arXiv:0807.1543v2 [cs.IT] 25 Sep 2008

# On the Capacity of MIMO Interference Channels

Xiaohu Shang
Department of EECS
Syracuse University
Email: xshang@syr.edu

Biao Chen
Department of EECS
Syracuse University
Email: bichen@syr.edu

Gerhard Kramer
Bell Laboratories
Alcatel-Lucent
Email: gkr@research.bell-labs.com

H. Vincent Poor
Department of EE
Princeton University
Email: poor@princeton.edu

***Abstract*— The capacity region of a multiple-input-multiple-output interference channel (MIMO IC) in which the channel matrices are square and invertible is studied. The capacity region for strong interference is established where the definition of strong interference parallels that of scalar channels. Moreover, the sum-rate capacity for Z interference, noisy interference, and mixed interference is established. These results generalize known results for the scalar Gaussian IC.**

## I. Introduction

The interference channel (IC) models the situation in which transmitters communicate with their respective receivers while generating interference to all other receivers. This channel model was mentioned in [1, Section 14] and its capacity region is still generally unknown.

In [2] Carleial showed that interference does not reduce capacity when it is very strong. This result follows because the interference can be decoded and subtracted at each receiver before decoding the desired message. Later Han and Kobayashi [3] and Sato [4] showed that the capacity region of the strong interference channel is the same as the capacity region of a compound multiple access channel. In both above cases, the interference is fully decoded at both receivers.

When the interference is not strong, the capacity region is unknown. The best inner bound is by Han and Kobayashi [3], which was later simplified by Chong *et al.* in [5], [6]. Etkin *et al.* and Telatar and Tse showed that Han and Kobayashi's inner bound is within one bit of the capacity region of scalar Gaussian ICs [7], [8]. Various outer bounds have been developed in [7]–[12].

Special ICs such as the degraded IC and the ZIC have been studied in [13], [14]. The sum-rate capacity for the ZIC was established in [13], [15], and Costa proved the equivalence of the ZIC and the degraded IC for the scalar Gaussian case [14]. A recent result in [10]–[12] has shown that if a simple condition is satisfied, then treating interference as noise can achieve the sum-rate capacity. [11] and [16] derived the sum-rate capacity for mixed interference, i.e., one receiver experiences strong interference and the other experiences weak interference.

In this paper, we study the sum-rate capacity of the two-user Gaussian multiple-input-multiple-output (MIMO) IC shown in Fig. 1. The received signals are defined as

[0]This research was supported in part by the National Science Foundation under Grants CCF-05-46491, ECS-05-01534, ANI-08-38807 and CNS-06-25637.

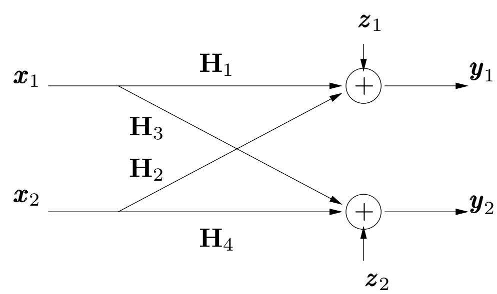


Fig. 1. The MIMO IC.

$$\begin{aligned} \boldsymbol{y}_1 &= \mathbf{H}_1\boldsymbol{x}_1 + \mathbf{H}_2\boldsymbol{x}_2 + \boldsymbol{z}_1 \\ \boldsymbol{y}_2 &= \mathbf{H}_3\boldsymbol{x}_1 + \mathbf{H}_4\boldsymbol{x}_4 + \boldsymbol{z}_2, \end{aligned} \tag{1}$$

where $\boldsymbol{x}_i, i = 1, 2$, is the transmitted signal of user $i$ which is subject to an average block power constraint $P_i$; $\boldsymbol{z}_i, i = 1, 2$ is a Gaussian random vector with zero mean and identity covariance matrix; and $\mathbf{H}_j, j = 1, \ldots, 4$, are the channel matrices. For simplicity, we assume that the $\mathbf{H}_j$'s are real and that $\mathbf{H}_1$ and $\mathbf{H}_4$ are invertible. However we remark that one can generalize our results to non-invertible or rectangular channel matrices (see Remark 1).

For the MIMO IC Telatar and Tse [8] showed that Han and Kobayashi's region is within one bit per receive antenna of the capacity region. Some upper bounds were discussed in [17] and some lower bounds on the sum-rate capacity based on Han and Kobayashi's region were discussed in [18]. However capacity results for the MIMO IC are still lacking. In our work, assuming the channel matrices are invertible, we derive the sum-rate capacity with noisy-interference, strong interference and mixed interference, as well as one-sided interference. The capacity region of the MIMO IC with strong interference is also obtained.

The rest of the paper is organized as follows: we present our main results and proofs in Section II and III; numerical results are given in Section IV, and we conclude in Section V.

Before proceeding we introduce some notation which will be used in the paper.

- Italic font $X$ denotes a scalar; and the bold fonts $\boldsymbol{x}$ and $\mathbf{X}$ denote vectors and matrices respectively.
- $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semi-definite.
- $\mathbf{I}$ denotes the identity matrix and $\mathbf{0}$ denotes the zero matrix.
- $|\mathbf{X}|$, $\mathbf{X}^T$, $\mathbf{X}^H$, $\mathbf{X}^{-1}$, $\mathbf{X}^{-T}$ denote respectively the de-

terminant, transpose, conjugate transpose, inverse, and transpose inverse of the matrix $\mathbf{X}$.

- $\boldsymbol{x}^n = \left[\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \ldots, \boldsymbol{x}_n^T\right]^T$ is a long vector which consists of a sequence of vectors $\boldsymbol{x}_i, i = 1, \ldots, n$.
- $||\mathcal{S}||$ denotes the size of the set $\mathcal{S}$.
- $abs(\cdot)$ denotes the absolute value.
- $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ means that the random vector $\boldsymbol{x}$ is Gaussian distributed with zero mean and covariance matrix $\mathbf{\Sigma}$.
- $E[\cdot]$ denotes expectation; $\mathrm{Cov}(\cdot)$ denotes covariance matrix; $I(\cdot;\cdot)$ denotes mutual information; $h(\cdot)$ denotes differential entropy with the logarithm base $e$ and $\log(\cdot) = \log_e(\cdot)$.

## II. Main Results

*Theorem 1:* For the MIMO IC defined in (1), and where the channel matrices $\mathbf{H}_1$ and $\mathbf{H}_4$ are square and invertible, the sum-rate capacity is achieved by treating interference as noise at both receivers if for any covariance matrices $\mathbf{S}_i, i = 1, 2$, with $\mathrm{tr}(\mathbf{S}_i) \leq P_i$, the following conditions are satisfied:

$$\max_{\boldsymbol{\alpha}^H \boldsymbol{\alpha} = 1} abs\left(\boldsymbol{\alpha}^H \mathbf{M}^{-\frac{1}{2}} \mathbf{W}_1 \mathbf{M}^{-\frac{1}{2}} \boldsymbol{\alpha}\right) \leq \frac{1}{2}, \tag{2}$$

and

$$\max_{\boldsymbol{\alpha}^H \boldsymbol{\alpha} = 1} abs\left(\boldsymbol{\alpha}^H \mathbf{M}^{-\frac{1}{2}} \mathbf{W}_2 \mathbf{M}^{-\frac{1}{2}} \boldsymbol{\alpha}\right) \leq \frac{1}{2}, \tag{3}$$

where

$$\begin{aligned}
\mathbf{M} &= \mathbf{I} - \mathbf{A}_1 \mathbf{A}_1^T - \mathbf{A}_2 \mathbf{A}_2^T, && (4)\\
\mathbf{W}_1 &= \mathbf{A}_1^T \mathbf{A}_2^T, && (5)\\
\mathbf{W}_2 &= \mathbf{A}_2^T \mathbf{A}_1^T, && (6)\\
\mathbf{A}_1 &= \left(\mathbf{I} + \mathbf{H}_2 \mathbf{S}_2 \mathbf{H}_2^T\right) \mathbf{H}_1^{-T} \mathbf{H}_3^T \quad \text{and} && (7)\\
\mathbf{A}_2 &= \left(\mathbf{I} + \mathbf{H}_3 \mathbf{S}_1 \mathbf{H}_3^T\right) \mathbf{H}_4^{-T} \mathbf{H}_2^T. && (8)
\end{aligned}$$

The sum-rate capacity is the solution of the following optimization problem

$$\begin{aligned}
\max \quad & \frac{1}{2} \log \left| \mathbf{I} + \mathbf{H}_1 \mathbf{S}_1 \mathbf{H}_1^T \left(\mathbf{I} + \mathbf{H}_2 \mathbf{S}_2 \mathbf{H}_2^T\right)^{-1} \right| \\
& + \frac{1}{2} \log \left| \mathbf{I} + \mathbf{H}_4 \mathbf{S}_2 \mathbf{H}_4^T \left(\mathbf{I} + \mathbf{H}_3 \mathbf{S}_1 \mathbf{H}_3^T\right)^{-1} \right| \\
\text{subject to} \quad & \mathrm{tr}(\mathbf{S}_1) \leq P_1, \mathrm{tr}(\mathbf{S}_2) \leq P_2, \\
& \mathbf{S}_1 \succeq \mathbf{0}, \mathbf{S}_2 \succeq \mathbf{0}.
\end{aligned} \tag{9}$$

*Remark 1:* Theorem 1 can be generalized to the MIMO ICs with the channel matrices $\mathbf{H}_1$ and $\mathbf{H}_4$ being non-invertible or rectangular. In those cases, two additional conditions must be satisfied such that the matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ exist. This result will be reported in a subsequent paper.

