On Marton's Inner Bound for the General Broadcast Channel

Amin Gohari, Abbas El Gamal, and Venkat Anantharam

***Abstract*— We establish several new results on Marton's inner bound on the capacity region of the general broadcast channel. Inspired by the fact that Marton's coding scheme without superposition coding is optimal in the Gaussian case, we consider the class of binary input degraded broadcast channels with no common message that have the same property. We characterize this class. We also establish new properties of Marton's inner bound that help restrict the search space for computing the Marton sum rate. In particular, we establish an extension of the XOR case of the binary inequality of Nair, Wang, and Geng.**

***Index Terms*— General broadcast channel, Marton's inner bound.**

I. Introduction

CONSIDER the two-receiver broadcast channel [2] with input alphabet $\mathcal{X}$, output alphabets $\mathcal{Y}$ and $\mathcal{Z}$, and conditional probability mass function $q(y, z|x)$. The capacity region of this channel is the set of rate triples (R_0, R_1, R_2) such that the sender X can reliably communicate a common message at rate R_0 to both receivers and two private messages at rates R_1 and R_2 to receivers Y and Z, respectively; see for example [3] for a detailed definition. The capacity region of this channel is known only for several special cases (including the vector Gaussian broadcast channel [22]) but is not known in general. The best known inner bound on the capacity region is due to Marton [4].

Marton's Inner Bound: The set of rate triples (R_0, R_1, R_2) such that

$$\begin{aligned} R_0 + R_1 &< I(U, W; Y), \\ R_0 + R_2 &< I(V, W; Z), \\ R_0 + R_1 + R_2 &< I(U, W; Y) + I(V; Z|W) \\ &\quad - I(U; V|W), \\ R_0 + R_1 + R_2 &< I(U; Y|W) + I(V, W; Z) \\ &\quad - I(U; V|W), \\ 2R_0 + R_1 + R_2 &\le I(U, W; Y) + I(V, W; Z) \\ &\quad - I(U; V|W) \end{aligned} \tag{1}$$

for some $(U, V, W, X, Y, Z) \sim p(u, v, w, x)q(y, z|x)$ constitutes an inner bound on the capacity of the two-receiver broadcast channel $q(y, z|x)$. Further, to compute this region it suffices to consider $|\mathcal{U}| \le |\mathcal{X}|, |\mathcal{V}| \le |\mathcal{X}|, |\mathcal{W}| \le |\mathcal{X}| + 4$, and $H(X|U, V, W) = 0$ [10]. Note that the constraint $H(X|U, V, W) = 0$ corresponds to a deterministic encoder for the code associated with joint probability mass function (pmf) $p(u, v, w, x)$ as one would expect (see Appendix IV of [12]). We denote by R_{sum} the maximum achievable sum-rate in Marton's inner bound, or the *Marton sum-rate* in short, that is, the maximum of $R_0 + R_1 + R_2$ over all (R_0, R_1, R_2) in Marton's inner bound. Note that

$$\begin{aligned} R_{\text{sum}} = \max_{p(u,v,w,x)} &\min\{I(W; Y), I(W; Z)\} \\ &+ I(U; Y|W) + I(V; Z|W) \\ &- I(U; V|W). \end{aligned} \tag{2}$$

It is not known if Marton's region is tight. Evaluation of Marton's inner bound in [10] has provided the possibility of checking whether it matches any of the known outer bounds (see [7], [10], [13], [16], [19], [24]). Furthermore it has motivated comparing multi-letter characterizations of Marton's inner bound with its single-letter version [14]. The following is a summary of some of these recent developments.

- It was originally shown that there is a gap between the Nair-El Gamal outer bound [19] and Marton's inner bound [7], [10], [13]. Thus either the inner bound, the outer bound, or both are loose.
- In [15], it was shown that the Nair-El Gamal outer bound is loose. The paper established a tighter outer bound for product broadcast channels and showed that this new outer bound coincides with Marton's inner bound for a new class of these channels.
- In [16], a new inequality was found for binary input broadcast channels. It was shown that for all random variables (U, V, X, Y, Z) such that $(U, V) \to X \to$

Manuscript received June 11, 2011; revised July 21, 2013; accepted January 16, 2014. Date of publication May 1, 2014; date of current version June 12, 2014. This work was supported in part by the National Science Foundation (NSF) under Grant CCF-0500234, Grant CCF- 0635372, Grant CNS-0627161, and Grant CNS-0910702, in part by the NSF Science and Technology Center under Grant CCF-0939370, in part by the Science of Information, and in part by Army Research Office, Multidisciplinary University Research Initiative, under Grant W911NF-08-1-0233 Tools for the Analysis and Design of Complex Multiscale Networks. This paper was presented in part at the 2010 IEEE International Symposium on Information Theory.

A. Gohari is with the Department of Electrical Engineering, Sharif University of Technology, Tehran 16846-13114, Iran (e-mail: aminzadeh@sharif.edu).

A. El Gamal is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: abbas@ee.stanford.edu).

V. Anantharam is with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: ananth@eecs.berkeley.edu).

Communicated by E. Erkip, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2014.2321384

(Y, Z) and $|\mathcal{X}| = 2$,

$$I(U;Y)+I(V;Z)-I(U;V)\leq \max\{I(X;Y), I(X;Z)\}. \tag{3}$$

To prove this the authors of [16] consider different mappings from $\mathcal{U}\times\mathcal{V}\mapsto\mathcal{X}$. Because of the cardinality bound of two on $\mathcal{U}$ and $\mathcal{V}$ ([10]) it suffices to argue, as the authors do, that the XOR mapping (i.e., $X = U \oplus V$) and the AND mapping (i.e., $X = U \wedge V$) cannot occur in any maximizer of $I(U;Y)+I(V;Z)-I(U;V)$. This inequality led to the simple representation of the Marton sum-rate for binary input broadcast channels as

$$\begin{aligned}\max_{p(w,x)} \min\ &\{I(W;Y), I(W;Z)\}\\ &+\mathbb{P}(W=0)I(X;Y|W=0)\\ &+\mathbb{P}(W=1)I(X;Z|W=1).\end{aligned}$$

Here $\mathcal{W} = \{0, 1\}$.

- In [23], extensions of inequality (3) for computing the entire Marton region were studied.
- New cardinality bounds for Marton's region in the private message case were derived in [21].

In this paper we establish the following results most of which are related to evaluating the Marton sum-rate. We believe that finding the correct extension of equation (3) to larger alphabets can be useful in computing the boundary of Marton's inner bound efficiently for a given channel, and comparing Marton's inner bound with its multi-letter characterizations to see if Marton's inner bound is optimal or not (see [14] for a discussion of this line of attack on determining the capacity region of the general broadcast channel).

1) (*Computing the Marton sum-rate*): To compute the Marton sum-rate, one has to solve a maximization problem over all $p(u, v, w, x)$. In Section II, we introduce an alternative form of this optimization problem (Lemma 1) and establish several restrictions on the optimizers that reduce the search space (Theorems 1 and 2). In particular we extend part of the result in [16], which is used to prove (3), to larger alphabets by showing that any $p(u, v, x)$ that maximizes $I(U;Y)+I(V;Z)-I(U;V)$ cannot satisfy $X = U \oplus V$ (i.e., X being the XOR of U and V) in a suitable sense. We also note that since the presentation of part of this work at the 2010 ISIT conference [5], our alternative form of expressing the Marton sum-rate in Lemma 1 has proved to very helpful in studying the Marton sum-rate, see [14].
2) (*Insufficiency of Marton's coding scheme without a superposition variable*): In Marton's inner bound (1), the auxiliary random variable W corresponds to the "superposition-coding" aspect of the bound, while U and V correspond to the "Marton-coding" aspect of the bound. Necessity of the "superposition-coding" aspect of the inner bound had previously been observed for a non-degraded broadcast channel [13]. For degraded channels, it is known that W is unnecessary for achieving the capacity region of Gaussian broadcast channels (through dirty paper coding) [17]. It is interesting to find out the extent to which this property extends to other degraded broadcast channels. To study this, we consider the class of binary input degraded broadcast channels. Theorem 3 shows that any channel in this class has to satisfy some restrictive conditions. In particular any $p(x)$ that maximizes $I(X;Y)$ must maximize $I(X;Z)$ as well.
3) (*A simple direct proof for optimality of superposition coding along certain directions*): For a general broadcast channel, the rate pair (R_1, R_2) is said to be achievable by superposition coding if we have

$$\begin{aligned}R_1 &\leq I(X;Y|U),\\ R_2 &\leq I(U;Z),\\ R_1+R_2 &\leq I(X;Y)\end{aligned} \tag{4}$$

for some $(U, X, Y, Z) \sim p(u, x)q(y, z|x)$, or we have the similar set of inequalities with the role of Y and Z interchanged, see [3, Th. 5.1].[1]

Consider the problem of computing the maximum of $\lambda_0R_0+\lambda_1R_1+\lambda_2R_2$ over all (R_0, R_1, R_2) in the capacity region of the general broadcast channel where λ_0, λ_1 and λ_2 are real numbers such that $\lambda_0 \geq \lambda_1 + \lambda_2$. This would characterize part of the boundary of the capacity region since any convex region can be expressed as the intersection of the half spaces formed by its supporting hyperplanes (see [20, pp. 50-51]). We observe in Theorem 4 that superposition coding is tight along directions corresponding to $\lambda_0 \geq \lambda_1 + \lambda_2$. Our contribution here is a simple direct argument based on the characterization of the capacity region of a degraded BC [9].

The following section describes the above results in detail. The proofs of these results are contained in Section III with some of the details relegated to the appendices.

II. Main Results

Let $\mathcal{C}(q(y, z|x))$ denote the capacity region of the broadcast channel $q(y, z|x)$, and $\mathcal{C}_M(q(y, z|x))$ denote Marton's inner bound as given in (1). We use the standard notation, $X^i = (X_1, X_2, \ldots X_i)$ and $X_i^n = (X_i, X_{i+1}, \ldots, X_n)$.

A. Computing the Marton Sum-Rate

We establish the following alternative representation of the Marton sum-rate, R_{sum} defined in eqn. (2).

Lemma 1: $R_{\text{sum}} = \min_{\lambda\in[0,1]} T_\lambda$, where for any $\lambda \in [0, 1]$,

$$\begin{aligned}T_\lambda = \max_{p(u,v,w,x)} \big(&\lambda I(W;Y) + (1-\lambda)I(W;Z)\\ &+I(U;Y|W) + I(V;Z|W) - I(U;V|W)\big).\end{aligned} \tag{5}$$

Remark 1: Since the presentation of part of this work at the 2010 ISIT conference [5], this lemma has proved very helpful in studying the Marton sum-rate, see [14]. Some interesting properties of T_λ such as its convexity in λ, its connection to the outer bound, and its factorization (for products of broadcast channels) have been investigated in [14] and [15]. An alternative proof of Lemma 1 using a theorem by Terkelsen is reported in [14].

[1]Further, we can take $|\mathcal{U}| \leq |\mathcal{X}| + 1$ without loss of generality in the definition of this region.

Observe that $\lambda I(W;Y)+(1-\lambda)I(W;Z)$ depends only on $p(w,x)$. The term $I(U;Y|W)+I(V;Z|W)-I(U;V|W)$ can be written as $\sum_w p(w)\big(I(U;Y|W=w)+I(V;Z|W=w)-I(U;V|W=w)\big)$. Then, we have

$$T_\lambda = \max_{p(w,x)} \Bigg[\lambda I(W;Y)+(1-\lambda)I(W;Z) + \sum_w p(w) \max_{p(u,v|w,x)} \big[I(U;Y|W=w) + I(V;Z|W=w) - I(U;V|W=w)\big]\Bigg].$$

One can think of this maximization as

$$\max_{p(w,x)} \lambda I(W;Y)+(1-\lambda)I(W;Z)+\sum_w p(w)T(p(x|w)),$$

where $T(p(x))$ is the maximum of $I(U;Y)+I(V;Z)-I(U;V)$ over all $p(u,v|x)$ with $p(x)$ and $H(X|U,V)=0$. The first theorem of this section proves results for the problem of maximizing $T(p(x))$, while the second theorem concerns the problem of maximizing T_λ.

To state our main result we need the following two definitions.

Definition 1: The input symbols x_0 and x_1 are said to be indistinguishable *by the channel if $q(y|x_0)=q(y|x_1)$ for all y, and $q(z|x_0)=q(z|x_1)$ for all z. A channel $q(y,z|x)$ is said to be* irreducible *if no two of its inputs symbols are indistinguishable by the channel.*

Definition 2: Let $\mathcal{U}=\{u_1,u_2,\ldots,u_{|\mathcal{U}|}\}$, $\mathcal{V}=\{v_1,\ldots,v_{|\mathcal{V}|}\}$ be finite sets, and ξ be a deterministic mapping from $\mathcal{U}\times\mathcal{V}$ to $\mathcal{X}$. One can represent the mapping by a table having $|\mathcal{U}|$ rows and $|\mathcal{V}|$ columns with the rows indexed by $u_1,u_2,\ldots,u_{|\mathcal{U}|}$ and the columns indexed by $v_1,v_2,\ldots,v_{|\mathcal{V}|}$. In cell (i,j), we write $\xi(u_i,v_j)$ for the symbol x that (u_i,v_j) is being mapped to. The profile *of the row i is defined as a vector of size $|\mathcal{X}|$ counting the number of occurrences of the elements of $\mathcal{X}$ in the row i. In other words if $\mathcal{X}=\{x_1,x_2,\ldots,x_{|\mathcal{X}|}\}$, the element k of the profile of row i is the number of times that x_k shows up in row i of the table. The* profile *of column j is defined similarly. Define the* profile *of the table to be a vector of size $(|\mathcal{U}|+|\mathcal{V}|)|\mathcal{X}|$ formed by concatenating the profile vectors of the rows and the columns of the table. Denote the profile vector of the mapping ξ by $\overrightarrow{v_\xi}$.*

We establish the following.

