The Rate Loss of Single-Letter Characterization: The "Dirty" Multiple Access Channel

Tal Philosof and Ram Zamir [†]

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

Tel-Aviv 69978, ISRAEL

talp,zamir@eng.tau.ac.il

Submitted to IEEE Trans. on Information Theory March 2008

Abstract

For general memoryless systems, the typical information theoretic solution - when exists - has a "single-letter" form. This reflects the fact that optimum performance can be approached by a random code (or a random binning scheme), generated using independent and identically distributed copies of some single-letter distribution. Is that the form of the solution of any (information theoretic) problem? In fact, some counter examples are known. The most famous is the "two help one" problem: Korner and Marton showed that if we want to decode the modulo-two sum of two binary sources from their independent encodings, then linear coding is better than random coding. In this paper we provide another counter example, the "doubly-dirty" multiple access channel (MAC). Like the Korner-Marton problem, this is a multi-terminal scenario where side information is distributed among several terminals; each transmitter knows part of the channel interference but the receiver is not aware of any part of it. We give an explicit solution for the capacity region of a binary version of the doubly-dirty MAC, demonstrate how the capacity region can be approached using a linear coding scheme, and prove that the "best known single-letter region" is strictly contained in it. We also state a conjecture regarding a similar rate loss of single letter characterization in the Gaussian case.

Index Terms

Multi-user information theory, random binning, linear lattice binning, dirty paper coding, lattice strategies, Korner-Marton problem.

I. Introduction

Consider the two-user / double-state memoryless multiple access channel (MAC) with transition and state probability distributions

$$P(y|x_1, x_2, s_1, s_2) \quad \text{and} \quad P(s_1, s_2), \tag{1}$$

[†]This research was partially supported by BSF grant No-2004398

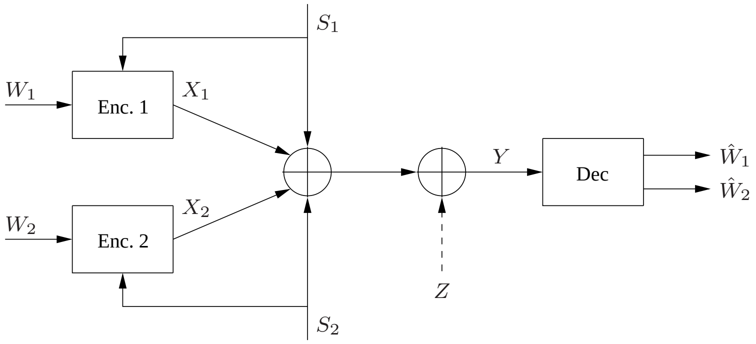

Fig. 1. Doubly-dirty MAC

respectively, where the states S_1 and S_2 are known non-causally to user 1 and user 2, respectively. A special case of (1) is the additive channel shown in Fig. 1. In this channel, called the *doubly-dirty MAC* (after Costa's "writing on dirty paper" [1]), the total channel noise consists of three independent components: S_1 and S_2, the interference signals, that are known to user 1 and user 2, respectively, and Z, the unknown noise, which is known to neither. The channel inputs X_1 and X_2 may be subject to some average cost constraint.

Neither the capacity region of (1) nor that of the special case of Fig. 1 are known. In this paper we consider a particular binary version of the doubly-dirty MAC of Fig. 1, where all variables are in $\mathbb{Z}_2$, i.e., $\{0, 1\}$, and the unknown noise $Z = 0$. The channel output of the binary doubly-dirty MAC is given by

$$Y = X_1 \oplus X_2 \oplus S_1 \oplus S_2, \tag{2}$$

where $\oplus$ denotes the mod 2 addition (xor), and S_1, S_2 are Bernoulli$(1/2)$ and independent. Each of the codewords $\mathbf{x}_i \in \mathbb{Z}_2^n$ is a function of the message W_i and the interference vector $\mathbf{s}_i \in \mathbb{Z}_2^n$, and must satisfy the input constraint, $\frac{1}{n} w_H(\mathbf{x}_i) \leq q_i,\ \ i = 1, 2$, where $0 \leq q_1, q_2 \leq 1/2$ and $w_H(\cdot)$ is the Hamming weight. The coding rates R_1 and R_2 of the two users are given as usual by $R_i = \frac{1}{n}\log|\mathcal{W}_i|$, where $\mathcal{W}_i$ is the set of messages of user i, and n is the length of the codeword.

The double state MAC (1) generalizes the point to point channel with side information (SI) at the transmitter considered by Gel'fand and Pinsker [2]. They prove their direct coding theorem using the framework of random binning, which is widely used in the analysis of multi-terminal source and channel coding problems [3]. They obtain a general capacity expression which involves an auxiliary random variable U:

$$C = \max_{P(u,x|s)} \{H(U|S) - H(U|Y)\} \tag{3}$$

where the maximization is over all the joint distributions of the form $p(u, s, y, x) = p(s)p(u, x|s)p(y|x, s)$.

The channel in (1) with only one informed encoder (i.e., where $S_2 = \{\emptyset\}$) was considered recently by Somekh-Baruch et al. [4] and Kotagiri and Laneman [5]. The common message ($W_1 = W_2$) capacity of this channel is known [4], and it involves using random binning by the informed user. For the binary "one dirty user" case (i.e.,

(2) with $S_2 = 0$), we show that Somekh-Baruch's common-message capacity becomes (see Appendix I)

$$C_{com} = H_b(q_1), \tag{4}$$

where $H_b(x) \triangleq -x\log_2(x) - (1-x)\log_2(1-x)$ is the binary entropy function. Clearly, the doubly-dirty individual-message case is harder. Thus, it follows from (4) that the rate-sum in the setting of Fig. 1 is upper bounded by

$$R_1 + R_2 \leq \min\Big\{H_b(q_1), H_b(q_2)\Big\}. \tag{5}$$

In Theorem 1 we show that this upper bound is in fact tight.

One approach to find *achievable rates* for the doubly-dirty MAC, is to extend the Gel'fand and Pinsker solution [2] to the two-user / double-state case. As shown by Jafar [6], this extension leads to the following pentagonal inner bound for the capacity region of (1):

$$\begin{aligned}\mathcal{R}(U_1,U_2) \triangleq \Bigg\{(R_1,R_2) : R_1 &\leq I(U_1,Y|U_2) - I(U_1;S_1)\\ R_2 &\leq I(U_2,Y|U_1) - I(U_2;S_2)\\ R_1 + R_2 &\leq I(U_1,U_2,Y) - I(U_1;S_1) - I(U_2;S_2)\Bigg\}\end{aligned} \tag{6}$$

for some $P(U_1,U_2,X_1,X_2|S_1,S_2) = P(U_1,X_1|S_1)P(U_2,X_2|S_2)$. In fact, by a standard time-sharing argument [3], the closure of the convex hull of the set of all rate pairs (R_1,R_2) satisfying (7),

$$\mathcal{R}_{BSL} \triangleq cl\ conv\ \Bigg\{(R_1,R_2) \in \mathcal{R}(U_1,U_2) : P(U_1,U_2,X_1,X_2|S_1,S_2) = P(U_1,X_1|S_1)P(U_2,X_2|S_2)\Bigg\}, \tag{7}$$

is also achievable[1]. To the best of our knowledge, the set $\mathcal{R}_{BSL}$ is the best currently known single-letter characterization for the rate region of the MAC with side information at the transmitters (1), and in particular, for the doubly-dirty MAC (2)[2]. The achievability of (7) can be proved, as usual, by an i.i.d random binning scheme [6].

A different method to cancel known interference is by "linear strategies", i.e, binning based on the cosets of a linear code [8], [9], [10]. In the sequel, we show that the outer bound (5) can indeed be achieved by a linear coding scheme. Hence, the set of rate pairs (R_1,R_2) satisfying (5) is the capacity region of the binary doubly-dirty MAC. In contrast, we show that the single-letter region (7) is *strictly contained* in this capacity region. Hence, a random binning scheme based on this extension of the Gel'fand-Pinsker solution [2] is not optimal for this problem.

A similar observation has been made by Korner-Marton [11] for the "two help one" source coding problem. For a specific binary version known as the "modulo-two sum" problem, they showed that the minimum possible

[1] As in the Gel'fand and Pinsker solution, for a finite alphabet system it is enough to optimize over auxiliary variables U_1 and U_2 whose alphabet size is bounded in terms of the size of the input and state alphabets.

[2] For the case where the users have also a common message W_0 to be transmitted jointly by both encoders, (7) can be improved by adding another auxiliary random variable U_0 which plays the role of the common auxiliary r.v. in Marton's inner bound for the non-degraded broadcast channel [7]. In this case, the joint distribution of (U_0,U_1,U_2) is given by $P(U_0,U_1,U_2) = P(U_0)P(U_1|U_0)P(U_2|U_0)$, i.e, U_1 and U_2 are conditionally independent given U_0.

rate sum is achieved by a linear coding scheme, while the best known single-letter expression for this problem is strictly higher. See the discussion in [11, Section IV] and in the end of Section III.

Although the "*single-letter characterization*" is a fundamental concept in information theory, it has not been generally defined [12, p.35]. Csiszar and Korner [13, p.259] suggested to define it through the notion of *computability*, i.e., a problem has a single-letter solution if there exists an algorithm which can decide if a point belongs to an ε-neighborhood of the achievable rate region with polynomial complexity in $1/\varepsilon$. Since we are not aware of any other computable solution to our problem, we shall refer to (7) as the *"best known single-letter characterization"*.