*Remark 2:* In the scalar case, if we have $\mathbf{H}_1 = \mathbf{H}_4 = 1$, $\mathbf{H}_2 = \sqrt{a}$, $\mathbf{H}_3 = \sqrt{b}$, from (2) and (3) we obtain

$$\sqrt{a}(1 + bP_1) + \sqrt{b}(1 + aP_2) \leq 1. \tag{10}$$

Therefore Theorem 1 is an extension of the noisy-interference sum-rate capacity of the scalar IC [10]–[12] to the MIMO IC.

*Remark 3:* Theorem 1 is also valid by replacing the power constraint with the covariance matrix constraint. This extension applies to all the following theorems.

*Theorem 2:* For the MIMO IC defined in (1), and where the channel matrices $\mathbf{H}_1$ and $\mathbf{H}_4$ are square and invertible, the sum-rate capacity is achieved by treating interference as noise at both receivers, if for any covariance matrices $\mathbf{S}_i, i = 1, 2$, with $\mathrm{tr}(\mathbf{S}_i) \leq P_i$, there exist symmetric positive definite matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ satisfying the following conditions

$$\begin{aligned}
\mathbf{\Sigma}_1 &\preceq \mathbf{I} - \mathbf{A}_2 \mathbf{\Sigma}_2^{-1} \mathbf{A}_2^T \quad \text{and} && (11)\\
\mathbf{\Sigma}_2 &\preceq \mathbf{I} - \mathbf{A}_1 \mathbf{\Sigma}_1^{-1} \mathbf{A}_1^T, && (12)
\end{aligned}$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are defined in Theorem 1.

Theorem 2 is another description of a sufficient condition for single-user detection to be sum-rate optimal. It can be shown that for the scalar case, (11) and (12) reduce to (10).

*Theorem 3:* For the MIMO IC defined in (1) with $\mathbf{H}_3 = \mathbf{0}$ and $\mathbf{H}_2$ and $\mathbf{H}_4$ square and invertible, the sum-rate capacity is

$$\begin{aligned}
C^* = \max_{\substack{\mathrm{tr}(\mathbf{S}_1) \leq P_1 \\ \mathrm{tr}(\mathbf{S}_2) \leq P_2}} & \frac{1}{2} \log \left| \mathbf{I} + \mathbf{H}_1 \mathbf{S}_1 \mathbf{H}_1^T \left(\mathbf{I} + \mathbf{H}_2 \mathbf{S}_2 \mathbf{H}_2^T\right)^{-1} \right| \\
& + \frac{1}{2} \log \left| \mathbf{I} + \mathbf{H}_4 \mathbf{S}_2 \mathbf{H}_4^T \right|,
\end{aligned} \tag{13}$$

if the following condition is satisfied

$$\mathbf{H}_2^T \mathbf{H}_2 \prec \mathbf{H}_4^T \mathbf{H}_4. \tag{14}$$

Furthermore,

$$C^* = \max_{\mathrm{tr}(\mathbf{S}_1) \leq P_1, \mathrm{tr}(\mathbf{S}_2) \leq P_2} \min \left\{ \begin{array}{c} \frac{1}{2} \log \left| \mathbf{I} + \mathbf{H}_1 \mathbf{S}_1 \mathbf{H}_1^T + \mathbf{H}_2 \mathbf{S}_2 \mathbf{H}_2^T \right|, \\ \frac{1}{2} \log \left| \mathbf{I} + \mathbf{H}_1 \mathbf{S}_1 \mathbf{H}_1^T \right| + \frac{1}{2} \log \left| \mathbf{I} + \mathbf{H}_4 \mathbf{S}_2 \mathbf{H}_4^T \right| \end{array} \right\} \tag{15}$$

if

$$\mathbf{H}_2^T \mathbf{H}_2 \succeq \mathbf{H}_4^T \mathbf{H}_4. \tag{16}$$

Theorem 3 gives the sum-rate capacity of a MIMO ZIC. Specifically, when $\mathbf{H}_2^T \mathbf{H}_2 \prec \mathbf{H}_4^T \mathbf{H}_4$ we consider the interference to be weak and the sum-rate capacity can be achieved by treating the interference as noise. When $\mathbf{H}_2^T \mathbf{H}_2 \succeq \mathbf{H}_4^T \mathbf{H}_4$ we consider the interference to be strong and the sum-rate capacity can be achieved by fully decoding the interference.

*Theorem 4:* For the MIMO IC defined in (1), and where the channel matrices are square and invertible, if $\mathbf{H}_2^T \mathbf{H}_2 \succeq \mathbf{H}_4^T \mathbf{H}_4$ and $\mathbf{H}_3^T \mathbf{H}_3 \succeq \mathbf{H}_1^T \mathbf{H}_1$, then the sum-rate capacity is

$$C^* = \max_{\mathrm{tr}(\mathbf{S}_1) \leq P_1, \mathrm{tr}(\mathbf{S}_2) \leq P_2} \min \left\{ \begin{array}{c} \frac{1}{2} \log \left| \mathbf{I} + \mathbf{H}_1 \mathbf{S}_1 \mathbf{H}_1^T + \mathbf{H}_2 \mathbf{S}_2 \mathbf{H}_2^T \right|, \\ \frac{1}{2} \log \left| \mathbf{I} + \mathbf{H}_3 \mathbf{S}_1 \mathbf{H}_3^T + \mathbf{H}_4 \mathbf{S}_2 \mathbf{H}_4^T \right|, \\ \frac{1}{2} \log \left| \mathbf{I} + \mathbf{H}_1 \mathbf{S}_1 \mathbf{H}_1^T \right| + \frac{1}{2} \log \left| \mathbf{I} + \mathbf{H}_4 \mathbf{S}_2 \mathbf{H}_4^T \right| \end{array} \right\} \tag{17}$$

Theorem 4 shows that if $\mathbf{H}_2^T\mathbf{H}_2 \succeq \mathbf{H}_4^T\mathbf{H}_4$ and $\mathbf{H}_3^T\mathbf{H}_3 \succeq \mathbf{H}_1^T\mathbf{H}_1$ is satisfied, then the receivers experience strong interference. Thus the channel acts as a compound MIMO multiple access channel and the sum-rate capacity is achieved by fully decoding the interference at both users.

*Theorem 5:* For the MIMO IC defined in (1), and where the channel matrices are square and invertible, if $\mathbf{H}_2^T\mathbf{H}_2 \prec \mathbf{H}_4^T\mathbf{H}_4$ and $\mathbf{H}_3^T\mathbf{H}_3 \succeq \mathbf{H}_1^T\mathbf{H}_1$, then the sum-rate capacity is

$$C^* = \max_{\mathrm{tr}(\mathbf{S}_1)\le P_1, \mathrm{tr}(\mathbf{S}_2)\le P_2} \min \left\{ \begin{array}{c} \frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_3\mathbf{S}_1\mathbf{H}_3^T+\mathbf{H}_4\mathbf{S}_2\mathbf{H}_4^T\right|, \\ \frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\mathbf{S}_1\mathbf{H}_1^T\left(\mathbf{I}+\mathbf{H}_2\mathbf{S}_2\mathbf{H}_2^T\right)^{-1}\right| \\ +\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_4\mathbf{S}_2\mathbf{H}_4^T\right| \end{array} \right\}. \quad (18)$$

Theorem 5 gives the sum-rate capacity of the MIMO IC with mixed interference $\mathbf{H}_2^T\mathbf{H}_2 \prec \mathbf{H}_4^T\mathbf{H}_4$ and $\mathbf{H}_3^T\mathbf{H}_3 \succeq \mathbf{H}_1^T\mathbf{H}_1$. The sum-rate capacity is achieved by treating interference as noise at the receiver that experiences weak interference and fully decoding the interference at the receiver that experiences strong interference.

*Theorem 6:* For the MIMO IC defined in (1) with $\mathbf{H}_3 = 0$ and all other channel matrices being square and invertible, if $\mathbf{H}_2^T\mathbf{H}_2 \succeq \mathbf{H}_4^T\mathbf{H}_4$, then the capacity region is

$$\bigcup_{\substack{\mathrm{tr}(\mathbf{S}_1)\le P_1\\ \mathrm{tr}(\mathbf{S}_2)\le P_2}} \left\{ \begin{array}{c} R_1 \le \frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\mathbf{S}_1\mathbf{H}_1^T\right| \\ R_2 \le \frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_4\mathbf{S}_2\mathbf{H}_4^T\right| \\ R_1+R_2 \le \frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\mathbf{S}_1\mathbf{H}_1^T+\mathbf{H}_2\mathbf{S}_2\mathbf{H}_2^T\right| \end{array} \right\}$$

*Theorem 7:* For the MIMO IC defined in (1), and where the channel matrices $\mathbf{H}_2$ and $\mathbf{H}_4$ are square and invertible, if $\mathbf{H}_2^T\mathbf{H}_2 \succeq \mathbf{H}_4^T\mathbf{H}_4$ and $\mathbf{H}_3^T\mathbf{H}_3 \succeq \mathbf{H}_1^T\mathbf{H}_1$, the capacity region is

$$\bigcup_{\substack{\mathrm{tr}(\mathbf{S}_1)\le P_1\\ \mathrm{tr}(\mathbf{S}_2)\le P_2}} \left\{ \begin{array}{c} R_1 \le \frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\mathbf{S}_1\mathbf{H}_1^T\right| \\ R_2 \le \frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_4\mathbf{S}_2\mathbf{H}_4^T\right| \\ R_1+R_2 \le \frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\mathbf{S}_1\mathbf{H}_1^T+\mathbf{H}_2\mathbf{S}_2\mathbf{H}_2^T\right| \\ R_1+R_2 \le \frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_3\mathbf{S}_1\mathbf{H}_3^T+\mathbf{H}_4\mathbf{S}_2\mathbf{H}_4^T\right| \end{array} \right\}$$

Theorems 6 and 7 give the capacity region of the MIMO ZIC and MIMO IC under strong interference.

