The Capacity Region of the Degraded Multiple-Input Multiple-Output Compound Broadcast Channel — Source link

H. Weingarten, Tie Liu, Shlomo Shamai, Yossef Steinberg ...+1 more authors

Institutions: Technion – Israel Institute of Technology, Texas A&M University, University of Illinois at Urbana–Champaign

Related papers:

- The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel

- The wire-tap channel

- An Extremal Inequality Motivated by Multiterminal Information-Theoretic Problems

- Broadcast channels with confidential messages

- The Secrecy Capacity Region of the Gaussian MIMO Multi-Receiver Wiretap Channel

The Capacity Region of the Degraded Multiple-Input Multiple-Output Compound Broadcast Channel

Hanan Weingarten, *Member, IEEE*, Tie Liu, *Member, IEEE*, Shlomo Shamai (Shitz), *Fellow, IEEE*, Yossef Steinberg, *Senior Member, IEEE*, and Pramod Viswanath, *Member, IEEE*

Abstract—The capacity region of a compound multiple-antenna broadcast channel is characterized when the users exhibit a certain degradedness order. The channel under consideration has two users, each user has a finite set of possible realizations. The transmitter transmits two messages, one for each user, in such a manner that regardless of the actual realizations, both users will be able to decode their messages correctly. An alternative view of this channel is that of a broadcast channel with two common messages, each common message is intended to a different set of users. The degradedness order between the two sets of realizations/users is defined through an additional, fictitious, user whose channel is degraded with respect to all realizations/users from one set while all realizations/users from the other set are degraded with respect to him.

Index Terms—Broadcast channel, capacity region, compound channel, enhancement, extremal inequalities, multiple-antenna.

I. INTRODUCTION

IN this paper we find the capacity region of a degraded compound multiple-antenna Gaussian broadcast channel (BC) with two users and two private messages. In a compound BC, each user has several possible realizations. In our case, users 1 and 2 have K_1 and K_2 possible realizations, respectively. At the receiver, each user has perfect knowledge of the actual realization but the transmitter does not. We require that regardless of the actual realizations of users 1 and 2, the messages should be received successfully. We shall also assume an order of degradedness between the users which we define later on.

An alternative view of this channel is that of a broadcast channel with common messages. The different realizations of the channel can be considered as different users to whom a common message is being transmitted. This is actually quite a realistic model as third-generation cellular systems transmit TV

broadcasts over the downlink channel [13]. In our scenario, we have two sets of users. One set is close to the transmitter while the other set is degraded with respect to (w.r.t.) the first set and is further away from the transmitter. We wish to transmit a different message to each set where each message is common to all users in the set. Interestingly, it is also possible to use the results we present here to give bounds on the sum-rate, with focus on the multiplexing gain, for the non-degraded compound BC [14].

We first consider a canonic version of the channel (which we also refer to as an aligned channel) such that

$$
\begin{aligned}
\boldsymbol{y}_i^1 &= \boldsymbol{x} + \boldsymbol{n}_i^1, & i &= 1, \ldots, K_1 \\
\boldsymbol{y}_j^2 &= \boldsymbol{x} + \boldsymbol{n}_j^2, & j &= 1, \ldots, K_2
\end{aligned}
\tag{1}
$$

where $\boldsymbol{x}$ is a real input vector (size $t \times 1$), $\boldsymbol{y}_i^1$, $i = 1, 2, \ldots, K_1$ and $\boldsymbol{y}_j^2$, $j = 1, 2, \ldots, K_2$ are real output vectors ($t \times 1$), and $\boldsymbol{n}_i^1$ and $\boldsymbol{n}_j^2$ are real Gaussian noise vectors with zero mean and covariance matrices $\boldsymbol{N}_i^1$ and $\boldsymbol{N}_j^2$.

We also assume a degradedness order such that there exists a covariance matrix $\boldsymbol{N}^*$ such that

$$
\boldsymbol{N}_i^1 \preceq \boldsymbol{N}^* \preceq \boldsymbol{N}_j^2, \quad i = 1, \ldots, K_1, \ \ j = 1, \ldots, K_2
\tag{2}
$$

where we use $\preceq$ to denote an order between two semi-definite matrices such that $\boldsymbol{A} \preceq \boldsymbol{B}$ means that $\boldsymbol{B} - \boldsymbol{A}$ is a nonnegative semi-definite matrix.

Note that we do not require that there will be any degradedness order within the groups of the first K_1 or second K_2 realizations. Furthermore, note that the requirement in (2) is not equivalent to the more general requirement defined by $\boldsymbol{N}_i^1 \preceq \boldsymbol{N}_j^2 \, \forall \, i, j$. A simple example that illustrates this appears in Appendix I. Nevertheless, our results here are limited to the case defined by (2).

Finally, we shall assume a matrix constraint on the input, $E[\boldsymbol{x}\boldsymbol{x}^T] \preceq \boldsymbol{S}$. This will allow us to give results on the capacity region of this channel under various power constraints, including the total power constraint.

Next, we broaden the scope of the discussion to the multiple-antenna channel where for each realization, the input vector is multiplied by a different linear channel matrix such that the channel is defined by

$$
\begin{aligned}
\boldsymbol{y}_i^1 &= \boldsymbol{H}_i^1 \boldsymbol{x} + \boldsymbol{n}_i^1, & i &= 1, \ldots, K_1 \\
\boldsymbol{y}_j^2 &= \boldsymbol{H}_j^2 \boldsymbol{x} + \boldsymbol{n}_j^2, & j &= 1, \ldots, K_2
\end{aligned}
\tag{3}
$$

where

- $\boldsymbol{y}_i^1$ and $\boldsymbol{y}_j^2$ are the received vectors of sizes $r_i^1 \times 1$ and $r_j^2 \times 1$, respectively.

- H_i^1 and H_j^2 are the linear channel matrices of sizes $r_i^1 \times t$ and $r_j^2 \times t$, respectively.
- $n_i^1 \sim \mathcal{N}(0, I_{r_i^1 \times r_i^1})$ and $n_j^2 \sim \mathcal{N}(0, I_{r_j^2 \times r_j^2})$ are random real Gaussian vectors.

In the above notation, we specifically mentioned the dimensions of the identity matrix. We shall omit this notation throughout the rest of the paper and the exact dimensions of the identity matrix will be clear from the context.

Note that the dimensions of matrices H may vary between group 1 and group 2 and within each group (i.e., we may have $r_j^2 \neq r_i^1, r_j^2 \neq r_k^2$, and $r_i^1 \neq r_k^1$). Though the case of realizations with different number of receive antennas might not be realistic for the compound case mentioned above, it is a possibility in the broadcast channel with common messages.

In the aligned channel (1) we determined that one channel output is degraded w.r.t. another by examining whether their noise covariances can be ordered correctly. However, in (3), all noise covariances are identity matrices and the receive vectors differ only in their linear channel matrices. Therefore, we shall use the following definition to determine a degradedness order.

Definition 1: A receive vector $y_a = H_a x + n_a$ of size $r_a \times 1$ is said to be degraded w.r.t. $y_b = H_b x + n_b$ of size $r_b \times 1$ if there exists a matrix D (of size $r_a \times r_b$) such that $DH_b = H_a$ and such that $DD^T \preceq I_{r_a \times r_a}$. Alternatively, we say that H_a is degraded w.r.t. H_b.

By examining this definition, it is not difficult to see that y_a can be emulated from y_b by multiplying Dy_b. The emulated channel only differs from the original channel by its additive noise which is now given by $Dn_b \sim \mathcal{N}(0, DD^T)$. However, as this approximated channel has less noise $(DD^T \preceq I_{r_a \times r_a})$ it is clear that any message that can be decoded by a receiver that gets y_a, can be deciphered by a receiver which has the approximated channel, or alternatively, y_b.

We can now state an equivalent degradedness requirement to that of the aligned channel: the second group of users is said to be degraded w.r.t. the first group of users if there exists a matrix H^* such that the receive vector $y^* = H^* x + n^*$ is degraded w.r.t. y_i^1 for all i and such that y_j^2 are degraded w.r.t. y^* for all j.

Indeed, in recent years, many theoretical aspects of the downlink channel (alternatively, BC) have been resolved [4], while the capacity region of the general discrete memoryless BC is still open. One important example, the capacity region of the multiple-antenna BC with a single realization for each user, has been settled [12], [16]. As we shall show, once a degradedness order is introduced we are also able to find the capacity region of the multiple-antenna compound BC by incorporating the enhancement technique from [16] and variations on the entropy extremal inequalities introduced in [12].

A related problem is that of a broadcast channel where on top of all of the individual messages there is one message which is common to all users. In prior art, this was considered part of the broadcast model (see [5] and references therein). However, in the multiple-antenna regime this problem was solved only for the case where there are no common messages. Progress towards a solution of this problem for the multiple-antenna case are reported in [11], [17].

In [9], Diggavi and Tse investigate diversity embedded codes. For that purpose, they characterize the capacity region of the degraded message set broadcast channel for the case of a parallel Gaussian channel. They assume there are two messages. One common message is transmitted to all users and a private message is either transmitted to the first user or to the first $N - 1$ (out of N) users. They further assume that there is either a degradedness order between the first and the last $N - 1$ users or between the first $N - 1$ users and the last user. This is, in fact, not too different from the problem investigated here. Due to the degradedness order in the case presented here, the first user can always decode the messages intended for the second user and therefore, the message of the second user is in fact a common message. In the following, we generalize the result given in [9] to the case of degraded vector Gaussian BCs which are not necessarily parallel and to a more general case where there are two sets of users or realizations such that no group is limited to just one realization or user.

In Section II, we briefly present the main results of this text. In Section III, we go over some important results needed for the proof of the main theorem which is given in Section IV. In Section V, we generalize the result of Section IV to the case of arbitrary linear channel matrices, and in Section VI, we give an illustrative example.

II. MAIN RESULTS

The main result of this paper is presented in the following theorem, proved in Section IV.

Theorem 1: The capacity region of the channel given by (1) and (2), $\mathcal{C}_{\mathrm{DC}}(S)$, is given by the set of all rate pairs (R_1, R_2) such that

$$R_1 = \min_{i=1,\ldots,K_1} \frac{1}{2} \log \frac{\left| Q + N_i^1 \right|}{\left| N_i^1 \right|}$$

$$R_2 = \min_{j=1,\ldots,K_2} \frac{1}{2} \log \frac{\left| S + N_j^2 \right|}{\left| Q + N_j^2 \right|}$$

for some $0 \preceq Q \preceq S$.

Note that typically, the above theorem will give us even lower rates than those obtained by a two-user BC with the worst pair of users (one from group 1 and the other from group 2). In fact, this feature resembles the one that exists in the standard compound channel where we contrast maxmin with minmax. This duality is not surprising as we require that no matter the actual realizations of the users, the messages should be decoded successfully.

As mentioned earlier, the characterization of the capacity region under a covariance constraint on the input allows us to give a general result for many types of constraints on the input, including the most practical ones such as the total power constraint and the per-antenna power constraint. The following corollary extends the result of the above theorem to the case of total power constraint.

Corollary 2: The capacity region of the channel in (1) under a total power constraint is given by

$$\mathcal{C}_{\mathrm{DC}}(P) = \bigcup_{\mathrm{tr}\{S\} \leq P} \mathcal{C}_{\mathrm{DC}}(S).$$

Proof: This is a direct result of [16, Lemma 1]. $\qquad\square$

The extension of Theorem 1 to the multiple-antenna channel defined by (3) is given in the following theorem and is proved in Section V.

Theorem 3: Let $\mathcal{C}_{\mathrm{ADC}}\left(\boldsymbol{H}_i^1, \boldsymbol{H}_j^2, \boldsymbol{S}\right)$ denote the set of all pairs (R_1, R_2) such that

$$R_1 = \min_{i=1,\ldots,K_1} \frac{1}{2}\log|\boldsymbol{H}_i^1 \boldsymbol{Q} \boldsymbol{H}_i^{1T} + \boldsymbol{I}|$$

$$R_2 = \min_{j=1,\ldots,K_2} \frac{1}{2}\log \frac{|\boldsymbol{H}_j^2 \boldsymbol{S} \boldsymbol{H}_j^{2T} + \boldsymbol{I}|}{|\boldsymbol{H}_j^2 \boldsymbol{Q} \boldsymbol{H}_j^{2T} + \boldsymbol{I}|}$$

for some $0 \preceq \boldsymbol{Q} \preceq \boldsymbol{S}$. If there exists a matrix $\boldsymbol{H}^*$ such that the receive vector $\boldsymbol{y}^* = \boldsymbol{H}^* \boldsymbol{x} + \boldsymbol{n}^*$ is degraded w.r.t. $\boldsymbol{y}_i^1$ for all i and such that $\boldsymbol{y}_j^2$ are degraded w.r.t. $\boldsymbol{y}^*$ for all j, then $\mathcal{C}_{\mathrm{ADC}}\left(\boldsymbol{H}_i^1, \boldsymbol{H}_j^2, \boldsymbol{S}\right)$ is the capacity region of the channel in (3).