An extension of these observations to continuous channels would be of interest. Costa [1] considered the single-user case of the dirty channel problem $Y = X + S + Z$, where the interference S and the noise Z are assumed to be i.i.d. Gaussian with variances Q and N, respectively, and the input X is subject to a power constraint P. He showed that in this case, the transmitter side-information capacity (3) coincides with the zero-interference capacity $\frac{1}{2}\log_2(1 + SNR)$, where $SNR = P/N$. Selecting the auxiliary random variable U in (3) such that

$$U = X + \alpha S, \tag{8}$$

where X and S are independent, and taking $\alpha = \frac{P}{P+N}$, the formula (3) and its associated random binning scheme are capacity achieving. The continuous (Gaussian) version of the doubly-dirty MAC of Fig. 1 was considered in [10]. It was shown that by using a linear structure, i.e., lattice strategies [8], the full capacity region is achieved in the limit of high SNR and high lattice dimension. In contrast, it was shown that for $Q \to \infty$ no positive rate is achievable by using the natural generalization of Costa's strategy (8) to the two user case, while a (scalar) modulo addition version of (8) looses ≈ 0.254 bit in the sum capacity. We shall further elaborate on this issue in Section IV.

Similar observations regarding the advantage of modulo-lattice modulation with respect to a separation based solution were made by Nazer and Gastpar [14], in the context of computation over linear Gaussian networks, and also by Krithivasan and Pradhan [15] for multi-terminal rate distortion problems.

The paper is organized as follows. In Section II the capacity region for the binary doubly-dirty MAC (2) is derived, and linear coding is shown to be optimal. Section III develops a closed form expression for the best known single-letter characterization (7) for this channel, and demonstrates that it is strictly contained in the the true capacity region. In Section IV we consider the Gaussian doubly-dirty MAC, and state a conjecture regarding the capacity loss of single-letter characterization in this case.

II. The Capacity Region of the Binary Doubly-Dirty MAC

The following theorem characterizes the capacity region of the binary doubly-dirty MAC of Fig. 1.

Theorem 1. *The capacity region of the binary doubly-dirty MAC (2) is the set of all rate pairs (R_1, R_2) satisfying*

$$\mathcal{C}(q_1, q_2) \triangleq \left\{ (R_1, R_2) : R_1 + R_2 \leq \min \left\{ H_b(q_1), H_b(q_2) \right\} \right\}. \tag{9}$$

Proof: ***The converse part:*** As explained in the Introduction (5), one way to derive an upper bound for the rate-sum is through the general one-dirty-user capacity formula [4], which we derive explicitly for the binary case in Appendix I. Here we show directly the converse part, which is similar to the proof of the outer bound for the Gaussian case in [16], [10]. We assume that user 1 and user 2 intend to transmit a common message W. An upper bound on the rate of this message clearly upper bounds the sum rate $R_1 + R_2$ in the individual messages case. Thus,

$$\begin{aligned}
n(R_1+R_2) &\leq H(W) \\
&= H(W|Y^n) + I(W;Y^n) \\
&\leq I(W;Y^n) + n\epsilon_n && (10)\\
&= H(Y^n) - H(Y^n|W) + n\epsilon_n \\
&= H(Y^n) - H(Y^n|W,S_1^n,S_2^n) - I(S_1^n,S_2^n;Y^n|W) + n\epsilon_n \\
&= H(Y^n) - I(S_1^n,S_2^n;Y^n|W) + n\epsilon_n && (11)\\
&= H(Y^n) - H(S_1^n,S_2^n|W) + H(S_1^n,S_2^n|W,Y^n) + n\epsilon_n \\
&\leq -n + H(S_1^n|W,Y^n) + H(S_2^n|W,Y^n,S_1^n) + n\epsilon_n && (12)\\
&\leq H(X_1^n \oplus X_2^n \oplus S_1^n|W,Y^n,S_1^n) + n\epsilon_n && (13)\\
&= H(X_2^n|W,Y^n,S_1^n) + n\epsilon_n && (14)\\
&\leq nH_b(q_2) + n\epsilon_n, && (15)
\end{aligned}$$

where (10) follows from Fano's inequality where $\epsilon_n \to 0$ as the error probability $P_e^{(n)}$ goes to zero for $n \to \infty$; (11) follows since Y is fully known given W, S_1 and S_2; (12) follows from the chain rule for entropy, and due to $H(Y^n) \leq n$ and $H(S_1^n, S_2^n|W) = H(S_1^n) + H(S_2^n) = 2n$ since W, S_1^n and S_2^n are mutually independent; (13) follows since $H(S_1^n|W,Y^n) \leq n$ and $Y^n = X_1^n \oplus X_2^n \oplus S_1^n \oplus S_2^n$; (14) follows since X_1^n is a function of (W, S_1^n), finally (15) follows since $H(X_2^n|W,Y^n,S_1^n) \leq H(X_2^n) \leq nH_b(q_2)$.

In the same way we can show that $R_1 + R_2 \leq H_b(q_1) + \epsilon_n$. The converse part follows since for $n \to \infty$ we have that $\epsilon_n \to 0$, thus $P_e^{(n)} \to 0$.

The direct part is based on the scheme for the point-to-point binary dirty paper channel [9]. We define $q \triangleq \min\{q_1, q_2\}$. In view of the converse part, it is sufficient to show achievability of the point $(R_1, R_2) = (H_b(q), 0)$, since the outer bound may be achieved by time sharing with the symmetric point $(R_1, R_2) = (0, H_b(q))$. The corner point $(R_1, R_2) = (H_b(q), 0)$ corresponds to the "helper problem", i.e., user 2 tries to help user 1 to transmit at its highest rate. The encoders and decoder are described using a binary linear code $\mathcal{C}(n,k)$ with parity check matrix H. Let $\mathbf{v} \in \mathbb{Z}_2^{n-k}$ be a syndrome of the code $\mathcal{C}$, where we note that each syndrome represents a different coset of the linear code $\mathcal{C}$. Let $f(\mathbf{v})$ denote the "leader" of (or the minimum weight vector in) the coset associated with the syndrome $\mathbf{v}$ [17, Chap. 6], hence $f : \{0,1\}^{n-k} \to \{0,1\}^n$. For $\mathbf{a} \in \mathbb{Z}_2^n$, we define the n-dimensional modulo operation over the code $\mathcal{C}$ as

$$\mathbf{a} \bmod \mathbb{C} \triangleq f(H\mathbf{a}),$$

which is the leader of the coset to which the vector $\mathbf{a}$ belongs.

- **Encoder of user 1:** Let the transmitted message $\mathbf{v}_1 \in \mathbb{Z}_2^{n-k}$ be a syndrome in $\mathcal{C}$, and let $\tilde{\mathbf{x}}_1 = f(\mathbf{v}_1)$ be its coset leader. In particular $\mathbf{v}_1 = H\tilde{\mathbf{x}}_1$. Transmit the modulo of the code $\mathcal{C}$ with respect to the difference between $\tilde{\mathbf{x}}_1$ and $\mathbf{s}_1$, i.e.,

$$\mathbf{x}_1 = (\tilde{\mathbf{x}}_1 \oplus \mathbf{s}_1) \text{ mod } \mathbb{C} = f(\mathbf{v}_1 \oplus H\mathbf{s}_1).$$

- **Encoder of user 2:** (functions as a "helper" for user 1). Transmit

$$\mathbf{x}_2 = \mathbf{s}_2 \text{ mod } \mathbb{C} = f(H\mathbf{s}_2).$$

- **Decoder:**

  1. Reconstruct $\tilde{\mathbf{x}}_1$ by $\hat{\tilde{\mathbf{x}}}_1 = \mathbf{y} \bmod \mathbb{C}$.

  2. Reconstruct the transmitted coset of user 1 by $\hat{\mathbf{v}}_1 = H\hat{\tilde{\mathbf{x}}}_1$.

  In fact, the transmitted coset can be reconstructed directly as $\hat{\mathbf{v}}_1 = H\hat{\tilde{\mathbf{x}}}_1 = H(\mathbf{y} \bmod \mathbb{C}) = \text{H}\mathbf{y}$, where the last equality follows since $\mathbf{y} \bmod \mathbb{C}$ and $\mathbf{y}$ are in the same coset.

It follows that the decoder correctly decodes the message coset $\mathbf{v}_1$, since

$$\begin{aligned}\hat{\mathbf{v}}_1 &= H \cdot \Big(\mathbf{y} \text{ mod } \mathbb{C}\Big)\\ &= H \cdot \Big([\tilde{\mathbf{x}}_1 \oplus \mathbf{s}_1 \oplus \mathbf{s}_2 \oplus \mathbf{s}_1 \oplus \mathbf{s}_2] \text{ mod } \mathbb{C}\Big)\\ &= H\tilde{\mathbf{x}}_1\\ &= \mathbf{v}_1,\end{aligned}$$

where the third equality follows since $\tilde{\mathbf{x}}_1$ and $\tilde{\mathbf{x}}_1 \bmod \mathbb{C}$ are in the same coset. It is left to relate the coding rate $R_1 = \frac{1}{n}\log\left(\left|\{0,1\}^{n-k}\right|\right) = 1 - k/n$ to the input constraint q. Form [18], there exists a binary linear code with covering radius ρ that satisfies $\frac{k}{n} \leq 1 - H_b(\rho/n) + \epsilon$ where $\epsilon \to 0$ as $n \to \infty$. The achievability of the point $(H_b(q), 0)$ follows by using $q = \rho/n$, thus $R_1 = 1 - k/n \geq H_b(q) - \epsilon$, while $w_H(\mathbf{x}_1) = w_H(f(\mathbf{v}_1 \oplus H\mathbf{s}_1)) \leq \rho$ and $w_H(\mathbf{x}_2) = w_H(f(H\mathbf{s}_2)) \leq \rho$, hence

$$\begin{aligned}\frac{1}{n}Ew_H\{\mathbf{x}_1\} &= \frac{1}{n}Ew_H\{f(\mathbf{v}_1 \oplus H\mathbf{s}_1)\} \leq q\\ \frac{1}{n}Ew_H\{\mathbf{x}_2\} &= \frac{1}{n}Ew_H\{f(H\mathbf{s}_2)\} \leq q.\end{aligned}$$

