An information inequality and evaluation of Marton's inner bound for binary input broadcast channels

Yanlin Geng, *Member, IEEE,* Varun Jog, Chandra Nair, *Member, IEEE,* and Zizhou Wang,

***Abstract*—We establish an information inequality concerning five random variables. This inequality is motivated by the sum-rate evaluation of Marton's inner bound for two receiver broadcast channels with a binary input alphabet. We establish that randomized time-division strategy achieves the sum-rate of Marton's inner bound for all binary input broadcast channels. We also obtain an improved cardinality bound for evaluating the maximum sum-rate given by Marton's inner bound for all broadcast channels. Using these tools we explicitly evaluate the inner and outer bounds for the binary skew-symmetric broadcast channel and demonstrate a gap between the bounds.**

***Index Terms*—information inequality, Marton's inner bound, binary input alphabet**

I. Introduction

A two-receiver broadcast channel models the communication scenario where two (independent) messages are to be transmitted from a sender X to two receivers Y, Z. Each receiver is interested in decoding its message. A transition probability matrix given by $\mathfrak{q}(y, z|x)$ models the stochastic nature of the errors introduced during the communication. For formal definitions and early results the reader can refer to [1], [2].

A. Background

The following region obtained by Marton [3] represents the best-known achievable region to-date:

Bound 1 (Marton's inner bound [3])**.** *The set of rate-pairs (R_1, R_2) satisfying the following constraints:*

$$\begin{aligned} R_1 &\leq I(U,W;Y) \\ R_2 &\leq I(V,W;Z) \\ R_1 + R_2 &\leq \min\{I(W;Y), I(W;Z)\} \\ &\quad + I(U;Y|W) + I(V;Z|W) - I(U;V|W) \end{aligned}$$

for any set of (U,V,W) such that $(U,V,W) \to X \to (Y,Z)$ forms a Markov chain is achievable.

Y. Geng and C. Nair are with The Chinese University of Hong Kong.
V. Jog is with University of California, Berkeley.
Z. Wang is with Altai Technologies Ltd., Hong Kong.

This work was partially supported by a grant from the University Grants Committee of the Hong Kong Special Administrative Region, China (Project No. AoE/E-02/08), and by a direct grant for Research at the Chinese University of Hong Kong. Part of this work was done when Varun Jog was a summer intern at CUHK. The paper was presented in part at ITA 2010 and ISIT 2010.

Recently Gohari and Ananthram [4] used a remarkable perturbation-based argument to establish that it suffices to consider (U,V,W) with alphabet sizes bounded by $|\mathcal{U}| \leq |\mathcal{X}|, |\mathcal{V}| \leq |\mathcal{X}|, |\mathcal{W}| \leq |\mathcal{X}| + 4$ to compute the extreme points of Bound 1. In general the computation of Marton's inner bound is difficult, and prior to [4], this bound was not strictly evaluable. Even with these bounds on cardinalities, explicit evaluation of the bounds is still a difficult task.

The following region represents the best-known outer bound to the capacity region of the broadcast channel with private messages.

Bound 2 (UV outer bound [5])**.** *The union of rate-pairs (R_1, R_2) satisfying the following constraints:*

$$\begin{aligned} R_1 &\leq I(U;Y) \\ R_2 &\leq I(V;Z) \\ R_1 + R_2 &\leq I(U;Y) + I(V;Z|U) \\ R_1 + R_2 &\leq I(V;Z) + I(U;Y|V) \end{aligned}$$

over all pairs of random variables (U,V) such that $(U,V) \to X \to (Y,Z)$ forms a Markov chain forms an outer bound to the capacity region of the broadcast channel.

The capacity regions of special classes of broadcast channels have been established and in every case it turns out that Bounds 1 and 2 agree. Though there have been attempts to improve the outer bound, it has been shown [6] that the new outer bounds reduced to the UV outer bound (for private messages). In order to study whether Bounds 1 and 2 correspond to different regions or whether they are alternate representations of the same region, [7] studied a particular channel called the binary skew-symmetric broadcast channel (BSSC) shown in Figure 1. The authors conjectured that for BSSC the following inequality

$$I(U;Y) + I(V;Z) - I(U;V) \leq \max\{I(X;Y), I(X;Z)\} \quad (1)$$

holds for all $(U,V) \to X \to (Y,Z)$ that forms a Markov chain. The authors further showed that, assuming (1) holds, Bounds 1 and 2 *differ* for BSSC.

In [4], the authors established that Bounds 1 and 2 were indeed different for BSSC without actually establishing that (1) was true. They verified that (1) was plausible by confirming it for a large number of (randomly-generated) samples from the cardinality constrained space.

In this paper we establish that (1) is true not only for BSSC but also for any binary input broadcast channel. This paper

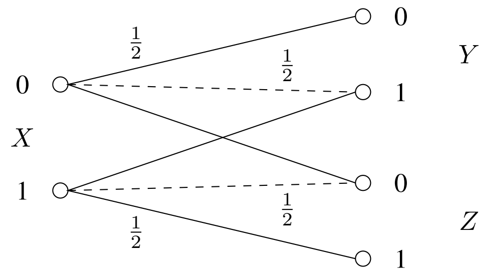

Fig. 1. Binary skew-symmetric broadcast channel (BSSC)

unifies the results in two papers [8], [9] and adds a few other results, along with providing complete details of proofs.

B. Summary of results

The main results of the paper are the following:

Theorem 1. *Consider a five tuple of random variables (U,V,X,Y,Z) such that $(U,V) \to X \to (Y,Z)$ forms a Markov chain and further let $|\mathcal{X}|=2$. Then the following inequality holds:*

$$I(U;Y)+I(V;Z)-I(U;V) \le \max\{I(X;Y),I(X;Z)\}. \quad (2)$$

Given an input probability distribution $p(x)$ and a fixed broadcast channel $\mathsf{q}(y,z|x)$, define[1]

$$SR(X) := \max_{\substack{p(u,v,w|x): \\ (U,V,W)\to X\to(Y,Z)}} \min\{I(W;Y),I(W;Z)\} + I(U;Y|W)+I(V;Z|W)-I(U;V|W).$$

Observe that $SR(X)$ is the maximum achievable sum rate using Marton's inner bound if one imposes the constraint that the codewords are generated using the distribution $p(x)$.

For a given broadcast channel $\mathsf{q}(y,z|x)$, define

$$SR^* = \max_{p(x)} SR(X). \quad (3)$$

Theorem 2. *It suffices to consider $|\mathcal{U}| \le |\mathcal{X}|$, $|\mathcal{V}| \le |\mathcal{X}|$, and $|\mathcal{W}| \le |\mathcal{X}|$ in the evaluation of $SR(X)$ for any input distribution $p(x)$, and in the evaluation of SR^*, the maximum value of the sum-rate achievable using Marton's inner bound.*

Remark 1. From (3), it is clear that if we establish the cardinality bounds in Theorem 2 for $SR(X)$ then the bounds for SR^* will automatically follow. Hence we will establish the bounds for the stronger statement, i.e. $SR(X)$.

Theorem 1 proves that the inequality in (1) is valid for *every* binary input broadcast channel. Combining this result with the cardinality bounds in Theorem 2, for binary input broadcast channels we establish that the maximum sum-rate given by Marton's coding strategy matches that given by the randomized time-division strategy [5], a much simpler achievable strategy.

[1] Note that $SR(X)$ is a function of $p(x)$, however an abuse of notation is made to be consistent with quantities such as $H(X), I(X;Y)$, etc.

Theorem 3. *The maximum value of the sum-rate for Marton's inner bound for any binary input broadcast channel is given by*

$$\max_{\substack{p(w,x): \\ W\to X\to(Y,Z)}} \min\{I(W;Y),I(W;Z)\} + \mathrm{P}(W{=}0)I(X;Y|W{=}0)+\mathrm{P}(W{=}1)I(X;Z|W{=}1)$$

where $|\mathcal{W}|=2$.

Theorem 4. *For the broadcast channel (BSSC) shown in Figure 1 the maximum sum-rate value for the various bounds is given by:*

- *Marton's inner bound (Bound 1) = 0.36164288... ,*
- *UV outer bound (Bound 2) = 0.3725562... ,*
- *Körner-Marton outer bound (Bound 3) = 0.3743955... .*

Remark 2. We will introduce the Körner-Marton outer bound (Bound 3) in a subsequent section.

1) Randomized time-division strategy: Randomized time-division (R-TD) strategy [5] corresponds to an achievable strategy for the following setting of (U,V,W) in Bound 1: $W=0$ implies that $U=X, V=\emptyset$; and $W=1$ implies that $V=X, U=\emptyset$ (where $\emptyset$ refers to the trivial random variable). Observe that this corresponds to a time-division strategy except that the time slots for which communication occurs to one receiver are drawn from a codebook (the time slots are used to convey additional information that can be decoded by both receivers).