III. PRELIMINARIES

In this section, we obtain intermediate results which will be used in the proof of Theorem 1 in Section IV. In that proof, we shall need an auxiliary lemma (Lemma 2) regarding the difference between the weighted sum of two sets of entropies.

We shall heavily rely on Fisher information and Fisher information inequalities as a tool to prove entropy inequalities. We shall use $J(\boldsymbol{x})$ to denote the Fisher information of the random vector $\boldsymbol{x}$ and it is defined by the following matrix:

$$J(\boldsymbol{x}) = E\left[\left(\nabla \log f_{\boldsymbol{x}}(x)\right)\left(\nabla \log f_{\boldsymbol{x}}(x)\right)^T\right].$$

One example is the Fisher Information matrix of a random Gaussian vector $\boldsymbol{n} \sim \mathcal{N}(0, \boldsymbol{N})$ which is given by $J(\boldsymbol{n}) = \boldsymbol{N}^{-1}$ ([6, the scalar case on p. 330]). We can relate the Fisher information matrix above to the covariance matrix of $\boldsymbol{x}$ through the Cramer–Rao inequality ([6, Theorem 16.6.1, p. 494], [8, Theorem 20])

$$E\left[(\boldsymbol{x} - E\boldsymbol{x})(\boldsymbol{x} - E\boldsymbol{x})^T\right] \succeq J(\boldsymbol{x})^{-1}.$$

We can also relate the Fisher information matrix to the differential entropy through the DeBruijn identity ([6, Theorem 16.6.2, pp. 494–497], [8, Theorem 14])

$$\frac{d}{d\epsilon}h(\boldsymbol{x} + \sqrt{\epsilon}\boldsymbol{n}) = \frac{1}{2}\mathrm{tr}\left\{J(\boldsymbol{x} + \sqrt{\epsilon}\boldsymbol{n})\right\}$$

where $\boldsymbol{n} \sim \mathcal{N}(0, \boldsymbol{I})$ is a standard Gaussian vector and $h(\boldsymbol{x})$ denotes the differential entropy of a random vector.

As Lemma 2 relies on a perturbation approach, as was described in [12], we first need the following Fisher information inequality result.

Lemma 1: Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two independent random vectors of length t. Then

$$J(\boldsymbol{x} + \boldsymbol{y}) \preceq \boldsymbol{A}J(\boldsymbol{x})\boldsymbol{A}^T + (\boldsymbol{I} - \boldsymbol{A})J(\boldsymbol{y})(\boldsymbol{I} - \boldsymbol{A})^T$$

for any matrix $\boldsymbol{A}$ of size $t \times t$.

Proof: The proof can be found in [12, Appendix II]. $\qquad\square$

We can now state and prove the main result of this section.

Lemma 2: Let $\boldsymbol{K}_X^*$ be a positive semi-definite matrix such that $0 \preceq \boldsymbol{K}_X^* \preceq \boldsymbol{S}$ and such that

$$\sum_{i=1}^{K_1} \lambda_i^1 \left(\boldsymbol{K}_X^* + \boldsymbol{N}_i^1\right)^{-1} = \mu \sum_{j=1}^{K_2} \lambda_j^2 \left(\boldsymbol{K}_X^* + \boldsymbol{N}_j^2\right)^{-1} + \boldsymbol{O}_2 \quad (4)$$

where $\mu > 0$, $\lambda_i^j \geq 0$, $\boldsymbol{O}_2 \succeq 0$, and $(\boldsymbol{S} - \boldsymbol{K}_X^*) \cdot \boldsymbol{O}_2 = 0$. Then, for every distribution of $\boldsymbol{x}$ such that $E[\boldsymbol{x}\boldsymbol{x}^T] \preceq \boldsymbol{S}$ we have

$$\sum_{i=1}^{K_1} \lambda_i^1 h\left(\boldsymbol{x} + \boldsymbol{n}_i^1\right) - \mu \sum_{j=1}^{K_2} \lambda_j^2 h\left(\boldsymbol{x} + \boldsymbol{n}_j^2\right)$$

$$\leq \sum_{i=1}^{K_1} \lambda_i^1 \frac{1}{2}\log\left|2\pi e\left(\boldsymbol{K}_X^* + \boldsymbol{N}_i^1\right)\right|$$

$$- \mu \sum_{j=1}^{K_2} \lambda_j^2 \frac{1}{2}\log\left|2\pi e\left(\boldsymbol{K}_X^* + \boldsymbol{N}_j^2\right)\right|. \quad (5)$$

Proof: In the following proof we use $\boldsymbol{n}^*$ to denote a random vector such that $\boldsymbol{n}^* \sim \mathcal{N}(0, \boldsymbol{N}^*)$. We wish to show that the distribution that optimizes $\sum_{i=1}^{K_1} \lambda_i^1 h\left(\boldsymbol{x} + \boldsymbol{n}_i^1\right) - \mu \sum_{j=1}^{K_2} \lambda_j^2 h(\boldsymbol{x} + \boldsymbol{n}_j^2)$ under a covariance constraint is $\boldsymbol{x} = \boldsymbol{x}^G \sim \mathcal{N}(0, \boldsymbol{K}_X^*)$. In order to show this, we consider the following function:

$$L(\gamma) = \sum_{i=1}^{K_1} \lambda_i^1 h\left(\boldsymbol{x}_\gamma + \boldsymbol{n}_i^1\right) - \mu \sum_{j=1}^{K_2} \lambda_j^2 h\left(\boldsymbol{x}_\gamma + \boldsymbol{n}_j^2\right)$$

where $\boldsymbol{x}_\gamma = \sqrt{1-\gamma}\boldsymbol{x} + \sqrt{\gamma}\boldsymbol{x}^G$. Note that $L(0)$ is equal to the left-hand side of (5) and $L(1)$ is equal to the right-hand side of (5). Therefore, it will be sufficient to show that for every distribution of $\boldsymbol{x}$, $L(\gamma)$ is a nondecreasing function in the segment $\gamma \in (0, 1)$. That is, we wish to show that $\frac{dL(\gamma)}{d\gamma} \geq 0$.

We calculate $\frac{dL(\gamma)}{d\gamma}$ in two steps. In the first step, we shall calculate the derivative of a generic term given by $h\left(\sqrt{1-\gamma}\boldsymbol{x} + \sqrt{\gamma}\boldsymbol{x}^G + \boldsymbol{n}\right)$ where $\boldsymbol{n} \sim \mathcal{N}(0, \boldsymbol{N})$ and $\boldsymbol{x}^G \sim \mathcal{N}(0, \boldsymbol{K}_x^*)$ are $t \times 1$ random variables (r.v.'s). In the second step, we shall add up all the appropriate derivatives as to correspond with the derivative of $L(\gamma)$. As $\boldsymbol{n}$ is a Gaussian r.v., we can rewrite an equivalent r.v. given by $\sqrt{1-\gamma}\boldsymbol{n}^a + \sqrt{\gamma}\boldsymbol{n}^b$ where $\boldsymbol{n}^a, \boldsymbol{n}^b \sim \mathcal{N}(0, \boldsymbol{N})$. Therefore

$$\frac{d}{d\gamma}h\left(\sqrt{1-\gamma}\boldsymbol{x} + \sqrt{\gamma}\boldsymbol{x}^G + \boldsymbol{n}\right)$$

$$= \frac{d}{d\gamma}h\left(\sqrt{1-\gamma}(\boldsymbol{x} + \boldsymbol{n}^a) + \sqrt{\gamma}(\boldsymbol{x}^G + \boldsymbol{n}^b)\right)$$

$$\stackrel{(a)}{=} \frac{d}{d\gamma}\left\{h\left((\boldsymbol{x} + \boldsymbol{n}^a) + \sqrt{\frac{\gamma}{1-\gamma}}(\boldsymbol{x}^G + \boldsymbol{n}^b)\right)\right.$$

$$\left. + \frac{t}{2}\log(1-\gamma)\right\}$$

$$\stackrel{(b)}{=} \frac{d}{d\gamma}\left\{h\left((\boldsymbol{K}^* + \boldsymbol{N})^{-\frac{1}{2}}(\boldsymbol{x} + \boldsymbol{n}^a) + \sqrt{\frac{\gamma}{1-\gamma}}\boldsymbol{z}\right)\right.$$

$$\left. + \frac{t}{2}\log(1-\gamma) + \frac{1}{2}\log|\boldsymbol{K}^* + \boldsymbol{N}|\right\}$$

$$\stackrel{(c)}{=} \frac{1}{2}\mathrm{tr}\left\{J\left((\boldsymbol{K}^* + \boldsymbol{N})^{-\frac{1}{2}}(\boldsymbol{x} + \boldsymbol{n}^a) + \sqrt{\frac{\gamma}{1-\gamma}}\boldsymbol{z}\right)\right\}$$

$$2(1-\gamma)\frac{dL(\gamma)}{d\gamma}$$

$$\geq \mathrm{tr}\left\{(\boldsymbol{K}_X^* + \boldsymbol{N}^*)\left(J(\boldsymbol{x}_\gamma + \boldsymbol{n}^*)(\boldsymbol{K}_X^* + \boldsymbol{N}^*) - \boldsymbol{I}\right)\cdot\left(\sum_{i=1}^{K_1}\lambda_i^1\left(\boldsymbol{K}_X^* + \boldsymbol{N}_i^1\right)^{-1} - \mu\sum_{j=1}^{K_2}\lambda_j^2\left(\boldsymbol{K}_X^* + \boldsymbol{N}_j^2\right)^{-1}\right)\right\}$$

$$\overset{(a)}{=} \mathrm{tr}\left\{(\boldsymbol{K}_X^* + \boldsymbol{N}^*)\left(J(\boldsymbol{x}_\gamma + \boldsymbol{n}^*)(\boldsymbol{K}_X^* + \boldsymbol{N}^*) - \boldsymbol{I}\right)\boldsymbol{O}_2\right\}$$

$$\overset{(b)}{\geq} \mathrm{tr}\left\{(\boldsymbol{K}_X^* + \boldsymbol{N}^*)\left((\boldsymbol{S} + \boldsymbol{N}^*)^{-1}(\boldsymbol{K}_X^* + \boldsymbol{N}^*) - \boldsymbol{I}\right)\boldsymbol{O}_2\right\}$$

$$= \mathrm{tr}\left\{(\boldsymbol{K}_X^* + \boldsymbol{N}^*)(\boldsymbol{S} + \boldsymbol{N}^*)^{-1}(\boldsymbol{K}_X^* - \boldsymbol{S})\boldsymbol{O}_2\right\} = 0 \qquad (8)$$

$$\cdot\frac{1}{(1-\gamma)^2} + \frac{t}{2(1-\gamma)}$$
$$\overset{(d)}{=} \frac{1}{2(1-\gamma)}\mathrm{tr}\left\{(\boldsymbol{K}^* + \boldsymbol{N})J(\boldsymbol{x}_\gamma + \boldsymbol{n}) + \boldsymbol{I}\right\} \qquad (6)$$

where $\boldsymbol{z} \sim \mathcal{N}(0, \boldsymbol{I})$ is a standard Gaussian vector. Steps (a) and (b) are due to the differential entropy scaling law. Step (c) is due to DeBruijn's identity ([8, Theorem 14]). Step (d) is due to the Fisher information scaling law and the equivalence of $\sqrt{1-\gamma}\boldsymbol{n}^a + \sqrt{\gamma}\boldsymbol{n}^b$ and $\boldsymbol{n}$.

