On Marton's inner bound and its optimality for classes of product broadcast channels*

Yanlin Geng, Amin Gohari, Chandra Nair, Yuanming Yu

Abstract

Marton's inner bound is the tightest known inner bound on the capacity region of the broadcast channel. It is not known, however, if this bound is tight in general. One approach to settle this key open problem in network information theory is to investigate the multi-letter extension of Marton's bound, which is known to be tight in general. This approach has become feasible only recently through the development of a new method for bounding cardinalities of auxiliary random variables by Gohari and Anantharam. This paper undertakes this long overdue approach to establish several new results, including (i) establishing the optimality of Marton's bound for new classes of product broadcast channels, (ii) showing that the best known outer bound by Nair and El Gamal is not tight in general, and (iii) finding sufficient conditions for a global maximizer of Marton's bound that imply that the 2-letter extension does not increase the achievable rate. Motivated by the new capacity results, we establish a new outer bound on the capacity region of product broadcast channels in general.

1 Introduction

Consider the broadcast channel $\mathfrak{q}(y, z|x)$ with private messages depicted in Figure 1.1. The sender X wishes to communicate a message M_1 at rate R_1 to receiver Y_1 and a message M_2 at rate R_2 to another receiver Y_2. What is the capacity region, that is, the closure of the set of achievable rate pairs (R_1, R_2)?

This question is one of the key open problems in network information theory. Since the introduction of this problem in the groundbreaking paper by Cover [1], several inner and outer bounds on the capacity region of this channel have been developed and shown to be tight in some special cases; see Chapters 5, 8, and 9 of [2] for a detailed discussion of previous works.

Marton's inner bound [3] and the UV outer bound [4](also sometimes referred to as the Nair–El Gamal outer bound) are the tightest known bounds on the capacity region of the broadcast channel. These bounds have been shown to coincide for all classes of broadcast channels with known capacity regions. Recently it has been shown [5, 6, 7] that there are channels for which these inner and outer bounds do not coincide. Therefore, clearly at least one of them is strictly sub-optimal.

In this paper we show that the UV outer bound is strictly suboptimal by establishing the capacity region for a new class of broadcast channels and showing that this capacity region coincides with Marton's inner bound but not with Nair–El Gamal's outer bound. This result is only one consequence of exploring an approach to establish the optimality (or lack thereof) of Marton's region by investigating its multi-letter extension. This approach, although conceptually simple, has only become interesting recently. This is due to the fact that cardinality bounds on the auxiliary random variables in Marton's inner bound were recently established in [6]; and only since then did Marton's inner bound become computable and hence amenable to numerical simulations for test channels.

*This paper contains the proofs of the results that were in part presented at ITA, UCSD and at ISIT, 2011.

1.1 Preliminaries

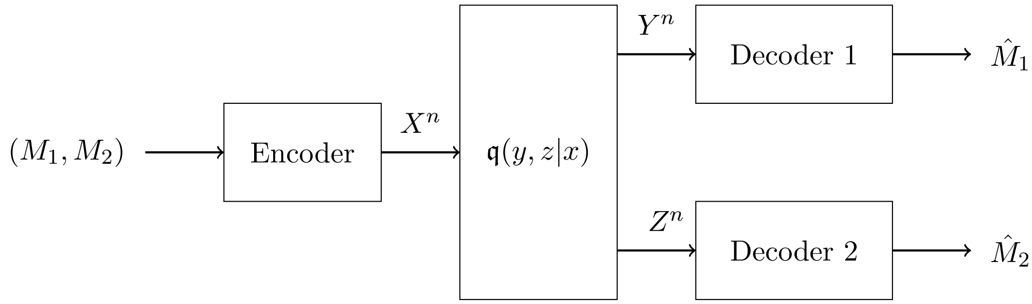

Figure 1: A broadcast channel

In the broadcast channel setting a sender X, who has messages M_1, M_2, wishes to communicate message M_1 to receiver Y and M_2 to receiver Z over a noisy discrete memoryless broadcast channel $\mathfrak{q}(y, z|x)$. A set of rate pairs (R_1, R_2) is said to be achievable for this broadcast channel, $\mathfrak{q}(y, z|x)$, if there is a sequence of codebooks, each consisting of:

- an encoder at the sender that maps the message pair (M_1, M_2) into a sequence X^n,
- a decoder at receiver Y that maps the received sequence Y^n into an estimate $\hat{M}_1$ of its intended message M_1, and
- a decoder at receiver Z that maps the received sequence Z^n into an estimate $\hat{M}_2$ of its intended message M_2

such that $\mathrm{P}(\hat{M}_1 \neq M_1), \mathrm{P}(\hat{M}_2 \neq M_2) \to 0$ as $n \to \infty$, when the messages M_1, M_2 are uniformly distributed in $[1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$. The capacity region is the closure of the set of all achievable rate pairs. An evaluable characterization of this capacity region is a well known open problem.

An inner bound to the capacity region refers to a set of rate pairs for which there is a strategy to achieve it. The best known inner bound to the capacity region of the two receiver broadcast channel is due to Marton [3]. It is not known if Marton's inner bound is optimal or not. Marton's inner bound for a general two-receiver discrete-memoryless broadcast channel with private messages is the following:

Inner bound: (Marton [3]) The union of rate pairs (R_1, R_2) satisfying the inequalities

$$\begin{aligned} R_1 &\leq I(U, W; Y) \\ R_2 &\leq I(V, W; Z) \\ R_1 + R_2 &\leq \min\{I(W; Y), I(W; Z)\} + I(U; Y|W) + I(V; Z|W) - I(U; V|W) \end{aligned} \tag{1}$$

over all $(U, V, W, X) : (U, V, W) \to X \to (Y, Z)$ forms a Markov chain constitutes an inner bound to the capacity region. Further to compute this region it suffices[6] to consider $|\mathcal{U}|, |\mathcal{V}| \leq |\mathcal{X}|, |\mathcal{W}| \leq |\mathcal{X}| + 4$.

One of the main results of this paper is computing the capacity region for new classes of product broadcast channels, and deducing that the best outer bound previously known is strictly sub-optimal. The best outer bound[1] for a general two-receiver discrete-memoryless broadcast channel with private messages is the following:

[1] Though there have been several proposed outer bounds since [4], it was shown in [8] that they reduced to the one in [4] for the private messages case.

Outer bound: (UV outer bound [4]) The union of rate pairs (R_1, R_2) satisfying the inequalities

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

over all $(U, V, X) : (U, V) \to X \to (Y, Z)$ forms a Markov chain constitutes an outer bound to the capacity region. Further to compute this region it suffices to consider $|\mathcal{U}|, |\mathcal{V}| \leq |\mathcal{X}| + 1$.

1.1.1 Definitions of some classes of broadcast channels

Definition 1. A broadcast channel $\mathfrak{q}(y, z|x)$ is said to be a *product broadcast channel* if $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2), \mathcal{Y} = (\mathcal{Y}_1, \mathcal{Y}_2), \mathcal{Z} = (\mathcal{Z}_1, \mathcal{Z}_2)$ and $\mathfrak{q}(y_1, y_2, z_1, z_2|x_1, x_2) = \mathfrak{q}_1(y_1, z_1|x_1)\mathfrak{q}_2(y_2, z_2|x_2)$. Here we denote $\mathfrak{q} = \mathfrak{q}_1 \times \mathfrak{q}_2$.

Definition 2. A product broadcast channel $\mathfrak{q} = \mathfrak{q}_1 \times \mathfrak{q}_2$ is said to be *reversely semi-deterministic* if the channel to one of the receivers in the first component is deterministic, and the channel to the other receiver in the second component is deterministic. That is either both $\mathfrak{q}_1(y_1|x_1), \mathfrak{q}_2(z_2|x_2) \in \{0, 1\}$ or both $\mathfrak{q}_1(z_1|x_1), \mathfrak{q}_2(y_2|x_2) \in \{0, 1\}$.

Definition 3. A product broadcast channel $\mathfrak{q} = \mathfrak{q}_1 \times \mathfrak{q}_2$ is said to be *reversely more capable* if one of the following two holds:

- $I(X_1; Y_1) \geq I(X_1; Z_1),\ \forall p(x_1)$, and $I(X_2; Z_2) \geq I(X_2; Y_2),\ \forall p(x_2)$,
- $I(X_1; Z_1) \geq I(X_1; Y_1),\ \forall p(x_1)$, and $I(X_2; Y_2) \geq I(X_2; Z_2),\ \forall p(x_2)$.

1.2 Organization and summary of results

The rest of the paper is organized as follows. In Section 2 we show that the UV outer bound is not tight (Claim 3). To do so, we need to introduce a quantity called the λ-sum-rate and use some of its properties to compare the inner and outer bounds. The quantity λ-sum-rate is by itself interesting and in Section 4.2 we will establish some additional properties of it. In Section 3 we establish a new outer bound (Claim 4) for product broadcast channels and use it to determine the capacity region of some new classes (Theorems 2 and 3). This outer bound is strictly better than the UV outer bound as it is optimal for the example where the UV outer bound is loose. In Section 4 we present a collection of results and observations that are interesting to note, and could potentially be very useful in guiding towards the optimality (or not) of Marton's inner bound. In particular we show (Theorem 4) that Marton's optimality can be determined by verifying whether a particular distribution of random variables forms a local maximum for a 2-letter version of the channel. We also show (Lemma 9) that Marton's region for a product of two non-identical broadcast channels can be strictly larger than the (Minkowski) sum of the individual regions. Section 5 deals with a particular coding strategy that is equivalent to Marton's inner bound (albeit much simpler) for binary input broadcast channels. If the generalization of the strategy described in Section 5 is indeed equivalent to Marton's inner bound for products of binary input broadcast channels, then one can deduce the optimality of Marton's coding scheme for binary input broadcast channels from a result (Theorem 5) we establish in this section. Finally, some of the technical arguments as well as other lengthy but routine arguments is relegated to the Appendices.

2 The UV outer bound is not tight

The flow of this section is as follows: we first introduce λ-*sum-rate*, a quantity that helps in the computation of Marton's inner bound. To explicitly compute the sum rate for product channels we introduce the notion of *factorization of* λ-*sum-rate*. Using this factorization idea, we show that Marton's sum rate is optimal for the product of reversely semi-deterministic channels. Having computed the optimal sum rate, the UV outer bound is shown to be strictly suboptimal (via a specifically constructed example) over this class of broadcast channels.

2.1 Preliminaries

Given a broadcast channel $\mathfrak{q}(y,z|x)$ we define the following quantities for $\lambda \in [0,1]$ and for auxiliary random variables (U,V,W) that satisfy the Markov chain $(U,V,W) \to X \to (Y,Z)$:

$$\lambda\text{-}SR_M(\mathfrak{q}, p(u,v,w,x)) := \lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W) \tag{2}$$

$$\lambda\text{-}SR_M(\mathfrak{q}, p(x)) := \max_{p(u,v,w|x)} \lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W) \tag{3}$$

$$\lambda\text{-}SR_M(\mathfrak{q}) := \max_{p(u,v,w,x)} \lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W) \tag{4}$$

Note the following relations:

$$\lambda\text{-}SR_M(\mathfrak{q}, p(x)) = \max_{p(u,v,w|x)} \lambda\text{-}SR_M(\mathfrak{q}, p(u,v,w,x)), \text{ and } \lambda\text{-}SR_M(\mathfrak{q}) = \max_{p(x)} \lambda\text{-}SR_M(\mathfrak{q}, p(x)).$$

Further one can verify that $\lambda\text{-}SR_M(\mathfrak{q}, p(x))$ is concave in $p(x)$ for a fixed λ.

Note that the maximum sum rate yielded by Marton's inner bound in (1) is given by

$$SR_M(\mathfrak{q}) := \max_{p(u,v,w,x)} \min\{I(W;Y), I(W;Z)\} + I(U;Y|W) + I(V;Z|W) - I(U;V|W).$$

Hence $SR_M(\mathfrak{q}) = \max_{p(u,v,w,x)} \min_{\lambda\in[0,1]} \lambda\text{-}SR_M(\mathfrak{q}, p(u,v,w,x))$.

The following lemma allows us to shift the discussion from Marton's sum rate to λ-sum-rate, and then return to Marton's sum rate at a later point to complete our arguments.

Lemma 1. *The following min-max theorem holds:*

$$\begin{aligned}\max_{p(u,v,w,x)} \min_{\lambda\in[0,1]} \lambda\text{-}SR_M(\mathfrak{q}, p(u,v,w,x)) &= \max_{p(x)} \min_{\lambda\in[0,1]} \max_{p(u,v,w|x)} \lambda\text{-}SR_M(\mathfrak{q}, p(u,v,w,x)) \\ &= \min_{\lambda\in[0,1]} \max_{p(u,v,w,x)} \lambda\text{-}SR_M(\mathfrak{q}, p(u,v,w,x)).\end{aligned}$$

This implies that the sum rate of Marton's inner bound can be calculated using any of the three above expressions.