2) Relationship between Theorem 1 and non-Shannon type inequalities: Recently there has been a lot of interest in information inequalities and the so-called Shannon-type and non-Shannon type inequalities. The space $\bar{\Gamma}_n^*$ (see Section 13.1 in [10]) refers to the closure of the space of entropic vectors formed using n discrete random variables. An entropic vector corresponding to a given n-tuple of discrete random variables is a point in $\mathbb{R}_+^{2^n-1}$ obtained by taking the entropy of each of the non-empty subsets of the given n-random variables. It is known that $\bar{\Gamma}_n^*$ is a closed convex cone.

Theorem 1 refers to a subset, $\mathcal{S}$, of points in $\bar{\Gamma}_5^*$: those corresponding to a five tuple of random variables (U,V,X,Y,Z) such that $(U,V) \to X \to (Y,Z)$ forms a Markov chain and with a binary constraint on the cardinality of X, i.e. $|\mathcal{X}|=2$. Theorem 1 shows that the points in $\mathcal{S}$ have to lie in the union of two half-spaces induced by the two hyperplanes:

$$I(U;Y)+I(V;Z)-I(U;V) \le I(X;Y)$$
$$I(U;Y)+I(V;Z)-I(U;V) \le I(X;Z).$$

Since the inequalities are tight, i.e. there are points lying on the boundary of these hyperplanes, $\mathcal{S}$ is not a convex region in general.

Consider the Blackwell channel shown in Figure 2.

For this channel, consider $U=Y, V=Z$ and $X \sim [\frac{1}{3},\frac{1}{3},\frac{1}{3}]$. Observe that $(U,V)\to X\to(Y,Z)$ is still Markov and that $I(U;Y)+I(V;Z)-I(U;V)=H(Y,Z)=\log_2 3 > 1 \ge \max\{I(X;Y),I(X;Z)\}$. Thus there are points in Γ_5^* that violate the inequality implied by Theorem 1 and hence this is an example of an inequality that cannot be deduced by even the knowledge of Γ_5^*. This is primarily due to the cardinality

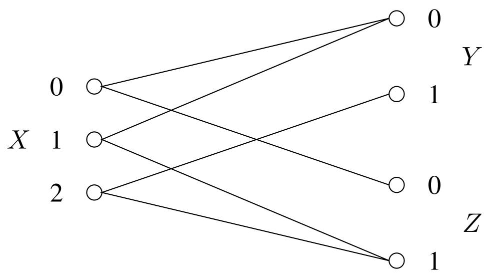

Fig. 2. Blackwell broadcast channel

constraint imposed on one of the random variables and thus demonstrates the existence of significant inequalities beyond the non-Shannon and Shannon-type ones usually considered in literature.

Organization of the paper: The proof of Theorem 1 uses a rather detailed perturbation analysis. To provide a gentler introduction to the proofs, we will first establish Theorem 2 in the next section as its proof uses a milder form of the perturbation arguments. We then show how Theorem 3 follows from Theorem 1 and Theorem 2. In the subsequent sections we will prove Theorem 1 and then establish Theorem 4.

II. Proofs of Theorem 2 and Theorem 3

A. Preliminaries

Given an input probability distribution $p(x)$ and a fixed broadcast channel $\mathfrak{q}(y,z|x)$, let[2]

$$T(X) := \max_{\substack{p(u,v|x): \\ (U,V)\to X \to (Y,Z)}} I(U;Y) + I(V;Z) - I(U;V). \tag{4}$$

Remark 3. From [4] (or see Fact 1 and Claim 1 in [8] for a self-contained shorter proof) we know that it suffices to consider $|\mathcal{U}| \leq |\mathcal{X}|$, $|\mathcal{V}| \leq |\mathcal{X}|$, $X = f(U,V)$ to evaluate $T(X)$, where $f(\cdot,\cdot)$ is a function that maps $\mathcal{U} \times \mathcal{V} \mapsto \mathcal{X}$.

From standard Fenchel-Bunt extension to Caratheodory's theorem, it follows that we can restrict ourselves to $|\mathcal{W}| \leq |\mathcal{X}|+1$ to compute $SR(X)$. The main contribution of Theorem 2 is that one may reduce the cardinality further to $|\mathcal{W}| \leq |\mathcal{X}|$. This (mild but nontrivial) improvement is very useful when one needs to explicitly evaluate the bound.

An equivalent version of the next claim is known in literature [11]. The claim asserts that in addition to the cardinality constraints $|\mathcal{U}| \leq |\mathcal{X}|$, $|\mathcal{V}| \leq |\mathcal{X}|$ and $|\mathcal{W}| \leq |\mathcal{X}|+1$, we can assume that a maximizer also satisfies $I(W;Y) = I(W;Z)$. We present a proof here for completeness.

Claim 1. *The function $SR(X)$ can be attained at a $p^*(u,v,w|x)$ that satisfies $I(W;Y) = I(W;Z)$, $|\mathcal{U}| \leq |\mathcal{X}|$, $|\mathcal{V}| \leq |\mathcal{X}|$, and $|\mathcal{W}| \leq |\mathcal{X}|+1$.*

Proof: Let $p^*(u,v,w|x)$ attain the value $SR(X)$. Suppose, at $p^*(u,v,w|x)$, we have $I(U,W;Y) \leq I(U,W;Z)$ then observe that

$$\begin{aligned} SR(X) &\leq I(U,W;Y) + I(V;Z|W) - I(U;V|W) \\ &\leq I(U,W;Y) + I(V;Z|U,W) \\ &\leq I(V,U,W;Z) \leq I(X;Z). \end{aligned}$$

Clearly by setting $V = X$, $U = W = \emptyset$ we obtain that $SR(X) \geq I(X;Z)$. Thus we have $SR(X) = I(X;Z)$. Since $V = X$, $U = W = \emptyset$ attains $SR(X)$ we have $I(W;Y) = I(W;Z) = 0$.

Similarly if $I(V,W;Z) \leq I(V,W;Y)$ then $SR(X) = I(X;Y)$, and we can set $U = X$, $V = W = \emptyset$ to achieve this $SR(X)$. In this case as well, $I(W;Y) = I(W;Z) = 0$. Thus for these two cases the claim is established.

The only case that remains is the following: at $p^*(u,v,w|x)$ we have $I(U,W;Y) > I(U,W;Z)$ and $I(V,W;Z) > I(V,W;Y)$. In this case we will show that there is a maximizer such that $I(W;Y) = I(W;Z)$.

Suppose that we have $I(W;Y) > I(W;Z)$ at a maximizing distribution. Let Q be Bernoulli(a) and independent of (U,V,W,X). When $Q=0$, set $(\tilde{U},\tilde{V},\tilde{W}) = (U,V,W)$ as before; and when $Q=1$ set $(\tilde{U},\tilde{V},\tilde{W}) = (U,\emptyset,(V,W))$. Let $W' = (\tilde{W},Q)$. Now $a \in (0,1)$ is chosen such that

$$\begin{aligned} I(W';Y) &= (1-a)I(W;Z) + aI(V,W;Z) \\ &= (1-a)I(W;Y) + aI(V,W;Y) = I(W';Z). \end{aligned}$$

Note that such an $a \in (0,1)$ always exists since $I(V,W;Z) > I(V,W;Y)$ and $I(W;Y) > I(W;Z)$.

For the new triple random variables $(W',\tilde{U},\tilde{V})$ note that

$$\begin{aligned} &\min\{I(W';Y), I(W';Z)\} + I(\tilde{U};Y|W') \\ &\quad + I(\tilde{V};Z|W') - I(\tilde{U};\tilde{V}|W') \\ &= I(W';Z) + I(\tilde{U};Y|W') \\ &\quad + I(\tilde{V};Z|W') - I(\tilde{U};\tilde{V}|W') \\ &= (1-a)\{I(W;Z) + I(U;Y|W) + I(V;Z|W) \\ &\quad - I(U;V|W)\} + a\{I(V,W;Z) + I(U;Y|V,W)\} \\ &= (1-a)SR(X) + a\{I(W;Z) + I(U;Y|W) \\ &\quad + I(V;Z|W) - I(U;V|W) + I(U;V|W,Y)\} \\ &= SR(X) + aI(U;V|W,Y) \\ &\geq SR(X). \end{aligned}$$

Thus the distribution $p(\tilde{u},\tilde{v},w'|x)$ also achieves $SR(X)$. Notice that here $|\mathcal{W}'| = 2|\mathcal{W}|$ and satisfies $I(W';Y) = I(W';Z)$.