Theorem 1: Consider an arbitrary irreducible broadcast channel $q(y,z|x)$, where $q(y|x)>0, q(z|x)>0$ for all x,y,z. Fix a pmf $p(x)$. Consider any $p(u,v|x)$ that maximizes $I(U;Y)+I(V;Z)-I(U;V)$, where X is a function of (U,V). Without loss of generality assume that $p(u)>0$ for all $u\in\mathcal{U}$ and $p(v)>0$ for all $v\in\mathcal{V}$. Let $x=\xi(u,v)$ denote the deterministic mapping from $\mathcal{U}\times\mathcal{V}$ to $\mathcal{X}$. Then the following conditions must hold:

1) *$p(u,v)>0$, $p(u,y)>0$, and $p(v,z)>0$ for all u,v,y and z.*
2) *The profile vector of the mapping ξ, $\overrightarrow{v_\xi}$, cannot be written as $\sum_{t=1}^M \alpha_t \overrightarrow{v_{\xi_t}}$, where ξ_t (for $t=1,2,3,\ldots,M$) are deterministic mappings from $\mathcal{U}\times\mathcal{V}$ to $\mathcal{X}$ not equal to ξ, and α_t, $t=1,\ldots,M$, are nonnegative numbers such that $\sum_{t=1}^M \alpha_t=1$.*
3) *Define the functions:*

$$f_u(x)=\sum_y q(y|x)\log p(u,y),$$

$$g_v(x)=\sum_z q(z|x)\log p(v,z),$$

$$h(x)=\min_{u'\in\mathcal{U},v'\in\mathcal{V}} \big(\log(p(u',v'))-f_{u'}(x)-g_{v'}(x)\big).$$

These definitions make sense because of the first part of this theorem. Then, for any u and v, the following two equations hold:

$$\log(p(u,v))=\max_x[f_u(x)+g_v(x)+h(x)],$$

and

$$p(x_0|u,v)=1 \text{ for some } x_0\in\mathcal{X} \Rightarrow x_0\in\arg\max_x[f_u(x)+g_v(x)+h(x)].$$

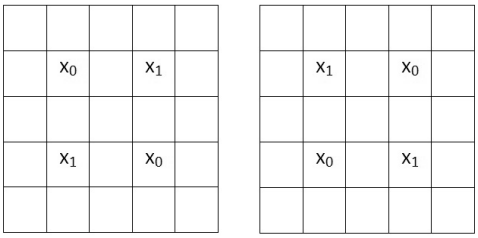

Fig. 1. If we have a mapping with the XOR structure, we can get another mapping with the same profile by switching x_0 and x_1 of four cells of the mappings.

Remark 2: These constraints imply restrictions on the maximizers. The second part of the theorem implies that one cannot find distinct $u_0,u_1\in\mathcal{U}$, distinct $v_0,v_1\in\mathcal{V}$ and distinct $x_0,x_1\in\mathcal{X}$ such that $p(x_0|u_0,v_0)=p(x_0|u_1,v_1)=p(x_1|u_1,v_0)=p(x_1|u_0,v_1)=1$. To see this, let the mapping ξ_1 be equal to ξ except that (u_0,v_0) and (u_1,v_1) are mapped to x_1 (instead of x_0), and (u_1,v_0) and (u_0,v_1) are mapped to x_0 (instead of x_1); see Figure 1. The mapping ξ_1 has the same profile vector as ξ. Thus we can write the original profile as a convex combination of other profiles (i.e., $\overrightarrow{v_\xi}=\sum_{t=1}^M \alpha_t\overrightarrow{v_{\xi_t}}$ holds for the choice of $M=1$, ξ_1, and $\alpha_1=1$). Thus the second part implies that it cannot happen. Similarly the mapping shown in Figure 2 cannot occur because there is another mapping with the same profile.

Remark 3: A special case of the result of the second part of the theorem for a binary X has been studied in [16], where the authors show that the optimizers of the expression $\max_{p(u,v,x)} I(U;Y)+I(V;Z)-I(U;V)$ are not of the form $X=U\oplus V$ (i.e., the XOR mapping from (U,V) to X). Their proof applies to binary input broadcast channels by considering the first order derivatives of $I(U;Y)+I(V;Z)-I(U;V)$ for local perturbations that preserve the alphabet size of $\mathcal{U}$ and $\mathcal{V}$. This proof technique, however, cannot be used to refute the XOR pattern for larger input alphabets. Our proof goes beyond theirs by considering perturbations that extend the

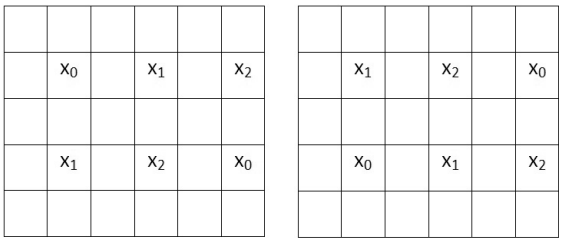

Fig. 2. Another mapping that cannot occur because one can find another mapping with the same profile.

alphabet of $\mathcal{U}$ and $\mathcal{V}$. The proof considers a certain $p(u,v,x)$ whose mapping contains such an XOR pattern. It explicitly *constructs a joint pmf $p(u',v',x)$ such that $I(U';Y) > I(U;Y)$, $I(V';Z) = I(V;Z)$, and $I(U';V') = I(U;V)$. In constructing $p(u',v',x)$, we extend the alphabet of U.*

Remark 4: The second part of the theorem holds more generally for any $p(u,v|x)$ maximizing the weighted expression $\lambda_1 I(U;Y) + \lambda_2 I(V;Z) - I(U;V)$, where $\lambda_1, \lambda_2 > 0$ and X is a function of (U,V). If the condition in the second part is violated, one can use the explicit construction given in the proof of the theorem to construct a new $p(u,v,x)$ such that the term $I(U;Y)$ increases while the terms $I(V;Z)$ and $I(U;V)$ remain constant. Thus, the weighted expression $\lambda_1 I(U;Y) + \lambda_2 I(V;Z) - I(U;V)$ also increases.

Remark 5: Assume that all we know about the mapping pattern is that $x_0 = \xi(u_0,v_0) = \xi(u_1,v_1)$ for some x_0. Then the third part of the theorem implies that $p(u_0,v_0)p(u_1,v_1) \le p(u_1,v_0)p(u_0,v_1)$. This holds since

$$\begin{aligned}\log p(u_0,v_0)+\log p(u_1,v_1) &= f_{u_0}(x_0)+g_{v_0}(x_0)+h(x_0)\\ &\quad +f_{u_1}(x_0)+g_{v_1}(x_0)+h(x_0)\\ &= f_{u_0}(x_0)+g_{v_1}(x_0)+h(x_0)\\ &\quad +f_{u_1}(x_0)+g_{v_0}(x_0)+h(x_0)\\ &\le \max_x f_{u_0}(x)+g_{v_1}(x)+h(x)\\ &\quad +\max_x f_{u_1}(x)+g_{v_0}(x)+h(x)\\ &= \log p(u_0,v_1)+\log p(u_1,v_0).\end{aligned}$$

Let us next turn to the evaluation of the entire Marton sum-rate expression (including the W terms). Recall the definition of T_λ in (5) for $\lambda \in [0,1]$. The next theorem restricts the search space for computing T_λ. For this theorem, we only deal with broadcast channels $q(y,z|x)$ with strictly positive transition matrices, i.e., when $q(y|x) > 0, q(z|x) > 0$ for all x, y, z. In order to evaluate T_λ when $q(y|x)$ or $q(z|x)$ become zero for some y or z, one can use the continuity of T_λ in $q(y,z|x)$ and take the limit of T_λ for a sequence of channels with positive entries converging to the desired channel. The reason for dealing with this class of broadcast channels should become clear from the following corollary to the first part of Theorem 1.

Corollary 1: Take an arbitrary broadcast channel $q(y,z|x)$ with strictly positive transition matrices (i.e. $q(y|x) > 0, q(z|x) > 0$ for all x,y,z). Let $p(u,v,w,x)$ be an arbitrary joint pmf maximizing T_λ for some $\lambda \in [0,1]$ where $H(X|U,V,W) = 0$. If $p(u,w)$ and $p(v,w)$ are positive for some triple (u,v,w), then it must be the case that $p(u,v,w) > 0$, $p(u,w,y) > 0$ and $p(v,w,z) > 0$ for all y and z.

We are now ready to state the following.

Theorem 2: Consider an arbitrary irreducible broadcast channel $q(y,z|x)$ with strictly positive transition matrices. In computing T_λ for some $\lambda \in [0,1]$, it suffices to take the maximum over auxiliary random variables $p(u,v,w,x)q(y,z|x)$ simultaneously satisfying the following constraints:

1) *$|\mathcal{U}| \le \min(|\mathcal{X}|,|\mathcal{Y}|)$, $|\mathcal{V}| \le \min(|\mathcal{X}|,|\mathcal{Z}|)$, $|\mathcal{W}| \le |\mathcal{X}|$.*
2) *$H(X|U,V,W) = 0$. Given w where $p(w) > 0$, we use $x = \xi^{(w)}(u,v)$ to denote the deterministic mapping from $\mathcal{U}_w \times \mathcal{V}_w$ to $\mathcal{X}$. Here $\mathcal{U}_w$ is the set of $u \in \mathcal{U}$ such that $p(u|w) > 0$ and $\mathcal{V}_w$ is the set of $v \in \mathcal{V}$ such that $p(v|w) > 0$.*
3) *For arbitrary w such that $p(w) > 0$, the profile vector of the mapping $\xi^{(w)}$, $\overrightarrow{v_{\xi^{(w)}}}$, cannot be written as $\sum_{t=1}^{M} \alpha_t \overrightarrow{v_{\xi_t^{(w)}}}$, where $\xi_t^{(w)}$ (for $t = 1,2,3,\ldots,M$) are deterministic mappings from $\mathcal{U}_w \times \mathcal{V}_w$ to $\mathcal{X}$ not equal to $\xi^{(w)}$, and α_t are non-negative numbers adding up to one, i.e. $\sum_{t=1}^{M} \alpha_t = 1$.*
4) *For arbitrary w such that $p(w) > 0$, define the functions*

$$\begin{aligned}f_{u,w}(x) &= \sum_y q(y|x)\log p(uy|w),\\ g_{v,w}(x) &= \sum_z q(z|x)\log p(vz|w),\\ h_w(x) &= \min_{u'\in\mathcal{U}_w, v'\in\mathcal{V}_w} \big(\log(p(u'v'|w)) - f_{u',w}(x)\\ &\qquad - g_{v',w}(x)\big).\end{aligned}$$

These definitions make sense because of Corollary 1. Then, for any $u \in \mathcal{U}_w$ and $v \in \mathcal{V}_w$, the following two equations hold:

$$\log(p(uv|w)) = \max_x [f_{u,w}(x) + g_{v,w}(x) + h_w(x)],$$

and

$$\begin{aligned}&p(x_0|u,v,w) = 1 \text{ for some } x_0 \in \mathcal{X}\\ &\quad \Rightarrow x_0 \in argmax_x [f_{u,w}(x) + g_{v,w}(x) + h_w(x)].\end{aligned}$$

5) *Given any w, random variables U_w, V_w, X_w, Y_w, Z_w distributed according to $p(u,v,x,y,z|w)$ satisfy the following:*

$$\begin{aligned}&I(\overline{U};Y_w) \ge I(\overline{U};V_w,Z_w) \text{ for any } \overline{U} \to U_w \to V_w X_w Y_w Z_w,\\ &I(\overline{V};Z_w) \ge I(\overline{V};U_w,Y_w) \text{ for any } \overline{V} \to V_w \to U_w X_w Y_w Z_w.\end{aligned}$$

Remark 6: The first part imposes cardinality bounds on $|\mathcal{U}|$ and $|\mathcal{V}|$ that are better than those reported in [10]. The improved cardinality bounds, however, are only for T_λ and not for the entire capacity region. The constraint of the second part is not new, and can be found in [10]. The other constraints are useful in restricting the search space due to the constraints imposed on $p(u,v,w,x)$. For instance, the third and fourth parts restrict the set of possible mappings, as discussed in Remarks 2 and 5. The constraint of the last part was inspired by studying the binary inequality $I(U;Y) + I(V;Z) - I(U;V) \le \max(I(X;Y), I(X;Z))$. This inequality can be expressed as $I(U;Y) + I(V;Z) -$

$I(U;V) \le \max(I(U,V;Y), I(U,V;Z))$ *or alternatively as* $I(U;Y) \le I(U;V,Z)$ *and* $I(V;Z) \le I(V;U,Y)$. *The last part shows that the channels* $p(y,z|u)$ *and* $p(y,z|v)$ *are less noisy channels in opposite directions. It has been recently shown [21] that this property can be further developed to establish the improved cardinality bound* $|\mathcal{U}| + |\mathcal{V}| \le |\mathcal{X}|$.[2]

B. Insufficiency of Marton's Coding Scheme Without W

When $R_0 = 0$ (private messages only) and $W = \emptyset$, Marton's inner bound (1) reduces to the set of rate pairs (R_1, R_2) such that

$$R_1 \le I(U;Y|Q), \tag{6}$$

$$R_2 \le I(V;Z|Q), \tag{7}$$

$$R_1 + R_2 \le I(U;Y|Q) + I(V;Z|Q) - I(U;V|Q) \tag{8}$$

for some random variables $(Q, U, V, X, Y, Z) \sim p(q)p(u,v,x|q)q(y,z|x)$. This inner bound corresponds to the "Marton-coding" aspect of the Marton bound.

It is known that this inner bound is tight for Gaussian broadcast channels (through dirty paper coding), implying that W is unnecessary for achieving the capacity region of this class of broadcast channels [17]. Thus, one might ask to what extent this property continues to hold for not-necessarily Gaussian degraded broadcast channels. For degraded broadcast channels, the Marton region with the superposition variable W equals the true capacity region. We are looking for conditions that imply achievability of the capacity region by using only the "Marton-coding" aspect of the bound. To study this question, we consider the class of binary-input degraded broadcast channels (receiver Z is a degraded version of receiver Y). Here W is unnecessary for achieving the sum-rate (which is $\max_{p(x)} I(X;Y)$). Thus we need consider the entire capacity region in order to answer this question. For simplicity, we restrict ourselves to the set of binary-input degraded broadcast channels where $q(y|x) > 0$ for all $(x, y) \in (\mathcal{X}, \mathcal{Y})$ and denote it by C_{bd}. Let C_{bd}^m be the set of broadcast channels in C_{bd} where W is unnecessary for achieving the capacity region (i.e., the inner bound given by (6)-(8) is tight). We show that C_{bd}^m is a very small subset of C_{bd}. In particular a broadcast channel would not belong to C_{bd}^m if the $p(x)$ that maximizes $I(X;Y)$ is different from the one that maximizes $I(X;Z)$.

To state our result let us further define C_{bd}^r to be the set of broadcast channels in C_{bd} whose private message capacity region is the simple time-division region, i.e., the capacity region is the set of rate pairs (R_1, R_2) such that $R_1/C_1 + R_2/C_2 \le 1$, where $C_1 = \max_{p(x)} I(X;Y)$ and $C_2 = \max_{p(x)} I(X;Z)$. We prove the following.

Theorem 3: We have $C_{bd}^m = C_{bd}^r$. *Further, any broadcast channel belonging to* $C_{bd}^r = C_{bd}^m$ *satisfies the following: any* $p(x)$ *maximizing* $I(X;Y)$ *is also a maximizer for* $I(X;Z)$. *More generally for any* $p(x)$, $I(X;Z)/C_2 \ge I(X;Y)/C_1$.

[2]Essentially, the idea of [21] is to consider the two subsets of the probability simplex on $\mathcal{X}$ that one would get by fixing $p(x|u)$ and $p(x|v)$ and varying $p(u)$ and $p(v)$, respectively. The less noisy property implies that the function $p(x) \mapsto H(Y) - H(Z)$ is convex on one of these subsets and concave on the other. This is used to prove the cardinality reduction statement.