We now use (6) to obtain (see also [12, Appendix C]):

$$\frac{dL(\gamma)}{d\gamma}$$
$$= \frac{1}{2(1-\gamma)}\mathrm{tr}\left\{\sum_{i=1}^{K_1}\lambda_i^1\left(\left(\boldsymbol{K}_X^* + \boldsymbol{N}_i^1\right)J\left(\boldsymbol{x}_\gamma + \boldsymbol{n}_i^1\right) - \boldsymbol{I}\right)\right.$$
$$\left. - \mu\sum_{j=1}^{K_2}\lambda_j^2\left(\left(\boldsymbol{K}_X^* + \boldsymbol{N}_j^2\right)J\left(\boldsymbol{x}_\gamma + \boldsymbol{n}_j^2\right) - \boldsymbol{I}\right)\right\}. \qquad (7)$$

For the following steps to be well defined, we shall require that the inequalities $\boldsymbol{N}_i^1 \preceq \boldsymbol{N}^* \preceq \boldsymbol{N}_j^2$ shall be strict for all i and j. If this is not the case, we can force it to be so by subtracting $\boldsymbol{N}_i^1 - \epsilon\boldsymbol{I}$ and adding $\boldsymbol{N}_j^2 + \epsilon\boldsymbol{I}$ where ϵ is chosen to be arbitrarily small (and where the dimension of the identity matrix depends on the dimensions of the respective matrices $\boldsymbol{N}_i^1$ and $\boldsymbol{N}_j^2$). For all $\epsilon > 0$, the above inequalities will remain strict and the rest of the proof will hold. By taking $\epsilon \to 0$ and relying on the contiguity of the differential entropy function with respect to variance of the added Gaussian noise, we prove the theorem also for the case where the inequality is not strict. Therefore, for the sake of brevity we shall assume that the above inequality is strict in the remainder of this proof.

We now lower-bound the summands of the first sum in the above equation. By assigning $\boldsymbol{A} = (\boldsymbol{K}_X^* + \boldsymbol{N}^*)^{-1}(\boldsymbol{K}_X^* + \boldsymbol{N}_i^1)$ in Lemma 1 we may write

$$J(\boldsymbol{x}_\gamma + \boldsymbol{n}_i^1)$$
$$\succeq \boldsymbol{A}^{-1}\left(J(\boldsymbol{x}_\gamma + \boldsymbol{n}^*) - (\boldsymbol{I} - \boldsymbol{A})J(\tilde{\boldsymbol{n}})(\boldsymbol{I} - \boldsymbol{A})^T\right)\boldsymbol{A}^{-T}$$
$$= \left(\boldsymbol{K}_X^* + \boldsymbol{N}_i^1\right)^{-1}\left((\boldsymbol{K}_X^* + \boldsymbol{N}^*)J(\boldsymbol{x}_\gamma + \boldsymbol{n}^*)(\boldsymbol{K}_X^* + \boldsymbol{N}^*)\right.$$
$$\left. - (\boldsymbol{N}^* - \boldsymbol{N}_i^1)\right)\left(\boldsymbol{K}_X^* + \boldsymbol{N}_i^1\right)^{-1}$$

where $\boldsymbol{n}^* \sim \mathcal{N}(0, \boldsymbol{N}^*)$, $\tilde{\boldsymbol{n}} \sim \mathcal{N}(0, \boldsymbol{N}^* - \boldsymbol{N}_i^1)$, and we used the fact that $\boldsymbol{K}_X^*$, $\boldsymbol{N}^*$, $\boldsymbol{N}_i^j$ are symmetric and the identity $J(\tilde{\boldsymbol{n}}) = (\boldsymbol{N}^* - \boldsymbol{N}_i^1)^{-1}$. Thus, the summands in the first sum in (7) may be lower-bounded by

$$\mathrm{tr}\left\{\left(\boldsymbol{K}_X^* + \boldsymbol{N}_i^1\right)J(\boldsymbol{x}_\gamma + \boldsymbol{n}_i^1) - \boldsymbol{I}\right\}$$
$$\geq \left\{\left[(\boldsymbol{K}_X^* + \boldsymbol{N}^*)J(\boldsymbol{x}_\gamma + \boldsymbol{n}^*)(\boldsymbol{K}_X^* + \boldsymbol{N}^*)\right.\right.$$
$$\left.\left. - (\boldsymbol{K}_X^* + \boldsymbol{N}^*)\right]\left(\boldsymbol{K}_X^* + \boldsymbol{N}_i^1\right)^{-1}\right\}.$$

Next, we use Lemma 1 to upper-bound the summands in the second sum in (7). Again, by assigning $\boldsymbol{A} = (\boldsymbol{K}_X^* + \boldsymbol{N}_j^{2-1})(\boldsymbol{K}_X^* + \boldsymbol{N}^*)$ we may write

$$J\left(\boldsymbol{x}_\gamma + \boldsymbol{n}_j^2\right)$$
$$\preceq \boldsymbol{A}J(\boldsymbol{x}_\gamma + \boldsymbol{n}^*)\boldsymbol{A}^T + (\boldsymbol{I} - \boldsymbol{A})J(\tilde{\boldsymbol{n}})(\boldsymbol{I} - \boldsymbol{A})^T$$
$$= \left(\boldsymbol{K}_X^* + \boldsymbol{N}_j^2\right)^{-1}\left((\boldsymbol{K}_X^* + \boldsymbol{N}^*)J(\boldsymbol{x}_\gamma + \boldsymbol{n}^*)(\boldsymbol{K}_X^* + \boldsymbol{N}^*)\right.$$
$$\left. + \left(\boldsymbol{N}_j^2 - \boldsymbol{N}^*\right)\right)\left(\boldsymbol{K}_X^* + \boldsymbol{N}_j^2\right)^{-1}$$

where this time $\tilde{\boldsymbol{n}} \sim \mathcal{N}(0, \boldsymbol{N}_j^2 - \boldsymbol{N}^*)$. Thus, the summands in the second sum may be upper-bounded by

$$\mathrm{tr}\left\{(\boldsymbol{K}_X^* + \boldsymbol{N}_j^2)J\left(\boldsymbol{x}_\gamma + \boldsymbol{n}_j^2\right) - \boldsymbol{I}\right\}$$
$$\leq \left\{\left[(\boldsymbol{K}_X^* + \boldsymbol{N}^*)J(\boldsymbol{x}_\gamma + \boldsymbol{n}^*)(\boldsymbol{K}_X^* + \boldsymbol{N}^*)\right.\right.$$
$$\left.\left. - (\boldsymbol{K}_X^* + \boldsymbol{N}^*)\right]\left(\boldsymbol{K}_X^* + \boldsymbol{N}_j^2\right)^{-1}\right\}.$$

Finally, by assigning the above bounds into (7) we obtain (8) at the top of the page, where (b) is due to the Cramer–Rao ([8, Theorem 20]) inequality and the covariance constraint on $\boldsymbol{K}_X^* \preceq \boldsymbol{S}$. (a) and the last equality are due to the conditions stated in (4). Therefore, $\frac{dL(\gamma)}{d\gamma}$ is nonnegative and the proof is complete. $\qquad\square$

The above lemma can be extended to the case of conditional entropies as stated in the following corollary.

Corollary 4: Let $\boldsymbol{K}_X^*$ be a positive semi-definite matrix such that $0 \preceq \boldsymbol{K}_X^* \preceq \boldsymbol{S}$ and such that (4) holds and let W be a random

variable independent of $\boldsymbol{n}_i^1$ and $\boldsymbol{n}_j^2$. Then, for every distribution of $\boldsymbol{x}$ such that $E[\boldsymbol{x}\boldsymbol{x}^T] \preceq \boldsymbol{S}$ we have

$$
\sum_{i=1}^{K_1} \lambda_i^1 h\left(\boldsymbol{x} + \boldsymbol{n}_i^1 | W\right) - \mu \sum_{j=1}^{K_2} \lambda_j^2 h\left(\boldsymbol{x} + \boldsymbol{n}_j^2 | W\right)
$$
$$
\leq \sum_{i=1}^{K_1} \lambda_i^1 \frac{1}{2} \log \left| 2\pi e \left(\boldsymbol{K}_X^* + \boldsymbol{N}_i^1\right)\right|
$$
$$
- \mu \sum_{j=1}^{K_2} \lambda_j^2 \frac{1}{2} \log \left| 2\pi e \left(\boldsymbol{K}_X^* + \boldsymbol{N}_j^2\right)\right|. \tag{9}
$$

Proof: The proof here is similar to that of Lemma 2. We first write the left-hand side of (9) as

$$
E\left[\sum_{i=1}^{K_1} \lambda_i^1 h\left(\boldsymbol{x} + \boldsymbol{n}_i^1 | W = w\right) - \mu \sum_{j=1}^{K_2} \lambda_j^2 h\left(\boldsymbol{x} + \boldsymbol{n}_j^2 | W = w\right)\right]
$$

where the expectation is taken with respect to W. We then rewrite (7) as

$$
2(1-\gamma) \frac{dL(\gamma)}{d\gamma}
$$
$$
= E\left[\operatorname{tr}\left\{\sum_{i=1}^{K_1} \lambda_i^1 \left(\left(\boldsymbol{K}_X^* + \boldsymbol{N}_i^1\right) J\left(\boldsymbol{x}_\gamma + \boldsymbol{n}_i^1 | W = w\right) - \boldsymbol{I}\right)\right.\right.
$$
$$
\left.\left. - \mu \sum_{j=1}^{K_2} \lambda_j^2 \left(\left(\boldsymbol{K}_X^* + \boldsymbol{N}_j^2\right) J\left(\boldsymbol{x}_\gamma + \boldsymbol{n}_j^2 | W = w\right) - \boldsymbol{I}\right)\right\}\right]
$$

where $J(\boldsymbol{x} | W = w)$ is the fisher information w.r.t. the distribution $f_{\boldsymbol{x}|W}(\boldsymbol{x}|w)$.

Using the same arguments as in Lemma 2, we may write the average form of the second equality in (8) as shown in the equation at the bottom of the page.

We now note that

$$
E\left[J\left(\boldsymbol{x}_\gamma + \boldsymbol{n}^* | W = w\right)\right]
$$
$$
\succeq E\left[E\left[\left(\boldsymbol{x}_\gamma + \boldsymbol{n}^*\right)\left(\boldsymbol{x}_\gamma + \boldsymbol{n}^*\right)^T | W = w\right]^{-1}\right]
$$
$$
\succeq E\left[\left(\boldsymbol{x}_\gamma + \boldsymbol{n}^*\right)\left(\boldsymbol{x}_\gamma + \boldsymbol{n}^*\right)^T\right]^{-1}
$$

where the first inequality is due to the Cramer–Rao inequality and the last inequality is due to the concavity of the matrix inverse function over positive semi-definite matrices (i.e., $\sum \lambda_i \boldsymbol{X}_i^{-1} \succeq \left(\sum \lambda_i \boldsymbol{X}_i\right)^{-1}$ where $\sum \lambda_i = 1$ and $\lambda_i \geq 0$; see [10, pp. 554–555]).

Again, using the same arguments as in the proof of Lemma 2, we obtain that $\frac{dL(\gamma)}{d\gamma} \geq 0$ and thus prove the corollary. $\square$

IV. PROOF OF THEOREM 1

We can now turn to prove Theorem 1. The rates stated in Theorem 1 can be obtained using standard arguments of Gaussian coding and successive decoding as is done in the degraded Gaussian broadcast channel [5]. In the following, we give just an outline of the direct part of the proof. The bulk of the proof deals with the converse. The interested reader is referred to [6, Sec. 14.6, pp. 418–428] and references therein, which discuss in detail the methods considered in the following proof outline of the direct part.

To transmit over this channel we begin by constructing two codebooks. Assuming each codeword contains n symbols (each symbol is now a vector as we are considering the multiple-input multiple-output case), the codebooks for user 1 and 2 will contain $2^{n(R_1-\epsilon)}$ and $2^{n(R_2-\epsilon)}$ codewords, respectively. Each vector in each codeword is drawn independently using a random vector generator with a Gaussian distribution given by $\mathcal{N}(0, \boldsymbol{Q}_1)$ for the first user and $\mathcal{N}(0, \boldsymbol{S} - \boldsymbol{Q}_1 - \delta\boldsymbol{I})$ for the second user. To transmit a message (W_1, W_2), the appropriate codewords are chosen from the codebook and their sum is transmitted over the channel. As $n \longrightarrow \infty$, we meet the covariance constraint on the input with probability arbitrarily close to 1 for arbitrarily small δ.

The further away user receives the combination of the two codewords. The interference due to the message sent to user 1 acts as an additional Gaussian noise. Therefore, from the point of view of user 2, it is a standard Gaussian compound channel with additive Gaussian noise given by $\boldsymbol{N}_j^2 + \boldsymbol{Q}_1$. Therefore, we can achieve a rate

$$
R_2 \leq \min_j \frac{1}{2} \log \frac{\left|\boldsymbol{S} + \boldsymbol{N}_j^2 - \delta\boldsymbol{I}\right|}{\left|\boldsymbol{Q}_1 + \boldsymbol{N}_j^2\right|} - \epsilon
$$

with arbitrarily small probability of decoding error for n sufficiently large and arbitrarily small ϵ. As the first user suffers from a smaller additive Gaussian noise (i.e., a degraded compound channel), it can always decode the messages transmitted to the second user and remove their effect. Therefore, the second user only suffers from the channel noise when attempting to decode its own messages (i.e., additive Gaussian noise given by $\boldsymbol{n}_i^1$). The rate achievable in this compound channel, following the removal of the interference from the second user, is given by

$$
R_1 \leq \min_i \frac{1}{2} \log \frac{\left|\boldsymbol{Q}_1 + \boldsymbol{N}_i^1\right|}{\left|\boldsymbol{N}_i^1\right|} - \epsilon
$$

with arbitrarily small probability of decoding error for n sufficiently large and arbitrarily small ϵ.