Now starting from a $p(x)$ and $p(\tilde{u},\tilde{v},w'|x)$ satisfying $I(W';Y) = I(W';Z)$, we can find (a standard application of Fenchel-Butt extension to Caratheodory's theorem) a W of cardinality at most $|\mathcal{X}|+1$, and a $q^*(u,v,w,x)$ such that $p(x)$, $-H(Y|W) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$, $-H(Z|W) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$ are preserved. Thus $I_{q^*}(W;Y) = I_{q^*}(W;Z)$, and $q^*(u,v,w,x)$ also attains $SR(X)$. This completes the proof. ∎

We now prove Theorem 2 using a perturbation argument. By Claim 1, the only interesting case left is when there is a maximizer $p^*(u,v,w|x)$ that satisfies: $I(W;Y) = I(W;Z)$ and $|\mathcal{W}| = |\mathcal{X}|+1$.

[2]Note that $T(X)$ is a function of $p(x)$ for a fixed channel $\mathfrak{q}(y,z|x)$.

Notation: Given a pair of random variables (X,Y) distributed according to $p(x,y)$ and a function $L(x,y)$, define

$$H_p^L(X,Y) := -\sum_{x,y} p(x,y)L(x,y)\log p(x,y),$$

as the L-weighted entropy function. When L is omitted we assume that $L(x,y)=1$ which reduces to the standard entropy function, and we also omit $p(\cdot)$ when the distribution under consideration is evident.

B. Proof of Theorem 2

Let $p^*(u,v,w|x)$ achieve $SR(X)$ and satisfy $|\mathcal{U}| \le |\mathcal{X}|$, $|\mathcal{V}| \le |\mathcal{X}|$, $I(W;Y) = I(W;Z)$ and $|\mathcal{W}| = |\mathcal{X}|+1$. Let

$$\begin{aligned} &r(u,v,w,x) = q^*(u,v,w,x)(1+\epsilon L(w)), \quad (5)\\ &\text{such that} \sum_{w,u,v} q^*(u,v,w,x)L(w) = 0, \forall x \in \mathcal{X}. \end{aligned}$$

Here ϵ is any real number such that $1+\epsilon L(w) \ge 0, \forall w$.

The conditions $\sum_{u,v,w} q^*(u,v,w,x)L(w) = 0, \forall x$ imply that the distribution of X is unchanged by the perturbation, i.e. $r(x) = q^*(x)$.

Observe that a nontrivial $L(w)$ exists whenever $|\mathcal{W}| > |\mathcal{X}|$, since $L(w)$ is $(|\mathcal{X}|+1)$-dimensional and there are only $|\mathcal{X}|$ linear constraints which equal zero.

Let A represent a generic random variable, initially distributed according to $q^*(a,w)$, and perturbed according to

$$r(a,w) = q^*(a,w)(1+\epsilon L(w))$$

where $L(w)$ is the same function in (5). Observe that

$$\begin{aligned} H_r(AW) &= -\sum_{a,w} r(a,w)\log r(a,w)\\ &= -\sum_{a,w} q^*(a,w)(1+\epsilon L(w))\log\big(q^*(a,w)(1+\epsilon L(w))\big)\\ &= H_{q^*}(AW) + \epsilon H^L(AW)\\ &\quad - \sum_{a,w} q^*(a,w)(1+\epsilon L(w))\log(1+\epsilon L(w))\\ &= H_{q^*}(AW) + \epsilon H^L(AW) \quad (6)\\ &\quad - \sum_{w} q^*(w)(1+\epsilon L(w))\log(1+\epsilon L(w)). \end{aligned}$$

We will set A to be an appropriate subset of (U,V,Y,Z) in the terms that appear below.

Since the perturbation $L(w)$ preserves the distribution of X and the channel transition probabilities $\mathsf{q}(y,z|x)$ are fixed, the perturbation also fixes the distributions of Y and Z, i.e. $r(y) = q^*(y)$, $r(z) = q^*(z)$.

Observe that

$$\begin{aligned} I_r(W;Y) &= H_r(Y) + H_r(W) - H_r(YW)\\ &= H_{q^*}(Y) + H_{q^*}(W) - H_{q^*}(YW)\\ &\quad + \epsilon(H^L(W) - H^L(YW))\\ &= I_q^*(W;Y) + \epsilon(H^L(W) - H^L(YW)). \end{aligned}$$

Here the first equality is obtained by two applications of (6): setting $A=\emptyset$ and $A=Y$ respectively. Note that the term $\sum_w q^*(w)(1+\epsilon L(w))\log(1+\epsilon L(w))$ cancels, leaving us with only a linear term in ϵ. We also used the fact that $r(y) = q^*(y)$.

Similarly we obtain

$$\begin{aligned} I_r(W;Z) &= I_{q^*}(W;Z) + \epsilon(H^L(W) - H^L(ZW))\\ I_r(U;Y|W) &= I_{q^*}(U;Y|W) + \epsilon(H^L(UW)\\ &\quad + H^L(YW) - H^L(UYW) - H^L(W))\\ I_r(V;Z|W) &= I_{q^*}(V;Z|W) + \epsilon(H^L(VW)\\ &\quad + H^L(ZW) - H^L(VZW) - H^L(W))\\ I_r(U;V|W) &= I_{q^*}(U;V|W) + \epsilon(H^L(UW)\\ &\quad + H^L(VW) - H^L(UVW) - H^L(W)). \end{aligned}$$

Therefore the sum-rate corresponding to the distribution $r(u,v,w,x)$ is given by

$$\begin{aligned} &\min\{I_{q^*}(W;Y) + \epsilon(H^L(W) - H^L(YW)),\\ &\quad I_{q^*}(W;Z) + \epsilon(H^L(W) - H^L(ZW))\}\\ &\quad + I_{q^*}(U;Y|W) + I_{q^*}(V;Z|W) - I_{q^*}(U;V|W)\\ &\quad + \epsilon(H^L(UW) + H^L(YW) - H^L(UYW) - H^L(W))\\ &\quad + \epsilon(H^L(VW) + H^L(ZW) - H^L(VZW) - H^L(W))\\ &\quad - \epsilon(H^L(UW) + H^L(VW) - H^L(UVW) - H^L(W))\\ &= I_{q^*}(W;Y)(= I_{q^*}(W;Z))\\ &\quad + I_{q^*}(U;Y|W) + I_{q^*}(V;Z|W) - I_{q^*}(U;V|W)\\ &\quad + \epsilon\Big(\min\{H^L(W) - H^L(YW), H^L(W) - H^L(ZW)\}\\ &\quad + H^L(YW) - H^L(UYW) - H^L(W)\\ &\quad + H^L(ZW) - H^L(VZW) + H^L(UVW)\Big). \end{aligned}$$

Since $q^*(u,v,w,x)$ is a global maximum of the sum-rate it implies that the factor multiplying ϵ must be zero. This immediately implies that for any ϵ satisfying $1+\epsilon L(w) \ge 0, \forall w$ the sum-rate corresponding to the distribution $r(u,v,w,x)$ matches that of $q^*(u,v,w,x)$. Choose ϵ such that $\min_w(1+\epsilon L(w)) = 0$. Let w_0 be the minimizer, i.e. $1+\epsilon L(w_0) = 0$, then observe that $r(w_0) = 0$. Thus there exists a distribution $r(u,v,w,x)$ which attains $SR(X)$ such that $|\mathcal{U}| \le |\mathcal{X}|$, $|\mathcal{V}| \le |\mathcal{X}|$, and $|\mathcal{W}| \le |\mathcal{X}|$.

Since $SR^* = \max_{p(x)} SR(X)$, these cardinality constraints also suffice to compute SR^*, the maximum value of the sum-rate achievable using Marton's inner bound for a given broadcast channel. □

C. Proof of Theorem 3

Let $\bar{R}$ be the maximum sum-rate obtained by the randomized time-division strategy and R be that by Marton's inner bound. Note that

$$\begin{aligned} \bar{R} = &\max_{\substack{p(w,x):\\ W\to X\to(Y,Z)}} \min\{I(W;Y), I(W;Z)\}\\ &+ \mathrm{P}(W{=}0)I(X;Y|W{=}0) + \mathrm{P}(W{=}1)I(X;Z|W{=}1), \end{aligned}$$

where W is a binary random variable. We need to show $R = \bar{R}$. Clearly, we have $R \ge \bar{R}$ as $\bar{R}$ is a restriction of the choice of (U,V,W).

From Theorem 2, to evaluate the Marton's sum-rate for a binary input broadcast channel it suffices to look at $|\mathcal{W}| \leq 2$. Consider a (U, V, W, X) that achieves the maximum sum-rate R. Without loss of generality we consider two cases below. The two remaining cases follow by interchanging the roles of Y and Z and hence is omitted.