Example 1: The binary symmetric broadcast channel, as defined in [3, p. 107], is often considered the discrete counterpart of the Gaussian BC. It turns out, however, that it does not belong to C_{bd}^m, *since its private message capacity region is not equal to the triangular time-division region.*

C. Optimality of Superposition Coding Along Certain Directions

In order to state the main result of this section, we need the following.

Definition 3: [9] Let $\mathcal{C}_{d_1}(q(y,z|x))$ and $\mathcal{C}_{d_2}(q(y,z|x))$ denote the *degraded message set capacity regions*, i.e., when $R_1 = 0$ and $R_2 = 0$, respectively. The capacity region $\mathcal{C}_{d_1}(q(y,z|x))$ is the set of of rate pairs (R_0, R_2) such that

$$R_0 \le I(W;Y),$$
$$R_2 \le I(X;Z|W),$$
$$R_0 + R_2 \le I(X;Z)$$

for some random variables $(W, X, Y, Z) \sim p(w,x)q(y,z|x)$. The capacity region $\mathcal{C}_{d_2}(q(y,z|x))$ is defined similarly.

We now state the result of this subsection.

Theorem 4: For a broadcast channel $q(y,z|x)$ *and real numbers* λ_0, λ_1 *and* λ_2 *such that* $\lambda_0 \ge \lambda_1 + \lambda_2$,

$$\max_{(R_0,R_1,R_2)\in\mathcal{C}(q(y,z|x))} (\lambda_0 R_0 + \lambda_1 R_1 + \lambda_2 R_2)$$
$$= \max\Big\{ \max_{(R_0,R_2)\in\mathcal{C}_{d_1}(q(y,z|x))} (\lambda_0 R_0 + \lambda_2 R_2), \max_{(R_0,R_1)\in\mathcal{C}_{d_2}(q(y,z|x))} (\lambda_0 R_0 + \lambda_1 R_1)\Big\}.$$

Corollary 2: The above observation essentially says that if $\lambda_0 \ge \lambda_1 + \lambda_2$, *then* a *maximum of* $\lambda_0 R_0 + \lambda_1 R_1 + \lambda_2 R_2$ *over triples* (R_0, R_1, R_2) *in the capacity region occurs when either* $R_1 = 0$ *or* $R_2 = 0$.

Remark 7. Since $\mathcal{C}_{d_1}(q(y,z|x)) \cup \mathcal{C}_{d_1}(q(y,z|x)) \subset \mathcal{C}_M(q(y,z|x)) \subset \mathcal{C}(q(y,z|x))$, *the above lemma implies that Marton's inner bound is tight along the direction of each such* $(\lambda_0, \lambda_1, \lambda_2)$, *i.e.,*

$$\max_{(R_0,R_1,R_2)\in\mathcal{C}(q(y,z|x))} (\lambda_0 R_0 + \lambda_1 R_1 + \lambda_2 R_2) = \max_{(R_0,R_1,R_2)\in\mathcal{C}_M(q(y,z|x))} (\lambda_0 R_0 + \lambda_1 R_1 + \lambda_2 R_2)$$

whenever $\lambda_0 \ge \lambda_1 + \lambda_2$.

Remark 8. One way to prove the theorem is to use a rate transfer argument to exchange between the common rate and the individual rates. As discussed in the next section, such a proof requires the use a result by Willems [11], which shows that the maximal *probability of error capacity region is equal to the* average *probability of error capacity region. Our contribution here is to provide a simple direct proof for optimality along these directions of superposition coding (without using the result of Willems, and without explicitly exchanging between the common rate and the individual rates).*

III. Proofs

A. Computing the Marton Sum-Rate

(Proof of Lemma 1). We would like to show that $R_{\text{sum}} = \min_{0\le\lambda\le 1} T_\lambda$. To do so, we need to argue that the following exchange of max and min is legitimate:

$$\begin{aligned}&\max_{p(u,v,w,x)} \min_{\lambda\in[0,1]} \lambda I(W;Y)+(1-\lambda)I(W;Z)+I(U;Y|W)\\&\qquad +I(V;Z|W)-I(U;V|W)\\&=\min_{\lambda\in[0,1]} \max_{p(u,v,w,x)} \lambda I(W;Y)+(1-\lambda)I(W;Z)+I(U;Y|W)\\&\qquad +I(V;Z|W)-I(U;V|W).\end{aligned}$$

Let $\mathcal{D}$ be the union over all $p(u,v,w,x)$ of real pairs (d_1,d_2) satisfying

$$\begin{aligned}d_1 &\le I(W;Y)+I(U;Y|W)+I(V;Z|W)-I(U;V|W),\\ d_2 &\le I(W;Z)+I(U;Y|W)+I(V;Z|W)-I(U;V|W).\end{aligned}$$

We claim that this region is convex. Take two points (d_1,d_2) and (d_1',d_2') in the region. Corresponding to these are joint pmfs $p(u_1,v_1,w_1,x_1)q(y_1,z_1|x_1)$ and $p(u_2,v_2,w_2,x_2)q(y_2,z_2|x_2)$. Take a uniform binary random variable Q independent of all the previously defined random variables. Set $U=U_Q$, $V=V_Q$, $W=(Q,W_Q)$, $X=X_Q$, $Y=Y_Q$, $Z=Z_Q$. We then have

$$\begin{aligned}&I(W;Y)+I(U;Y|W)+I(V;Z|W)-I(U;V|W)\\&\quad=I(W_Q,Q;Y_Q)+I(U_Q;Y_Q|W_Q,Q)\\&\qquad+I(V_Q;Z_Q|W_Q,Q)-I(U_Q;V_Q|W_Q,Q)\\&\quad\ge I(W_Q;Y_Q|Q)+I(U_Q;Y_Q|W_Q,Q)\\&\qquad+I(V_Q;Z_Q|W_Q,Q)-I(U_Q;V_Q|W_Q,Q)\\&\quad=\frac{1}{2}\big(I(W_1;Y_1)+I(U_1;Y_1|W_1)+I(V_1;Z_1|W_1)\\&\qquad-I(U_1;V_1|W_1)\big)+\frac{1}{2}\big(I(W_2;Y_2)+I(U_2;Y_2|W_2)\\&\qquad+I(V_2;Z_2|W_2)-I(U_2;V_2|W_2)\big)\ge\frac{1}{2}(d_1+d_1').\end{aligned}$$

Similarly,

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

is greater than or equal to $(d_2+d_2')/2$. Thus, the point $\big((d_1+d_1')/2,(d_2+d_2')/2\big)$ is in the region, and $\mathcal{D}$ is convex.

Next, note that the point $(R_{\text{sum}},R_{\text{sum}})\in\mathcal{D}$. We claim that it is a boundary point of $\mathcal{D}$. If it is an interior point, there must exist an $\epsilon>0$ such that $(R_{\text{sum}}+\epsilon,R_{\text{sum}}+\epsilon)$ is in $\mathcal{D}$. This implies the existence of some $p(u,v,w,x)$ where

$$\begin{aligned}R_{\text{sum}}+\epsilon &\le I(W;Y)+I(U;Y|W)+I(V;Z|W)-I(U;V|W),\\ R_{\text{sum}}+\epsilon &\le I(W;Z)+I(U;Y|W)+I(V;Z|W)-I(U;V|W).\end{aligned}$$

This implies that

$$\begin{aligned}R_{\text{sum}}+\epsilon \le{}&\min(I(W;Y),I(W;Z))+I(U;Y|W)\\&+I(V;Z|W)-I(U;V|W)\end{aligned}$$

for some $p(u,v,w,x)$, which is a contradiction.

Using the supporting hyperplane theorem (see [20, pp. 50-51]) and the fact that $\mathcal{D}$ is convex and closed, one can conclude that there exists a supporting hyperplane to $\mathcal{D}$ at the boundary point $(R_{\text{sum}},R_{\text{sum}})$. We claim that this supporting hyperplane must satisfy the equation $\lambda^*d_1+(1-\lambda^*)d_2=T(\lambda^*)$ for some $\lambda^*\in[0,1]$. The proof is as follows: any supporting hyperplane must satisfy $\lambda^*d_1+(1-\lambda^*)d_2=k$ for some real λ^* and real k. We claim that λ^* must be in $[0,1]$ and $k=T(\lambda^*)$. Assume that $\lambda^*<0$. We know that $\mathcal{D}$ must be entirely contained in one of the two closed half-spaces determined by the hyperplane. Note that the points $(0,0)$, $(-\infty,0)$, and $(0,-\infty)$ are all in $\mathcal{D}$ (take $p(u,v,w,x)$ satisfying $I(U;V|W)=0$ in the definition of $\mathcal{D}$). The value of $\lambda^*d_1+(1-\lambda^*)d_2$ at these points is equal to 0, $+\infty$ and $-\infty$, respectively. Thus, $\mathcal{D}$ cannot possibly be entirely contained in one of the two closed half-spaces determined by the hyperplane. The case $1-\lambda^*<0$ can be similarly refuted. Therefore λ^* must be in $[0,1]$. Since the points $(-\infty,0)$ and $(0,-\infty)$ are in $\mathcal{D}$, the half-space determined by the hyperplane that contains $\mathcal{D}$ is the one determined by the equation $\lambda^*d_1+(1-\lambda^*)d_2\le k$ for some k. Since the half-space has at least one point in $\mathcal{D}$, the value of k must be equal to $\max_{(d_1,d_2)\in\mathcal{R}}\lambda^*d_1+(1-\lambda^*)d_2$. The latter is equal to $T(\lambda^*)$. Thus, the supporting hyperplane at the boundary point $(R_{\text{sum}},R_{\text{sum}})$ satisfies the equation $\lambda^*d_1+(1-\lambda^*)d_2=T(\lambda^*)$ for some $\lambda^*\in[0,1]$.

Since $(R_{\text{sum}},R_{\text{sum}})$ lies on this hyperplane, $\lambda^*R_{\text{sum}}+(1-\lambda^*)R_{\text{sum}}=T(\lambda^*)$ implies that $R_{\text{sum}}=T(\lambda^*)$ for some $\lambda^*\in[0,1]$. Therefore

$$\min_{0\le\lambda\le1} T_\lambda \le R_{\text{sum}}.$$

On the other hand, for every λ, $T_\lambda\ge R_{\text{sum}}$. Therefore, $\min_{0\le\lambda\le1}T_\lambda\ge R_{\text{sum}}$. □

Proof of Theorem 1:

1) Note that $p(u,y)>0$ for all (u,y) because there must exist some x such that $p(u,x)>0$. Since the transition matrices have positive entries and $p(u,y)\ge p(u,x)q(y|x)$, $p(u,y)$ is positive for all y. A similar argument shows that $p(v,z)>0$ for all (v,z). Next assume that $p(u,v)=0$ for some (u,v). Take some u',v' such that $p(u',v')>0$. Let us reduce $p(u',v')$ by ϵ and increase $p(u,v)$ by ϵ. Furthermore, let us have (u,v) mapped to the same x that (u',v') is mapped to; this ensures that $p(x)$ is not changed. One can write

$$\begin{aligned}&I(U;Y)+I(V;Z)-I(U;V)\\&\quad=H(Y)+H(Z)+H(U,V)-H(U,Y)-H(V,Z).\end{aligned}$$

The only change in this expression comes from the change in $H(U,V)-H(U,Y)-H(V,Z)$. The derivative of $H(U,V)$ with respect to ϵ at $\epsilon=0$ is infinite. But the derivatives of $H(U,Y)$ and $H(V,Z)$ are finite since $p(u,y)$, $p(u',y)$, $p(v,z)$ and $p(v',z)$ are positive for all y and z. So, the first derivative of $H(U,V)-H(U,Y)-H(V,Z)$ with respect to ϵ at $\epsilon=0$ is positive. This is a contradiction since $p(u,v|x)$ is assumed to maximize $I(U;Y)+I(V;Z)-I(U;V)$.

2) Assume that $\mathcal{U}=\{u_1,u_2,\dots,u_{|\mathcal{U}|}\}$ and $\mathcal{V}=\{v_1,v_2,\dots,v_{|\mathcal{V}|}\}$. Let $\pi_{i,j}=p(u_i,v_j)$ for $i=1,\dots,|\mathcal{U}|$, $j=1,\dots,|\mathcal{V}|$. From the first part we know that $\pi_{i,j}>0$ for all i and j. Let $\overline{\epsilon}=\min_{i,j}\pi_{i,j}$.

Take some $\epsilon \in (0, \overline{\epsilon})$. Let $x = \xi_0(u, v)$ denote the deterministic mapping from $\mathcal{U} \times \mathcal{V}$ to $\mathcal{X}$.

We prove the statement by contradiction. Assume that $\overrightarrow{v_{\xi_0}} = \sum_{t=1}^{M} \alpha_t \overrightarrow{v_{\xi_t}}$, for some mappings ξ_t ($t = 1, 2, \ldots, M$) distinct from ξ_0 and non-negative numbers α_t adding up to one.

Let random variables $T_{i,j}$ (for $i = 1, \ldots, |\mathcal{U}|$, $j = 1, 2, 3, \ldots, |\mathcal{V}|$) be $(M+1)$-ary random variables mutually independent of each other and of U, V, X, Y, Z, satisfying:

- $\mathbb{P}(T_{i,j} = 0) = 1 - \frac{\epsilon}{\pi_{i,j}}$,
- $\mathbb{P}(T_{i,j} = 1) = \frac{\epsilon}{\pi_{i,j}} \alpha_1$,
- $\mathbb{P}(T_{i,j} = 2) = \frac{\epsilon}{\pi_{i,j}} \alpha_2$,
- $\mathbb{P}(T_{i,j} = 3) = \frac{\epsilon}{\pi_{i,j}} \alpha_3$,
- $\ldots$
- $\mathbb{P}(T_{i,j} = M) = \frac{\epsilon}{\pi_{i,j}} \alpha_M$.

Let $\widetilde{X}$ be defined as follows:

- On the event $\{(U, V) = (u_i, v_j)\}$, let $\widetilde{X}$ be equal to $\xi_{T_{i,j}}(u_i, v_j)$. In other words, if $T_{i,j} = 0$, $\widetilde{X} = \xi_0(u_i, v_j)$; if $T_{i,j} = 1$, $\widetilde{X} = \xi_1(u_i, v_j)$, etc.

We claim that $\mathbb{P}(\widetilde{X} = x | U = u_i) = \mathbb{P}(X = x | U = u_i)$ for all $i = 1, 2, 3, \ldots, |\mathcal{U}|$ and x; and similarly $\mathbb{P}(\widetilde{X} = x | V = v_j) = \mathbb{P}(X = x | V = v_j)$ for all $j = 1, 2, 3, \ldots, |\mathcal{V}|$ and x. This is proved in Appendix III. Note that the above property implies that $\widetilde{X}$ and X have the same marginal pmfs.

Let $\widetilde{Y}$ and $\widetilde{Z}$ be defined such that $U, V, (T_{i,j})_{i:1,2,\ldots,j=1,2,\ldots} \to \widetilde{X} \to \widetilde{Y}\widetilde{Z}$, and the conditional law of $(\widetilde{y}, \widetilde{z})$ given $\widetilde{x}$ is the same as $q(y, z|x)$. Here $(T_{i,j})_{i:1,2,..,j=1,2,..}$ denotes the collection of all $T_{i,j}$ for all i and j.

