) \in \mathbb{S}_+^2 \quad \text{s.t.} \\ \boldsymbol{\Sigma}(S_1, S_2) = \begin{bmatrix} \boldsymbol{g}_1^\mathsf{H}(S_1) \\ \boldsymbol{g}_2^\mathsf{H}(S_2) \end{bmatrix} \begin{bmatrix} \boldsymbol{g}_1(S_1) & \boldsymbol{g}_2(S_2) \end{bmatrix}, \\ \boldsymbol{g}_1(S_1), \boldsymbol{g}_2(S_2) \in \mathbb{C}^{d'}, \ d' \leq \tilde{d} \end{cases} \quad (21)$$

Clearly, $\mathcal{G}(d_1) \subseteq \mathcal{G}(d_2)$, for $d_1 \leq d_2$. The following lemma holds:

Lemma 2. $\boldsymbol{\Sigma}^\star(S_1, S_2) \in \mathcal{G}(3)$, and $\boldsymbol{\Sigma}^\star(S_1, S_2) \notin \mathcal{G}(2)$.

Proof: For $\boldsymbol{\Sigma}^\star(S_1, S_2)$ to be achievable, we need to find precoders $\boldsymbol{g}_k(S_k)$ s.t.

$$\begin{cases} \boldsymbol{g}_1^\mathsf{H}(0)\boldsymbol{g}_2(0) = 0, \quad \boldsymbol{g}_1^\mathsf{H}(0)\boldsymbol{g}_2(1) = 0.6, \\ \boldsymbol{g}_1^\mathsf{H}(1)\boldsymbol{g}_2(0) = 0, \quad \boldsymbol{g}_1^\mathsf{H}(1)\boldsymbol{g}_2(1) = 0.8, \\ \|\boldsymbol{g}_1(0)\| = \|\boldsymbol{g}_1(1)\| = \|\boldsymbol{g}_2(0)\| = \|\boldsymbol{g}_2(1)\| = 1. \end{cases}$$

For $\boldsymbol{g}_k(S_k)$ of dimension $d' = 2$, the above system has no solution. In fact, we need to simultaneously satisfy

$$\begin{cases} \boldsymbol{g}_1(0) \perp \boldsymbol{g}_2(0), \\ \boldsymbol{g}_1(1) \perp \boldsymbol{g}_2(0), \\ \|\boldsymbol{g}_1(0)\| = \|\boldsymbol{g}_1(1)\| = 1, \end{cases}$$

which, for $d' = 2$, implies $\boldsymbol{g}_1(0) = \pm \boldsymbol{g}_1(1)$, and hence leads to the following contradiction $0.6 = \boldsymbol{g}_1^\mathsf{H}(0)\boldsymbol{g}_2(1) =$

$\pm \boldsymbol{g}_1^\mathsf{H}(1)\boldsymbol{g}_2(1) = \pm 0.8$. Instead, one can check that $\boldsymbol{\Sigma}^\star(S_1, S_2)$ is readily obtained by letting $d' = 3$ and

$$\begin{aligned}
\boldsymbol{g}_1(0) &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{g}_1(1) = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}, \\
\boldsymbol{g}_2(0) &= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{g}_2(1) = \begin{bmatrix} 0.6 & 0.8 & 0 \end{bmatrix}.
\end{aligned}$$

$\blacksquare$

Step 2: Consider the following rewriting of Problem (4), by letting again $\mathcal{S}_1 = \mathcal{S}_2 = \{0, 1\}$ (hence $d = 4$), and unitary power constraint $P_1 = P_2 = 1$:

$$C_0 = \max_{\boldsymbol{\Sigma} \in \mathcal{P} \cap \mathcal{G}(4)} \boldsymbol{\mathcal{E}} \left[\log \det \left(\mathbf{I} + \mathbb{S}\boldsymbol{\Sigma}(S_1, S_2)\mathbb{S}^\mathsf{H} \right) \right], \quad (22)$$

where $\mathcal{G}(4)$ is given by (21), and where

$$\mathcal{P} \triangleq \{ \boldsymbol{\Sigma}(S_1, S_2) \in \mathbb{S}_+^2 \mid \boldsymbol{\mathcal{E}}[\Sigma_{k,k}(S_1, S_2)] \leq 1, \ k = 1, 2 \}$$

is the per-TX power constraint. Note that $\boldsymbol{\Sigma}^\star(S_1, S_2)$ belongs to the feasible set, i.e. $\boldsymbol{\Sigma}^\star(S_1, S_2) \in \mathcal{P} \cap \mathcal{G}(4)$.

