On the Fundamental Limits of Cooperative Multiple-Access Channels with Distributed CSIT

Lorenzo Miretti, Paul de Kerret, and David Gesbert
Communication Systems Department, EURECOM
Sophia Antipolis, France
{miretti,dekerret,gesbert}@eurecom.fr

***Abstract*—The availability of accurate and, most importantly, shared channel state information at the transmitter (CSIT) is one of the key factors that enable transmitters cooperation in decentralized wireless systems. However, in some cases, channel information may not be easily or perfectly shared among the transmitters, thus limiting their coordination capabilities. In this paper we shed some light on the fundamental limits of networks with cooperating transmitters impaired by a general *distributed CSIT* assumption. To this end, we consider a state-dependent memory-less multiple-access channel with common message, and with noisy causal CSIT and noisy channel state information at the receiver (CSIR). Perhaps surprisingly, and in contrast to the same setting in absence of common message, we show that distributed precoding based on current CSIT only (namely, a Shannon strategy) achieves the sum-rate capacity of this channel, for every degree of CSIT and CSIR. By focusing on the transmission of a common message only, we then illustrate this result in a practically relevant Gaussian setting.**

***Index Terms*—distributed CSIT, cooperative communication, state-dependent channels, MAC, sum-rate capacity.**

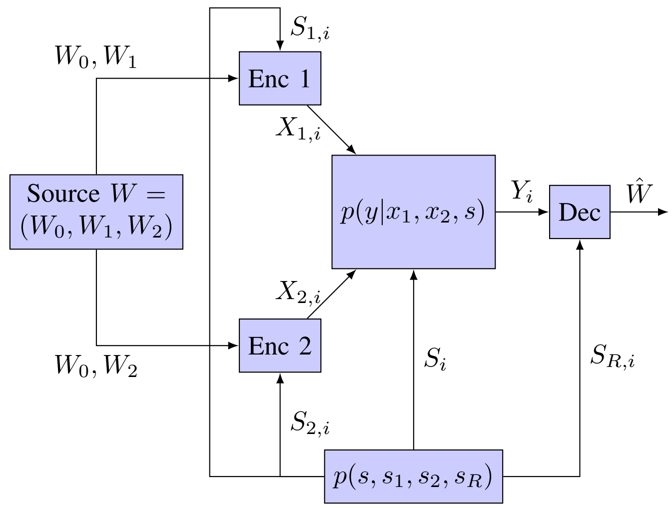

Fig. 1. State-dependent multiple-access channel with common message W_0, independent messages W_1, W_2, D-CSIT S_1, S_2 and imperfect CSIR S_R.

I. Introduction

The aim of this paper is to shed some light on the fundamental limits of systems with multiple cooperating transmitters (TXs) under the *distributed CSIT* (D-CSIT) assumption, that is when the TXs have access to different noisy versions of the channel state information (CSI) [1].

Known joint coding techniques are usually designed by assuming that the available CSIT is fully shared among the TXs, thus allowing for solutions to be directly derived from the centralized case. However, in many practical scenarios, the TXs do not have enough time and/or resources to share their available CSI, or an information exchange does exist but incurs latency or quantization noise, thus limiting their coordination capabilities [1]. This distributed nature of the CSI gives rise to many interesting, yet challenging, problems. Most of the efforts in the literature on this matter have been focused on precoders optimization [1] and asymptotic rate analysis [2], [3] for cooperative multi-user networks.

In this work, to better focus on the effects of distributed CSIT, we consider a simple setting without interference where two TXs cooperatively transmit to a single receiver (RX). Furthermore, we consider only the notion of *causal* state information at the encoders, and we do not consider coding over any cooperation link between the TXs, as in the works based on *conferencing encoders* [4]. The cooperation among the TXs is here assumed to be in the form of a preliminary message sharing phase, for example through offline caching techniques, and any online communication among the TXs is limited to a predefined CSIT sharing mechanism.