Without loss of generality, assume that $\alpha_1 \neq 0$. Since the mapping $\xi_0(\cdot, \cdot) = \xi_1(\cdot, \cdot)$, there must exist (i, j) such that $\xi_0(u_i, v_j) \neq \xi_1(u_i, v_j)$. Let us label the input symbol $\xi_0(u_i, v_j)$ by x_0, and the input symbol $\xi_1(u_i, v_j)$ by x_1 (the channel is irreducible). Let us then assume that there is some y such that $q(y|x_0) \neq q(y|x_1)$; the proof for the case when there is some z such that $q(z|x_0) \neq q(z|x_1)$ is similar. Let $\widetilde{U} = (U, T_{i,j})$ and $\widetilde{V} = V$.

Since $\mathbb{P}(\widetilde{X} = x | U = u) = \mathbb{P}(X = x | U = u)$ for all u and x, and $\mathbb{P}(\widetilde{X} = x | V = v) = \mathbb{P}(X = x | V = v)$ for all v and x, we have

- $I(U; \widetilde{Y}) = I(U; Y)$,
- $I(V; \widetilde{Z}) = I(V; Z)$.

Therefore $I(\widetilde{V}; \widetilde{Z}) = I(V; Z)$ and $I(\widetilde{U}; \widetilde{Y}) = I(U; Y) + I(T_{i,j}; \widetilde{Y}|U)$. Furthermore, since $T_{i,j}$ is independent of (U, V), we have $I(\widetilde{U}; \widetilde{V}) = I(U; V)$. Therefore

$$I(\widetilde{U}; \widetilde{Y}) + I(\widetilde{V}; \widetilde{Z}) - I(\widetilde{U}; \widetilde{V}) - \big(I(U; Y) + I(V; Z) - I(U; V)\big) = I(T_{i,j}; \widetilde{Y}|U).$$

Since $p(u, v, x)$ is maximizing $I(U; Y) + I(V; Z) - I(U; V)$ under the fixed $p(x)$, we must have $I(T_{i,j}; \widetilde{Y}|U) = 0$. Therefore $I(T_{i,j}; \widetilde{Y}|U = u_i) = 0$ holds as well.

In Appendix III, we prove the following:

$$\mathbb{P}(\widetilde{X} = x_0 | U = u_i, T_{i,j} = 0) \neq \mathbb{P}(\widetilde{X} = x_0 | U = u_i, T_{i,j} = 1),$$
$$\mathbb{P}(\widetilde{X} = x_1 | U = u_i, T_{i,j} = 0) \neq \mathbb{P}(\widetilde{X} = x_1 | U = u_i, T_{i,j} = 1),$$

but for any $x \notin \{x_0, x_1\}$,

$$\mathbb{P}(\widetilde{X} = x | U = u_i, T_{i,j} = 0) = \mathbb{P}(\widetilde{X} = x | U = u_i, T_{i,j} = 1).$$

Recall that we assumed there is some y such that $q(y|x_0) \neq q(y|x_1)$. In Appendix III, we show that

$$\mathbb{P}(\widetilde{Y} = y | U = u_i, T_{i,j} = 0) \neq \mathbb{P}(\widetilde{Y} = y | U = u_i, T_{i,j} = 1).$$

This implies that $\widetilde{Y}$ and $T_{i,j}$ are not conditionally independent given $U = u_i$. Therefore $I(T_{i,j}; \widetilde{Y}|U = u_i) \neq 0$, which is a contradiction.

3) The proof of this part begins by noting that the definition of $h(x)$ implies that for any (u, v, x),

$$h(x) \leq \log(p(u, v)) - f_u(x) - g_v(x).$$

Therefore, for any (u, v, x),

$$\log(p(u, v)) \geq f_u(x) + g_v(x) + h(x).$$

Thus,

$$\log(p(u, v)) \geq \max_x \big(f_u(x) + g_v(x) + h(x)\big). \quad (9)$$

Note that the first partial derivative of $H(U, V) - H(U, Y) - H(V, Z)$ with respect to $p(u, v, x)$ is proportional to

$$-\log p(u, v) - 1 + \sum_y q(y|x) \log p(u, y) + 1 + \sum_z q(z|x) \log p(v, z) + 1 = -\log p(u, v) + f_u(x) + g_v(x) + 1.$$

Assume that the triple (u, v, x) is such that $p(u, v, x) > 0$. Take some arbitrary u' and v'. Reducing $p(u, v, x)$ by an $\epsilon > 0$ and increasing $p(u', v', x)$ by the same ϵ does not affect $p(x)$, hence should not increase $H(U, V) - H(U, Y) - H(V, Z)$. After such a perturbation X is no longer a deterministic function of (U, V). Nevertheless $H(U, V) - H(U, Y) - H(V, Z)$ cannot increase. Therefore the first derivative of $H(U, V) - H(U, Y) - H(V, Z)$ with respect to $p(u, v, x)$ must be greater than or equal to the first derivative of $H(U, V) - H(U, Y) - H(V, Z)$ with respect to $p(u', v', x)$. Thus,

$$-\log p(u, v) + f_u(x) + g_v(x) + 1 \geq -\log p(u', v') + f_{u'}(x) + g_{v'}(x) + 1.$$

In other words, for any arbitrary u' and v', we have

$$\log p(u, v) - f_u(x) - g_v(x) \leq \log p(u', v') - f_{u'}(x) - g_{v'}(x).$$

Therefore

$$\log p(u,v) - f_u(x) - g_v(x) \leq \min_{u',v'} (\log p(u',v') - f_{u'}(x) - g_{v'}(x)) = h(x).$$

Thus, $\log p(u,v) \leq f_u(x) + g_v(x) + h(x)$ whenever $p(u,v,x) > 0$. This together with (9) imply that

$$\log(p(u,v)) = \max_x [f_u(x) + g_v(x) + h(x)],$$

and

$$p(x_0|u,v) = 1 \text{ for some } x_0 \in \mathcal{X} \Rightarrow x_0 \in argmax_x f_u(x) + g_v(x) + h(x).$$

Proof of Theorem 2: From the set of pmfs $p(u,v,w,x)$ that maximize the expression $\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$, let $p_0(u,v,w,x)$ be the one that achieves the largest value of $I(W;Y) + I(W;Z)$. In Appendix III, we show that one can find $p(\widehat{u},\widehat{v},\widehat{w},\widehat{x})$ such that

- $\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$ is equal to $\lambda I(\widehat{W};\widehat{Y}) + (1-\lambda) I(\widehat{W};\widehat{Z}) + I(\widehat{U};\widehat{Y}|\widehat{W}) + I(\widehat{V};\widehat{Z}|\widehat{W}) - I(\widehat{U};\widehat{V}|\widehat{W})$,
- $I(W;Y) + I(W;Z)$ is equal to $I(\widehat{W};\widehat{Y}) + I(\widehat{W};\widehat{Z})$,
- $|\widehat{\mathcal{U}}| \leq \min(|\mathcal{X}|, |\mathcal{Y}|)$,
- $|\widehat{\mathcal{V}}| \leq \min(|\mathcal{X}|, |\mathcal{Z}|)$,
- $|\widehat{\mathcal{W}}| \leq |\mathcal{X}|$,
- $H(\widehat{X}|\widehat{U},\widehat{V},\widehat{W}) = 0$.

Thus the constraints in the first and second parts are satisfied by $p(\widehat{u},\widehat{v},\widehat{w},\widehat{x})$. The second and third parts of Theorem 1 imply that $p(\widehat{u},\widehat{v},\widehat{w},\widehat{x})$ automatically satisfies the third and fourth part of Theorem 2. In Appendix IV, we show that the fifth part of Theorem 2 holds for any joint pmf that maximizes the expression $\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$, and at the same time has the largest possible value of $I(W;Y) + I(W;Z)$. Thus it must also hold for $p(\widehat{u},\widehat{v},\widehat{w},\widehat{x})$.

B. Insufficiency of Marton's Coding Scheme Without a Superposition Variable

Proof of Theorem 3: The direction $C_{bd}^r \subseteq C_{bd}^m$ is trivial, since the corner points of the time-division region are achievable by $U = X$, $V = \emptyset$ and $U = \emptyset$, $V = X$, respectively. Thus it remains to show that $C_{bd}^m \subseteq C_{bd}^r$ Consider a binary-input degraded broadcast channel in C_{bd}^m. The maximum of $R_1 + \lambda R_2$ (for $\lambda \geq 1$) over the region given by equations (6)-(8) is equal to $\max_{p(u,v,x)} I(U;Y) + \lambda I(V;Z) - I(U;V)$. We claim that this is equal to $\max(C_1, \lambda C_2)$, where $C_1 = \max_{p(x)} I(X;Y)$ and $C_2 = \max_{p(x)} I(X;Z)$. This would establish the first part of the claim since the maximum of $R_1 + \lambda R_2$ for $\lambda < 1$ (when the weight of the weaker receiver is smaller than the weight of the stronger receiver) is clearly $C_1 = \max(C_1, \lambda C_2)$.

To show that $\max_{p(u,v,x)} I(U;Y) + \lambda I(V;Z) - I(U;V) = \max(C_1, \lambda C_2)$ when $\lambda \geq 1$, let $p(u,v,x)$ be a maximizer for the expression $\max_{p(u,v,x)} I(U;Y) + \lambda I(V;Z) - I(U;V)$. Without loss of generality we can assume that $|\mathcal{U}| = |\mathcal{V}| = 2$ and that X is a function of (U,V). Since X is a function of (U,V) without loss of generality we can assume that either $X = U$, $X = V$, $X = U \oplus V$ (the XOR mapping) or $X = U \wedge V$ (the AND mapping). If $X = U$, then $I(U;Y) + \lambda I(V;Z) - I(U;V) \leq I(X;Y) \leq C_1$. Similarly when $X = V$, $I(U;Y) + \lambda I(V;Z) - I(U;V) \leq \lambda I(X;Z) \leq \lambda C_2$. The next case is when $X = U \oplus V$. But Remark 4 shows that this pattern cannot happen.

The only remaining case is when $X = U \wedge V$. Thus, $\mathbb{P}(X = 0) = \mathbb{P}(U = 0, V = 0) + \mathbb{P}(U = 1, V = 0) + \mathbb{P}(U = 0, V = 1)$. Let us define

$$\alpha_{ij} = \frac{\mathbb{P}(U = i, V = j)}{\mathbb{P}(X = 0)} \quad \text{for } (i,j) = (0,0), (0,1) \text{ and } (1,0).$$

Note that

$$\begin{aligned} I(U;Y) + \lambda I(V;Z) - I(U;V) &= I(U;Y|V) + \lambda I(V;Z) - I(U;V|Y) \\ &= \left[I(X;Y|V) + \lambda I(V;Z)\right] - I(U;V|Y). \end{aligned}$$

Since we are assuming that the inner bound given by equations (6)-(8) is equal to the true capacity region, it has to be the case that $I(U;V|Y) = 0$. Otherwise, the true capacity region attains a larger value for the maximum of $R_1 + \lambda R_2$. Thus $I(U;V|Y = y) = 0$ for all y (note that we had assumed that $p(y|x) > 0$ for all x, y and hence $p(y) > 0$ for all y). The joint pmf of (U,V) conditioned on $Y = y$ is as follows: $\mathbb{P}(U = i, V = j | Y = y) = \alpha_{ij} \mathbb{P}(X = 0 | Y = y)$ for $(i,j) = (0,0), (0,1)$ and $(1,0)$. Further, $\mathbb{P}(U = 1, V = 1 | Y = y) = \mathbb{P}(X = 1 | Y = y)$. Since (U,V) are conditionally independent, we have $\mathbb{P}(U = 0, V = 0 | Y = y)\mathbb{P}(U = 1, V = 1 | Y = y) = \mathbb{P}(U = 0, V = 1 | Y = y)\mathbb{P}(U = 1, V = 0 | Y = y)$. Thus,

$$\begin{aligned} &\alpha_{00}\mathbb{P}(X = 0|Y = y)\mathbb{P}(X = 1|Y = y) \\ &\quad = \alpha_{01}\alpha_{10}\mathbb{P}(X = 0|Y = y)^2. \end{aligned}$$

Since $p(y|x) > 0$ for all $(x,y) \in (\mathcal{X}, \mathcal{Y})$, $\mathbb{P}(X = 0|Y = y) > 0$. Therefore $\alpha_{00}\mathbb{P}(X = 1|Y = y) = \alpha_{01}\alpha_{10}\mathbb{P}(X = 0|Y = y)$ or in other words

$$\mathbb{P}(X = 0|Y = y) = \frac{\alpha_{00}}{\alpha_{01}\alpha_{10} + \alpha_{00}}.$$

Therefore $\mathbb{P}(X = 0|Y = y)$ has to be equal to the above value for all y. Therefore X is independent of Y. Thus, $I(U;Y) \leq I(X;Y) = 0$ and $I(V;Z) \leq I(V;Y) \leq I(X;Y) = 0$. But this is a contradiction. This completes the proof for the first part of the claim.

We now prove the second part of the claim, i.e., if the private message capacity region is the time-sharing region, it has to be the case that any $p(x)$ maximizing $I(X;Y)$ has to be also a maximizer for $I(X;Z)$. More generally if we denote $p_0 = \mathbb{P}(X = 0)$, then for any $p_0 \in [0,1]$, $I(X;Z)/C_2 \geq I(X;Y)/C_1$, where $C_1 = \max_{p(x)} I(X;Y)$ and $C_2 = \max_{p(x)} I(X;Z)$.

To show this, take some $p_0 \in [0,1]$ where $I(X;Z)/C_2 < I(X;Y)/C_1$. Let $\lambda = C_1/C_2$. Note that for this value of λ, the maximum of $R_1 + \lambda R_2$ is equal to $\max(C_1, \lambda C_2) = C_1 = \lambda C_2$. We want to show that $\max_{p(v,x)} [I(X;Y|V) + \lambda I(V;Z)] > C_1$,

and this is a contradiction. Note that

$$\begin{aligned}\max_{p(v,x)} & I(X;Y|V)+\lambda I(V;Z)\\ &= \max_{p(v,x)} \lambda I(X;Z)+[I(X;Y|V)-\lambda I(X;Z|V)]\\ &= \max_{p(x)} \lambda I(X;Z)+\max_{p(v|x)}[I(X;Y|V)-\lambda I(X;Z|V)]\\ &= \max_{p(x)} \lambda I(X;Z)+\mathfrak{C}[I(X;Y)-\lambda I(X;Z)],\end{aligned}$$

where $\mathfrak{C}$ is the upper concave envelope operator; the upper concave envelope of a function f, i.e., $\mathfrak{C}[f]$ is the smallest concave function that lies above f throughout the domain of f.