As ϵ and δ can be made arbitrarily small, we obtain the rate region defined in Theorem 1. Therefore, we only need to prove that all rates outside the Gaussian coding region are not achievable.

We now turn to prove the converse part. Assume that (R_1^o, R_2^o) is an achievable rate pair that lies outside the rate

$$
2(1-\gamma) \frac{dL(\gamma)}{d\gamma} \geq \operatorname{tr}\left\{\left(\boldsymbol{K}_X^* + \boldsymbol{N}^*\right) \cdot \left(E\left[J\left(\boldsymbol{x}_\gamma + \boldsymbol{n}^* | W = w\right)\right]\left(\boldsymbol{K}_X^* + \boldsymbol{N}^*\right) - \boldsymbol{I}\right)\boldsymbol{O}_2\right\}.
$$

region, $\mathcal{C}_{\mathrm{DC}}(\boldsymbol{S})$. It is well known that for the single-user compound channel [7, p. 173] we can bound R_1^o by

$$R_1^o \leq \max_{P_x} \min_i I\left(\boldsymbol{x}; \boldsymbol{y}_i^1\right) = \min_i \frac{1}{2}\log\frac{\left|\boldsymbol{S}+\boldsymbol{N}_i^1\right|}{\left|\boldsymbol{N}_i^1\right|} = R_1^{\max}.$$

Therefore, if $R_2^o > 0$,[1] we can find a point (R_1^o, R_2^*) on the boundary of $\mathcal{C}_{\mathrm{DC}}(\boldsymbol{S})$ such that $R_2^o = R_2^* + b$ for some $b > 0$. Furthermore, R_2^* is the solution of the following program:

$$R_2^* = \begin{array}{c} \max \\ 0 \preceq \boldsymbol{Q}_1 \preceq \boldsymbol{S} \\ \frac{1}{2}\log\frac{|\boldsymbol{Q}_1+\boldsymbol{N}_1^1|}{|\boldsymbol{N}_1^1|} \geq R_1^o \\ \vdots \\ \frac{1}{2}\log\frac{|\boldsymbol{Q}_1+\boldsymbol{N}_{K_1}^1|}{|\boldsymbol{N}_{K_1}^1|} \geq R_1^o \end{array} \qquad \psi(\boldsymbol{Q}_1) \qquad (10)$$

where

$$\psi(\boldsymbol{Q}_1) = \min_{j=1,\ldots,K_2} \frac{1}{2}\log\frac{\left|\boldsymbol{S}+\boldsymbol{N}_j^2\right|}{\left|\boldsymbol{Q}_1+\boldsymbol{N}_j^2\right|}.$$

We can use the following lemma to obtain the set of conditions which must hold at the solution points.

Lemma 3: Let $\boldsymbol{Q}_1^*$ be the optimizing solution of the optimization problem in (10). The following conditions must hold:

$$\sum_{i=1}^{K_1} \lambda_i^1 \left(\boldsymbol{Q}_1^* + \boldsymbol{N}_i^1\right)^{-1} + \boldsymbol{O}_1 = \mu \sum_{j=1}^{K_2} \lambda_j^2 \left(\boldsymbol{Q}_1^* + \boldsymbol{N}_j^2\right)^{-1} + \boldsymbol{O}_2$$

$$(11)$$

where
1. $\boldsymbol{O}_1$ and $\boldsymbol{O}_2$ are positive semi-definite matrices such that $\boldsymbol{Q}_1^* \cdot \boldsymbol{O}_1 = 0$ and $(\boldsymbol{S} - \boldsymbol{Q}_1^*) \cdot \boldsymbol{O}_2 = 0$.
2. $\sum_{i=1}^{K_1} \lambda_i^1 = 1$ and for all $i = 1\ldots K_1$, $\lambda_i^1 \geq 0$ with equality if $\frac{1}{2}\log\frac{|\boldsymbol{Q}_1^*+\boldsymbol{N}_i^1|}{|\boldsymbol{N}_i^1|} > R_1^o$.
3. $\sum_{j=1}^{K_2} \lambda_j^2 = 1$ and for all $j = 1\ldots K_2$, $\lambda_j^2 \geq 0$ with equality if $\frac{1}{2}\log\frac{|\boldsymbol{S}+\boldsymbol{N}_j^2|}{|\boldsymbol{Q}_1^*+\boldsymbol{N}_k|} > R_2^*$.
4. $\mu \geq 1$.

Proof: We can rewrite the above optimization problem (10) as follows:

$$R_2^o > R_2^* = \begin{array}{c} \max \\ 0 \preceq \boldsymbol{Q}_1 \preceq \boldsymbol{S} \\ \frac{1}{2}\log\frac{|\boldsymbol{Q}_1+\boldsymbol{N}_i^1|}{|\boldsymbol{N}_i^1|} \geq R_1^o \quad i = 1\ldots K_1 \\ \frac{1}{2}\log\frac{|\boldsymbol{S}+\boldsymbol{N}_j^2|}{|\boldsymbol{Q}_1+\boldsymbol{N}_j^2|} \geq a \quad j = 1\ldots K_2 \end{array} \quad a \qquad (12)$$

[1] The case of $R_2^o = 0$ is trivial, as then $R_1 = R_1^{\max}$ is achievable.

The above maximization problem may be rewritten as a minimization problem as follows:

$$\min \quad -a$$
$$\text{s.t.}$$
(a) $\quad -\boldsymbol{Q}_1 \preceq 0$
(b) $\quad \boldsymbol{Q}_1 - \boldsymbol{S} \preceq 0$
(c) $\quad R_1^o - \frac{1}{2}\log\frac{\left|\boldsymbol{Q}_1+\boldsymbol{N}_i^1\right|}{\left|\boldsymbol{N}_i^1\right|} \leq 0 \quad i = 1\ldots K_1$
(d) $\quad a - \frac{1}{2}\log\frac{\left|\boldsymbol{S}+\boldsymbol{N}_j^2\right|}{\left|\boldsymbol{Q}_1+\boldsymbol{N}_j^2\right|} \leq 0 \quad j = 1\ldots K_2.$

The above optimization problem contains both real inequalities and semi-definite inequalities. Furthermore, the optimization problem has one semi-definite variable, $\boldsymbol{Q}_1$, which is constrained by (a) and (b) and one scalar variable which is not directly constrained. This is in fact an optimization problem with generalized inequalities in the form of semi-definite inequalities (see [3, Subsec. 5.9.2, p. 267]). Therefore, the Karush–Kuhn–Tucker (KKT) conditions state that the derivative with respect to both variables of the Lagrangian

$$\mathcal{L} = -a + \mathrm{tr}\{(-\boldsymbol{Q}_1)\boldsymbol{O}_1\} + \mathrm{tr}\{(\boldsymbol{Q}_1-\boldsymbol{S})\boldsymbol{O}_2\}$$
$$+ \sum_{i=1}^{K_1} \lambda_i^1 \left(R_1^o - \frac{1}{2}\log\frac{\left|\boldsymbol{Q}_1+\boldsymbol{N}_i^1\right|}{\left|\boldsymbol{N}_i^1\right|}\right)$$
$$+ \sum_{j=1}^{K_2} \lambda_j^2 \left(a - \frac{1}{2}\log\frac{\left|\boldsymbol{S}+\boldsymbol{N}_j^2\right|}{\left|\boldsymbol{Q}_1+\boldsymbol{N}_j^2\right|}\right)$$

must vanish at the optimal solution (see also [16, Sec. 3 and Appendix D]). Furthermore, $\boldsymbol{O}_1$ and $\boldsymbol{O}_2$ are positive semi-definite matrices such that $\boldsymbol{Q}_1^* \cdot \boldsymbol{O}_1 = 0$ and $(\boldsymbol{S} - \boldsymbol{Q}_1^*) \cdot \boldsymbol{O}_2 = 0$ and $\lambda_i^1, \lambda_j^2 \geq 0 \ \forall\ i = 1\ldots K_1, j = 1\ldots K_2$ with equality if $\frac{1}{2}\log\frac{|\boldsymbol{Q}_1^*+\boldsymbol{N}_i^1|}{|\boldsymbol{N}_i^1|} > R_1^o$ or $\frac{1}{2}\log\frac{|\boldsymbol{S}+\boldsymbol{N}_j^2|}{|\boldsymbol{Q}_1^*+\boldsymbol{N}_j^2|} > R_2^*$, respectively.

Therefore, if $\boldsymbol{Q}_1^*$ and $a = R_2^*$ solve the above optimization problem, the KKT conditions of this program can be written as follows:[2] see (13) at the bottom of the page, and (14) at the bottom of the following page.

By writing (13) explicitly we obtain

$$\sum_{i=1}^{K_1} \lambda_i^1 \frac{1}{2}(\boldsymbol{Q}_1^*+\boldsymbol{N}_i^1)^{-1} - \sum_{j=1}^{K_2} \lambda_j^2 \frac{1}{2}(\boldsymbol{Q}_1^*+\boldsymbol{N}_j^2)^{-1} + \frac{1}{2}\boldsymbol{O}_1 - \frac{1}{2}\boldsymbol{O}_2$$
$$= 0_{t\times t}.$$

[2] As the program in (12) is not convex, a set of constraint qualifications (CQs) should be checked to make sure that the KKT conditions indeed hold. The CQs stated in [16, Appendix D] hold in a trivial manner for this program. See also [1, Ch. 4]

$$\nabla_{\boldsymbol{Q}_1}\left(a + \sum_{i=1}^{K_1} \lambda_i^1\left(\frac{1}{2}\log\frac{\left|\boldsymbol{Q}_1+\boldsymbol{N}_i^1\right|}{\left|\boldsymbol{N}_i^1\right|} - R_1^o\right) + \sum_{j=1}^{K_2} \lambda_j^2\left(\frac{1}{2}\log\frac{\left|\boldsymbol{S}+\boldsymbol{N}_j^2\right|}{\left|\boldsymbol{Q}_1+\boldsymbol{N}_j^2\right|} - a\right)\right)\Bigg|_{\boldsymbol{Q}_1=\boldsymbol{Q}_1^*, a=R_2^*} + \frac{1}{2}\boldsymbol{O}_1 - \frac{1}{2}\boldsymbol{O}_2 = 0_{t\times t} \qquad (13)$$

Furthermore, if $\boldsymbol{Q}_1^* \neq \boldsymbol{S}(R_2^* > 0)$, we can find a vector $\boldsymbol{v} \neq 0$ such that $\boldsymbol{v}^T \boldsymbol{O}_2 \boldsymbol{v} = 0$. By multiplying both sides of the equation above by $\boldsymbol{v}^T$ on the left and $\boldsymbol{v}$ on the right and by noting that $\boldsymbol{N}_j^2 \succeq \boldsymbol{N}^* \succeq \boldsymbol{N}_i^1$ for all i and j, we obtain that $\sum_{i=1}^{K_1} \lambda_i^1 \leq \sum_{j=1}^{K_2} \lambda_j^2$.

If $\boldsymbol{Q}_1^* = \boldsymbol{S}$ and there is no vector $\boldsymbol{v} \neq 0$ such that $\boldsymbol{v}^T \boldsymbol{O}_2 \boldsymbol{v} = 0$ (i.e., $\boldsymbol{O}_2 \succ 0_{t \times t}$), then $\boldsymbol{O}_2$ and λ_j^2 can be reassigned such that the above equation holds and such that $\sum_{i=1}^{K_1} \lambda_i^1 \leq \sum_{j=1}^{K_2} \lambda_j^2$. More specifically, when $\boldsymbol{O}_2 \succ 0_{t \times t}$ in the above equation we can multiply all λ_j^2 by the same factor, larger than 1 (regardless of any previous equations), and at the same time we modify $\boldsymbol{O}_2$ such that the above equation still holds. The scaling factor is chosen such that the resulting $\boldsymbol{O}_2$ is nonnegative but has at least one zero eigenvalue. With the new choice of λ_j^2, we obtain $\sum_{i=1}^{K_1} \lambda_i^1 \leq \sum_{j=1}^{K_2} \lambda_j^2$ as before.

