The Cost of Uncorrelation and Non-Cooperation in MIMO Channels †

Tal Philosof and Ram Zamir

Dept. of Electrical Engineering - Systems, Tel-Aviv University

Tel-Aviv 69978, ISRAEL

talp,zamir@eng.tau.ac.il

Abstract

We investigate the capacity loss for using uncorrelated Gaussian input over a multiple-input multiple-output (MIMO) linear additive-noise (not necessarily Gaussian) channel. We upper bound the capacity loss by a universal constant, C^*, which is independent of the channel matrix and the noise distribution. For a single-user MIMO channel with n_t inputs and n_r outputs, $C^* = \min\{\frac{1}{2}, \frac{n_r}{2n_t}\log_2(1+\frac{n_t}{n_r})\}$ bit per input dimension (or $2C^*$ bit per transmit antenna per second per Hertz), under both total and per-input power constraints. If we restrict attention to (colored) Gaussian noise, then the capacity loss is upper bounded by a smaller constant, $C_G = \frac{n_r}{2n_t}\log_2(\frac{n_t}{n_r})$ for $n_r \geq n_t/e$, and $C_G = 0.265$ otherwise, and this bound is tight for certain cases of channel matrix and noise covariance. We also derive similar bounds for the sum-capacity loss in multi-user MIMO channels. This includes in particular uncorrelated Gaussian transmission in a MIMO multiple access channel, and "flat" Gaussian dirty-paper coding in a MIMO broadcast channel. In the context of wireless communication, our results imply that the benefit of beamforming and spatial water filling over simple isotropic transmission is limited. Moreover, the excess capacity of a point-to-point MIMO channel over the same MIMO channel in a multi-user configuration is bounded by a universal constant.

Index Terms

Multiple-input multiple-output (MIMO) channel, multiple-input multiple-output broadcast channel (MIMO-BC), multiple-input multiple-output multiple-access channel (MIMO-MAC), capacity loss, uncorrelation loss, non-cooperation loss, robust input.

I. Introduction

Shannon's channel capacity is given by the maximum mutual information over all possible input distributions,

$$C = C(W) = \max_{P\in\mathcal{P}} I(P, W), \tag{1}$$

†Parts of this work were presented at the 42th Annual Allerton Conference, University of Illinois, USA, October 2004, and at ISIT2005 Adelaide, Australia September 2005

where $I(P,W)$ is the mutual information associated with input P and channel W, and $\mathcal{P}$ is the set of allowed input distributions [1]. In transmission over power-constrained linear MIMO channels, the optimum input distribution in (1) depends on the power constraint, the channel matrix and the joint statistics of the additive noise. Therefore, knowledge of these parameters is required not only at the receiver (which is usually the case for slow fading channels) but also at the transmitter. In this paper we examine the capacity loss for using a fixed (channel independent) input, namely, uncorrelated Gaussian input, in both single user and multi user configurations. The term *capacity loss* means here the difference between the true capacity (or capacity-sum) of the channel and the mutual information achieved by uncorrelated Gaussian input. In the context of multi-antenna wireless communication, uncorrelated input amounts to isotropic transmission, as happens in many forms of space-time coding [2], [3], as opposed to channel optimized beamforming [4].

Our main motivation is the lack of channel state information (CSI) at the transmitter [5]. The classical model for transmission over channels with uncertainty is the *compound channel* [6], [7]. Assuming the channel can be any member from a family $\mathcal{W}$, and the input distribution is constrained to a set $\mathcal{P}$, the compound channel capacity is given by

$$\begin{aligned} C_{compound} &= \max_{P\in\mathcal{P}} \min_{W\in\mathcal{W}} I(P,W) && (2) \\ &= \min_{W\in\mathcal{W}} C(W) && (3) \end{aligned}$$

where the second equality holds by the Minmax Theorem whenever the sets $\mathcal{P}$ and $\mathcal{W}$ are convex [8, p.214], [6]. The resulting optimum input P^* guarantees rate of $C_{compound}$ no matter what the actual channel is, hence it represents a robust codebook design for transmission over a channel with uncertainty.

This approach is, however, quite pessimistic, as it is dominated by the worst channel in the family. For example, the compound capacity of the family of power constrained linear channels is *zero*, because the worst member in that class has zero channel gain and/or infinite noise power. One way to overcome this difficulty in the linear MIMO case is to assume that the channel matrix belongs to an "isotropic" set, and that the total noise power is bounded. For this restricted class, it was shown in [9] that the compound channel capacity is achieved using an i.i.d Gaussian input. A similar assumption is that the channels in $\mathcal{W}$ have some a-priori distribution, so one can look for an input which is good on the average [10], [5].

A more optimistic approach, very common in the areas of universal source coding [11] and universal channel decoding [12], [13], is that of *competitive optimality*. Instead of trying to guarantee a certain rate, we look for a robust input P^{**} which for any channel $W \in \mathcal{W}$ does not loose "too much" mutual information relative to the channel capacity, i.e.,

$$C(W) - I(P^{**},W) \leq Const. \quad \forall\ W, \tag{4}$$

where the universal constant in the right-hand side is hopefully small. Thus, we want to enjoy the possibility for transmission at higher rates if the channel happens to be good, but still use the same codebook generating distribution for all channels in the class. As we shall show, i.i.d. Gaussian distribution provides such a competitive input for linear MIMO channels.

A natural question to ask, though, is what is the operational meaning of mutual information if the transmitter does not know the channel? Limited feedback provides one possible answer; in slow fading channels, it is practically reasonable to assume that while the receiver cannot describe all channel parameters to the transmitter, it can inform it about the achievable rate or the "effective SNR". Thus, the transmitter can translate extra mutual information into extra coding rate. An alternative model where mutual information may be a meaningful measure even without feedback is that of rateless codes; here the effective communication time is determined by the receiver which "quits" the channel when it received enough information to decode [14], [15].

Although an input with a uniformly bounded capacity-loss as in (4) above does not exist in general [16], it does for *additive noise* channels. Previous work [17] considered the class of single-input single-output (SISO) power-constrained channels of the form

$$Y_i = X_i + N_i, \quad i = 1, \ldots n, \tag{5}$$

where the additive noise $\{N_i\}$ may have arbitrary distribution, possibly with memory. It was shown that the capacity-loss for using white (i.i.d.) Gaussian input $\mathbf{X}^*$ instead of the optimum power constrained input is bounded by *half a bit per channel use*:

$$C - \frac{1}{n} I(\mathbf{X}^*; \mathbf{X}^* + \mathbf{N}) \leq \frac{1}{2} \text{ bit}, \ \forall \, \mathbf{N}. \tag{6}$$

If we restrict attention to the class of *Gaussian* noise channels, the loss for not performing the water filling optimization is at most $\log_2(e)/2e \approx 0.265$ bit per channel use, and this bound can be achieved for some two-step noise spectrum [17].

The half a bit bound (6) is interesting in general, though not always useful. It is loose at low signal-to-noise ratio (SNR), when the capacity is smaller than half a bit; it is loose asymptotically at high SNR, in which case white Gaussian input is approximately optimal for any noise distribution; and it's loose when the noise is approximately white Gaussian (see Appendix A). In fact, it was conjectured in [17] that this bound is never tight, and the worst loss is somewhere between 0.265 and 0.5 bit.

The two bounds above can be easily extended to point-to-point communication over *symmetric* vector channels, i.e., MIMO channels with equal number of inputs and outputs: the capacity loss for using i.i.d. Gaussian input is always less than $1/2$ bit per dimension, and in the case of Gaussian noise (with arbitrary spatial correlation) is at most $\log_2(e)/2e \approx 0.265$ bit per dimension (where the latter loss is achieved for certain combinations of Gaussian noise and channel matrix). In the context of wireless communication, these results can be expressed as follows: if the number of transmit antennas is equal to the number of receive antennas, then the capacity loss for using "isotropic" transmission is at most 1 bit per antenna per second per Hertz (where each antenna represents two dimensions: in-phase (I) and quadrature (Q)). When restricting the background noise and interference to be Gaussian, the bound becomes $\approx 2 \times 0.265 = 0.53$ bit per antenna per second per Hertz.

Unfortunately, the extension of these bounds to *asymmetric* MIMO channels is not straightforward, as it is not clear if the right normalization is per output dimension or per input dimension. Such channels are very common in cellular communication, where the base station usually has more antennas than the mobile station. In the SISO case, this corresponds to a multi-rate channel (or a poly-phase filter channels), i.e., a channel where the input sampling rate is not equal to the output sampling rate.

As we show in this work, the behavior of the capacity loss in the asymmetric case is quite different. To get some insight, consider the two simple examples of a single input multiple output (SIMO) and a multiple input single output (MISO) channels. Gaussian input is clearly optimal for the former case. A SIMO channel is defined as

$$Y_i = X + N_i, \quad i = 1, \ldots, n_r,$$

where X satisfies the power constraint $EX^2 \leq P$, and the N_i's are independent Gaussian noises. In this case white Gaussian input is optimal, and the capacity-loss is *zero*. In contrast, optimum transmission over a MISO channel

$$Y = \sum_{i=1}^{n_t} h_i X_i + N, \quad N \sim \mathcal{N}(0, \sigma^2),$$

under the sum power constraint $\sum_{i=1}^{n_t} EX_i^2 \leq P$, is given by beamforming or "maximum ratio combiner", [18], [4],

$$X_i = \frac{h_i}{||\mathbf{h}||} X, \quad \text{with } X \sim \mathcal{N}(0, P),$$

where $\|\mathbf{h}\|$ is the norm of the channel coefficients vector, and the resulting capacity is

$$C = \frac{1}{2} \log \left(1 + \frac{||\mathbf{h}||^2 P}{\sigma^2} \right) \tag{7}$$

bit per input vector. I.i.d. Gaussian input $X_i \sim \mathcal{N}(0, P/n_t)$, on the other hand, gives mutual information of

$$I(X_1, \ldots, X_n; Y) = \frac{1}{2} \log \left(1 + \frac{||\mathbf{h}||^2 P}{n_t \sigma^2} \right). \tag{8}$$

Comparing (7) and (8) we see that at high SNR the capacity loss per input vector is $\frac{1}{2}\log_2(n_t)$, i.e., sub-linear in the number of inputs n_t but more than 0.265 bit per output.

In Section II we develop bounds for the capacity loss of i.i.d Gaussian input for general linear additive noise MIMO channels, which go roughly like $O\left(\frac{n_r}{n_t}\log(\frac{n_t}{n_r})\right)$ bit per channel input dimension (or $O\left(n_r \log(\frac{n_t}{n_r})\right)$ bit per input vector). In agreement with [17], the bound for general noise becomes $\frac{1}{2}$ bit per input (or 1 bit per transmit antenna per second per Hertz) for $n_r \geq n_t$, and it is apparently not tight. If we restrict attention to Gaussian noise but still allow arbitrary correlation structure, the bound becomes $\frac{\log_2(e)}{2e}$ bit per input for $n_r = n_t$, and it is sharp, i.e, it can be approached for certain channel matrices and (Gaussian) noise. Section II also shows that the same bounds hold for the case of a per-antenna power constraint [19].

Section III extends the discussion to multi-user configurations: the MIMO multiple access channel (MAC) and the MIMO broadcast channel (BC). We consider two types of rate-sum losses: (i) loss due to separate processing by the users (called the "cost of non-cooperation"), and (ii) loss due to using uncorrelated Gaussian inputs. It turns out that bounds similar to those developed for the single user case apply also to these types of capacity losses. The formulation is, however, different in the two channel configurations. In the MIMO-MAC, a single constant bounds the combined effect of both types of capacity loss. In fact, the problem is equivalent to a MIMO point-to-point channel with a per-antenna power constraint.

The case of MIMO-BC differs in two aspects. First, although the input distribution is channel independent, the transmitter must know the exact channel matrix in order to achieve the promised rates [20]. Hence the meaning of robustness is weaker in this case, and is mainly about eliminating the need to perform complex water-filling computations in real-time [21]. Secondly, each type of loss in the BC case is bounded by a different constant. The loss for using white Gaussian transmission is bounded by a constant of the form $O(\frac{n_r}{n_t}\log(\frac{n_t}{n_r}))$, while the loss for the users being unable to jointly decode their receptions is bounded by a constant of the form $O(\frac{n_t}{n_r}\log(\frac{n_r}{n_t}))$ (i.e, with the roles of n_t and n_r switched). Finally, Section IV concludes the paper and adds some remarks. Preliminary version of these results appeared in [22], [23]

II. The Single User (Point to Point) Case

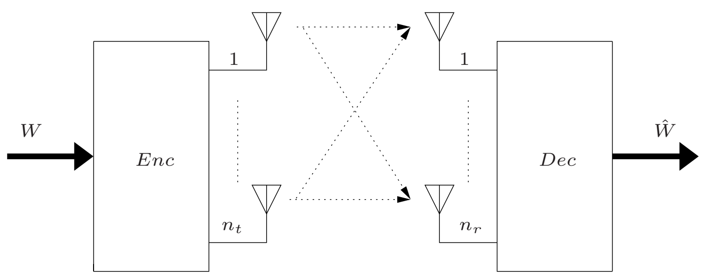

Fig. 1. The MIMO point-to-point channel.

Consider the real valued MIMO channel shown in Figure 1

$$\mathbf{Y} = H\mathbf{X} + \mathbf{N}, \tag{9}$$

where for the sake of illustration inputs and outputs are drawn as antennas. Here $H \in \mathbb{R}^{n_r \times n_t}$ is the channel matrix, n_t and n_r are the number of inputs and outputs respectively, $\mathbf{X} \in \mathbb{R}^{n_t}$ and $\mathbf{Y} \in \mathbb{R}^{n_r}$ are the transmitted and received vectors, respectively, and $\mathbf{N} \in \mathbb{R}^{n_r}$ is a general (not necessarily Gaussian nor component-wise independent) additive noise. As we shall discuss after presenting our main Theorem below, the generalization of our results to complex channel is straight forward. This is a single user (point to point) setting in the sense that a single transmitter controls all channel inputs $X_1 \dots X_{n_t}$ and a single receiver observes all channel outputs $Y_1 \dots Y_{n_r}$. The sum power constraint is

$$tr(R_x) = E||\mathbf{X}||^2 \leq P, \tag{10}$$

where tr is the trace operator and $R_x = E\mathbf{X}\mathbf{X}^\mathbf{T}$ is the covariance matrix of $\mathbf{X}$.