Proof. The proof is presented in Appendix A and can be considered as an application of a min-max theorem of Terkelsen[10]. In particular Corollary 4 established in the Appendix can also be used in other instances where a max-min occurs, such as compound channels. The fact that

$$SR_M(\mathfrak{q}) = \max_{p(u,v,w,x)} \min_{\lambda\in[0,1]} \lambda\text{-}SR_M(\mathfrak{q}, p(u,v,w,x)) = \min_{\lambda\in[0,1]} \max_{p(u,v,w,x)} \lambda\text{-}SR_M(\mathfrak{q}, p(u,v,w,x))$$

was established in section 3.1.1 of [9]. □

Definition 4. For a given product channel $\mathfrak{q}_1(y_1,z_1|x_1) \times \mathfrak{q}_2(y_2,z_2|x_2)$ we say that the λ-*sum-rate factorizes* if for all $p(x_1,x_2)$ we have

$$\lambda\text{-}SR_M(\mathfrak{q}_1 \times \mathfrak{q}_2, p_{X_1,X_2}(x_1,x_2)) \le \lambda\text{-}SR_M(\mathfrak{q}_1, p_{X_1}(x_1)) + \lambda\text{-}SR_M(\mathfrak{q}_2, p_{X_2}(x_2)). \tag{5}$$

Sufficient conditions for factorization of λ-sum-rate

In this section, we derive sufficient conditions under which (5) holds. The following claim is key to the arguments in this section.

Claim 1. *Let $U_1 = U_2 = U, V_1 = V_2 = V, W_1 = (W, Z_2), W_2 = (W, Y_1)$. Then the following holds:*

$$\begin{aligned}
&\lambda\text{-}SR_M(\mathfrak{q}_1 \times \mathfrak{q}_2, p(u, v, w, x_1, x_2)) \\
&\quad = \lambda\text{-}SR_M(\mathfrak{q}_1, p(u_1, v_1, w_1, x_1)) + \lambda\text{-}SR_M(\mathfrak{q}_2, p(u_2, v_2, w_2, x_2)) + I(U; V|W, Y_1, Z_2) \\
&\qquad - \lambda I(Y_1; Y_2) - (1 - \lambda)I(Z_1; Z_2) - I(Y_1; Z_2|U, V, W).
\end{aligned}$$

Proof.

$$\begin{aligned}
&\lambda\text{-}SR_M(\mathfrak{q}_1 \times \mathfrak{q}_2, p(u, v, w, x_1, x_2)) \\
&\quad = \lambda I(W; Y_1, Y_2) + (1 - \lambda)I(W; Z_1, Z_2) + I(U; Y_1, Y_2|W) + I(V; Z_1, Z_2|W) - I(U; V|W) \\
&\quad = \lambda I(W, Z_2; Y_1) + (1 - \lambda)I(W, Z_2; Z_1) + I(U; Y_1|W, Z_2) + I(V; Z_1|W, Z_2) - I(U; V|W, Z_2) \\
&\qquad + \lambda I(W, Y_1; Y_2) + (1 - \lambda)I(W, Y_1; Z_2) + I(U; Y_2|W, Y_1) + I(V; Z_2|W, Y_1) - I(U; V|W, Y_1) \\
&\qquad + I(U; V|W, Y_1, Z_2) - \lambda I(Y_1; Y_2) - (1 - \lambda)I(Z_1; Z_2) - I(Y_1; Z_2|U, V, W). \qquad \square
\end{aligned}$$

Thus the excess term one needs to cancel (by using a different choice of (U_1, V_1, W_1) or (U_2, V_2, W_2) or both) to ensure factorization, is at most $I(U; V|W, Y_1, Z_2)$.

Also observe that one can get a similar identity by interchanging $Y_1 \leftrightarrow Z_1$ and $Z_2 \leftrightarrow Y_2$. Here $W_1 = (W, Y_2)$ and $W_2 = (W, Z_1)$. This will yield the term $I(U; V|W, Y_2, Z_1)$ instead of $I(U; V|W, Y_1, Z_2)$.

Theorem 1. *The λ-sum-rate factorizes (as in (5)) if either of the conditions below hold:*

1. *Any one of the four channels $X_1 \to Y_1; X_1 \to Z_1; X_2 \to Y_2$ or $X_2 \to Z_2$ is deterministic.*
2. *In any one of the two components, one channel is more capable than the other.*

Proof. Assume the first condition holds. In particular let $X_2 \to Z_2$ be deterministic. Then we will show that

$$\lambda\text{-}SR_M(\mathfrak{q}_1 \times \mathfrak{q}_2, p(u, v, w, x_1, x_2)) \le \lambda\text{-}SR_M(\mathfrak{q}_1, p(u_1, v_1, w_1, x_1)) + \lambda\text{-}SR_M(\mathfrak{q}_2, p(u_2, v_2, w_2, x_2))$$

where $U_1 = U_2 = U, V_1 = V, V_2 = Z_2, W_1 = (W, Z_2), W_2 = (W, Y_1)$. To show this, from Claim 1 it suffices to show that

$$\lambda\text{-}SR_M(\mathfrak{q}_2, p(u, v, (w, y_1), x_2)) + I(U; V|W, Y_1, Z_2) \le \lambda\text{-}SR_M(\mathfrak{q}_2, p(u, z_2, (w, y_1), x_2)).$$

Observe that

$$\begin{aligned}
&\lambda\text{-}SR_M(\mathfrak{q}_2, p(u, v, (w, y_1), x_2)) + I(U; V|W, Y_1, Z_2) \\
&\quad = \lambda I(W, Y_1; Y_2) + (1 - \lambda)I(W, Y_1; Z_2) + I(U; Y_2|W, Y_1) \\
&\qquad + I(V; Z_2|W, Y_1) - I(U; V|W, Y_1) + I(U; V|W, Y_1, Z_2) \\
&\quad = \lambda I(W, Y_1; Y_2) + (1 - \lambda)I(W, Y_1; Z_2) + I(U; Y_2|W, Y_1) + I(V; Z_2|U, W, Y_1) \\
&\quad \le \lambda I(W, Y_1; Y_2) + (1 - \lambda)I(W, Y_1; Z_2) + I(U; Y_2|W, Y_1) + H(Z_2|U, W, Y_1) \\
&\quad = \lambda I(W, Y_1; Y_2) + (1 - \lambda)I(W, Y_1; Z_2) + I(U; Y_2|W, Y_1) + I(Z_2; Z_2|W, Y_1) - I(Z_2; U|W, Y_1) \\
&\quad = \lambda\text{-}SR_M(\mathfrak{q}_2, p(u, z_2, (w, y_1), x_2)).
\end{aligned}$$

Similar reasoning can deal with the case where $X_1 \to Y_1$ is a deterministic channel.

Note that if $X_2 \to Y_2$ is deterministic, then one must start with the interchanged W_1, W_2, i.e. $W_1 = (W, Y_2), W_2 = (W, Z_1)$, and similarly show that

$$\lambda\text{-}SR_M(\mathfrak{q}_2, p(u, v, (w, z_1), x_2)) + I(U; V|W, Z_1, Y_2) \leq \lambda\text{-}SR_M(\mathfrak{q}_2, p(y_2, v, (w, z_1), x_2)).$$

Finally, the case when $X_1 \to Z_1$ is deterministic can be dealt with similarly.

Proceeding to the second condition, let us assume that the channel $X_2 \to Y_2$ is more capable than the channel $X_2 \to Z_2$, i.e. for all $p(x_2), I(X_2; Y_2) \geq I(X_2; Z_2)$. Then observe that

$$\begin{aligned}
&\lambda\text{-}SR_M(\mathfrak{q}_2, p(u, v, (w, y_1), x_2)) + I(U; V|W, Y_1, Z_2) \\
&\quad = \lambda I(W, Y_1; Y_2) + (1-\lambda) I(W, Y_1; Z_2) + I(U; Y_2|W, Y_1) \\
&\qquad + I(V; Z_2|W, Y_1) - I(U; V|W, Y_1) + I(U; V|W, Y_1, Z_2) \\
&\quad = \lambda I(W, Y_1; Y_2) + (1-\lambda) I(W, Y_1; Z_2) + I(U; Y_2|W, Y_1) + I(V; Z_2|U, W, Y_1) \\
&\quad \leq \lambda I(W, Y_1; Y_2) + (1-\lambda) I(W, Y_1; Z_2) + I(U; Y_2|W, Y_1) + I(X_2; Z_2|U, W, Y_1) \\
&\quad \leq \lambda I(W, Y_1; Y_2) + (1-\lambda) I(W, Y_1; Z_2) + I(U; Y_2|W, Y_1) + I(X_2; Y_2|U, W, Y_1) \\
&\quad = \lambda I(W, Y_1; Y_2) + (1-\lambda) I(W, Y_1; Z_2) + I(X_2; Y_2|W, Y_1) \\
&\quad = \lambda\text{-}SR_M(\mathfrak{q}_2, p(x_2, \emptyset, (w, y_1), x_2)).
\end{aligned}$$

Thus from Claim 1 we have the factorization of $\lambda\text{-}SR_M(\mathfrak{q}_1 \times \mathfrak{q}_2)$.

Similar reasoning works for the other three cases. Again observe that when Z_2 is more capable than Y_2 or Y_1 is more capable than Z_1, one should start with start with the interchanged W_1, W_2, i.e. $W_1 = (W, Y_2), W_2 = (W, Z_1)$. This completes the proof of the lemma. □

2.2 Optimal sum rate for product of reversely semi-deterministic channels

Claim 2. *Marton's sum rate is optimal for product of reversely semi-deterministic channels. Moreover the sum rate of such a product channel $\mathfrak{q}_1 \times \mathfrak{q}_2$ is given by*

$$\min_{\lambda \in [0,1]} \big(\lambda\text{-}SR_M(\mathfrak{q}_1) + \lambda\text{-}SR_M(\mathfrak{q}_2)\big).$$

Proof. Take two semi-deterministic channels $\mathfrak{q}_1(y_1, z_1|x_1)$ and $\mathfrak{q}_2(y_2, z_2|x_2)$ where Y_1 is a deterministic function of X_1 and Z_2 is a deterministic function of X_2.

Consider the n-letter λ-sum-rate of the product channel $\mathfrak{q}_1 \times \mathfrak{q}_2$. Using Theorem 1 the n-letter product channel factorizes into λ-sum-rate of two n-letter sub channels. Each term again factorizes by repeated application of Theorem 1. More precisely,

$$\begin{aligned}
\lambda\text{-}SR_M(\mathfrak{q}_1 \otimes_n \times \mathfrak{q}_2 \otimes_n) &= \lambda\text{-}SR_M(\mathfrak{q}_1 \otimes_n) + \lambda\text{-}SR_M(\mathfrak{q}_2 \otimes_n) \\
&= n \cdot \lambda\text{-}SR_M(\mathfrak{q}_1) + n \cdot \lambda\text{-}SR_M(\mathfrak{q}_2).
\end{aligned}$$

Marton's inner bound sum rate for the n-letter of the product channel $\mathfrak{q}_1 \times \mathfrak{q}_2$ is equal to

$$\min_{\lambda \in [0,1]} \big(\lambda\text{-}SR_M(\mathfrak{q}_1 \otimes_n \times \mathfrak{q}_2 \otimes_n)\big).$$

We can write the above expression as

$$n \cdot \min_{\lambda \in [0,1]} \big(\lambda\text{-}SR_M(\mathfrak{q}_1) + \lambda\text{-}SR_M(\mathfrak{q}_2)\big).$$

Therefore, the actual sum rate satisfies[2]

$$SR^*(\mathfrak{q}_1 \times \mathfrak{q}_2) \leq \min_{\lambda\in[0,1]} \big(\lambda\text{-}SR_M(\mathfrak{q}_1) + \lambda\text{-}SR_M(\mathfrak{q}_2)\big).$$

On the other hand, this sum rate is achievable since it is equal to the single letter Marton's inner bound for $\mathfrak{q}_1 \times \mathfrak{q}_2$, i.e.

$$SR^*(\mathfrak{q}_1 \times \mathfrak{q}_2) = \min_{\lambda\in[0,1]} \lambda\text{-}SR_M(\mathfrak{q}_1 \times \mathfrak{q}_2) = \min_{\lambda\in[0,1]} \big(\lambda\text{-}SR_M(\mathfrak{q}_1) + \lambda\text{-}SR_M(\mathfrak{q}_2)\big).$$

□

2.3 The UV outer bound is strictly suboptimal

From UV outer bound the sum rate of a general broadcast channel can be bounded from above by

$$SR_{UV}(\mathfrak{q}) = \max_{p(u,v,x)} \min\{I(U;Y) + I(V;Z), I(U;Y) + I(X;Z|U), I(V;Z) + I(X;Y|V)\}. \tag{6}$$

In this example we will demonstrate a product of reversely semi-deterministic channel, $\mathfrak{q} = \mathfrak{q}_1 \times \mathfrak{q}_2$, such that the optimal sum rate $SR^*(\mathfrak{q}_1 \times \mathfrak{q}_2)$ satisfies

$$SR_{MIB}(\mathfrak{q}_1 \times \mathfrak{q}_2) = SR^*(\mathfrak{q}_1 \times \mathfrak{q}_2) < SR_{UV}(\mathfrak{q}_1 \times \mathfrak{q}_2).$$

This unequivocally shows that the UV outer bound is *strictly* suboptimal for the general broadcast channel.

