Capacity Bounds for Gaussian Vector Broadcast Channels

S. Vishwanath, G. Kramer, S. Shamai (Shitz), S. Jafar, and A. Goldsmith

Abstract. An outer bound on the capacity region of broadcast channels is presented called the degraded, same marginals (DSM) bound. This bound includes and improves on Sato's sum-rate bound. The DSM bound is applied to Gaussian vector broadcast channels, and it is found that the Gel'fand/Pinsker/Costa dirty-paper coding technique achieves all points inside the bound if only Gaussian code books are permitted. This result suggests that dirty paper coding gives the capacity region of Gaussian vector broadcast channels, but the optimality of Gaussian code books remains to be verified.

1. Introduction

A broadcast channel [1] with K receivers is defined by an input alphabet $\mathcal{X}$, K output alphabets $\mathcal{Y}_k$ for $1 \leq k \leq K$, and a conditional probability distribution

$$P(y_1, y_2, \dots, y_K|x) \tag{1.1}$$

where $x \in \mathcal{X}$ and $y_k \in \mathcal{Y}_k$ for $1 \leq k \leq K$. The channel is said to be *physically degraded* if (1.1) factors as

$$P(y_{\pi(1)}|x)\, P(y_{\pi(2)}|y_{\pi(1)}) \,\cdots\, P(y_{\pi(K)}|y_{\pi(K-1)}) \tag{1.2}$$

for some permutation $\pi(\cdot)$ on the integers $1, 2, \dots, K$. For simplicity, we assume the random variables Y_k to be labeled so that we can ignore the permutation, i.e., $\pi(k) = k$. Thus, the channel is physically degraded if

$$X \to Y_1 \to Y_2 \to \dots \to Y_K$$

forms a Markov chain.

The capacity region of the physically degraded broadcast channel is known and was determined by Bergmans [2] (coding theorem) and Gallager [3] (converse). These two papers were concerned mainly with giving an information-theoretic *expression* for the capacity region, and did not attempt to find a channel input distribution that optimizes this expression. For the Gaussian scalar channel this optimization was done by Bergmans [4] who showed that Gaussian code books are optimal. After this early work there have been many papers addressing general

Key words and phrases. broadcast channels, multiple antennas.

At the time of the workshop we thought we had a proof that Gaussian input distributions are optimal, and we presented the main arguments of this purported "proof". Unfortunately, in the time after the workshop and before this document was due, we found a step that we could not verify. Thus, despite the progress reported here showing that dirty-paper coding is optimal for Gaussian code books, the problem of determining the *capacity region* remains open.

broadcast channels. A review of the known results until 1977 is given in [5, Sec. V]. A recent review can be found in [6].

We are interested primarily in the recent work of Caire and Shamai [7] on Gaussian *vector* broadcast channels. Our interest stems in part from the practical importance of such channels for wireless communication, and in part because they are not degraded. The paper [7] exploits a coding technique of Gel'fand and Pinsker [8] that was applied to Gaussian scalar channels by Costa [9] and to Gaussian vector channels by Yu [10]. Costa called this coding scheme *writing on dirty paper* and we will call it *dirty paper coding.* It turns out that dirty paper coding, as applied to broadcast channels, is a special case of Marton's celebrated coding theorem [11] (see [8, 12]).

It was found in [7] that dirty paper coding achieves the *sum-rate* capacity of a Gaussian broadcast channel. The paper [7] considers only the two-receiver problem where the transmitter has two antennas and the receivers each have one antenna. This work was subsequently generalized to any number of receivers with any number of antennas; [12] addresses the problem with receivers having one antenna and [13, 14] address the general problem.

We are here concerned with the problem of determining the entire capacity *region* of Gaussian vector broadcast channels. Our approach will be to apply a genie-aided bound we call the *degraded, same-marginals* (DSM) bound that converts a non-degraded channel into a (generally better) degraded channel.[1] We find that this bound gives the best possible region *if the code books are restricted to be Gaussian.* It now seems natural to suspect that the capacity region must be the achievable region we have given here. However, until now we have been unable to prove that Gaussian code books are optimal.

This paper is organized as follows. Section 2 reviews the information-theoretic expression for the capacity region of a degraded broadcast channel. Section 3 reviews known outer bounds on the capacity region and introduces the DSM bound. Section 4 describes the Gaussian models we consider and gives achievable regions when one uses Gaussian code books. Section 5 shows that the dirty paper coding region is the same as the DSM outer bound when using Gaussian distributions.

2. A Capacity Expression

Suppose the transmitter wants to communicate a message W_k to receiver k for $k = 1, 2, \dots, K$, and that these K messages are statistically independent. If W_k has B_k bits and the channel is used N times, then the rate for receiver k is $R_k = B_k/N$ bits per channel use. The *capacity region* $\mathcal{C}$ of the broadcast channel is the closure of the set of K-tuples $(R_1, R_2, \dots, R_K)$ at which the communication can be made *reliable*, i.e., receiver k's estimate $\widehat{W}_k$ can be made to satisfy $\Pr(\widehat{W}_k \neq W_k) \leq \epsilon$ for any $\epsilon > 0$ and all k.

Broadcast channels turn out to have the following exceedingly useful property.

LEMMA 2.1 (Cover [1]). *Broadcast channels that have the same marginal distributions $P(y_k|x)$ for $k = 1, 2, \dots, K$ have the same capacity region $\mathcal{C}$.*

The import of this lemma is that we can consider any joint distribution $P(y_1, y_2, \dots, y_K|x)$ for which the marginals are the right ones. We can now design encoding schemes for

[1] The DSM bound was derived independently by Viswanath and Tse [16], who also presented this bound at the workshop.

this new channel and they will perform the same (in terms of rate) as for our original channel. Furthermore, any outer bound on the capacity region of the new channel will be an outer bound on the capacity region of the original channel. We will here consider the class of *degraded* (or stochastically degraded) broadcast channels which have the same marginal distribution as the physically degraded broadcast channel [1, p. 4].