This completes the proof of the direct part of the theorem. □

As stated above, the achievability for the capacity region follows by time sharing the corner points $(H_b(q), 0)$ and $(0, H_b(q))$ where $q = \min\{q_1, q_2\}$. It is also interesting to see how to achieve the rate sum $H_b(q)$ for an arbitrary rate pair (R_1, R_2) without time sharing. For that, let the message of user 1 be $\mathbf{m}_1 \in \mathbb{Z}_2^{l_1}$ and the message of user

2 be $\mathbf{m}_2 \in \mathbb{Z}_2^{l_2}$ where $l_1 + l_2 = n - k$. We define the following syndromes in $\mathcal{C}$

$$\mathbf{v}_1 \triangleq [\mathbf{m}_1 \; \underbrace{0\,0\,\ldots\,0}_{l_2}] \in \mathbb{Z}_2^{n-k}$$
$$\mathbf{v}_2 \triangleq [\underbrace{0\,0\,\ldots\,0}_{l_1} \; \mathbf{m}_2] \in \mathbb{Z}_2^{n-k}$$
$$\mathbf{v} \triangleq \mathbf{v}_1 \oplus \mathbf{v}_2.$$

Clearly, given the syndrome $\mathbf{v}$ the syndromes $\mathbf{v}_1$ and $\mathbf{v}_2$ are fully known and the messages $\mathbf{m}_1$ and $\mathbf{m}_2$ as well. Let $\tilde{\mathbf{x}}_i = f(\mathbf{v}_i)$ be the coset leader of $\mathbf{v}_i$ for $i = 1, 2$. In this case the transmission scheme is as follow:

- Encoder of user 1: transmit $\mathbf{x}_1 = (\tilde{\mathbf{x}}_1 \oplus \mathbf{s}_1) \bmod \mathbb{C} = f(\mathbf{v}_1 \oplus H\mathbf{s}_1)$.
- Encoder of user 2: transmit $\mathbf{x}_2 = (\tilde{\mathbf{x}}_2 \oplus \mathbf{s}_2) \bmod \mathbb{C} = f(\mathbf{v}_2 \oplus H\mathbf{s}_2)$.
- Decoder: reconstruct $\hat{\mathbf{v}} = H \cdot \Big(\mathbf{y} \bmod \mathbb{C}\Big)$.

Therefore, we have that

$$\begin{aligned}\hat{\mathbf{v}} &= H \cdot \Big(\mathbf{y} \bmod \mathbb{C}\Big) \\ &= H \cdot \Big(\tilde{\mathbf{x}}_1 \oplus \tilde{\mathbf{x}}_2\Big) = \mathbf{v}_1 \oplus \mathbf{v}_2 = \mathbf{v}.\end{aligned}$$

The sum capacity is achieved since $R_1 + R_2 = \frac{l_1+l_2}{n} = \frac{n-k}{n} \geq H_b(q) - \epsilon$ where $\epsilon \to 0$ as $n \to \infty$ which satisfies the input constraints.

III. A Single-Letter Characterization for the Capacity Region

In this section we characterize the best known single-letter region (7) for the binary doubly-dirty MAC (2), and show that it is strictly contained in the capacity region (9). For simplicity, we shall assume identical input constraints, i.e., $q_1 = q_2 = q$.

Definition 1. *For a given q, the best known single-letter rate region for the binary doubly-dirty MAC* (2)*, denoted by $\mathcal{R}_{BSL}(q)$, is the set of all rate pairs (R_1, R_2) satisfying* (7) *with the additional constraints that $E\mathbf{X}_1, E\mathbf{X}_2 \leq q$.*

In the following theorem we give a closed form expression for $\mathcal{R}_{BSL}(q)$.

Theorem 2. *The best known single-letter rate region for the binary doubly-dirty MAC* (2) *is a triangular region given by*

$$\mathcal{R}_{BSL}(q) = \Bigg\{(R_1, R_2) : R_1 + R_2 \leq u.c.e\Big[2H_b(q) - 1\Big]^+\Bigg\}, \tag{16}$$

where $u.c.e$ is the upper convex envelope with respect to q, and $[x]^+ \triangleq \max\{0, x\}$.

Fig. 2 shows the sum capacity of the binary doubly-dirty MAC (9) versus the best known single-letter rate sum (16) for equal input constraints. The latter is strictly contained in the capacity region which is achieved by a linear code. The quantity $[2H_b(q) - 1]^+$ is not a convex - $\cap$ function with respect to q. The upper convex envelope of

$[2H_b(q)-1]^+$ is achieved by time-sharing between the points $q=0$ and $q=q^* \triangleq 1-1/\sqrt{2}$, therefore it is given by

$$R_1+R_2 \leq \begin{cases} 2H_b(q)-1, & q^* \leq q \leq 1/2 \\ C^* q, & 0 \leq q \leq q^* \end{cases}, \tag{17}$$

where $C^* \triangleq \frac{2H_b(q^*)-1}{q^*}$.

Proof: ***The direct part*** is shown by choosing in (6) $U_1 = S_1 \oplus X_1$ and $U_2 = S_2 \oplus X_2$, where $X_1, X_2 \sim$ Bernoulli(q) and X_1, X_2, S_1, S_2 are independent. From (6) the achievable rate sum is given by

$$\begin{aligned} R_1+R_2 &= I(U_1,U_2;Y) - I(U_1,U_2;S_1,S_2) \\ &= H(U_1|S_1) + H(U_2|S_2) - H(U_1,U_2|U_1 \oplus U_2) && (18) \\ &= H(U_1|S_1) + H(U_2|S_2) - H(U_1|U_1 \oplus U_2) - H(U_2|U_1 \oplus U_2, U_1) && (19) \\ &= H(X_1) + H(X_2) - H(U_1|U_1 \oplus U_2) && (20) \\ &= 2H_b(q) - 1, && (21) \end{aligned}$$

where (18) follows since $Y = U_1 \oplus U_2$; (19) follows from the chain rule for entropy; (20) follows since U_2 is fully known given $U_1 \oplus U_2, U_1$ thus $H(U_2|U_1 \oplus U_2, U_1) = 0$; (21) follows since $H(X_i) \leq H_b(q)$ and since U_1, U_2 are independent with $P(U_i = 1) = 1/2$ thus $H(U_1|U_1 \oplus U_2) = H(U_1) = 1$.

The converse part of the proof is given in Appendix II. □

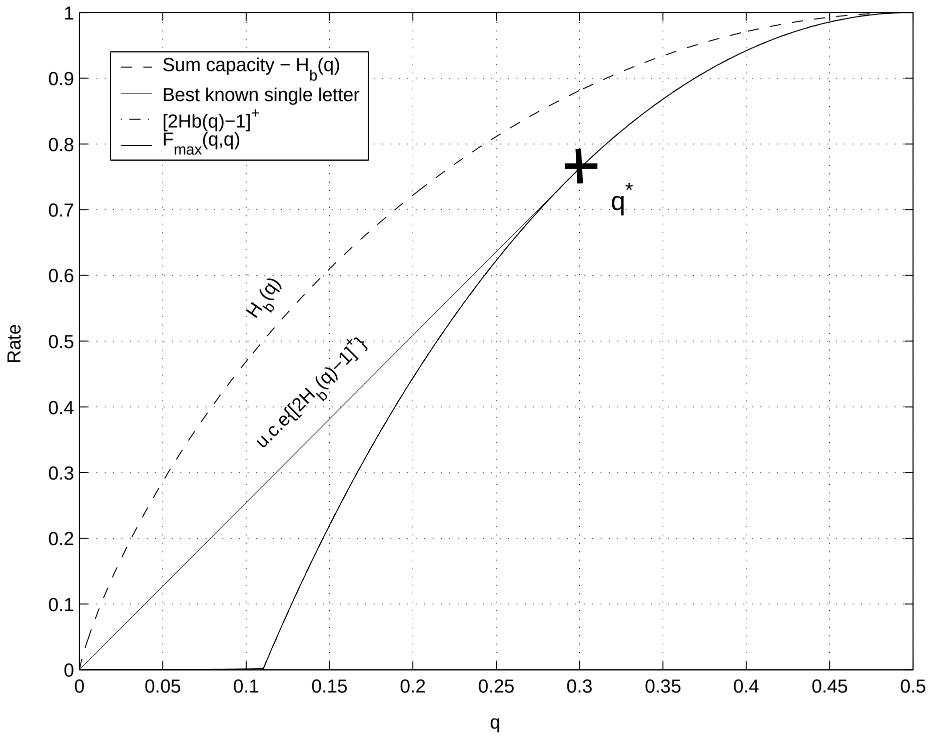

Fig. 2. The rate sum of binary doubly-dirty MAC vs. best known single-letter rate sum with input constraints $EX_1, EX_2 \leq q$.

We see that the binary doubly-dirty MAC is a memoryless channel coding problem, where the capacity region is achievable by a linear code, while the best known single-letter rate region is strictly contained in the capacity region. This may be explained by the fact that each user has only *partial* side information, and distributed random binning is unable to capture the linear structure of the channel.