Lemma 3. *There exist some $p(\mathbf{S}, s_1, s_2)$ such that $\boldsymbol{\Sigma}^\star(S_1, S_2)$ given by* (20) *is the unique optimal solution for problem* (22).

Proof: The main idea is to build such CSI distribution by "reversing" a spatio-temporal water-filling algorithm which gives as unique optimal solution the conditional input covariance $\boldsymbol{\Sigma}^\star(S_1, S_2)$. We now provide the details.

Define a uniformly distributed random state $\mathbb{S}$ taking values in the finite alphabet $\mathcal{S} = \{\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \mathbf{S}_4\}$, and let the CSIT be given by the functions

$$s_1 = q_1(\mathbf{S}) = \begin{cases} 0 \text{ for } \mathbf{S} \in \{\mathbf{S}_1, \mathbf{S}_2\} \\ 1 \text{ otherwise} \end{cases}$$

$$s_2 = q_2(\mathbf{S}) = \begin{cases} 0 \text{ for } \mathbf{S} \in \{\mathbf{S}_1, \mathbf{S}_3\} \\ 1 \text{ otherwise} \end{cases}$$

The capacity of such a channel can be upper bounded by

$$C_0 = \max_{\boldsymbol{\Sigma} \in \mathcal{P} \cap \mathcal{G}(4)} \boldsymbol{\mathcal{E}} \left[\log \det \left(\mathbf{I} + \mathbb{S}\boldsymbol{\Sigma}(S_1, S_2)\mathbb{S}^\mathsf{H} \right) \right], \quad (23)$$

$$\leq \max_{\boldsymbol{\Sigma} \in \mathcal{P}} \boldsymbol{\mathcal{E}} \left[\log \det \left(\mathbf{I} + \mathbb{S}\boldsymbol{\Sigma}(S_1, S_2)\mathbb{S}^\mathsf{H} \right) \right], \quad (24)$$

$$\leq \max_{\boldsymbol{\Sigma} \in \mathcal{P}'} \boldsymbol{\mathcal{E}} \left[\log \det \left(\mathbf{I} + \mathbb{S}\boldsymbol{\Sigma}(S_1, S_2)\mathbb{S}^\mathsf{H} \right) \right], \quad (25)$$

where

$$\mathcal{P}' \triangleq \{ \boldsymbol{\Sigma}(S_1, S_2) \in \mathbb{S}_+^2 \mid \operatorname{tr}\{\boldsymbol{\mathcal{E}}[\boldsymbol{\Sigma}(S_1, S_2)]\} \leq 2 \},$$

is the set obtained by relaxing the per-TX power constraint $\mathcal{P}$ to a total power constraint ($\mathcal{P} \subseteq \mathcal{P}'$). Inequalities (24) and (25) are obtained respectively by relaxing the achievability via distributed linear precoding and the power constraint.

Problem (25) turns out to be an instance of a classical (centralized) MIMO capacity problem, where the optimal solution is given by the well-known spatio-temporal water-filling algorithm. More precisely, let us rewrite (25) as

$$\max_{\boldsymbol{\Sigma}(S_1, S_2) \in \mathcal{P}'} \sum_{i=1}^{4} p(\mathbf{S}_i) \log \det \left(\mathbf{I} + \mathbf{S}_i \boldsymbol{\Sigma}(q_1(\mathbf{S}_i), q_2(\mathbf{S}_i))\mathbf{S}_i^\mathsf{H} \right)$$

$$= \max_{\{\boldsymbol{\Sigma}_i\} \in \mathcal{P}'} \frac{1}{4} \sum_{i=1}^{4} \log \det \left(\mathbf{I} + \mathbf{S}_i \boldsymbol{\Sigma}_i \mathbf{S}^\mathsf{H} \right),$$

$$(26)$$

where we defined $\boldsymbol{\Sigma}_i \triangleq \boldsymbol{\Sigma}(q_1(\mathbf{S}_i), q_2(\mathbf{S}_i))$, and where

$$\tilde{\mathcal{P}}' = \left\{ \boldsymbol{\Sigma}_i \in \mathbb{S}_+^2 \mid \frac{1}{4} \sum_{i=1}^{4} \mathrm{tr}\{\boldsymbol{\Sigma}_i\} \leq 2 \right\}.$$