The capacity region of the physically degraded broadcast channel is known to be the convex hull of the closure of the $(R_1, R_2, \ldots, R_K)$ satisfying

$$0 \leq R_k \leq I(U_k; Y_k | U_{k+1} U_{k+2} \ldots U_K) \tag{2.1}$$

for $k = 1, 2, \ldots, K$, where $P(u_1, \ldots, u_K, x, y_1, \ldots, y_K)$ factors as

$$P(u_1, \ldots, u_K, x)\, P(y_1|x) \prod_{k=2}^{K} P(y_k|y_{k-1}). \tag{2.2}$$

The above region was described by Bergmans [2, Sec. III.B] and is identical to his expression if we set $X_k = [U_k\ U_{k+1}\ \ldots\ U_K]$. Observe that we have

$$I(U_1; Y_1 | U_2 \ldots U_K) = I(X_1; Y_1 | X_2) \leq I(X; Y_1 | X_2) \tag{2.3}$$

where the inequality follows by (2.2). This means that we can remove the random variable U_1 in (2.1) and (2.2) if we so desire. For example, for $K = 2$ the capacity region is the convex hull of the closure of the set of (R_1, R_2) satisfying

$$0 \leq R_1 \leq I(X; Y_1 | U_2) \tag{2.4}$$

$$0 \leq R_2 \leq I(U_2; Y_2) \tag{2.5}$$

where $P(u_2, x, y_1, y_2) = P(u_2, x)\, P(y_1|x)\, P(y_2|y_1)$. However, we will sometimes find it useful to keep U_1 because it represents the first receiver's message.

3. Capacity Outer Bounds

3.1. Basic Bounds. The first capacity outer bound for broadcast channels was given in the paper that introduced such channels.

LEMMA 3.1 (Cover [1, Sec. VIII], Bergmans [2, Sec. IV]).
The capacity region of a broadcast channel $P(y_1, \ldots, y_K|x)$ satisfies, for all k,

$$R_k \leq \max_{P_X} I(X; Y_k). \tag{3.1}$$

A simple extension of this bound for degraded channels appeared soon after [1], and we include it here for completeness.

LEMMA 3.2 (Wyner [2, Ack.]). *The capacity region of a degraded broadcast channel satisfies, for all k,*

$$\sum_{i=k}^{K} R_i \leq \max_{P_X} I(X; Y_k). \tag{3.2}$$

Of course, this bound was soon superseded by Gallager's capacity characterization in [3]. An outer bound for *general* broadcast channels was given by Marton in [11].

LEMMA 3.3 (Körner-Marton [11]). *The capacity region of a two receiver broadcast channel $P(y_1, y_2|x)$ satisfies*

$$\begin{aligned} R_1 &\le I(X;Y_1) \\ R_2 &\le I(U;Y_2) \\ R_1 + R_2 &\le I(X;Y_1|U) + I(U;Y_2) \end{aligned} \tag{3.3}$$

for some $P(u,x)$ where $P(u,x,y_1,y_2) = P(u,x)P(y_1,y_2|x)$. One can further add the bound obtained by swapping Y_1 and Y_2.

REMARK 3.4. Lemma 3.3 is currently the best known outer bound on the capacity of general broadcast channels. The bounds we give in the next section are in fact *included* in Lemma 3.3. However, there are reasons for using these other bounds, as explained below.

3.2. Bounds That Exploit Lemma 2.1. We now consider outer bounds that exploit Lemma 2.1 in the sense that *classes* of channels are considered. The first bound of this type was due to Sato [17], and the bound's two main steps are

1. let the receivers co-operate to get a point-to-point channel, and
2. minimize the capacity of this point-to-point channel over all joint channel distributions with the same marginals.

A descriptive name for Sato's bound might be the *co-operative, same marginals* (CSM) bound because of the two main ingredients involved in this bound. A formal statement of Sato's bound is the following.

LEMMA 3.5 (Sato [17]). *The capacity region $\mathcal{C}$ of a broadcast channel $P(y_1, \ldots, y_K|x)$ satisfies*

$$\sum_{k=1}^{K} R_k \le \min_{P_{\widetilde{Y}_1 \ldots \widetilde{Y}_K|X} \in \mathcal{P}} \max_{P_X} I(X; \widetilde{Y}_1 \widetilde{Y}_2 \ldots \widetilde{Y}_K) \tag{3.4}$$

where $\mathcal{P}$ is the set of channels $P(\widetilde{y}_1 \ldots \widetilde{y}_K|x)$ that have the same marginals as $P(y_1, \ldots, y_K|x)$, i.e., $P_{\widetilde{Y}_k|X}(y|x) = P_{Y_k|X}(y|x)$ for all k, x and y.

Note that Sato's bound can be used for any broadcast channel, but is weaker than Lemma 3.3.[2] Sato applied this bound to the Gaussian scalar channel in [17] where he found that it is tight for the sum-rate. However, this result was hardly surprising since it was known that maximizing the sum rate for degraded channels is the same as sending to the better user only. Much more surprising was the discovery of Caire and Shamai [7] that this bound is tight for certain Gaussian *vector* broadcast channels that are *not* degraded. This discovery led to a number of papers that showed Sato's bound to be tight in the general Gaussian setting, i.e., any number of receivers and any number of transmit and receive antennas [12–14]. A key to proving these results was the multiple-access/broadcast channel duality of [15].

We here wish to make some progress on finding the entire capacity region $\mathcal{C}$. To this end, we combine Sato's "same marginals" insight with a "degraded" technique first used by Ozarow and Leung-Yan-Cheong in the context of broadcast channels with feedback [18]. The resulting bound is as follows.

[2] We have used "min" and "max" in Lemma 3.5, which suffices for the channels and distributions we consider. A bound for general channels and distributions would instead require an "inf" and a "sup", respectively.

1. Create a degraded broadcast channel by giving receiver k the channel outputs of receivers $k+1, k+2, \ldots, K$.
2. Minimize the capacity of this degraded channel over all joint channel distributions with the same marginals.