In order to understand the limitation of random binning versus a linear code, we consider these two schemes for high enough q, that is $2H_b(q) - 1 \geq 0$. The random binning scheme uses $U_i = X_i \oplus S_i$ where $X_i \sim$ Bernoulli(q) and $S_i \sim$ Bernoulli$(1/2)$ are independent, therefore $Y = U_1 \oplus U_2$ where $U_i \sim$ Bernoulli$(1/2)$ for $i = 1, 2$. Each transmitter maps the message (bin) W_i into a codeword $\mathbf{u}_i$ which is with high probability at a Hamming distance of nq from $\mathbf{s}_i$. Therefore, given the vectors $(\mathbf{s}_1^n, \mathbf{s}_2^n)$, the available *input space* is approximately $2^{nH(U_1,U_2|S_1,S_2)} = 2^{nH(X_1,X_2)} = 2^{2nH_b(q)}$. Given the received vector $\mathbf{y}$, the *residual ambiguity* is given by $2^{nH(U_1,U_2|Y)} = 2^{n[H(U_1|Y)+H(U_2|Y,U_1)]} = 2^n$, since $H(U_1|Y) = 1$ and $H(U_2|Y, U_1) = 0$. As a result, the achievable rate sum is given by

$$R_1 + R_2 = \frac{1}{n}\log_2\left(\frac{|\text{input space}|}{|\text{residual ambiguity space}|}\right) \approx 2H_b(q) - 1.$$

The linear coding scheme shown in Theorem 1 has the same input space size as the random binning scheme, i.e., $2^{2nH_b(q)}$, since each user has $2^{nH_b(q)}$ cosets. However, given the received vector $\mathbf{y}$ there are $2^{nH_b(q)}$ possible pairs of cosets, i.e., the residual ambiguity is only $2^{nH_b(q)}$. Therefore, the linear code achieves rate sum of $R_1 + R_2 \approx 2H_b(q) - H_b(q) = H_b(q)$. The advantage of the linear coding scheme results from the "ordered structure" of the linear code, which decreases the residual ambiguity from 1 bit in random coding to $H_b(q)$.

The following example illustrates the above arguments for the case that user 2 is a "helper" for user 1, i.e, $R_2 = 0$, and user 1 transmits at his highest rate for each technique (random binning or linear coding). Table I summarizes the rates and codebooks sizes for each user for $q = 0.3$, that is $H_b(q) \approx 0.88$ bit.

	Random binning	Linear code
Rate sum	$2H_b(q) - 1 = 0.76$ bit	$H_b(q) = 0.88$ bit
Codewords per bin/coset	$2^{nI(U_i;S_i)} = 2^{n[1-H_b(q)]} = 2^{0.12n}$	$2^{n[1-H_b(q)]} = 2^{0.12n}$
Helper (user 2) - codebook size	$2^{nI(U_2;S_2)} = 2^{n[1-H_b(q)]} = 2^{0.12n}$	$2^{n[1-H_b(q)]} = 2^{0.12n}$
User 1 - codebook size	$2^{0.76n}2^{0.12n} = 2^{0.88n}$	$2^{0.12n}2^{0.88n} = 2^n$
Number of possible codeword pairs	$2^{0.88n}2^{0.12n} = 2^n$	$2^n 2^{0.12n} = 2^{1.12n}$

TABLE I

RANDOM BINNING AND LINEAR CODING SCHEMES CODEBOOKS SIZES FOR THE HELPER PROBLEM WITH $q = 0.3$.

Korner and Marton [11] observed a similar behavior for the "two help one" source coding problem shown in Fig. 3. In this problem, there are three binary sources X, Y, Z, where $Z = X \oplus Y$, and the joint distribution of X and Y is symmetric with $P(X \neq Y) = \theta$. The goal is to encode the sources X and Y separately such that Z can

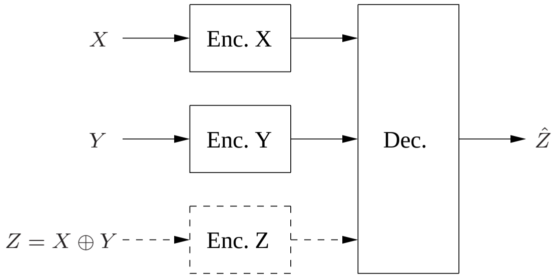

Fig. 3. The Korner-Marton configuration.

be reconstructed losslessly. Korner and Marton showed that the rate sum required is at least

$$R_x + R_y \geq 2H(Z), \tag{22}$$

and furthermore, this rate sum can be achieved by a linear code: each encoder transmits the syndrome of the observed source relative to a good linear binary code for a BSC with crossover probability θ.

In contrast, the "one help one" problem [19], [20] has a closed single-letter expression for the rate region, which corresponds to a random binning coding scheme. Korner and Marton [11] generalize the expression of [19], [20] to the "two help one" problem, and show that the minimal rate sum required using this expression is given by

$$R_x + R_y \geq H(X, Y). \tag{23}$$

The region (23) corresponds to Slepian-Wolf encoding of X and Y, and it can also be derived from the Burger-Tung achievable region [21] for distributed coding for X and Y with one reconstruction $\hat{Z}$ under the distortion measure $d(X, Y, \hat{Z}) \triangleq X \oplus Y \oplus \hat{Z}$. Clearly, the region (6) is strictly contained in the Korner-Marton region $R_x + R_y \geq 2H(Z)$ (22) (since $H(X, Y) = 1 + H(Z) > 2H(Z)$ for $Z \sim$ Bernoulli(θ), where $\theta \neq \frac{1}{2}$). For further background on related source coding problems, see [15].

IV. The Gaussian Doubly-Dirty MAC

In this section we introduce our conjecture regarding the rate loss of the best known single-letter characterization for the capacity region of the two-user Gaussian doubly-dirty MAC at high SNR. The Gaussian doubly-dirty MAC [10] is given by

$$Y = X_1 + X_2 + S_1 + S_2 + Z, \tag{24}$$

where $Z \sim \mathcal{N}(0, N)$ is independent of X_1, X_2, S_1, S_2, and where user 1 and user 2 must satisfy the power constraints, $\frac{1}{n}\sum_{i=1}^{n} X_{1_i}^2 \leq P_1$ and $\frac{1}{n}\sum_{i=1}^{n} X_{2_i}^2 \leq P_2$ see Fig. 1. The interference signals S_1 and S_2 are known non-causally to the transmitters of user 1 and user 2, respectively. We shall assume that S_1 and S_2 are independent Gaussian with variances going to infinity, i.e., $S_i \sim \mathcal{N}(0, Q_i)$ where $Q_i \rightarrow \infty$ for $i = 1, 2$. The signal to noise ratios for the two users are $SNR_1 = \frac{P_1}{N}$ and $SNR_2 = \frac{P_2}{N}$.

The capacity region at high SNR, i.e., $SNR_1, SNR_2 \gg 1$, is given by [10],

$$R_1 + R_2 \leq \frac{1}{2}\log_2\left(\frac{\min\{P_1, P_2\}}{N}\right), \tag{25}$$

and it is achievable by a modulo lattice coding scheme of dimension going to infinity. In contrast, it was shown in [10] that at high SNR and strong independent Gaussian interferences, the natural generalization of Costa's strategy (8) for the two users case, i.e., with auxiliary random variables $U_1 = X_1 + S_1$ and $U_2 = X_2 + S_2$, is not able to achieve *any positive rate*. A better choice for U_1 and U_2 suggested in [10] is a modulo version of Costa's strategy (8),

$$U_i^* = [X_i + S_i] \bmod \Delta_i, \tag{26}$$

where $\Delta_i = \sqrt{12P_i}$, and where $X_i \sim \text{Unif}\left([-\frac{\Delta_i}{2}, \frac{\Delta_i}{2})\right)$ is independent of S_i, for $i = 1, 2$. In this case the rate loss with respect to (25) is $\frac{1}{2}\log_2\left(\frac{\pi e}{6}\right) \approx 0.254$ bit.

The best known single-letter capacity region for the Gaussian doubly-dirty MAC (24) is defined as the set of all rate pairs (R_1, R_2) satisfying (7), where X_1 and X_2 are restricted to the power constraints $EX_1^2 \leq P_1$ and $EX_2^2 \leq P_2$. We believe that for high SNR and strong interference, the modulo-Δ strategy (26) is an optimum choice for (X_1, X_2, U_1, U_2) in (7) for the Gaussian doubly-dirty MAC. This implies the following conjecture about the rate loss of the best known single-letter characterization.

Conjecture 1. *For the Gaussian doubly-dirty MAC, at high SNR and strong interference, the best known single-letter expression* R_{BSL}^{sum} (7) *looses*

$$C^{sum} - R_{BSL}^{sum} = \frac{1}{2}\log_2\left(\frac{\pi e}{6}\right) \approx 0.254\ \textit{bit}, \tag{27}$$

with respect to the sum capacity C^{sum} (25).

Note that the right hand side of (27) is the well known "shaping loss" [22] (equivalent to a $1.53dB$ power loss).

A heuristic approach to attack the proof of this conjecture is to follow the steps of the proof of the converse part in the binary case (Theorem 2). First, in Lemma 6 we derive a simplified single-letter formula, $\overline{G}_{max}(P_1, P_2)$, which is analogous to Lemma 1 in the binary case. The next step would be to optimize this expression. However, an optimal choice for the auxiliary random variables V_1, V_1', V_2, V_2' (provided in the binary case by Lemma 2 and Lemma 3) is unfortunately still missing for the Gaussian case. The expression in Lemma 6 is close in spirit to the point-to-point *dirty tape* capacity for high SNR and strong interference [8]. In [8] it is shown that optimizing the capacity is equivalent to minimum *entropy-constrained scalar quantization* in high resolution, which is achieved by a lattice quantizer. Clearly, if we could show a similar lemma for the two variable pairs in the maximization of Lemma 6, i.e., that it is achieved by a pair of lattice quantizers, then the conjecture would be an immediate consequence.

It should be noted that the above discussion is valid only for strong interferences S_1 and S_2. For interference with finite power, it seems that cancelling the interference part of the time and staying silence the rest of the time (like in the time-sharing region $0 \leq q \leq q^*$ in the binary case) may achieve better rates.

V. Summary

A memoryless information theoretic problem is considered open as long as we are missing a general single-letter characterization for its information performance. This goes hand in hand with the optimality of the random coding approach for those problems which are currently solved. We examined this traditional view for the memoryless doubly-dirty MAC.

In the binary case, we showed that the best known single letter characterization is strictly contained in the region achievable by linear coding, and that the latter is in fact the full capacity region of the problem. In the Gaussian case, we conjectured that the best known single-letter characterization suffers an inherent rate loss (equal to the well known "shaping loss" $0.5\log(\pi e/6)$), and we provide a partial proof. This is in contrast to the asymptotic optimality (dimension $\rightarrow \infty$) of lattice strategies, as recently shown in [10].