Finally we connect the MIMO IC with the parallel Gaussian interference channel (PGIC), which is a special case of (1) with all $\mathbf{H}_i$'s being diagonal matrices. In [19] we present conditions for which single user detection for each sub-channel is sum-rate optimal under the assumption that the coding and decoding is independent across sub-channels. The following theorem proves that independent coding and decoding is indeed sum-rate optimal under noisy-interference if some conditions are satisfied.

*Theorem 8:* For the MIMO IC defined in (1) with $\mathbf{H}_i = diag\left(h_{i1},\ldots,h_{it}\right), i = 1,\ldots,4$, let $P_{1i}^*$ and $P_{2i}^*$ be the optimal solution of the following optimization problem

$$\begin{aligned} \max \quad & \frac{1}{2}\sum_{i=1}^t \left[\log\left(1+\frac{h_{1i}^2P_{1i}}{1+h_{2i}^2P_{2i}}\right) + \log\left(1+\frac{h_{4i}^2P_{2i}}{1+h_{3i}^2P_{1i}}\right)\right] \\ \text{subject to} \quad & \sum_{1=1}^t P_{1i} \le P_1, \quad \sum_{1=1}^t P_{2i} \le P_2 \\ & P_{1i}\ge 0, \quad P_{2i} \ge 0. \end{aligned} \quad (19)$$

Then the sum-rate capacity is

$$\begin{aligned} C^* &= \sum_{i=1}^t C_i(P_{1i}^*, P_{2i}^*) \quad (20) \\ &= \frac{1}{2}\sum_{i=1}^t \left[\log\left(1+\frac{h_{1i}^2P_{1i}^*}{1+h_{2i}^2P_{2i}^*}\right) + \log\left(1+\frac{h_{4i}^2P_{2i}^*}{1+h_{3i}^2P_{1i}^*}\right)\right] \end{aligned}$$

if

$$\begin{aligned} abs\left(h_{1i}h_{2i}\right)\left(1+h_{3i}^2P_{1i}^*\right) + abs\left(h_{3i}h_{4i}\right)\left(1+h_{2i}^2P_{2i}^*\right) \\ \le abs\left(h_{1i}h_{4i}\right), \end{aligned} \quad (21)$$

and

$$\bigcap_{i=1}^t \partial C_i(P_{1i}^*, P_{2i}^*) \ne \phi, \quad (22)$$

for all $i = 1,\ldots,t$, where $\partial C_i(P_{1i}^*, P_{2i}^*)$ denotes the subdifferential of $C_i(\cdot,\cdot)$ at point $(P_{1i}^*, P_{2i}^*)$, and $\phi$ denotes the empty set. The notion of subdifferential follows that in [20].

Theorem 8 illustrates that if each sub-channel (each antenna pair in MIMO IC) satisfies the noisy-interference condition, then independent decoding at each sub-channel with single-user detection achieves the sum-rate capacity. Theorem 8 shows the conditions for *independent* coding and single-user detection across sub-channels to be optimal.

## III. Proof of the Main Results

### A. Preliminaries

We introduce some lemmas that we use to prove our main results.

*Lemma 1:* [21, Lemma 1] Let $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ be zero-mean random vectors and denote the covariance matrix of the stacked vector $\left[\boldsymbol{x}_1^T,\ldots,\boldsymbol{x}_n^T\right]^T$ as $\mathbf{K}$. Let $\mathcal{S}$ be a subset of $\{1,2,\ldots,n\}$ and $\bar{\mathcal{S}}$ be its complement. Then we have

$$h\left(\boldsymbol{x}_{\mathcal{S}}\left|\boldsymbol{x}_{\bar{\mathcal{S}}}\right.\right) \le h\left(\boldsymbol{x}_{\mathcal{S}}^*\left|\boldsymbol{x}_{\bar{\mathcal{S}}}^*\right.\right), \quad (23)$$

where $\left[\boldsymbol{x}_1^T,\ldots,\boldsymbol{x}_n^T\right]^T \sim \mathcal{N}\left(\mathbf{0},\mathbf{K}\right)$.

*Lemma 2:* Let $\boldsymbol{x}_i^n = \left[\boldsymbol{x}_{i,1}^T,\ldots,\boldsymbol{x}_{i,n}^T\right]^T, i = 1,\ldots,k$, be $k$ stacked random vectors each of which consists of $n$ vectors. Let $\boldsymbol{y}^n = \left[\boldsymbol{y}_1^T,\ldots,\boldsymbol{y}_n^T\right]^T$ be $n$ Gaussian random vectors with covariance matrix

$$\sum_{i=1}^k \lambda_i \mathrm{Cov}\left(\boldsymbol{x}_i^n\right) = \mathrm{Cov}\left(\boldsymbol{y}^n\right), \quad (24)$$

where $\sum_{i=1}^{k}\lambda_i=1,\lambda_i\geq 0$. Let $\mathcal{S}$ be a subset of $\{1,2,\ldots,n\}$ and $\bar{\mathcal{S}}$ be its complement. Then we have

$$\sum_{i=1}^{k}\lambda_i h\left(\boldsymbol{x}_{i,\mathcal{S}}\left|\boldsymbol{x}_{i,\bar{\mathcal{S}}}\right.\right)\leq h\left(\boldsymbol{y}_{\mathcal{S}}\left|\boldsymbol{y}_{\bar{\mathcal{S}}}\right.\right). \tag{25}$$

The proof of Lemma 2 is given in the appendix. Lemma 2 shows the concave-like property of the conditional entropy $h\left(\boldsymbol{x}_{\mathcal{S}}\left|\boldsymbol{x}_{\bar{\mathcal{S}}}\right.\right)$ over the covariance matrix $\mathrm{Cov}\left(\boldsymbol{x}^n\right)$.

Consider a special case of Lemma 2 with $n=1$, $\bar{\mathcal{S}}$ being the empty set and $\lambda_i=1/k$. We obtain the following lemma.

*Lemma 3:* Let $\boldsymbol{x}^k$ be a set of $k$ random vectors. Then

$$h\left(\boldsymbol{x}^k\right)\leq k\cdot h\left(\widehat{\boldsymbol{x}}^*\right), \tag{26}$$

where $\widehat{\boldsymbol{x}}^*$ is a Gaussian vector with the covariance matrix

$$\mathrm{Cov}\left(\widehat{\boldsymbol{x}}^*\right)=\frac{1}{k}\sum_{i=1}^{k}\mathrm{Cov}\left(\boldsymbol{x}_i\right). \tag{27}$$

Let $n=2,||\mathcal{S}||=||\bar{\mathcal{S}}||=1$ and $\lambda_i=1/k$. We obtain another special case of Lemma 2.

*Lemma 4:* Let $\boldsymbol{x}^k$ and $\boldsymbol{y}^k$ be two sequences of random vectors. Then we have

$$h\left(\boldsymbol{y}^k\left|\boldsymbol{x}^k\right.\right)\leq k\cdot h\left(\widehat{\boldsymbol{y}}^*\left|\widehat{\boldsymbol{x}}^*\right.\right), \tag{28}$$

where $\widehat{\boldsymbol{x}}^*$ and $\widehat{\boldsymbol{y}}^*$ are Gaussian vectors with the joint covariance matrix

$$\mathrm{Cov}\begin{bmatrix}\widehat{\boldsymbol{x}}^*\\ \widehat{\boldsymbol{y}}^*\end{bmatrix}=\frac{1}{k}\sum_{i=1}^{k}\mathrm{Cov}\begin{bmatrix}\boldsymbol{x}_i\\ \boldsymbol{y}_i\end{bmatrix}. \tag{29}$$

The proof is straightforward from Lemma 2 by noticing that $h\left(\boldsymbol{y}^k\left|\boldsymbol{x}^k\right.\right)\leq\sum_{i=1}^{k}h\left(\boldsymbol{y}_i\left|\boldsymbol{x}_i\right.\right)$.

*Lemma 5:* [22, Lemma II.2] Let $\boldsymbol{x}^*\sim\mathcal{N}\left(\mathbf{0},\mathbf{K}_x\right)$, and let $\boldsymbol{z}$ and $\boldsymbol{z}^*$ be real random vectors (independent of $\boldsymbol{x}^*$) with the same covariance matrix $\mathbf{K}_z$. If $\boldsymbol{z}^*\sim\mathcal{N}\left(\mathbf{0},\mathbf{K}_z\right)$, and $\boldsymbol{z}$ has any other distribution with covariance matrix $\mathbf{K}_z$ then

$$I\left(\boldsymbol{x}^*;\boldsymbol{x}^*+\boldsymbol{z}\right)\geq I\left(\boldsymbol{x}^*;\boldsymbol{x}^*+\boldsymbol{z}^*\right). \tag{30}$$

If $\mathbf{K}_z\succ\mathbf{0}$, then equality is achieved if and only if $\boldsymbol{z}\sim\mathcal{N}\left(\mathbf{0},\mathbf{K}_z\right)$.

*Lemma 6:* Let $\boldsymbol{x}^n$ be a sequence of $n$ zero mean random vectors. Let $\boldsymbol{z}$ and $\tilde{\boldsymbol{z}}$ be two independent Gaussian random vectors and $\boldsymbol{z}^n$ and $\tilde{\boldsymbol{z}}^n$ be two sequences of random vectors each independent and identically distributed (i.i.d.) as $\boldsymbol{z}$ and $\tilde{\boldsymbol{z}}$ respectively, then

$$\begin{aligned}&h\left(\boldsymbol{x}^n+\boldsymbol{z}^n\right)-h\left(\boldsymbol{x}^n+\boldsymbol{z}^n+\tilde{\boldsymbol{z}}^n\right)\\ &\qquad\leq nh\left(\widehat{\boldsymbol{x}}^*+\boldsymbol{z}\right)-nh\left(\widehat{\boldsymbol{x}}^*+\boldsymbol{z}+\tilde{\boldsymbol{z}}\right),\end{aligned} \tag{31}$$

where $\widehat{\boldsymbol{x}}^*$ is a zero mean Gaussian random vector with covariance matrix

$$\mathrm{Cov}\left(\widehat{\boldsymbol{x}}^*\right)=\frac{1}{n}\sum_{i=1}^{n}\mathrm{Cov}\left(\boldsymbol{x}_i\right). \tag{32}$$

The proof is given in the Appendix.