To this end, we make use of, and partially extend, available information theoretical results on the capacity of state-dependent channels with causal CSIT. In [5], Shannon characterized the capacity of a state-dependent point-to-point channel with perfect CSIT and absent CSI at the receiver (CSIR), by means of an optimal scheme based on coding over the alphabet of functions mapping the *current* CSIT to the channel input, called *Shannon strategies*. Coding over Shannon strategies constitutes the theoretical foundation of one of the fundamental building blocks of modern wireless communication architectures, where a coded information stream is usually fed to the channel input after a *precoding* stage matching the channel input to the current CSIT. By using the classical wireless communication theory terminology, the capacity in [5] has the operational meaning of maximum achievable *ergodic* rate [6], [7]. This notion is particularly useful for applications where the delay constraints are sufficiently loose to allow the

L. Miretti, P. de Kerret, and D. Gesbert are supported by the European Research Council under the European Unions Horizon 2020 research and innovation program (Agreement no. 670896).

codewords to span multiple i.i.d. fading realizations.

Shannon approach has been successfully extended to arbitrary CSI structures and to many multi-user settings, for example in [8], [9]. For a state-dependent multiple-access channel (MAC), Shannon strategies have been proved to be optimal only for some special cases [10]–[12]. In particular, Shannon strategies are shown to achieve the sum-rate capacity if the CSIT sequences are mutually independent [10], or for perfect CSIR [12]. In case the CSIT sequences are functions of the CSIR, the full capacity region is characterized in terms of Shannon strategies in [11]. In [13], [14] a cooperative MAC with degraded message sets is considered, and Shannon strategies are shown to achieve the capacity region if the CSIT is available at one TX only [13] or perfectly shared [14]. However, interestingly, in [15] Shannon strategies are shown to be generally insufficient to exhaust the capacity region of a MAC with independent messages only, and that more complex schemes that exploit coding also over the past CSIT sequences can lead to larger achievable regions.

In this work, we extend these works by considering a cooperative MAC with a general imperfect CSIT and imperfect CSIR as illustrated in Fig. 1. Our main results read as:

- We show that the transmission of a common message allows to achieve the sum-rate capacity by means of Shannon strategies, i.e. by coding over current CSIT only, for any degree of CSIT and CSIR.
- We specialize the results to a particular cooperative MIMO Gaussian channel with fading and discuss optimal transmission strategies.

II. Cooperative Multiple Access Channels

We start by giving the channel model and the basic definitions adopted throughout this work, by following closely the classical conventions given for example in [7].

a) Channel Model: We consider the state-dependent multiple-access channel (MAC) in Fig. 1, with common message W_0, 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 $S \in \mathcal{S}$, memory-less channel law $p(y_1, y_2|x_1, x_2, s)$, *distributed* CSIT (D-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 $\{(S_i, S_{1,i}, S_{2,i}, S_{R,i})\}$ is assumed to follow a generic memory-less law $p(s, s_1, s_2, s_R)$. All alphabets are assumed to be finite. An n-sequence of inputs, output and states is then governed by the law

$$p(y^n|x_1^n, x_2^n, s^n) = \prod_{i=1}^{n} p(y_i|x_{1,i}, x_{2,i}, s_i),$$

$$p(s^n, s_1^n, s_2^n, s_R^n) = \prod_{i=1}^{n} p(s_i, s_{1,i}, s_{2,i}, s_{R,i}).$$

We assume the messages W_0, W_1, W_2 to be independent and uniformly distributed over the sets $\mathcal{W}_j := \{1, \dots, 2^{\lceil nR_j \rceil}\}$, $j = 0, 1, 2$, where $R_j \geq 0$ is the rate of the message W_j.