A. *Capacity Loss for Arbitrary Noise*

The capacity in bit per input dimension of the MIMO channel (9) is given by

$$C(H,\mathbf{N},P) \triangleq \frac{1}{n_t} \sup_{\mathbf{X}:E||\mathbf{X}||^2 \leq P} I(\mathbf{X}; H\mathbf{X}+\mathbf{N}), \tag{11}$$

where the maximization is over all possible joint distributions of the input vector $\mathbf{X}$ satisfying the power constraint (10). We shall assume in the sequel that $C(H,\mathbf{N},P)$ is finite. In view of (1), for a given number of inputs n_t and outputs n_r, we shall be interested in the worst capacity loss

$$L_{max}(n_t,n_r) \triangleq \sup_{H,\mathbf{N}} \left\{ C(H,\mathbf{N},P) - \frac{1}{n_t} I(\mathbf{X}^*; H\mathbf{X}^*+\mathbf{N}) \right\} \tag{12}$$

for using zero mean i.i.d Gaussian input $\mathbf{X}^*$, when the maximization is over all channel matrices and noise distributions. Although a complete characterization of L_{max} is still missing, we shall develop uniform upper and lower bounds which in some cases are quite tight.

Following the derivation in the SISO case [17], we start with a lemma regarding a general input.

Lemma 1: Let $\mathbf{X}'$ be an arbitrary input to the MIMO channel (9). Then

$$C(H,\mathbf{N},P) - \frac{1}{n_t} I(\mathbf{X}'; H\mathbf{X}'+\mathbf{N}) \leq C^*(H,\mathbf{X}'), \tag{13}$$

where

$$C^*(H,\mathbf{X}') \triangleq \frac{1}{n_t} \sup_{\mathbf{Z}:E||\mathbf{Z}||^2 \leq P} I(\mathbf{Z}; H\mathbf{Z}+H\mathbf{X}'). \tag{14}$$

Proof: The proof is given in Appendix B. □

Note that the quantity $C^*(H,\mathbf{X}')$ is independent of the distribution of the noise $\mathbf{N}$, hence it bounds the capacity loss for using input $\mathbf{X}'$ for any additive noise channel with channel matrix H. The worst capacity loss for using $\mathbf{X}'$ is bounded from above by the maximum of this quantity over all channel matrices, i.e, $\sup_H C^*(H,\mathbf{X}')$. The smallest bound is achieved by using an input $\mathbf{X}'$ that minimizes the maximum value of $C^*(H,\mathbf{X}')$ over H, i.e, $\mathbf{X}'$ which achieves

$$C^* \triangleq \inf_{\mathbf{X}':E||\mathbf{X}'||^2 \leq P} \sup_H C^*(H,\mathbf{X}'). \tag{15}$$

Lemma 2: The minimum in (15) is achieved by $\mathbf{X}' = \mathbf{X}^*$, where $\mathbf{X}^* \sim \mathcal{N}(0, \frac{P}{n_t} I_{n_t})$ is an i.i.d Gaussian vector and I_{n_t} is the $n_t \times n_t$ dimensional identity matrix.

Proof: The proof is given in Appendix C. □

Using these two lemmas we arrive at the following theorem.

Theorem 1: **(Uniform bound for arbitrary noise)** For any noise $\mathbf{N}$ and $n_r \times n_t$ channel matrix H, the capacity loss of uncorrelated Gaussian input $\mathbf{X}^* \sim \mathcal{N}(0, \frac{P}{n_t} I_{n_t})$, to the channel (9), is upper bounded by

$$C(H,\mathbf{N},P) - \frac{1}{n_t} I(\mathbf{X}^*; H\mathbf{X}^*+\mathbf{N}) \leq C^*, \tag{16}$$

where

$$C^* = \begin{cases} \frac{n_r}{2n_t} \log_2 (1+\frac{n_t}{n_r}), & n_r \leq n_t \\ \frac{1}{2}, & n_r \geq n_t \end{cases} \tag{17}$$

bit per input dimension.

Proof: The proof is given in Appendix D. □

This bound may also be expressed as $C^* = \min\{1, \frac{n_r}{n_t}\log_2(1+\frac{n_t}{n_r})\}$ $[\text{bit/sec}\cdot\text{Hz}\cdot\text{transmit antenna}]$ for a wireless channel with n_t transmit antennas and n_r receive antennas; this follows since each antenna amounts to a complex input/output, i.e., it is not better than two real inputs/outputs. In this case, i.i.d input amounts to *isotropic transmission*, as opposed to a channel optimized beamforming.

In Figure 2, we illustrate the behavior of $C^* = C^*(n_t, n_r)$ for fixed n_t. For $n_r \geq n_t$, i.e, when the number of receive antennas is the same or greater than the number of transmit antennas, the loss for using i.i.d Gaussian input is at most $\frac{1}{2}$ bit per channel use per transmit antenna, similarly to the result in [17]. However, for $n_r < n_t$, i.e, when there are less receive antennas than transmit antennas, the bound is $\frac{n_r}{2}\log_2(1+\frac{n_t}{n_r})$ bit per input vector, which is less than $\frac{n_t}{2}$ bits but more than $\frac{n_r}{2}$ bits. Therefore, the bound on the loss in this case is worse than half a bit per degree of freedom, i.e, greater than $\frac{1}{2}\min(n_t, n_r) = \frac{n_r}{2}$.

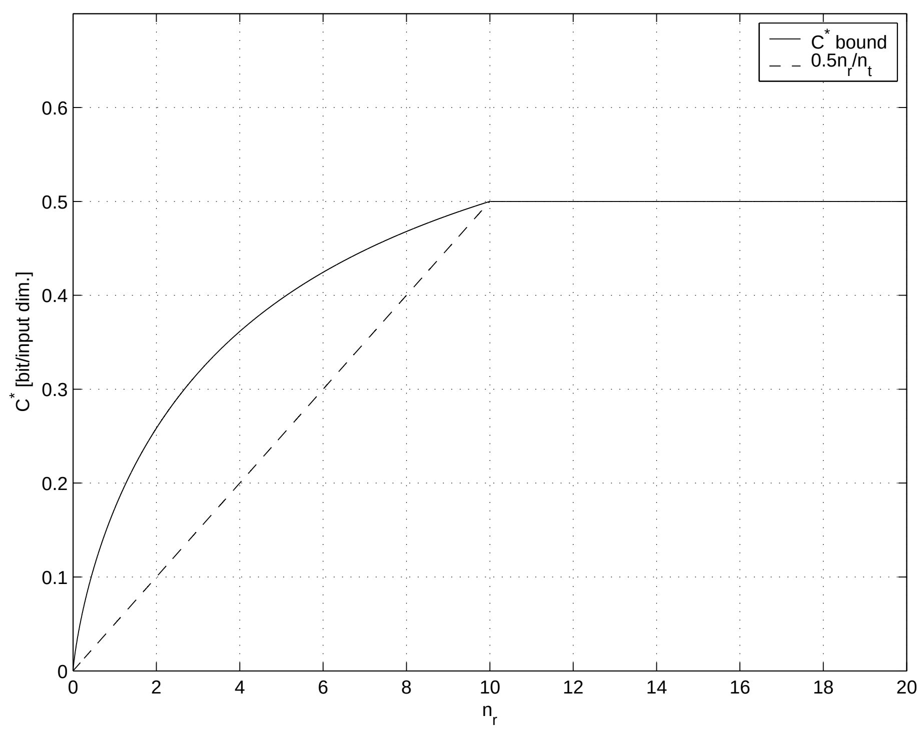

Fig. 2. The Capacity loss for a MIMO channel with arbitrary noise for $n_t = 10$.

We believe the bound of Theorem 1is not tight, and the distance from the true curve of the *worst* loss of (12) is still unknown. Nevertheless, for the case of Gaussian noise, we show below a tighter bound which can actually be achieved, and hence can be considered as a *lower bound* for the worst loss of (12).

B. Capacity Loss for Gaussian Noise

The bound in Theorem 1 takes into account two effects. One is the loss of "shaping gain" due to Gaussian input being mismatched to the higher order statistics of the noise, and the other is the loss of "beamforming gain" (or the "isotropic transmission loss") due to uncorrelated input being mismatched to the matrix H and to the noise covariance. In this section we focus on the second effect by restricting attention to channels with Gaussian noise.

Theorem 2: **(Bound for Gaussian noise)** If the additive noise $\mathbf{N}_G$ in the MIMO channel (9) is Gaussian with arbitrary correlation, then

$$C(H, \mathbf{N}_G, P) - \frac{1}{n_t} I(\mathbf{X}^*; H\mathbf{X}^* + \mathbf{N}_G) \le C_G^*, \tag{18}$$

where

$$C_G^* = \begin{cases} \frac{n_r}{2n_t} \log_2\left(\frac{n_t}{n_r}\right), & n_r \le [[n_t/e]] \\ g(n_t), & n_r \ge [[n_t/e]] \end{cases} \tag{19}$$

bit per input dimension; here $g(\cdot)$ is given by the integer maximization

$$g(n_t) \triangleq \frac{1}{2n_t} \max_{n \in \mathbb{Z}} \left\{ n \log_2 \left(\frac{n_t}{n} \right) \right\}, \tag{20}$$

and $[[n_t/e]]$ is the $n \in \mathbb{Z}$ that maximizes (20)[1]. Equality in (18) can be arbitrary approached for any channel matrix with rank $\min\{n_r, [[n_t/e]]\}$ (i.e, full rank for $n_r \le [[n_t/e]]$, and rank $[[n_t/e]]$ otherwise) in the limit of high SNR.

Proof: The proof is given in Appendix E. □

A simple example for equality in Theorem 2 is the MISO case discussed in the Introduction, when the loss at high SNR is $\frac{1}{2n_t}\log(n_t)$ bit per dimension; see (7) and (8). In the general MIMO case, equality in Theorem 2 is achievable if we choose the rank of the channel matrix H (the number of effective non-zero sub-channels) as close as possible to the integer $[[n_t/e]]$ which maximizes (20). Since the rank of H is an integer in the range $0, \ldots, \min\{n_r, n_t\}$, we arrive at the two cases of (19). For example, assume that $n_r \le [[n_t/e]]$ and the channel matrix H has n_r identical non-zero values on the main diagonal and zero elsewhere. Assume further that the noise is white and weak (the SNR is high). Thus, the capacity per input vector is $\approx \frac{n_r}{2}\log_2\frac{P}{n_r}$ and it is achieved by dividing the power P equally among the n_r non-zero sub-channels. On the other hand, the i.i.d. input $\mathbf{X}^*$ spreads its power evenly among all the n_t transmit antennas, hence its rate is $\approx \frac{n_r}{2}\log_2\frac{P}{n_t}$, and the capacity loss is $\approx \frac{n_r}{2}\log_2\frac{n_t}{n_r}$ bit per input vector as claimed in the Theorem.

Figure 3 illustrates the bound C_G^* with respect to the number of channel outputs n_r (receive antennas), compared to the bound C^* of Theorem 1. It can be seen that the bound for Gaussian noise is strictly less than that for arbitrary noise. Yet, for $n_r \ll n_t$ the bounds are quite close, so they provide a tight characterization for the worst loss L_{max} in (12).

If we omit the integer restriction in (20), then the maximum in the second case of (19) is achieved by n_t/e, and C_G^* takes a simpler form, which is slightly less tight for small number of transmit antennas:

$$C_G = \begin{cases} \frac{n_r}{2n_t} \log_2\left(\frac{n_t}{n_r}\right), & n_r \le n_t/e \\ \frac{1}{2}\frac{\log_2(e)}{e}, & n_r \ge n_t/e. \end{cases} \tag{21}$$

[1]Numerically it was found that for any practical purpose ($n_t \le 10^6$) the maximum is achieved by $[n_t/e]$, i.e., the integer nearest to n_t/e. However, a proof for any n_t is missing.

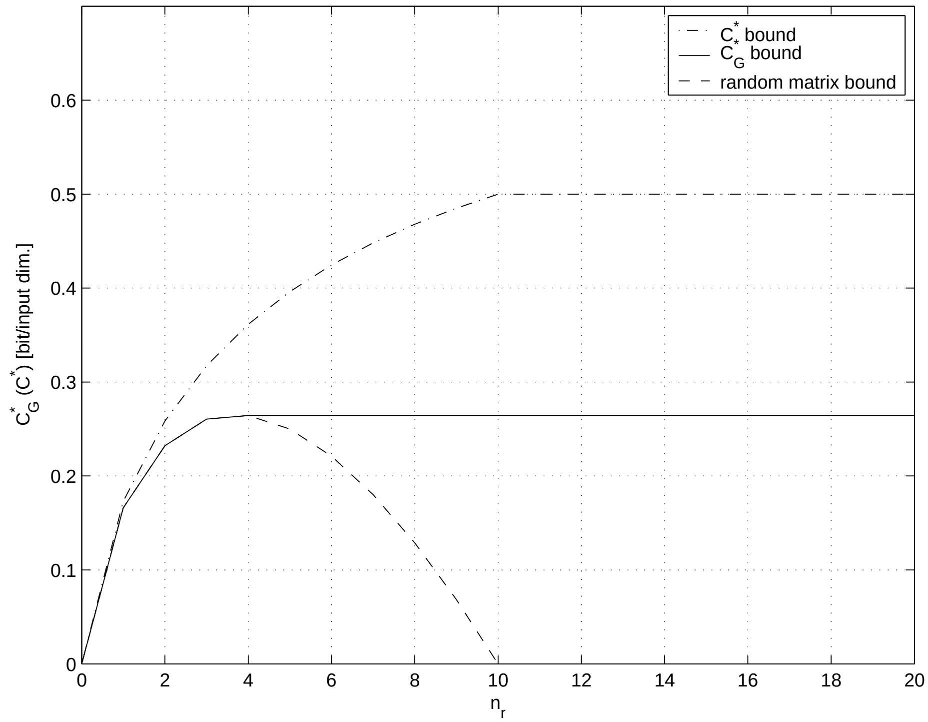

Fig. 3. The bounds on the capacity loss versus the number of outputs n_r for $n_t = 10$.

From (21) it is apparent that for a Gaussian MIMO channel the capacity loss of i.i.d Gaussian input is at most $\frac{\log_2(e)}{2e} \simeq 0.265$ bit per channel use per input dimension, similarly to the result of [17].