Case 1:

$$\begin{aligned} &I(X;Y|W{=}0) \geq I(X;Z|W{=}0) \text{ and} \\ &I(X;Y|W{=}1) \geq I(X;Z|W{=}1). \end{aligned} \tag{7}$$

Clearly

$$\begin{aligned} R &= \min\{I(W;Y), I(W;Z)\} \\ &\quad + I(U;Y|W) + I(V;Z|W) - I(U;V|W) \\ &= \min\{I(W;Y), I(W;Z)\} \\ &\quad + \mathrm{P}(W{=}0)\big(I(U;Y|W{=}0) + I(V;Z|W{=}0) - I(U;V|W{=}0)\big) \\ &\quad + \mathrm{P}(W{=}1)\big(I(U;Y|W{=}1) + I(V;Z|W{=}1) - I(U;V|W{=}1)\big) \\ &\overset{(a)}{\leq} \min\{I(W;Y), I(W;Z)\} \\ &\quad + \mathrm{P}(W{=}0)I(X;Y|W{=}0) + \mathrm{P}(W{=}1)I(X;Y|W{=}1) \\ &\leq \min\{I(W;Y), I(W;Z)\} + I(X;Y|W) \leq I(X;Y) \leq \bar{R}, \end{aligned}$$

where (a) follows from Theorem 1 and equations (7).

Case 2:

$$\begin{aligned} &I(X;Y|W{=}0) \geq I(X;Z|W{=}0) \text{ and} \\ &I(X;Y|W{=}1) \leq I(X;Z|W{=}1). \end{aligned} \tag{8}$$

Observe that

$$\begin{aligned} R &= \min\{I(W;Y), I(W;Z)\} \\ &\quad + I(U;Y|W) + I(V;Z|W) - I(U;V|W) \\ &= \min\{I(W;Y), I(W;Z)\} \\ &\quad + \mathrm{P}(W{=}0)\big(I(U;Y|W{=}0) + I(V;Z|W{=}0) - I(U;V|W{=}0)\big) \\ &\quad + \mathrm{P}(W{=}1)\big(I(U;Y|W{=}1) + I(V;Z|W{=}1) - I(U;V|W{=}1)\big) \\ &\overset{(b)}{\leq} \min\{I(W;Y), I(W;Z)\} \\ &\quad + \mathrm{P}(W{=}0)I(X;Y|W{=}0) + \mathrm{P}(W{=}1)I(X;Z|W{=}1) \\ &\leq \bar{R}, \end{aligned}$$

where (b) follows from Theorem 1 and equations (8).

Thus $R \leq \bar{R}$ and completes the proof of Theorem 3. □

III. Proof of Theorem 1

Theorem 1 is an inequality concerning random variables; however we will interpret the inequality in the language of broadcast channels. First fix a broadcast channel $\mathfrak{q}(y, z|x)$ and then an input distribution $p(x)$. Observe that Theorem 1 is equivalent to showing that when X is binary,

$$\begin{aligned} T(X) = &\max_{\substack{p(u,v|x) \\ (U,V)\to X\to(Y,Z)}} I(U;Y) + I(V;Z) - I(U;V) \\ \leq &\quad \max\{I(X;Y), I(X;Z)\}. \end{aligned}$$

Remark 4. As $T(X)$, $I(X;Y)$, and $I(X;Z)$ are continuous in the transition probabilities $\mathfrak{q}(y|x)$ and $\mathfrak{q}(z|x)$, it suffices to prove the inequality when $\mathfrak{q}(y|x)$ and $\mathfrak{q}(z|x)$ are positive.

Having transformed Theorem 1 into a maximization problem of computing $T(X)$, we use Remark 3 and restrict ourselves to random variables (U, V) such that $|\mathcal{U}| \leq |\mathcal{X}|, |\mathcal{V}| \leq |\mathcal{X}|$ and $X = f(U, V)$, a deterministic function of (U, V).

The following claim is a statement concerning any maximizer $p(u, v|x)$ that attains $T(X)$. For this claim we only assume that X is finite valued, not necessarily binary.

Claim 2. *Let $\mathfrak{q}(y|x) > 0$ and $\mathfrak{q}(z|x) > 0$ for every x, y, z. Let $X = f(U, V)$ and p.m.f. $p(u, v|x)$ attain $T(X)$. If $p(u) > 0$ and $p(v) > 0$ for a pair (u, v), then $p(u, v) > 0$.*

Proof: The proof uses perturbation to show that one can increase

$$I(U;Y) + I(V;Z) - I(U;V)$$

if the conditions in the theorem does not hold, thus contradicting the maximality of $p(u, v|x)$. To make our calculations simpler we assume that all logarithms are to the base e.

Suppose $p(u_1, v_1) = 0$ and $p(u_1) > 0$, $p(v_1) > 0$. Then we must have some $v_2 \neq v_1$ such that $p(u_1, v_2) > 0$ (otherwise $p(u_1) = 0$). Let $f(u_1, v_2) = x_1$. Perturb p at two points

$$q(u, v, x) = \begin{cases} p(u, v, x) - \epsilon & (u, v, x) = (u_1, v_2, x_1) \\ \epsilon & (u, v, x) = (u_1, v_1, x_1) \\ p(u, v, x) & \text{otherwise} \end{cases}$$

Notice that $q(u, x) = p(u, x)\ \forall u, x$ and $p(v, x) = q(v, x), v \notin \{v_1, v_2\}$. Now we perform the following manipulations

$$\begin{aligned} &I_q(U;Y) + I_q(V;Z) - I_q(U;V) \\ &\quad - \big(I_p(U;Y) + I_p(V;Z) - I_p(U;V)\big) \\ &= H_q(Y) + H_q(Z) + H_q(UV) - H_q(UY) - H_q(VZ) \\ &\quad - \big(H_p(Y) + H_p(Z) + H_p(UV) - H_p(UY) - H_p(VZ)\big) \\ &\overset{(a)}{=} H_q(UV) - H_q(VZ) - H_p(UV) + H_p(VZ) \\ &\overset{(b)}{=} -\epsilon \log \epsilon - (p(u_1, v_2) - \epsilon) \log(p(u_1, v_2) - \epsilon) \\ &\qquad + p(u_1, v_2) \log p(u_1, v_2) + \sum_{v,z} p(v, z) \log p(v, z) \\ &\qquad - \sum_{v,z} q(v, z) \log q(v, z) \\ &= -\epsilon \log \epsilon - \epsilon(1 + \log p(u_1, v_2)) + o(\epsilon) \\ &\qquad + \sum_{v,z} p(v, z) \log p(v, z) - \sum_{v,z} q(v, z) \log q(v, z). \end{aligned} \tag{9}$$

Here (a) follows since $q(u, x) = p(u, x)$ and this implies that $q(u, y) = p(u, y)$ and $q(z) = p(z)$ since the broadcast channel $\mathfrak{q}(y, z|x)$ remains fixed. Equality (b) follows from the following observation: distribution on (U, V) only changes for the pairs (u_1, v_1) and (u_1, v_2).

Express the induced perturbation on the pair (V, Z) as

$$q(v, z) = p(v, z) + \epsilon\lambda(v, z).$$

Note that $\sum_{v,z} \lambda(v, z) = 0$. Since the channels have positive transition probabilities we have $p(v, z) > 0\ \forall v, z$. Thus we can expand $H_q(V,Z)$, using Taylor series, as

$$-\sum_{v,z} q(v,z)\log q(v,z) = -\sum_{v,z} p(v,z)\log p(v,z) - \epsilon\sum_{v,z}\lambda(v,z)\log p(v,z) + o(\epsilon). \quad (10)$$

Combining equations (9) and (10) we have

$$\begin{aligned} &I_q(U;Y) + I_q(V;Z) - I_q(U;V) \\ &\quad - \big(I_p(U;Y) + I_p(V;Z) - I_p(U;V)\big) \\ &= -\epsilon\log\epsilon - \epsilon\big(1 + \log p(u_1,v_2) \\ &\quad + \sum_{v,z}\lambda(v,z)\log p(v,z)\big) + o(\epsilon). \end{aligned}$$

Clearly, this difference becomes positive as $\epsilon \to 0$, contradicting the optimality of $p(u,v)$. ■

In the rest of the proof we will seek maximizers $p(u,v|x)$, that attain $T(X)$, of the form: $\mathcal{U} = \mathcal{V} = \{0,1\}$, and $X = f(U,V)$. Further we also assume that if $p(u) > 0$ and $p(v) > 0$ for a pair (u,v), then $p(u,v) > 0$.