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 the encoding functions $\phi_{k,i} : \mathcal{W}_0 \times \mathcal{W}_k \times \mathcal{S}_k^i \to \mathcal{X}_k$, for $i = 1, \dots, n$ and $k = 1, 2$, such that $x_{k,i} = \phi_{k,i}(w_0, w_k, s_k^i)$. The decoding function is $\psi : \mathcal{Y}^n \times \mathcal{S}_R^n \to \mathcal{W}_0 \times \mathcal{W}_1 \times \mathcal{W}_2$, such that the decoded messages are $(\hat{w}_0, \hat{w}_1, \hat{w}_2) = \psi(y^n, s_R^n)$. A rate tuple (R_0, R_1, R_2) is said to be achievable if, for the considered channel, there exists a block code of length n defined as before such that the average probability of error $P_e^{(n)} := P((\hat{W}_0, \hat{W}_1, \hat{W}_2) \neq (W_0, W_1, W_2))$ vanishes as $n \to \infty$. The closure of the set of all achievable tuples (R_0, R_1, R_2) is the capacity region $\mathscr{C}$ of the considered channel. We mostly consider the sum-rate $R_{\text{sum}} := R_0 + R_1 + R_2$, which corresponds to the rate of the aggregate message $W := (W_0, W_1, W_2)$ in Fig. 1, and the corresponding sum-rate capacity $C_{\text{sum}} := \max\{R_{\text{sum}} : (R_0, R_1, R_2) \in \mathscr{C}\}$.

We now provide the main result of this section, which states that an achievable scheme based on Slepian-Wolf coding [16] over the alphabets of Shannon strategies [5] is sufficient to achieve the sum-rate capacity of the considered channel, for every state law $p(s, s_1, s_2, s_R)$. We will also discuss how this result relates to the particular cases analyzed in [10]–[13], [15].

Theorem 1. *The sum-rate capacity of the channel in Fig. 1 is given by*

$$C_{\text{sum}} = \max_{\substack{p(u_0)p(u_1|u_0)p(u_2|u_0) \\ x_k = f_k(u_k, s_k)}} I(U_1, U_2; Y|S_R), \tag{1}$$

where $(U_0, U_1, U_2) \in \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2$ are auxiliary variables of finite cardinality, independent of (S, S_1, S_2, S_R), and where f_k, $k = 1, 2$, are two deterministic functions $\mathcal{U}_k \times \mathcal{S}_k \to \mathcal{X}_k$.

Proof (Achievability): For the channel in Fig. 1, the following rate-region is achievable

$$\begin{aligned} R_1 &\leq I(U_1; Y|U_2, U_0, S_R), \\ R_2 &\leq I(U_2; Y|U_1, U_0, S_R), \\ R_1 + R_2 &\leq I(U_1, U_2; Y|U_0, S_R), \\ R_0 + R_1 + R_2 &\leq I(U_1, U_2; Y|S_R), \end{aligned} \tag{2}$$

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 probability mass function factorizing as $p(u_0)p(u_1|u_0)p(u_2|u_0)$, and for some functions $f_k : \mathcal{U}_k \times \mathcal{S}_k \to \mathcal{X}_k$, $x_k = f_k(u_k, s_k)$, $k = 1, 2$.

The detailed analysis of the the achievability of (2) is omitted for space limitations. We here provide only a sketch of the proof. By fixing the functions f_k, we can consider a new state-less and memory-less MAC with common and independent messages, inputs U_1, U_2 and output (Y, S_R) (this is a simple extension of the classical *physical device* argument of Shannon [5], [7]). The region in (2) is nothing but the Slepian-Wolf region [16] achieving the capacity region of this new MAC. Note that $I(U_1, U_2; Y, S_R) = I(U_1, U_2; Y|S_R)$, since S_R is independent of U_1, U_2. The finite cardinality of $\mathcal{U}_k$, $k = 1, 2$ follows trivially by the finite cardinality of the sets of functions $S_k \to \mathcal{X}_k$, while the finite cardinality of $\mathcal{U}_0$ follows by a simple application of the support lemma [7]. ■

Proof (Converse): Let us define $U_{0,i} = (W_0, S_1^{i-1}, S_2^{i-1})$, $U_{1,i} = (W_1, U_{0,i})$ and $U_{2,i} = (W_2, U_{0,i})$. Note that this choice of auxiliary random variables satisfies the Markov chain $U_{1,i} \to U_{0,i} \to U_{2,i}$. 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 $X_{1,i}$ and $X_{2,i}$ are functions of $(W_0, W_1, S_1^i, S_2^{i-1}) = (U_{1,i}, S_{1,i})$ and $(W_0, W_2, S_2^i, S_1^{i-1}) = (U_{2,i}, S_{2,i})$ respectively. Note that $U_{1,i}, U_{2,i}$ are independent of $(S_i, S_{1,i}, S_{2,i}, S_{R,i})$ and that $(U_{1,i}, U_{2,i}) \to (X_{1,i}, S_{1,i}, X_{2,i}, S_{2,i}) \to (Y_i, S_{R,i})$ forms a Markov chain. We then have:

$$
\begin{aligned}
nR_{\mathrm{sum}} \leq & H(W) \\
=& I(W; Y^n, S_R^n) + H(W|Y^n, S_R^n) \\
\stackrel{(a)}{\leq} & I(W; Y^n, S_R^n) + n\epsilon_n \\
=& \sum_{i=1}^{n} I(W; Y_i, S_{R,i}|Y^{i-1}, S_R^{i-1}) + n\epsilon_n \\
=& \sum_{i=1}^{n} H(Y_i, S_{R,i}|Y^{i-1}, S_R^{i-1}) \\
& - H(Y_i, S_{R,i}|W, Y^{i-1}, S_R^{i-1}) + n\epsilon_n \\
\leq & \sum_{i=1}^{n} H(Y_i, S_{R,i}|Y^{i-1}, S_R^{i-1}) \\
& - H(Y_i, S_{R,i}|W, S_1^{i-1}, S_2^{i-1}, Y^{i-1}, S_R^{i-1}) + n\epsilon_n \\
\stackrel{(b)}{=} & \sum_{i=1}^{n} H(Y_i, S_{R,i}|Y^{i-1}, S_R^{i-1}) \\
& - H(Y_i, S_{R,i}|U_{1,i}, U_{2,i}) + n\epsilon_n \\
\leq & \sum_{i=1}^{n} H(Y_i, S_{R,i}) - H(Y_i, S_{R,i}|U_{1,i}, U_{2,i}) + n\epsilon_n \\
=& \sum_{i=1}^{n} I(U_{1,i}, U_{2,i}; Y_i, S_{R,i}) + n\epsilon_n \\
\leq & nC_{\mathrm{sum}} + n\epsilon_n.
\end{aligned}
$$

where (a) follows from Fano's inequality ($\lim_{n\to\infty} \epsilon_n = 0$), and (b) follows from the Markov chain $(W, Y^{i-1}, S_R^{i-1}) \to (W, S_1^{i-1}, S_2^{i-1}) \to (Y_i, S_{R,i})$ corresponding to the memoryless property of the channel. Note that $I(U_1, U_2; Y, S_R) = I(U_1, U_2; Y|S_R)$ because S_R is independent of U_1, U_2. ■

The main message of (1) is that sum-rate capacity can be achieved by Shannon strategies, i.e. by precoding on current CSIT $S_{1,i}, S_{2,i}$ only and by neglecting the past CSIT sequences S_1^{i-1}, S_2^{i-1}. The converse proof shows also that providing to both encoders the entire past CSIT sequences S_1^{i-1}, S_2^{i-1} does not increase the sum-rate capacity.

We now consider some important special cases of the channel in Fig. 1.

A. Full Message Sharing

We consider transmission of a common message W_0 only, i.e. $W_1 = W_2 = \emptyset$. This special case corresponds to the interesting cooperative scenario where the TXs perfectly share the messages, but where their cooperation capabilities are still impaired by the D-CSIT assumption. Note that due to the D-CSIT assumption we cannot in general apply the results for centralized encoding. It then holds:

Corollary 1. *For the channel in Fig. 1, the maximum achievable common rate R_0 is given by*

$$C_0 = \max_{\substack{p(u) \\ x_k = f_k(u, s_k)}} I(U; Y|S_R), \tag{3}$$

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, $k = 1, 2$, are two deterministic functions $\mathcal{U} \times \mathcal{S}_k \to \mathcal{X}_k$.

Proof (sketch): The proof follows by applying similar steps as for Theorem 1, but by considering a single auxiliary variable $U := U_0 = U_1 = U_2$, as there are no independent messages to transmit. ■

We point out that the above corollary includes the particular case analyzed in [13, Corollary 3]. This result shows that, if the cooperating transmitters can fully share the messages, Shannon strategies are capacity achieving.