For obvious reasons, we call this bound the *degraded, same marginals* (DSM) outer bound. Of course, the first step could have been done in any order, so that one can further take the intersection of regions obtained with different degraded orderings. A formal statement of the DSM bound is the following lemma. For this lemma, we consider a broadcast channel $P(y_1, \ldots, y_K|x)$ and let $\mathcal{R}_D(\pi, P(y_1, \ldots, y_K|x))$ be the capacity region of the degraded broadcast channel created by giving receiver $\pi(k)$ the outputs $Y_{\pi(k)}, Y_{\pi(k+1)}, \ldots, Y_{\pi(K)}$. This capacity region is given by (2.1) and (2.2).

LEMMA 3.6 (DSM Bound). *The capacity region $\mathcal{C}$ of a broadcast channel $P(y_1, \ldots, y_K|x)$ satisfies*

$$\mathcal{C} \subseteq \bigcap_{\pi} \left\{ \bigcap_{P_{\widetilde{Y}_1 \ldots \widetilde{Y}_K|X} \in \mathcal{P}} \mathcal{R}_D(\pi, P_{\widetilde{Y}_1 \ldots \widetilde{Y}_K|X}) \right\}. \tag{3.5}$$

where $\mathcal{P}$ is the set of channels $P(\widetilde{y}_1 \ldots \widetilde{y}_K|x)$ that have the same marginals as $P(y_1, \ldots, y_K|x)$, i.e., $P_{\widetilde{Y}_k|X}(y|x) = P_{Y_k|X}(y|x)$ for all k, x and y.

REMARK 3.7. The DSM bound applies to *general* broadcast channels.

REMARK 3.8. The DSM bound is *stronger* than Sato's bound because the sum rate cannot be more than the capacity of the best channel.

REMARK 3.9. Lemma 3.3 includes the DSM bound, as can be seen by weakening (3.3) by replacing Y_1 with (Y_1, Y_2). However, just as Sato's bound seems easier to use than Lemma 3.3, so it seems that Lemma 3.6 is easier to use than Lemma 3.3. It is unclear whether one can obtain the results we present in Section 5 without going through Lemma 3.6.

4. Gaussian Scalar and Vector Channels

This section introduces the Gaussian channel models we wish to consider. The DSM bound is later applied to such channels.

4.1. Gaussian Scalar Channels. The Gaussian scalar broadcast channel has

$$Y_k = X + Z_k \tag{4.1}$$

for $1 \leq k \leq K$, where X is a complex random variable and Z_k is zero-mean, complex, Gaussian noise that is *circularly symmetric* in the sense that its real and imaginary parts are independent and have the same variance $N_k/2$. There is further a power constraint $\mathrm{E}[|X|^2] \leq P$ on the input.

Consider the region defined by (2.1) and (2.2). We choose the U_k to be independent, zero-mean, complex, circularly symmetric, Gaussian random variables with variance $\mathrm{E}[|U_k|^2] = Q_k$ and set $X = \sum_{k=1}^{K} U_k$. We thus have $\sum_{k=1}^{K} Q_k \leq P$ and obtain as our achievable region the set of rate-tuples satisfying

$$0 \leq R_k \leq \log\left(N_k + \sum_{i=1}^{k} Q_i\right) - \log\left(N_k + \sum_{i=1}^{k-1} Q_i\right) \tag{4.2}$$

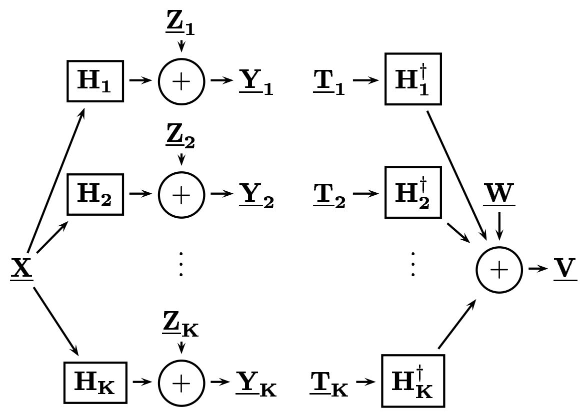

FIGURE 1. System models for the vector Gaussian broadcast channel (left) and its *dual* vector Gaussian multiple access channels (right).

where the units are nats per channel use (1 nat is $\log_2(e)$ bits). Bergmans [4] showed that (4.2) gives the capacity region by varying over all $(Q_1, \dots, Q_K)$ with $\sum_{k=1}^{K} Q_k \leq P$. This implies that Gaussian input distributions are optimal.

4.2. Gaussian Vector Channels. The Gaussian vector broadcast channel has vectors as inputs and outputs. The output of receiver k is

$$\underline{Y}_k = H_k \, \underline{X} + \underline{Z}_k \tag{4.3}$$

where $\underline{X}$ is a complex column vector of length t and $\underline{Y}_k$ is a complex column vector of length r_k. The channel matrix H_k is a $r_k \times t$ complex matrix and $\underline{Z}_k$ is a vector of r_k zero-mean, complex, circularly symmetric, Gaussian random variables. Let $Q_{\underline{X}} = \mathrm{E}[\underline{X}_k \underline{X}_k^\dagger]$ be the covariance matrix of $\underline{X}$, where $\dagger$ denotes complex conjugation and transposition. The transmitter power constraint is

$$\mathrm{Tr}(Q_{\underline{X}}) \leq P \tag{4.4}$$

where $\mathrm{Tr}(\cdot)$ is the trace operator. The Gaussian vector broadcast channel is *not* degraded in general, as can be seen for $K = 2$, $t = 2$, $r_1 = r_2 = 1$, $H_1 = [1 \ 0]$ and $H_2 = [0 \ 1]$.

The dirty paper coding region where the messages are encoded in the order $K, K-1, \dots, 1$ is the convex hull of

$$C_{BC}^{DP}(P) = \bigcup_{Q_1,\dots,Q_K} \left\{ \begin{array}{l} (R_1, \dots, R_K) : 0 \leq R_k \leq \\ \log \left| N_k + H_k \left(\sum_{i=1}^{k} Q_i \right) H_k^\dagger \right| \\ - \log \left| N_k + H_k \left(\sum_{i=1}^{k-1} Q_i \right) H_k^\dagger \right|, \text{ all } k \end{array} \right\} \tag{4.5}$$

where $|A|$ is the determinant of A, Q_k is now the covariance matrix of User k and $\sum_k Q_k = Q_{\underline{X}}$. Of course, one can perform the encoding in any of the $K!$ orderings.