Remark 1. Even if one were to consider the best outer bound with a common message requirement, the UVW outer bound [8], the fact that we are showing that the sum rate is strictly weak for the UV outer bound immediately implies the *strict* sub optimality of the UVW outer bound as well. To note this, observe that the projection of the UVW outer bound on the plane $R_0 = 0$ (which is shown in [8] to be the UV outer bound) is *strictly* suboptimal.

Claim 3. *Consider the reversely semi-deterministic channel in Figure 2. Assume that the transition probabilities are uniform across the possible outputs, i.e the red edges have a probability $\frac{1}{3}$ in the first component and the blue edges have a probability $\frac{1}{3}$ in the second component. Then Marton's sum rate (the optimal sum rate) is given by $\frac{8}{3} = 3 - \frac{1}{3}$, while the UV sum rate is at least by $3 - \frac{1}{15}$.*

Proof. We begin by showing that Marton's sum rate (the optimal sum rate) is given by $\frac{8}{3}$. Claim 2 shows that the sum rate of $\mathfrak{q}_1 \times \mathfrak{q}_2$ is

$$\min_{\lambda\in[0,1]} \big(\lambda\text{-}SR_M(\mathfrak{q}_1) + \lambda\text{-}SR_M(\mathfrak{q}_2)\big).$$

The result of Appendix B.1 implies that for any $\lambda \in [0,1]$, $\lambda\text{-}SR_M(\mathfrak{q}_1)$ is equal to $\lambda\text{-}SR_M(\mathfrak{q}_1, u(x_1))$ where u is the uniform distribution on $\mathcal{X}_1$. A similar statement holds for $\lambda\text{-}SR_M(\mathfrak{q}_2)$. Therefore the sum rate of $\mathfrak{q}_1 \times \mathfrak{q}_2$ is equal to

$$\min_{\lambda\in[0,1]} \big(\lambda\text{-}SR_M(\mathfrak{q}_1, u(x_1)) + \lambda\text{-}SR_M(\mathfrak{q}_2, u(x_2))\big). \tag{7}$$

[2] We utilize the known fact that Marton's inner bound sum rate for the n-letter version of the channel approaches the optimal sum rate as n goes to infinity.

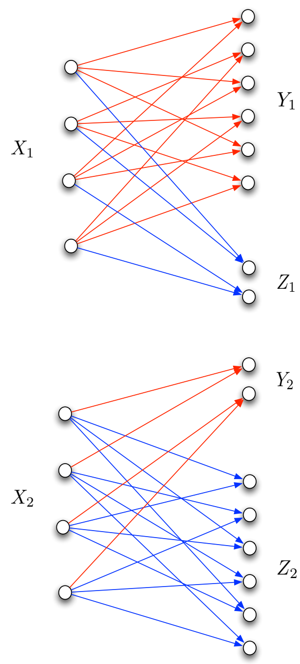

Figure 2: A reversely semi-deterministic channel

By symmetry, λ-$SR_M(\mathfrak{q}_2, u(x_2)) = (1-\lambda)$-$SR_M(\mathfrak{q}_1, u(x_1))$. Therefore we can express the sum rate as

$$\min_{\lambda\in[0,1]} \big(\lambda\text{-}SR_M(\mathfrak{q}_1, u(x_1)) + (1-\lambda)\text{-}SR_M(\mathfrak{q}_1, u(x_1))\big).$$

In Appendix B.2 we show that λ-$SR_M(\mathfrak{q}_1, u(x_1))$ is equal to

$$\lambda\text{-}SR_M(\mathfrak{q}_1, u(x_1)) = \begin{cases} \frac{5}{3} - \frac{2}{3}\lambda & \lambda \in [0, \frac{1}{2}] \\ \frac{4}{3} & \lambda \in [\frac{1}{2}, 1] \end{cases}.$$

Substituting this function into (7) we see that the minimum occurs uniquely at $\lambda = 0.5$ and the optimum sum rate is equal to $\frac{8}{3}$.

To compute a lower bound on the UV sum rate, let $p(x_1, x_2) = u(x_1)u(x_2)$, i.e. independent uniform distribution on $\mathcal{X}_1$ and $\mathcal{X}_2$. We define $U_1, V_1, X_1, U_2, V_2, X_2$ having a joint distribution of the form $p(u_1, v_1, x_1)p(u_2, v_2, x_2)$ as follows. Let $U_1 = Y_1$ and $p(u_2, x_2)$ satisfy

$$\mathrm{P}(X_2 = 1|U_2 = 1) = \mathrm{P}(X_2 = 3|U_2 = 1) = \frac{1}{2}, \text{ and } \mathrm{P}(X_2 = 2|U_2 = 1) = \mathrm{P}(X_2 = 4|U_2 = 1) = \frac{1}{2},$$
$$\mathrm{P}(U_2 = 1) = \mathrm{P}(U_2 = 2) = \frac{1}{2}.$$

Similarly, let $V_2 = Z_2$ and $p(v_1, x_1)$ satisfy

$$\mathrm{P}(X_1 = 1|V_1 = 1) = \mathrm{P}(X_1 = 3|V_1 = 1) = \frac{1}{2}, \text{ and } \mathrm{P}(X_1 = 2|V_1 = 1) = \mathrm{P}(X_1 = 4|V_1 = 1) = \frac{1}{2},$$
$$\mathrm{P}(V_1 = 1) = \mathrm{P}(V_1 = 2) = \frac{1}{2}.$$

Let Q_1 and Q_2 to binary random variable be mutually independent of each other, and of $U_1, V_1, X_1, U_2, V_2, X_2$. Furthermore assume that $\mathrm{P}(Q_1 = 0) = \mathrm{P}(Q_2 = 0) = \frac{4}{5}$. Define V_1' and U_2' as follows: When $Q_1 = 0$ set $V_1' = V_1$ and else set $V_1' = X_1$. When $Q_2 = 0$ set $U_2' = U_2$ and else set $U_2' = X_2$. Lastly set $\tilde{V}_1 = (V_1', Q_1)$ $\tilde{U}_2 = (U_2', Q_2)$.

We consider the UV region for the choice of $(U_1, \tilde{U}_2)$, $(\tilde{V}_1, V_2)$, (X_1, X_2). Note that

$$\begin{aligned} R_1 &\leq I(U_1, \tilde{U}_2; Y_1, Y_2) \\ &= I(U_1; Y_1) + I(\tilde{U}_2; Y_2) \\ &= H(Y_1) + \frac{4}{5} I(U_2; Y_2) + \frac{1}{5} I(X_2; Y_2) \\ &= 1 + \frac{4}{5} \cdot \frac{1}{3} + \frac{1}{5} \cdot 1 \\ &= \frac{22}{15}. \end{aligned}$$

Similarly, one can show that

$$\begin{aligned} R_2 &\leq I(\tilde{V}_1, V_2; Z_1, Z_2) \\ &= \frac{22}{15}. \end{aligned}$$

The sum rate constraint on $R_1 + R_2$ is as follows:

$$\begin{aligned} R_1 + R_2 &\leq I(U_1, \tilde{U}_2; Y_1, Y_2) + I(X_1, X_2; Z_1, Z_2 | U_1, \tilde{U}_2) \\ &= I(U_1; Y_1) + I(X_1; Z_1 | U_1) + I(\tilde{U}_2; Y_2) + I(X_2; Z_2 | \tilde{U}_2) \\ &= H(Y_1) + I(X_1; Z_1 | Y_1) + \frac{4}{5} I(U_2; Y_2) + \frac{1}{5} I(X_2; Y_2) + \frac{4}{5} H(Z_2 | U_2) \\ &= 1 + \frac{2}{3} + \frac{4}{5} \cdot \frac{1}{3} + \frac{1}{5} \cdot 1 + \frac{4}{5} \cdot 1 \\ &= \frac{44}{15}. \end{aligned}$$

Similarly, one can show that

$$\begin{aligned} R_1 + R_2 &\leq I(\tilde{V}_1, V_2; Z_1, Z_2) + I(X_1, X_2; Y_1, Y_2 | \tilde{V}_1, V_2) \\ &= \frac{44}{15}. \end{aligned}$$

Therefore the point $(R_1, R_2) = (\frac{22}{15}, \frac{22}{15})$ is in this region. Hence the UV sum rate is at least $\frac{44}{15} = 3 - \frac{1}{15}$. Thus for the product channel under consideration

$$\frac{8}{3} = SR_{MIB}(\mathfrak{q}_1 \times \mathfrak{q}_2) = SR^*(\mathfrak{q}_1 \times \mathfrak{q}_2) < \frac{44}{15} \leq SR_{UV}(\mathfrak{q}_1 \times \mathfrak{q}_2).$$

This shows that the UV outer bound is strictly suboptimal in general. □

3 Capacity regions for classes of product broadcast channels

In this section we establish the capacity region for classes of product broadcast channels. Here we consider a more general setting where in addition to the private messages, the receivers also wish to decode a common message M_0. Hence we are interested in the achievable rate triples (R_0, R_1, R_2). The capacity region is defined in a similar fashion as in the case without common message.

3.1 An outer bound for product channels

We present a new outer bound for the product of two broadcast channels. The manipulations here are inspired by the manipulations in the proof of Theorem 1. This outer bound matches the capacity region for a variety of product channels, including product of two reversely semideterministic and product of two reversely more-capable channels. Hence, from Claim 3, it follows that this is a *strictly* better bound for product broadcast channels as compared to the UVW outer bound.

Claim 4. *Given a product channel* $\mathfrak{q}(y_1y_2, z_1z_2|x_1x_2) = \mathfrak{q}(y_1, z_1|x_1)\mathfrak{q}(y_2, z_2|x_2)$*, the union over all* $p_1(w_1, u_1, v_1, x_1)p_2(w_2, u_2, v_2, x_2)$ *of triples* (R_0, R_1, R_2) *satisfying*

$$
\begin{aligned}
R_0 &\leq \min\{I(W_1;Y_1) + I(W_2;Y_2), I(W_1;Z_1) + I(W_2;Z_2)\}\\
R_0 + R_1 &\leq \min\{I(W_1;Y_1) + I(W_2;Y_2), I(W_1;Z_1) + I(W_2;Z_2)\} + I(U_1;Y_1|W_1) + I(U_2;Y_2|W_2)\\
R_0 + R_2 &\leq \min\{I(W_1;Y_1) + I(W_2;Y_2), I(W_1;Z_1) + I(W_2;Z_2)\} + I(V_1;Z_1|W_1) + I(V_2;Z_2|W_2)\\
R_0 + R_1 + R_2 &\leq \min\{I(W_1;Y_1) + I(W_2;Y_2), I(W_1;Z_1) + I(W_2;Z_2)\}\\
&\quad + I(U_2;Y_2|W_2) + I(X_2;Z_2|U_2,W_2)\\
&\quad + \min\big\{I(U_1;Y_1|W_1) + I(X_1;Z_1|U_1,W_1), I(V_1;Z_1|W_1) + I(X_1;Y_1|V_1,W_1)\big\},\\
R_0 + R_1 + R_2 &\leq \min\{I(W_1;Y_1) + I(W_2;Y_2), I(W_1;Z_1) + I(W_2;Z_2)\}\\
&\quad + \min\big\{I(U_2;Y_2|W_2) + I(X_2;Z_2|U_2,W_2), I(V_2;Z_2|W_2) + I(X_2;Y_2|V_2,W_2)\big\}\\
&\quad + I(V_1;Z_1|W_1) + I(X_1;Y_1|V_1,W_1),
\end{aligned}
$$

forms an outer bound to the general broadcast channel.

Remark 2. Note that setting $X_2, Y_2, Z_2 = \emptyset$ reduces this bound to UVW outer bound [8]. Additionally, one can interchange the roles of Y_2 and Z_1 with Z_2 and Y_1 respectively to get another set of similar constraints. These constraints will be over different auxiliaries distributed as $p_1(\tilde{w}_1, \tilde{u}_1, \tilde{v}_1, x_1)p_2(\tilde{w}_2, \tilde{u}_2, \tilde{v}_2, x_2)$ (observe that the distributions on X_1, X_2 are preserved), and we can take the intersection of these two constraints. Finally one can take union of these two sets of constraints over all $p_1(w_1, u_1, v_1, \tilde{w}_1, \tilde{u}_1, \tilde{v}_1, x_1)p_2(w_2, u_2, v_2, \tilde{w}_2, \tilde{u}_2, \tilde{v}_2, x_2)$ to get another, possibly better, outer bound.

Proof. The proof of this claim is given in the Appendix C. □

Remark 3. The above outer bound is *also strictly sub-optimal.* To see this first note that when one of the product channels is trivial, this outer bound does not give us anything beyond the UVW-outer bound [8]. Now, consider a product of three channels, first one is trivial, the collection of two and three forms a reversely semi-deterministic pair. The new outer bound reduces to the UVW bound on the reversely semi-deterministic, and therefore it is strictly sub-optimal. However, one could argue that in order to write the outer bound, one should take the intersection of all possible outer bounds one can write by breaking up the broadcast channel into product forms. To deal with this objection one can consider the product of three channels as above and then slightly perturb the channel to destroy the product form structure of the channel. Because the above outer bound is continuous in the underlying channel, this outer bound must be loose for this channel. In fact, we still don't know of "the correct way" to write an outer bound that fully captures the spirit of the counterexample discussed earlier. We have thought of alternative expressions but none seemed satisfactory.