Remark 1. *Since from the achievable region in* (2) *the only bound for R_0 is C_{sum}, we have that $C_0 = C_{\mathrm{sum}}$ holds. Note that this does not imply that there are no other tuples (R_0, R_1, R_2) achieving C_{sum} with Shannon strategies.*

B. CSITs as functions of the CSIR

We assume the CSIT sequences to be deterministic functions of the CSIR. This assumption is suitable for frequency-division duplex (FDD) systems, where the TXs acquire channel knowledge in form of quantized feedback from the RX.

Corollary 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 sum-rate capacity given by Theorem 1 reduces to*

$$C_{\mathrm{sum}} = \max_{\substack{p(u_0)p(u_1|u_0)p(u_2|u_0) \\ x_k = f_k(u_k, s_k)}} I(X_1, X_2; Y|S_R), \tag{4}$$

for some pmf $p(u_0)p(u_1|u_0)p(u_2|u_0)$ and deterministic functions $f_k : \mathcal{U}_k \times \mathcal{S}_k \to \mathcal{X}_k$, $x_k = f_k(u_k, s_k)$, $k = 1, 2$.

Proof: The proof follows the same lines as in [10]. From Theorem 1, we observe that

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

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 because of the Markov chain $(S_1, S_2, U_1, U_2) \to (X_1, X_2, S_R) \to Y$. ■

C. Independent messages only

We consider the rate region obtained by letting $R_0 = 0$, i.e. by transmitting independent messages only ($W_0 = \emptyset$). In such case, the region in (2) reduces to the rate region achieved by Shannon strategies for a state-dependent MAC with independent messages and general CSIT and CSIR structure [10], which is obtained by considering only the bounds

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

where U_0 plays the role of a simple time-sharing variable.

By letting $R_0 = 0$, we cannot use anymore Theorem 1 to characterize the performance limits in terms of Shannon strategies, as this result relies on the fact that the rates (R_0, R_1, R_2) can take arbitrary values. In fact, for the case of independent messages, it has been shown in [15] that Shannon strategies are in general suboptimal, except for some particular CSI configurations [10]–[12]. This implies that for achieving the performance limits of this channel, in absence of a common message, a more complex coding scheme that takes into account past CSIT realizations may be required. However, such schemes are often too complicated in practice, and solutions based on simple Shannon strategies are usually preferred. An optimal scheme for the independent messages case with general CSI structure is also unknown.

Nevertheless, Theorem 1 suggests a possible alternative to circumvent the aforementioned issues. In fact, it states that the sum-rate capacity can be achieved via Shannon strategies at the cost of sufficient message sharing, i.e., R_0 should have a minimum value, which in the worst case is $R_0 = C_0$ given by (3). Finding the minimal value of R_0 so that Shannon strategies achieves the sum-rate capacity in Theorem 1 is a very interesting problem that will be discussed in the extended version of this paper.

III. Cooperative MIMO AWGN Channel with Fading and Full Message Sharing

In this section we specialize the channel in Fig. 1 by considering a 2×2 cooperative MIMO channel defined by the following input-output relation:

$$Y = SX + Z = S\begin{bmatrix}X_1\\X_2\end{bmatrix} + Z,$$

where the state $S \in \mathbb{C}^{2\times 2}$ is a matrix of random fast-fading coefficients, X_k is the signal transmitted by TX k, subject to an average power constraint $\mathbb{E}[|X_k|^2] \leq P_k$, and where $Z \sim \mathcal{CN}(0,\mathrm{I})$ is independent of S. We assume a system where the RX has perfect CSIR $S_R = S$, and where the CSIT is a quantized version of the CSIR, i.e. $S_1 = q_1(S)$ and $S_2 = q_2(S)$. The alphabets $\mathcal{S}_k$, $k = 1,2$ are finite. We recall that, as discussed in Section II-B, this assumption is particularly suitable for FDD systems. Finally, we assume full message sharing, i.e. we consider only the joint transmission of a common message W_0. Although restrictive, all these assumptions allow us to isolate the fundamental limitations of the D-CSIT assumption.