Since the proof strategy is slightly unconventional, we first outline it here. There are two equivalent forms and we will use both forms interchangeably. The first form is the original form which states the result as the information inequality

$$I(U;Y) + I(V;Z) - I(U;V) \le \max\{I(X;Y), I(X;Z)\}. \quad (11)$$

The second formulation is that of an optimization problem of computing

$$T(X) = \max_{\substack{p(u,v|x) \\ (U,V)\to X\to(Y,Z)}} I(U;Y) + I(V;Z) - I(U;V)$$

and then showing that $T(X) \le \max\{I(X;Y), I(X;Z)\}$.

The outline of the proof is as follows:

1) We first prove the inequality (11) for some special settings, or "trivial" cases. (Section III-A)
2) We then consider all possible functions $f(U,V)$ and show that it can be partitioned into 5 equivalent classes. We then show the inequality (11) for two of these classes and deduce that a nontrivial maximizer $p(u,v|x)$ that attains $T(X)$ cannot exist in the other classes.

A. *Proof under special settings*

For binary input X, for brevity let

$$a_i = \mathrm{P}(Y{=}i|X{=}0), \quad \hat{a}_i = \mathrm{P}(Y{=}i|X{=}1), \quad i = 1,\dots,|\mathcal{Y}|$$
$$b_i = \mathrm{P}(Z{=}i|X{=}0), \quad \hat{b}_i = \mathrm{P}(Z{=}i|X{=}1), \quad i = 1,\dots,|\mathcal{Z}|.$$

Since $U \to X \to Y$ and $V \to X \to Z$ are Markov chains, from data processing inequality, we know that

$$\begin{aligned} I(U;Y) \le I(X;Y), \quad I(U;Y) \le I(U;X), \\ I(V;Z) \le I(X;Z), \quad I(V;Z) \le I(V;X). \end{aligned} \quad (12)$$

With these inequalities, we first prove Theorem 1 for some special settings. Denote $X \perp\!\!\!\perp Y$ to mean independence of random variables.

SS1. $a_i \equiv \hat{a}_i$. This implies that $X \perp\!\!\!\perp Y$, thus $I(U;Y) = I(X;Y) = 0$. From (12) and the non-negativity of $I(U;V)$ we have $I(V;Z) - I(V;U) \le I(X;Z)$, i.e. Theorem 1 holds. Similarly Theorem 1 holds when $b_i \equiv \hat{b}_i$.

SS2. $U \perp\!\!\!\perp X$. Then $I(U;Y) = I(U;X) = 0$. Again from (12) and the non-negativity of $I(U;V)$ Theorem 1 holds. Similarly when $V \perp\!\!\!\perp X$, Theorem 1 also holds.

B. *Reduction to two nontrivial cases*

Notation: We use the notation: $U \wedge V$ (and), $U \vee V$ (or), $U \oplus V$ (xor), $\bar{U}$ (not).

Since U, V, and X are binary, there are 16 possible functions $X = f(U,V)$, and they can be classified into the following equivalent groups

G_1: $X = 0, X = 1$
G_2: $X = U, X = \bar{U}, X = V, X = \bar{V}$
G_3: $X = U \wedge V, X = \bar{U} \wedge V, X = U \wedge \bar{V}, X = \bar{U} \wedge \bar{V}$
G_4: $X = U \vee V, X = \bar{U} \vee V, X = U \vee \bar{V}, X = \bar{U} \vee \bar{V}$
G_5: $X = U \oplus V, X = \bar{U} \oplus V$

The reason that these are equivalent groups is that, in each group, all the cases can be reduced to the first case using bijections. Since bijections preserve mutual information, we just need to prove Theorem 1 for the first function in each group.

The case $X = U \vee V$ with $q(u,v)$, can be mapped to the case $X = U \wedge V$ with $p(u,v)$, using the bijection $p_{00} \leftrightarrow q_{11}$, $p_{01} \leftrightarrow q_{01}$, $p_{10} \leftrightarrow q_{10}$, $p_{11} \leftrightarrow q_{00}$. That is, we use $X = U \vee V \Leftrightarrow \bar{X} = \bar{U} \wedge \bar{V}$ to reduce the proof of the "or" case of one channel to the "and" case of another broadcast channel obtained by flipping U, V, and X.

So it remains to consider the first case of all the groups except G_4.

The first two cases are trivial. When $X = 0$, inequality (11) reduces to $-I(U;V) \le 0$, which is true.

When $X = U$, inequality (11) follows from $I(U;Y) = I(X;Y)$ and data processing inequality, $I(V;Z) \le I(V;U) = I(V;X)$ (see (12)).

Now for cases in G_3 and G_5, if $p(x) = 0$ for some x, then it reduces to G_1; if $p(u) = 0$ (or $p(v) = 0$) for some u (or v), then it reduces to cases in G_1 or G_2. Thus we can assume that $p(u) > 0$ for $u \in \{0,1\}$ and similarly $p(v) > 0$ for $v \in \{0,1\}$. Thus from Claim 2 our search for a nontrivial maximizer of $T(X)$ can be restricted to:

C_3: $X = U \wedge V$ with $p(x) > 0$ and $p(u,v) > 0$ for all u, v, x.
C_5: $X = U \oplus V$ with $p(x) > 0$ and $p(u,v) > 0$ for all u, v, x.

We are going to prove that there is no nontrivial local maximum (by a nontrivial local maximum we mean a distribution that does not reduce to one of the special settings discussed above) for these two cases. Hence $T(X)$ cannot be achieved in either of these cases and since the inequality is true in all other cases, we are done.

C. *Proof of XOR case*

Consider an additive perturbation

$$q(u,v,x) = p(u,v,x) + \varepsilon\lambda(u,v,x)$$

for some $\varepsilon \geq 0$.

Let $p_{uv} = \mathrm{P}(U{=}u, V{=}v)$, $p_{y|x} = \mathrm{P}(Y{=}y|X{=}x)$, $p_{z|x} = \mathrm{P}(Z{=}z|X{=}x)$, $\lambda_{uvx} = \lambda(u,v,x)$, and a similar short hand is used for other variables as well.

For a valid perturbation, we require that $\lambda_{uvx} \geq 0$ if the corresponding $p(u,v,x)$ is zero, i.e.

$$\lambda_{uvx} \geq 0, \quad \text{if } f(u,v) \neq x. \tag{13}$$

Further let us require the perturbation maintains $p(x)$ (hence $H(Y)$ and $H(Z)$), that is

$$\sum_{uv} \lambda_{uvx} = 0, \quad \forall x \in \mathcal{X}. \tag{14}$$

The first derivative of $I_q(U;Y) + I_q(V;Z) - I_q(U;V)$ with respect to ε can be expressed as $\sum_{uvx} \lambda_{uvx} C_{uvx}$, where

$$C_{uvx} = -\log p_{uv} + \sum_y p_{y|x} \log p_{uy} + \sum_z p_{z|x} \log p_{vz}. \tag{15}$$

At a local maximum $p(u,v,x)$ it must be true that the first derivative cannot be positive, i.e.

$$\sum_{uvx} \lambda_{uvx} C_{uvx} \leq 0, \tag{16}$$

for all valid perturbations λ_{uvx} satisfying (13) and (14).

For $x \in \mathcal{X}$, choose one pair (u_x, v_x) such that $f(u_x, v_x) = x$. This is possible since $p(x) > 0$. Using (14), we express $\lambda_{u_x v_x x}$ using other λ_{uvx} variables as

$$\lambda_{u_x v_x x} = - \sum_{uv \neq u_x v_x} \lambda_{uvx}.$$

The equation (16) can be now written as

$$\sum_{uvx: uv \neq u_x v_x} \lambda_{uvx} (C_{uvx} - C_{u_x v_x x}) \leq 0.$$

As the above inequality needs to hold for any signed $\{\lambda_{uvx} : f(u,v) = x, (u,v) \neq (u_x, v_x)\}$ and any nonnegative $\{\lambda_{uvx} : f(u,v) \neq x\}$, we must have

$$C_{uvx} = C_{u_x v_x x}, \quad \text{if } f(u,v) = x,$$
$$C_{uvx} \leq C_{u_x v_x x}, \quad \text{if } f(u,v) \neq x.$$

The conditions above can be re-written as follows:

Claim 3. *Let $f(u,v) = x$. For any (u_1, v_1) we have $C_{u_1 v_1 x} \leq C_{uvx}$, that is*

$$\log \frac{p_{u_1 v_1}}{p_{uv}} \geq \sum_y p_{y|x} \log \frac{p_{u_1 y}}{p_{uy}} + \sum_z p_{z|x} \log \frac{p_{v_1 z}}{p_{vz}}$$

We are now ready to prove a claim that is central to the proof of XOR case.