3.1.1 An achievable region for a product broadcast channel

Given a product channel $\mathfrak{q}(y_1 y_2, z_1 z_2|x_1 x_2) = \mathfrak{q}(y_1, z_1|x_1)\mathfrak{q}(y_2, z_2|x_2)$ the union of rate triples satisfying

$$\begin{aligned} R_0 &\leq \min\{I(W_1;Y_1) + I(W_2;Y_2), I(W_1;Z_1) + I(W_2;Z_2)\} \\ R_0 + R_1 &\leq I(W_1;Y_1) + I(W_2;Y_2) + I(U_1;Y_1|W_1) + I(U_2;Y_2|W_2) \\ R_0 + R_2 &\leq I(W_1;Z_1) + I(W_2;Z_2) + I(V_1;Z_1|W_1) + I(V_2;Z_2|W_2) \\ R_0 + R_1 + R_2 &\leq \min\{I(W_1;Y_1) + I(W_2;Y_2), I(W_1;Z_1) + I(W_2;Z_2)\} \\ &\quad + I(U_1;Y_1|W_1) + I(U_2;Y_2|W_2) + I(V_1;Z_1|W_1) \\ &\quad + I(V_2;Z_2|W_2) - I(U_1;V_1|W_1) - I(U_2;V_2|W_2) \end{aligned} \tag{8}$$

over all $p_1(w_1, v_1, u_1, x_1)p_2(w_2, v_2, u_2, x_2)$ constitutes an inner bound to the capacity region. The achievablility of these points are immediate from Marton's inner bound by letting $U = (U_1, U_2), V = (V_1, V_2), W = (W_1, W_2)$ and $p(u, v, w) \sim p_1(w_1, v_1, u_1, x_1)p_2(w_2, v_2, u_2, x_2)$.

3.2 Capacity regions for new classes of product broadcast channels

Theorem 2. *The capacity region for a product of reversely semi-deterministic (say, channels $X_1 \to Y_1, X_2 \to Z_2$ are deterministic) broadcast channel is given by the union of rate triples satisfying*

$$\begin{aligned} R_0 &\leq \min\{I(W_1;Y_1) + I(W_2;Y_2), I(W_1;Z_1) + I(W_2;Z_2)\} \\ R_0 + R_1 &\leq I(W_1;Y_1) + I(W_2;Y_2) + H(Y_1|W_1) + I(U_2;Y_2|W_2) \\ R_0 + R_2 &\leq I(W_1;Z_1) + I(W_2;Z_2) + I(V_1;Z_1|W_1) + H(Z_2|W_2) \\ R_0 + R_1 + R_2 &\leq \min\{I(W_1;Y_1) + I(W_2;Y_2), I(W_1;Z_1) + I(W_2;Z_2)\} \\ &\quad + I(V_1;Z_1|W_1) + H(Y_1|V_1, W_1) + I(U_2;Y_2|W_2) + H(Z_2|U_2, W_2) \end{aligned}$$

over all $p_1(w_1, v_1, x_1)p_2(w_2, u_2, x_2)$.

Proof. The achievability is immediate by setting $U_1 = Y_1$ and $V_2 = Z_2$ in (8). Note that these two choices of auxiliary random variables are possible since channels $X_1 \to Y_1, X_2 \to Z_2$ are deterministic.

The converse is also immediate from the outer bound in 4. Observe that for any $p_1(w_1, v_1, u_1, x_1)$, $p_2(w_2, v_2, u_2, x_2)$ we have

$$I(U_1;Y_1|W_1) \leq H(Y_1|W_1), \quad I(V_2;Z_2|W_2) \leq H(Z_2|W_2),$$

and each of the two sum rate terms is bounded by

$$\begin{aligned} &\min\{I(W_1;Y_1) + I(W_2;Y_2), I(W_1;Z_1) + I(W_2;Z_2)\} + I(V_1;Z_1|W_1) \\ &\quad + H(Y_1|V_1, W_1) + I(U_2;Y_2|W_2) + H(Z_2|U_2, W_2). \end{aligned}$$

Thus the outer bound is contained in the inner bound (and hence they coincide). □

Theorem 3. *The capacity region for a product of reversely more-capable (say, receiver Z_1 is more capable than Y_1, and receiver Y_2 is more capable than Z_2) broadcast channel is given by the union* *of rate triples satisfying*

$$\begin{aligned}
R_0 &\le \min\{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}\\
R_0+R_1 &\le \min\{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}+I(U_1;Y_1|W_1)+I(X_2;Y_2|W_2)\\
R_0+R_2 &\le \min\{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}+I(X_1;Z_1|W_1)+I(V_2;Z_2|W_2)\\
R_0+R_1+R_2 &\le \min\{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}+I(X_2;Y_2|W_2)\\
&\quad + \min\big\{I(U_1;Y_1|W_1)+I(X_1;Z_1|U_1,W_1), I(X_1;Z_1|W_1)\big\},\\
R_0+R_1+R_2 &\le \min\{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}\\
&\quad + \min\big\{I(X_2;Y_2|W_2), I(V_2;Z_2|W_2)+I(X_2;Y_2|V_2,W_2)\big\}+I(X_1;Z_1|W_1)
\end{aligned}$$

over all $p_1(w_1,v_1,x_1)p_2(w_2,u_2,x_2)$.

Proof. The achievability is immediate by setting $W_1'=(U_1,W_1), U_1'=\emptyset, V_1'=X_1$ and $W_2'=(V_2,W_2), U_2'=X_2', V_2'=\emptyset$ in (8). Plugging these choices into (8) we obtain that one can achieve rate triples satisfying

$$\begin{aligned}
R_0 &\le \min\{I(U_1,W_1;Y_1)+I(V_2,W_2;Y_2), I(U_1,W_1;Z_1)+I(V_2,W_2;Z_2)\}\\
R_0+R_1 &\le I(U_1,W_1;Y_1)+I(V_2,W_2;Y_2)+I(X_2;Y_2|V_2,W_2)\\
&= I(W_1;Y_1)+I(W_2;Y_2)+(U_1;Y_1|W_1)+I(X_2;Y_2|W_2)\\
R_0+R_2 &\le I(U_1,W_1;Z_1)+I(V_2,W_2;Z_2)+I(X_1;Z_1|U_1,W_1)\\
&= I(W_1;Z_1)+I(W_2;Z_2)+I(X_1;Z_1|W_1)+I(V_2;Z_2|W_2)\\
R_0+R_1+R_2 &\le \min\{I(U_1,W_1;Y_1)+I(V_2,W_2;Y_2), I(U_1,W_1;Z_1)+I(V_2,W_2;Z_2)\}\\
&\quad + I(X_2;Y_2|V_2,W_2)+I(X_1;Z_1|U_1,W_1).
\end{aligned}$$

The last sum rate term can be split into two terms as follows

$$\begin{aligned}
R_0+R_1+R_2 &\le I(U_1,W_1;Y_1)+I(V_2,W_2;Y_2)+I(X_2;Y_2|V_2,W_2)+I(X_1;Z_1|U_1,W_1)\\
&= I(W_1;Y_1)+I(W_2;Y_2)+I(X_2;Y_2|W_2)+I(U_1;Y_1|W_1)+I(X_1;Z_1|U_1,W_1)\\
R_0+R_1+R_2 &\le I(U_1,W_1;Z_1)+I(V_2,W_2;Z_2)+I(X_2;Y_2|V_2,W_2)+I(X_1;Z_1|U_1,W_1)\\
&= I(W_1;Z_1)+I(W_2;Z_2)+I(X_1;Z_1|W_1)+I(V_2;Z_2|W_2)+I(X_2;Y_2|V_2,W_2).
\end{aligned}$$

Thus we see, by comparing term by term, that this achievable region is at least as large as the region stated in Theorem 3, and hence the region in Theorem 3 is achievable.

The converse is also reasonably immediate from the outer bound in 4. Observe the following:

$$\begin{aligned}
&\min\{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}+I(U_1;Y_1|W_1)+I(U_2;Y_2|W_2)\\
&\quad\le \min\{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}+I(U_1;Y_1|W_1)+I(X_2;Y_2|W_2),\\
&\min\{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}+I(V_1;Z_1|W_1)+I(V_2;Z_2|W_2)\\
&\quad \{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}+I(X_1;Z_1|W_1)+I(V_2;Z_2|W_2),\\
&\min\{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}+I(U_2;Y_2|W_2)+I(X_2;Z_2|U_2,W_2)\\
&\quad + \min\big\{I(U_1;Y_1|W_1)+I(X_1;Z_1|U_1,W_1), I(V_1;Z_1|W_1)+I(X_1;Y_1|V_1,W_1)\big\}\\
&\quad\le \min\{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}+I(X_2;Y_2|W_2)\\
&\qquad + \min\big\{I(U_1;Y_1|W_1)+I(X_1;Z_1|U_1,W_1), I(X_1;Z_1|W_1)\big\},
\end{aligned}$$

and finally,

$$\begin{aligned}
&\min\{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}+I(V_1;Z_1|W_1)+I(X_1;Y_1|V_1,W_1)\\
&\quad+\min\left\{I(U_2;Y_2|W_2)+I(X_2;Z_2|U_2,W_2), I(V_2;Z_2|W_2)+I(X_2;Y_2|V_2,W_2)\right\}\\
&\quad\le \{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}+I(X_1;Z_1|W_1)\\
&\qquad+\min\left\{I(X_2;Y_2|W_2), I(V_2;Z_2|W_2)+I(X_2;Y_2|V_2,W_2)\right\}.
\end{aligned}$$

Thus we see, by comparing term by term, that the region stated in Theorem 3 is at least as large as the outer bound in Claim 4. Hence the region in Theorem 3 is an outer bound, thus completing the converse. □

Remark 4. The achievable region in (8) also matches the outer bound in Claim 4 for a variety of other classes. For instance, say Z_1 is more capable than Y_1 and Y_2 is a deterministic function of X_2. In this case, one can show that the capacity region is given by the union of rate triples satisfying

$$\begin{aligned}
R_0 &\le \min\{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}\\
R_0+R_1 &\le \min\{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}+I(U_1;Y_1|W_1)+H(Y_2|W_2)\\
R_0+R_2 &\le \min\{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}+I(X_1;Z_1|W_1)+I(V_2;Z_2|W_2)\\
R_0+R_1+R_2 &\le \min\{I(W_1;Y_1)+I(W_2;Y_2), I(W_1;Z_1)+I(W_2;Z_2)\}\\
&\qquad+I(V_2;Z_2|W_2)+H(Y_2|V_2,W_2)+I(X_1;Z_1|W_1).
\end{aligned}$$

The details are left to the reader.

4 On Marton's inner bound and λ-sum-rate

In this section we prove a collection of results regarding Marton's inner bound and also about the quantity we introduced earlier, the λ-sum-rate.

4.1 Two letter Marton's inner bound

This section considers the two letter Marton's inner bound and the role it plays in determining the optimality of the traditional Marton's inner bound. To simplify our analysis and for the ease of exposition we will focus on the sum rate, but some of the insights that we obtained have already been useful beyond just the sum rate.

Given a broadcast channel $\mathfrak{q}(y,z|x)$ the maximum sum rate achievable via Marton's strategy is given by

$$SR_M(\mathfrak{q}) = \max_{p(u,v,w,x)} \min\{I(W;Y), I(W;Z)\}+I(U;Y|W)+I(V;Z|W)-I(U;V|W). \tag{9}$$

The maximum is taken over distributions $p(u,v,w,x)$ where the auxiliary random variables (U,V,W) satisfy the Markov chain $(U,V,W)\to X\to(Y,Z)$.

Consider a product broadcast channel $\mathfrak{q}(y_1,z_1|x_1)\times\mathfrak{q}(y_2,z_2|x_2)$ obtained by taking identical copies of the original channel. One can obtain the maximum sum rate achievable via Marton's strategy for this new channel as

$$\begin{aligned}
SR_M(\mathfrak{q}\times\mathfrak{q}) = \max_{p(u,v,w,x_1,x_2)} &\min\{I(W;Y_1,Y_2), I(W;Z_1,Z_2)\}+I(U;Y_1,Y_2|W)\\
&+I(V;Z_1,Z_2|W)-I(U;V|W).
\end{aligned}\tag{10}$$

Here the maximum is taken over distributions $p(u,v,w,x_1,x_2)$ where the auxiliary random variables (U,V,W) satisfy the Markov chain: $(U,V,W)\to X_1,X_2\to(Y_1,Y_2,Z_1,Z_2)$, and the channel has a product nature given by $\mathfrak{q}(y_1, y_2, z_1, z_2|x_1, x_2) = \mathfrak{q}(y_1, z_1|x_1)\mathfrak{q}(y_2, z_2|x_2)$. Define $SR_{2M}(\mathfrak{q}) := \frac{1}{2}SR_M(\mathfrak{q} \times \mathfrak{q})$ to be the two-letter sum rate yielded by Marton's inner bound.

Here we state a (folk-lore) lemma that relates the optimality of Marton's achievable strategy and the relationship between $SR_{2M}(\mathfrak{q})$ and $SR_M(\mathfrak{q})$.