The underlying reason for these performance gaps is that *random* binning is in general not optimal when side information is distributed among more than one terminal in the network. In the specific case of the doubly-dirty MAC (like in Korner-Marton's modulo-two sum problem [11] and similar settings [14], [15]), the linear structure of the network allows to show that *linear binning* is not only better, but it is capacity achieving.

Appendix I
A Closed Form Expression for the Capacity of the Binary MAC with One Dirty User

We consider the binary dirty MAC (2) with $S_2 = 0$,

$$Y = X_1 \oplus X_2 \oplus S_1, \tag{28}$$

where $S_1 \sim$ Bernoulli(1/2) is known non-causally at the encoder of user 1 with the input constraints $\frac{1}{n}W_H(\mathbf{x}_i) \leq q_i$ for $i = 1, 2$. We show that the common message ($W_1 = W_2 = W$) capacity of this channel is given by

$$C_{com} = H_b(q_1). \tag{29}$$

To prove (29), consider the general expression for the common message capacity of the MAC with one informed user [4], given by

$$C_{com} = \max_{U_1, X_1, X_2} \{I(U_1, X_2; Y) - I(U_1, X_2; S_1)\}, \tag{30}$$

where the maximization is over al the joint distributions

$$P(S_1, X_1, X_2, U_1, Y) = P(S_1)P(X_2)P(U_1|X_2, S_1)P(X_1|S_1, U_1)P(Y|X_1, X_2, S_1).$$

The converse part of (29) follows since for any U_1, X_1, X_2, the common message rate R_{com} can be upper bounded by

$$
\begin{aligned}
R_{com} &= I(U_1, X_2; Y) - I(U_1, X_2; S_1) \\
&= H(S_1|U_1, X_2) - H(Y|U_1, X_2) + H(Y) - H(S_1) \\
&\leq H(S|U_1, X_2) - H(Y|U_1, X_2) && (31) \\
&= H(S_1|U_1, X_2) - H(X_1 \oplus S_1|U_1, X_2) && (32) \\
&= H(S_1|T) - H(X_1 \oplus S_1|T) && (33) \\
&= E_T\Big\{H(S_1|T=t) - H(X_1 \oplus S_1|T=t)\Big\} && (34) \\
&= E_T\Big\{H_b(\alpha_t) - H_b(\beta_t)\Big\}, && (35)
\end{aligned}
$$

where (31) follows since $H(Y) \leq 1$ and $H(S_1) = 1$; (32) follows since $Y = X_1 \oplus X_2 \oplus S_1$; (33) follows the definition $T \triangleq (U_1, X_2)$; (34) follows from the definition of the conditional entropy; (35) follows from the following definitions $\alpha_t \triangleq P(S_1 = 1|T = t)$ and $\beta_t \triangleq P(S_1 \oplus X_1 = 1|T = t)$ for any $t \in T$. We also define $q_{1|t} \triangleq P(X_1 = 1|T = t) = E\{X_1|T = t\}$, therefore the input constraint of user 1 can be written as

$$
EX_1 = E_T E\{X_1|T=t\} = E_T\{q_{1|t}\} \leq q_1. \tag{36}
$$

Without loss of generality, we can only consider $\alpha_t, \beta_t, q_{1|t} \in [0, 1/2]$ in (35) for any $t \in T$. Thus,

$$
\begin{aligned}
R_{com} &\leq E_T\Big\{H_b(\alpha_t) - H_b\Big([\alpha_t - q_{1|t}]^+\Big)\Big\} && (37) \\
&\leq E_T\Big\{H_b(q_{1|t})\Big\} && (38) \\
&\leq H_b\Big(E_T\{q_{1|t}\}\Big) && (39) \\
&\leq H_b(q_1), && (40)
\end{aligned}
$$

where (37) follows from (35) and since $H_b(\beta_t) \geq H_b\Big([\alpha_t - q_{1|t}]^+\Big)$, where $[x]^+ = max\{x, 0\}$; (38) follows since $H_b(\alpha_t) - H_b\Big([\alpha_t - q_{1|t}]^+\Big)$ is increasing in α_t for $\alpha_t \leq q_{1|t} \leq 1/2$ and decreasing in α_t for $q_{1|t} < \alpha_t \leq 1/2$, thus the maximum is for $\alpha_t = q_{1|t}$; (39) follows from Jensen's inequality since $H_b(\cdot)$ is convex-$\cap$; (40) follows from the input constraint for user 1 (36). The converse part follows since the outer bound is valid for any U_1 and X_1, X_2 that satisfy the input constraints.

The direct part is shown by using $U_1 = X_1 \oplus S_1$ where X_1 and S_1 are independent with $X_1 \sim$ Bernoulli(q_1), thus $U_1 \sim$ Bernoulli$(1/2)$. Furthermore, $X_2 \sim$ Bernoulli(q_2) which is independent of X_1, U_1, S_1. In this case $Y = U_1 \oplus X_2$, hence $Y \sim$ Bernoulli$(1/2)$. Using this choice for U_1, X_1, X_2, the achievable common message rate is given by

$$
\begin{aligned}
R_{com} &= I(U_1, X_2; Y) - I(U_1, X_2; S_1) \\
&= H(S_1|U_1, X_2) - H(Y|U_1, X_2) + H(Y) - H(S_1) \\
&= H(X_1) && (41) \\
&= H_b(q_1),
\end{aligned}
$$

where (41) follows since $H(S_1|U_1,X_2)=H(S_1|U_1)=H(X_1)$, $H(Y|U_1,X_2)=0$, $H(Y)=1$ and $H(S_1)=1$.

Appendix II
Proof of the Converse Part of Theorem 2

The proof of the converse part follows from Lemma 1, Lemma 2 and Lemma 3, whereas Lemma 5 and Lemma 4 are technical results which assist in the derivation of Lemma 3.

Let us define the following functions:

$$F(P_{V_1,V_1'},P_{V_2,V_2'}) \triangleq \Big[H(V_1)+H(V_2)-H(V_1'\oplus V_2')-1\Big]^+, \tag{42}$$

where $[x]^+=\max(0,x)$; its (q_1,q_2)-constrained maximization with respect to $V_1,V_1',V_2,V_2'\in\mathbb{Z}_2$ where (V_1,V_1') and (V_2,V_2') are independent, i.e.,

$$F_{max}(q_1,q_2) \triangleq \max_{V_1,V_1',V_2,V_2'} F(P_{V_1,V_1'},P_{V_2,V_2'}) \tag{43}$$
$$\text{s.t } \; P(V_i\neq V_i')\leq q_i, \text{ for } \; i=1,2;$$

and the upper convex envelope of $F_{max}(q_1,q_2)$ with respect to q_1,q_2

$$\overline{F}_{max}(q_1,q_2) \triangleq u.c.e\Big\{F_{max}(q_1,q_2)\Big\}. \tag{44}$$

In the following lemma we give an outer bound for the single-letter region (7) of the binary doubly-dirty MAC in the spirit of [23, Lemma 3] and [8, Proposition 1].

Lemma 1. *The best known single-letter rate sum* (7) *of the binary doubly-dirty MAC* (2) *with input constraint* q_1 *and* q_2 *is upper bounded by*

$$R_1+R_2\leq \overline{F}_{max}(q_1,q_2). \tag{45}$$

Proof: An outer bound on the best known single-letter region (7) is given by

$$\begin{aligned}
R_{BSL}^{sum}(U_1,U_2) &\triangleq \Big[I(U_1,U_2;Y)-I(U_1,U_2;S_1,S_2)\Big]^+ && (46)\\
&= \Big[H(S_1|U_1)+H(S_2|U_2)-H(Y|U_1,U_2)+H(Y)-H(S_1)-H(S_2)\Big]^+ && (47)\\
&\leq \Big[H(S_1|U_1)+H(S_2|U_2)-H(Y|U_1,U_2)-1\Big]^+ && (48)\\
&= \bigg[E_{U_1,U_2}\Big\{H(S_1|U_1=u_1)+H(S_2|U_2=u_2)-H(Y|U_1=u_1,U_2=u_2)-1\Big\}\bigg]^+ && (49)\\
&\leq E_{U_1,U_2}\bigg\{\Big[H(S_1|U_1=u_1)+H(S_2|U_2=u_2)-H(Y|U_1=u_1,U_2=u_2)-1\Big]^+\bigg\} && (50)\\
&\leq E_{U_1,U_2}\bigg\{F\Big(P_{S_1,S_1\oplus X_1|U_1=u_1},P_{S_2,S_2\oplus X_2|U_2=u_2}\Big)\bigg\} && (51)\\
&\leq E_{U_1,U_2}\Big\{\overline{F}_{max}\big(q_{1|u_1},q_{2|u_2}\big)\Big\} && (52)\\
&\leq \overline{F}_{max}\big(E_{U_1}q_{1|u_1},E_{U_2}q_{2|u_2}\big) && (53)\\
&\leq \overline{F}_{max}\big(q_1,q_2\big), && (54)
\end{aligned}$$

where (48) follows since $H(S_1) = H(S_2) = 1$ and $H(Y) \leq 1$; (49) follows from the definition of the conditional entropy; (50) follows since $[Ex]^+ \leq E\{x^+\}$; (51) follows from the definition of the function $F(P_{V_1,V_1'}, P_{V_2,V_2'})$ (42), likewise (52) follows from the definition of the function $\overline{F}_{max}(q_1, q_2)$ (44), and from the definition

$$q_{i|u_i} \triangleq P(S_i \neq X_i \oplus S_i | U_i = u_i) = P(X_i = 1 | U_i = u_i), \ for \ i = 1, 2;$$

(53) follows from Jensen's inequality since $\overline{F}_{max}(q_1, q_2)$ is a concave function; (54) follows from the input constraints where

$$\begin{aligned} EX_i &= E_{U_i} P(X_i = 1 | U_i = u_i) \\ &= \sum_{u_i \in U_i} P(u_i) P(X_i = 1 | U_i = u_i) \\ &= \sum_{u_i \in U_i} P(u_i) q_{i|u_i} \leq q_i, \text{ for } \ i = 1, 2. \end{aligned} \tag{55}$$

The lemma now follows since the upper bound (54) for the rate sum is independent of U_1 and U_2, hence it also bounds the single-letter region $\mathcal{R}_{BSL}(q)$. □

A simplified expression for the function $F_{max}(q_1, q_2)$ of (43) is shown in the following lemma.