In order to understand the behavior of the bound C_G^* for small number of transmit antennas n_t, we have drawn in Figure 4 the bounds $n_t C_G^*$ and $n_t C_G$ (in bits per input vector) with respect to n_t for $n_r = 2$. We observe that the graph of $n_t C_G$ in (21) is monotonic and it has two regions: (i) for $n_t \leq e n_r$ it is linear with n_t, (ii) for $n_t \geq e n_r$ it is logarithmic with n_t. The graph of $n_t C_G^*$ has a ripple in region (i) due to the integer constraint on the rank of H; it increases logarithmically inside the intervals where $[[n_t/e]]$ is fixed, and it "jumps" between these intervals.

The bound above holds for an arbitrary channel matrix and noise correlation. In some scenarios we may have some prior knowledge regarding the channel parameters. In these cases the capacity loss may be smaller, or we can find a better robust input. One example is when the SNR is high and the matrix H is full rank. A tight characterization for the loss in this case is given in the following Proposition

Proposition 1: **(Loss for full rank channel matrix)** If the channel matrix H in (9) has full rank, and the noise is Gaussian, then the capacity loss is approximately

$$C(H, N_G, P) - \frac{1}{n_t} I(\mathbf{X}^*; H\mathbf{X}^* + \mathbf{N}_G) \simeq \begin{cases} \frac{n_r}{2n_t} \log_2\left(\frac{n_t}{n_r}\right), & n_r \leq n_t \\ 0, & n_r \geq n_t \end{cases} \tag{22}$$

bit per input dimension, and the approximation becomes tight as $P \to \infty$.

Proof: The proof is given in Appendix F. □

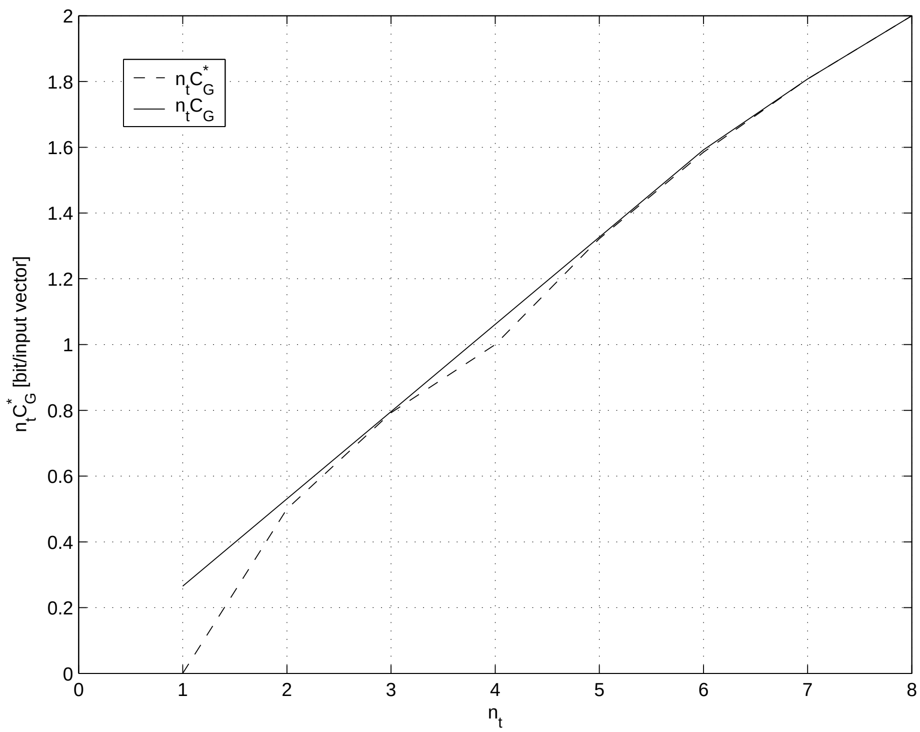

Fig. 4. The bounds (19) and its approximation (21) as function of n_t for $n_r = 2$.

It should be noted that the capacity loss in the region $n_r \geq n_t$ approaches zero because for high SNR and full rank channel matrix, i.e, when $rank(H) = n_t$, the optimum (water filling) input becomes i.i.d Gaussian input becomes the optimal.

In many communication scenarios the entries of the channel matrix are drawn randomly and independently, thus the channel matrix has full rank with probability one. It follows from the discussion above, that at high SNR the capacity loss is equal with high probability to that in the deterministic case of (22). The dashed curve in Figure 3 shows this capacity loss. For $n_r > [[n_t/e]]$, the loss in the random channel case is clearly much lower than the worse case of Theorem 2 because the number of degrees of freedom in the channel is most likely larger than the worst value of $[[n_t/e]]$. As for the general SNR case, we can still make a high probability claim in the random matrix case if the number of receive antennas is much larger than the number of transmit antennas. By the weak law of large numbers, vectors with i.i.d components become orthogonal with equal norm in the limit of infinite dimension. Hence, the columns of the channel matrix become orthonormal up to a scale factor, and i.i.d Gaussian input becomes optimal. Thus, for any SNR the capacity loss goes to zero as the number of receive antennas increases to infinity.

It follows from Theorem 2 that the bound C_G^* on the capacity loss is tight, in the sense that there exists a “worst” channel matrix and noise correlation such that the loss is exactly C_G^*. On the other hand, in Theorem 1 we bounded from above the capacity loss for any linear additive noise MIMO channel by C^*. Hence, the maximal capacity loss L_{max} of (12) must be between these two constants (C_G^* and C^*).

Corollary 1: The worst capacity loss of i.i.d Gaussian input (12) is limited to the region

$$C_G^* \leq L_{max} \leq C^*. \tag{23}$$

C. Per-Antenna (Subset) Power Constraint

We now show that the bounds above apply also to the case of individual power constraints [19]. Specifically, assume that the transmit antennas are partitioned into m subsets with n_{ti} antennas in subset i, where $\sum_{i=1}^{m} n_{ti} = n_t$. In this scenario we assume a separate power constraint P_i for each subset i. If $n_{ti} = 1$ for all i, then each antenna has an individual power constraint. For $m = 1$ there is only one subset, and the problem reduces to the sum power constraint considered in Theorems 1 and 2 above. We investigate the capacity loss of Gaussian input which is uncorrelated and uniform within the antennas in each subset. This is in contrast to the optimal input, which is not necessarily Gaussian (if the noise is not Gaussian) and it allows any correlation and any power allocation between antennas in the subset as long as the power constraint is satisfied. The motivation for this setting is two fold: (i) it generalizes the sum power constraint considered in Theorems 1 and 2, and (ii) it can easily be applied to the case of MIMO-MAC for which the subsets represent the users, as we shall see in Section III-A.

Specifically, assume the channel model

$$\mathbf{Y} = H\mathbf{X} + \mathbf{N}, \tag{24}$$

like in (9), where $\mathbf{X} = [\mathbf{X}_1^T \ \mathbf{X}_2^T \ \ldots \ \mathbf{X}_m^T]^T \in \mathbb{R}^{n_t \times 1}$ and the i-th subset channel input is $\mathbf{X}_i \in \mathbb{R}^{n_{ti} \times 1}$. The power constraints are given by

$$E||\mathbf{X}_i||^2 \leq P_i, i = 1 \cdots m. \tag{25}$$

The capacity is given in this case by

$$C(H, \mathbf{N}, \mathbf{P}) \triangleq \frac{1}{n_t} \sup_{\{\mathbf{X}: E||\mathbf{X}_i||^2 \leq P_i, i=1,\ldots m\}} I(\mathbf{X}; H\mathbf{X} + \mathbf{N}) \tag{26}$$

bit per input dimension, where $\mathbf{P} = [P_1, \ldots, P_m]^T$. The uncorrelated piecewise-uniform input which we want to examine is denoted by $\mathbf{X}^D$; it has a diagonal covariance matrix, with diagonal elements which are constant over subsets. Thus, $\mathbf{X}^D \sim \mathcal{N}(\mathbf{0}, R_{x^D})$, where

$$R_{x^D} = diag\left\{ \underbrace{\frac{P_1}{n_{t1}}, \ldots, \frac{P_1}{n_{t1}}}_{n_{t1}}, \underbrace{\frac{P_2}{n_{t2}}, \ldots, \frac{P_2}{n_{t2}}}_{n_{t2}}, \ldots, \underbrace{\frac{P_m}{n_{tm}}, \ldots, \frac{P_m}{n_{tm}}}_{n_{tm}} \right\}. \tag{27}$$

Lemma 3: **(Bounds for per-antenna power constraint)** The capacity loss of uncorrelated piecewise-uniform Gaussian input in the MIMO channel (24) under individual (per-subset) power constraint (25) is bounded by the same universal constants as in the case of a sum power constraint. That is, for arbitrary noise

$$C(H, \mathbf{N}, \mathbf{P}) - \frac{1}{n_t} I(\mathbf{X}^{\mathbf{D}}; H\mathbf{X}^{\mathbf{D}} + \mathbf{N}) \leq C^*, \tag{28}$$

and for Gaussian noise

$$C(H, \mathbf{N}_G, \mathbf{P}) - \frac{1}{n_t} I(\mathbf{X}^{\mathbf{D}}; H\mathbf{X}^{\mathbf{D}} + \mathbf{N}_G) \leq C_G^*, \tag{29}$$

bit per input dimension, where $C^* = C^*(n_t, n_r)$ and $C_G^* = C_G^*(n_t, n_r)$ are given in (17) and (19), respectively. Equality in (29) holds in the limit of high SNR for any channel matrix whose rank is $\min\{n_r, [[n_t/e]]\}$.

Proof: The proof is given in Appendix G.

□

Interestingly, the bounds of Lemma 3 are independent of the partition into subsets and the power constraints.

D. Multi-rate SISO inter-symbol interference channels

As discussed in the Introduction, the *worst* capacity loss for SISO additive colored Gaussian noise channels is at least $\log_2(e)/2e$ bit per channel use, as shown in [17]. In view of the duality between frequency, time and spatial domains, the above results can be applied to extend the result of [17] to multi-rate SISO inter-symbol interference (ISI) channels with additive colored Gaussian noise. Consider a complex time domain channel where its (complex) input sampling rate is qf_s and its (complex) output sampling rate is pf_s, where q and p are integers and f_s is some common reference sampling rate. The system consists of a p-factor up-sampling block (interpolator), factored by a discrete time channel at sampling rate of qpf_s, and a q-factor down-sampling block (decimator). The output is contaminated by i.i.d (complex) Gaussian noise at sampling rate pf_s.

The duality between spatial domain and frequency domain takes the following form: the input frequencies and output frequencies in this SISO ISI channel play the role of the multiple inputs and multiple outputs in a MIMO channel. Using this duality, and by the linear relation between input and output, the multi-rate ISI channel may be modelled as a complex input complex output MIMO channel $\mathbf{Y} = H\mathbf{X} + \mathbf{N}$, where $\mathbf{H} \in \mathbb{C}^{\mathbf{r}\times\mathbf{t}}$. Since the frequency domain is continuous, the channel matrix has infinite dimensions ($t, r \to \infty$) with fixed ratio, i.e, $\frac{r}{t} = \frac{p}{q}$. Applying the result of Theorem 2 for the case where the input and output are complex and due to the infinite dimensions of H (the integer effects in (19) are negligible), we conclude that the capacity loss of i.i.d Gaussian input is bounded by

$$C_G^*(q,p) = \begin{cases} \frac{p}{q}\log_2\left(\frac{q}{p}\right), & p \le q/e \\ \frac{\log_2 e}{e}, & p \ge q/e \end{cases} \tag{30}$$

bit per second per Hertz. This bound is tight and can be arbitrarily approached by a channel filter whose equivalent matrix H has rank $\min\{r, t/e\}$ in the limit of high SNR. For $p = q$ (a single rate channel) the loss is bounded by $\frac{\log_2 e}{e}$ bit per second per Hertz in agreement with [17, Proposition 1].

III. The Multi-User Case

In this section we present upper bounds on the sum-capacity loss for using uncorrelated Gaussian input over a MIMO multiple access channel (MAC) (with arbitrary noise), and for using i.i.d Gaussian input over a *Gaussian* MIMO broadcast channel (BC). As discussed in the Introduction, we have two types of sum-rate losses in these scenarios: (i) the uncorrelation loss due to the transmission being uncorrelated between the inputs, and (ii) the non-cooperation loss due to lack of cooperation between the users (see, e.g., [24]). The latter loss is measured with respect to the same channel in a point-to-point configuration. We emphasize again that in the MIMO-BC case the implication of the bound is weaker, as the transmitter is assumed to know the exact channel matrix in order to achieve the target mutual information.

A. The Sum-Capacity Loss for a MIMO Multiple Access Channel

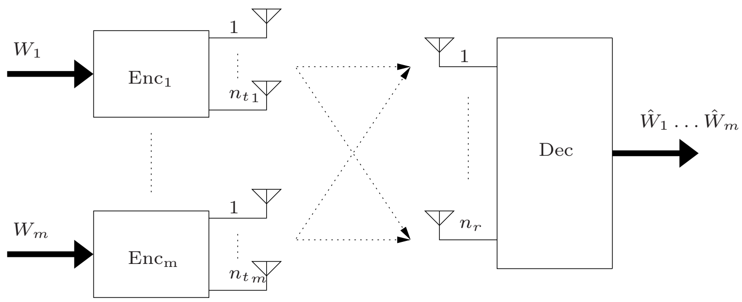

Fig. 5. The MIMO-MAC model.

Consider the m users MIMO-MAC shown in Figure 5. User i has n_{ti} transmit antennas, so the total number of transmit antennas over all users is

$$n_t = \sum_{i=1}^{m} n_{ti}. \tag{31}$$

The channel output is given by

$$\mathbf{Y} = \sum_{i=1}^{m} H_i \mathbf{X}_i + \mathbf{N}, \tag{32}$$

where $\mathbf{Y} \in \mathbb{R}^{n_r \times 1}$, $\mathbf{N} \in \mathbb{R}^{n_r \times 1}$ is the noise vector (which is not necessarily Gaussian) and where $\mathbf{X}_i \in \mathbb{R}^{n_{ti} \times 1}$ and $H_i \in \mathbb{R}^{n_r \times n_{ti}}$ for $i = 1, \dots m$, are the channel inputs and channel matrices, respectively. Each user has a power constraint $E||\mathbf{X}_i||^2 \leq P_i,\ i = 1, \dots, m$. The same model (with Gaussian noise) was considered in [25] for information rate maximization using an "iterative water filling algorithm".