Thus, we need only to scale λ_i^1, λ_j^2, $\boldsymbol{O}_1$, and $\boldsymbol{O}_2$ (with the same scale factor for all) such that $\sum_{i=1}^{K_1} \lambda_i^1 = 1$. Then, by substituting λ_j^2 by their normalized version $\frac{\lambda_j^2}{\mu}$ and multiplying the sum by $\mu = \sum_{j=1}^{K_2} \lambda_j^2 \geq 1$ we complete the proof.

We shall now show that any point which observes the above KKT conditions must be a point on the boundary of the capacity region. We shall initially assume that $\boldsymbol{O}_1 = 0$ and prove by contradiction that the point (R_1^o, R_2^o) cannot be achievable. In the last part of the section, we shall extend this proof to the case where $\boldsymbol{O}_1 \succeq 0$.

In addition, we shall assume, without loss of generality, that λ_i^1 and λ_j^2 are strictly positive for all i and j. If that is not the case, we may consider the channel which only contains those outputs which correspond to those λ's which are strictly positive. Clearly, the KKT conditions in (11) hold for this new channel for the same choice of $\boldsymbol{Q}_1^*$. Further, note that the capacity region of this new channel contains that of the original channel. Showing that R_1^o and R_2^o are outside the capacity region of this new channel will also ensure that they are outside the region of the original channel.

Next, we make use of a single-letter result on the capacity region of a discrete memoryless and degraded compound BC. In Appendix III, we give an alternative proof based on Lemma 2, Corollary 4, and multiletter Fano bounds instead of the following lemma.

Lemma 4: Consider a memoryless compound broadcast channel with input X and outputs Y_i^1, $i = 1 \ldots K_1$ for the first user, Y_j^2, $j = 1 \ldots K_2$ for the second user, and an auxiliary output Y^*. All outputs are defined by their conditional probability functions: $P_{Y_i^1|X}$, $P_{Y_j^2|X}$, and $P_{Y^*|X}$. Furthermore, assume that these outputs are stochastically degraded such that there exists some distribution $P_{Y_1^1,\ldots,Y_{K_1}^1,Y_1^2,\ldots,Y_{K_2}^2,Y^*|X}$ with marginal distributions $P_{Y_i^1|X}$, $P_{Y_j^2|X}$, and $P_{Y^*|X}$ such that

$X \to Y_i^1 \to Y^* \to Y_j^2$ form a Markov chain for every choice of i and j. The capacity region of this channel is given by

$$R_1 \leq \min_{i=1,\ldots,K_1} I(X; Y_i^1 | V)$$
$$R_2 \leq \min_{j=1,\ldots,K_2} I(V; Y_j^2) \qquad (15)$$

for some auxiliary random variable V such that $V \to X \to Y_i^1 \to Y^* \to Y_j^2$ form a Markov chain.

Proof: The proof is deferred to Appendix II. $\square$

We shall show that (R_1^o, R_2^o) are not achievable by showing that any rate pair (R_1, R_2) that satisfy (15) must also obey $R_1 + \mu R_2 < R_1^o + \mu R_2^o$ where μ is defined in (11). By (15), there exists some distribution $P_{V,X}$ such that

$$
\begin{aligned}
&R_1 + \mu R_2 \\
&\leq \min_{i=1\ldots K_1} I(X; Y_i^1 | V) + \mu \min_{j=1\ldots K_2} I(V; Y_j^2) \\
&\overset{(a)}{\leq} \sum_{i=1}^{K_1} \lambda_i^1 I(X; Y_i^1 | V) + \mu \sum_{j=1}^{K_2} \lambda_j^2 I(V; Y_j^2) \\
&= \mu \sum_{j=1}^{K_2} \lambda_j^2 h(Y_j^2) - \sum_{i=1}^{K_1} \lambda_i^1 h(Y_1^i | X, V) \\
&\quad + \sum_{i=1}^{K_1} \lambda_i^1 h(Y_1^i | V) - \mu \sum_{j=1}^{K_2} \lambda_j^2 h(Y_j^2 | V) \\
&\overset{(b)}{\leq} \mu \sum_{j=1}^{K_2} \lambda_j^2 \frac{1}{2} \log \left| 2\pi e (\boldsymbol{S} + \boldsymbol{N}_j^2) \right| - \sum_{i=1}^{K_1} \lambda_i^1 \frac{1}{2} \log \left| 2\pi e \boldsymbol{N}_i^1 \right| \\
&\quad + \sum_{i=1}^{K_1} \lambda_i^1 h(Y_i^1 | V) - \mu \sum_{j=1}^{K_2} \lambda_j^2 h(Y_j^2 | V) \qquad (16)
\end{aligned}
$$

where λ_i^1 and λ_j^2 are defined in (11). Inequality (a) follows from the non-negativity of λ_i^1 and λ_j^2 and the fact that $\sum \lambda_i^2 = \sum \lambda_j^2 = 1$. Inequality (b) is due to the optimality of the Gaussian distribution under a covariance constraint and the Markov chain $V \to X \to Y_i^1$.

We can now use Lemma 2, Corollary 4, and (11) (under the assumption that $\boldsymbol{O}_1 = 0$) to upper-bound the last two summands in the last inequality of (16) and write

$$
\begin{aligned}
&R_1 + \mu R_2 \\
&\leq \mu \sum_{j=1}^{K_2} \lambda_j^2 \frac{1}{2} \log \left| 2\pi e (\boldsymbol{S} + \boldsymbol{N}_j^2) \right| - \sum_{i=1}^{K_1} \lambda_i^1 \frac{1}{2} \log \left| 2\pi e \boldsymbol{N}_i^1 \right| \\
&\quad + \sum_{i=1}^{K_1} \lambda_i^1 \frac{1}{2} \log \left| 2\pi e (\boldsymbol{Q}_1^* + \boldsymbol{N}_i^1) \right| \\
&\quad - \mu \sum_{j=1}^{K_2} \lambda_j^2 \frac{1}{2} \log \left| 2\pi e \left(\boldsymbol{Q}_1^* + \boldsymbol{N}_j^2 \right) \right|
\end{aligned}
$$

$$
\frac{d}{da} \left(a + \sum_{i=1}^{K_1} \lambda_i^1 \left(\frac{1}{2} \log \frac{\left| \boldsymbol{Q}_1 + \boldsymbol{N}_i^1 \right|}{\left| \boldsymbol{N}_i^1 \right|} - R_1^o \right) + \sum_{j=1}^{K_2} \lambda_j^2 \left(\frac{1}{2} \log \frac{\left| \boldsymbol{S} + \boldsymbol{N}_j^2 \right|}{\left| \boldsymbol{Q}_1 + \boldsymbol{N}_j^2 \right|} - a \right) \right) \Bigg|_{\boldsymbol{Q}_1 = \boldsymbol{Q}_1^*, a = R_2^*} = 0. \qquad (14)
$$

$$= \sum_{i=1}^{K_1} \lambda_i^1 \frac{1}{2} \log \left(\frac{|Q_1^* + N_i^1|}{|N_i^1|} \right)$$

$$+ \mu \sum_{j=1}^{K_2} \lambda_j^2 \frac{1}{2} \log \left(\frac{|S + N_j^2|}{|Q_1^* + N_j^2|} \right)$$

$$\overset{(a)}{\leq} R_1^o + \mu R_2^*$$

$$\overset{(b)}{<} R_1^o + \mu R_2^o$$

where (a) is due to the definition of R_2^* and (b) is due to our original assumption that $R_2^o = R_2^* + b$ for some $b > 0$.

To complete the proof and extend it to the case where $O_1 \succeq 0$, we borrow the idea of enhancement from [16]. Instead of investigating the original channel, we consider the enhanced channel where n_i^1 are replaced by $\tilde{n}_i^1$ such that $\tilde{N}_i^1 \preceq N_i^1$ and are defined such that $(Q_1^* + N_i^1)^{-1} + O_1 = (Q_1^* + \tilde{N}_i^1)^{-1}$. In [16] it is shown that with $Q_1 = Q_1^*$, the exact same rates (R_1^o, R_2^*) are obtained in the enhanced channel as in the original channel. In addition, the KKT conditions in (11) hold for the enhanced channel as well. Therefore, we can follow all the previous steps to show that (R_1^o, R_2^o) is not achievable in the enhanced channel. However, as $\tilde{N}_i^1 \preceq N_i^1$, it is clear that the capacity region of the enhanced channel contains that of the original channel. Therefore, (R_1^o, R_2^o) must lie outside the capacity region of the original channel as well. □

V. PROOF OF THE CAPACITY REGION OF THE DEGRADED MIMO COMPOUND BROADCAST CHANNEL WITH ARBITRARY LINEAR CHANNEL MATRICES

We can now extend the proof of Theorem 1 to the multiple-antenna case and prove Theorem 3.

Proof: In the following, we borrow the main ideas from [16, Sec. 5]. Due to the degradedness order, it is quite clear that $\mathcal{C}_{\mathrm{ADC}}$ is indeed achievable using Gaussian coding and successive decoding. We therefore concentrate on the proof of the converse. The proof relies on Theorem 1. The case where all linear channel matrices are square and nonsingular can be easily transformed into the channel presented in (1) and thus, Theorem 1 completes the proof. Our goal will be to approximate a channel with singular linear channel matrices, by a channel with invertible linear channel matrices that maintains the degradedness order.

We may assume without loss of generality that all linear channel matrices H_i^1, H_j^2, and H^* are $t \times t$ square matrices.[3] If that is not the case, we can use singular value decomposition (SVD) and follow the same steps that were carried out in [16, Sec. 5] to show that there is an equivalent channel which does have $t \times t$ square linear channel matrices. That is, we may find a new channel with square linear channel matrices which are derived from the original channel matrices via a matrix multiplication. The new channel is equivalent to the original one in the sense that for any receiver working on one channel with any given codebook, we may find a modified receiver which will work equally well on the other channel for the same

---

[3] H_i^1, H_j^2, and H^* may be singular.

codebook. The linear channel matrices of the new channel still observe the degradedness order. For more details, the reader is referred to [16, Sec. 5].

We now define a new degraded compound channel which has invertible linear channel matrices. Furthermore, each of the received vectors in our original channel will be degraded w.r.t. a corresponding receive vector in the new channel. We first use SVD to rewrite the linear channel matrices as follows:

$$H_i^1 = U_i^1 \Lambda_i^1 V_i^1, \quad H_j^2 = U_j^2 \Lambda_j^2 V_j^2$$

and

$$H^* = U^* \Lambda^* V^*$$

where U and V are $t \times t$ unitary matrices and Λ is diagonal. We define the linear channel matrices of the new channel as follows:

$$\tilde{H}_i^1 = U_i^1 (\Lambda_i^1 + \alpha I) V_i^1, \quad \tilde{H}_j^2 = U_j^2 (\Lambda_j^2 + \beta I) V_j^2$$

and

$$\tilde{H}^* = U^* (\Lambda^* + \gamma I) V^*$$

where $\alpha \geq \gamma \geq \beta > 0$.

Note that we can write $H_i^1 = D_i^{\tilde{1}1} \tilde{H}_i^1$ where

$$D_i^{\tilde{1}1} = U_i^1 \Lambda_i^1 (\Lambda_i^1 + \alpha I)^{-1} U_i^{1T}.$$

As $D_i^{\tilde{1}1} D_i^{\tilde{1}1T} \prec I$, we conclude that H_i^1 is degraded w.r.t. $\tilde{H}_i^1$. Similarly, H_j^2 ($D_j^{\tilde{2}2} = U_j^2 \Lambda_j^2 (\Lambda_j^2 + \beta I)^{-1} U_j^{2T}$) and H^* ($D^{\tilde{*}*} = U^* \Lambda^* (\Lambda^* + \gamma I)^{-1} U^{*T}$) are degraded w.r.t. $\tilde{H}_j^2$ and $\tilde{H}^*$. Therefore, the capacity region of the new channel contains that of the original channel.

Next, we write

$$\tilde{H}^* = U^* \Lambda^* V^* + \gamma U^* V^* = D_i^{\tilde{1}*} \tilde{H}_i^1$$

where

$$D_i^{\tilde{1}*} = \left(D_i^{1*} D_i^{\tilde{1}1} + \gamma \left(U^* V^* V_i^{1T} (\Lambda_i^1 + \alpha I)^{-1} U_i^{1T} \right) \right)$$

and where D_i^{1*} is the matrix for which $H^* = D_i^{1*} H_i^1$ and $D_i^{\tilde{1}1}$ was defined above.