Lemma 2. *(Folklore) The following two statements are equivalent:*

1. *Marton's achievable strategy achieves the optimal sum rate, $SR^*(\mathfrak{q})$, for all broadcast channels $\mathfrak{q}(y, z|x)$, i.e. $SR_M(\mathfrak{q}) = SR^*(\mathfrak{q})$.*
2. *$SR_{2M}(\mathfrak{q}) = SR_M(\mathfrak{q})$ for all $\mathfrak{q}(y, z|x)$.*

Proof. We present an argument here for completeness.

($1 \implies 2$) This follows from two facts: first, $SR_{2M}(\mathfrak{q})$ yields an achievable sum rate for the broadcast channel $\mathfrak{q}(y, z|x)$, i.e. $SR_{2M}(\mathfrak{q}) \leq SR^*(\mathfrak{q})$; and second, $SR_{2M}(\mathfrak{q}) \geq SR_M(\mathfrak{q})$ for all $\mathfrak{q}(y, z|x)$. To see the first, observe that a codebook of block length n for the product channel $\mathfrak{q}(y_1, z_1|x_1)\mathfrak{q}(y_2, z_2|x_2)$ yields a codebook of block length $2n$ for the original channel $\mathfrak{q}(y, z|x)$, since the mapping from $(x_1, ..., x_{2n})$ to the pairs $(y_1, \dots y_{2n})$, $(z_1, \dots, z_{2n})$ by the channel $\mathfrak{q}(y, z|x)$ is same as the mapping from $((x_1, x_2), ..., (x_{2n-1}, x_{2n}))$ to the pairs $((y_1, y_2), ..., (y_{2n-1}, y_{2n}))$, and $((z_1, z_2), ..., (z_{2n-1}, z_{2n}))$ by the channel $\mathfrak{q}(y_1, z_1|x_1)\mathfrak{q}(y_2, z_2|x_2)$. Hence any rate achievable for the product channel $\mathfrak{q}(y_1, z_1|x_1)\mathfrak{q}(y_2, z_2|x_2)$ (normalized by factor $\frac{1}{2}$) is also achievable for the single channel $\mathfrak{q}(y, z|x)$.

Let $p^*(u, v, w, x)$ achieve the maximum sum rate in (9). Choose $\tilde{U} = (U_1, U_2), \tilde{V} = (V_1, V_2), \tilde{W} = (W_1, W_2)$ and let $p(\tilde{u}, \tilde{v}, \tilde{w}, x_1, x_2) = p^*(u_1, v_1, w_1, x_1)p^*(u_2, v_2, w_2, x_2)$, i.e. take a product distribution by taking two i.i.d. copies of the single letter optimal distribution. Now observe that

$$\begin{aligned}
2SR_{2M}(\mathfrak{q}) &\geq \min\{I(\tilde{W}; Y_1, Y_2), I(\tilde{W}; Z_1, Z_2)\} + I(\tilde{U}; Y_1, Y_2|\tilde{W}) + I(\tilde{V}; Z_1, Z_2|\tilde{W}) - I(\tilde{U}; \tilde{V}|\tilde{W}) \\
&= \min\{I(W_1; Y_1), I(W_1; Z_1)\} + I(U_1; Y_1|W_1) \quad + I(V_1; Z_1|W_1) - I(U_1; V_1|W_1) \\
&\quad + \min\{I(W_2; Y_2), I(W_2; Z_2)\} + I(U_2; Y_2|W_2) + I(V_2; Z_2|W_2) - I(U_2; V_2|W_2) \\
&= 2SR_M(\mathfrak{q}).
\end{aligned}$$

This shows that if $SR_M(\mathfrak{q})$ is the maximum achievable sum rate then $SR_{2M}(\mathfrak{q}) = SR_M(\mathfrak{q})$ for all $\mathfrak{q}(y, z|x)$.

($2 \implies 1$) Let $\mathfrak{q} \otimes_n (y_1^n, z_1^n|x_1^n) = \prod_{i=1}^n \mathfrak{q}(y_i, z_i|x_i)$ denote the n-fold product channel. If (2) holds then, by induction, for any $k \geq 1$ the 2^k-fold product channel satisfies

$$\frac{1}{2^k} SR_M(\mathfrak{q} \otimes_{2^k}) = SR_M(\mathfrak{q}).$$

However for any n, we know from Fano's inequality that for any sequence of good codebooks

$$\begin{aligned}
&n(R_1 + R_2) \\
&\quad \leq I(M_1; Y_1^n) + I(M_2; Z_1^n) + n(R_1 + R_2)\epsilon_n + 1 \\
&\quad \leq SR_M(\mathfrak{q} \otimes_n) + n(R_1 + R_2)\epsilon_n + 1.
\end{aligned}$$

where $SR_M(\mathfrak{q} \otimes_n)$ is the maximum sum rate by Marton's strategy for the n-fold product channel, as setting $U = M_1, V = M_2, W = \emptyset$ is a particular choice of the auxiliary random variables for the n-fold product channel. Further we also know that $\epsilon_n \to 0$ as $n \to \infty$. This implies that the optimal sum rate, $SR^*(\mathfrak{q})$, for the broadcast channel $\mathfrak{q}(y, z|x)$ satisfies

$$SR^*(\mathfrak{q}) \leq \liminf_n \frac{1}{n} SR_M(\mathfrak{q} \otimes_n) \leq \lim_{k \to \infty} \frac{1}{2^k} SR_M(\mathfrak{q} \otimes_{2^k}) = SR_M(\mathfrak{q}).$$

On the other hand $SR_M(\mathfrak{q}) \leq SR^*(\mathfrak{q})$ since $SR_M(\mathfrak{q})$ is the rate given by Marton's achievable strategy. Hence we have $SR_M(\mathfrak{q}) = SR^*(\mathfrak{q})$. $\square$

Remark 5. Lemma 2 is an attempt at answering the question of whether Marton'd inner bound is optimal or not. If one can find a channel for which $SR_{2M}(\mathfrak{q}) > SR_M(\mathfrak{q})$ then Marton's inner bound is strictly sub-optimal, otherwise (i.e. for all channels $\mathfrak{q}$ we have $SR_{2M}(\mathfrak{q}) = SR_M(\mathfrak{q})$) Marton's inner bound is optimal and would yield the capacity region. The advantage of just having to look at 2-letter extensions is that with the recently established cardinality bounds [6] one can numerically search over channels $\mathfrak{q}$, to try and determine a channel where $SR_{2M}(\mathfrak{q}) > SR_M(\mathfrak{q})$. So far, our searches have yielded evidence to the contrary, i.e. they point towards a potential optimality of Marton's coding scheme.

4.2 Properties of λ-sum-rate

In this section, we state some results about the λ-sum-rate as this quantity seems to possess properties (such as factorizations over $\mathfrak{q}_1 \times \mathfrak{q}_2$) which we will show that $SR_M(\mathfrak{q})$ does not possess. Further λ-sum-rate also gives us a lot of insight into evaluations of the various bounds and in the search for potential counterexamples to optimality of Marton.

Lemma 3. *For a given channel $\mathfrak{q}(y,z|x)$, λ-$SR_M(\mathfrak{q})$ and λ-$SR_M(\mathfrak{q}, p(x))$ are convex in λ for $\lambda \in [0,1]$.*

Proof. To show that $\lambda \mapsto \lambda\text{-}SR_M(\mathfrak{q}, p(x))$ is convex, take arbitrary $\lambda_1, \lambda_2, \lambda_3$ satisfying $\lambda_2 = \frac{\lambda_1+\lambda_3}{2}$. Take some $p(w^*, u^*, v^*|x)$ maximizing $\lambda_2\text{-}SR_M(\mathfrak{q}, p(x))$. Note that

$$\begin{aligned}
&\lambda_2\text{-}SR_M(\mathfrak{q}, p(x)) = \\
&\big\{\lambda_2 I(W^*;Y) + (1-\lambda_2) I(W^*;Z) + I(U^*;Y|W^*) + I(V^*;Z|W^*) - I(U^*;V^*|W^*)\big\} = \\
&\frac{1}{2}\Big[\big\{\lambda_1 I(W^*;Y) + (1-\lambda_1) I(W^*;Z) + I(U^*;Y|W^*) + I(V^*;Z|W^*) - I(U^*;V^*|W^*)\big\} + \\
&\big\{\lambda_3 I(W^*;Y) + (1-\lambda_3) I(W^*;Z) + I(U^*;Y|W^*) + I(V^*;Z|W^*) - I(U^*;V^*|W^*)\big\}\Big] \leq \\
&\frac{1}{2}\big[\lambda_1\text{-}SR_M(\mathfrak{q}, p(x)) + \lambda_3\text{-}SR_M(\mathfrak{q}, p(x))\big].
\end{aligned}$$

To show that $\lambda \mapsto \lambda\text{-}SR_M(\mathfrak{q})$ is convex, let $p^*(x)$ be the maximizing input distribution, i.e. $\lambda\text{-}SR_M(\mathfrak{q}, p^*(x)) = \lambda\text{-}SR_M(\mathfrak{q})$. Note that

$$\begin{aligned}
\lambda\text{-}SR_M(\mathfrak{q}) &= \lambda\text{-}SR_M(\mathfrak{q}, p^*(x)) \\
&\leq \frac{1}{2}\big[\lambda_1\text{-}SR_M(\mathfrak{q}, p^*(x)) + \lambda_3\text{-}SR_M(\mathfrak{q}, p^*(x))\big] \\
&\leq \max_{p(x)} \frac{1}{2}\lambda_1\text{-}SR_M(\mathfrak{q}, p(x)) + \max_{p(x)} \frac{1}{2}\lambda_3\text{-}SR_M(\mathfrak{q}, p(x)) \\
&= \frac{1}{2}\big[\lambda_1\text{-}SR_M(\mathfrak{q}) + \lambda_3\text{-}SR_M(\mathfrak{q})\big].
\end{aligned}$$

$\square$

Lemma 4. *λ-$SR_M(\mathfrak{q})$ is related to the optimal sum rate as follows:*

$$\min_{\lambda \in \{0,1\}} \lambda\text{-}SR_M(\mathfrak{q}) \geq SR^*(\mathfrak{q}),$$

i.e. the minimum value of λ-SR at the boundaries, i.e. $\lambda = 0, 1$, yields an upper bound on the optimal sum rate, $SR^*(\mathfrak{q})$.

Proof. We prove the statement for $\lambda = 0$; the proof for $\lambda = 1$ is similar. We begin by showing that for any $p(x)$, $0\text{-}SR_M(\mathfrak{q}, p(x)) = \max_{p(w|x)} I(W;Z) + I(X;Y|W)$. This implies that $0\text{-}SR_M(\mathfrak{q}) = \max_{p(w,x)} I(W;Z) + I(X;Y|W)$ which is in turn an upper bound on the optimal sum rate by the UV outer bound (replace W by V). To see this first note that by setting $V = \emptyset, U = X$ we obtain

$$0\text{-}SR_M(\mathfrak{q}, p(x)) \geq \max_{p(w|x)} I(W;Z) + I(X;Y|W)$$

To obtain the other direction, observe that

$$\begin{aligned}
0\text{-}SR_M(\mathfrak{q}, p(x)) &= \max_{p(u,v,w|x)} \big\{ I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W) \big\} \\
&= \max_{p(u,v,w|x)} \big\{ I(VW;Z) + I(U;Y|W) - I(U;V|W) \big\} \\
&= \max_{p(u,v,w|x)} \big\{ I(VW;Z) + I(U;Y|VW) - I(U;V|WY) \big\} \\
&\leq \max_{p(u,v,w|x)} \big\{ I(VW;Z) + I(X;Y|VW) \big\} \\
&= \max_{p(w'|x)} I(W';Z) + I(X;Y|W').
\end{aligned}$$

Note that in the last step we replace (V, W) by W'.