Let $S \subseteq \{1, 2 \dots, m\}$. Let S^c denote the complementary subset, and let $\mathbf{X}(S) = \{\mathbf{X}_i : i \in S\}$. The capacity region of a general MIMO-MAC [1] is the closure of the convex hull of the rate vectors satisfying,

$$\sum_{i \in S} R_i \leq I(\mathbf{X}(S); \mathbf{Y} | \mathbf{X}(S^c)),\ \forall S \subseteq \{1, 2, \dots, m\}, \tag{33}$$

for some distribution $P(\mathbf{X}) = \prod_{i=1}^{m} P(X_i)$. Even in the Gaussian noise case there is no closed form expression for this region [25].

The *sum* capacity of the MIMO-MAC (32) can, nevertheless, be expressed in a simple form

$$\begin{aligned} C_{MAC}^{sum}(H, \mathbf{N}, \mathbf{P}) &\triangleq \sup_{\{\mathbf{X}: E||\mathbf{X}_i||^2 \leq P_i, E\mathbf{X}_i \mathbf{X}_j^T = 0, \forall i \neq j\}} I(\mathbf{Y}; \mathbf{X}_1, \dots, \mathbf{X}_m) \\ &= \sup_{\{\mathbf{X}: E||\mathbf{X}_i||^2 \leq P_i, E\mathbf{X}_i \mathbf{X}_j^T = 0, \forall i \neq j\}} I(\mathbf{X}; H\mathbf{X} + \mathbf{N}), \end{aligned} \tag{34}$$

bit per (vector) channel use. Here $\mathbf{X} = [\mathbf{X}_1^T \ \mathbf{X}_2^T \ \ldots \ \mathbf{X}_m^T]^T \in \mathbb{R}^{n_t \times 1}$, where $\mathbf{X}_i$ and $\mathbf{X}_j$ are independent for $i \neq j$ ($i, j \in \{1 \ldots m\}$), and H is the equivalent channel matrix, that is $H = [H_1 \ H_2 \ \ldots \ H_m]$.

Compare the MIMO-MAC (32) with a point-to-point MIMO channel with a per subset power constraints (25), assuming that both have the same channel matrix H. Assume further that the partition into input subsets of the corresponding MIMO point-to-point channel is identical to the partition of the MIMO-MAC users with the same individual power constraints. Since the power allocation between subsets (users) is inflexible in both problems, the loss in MIMO-MAC is due on one hand to users being unable to cross-correlate their transmissions which we call "non-cooperation loss", and on the other hand due to restricting each user to i.i.d (Gaussian) transmission over its own inputs, which we call "uncorrelation loss". Specifically, the non-cooperation loss $L_{non-coop}$ is defined as

$$L_{non-coop} \triangleq C_{P2P}(H, \mathbf{N}, \mathbf{P}) - C_{MAC}^{sum}(H, \mathbf{N}, \mathbf{P}), \tag{35}$$

bit per channel use, where

$$C_{P2P}(H, \mathbf{N}, \mathbf{P}) = \sup_{\{\mathbf{X}: E||\mathbf{X}_i||^2 \leq P_i, i=1,\ldots m\}} I(\mathbf{X}; H\mathbf{X} + \mathbf{N}), \tag{36}$$

is the capacity in bit per channel use of the corresponding MIMO point-to-point channel with per subset power constraint (26), and where the subscript P2P was added to distinguish between the two cases. The uncorrelation loss $L_{un-corr}$ is defined as

$$L_{un-corr} \triangleq C_{MAC}^{sum}(H, \mathbf{N}, \mathbf{P}) - I(\mathbf{X}^D; H\mathbf{X}^D + \mathbf{N}), \tag{37}$$

bit per channel use, where $\mathbf{X}^D$ is given in (27). The total loss (the sum of these two losses) is the information loss of MIMO-MAC with uncorrelated input with respect to the corresponding MIMO point-to-point capacity, i.e.,

$$L_{un-corr} + L_{non-coop} = C_{P2P}(H, \mathbf{N}, \mathbf{P}) - I(\mathbf{X}^D; H\mathbf{X}^D + \mathbf{N}). \tag{38}$$

Since the rate achieved by uncorrelated input $\mathbf{X}^D$ is the same for MIMO-MAC and MIMO point-to-point channel with per subset power constraint, from Lemma 3 we have the following bounds.

Theorem 3: **(Bounds for MIMO-MAC)** The total loss, i.e, the sum of uncorrelation loss (37) and non-cooperation loss (35), for MIMO-MAC with arbitrary noise is bounded by

$$L_{un-corr} + L_{non-coop} \leq n_t C^*. \tag{39}$$

For MIMO-MAC with Gaussian noise $\mathbf{N}_G$, we have a tighter bound:

$$L_{un-corr} + L_{non-coop} \leq n_t C_G^*. \tag{40}$$

In both cases the bounds are in bit per (vector) channel use, and C_G^* and C^* are given in (19) and (17), respectively, with $n_t = \sum_{i=1}^m n_{ti}$. Equality in (40) holds for any channel matrix with rank $\min\{n_r, [[n_t/e]]\}$ in the limit of high SNR.

Proof: The proof is given in Appendix H. □

Since the losses $L_{un-corr}$ and $L_{non-coop}$ are non negative values, the bounds apply to each of them separately. For a fixed n_t, the larger the number of users, the larger is the portion of the total loss due to non-cooperation; the larger the number of antennas per each user, the larger is the portion of the total loss due to uncorrelation. For a single user ($m = 1$) $L_{un-coor}$ is maximal, while for single antenna per user ($n_{ti} = 1$) $L_{non-coop}$ is maximal.

B. The Sum-Capacity Loss for a Gaussian MIMO Broadcast Channel

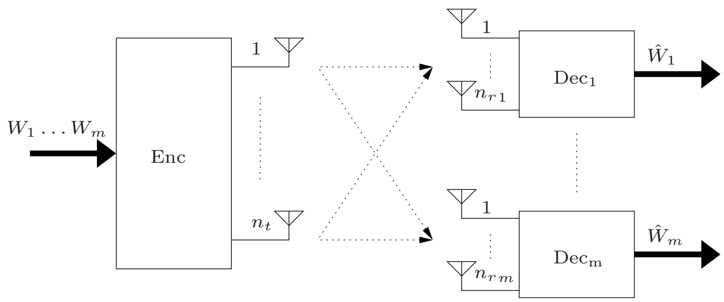

Fig. 6. The MIMO-BC model.

The MIMO-BC is a non-degraded broadcast channel for which the capacity region is still unknown in general. Hence, we only consider here the *Gaussian* MIMO-BC, whose capacity region was found recently in [26]. As in the MIMO-MAC, we consider two types of losses: the uncorrelation loss and the non-cooperation loss; however, in a MIMO-BC these losses have a slightly different meaning. To get any meaningful capacity, we must assume knowledge of the channel matrix at the transmitter. So in this sense our motivation here is not transmission to an unknown channel, but evaluation of a simple signaling scheme.

Consider the Gaussian MIMO-BC with m users, as shown in Figure 6. User i has n_{ri} receive antennas, so there are a total of $n_r = \sum_{i=1}^{m} n_{ri}$ receive antennas over all users. The channel model is given by

$$\mathbf{Y}_i = H_i\mathbf{X} + \mathbf{N}_i,\ i = 1, \dots, m \ , \tag{41}$$

where $\mathbf{Y}_i \in \mathbb{R}^{n_{ri} \times 1}$ and $\mathbf{N}_i \in \mathbb{R}^{n_{ri} \times 1}$, are the channel output and the Gaussian noise, respectively, associated with user i. Without loss of generality we can assume that $E\mathbf{N_i}\mathbf{N_i^T} = I_{n_{ri}}$. The channel input is $\mathbf{X} \in \mathbb{R}^{n_t \times 1}$, the channel matrices are $H_i \in \mathbb{R}^{n_{ri} \times n_t}$ $i = 1 \dots m$, and the power constraint is $E||\mathbf{X}||^2 \leq P$. We bound the sum-capacity loss due to using i.i.d Gaussian input, relative to the optimal sum-capacity that was shown in [21], [27], [28].

Both the optimal scheme that achieves the sum capacity and the sub-optimal scheme that uses i.i.d Gaussian input can be realized by the dirty paper coding (DPC) scheme shown in Figure 7 (for simplicity we consider a two users scheme, i.e, $m = 2$). In this scheme, the first user eliminates the interference induced by the second user using a DPC technique, while the second user considers the interference from the first user as additive noise, in which case we have a standard MIMO point-to-point encoder and decoder.

The transmitted vector $\mathbf{X}$ is a linear combination of the vectors $\mathbf{U}_1 \in \mathbb{R}^{n_{u_1}}$ and $\mathbf{U}_2 \in \mathbb{R}^{n_{u_2}}$, i.e, $\mathbf{X} = B_1\mathbf{U}_1 + B_2\mathbf{U}_2$, where B_1 is an $n_t \times n_{u1}$ matrix and B_2 is an $n_t \times n_{u2}$ matrix. The information for the first user and for the second user are carried by the vectors $\mathbf{U}_1$ and $\mathbf{U}_2$ respectively, where $\mathbf{U}_1 \sim \mathcal{N}(0, I_{n_{u_1}})$ and $\mathbf{U}_2 \sim \mathcal{N}(0, I_{n_{u_2}})$ are independent vectors. We define the matrix B which is given by $B = [B_1 \; B_2]$. The channel outputs are given by

$$\mathbf{Y}_1 = H_1\mathbf{X} + \mathbf{N}_1 = H_1B_1\mathbf{U}_1 + H_1B_2\mathbf{U}_2 + \mathbf{N}_1 \tag{42}$$

$$\mathbf{Y}_2 = H_2\mathbf{X} + \mathbf{N}_2 = H_2B_1\mathbf{U}_1 + H_2B_2\mathbf{U}_2 + \mathbf{N}_2. \tag{43}$$

The second user considers $H_2B_1\mathbf{U}_1 + \mathbf{N}_2$ as additive noise, while for the first user the transmitter performs DPC and pre-cancels $H_1B_2\mathbf{U}_2$ treating it as known interference.

The above description is common for both the optimal and the sub-optimal schemes. We now consider each scheme separately. In the sum capacity achieving scheme the matrices B_1 and B_2 should be optimized such that the input covariance matrix R_x achieves the sum capacity [21]. We assume that the channel matrices H_1 and H_2 have rank n_{r_1} and n_{r_2}, respectively. We also assume that $n_t \geq n_{r_1} + n_{r_2} = n_r$ (the extension for the general dimensions is straightforward). The vector $\mathbf{U}_1$ has $n_{u_1} = n_{r_1}$ elements and $\mathbf{U}_2$ has $n_{u_2} = n_{r_2}$ elements. The matrix B_1 is an $n_t \times n_{r_1}$ matrix and B_2 is an $n_t \times n_{r_2}$ matrix, therefore B is an $n_t \times n_r$ matrix. Using such B, we can generate an input covariance matrix with rank n_r at most, as required by the optimal solution.

The sub-optimal scheme generates i.i.d Gaussian transmission, hence the matrix B is determined such that $R_x = \frac{P}{n_t} I_{n_t}$. The vectors $\mathbf{U}_1$ and $\mathbf{U}_2$ have n_t elements each, i.e, $n_{u_1} = n_{u_2} = n_t$. The matrices B_1 and B_2 are $n_t \times n_t$ matrices, therefore B is an $n_t \times 2n_t$ matrix (for m users B is an $n_t \times mn_t$ matrix). In this case the matrix B is not optimal which results in transmission in redundant dimensions.

Unlike in the MIMO point-to-point channel and the MIMO-MAC cases, here the transmitter must know the channel matrices H_i of the users to pre-cancel the cross interferences between them [21], [20], [27]. In fact, for certain channels the BC capacity without knowing the channel matrix at the transmitter is close to zero. Even when the channel matrix is known, the sum-capacity achieving input distribution requires complex computations [21]. Furthermore, it also requires knowing the variances of the receivers' noises. This motivates us to bound the sum-capacity loss for using a fixed i.i.d Gaussian codebook.

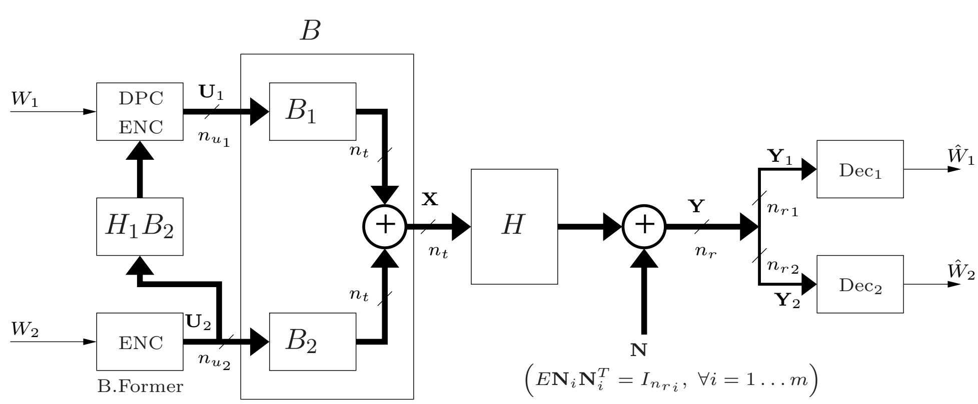

Fig. 7. The two users transmission scheme using dirty paper coding scheme.