Lemma 2. *The function* $F_{max}(q_1, q_2)$ (43) *is given by*

$$F_{max}(q_1, q_2) = \max_{\alpha_1, \alpha_2 \in [0, 1/2]} \Big[H_b(\alpha_1) + H_b(\alpha_2) - H_b\Big([\alpha_1 - q_1]^+ * [\alpha_2 - q_2]^+ \Big) - 1 \Big]^+, \tag{56}$$

where $*$ *is the binary convolution, i.e.,* $x * y \triangleq (1 - x)y + (1 - y)x$.

Proof: The function $F_{max}(q_1, q_2)$ is defined in (42) and (43) where V_1, V_1', V_2, V_2' are binary random variables. Let us define the following probabilities:

$$\begin{aligned} \alpha_i &\triangleq P(V_i = 1) \\ \delta_i &\triangleq P(V_i' = 1 | V_i = 0) \\ \gamma_i &\triangleq P(V_i' = 0 | V_i = 1), \end{aligned}$$

for $i = 1, 2$. We thus have

$$\begin{aligned} P(V_i' = 1) &= (1 - \alpha_i)\delta_i + \alpha_i(1 - \gamma_i) \triangleq g(\alpha_i, \delta_i \gamma_i) \\ P(V_i \neq V_i') &= \alpha_i \gamma_i + (1 - \alpha_i)\delta_i \triangleq h(\alpha_i, \delta_i, \gamma_i), \end{aligned}$$

for $i = 1, 2$. The maximization (43) can be written as

$$F_{max}(q_1, q_2) = \max_{\alpha_1, \alpha_2} \Big[H_b(\alpha_1) + H_b(\alpha_2) - \min_{\substack{\gamma_1, \delta_1, \gamma_2, \delta_2 \\ h(\alpha_i, \delta_i, \gamma_i) \leq q_i, \ i = 1, 2}} H_b\Big(g(\alpha_1, \delta_1, \gamma_1) * g(\alpha_2, \delta_2, \gamma_2) \Big) - 1 \Big]^+. \tag{57}$$

This maximization has two equivalent solutions (α_1^o, α_2^o) and $(1 - \alpha_1^o, 1 - \alpha_2^o)$ where $0 \leq \alpha_1^o, \alpha_2^o \leq 0.5$, since any other (α_1, α_2) can only increase the inner minimization in (57) which results in a lower $F_{max}(q_1, q_2)$. Therefore, without loss of generality we may assume that $0 \leq \alpha_1, \alpha_2 \leq 0.5$.

To prove the lemma we need to show that for any α_i the inner minimization is achieved by

$$\delta_i = 0, \gamma_i = \min\{1, q_i/\alpha_i\}, \ i = 1, 2.$$

In other words, V_i' has the smallest possible probability for 1 under the constraint that $P(V_i \neq V_i') \leq q_i$, implying that the transition from V_i to V_i' is a "Z channel". The inner minimization requires that $P(V_i' = 1)$ will be minimized restricted to the constraint $P(V_i \neq V_i') \leq q_i$, therefore it is equivalent to the following minimization

$$\min_{\substack{\gamma_i, \delta_i \\ h(\alpha_i, \delta_i \gamma_i) \leq q_i}} g(\alpha_i, \delta_i \gamma_i), \ i = 1, 2.$$

For $\alpha_i \leq q$, the solution is $\delta_i = 0$ and $\gamma_i = 1$ since in this case $g(\alpha_i, \gamma_i, \delta_i) = 0$ and the constraint is satisfied. For $q \leq \alpha_i \leq 0.5$, in order to minimize $g(\alpha_i, \gamma_i, \delta_i)$, it is required that $\delta_i \in [0, q/(1 - a_i)]$ will be minimal and $\gamma_i \in [0, q/\alpha_i]$ will be maximal such that the constraint is satisfied. Clearly, the best choice is for $\delta_i = 0$ and $\gamma_i = q/\alpha_i$, in this case the constraint is satisfies and $g(\alpha_i, \gamma_i, \delta_i) = \alpha_i - q$. □

The next lemma gives an explicit upper bound for $F_{max}(q_1, q_2)$ (43) for the case that $q_1 = q_2$. Let

$$f(x) = x - \frac{1}{1 + \left(\frac{1}{x} - 1\right)^2}, \tag{58}$$

and let

$$q_c \triangleq \max_{x \in [0, 1/2]} f(x). \tag{59}$$

Since $f(x)$ is differentiable, we can characterize q_c by differentiating $f(x)$ with respect to x and equating to zero, thus we get that

$$4x^4 - 8x^3 + 10x^2 - 6x + 1 = 0.$$

This fourth order polynomial has two complex roots and two real roots, where one of its real roots is a local minimum and the other root is a local maximum. Specifically, this local maximum maximizes $f(x)$ for the interval $x \in [0, 1/2]$ and it achieves $q_c \simeq 0.1501$ which occurs at $x \simeq 0.257$.

Lemma 3. *For $q_1 = q_2 = q$, we have that:*

$$\begin{aligned} & F_{max}(q, q) = 2H_b(q) - 1, && q_c \leq q \leq 1/2 \\ & F_{max}(q, q) \leq C^* q, && 0 < q < q_c \\ & F_{max}(0, 0) = 0, && q = 0, \end{aligned} \tag{60}$$

where q_c is defined in (59)*, while $C^* = \frac{2H_b(q^*) - 1}{q^*}$ and $q^* \triangleq 1 - 1/\sqrt{2} \simeq 0.3$ are defined in* (17).

Note that in the first case ($q_c \leq q \leq 1/2$) in (56) is achieved by $\alpha_1 = \alpha_2 = q$, while in the third case ($q = 0$) (56) is achieved by $\alpha_1 = \alpha_2 = 1/2$ as shown in Fig. 5. Although, we do not have an explicit expression for $F_{max}(q, q)$ in the range $0 < q < q_c$, the bound $F_{max}(q, q) \leq C^* q$ is sufficient for the purpose of proving Theorem 2 because $q_c \leq q^*$. In Fig. 4 a numerical characterization of $F_{max}(q, q)$ is plotted.

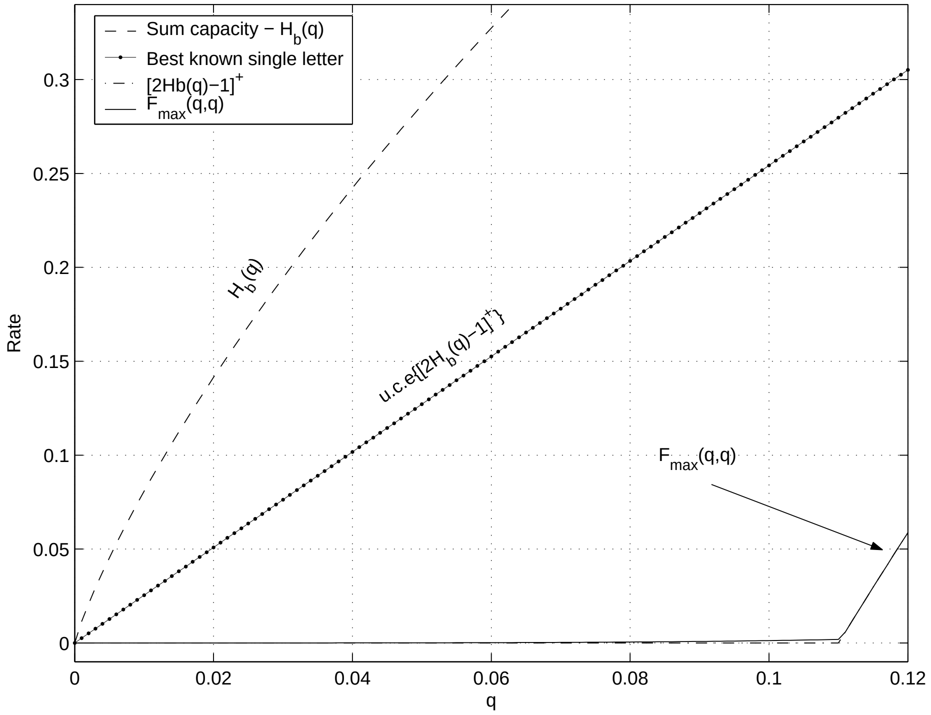

Fig. 4. Numerical results of $F_{max}(q,q)$ (56) for $q \in [0, 0.12]$ (Fig. 2 is the same plot for $q \in [0, 0.5]$) .

Proof: Define

$$F(\alpha_1, \alpha_2, q) \triangleq H_b(\alpha_1) + H_b(\alpha_2) - H_b\big([\alpha_1 - q]^+ * [\alpha_2 - q]^+\big) - 1.$$

From the discussion above about the cases of equality in (60), Lemma 3 will follow by showing that $F(\alpha_1, \alpha_2, q)$ is otherwise smaller, i.e.,

$$F(\alpha_1, \alpha_2, q) \leq \begin{cases} C^* q, & 0 \leq q \leq q_c \\ 2H_b(q) - 1, & q_c \leq q \leq 1/2 \end{cases} \tag{61}$$

for all $0 \leq \alpha_1, \alpha_2 \leq 1/2$. It is easy to see that for $\alpha_1, \alpha_2 \leq q$ the function $F(\alpha_1, \alpha_2, q)$ is monotonically increasing with α_1, α_2, and thus $F(\alpha_1, \alpha_2, q) \leq F(q, q, q) = 2H_b(q) - 1$. For $\alpha_1 \leq q$ and $q < \alpha_2 \leq 1/2$, $F(\alpha_1, \alpha_2, q)$ is increasing with α_1 and decreasing with α_2, and thus $F(\alpha_1, \alpha_2, q) \leq F(q, q, q) = 2H_b(q) - 1$. Clearly, from symmetry, also for $\alpha_2 \leq q$ and $q \leq \alpha_1 \leq 1/2$, $F(\alpha_1, \alpha_2, q) \leq 2H_b(q) - 1$. As a consequence, we have to show that (61) is satisfied only for $q \leq \alpha_1, \alpha_2 \leq 1/2$. Likewise, in the sequel we may assume without loss of generality that $q \leq \alpha_2 \leq \alpha_1 \leq 1/2$.