A well-known application of the Hadamard's inequality gives the following upper bound in terms of the channel eigen-decompositions $\mathbf{S}_i^{\mathsf{H}} \mathbf{S}_i = \mathbf{V}_i \boldsymbol{\Lambda}_i \mathbf{V}_i^{\mathsf{H}}$, $\boldsymbol{\Lambda}_i = \mathrm{diag}(\lambda_{i,1}, \lambda_{i,2})$

$$\max_{\{\boldsymbol{\Sigma}_i\} \in \tilde{\mathcal{P}}'} \frac{1}{4} \sum_{i=1}^{4} \log \det \left(\mathbf{I} + \mathbf{S}_i \boldsymbol{\Sigma}_i \mathbf{S}_i^{\mathsf{H}} \right)$$

$$\leq \max_{\substack{\xi_{i,k} \geq 0 \\ \frac{1}{4} \sum_{i,k} \xi_{i,k} \leq 2}} \frac{1}{4} \sum_{i=1}^{4} \sum_{k=1}^{2} \log(1 + \lambda_{i,k} \xi_{i,k}),$$

where the optimal $\xi_{i,k}$ are given by the water-filling conditions

$$\xi_{i,k} = \max\left\{ \nu - \frac{1}{\lambda_{i,k}}, 0 \right\}, \quad i = 1, \dots, 4, \quad k = 1, 2,$$

$$\sum_{i,k} \max\left\{ \nu - \frac{1}{\lambda_{i,k}}, 0 \right\} = 8,$$

and where equality is achieved for

$$\boldsymbol{\Sigma}_i = \mathbf{V}_i \boldsymbol{\Xi}_i \mathbf{V}_i^{\mathsf{H}}, \quad \boldsymbol{\Xi}_i = \mathrm{diag}(\xi_{i,1}, \xi_{i,2}), \quad i = 1, \dots, 4.$$

Consider the conditional covariance $\boldsymbol{\Sigma}^{\star}(S_1, S_2)$ given by (20). Note that $\boldsymbol{\Sigma}^{\star}(S_1, S_2) \in \mathcal{P}'$, i.e., it satisfies the total power constraint. We wish to construct $\mathcal{S} = \{\mathbf{S}_i\}$ such that $\boldsymbol{\Sigma}^{\star}(S_1, S_2)$ is the unique optimal solution for (25). This can be done by "reversing" the MIMO water-filling algorithm described above. More precisely, let us consider $\boldsymbol{\Sigma}_i^{\star} \triangleq \boldsymbol{\Sigma}^{\star}(q_1(\mathbf{S}_i), q_2(\mathbf{S}_i))$ and their eigen-decompositions

$$\boldsymbol{\Sigma}_i^{\star} = \mathbf{V}_i^{\star} \boldsymbol{\Xi}_i^{\star} \mathbf{V}_i^{\star \mathsf{H}}, \quad \boldsymbol{\Xi}_i^{\star} = \mathrm{diag}(\xi_{i,1}^{\star}, \xi_{i,2}^{\star}).$$

We construct now $\mathcal{S}$ by letting

$$\mathbf{S}_i = (\mathbf{V}_i^{\star} \boldsymbol{\Lambda}_i^{\star} \mathbf{V}_i^{\star \mathsf{H}})^{\frac{1}{2}}, \quad i = 1, \dots, 4$$

where the eigenvalues $\boldsymbol{\Lambda}_i^{\star} = \mathrm{diag}(\lambda_{i,1}^{\star}, \lambda_{i,2}^{\star})$ are given by

$$\lambda_{i,k}^{\star} = \frac{1}{\nu^{\star} - \xi_{i,k}^{\star}}, \quad i = 1, \dots, 4, \quad k = 1, 2,$$

and any choice of $\infty > \nu^{\star} > \max_{i,k} \xi_{i,k}^{\star} = 1.8$.