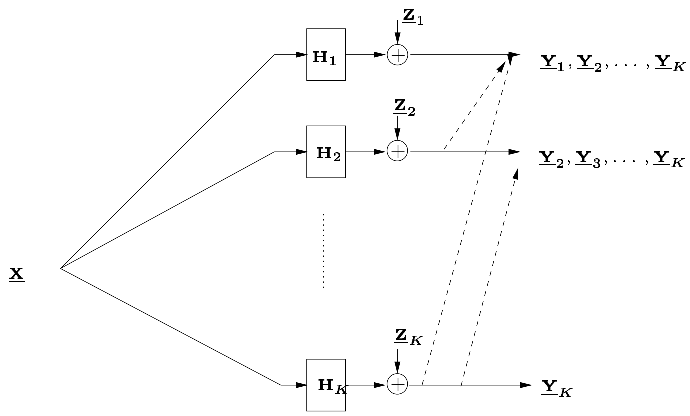

FIGURE 2. Model for the DSM bound.

4.3. The DSM Bound for Gaussian Vector Channels. We set $H_{Dk} = [H_k^\dagger \ldots H_K^\dagger]^\dagger$ and $\underline{Z}_{Dk} = [\underline{Z}_k^\dagger \ldots \underline{Z}_K^\dagger]^\dagger$ for all k. The channel outputs after step 1 of the DSM bound are thus

$$\underline{Y}_{Dk} = H_{Dk}\,\underline{X} + \underline{Z}_{Dk} \tag{4.6}$$

so that receiver k sees vectors of length $\sum_{i=k}^{K} r_i$. The second step of the DSM bound permits $\underline{Z}_{DK}$ to have *any* covariance matrix as long as the covariance matrices of the sub-vectors $\underline{Z}_k$ are the same as before.

Consider next the capacity region for the DSM bound given by (2.1) and (2.2). Suppose we choose the U_k, which we will now write as $\underline{U}_k$, to be independent, zero-mean, complex, Gaussian random vectors with covariance matrices $\mathrm{E}[\underline{U}_k\underline{U}_k^\dagger] = T_k$. We set $\underline{X} = \sum_{k=1}^{K}\underline{U}_k$ so that $\sum_{k=1}^{K}\mathrm{Tr}(T_k) \le P$. Thus, the DSM bound restricted to Gaussian inputs and the degraded ordering given in (4.6) is

$$C_{BC}^{DSM}(P) = \bigcap_{N_{DK}}\bigcup_{T_k}\left\{\begin{array}{l}(R_1,\ldots,R_K):\\ 0 \le R_k \le \log\left|N_{Dk} + H_{Dk}(\sum_{i=1}^{k}T_i)H_{Dk}^\dagger\right|\\ -\log\left|N_{Dk} + H_{Dk}\left(\sum_{i=1}^{k-1}T_i\right)H_{Dk}^\dagger\right|,\ \text{all } k\end{array}\right\} \tag{4.7}$$

where $N_{Dk} = \mathrm{E}[\underline{Z}_{Dk}\underline{Z}_{Dk}^\dagger]$. The notation $\cap_{N_{DK}}$ is used as a shorthand for the intersection over all N_{DK} that keep the covariance matrices of $\underline{Z}_k$ the same as the original problem. Similarly, the notation $\cup_{T_k}$ is used as a shorthand for $\cup_{T_1,\ldots,T_K}$. The next section shows that (4.7) is the same as (4.5).

5. Optimization of the DSM Bound

The region (4.7) can be rewritten as

$$\bigcap_{N_{DK}}\bigcup_{T_k}\left\{\begin{array}{l}(R_1,\ldots,R_K): 0 \le R_k \le \log|I+\\ \left(I + N_{Dk}^{-1/2}H_{Dk}\left(\sum_{i=1}^{k-1}T_i\right)H_{Dk}^\dagger N_{Dk}^{-1/2}\right)^{-1}\\ N_{Dk}^{-1/2}H_{Dk}T_kH_{Dk}^\dagger N_{Dk}^{-1/2}\Big|,\ \text{all } k\end{array}\right\} \tag{5.1}$$

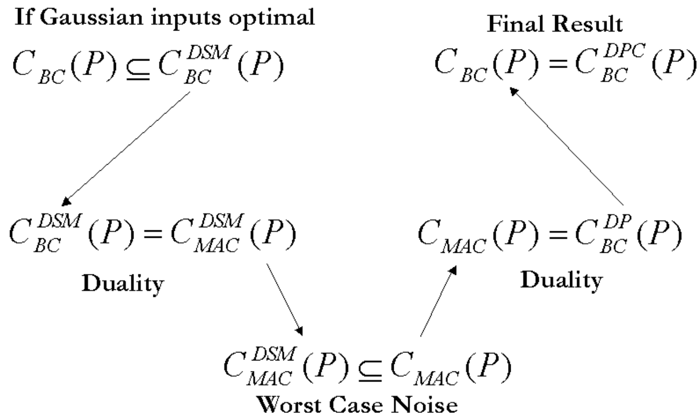

FIGURE 3. Pictorial representation of the steps of the proof.