The bound for the interval $q_c < q \leq 1/2$: in this case (61) is equivalent to the following bound

$$H_b\Big((\alpha_1 - q) * (\alpha_2 - q)\Big) - H_b(\alpha_1) - H_b(\alpha_2) + 2H_b(q) \geq 0, \text{ for } q_c \leq q \leq \alpha_2 \leq \alpha_1 \leq 1/2. \tag{62}$$

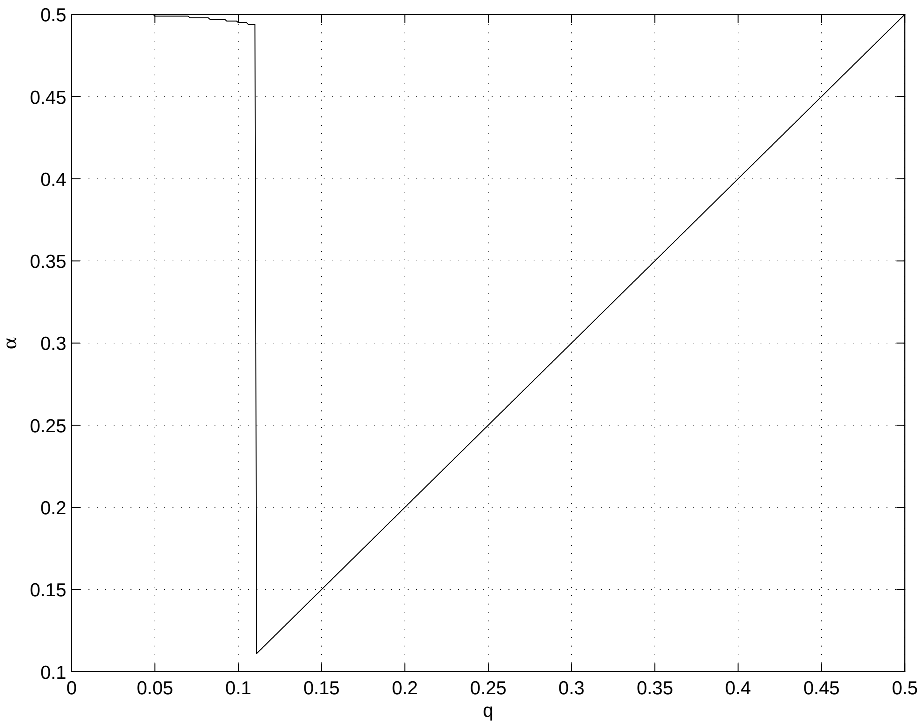

Fig. 5. The optimal $\alpha_1 = \alpha_2 = \alpha(q)$ which maximizes (56).

The LHS is lower bounded by

$$H_b\Big((\alpha_1 - q) * (\alpha_2 - q)\Big) - H_b(\alpha_1) - H_b(\alpha_2) + 2H_b(q)$$

$$\geq H_b(\alpha_1 - q) - H_b(\alpha_1) - H_b(\alpha_2) + 2H_b(q) \tag{63}$$

$$\geq H_b(\alpha_1 - q) - 2H_b(\alpha_1) + 2H_b(q) \tag{64}$$

$$\geq 0, \tag{65}$$

where (63) follows since $H_b\Big((\alpha_1 - q) * (\alpha_2 - q)\Big) \geq H_b(\alpha_1 - q)$; (64) follows since $\alpha_2 \leq \alpha_1 \leq 1/2$; (65) follows from Lemma 4 below.

The bound for the interval $0 \leq q \leq q_c$: in this case (61) is equivalent to the following bound

$$H_b\Big((\alpha_1 - q) * (\alpha_2 - q)\Big) \geq H_b(\alpha_1) + H_b(\alpha_2) - 1 - C^* \cdot q, \ for\ 0 \leq q \leq \alpha_2 \leq \alpha_1 \leq q_c. \tag{66}$$

For fixed α_1 and α_2, let us denote the RHS and the LHS of (66) as

$$g_l(q) \triangleq H_b\Big((\alpha_1 - q) * (\alpha_2 - q)\Big)$$

$$g_r(q) \triangleq H_b(\alpha_1) + H_b(\alpha_2) - 1 - C^* \cdot q.$$

The function $g_l(q)$ is convex-$\cap$ in q, since it is a composition of the function $H_b(x)$ which is non-decreasing convex-$\cap$ in the range $[0, 1/2]$ and the function $[\alpha_1 - q] * [\alpha_2 - q]$ which is convex-$\cap$ in q [24]. Since $g_r(q)$ is linear function in q and $g_l(q)$ is convex-$\cap$ function in q, the bound (66) is satisfied if the interval edges ($q = 0$ and

$q = \alpha_2$) satisfy this bound. For $q = 0$, (66) holds since

$$\begin{aligned} g_l(q=0) &= H_b(\alpha_1 * \alpha_2) \\ &\geq \max\{H_b(\alpha_1), H_b(\alpha_2)\} \\ &\geq \min\{H_b(\alpha_1), H_b(\alpha_2)\} \\ &\geq H_b(\alpha_1) + H_b(\alpha_2) - 1 \\ &= g_r(q=0). \end{aligned}$$

For $q = \alpha_2$ where $0 \leq q \leq q_c$, the bound (66) is satisfied since

$$\begin{aligned} g_r(q=\alpha_2) &= H_b(\alpha_1) + H_b(\alpha_2) - 1 - C^* \cdot \alpha_2 && (67) \\ &\leq H_b(\alpha_1) - H_b(q^*) + H_b(0.5q^*) - 0.5 && (68) \\ &\leq H_b(\alpha_1) - H_b(q_c) && (69) \\ &\leq H_b(\alpha_1) - H_b(\alpha_2) && (70) \\ &\leq H_b(\alpha_1 - \alpha_2) && (71) \\ &= g_l(q=\alpha_2), && (72) \end{aligned}$$

where (68) follows from Lemma 5 since $\arg\max_{\alpha_2 \in [0,1/2]} g_r(\alpha_2) = 0.5q^*$, and since $C^* = \frac{2H_b(q^*)-1}{q^*}$;(69) follows since for $q^* = 1-1/\sqrt{2}$ and q_c defined in (59), we have $H_b\big(1-1/\sqrt{2}\big) - H_b\big(0.5(1-1/\sqrt{2})\big) + 0.5 \simeq 0.68... \geq H_b(q_c)$; (70) follows since $q_c \geq \alpha_2$, thus $H_b(q_c) \geq H_b(\alpha_2)$; (71) follows since $H_b(\alpha_1) - H_b(\alpha_1 - \alpha_2)$ is decreing in α_1, thus $H_b(\alpha_1) - H_b(\alpha_1 - \alpha_2) \leq H_b(\alpha_2)$ for $\alpha_2 \leq \alpha_1 \leq 1/2$. Therefore, the bound (66) follows which completes the proof. □

Lemma 4 and Lemma 5 are auxiliary lemmas used in the proof of Lemma 3.

Lemma 4. *For $q_c \leq q \leq \alpha_1 \leq 1/2$, the following inequality is satisfied*

$$f_1(\alpha_1) \triangleq H_b(\alpha_1 - q) - 2H_b(\alpha_1) + 2H_b(q) \geq 0. \tag{73}$$

Proof: Since $f_1(\alpha_1 = q) = 0$, it is sufficient to show that $f_1(\alpha_1)$ is non-decreasing function in α_1, i.e., $\frac{d}{d\alpha_1} f_1(\alpha_1) \geq 0$ for $q_c \leq q \leq \alpha_1 \leq 1/2$, therefore

$$\frac{d}{d\alpha_1} f_1(\alpha_1) = \log_2\Big(\frac{1}{\alpha_1 - q} - 1\Big) - 2\log_2\Big(\frac{1}{\alpha_1} - 1\Big) \geq 0. \tag{74}$$

Due to monotonicity of the log function (74) is equivalent to

$$q \geq \alpha_1 - \frac{1}{1 + \Big(\frac{1}{\alpha_1} - 1\Big)^2} = f(\alpha_1), \tag{75}$$

where $f(\cdot)$ was defined in (58). Since by the definition of q_c (59) $f(x) \leq q_c \; \forall x \in [0, 1/2]$, it follows that $f(\alpha_1) \leq q \; \forall \; \alpha_1$ if $q_c \leq q$, and in particular for $q_c \leq q \leq \alpha_1$, which implies (75) as desired. □

Lemma 5. *Let*

$$f_2(x) = H_b(x) - 1 - C^* \cdot x, \tag{76}$$

where $x \in [0, 1/2]$, *and* $C^* = \frac{2H_b(q^*)-1}{q^*}$ *where* $q^* = 1 - 1/\sqrt{2}$. *The maximum of* $f_2(x)$ *is achieved by*

$$\arg\max_x f_2(x) = 0.5q^* = \frac{1}{2}(1 - 1/\sqrt{2}). \tag{77}$$

Proof: By differentiating $f_2(x)$ with respect to x and comparing to zero, we get that

$$0 = \frac{d}{dx} f_2(x) = \log_2\left(\frac{1-x}{x}\right) - C^*, \tag{78}$$

thus $x^o = \frac{1}{2^{C^*}+1}$ maximizes $f_2(x)$ since the second derivative is negative, i.e., $\frac{d^2}{x^2} f_2(x)|_{x=x^o} < 0$. The lemma is followed since $x^o = \frac{1}{2^{C^*}+1} = 0.5q^*$. □

We are now in a position to summarize the proof of Theorem 2.