By construction, $\boldsymbol{\Sigma}^{\star}(S_1, S_2)$ is an optimal solution for (25). Uniqueness of the solution can be proven by contradiction as in [38, Section III.A], or directly by the strict concavity of $\sum_{i=1}^{4} \log \det \left(\mathbf{I} + \mathbf{S}_i \boldsymbol{\Sigma}_i \mathbf{S}_i^{\mathsf{H}} \right)$ in $\{\boldsymbol{\Sigma}_i \succeq 0\}$, which is a direct consequence of the strict concavity of $\log \det(\boldsymbol{A})$ in $\boldsymbol{A} \succ \mathbf{0}$ and of the positive definiteness of $\mathbf{S}_i$ by construction. Finally, since $\boldsymbol{\Sigma}^{\star}(S_1, S_2) \in \mathcal{P} \cap \mathcal{G}(4)$, (25) and (24) are satisfied with equality. ∎

Step 3: The proof is now concluded by combining Lemma 2 and Lemma 3, yielding $\exists\, p(\mathbf{S}, s_1, s_2)$ such that

$$\arg \max_{\boldsymbol{\Sigma}(S_1, S_2) \in \mathcal{P} \cap \mathcal{G}(4)} \boldsymbol{\mathcal{E}} \left[\log \det \left(\mathbf{I} + \mathbb{S} \boldsymbol{\Sigma}(S_1, S_2) \mathbb{S}^{\mathsf{H}} \right) \right]$$

$$\overset{(a)}{=} \{\boldsymbol{\Sigma}^{\star}(S_1, S_2)\} \overset{(b)}{\notin} \mathcal{G}(2),$$

where (a) follows from Lemma 3, and (b) from Lemma 2, which implies that $\exists\, p(\mathbf{S}, s_1, s_2)$ such that

$$\max_{\boldsymbol{\Sigma} \in \mathcal{P} \cap \mathcal{G}(4)} \boldsymbol{\mathcal{E}} \left[\log \det \left(\mathbf{I} + \mathbb{S} \boldsymbol{\Sigma}(S_1, S_2) \mathbb{S}^{\mathsf{H}} \right) \right]$$

$$> \max_{\boldsymbol{\Sigma} \in \mathcal{P} \cap \mathcal{G}(2)} \boldsymbol{\mathcal{E}} \left[\log \det \left(\mathbf{I} + \mathbb{S} \boldsymbol{\Sigma}(S_1, S_2) \mathbb{S}^{\mathsf{H}} \right) \right].$$

F. Proof of Proposition 5

The proof follows by manipulating an off-diagonal entry of the sub-matrices $\mathbf{Q}_1, \mathbf{Q}_2$ given by (13) until $\mathbf{Q}$ becomes rank-deficient. We recall that varying these entries has no influence on the achievable rates, provided that the positive semi-definiteness of $\mathbf{Q}$ is maintained. Consider the symmetric matrix

$$\tilde{\mathbf{Q}}(t) \triangleq \mathbf{Q} + t \begin{bmatrix} 0 & 1 & \mathbf{0} \\ 1 & 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad t \in \mathbb{R}, \quad \mathbf{Q} \in \mathbb{S}_+^d.$$

Let $\lambda_{\min} : \mathbb{S}^d \to \mathbb{R}$ be the minimum eigenvalue of a Hermitian symmetric matrix (not necessarily positive semi-definite). By definition, $\lambda_{\min}(\tilde{\mathbf{Q}}(0)) \geq 0$. In addition, $\exists\, t_1 > 0$ such that $\lambda_{\min}(\tilde{\mathbf{Q}}(t_1)) < 0$. Furthermore, by the continuity of the map $\lambda_{\min}$ in the matrix entries [39, Theorem 5.2], $\lambda_{\min}(\tilde{\mathbf{Q}}(t))$ is a continuous function of t. Hence, by the intermediate value theorem, $\exists\, t_0 \in [0, t_1]$ such that $\lambda_{\min}(\tilde{\mathbf{Q}}(t_0)) = 0$, i.e., such that $\tilde{\mathbf{Q}}(t_0)$ is positive semi-definite, low rank, and achieves the same rates as $\mathbf{Q}$. ∎

REFERENCES

[1] D. Gesbert, S. Hanly, H. Huang, S. Shamai (Shitz), O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Sept. 2010.

[2] O. Simeone, N. Levy, A. Sanderovich, O. Somekh, B. M. Zaidel, H. V. Poor, S. Shamai (Shitz) *et al.*, "Cooperative wireless cellular systems: An information-theoretic view," *Foundations and Trends® in Communications and Information Theory*, vol. 8, no. 1-2, pp. 1–177, 2012.