As H^* is degraded w.r.t. H_i^1, the eigenvalues of D_i^{1*} are less than or equal to 1. Furthermore, as the eigenvalues of $D_i^{\tilde{1}1}$ are strictly smaller than 1, we conclude that the eigenvalues of $D_i^{1*} D_i^{\tilde{1}1}$ are also strictly smaller than 1. Therefore, for every choice of $\alpha > 0$, we can set $\gamma > 0$ to be small enough such that $D_i^{\tilde{1}*} D_i^{\tilde{1}*T} \preceq I$. Therefore, we conclude that γ can be set small enough such that $\tilde{H}^*$ is degraded w.r.t. $\tilde{H}_i^1$ for all i. Similarly, for every choice of γ, β can be set small enough such that $\tilde{H}_j^2$ is degraded w.r.t. $\tilde{H}^*$ for all j. Thus, we can construct a new channel which preserves the degradedness order but for which linear channel matrices are invertible.

As $\tilde{H}_i^1$, $\tilde{H}^*$, and $\tilde{H}_j^2$ are all invertible, we can apply Theorem 1 in order to show that $\mathcal{C}_{\mathrm{ADC}}(\tilde{H}_i^1, \tilde{H}_j^2, S) \supseteq \mathcal{C}_{\mathrm{ADC}}(H_i^1, H_j^2, S)$ is the capacity region of this new channel. Finally, we note that as $\alpha, \beta, \gamma \to 0$

$$\mathcal{C}_{\mathrm{ADC}}(\tilde{H}_i^1, \tilde{H}_j^2, S) \to \mathcal{C}_{\mathrm{ADC}}(H_i^1, H_j^2, S)$$

and thus, complete the proof. □

VI. An Illustrative Example

In this section, we give an example that illustrates the capacity region of the compound broadcast channel. We consider a multiple-antenna channel of the type discussed in Section V with two transmit antennas. In the following example, there are two users, the stronger user has only one realization with two receive antennas while the weaker user has two possible realizations, each with one receive antenna, as defined by the following linear channel matrices:

$$\boldsymbol{H}_1^1 = \begin{pmatrix} 2 & \\ & 3 \end{pmatrix}$$
$$\boldsymbol{h}_1^2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \end{pmatrix}, \quad \boldsymbol{h}_2^2 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & 2 \end{pmatrix}. \tag{17}$$

We shall assume that there is a total power constraint $E\boldsymbol{x}^T\boldsymbol{x} \leq P = 10$. As in (3), the additive noise in our example is also normalized to unity at each receive antenna.

We can associate this channel with a realistic scenario where there is one user who is close to the transmitter and enjoys full diversity, while there are two far away users who suffer from a keyhole effect and who receive a common message. A keyhole effect [2] is a situation where all multipath components from the transmitter merge before they split up into received multipath components. Hence, the far away users observe a channel with a reduced degree of freedom. One can easily verify that $\boldsymbol{h}_1^2$ and $\boldsymbol{h}_2^2$ are degraded with respect to $\boldsymbol{H}_1^1$ as defined in Section V. Furthermore, as there is only one realization for user 1, we may associate $\boldsymbol{H}^*$ with $\boldsymbol{H}_1^1$. Therefore, we can use Theorem 3 to calculate the capacity region of this channel.

By applying Theorem 3 to our scenario with a total power constraint (i.e., $\text{tr}\{\boldsymbol{Q}\} \leq P$) we can now find the capacity region of this problem. As we now have a total power constraint and not a covariance matrix constraint, the covariance matrix allocated for transmission to the second user is no longer trivially reduced from that of the first user (that is, we no longer have $\boldsymbol{Q}_2 = \boldsymbol{S} - \boldsymbol{Q}_1$). Therefore, the second user covariance matrix is now an additional optimization parameter. By solving the boundary point of R_1 for every given R_2 and by acknowledging that $\boldsymbol{h}_j^2$ are vectors and not matrices, we can show that every point on the boundary of the capacity region is a solution of the following optimization problem:

$$R_1 = \max_{\boldsymbol{Q}_1, \boldsymbol{Q}_2} \frac{1}{2}\log\left|\boldsymbol{H}_1\boldsymbol{Q}_1\boldsymbol{H}_1^T + \boldsymbol{I}\right|$$
$$\text{s.t.} \quad \boldsymbol{Q}_1 \succeq 0,$$
$$\boldsymbol{Q}_2 \succeq 0,$$
$$\text{tr}\{\boldsymbol{Q}_1 + \boldsymbol{Q}_2\} \leq P,$$
$$\boldsymbol{h}_2^1(\boldsymbol{Q}_1 + \boldsymbol{Q}_2)\boldsymbol{h}_2^{1^T} + 1 \geq e^{2R_2} \cdot \left(\boldsymbol{h}_2^1\boldsymbol{Q}_1\boldsymbol{h}_2^{1^T} + 1\right)$$
$$\boldsymbol{h}_2^2(\boldsymbol{Q}_1 + \boldsymbol{Q}_2)\boldsymbol{h}_2^{2^T} + 1 \geq e^{2R_2} \cdot \left(\boldsymbol{h}_2^2\boldsymbol{Q}_1\boldsymbol{h}_2^{2^T} + 1\right).$$

As the object of the above optimization problem is convex (the $\log|\cdot|$ function is convex w.r.t. semi-definite matrices) and as the constraints are linear, the above semi-definite optimization program is a convex one and can be numerically solved using standard optimization tools (see [3, Ch. 11, pp. 561–622]).
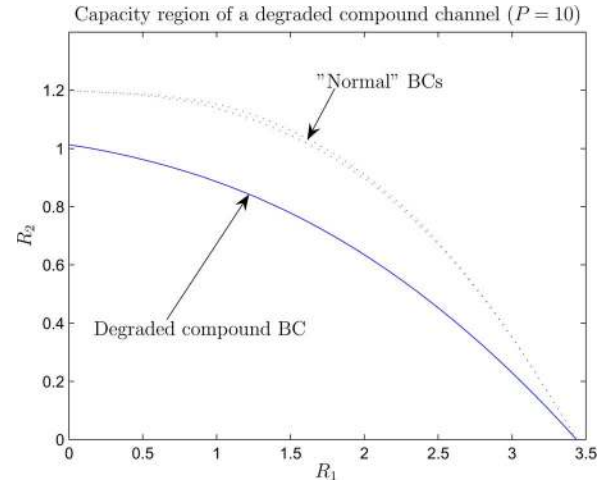

Fig. 1. The capacity region of the compound BC in (17).

In Fig. 1, we plotted the capacity region of the compound BC in (17) as well as the two capacity regions of the "normal" BCs that are created when we transmit the second message just to one of the realizations of the second user. As can be seen, the capacity region of the compound channel is contained within the other two regions. One could wonder whether the compound BC region may be obtained by switching between the two "normal" BC schemes and hence, "halving" rate R_2. However, Fig. 1 illustrates that simultaneously transmitting to both realizations is still far superior.

VII. Summary

In this paper, we give an expression for the capacity region of a compound multiple-antenna BC with two users where one user is degraded with respect to the other. In this compound channel, each of the two users has several possible realizations and at the receiver, each user has perfect knowledge of the actual realization. We require that regardless of the actual channel realizations, the messages should be received successfully. Alternatively, this channel may be viewed as a broadcast channel with common messages. The different realizations of the channel can be considered as different users to whom a common message is being transmitted.

The degradedness order between the two users was defined through a third, fictitious, user which is degraded with respect to all realizations of one user while all realizations of the second user are degraded with respect to him.

In this work, we brought to bare two, relatively, new tools. The first is an extremal inequality as it appears in Lemma 2 and Corollary 4. This extremal inequality is an extension of a result which appeared in [12] and was useful in the non-compound Gaussian BC. The second tool is the enhancement technique which was first used in [15], [16] to characterize the capacity region of the multiple-antenna Gaussian broadcast channel with private messages. Later on [12], it was also shown that the extremal inequality could also be used to characterize the capacity region of the multiple-antenna Gaussian broadcast channel in the two-user case.

APPENDIX I
A DEGRADED BC WITH COMMON MESSAGES THAT DOES NOT MEET THE REQUIREMENTS IN (2)

In the following, we give a simple example in which we have a compound BC with common messages such that $N_i^1 \preceq N_j^2 \; \forall i,j$ but for which we cannot find a matrix N^* such that $N_i^1 \preceq N^* \preceq N_j^2 \; \forall i,j$. Consider an aligned compound BC with two transmit antennas where

$$N_1^1 = \frac{1}{3}\begin{pmatrix} 5 & 2 \\ 2 & 5 \end{pmatrix}, \quad N_2^1 = \frac{1}{3}\begin{pmatrix} 5 & -2 \\ -2 & 5 \end{pmatrix}$$

and

$$N_1^2 = \begin{pmatrix} 2 & \\ & 3 \end{pmatrix}, \quad N_2^2 = \begin{pmatrix} 3 & \\ & 2 \end{pmatrix}.$$

One can verify that indeed $N_i^1 \preceq N_j^2$ for $i,j = 1,2$. In addition, due to the choice of N_i^1 and N_j^2, if $N_i^1 \preceq N^* \preceq N_j^2$ (i.e., (2) holds) neither inequalities hold with equality. We now note that $N_1^2 - N_1^1 = \frac{1}{3}\binom{1}{-2}(1 \quad -2)$. That is, the difference is a matrix of rank one. Therefore, if $N_1^1 \preceq N^* \preceq N_1^2$ then

$$N^* = N_1^1 + \alpha \begin{pmatrix} 1 \\ -2 \end{pmatrix} (1 \quad -2) \tag{18}$$

for some $\alpha > 0$. This is the only possibility as the difference $N_1^2 - N_1^1$ is a rank one matrix. Adding any other positive semi-definite matrix (with different or additional eigenvectors) would immediately contradict $N^* \preceq N_1^2$. On the other hand

$$N_2^2 - N_1^1 = \frac{1}{3}\begin{pmatrix} 2 \\ 1 \end{pmatrix} (2 \quad 1).$$

Therefore

$$N^* = N_1^1 + \beta \begin{pmatrix} 2 \\ -1 \end{pmatrix} (2 \quad -1)$$

for some $\beta > 0$. However, as the added matrix here is a different rank one matrix, this contradicts (18). A simple way to observe this is to multiply (18) by $(1 \quad 2)$ on the left and $\binom{1}{2}$ on the right. This would yield $(1 \quad 2) N_1^1 \binom{1}{2} + 9\alpha$. Yet, performing the same multiplication on $N_1^1 + \beta \binom{2}{-1}(2 \quad -1)$ would yield $(1 \quad 2) N_1^1 \binom{1}{2}$, proving that the two options are necessarily different.

APPENDIX II
PROOF OF LEMMA 4

The proof of this lemma is very similar to the proof of the capacity region of a degraded broadcast channel in [6, Sec. 14.6, pp. 418–428]. As the converse of the proof is more relevant to our case, we only detail this part here. The proof of the direct part relies on successive decoding at the stronger user and is practically identical to that found in [6].

Let $\bar{Y}_i^1$ denote a sequence of n channel outputs of the ith realization of user 1 and let W_1 and W_2 be the message indices. Furthermore, let $Y_i^1(l)$ be the lth sample of $\bar{Y}_i^1$ and $Y_i^1(1 \ldots l-1)$ be the set of all samples up to $l-1$ (and including). We use similar notations for Y_j^2, Y^*, and X. As the capacity region

depends only on the marginals $P_{Y_i^1|X}$ and $P_{Y_j^2|X}$, we may assume without loss of generality that indeed the mutual distribution $P_{Y_1^1,\ldots,Y_{K_1}^1,Y_1^2,\ldots,Y_{K_2}^2,Y^*|X}$ is such that $(W_1,W_2) \to X \to Y_i^1 \to Y^* \to Y_j^2$ form a Markov chain for every choice of i and j. Using Fano's inequality and the fact that W_1 and W_2 are independent we can write for every $i = 1,\ldots,K_1$, expressions (19) at the top of the following page, where $V(l) = (W_2, Y^*(1 \ldots l-1))$ and where $\delta(n) \to 0$ as $n \to \infty$. The equality in (a) is due to the chain rule of mutual information. (b) is due to the Markov chain $(W_1,W_2) \to X \to Y_i^1 \to Y^*$ and the memoryless nature of the channel, i.e.,

$$P(Y^*(1\ldots k)|Y_i^1(1\ldots k), X(1\ldots k), W_1, W_2)$$
$$= \pi_{j=1}^k P(Y^*(j)|Y_i^1(j)),$$

as can be seen in the following identity, also given at the top of the following page. As (W_1,W_2) explicitly determine $\overline{X}$ step (c) follows. (d) and (e) follow, again, from the Markov chain $(W_1,W_2) \to X \to Y_i^1 \to Y^*$ and the memoryless nature of the channel, i.e.,

$$P(Y_i^1(l)|W_1, W_2, X(1\ldots l), Y_i^1(1\ldots l-1))$$
$$= P(Y_i^1(l)|W_1, W_2, X(1\ldots l))$$
$$= P(Y_i^1(l)|W_1, W_2, X(1\ldots l), Y^*(1\ldots l-1)).$$

Step (f) follows from the fact that conditioning decreases entropy.