Proof of Theorem 2 - Converse Part. The rate sum is upper bounded by

$$\begin{aligned} R_1 + R_2 &\leq u.c.e\Big\{F_{max}(q, q)\Big\} && (79)\\ &\leq u.c.e\left\{\begin{array}{cc} C^* \cdot q, & 0 \leq q \leq q_c \\ 2H_b(q) - 1, & q_c < q \leq 1/2 \end{array}\right\} && (80)\\ &= u.c.e\Big\{[2H_b(q) - 1]^+\Big\}, && (81)\end{aligned}$$

where (79) follows from Lemma 1; (80) follows from Lemma 3; and (81) follows since (80) is equal to the upper convex envelope of $[2Hb(q) - 1]^+$.

Appendix III

A simplified Outer Bound for the Sum Capacity in the Strong Interference Gaussian Case

Lemma 6. *The best known single-letter sum capacity* (7) *of the Gaussian doubly-dirty MAC* (24) *with power constraints* P_1, P_2, *and strong interferences* ($Q_1, Q_2 \to \infty$) *is upper bounded by*

$$R_1 + R_2 \leq u.c.e\bigg\{\sup_{V_1, V_1', V_2, V_2'} \Big[h(V_1) + h(V_2) - h\big(V_1' + V_2' + Z\big) + h(S_1 + S_2) - h(S_1) - h(S_2)\Big]^+\bigg\}, \tag{82}$$

where $u.c.e$ *is the upper convex envelope operation with respect to* P_1 *and* P_2, *and* $[x]^+ = \max(0, x)$. *The supremum is over all* V_1, V_1', V_2, V_2' *such that* (V_1, V_1') *is independent of* (V_2, V_2'), *and*

$$\begin{aligned} &E\Big\{(V_i - V_i')^2\Big\} \leq P_i,\\ &h(V_i) \leq h(S_i), \end{aligned}$$

for $i = 1, 2$.

Proof: Let us define the following functions (corresponds to $F(P_{V_1,V_1'}, P_{V_2,V_2'})$ of (42))

$$G\big(f_{V_1,V_1'}, f_{V_2,V_2'}\big) \triangleq \Big[h(V_1) + h(V_2) - h\big(V_1' + V_2' + Z\big) + h(S_1 + S_2) - h(S_1) - h(S_2)\Big]^+. \tag{83}$$

The second function is the following maximization of (83) with respect to V_1, V_1', V_2, V_2'.

$$G_{max}(P_1, P_2) \triangleq \sup_{V_1, V_1', V_2, V_2'} G\big(f_{V_1,V_1'}, f_{V_2,V_2'}\big) \quad (84)$$
$$\text{s.t } E\Big\{(V_i - V_i')^2\Big\} \le P_i, \quad h(V_i) \le h(S_i), \quad \text{for } i = 1, 2.$$

Finally, we define the upper convex envelope of $G_{max}(P_1, P_2)$ with respect to P_1 and P_2:

$$\overline{G}_{max}(P_1, P_2) \triangleq u.c.e\Big\{G_{max}(P_1, P_2)\Big\}. \quad (85)$$

Clearly if we take only the rate sum equation in (6) we get an outer bound on the best known single-letter region,

$$R^{sum}_{BSL}(U_1, U_2) \triangleq \Big[I(U_1, U_2; Y) - I(U_1, U_2; S_1, S_2)\Big]^+ \quad (86)$$
$$= \Big[h(S_1|U_1) + h(S_2|U_2) - h(Y|U_1, U_2) + h(Y) - h(S_1) - h(S_2)\Big]^+ \quad (87)$$
$$\le \Big[h(S_1|U_1) + h(S_2|U_2) - h(Y|U_1, U_2) + h(S_1 + S_2) - h(S_1) - h(S_2)\Big]^+ + o(1) \quad (88)$$
$$= \Bigg[E_{U_1,U_2}\Big\{h(S_1|U_1 = u_1) + h(S_2|U_2 = u_2) - h(Y|U_1 = u_1, U_2 = u_2) + h(S_1 + S_2) - h(S_1) - h(S_2)\Big\}\Bigg]^+ + o(1) \quad (89)$$
$$\le E_{U_1,U_2}\Bigg\{\Big[h(S_1|U_1 = u_1) + h(S_2|U_2 = u_2) - h(X_1 + S_1 + X_2 + S_2 + Z|U_1 = u_1, U_2 = u_2)$$
$$+ h(S_1 + S_2) - h(S_1) - h(S_2)\Big]^+\Bigg\} + o(1) \quad (90)$$
$$= E_{U_1,U_2}\Bigg\{G\Big(f_{S_1,S_1+X_1|U_1=u_1} f_{S_2,S_2+X_2|U_2=u_2}\Big)\Bigg\} + o(1) \quad (91)$$
$$\le E_{U_1,U_2}\Big\{\overline{G}_{max}\big(P_{1|u_1}, P_{2|u_2}\big)\Big\} + o(1) \quad (92)$$
$$\le \overline{G}_{max}\big(E_{U_1} P_{1|u_1}, E_{U_2} P_{2|u_2}\big) + o(1) \quad (93)$$
$$\le \overline{G}_{max}\big(P_1, P_2\big) + o(1), \quad (94)$$

where (88) follows since $h(Y) \le h(S_1 + S_2) + o(1)$ where $o(1) \to 0$ as $Q_1, Q_2 \to \infty$; (89) follows from the definition of the conditional entropy; (90) follows since $[Ex]^+ \le E\{x^+\}$ and since $Y = X_1 + S_1 + X_2 + S_2 + Z$; (91) follows from the definition of the function $G\big(f_{V_1,V_1'}, f_{V_2,V_2'}\big)$ (83), likewise (92) follows from the definition of the function $\overline{G}_{max}(P_1, P_2)$ (85), and since $h(S_i|U_i) \le h(S_i)$ and from the definition

$$P_{i|u_i} \triangleq E\Big\{X_i^2|U_i = u_i\Big\}, \ for \ i = 1, 2;$$

(93) follows from Jensen's inequality since $\overline{G}_{max}(P_1, P_2)$ is a concave function; (94) follows from the input constraints where

$$EX_i^2 = E_{U_i} E\big\{X_i^2|U_i = u_i\big\} = E_{U_i} P_{i|u_i} \le P_i, \ \text{for } \ i = 1, 2. \quad (95)$$

The lemma follows since the upper bound (94) for the rate sum is now independent of U_1 and U_2, hence it also bound the single-letter region $\mathcal{R}_{BSL}(P_1, P_2)$. □

Acknowledgment

The authors wish to thank Ashish Khisti for earlier discussions on the binary case. The authors also would like to thank Uri Erez for helpful comments.

References

[1] M. Costa, "Writing on dirty paper," *IEEE Trans. Information Theory*, vol. IT-29, pp. 439–441, May 1983.

[2] S. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Problemy Pered. Inform. (Problems of Inform. Trans.)*, vol. 9, No. 1, pp. 19–31, 1980.

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

[4] A. Somekh-Baruch, S. Shamai, and S. Verdu, "Cooperative encoding with asymmetric state information at the transmitters," in *Proceedings 44th Annual Allerton Conference on Communication, Control, and Computing, Univ. of Illinois, Urbana, IL, USA*, Sep. 2006.

[5] S. Kotagiri and J. N. Laneman, "Multiple access channels with state information known at some encoders," *IEEE Trans. Information Theory*, July 2006, submitted for publication.

[6] S. A. Jafar, "Capacity with causal and non-causal side information - a unified view," *IEEE Trans. Information Theory*, vol. IT-52, pp. 5468–5475, Dec. 2006.

[7] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Information Theory*, vol. IT–22, pp. 374–377, May 1979.

[8] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Trans. Information Theory*, vol. IT-51, pp. 3820–3833, Nov. 2005.

[9] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Information Theory*, vol. IT-48, pp. 1250–1276, June 2002.

[10] T. Philosof, A. Khisti, U. Erez, and R. Zamir, "Lattice strategies for the dirty multiple access channel," in *Proceedings of IEEE International Symposium on Information Theory, Nice, France*, June 2007.

[11] J. Korner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Trans. Information Theory*, vol. IT-25, pp. 219–221, March 1979.

[12] T. M. Cover and B. Gopinath, *Open Problems in Communication and Computation*. New York: Springer-Verlag, 1987.

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

[14] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Information Theory*, vol. IT-53, pp. 3498–3516, Oct. 2007.

[15] D. Krithivasan and S. S. Pradhan, "Lattices for distributed source coding: Jointly Gaussian sources and reconstruction of a linear function," *arXiv:cs.IT/0707.3461V1*.

[16] A. Khisti, "Private communication."

[17] R. G. Gallager, *Information Theory and Reliable Communication*. New York, N.Y.: Wiley, 1968.

[18] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein, *Covering Codes*. Amsterdam, The Netherlands: North Holland Publishing, 1997.

[19] R. Ahlswede and J. Korner, "Source coding with side information and a converse for the degraded broadcast channel," *IEEE Trans. Information Theory*, vol. 21, pp. 629–637, 1975.

[20] A. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Information Theory*, vol. IT-21, pp. 294–300, 1975.

[21] T. Berger, *Multiterminal Source Coding*. New York: In G.Longo, editor, the Information Theory Approach to Communications, Springer-Verlag, 1977.

[22] L. F. Wei and G. D. Forney, "Multidimensional constellation - part I: Introduction, figures of merit, and generalized cross constellations," vol. 7, pp. 877–892, Aug. 1989.

[23] A. Cohen and R. Zamir, “Entropy amplification property and the loss for writing on dirty paper,” *IEEE Trans. Information Theory*, To appear, April 2008.

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge: Cambridge University Press, 2004.