Claim 4. *If $f(u_1, v_1) = f(u_2, v_2) = x$, then*

$$p_{u_1 v_1} p_{u_2 v_2} \leq p_{u_1 v_2} p_{u_2 v_1}$$

where the equality holds iff $C_{u_1 v_2 x} = C_{u_2 v_1 x} = C_{u_1 v_1 x} (= C_{u_2 v_2 x})$.

Proof: Two uses of Claim 3 yield $C_{u_1 v_1 x} + C_{u_2 v_2 x} \geq C_{u_1 v_2 x} + C_{u_2 v_1 x}$. Substituting for the terms from (15) and canceling common terms we obtain

$$-\log p_{u_1 v_1} - \log p_{u_2 v_2} \geq -\log p_{u_1 v_2} - \log p_{u_2 v_1}.$$

From Claim 3 and above it is clear that equality holds iff $C_{u_1 v_2 x} = C_{u_2 v_1 x} = C_{u_1 v_1 x} (= C_{u_2 v_2 x})$. ■

Remark 5. Note that the two claims above hold for any maximizer of $T(X)$ and does not require that X is binary.

Now return to $X = U \oplus V$, notice that $f(0,0) = f(1,1) = 0$, hence by Claim 4 we have for p_{uv} that $p_{00} p_{11} \leq p_{01} p_{10}$; also $f(0,1) = f(1,0) = 1$, hence $p_{00} p_{11} \geq p_{01} p_{10}$. Thus we have

$$p_{00} p_{11} = p_{01} p_{10} \tag{17}$$

and by Claim 4, this holds iff $C_{010} = C_{100} = C_{000} = C_{110}$ and $C_{001} = C_{111} = C_{011} = C_{101}$. In particular, $C_{000} = C_{010}$ and $C_{001} = C_{011}$ imply that

$$\log \frac{p_{00}}{p_{01}} = \sum b_i \log \frac{b_i p_{00} + \hat{b}_i p_{10}}{b_i p_{11} + \hat{b}_i p_{01}} = \sum \hat{b}_i \log \frac{b_i p_{00} + \hat{b}_i p_{10}}{b_i p_{11} + \hat{b}_i p_{01}}$$

Taking a weighted sum, we get

$$(p_{00} + p_{10}) \log \frac{p_{00}}{p_{01}} = \sum (b_i p_{00} + \hat{b}_i p_{10}) \log \frac{b_i p_{00} + \hat{b}_i p_{10}}{b_i p_{11} + \hat{b}_i p_{01}}$$

From above and using K-L divergence, we have

$$\log \frac{p_{00}}{p_{01}} \geq \log \frac{p_{00} + p_{10}}{p_{11} + p_{01}} = \log \frac{p_{00}}{p_{01}}$$

where the last step holds since $p_{00} p_{11} = p_{01} p_{10}$. Now that the K-L divergence inequality is indeed an equality, we require

$$\frac{b_i p_{00} + \hat{b}_i p_{10}}{b_i p_{11} + \hat{b}_i p_{01}} \equiv \frac{p_{00}}{p_{01}}.$$

From the above we obtain

$$(p_{01} - p_{11})(b_i - \hat{b}_i) \equiv 0.$$

Similarly from $C_{100} = C_{110}$ and $C_{101} = C_{111}$, we can obtain

$$(p_{10} - p_{11})(a_i - \hat{a}_i) \equiv 0.$$

Now we have two cases

1) $b_i \equiv \hat{b}_i$, or $a_i \equiv \hat{a}_i$. In this case the Theorem holds (special setting SS1).
2) $p_{01} = p_{11}$, $p_{10} = p_{11}$. Combining this with $p_{00} p_{11} = p_{01} p_{10}$ (from (17)) one obtains that $p_{uv} = 1/4$, and as a result U, V and X are mutually independent. The Theorem holds (special setting SS2).

If neither of these two cases is satisfied, there would be no local maxima.

D. Proof of AND case

We will show that nontrivial local maxima can't be achieved when $p(x) > 0$ and $p(u,v) > 0$. In this case, $\mathrm{P}(X{=}1) = p_{11}$. Now we fix $p_{11} \in (0,1)$. Take (p_{10}, p_{01}) as the free variables, with $p_{00} = 1 - p_{11} - p_{01} - p_{10}$. Notice that $H(Y)$ and $H(Z)$

are fixed, the local maxima of $I(U;Y)+I(V;Z)-I(U;V)$ is the same as that of

$$
\begin{aligned}
&J(p_{10},p_{01}) := H(U,V)-H(U,Y)-H(V,Z)\\
&= -p_{00}\log p_{00}-p_{01}\log p_{01}-p_{10}\log p_{10}-p_{11}\log p_{11}\\
&\quad+\sum a_i(p_{00}+p_{01})\log[a_i(p_{00}+p_{01})]\\
&\quad+\sum(a_ip_{10}+\hat a_ip_{11})\log[a_ip_{10}+\hat a_ip_{11}]\\
&\quad+\sum b_i(p_{00}+p_{10})\log[b_i(p_{00}+p_{10})]\\
&\quad+\sum(b_ip_{01}+\hat b_ip_{11})\log[b_ip_{01}+\hat b_ip_{11}].
\end{aligned}
$$

At any local interior maximum, the gradient ∇J and Hessian matrix $\nabla^2 J$ must satisfy

$$\nabla J=\vec{0},\ \nabla^2 J\preceq \mathbf{0}, \tag{18}$$

where $\nabla^2 J\preceq\mathbf{0}$ denotes that $\nabla^2 J$ is negative semi-definite. We now compute the gradient and the Hessian to investigate locations of the local maxima.

1. First Derivative:

Differentiating w.r.t. the free variables we obtain:

$$
\begin{aligned}
\frac{\partial J}{\partial p_{10}}&=\log\frac{p_{00}}{p_{10}}-\sum a_i\log\frac{a_i(p_{00}+p_{01})}{a_ip_{10}+\hat a_ip_{11}}\\
\frac{\partial J}{\partial p_{01}}&=\log\frac{p_{00}}{p_{01}}-\sum b_i\log\frac{b_i(p_{00}+p_{10})}{b_ip_{01}+\hat b_ip_{11}}.
\end{aligned}
$$

The condition $\nabla J=\vec{0}$ implies that

$$\log\frac{p_{00}}{p_{10}}=\sum a_i\log\frac{a_i(p_{00}+p_{01})}{a_ip_{10}+\hat a_ip_{11}} \tag{19}$$

$$\log\frac{p_{00}}{p_{01}}=\sum b_i\log\frac{b_i(p_{00}+p_{10})}{b_ip_{01}+\hat b_ip_{11}}. \tag{20}$$

Remark 6. Equalities above are obvious from Claim 3 by noticing that $0\wedge 0=0\wedge 1=1\wedge 0=0$. This is expected as Claim 3 is a result from first derivative.

Using the concavity of logarithm, we have

$$\frac{p_{00}}{p_{10}}\le\sum\frac{a_i^2(p_{00}+p_{01})}{a_ip_{10}+\hat a_ip_{11}},\quad \frac{p_{00}}{p_{01}}\le\sum\frac{b_i^2(p_{00}+p_{10})}{b_ip_{01}+\hat b_ip_{11}}, \tag{21}$$

where the equalities hold iff

$$a_i\equiv c_a\hat a_i,\ b_i\equiv c_b\hat b_i,$$

for some constants c_a,c_b respectively. However since $\sum_i a_i=\sum_i\hat a_i=1$ we obtain that $c_a=1$ (similarly $c_b=1$). Thus equalities hold iff

$$a_i\equiv\hat a_i,\ b_i\equiv\hat b_i. \tag{22}$$

2. Second Derivative:

We now compute the Hessian $G:=\nabla^2 J$, The second derivatives are

$$
\begin{aligned}
G_{11}&=\frac{\partial^2 J}{\partial p_{10}^2}=-\frac{1}{p_{00}}-\frac{1}{p_{10}}+\frac{1}{p_{00}+p_{01}}+\sum\frac{a_i^2}{a_ip_{10}+\hat a_ip_{11}}\\
G_{12}&=G_{21}=-\frac{1}{p_{00}}\\
G_{22}&=\frac{\partial^2 J}{\partial p_{01}^2}=-\frac{1}{p_{00}}-\frac{1}{p_{01}}+\frac{1}{p_{00}+p_{10}}+\sum\frac{b_i^2}{b_ip_{01}+\hat b_ip_{11}}
\end{aligned}
$$

As $p_{01}>0$, we have

$$G_{11}\le-\frac{1}{p_{00}}-\frac{1}{p_{10}}+\frac{1}{p_{00}+p_{01}}+\frac{1}{p_{10}}<0.$$

Similarly we have $G_{22}<0$. For G with $G_{11}<0$ and $G_{22}<0$ to be negative semi-definite, it is necessary and sufficient that $\det(G)\ge 0$.