In a similar manner we use Fano's inequality to bound the rate of the second user

$$R_2 \le \frac{1}{n} I(W_2; \bar{Y}_j^2) + \delta(n)$$
$$\stackrel{(d)}{=} \frac{1}{n} \sum_{l=1}^n I(W_2; Y_j^2(l)|Y_j^2(1\ldots l-1)) + \delta(n)$$
$$\stackrel{(e)}{\le} \frac{1}{n} \sum_{l=1}^n I(W_2, Y_j^2(1\ldots l-1); Y_j^2(l)) + \delta(n)$$
$$\stackrel{(f)}{\le} \frac{1}{n} \sum_{l=1}^n I(W_2, Y^*(1\ldots l-1); Y_j^2(l)) + \delta(n)$$
$$= \frac{1}{n} \sum_{l=1}^n I(V(l); Y_j^2(l)) + \delta(n). \tag{20}$$

The equality in (d) is due to the chain rule of mutual information. The inequality in (e) is due to the fact that conditioning reduces entropy, and the inequality in (f) is a result of the channel being memoryless and the Markov chain: $(W_1,W_2) \to X \to Y_i^1 \to Y^* \to Y_j^2$.

Next, we replace the index l with a random variable I which is uniformly distributed over the integers $1,\ldots,n$ and define $V = (V(I),I)$, $X = X(I)$, $Y_i^1 = Y_i^1(I)$, and $Y_j^2 = Y_j^2(I)$. As the channel is memoryless, we get

$$R_1 \le I(X; Y_i^1|V) + \delta(n)$$
$$R_2 \le I(V; Y_j^2) + \delta(n)$$

for all i and j where $P_{Y_i^1,X,V} = P_{Y_i^1|X} \cdot P_{X,V}$ and $P_{Y_j^2,X,V} = P_{Y_j^2|X} \cdot P_{X,V}$. Note that as the channel is memoryless we indeed

$$R_1 \leq \frac{1}{n} I(W_1; \bar{Y}_i^1 | W_2) + \delta(n)$$

$$\overset{(a)}{=} \frac{1}{n} \sum_{l=1}^n I(W_1; Y_i^1(l) | W_2, Y_i^1(1 \ldots l-1)) + \delta(n)$$

$$= \frac{1}{n} \sum_{l=1}^n \left(h(Y_i^1(l) | W_2, Y_i^1(1 \ldots l-1)) - h(Y_i^1(l) | W_1, W_2, Y_i^1(1 \ldots l-1)) \right) + \delta(n)$$

$$\overset{(b)}{=} \frac{1}{n} \sum_{l=1}^n \left(h(Y_i^1(l) | W_2, Y_i^1(1 \ldots l-1), Y^*(1 \ldots l-1)) - h(Y_i^1(l) | W_1, W_2, Y_i^1(1 \ldots l-1)) \right) + \delta(n)$$

$$\overset{(c)}{=} \frac{1}{n} \sum_{l=1}^n \left(h(Y_i^1(l) | W_2, Y_i^1(1 \ldots l-1), Y^*(1 \ldots l-1)) - h(Y_i^1(l) | W_1, W_2, X(1 \ldots l), Y_i^1(1 \ldots l-1)) \right) + \delta(n)$$

$$\overset{(d)}{=} \frac{1}{n} \sum_{l=1}^n \left(h(Y_i^1(l) | W_2, Y_i^1(1 \ldots l-1), Y^*(1 \ldots l-1)) - h(Y_i^1(l) | W_1, W_2, X(1 \ldots l)) \right) + \delta(n)$$

$$\overset{(e)}{=} \frac{1}{n} \sum_{l=1}^n \left(h(Y_i^1(l) | W_2, Y_i^1(1 \ldots l-1), Y^*(1 \ldots l-1)) - h(Y_i^1(l) | W_1, W_2, X(1 \ldots l), Y^*(1 \ldots l-1)) \right) + \delta(n)$$

$$= \frac{1}{n} \sum_{l=1}^n \left(h(Y_i^1(l) | W_2, Y_i^1(1 \ldots l-1), Y^*(1 \ldots l-1)) - h(Y_i^1(l) | W_1, W_2, X(l), Y^*(1 \ldots l-1)) \right) + \delta(n)$$

$$\overset{(f)}{\leq} \frac{1}{n} \sum_{l=1}^n \left(h(Y_i^1(l) | W_2, Y^*(1 \ldots l-1)) - h(Y_i^1(l) | W_1, W_2, X(l), Y^*(1 \ldots l-1)) \right) + \delta(n)$$

$$= \frac{1}{n} \sum_{l=1}^n I(X(l); Y_i^1(l) | W_2, Y^*(1 \ldots l-1)) + \delta(n)$$

$$= \frac{1}{n} \sum_{l=1}^n I\left(X(l); Y_i^1(l) | V(l) \right) + \delta(n) \tag{19}$$

$$P(Y_i^1(l) | W_2, Y_i^1(1 \ldots l-1), Y^*(1 \ldots l-1))$$

$$= \frac{P(Y^*(1 \ldots l-1) | W_2, Y_i^1(1 \ldots l-1), Y_i^1(l)) \times P(Y_i^1(l) | W_2, Y_i^1(1 \ldots l-1))}{P(Y^*(1 \ldots l-1) | W_2, Y_i^1(1 \ldots l-1))}$$

$$= \frac{P(Y^*(1 \ldots l-1) | Y_i^1(1 \ldots l-1)) \times P(Y_i^1(l) | W_2, Y_i^1(1 \ldots l-1))}{P(Y^*(1 \ldots l-1) | Y_i^1(1 \ldots l-1))}$$

$$= P(Y_i^1(l) | W_2, Y_i^1(1 \ldots l-1)).$$

---

have $V \rightarrow X \rightarrow Y_i^1 \rightarrow Y^* \rightarrow Y_j^2$. Finally, as the above inequalities hold for every $i = 1, \ldots, K_1$ and $j = 1, \ldots, K_2$, we complete the proof by taking n to infinity.

APPENDIX III
PROOF OF THEOREM 1 USING MULTIPLE-LETTER FANO BOUNDS

As we assume that (R_1^o, R_2^o) are achievable, we can find a sequence of codebooks, increasing in length n, that map the set of messages $W_1 \in \{1, \ldots, e^{nR_1}\}$ and $W_2 \in \{1, \ldots, e^{nR_2}\}$ onto a sequence of n channel inputs denoted by $\bar{\boldsymbol{x}}(W_1, W_2)$ and[4] obtain a probability of decoding error $P_e(n)$ such that $P_e(n) \rightarrow 0$ as $n \rightarrow \infty$.

[4]$\bar{\boldsymbol{x}}(W_1, W_2)$ is a matrix of size $t \times n$. Throughout the rest of this Appendix, we shall omit (W_1, W_2) and remember that $\bar{\boldsymbol{x}}$ is a function of W_1 and W_2.

As we assume a power limitation on the input, we shall be interested in the second moments of $\bar{\boldsymbol{x}}$. We define the covariance of $\bar{\boldsymbol{x}}$ as $\boldsymbol{K}_{\bar{\boldsymbol{x}}} = E[\text{concat}(\bar{\boldsymbol{x}}) \cdot \text{concat}(\bar{\boldsymbol{x}})^T]$ where $\text{concat}(\bar{\boldsymbol{x}})$ is the concatenation of all the columns in $\bar{\boldsymbol{x}}$ into one big vector of size $nt \times 1$ and $\boldsymbol{K}_{\bar{\boldsymbol{x}}}$ is a matrix of size $nt \times nt$.

Due to the power constraint requirements in Lemma 2, it will be important for us to show that we are able not only to find one codebook for each length n, but also can find an entire equiprobable ensemble of such codebooks such that when we average over the entire ensemble, $\boldsymbol{K}_{\bar{\boldsymbol{x}}} = \boldsymbol{I} \otimes \boldsymbol{K}$ where $\boldsymbol{K} \preceq \boldsymbol{S}$ is of size $t \times t$ and $\otimes$ is the Kronecker product.

In order to construct such an ensemble, we first note that as the Gaussian noise is symmetric w.r.t. its average, we can easily create new codebooks with the same length which obtain the same error probabilities by multiplying any of the column vectors in $\bar{\boldsymbol{x}}$ by -1. Thus, we can create 2^n codebooks, numbered $k = 0, \ldots, 2^n - 1$, by multiplying column $i = 1 \ldots n$ by $(-1)^{b_i}$

where $b_n \ldots b_2 b_1$ is the binary representation of k. If we denote by $\boldsymbol{K}_{\bar{\boldsymbol{x}}}^k$ the covariance of codebook k, this collection of codebooks has the interesting property that $\frac{1}{2^n} \sum_{k=0}^{2^n-1} \boldsymbol{K}_{\bar{\boldsymbol{x}}}^k$ is block diagonal with n blocks of size $t \times t$ on its diagonal.

We can also use the fact that the channel is memoryless to construct from every codebook we found till now additional n codebooks by cyclically shifting the codebooks by $i = 0, \ldots, n-1$ time samples. Thus, we can create $k = 0, \ldots, n2^n - 1$ codebooks of length n and the same performance. Furthermore, the average of the covariances of these codebooks is given by $\frac{1}{n2^n} \sum_{k=1}^{n2^n-1} \boldsymbol{K}_{\bar{\boldsymbol{x}}}^k = \boldsymbol{I} \otimes \boldsymbol{K}$, where $\boldsymbol{K}$ is of size $t \times t$ and $\otimes$ is the Kronecker product. Finally, as we assume an average covariance constraint, we must have $\boldsymbol{K} \preceq \boldsymbol{S}$.

We now apply Fano's inequality to show that (R_1^o, R_2^o) cannot be achieved. We denote by $\bar{\boldsymbol{y}}_i^1$, $i = 1, \ldots, K_1$, and $\bar{\boldsymbol{y}}_j^2$, $j = 1, \ldots, K_2$, n time samples of the channel outputs. As the messages W_1 and W_2 are independent, we may write Fano's inequality as follows:

$$\frac{1}{n} I(W_1; \bar{\boldsymbol{y}}_i^1 | W_2) \geq R_1^o - \delta(n), \qquad \forall i = 1, \ldots, K_1$$
$$\frac{1}{n} I(W_2; \bar{\boldsymbol{y}}_j^2) \geq R_2^* + b - \delta(n), \quad \forall j = 1, \ldots, K_2.$$

where $\delta(n) \to 0$ as $n \to \infty$. These inequalities hold for all the codebooks in the ensemble obtained above. Assume that the codebooks are chosen randomly[5] and let V be a random variable which uniformly takes values between 0 and $n2^n - 1$ according to the codebook being chosen. As the codebooks have the same performance, Fano's inequality holds also when we condition the mutual information on V

$$\frac{1}{n} I(W_1; \bar{\boldsymbol{y}}_i^1 | W_2, V) \geq R_1^o - \delta(n), \qquad \forall i = 1 \ldots K_1 \quad (21)$$
$$\frac{1}{n} I(W_2; \bar{\boldsymbol{y}}_j^2 | V) \geq R_2^* + b - \delta(n), \quad \forall j = 1 \ldots K_2. \quad (22)$$

As we assume that $\frac{1}{2} \log \frac{|\boldsymbol{Q}_1^* + \boldsymbol{N}_i^1|}{|\boldsymbol{N}_i^1|} = R_1^o$ for all i,[6] we can use (21) to write

$$\frac{1}{n} I(W_1; \bar{\boldsymbol{y}}_i^1 | W_2, V) \geq R_1^o - \delta(n)$$
$$= \frac{1}{2} \log \frac{|\boldsymbol{Q}_1^* + \boldsymbol{N}_i^1|}{|\boldsymbol{N}_i^1|} - \delta(n), \quad \forall i.$$

Therefore, as $\frac{1}{n} h(\bar{\boldsymbol{y}}_i^1 | W_1, W_2, V) = \frac{1}{2} \log |2\pi e \boldsymbol{N}_i^1|$, we can write

$$\frac{1}{n} \sum_{i=1}^{K_1} \lambda_i^1 h(\bar{\boldsymbol{y}}_i^1 | W_2, V) \geq \sum_{i=1}^{K_1} \lambda_i^1 \frac{1}{2} \log |2\pi e (\boldsymbol{Q}_1^* + \boldsymbol{N}_i^1)| - \delta(n). \quad (23)$$

We now use Lemma 2 and Corollary 4. We can use these results for the following reasons.