Proposition 1. *The capacity of this channel is given by*

$$C_0 = \max_G \mathbb{E}\left[\log\det\left(\mathrm{I} + SG^{\mathsf{H}}GS^{\mathsf{H}}\right)\right], \tag{5}$$

where, $\forall\ s_1, s_2$, the maximization is over the matrices

$$G = \begin{bmatrix}g_1(s_1) & g_2(s_2)\end{bmatrix} \in \mathbb{C}^{d\times 2}$$

satisfying the constraint $\mathbb{E}\left[\|g_k(S_k)\|^2\right] \leq P_k$, and where $d \leq (|\mathcal{S}_1| + |\mathcal{S}_2|)$. Furthermore, C_0 can be achieved by letting

$$\begin{bmatrix}X_1\\X_2\end{bmatrix} = \begin{bmatrix}g_1^H(S_1)\\g_2^H(S_2)\end{bmatrix}U, \quad U \sim \mathcal{CN}(0,\mathrm{I}_d).$$

Remark 2. *Proposition 1 shows that, in contrast to the corresponding centralized setting, linear precoding over a number of data streams $d = 2$, i.e. equal to the total number of TX antennas, may be suboptimal. This surprising observation is further elaborated through an example in Section III-A.*

Proof: By extending equation (4) and Corollary 1 to continuous alphabets and to input cost constraints, for example similarly to [6], [8], we write

$$\begin{aligned}C_0 &= \sup_{\substack{x_1=f_1(s_1,u)\\x_2=f_2(s_2,u)\\p(u),\ \mathbb{E}[|X_k|^2]\leq P_k}} I(X_1,X_2;Y|S)\\ &= \sup_{\substack{x_1=f_1(s_1,u)\\x_2=f_2(s_2,u)\\p(u),\ \mathbb{E}[|X_k|^2]\leq P_k}} \sum_{\substack{s_1\in\mathcal{S}_1\\s_2\in\mathcal{S}_2}} I(X_1,X_2;Y|S,s_1,s_2)p(s_1,s_2).\end{aligned}$$

Rewriting the mutual information term yields (e.g. see [7])

$$\begin{aligned}&I(X_1,X_2;Y|S,s_1,s_2)\\ &\quad= h(Y|S,s_1,s_2) - h(Y|X_1,X_2,S,s_1,s_2)\\ &\quad= h(SX+Z|S,s_1,s_2) - h(Z)\\ &\quad\leq \mathbb{E}\left[\log\det\left(\mathrm{I} + S\Sigma_X(s_1,s_2)S^{\mathsf{H}}\right)|s_1,s_2\right],\end{aligned}$$

with equality for conditionally Gaussian inputs, and where

$$\Sigma_X(s_1,s_2) := \mathbb{E}_U[XX^{\mathsf{H}}|s_1,s_2]. \tag{6}$$

Taking the supremum gives the upper bound $C_0 \leq C_{\text{upp}}$, with

$$C_{\text{upp}} := \sup \sum_{\substack{s_1\in\mathcal{S}_1\\s_2\in\mathcal{S}_2}} \mathbb{E}\left[\log\det\left(\mathrm{I} + S\Sigma_X(s_1,s_2)S^{\mathsf{H}}\right)|s_1,s_2\right] \times p(s_1,s_2). \tag{7}$$

where the maximization is over all distributions $p(u)$ of U and over all functions f_1,f_2 such that $X_1 = f_1(U,S_1)$, $X_2 = f_2(U,S_2)$, satisfying the power constraints.

We now show that any conditional covariance $\Sigma_X(s_1,s_2)$ (in particular, also the optimal in (7)) can be achieved via linear precoding. To this end, we first define the shorthand

$$\langle f,g\rangle := \int_{\mathcal{U}} f(u)g^\star(u)p(u)du = \mathbb{E}_U[f(U)g^\star(U)], \tag{8}$$

where $p(u) > 0$ is the distribution of U, and we point out that any $\Sigma_X(s_1,s_2)$ is given by the following sets of scalars