*Lemma 7:* $\begin{bmatrix}\mathbf{I} & \mathbf{A}\\ \mathbf{A}^T & \mathbf{B}\end{bmatrix}\succeq\mathbf{0}$ if and only if $\mathbf{B}\succeq\mathbf{A}^T\mathbf{A}$.

The proof is omitted.

*Lemma 8:* [23, Theorem 5.2] Suppose $\mathbf{W}$ is nonsingular and $\mathbf{M}$ is positive definite. Then the matrix equation

$$\mathbf{X}+\mathbf{W}^H\mathbf{X}^{-1}\mathbf{W}=\mathbf{M} \tag{33}$$

has a positive definite solution $\mathbf{X}$ if and only if

$$\max_{\boldsymbol{\alpha}^H\boldsymbol{\alpha}=1} abs\left(\boldsymbol{\alpha}^H\mathbf{M}^{-\frac{1}{2}}\mathbf{W}\mathbf{M}^{-\frac{1}{2}}\boldsymbol{\alpha}\right)\leq\frac{1}{2}. \tag{34}$$

### B. Proof of Theorem 1

Suppose the channel is used $n$ times. The transmit and receive vector sequences are denoted by $\boldsymbol{x}_i^n$ and $\boldsymbol{y}_i^n$ for user $i$, $i=1,2$. For the $j$th use of the channel, the covariance matrix of $\boldsymbol{x}_{i,j}$ is denoted as $\mathbf{S}_{i,j}$, $j=1,\ldots,n$, and we use the power constraints

$$\sum_{j=1}^{n}\mathrm{tr}\left(\mathbf{S}_{i,j}\right)\leq nP_i. \tag{35}$$

From Fano's inequality we have that the achievable sum rate $R_1+R_2$ must satisfy

$$\begin{aligned}&n(R_1+R_2)-n\epsilon\\ &\leq I\left(\boldsymbol{x}_1^n;\boldsymbol{y}_1^n\right)+I\left(\boldsymbol{x}_2^n;\boldsymbol{y}_2^n\right)\\ &\leq I\left(\boldsymbol{x}_1^n;\boldsymbol{y}_1^n,\mathbf{H}_3\boldsymbol{x}_1^n+\boldsymbol{n}_1^n\right)+I\left(\boldsymbol{x}_2^n;\boldsymbol{y}_2^n,\mathbf{H}_2\boldsymbol{x}_2^n+\boldsymbol{n}_2^n\right)\\ &=h\left(\mathbf{H}_3\boldsymbol{x}_1^n+\boldsymbol{n}_1^n\right)-h(\boldsymbol{n}_1^n)+h\left(\boldsymbol{y}_1^n\left|\mathbf{H}_3\boldsymbol{x}_1^n+\boldsymbol{n}_1^n\right.\right)\\ &\quad-h\left(\mathbf{H}_2\boldsymbol{x}_2^n+\boldsymbol{z}_1^n\left|\boldsymbol{n}_1^n\right.\right)+h\left(\mathbf{H}_2\boldsymbol{x}_2^n+\boldsymbol{n}_2^n\right)-h(\boldsymbol{n}_2^n)\\ &\quad+h\left(\boldsymbol{y}_2^n\left|\mathbf{H}_2\boldsymbol{x}_2^n+\boldsymbol{n}_2^n\right.\right)-h\left(\mathbf{H}_3\boldsymbol{x}_1^n+\boldsymbol{z}_2^n\left|\boldsymbol{n}_2^n\right.\right)\end{aligned} \tag{36}$$

where $\boldsymbol{z}_i^n=\left[\boldsymbol{z}_{i,1}^T,\boldsymbol{z}_{i,2}^T,\ldots,\boldsymbol{z}_{i,n}^T\right]^T, i=1,2$, with all the $\boldsymbol{z}_{i,j}, j=1,\ldots,n$ independent of each other. $\boldsymbol{n}_i^n=\left[\boldsymbol{n}_{i,1}^T,\boldsymbol{n}_{i,2}^T,\ldots,\boldsymbol{n}_{i,n}^T\right]$, and $\boldsymbol{n}_{i,j}$ are i.i.d. Gaussian vectors with zero mean and covariance matrices $\mathbf{\Sigma}_i$. We further let $\boldsymbol{n}_i$ to be correlated with $\boldsymbol{z}_i$, and $E\left[\boldsymbol{z}_i\boldsymbol{n}_i^T\right]=\mathbf{A}_i$. We can write the joint distribution of $\boldsymbol{z}_i$ and $\boldsymbol{n}_i$ as

$$\begin{bmatrix}\boldsymbol{z}_i\\ \boldsymbol{n}_i\end{bmatrix}\sim\mathcal{N}\left(\mathbf{0},\begin{bmatrix}\mathbf{I} & \mathbf{A}_i\\ \mathbf{A}_i^T & \mathbf{\Sigma}_i\end{bmatrix}\right),\qquad i=1,2, \tag{37}$$

and we have

$$\mathrm{Cov}\left(\boldsymbol{z}_i\left|\boldsymbol{n}_i\right.\right)=\mathbf{I}-\mathbf{A}_i\mathbf{\Sigma}_i^{-1}\mathbf{A}_i^T. \tag{38}$$

Let

$$\mathbf{\Sigma}_1=\mathbf{I}-\mathbf{A}_2\mathbf{\Sigma}_2^{-1}\mathbf{A}_2^T; \tag{39}$$

so we have

$$\mathrm{Cov}\left(\boldsymbol{n}_1\right)=\mathrm{Cov}\left(\boldsymbol{z}_2\left|\boldsymbol{n}_2\right.\right). \tag{40}$$

Since $\boldsymbol{n}_{1,j}$ is independent of $\boldsymbol{n}_{1,k}$ and $\boldsymbol{z}_{2,j}$ is independent of $\boldsymbol{n}_{2,k}$ for any $j\neq k$, we have

$$\mathrm{Cov}\left(\boldsymbol{n}_1^n\right)=\mathrm{Cov}\left(\boldsymbol{z}_2^n\left|\boldsymbol{n}_2^n\right.\right). \tag{41}$$

Therefore we have

$$h\left(\mathbf{H}_3\boldsymbol{x}_1^n+\boldsymbol{n}_1^n\right)-h\left(\mathbf{H}_3\boldsymbol{x}_1^n+\boldsymbol{z}_2^n\left|\boldsymbol{n}_2^n\right.\right)=0. \tag{42}$$

Similarly, let

$$\boldsymbol{\Sigma}_2=\mathbf{I}-\mathbf{A}_1\boldsymbol{\Sigma}_1^{-1}\mathbf{A}_1^T; \tag{43}$$

so we have

$$h\left(\mathbf{H}_4\boldsymbol{x}_2^n+\boldsymbol{n}_2^n\right)-h\left(\mathbf{H}_2\boldsymbol{x}_2^n+\boldsymbol{z}_1^n\left|\boldsymbol{n}_1^n\right.\right)=0. \tag{44}$$

Therefore if (39) and (43) hold, (42) and (44) are constants regardless of the distribution of $\boldsymbol{x}_1^n$ and $\boldsymbol{x}_2^n$. Then we can write

$$\begin{aligned}
&h\left(\mathbf{H}_3\boldsymbol{x}_1^n+\boldsymbol{n}_1^n\right)-h\left(\mathbf{H}_3\boldsymbol{x}_1^n+\boldsymbol{z}_2^n\left|\boldsymbol{n}_2^n\right.\right)\\
&\quad=nh\left(\mathbf{H}_3\widehat{\boldsymbol{x}}_1^*+\boldsymbol{n}_1\right)-nh\left(\mathbf{H}_3\widehat{\boldsymbol{x}}_1^*+\boldsymbol{z}_2\left|\boldsymbol{n}_2\right.\right) \qquad (45)\\
&h\left(\mathbf{H}_2\boldsymbol{x}_2^n+\boldsymbol{n}_2^n\right)-h\left(\mathbf{H}_2\boldsymbol{x}_2^n+\boldsymbol{z}_1^n\left|\boldsymbol{n}_1^n\right.\right)\\
&\quad=nh\left(\mathbf{H}_2\widehat{\boldsymbol{x}}_2^*+\boldsymbol{n}_2\right)-nh\left(\mathbf{H}_2\widehat{\boldsymbol{x}}_2^*+\boldsymbol{z}_1\left|\boldsymbol{n}_1\right.\right), \qquad (46)
\end{aligned}$$

where $\widehat{\boldsymbol{x}}_1^*$ and $\widehat{\boldsymbol{x}}_2^*$ are zero mean Gaussian vectors with respective covariance matrices

$$\mathrm{Cov}\left(\widehat{\boldsymbol{x}}_1^*\right)=\frac{1}{n}\sum_{i=1}^n\mathbf{S}_{1,i}\triangleq\widehat{\mathbf{S}}_1^*, \tag{47}$$

and

$$\mathrm{Cov}\left(\widehat{\boldsymbol{x}}_2^*\right)=\frac{1}{n}\sum_{i=1}^n\mathbf{S}_{2,i}\triangleq\widehat{\mathbf{S}}_2^*. \tag{48}$$