A general input covariance matrix R_x for a Gaussian MIMO-BC has the following operational meaning [21]. For any R_x, there exists Gaussian noise called *least favorable noise* $\mathbf{N}_{LFN}$, so that the mutual information $I(\mathbf{X}; H\mathbf{X} + \mathbf{N}_{LNF})$ associated with a Gaussian input with covariance matrix R_x can be achieved over the BC using the above DPC scheme. In particular, for the optimal R_x this inner bound and the Sato outer bound [29] for the MIMO Gaussian case become tight. The sum-capacity of the Gaussian MIMO-BC [21] is

$$C_{BC}^{sum}(H, \mathbf{N}, P) = \min_{\substack{R_n \succeq 0 \\ R_{n_{ii}} = I_{n_{ri}}}} \max_{\substack{R_x \succeq 0 \\ tr(R_x) \leq P}} \frac{1}{2} \log_2 \left(\frac{det(HR_xH^T + R_n)}{det(R_n)} \right), \tag{44}$$

where R_x is the auto-correlation matrix of $\mathbf{X}$, and H is a concatenation of the users' channel matrices H_i, i.e, $H = [H_1^T \; H_2^T \; \ldots \; H_m^T]^T$. The noise is a concatenation of the user noises $\mathbf{N}_i, \; i = 1, \ldots, m$, that is $\mathbf{N} = [\mathbf{N}_1^T \; \mathbf{N}_2^T \; \ldots \; \mathbf{N}_m^T]^T$ and R_n is the auto-correlation matrix of $\mathbf{N}$. In [21] the sum-rate for *any* input was derived. For i.i.d Gaussian input $\mathbf{X}^* \sim \mathcal{N}(\mathbf{0}, \frac{P}{n_t} I_{nt})$ the sum-rate in bit per input dimension achievable by the input is given by,

$$R_{BC}^{sum}(H, \mathbf{N}, \mathbf{X}^*) \triangleq \min_{\substack{R_n \succeq 0 \\ R_{n_{ii}} = I_{n_{ri}}}} \frac{1}{2} \log_2 \left(\frac{\det(HR_{x^*}H^T + R_n)}{\det(R_n)} \right). \tag{45}$$

An equivalent expression for the sum-capacity (44) using Gaussian BC and Gaussian MAC duality was given in [27],

$$C_{BC}^{sum}(H, \mathbf{N}, P) = \sup_{\substack{\sum_{i=1}^m tr(R_{x_i}) \leq P \\ R_{x_i} \succeq 0}} \frac{1}{2} \log_2 det \left(I + \sum_{i=1}^m H_i^T R_{x_i} H_i \right). \tag{46}$$

The capacity region of BC depends only on the noises marginal distribution. In our model the noises are Gaussian with $E\mathbf{N}_i\mathbf{N}_i^T = I_{n_{ri}}, \; i = 1 \ldots, m$. Therefore, for Gaussian noises with $\mathbf{N}_W \sim \mathcal{N}(0, I_{n_r})$ the sum-capacity and the rate region is not changed. On the other hand, if we let the noises to have a correlation, then the point-to-point capacity can become infinite while the MIMO-BC sum-capacity remains constant. Therefore, it is meaningless to compare the sum capacity of the Gaussian MIMO-BC to the capacity of MIMO point-to-point channel with correlated Gaussian noise, and we define the non-cooperation loss as

$$L_{non-coop} \triangleq C_{P2P}(H, \mathbf{N}_W, P) - C_{BC}^{sum}(H, \mathbf{N}, P), \tag{47}$$

bit per channel use, where

$$C_{P2P}(H, \mathbf{N}_W, P) = \sup_{\mathbf{X}: E||\mathbf{X}||^2 \leq P} I(\mathbf{X}; H\mathbf{X} + \mathbf{N}_W). \tag{48}$$

Therefore, the non-cooperation loss amounts to the loss of sum capacity of Gaussian MIMO-BC with respect to the same MIMO channel in a point-to-point configuration with uncorrelated Gaussian noise (note that $\mathbf{N}$, $\mathbf{N}_W$ have the same covariance matrix diagonals). The uncorrelation loss is defined as the loss of i.i.d Gaussian input with respect to the sum capacity of Gaussian MIMO-BC, i.e,

$$L_{un-corr} \triangleq C_{BC}^{sum}(H, \mathbf{N}, \mathbf{P}) - R_{BC}^{sum}(H, \mathbf{N}, \mathbf{X}^*), \tag{49}$$

bit per channel use. This quantity represents the loss due to lack of correlation between transmit antennas. In the following Theorem we use the notation $C_G^*(n_t, n_r) = C_G^*$ which shows the dependence on the number of transmit and receive antennas.

Theorem 4: **(Bound for MIMO-BC)** The uncorrelation loss (49) in Gaussian MIMO-BC (41), is bounded by

$$L_{un-corr} \leq n_t C_G^*(n_t, n_r), \tag{50}$$

bit per (vector) channel use, where C_G^* is defined in (19). Equality in (50) holds for any channel matrix with rank $\min\{n_r, [[n_t/e]]\}$ in the limit of high SNR. The non-cooperation loss (47) is bounded by

$$L_{non-coop} \leq n_r C_G^*(n_r, n_t) = \begin{cases} \frac{n_t}{2}\log_2(\frac{n_r}{n_t}), & n_t \leq [[n_r/e]] \\ n_r g(n_r), & n_t \geq [[n_r/e]] \end{cases} \tag{51}$$

bit per channel use. Equality in (51) holds for any channel matrix with rank $\min\{n_t, [[n_r/e]]\}$ in the limit of high SNR.

Proof: The proof is given in Appendix I. □

Note that in the non-cooperation loss the number of transmit and receive antennas switch roles. The uncorrelation loss of Gaussian MIMO-BC is the same as the capacity loss of Gaussian MIMO point-to-point for the composite channel matrix $H \in \mathbb{R}^{n_r \times n_t}$. The sum-rate loss of i.i.d Gaussian input over Gaussian MIMO-BC with respect to the capacity of MIMO point-to-point with i.i.d Gaussian noise is given by $L_{non-coop} + L_{un-corr}$, therefore we have that,

$$L_{non-coop} + L_{un-corr} = C_{P2P}(H, \mathbf{N}_W, \mathbf{P}) - R_{BC}^{sum}(H, \mathbf{N}, \mathbf{X}^*) \leq n_t C_G^*(n_t, n_r) + n_r C_G^*(n_r, n_t), \tag{52}$$

bit per channel use. Although, as shown in Theorem 4, each of the bounds for $L_{non-coop}$ and $L_{un-corr}$ can be tight separately, we are not certain whether the bound for total loss $L_{non-coop} + L_{un-corr}$ can ever be tight.

In Figure 8 the bounds $L_{non-coop}$, $L_{un-corr}$ and $L_{non-coop} + L_{un-corr}$ in bit per input dimension are shown for $n_t = 10$ with respect to the number of receive antennas n_r. We are already familiar with the uncorrelation loss from the Gaussian MIMO-MAC and from the Gaussian MIMO point-to-point channel. The non-cooperation loss in the region of $[[n_r/e]] \leq n_t$ is almost linear with n_r. (This is the same as the curve in Figure 4, only here n_r and n_t switch roles.) For $[[n_r/e]] \geq n_t$ the non-cooperation loss per input dimension increases logarithmically with n_r.

The same scenario as in Figure 8 is shown in Figure 9, however now the losses are normalized to $\min\{n_r, n_t\}$ (number of degrees of freedom) and are drawn with respect to n_r (i.e, the losses drawn in bit per degree of freedom). For $n_r \leq n_t$ the degree of freedom is dominated by n_r, although the uncorrelation loss (in bit) increases logarithmic with n_r, the uncorrelation loss per degree of freedom decreases until it becomes fixed loss per degree of freedom in $n_r = n_t$. On the other hand, the non-cooperation loss for $n_r \leq n_t$ has quite fixed contribution for each degree of freedom. However, for $n_r \geq n_t$ each additional degree of freedom increases the loss logarithmic.

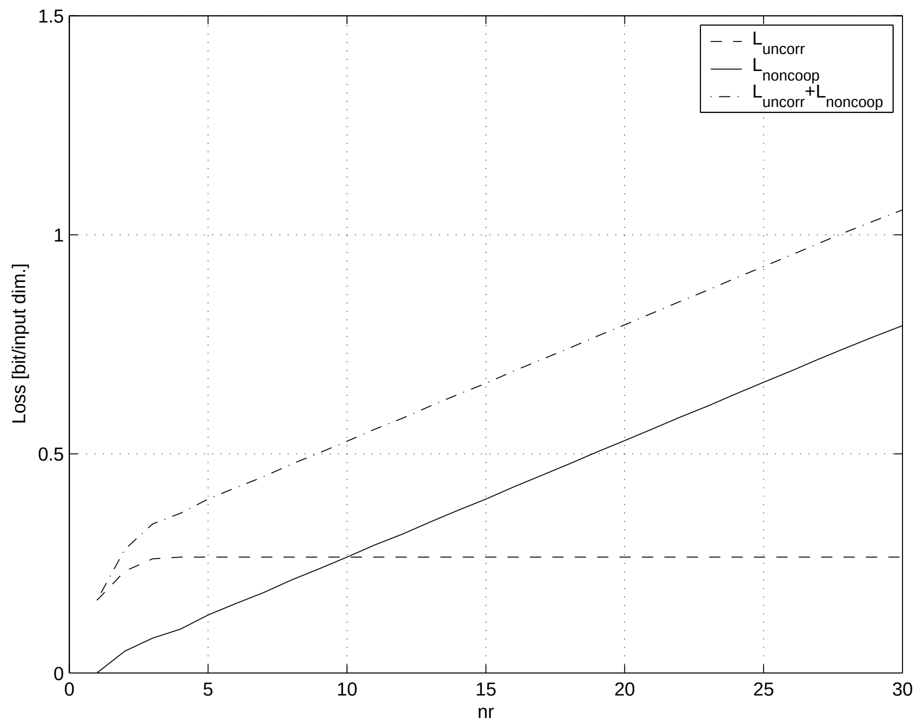

Fig. 8. The Gaussian MIMO-BC losses per input dimension for $n_t = 10$.

IV. DISCUSSION

We investigated the robustness of uncorrelated Gaussian input for linear additive noise MIMO channels. We presented two uniform upper bounds on the capacity loss for using this input. The first bound holds for any noise (not necessarily Gaussian) while the second bound holds for (correlated) Gaussian noise. We also extended these bounds to the sum-capacity loss in the MIMO multiuser case.

These bounds imply that the capacity loss is additive, meaning that we do not lose degrees of freedom due to using the robust input. Furthermore, since the bounds are independent of the SNR, the fractional capacity loss becomes negligible as the SNR and hence the capacity increase.

These results may be extended in several directions. The first direction is a bound on the loss in the capacity *region* in the multiuser case; this can be accomplished by applying the same techniques of Theorems 3 and 4 for the sum-rate loss to all possible subsets of users with respect to the capacity region of Gaussian MIMO-MAC and Gaussian MIMO-BC, respectively. Another interesting direction is the capacity loss of robust input for the case where there is partial side information on the channel at the transmitter.

In this work we limited our discussion to capacity as a measure for the channel quality, and as we saw the results are quite optimistic. However, the decoder error probability and the outage capacity seem to be more sensitive of having the CSI at the transmitter. In particular, efficient coding techniques can use the CSI at the transmitter, in order to improve the error probability and the outage capacity.

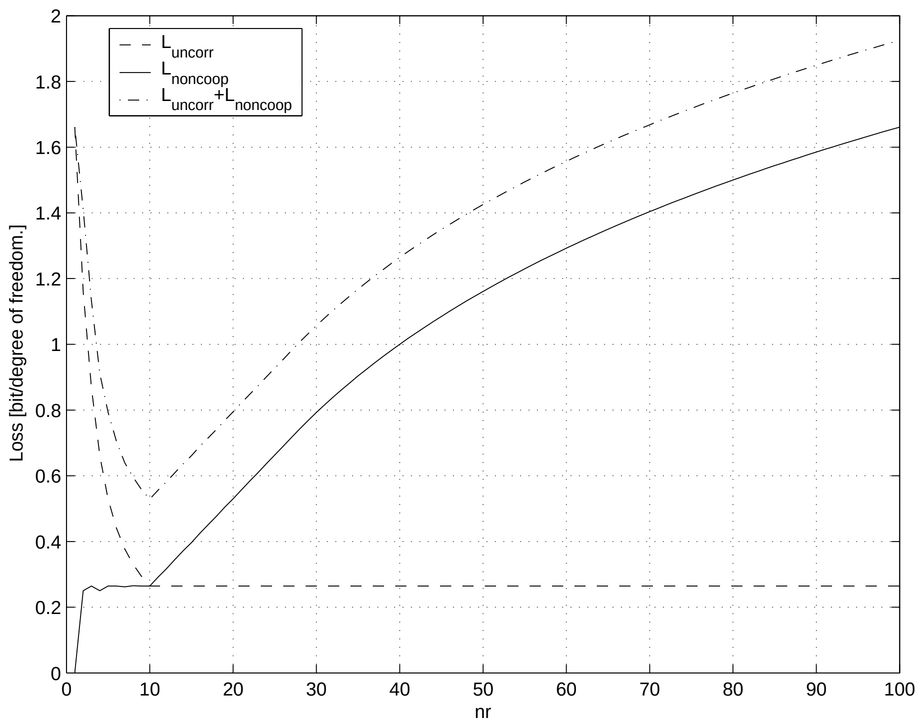

Fig. 9. The loss in the Gaussian MIMO-BC w.r.t $\min(n_t, n_r)$ for $n_t = 10$.

ACKNOWLEDGMENT

The authors wish to thank Uri Erez for helpful discussions.

APPENDIX

A. Bounds for the additive noise SISO channels

In this Appendix we complement the result of [17], and give tighter bounds for the capacity loss in a few special cases. Consider the additive (general) noise SISO channel $Y = X + N$ with power constraint $EX^2 \leq P$ and noise variance σ_n^2, where C denotes the channel capacity and the signal to noise ratio is $SNR = \frac{P}{\sigma_n^2}$. Let R_G denote the achievable rate by a Gaussian input $X_G \sim \mathcal{N}(0, P)$, i.e, $R_G = I(X_G; X_G + N)$. The following bounds hold:

1. $C - R_G \leq D(N||N_G)$.
2. $C - R_G \leq \frac{1}{2}\log_2(1 + 1/SNR)$.
3. $C - R_G \leq C$

where $D(\cdot||\cdot)$ is Kullback Leibler distance (divergence) and N_G is a Gaussian noise with the same variance as N.

Proof:

1. The capacity C is given by

$$
\begin{aligned}
C &= \max_{X:EX^2\leq P} I(X;X+N) && (53)\\
&= \max_{X:EX^2\leq P} h(X+N)-h(N) && (54)\\
&\leq \frac{1}{2}\log_2\left(2\pi e(P+\sigma_n^2)\right)-h(N) && (55)\\
&= \frac{1}{2}\log_2\left(2\pi e(P+\sigma_n^2)\right)-h(N_G)+h(N_G)-h(N) && (56)\\
&= \frac{1}{2}\log_2\left(1+SNR\right)+D(N||N_G) && (57)\\
&\leq R_G+D(N||N_G), && (58)
\end{aligned}
$$

   where (55) follows since $X\sim\mathcal{N}(0,P)$ maximizes $h(X+N)$; (57) follows directly from the identity $D(N||N_G)=h(N_G)-h(N)$. Finally, (58) follows from the inequality $R_G=I(X_G;X_G+N)\geq I(X_G;X_G+N_G)$ [1, p263].