Thus we have, as desired,

$$0\text{-}SR_M(\mathfrak{q}, p(x)) = \max_{p(w|x)} I(W;Z) + I(X;Y|W).$$

□

Corollary 1. *If the minimum value of λ-$SR_M(\mathfrak{q})$ is attained at $\lambda = 0$ or $\lambda = 1$ then $SR_M(\mathfrak{q}) = SR^*(\mathfrak{q})$, i.e. Marton's strategy achieves the optimal sum rate.*

Proof. This follows from the relationships

$$\min_{\lambda \in [0,1]} \lambda\text{-}SR_M(\mathfrak{q}) = SR_M(\mathfrak{q}) \leq SR^*(\mathfrak{q}) \leq \min_{\lambda \in \{0,1\}} \lambda\text{-}SR_M(\mathfrak{q}).$$

□

Lemma 5. *To compute the maximum sum rate in (4), it suffices to consider auxiliary random variables that satisfy $|\mathcal{U}|, |\mathcal{V}|, |\mathcal{W}| \leq |\mathcal{X}|$.*

Proof. This is proved in Theorem 2 of [9]. □

Lemma 6. *Take some arbitrary $p(x)$ and real λ^*. Then for any $p(w^*, u^*, v^*|x)$ maximizing λ-$SR_M(\mathfrak{q}, p(x))$, the line $\lambda \mapsto (\lambda - \lambda^*)(I(W^*;Y) - I(W^*;Z)) + \lambda^*\text{-}SR_M(\mathfrak{q}, p(x))$ is a supporting hyperplane to the convex curve $\lambda \mapsto \lambda\text{-}SR_M(\mathfrak{q}, p(x))$.*

Proof. At $\lambda = \lambda^*$, the expression $(\lambda - \lambda^*)(I(W^*;Y) - I(W^*;Z)) + \lambda^*\text{-}SR_M(\mathfrak{q}, p(x))$ is equal to $\lambda^*\text{-}SR_M(\mathfrak{q}, p(x))$ which is a point on the curve $\lambda \mapsto \lambda\text{-}SR_M(\mathfrak{q}, p(x))$. We need to show that for any arbitrary λ,

$$\lambda\text{-}SR_M(\mathfrak{q}, p(x)) \geq (\lambda - \lambda^*)(I(W^*;Y) - I(W^*;Z)) + \lambda^*\text{-}SR_M(\mathfrak{q}, p(x)).$$

The above inequality holds because it is equivalent to

$$\lambda\text{-}SR_M(\mathfrak{q}, p(x)) \geq \lambda I(W^*; Y) + (1-\lambda) I(W^*; Z) + I(U^*; Y|W^*) + I(V^*; Z|W^*) - I(U^*; V^*|W^*).$$

$\square$

Lemma 7. *λ-$SR_M(\mathfrak{q}, p(x))$ is constant in λ for less noisy[3] channels, deterministic channels, and linear in λ for more capable channels.*

Proof. Less Noisy: Assume that Y is less noisy than Z.

$$\begin{aligned}
\lambda\text{-}SR_M(\mathfrak{q}, p(x)) &= \max_{p(u,v,w|x)} \big[\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)\big] \\
&\leq \max_{p(u,v,w|x)} \big[\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|U,W)\big] \\
&\overset{(a)}{\leq} \max_{p(u,v,w|x)} \big[\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Y|U,W)\big] \\
&\overset{(a)}{\leq} \max_{p(w|x)} \big[\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(X;Y|W)\big] \\
&\leq \max_{p(w|x)} \big[I(W;Y) + I(X;Y|W)\big] = I(X;Y).
\end{aligned}$$

The inequalities marked by (a) are justified using the less noisy assumption. On the other hand setting[4] $W = \emptyset$, $U = X$, $V = \emptyset$ shows that λ-$SR_M(\mathfrak{q}, p(x)) \geq I(X;Y)$. Hence λ-$SR_M(\mathfrak{q}, p(x))$ is a constant.

Deterministic:

$$\begin{aligned}
\lambda\text{-}SR_M(\mathfrak{q}, p(x)) &= \max_{p(u,v,w|x)} \big[\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)\big] \\
&\leq \max_{p(u,v,w|x)} \big[\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|U,W)\big] \\
&\leq \max_{p(u,v,w|x)} \big[\lambda I(W;Y,Z) + (1-\lambda) I(W;Y,Z) + I(U;Y,Z|W) + I(V;Y,Z|U,W)\big] \\
&\leq I(X;Y,Z) = H(Y,Z).
\end{aligned}$$

One the other hand setting $W = \emptyset$, $U = Y$, $V = Z$ shows that λ-$SR_M(\mathfrak{q}, p(x)) \geq H)Y, Z)$. Hence λ-$SR_M(\mathfrak{q}, p(x))$ is a constant. (Note that these choices of auxiliaries, i.e. $U = Y$, $V = Z$, are permissible for deterministic channels since $(U,V) \rightarrow X \rightarrow (Y,Z)$ is a Markov chain.)

More capable: Assume that Y is more capable than Z.

$$\begin{aligned}
\lambda\text{-}SR_M(\mathfrak{q}, p(x)) &= \max_{p(u,v,w|x)} \big[\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)\big] \\
&\leq \max_{p(u,v,w|x)} \big[\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|U,W)\big] \\
&\leq \max_{p(u,w|x)} \big[\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(X;Z|U,W)\big] \\
&\overset{(a)}{\leq} \max_{p(u,w|x)} \big[\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(X;Y|U,W)\big] \\
&\leq \max_{p(w|x)} \big[\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(X;Y|W)\big] \\
&= I(X;Y) + (1-\lambda) \max_{p(w|x)} (I(W;Z) - I(W;Y)).
\end{aligned}$$

[3]For the definitions of less noisy broadcast channel or more capable broadcast channel, please refer to [11].

[4]We denote $X = \emptyset$ if the random variable X takes a constant value with probability one.

The inequality marked (a) is justified by the more-capable assumption. On the other hand setting $U = X$, $V = \emptyset$ shows that $\lambda\text{-}SR_M(\mathfrak{q}, p(x)) \geq I(X;Y) + (1-\lambda)\max_{p(w|x)}(I(W;Z) - I(W;Y))$. Hence $\lambda\text{-}SR_M(\mathfrak{q}, p(x))$ is linear in λ. $\square$

Remark 6. In each of the cases above, it is clear that the minimizing λ for the $\lambda\text{-}SR_M(\mathfrak{q})$ lies on $\lambda \in \{0, 1\}$. Thus the optimality of SR_M could be deduced alternately using Corollary 1.

4.3 Global optimizers of λ-SR

We now prove an important property regarding any global maximizer of (4).

Lemma 8. *Let $p_\lambda^*(u,v,w,x)$ be any maximizer of the expression in (4), then the following holds: For all (u,v,w,x) such that $p_\lambda^*(u,v,w,x) > 0$,*

$$\sum_{y,z} \mathfrak{q}(y,z|x) \log \frac{p_\lambda^*(uyw)p_\lambda^*(vzw)}{p_\lambda^*(uvw)(p_\lambda^*(z)p_\lambda^*(wy))^{(1-\lambda)}(p_\lambda^*(y)p_\lambda^*(wz))^{\lambda}} = \lambda\text{-}SR_M(\mathfrak{q}),$$

For all (u,v,w,x) such that $p_\lambda^*(u,v,w) > 0$ but $p_\lambda^*(u,v,w,x) = 0$,

$$\sum_{y,z} \mathfrak{q}(y,z|x) \log \frac{p_\lambda^*(uyw)p_\lambda^*(vzw)}{p_\lambda^*(uvw)(p_\lambda^*(z)p_\lambda^*(wy))^{(1-\lambda)}(p_\lambda^*(y)p_\lambda^*(wz))^{\lambda}} \leq \lambda\text{-}SR_M(\mathfrak{q}).$$

Proof. Let

$$\Psi_{p_\lambda^*}(u,v,w,x) = \sum_{y,z} \mathfrak{q}(y,z|x) \log \frac{p_\lambda^*(uyw)p_\lambda^*(vzw)}{p_\lambda^*(uvw)(p_\lambda^*(z)p_\lambda^*(wy))^{(1-\lambda)}(p_\lambda^*(y)p_\lambda^*(wz))^{\lambda}}.$$

We begin with the proof of the first statement. The proof follows from the first derivative condition that any local maximizer has to satisfy. Suppose $p_\lambda^*(u_1,v_1,w_1,x_1), p_\lambda^*(u_2,v_2,w_2,x_2) > 0$ be any two non-zero elements of $p_\lambda^*(u,v,w,x)$. Then, for $0 < \epsilon < \min\{p_\lambda^*(u_1,v_1,w_1,x_1), p_\lambda^*(u_2,v_2,w_2,x_2)\}$ define a new distribution

$$p_1(u,v,w,x) = \begin{cases} p_\lambda^*(u_1,v_1,w_1,x_1) - \epsilon, & (u,v,w,x) = (u_1,v_1,w_1,x_1) \\ p_\lambda^*(u_2,v_2,w_2,x_2) + \epsilon, & (u,v,w,x) = (u_2,v_2,w_2,x_2) \\ p_\lambda^*(u_2,v_2,w_2,x_2) & \text{otherwise} \end{cases}.$$

Expanding with respect to ϵ, one can see that

$$\lambda\text{-}SR_M(\mathfrak{q}, p_1) = \lambda\text{-}SR_M(\mathfrak{q}, p_\lambda^*) + \epsilon\big(\Psi_{p_\lambda^*}(u_2,v_2,w_2,x_2) - \Psi_{p_\lambda^*}(u_1,v_1,w_1,x_1)\big) + o(\epsilon),$$

where the notation $\lambda\text{-}SR_M(\mathfrak{q}, p_1)$ denotes the λ-SR evaluated that the distribution $p_1(u,v,w,x)$.

Hence for $p_\lambda^*(u,v,w,x)$ to be a local-maximum, it must be that $\Psi_{p_\lambda^*}(u_2,v_2,w_2,x_2) = \Psi_{p_\lambda^*}(u_1,v_1,w_1,x_1)$ whenever $p_\lambda^*(u_1,v_1,w_1,x_1), p_\lambda^*(u_2,v_2,w_2,x_2) > 0$. This in-particular implies that when $p_\lambda^*(u,v,w,x) > 0$ then $\Psi(p^*(u,v,w,x))$ takes a constant value. On the other hand note that

$$\sum_{u,v,w,x} p_\lambda^*(u,v,w,x)\Psi_{p_\lambda^*}(u,v,w,x) = \lambda\text{-}SR_M(\mathfrak{q}).$$

Hence when $p_\lambda^*(u,v,w,x) > 0$ then $\Psi_{p_\lambda^*}(u,v,w,x) = \lambda\text{-}SR_M(\mathfrak{q})$.

We continue with the proof of the second statement. Take some (u_0, v_0, w_0) such that $p^*_\lambda(u_0, v_0, w_0) > 0$. Let x_0, x_1 be such that $p^*_\lambda(u_0, v_0, w_0, x_0) > 0$ and $p^*_\lambda(u_0, v_0, w_0, x_1) = 0$. Take some $x_1 \neq x_0$. We would like to prove that $\Psi(u_0, v_0, w_0, x_1) \leq \lambda\text{-}SR_M(\mathfrak{q})$. Define a distribution according to

$$p_1(u, v, w, x) = \begin{cases} p^*_\lambda(u_1, v_1, w_1, x_0) - \epsilon, & (u, v, w, x) = (u_0, v_0, w_0, x_0) \\ \epsilon, & (u, v, w, x) = (u_0, v_0, w_0, x_1) \\ p^*_\lambda(u, v, w, x) & \text{otherwise} \end{cases}$$

where $0 < \epsilon < p^*_\lambda(u_1, v_1, w_1, x_0)$.

Expanding with respect to ϵ, one can see that

$$\lambda\text{-}SR_M(\mathfrak{q}, p_1) = \lambda\text{-}SR_M(\mathfrak{q}, p^*_\lambda) + \epsilon\big(\Psi_{p^*_\lambda}(u_0, v_0, w_0, x_1) - \Psi_{p^*_\lambda}(u_0, v_0, w_0, x_0)\big) + o(\epsilon).$$

Since for every $\epsilon > 0$ we have $\lambda\text{-}SR_M(\mathfrak{q}, p_1) \leq \lambda\text{-}SR_M(\mathfrak{q}, p^*_\lambda)$ it follows that

$$\Psi_{p_1}(u_0, v_0, w_0, x_1) \leq \Psi_{p^*_\lambda}(u_0, v_0, w_0, x_0) = \lambda\text{-}SR_M(\mathfrak{q})$$

as desired. □

Definition 5. Let $p(u, v, w, x)$ be a given distribution with $|\mathcal{W}| \leq |\mathcal{X}|$. Define $\mathcal{W} = \{1, ..., m\}, m \leq |\mathcal{X}|$ to be the alphabets taken by W and let $p^\epsilon_\mu(u, v, w, x)$ be a distribution defined according to

$$p^\epsilon_\mu(u, v, w, x) = \begin{cases} (1-\epsilon)p(u, v, w, x) & w \in \mathcal{W} \\ \epsilon\mu(u, v, x) & w = m+1 \end{cases},$$

where $\mu(u, v, x)$ is any probability distribution on $\mathcal{U} \times \mathcal{V} \times \mathcal{X}$. Observe that $d_{TV}(p, p_\lambda) = \epsilon$, where $d_{TV}(\cdot, \cdot)$ is the total-variation distance between probability distributions. We say that $p(u, v, w, x)$ is an *enhanced-local-maximum* if there is an $\epsilon > 0$ such that

$$\lambda\text{-}SR_M(\mathfrak{q}, p) \geq \lambda\text{-}SR_M(\mathfrak{q}, p^\epsilon_\mu), \forall \mu(u, v, x).$$

Remark 7. We use the term enhanced to denote that we allow the cardinality of W to increase by one. If the underlying space is the space of all probability distributions $p(u, v, w, x)$, then any local maximum is also an enhanced local maximum. However if the underlying space is considered to be the just the set of distributions that have the same support, then a local maximum need not be an enhanced local maximum.