Next by Lemma 4 we have

$$\begin{aligned}
&h\left(\boldsymbol{y}_1^n\left|\mathbf{H}_3\boldsymbol{x}_1^n+\boldsymbol{n}_1^n\right.\right)\\
&=h\left(\mathbf{H}\boldsymbol{x}_1^n+\mathbf{H}_2\boldsymbol{x}_2^n+\boldsymbol{z}_1^n\left|\mathbf{H}_3\boldsymbol{x}_1^n+\boldsymbol{n}_1^n\right.\right)\\
&\le nh\left(\mathbf{H}_1\widehat{\boldsymbol{x}}_1^*+\mathbf{H}_2\widehat{\boldsymbol{x}}_2^*+\boldsymbol{z}_1\left|\mathbf{H}_3\widehat{\boldsymbol{x}}_1^*+\boldsymbol{n}_1\right.\right)\\
&=\frac{n}{2}\log\left|\mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T+\mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T+\mathbf{I}-\left(\mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_3^T+\mathbf{A}_1\right)\right.\\
&\quad\left.\cdot\left(\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_3^T+\boldsymbol{\Sigma}_1\right)^{-1}\left(\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T+\mathbf{A}_1^T\right)\right|+\frac{n}{2}\log 2\pi e. \qquad (49)
\end{aligned}$$

Similarly, we obtain

$$\begin{aligned}
&h\left(\boldsymbol{y}_2^n\left|\mathbf{H}_2\boldsymbol{x}_2^n+\boldsymbol{n}_2^n\right.\right)\\
&\le\frac{n}{2}\log\left|\mathbf{H}_4\widehat{\mathbf{S}}_2^*\mathbf{H}_4^T+\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_3^T+\mathbf{I}-\left(\mathbf{H}_4\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T+\mathbf{A}_2\right)\right.\\
&\quad\left.\cdot\left(\mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T+\boldsymbol{\Sigma}_2\right)^{-1}\left(\mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_4^T+\mathbf{A}_2^T\right)\right|+\frac{n}{2}\log 2\pi e. \qquad (50)
\end{aligned}$$

On substituting (45)-(50) into (36) we have

$$\begin{aligned}
&R_1+R_2-\epsilon\\
&\le\frac{1}{2}\log\left|\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_3^T+\boldsymbol{\Sigma}_1\right|-\frac{1}{2}\log\left|\boldsymbol{\Sigma}_1\right|\\
&\quad-\frac{1}{2}\log\left|\mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T+\mathbf{I}-\mathbf{A}_1\boldsymbol{\Sigma}_1^{-1}\mathbf{A}_1^T\right|\\
&\quad+\frac{1}{2}\log\left|\mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T+\mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T+\mathbf{I}-\left(\mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_3^T+\mathbf{A}_1\right)\right.\\
&\quad\left.\cdot\left(\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_3^T+\boldsymbol{\Sigma}_1\right)^{-1}\left(\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T+\mathbf{A}_1^T\right)\right|\\
&\quad+\frac{1}{2}\log\left|\mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T+\boldsymbol{\Sigma}_2\right|-\frac{1}{2}\log\left|\boldsymbol{\Sigma}_2\right|\\
&\quad-\frac{1}{2}\log\left|\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_3^T+\mathbf{I}-\mathbf{A}_2\boldsymbol{\Sigma}_2^{-1}\mathbf{A}_2^T\right|\\
&\quad+\frac{1}{2}\log\left|\mathbf{H}_4\widehat{\mathbf{S}}_2^*\mathbf{H}_4^T+\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_3^T+\mathbf{I}-\left(\mathbf{H}_4\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T+\mathbf{A}_2\right)\right.\\
&\quad\left.\cdot\left(\mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T+\boldsymbol{\Sigma}_2\right)^{-1}\left(\mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_4^T+\mathbf{A}_2^T\right)\right|\\
&\overset{(a)}{=}\frac{1}{2}\log\left|\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_3^T+\boldsymbol{\Sigma}_1\right|-\frac{1}{2}\log\left|\boldsymbol{\Sigma}_1\right|\\
&\quad-\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T\right|-\frac{1}{2}\log\left|\mathbf{I}-\mathbf{H}_1^{-T}\mathbf{H}_3^T\boldsymbol{\Sigma}_1^{-1}\mathbf{A}_1^T\right|\\
&\quad+\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T+\mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T\right|+\frac{1}{2}\log\left|\mathbf{I}\right.\\
&\quad\left.-\mathbf{H}_1^{-T}\mathbf{H}_3^T\left(\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_3^T+\boldsymbol{\Sigma}_1^T\right)^{-1}\left(\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T+\mathbf{A}_1^T\right)\right|\\
&\quad+\cdots\\
&\overset{(b)}{=}\frac{1}{2}\log\left|\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_3^T+\boldsymbol{\Sigma}_1\right|-\frac{1}{2}\log\left|\boldsymbol{\Sigma}_1\right|\\
&\quad-\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2\right|-\frac{1}{2}\log\left|\mathbf{I}-\boldsymbol{\Sigma}_1^{-1}\mathbf{A}_1^T\mathbf{H}_1^{-T}\mathbf{H}_3^T\right|\\
&\quad+\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T+\mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T\right|+\frac{1}{2}\log\left|\mathbf{I}\right.\\
&\quad\left.-\left(\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_3^T+\boldsymbol{\Sigma}_1^T\right)^{-1}\left(\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T+\mathbf{A}_1^T\right)\mathbf{H}_1^{-T}\mathbf{H}_3^T\right|\\
&\quad+\cdots\\
&=\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T\left(\mathbf{I}+\mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T\right)^{-1}\right|\\
&\qquad+\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_4\widehat{\mathbf{S}}_2^*\mathbf{H}_4^T\left(\mathbf{I}+\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_3^T\right)^{-1}\right|\\
&\overset{(c)}{\le}\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\mathbf{S}_1^*\mathbf{H}_1^T\left(\mathbf{I}+\mathbf{H}_2\mathbf{S}_2^*\mathbf{H}_2^T\right)^{-1}\right|\\
&\qquad+\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_4\mathbf{S}_2^*\mathbf{H}_4^T\left(\mathbf{I}+\mathbf{H}_3\mathbf{S}_1^*\mathbf{H}_3^T\right)^{-1}\right|, \qquad (51)
\end{aligned}$$

where in (a) we let

$$\mathbf{A}_1=\left(\mathbf{I}+\mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T\right)\mathbf{H}_1^{-T}\mathbf{H}_3^T, \tag{52}$$

and

$$\mathbf{A}_2=\left(\mathbf{I}+\mathbf{H}_3\widehat{\mathbf{S}}_1^*\mathbf{H}_3^T\right)\mathbf{H}_4^{-T}\mathbf{H}_2^T. \tag{53}$$

Equality (b) is from the fact $|\mathbf{I}-\mathbf{UV}|=|\mathbf{I}-\mathbf{VU}|$. Inequality (c) is from the assumption that $\mathbf{S}_1^*$ and $\mathbf{S}_2^*$ optimize (9) and the equality holds when $\widehat{\mathbf{S}}_1^*=\mathbf{S}_1^*$ and $\widehat{\mathbf{S}}_2^*=\mathbf{S}_2^*$.

The above sum rate in (51) is also achievable by treating interference as noise at each receiver, therefore the sum-rate capacity is (51), if there exist Gaussian vectors $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ with distribution in (37) that satisfies (39), (43), (52) and (53).

We consider the existence of $\boldsymbol{n}_1$. From Lemma 7, $\boldsymbol{n}_1$ exists if and only if

$$\boldsymbol{\Sigma}_1\succeq\mathbf{A}_1^T\mathbf{A}_1, \tag{54}$$

with $\mathbf{A}_1$ defined in (52).

From (43) and Woodbury identity [24]:

$$\begin{aligned}
&\left(\mathbf{A}+\mathbf{CBC}^T\right)^{-1}\\
&\quad=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{C}\left(\mathbf{B}^{-1}+\mathbf{C}^T\mathbf{A}^{-1}\mathbf{C}\right)^{-1}\mathbf{C}^T\mathbf{A}^{-1}, \qquad (55)
\end{aligned}$$

we have

$$\mathbf{\Sigma}_2^{-1} = \mathbf{I} - \mathbf{A}_1\left(-\mathbf{\Sigma}_1 + \mathbf{A}_1^T\mathbf{A}_1\right)^{-1}\mathbf{A}_1^T. \quad (56)$$

On substituting (56) into (39) we have

$$\mathbf{\Sigma}_1 = \mathbf{I} - \mathbf{A}_2\mathbf{A}_2^T + \mathbf{A}_2\mathbf{A}_1\left(\mathbf{A}_1^T\mathbf{A}_1 - \mathbf{\Sigma}_1\right)^{-1}\mathbf{A}_1^T\mathbf{A}_2^T. \quad (57)$$

Define

$$\mathbf{X} = \mathbf{\Sigma}_1 - \mathbf{A}_1^T\mathbf{A}_1 \quad (58)$$

and substitute (4) and (5) into (57). We then have the following matrix equation:

$$\mathbf{X} + \mathbf{W}_1^T\mathbf{X}^{-1}\mathbf{W}_1 = \mathbf{M}. \quad (59)$$

Equation (59) is a special case of a discrete algebraic Ricatti equation [23]. From Lemma 8, with $\mathbf{M}$ symmetric and positive definite, (59) has symmetric positive definite solution $\mathbf{X}$ if and only if (2) holds. Therefore $\boldsymbol{n}_1$ exists with condition (2). Similarly, $\boldsymbol{n}_2$ exists with condition (3).

Therefore if (2) and (3) hold for any $\mathbf{S}_i$ satisfying the power constraint, for any choice of $\mathbf{S}_{ij}, i=1,2, j=1,\ldots,n$, the sum rate must satisfy (51). This completes our proof.