2. From (57) we know that,

$$C\leq\log_2\left(1+SNR\right)+D(N||N_G). \tag{59}$$

   On the other hand, R_G can be written as,

$$
\begin{aligned}
R_G &= h(X_G+N)-h(N) && (60)\\
&\geq h(X_G)-h(N) && (61)\\
&= h(X_G)-h(N_G)+h(N_G)-h(N) && (62)\\
&= \frac{1}{2}\log_2\left(SNR\right)+D(N||N_G), && (63)
\end{aligned}
$$

   where (61) follows again from $h(X_G+N)\geq h(X_G)$; (63) follows from $D(N||N_G)=h(N_G)-h(N)$. Therefore, using (59) and (63) we have that,

$$C-R_G\leq\frac{1}{2}\log_2\left(1+\frac{1}{SNR}\right) \tag{64}$$

3. The bound is true because $R_G\geq 0$.

□

Similar bounds can be shown for the MIMO point-to-point channels.

B. Proof of Lemma 1

For any independent random variables W,T,Z, it was shown in [17] that

$$I(W;W+Z)\leq I(T;T+Z)+I(W;W+T). \tag{65}$$

This inequality can be generalized to random vectors as well. Specifically, replacing W,T,Z by the following quantities $W\rightarrow H\mathbf{X}$, $Z\rightarrow\mathbf{N}$, $T\rightarrow H\mathbf{X}'$, we have that

$$I(H\mathbf{X};H\mathbf{X}+\mathbf{N}) \quad\leq\quad I(H\mathbf{X}';H\mathbf{X}'+\mathbf{N})+I(H\mathbf{X};H\mathbf{X}+H\mathbf{X}'). \tag{66}$$

Since $I(\mathbf{X}; A\mathbf{X}+\mathbf{N}) = I(A\mathbf{X}; A\mathbf{X}+\mathbf{N})$ for any $A \in \mathbb{R}^{n_r \times n_t}$, (66) can be written as

$$I(\mathbf{X}'; H\mathbf{X}'+\mathbf{N}) \geq I(\mathbf{X}; H\mathbf{X}+\mathbf{N}) - I(\mathbf{X}; H\mathbf{X}+H\mathbf{X}'). \tag{67}$$

Noting that $I(\mathbf{X}; H\mathbf{X}+H\mathbf{X}') \leq \sup_{\mathbf{Z}: E||\mathbf{Z}||^2 \leq P} I(\mathbf{Z}; H\mathbf{Z}+H\mathbf{X}')$ and using the definition in (14), inequality (67) becomes

$$I(\mathbf{X}'; H\mathbf{X}'+\mathbf{N}) \geq I(\mathbf{X}; H\mathbf{X}+\mathbf{N}) - n_t C^*(H, \mathbf{X}'), \tag{68}$$

for any $\mathbf{X}$. In particular, it holds for an input $\mathbf{X}$ that achieves the capacity $C(H, \mathbf{N}, P)$, and the Lemma follows.

C. Proof of Lemma 2

Without loss of generality we can assume that the $n_r \times n_t$ channel matrix H is full rank with orthonormal rows and $n_r \leq n_t$. To see that, let G be an arbitrary $n_r \times n_t$ matrix with rank $r \leq \min(n_t, n_r)$. Then, there exists a transformation $T \in \mathbb{R}^{r \times n_r}$ which when applied to any vector $G\mathbf{Z}$ is information lossless. In particular:

$$I(\mathbf{Z}; G\mathbf{Z}+G\mathbf{X}') \quad = \quad I(\mathbf{Z}; TG\mathbf{Z}+TG\mathbf{X}'). \tag{69}$$

Note that the matrix T is not unique; a natural selection of T can be a matrix that chooses r rows of G that are linearly independent. The equivalent channel matrix $\tilde{G} = TG$ is an $r \times n_t$ full rank matrix. Using a Gram Schmidt process, $\tilde{G}$ can be written as $\tilde{G} = RH$, where $R \in \mathbb{R}^{r \times r}$ is a non-singular lower triangular matrix and $H \in \mathbb{R}^{r \times n_t}$ is full rank with orthonormal rows; thus

$$\begin{aligned} I(\mathbf{Z}; G\mathbf{Z}+G\mathbf{X}') &= I(\mathbf{Z}; RH\mathbf{Z}+RH\mathbf{X}') && (70)\\ &= I(\mathbf{Z}; H\mathbf{Z}+H\mathbf{X}'), && (71)\end{aligned}$$

where (70) follows from (69), and (71) follows from the fact that multiplication by non-singular square matrix R^{-1} does not change the mutual information. We see that we can always transform the channel to be orthonormal with $n_r \leq n_t$, without loss of information.

We now turn to show the following lower bound:

$$\frac{n_r}{2n_t} \log_2(1+\frac{n_t}{n_r}) \leq \sup_H C^*(H, \mathbf{X}'),\ \forall \mathbf{X}'. \tag{72}$$

Among all the eigenvalue decompositions of $R_{x'}$, we will be interested in the following decomposition: $R_{x'} = Q\Lambda_{\mathbf{x}'}Q^T$ where Q is an $n_t \times n_t$ unitary matrix, and $\Lambda_{\mathbf{x}'}$ is an $n_t \times n_t$ diagonal matrix with positive elements with increasing order on the diagonal. Let $\tilde{H} = WQ^T$ where W is an $n_r \times n_t$ diagonal matrix with unit elements. Note that $\tilde{H}$ is full rank with orthonormal rows. We have

$$
\begin{aligned}
n_t \sup_H C^*(H,\mathbf{X}') &\geq n_t C^*(\tilde{H},\mathbf{X}') && (73)\\
&= \sup_{\mathbf{Z}:E||\mathbf{Z}||^2\leq P} I(\tilde{H}\mathbf{Z};\tilde{H}\mathbf{Z}+\tilde{H}\mathbf{X}') && (74)\\
&\geq I(\tilde{H}\mathbf{Z}';\tilde{H}\mathbf{Z}'+\tilde{H}\mathbf{X}') && (75)\\
&= I(\tilde{\mathbf{Z}};\tilde{\mathbf{Z}}+\tilde{\mathbf{X}}) && (76)\\
&\geq \sum_{i=1}^{n_r} I(\tilde{Z}_i;\tilde{Z}_i+\tilde{X}_i) && (77)\\
&\geq \frac{1}{2}\sum_{i=1}^{n_r}\log_2\Big(1+\frac{P/n_r}{\sigma^2_{\tilde{X}_i}}\Big) && (78)\\
&= \frac{n_r}{2}\sum_{i=1}^{n_r}\frac{1}{n_r}\log_2\Big(1+\frac{P/n_r}{\sigma^2_{\tilde{X}_i}}\Big) && (79)\\
&\geq \frac{n_r}{2}\log_2\Big(1+\frac{P/n_r}{\frac{1}{n_r}\sum_{i=1}^{n_r}\sigma^2_{\tilde{X}_i}}\Big) && (80)\\
&\geq \frac{n_r}{2}\log_2\Big(1+\frac{P/n_r}{\frac{1}{n_r}P\frac{n_r}{n_t}}\Big) && (81)\\
&= \frac{n_r}{2}\log_2\Big(1+\frac{n_t}{n_r}\Big), && (82)
\end{aligned}
$$

where in (73) we replaced the maximization by a specific $H=\tilde{H}$;

(74) is equivalent to the definition of $C^*(H,\mathbf{X}')$ given in (14);

in (75) we replaced the maximization by a specific $\mathbf{Z}=\mathbf{Z}'\sim\mathcal{N}(0,\frac{P}{n_t}I_{n_t})$;

in (76) we defined $\tilde{\mathbf{X}}\triangleq\tilde{H}\mathbf{X}'$, and $\tilde{\mathbf{Z}}\triangleq\tilde{H}\mathbf{Z}'$; we observe that the covariance matrix $R_{\tilde{x}}=W\Lambda_{\mathbf{x}'}W^T$ is an $n_r\times n_r$ diagonal matrix with diagonal elements $R_{\tilde{x}_{i,i}}=\Lambda_{x'_{i,i}}\triangleq\sigma^2_{\tilde{\mathbf{x}}_\mathbf{i}}$ for $\mathrm{i}=1\ldots\mathrm{n_r}$ (where we use $A_{i,j}$ to denote the i,j-th element of a matrix A), and that $\tilde{\mathbf{Z}}\sim\mathcal{N}(0,\frac{P}{n_r}I_{n_r})$ due to the orthonormality of Q and the structure of W;

(77) follows since the mutual information over an additive noise channel with memoryless input is minimized when the noise is also memoryless, in which case we can write the vector mutual information as the sum of scalar mutual informations;

(78) follows from the fact that Gaussian noise has the lowest capacity over all additive noise channels [1, p.488];

(80) follows from Jensen's Inequality since $\log(1+1/x)$ is convex ($\cup$) with respect to x;

finally, (81) follows because $tr(R_{x'})\leq P$ from (15), and since $\sigma_{\tilde{x}_1},\ldots,\sigma_{\tilde{x}_{n_r}}$ are the lowest n_r out of n_t eigenvalues of $R_{x'}$, which together imply that $\sum_{i=1}^{n_r}\sigma^2_{\tilde{\mathbf{x}}_\mathbf{i}}\leq P\frac{n_r}{n_t}$. Hence (72) is proved.

Now, we show that equality in (72) is achieved for $\mathbf{X}'=\mathbf{X}^*$, i.e.,

$$\sup_H C^*(H,\mathbf{X}^*)=\frac{n_r}{2n_t}\log_2(1+\frac{n_t}{n_r}), \tag{83}$$

which proves Lemma 2. For $\mathbf{X}'=\mathbf{X}^*$, (14) becomes

$$C^*(H,\mathbf{X}^*)=\frac{1}{n_t}\sup_{\mathbf{Z}:E||\mathbf{Z}||^2\leq P} I(\mathbf{Z};H\mathbf{Z}+H\mathbf{X}^*). \tag{84}$$

Since $\mathbf{X}^* \sim \mathcal{N}(0, \frac{P}{n_t} I_{n_t})$, the optimum is achieved by $\mathbf{Z}$ with Gaussian distribution, thus

$$\begin{aligned} C^*(H, \mathbf{X}^*) = \quad & \frac{1}{n_t} \sup \left\{ \frac{1}{2} \log_2 \frac{|H R_z H^T + \frac{P}{n_t} H H^T|}{|\frac{P}{n_t} H H^T|} \right\}. \\ s.t: \quad & R_z \succeq 0 \\ & tr(R_z) \leq P \end{aligned} \tag{85}$$

Let us define $D \triangleq H R_z H^T + \frac{P}{n_t} I_{n_r}$. The rows of H are orthonormal, i.e, $HH^T = I_{n_r}$, hence the denominator in (85) is constant, while the numerator is given by

$$|H R_z H^T + \frac{P}{n_t} I_{n_r}| \leq \prod_{i=i}^{n_r} D_{ii} \tag{86}$$

$$\leq \left(\frac{1}{n_r} \sum_{i=1}^{n_r} D_{ii} \right)^{n_r} \tag{87}$$

$$= \left(\frac{tr(H R_z H^T) + \frac{P}{n_t} n_r}{n_r} \right)^{n_r} \tag{88}$$

$$\leq \left(\frac{P + P \frac{n_r}{n_t}}{n_r} \right)^{n_r} \tag{89}$$

$$= \left(\frac{P}{n_r} + \frac{P}{n_t} \right)^{n_r}, \tag{90}$$

where (86) follows from the Hadamard Inequality [1, p.502]; (87) follows from the Arithmetic-Geometric Mean Inequality; (89) follows from the fact that $tr(H R_z H^T) \leq P$, since H has orthonormal rows. Substituting (90) in (85) and using that $HH^T = I_{n_r}$, we have

$$C^*(H, \mathbf{X}^*) \leq \frac{1}{2n_t} \log_2 \left(\frac{\frac{P}{n_r} + \frac{P}{n_t}}{\frac{P}{n_t}} \right)^{n_r} \tag{91}$$

$$= \frac{n_r}{2n_t} \log_2 \left(1 + \frac{n_t}{n_r} \right). \tag{92}$$

Therefore (83) follows from (92) and (72), and Lemma 2 as well.

D. Proof of Theorem 1

In the case $n_r \leq n_t$, the theorem follows directly from (83). As for the case $n_r > n_t$, from (71) in the proof of Lemma 2, we showed that with no loss of information, any channel can be transformed into a channel whose number of outputs n_r is not more than the number of inputs n_t. Since the bound C^* is monotonically increasing in the number of outputs in the range $1 \leq n_r \leq n_t$, the capacity loss is no more than the value of the bound at $n_r = n_t$, which is $C^* = \frac{1}{2}$ bit per input dimension.

E. Proof of Theorem 2

We assume a MIMO channel model where $\mathbf{Y} = H\mathbf{X} + \mathbf{N}_G$. Since the capacity of this channel is finite the correlation matrix of $\mathbf{N}_G$ is nonsingular, therefore there is an equivalent information lossless model $\mathbf{Y} = \tilde{H}\mathbf{X} + \mathbf{N}_G^*$ where the noise distribution is $N_G^* \sim \mathcal{N}(0, I_{n_r})$.