We claim that the upper concave envelope of the curve $\mathbb{P}(X=0)\mapsto \mathfrak{C}[I(X;Y)-\lambda I(X;Z)]$ is strictly positive throughout the interval $(0,1)$. To see this, observe that there is some $p_0\in[0,1]$ where $I(X;Y)>\lambda I(X;Z)$, thus the curve of $\mathbb{P}(X=0)\mapsto I(X;Y)-\lambda I(X;Z)$ is strictly positive at $p_0\in[0,1]$. Next note that the upper concave envelope of a curve is always greater than or equal to the curve itself; thus the upper concave envelope of $\mathbb{P}(X=0)\mapsto I(X;Y)-\lambda I(X;Z)$ is non-negative at 0 and 1 and strictly positive at p_0. Any concave function that is non-negative at 0 and 1, and strictly positive at p_0 is strictly positive throughout the interval $(0,1)$.

Consider the $p(x)$ that maximizes $\lambda I(X;Z)$. At this $p(x)$, we have $\mathfrak{C}[I(X;Y)-\lambda I(X;Z)]>0$. Therefore

$$\begin{aligned}\max_{p(v|x)} & I(X;Y|V)+\lambda I(V;Z)\\ &= \lambda I(X;Z)+\mathfrak{C}[I(X;Y)-\lambda I(X;Z)]\\ &= \lambda C_2+\mathfrak{C}[I(X;Y)-\lambda I(X;Z)]\\ &> \lambda C_2.\end{aligned}$$

Thus,

$$\max_{p(v,x)}[I(X;Y|V)+\lambda I(V;Z)]>\lambda C_2=C_1.$$

This contradiction completes the proof.

C. A Simple Direct Proof for Optimality of Superposition Coding Along Certain Directions

Proof of Theorem 4: We show that

$$\begin{aligned}\max_{(R_0,R_1,R_2)\in\mathcal{C}(q(y,z|x))} & (\lambda_0R_0+\lambda_1R_1+\lambda_2R_2)\\ &\le \max\{\max_{(R_0,R_2)\in\mathcal{C}_{d_1}(q(y,z|x))}(\lambda_0R_0+\lambda_2R_2),\\ &\qquad \max_{(R_0,R_1)\in\mathcal{C}_{d_2}(q(y,z|x))}(\lambda_0R_0+\lambda_1R_1)\}.\end{aligned}$$

Take an arbitrary code $(M_0,M_1,M_2,X^n,\epsilon)$. Assume without loss of generality that $H(M_2)\le H(M_1)$, i.e., $R_2\le R_1$. Let $\widehat{W}=M_0M_2$, $\widehat{X}=X^n$, $\widehat{Y}=Y^n$, $\widehat{Z}=Z^n$. Note that $q(\widehat{y},\widehat{z}|\widehat{x})$ is the n-fold version of $q(y,z|x)$. Let us look at $\mathcal{C}_{d_2}(q(\widehat{y},\widehat{z}|\widehat{x}))$, evaluated at the joint pmf $p(\widehat{w},\widehat{x})$:

$$\begin{aligned}\widehat{R_0}&\le I(\widehat{W};\widehat{Z}),\\ \widehat{R_1}&\le I(\widehat{X};\widehat{Y}|\widehat{W}),\\ \widehat{R_0}+\widehat{R_1}&\le I(\widehat{X};\widehat{Y}).\end{aligned}$$

Note that by Fano's inequality,

$$\begin{aligned}I(\widehat{W};\widehat{Z}) &= I(M_0,M_2;Z^n)\\ &= H(M_0)+H(M_2)-n\epsilon_n,\\ I(\widehat{X};\widehat{Y}|\widehat{W}) &= I(X^n;Y^n|M_0,M_2)\\ &= H(M_1)-n\epsilon_n,\\ I(\widehat{X};\widehat{Y}) &= I(X^n;Y^n)\\ &\ge I(M_0,M_1;Y^n)\\ &= H(M_0)+H(M_1)-H(M_0,M_1|Y^n)\\ &\ge H(M_0)+H(M_1)-n\epsilon_n\end{aligned}$$

for a sequence ϵ_n that tends to zero as n approaches infinity. Therefore, $\widehat{R_0}=H(M_0)+H(M_2)-n\epsilon_n=n(R_0+R_2)-n\epsilon_n$ and $\widehat{R_1}=H(M_1)-H(M_2)=n(R_1-R_2)-n\epsilon_n$ is in $\mathcal{C}_{d_2}(q(\widehat{y},\widehat{z}|\widehat{x}))$. Since $q(\widehat{y},\widehat{z}|\widehat{x})$ is the n-fold version of $q(y,z|x)$ and $\mathcal{C}_{d_2}(q(\widehat{y},\widehat{z}|\widehat{x}))$ is the degraded message set capacity region for $q(\widehat{y},\widehat{z}|\widehat{x})$, we must have: $\mathcal{C}_{d_2}(q(\widehat{y},\widehat{z}|\widehat{x}))=n\cdot\mathcal{C}_{d_2}(q(y,z|x))$, where the multiplication here is pointwise. Thus, $(\widehat{R_0}/n,\widehat{R_1}/n)\in\mathcal{C}_{d_2}(q(y,z|x))$. Letting $n\to\infty$ we conclude that $(R_0+R_2,R_1-R_2,0)\in\mathcal{C}_{d_2}(q(y,z|x))$. Furthermore $\lambda_0R_0+\lambda_1R_1+\lambda_2R_2\le\lambda_0(R_0+R_2)+\lambda_1(R_1-R_2)$ since $\lambda_0-\lambda_1\ge\lambda_2$. This completes the proof.

An Alternative Proof: We note that one can prove the theorem using a rate transfer argument to exchange between the common rate and the individual rates. In other words if (R_0,R_1,R_2) is in the capacity region of a broadcast channel, then $(R_0+\min\{R_1,R_2\},R_1-\min\{R_1,R_2\},R_2-\min\{R_1,R_2\})$ is also in the capacity region. Since $\lambda_0\ge\lambda_1+\lambda_2$, we have that $\lambda_0(R_0+\min\{R_1,R_2\})+\lambda_1(R_1-\min\{R_1,R_2\})+\lambda_2(R_2-\min\{R_1,R_2\})\ge\lambda_0R_0+\lambda_1R_1+\lambda_2R_2$. Thus if (R_0,R_1,R_2) maximizes $\lambda_0R_0+\lambda_1R_1+\lambda_2R_2$, so does $(R_0+\min\{R_1,R_2\},R_1-\min\{R_1,R_2\},R_2-\min\{R_1,R_2\})$. This completes the proof since either $R_1-\min\{R_1,R_2\}=0$ or $R_2-\min\{R_1,R_2\}=0$. The idea is to basically use $\lambda_0\ge\lambda_1+\lambda_2$ to transfer one of individual message rates completely to the common message rate. To do this, one requires a code with small *maximum* error probability, rather than one with small *average* probability of error. To show this one can apply a result of Willems [11] who shows that the *maximal* probability of error capacity region is equal to the *average* probability of error capacity region. Willems's proof of his result, however, is rather involved. The first proof is a simple direct argument based on the characterization of the capacity region of a degraded BC [9].

Appendix I

Suppose $p_0(u,v,w,x)$ is a joint pmf that maximizes $\lambda I(W;Y)+(1-\lambda)I(W;Z)+I(U;Y|W)+I(V;Z|W)-I(U;V|W)$, and among all such joint pmfs has the largest value of $I(W;Y)+I(W;Z)$. In this appendix, we show that there exists a pmf $p(\widehat{u},\widehat{v},\widehat{w},\widehat{x})$ such that

- $\lambda I(W;Y)+(1-\lambda)I(W;Z)+I(U;Y|W)+I(V;Z|W)-I(U;V|W)$ is equal to $\lambda I(\widehat{W};\widehat{Y})+(1-\lambda)I(\widehat{W};\widehat{Z})+I(\widehat{U};\widehat{Y}|\widehat{W})+I(\widehat{V};\widehat{Z}|\widehat{W})-I(\widehat{U};\widehat{V}|\widehat{W})$,
- $I(W;Y)+I(W;Z)$ is equal to $I(\widehat{W};\widehat{Y})+I(\widehat{W};\widehat{Z})$,
- $|\widehat{\mathcal{U}}|\le\min(|\mathcal{X}|,|\mathcal{Y}|)$,

- $|\widehat{\mathcal{V}}| \leq \min(|\mathcal{X}|, |\mathcal{Z}|)$,
- $|\widehat{\mathcal{W}}| \leq |\mathcal{X}|$,
- $H(\widehat{X}|\widehat{U}, \widehat{V}, \widehat{W}) = 0$.

We begin by reducing the cardinality of W. Assume that $|\mathcal{W}| > |\mathcal{X}|$ and $p(w) \neq 0$ for all w. Then, there must exist a non-zero function $L : \mathcal{W} \to \mathbb{R}$ where $\mathbb{E}[L(W)|X] = 0$. Let us perturb $p_0(u, v, w, x)$ along L as follows:

$$p_\epsilon(u, v, w, x, y, z) = p_0(u, v, w, x, y, z) \cdot [1 + \epsilon L(w)],$$

where ϵ is in some interval $[-\overline{\epsilon}_1, \overline{\epsilon}_2]$ where

$$\overline{\epsilon}_1 = \min_{w:L(w)>0} \frac{1}{|L(w)|}, \qquad \overline{\epsilon}_2 = \min_{w:L(w)<0} \frac{1}{|L(w)|}.$$

Observe that $p_\epsilon(u, v, x, y, z|w) = p_0(u, v, x, y, z|w)$, and we are only perturbing the marginal pmf of W. Further note that

$$p_\epsilon(x, y, z) = p_0(x, y, z) \cdot [1 + \epsilon \mathbb{E}[L(W)|X = x]],$$

and thus the constraint $\mathbb{E}[L(W)|X] = 0$ implies that the marginal pmf of (X, Y, Z) remains constant as we vary ϵ. We will use lemmas from [10] to compute derivatives of entropy expressions as a function of ϵ.

Consider the expression $\lambda I(W; Y) + (1 - \lambda)I(W; Z) + I(U; Y|W) + I(V; Z|W) - I(U; V|W)$ at $p_\epsilon(u, v, w, x, y, z)$. It can be verified that the expression is a linear function of ϵ under this perturbation.[3] Since a maximum of this expression occurs at $\epsilon = 0$, which is a point strictly inside the interval $[-\overline{\epsilon}_1, \overline{\epsilon}_2]$, it must be the case that this expression is a constant function of ϵ. Next consider the expression $I(W; Y) + I(W; Z)$ at $p_\epsilon(u, v, w, x, y, z)$. It can be verified that the expression is a linear function of ϵ under this perturbation.[4] Note that $p_0(u, v, w, x)$ is a joint pmf that has the largest value of $I(W; Y) + I(W; Z)$ among all joint pmfs that maximize $\lambda I(W; Y) + (1 - \lambda)I(W; Z) + I(U; Y|W) + I(V; Z|W) - I(U; V|W)$. Thus a maximum of $I(W; Y) + I(W; Z)$ occurs at $\epsilon = 0$, which is a point strictly inside the interval $[-\overline{\epsilon}_1, \overline{\epsilon}_2]$. But this can only happen when $I(W; Y) + I(W; Z)$ is a constant function of ϵ. Now, taking $\epsilon = -\overline{\epsilon}_1$ or $\epsilon = \overline{\epsilon}_2$ gives us a joint pmf with the same values of $\lambda I(W; Y) + (1-\lambda)I(W; Z) + I(U; Y|W) + I(V; Z|W) - I(U; V|W)$ and $I(W; Y) + I(W; Z)$, but with a smaller support on $\mathcal{W}$. Using this argument, one can reduce the cardinality of W to $|\mathcal{X}|$.

Next, we show how one can reduce the cardinality of U to find $p(\widehat{u}, \widehat{v}, \widehat{w}, \widehat{x})$ such that

- $\lambda I(W; Y) + (1-\lambda)I(W; Z) + I(U; Y|W) + I(V; Z|W) - I(U; V|W)$ is equal to $\lambda I(\widehat{W}; \widehat{Y}) + (1 - \lambda)I(\widehat{W}; \widehat{Z}) + I(\widehat{U}; \widehat{Y}|\widehat{W}) + I(\widehat{V}; \widehat{Z}|\widehat{W}) - I(\widehat{U}; \widehat{V}|\widehat{W})$,
- $I(W; Y) + I(W; Z)$ is equal to $I(\widehat{W}; \widehat{Y}) + I(\widehat{W}; \widehat{Z})$,
- $|\widehat{\mathcal{U}}| \leq \min(|\mathcal{X}|, |\mathcal{Y}|)$,
- $|\widehat{\mathcal{W}}| \leq |\mathcal{X}|$.

We can repeat a similar procedure to impose the constraint $|\widehat{\mathcal{V}}| \leq \min(|\mathcal{X}|, |\mathcal{Z}|)$. Imposing the extra constraint $H(\widehat{X}|\widehat{U}, \widehat{V}, \widehat{W}) = 0$ is discussed at the end.

[3]To see this, note that $I(W; Y) = H(Y) - H(Y|W)$. The term $H(Y)$ is fixed because the marginal pmf of Y is fixed. The term $H(Y|W) = \sum_w p_\epsilon(w)H(Y|W = w)$ is linear in ϵ since $H(Y|W = w)$ is invariant under the perturbation and $p_\epsilon(w)$ is linear in ϵ. Thus $I(W; Y)$ is linear in ϵ. All the other terms $I(U; Y|W)$, $I(V; Z|W)$, $I(U; V|W)$ that are conditioned on W are linear in ϵ for a similar reason.

[4]The reason is similar to that discussed in the previous footnote.

If $|\mathcal{X}| \leq |\mathcal{Y}|$, establishing the cardinality bound of $|\mathcal{X}|$ on $\mathcal{U}$ suffices. This cardinality bound is proved in Theorem 1 of [10] using perturbations of the type $L : \mathcal{U} \times \mathcal{W} \to \mathbb{R}$ where $\mathbb{E}[L(U, W)|W, X] = 0$. Note that these perturbations preserve the marginal pmf of $p(w, x)$, and thus also $I(W; Y) + I(W; Z)$. The interesting case is therefore when $|\mathcal{X}| > |\mathcal{Y}|$. Assume that $|\mathcal{U}| > |\mathcal{Y}|$. If for every $w \in \mathcal{W}$, $p(u|w) \neq 0$ for at most $|\mathcal{Y}|$ elements u, we are done, since we can relabel the elements in $\mathcal{U}$ to ensure that only an alphabet of size at most $|\mathcal{Y}|$ is used without affecting any of the mutual information terms in the expression of interest. Hence, there must exist a function $L : \mathcal{U} \times \mathcal{W} \to \mathbb{R}$, where

$$\mathbb{E}[L(U, W)|W, Y] = 0, \exists(u, w) : p_0(u, w) \neq 0,\ L(u, w) \neq 0.$$

Let us perturb $p_0(u, v, w, x)$ along the random variable $L : \mathcal{U} \times \mathcal{W} \to \mathbb{R}$. Random variables $\widetilde{U}, \widetilde{V}, \widetilde{W}, \widetilde{X}, \widetilde{Y}, \widetilde{Z}$ are distributed according to $p_\epsilon(\widetilde{u}, \widetilde{v}, \widetilde{w}, \widetilde{x}, \widetilde{y}, \widetilde{z})$ defined as follows:

$$p_\epsilon(\widetilde{u}, \widetilde{v}, \widetilde{w}, \widetilde{x}, \widetilde{y}, \widetilde{z}) = p_0(\widetilde{u}, \widetilde{v}, \widetilde{w}, \widetilde{x}, \widetilde{y}, \widetilde{z}) \cdot [1 + \epsilon L(\widetilde{u}, \widetilde{w})],$$

where $\epsilon \in [-\overline{\epsilon}_1, \overline{\epsilon}_2]$.