### C. Proof of Theorem 2

In the proof of Theorem 1, we let (39) and (43) hold, and obtain (45) and (46). On the other hand, by Lemma 6, if (11) and (12) hold then we can still obtain (45) and (46). The rest of the proof of Theorem 2 is the same as the proof of Theorem 1. Therefore, treating interference as noise is sum-rate capacity achieving if there exist $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ that satisfy (11) and (12).

### D. Proof of Theorem 3

We provide two proofs of the first part of Theorem 3, i.e., $\mathbf{H}_2^T\mathbf{H}_2 \prec \mathbf{H}_4^T\mathbf{H}_4$. The first proof applies the same genie-aided method we used in the proof of Theorem 1. The second proof does not need a genie and is based on Lemma 6.

*1) Genie-aided proof:* This proof is similar to the proof of Theorem 1 but much simpler. Assume a Gaussian vector $\boldsymbol{n}$ which has joint distribution with $\boldsymbol{z}_1$ as

$$\begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{n} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{\Sigma} \end{bmatrix}\right). \quad (60)$$

Let $\boldsymbol{n}^n$ be a sequence of $n$ column random vectors with each $\boldsymbol{n}_i$ being i.i.d. Then from Fano's inequality we have

$$\begin{aligned}
& n(R_1+R_2) - n\epsilon \\
&\le I\left(\boldsymbol{x}_1^n;\boldsymbol{y}_1^n\right) + I\left(\boldsymbol{x}_2^n;\boldsymbol{y}_2^n\right) \\
&\le I\left(\boldsymbol{x}_1^n;\boldsymbol{y}_1^n, \mathbf{H}_1\boldsymbol{x}_1^n + \boldsymbol{n}^n\right) + I\left(\boldsymbol{x}_2^n;\boldsymbol{y}_2^n\right) \\
&= h\left(\mathbf{H}_1\boldsymbol{x}_1^n + \boldsymbol{n}^n\right) + h\left(\mathbf{H}_1\boldsymbol{x}_1^n + \mathbf{H}_2\boldsymbol{x}_2^n + \boldsymbol{z}_1^n \left|\mathbf{H}_1\boldsymbol{x}_1^n + \boldsymbol{n}^n\right.\right) \\
&\quad - h\left(\mathbf{H}_2\boldsymbol{x}_2^n + \boldsymbol{z}_1^n \left|\boldsymbol{n}^n\right.\right) + h\left(\mathbf{H}_4\boldsymbol{x}_2^n + \boldsymbol{z}_2^n\right) - h\left(\boldsymbol{n}^n\right) - h\left(\boldsymbol{z}_2^n\right) \\
&\overset{(a)}{\le} nh\left(\mathbf{H}_1\widehat{\boldsymbol{x}}_1^* + \boldsymbol{n}\right) + nh\left(\mathbf{H}_1\widehat{\boldsymbol{x}}_1^* + n\mathbf{H}_2\widehat{\boldsymbol{x}}_2^* + \boldsymbol{z}_1 \left|\mathbf{H}_1\widehat{\boldsymbol{x}}_1^* + \boldsymbol{n}\right.\right) \\
&\quad - h\left(\mathbf{H}_2\boldsymbol{x}_2^n + \boldsymbol{z}_1^n \left|\boldsymbol{n}^n\right.\right) + h\left(\mathbf{H}_4\boldsymbol{x}_2^n + \boldsymbol{z}_2^n\right) - nh\left(\boldsymbol{n}\right) - nh\left(\boldsymbol{z}_2\right) \\
&\overset{(b)}{\le} nh\left(\mathbf{H}_1\widehat{\boldsymbol{x}}_1^* + \boldsymbol{n}\right) + nh\left(\mathbf{H}_1\widehat{\boldsymbol{x}}_1^* + n\mathbf{H}_2\widehat{\boldsymbol{x}}_2^* + \boldsymbol{z}_1 \left|\mathbf{H}_1\widehat{\boldsymbol{x}}_1^* + \boldsymbol{n}\right.\right) \\
&\quad - nh\left(\mathbf{H}_2\widehat{\boldsymbol{x}}_2^* + \boldsymbol{z}_1 \left|\boldsymbol{n}\right.\right) + nh\left(\mathbf{H}_4\widehat{\boldsymbol{x}}_2^* + \boldsymbol{z}_2\right) - nh\left(\boldsymbol{n}\right) - nh\left(\boldsymbol{z}_2\right) \\
&\overset{(c)}{=} \frac{n}{2}\Big(\log\left|\mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T + \mathbf{\Sigma}\right| - \log\left|\mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T + \mathbf{I} - \mathbf{A}\mathbf{\Sigma}^{-1}\mathbf{A}^T\right| \\
&\quad + \log\left|\mathbf{H}_4\widehat{\mathbf{S}}_2^*\mathbf{H}_4^T + \mathbf{I}\right| + \log\Big|\mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T + \mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T + \mathbf{I} \\
&\quad - \left(\mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T + \mathbf{A}\right)\left(\mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T + \mathbf{\Sigma}\right)^{-1}\left(\mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T + \mathbf{A}^T\right)\Big| \\
&\quad - \log\left|\mathbf{\Sigma}\right|\Big) \\
&= \frac{n}{2}\left(\log\left|\mathbf{I} + \mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T\mathbf{A}^{-1}\right| + \log\left|\mathbf{I} + \mathbf{H}_4\widehat{\mathbf{S}}_2^*\mathbf{H}_4^T\right|\right) \\
&= \frac{n}{2}\Big(\log\left|\mathbf{I} + \mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T\left(\mathbf{I} + \mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2^T\right)^{-1}\right| \\
&\qquad\qquad + \log\left|\mathbf{I} + \mathbf{H}_4\widehat{\mathbf{S}}_2^*\mathbf{H}_4^T\right|\Big) \\
&\le \frac{n}{2}\cdot\max_{\substack{\mathrm{tr}(\mathbf{S}_1)\le P_1 \\ \mathrm{tr}(\mathbf{S}_2)\le P_2}}\Big(\log\left|\mathbf{I} + \mathbf{H}_1\mathbf{S}_1\mathbf{H}_1^T\left(\mathbf{I} + \mathbf{H}_2\mathbf{S}_2\mathbf{H}_2^T\right)^{-1}\right| \\
&\qquad\qquad + \log\left|\mathbf{I} + \mathbf{H}_4\mathbf{S}_2\mathbf{H}_4^T\right|\Big)
\end{aligned} \quad (61)$$

where, (a) is from Lemmas 3 and 4, and $\widehat{\boldsymbol{x}}_i^*$ is zero mean Gaussian vector with $\mathrm{Cov}\left(\boldsymbol{x}_i^*\right) = \frac{1}{n}\sum_{j=1}^n \mathrm{Cov}\left(\boldsymbol{x}_{i,j}\right)$, $i=1,2$; in (b) we let

$$\mathbf{H}_4^{-1}\mathbf{H}_4^{-T} = \mathbf{H}_2^{-1}\left(\mathbf{I} - \mathbf{A}\mathbf{\Sigma}^{-1}\mathbf{A}^T\right)\mathbf{H}_2^{-T}, \quad (62)$$

and thus

$$\begin{aligned}
& -h\left(\mathbf{H}_2\boldsymbol{x}_2^n + \boldsymbol{z}_1^n \left|\boldsymbol{n}\right.\right) + h\left(\mathbf{H}_4\boldsymbol{x}_2^n + \boldsymbol{z}_2^n\right) \\
&= -n\log\left(abs\left|\mathbf{H}_2\right|\right) + n\log\left(abs\left|\mathbf{H}_4\right|\right) \\
&= -nh\left(\mathbf{H}_2\widehat{\boldsymbol{x}}_2^* + \boldsymbol{z}_1\left|\boldsymbol{n}\right.\right) + nh\left(\mathbf{H}_4\widehat{\boldsymbol{x}}_2^* + \boldsymbol{z}_2\right);
\end{aligned} \quad (63)$$

in (c) we let

$$\mathbf{A} = \mathbf{I} + \mathbf{H}_2\widehat{\mathbf{S}}_2^*\mathbf{H}_2. \quad (64)$$

In order that all the equalities in (61) hold, there must exist $\boldsymbol{n}$ such that the covariance matrix in (60) satisfies (62) and (64). From (62) and (64) we have

$$\mathbf{\Sigma} = \mathbf{A}^T\left(\mathbf{I} - \mathbf{H}_2\mathbf{H}_4^{-1}\mathbf{H}_4^{-T}\mathbf{H}_2^T\right)\mathbf{A}. \quad (65)$$