Note that (5.1) is the dirty paper coding region of a broadcast channel with channel matrices $N_{Dk}^{-1/2}H_{Dk}$ and power constraint P. We use the duality result of [14] and replace (5.1) with the capacity region of the dual vector multiple access (MAC) with channel $H_{Dk}^{\dagger}N_{Dk}^{-1/2}$. The result is

$$C_{MAC}^{DSM}(P) = \tag{5.2}$$
$$\bigcap_{N_{DK}}\bigcup_{C_k}\left\{\begin{array}{l}(R_1,\dots,R_K): 0 \le \sum_{i=k}^{K} R_i \le \\ \log|I + \sum_{i=k}^{K} H_{Di}N_{Di}^{-1/2}C_iN_{Di}^{-1/2}H_{Di}^{\dagger}|, \\ \text{all } k\end{array}\right\}$$

where C_k is positive semidefinite for all k and $\sum_k \mathrm{Tr}(C_k) \le P$. Note that the numbering of users $\{1,\dots,K\}$ is arbitrary, and we can obtain a tighter upper bound by taking an intersection over all $K!$ outer bounds of the form (5.2). In subsection 5.1, we show that this intersection is inside the capacity region of the K user vector multiple access (MAC) with channel $H_k^{\dagger}$ and *joint* power constraint P. This region is given by:

$$C_{MAC}(P) = \tag{5.3}$$
$$\bigcup_{S_k}\left\{\begin{array}{l}(R_1,\dots,R_K): \\ 0 \le \sum_{i\in E} R_i \le \log\left|I + \sum_{i\in E} H_iS_iH_i^{\dagger}\right|,\end{array}\right\}$$

where E is any subset of $\{1,\dots,K\}$, S_k is positive semidefinite for all k and $\sum_k \mathrm{Tr}(S_k) \le P$. Employing duality [14] once again, we find that (5.3) is equivalent to the dirty paper coding region of the K user BC with channels H_k and power constraint P. Thus, after taking the intersection over all possible orderings, the DSM upper bound is the dirty paper coding region. The steps of our proof are depicted in Fig. 3, where $C_{BC}(P)$ is the "capacity" region of the vector Gaussian broadcast channel with Gaussian input distributions.

REMARK 5.1. One can show that there are interesting special cases where Gaussian input distributions are optimal, e.g., if the covariance matrices of the $\underline{U}_k$ are restricted to have unit rank.

5.1. Optimization Problems. Consider the K user MAC with channels $H_k^\dagger$ whose capacity region is (5.3). We later extend the analysis to the DSM upper bound, and finally show that (5.3) includes (5.2). Note that any point on the boundary of (5.3) is the solution of an optimization problem parameterized by a real nonnegative vector $(\mu_0 = 0, \mu_1 = 1, \ldots, \mu_K)$ [19]. Since the ordering of users $1, \ldots, K$ is arbitrary, it is sufficient to consider the case when $\mu_0 = 0, \mu_1 = 1, \mu_2 \leq \mu_3 \leq \ldots \leq \mu_K$. The optimization problem is

$$\max_{P_1,\ldots,P_K} \max_{S_1,\ldots,S_K} \sum_k (\mu_k - \mu_{k-1}) \log \left| I + \sum_{i=k}^{K} H_i^\dagger S_i H_i \right| \tag{5.4}$$

subject to

$$\begin{aligned} &\sum_k P_k = P \\ &\mathrm{Tr}(S_k) \leq P_k \ \forall\ k \\ &P_k, S_k \geq 0 \ \forall\ k. \end{aligned}$$

This optimization problem is convex. We can therefore formulate a *Lagrangian* dual problem for which the dual and primal optimum values are the same [21].

Consider first S_1 and note that only the first term of the outer sum in (5.4) is a function of S_1. We thus begin with the dual problem of

$$\max_{S_1} \log \left| I + \sum_{k=1}^{K} H_k^\dagger S_k H_k \right| \tag{5.5}$$

subject to

$$\begin{aligned} &\mathrm{Tr}(S_1) \leq P_1 \\ &S_1 \geq 0. \end{aligned}$$

Using results from the Appendix, the dual problem to (5.5) is

$$\min_{\Gamma_1,\lambda_1} -\log|\Gamma_1| - t + \mathrm{Tr}\left(\Gamma_1 (I + \sum_{k=2}^{K} H_k^\dagger S_k H_k)\right) + \lambda_1 P_1 \tag{5.6}$$

subject to

$$\begin{aligned} &\lambda_1 I \geq H_1 \Gamma_1 H_1^\dagger \\ &\Gamma_1 \geq 0. \end{aligned}$$

Substituting (5.6) into (5.4), the boundary point maximization problem becomes

$$\begin{aligned} \max_{P_1,\ldots,P_K} \max_{S_2,\ldots,S_K} \min_{\Gamma_1,\lambda_1} & -\log|\Gamma_1| - t + \mathrm{Tr}\left(\Gamma_1 (I + \sum_{k=2}^{K} H_k^\dagger S_k H_k)\right) + \lambda_1 P_1 \\ & + \sum_{k=2}^{K} (\mu_k - \mu_{k-1}) \log \left| I + \sum_{i=k}^{K} H_i^\dagger S_i H_i \right| \end{aligned} \tag{5.7}$$

subject to

$$\sum_k P_k = P$$
$$\mathrm{Tr}(S_k) \le P_k \text{ for } k = 2, \ldots, K$$
$$\lambda_1 I \ge H_1 \Gamma_1 H_1^\dagger$$
$$P_k, S_k, \Gamma_1 \ge 0 \ \forall\ k.$$

Using Ky Fan's min-max switching theorem [20], the objective function in (5.7) can be rewritten as

$$\min_{\Gamma_1,\lambda_1} \max_{P_1,\ldots,P_K} \max_{S_2,\ldots,S_K} -\log|\Gamma_1| - t + \mathrm{Tr}\left(\Gamma_1(I + \sum_{k=2}^K H_k^\dagger S_k H_k)\right) + \lambda_1 P_1$$
$$+ \sum_{k=2}^K (\mu_k - \mu_{k-1}) \log \left| I + \sum_{i=k}^K H_i^\dagger S_i H_i \right|. \tag{5.8}$$