The first derivative of $\lambda I(W; Y) + (1 - \lambda)I(W; Z) + I(U; Y|W) + I(V; Z|W) - I(U; V|W)$ with respect to ϵ at $\epsilon = 0$ should be zero. Note that

$$\begin{aligned}
&\lambda I(W; Y) + (1 - \lambda)I(W; Z) + I(U; Y|W) \\
&\qquad + I(V; Z|W) - I(U; V|W) \\
&\quad = \lambda\big(H(W) + H(Y) - H(W, Y)\big) \\
&\qquad + (1 - \lambda)\big(H(W) + H(Z) - H(W, Z)\big) \\
&\qquad + H(Y, W) + H(Z, W) - H(U, Y, W) \\
&\qquad - H(V, Z, W) + H(U, V, W) - H(W).
\end{aligned}$$

We can compute the first derivative of this expression using part one of Lemma 2 of [10] and set it to zero:

$$\begin{aligned}
&\lambda\big(H_L(W) + H_L(Y) - H_L(W, Y)\big) \\
&\qquad + (1 - \lambda)\big(H_L(W) + H_L(Z) - H_L(W, Z)\big) \\
&\qquad + H_L(Y, W) + H_L(Z, W) - H_L(U, Y, W) \\
&\qquad - H_L(V, Z, W) + H_L(U, V, W) - H_L(W) = 0,
\end{aligned}$$

where $H_L(W)$ denotes $\sum_w \mathbb{E}[L|W = w]p(w)\log(1/p(w))$ and similarly for the other terms. Using part two of Lemma 2 of [10], we have:

$$\begin{aligned}
&\lambda I(\widetilde{W}; \widetilde{Y}) + (1 - \lambda)I(\widetilde{W}; \widetilde{Z}) + I(\widetilde{U}; \widetilde{Y}|\widetilde{W}) \\
&\quad + I(\widetilde{V}; \widetilde{Z}|\widetilde{W}) - I(\widetilde{U}; \widetilde{V}|\widetilde{W}) \\
&\quad = \lambda I(W; Y) + (1 - \lambda)I(W; Z) + I(U; Y|W) \\
&\qquad + I(V; Z|W) - I(U; V|W) \\
&\qquad + \lambda\big(-\mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|W]\big)\big] - \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|Y]\big)\big] \\
&\qquad\quad + \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|WY]\big)\big]\big) + (1-\lambda)\big(-\mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|W]\big)\big] \\
&\qquad\quad - \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|Z]\big)\big] + \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|WZ]\big)\big]\big) \\
&\qquad - \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|YW]\big)\big] - \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|ZW]\big)\big] \\
&\qquad + \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|UYW]\big)\big] + \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|VWZ]\big)\big] \\
&\qquad - \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|UVW]\big)\big] + \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|W]\big)\big],
\end{aligned}$$

where $r(x) = (1+x)\log(1+x)$. Since $\mathbb{E}[L(U,W)|WY] = 0$ and L is a function of (U,W), we have:

$$\begin{aligned}
&\lambda I(\widetilde{W};\widetilde{Y}) + (1-\lambda) I(\widetilde{W};\widetilde{Z}) + I(\widetilde{U};\widetilde{Y}|\widetilde{W}) \\
&\quad + I(\widetilde{V};\widetilde{Z}|\widetilde{W}) - I(\widetilde{U};\widetilde{V}|\widetilde{W}) \\
&= \lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) \\
&\quad + I(V;Z|W) - I(U;V|W) \\
&\quad + (1-\lambda)\big(- \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|Z]\big)\big] + \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|WZ]\big)\big]\big) \\
&\quad - \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|ZW]\big)\big] + \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|VWZ]\big)\big].
\end{aligned}$$

To see this observe that $\mathbb{E}[L|WY] = 0$ implies $\mathbb{E}[L|W] = \mathbb{E}[L|Y] = 0$ so the terms $\mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|W]\big)\big]$, $\mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|Y]\big)\big]$ and $\mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|WY]\big)\big]$ vanish. Since L is a function of (U,W) we have

$$\begin{aligned}
\mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|UYW]\big)\big] &= \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|UW]\big)\big] \\
&= \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|UVW]\big)\big],
\end{aligned}$$

so these terms cancel out each other. Since $r(x) = (1+x)\log(1+x)$ is a convex function, we can use Jensen's inequality to obtain

$$\begin{aligned}
\mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|WZ]\big)\big] &\geq \mathbb{E}_Z\big[r\big(\mathbb{E}_W\big[\epsilon \cdot \mathbb{E}[L|WZ]\big]\big)\big] \\
&= \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|Z]\big)\big].
\end{aligned}$$

Thus,

$$\begin{aligned}
-\mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|Z]\big)\big] + \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|WZ]\big)\big] \geq 0, \\
-\mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|WZ]\big)\big] + \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|VWZ]\big)\big] \geq 0.
\end{aligned}$$

Therefore for any $\epsilon \in [-\overline{\epsilon}_1, \overline{\epsilon}_2]$, we have

$$\begin{aligned}
&\lambda I(\widetilde{W};\widetilde{Y}) + (1-\lambda) I(\widetilde{W};\widetilde{Z}) + I(\widetilde{U};\widetilde{Y}|\widetilde{W}) \\
&\quad + I(\widetilde{V};\widetilde{Z}|\widetilde{W}) - I(\widetilde{U};\widetilde{V}|\widetilde{W}) \\
&\geq \lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) \\
&\quad + I(V;Z|W) - I(U;V|W).
\end{aligned}$$

This implies that $\lambda I(\widetilde{W};\widetilde{Y}) + (1-\lambda) I(\widetilde{W};\widetilde{Z}) + I(\widetilde{U};\widetilde{Y}|\widetilde{W}) + I(\widetilde{V};\widetilde{Z}|\widetilde{W}) - I(\widetilde{U};\widetilde{V}|\widetilde{W})$ is a constant function of ϵ. The maximum of $I(\widetilde{W};\widetilde{Y}) + I(\widetilde{W};\widetilde{Z})$ as a function of ϵ occurs at $\epsilon = 0$. Therefore $I_L(W;Y) + I_L(W;Z) = 0$, where

$$I_L(W;Y) = \sum_{u,w,y} p(u,w,y) L(u,w) \log \frac{p(w,y)}{p(w)p(y)},$$

and similarly for other terms (see [10, Lemma 2]).

Using Lemma 2 of [10] again, one can observe that

$$\begin{aligned}
&[I(\widetilde{W};\widetilde{Y}) + I(\widetilde{W};\widetilde{Z})] - [I(W;Y) + I(W;Z)] \\
&\quad = -\mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|Z]\big)\big] + \mathbb{E}\big[r\big(\epsilon \cdot \mathbb{E}[L|WZ]\big)\big] \geq 0.
\end{aligned}$$

But this can only happen if $I(\widetilde{W};\widetilde{Y}) + I(\widetilde{W};\widetilde{Z})$ is a constant function of ϵ. Now, taking $\epsilon = -\overline{\epsilon}_1$ or $\epsilon = \overline{\epsilon}_2$ gives us an auxiliary random variable pair $(\widetilde{U},\widetilde{W})$ with smaller support than that of (U,W). We can continue this process as long as there exists $w \in \mathcal{W}$ such that $p(u|w) \neq 0$ for more than $|\mathcal{Y}|$ elements u.

It remains to show that one can impose the extra constraint $H(\widehat{X}|\widehat{U},\widehat{V},\widehat{W}) = 0$. Fix $p(u,v,w)$. Consider the expressions $\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$ and $I(W;Y) + I(W;Z)$ as functions of the conditional pmf of X given (U,V,W). Denote it by $r(x|u,v,w)$. We know that for instance the former expression is maximized at $p(x|u,v,w)$. Further, the extreme points of the corresponding region for $r(x|u,v,w)$ satisfy $r(x|u,v,w) \in \{0,1\}$. Both expressions are convex functions of $r(x|u,v,w)$ because $I(W;Y)$ is convex in the conditional pmf $r(y|w)$; similarly $I(U;Y|W=w)$ is convex for any fixed value of w. The term $I(U;V|W)$ that appears with a negative sign is constant since the joint pmf $p(u,v,w)$ is fixed.

We can express $p(x|u,v,w)$ as a linear combination of the extreme points of the region formed by all conditional pmfs $r(x|u,v,w)$. Since the maximum of $\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$ occurs at some $p(x|u,v,w)$ and the expression is convex in $r(x|u,v,w)$, the maximum must also occur at all the extreme points showing up in the linear combination. One can use the convexity of $I(W;Y) + I(W;Z)$ in $r(x|u,v,w)$ to show that the value of $I(W;Y) + I(W;Z)$ at all these extreme points must be also equal to that at $p(x|u,v,w)$.

Appendix II

In this Appendix we close a gap in the proof of Theorem 1 by proving that $\mathbb{P}(\widetilde{X} = x|U = u_i) = \mathbb{P}(X = x|U = u_i)$ for all $i = 1,2,3,\ldots,|\mathcal{U}|$ and x, and similarly $\mathbb{P}(\widetilde{X} = x|V = v_j) = \mathbb{P}(X = x|V = v_j)$ for all $j = 1,2,3,\ldots,|\mathcal{V}|$ and x.

Note that $\mathbb{P}(\widetilde{X} = x|U = u_i)$ is equal to

$$\begin{aligned}
&\sum_j \mathbb{P}(V = v_j|U = u_i)\mathbb{P}(\widetilde{X} = x|U = u_i, V = v_j) \\
&\quad = \sum_j \mathbb{P}(V = v_j|U = u_i) \sum_{k=0}^{M} \mathbb{P}(T_{i,j} = k)\mathbf{1}[\xi_k(u_i,v_j) = x] \\
&\quad = \sum_j \mathbb{P}(V = v_j|U = u_i)\Big(1 - \frac{\epsilon}{\pi_{i,j}}\Big)\mathbf{1}[\xi_0(u_i,v_j) = x] \\
&\qquad + \sum_j \mathbb{P}(V = v_j|U = u_i) \sum_{k=1}^{M} \frac{\epsilon}{\pi_{i,j}} \alpha_k \mathbf{1}[\xi_k(u_i,v_j) = x] \\
&\quad = \sum_j \mathbb{P}(V = v_j|U = u_i)\Big(\frac{\pi_{i,j} - \epsilon}{\pi_{i,j}}\Big)\mathbf{1}[\xi_0(u_i,v_j) = x] \\
&\qquad + \sum_{k=1}^{M} \sum_j \mathbb{P}(V = v_j|U = u_i) \frac{\epsilon}{\pi_{i,j}} \alpha_k \mathbf{1}[\xi_k(u_i,v_j) = x].
\end{aligned}$$

Also, note that

$$\mathbb{P}(V = v_j|U = u_i) = \frac{\mathbb{P}(V = v_j, U = u_i)}{\mathbb{P}(U = u_i)} = \frac{\pi_{i,j}}{\mathbb{P}(U = u_i)}.$$

Therefore,

$$\begin{aligned}
\mathbb{P}(\widetilde{X} = x|U = u_i) &= \sum_j \frac{\pi_{i,j} - \epsilon}{\mathbb{P}(U = u_i)} \mathbf{1}[\xi_0(u_i,v_j) = x] \\
&\quad + \sum_{k=1}^{M} \sum_j \frac{\epsilon}{\mathbb{P}(U = u_i)} \alpha_k \mathbf{1}[\xi_k(u_i,v_j) = x]
\end{aligned}$$

$$= \sum_j \frac{\pi_{i,j}}{\mathbb{P}(U=u_i)} \mathbf{1}[\xi_0(u_i, v_j)=x] - \frac{\epsilon}{\mathbb{P}(U=u_i)} \sum_j \mathbf{1}[\xi_0(u_i, v_j)=x] + \frac{\epsilon}{\mathbb{P}(U=u_i)} \sum_{k=1}^{M} \alpha_k \sum_j \mathbf{1}[\xi_k(u_i, v_j)=x].$$

But since $\overrightarrow{v_{\xi_0}} = \sum_{t=1}^{M} \alpha_t \overrightarrow{v_{\xi_t}}$, the profiles of the row i's must also satisfy the same property

$$\sum_j \mathbf{1}[\xi_0(u_i, v_j) = x] = \sum_{k=1}^{M} \alpha_k \sum_j \mathbf{1}[\xi_k(u_i, v_j) = x].$$

Therefore, $\mathbb{P}(\widetilde{X} = x|U = u_i)$ is equal to

$$\sum_j \frac{\pi_{i,j}}{\mathbb{P}(U = u_i)} \mathbf{1}[\xi_0(u_i, v_j) = x] + 0 - 0 = \sum_j \frac{\pi_{i,j}}{\mathbb{P}(U = u_i)} \mathbf{1}[\xi_0(u_i, v_j) = x] = \mathbb{P}(X = x|U = u_i).$$

The equation $\mathbb{P}(\widetilde{X} = x|V = v_j) = \mathbb{P}(X = x|V = v_j)$ for all $j = 1, 2, 3, \ldots, |\mathcal{V}|$ and x can be proved similarly.