- Equations (11) holds also when we pre-multiply (using Kronecker's product) each of the matrices by $\bar{\boldsymbol{I}}$.

[5] The codebook is known both to the transmitter and the receiver.

[6] Recall that earlier in the paper in Section IV we assume that $\lambda_i^1 > 0 \; \forall i$ and therefore the rate for each i is obtained with equality.

- We assume (for now) that $\boldsymbol{O}_1 = 0$.
- The Gaussian codes meet the covariance constraint on the ensemble of codes (i.e., $\boldsymbol{I} \otimes \boldsymbol{S}$).

Therefore, we may write

$$\frac{1}{n} \sum_{i=1}^{K_1} \lambda_i^1 h(\bar{\boldsymbol{y}}_i^1 | W_2, V) - \frac{1}{n} \mu \left(\sum_{j=1}^{K_2} \lambda_j^2 h(\bar{\boldsymbol{y}}_j^2 | W_2, V) \right)$$
$$\leq \sum_{i=1}^{K_1} \lambda_i^1 \frac{1}{2} \log |2\pi e (\boldsymbol{Q}_1^* + \boldsymbol{N}_i^1)|$$
$$- \mu \left(\sum_{j=1}^{K_2} \lambda_j^2 \frac{1}{2} \log |2\pi e (\boldsymbol{Q}_1^* + \boldsymbol{N}_j^2)| \right).$$

Thus, using the above equation and (23) we obtain

$$\frac{1}{n} \sum_{j=1}^{K_2} \lambda_j^2 h(\bar{\boldsymbol{y}}_j^2 | W_2, V)$$
$$\geq \sum_{j=1}^{K_2} \lambda_j^2 \frac{1}{2} \log |2\pi e (\boldsymbol{Q}_1^* + \boldsymbol{N}_j^2)| - \delta(n). \quad (24)$$

We can now use the assumption that $\frac{1}{2} \log \frac{|\boldsymbol{S} + \boldsymbol{N}_j^2|}{|\boldsymbol{Q}_1^* + \boldsymbol{N}_j^2|} = R_2^*$ for all j[7] to write

$$\frac{1}{n} \sum_{j=1}^{K_2} \lambda_j^2 h(\bar{\boldsymbol{y}}_j^2 | V)$$
$$= \frac{1}{n} \sum_{j=1}^{K_2} \lambda_j^2 I(W_2; \bar{\boldsymbol{y}}_j^2 | V) + \frac{1}{n} \sum_{j=1}^{K_2} \lambda_j^2 h(\bar{\boldsymbol{y}}_j^2 | W_2, V)$$
$$\overset{(a)}{\geq} \sum_{j=1}^{K_2} \lambda_j^2 R_2^* + b - \delta(n) + \frac{1}{n} \sum_{j=1}^{K_2} \lambda_j^2 h(\bar{\boldsymbol{y}}_j^2 | W_2, V)$$
$$= \frac{1}{2} \sum_{j=1}^{K_2} \lambda_j^2 \log \frac{|\boldsymbol{S} + \boldsymbol{N}_j^2|}{|\boldsymbol{Q}_1^* + \boldsymbol{N}_j^2|} + b - \delta(n)$$
$$+ \frac{1}{n} \sum_{j=1}^{K_2} \lambda_j^2 h(\bar{\boldsymbol{y}}_j^2 | W_2, V)$$
$$\overset{(b)}{\geq} \frac{1}{2} \sum_{j=1}^{K_2} \lambda_j^2 \log \frac{|\boldsymbol{S} + \boldsymbol{N}_j^2|}{|\boldsymbol{Q}_1^* + \boldsymbol{N}_j^2|} + b - \delta(n)$$
$$+ \sum_{j=1}^{K_2} \lambda_j^2 \frac{1}{2} \log |2\pi e (\boldsymbol{Q}_1^* + \boldsymbol{N}_j^2)| - \delta(n)$$
$$= \sum_{j=1}^{K_2} \lambda_j^2 \frac{1}{2} \log |2\pi e (\boldsymbol{S} + \boldsymbol{N}_j^2)| + b - 2\delta(n) \quad (25)$$

where (a) is due to (22) and (b) is due to (24).

On the other hand, using the upper bound on the entropy of a random vector with a covariance constraint we may write

$$\frac{1}{n} \sum_{j=1}^{K_2} \lambda_j^2 h(\bar{\boldsymbol{y}}_j^2 | V)$$
$$= \frac{1}{n} \sum_{j=1}^{K_2} \lambda_j^2 \left(\sum_{k=0}^{n2^n-1} \frac{1}{n2^n} h(\bar{\boldsymbol{y}}_j^2 | V = k) \right)$$

[7] Recall that earlier in the paper in Section IV we assume that $\lambda_j^2 > 0 \; \forall j$ and therefore the rate for each j is obtained with equality.

$$\leq \frac{1}{n} \sum_{j=1}^{K_2} \lambda_j^2 \left(\sum_{k=0}^{n2^n-1} \frac{1}{n2^n} \frac{1}{2} \log |2\pi e \boldsymbol{K}_{\bar{\boldsymbol{y}}_j^2}^k| \right)$$

$$\overset{(a)}{\leq} \frac{1}{n} \sum_{j=1}^{K_2} \lambda_j^2 \frac{1}{2} \log \left| 2\pi e \sum_{k=0}^{n2^n-1} \frac{1}{n2^n} \boldsymbol{K}_{\bar{\boldsymbol{y}}_j^2}^k \right|$$

$$\overset{(b)}{\leq} \sum_{j=1}^{K_2} \lambda_j^2 \frac{1}{2} \log \left| 2\pi e (\boldsymbol{S} + \boldsymbol{N}_j^2) \right|$$

where $\boldsymbol{K}_{\bar{\boldsymbol{y}}_j^2}^k$ denotes the covariance matrix of $\bar{\boldsymbol{y}}_j^2$ given codebook k in the same way $\boldsymbol{K}_{\bar{\boldsymbol{x}}}^k$ is defined. (a) holds due to the concavity of the $\log \det$ function and (b) holds due to the covariance constraint on the ensemble of codes.

However, the above equation contradicts (25) as we can make $\delta(n)$ arbitrarily smaller than b by taking n to be large enough. Thus, we conclude that if $\boldsymbol{O}_1 = 0$, (R_1^o, R_2^*) lie on the boundary of the capacity region and therefore (R_1^o, R_2^o) cannot be achievable. The extension of this result to the case of $\boldsymbol{O}_1 \succeq 0$ follows as detailed in Section IV.

REFERENCES

[1] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA: Athena Scientific, 2003.

[2] E. Bonek, M. Herdin, W. Weichselberger, and H. Ozcelik, "MIMO—Study propagation first," in *ISSPIT 2003. Proc. 3rd IEEE Int. Symp. Signal Processing and Information Technology, 2003.*, Darmstadt, Germany, Dec. 2003, pp. 150–153.

[3] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[4] G. Caire, S. Shamai, Y. Steinberg, and H. Weingarten, "On Information Theoretic Aspects of MIMO-Broadcast Channel," in *Space–Time Wireless Systems: From Array Processing to MIMO Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[5] T. M. Cover, "Comments on broadcast channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2524–2530, Sep. 1998.

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

[7] C. Imre and K. János, *Information Theory: Coding Theorems for Discrete Memoeyless Systems*. New York: Academic, 1981.

[8] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1501–1518, Nov. 1991.

[9] S. N. Diggavi and D. N. C. Tse, "On opportunistic codes and broadcast codes with degraded message sets," in *Proc. 2006 IEEE Information Theory Workshop*, Punta del Este, Uruguay, Mar. 2006, pp. 227–231.

[10] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. New York: Cambridge Univ. Press, 1991.

[11] N. Jindal and A. Goldsmith, "Optimal power allocation for parallel broadcast channels with independent and common information," in *Proc.IEEE Int. Symp. Information Theory (ISIT 2004)*, Chicago, IL, Jun./Jul. 2004, p. 215.

[12] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information theoretic problems," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1839–1851, May 2007.

[13] DVB-H Mobile Digital TV TI [Online]. Available: http://focus.ti.com/pdfs/wtbu/ti_dvbh_overview.pdf

[14] H. Weingarten, G. Kramer, and S. Shamai (Shitz), "On the compound MIMO broadcast channel," in *Proc. Information Theory and Applications (ITA 2007)*, UCSD, Palo Alto, CA, Feb. 2007, submitted for publication to *IEEE Trans. Inf. Theory.*.

[15] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian MIMO broadcast channel," in *Proc. Conf. Information Sciences and Systems (CISS 2004)*, Princeton, NJ, Mar. 2004, pp. 7–12.

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

[17] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "On the capacity region of the multiple-antenna broadcast channel with common messages," in *Proc. IEEE Int. Symp. Information Theory (ISIT2006)*, Seattle, WA, Jul. 2006, pp. 2195–2199.

Hanan Weingarten (S'06–M'07) received his B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion–Israel Institute of Technology, Haifa, Israel, in 1995, 2002, and 2007 respectively.

During the Summer of 2005 he was an intern at Bell Labs, Murray Hill, NJ. He is interested in several topics in information theory including, multiuser systems, fading channels, and coding theory. Currently, he is a cofounder of a new startup company in Haifa, Israel.

Dr. Weingarten has won several prizes during his Ph.D. work including the IEEE TRANSACTIONS ON INFORMATION THEORY 2007 Best Paper award, the Advanced Communications Center Student Competition Award (2007), and the Jacobs Prize for outstanding publication (2007), awarded by the Technion Graduate School.

Tie Liu (S'99–M'06) received the B.S. (1998) and M.S. (2000) degrees, both in electrical engineering, from the Tsinghua University, Beijing, China, and the M.S. degree in mathematics (2004) and the Ph.D. degree in electrical and computer engineering (2006) from the University of Illinois at Urbana-Champaign.

Since August 2006, he has been with the Texas A&M University, College Station, where he is currently an Assistant Professor in Electrical and Computer Engineering. His research interests are in the field of information theory, wireless communication, and statistical signal processing.

Prof. Liu is a recipient of the M. E. Van Valkenburg Graduate Research Award (2006) from the University of Illinois at Urbana-Champaign, the Best Paper Award (2008) from the 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications, and the Faculty Early Career Development (CAREER) Award (2009) from the National Science Foundation.

Shlomo Shamai (Shitz) (S'80–M'82–SM'88–F'94) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion–Israel Institute of Technology, Haifa, Israel, in 1975, 1981, and 1986 respectively.

During 1975–1985, he was with the Communications Research Labs in the capacity of a Senior Research Engineer. Since 1986, he has been with the Department of Electrical Engineering, Technion–Israel Institute of Technology, where he is now the William Fondiller Professor of Telecommunications. His research interests encompasses a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is a member of the Union Radio Scientifique Internationale (URSI). He is the recipient of the 1999 van der Pol Gold Medal of URSI, and a corecipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 joint IT/COM societies paper award, and the 2007 IEEE Information Theory Society Paper Award. He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY, and also has served on the Board of Governors of the IEEE Information Theory Society.

Yossef Steinberg (M'96–SM'09) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering in 1983, 1986, and 1990, respectively, all from Tel-Aviv University, Tel-Aviv, Israel.

He was a Lady Davis Fellow in the Department of Electrical Engineering, Technion, and held visiting appointments in the Department of Electrical Engineering at Princeton University, Princeton, NJ, and at the C I Center, George Mason University, Fairfax, VA. From 1995 to 1999, he was with the Department of Electrical Engineering, Ben Gurion University, Beer-Sheva, Israel. In 1999, he joined the Department of Electrical Engineering at the Technion.

Prof. Steinberg served as Associate Editor for Shannon Theory in the IEEE TRANSACTIONS ON INFORMATION THEORY, and won the 2007 best paper award, jointly with Hanan Weingarten and Shlomo Shamai.

Pramod Viswanath (S'88–M'03) received the Ph.D. degree in electrical engineering and computer science (EECS) from the University of California at Berkeley in 2000.

He was a member of technical staff at Flarion Technologies until August 2001 before joining the Electrical and Computer engineering Department at the University of Illinois, Urbana-Champaign.

Prof. Viswanath is a recipient of the Eliahu Jury Award from the EECS Department of University of California, Berkeley (2000), the Bernard Friedman Award from the Mathematics department of UC Berkeley (2000), and the NSF CAREER Award (2003). He was an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY for the period 2006–2008.