Now observe that in (5.8) there are two terms that depend on S_2. Isolating these terms, we formulate a primal problem as

$$\max_{S_2} \log \left| I + \sum_{k=2}^K H_k^\dagger S_k H_k \right| + \frac{1}{\mu_2 - 1} \mathrm{Tr}\left(\Gamma_1(I + \sum_{k=2}^K H_k^\dagger S_k H_k)\right) \tag{5.9}$$

subject to

$$\mathrm{Tr}(S_2) \le P_2.$$

Again using results from the Appendix, the dual problem to (5.9) is

$$\min_{\Gamma_2,\lambda_2} -\log\left|\Gamma_2 - \frac{\Gamma_1}{\mu_2 - 1}\right| - t + \mathrm{Tr}\left(\Gamma_2(I + \sum_{k=3}^K H_k^\dagger S_k H_k)\right) + \lambda_2 P_2 \tag{5.10}$$

subject to

$$\lambda_2 I \ge H_2 \Gamma_2 H_2^\dagger$$
$$\Gamma_2 \ge 0.$$

Continuing in this fashion, at the j^{th} step we find that the optimization problem in terms of S_j is

$$\max_{S_j} \log \left| I + \sum_{k=j}^K H_k^\dagger S_k H_k \right| + \frac{\mu_{j-1} - \mu_{j-2}}{\mu_j - \mu_{j-1}} \mathrm{Tr}\left(\Gamma_{j-1}(I + \sum_{k=j}^K H_k^\dagger S_k H_k)\right) \tag{5.11}$$

subject to

$$\mathrm{Tr}(S_j) \le P_j.$$

The corresponding dual problem to (5.11) is

$$\min_{\Gamma_j,\lambda_j} -\log\left|\Gamma_j - \frac{\mu_{j-1} - \mu_{j-2}}{\mu_j - \mu_{j-1}}\Gamma_{j-1}\right| - t + \mathrm{Tr}\left(\Gamma_j(I + \sum_{k=j+1}^K H_k^\dagger S_k H_k)\right) + \lambda_j P_j$$

subject to

$$\lambda_j I \ge H_j \Gamma_j H_j^\dagger$$
$$\Gamma_j \ge 0.$$

After the final (K^{th}) step, the overall dual problem is

$$\min_{\Gamma_1,\dots,\Gamma_K} \min_{\lambda_1,\dots,\lambda_K} \max_{P_1,\dots,P_K} -\log|\Gamma_1| - \left[\sum_{j=2}^{K}(\mu_j-\mu_{j-1})\log\left|\Gamma_j - \frac{\mu_{j-1}-\mu_{j-2}}{\mu_j-\mu_{j-1}}\Gamma_{j-1}\right|\right]$$

$$-\mu_K t + (\mu_K-\mu_{K-1})\mathrm{Tr}(\Gamma_K) + \sum_k(\mu_k-\mu_{k-1})\lambda_k P_k \tag{5.12}$$

subject to

$$\sum_k P_k = P$$
$$\lambda_k I \geq H_k\Gamma_k H_k^\dagger \ \forall\ k$$
$$P_k, \Gamma_k \geq 0\ \forall\ k.$$

Note that the dual problem (5.12) is linear in $\{P_1,\dots,P_K\}$ and in $\{\lambda_1,\dots,\lambda_K\}$.

The same analysis as for the K user MAC can be repeated to obtain the dual problem for (5.2) by replacing H_k with $N_{Dk}^{-1/2}H_{Dk}$. This dual problem is

$$\min_{N_{D1},\dots,N_{DK}} \min_{\lambda_1,\dots,\lambda_K} \min_{\Gamma_1,\dots,\Gamma_K} \max_{P_1,\dots,P_K} -\log|\Gamma_1| - \sum_{j=2}^{K}(\mu_j-\mu_{j-1})\log\left|\Gamma_j - \frac{\mu_{j-1}-\mu_{j-2}}{\mu_j-\mu_{j-1}}\Gamma_{j-1}\right|$$

$$-\mu_K t + (\mu_K-\mu_{K-1})\mathrm{Tr}(\Gamma_K) + \sum_k(\mu_k-\mu_{k-1})\lambda_k P_k$$

subject to

$$\sum_k P_k = P$$
$$\lambda_k N_{Dk} \geq H_{Dk}\Gamma_k H_{Dk}^\dagger \ \forall\ k$$
$$\Gamma_k \geq 0\ \forall\ k.$$

The above problem is linear in $\{N_{D1},\dots,N_{DK}\}$, so we can use Ky-Fan's min-max theorem [20] to obtain

$$\min_{\lambda_1,\dots,\lambda_K} \min_{\Gamma_1,\dots,\Gamma_K} \max_{P_1,\dots,P_K} \min_{N_{D1},\dots,N_{DK}} -\log|\Gamma_1| - \sum_{j=2}^{K}(\mu_j-\mu_{j-1})\log\left|\Gamma_j - \frac{\mu_{j-1}-\mu_{j-2}}{\mu_j-\mu_{j-1}}\Gamma_{j-1}\right|$$

$$-\mu_K t + (\mu_K-\mu_{K-1})\mathrm{Tr}(\Gamma_K) + \sum_k(\mu_k-\mu_{k-1})\lambda_k P_k \tag{5.13}$$

subject to

$$\sum_k P_k = P$$
$$\lambda_k N_{Dk} \geq H_{Dk}\Gamma_k H_{Dk}^\dagger \ \forall\ k$$
$$\Gamma_k \geq 0\ \forall\ k.$$

To obtain the final result, we show that (5.13) is at most (5.12). We show this by constructing a set of K MACs having $K+1$ users each. The first MAC has channel H_1 for user 1, channels $N_{Dj}^{-1/2}H_j$ for user j, $j \in \{2,\dots,K\}$, and channel $N_{D2}^{-1/2}H_2$ for user $K+1$. Let consider a point on the boundary of the capacity region of this channel, where users 1 and $K+1$ have priorities $\mu_1 = \mu_{K+1} = 1$, and user j has priority μ_j, $j \in \{2,\dots,K\}$. Extending the argument presented earlier, one can show that the dual problem to the boundary maximization problem for the point $(\mu_0, \mu_1, \mu_2, \ldots, \mu_K, \mu_{K+1}) = (0, 1, \mu_2, \ldots, \mu_K, 1)$ on the MAC with channels $H_1, N_{D2}^{-1/2} H_2, \ldots, H_K, N_{D2}^{-1/2} H_2$ is given by

$$\min_{\lambda_1,\ldots,\lambda_K} \min_{\Gamma_1,\ldots,\Gamma_K} \max_{P_1,\ldots,P_K} \min_{N_{D2},\ldots,N_{DK}} -\log|\Gamma_1| - \sum_{j=2}^{K} (\mu_j - \mu_{j-1}) \log \left| \Gamma_j - \frac{\mu_{j-1} - \mu_{j-2}}{\mu_j - \mu_{j-1}} \Gamma_{j-1} \right|$$

$$- \mu_K t + (\mu_K - \mu_{K-1}) \mathrm{Tr}(\Gamma_K) + \sum_k (\mu_k - \mu_{k-1}) \lambda_k P_k \tag{5.14}$$

subject to

$$\sum_k P_k = P$$
$$\lambda_k N_{Dk} \geq H_{Dk} \Gamma_k H_{Dk}^\dagger \text{ for } k = 2, \ldots, K$$
$$\lambda_1 I \geq H_1 \Gamma_1 H_1^\dagger$$
$$\lambda_1 N_{D2} \geq H_{D2} \Gamma_1 H_{D2}^\dagger$$
$$\Gamma_k \geq 0 \ \forall \ k.$$