$$\langle f_1(s_{1,i}), f_1(s_{1,i})\rangle, \quad i = 1,\dots,|\mathcal{S}_1|, \tag{9}$$
$$\langle f_2(s_{2,j}), f_2(s_{2,j})\rangle, \quad j = 1,\dots,|\mathcal{S}_2|, \tag{10}$$
$$\langle f_1(s_{1,i}), f_2(s_{2,j})\rangle, \quad i = 1,\dots,|\mathcal{S}_1|, \quad j = 1,\dots,|\mathcal{S}_2|, \tag{11}$$

where $s_{k,l}$ denotes the l-th value of the random state S_k. The sets (9) and (10) describe the diagonal elements of $\Sigma_X(s_1,s_2)$, while (11) describes the off-diagonal elements.

We then build the following square matrix

$$Q := \begin{bmatrix}\{\langle f_1(s_{1,i}), f_1(s_{1,i'})\rangle\}_{ii'} & \{\langle f_1(s_{1,i}), f_2(s_{2,j})\rangle\}_{ij}\\ \{\langle f_2(s_{2,j}), f_1(s_{1,i})\rangle\}_{ji} & \{\langle f_2(s_{2,j}), f_2(s_{2,j'})\rangle\}_{jj'}\end{bmatrix}$$

of dimension $d := |\mathcal{S}_1| + |\mathcal{S}_2|$, which contains (9), (10), and (11), and hence it completely describes $\Sigma_X(s_1,s_2)$. In fact,

Q is the Gram matrix of the $|\mathcal{S}_1|+|\mathcal{S}_2|$ vectors $f_k(\cdot, s_{k,l})$ belonging to the infinite-dimensional Hilbert space $\mathcal{H}$ of square-integrable functions $\mathcal{U} \to \mathbb{C}$ equipped with the inner product $\langle\cdot,\cdot\rangle$ in (8). Note that, due to the power constraint

$$\sum_{l=1}^{|\mathcal{S}_k|} \|f_k(s_{k,l})\|^2 p(S_k = s_{k,l}) \leq P_k < \infty,$$

where $\|\cdot\|$ is the norm induced by the inner product $\langle\cdot,\cdot\rangle$, the feasible functions $f_k(\cdot, s_{k,l})$ must be square integrable, hence we are not loosing generality by restricting ourselves to $\mathcal{H}$.

From standard properties of Gram matrices (see e.g. [17, Th. 7.2.10]), Q is positive semi-definite, hence there exists a square matrix F of the same dimension d such that $F^{\mathsf{H}}F = Q$. Denote now the column vectors of $F \in \mathbb{C}^{d\times d}$ as

$$F =: \begin{bmatrix} g_1(s_{1,1}) & \dots & g_1(s_{1,|\mathcal{S}_1|}) & g_2(s_{2,1}) & \dots & g_2(s_{2,|\mathcal{S}_2|}) \end{bmatrix},$$

where the ordering of $g_k(s_{k,l})$ is consistent with the ordering of the inner products in Q. By letting $X_k = g_k^H(S_k)U$, where $U \sim \mathcal{CN}(0, \mathrm{I}_d)$, and by applying (6), we obtain exactly the original conditional covariance matrix

$$\Sigma_X(s_1, s_2) = \begin{bmatrix} g_1(s_1) & g_2(s_2) \end{bmatrix}^{\mathsf{H}} \begin{bmatrix} g_1(s_1) & g_2(s_2) \end{bmatrix}$$

from which g_1, g_2 are constructed. Finally, since X is conditionally Gaussian, we can also achieve C_{upp} in (7). ∎

A. Example

In this section, we show through an example that to span the whole feasible set of conditional covariance matrices $\Sigma_X(s_1, s_2) = G^{\mathsf{H}}G$ of problem (5), restricting the dimensionality of $g_k(s_k)$ to $d = 2$ is not in general sufficient. Assume $\mathcal{S}_1 = \mathcal{S}_2 = \{0, 1\}$, and $\Sigma_X(s_1, s_2) = G^{\mathsf{H}}G$ such that

$$\Sigma_X(0,0) = \mathrm{I}, \quad \Sigma_X(1,0) = \mathrm{I},$$
$$\Sigma_X(0,1) = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}, \quad \Sigma_X(1,1) = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}.$$