The Singular Value Decomposition (SVD) [30] of $\tilde{H}$ is

$$\tilde{H} = Q_2 \Delta Q_1^T, \tag{93}$$

where $Q_2 \in \mathbb{R}^{n_r \times n_r}$ and $Q_1 \in \mathbb{R}^{n_t \times n_t}$ are unitary matrices and $\Delta \in \mathbb{R}^{n_r \times n_t}$ is a diagonal matrix. Using the unitary transformations $\tilde{\mathbf{Y}} = Q_2^T \mathbf{Y}$ and $\mathbf{X} = Q_1 \tilde{\mathbf{X}}$ at the decoder and encoder respectively, the capacity is given by

$$C(\tilde{H}, \mathbf{N}_G^*, P) = \frac{1}{n_t} \max_{E||\mathbf{X}||^2 \leq P} I(\mathbf{X}; \tilde{H}\mathbf{X} + \mathbf{N}_G^*) = \frac{1}{n_t} \max_{E||\tilde{\mathbf{X}}||^2 \leq P} I(\tilde{\mathbf{X}}; \Delta\tilde{\mathbf{X}} + \tilde{\mathbf{N}}), \tag{94}$$

where $\tilde{\mathbf{N}} = Q_2^T \mathbf{N}_G^*$, hence $\tilde{\mathbf{N}} \sim \mathcal{N}(0, I_{n_r})$. The power constraint is $E||\mathbf{X}||^2 = E||\tilde{\mathbf{X}}||^2 \leq P$. Let r be the rank of the matrix $\tilde{H}$, i.e $r = rank(H)$. We assume that the r non-zero elements on the diagonal of the matrix Δ are $\Delta_{11}, \ldots, \Delta_{rr}$. The capacity is given by

$$C(\tilde{H}, \mathbf{N}_G^*, P) = \frac{1}{n_t} \max_{E||\mathbf{X}'||^2 \leq P} I(\mathbf{X}'; R\mathbf{X}' + \mathbf{N}'), \tag{95}$$

where

$$\begin{aligned} R_{ij} &= \Delta_{ij}, i, j = 1, \ldots, r \\ X_i' &= \tilde{X}_i, i = 1, \ldots, r \\ N_i' &= \tilde{N}_i, i = 1, \ldots, r. \end{aligned}$$

Note that the R is diagonal matrix and $\mathbf{N}' \sim N(0, I_r)$. By multiplying the receiver input with R^{-1}, the equivalent model corresponds to a parallel channels model with colored noise, where the noise covariance matrix is $\Lambda = diag(\lambda_1 \ldots \lambda_r)$ where $\lambda_i = \frac{1}{R_{ii}^2} \geq 0$. The capacity of parallel channels is achieved using water filling optimization [1], which is given by

$$\begin{aligned} C(\tilde{H}, \mathbf{N}_G^*, P) &= \frac{1}{2n_t} \sum_{i=1}^{r} \log_2 \left(1 + \frac{(\nu - \lambda_i)^+}{\lambda_i}\right) \\ s.t: \quad & \sum_{i=1}^{r} (\nu - \lambda_i)^+ \leq P. \end{aligned} \tag{96}$$

On the other hand, the rate achieved using i.i.d Gaussian input $\mathbf{X}^* \sim \mathcal{N}(0, \frac{P}{n_t} I_{n_t})$ is given by,

$$\begin{aligned} I(\mathbf{X}^*; \tilde{H}\mathbf{X}^* + \mathbf{N}_G^*) &= \frac{1}{2} \log_2 \frac{|\tilde{H} \frac{p}{n_t} I_{n_t} \tilde{H}^T + I_{n_r}|}{|I_{n_r}|} && (97) \\ &= \frac{1}{2} \log_2 |\frac{p}{n_t} H H^T + I_{n_r}| && (98) \\ &= \frac{1}{2} \log_2 |\frac{p}{n_t} Q_2 \Delta \Delta^T Q_2^T + I_{n_r}| && (99) \\ &= \frac{1}{2} \log_2 |Q_2 (\frac{p}{n_t} \Delta \Delta^T + I_{n_r}) Q_2^T| && (100) \\ &= \frac{1}{2} \log_2 |\frac{p}{n_t} \Delta \Delta^T + I_{n_r})|, && (101) \end{aligned}$$

where (99) follows from (93). Since $\Delta\Delta^T = R^2$, we have that

$$I(\mathbf{X}^*; \tilde{H}\mathbf{X}^* + \mathbf{N}_G^*) = \frac{1}{2} \log_2 \prod_{i=1}^{r} (1 + \frac{P}{n_t} R_{ii}^2). \tag{102}$$

Since $R_{ii}^2 = \frac{1}{\lambda_i}$ for $i = 1 \dots r$, we have

$$I(\mathbf{X}^*; \tilde{H}\mathbf{X}^* + \mathbf{N}_G^*) = \frac{1}{2}\sum_{i=1}^{r} \log_2 \frac{\frac{P}{n_t} + \lambda_i}{\lambda_i}. \tag{103}$$

The capacity loss is given by the difference between (96) and (103), therefore

$$\begin{aligned}
n_t C(\tilde{H}, \mathbf{N}_G^*, P) - I(\mathbf{X}^*; \tilde{H}\mathbf{X}^* + \mathbf{N}_G^*) &= \frac{1}{2}\sum_{i=1}^{r} \left\{ \log_2 \left(\frac{\lambda_i + (\nu - \lambda_i)^+}{\lambda_i} \right) - \log_2 \left(\frac{\frac{P}{n_t} + \lambda_i}{\lambda_i} \right) \right\} && (104)\\
&= \frac{1}{2}\sum_{i=1}^{r} \left\{ \log_2 \left(\frac{\max(\lambda_i, \nu)}{\lambda_i} \cdot \frac{\lambda_i}{\lambda_i + \frac{P}{n_t}} \right) \right\} && (105)\\
&= \frac{1}{2}\sum_{i=1}^{r} \log_2 \left(\frac{\max(\lambda_i, \nu)}{\lambda_i + \frac{P}{n_t}} \right). && (106)
\end{aligned}$$

Let $\mathcal{I} \triangleq \{i : \nu - \lambda_i > P/n_t,\ i = 1, \dots, r\}$, where $|\mathcal{I}|$ is the cardinality of $\mathcal{I}$, thus

$$\begin{aligned}
n_t C(\tilde{H}, \mathbf{N}_G^*, P) - I(\mathbf{X}^*; \tilde{H}\mathbf{X}^* + \mathbf{N}_G^*) &\leq \frac{1}{2}\sum_{i\in\mathcal{I}} \log_2 \frac{\nu}{\lambda_i + \frac{P}{n_t}} && (107)\\
&\leq \frac{1}{2}\sum_{i\in\mathcal{I}} \log_2 \frac{\nu - \lambda_i}{\frac{P}{n_t}} && (108)\\
&= \frac{1}{2}|\mathcal{I}| \cdot \sum_{i\in\mathcal{I}} \frac{1}{|\mathcal{I}|} \log_2 \frac{\nu - \lambda_i}{\frac{P}{n_t}} && (109)\\
&\leq \frac{1}{2}|\mathcal{I}| \cdot \log_2 \Big(\frac{1}{|\mathcal{I}|} \sum_{i\in\mathcal{I}} \frac{\nu - \lambda_i}{\frac{P}{n_t}} \Big) && (110)\\
&\leq \frac{1}{2}|\mathcal{I}| \cdot \log_2 \frac{n_t}{|\mathcal{I}|} && (111)\\
&= \frac{n_t}{2} \cdot \frac{|\mathcal{I}|}{n_t} \cdot \log_2 \frac{n_t}{|\mathcal{I}|}, && (112)
\end{aligned}$$

where (108) follows from the property that $b \geq c$ implies that $\frac{a+b}{a+c} \leq b/c$ for positive a, b, c; (110) follows from Jensen's inequality; (111) follows from the power constraint in (96). Inspecting the function $\frac{x}{n_t} \cdot \log_2 \frac{n_t}{x}$ where x can take non-negative real values shows that the function is concave with respect to x, and the global maximum is achieved for $x = \frac{n_t}{e}$. In our case $x = |\mathcal{I}|$ which can take only non-negative integers, the maximum is achieved for $|\mathcal{I}| = [[n_t/e]]$ where $[[\cdot]]$ is defined in Theorem 2.

First we consider the case $n_r < [[n_t/e]]$, therefore $|\mathcal{I}| \leq r \leq n_r$. The function $\frac{|\mathcal{I}|}{n_t} \cdot \log_2 \frac{n_t}{|\mathcal{I}|}$ is increasing in $|I|$ at the interval $|\mathcal{I}| \in [0, [[n_t/e]])$, therefore the capacity loss is maximized by $|\mathcal{I}| = n_r$, and it is given by

$$n_t C_G^* = \frac{n_r}{2} \cdot \log_2(\frac{n_t}{n_r}). \tag{113}$$

In the case $n_r \geq [[n_t/e]]$, the capacity loss is bounded by the loss at the global maximum , i.e, $|\mathcal{I}| = [[n_t/e]]$, therefore

$$n_t C_G^* = \frac{1}{2}[[n_t/e]] \log_2 \Big(\frac{n_t}{[[n_t/e]]} \Big). \tag{114}$$

The equality in (18) requires channel matrix H with rank $r = \min\{n_r, [[n_t/e]]\}$ and high SNR. In this case $\nu >> \lambda_i,\ i = 1 \ldots r$, the capacity loss (106) becomes

$$n_t C(\tilde{H}, \mathbf{N}_G^*, P) - I(\mathbf{X}^*; \tilde{H}\mathbf{X}^* + \mathbf{N}_G^*) \simeq \frac{1}{2}\sum_{i=1}^{r} \log_2\left(\frac{\nu}{P/n_t}\right) \tag{115}$$

$$= \frac{\min\{n_r, [[n_t/e]]\}}{2} \log_2\left(\frac{n_t}{\min\{n_r, [[n_t/e]]\}}\right), \tag{116}$$

where (116) follows from substituting $r = \min\{n_r, [[n_t/e]]\}$ and $\nu = P/r$. Hence Theorem 2 follows.

F. *Proof of Proposition 1*

The proof is equivalent to the proof of the equality in Theorem 2 as shown in Appendix E, equations (115)-(116). Note that in this case the channel matrix H is full rank, i.e, $r = \min\{n_t, n_r\}$, the capacity loss becomes

$$n_t C(\tilde{H}, \mathbf{N}_G^*, P) - I(\mathbf{X}^*; \tilde{H}\mathbf{X}^* + \mathbf{N}_G^*) \simeq \frac{1}{2}\sum_{i=1}^{r} \log_2\left(\frac{\nu}{P/n_t}\right) \tag{117}$$

$$= \frac{\min\{n_r, n_t\}}{2} \log_2\left(\frac{n_t}{\min\{n_r, n_t\}}\right). \tag{118}$$

Hence, Proposition 1 follows.

G. *Proof of Lemma 3*

Arbitrary noise - for any input $\mathbf{X}$ restricted to individual power constraint (25), we have by [17, Lemma 1],

$$I(\mathbf{X}; H\mathbf{X} + \mathbf{N}) - I(\mathbf{X}^D; H\mathbf{X}^D + \mathbf{N}) \leq I(H\mathbf{X}; H\mathbf{X} + H\mathbf{X}^D) = I(\mathbf{X}; H\mathbf{X} + H\mathbf{X}^D). \tag{119}$$

In particular, the last inequality holds for the capacity achieving input, that is

$$n_t C(H, \mathbf{N}, \mathbf{P}) - I(\mathbf{X^D}; H\mathbf{X^D} + \mathbf{N}) \leq I(\mathbf{X}; H\mathbf{X} + H\mathbf{X}^D) \tag{120}$$

$$\leq \sup_{\{\mathbf{Z}: E||\mathbf{Z}_i||^2 \leq P_i,\ i=1,\ldots m\}} I(\mathbf{Z}; H\mathbf{Z} + H\mathbf{X}^D), \tag{121}$$

where the last inequality follows by maximizing $I(\mathbf{X}; H\mathbf{X} + H\mathbf{X}^D)$. Let us define $\tilde{H} \triangleq HR_{x^D}^{1/2}$ and $\tilde{\mathbf{X}}^D \triangleq R_{x^D}^{-1/2}\mathbf{X}^D$, and $\tilde{\mathbf{Z}} \triangleq R_{x^D}^{-1/2}\mathbf{Z}$, where $R_{x^D}^{-1/2}$ is the inverse square root of R_{x^D} defined in (24). It can be assumed that R_{x^D} is positive definite matrix ,i.e, $P_i/n_{ti} > 0,\ i = 1 \ldots m$, otherwise all the groups with $P_i = 0$ can be omitted. In this case $\tilde{\mathbf{X}}^D \sim \mathcal{N}(\mathbf{0}, I_{n_t})$ and $\tilde{\mathbf{Z}} \sim \mathcal{N}(\mathbf{0}, R_{\tilde{z}})$ where $E||\tilde{\mathbf{Z}}_i||^2 \leq n_{ti}, i = 1 \ldots m$. The capacity loss is given by

$$n_t C(H, \mathbf{N}, \mathbf{P}) - I(\mathbf{X^D}; H\mathbf{X^D} + \mathbf{N}) \leq \sup_{\{\tilde{\mathbf{Z}}: E||\tilde{\mathbf{Z}}_i||^2 n_{ti},\ i=1\ldots m\}} I(\tilde{\mathbf{Z}}; H\tilde{\mathbf{Z}} + \tilde{H}\tilde{\mathbf{X}}^D) \tag{122}$$

$$\leq \sup_{\{\tilde{\mathbf{Z}}: E||\tilde{\mathbf{Z}}||^2 \leq n_t\}} I(\tilde{\mathbf{Z}}; H\tilde{\mathbf{Z}} + \tilde{H}\tilde{\mathbf{X}}^D) \tag{123}$$

$$\leq n_t C^*, \tag{124}$$

where (123) follows since the maximization is taken over a larger set; (124) follows from Theorem 1.