From (21) we have

$$
\begin{aligned}
G_{11}&\ge-\frac{1}{p_{00}}-\frac{1}{p_{10}}+\frac{1}{p_{00}+p_{01}}+\frac{p_{00}}{p_{10}(p_{00}+p_{01})}\\
&=-\frac{p_{01}(p_{00}+p_{10})}{p_{00}p_{10}(p_{00}+p_{01})}.
\end{aligned}
$$

And similarly

$$G_{22}\ge-\frac{p_{10}(p_{00}+p_{01})}{p_{00}p_{01}(p_{00}+p_{10})}.$$

It is clear that equalities in the above two inequalities hold iff (22) holds.

Since $G_{11},G_{22}<0$ we have

$$
\begin{aligned}
G_{11}G_{22}&\le\frac{p_{01}(p_{00}+p_{10})}{p_{00}p_{10}(p_{00}+p_{01})}\cdot\frac{p_{10}(p_{00}+p_{01})}{p_{00}p_{01}(p_{00}+p_{10})}\\
&=\frac{1}{p_{00}^2}=G_{12}^2,
\end{aligned}
$$

with equality holding only if (22) holds.

Thus a local maximum may exist only when the channel parameters satisfy (22). However when (22) holds, the inequality is true from the special setting SS1.

This completes the proof of Theorem 1.

IV. Sum-rate evaluations of inner and outer bounds for BSSC

We shall evaluate the inner and outer bounds for the BSSC (Figure 1). Apart from completeness, this section serves some purposes:

- We correct a minor typo in the evaluation of the maximum sum-rate of the outer bound [5].
- We also explicitly compute the maximum sum-rate obtained via the Körner-Marton outer bound for the BSSC.

A. Sum-rate evaluation of Marton's inner bound

We are evaluting the optimization problem described in Theorem 3 for BSSC. A simple calculation shows that $I(X;Y)\ge I(X;Z)$ iff $\mathrm{P}(X{=}0)\le\frac{1}{2}$.

If $\mathrm{P}(X{=}0|W{=}0),\mathrm{P}(X{=}0|W{=}1)\le\frac{1}{2}$ then

$$
\begin{aligned}
&\min\{I(W;Y),I(W;Z)\}+\mathrm{P}(W{=}0)I(X;Y|W{=}0)\\
&\quad+\mathrm{P}(W{=}1)I(X;Z|W{=}1)\\
&\le\min(I(W;Y),I(W;Z))+I(X;Y|W)\\
&\le I(X;Y)\\
&\le C
\end{aligned}
$$

where $C=h(0.2)-0.4=0.321928...$ is the single channel capacity. Similarly if $\frac{1}{2}\le\mathrm{P}(X{=}0|W{=}0),\ \mathrm{P}(X{=}0|W{=}1)\le 1$ then the sum-rate will be bounded by C.

Assume $0 \leq \mathrm{P}(X{=}0|W{=}0) \leq \frac{1}{2} \leq \mathrm{P}(X{=}0|W{=}1) \leq 1$. Let

$$\begin{aligned} d &:= \max_{p(x)} I(X;Y) - I(X;Z) \\ &= \max_{x\in[0,1]} h\left(\frac{x}{2}\right) - h\left(\frac{1-x}{2}\right) + 1 - 2x, \end{aligned}$$

where $\mathrm{P}(X=0)=x$ and $h(\cdot)$ is the binary entropy function. We can solve the above explicitly and obtain $x_{opt} = 0.5 - \sqrt{105}/30$ and $d = 0.1007295...$. Now observe that

$$\begin{aligned} &\min\{I(W;Y), I(W;Z)\} + \mathrm{P}(W{=}0)I(X;Y|W{=}0) \\ &\quad + \mathrm{P}(W{=}1)I(X;Z|W{=}1) \\ &\leq I(X;Y) + \mathrm{P}(W{=}1)(I(X;Z|W{=}1) \\ &\quad - I(X;Y|W{=}1)) \\ &\leq I(X;Y) + \mathrm{P}(W{=}1)d. \end{aligned}$$

Similarly

$$\begin{aligned} &\min\{I(W;Y), I(W;Z)\} + \mathrm{P}(W{=}0)I(X;Y|W{=}0) \\ &\quad + \mathrm{P}(W{=}1)I(X;Z|W{=}1) \\ &= I(X;Z) + \mathrm{P}(W{=}0)(I(X;Y|W{=}0) \\ &\quad - I(X;Z|W{=}0)) \\ &\leq I(X;Z) + \mathrm{P}(W{=}0)d. \end{aligned}$$

By adding the above two inequalities we obtain

$$\begin{aligned} &\min\{I(W;Y), I(W;Z)\} + \mathrm{P}(W{=}0)I(X;Y|W{=}0) \\ &\quad + \mathrm{P}(W{=}1)I(X;Z|W{=}1) \\ &\leq \frac{1}{2}\big(I(X;Y) + I(X;Z) + d\big). \end{aligned}$$

Thus from Theorem 3 we have that SR^*, the maximum value of sum-rate given by Marton's inner bound, satisfies

$$SR^* \leq \max_{p(x)} \frac{1}{2}\big(I(X;Y) + I(X;Z) + d\big).$$

The maximum of $I(X;Y)+I(X;Z) = 0.6225562...$ occurs when $\mathrm{P}(X{=}0) = \frac{1}{2}$ and hence substituting we obtain that $SR^* \leq 0.36164288...$.

To show that it is indeed on the boundary of the achievable region consider the joint distribution on X and W as follows:

$$\begin{aligned} \mathrm{P}(W{=}0) = \mathrm{P}(W{=}1) &= \frac{1}{2} \\ \mathrm{P}(X{=}0|W{=}0) = \mathrm{P}(X{=}1|W{=}1) &= 0.5 - \sqrt{105}/30. \end{aligned}$$

For this distribution we achieve the above value and hence $SR^* = 0.36164288...$.

B. Sum-rate evaluations of the outer bounds for BSSC

1) Bound 2: To evaluate maximum of the sum-rate of Bound 2 it was shown [5] that it suffices to consider $\mathrm{P}(X{=}0) = \frac{1}{2}$. (It is immediate using the skew-symmetry of the channel and the inherent symmetry of the outer bound expressions.)

The sum-rate maximum is hence given by

$$\max_{\substack{p(u,x), \mathrm{P}(x=0)=\frac{1}{2} \\ U\to X\to(Y,Z)}} I(U;Y) + I(X;Z|U)$$

or in other words maximizing

$$\max_{\substack{p(u,x), \mathrm{P}(x=0)=\frac{1}{2} \\ U\to X\to(Y,Z)}} I(X;Y) + I(X;Z|U) - I(X;Y|U).$$

Let $\mathrm{P}(X{=}0) = x$. In [7] it was shown that the curve

$$f(x) = I(X;Y) - I(X;Z) = h\left(\frac{x}{2}\right) - h\left(\frac{1-x}{2}\right) + 1 - 2x$$

is concave when $x \in [0, \frac{1}{2}]$ and convex when $x \in [\frac{1}{2}, 1]$. Further it was also shown that the lower convex envelope[3] was given by

$$g(x) = \begin{cases} \frac{5x}{4} f\left(\frac{4}{5}\right) & 0 \leq x \leq \frac{4}{5} \\ f(x) & \frac{4}{5} \leq x \leq 1 \end{cases} .$$

From the definition of the lower convex envelope, we know that when $x = \frac{1}{2}$

$$I(X;Y|U) - I(X;Z|U) \geq g\left(\frac{1}{2}\right)$$

and it easy to see that there is a binary U that achieves the equality.

Therefore

$$\begin{aligned} &\max_{\substack{p(u,x), \mathrm{P}(x=0)=\frac{1}{2} \\ U\to X\to(Y,Z)}} I(X;Y) + I(X;Z|U) - I(X;Y|U) \\ &\qquad = h\left(\frac{1}{4}\right) - 0.5 - g\left(\frac{1}{2}\right) = 0.3725562... \ . \end{aligned}$$

This is a correction to the implicit error we made in [5] while calculating the lower convex envelope and obtained a bound of 0.37111....

2) Körner-Marton outer bound:

Bound 3 (Körner-Marton outer bound [3]). *The union of rate-pairs (R_1, R_2) satisfying the following constraints:*

$$\begin{aligned} R_1 &\leq I(U;Y) \\ R_2 &\leq I(X;Z) \\ R_1 + R_2 &\leq I(U;Y) + I(X;Z|U) \end{aligned}$$

over all random variables such that $U \to X \to (Y,Z)$ forms a Markov chain forms an outer bound to the capacity region of the broadcast channel.