Therefore $\boldsymbol{n}$ exists if and only if

$$\mathbf{I} - \mathbf{H}_2\mathbf{H}_4^{-1}\mathbf{H}_4^{-T}\mathbf{H}_2^T \succ \mathbf{0}, \quad (66)$$

which is equivalent to

$$\mathbf{H}_2^T\mathbf{H}_2 \prec \mathbf{H}_4^T\mathbf{H}_4. \quad (67)$$

*2) Proof based on Lemma 6:* Starting from Fano's inequality we have

$$\begin{aligned}
& n(R_1+R_2) - n\epsilon \\
&\le I\left(\boldsymbol{x}_1^n;\boldsymbol{y}_1^n\right) + I\left(\boldsymbol{x}_2^n;\boldsymbol{y}_2^n\right) \\
&= h\left(\mathbf{H}_1\boldsymbol{x}_1^n + \mathbf{H}_2\boldsymbol{x}_2^n + \boldsymbol{z}_1^n\right) - h\left(\mathbf{H}_2\boldsymbol{x}_2^n + \boldsymbol{z}_1^n\right) \\
&\quad + h\left(\mathbf{H}_4\boldsymbol{x}_2^n + \boldsymbol{z}_2^n\right) - h\left(\boldsymbol{z}_2^n\right) \\
&\overset{(a)}{\le} nh\left(\mathbf{H}_1\widehat{\boldsymbol{x}}_1^* + \mathbf{H}_2\widehat{\boldsymbol{x}}_2^* + \boldsymbol{z}_1\right) \\
&\quad - h\left(\boldsymbol{x}_2^n + \mathbf{H}_2^{-1}\boldsymbol{z}_1^n\right) + h\left(\boldsymbol{x}_2^n + \mathbf{H}_4^{-1}\boldsymbol{z}_2^n\right) - h\left(\boldsymbol{z}_2^n\right) \\
&\quad - n\log\left(abs\left|\mathbf{H}_2\right|\right) + n\log\left(abs\left|\mathbf{H}_4\right|\right) \\
&\overset{(b)}{\le} nh\left(\mathbf{H}_1\widehat{\boldsymbol{x}}_1^* + \mathbf{H}_2\widehat{\boldsymbol{x}}_2^* + \boldsymbol{z}_1^n\right) - nh\left(\widehat{\boldsymbol{x}}_2^* + \mathbf{H}_2^{-1}\boldsymbol{z}_1\right)
\end{aligned}$$

$$+nh\left(\widehat{\boldsymbol{x}}_2^*+\mathbf{H}_4^{-1}\boldsymbol{z}_2\right)-h\left(\boldsymbol{z}_2\right)$$
$$-n\log\left(abs\left|\mathbf{H}_2\right|\right)+n\log\left(abs\left|\mathbf{H}_4\right|\right)$$
$$=\frac{n}{2}\log\left|\mathbf{I}+\mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T\left(\mathbf{I}+\mathbf{H}_2\widehat{\mathbf{S}}_2\mathbf{H}_2^T\right)^{-1}\right|$$
$$+\frac{n}{2}\log\left|\mathbf{I}+\mathbf{H}_4\widehat{\mathbf{S}}_2^*\mathbf{H}_4^T\right|$$
$$\leq\max_{\substack{\mathrm{tr}(\mathbf{S}_1)\leq P_1\\ \mathrm{tr}(\mathbf{S}_2)\leq P_2}}\left\{\frac{n}{2}\log\left|\mathbf{I}+\mathbf{H}_1\mathbf{S}_1\mathbf{H}_1^T\left(\mathbf{I}+\mathbf{H}_2\mathbf{S}_2\mathbf{H}_2^T\right)^{-1}\right|\right.$$
$$\left.+\frac{n}{2}\log\left|\mathbf{I}+\mathbf{H}_4\mathbf{S}_2\mathbf{H}_4^T\right|\right\}, \qquad (68)$$

where (a) is from Lemma 3 and $\widehat{\boldsymbol{x}}_i^*$ is zero mean Gaussian vector with $\mathrm{Cov}\left(\boldsymbol{x}_i^*\right) = \frac{1}{n}\sum_{j=1}^n \mathrm{Cov}\left(\boldsymbol{x}_{i,j}\right)$; (b) is from Lemma 6.

Next we prove the second part of Theorem 3. The achievability of the sum rate is straightforward by letting the first receiver decode both messages. We need only to show the converse. Start from Fano's inequality and notice that $\mathbf{H}_2^{-1}\mathbf{H}_2^{-T} \preceq \mathbf{H}_4^{-1}\mathbf{H}_4^{-T}$, then the second and third terms of (b) in (68) become

$$\begin{aligned}
&-h\left(\boldsymbol{x}_2^n+\mathbf{H}_2^{-1}\boldsymbol{z}_1^n\right)+h\left(\boldsymbol{x}_2^n+\mathbf{H}_4^{-1}\boldsymbol{z}_2^n\right)\\
&=-h\left(\boldsymbol{x}_2^n+\mathbf{H}_2^{-1}\boldsymbol{z}_1^n\right)+h\left(\boldsymbol{x}_2^n+\mathbf{H}_2^{-1}\boldsymbol{z}_1^n+\tilde{\boldsymbol{z}}^n\right)\\
&=I\left(\tilde{\boldsymbol{z}}^n;\boldsymbol{x}_2^n+\mathbf{H}_2^{-1}\boldsymbol{z}_1^n+\tilde{\boldsymbol{z}}^n\right)\\
&\leq I\left(\tilde{\boldsymbol{z}}^n;\mathbf{H}_2^{-1}\boldsymbol{z}_1^n+\tilde{\boldsymbol{z}}^n\right)\\
&=-h\left(\mathbf{H}_2^{-1}\boldsymbol{z}_1^n\right)+h\left(\mathbf{H}_4^{-1}\boldsymbol{z}_2^n\right), \qquad (69)
\end{aligned}$$

where $\tilde{\boldsymbol{z}} \sim \mathcal{N}\left(\mathbf{0},\mathbf{H}_4^{-1}\mathbf{H}_4^{-T}-\mathbf{H}_2^{-1}\mathbf{H}_2^{-T}\right)$. On substituting (69) back into (68) we have

$$R_1+R_2-\epsilon\leq\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T+\mathbf{H}_2\widehat{\mathbf{S}}_2\mathbf{H}_2^T\right|. \qquad (70)$$

On the other hand, we have

$$\begin{aligned}
&n(R_1+R_2)-n\epsilon\\
&\leq I\left(\boldsymbol{x}_1^n;\boldsymbol{y}_1^n\left|\boldsymbol{x}_2^n\right.\right)+I\left(\boldsymbol{x}_2^n;\boldsymbol{y}_2^n\right)\\
&\leq\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\widehat{\mathbf{S}}_1^*\mathbf{H}_1^T\right|+\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_4\widehat{\mathbf{S}}_2^*\mathbf{H}_4^T\right|. \quad (71)
\end{aligned}$$

From (70) and (71) we have

$$\begin{aligned}
&R_1+R_2-\epsilon\\
&\leq\min\left\{\begin{array}{c}\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\widehat{\mathbf{S}}_1\mathbf{H}_1^T+\mathbf{H}_2\widehat{\mathbf{S}}_2\mathbf{H}_2^T\right|\\ \frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\widehat{\mathbf{S}}_1\mathbf{H}_1^T\right|+\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_4\widehat{\mathbf{S}}_2\mathbf{H}_4^T\right|\end{array}\right\}\\
&\leq\max_{\substack{\mathrm{Cov}(\mathbf{S}_1)\leq P_1\\ \mathrm{Cov}(\mathbf{S}_2)\leq P_2}}\min\\
&\left\{\begin{array}{c}\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\mathbf{S}_1\mathbf{H}_1^T+\mathbf{H}_2\mathbf{S}_2\mathbf{H}_2^T\right|\\ \frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\mathbf{S}_1\mathbf{H}_1^T\right|+\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_4\mathbf{S}_2\mathbf{H}_4^T\right|\end{array}\right\}. \quad (72)
\end{aligned}$$

### E. Proof of Theorem 4

The achievability is straightforward by letting both receivers decode both messages. We need only to show the converse, which can be shown by setting $\mathbf{H}_2 = \mathbf{0}$ and $\mathbf{H}_3 = \mathbf{0}$, respectively, and using Theorem 3.

### F. Proof of Theorem 5

The achievability part is straightforward by letting user 2 first decode message from user 1 and then decode its own message, and user 1 treat signals from user 2 as noise.

To prove the converse, we first let $\mathbf{H}_3 = \mathbf{0}$ and use the first part of Theorem 3, and then let $\mathbf{H}_2 = \mathbf{0}$ and use the second part of Theorem 3. We obtain

$$R_1+R_2\leq\max_{\mathrm{tr}(\mathbf{S}_1)\leq P_1,\mathrm{tr}(\mathbf{S}_2)\leq P_2}\min$$
$$\left\{\begin{array}{c}\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_3\mathbf{S}_1\mathbf{H}_3^T+\mathbf{H}_4\mathbf{S}_2\mathbf{H}_4^T\right|,\\ \frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\mathbf{S}_1\mathbf{H}_1^T\left(\mathbf{I}+\mathbf{H}_2\mathbf{S}_2\mathbf{H}_2^T\right)^{-1}\right|\\ \qquad+\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_4\mathbf{S}_2\mathbf{H}_4^T\right|,\\ \frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_1\mathbf{S}_1\mathbf{H}_1^T\right|+\frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}_4\mathbf{S}_2\mathbf{H}_4^T\right|\end{array}\right\}. \quad (73)$$

We complete the proof by pointing out that the last line of (73) is redundant because of the second line.

### G. Proof of Theorems 6, 7 and 8

Theorems 6 and 7 are consequences of Theorems 3 and 4 respectively. The proof is straightforward and hence is omitted.

The proof of Theorem 8 is also omitted due to the lack of space.

## IV. Numerical Result

Consider a symmetric MIMO IC with two transmit antennas and two receive antennas. Let $\mathbf{H}_1 = \mathbf{H}_2 = \mathbf{I}$, $\mathbf{H}_2 = \mathbf{H}_3 = \sqrt{a}\begin{bmatrix}\lambda_1 & \rho\\ \rho & \lambda_2\end{bmatrix}$, where $a$ varies from 0 to 1. Fig. 2 shows the noisy-interference sum-rate capacity v.s. $a$, for different $\lambda_1, \lambda_2$ and $\rho$. There is a range of $a$, within which the channel has noisy interference. Fig. 2 shows that the range of $a$ and the sum-rate capacity decrease as the norm of $\mathbf{H}_2$ and $\mathbf{H}_3$ increases.

## V. Conclusion

We have extended the capacity results on scalar ICs to MIMO ICs and have obtained the sum-rate capacity of the MIMO IC with noisy-interference, strong interference, and Z-interference, and the capacity region of the MIMO IC with strong interference.