Note that the objective functions of (5.13) and (5.14) are the same. The only difference between them lies in the constraints.

We now show that (5.13) is at most (5.14). Given that Γ_1 and λ_1 satisfy the constraints of (5.14), we choose N_{D1} for (5.13) to be

$$N_{D1} = \begin{bmatrix} I & \frac{H_1 \Gamma_1 H_{D2}^\dagger}{\lambda_1} \\ \frac{H_{D2} \Gamma_1 H_1^\dagger}{\lambda_1} & N_{D2} \end{bmatrix} \tag{5.15}$$

which satisfies the constraints on N_{D1} in (5.13). Also note that, for the point under consideration in the $K+1$ user MAC, it is best to allocate zero power to user $K+1$.

For the $(j-1)^{\text{th}}$ step, we consider a $K+1$ user MAC where users $1, \ldots, j-1$ have channels $H_1, \ldots, H_{j-1}$, users $j, \ldots, K$ have channels $N_{Dj}^{-1/2} H_{Dj}, \ldots, N_{DK}^{-1/2} H_{DK}$ and user $K+1$ has channel $N_{Dj}^{-1/2} H_{Dj}$. Suppose the priorities are μ_k for user k, $k \in \{1, \ldots, K\}$ and μ_{j-1} for User $K+1$. It can again be shown that the dual problem corresponding to this point in the capacity region is

$$\min_{\lambda_1,\ldots,\lambda_K} \min_{\Gamma_1,\ldots,\Gamma_K} \max_{P_1,\ldots,P_K} \min_{N_{Dj},\ldots,N_{DK}} -\log|\Gamma_1| - \sum_{j=2}^{K} (\mu_j - \mu_{j-1}) \log \left| \Gamma_j - \frac{\mu_{j-1} - \mu_{j-2}}{\mu_j - \mu_{j-1}} \Gamma_{j-1} \right|$$

$$- \mu_K t + (\mu_K - \mu_{K-1}) \mathrm{Tr}(\Gamma_K) + \sum_k (\mu_k - \mu_{k-1}) \lambda_k P_k \tag{5.16}$$

subject to

$$\sum_k P_k = P$$
$$\lambda_k N_{Dk} \geq H_{Dk} \Gamma_k H_{Dk}^\dagger \text{ for } k = j, \ldots, K$$
$$\lambda_k I \geq H_k \Gamma_k H_k^\dagger \text{ for } k = 1, \ldots, j-1$$
$$\lambda_{j-1} N_{Dj} \geq H_{Dj} \Gamma_{j-1} H_{Dj}^\dagger$$
$$\Gamma_k \geq 0 \ \forall \ k.$$

We choose

$$N_{D2} = \begin{bmatrix} I & \frac{H_2\Gamma_2 H_{D3}^\dagger}{\lambda_2} \\ \frac{H_{D3}\Gamma_2 H_2^\dagger}{\lambda_2} & N_{D3} \end{bmatrix} \tag{5.17}$$

where Γ_2 and λ_2 satisfy the constraints of (5.16) for $j = 3$. This choice of N_{D2} sets (5.14) equal to (5.16) for $j = 3$. Since (5.14) is a minimization over all N_{D2}, the optimal value of (5.16) for $j = 3$ is at least (5.14). In general, by using

$$N_{Dk} = \begin{bmatrix} I & \frac{H_k\Gamma_k H_{D(k+1)}^\dagger}{\lambda_k} \\ \frac{H_{D(k+1)}\Gamma_k H_k^\dagger}{\lambda_k} & N_{D(k+1)} \end{bmatrix} \tag{5.18}$$

where Γ_k and λ_k satisfy the constraints of (5.16) for $j = k+1$, we find that the optimum value of (5.16) for $j = k$ is at most that of (5.16) for $j = k+1$. Thus, the optimum value of (5.13) is at most that of (5.16) for $j = K$. But the optimum value of (5.16) for $j = K$ equals that of (5.12). Hence, we have the result that (5.13) is at most (5.12). □

Corollary: The worst case noise is given by:

$$\begin{aligned} N_{D1(i,i)} &= I \\ N_{D1(i,j)} &= \frac{H_i\Gamma_{i0}H_j^\dagger}{\lambda_{i0}} \text{ for } i < j \end{aligned} \tag{5.19}$$

where $N_{D1(i,j)}$ denotes the $(i,j)^{\text{th}}$ matrix entry of N_{D1}, and $\{\Gamma_{10}, \dots, \Gamma_{K0}\}$ and $\{\lambda_{10}, \dots, \lambda_{K0}\}$ are values of $\{\Gamma_1, \dots, \Gamma_K\}$ and $\{\lambda_1, \dots, \lambda_K\}$ corresponding to an extremizing solution of (5.12).

6. Conclusions

A new outer bound, termed the DSM bound, is obtained for the broadcast channel. It is shown that, for the vector broadcast channel constrained to Gaussian inputs, the DSM outer bound equals the Marton achievable region.