[3] T. Q. Quek, M. Peng, O. Simeone, and W. Yu, *Cloud radio access networks: Principles, technologies, and applications.* Cambridge University Press, 2017.

[4] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[5] N. Naderializadeh, M. A. Maddah-Ali, and A. S. Avestimehr, "Cache-aided interference management in wireless cellular networks," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3376–3387, May 2019.

[6] F. Willems, "Information Theoretical Results for the Discrete Memoryless Multiple Access Channel," *Ph. D. thesis, Katholieke Universiteit Leuven, Belgium*, 1989.

[7] H. H. Permuter, S. Shamai (Shitz), and A. Somekh-Baruch, "Message and state cooperation in multiple access channels," *IEEE Trans. Info. Theory*, vol. 57, no. 10, pp. 6379–6396, Oct. 2011.

[8] D. Gesbert and P. de Kerret, "Team methods for device cooperation in wireless networks," in *Cooperative and Graph Signal Processing.* Elsevier, 2018, pp. 469–487.

[9] P. de Kerret and D. Gesbert, "Degrees of freedom of the network MIMO channel with distributed CSI," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6806–6824, Nov. 2012.

[10] A. Bazco Nogueras, L. Miretti, P. de Kerret, D. Gesbert, and N. Gresset, "Achieving vanishing rate loss in decentralized network MIMO," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT).* IEEE, 2019.

[11] C.-S. Hwang, M. Malkin, A. El Gamal, and J. M. Cioffi, "Multiple-access channels with distributed channel state information," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT).* IEEE, 2007, pp. 1561–1565.

[12] M. Li, O. Simeone, and A. Yener, "Multiple access channels with states causally known at transmitters," *IEEE Trans. Info. Theory*, vol. 59, no. 3, pp. 1394–1404, 2012.

[13] A. Zaidi, P. Piantanida, and S. Shamai (Shitz), "Capacity region of co-operative multiple-access channel with states," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6153–6174, Oct. 2013.

[14] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Devp.*, vol. 2, no. 4, pp. 289–293, 1958.

[15] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge university press, 2011.

[16] G. Caire and S. Shamai (Shitz), "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, June 1999.

[17] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2867–2877, Aug. 2005.

[18] S. Sigurjónsson and Y.-H. Kim, "On multiple user channels with state information at the transmitters," pp. 72–76, 2005.

[19] M. N. Khormuji and M. Skoglund, "The relay channel with partial causal state information," in *Proc. IEEE Int. Symp. Inf. Theory and App. (ISITA)*, Dec. 2008.

[20] A. Das and P. Narayan, "Capacities of time-varying multiple-access channels with side information," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 4–25, Jan 2002.

[21] S. Jafar, "Capacity with causal and noncausal side information: A unified view," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5468–5474, Dec. 2006.

[22] G. Como and S. Yuksel, "On the capacity of memoryless finite-state multiple-access channels with asymmetric state information at the encoders," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1267–1273, March 2011.

[23] N. Şen, F. Alajaji, S. Yüksel, and G. Como, "Memoryless multiple access channel with asymmetric noisy state information at the encoders," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7052–7070, 2013.

[24] A. Lapidoth and Y. Steinberg, "The multiple-access channel with causal side information: Common state," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 32–50, Jan. 2013.

[25] ——, "The multiple-access channel with causal side information: Double state," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1379–1393, March 2013.

[26] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdu, "Cooperative multiple-access encoding with states available at one transmitter," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4448–4469, Oct. 2008.

[27] A. Zaidi and S. Shamai (Shitz), "On cooperative multiple access channels with delayed CSI at transmitters," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6204–6230, Oct. 2014.

[28] T. Cover and C. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Info. Theory*, vol. 27, no. 3, pp. 292–298, 1981.

[29] F. Willems, "The feedback capacity region of a class of discrete memoryless multiple access channels (corresp.)," *IEEE Trans. Info. Theory*, vol. 28, no. 1, pp. 93–95, 1982.

[30] C. Choudhuri, Y.-H. Kim, and U. Mitra, "Causal state communication," *IEEE Trans. Info. Theory*, vol. 59, no. 6, pp. 3709–3719, June 2013.