To achieve $\Sigma_X(s_1, s_2)$, we need to find precoders $g_k(s_k)$ s.t.

$$\begin{cases} g_1^{\mathsf{H}}(0)g_2(0) = 0, \quad g_1^{\mathsf{H}}(0)g_2(1) = 0.6, \\ g_1^{\mathsf{H}}(1)g_2(0) = 0, \quad g_1^{\mathsf{H}}(1)g_2(1) = 0.8, \\ \|g_1(0)\| = \|g_1(1)\| = \|g_2(0)\| = \|g_2(1)\| = 1. \end{cases}$$

For $g_k(s_k)$ of dimension $d = 2$, the above system has no solution, as we need unit norm $g_1(0), g_1(1) \perp g_2(0) \implies g_1(0) = \pm g_1(1)$, which gives $g_1^{\mathsf{H}}(0)g_2(1) = \pm g_1^{\mathsf{H}}(1)g_2(1) = \pm 0.8 \neq 0.6$. Instead, $\Sigma_X(s_1, s_2)$ is obtained by letting

$$g_1(0) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \quad g_1(1) = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix},$$
$$g_2(0) = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}, \quad g_2(1) = \begin{bmatrix} 0.6 & 0.8 & 0 \end{bmatrix}.$$

B. Discussion

This is the first information theoretical result where the extension of the classical linear precoding stage over Gaussian codewords from centralized to distributed settings is proven to be rate optimal. This result suggests that the fundamental limits of distributed transmission, in terms of sum-rate, may mirror the limits of centralized transmission from a coding architecture point of view. However, this result also unveils a fundamental difference of cooperative systems with D-CSIT. In fact, from an optimization point of view, the situation becomes significantly more complex. Firstly, in contrast with the centralized case [6] where the optimal precoder G has dimension $d \times 2 = 2 \times 2$ (obtained by any matrix square root of the optimal conditional covariance), in the distributed case a larger d may be required. Clearly, if the cardinalities of the CSIT alphabets are large, problem (5) may become quickly intractable. Finally, the design of capacity achieving linear precoders requires the solution of a Team Decision problem [1], which is non-convex due to the D-CSIT assumption.

In the extended version of this paper, we plan to repeat the analysis in Proposition 1 to include also independent messages. Another interesting line of work is to study the effect of a similar increase in dimensionality of distributed linear precoders in systems with interference.

Acknowledgment

The authors would like to thank Mari Kobayashi (TUM/Centrale Supelec) for the fruitful discussions that improved the final quality of this manuscript.

References

[1] 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.

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

[3] A. Bazco Nogueras, L. Miretti, P. de Kerret, D. Gesbert, and N. Gresset, "Achieving vanishing rate loss in decentralized network MIMO," *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019.

[4] F. M. Willems, "Information theoretical results for the discrete memoryless multiple access channel," Ph.D. dissertation, Katholieke Universiteit Leuven, 1982.

[5] C. E. Shannon, "Channels with side information at the transmitter," *IBM journal of Research and Development*, vol. 2, no. 4, pp. 289–293, 1958.

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

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

[8] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 2007–2019, 1999.

[9] S. Sigurjónsson and Y.-H. Kim, "On multiple user channels with state information at the transmitters," in *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*. IEEE, 2005, pp. 72–76.

[10] S. Jafar, "Capacity with causal and noncausal side information: A unified view," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5468–5474, 2006.

[11] G. Como and S. Yüksel, "On the capacity of memoryless finite-state multiple-access channels with asymmetric state information at the encoders," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1267–1273, 2011.

[12] N. Şen, G. Como, S. Yüksel, and F. Alajaji, "On the capacity of memoryless finite-state multiple access channels with asymmetric noisy state information at the encoders," in *2011 49th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2011.

[13] A. Somekh-Baruch, S. Shamai, and S. Verdu, "Cooperative multiple-access encoding with states available at one transmitter," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4448–4469, Oct 2008.

[14] A. Zaidi and S. S. Shitz, "On cooperative multiple access channels with delayed CSI at transmitters," *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6204–6230, 2014.

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

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

[17] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.