Denote this region as $\mathcal{OB}_1$. Similarly one can interchange the roles of the receivers Y and Z and will lead to yet another outer bound, and let this region be $\mathcal{OB}_2$. The intersection of these two regions is normally termed as the Körner-Marton outer bound.

Note that if $(a,b) \in \mathcal{OB}_1 \cap \mathcal{OB}_2$ then by skew symmetry of BSSC we will have that $(b,a) \in \mathcal{OB}_1 \cap \mathcal{OB}_2$. Since the region $\mathcal{OB}_1 \cap \mathcal{OB}_2$ is convex, this implies that $\left(\frac{a+b}{2}, \frac{a+b}{2}\right) \in \mathcal{OB}_1 \cap \mathcal{OB}_2$. Thus, to compute the maximum sum-rate, it suffices to consider the points of the form $(a,a) \in \mathcal{OB}_1 \cap \mathcal{OB}_2$.

Repeating the above observation, if a point $(R_1, R_2) = (a,a)$ belongs to $\mathcal{OB}_1$, by the skew-symmetry of BSSC, it

[3] more precisely, in [7] the upper concave envelope was characterized, and the characterization of the lower convex envelope follows by symmetry.

will also belong to $\mathcal{OB}_2$, hence to the intersection of the two regions.

Suppose we wish to compute

$$\max_{\substack{p(u,x):\\ U\to X\to(Y,Z)}} I(X;Y)+I(X;Z|U)-I(X;Y|U)$$

then from the earlier discussion, this will be the maximum over $x\in[0,1]$ of

$$h\left(\frac{x}{2}\right)-x-g(x)$$

It is easy to see that the global maximum occurs when $x\in[0,\frac{4}{5}]$ (otherwise maximum occurs when U is trivial and equals $I(X;Z)$). Taking derivatives we obtain that maximum occurs when

$$\frac{1}{2}\log_2\frac{2-x}{x}-1-\frac{5}{4}f\left(\frac{4}{5}\right)=0,$$

i.e.

$$x^*=\frac{2}{1+2^c}=0.4571429...,$$

where $c=2\left(1+\frac{5}{4}f\left(\frac{4}{5}\right)\right)=1.7548875...$

Thus the maximum sum-rate of Körner-Marton outer bound, SR_{KMOB} satisfies

$$\begin{aligned} SR_{KMOB} &\le \max_{\substack{p(u,x):\\ U\to X\to(Y,Z)}} I(X;Y)+I(X;Z|U)-I(X;Y|U)\\ &=0.3743955...\ . \end{aligned} \tag{23}$$

Consider a pair (U,X) such that

$$\begin{aligned} &\mathrm{P}(U{=}0)=1-a, && \mathrm{P}(X{=}0|U{=}0)=0,\\ &\mathrm{P}(U{=}1)=a, && \mathrm{P}(X{=}0|U{=}1)=\frac{4}{5} \end{aligned}$$

where $0.8*a=x^*$(defined above) or $a=0.5714286...$.

Observe that for this choice

$$\begin{aligned} I(U;Y)&=h\left(\frac{x^*}{2}\right)-ah(0.4)=0.2206837...\ ,\\ I(X;Z|U)&=0.1537118...\ ,\\ I(X;Z)&=0.3006499...\ . \end{aligned}$$

Hence $(R_1,R_2)=(0.1871978...,0.1871978...)$ lies on the boundary of the Körner-Marton outer bound. Note that the above point matches the bound given by (23). In summary, the maximum sum-rate given by Körner-Marton outer bound for the BSSC is $0.3743955...$.

Historical remarks

Perturbation method as a tool in computing bounds on cardinalities of auxiliary random variables were used by Amin Gohari and Venkat Anantharam [4]. The perturbations used in their work were support-preserving (or multiplicative) in nature. In [8], the authors used the perturbation technique (including the additive perturbation) to compute the local maximas of $I(U;Y)+I(V;Z)-I(U;V)$ for the BSSC channel. Using this technique they established the inequality in Theorem 1 for the BSSC channel. Working on a related problem, one of the authors realized that the inequality may be more generally true for all binary input broadcast channels. A (nontrivial) modification of the arguments in [8] yielded a proof for this fact, which was then presented in [9]. To present a complete picture to the community, it was decided to combine the related proofs in [8], [9] into a single paper.

V. Conclusion

An information theoretic inequality is established for some collections of five random variables. This inequality is used to show that the sum-rate given by Marton's inner bound is indeed equivalent to that given by the randomized time-division strategy. The inequality fails when $|\mathcal{X}|\ge 3$ so a natural question is whether there is a correct generalization for higher cardinality input-alphabets. It would also be useful to find a more intuitive (geometric) argument to shed more light into the actual counting of the sizes of typical sets. Here is an equivalent formulation which is related to the sizes of certain typical sets. It can be shown that the information inequality is equivalent to showing that

$$H(U|Y)+H(V|Z)\ge\min\{H(U,V|Y),H(U,V|Z)\}$$

whenever $(U,V)\to X\to(Y,Z)$ forms a Markov chain, $X=f(U,V)$ and $|\mathcal{X}|=2$.

Acknowedgements

The authors are grateful to a number of comments on this work from Abbas El Gamal and Amin Gohari. The authors are also thankful for valuable suggestions from the anonymous reviewers as well as Wei Yu, the associate editor.

References

[1] T. Cover, "Broadcast channels," *IEEE Trans. Info. Theory*, vol. IT-18, pp. 2–14, January, 1972.
[2] ——, "Comments on broadcast channels," *IEEE Trans. Info. Theory*, vol. IT-44, pp. 2524–2530, October, 1998.
[3] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Info. Theory*, vol. IT-25, pp. 306–311, May, 1979.
[4] A. A. Gohari and V. Anantharam, "Evaluation of Marton's inner bound for the general broadcast channel," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 608–619, 2012.
[5] C. Nair and A. El Gamal, "An outer bound to the capacity region of the broadcast channel," *IEEE Trans. Info. Theory*, vol. IT-53, pp. 350–355, January, 2007.
[6] C. Nair, "A note on outer bounds for broadcast channel," *Presented at International Zurich Seminar*, 2010.
[7] C. Nair and Z. V. Wang, "On the inner and outer bounds for 2-receiver discrete memoryless broadcast channels," *Proceedings of the ITA Workshop*, 2008.
[8] V. Jog and C. Nair, "An information inequality for the bssc channel," *Proceedings of the ITA Workshop*, 2010.
[9] C. Nair, Z. V. Wang, and Y. Geng, "An information inequality and evaluation of Marton's inner bound for binary input broadcast channels," *International Symposium on Information Theory*, 2010.
[10] R. W. Yeung, *Information Theory and Network Coding*, 1st ed. Springer Publishing Company, Incorporated, 2008.
[11] S. I. Gelfand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," *Probl. Inform. Transm.*, vol. 16(1), pp. 17–25, Jan. - Mar., 1980.

Yanlin Geng Yanlin Geng received his B.Sc. (mathematics) and M.Eng. (signal and information processing) from Peking University, and Ph.D. (information engineering) from The Chinese University of Hong Kong in 2006, 2009, and 2012, respectively. He is currently a postdoctoral researcher in the information engineering department at The Chinese University of Hong Kong.

Varun Jog Varun Jog received his Bachelor of Technology (B.Tech) degree in Electrical Engineering from the Indian Institute of Technology (IIT), Bombay in 2010. Since then he has been with the EECS department in University of California, Berkeley as a graduate student. His research interests are in fields of information theory and optimal transport.

Chandra Nair Chandra Nair received his Bachelor of Technology (B.Tech) degree in Electrical Engineering from the Indian Institute of Technology (IIT), Madras in 1999. Concurrently, he also completed a four year nurture program in Mathematics at the Institute of Mathematical Sciences (IMSc) under the auspices of the National Board of Higher Mathematics (NBHM). He received a Masters (2002) and PhD (2005) in electrical engineering from Stanford University. Subsequently he was a postdoctoral fellow at the theory group in Microsoft Research (Redmond) for two years. Following this he joined the IE department, CUHK, as an assistant professor in Fall 2007. His research interests are on fundamental problems in various interdisciplinary pursuits involving information theory, combinatorial optimization, statistical physics, and algorithms.

Zizhou Vincent Wang Vincent Wang is a system engineer at Altai Technologies located in the Hong Kong Science and Technology Park. He obtained his PhD from the department of Information Engineering at the Chinese University of Hong Kong in 2010. His research work mainly consisted of studying inner and outer bounds for two and three receiver broadcast channels.