Appendix III

Note that

$$\begin{aligned}\mathbb{P}(\widetilde{X} = x_0|U = u_i, T_{i,j} = 0) &= \mathbb{P}(\widetilde{X} = x_0|U = u_i, T_{i,j} = 0, V = v_j) \times \mathbb{P}(V = v_j|U = u_i, T_{i,j} = 0) \\ &\quad + \mathbb{P}(\widetilde{X} = x_0|U = u_i, T_{i,j} = 0, V \neq v_j) \times \mathbb{P}(V \neq v_j|U = u_i, T_{i,j} = 0).\end{aligned}$$

Since under the event $\{(U, V) = (u_i, v_j)\}$ and $T_{i,j} = 0$, $\widetilde{X} = x_0$, the term $\mathbb{P}(\widetilde{X} = x_0|U = u_i, T_{i,j} = 0, V = v_j) = 1$. Since (U, V) is independent of $T_{i,j}$, we have

$$\mathbb{P}(V = v_j|U = u_i, T_{i,j} = 0) = \mathbb{P}(V = v_j|U = u_i),$$
$$\mathbb{P}(V \neq v_j|U = u_i, T_{i,j} = 0) = \mathbb{P}(V \neq v_j|U = u_i).$$

Lastly $\mathbb{P}(\widetilde{X} = x_0|U = u_i, T_{i,j} = 0, V \neq v_j)$ is equal to $\mathbb{P}(\widetilde{X} = x_0|U = u_i, V \neq v_j)$ since under the event $\{(U = u_i, V \neq v_j)\}$, $\widetilde{X}$ will be independent of $T_{i,j}$ (note that the random variables $T_{\cdot,\cdot}$ are mutually independent of each other). Therefore,

$$\begin{aligned}\mathbb{P}(\widetilde{X} = x_0|U = u_i, T_{i,j} = 0) &= \mathbb{P}(\widetilde{X} = x_0|U = u_i, V \neq v_j)\mathbb{P}(V \neq v_j|U = u_i) \\ &\quad + \mathbb{P}(V = v_j|U = u_i). \qquad (10)\end{aligned}$$

Next, note that

$$\begin{aligned}\mathbb{P}(\widetilde{X} = x_0|U = u_i, T_{i,j} = 1) &= \mathbb{P}(\widetilde{X} = x_0|U = u_i, T_{i,j} = 1, V = v_j) \times \mathbb{P}(V = v_j|U = u_i, T_{i,j} = 1) \\ &\quad + \mathbb{P}(\widetilde{X} = x_0|U = u_i, T_{i,j} = 1, V \neq v_j) \times \mathbb{P}(V \neq v_j|U = u_i, T_{i,j} = 1).\end{aligned}$$

Since under the event $\{(U, V) = (u_i, v_j)\}$ and $T_{i,j} = 1$, $\widetilde{X}$ is equal to x_1, the term $\mathbb{P}(\widetilde{X} = x_0|U = u_i, T_{i,j} = 1, V = v_j) = 0$. Following an argument like above, one can show that

$$\begin{aligned}&\mathbb{P}(\widetilde{X} = x_0|U = u_i, T_{i,j} = 1) \\ &\quad = 0 + \mathbb{P}(\widetilde{X} = x_0|U = u_i, V \neq v_j)\mathbb{P}(V \neq v_j|U = u_i). \qquad (11)\end{aligned}$$

Comparing equations (10) and (11), and noting that $\mathbb{P}(V = v_j|U = u_i) > 0$, we conclude that

$$\mathbb{P}(\widetilde{X} = x_0|U = u_i, T_{i,j} = 0) \neq \mathbb{P}(\widetilde{X} = x_0|U = u_i, T_{i,j} = 1).$$

The proof for

$$\mathbb{P}(\widetilde{X} = x_1|U = u_i, T_{i,j} = 0) \neq \mathbb{P}(\widetilde{X} = x_1|U = u_i, T_{i,j} = 1)$$

is similar.

It remains to show that for any $x \notin \{x_0, x_1\}$,

$$\mathbb{P}(\widetilde{X} = x|U = u_i, T_{i,j} = 0) = \mathbb{P}(\widetilde{X} = x|U = u_i, T_{i,j} = 1).$$

Observe that

$$\begin{aligned}\mathbb{P}(\widetilde{X} = x|U = u_i, T_{i,j} = 1) &= \mathbb{P}(\widetilde{X} = x|U = u_i, T_{i,j} = 1, V = v_j) \times \mathbb{P}(V = v_j|U = u_i, T_{i,j} = 1) \\ &\quad + \mathbb{P}(\widetilde{X} = x|U = u_i, T_{i,j} = 1, V \neq v_j) \times \mathbb{P}(V \neq v_j|U = u_i, T_{i,j} = 1) \\ &= 0 + \mathbb{P}(\widetilde{X} = x|U = u_i, V \neq v_j)\mathbb{P}(V \neq v_j|U = u_i) \\ &= \mathbb{P}(\widetilde{X} = x|U = u_i, T_{i,j} = 0).\end{aligned}$$

Appendix IV

We prove the statement by contradiction. Assume that

$$\mathbb{P}(\widetilde{Y} = y|U = u_i, T_{i,j} = 0) = \mathbb{P}(\widetilde{Y} = y|U = u_i, T_{i,j} = 1).$$

We have that $\mathbb{P}(\widetilde{Y} = y|U = u_i, T_{i,j} = 0)$ is equal to

$$\begin{aligned}&\mathbb{P}(\widetilde{Y}=y|U=u_i, T_{i,j}=0, \widetilde{X}=x_0)\mathbb{P}(\widetilde{X}=x_0|U=u_i, T_{i,j}=0) \\ &\quad + \mathbb{P}(\widetilde{Y}=y|U=u_i, T_{i,j}=0, \widetilde{X}=x_1) \\ &\quad \times \mathbb{P}(\widetilde{X}=x_1|U=u_i, T_{i,j}=0) \\ &\quad + \sum_{x\in\mathcal{X}, x\notin\{x_0,x_1\}} \big(\mathbb{P}(\widetilde{Y}=y|U=u_i, T_{i,j}=0, \widetilde{X}=x) \\ &\quad \times \mathbb{P}(\widetilde{X}=x|U=u_i, T_{i,j}=0)\big) \\ &= \mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x_0)\mathbb{P}(\widetilde{X}=x_0|U=u_i, T_{i,j}=0) \\ &\quad + \mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x_1)\mathbb{P}(\widetilde{X}=x_1|U=u_i, T_{i,j}=0) \\ &\quad \sum_{x\in\mathcal{X}, x\notin\{x_0,x_1\}} \big(\mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x)\mathbb{P}(\widetilde{X}=x|U=u_i, T_{i,j}=0)\big).\end{aligned}$$

Similarly, $\mathbb{P}(\widetilde{Y} = y|U = u_i, T_{i,j} = 1)$ is equal to

$$\begin{aligned}&\mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x_0)\mathbb{P}(\widetilde{X}=x_0|U=u_i, T_{i,j}=1) \\ &\quad + \mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x_1)\mathbb{P}(\widetilde{X}=x_1|U=u_i, T_{i,j}=1) \\ &\quad + \sum_{x\in\mathcal{X}, x\notin\{x_0,x_1\}} \big(\mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x)\mathbb{P}(\widetilde{X}=x|U=u_i, T_{i,j}=1)\big).\end{aligned}$$

It was shown in Appendix III that

$$\mathbb{P}(\widetilde{X}=x_0|U=u_i, T_{i,j}=0) \neq \mathbb{P}(\widetilde{X}=x_0|U=u_i, T_{i,j}=1),$$
$$\mathbb{P}(\widetilde{X}=x_1|U=u_i, T_{i,j}=0) \neq \mathbb{P}(\widetilde{X}=x_1|U=u_i, T_{i,j}=1).$$

However, for any $x \notin \{x_0, x_1\}$,

$$\mathbb{P}(\widetilde{X}=x|U=u_i, T_{i,j}=0)=\mathbb{P}(\widetilde{X}=x|U=u_i, T_{i,j}=1). \quad (12)$$

Thus, we must have

$$\begin{aligned}\mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x_0)&\mathbb{P}(\widetilde{X}=x_0|U=u_i, T_{i,j}=0)\\ &+\mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x_1)\mathbb{P}(\widetilde{X}=x_1|U=u_i, T_{i,j}=0)\\ =\mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x_0)&\mathbb{P}(\widetilde{X}=x_0|U=u_i, T_{i,j}=1)\\ &+\mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x_1)\mathbb{P}(\widetilde{X}=x_1|U=u_i, T_{i,j}=1).\end{aligned}$$

This implies that

$$\begin{aligned}&\frac{\mathbb{P}(\widetilde{X}=x_0|U=u_i, T_{i,j}=1)-\mathbb{P}(\widetilde{X}=x_0|U=u_i, T_{i,j}=0)}{\mathbb{P}(\widetilde{X}=x_1|U=u_i, T_{i,j}=0)-\mathbb{P}(\widetilde{X}=x_1|U=u_i, T_{i,j}=1)}\\ &\quad=\frac{\mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x_1)}{\mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x_0)}.\end{aligned}$$

Note that the numerator and denominator are negative by what was proved in Appendix III.

On the other hand, we also have by equation (12):

$$\begin{aligned}1-&\sum_{x:x\notin\{x_0,x_1\}}\mathbb{P}(\widetilde{X}=x|U=u_i, T_{i,j}=0)\\ &=1-\sum_{x:x\notin\{x_0,x_1\}}\mathbb{P}(\widetilde{X}=x|U=u_i, T_{i,j}=1).\end{aligned}$$

Thus,

$$\begin{aligned}&\mathbb{P}(\widetilde{X}=x_0|U=u_i, T_{i,j}=0)+\mathbb{P}(\widetilde{X}=x_1|U=u_i, T_{i,j}=0)\\ &=\mathbb{P}(\widetilde{X}=x_0|U=u_i, T_{i,j}=1)+\mathbb{P}(\widetilde{X}=x_1|U=u_i, T_{i,j}=1).\end{aligned}$$

This implies that

$$\frac{\mathbb{P}(\widetilde{X}=x_0|U=u_i, T_{i,j}=1)-\mathbb{P}(\widetilde{X}=x_0|U=u_i, T_{i,j}=0)}{\mathbb{P}(\widetilde{X}=x_1|U=u_i, T_{i,j}=0)-\mathbb{P}(\widetilde{X}=x_1|U=u_i, T_{i,j}=1)}$$

is equal to one. Hence,

$$\frac{\mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x_1)}{\mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x_0)}=1.$$

But we know that $\mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x_0)\neq\mathbb{P}(\widetilde{Y}=y|\widetilde{X}=x_1)$ since the input values x_0 and x_1 are distinguishable by the Y receiver, which is a contradiction.

APPENDIX V

The proof follows from the following two Lemmas.

Lemma 2. Assume that $p^*(u,v,w,x)$ is an arbitrary pmf that maximizes $\lambda I(W;Y)+(1-\lambda)I(W;Z)+I(U;Y|W)+I(V;Z|W)-I(U;V|W)$ and achieves the largest value of $I(W;Y)+I(W;Z)$ among all maximizing joint pmfs. For every w, $p^*(x|w)$ must belong to the set $\mathcal{T}(q(y,z|x))$ defined as follows. Let $\mathcal{T}(q(y,z|x))$ be the set of pmfs on $\mathcal{X}$, $t(x)$, such that

$$\begin{aligned}\max_{p(\widehat{u},\widehat{v},\widehat{w}|\widehat{x})t(\widehat{x})q(\widehat{y},\widehat{z}|\widehat{x})}&\big\{\lambda I(\widehat{W};\widehat{Y})+(1-\lambda)I(\widehat{W};\widehat{Z})\\ &+I(\widehat{U};\widehat{Y}|\widehat{W})+I(\widehat{V};\widehat{Z}|\widehat{W})-I(\widehat{U};\widehat{V}|\widehat{W})\big\}\\ =\max_{p(u,v|x)t(x)q(y,z|x)}&(I(U;Y)+I(V;Z)-I(U;V)),\end{aligned}$$

and $I(\widehat{W};\widehat{Y})=I(\widehat{W};\widehat{Z})=0$ for any pmf $p(\widehat{u},\widehat{v},\widehat{w}|\widehat{x})t(\widehat{x})$ that maximizes the expression $\lambda I(\widehat{W};\widehat{Y})+(1-\lambda)I(\widehat{W};\widehat{Z})+I(\widehat{U};\widehat{Y}|\widehat{W})+I(\widehat{V};\widehat{Z}|\widehat{W})-I(\widehat{U};\widehat{V}|\widehat{W})$. Please note that the random variables $\widehat{U},\widehat{V},\widehat{W}$ used in the definition of $\mathcal{T}(q(y,z|x))$ have nothing to do with U,V,W; their alphabets may be different. However, the random variables $\widehat{X},\widehat{Y},\widehat{Z}$ take values from the same sets as X,Y,Z.

Remark 9. Note that a pmf $p(\widehat{u},\widehat{v},\widehat{w}|\widehat{x})t(\widehat{x})$ that maximizes the expression $\lambda I(\widehat{W};\widehat{Y})+(1-\lambda)I(\widehat{W};\widehat{Z})+I(\widehat{U};\widehat{Y}|\widehat{W})+I(\widehat{V};\widehat{Z}|\widehat{W})-I(\widehat{U};\widehat{V}|\widehat{W})$ may not unique. Also we have used maximum and not supremum since cardinality bounds on the auxiliary random variables exist [10].

Lemma 3. Let $q(y,z|x)$ be a general broadcast channel and $t(x)\in\mathcal{T}(q(y,z|x))$. Consider the maximization problem: $\max_{p(u,v|x)t(x)q(y,z|x)}(I(U;Y)+I(V;Z)-I(U;V))$. Assume that a maximum occurs at $p^*(u,v|x)$. Then the following holds for random variables $(U,V,X,Y,Z)\sim p^*(u,v|x)t(x)q(y,z|x)$:

- *$I(\overline{U};Y)\geq I(\overline{U};V,Z)$ for every $\overline{U}\to U\to VXYZ$;*
- *$I(\overline{V};Z)\geq I(\overline{V};U,Y)$ for every $\overline{V}\to V\to UXYZ$.*

A. Proof of Lemma 2

Assume that the marginal pmf of X given $W=w$ does not belong to $\mathcal{T}$ for some w. By the definition then, at least one of the following must hold:

Case 1: Corresponding to $p^*_{X|W=w}(x|w)$ is the conditional pmf $p(\widehat{u},\widehat{v},\widehat{w}|\widehat{x})$ such that

$$\begin{aligned}I(U;Y|W=w)+I(V;Z|W=w)-I(U;V|W=w)&\\ <\lambda I(\widehat{W};\widehat{Y})+(1-\lambda)I(\widehat{W};\widehat{Z})+I(\widehat{U};\widehat{Y}|\widehat{W})&\\ +I(\widehat{V};\widehat{Z}|\widehat{W})-I(\widehat{U};\widehat{V}|\widehat{W}),&\quad(13)\end{aligned}$$

where $p(\widehat{u},\widehat{v},\widehat{w},\widehat{x},\widehat{y},\widehat{z})=p(\widehat{u},\widehat{v},\widehat{w}|\widehat{x})p^*_{X|W=w}(\widehat{x})q(\widehat{y},\widehat{z}|\widehat{x})$.

Case 2: Corresponding to $p^*_{X|W=w}(x|w)$ is the conditional pmf $p(\widehat{u},\widehat{v},\widehat{w}|\widehat{x})$ such that

$$\begin{aligned}I(U;Y|W=w)+I(V;Z|W=w)-I(U;V|W=w)&\\ =\lambda I(\widehat{W};\widehat{Y})+(1-\lambda)I(\widehat{W};\widehat{Z})+I(\widehat{U};\widehat{Y}|\widehat{W})&\\ +I(\widehat{V};\widehat{Z}|\widehat{W})-I(\widehat{U};\widehat{V}|\widehat{W}),&\end{aligned}$$

but $I(\widehat{W};\widehat{Y})+I(\widehat{W};\widehat{Z})>0$, where $p(\widehat{u},\widehat{v},\widehat{w},\widehat{x},\widehat{y},\widehat{z})=p(\widehat{u},\widehat{v},\widehat{w}|\widehat{x})p^*_{X|W=w}(\widehat{x})q(\widehat{y},\widehat{z}|\widehat{x})$.