Note that by expanding with respect to ϵ we obtain

$$\begin{aligned}\lambda\text{-}SR_M(\mathfrak{q}, p^\epsilon_\mu) = (1-\epsilon)\lambda\text{-}SR_M(\mathfrak{q}, p) + \epsilon \sum_{\mu(u,v,x)} \Psi_{1,\mu}(u, v, x) \\ + \epsilon \sum_{u,v,x,y,z} \mu(u, v, x)\mathfrak{q}(y, z|x) \log \frac{\mu(z)^{(1-\lambda)}\mu(y)^\lambda}{p(z)^{(1-\lambda)}p(y)^\lambda} + o(\epsilon),\end{aligned}$$

Here

$$\Psi_{1,\mu}(u, v, x) = \sum_{y,z} \mathfrak{q}(y, z|x) \log \frac{\mu(uy)\mu(vz)}{\mu(uv)\mu(z)\mu(y)}.$$

Therefore, for $p(u, v, w, x)$ to be an enhanced local maximum, it is necessary that

$$\lambda\text{-}SR_M(\mathfrak{q}, p) \geq \sum_{\mu(u,v,x)} \mu(u, v, x)\Psi_{1,\mu}(u, v, x) + \sum_{u,v,x,y,z} \mu(u, v, x)\mathfrak{q}(y, z|x) \log \frac{\mu(z)^{(1-\lambda)}\mu(y)^\lambda}{p(z)^{(1-\lambda)}p(y)^\lambda}. \quad (11)$$

Theorem 4. *Any enhanced local maximum of the λ-$SR_M(\mathfrak{q})$ is also a global maximum.*

Proof. The proof follows by contradiction. Let $p(u,v,w,x)$ be an enhanced local maximum and $p_\lambda^*(u,v,w,x)$ be any global maximum, such that $\lambda\text{-}SR_M(\mathfrak{q},p) < \lambda\text{-}SR_M(\mathfrak{q},p_\lambda^*) = \lambda\text{-}SR_M(\mathfrak{q})$.

Note that for every w,

$$\Psi_{p_\lambda^*}(u,v,w,x) = \Psi_{1,p_\lambda^*(u,v,x|w)}(u,v,x) + \sum_{y,z} \mathfrak{q}(y,z|x) \log \frac{p_\lambda^*(y|w)^\lambda p_\lambda^*(z|w)^{(1-\lambda)}}{p_\lambda^*(y)^\lambda p_\lambda^*(z)^{(1-\lambda)}}.$$

Since $p(u,v,w,x)$ is an enhanced local maximum it follows from (11) that for every w such that $p_\lambda^*(w) > 0$, we have

$$\begin{aligned}
&\lambda\text{-}SR_M(\mathfrak{q},p)\\
&\geq \sum_{u,v,x} p_\lambda^*(u,v,x|w)\Psi_{1,p_\lambda^*(u,v,x|w)}(u,v,x) + \sum_{u,v,x,y,z} p_\lambda^*(u,v,x|w)\mathfrak{q}(y,z|x)\log\frac{p_\lambda^*(z|w)^{(1-\lambda)}p_\lambda^*(y|w)^\lambda}{p(z)^{(1-\lambda)}p(y)^\lambda}\\
&= \sum_{u,v,x} p_\lambda^*(u,v,x|w)\Psi_{p_\lambda^*}(u,v,w,x) + \sum_{u,v,x,y,z} p_\lambda^*(u,v,x|w)\mathfrak{q}(y,z|x)\log\frac{p_\lambda^*(z)^{(1-\lambda)}p_\lambda^*(y)^\lambda}{p(z)^{(1-\lambda)}p(y)^\lambda}.
\end{aligned}$$

Now multiply both sides by $p_\lambda^*(w)$ and sum over w such that $p_\lambda^*(w) > 0$ to obtain

$$\begin{aligned}
\lambda\text{-}SR_M(\mathfrak{q},p) &\geq \sum_{u,v,w,x} p_\lambda^*(u,v,w,x)\Psi_{p_\lambda^*}(u,v,w,x) + \sum_{x,y,z} p_\lambda^*(x)\mathfrak{q}(y,z|x)\log\frac{p_\lambda^*(z)^{(1-\lambda)}p_\lambda^*(y)^\lambda}{p(z)^{(1-\lambda)}p(y)^\lambda}\\
&= \lambda\text{-}SR_M(\mathfrak{q},p_\lambda^*) + \lambda D(p_\lambda^*(y)||p(y)) + (1-\lambda)D(p_\lambda^*(z)||p(z))\\
&\geq \lambda\text{-}SR_M(\mathfrak{q},p_\lambda^*).
\end{aligned}$$

This yields a contradiction to the assumption that $\lambda\text{-}SR_M(\mathfrak{q},p) < \lambda\text{-}SR_M(\mathfrak{q},p_\lambda^*) = \lambda\text{-}SR_M(\mathfrak{q})$, and completes the proof. $\square$

Consider the two-letter expression corresponding to the λ-sum-rate in (4) given by

$$\begin{aligned}
\lambda\text{-}SR_{2M}(\mathfrak{q}) &= \frac{1}{2}\lambda\text{-}SR_M(\mathfrak{q}\times\mathfrak{q}) &(12)\\
&= \frac{1}{2}\max_{p(u,v,w,x_1,x_2)} \lambda I(W;Y_1,Y_2) + (1-\lambda)I(W;Z_1,Z_2) + I(U;Y_1,Y_2|W)\\
&\qquad + I(V;Z_1,Z_2|W) - I(U;V|W). &(13)
\end{aligned}$$

Lemma 5 implies that to compute the global maximum of the product channel $\mathfrak{q}\times\mathfrak{q}$ in (13) it suffices to consider $|\mathcal{U}|,|\mathcal{V}|,|\mathcal{W}| \leq |\mathcal{X}|^2$. If one wishes to verify $\lambda\text{-}SR_{2M}(q) = \lambda\text{-}SR_M(q)$, then computations can be significantly reduced by using the following corollary to Theorem 4.

Corollary 2. *To verify that $\lambda\text{-}SR_{2M}(\mathfrak{q}) = \lambda\text{-}SR_M(\mathfrak{q})$ it suffices to verify that the product distribution $p_\lambda(\tilde{u},\tilde{v},\tilde{w},x_1,x_2) = p_\lambda^*(u_1,v_1,w_1,x_1)p_\lambda^*(u_2,v_2,w_2,x_2)$, where $p_\lambda^*(u,v,w,x)$ maximizes $\lambda\text{-}SR_M(\mathfrak{q})$, is an enhanced local maximum.*

We consider the two-letter factorization of the $\lambda\text{-}SR_M(\mathfrak{q})$ in this paper than the Marton's sum rate. The reason for this will become clear in the next section. However the following simple lemma shows that if the two-letter $\lambda\text{-}SR_M(\mathfrak{q})$ factorizes then so does $SR_M(\mathfrak{q})$.

Corollary 3. *If $\lambda\text{-}SR_{2M}(\mathfrak{q}) = \lambda\text{-}SR_M(\mathfrak{q})$ for all $\lambda \in [0,1]$ then $SR_{2M}(\mathfrak{q}) = 2SR_M(\mathfrak{q})$.*

Proof. The proof is immediate from Lemma 1. Observe that, under the assumption $\lambda\text{-}SR_{2M}(\mathfrak{q}) = \lambda\text{-}SR_M(\mathfrak{q})$, we have

$$SR_{2M}(\mathfrak{q}) = \min_{\lambda\in[0,1]} \lambda\text{-}SR_{2M}(\mathfrak{q}) = \min_{\lambda\in[0,1]} \lambda\text{-}SR_M(\mathfrak{q}) = SR_M(\mathfrak{q}).$$

□

4.4 On SR_M for product channels

In this section we consider the behavior of $SR_M(\mathfrak{q})$ for the product of two non-identical channels. An interested reader may wonder why we considered *factorization of* $\lambda\text{-}SR_M(\mathfrak{q})$ as opposed to *factorization of* $SR_M(\mathfrak{q})$. Indeed we will show that there are channels $\mathfrak{q}_1, \mathfrak{q}_2$ such that

$$SR_M(\mathfrak{q}_1 \times \mathfrak{q}_2) > SR_M(\mathfrak{q}_1) + SR_M(\mathfrak{q}_2).$$

Lemma 9. *Let $p = 0.1, e = H(0.1) = \log_2 10 - 0.9 \log_2 9$. Consider a product channel formed by the following components: Let the channels $X_1 \to Y_1$ and $X_2 \to Z_2$ be $BEC(e)$ and the channels $X_1 \to Z_1$ and $X_2 \to Y_2$ be $BSC(p)$. For this product channel*

$$SR(\mathfrak{q}_1 \times \mathfrak{q}_2) > SR_M(\mathfrak{q}_1) + SR_M(\mathfrak{q}_2).$$

Proof. From [12], since $1 - e = 1 - H(p)$ we know that Y_1 is more capable than Z_1 and Z_2 is more capable than Y_2. Thus, from Theorem 3, we know that Marton's inner bound is optimal for this channel. Hence from Lemma 1 and Theorem 1, we have that

$$SR(\mathfrak{q}_1 \times \mathfrak{q}_2) = \min_{\lambda\in[0,1]} \lambda\text{-}SR_M(\mathfrak{q}_1) + \lambda\text{-}SR_M(\mathfrak{q}_2).$$

By the skew-symmetry we know that $\lambda\text{-}SR_M(\mathfrak{q}_2) = (1-\lambda)\text{-}SR_M(\mathfrak{q}_1)$. Further, from the symmetry, it is easy to show that it suffices to consider $\mathrm{P}(X = 0) = \frac{1}{2}$ to compute $\lambda\text{-}SR_M(\mathfrak{q}_1)$. In particular one can show that

$$\lambda\text{-}SR_M(\mathfrak{q}_1) = C + (1 - \lambda)d^*,$$

where C is the common capacity of the $BSC(p)$ and $BEC(e)$, and $d^* = \max_{p(x)} I(X;Y) - I(X;Z)$. For the chosen parameters $d^* \approx 0.03877$. The maximum sum rate of the channel $\mathfrak{q}_1(y_1, z_1|x_1)$, since Y_1 is more capable than Z_1, is given by the capacity to receiver Y_1; hence $SR_M(\mathfrak{q}_1) = C$, the common capacity.

Thus $SR(\mathfrak{q}_1 \times \mathfrak{q}_2) - SR_M(\mathfrak{q}_1) - SR_M(\mathfrak{q}_2)$ is given by

$$\min_{\lambda\in[0,1]} \big(C + (1-\lambda)d^* + C + \lambda d^*\big) - C - C = d^* > 0.$$

□

5 Randomized time-division strategy

Randomized time-division refers to a strategy that generalizes the simple time-division strategy. In time-division, the sender X transmits exclusively to receiver Y for a predetermined α fraction of the time, and transmits exclusively to receiver Z for the remaining $(1-\alpha)$ fraction of the time. In randomized time-division, the sender chooses the α fraction of the time that it wants to transmit to Y using a codebook, thus conveying some commonly decodable information to the receivers when they decode the proper $(\alpha, 1-\alpha)$ division of slots. This strategy can be shown to improve on naive time division for some broadcast channels. For more details, an interested reader reader can refer to [2, pg. 216].

This is a indeed a special (and much simpler) instance of Marton's coding strategy that sets $U = X, V = \emptyset$ when $W \in \mathcal{A}$ and $V = X, U = \emptyset$ when $W = \mathcal{A}^c$. This strategy yields a λ-sum-rate given by

$$\lambda\text{-}SR_{RTD}(\mathfrak{q}) = \max_{p(w,x)} \lambda I(W;Y) + (1-\lambda) I(W;Z) + \sum_{w \in \mathcal{A}} \mathrm{P}(W=w) I(X;Y|W=w) + \sum_{w \in \mathcal{A}^c} \mathrm{P}(W=w) I(X;Z|W=w).$$

Using standard arguments it follows that it suffices to consider $|\mathcal{W}| \le |\mathcal{X}|$ to compute the λ-sum-rate.

It was shown [13] that for all binary input broadcast channels the sum rate obtained using the simple randomized time division strategy matches the sum rate obtained using Marton's coding strategy, i.e. $SR_M(\mathfrak{q}) = SR_{RTD}(\mathfrak{q})$ when $|\mathcal{X}| = 2$. This result is based on the inequality that whenever $|\mathcal{X}| = 2$ and $(U,V) \to X \to (Y,Z)$ is Markov we have

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

Using this inequality it also immediately follows that $\lambda\text{-}SR_{RTD}(\mathfrak{q}) = \lambda\text{-}SR_M(\mathfrak{q})$.