Appendix A
A Primal Problem and its Dual

We wish to find the dual problem for

$$\begin{aligned} &\max_{S_1,\dots,S_K} \log \left| A + \sum_k H_k^\dagger S_k H_k \right| \\ &+ \operatorname{Tr}\left(B(A + \sum_k H_k^\dagger S_k H_k) \right) \end{aligned} \tag{A.1}$$

subject to

$$\begin{aligned} &\sum_k \operatorname{Tr}(S_k) \leq P \\ &S_k \geq 0 \;\forall\, k. \end{aligned}$$

Consider first the optimization problem

$$\min_T - \log|T| - \operatorname{Tr}(BT) \tag{A.2}$$

subject to

$$\begin{aligned} &T \leq A + \sum_k H_k^\dagger S_k H_k \\ &\mathrm{Tr}(S_k) \leq P \\ &S_k \geq 0 \ \forall\ k. \end{aligned}$$

Note that (A.2) is obtained by negating the objective of (A.1). A Lagrangian for (A.2) is

$$\begin{aligned} L = &-\log|T| - \mathrm{Tr}(BT) + \mathrm{Tr}\left(\Gamma(T - A - \sum_k H_k^\dagger S_k H_k)\right) \\ &+ \lambda\left(\sum_k \mathrm{Tr}(S_k) - P\right) - \sum_k \mathrm{Tr}(\Psi_k S_k). \end{aligned} \tag{A.3}$$

The saddle point of this Lagrangian gives the optimum value of (A.2), and extremizing with respect to T and S_k gives us the dual problem. Rewriting (A.3), we have

$$\begin{aligned} L = &-\log|T| - \mathrm{Tr}(BT) + \mathrm{Tr}(\Gamma T) - \mathrm{Tr}(\Gamma A) - \lambda P \\ &- \sum_k \mathrm{Tr}\left(S_k(\Psi_k - \lambda I + H_k \Gamma H_k^\dagger)\right). \end{aligned} \tag{A.4}$$

Taking the derivative with respect to S_k and setting it to zero, we have $\Psi_k - \lambda I + H_k \Gamma H_k^\dagger = 0$. Taking the derivative with respect to T and setting it to zero, we have

$$-T^{-1} - B + \Gamma = 0.$$

Substituting the resulting Ψ_k and T into (A.4) we find that

$$L = \log|\Gamma - B| + m - \mathrm{Tr}(\Gamma A) - \lambda P. \tag{A.5}$$

The slackness variable Ψ_k is semidefinite, so the dual problem for (A.2) is

$$\max_{\Gamma,\lambda} \log|\Gamma - B| + m - \mathrm{Tr}(\Gamma A) - \lambda P \tag{A.6}$$

subject to

$$\begin{aligned} &\lambda I \geq H_k \Gamma H_k^\dagger \\ &\lambda, \Gamma \geq 0. \end{aligned}$$

Hence, the dual problem for (A.1) is

$$\min_{\Gamma,\lambda} -\log|\Gamma - B| - t + \mathrm{Tr}(\Gamma A) + \lambda P$$

subject to

$$\begin{aligned} &\lambda I \geq H_k \Gamma H_k^\dagger \\ &\lambda, \Gamma \geq 0. \end{aligned}$$

References

[1] T.M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 2-14, Jan. 1972.

[2] P.P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. 19, no. 2, pp. 197-207, March 1973.

[3] R.G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, vol. 10, no. 3, pp. 3-14, July-Sept. 1974.

[4] P.P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 20, no. 2, pp. 279-280, March 1974.

[5] E.C. van der Meulen, "A survey of multi-way channels in information theory: 1961-1976," *IEEE Trans. Inform. Theory*, vol. 23, no. 1, pp. 1-37, Jan. 1977.

[6] T.M. Cover, "Comments on broadcast channels," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2524-2530, Oct. 1998.

[7] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," submitted to *IEEE Trans. Inform. Theory*, July 2001.

[8] S.I. Gel'fand and M.S. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19-31, 1980.

[9] M.H.M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439-441, May 1983.

[10] W. Yu and J. Cioffi, "Trellis precoding for the broadcast channel," *Proc. IEEE Global Telecomm. Conf. (Globecom)*, pp. 1338-1344, San Antonio. TX, USA Nov. 25-29, 2001.

[11] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 3, pp. 306-311, May 1979.

[12] P. Viswanath and D. Tse, "Sum capacity of the multi-antenna broadcast channel," *Proc. IEEE Int. Symp. Inform. Theory*, pp. 497, Lausanne, Switzerland, June 30 - July 5, 2002.

[13] W. Yu and J. Cioffi, "Sum capacity of a Gaussian vector broadcast channel," *Proc. IEEE Int. Symp. Inform. Theory*, pp. 498, Lausanne, Switzerland, June 30 - July 5, 2002.

[14] S. Viswanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum capacity of Gaussian MIMO broadcast channels," submitted to *IEEE Trans. on Inform. Theory*, Aug. 2002.

[15] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," submitted to *IEEE Trans. on Inform. Theory*, Aug. 2002.

[16] P. Viswanath and D. N. C. Tse, "Capacity of the Multiantenna Gaussian Broadcast Channel" *Proc. DIMACS workshop on Signal Processing for Wireless Transmission*, Rutgers, NJ, Oct. 7-9 2002.

[17] H. Sato, "An outer bound to the capacity region of broadcast channels," *IEEE Trans. Inform. Theory*, vol. 24, no. 3, pp. 374-377, May 1978.

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

[19] D. Tse and S. Hanly, "Multi-Access Fading Channels: Part I: Polymatroid Structure, Optimal Resource Allocation and Throughput Capacities", IEEE Trans. Inf. Theory, Nov. 1998, pp.2796-2815.

[20] K. Fan, " Minimax Theorems", Proc. Nat. Acad. Sci, 39, 1953, pp.42-47.

[21] S. Boyd, " Convex Optimization", Lecture Notes, Stanford University.

Stanford University, Stanford, CA USA.
E-mail address: sriram@wsl.stanford.edu

Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA
E-mail address: gkr@bell-labs.com

Technion–Israel Institute of Technology, Haifa, Israel
E-mail address: sshlomo@ee.technion.ac.il

Stanford University, Stanford, CA USA.
E-mail address: syed@wsl.stanford.edu

Stanford University, Stanford, CA USA.
E-mail address: andrea@wsl.stanford.edu