Define $\widetilde{U},\widetilde{V},\widetilde{W}$ jointly distributed with (U,V,W,X,Y,Z) as follows: whenever $W\neq w$, the random variables $\widetilde{U}=U$, $\widetilde{V}=V$, $\widetilde{W}=W$. For $W=w$, the Markov chain $\widetilde{U}\widetilde{V}\widetilde{W}\to X\to U,V,W,Y,Z$ holds, and $p(\widetilde{u},\widetilde{v},\widetilde{w}|x)=p(\widehat{u},\widehat{v},\widehat{w}|\widehat{x})$. Next, assume that $U'=\widetilde{U}$, $V'=\widetilde{V}$, $W'=W\widetilde{W}$.

If case 1 holds, we prove that $\lambda I(W';Y)+(1-\lambda)I(W';Z)+I(U';Y|W')+I(V';Z|W')-I(U';V'|W')>\lambda I(W;Y)+(1-\lambda)I(W;Z)+I(U;Y|W)+I(V;Z|W)-I(U;V|W)$, which results in a contradiction. If case 2 holds, we prove that $\lambda I(W';Y)+(1-\lambda)I(W';Z)+I(U';Y|W')+I(V';Z|W')-I(U';V'|W')=\lambda I(W;Y)+(1-\lambda)I(W;Z)+I(U;Y|W)+I(V;Z|W)-I(U;V|W)$ but that $I(W';Y)+I(W';Z)>I(W;Y)+I(W;Z)$, which results in a contradiction.

Assume that case 1 holds. Since $W'=W\widetilde{W}$, $I(W';Y)=I(W;Y)+I(\widetilde{W};Y|W)$ and $I(W';Z)=I(W;Z)+$

$I(\widetilde{W}; Z|W)$. We need to show that

$$\begin{aligned}&\lambda I(\widetilde{W}; Y|W) + (1-\lambda) I(\widetilde{W}; Z|W) + I(\widetilde{U}; Y|W, \widetilde{W}) \\ &\quad + I(\widetilde{V}; Z|W, \widetilde{W}) - I(\widetilde{U}; \widetilde{V}|W, \widetilde{W}) \\ &\quad > I(U; Y|W) + I(V; Z|W) - I(U; V|W).\end{aligned}$$

Recall that whenever $W \neq w$, the random variables $\widetilde{U}$, $\widetilde{V}$, and $\widetilde{W}$ are defined to be equal to U, V, and W, respectively. Therefore we need to show that

$$\begin{aligned}&\mathbb{P}(W=w)\big[\lambda I(\widetilde{W}; Y|W=w) + (1-\lambda) I(\widetilde{W}; Z|W=w) \\ &\qquad + I(\widetilde{U}; Y|W=w, \widetilde{W}) I(\widetilde{V}; Z|W=w, \widetilde{W}) \\ &\qquad - I(\widetilde{U}; \widetilde{V}|W=w, \widetilde{W})\big] > \\ &\mathbb{P}(W=w)\big[I(U; Y|W=w) + I(V; Z|W=w) \\ &\qquad - I(U; V|W=w)\big].\end{aligned}$$

On the event $\{W=w\}$, the random variables $\widetilde{U}$, $\widetilde{V}$, $\widetilde{W}$ are defined so that $p(\widetilde{u}, \widetilde{v}, \widetilde{w}|x) = p(\widehat{u}, \widehat{v}, \widehat{w}|\widehat{x})$. Furthermore the marginal pmf of $\widehat{X}$ is $p^*(x|W=w)$. Therefore, $I(\widetilde{W}; Y|W=w) = I(\widehat{W}; \widehat{Y})$, $I(\widetilde{W}; Z|W=w) = I(\widehat{W}; \widehat{Z})$, $I(\widetilde{U}; Y|W=w, \widetilde{W}) = I(\widehat{U}; \widehat{Y}|\widehat{W})$, etc. Thus it remains to show that

$$\begin{aligned}&\lambda I(\widehat{W}; \widehat{Y}) + (1-\lambda) I(\widehat{W}; \widehat{Z}) + I(\widehat{U}; \widehat{Y}|\widehat{W}) \\ &\quad + I(\widehat{V}; \widehat{Z}|\widehat{W}) - I(\widehat{U}; \widehat{V}|\widehat{W}) \\ &> I(U; Y|W=w) + I(V; Z|W=w) - I(U; V|W=w).\end{aligned}$$

This holds because of equation (13). This concludes the proof for case 1.

Now, assume that case 2 holds. Following the above proof for case 1, we obtain

$$\begin{aligned}&\lambda I(W'; Y) + (1-\lambda) I(W'; Z) + I(U'; Y|W') \\ &\qquad + I(V'; Z|W') - I(U'; V'|W') \\ &\quad \geq \lambda I(W; Y) + (1-\lambda) I(W; Z) + I(U; Y|W) \\ &\qquad + I(V; Z|W) - I(U; V|W).\end{aligned}$$

Note that $I(W'; Y) + I(W'; Z) = I(W; Y) + I(\widetilde{W}; Y|W) + I(W; Z) + I(\widetilde{W}; Z|W)$. Thus, we need to show that $I(\widetilde{W}; Y|W) + I(\widetilde{W}; Z|W) > 0$. Note that

$$\begin{aligned}&I(\widetilde{W}; Y|W) + I(\widetilde{W}; Z|W) \\ &\quad = \mathbb{P}(W=w)\big(I(\widetilde{W}; Y|W=w) + I(\widetilde{W}; Z|W=w)\big) \\ &\quad = \mathbb{P}(W=w)\big(I(\widehat{W}; \widehat{Y}) + I(\widehat{W}; \widehat{Z})\big) > 0.\end{aligned}$$

B. Proof of Lemma 3

Take an arbitrary $\overline{U}$ satisfying $\overline{U} \to U \to (V, X, Y, Z)$. Let $\widehat{W} = \overline{U}$, $\widehat{U} = U$, $\widehat{V} = V$. Since $t(x) \in \mathcal{T}(q(y, z|x))$, and $p^*(u, v|x)$ maximizes $I(U; Y) + I(V; Z) - I(U; V)$, we can write:

$$\begin{aligned}&I(U; Y) + I(V; Z) - I(U; V) \\ &\quad \geq \lambda I(\widehat{W}; Y) + (1-\lambda) I(\widehat{W}; Z) + I(\widehat{U}; Y|\widehat{W}) \\ &\qquad + I(\widehat{V}; Z|\widehat{W}) - I(\widehat{U}; \widehat{V}|\widehat{W}),\end{aligned} \tag{14}$$

and furthermore if equality holds, we must have $I(\widehat{W}; Y) = I(\widehat{W}; Z) = 0$. We prove that this implies that $I(\overline{U}; Y) \geq I(\overline{U}; V, Z)$.

We can write:

$$\begin{aligned}&I(U; Y) + I(V; Z) - I(U; V) \\ &\quad \geq \lambda I(\widehat{W}; Y) + (1-\lambda) I(\widehat{W}; Z) + I(\widehat{U}; Y|\widehat{W}) \\ &\qquad + I(\widehat{V}; Z|\widehat{W}) - I(\widehat{U}; \widehat{V}|\widehat{W}) \\ &\quad = \lambda I(\overline{U}; Y) + (1-\lambda) I(\overline{U}; Z) + I(U; Y|\overline{U}) \\ &\qquad + I(V; Z|\overline{U}) - I(U; V|\overline{U}).\end{aligned}$$

Since $\overline{U} \to U \to VXYZ$, we have $I(U; Y) = I(\overline{U}U; Y)$ and $I(U; V) = I(\overline{U}U; V)$. This implies that

$$\begin{aligned}&I(\overline{U}; Y) + I(V; Z) - I(\overline{U}; V) \\ &\qquad \geq \lambda I(\overline{U}; Y) + (1-\lambda) I(\overline{U}; Z) + I(V; Z|\overline{U}),\end{aligned}$$

or

$$I(\overline{U}; Y) + I(V; Z) \geq \lambda I(\overline{U}; Y) + (1-\lambda) I(\overline{U}; Z) + I(V; Z, \overline{U}),$$

or

$$(1-\lambda) I(\overline{U}; Y) \geq (1-\lambda) I(\overline{U}; Z) + I(V; \overline{U}|Z).$$

In other words

$$(1-\lambda) I(\overline{U}; Y) \geq (1-\lambda) I(\overline{U}; V, Z) + \lambda I(V; \overline{U}|Z). \tag{15}$$

Let us consider the following two cases:

- $\lambda < 1$: In this case, equation (15) implies that
$$I(\overline{U}; Y) \geq I(\overline{U}; V, Z) + \frac{\lambda}{1-\lambda} I(V; \overline{U}|Z).$$
This inequality implies the desired inequality $I(\overline{U}; Y) \geq I(\overline{U}; V, Z)$.
- $\lambda = 1$: In this case, equation (15) implies that $I(V; \overline{U}|Z) = 0$. Furthermore equation (14) holds with equality. Since $t(x) \in \mathcal{T}$, we must have $I(\overline{U}; Y) = I(\overline{U}; Z) = 0$. The fact that $I(V; \overline{U}|Z) = I(\overline{U}; Y) = I(\overline{U}; Z) = 0$ implies that $I(\overline{U}; Y) = I(\overline{U}; Z, V) = 0$. Therefore the inequality $I(\overline{U}; Y) \geq I(\overline{U}; Z, V)$ also holds in this case.

In each case, we are done. The inequality $I(\overline{V}; Z) \geq I(\overline{V}; Y, U)$ can be proved similarly.

ACKNOWLEDGMENT

The authors would like to thank anonymous reviewers for helpful comments.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.
[2] T. M. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 2–14, Jan. 1972.
[3] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[4] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306–311, May 1979.
[5] A. Gohari, A. E. Gamal, and V. Anantharam, "On an outer bound and an inner bound for the general broadcast channel," in *Proc. IEEE Int. Symp. Inform. Theory*, Austin, TX, USA, Jun. 2010, pp. 540–544.
[6] S. I. Gelfand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," *Problems Inform. Trans.*, vol. 16, no. 1, pp. 17–25, 1980.

[7] C. Nair and V. W. Zizhou, "On the inner and outer bounds for 2-receiver discrete memoryless broadcast channels," in *Proc. ITA Workshop*, San Diego, CA, USA, Feb. 2008, pp. 226–229.
[8] Y. Liang, G. Kramer, and H. V. Poor, "On the equivalence of two achievable regions for the broadcast channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 95–100, Jan. 2011.
[9] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 23, no. 1, pp. 60–64, Jan. 1977.
[10] A. A. Gohari and V. Anantharam, "Evaluation of Marton's inner bound for the general broadcast channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 608–619, Feb. 2012.
[11] F. M. J. Willems, "The maximal-error and average-error capacity region of the broadcast channel are identical: A direct proof," *Problems Control Inform. Theory*, vol. 19, no. 4, pp. 339–347, 1990.
[12] F. M. J. Willems and E. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 313–327, May 1985.
[13] V. Jog and C. Nair, "An information inequality for the BSSC channel," in *Proc. ITA Workshop*, San Diego, CA, USA, Jan./Feb. 2010, pp. 1–8.
[14] Y. Geng, A. Gohari, C. Nair, and Y. Yu, "On Marton's inner bound for two receiver broadcast channels," in *Proc. ITA Workshop*, San Diego, CA, USA, 2011.
[15] Y. Geng, A. Gohari, C. Nair, and Y. Yu, "The capacity region for two classes of product broadcast channels," in *Proc. IEEE ISIT*, St. Petersburg, Russia, Jul./Aug. 2011, pp. 1549–1553.
[16] C. Nair, Z. V. Wang, and Y. Geng, "An information inequality and evaluation of Marton's inner bound for binary input broadcast channels," in *Proc. IEEE ISIT*, Austin, TX, USA, Jun. 2010, pp. 550–554.
[17] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.
[18] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inf. Theory*, vol. 25, no. 2, pp. 166–169, Mar. 1979.
[19] C. Nair and A. El Gamal, "An outer bound to the capacity region of the broadcast channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 350–355, Jan. 2007.
[20] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[21] V. Anantharam, A. Gohari, and C. Nair, "Improved cardinality bounds on the auxiliary random variables in Marton's inner bound," in *Proc. IEEE ISIT*, Istanbul, Turkey, Jul. 2013, pp. 1272–1276.
[22] Y. Geng and C. Nair, "The capacity region of the two-receiver vector Gaussian broadcast channel with private and common messages," in *Proc. IEEE ISIT*, Cambridge, MA, USA, Jul. 2012, pp. 586–590.
[23] A. Gohari, C. Nair, and V. Anantharam, "On Marton's inner bound for broadcast channels," in *Proc. IEEE ISIT*, Cambridge, MA, USA, Jul. 2012, pp. 581–585.
[24] Y. Geng, A. Gohari, C. Nair, and Y. Yu, "The capacity region of classes of product broadcast channels," in *Proc. IEEE ISIT*, St. Petersburg, Russia, Jul./Aug. 2011, pp. 1549–1553.

Amin Gohari received his M.Sc. and Ph.D. degree in electrical engineering in 2010 from the University of California, Berkeley, and his B.Sc. degree in 2004 from Sharif University of Technology, Iran. He is an Assistant Professor at Sharif University of Technology, Tehran, Iran.

Dr. Gohari received the 2010 Eli Jury Award from UC Berkeley, Department of Electrical Engineering and the 2009–2010 Bernard Friedman Memorial Prize in Applied Mathematics from UC Berkeley, Department of Mathematics. He also received the Gold Medal from the 41st International Mathematical Olympiad (IMO 2000) and the First Prize from the 9th International Mathematical Competition for University Students (IMC 2002).

Abbas El Gamal received the B.Sc. (honors) degree in electrical engineering from Cairo University in 1972 and the M.S. degree in statistics and the Ph.D. degree in electrical engineering from Stanford University in 1977 and 1978, respectively. From 1978 to 1980, he was an Assistant Professor in the Department of Electrical Engineering at the University of Southern California (USC). He has been on the Stanford faculty since 1981, where he is currently the Hitachi America Professor in the School of Engineering and Chair of the Department of Electrical Engineering. He was director of the Information Systems laboratory from 2004 to 2009. His current research interests and contributions have spanned the areas of information theory, wireless networks, imaging sensors and systems, and integrated circuit design and design automation. He has authored or coauthored over 200 papers and 30 patents in these areas. He is coauthor of the book Network Information Theory (Cambridge Press 2011). He has won several honors and awards, including the 2012 Claude E. Shannon Award, the 2009 Padovani Lecture, and the 2004 Infocom best paper award. He is a member of the National Academy of Engineering. He has been active in several IEEE societies, including serving on the Board on Governors of the IT society where he is currently President.

Venkat Anantharam is on the faculty of the EECS Department at the University of California at Berkeley. He received the Philips India Medal and the President of India Gold Medal from IIT Madras in 1980 and an NSF Presidential Young Investigator award in 1988. He is a co-recipient of the 1998 Prize Paper Award of the IEEE Information Theory Society, and a co-recipient of the 2000 Stephen O. Rice Prize Paper Award of the IEEE Communications Theory Society. He received the Distinguished Alumnus Award from IIT Madras in 2008.