[31] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, no. 7, pp. 1037–1076, Sep. 1973.

[32] G. Caire and K. R. Kumar, "Information theoretic foundations of adaptive coded modulation," *Proc. IEEE*, vol. 95, no. 12, pp. 2274–2298, Dec. 2007.

[33] J. Marschak and R. Radner, *Economic Theory of Teams*. Yale University Press, 1972.

[34] N. Liu and S. Ulukus, "Capacity region and optimum power control strategies for fading Gaussian multiple access channels with common data," *IEEE Trans. Commun.*, vol. 54, no. 10, pp. 1815–1826, Oct. 2006.

[35] E. J. Candes and T. Tao, "The power of convex relaxation: near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.

[36] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[37] G. Keshet, Y. Steinberg, N. Merhav *et al.*, "Channel coding in the presence of side information," *Foundations and Trends® in Communications and Information Theory*, vol. 4, no. 6, pp. 445–586, 2008.

[38] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for gaussian vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[39] T. Kato, *A short introduction to perturbation theory for linear operators*. Springer-Verlag New York, 1982.

Lorenzo Miretti (Student Member, IEEE) received the BSc degree in telecommunication engineering from Politecnico di Torino, Turin, Italy, in 2015 and the MSc degree in Communications and Computer Networks Engineering from Politecnico di Torino and Télécom ParisTech in 2018, both cum laude. He is currently working towards the Ph.D degree with the Department of Communication Systems, EURECOM, Sophia Antipolis, France. He studies the physical layer of wireless networks, signal processing, and multi-user information theory.

Mari Kobayashi (Senior Member, IEEE) received the B.E. degree in electrical engineering from Keio University, Yokohama, Japan, in 1999, and the M.S. degree in mobile radio and the Ph.D. degree from École Nationale Supérieure des Télécommunications, Paris, France, in 2000 and 2005, respectively. From November 2005 to March 2007, she was a postdoctoral researcher at the Centre Tecnològic de Telecomunicacions de Catalunya, Barcelona, Spain. In May 2007, she joined the Telecommunications department at Centrale Supélec, Gif-sur-Yvette, France, where she is now a professor. She is the recipient of the Newcom++ Best Paper Award in 2010, and IEEE Comsoc/IT Joint Society Paper Award in 2011, and ICC Best Paper Award in 2019. She was an Alexander von Humboldt Experienced Research Fellow (September 2017- April 2019) and an August-Wihelm Scheer Visiting Professor (August 2019-April 2020) at Technical University of Munich (TUM).

David Gesbert (Fellow, IEEE) is Professor and Head of the Communication Systems Department, EURECOM. 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. He was then a founding engineer of Iospan Wireless Inc, a Stanford spin off pioneering MIMO-OFDM (now Intel). Before joining EURECOM in 2004, he has been with the Department of Informatics, University of Oslo as an adjunct professor. D. Gesbert has published about 340 papers and 25 patents, some of them winning 2019 ICC Best Paper Award, 2015 IEEE Best Tutorial Paper Award (Communications Society), 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. He has been a Technical Program Co-chair for ICC2017. He was named a Thomson-Reuters Highly Cited Researchers in Computer Science. In 2015, he was awarded the ERC Advanced Grant "PERFUME" on the topic of smart device Communications in future wireless networks. He is a Board member for the OpenAirInterface (OAI) Software Alliance. Since early 2019, he heads the Huawei-funded Chair on Advanced Wireless Systems Towards 6G Networks. He sits on the Advisory Board of Huawei European Research Institute. In 2020, he was awarded funding by the French Interdisciplinary Institute on Artificial Intelligence for a Chair in the area of AI for the future IoT.

Paul de Kerret (Member, IEEE) has an Engineering degree from both the French Graduate School, IMT Atlantique, and from the Munich University of Technology through a double-degree program. In 2013, he obtained his doctorate from the French Graduate School, Telecom Paris. From 2015 to 2019, he was a Senior Researcher at EURECOM. In September 2019, he joined the Mantu Artificial Intelligence Laboratory to lead the Research Group and he is currently working as Lead Data Scientist within the Startup Company Greenly where he puts his skills to measure and reduce our carbon footprints. He has been involved in several European collaborative projects on mobile communications and co-presented several tutorials at major IEEE international conferences. He has authored more than 30 articles in the IEEE flagship conferences.