## Acknowledgment

We gratefully acknowledge S. Annapureddy and V. Veeravalli from the University of Illinois at Urbana Champaign for pointing out an error in the original manuscript related to Theorems 1 and 2.

## Appendix

### A. Proof of Lemma 2

Define a discrete random variable $E$ with the distribution

$$P_E(E=i)=\lambda_i, \quad i=1,\ldots,k. \qquad (74)$$

Let the conditional distribution of $\boldsymbol{x}^n$ be

$$p_{\boldsymbol{x}^n|E}\left(\boldsymbol{x}^n\left|E=i\right.\right)=p_{\boldsymbol{x}_i^n}\left(\boldsymbol{x}_i^n\right). \qquad (75)$$

Then the probability density function of $\boldsymbol{x}^n$ is

$$p_{\boldsymbol{x}^n} = \sum_{i=1}^{k} \left( \lambda_i \cdot p_{\boldsymbol{x}_i^n} \right). \tag{76}$$

Therefore

$$\mathrm{Cov}\left(\boldsymbol{x}^n\right) = \sum_{i=1}^{k} \lambda_i \mathrm{Cov}\left(\boldsymbol{x}_i^n\right) = \mathrm{Cov}\left(\boldsymbol{y}^n\right). \tag{77}$$

Then from Lemma 1 we have

$$h\left(\boldsymbol{x}_{\mathcal{S}} \left| \boldsymbol{x}_{\bar{\mathcal{S}}} \right.\right) \le h\left(\boldsymbol{y}_{\mathcal{S}} \left| \boldsymbol{y}_{\bar{\mathcal{S}}} \right.\right). \tag{78}$$

From (75) we have

$$\begin{aligned} & h\left(\boldsymbol{x}_{\mathcal{S}} \left| \boldsymbol{x}_{\bar{\mathcal{S}}}, E \right.\right) \\ & = \sum_{i=1}^{k} P_E(E=i) h\left(\boldsymbol{x}_{i,\mathcal{S}} \left| \boldsymbol{x}_{i,\bar{\mathcal{S}}} \right.\right) \\ & = \sum_{i=1}^{k} \lambda_i h\left(\boldsymbol{x}_{i,\mathcal{S}} \left| \boldsymbol{x}_{i,\bar{\mathcal{S}}} \right.\right) \\ & \le h\left(\boldsymbol{x}_{\mathcal{S}} \left| \boldsymbol{x}_{\bar{\mathcal{S}}} \right.\right). \end{aligned} \tag{79}$$

Therefore we have

$$\sum_{i=1}^{k} \lambda_i h\left(\boldsymbol{x}_{i,\mathcal{S}} \left| \boldsymbol{x}_{i,\bar{\mathcal{S}}} \right.\right) \le h\left(\boldsymbol{y}_{\mathcal{S}} \left| \boldsymbol{y}_{\bar{\mathcal{S}}} \right.\right). \tag{80}$$

### *B. Proof of Lemma 6*

$$\begin{aligned} & h\left(\boldsymbol{x}^n + \boldsymbol{z}^n\right) - h\left(\boldsymbol{x}^n + \boldsymbol{z}^n + \tilde{\boldsymbol{z}}^n\right) \\ & = -I\left(\tilde{\boldsymbol{z}}^n; \boldsymbol{x}^n + \boldsymbol{z}^n + \tilde{\boldsymbol{z}}^n\right) \\ & \overset{(a)}{\le} -I\left(\tilde{\boldsymbol{z}}^n; \boldsymbol{x}^{*n} + \boldsymbol{z}^n + \tilde{\boldsymbol{z}}^n\right) \\ & = -h\left(\tilde{\boldsymbol{z}}^n\right) + h\left(\tilde{\boldsymbol{z}}^n \left| \boldsymbol{x}^{*n} + \boldsymbol{z}^n + \tilde{\boldsymbol{z}}^n \right.\right) \\ & \overset{(b)}{\le} -nh\left(\tilde{\boldsymbol{z}}\right) + nh\left(\tilde{\boldsymbol{z}} \left| \widehat{\boldsymbol{x}}^* + \boldsymbol{z} + \tilde{\boldsymbol{z}} \right.\right) \\ & = -nI\left(\tilde{\boldsymbol{z}}; \widehat{\boldsymbol{x}}^* + \boldsymbol{z} + \tilde{\boldsymbol{z}}\right) \\ & = nh\left(\widehat{\boldsymbol{x}}^* + \boldsymbol{z}\right) - nh\left(\widehat{\boldsymbol{x}}^* + \boldsymbol{z} + \tilde{\boldsymbol{z}}\right), \end{aligned} \tag{81}$$

where (a) is from Lemma 5 and $\boldsymbol{x}^{*n}$ has the same covariance matrix as $\boldsymbol{x}^n$, and (b) is from Lemma 4.

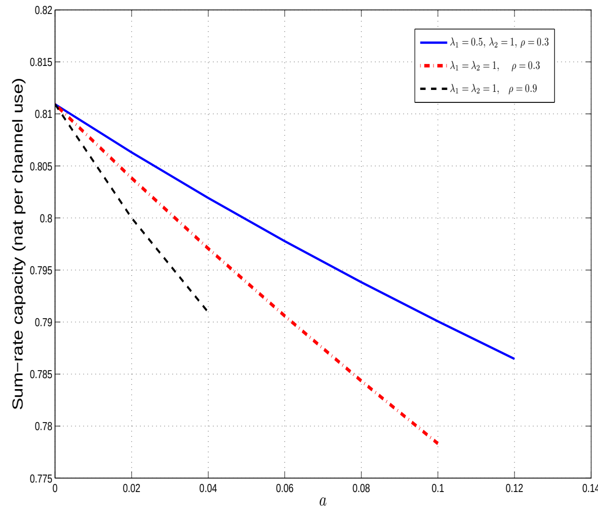


Fig. 2. Sum-rate capacity v.s. $a$

## References


[1] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Stat. and Prob.*, Berkeley, CA, 1961, vol. 1, pp. 611–644, Also available in *Claude E. Shannon: Collected Papers*, IEEE Press, New York, 1993.

[2] A. B. Carleial, "A case where interference does not reduce capacity," *IEEE Trans. Inf. Theory*, vol. 21, pp. 569–570, Sep. 1975.

[3] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, pp. 49–60, Jan. 1981.

[4] H. Sato, "The capacity of the Gaussian interference channel under strong interference," *IEEE Trans. Inf. Theory*, vol. 27, pp. 786–788, Nov. 1981.

[5] H. F. Chong, M. Motani, H. K. Garg, and H. El. Gamal, "On the Han-Kobayashi Region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3188–3195, Jul. 2008.

[6] G. Kramer, "Review of rate regions for interference channels," in *Proc. International Zurich Seminar*, Zurich, Feb.22-24 2006, pp. 162–165.

[7] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," submitted to *IEEE Trans. Inf. Theory*, 2007.

[8] E. Telatar and D. Tse, "Bounds on the capacity region of a class of interference channels," in *Proc. IEEE International Symposium on Information Theory 2007*, Nice, France, Jun. 2007, pp. 2871–2874.

[9] G. Kramer, "Outer bounds on the capacity of Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 50, pp. 581–586, Mar. 2004.

[10] X. Shang, G. Kramer, and B. Chen, "A new outer bound and the noisy-interference sum-rate capacity for Gaussian interference channels," submitted to *IEEE Trans. Inf. Theory. http://arxiv.org/abs/0712.1987*, Dec. 2007.

[11] A. S. Motahari and A. K. Khandani, "Capacity bounds for the Gaussian interference channel," submitted to *IEEE Trans. Inf. Theory. http://arxiv.org/abs/0801.1306*, Jan. 2008.

[12] V. S. Annapureddy and V. Veeravalli, "Gaussian interference networks: Sum capacity in the low interference regime and new outer bounds on the capacity region," submitted to *IEEE Trans. Inf. Theory. http://arxiv.org/abs/0802.3495*, Feb. 2008.

[13] H. Sato, "On degraded Gaussian two-user channels," *IEEE Trans. Inf. Theory*, vol. 24, pp. 634–640, Sep. 1978.

[14] M. H. M. Costa, "On the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 31, pp. 607–615, Sept. 1985.

[15] I. Sason, "On achievable rate regions for the Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 50, pp. 1345–1356, June 2004.

[16] Y. Weng and D. Tuninetti, "On Gaussian interference channels with mixed interference," in *Proc. Information Theory and Applications Workshop*, San Diego, CA, Jan. 2008.

[17] S. Vishwanath and S. A. Jafar, "On the capacity of vector Gaussian interference channels," in *Proc. IEEE Information Theory Workshop*, San Antonio, TX, Oct. 2004.

[18] X. Shang, B. Chen, and M. J. Gans, "On achievable sum rate for MIMO interference channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4313–4320, Sep. 2006.

[19] X. Shang, B. Chen, and G. Kramer, "On sum-rate capacity of parallel Gaussian symmetric interference channels," in *Proc. Globecom*, New Orleans, LA, Nov. 2008, to appear.

[20] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1997.

[21] J. A. Thomas, "Feedback can at most double Gaussian multiple access channel capacity," *IEEE Trans. Inf. Theory*, vol. 33, no. 5, pp. 711–716, Sep. 1987.

[22] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 3072–3081, Nov. 2001.

[23] J. C. Engwerda, A. C. M. Ran, and A. L. Rijkeboer, "Necessary and sufficient conditions for the existence of a positive definite solution of the matrix equation $\mathbf{X} + \mathbf{A}^*\mathbf{X}^{-1}\mathbf{A} = \mathbf{Q}$," *Linear Algebra and Its Applications*, vol. 186, pp. 255–275, 1993.

[24] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 3rd edition, 1996.