Gaussian noise - for any input $\mathbf{X}$ we have that

$$n_t C(H, \mathbf{N}_G, \mathbf{P}) - I(\mathbf{X^D}; H\mathbf{X^D} + \mathbf{N}_G) =$$

$$\sup_{\{\mathbf{X}: E||\mathbf{X}_i||^2 \leq P_i,\ i=1...m\}} I(\mathbf{X}; H\mathbf{X} + \mathbf{N}_G) - I(\mathbf{X^D}; H\mathbf{X^D} + \mathbf{N}_G), \tag{125}$$

where (125) follows from (26). Since $\mathbf{N}_G$ has Gaussian distribution, the supremum in (125) is achieved by $\mathbf{X}$ with Gaussian distribution [1, p.488]. Using the definitions of $\tilde{H}$, $\tilde{\mathbf{X}}^D$ as used for the arbitrary noise, and using $\tilde{\mathbf{X}} \triangleq R_{x^D}^{-1/2}\mathbf{X}$, thus $\tilde{\mathbf{X}} \sim \mathcal{N}(\mathbf{0}, R_{\tilde{x}})$ where $tr(R_{\tilde{x}}) = n_t$. Therefore, (125) can be written as

$$\begin{aligned}
&\sup_{\{\tilde{\mathbf{X}}: E||\tilde{\mathbf{X}}_i||^2 \leq n_{t\,i}, i=1...m\}} I(\tilde{\mathbf{X}}; \tilde{H}\tilde{\mathbf{X}} + \mathbf{N}_G) - I(\tilde{\mathbf{X}}^D; \tilde{H}\tilde{\mathbf{X}}^D + \mathbf{N}_G) \\
&\leq \sup_{\{\tilde{\mathbf{X}}: E||\tilde{\mathbf{X}}||^2 \leq n_t\}} I(\tilde{\mathbf{X}}; \tilde{H}\tilde{\mathbf{X}} + \mathbf{N}_G) - I(\tilde{\mathbf{X}}^D; \tilde{H}\tilde{\mathbf{X}}^D + \mathbf{N}_G) && (126) \\
&= n_t C(\tilde{H}, \mathbf{N}, P = n_t) - I(\tilde{\mathbf{X}}^D; H\tilde{\mathbf{X}}^D + \mathbf{N}_G) && (127) \\
&\leq n_t C_G^*, && (128)
\end{aligned}$$

where (126) follows since the maximum is taken over a larger set; (127) follows from the capacity definition; finally, (128) follows from Theorem 2.

The equality in (29) requires channel matrix H with rank $\min\{n_r, [[n_t/e]]\}$ and high SNR. The inequality (126) is achieved with equality for high SNR, since i.i.d Gaussian input maximizes the mutual information $I(\tilde{\mathbf{X}}; \tilde{H}\tilde{\mathbf{X}}+\mathbf{N}_G)$; the equality in (128) follows from the equality conditions in Theorem 2.

H. *Proof of Theorem 3*

The proof of Theorem 3 is followed directly from Lemma 3.

Arbitrary noise - the sum of the uncorrelation loss and noncooperation loss can be written as

$$\begin{aligned}
L_{un-corr} + L_{non-coop} &= C_{P2P}(H, \mathbf{N}, \mathbf{P}) - I(\mathbf{X}^D; H\mathbf{X}^D + \mathbf{N}) && (129) \\
&\leq n_t C^*, && (130)
\end{aligned}$$

where (129) follows from the definition in (38); (130) follows from Lemma 3.

Gaussian noise - the same arguments hold for the Gaussian noise, thus

$$\begin{aligned}
L_{un-corr} + L_{non-coop} &= C_{P2P}(H, \mathbf{N}_G, \mathbf{P}) - I(\mathbf{X}^D; H\mathbf{X}^D + \mathbf{N}_G) && (131) \\
&\leq n_t C_G^*. && (132)
\end{aligned}$$

The equality in (50) requires channel matrix H with rank $\min\{n_r, [[n_t/e]]\}$ and high SNR. In this case the inequality (132) is achieved with equality due to Lemma 3.

I. Proof of Theorem 4

Uncorrelation loss - the sum-capacity of Gaussian MIMO-BC is given in (44) where the input and the noise distributions are Gaussian. Let us define the optimal input $\mathbf{X}^{opt} \sim \mathcal{N}(\mathbf{0}, R_{x^{opt}})$ and the least favorable noise $\mathbf{N}^{opt} \sim \mathcal{N}(\mathbf{0}, R_{n^{opt}})$ which are the min-max solution of (44). The solution of (44) $\mathbf{X}^{opt}, \mathbf{N}^{opt}$ is a saddle point [21], thus for any input $\mathbf{X}$ satisfies the power constraint

$$\begin{aligned} I(\mathbf{X}; H\mathbf{X} + \mathbf{N}^{opt}) &\leq I(\mathbf{X}^{opt}; H\mathbf{X}^{opt} + \mathbf{N}^{opt}) \\ &\leq I(\mathbf{X}^{opt}; H\mathbf{X}^{opt} + \mathbf{N}). \end{aligned} \tag{133}$$

Let $\mathbf{N}^* \sim \mathcal{N}(\mathbf{0}, R_{n^*})$ be the solution of

$$R_{n^*} = \arg\min_{R_n} I(\mathbf{X}^*; H\mathbf{X}^* + \mathbf{N}), \tag{134}$$

where $\mathbf{X}^* \sim \mathcal{N}(\mathbf{0}, \frac{P}{n_t} I_{n_t})$, which follows from [21]. The uncorrelation loss of MIMO-BC is given by

$$\begin{aligned} L_{un-coor} &= C_{BC}^{sum}(H, \mathbf{N}, P) - I(\mathbf{X}^*; H\mathbf{X}^* + \mathbf{N}^*) \\ &= I(\mathbf{X}^{opt}; H\mathbf{X}^{opt} + \mathbf{N}^{opt}) - I(\mathbf{X}^*; H\mathbf{X}^* + \mathbf{N}^*) \qquad (135) \\ &\leq I(\mathbf{X}^{opt}; H\mathbf{X}^{opt} + \mathbf{N}^*) - I(\mathbf{X}^*; H\mathbf{X}^* + \mathbf{N}^*) \qquad (136) \\ &\leq \sup_{\{\mathbf{X}: E||\mathbf{X}||^2 \leq P\}} I(\mathbf{X}; H\mathbf{X} + \mathbf{N}^*) - I(\mathbf{X}^*; H\mathbf{X}^* + \mathbf{N}^*) \qquad (137) \\ &\leq n_t C_G^*(n_t, n_r), \qquad (138) \end{aligned}$$

where (136) follows from the right hand side of (133); (137) follows from replacing $\mathbf{X}^{opt}$ by the optimal input with respect to $\mathbf{N}^*$; finally, (138) follows from Theorem 2.

The equality in (50) requires channel matrix H with rank $\min\{n_r, [[n_t/e]]\}$ and high SNR. In high SNR $|HR_xH^T + R_n| \approx |HR_xH^T|$, and for $R_n = I_{n_r}$ the uncorrelated noise achieves the minimal mutual information for any input. specifically, $R_{n^{opt}} = R_{n^*} = I_{n_r}$ (the Sato noise in high SNR is i.i.d Gaussian noise), thus

$$\begin{aligned} I(\mathbf{X}^{opt}; H\mathbf{X}^{opt} + \mathbf{N}^{opt}) &= \inf_{R_n} \sup_{R_x} I(\mathbf{X}; H\mathbf{X} + \mathbf{N}) \qquad (139) \\ &= \sup_{R_x} I(\mathbf{X}; H\mathbf{X} + \mathbf{N}^*), \qquad (140) \end{aligned}$$

which achieves the equality in (136) and (137). Finally for H with rank $\min\{n_r, [[n_t/e]]\}$ and high SNR, the equality in (138) is achieved from Theorem 2.

Non-cooperation loss: Let $C_{MAC}^{sum}(H, N_G, P)$ be the sum capacity of Gaussian MIMO-MAC with sum power constraint P, i.e,

$$C_{MAC}^{sum}(H, \mathbf{N}_G, P) = \sup_{\mathbf{P}: \sum_{i=1}^m P_i \leq P} C_{MAC}^{sum}(H, \mathbf{N}_G, \mathbf{P}). \tag{141}$$

Let $\mathbf{N}_W \sim \mathcal{N}(0, I_{n_r})$. Using the MAC-BC duality [27], [28], the sum capacity of Gaussian MIMO-BC is given by,

$$
\begin{aligned}
C_{BC}^{sum}(H,\mathbf{N},P) &= C_{MAC}^{sum}(H^T,\mathbf{N}_W,P) && (142)\\
&\geq C_{P2P}(H^T,\mathbf{N}_W,P) - n_r C_G^*(n_r,n_t) && (143)\\
&= C_{P2P}(H,\mathbf{N}_W,P) - n_r C_G^*(n_r,n_t), && (144)
\end{aligned}
$$

where (143) follows from Theorem 3; (144) follows from *reciprocity* property [18]. Therefore, the non-cooperation loss is given by,

$$
L_{non-coop} = C_{P2P}(H,\mathbf{N}_W,P) - C_{BC}^{sum}(H,\mathbf{N},P) \leq n_r C_G^*(n_r,n_t). \qquad (145)
$$

The equality is shown for channel matrix H with rank $\min\{n_t, [[n_r/e]]\}$ and high SNR, since in this case (143) is achieved with equality from Theorem 3.

References

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

[2] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multiple antennas,” *Bell Labs Technical Journal*, vol. Vol. 1, pp. 41–59, Autumn 1996.

[3] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, “Space-time block codes from orthogonal design,” *IEEE Trans. Information Theory*, vol. IT-45, pp. 1456–1567, Jul. 1999.

[4] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. U.K: Cambridge Univ. Press, 2005.

[5] E. Biglieri, J. Proakis, and S. S. (Shitz), “Fading channels: Information-theoretic and communications aspects,” *IEEE Trans. Information Theory*, vol. IT-44, pp. 2619–2693, Oct. 1998.

[6] J. Wolfowitz, *Coding Theorems of Information Theory*. New York: Springer-Verlag, 1964.

[7] A. Lapidoth and P. Narayan, “Reliable communication under channel uncertainy,” *IEEE Trans. Information Theory*, vol. IT-44, pp. 2148–2177, Oct. 1998.

[8] I. Csiszar and J. Korner, *Information Theory - Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.

[9] D. P. Palmoar, J. M. Cioffi, and M. Lagunas, “Uniform power allocation in MIMO channels: A game-theortic approach,” *IEEE Trans. Information Theory*, vol. IT-49, pp. 1707–1727, Jul. 2003.

[10] D. Hosli and A. Lapidoth, “How good is an isotropic Gaussian input on a MIMO Ricean channel?” *Presented at the ISIT 2004, Chicago, USA*, p. 291, Jul. 2004.

[11] L. D. Davisson, “Universal lossless coding,” *IEEE Trans. Information Theory*, vol. IT-19, pp. 783–795, Nov. 1973.

[12] J. Ziv, “Universal decoding for finite-state channels,” *IEEE Trans. Information Theory*, vol. IT-31, pp. 453–460, Jul. 1985.

[13] M. Feder and A. Lapidoth, “Universal decoding for channels with memory,” *IEEE Trans. Information Theory*, vol. IT-44, pp. 1726–1745, Sep. 1998.

[14] N. Shulman, “Communication over an unknown channel via common broadcasting,” Ph.D. dissertation, Tel-Aviv University, July 2003.

[15] U. Erez, G. W. Wornell, and M. D. Trott, “Faster-than-nyquist coding: The merits of a regime change,” in *Proceedings 42th Annual Allerton Conference on Communication, Control, and Computing, Univ. of Illinois, Urbana, IL, USA*, 2004.

[16] N. Shulman and M. Feder, “The uniform distribution as a universal prior,” *IEEE Trans. Information Theory*, vol. IT-50, pp. 1356–1362, 2004.

[17] R. Zamir and U. Erez, “A Gaussian input is not too bad,” *IEEE Trans. Information Theory*, vol. IT-50, pp. 1362–1367, Jun. 2004.

[18] E. Telatar, “Capacity of the multiple antenna Gaussian channel,” *Europ. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.

[19] T. Lan and W. Yu, “Downlink beamforming with per-antenna power constraints,” in *Sixth IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2005.

[20] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Information Theory*, vol. 49, pp. 1691–1706, Jul. 2003.

[21] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Information Theory*, vol. IT-50, pp. 1875–1892, Sep, 2004.

[22] T. Philosof and R. Zamir, "The capacity loss of uncorrelated equi-power Gaussian input over a MIMO channel," in *Proceedings 42th Annual Allerton Conference on Communication, Control, and Computing, Univ. of Illinois, Urbana, IL, USA*, Oct. 2004.

[23] ——, "The cost of uncorrelation and non-cooperation in MIMO channels," in *Proceedings of IEEE International Symposium on Information Theory, Adelaide, Australia*, Sep. 2005.

[24] C. T. K. Ng and A. Goldsmith, "Capacity gain from transmitter and receiver cooperation," in *Proceedings of IEEE International Symposium on Information Theory, Adelaide, Australia*, Sep. 2005.

[25] W. Yu, W. Rhee, S. Boyd, and J. Cioffi, "Iterative water-filling for Gaussian vector multiple access channels," *IEEE Trans. Information Theory*, vol. IT-50, pp. 145–151, Jan. 2004.

[26] H. Weingarten, Y. Steinberg, and S. S. (Shitz), "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Information Theory*, vol. IT-52, pp. 3936–3964, Sep. 2006.

[27] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Information Theory*, vol. IT-49, pp. 1912–1921, Aug, 2003.

[28] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Information Theory*, vol. IT-49, pp. 2658–2668, Oct, 2003.

[29] H. Sato, "An outer bound on the capacity region of broadcast channels," *IEEE Trans. Information Theory*, vol. IT-24, pp. 374–377, May 1978.

[30] R. A. Horn and C. R. Johnson, *Matrix analysis*. U.K.: Cambridge Univ. Press, 1985.

List of Figures

Tal Philosof received the B.Sc. degree in Electrical Engineering from Israel Institute of Technology (Technion) and the M.Sc degree in Electrical Engineering from Tel-Aviv University in 1997 and 2003, respectively. He is currently working toward the Ph.D. degree at the department of Electrical Engineering-Systems at Tel-Aviv University. His research interests include information theory, digital communications and signal processing.

Ram Zamir was born in Ramat-Gan, Israel in 1961. He received the B.Sc., M.Sc. (summa cum laude) and D.Sc. (with distinction) degrees from Tel-Aviv University, Israel, in 1983, 1991, and 1994, respectively, all in electrical engineering. In the years 1994 - 1996 he spent a post-doctoral period at Cornell University, Ithaca, NY, and at the University of California, Santa Barbara. In 2002 he spent a Sabbatical year at MIT. Since 1996 he has been with the department of Elect. Eng. - Systems at Tel Aviv University.

Dr. Zamir has been consulting in the areas of radar and communications, mainly in developing algorithms and in the design of signals, and has been teaching information theory, data compression, random processes, communications systems and communications circuits at Tel Aviv University.

Dr. Zamir received the Israel Ministry of Communications Award in 1993, and the Wolfson Post-Doctoral Research Award in 1994, and visited the Technical University of Budapest under a joint program of the Israeli and Hungarian academies of science in summer 1995. He served as an Associate Editor for Source Coding in the IEEE transactions on Information Theory (2001-2003), and headed the Information Theory Chapter of the Israeli IEEE society (2000-2005). His research interests include information theory, communication and remote sensing systems, and signal processing.