For product of two channels $\mathfrak{q}_1 \times \mathfrak{q}_2$ one can define a slight generalization of the RTD strategy (equivalently this is a natural generalization of RTD for the 2-letter channel $\mathfrak{q} \times \mathfrak{q}$). This is again a special instance of Marton's coding strategy that sets

$$(U,V) := \begin{cases} U = (X_1 X_2), V = \emptyset & w \in \mathcal{A}_1 \\ U = X_1, V = X_2 & w \in \mathcal{A}_2 \\ U = X_2, V = X_1 & w \in \mathcal{A}_3 \\ U = \emptyset, V = (X_1, X_2) & w \in \mathcal{A}_4 \end{cases},$$

where $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4$ denotes a partition of $\mathcal{W}$. Let this scheme be called 2-RTD. We define

$$\begin{aligned} &\lambda\text{-}SR_{2\text{-}RTD}(\mathfrak{q}_1 \times \mathfrak{q}_2) \\ &\quad = \max_{p(w,x_1,x_2)} \lambda I(W;Y_1,Y_2) + (1-\lambda) I(W;Z_1,Z_2) + \sum_{w \in \mathcal{A}_1} \mathrm{P}(W=w) I(X_1,X_2;Y_1,Y_2|W=w) \\ &\qquad + \sum_{w \in \mathcal{A}_2} \mathrm{P}(W=w) \big(I(X_1;Y_1,Y_2|W=w) + I(X_2;Z_1,Z_2|W=w) - I(X_1;X_2|W=w) \big) \\ &\qquad + \sum_{w \in \mathcal{A}_3} \mathrm{P}(W=w) \big(I(X_1;Y_1,Y_2|W=w) + I(X_2;Z_1,Z_2|W=w) - I(X_1;X_2|W=w) \big) \\ &\qquad + \sum_{w \in \mathcal{A}_4} \mathrm{P}(W=w) I(X_1,X_2;Z_1,Z_2|W=w). \end{aligned}$$

Similarly define

$$
\begin{aligned}
&SR_{2\text{-}RTD}(\mathfrak{q}_1 \times \mathfrak{q}_2) \\
&\quad = \max_{p(w,x_1,x_2)} \min\{I(W;Y_1,Y_2), I(W;Z_1,Z_2)\} + \sum_{w \in \mathcal{A}_1} \mathrm{P}(W=w) I(X_1,X_2;Y_1,Y_2|W=w) \\
&\qquad + \sum_{w \in \mathcal{A}_2} \mathrm{P}(W=w)\big(I(X_1;Y_1,Y_2|W=w) + I(X_2;Z_1,Z_2|W=w) - I(X_1;X_2|W=w)\big) \\
&\qquad + \sum_{w \in \mathcal{A}_3} \mathrm{P}(W=w)\big(I(X_1;Y_1,Y_2|W=w) + I(X_2;Z_1,Z_2|W=w) - I(X_1;X_2|W=w)\big) \\
&\qquad + \sum_{w \in \mathcal{A}_4} \mathrm{P}(W=w) I(X_1,X_2;Z_1,Z_2|W=w).
\end{aligned}
$$

In a similar fashion to the proof of Lemma 1 one can show the following lemma.

Lemma 10. *The following holds:*

$$\min_{\lambda \in [0,1]} \lambda\text{-}SR_{2\text{-}RTD}(\mathfrak{q} \times \mathfrak{q}) = SR_{2\text{-}RTD}(\mathfrak{q}).$$

The proof is given in Appendix D.

Remark 8. Suppose there is a binary input channel $\mathfrak{q}(y,z|x)$ such that $SR_{2\text{-}RTD}(\mathfrak{q}\times\mathfrak{q}) > 2SR_{RTD}(\mathfrak{q})$ then it would immediately imply that

$$SR_{2M}(\mathfrak{q}) \geq \frac{1}{2} SR_{2\text{-}RTD}(\mathfrak{q} \times \mathfrak{q}) > SR_{RTD}(\mathfrak{q}) = SR_M(\mathfrak{q}),$$

where the last equality follows from the result about binary input broadcast channels. This would have been an easy technique to establish the strict sub-optimality of Marton's coding scheme if it had worked. However the next lemma shows that this cannot happen. Indeed we show that $\lambda\text{-}SR_{2\text{-}RTD}(\mathfrak{q}_1 \times \mathfrak{q}_2) = \lambda\text{-}SR_{RTD}(\mathfrak{q}_1) + \lambda\text{-}SR_{RTD}(\mathfrak{q}_2)$ for channels with arbitrary input cardinality. Hence from Lemma 10 it will immediately follow that $SR_{2\text{-}RTD}(\mathfrak{q} \times \mathfrak{q}) = 2SR_{RTD}(\mathfrak{q})$.

Theorem 5. *The following holds:*

$$\lambda\text{-}SR_{2\text{-}RTD}(\mathfrak{q}_1 \times \mathfrak{q}_2) = \lambda\text{-}SR_{RTD}(\mathfrak{q}_1) + \lambda\text{-}SR_{RTD}(\mathfrak{q}_2).$$

Proof. By taking the product of the optimizing distributions for $\lambda\text{-}SR_{RTD}(\mathfrak{q}_1), \lambda\text{-}SR_{RTD}(\mathfrak{q}_2)$ one can immediately see that

$$\lambda\text{-}SR_{2\text{-}RTD}(\mathfrak{q}_1 \times \mathfrak{q}_2) \geq \lambda\text{-}SR_{RTD}(\mathfrak{q}_1) + \lambda\text{-}SR_{RTD}(\mathfrak{q}_2).$$

Hence it suffices to show that

$$\lambda\text{-}SR_{2\text{-}RTD}(\mathfrak{q}_1 \times \mathfrak{q}_2) \leq \lambda\text{-}SR_{RTD}(\mathfrak{q}_1) + \lambda\text{-}SR_{RTD}(\mathfrak{q}_2).$$

Observe that

$$
\begin{aligned}
&\lambda I(W;Y_1,Y_2)+(1-\lambda)I(W;Z_1,Z_2)+\sum_{w\in\mathcal{A}_1}\mathrm{P}(W=w)I(X_1,X_2;Y_1,Y_2|W=w)\\
&\quad+\sum_{w\in\mathcal{A}_2}\mathrm{P}(W=w)\big(I(X_1;Y_1,Y_2|W=w)+I(X_2;Z_1,Z_2|W=w)-I(X_1;X_2|W=w)\big)\\
&\quad+\sum_{w\in\mathcal{A}_3}\mathrm{P}(W=w)\big(I(X_1;Y_1,Y_2|W=w)+I(X_2;Z_1,Z_2|W=w)-I(X_1;X_2|W=w)\big)\\
&\quad+\sum_{w\in\mathcal{A}_4}\mathrm{P}(W=w)I(X_1,X_2;Z_1,Z_2|W=w)\\
&=\lambda H(Y_1,Y_2)+(1-\lambda)H(Z_1,Z_2)\\
&\quad+\sum_{w\in\mathcal{A}_1}\mathrm{P}(W=w)\big(I(X_1,X_2;Y_1,Y_2|W=w)-\lambda H(Y_1,Y_2|W=w)-(1-\lambda)H(Z_1,Z_2|W=w)\big)\\
&\quad+\sum_{w\in\mathcal{A}_2}\mathrm{P}(W=w)\big(I(X_1;Y_1,Y_2|W=w)+I(X_2;Z_1,Z_2|W=w)-I(X_1;X_2|W=w)\\
&\qquad\qquad-\lambda H(Y_1,Y_2|W=w)-(1-\lambda)H(Z_1,Z_2|W=w)\big) \qquad\qquad (14)\\
&\quad+\sum_{w\in\mathcal{A}_3}\mathrm{P}(W=w)\big(I(X_1;Y_1,Y_2|W=w)+I(X_2;Z_1,Z_2|W=w)-I(X_1;X_2|W=w)\\
&\qquad\qquad-\lambda H(Y_1,Y_2|W=w)-(1-\lambda)H(Z_1,Z_2|W=w)\big)\\
&\quad+\sum_{w\in\mathcal{A}_4}\mathrm{P}(W=w)\big(I(X_1,X_2;Z_1,Z_2|W=w)-\lambda H(Y_1,Y_2|W=w)-(1-\lambda)H(Z_1,Z_2|W=w)\big).
\end{aligned}
$$

The idea of the proof is to factorize each of the four summation terms in (14) separately.

Consider the following manipulations of the terms.

$$
\begin{aligned}
&I(X_1,X_2;Y_1,Y_2|W=w)-\lambda H(Y_1,Y_2|W=w)-(1-\lambda)H(Z_1,Z_2|W=w)\\
&\quad=I(X_1;Y_1|W=w,Y_2)+I(X_2;Y_2|W=w,Z_1)-\lambda H(Y_1|W=w,Y_2) \qquad (15)\\
&\qquad-\lambda H(Y_2|W=w,Z_1)-(1-\lambda)H(Z_1|W=w,Y_2)-(1-\lambda)H(Z_2|W=w,Z_1),
\end{aligned}
$$

$$
\begin{aligned}
&I(X_1;Y_1,Y_2|W=w)+I(X_2;Z_1,Z_2|W=w)-I(X_1;X_2|W=w)\\
&\qquad-\lambda H(Y_1,Y_2|W=w)-(1-\lambda)H(Z_1,Z_2|W=w)\\
&\quad=I(X_1;Y_1|W=w,Y_2)+I(X_2;Z_2|W=w,Z_1)-\lambda H(Y_1|W=w,Y_2)\\
&\qquad-\lambda H(Y_2|W=w,Z_1)-(1-\lambda)H(Z_1|W=w,Y_2)-(1-\lambda)H(Z_2|W=w,Z_1)\\
&\qquad+I(X_2;Z_1|W=w)+I(X_1;Y_2|W=w,Z_1)-I(X_1;X_2|W=w)\\
&\quad\le I(X_1;Y_1|W=w,Y_2)+I(X_2;Z_2|W=w,Z_1)-\lambda H(Y_1|W=w,Y_2) \qquad (16)\\
&\qquad-\lambda H(Y_2|W=w,Z_1)-(1-\lambda)H(Z_1|W=w,Y_2)-(1-\lambda)H(Z_2|W=w,Z_1),
\end{aligned}
$$

where the last inequality follows since $I(X_1;X_2|W=w)=I(Z_1,X_1;X_2|W=w)=I(Z_1;X_2|W=w)+I(X_1;X_2|W=w,Z_1)\ge I(Z_1;X_2|W=w)+I(X_1;Y_2|W=w,Z_1)$. Here we use the fact that $(W,X_2)\to X_1\to Z_1$ is Markov and $(X_1,Z_1,W)\to X_2\to Y_2$ is Markov.

In a similar fashion we have

$$
\begin{aligned}
&I(X_2;Y_1,Y_2|W=w)+I(X_1;Z_1,Z_2|W=w)-I(X_1;X_2|W=w)\\
&\quad\le I(X_1;Z_1|W=w,Z_2)+I(X_2;Y_2|W=w,Y_1)-\lambda H(Y_1|W=w,Z_2) \qquad (17)\\
&\qquad-\lambda H(Y_2|W=w,Y_1)-(1-\lambda)H(Z_1|W=w,Z_2)-(1-\lambda)H(Z_2|W=w,Y_1)
\end{aligned}
$$

Finally

$$
\begin{aligned}
&I(X_1, X_2; Z_1, Z_2|W = w) - \lambda H(Y_1, Y_2|W = w) - (1 - \lambda)H(Z_1, Z_2|W = w)\\
&\quad = I(X_1; Z_1|W = w, Z_2) + I(X_2; Z_2|W = w, Y_1) - \lambda H(Y_1|W = w, Z_2) \\
&\qquad - \lambda H(Y_2|W = w, Y_1) - (1 - \lambda)H(Z_1|W = w, Z_2) - (1 - \lambda)H(Z_2|W = w, Y_1)
\end{aligned} \tag{18}
$$

Define new random variables W_1, W_2 having alphabets given by

$$
\mathcal{W}_1 = \begin{cases} (w, z_2) & w \in \mathcal{A}_1 \cup \mathcal{A}_2, z_2 \in \mathcal{Z} \\ (w, y_2) & w \in \mathcal{A}_3 \cup \mathcal{A}_4, y_2 \in \mathcal{Y} \end{cases} \quad \text{and} \quad \mathcal{W}_2 = \begin{cases} (w, y_1) & w \in \mathcal{A}_1 \cup \mathcal{A}_2, y_1 \in \mathcal{Y} \\ (w, z_1) & w \in \mathcal{A}_3 \cup \mathcal{A}_4, z_1 \in \mathcal{Z} \end{cases} .
$$

Further partition $\mathcal{W}_1$ into two sets $\mathcal{B}$ and $\mathcal{B}^c$ according to $\mathcal{B} = \{(w, z_2) : w \in \mathcal{A}_1 \cup \mathcal{A}_2, z_2 \in \mathcal{Z}\}$, and partition $\mathcal{W}_2$ into two sets $\mathcal{C}$ and $\mathcal{C}^c$ according to $\mathcal{C} = \{(w, y_1) : w \in \mathcal{A}_1, y_1 \in \mathcal{Y}\} \cup \{(w, z_1) : w \in \mathcal{A}_3, z_1 \in \mathcal{Z}\}$.

Using (15), (16), (17), (18), and the definitions of $W_1, W_2, \mathcal{B}, \mathcal{C}$ we can bound the expression in (14) by

$$
\begin{aligned}
&\lambda I(W_1; Y_1) + (1 - \lambda)I(W_1; Z_1) + \sum_{w_1 \in \mathcal{B}} \mathrm{P}(W_1 = w_1)I(X_1; Y_1|W_1 = w_1) \\
&\quad + \sum_{w_1 \in \mathcal{B}^c} \mathrm{P}(W_1 = w_1)I(X_1; Z_1|W_1 = w_1) + \lambda I(W_2; Y_2) + (1 - \lambda)I(W_2; Z_2) \\
&\quad + \sum_{w_2 \in \mathcal{C}} \mathrm{P}(W_2 = w_2)I(X_2; Y_2|W_2 = w_2) + \sum_{w_2 \in \mathcal{C}^c} \mathrm{P}(W_2 = w_2)I(X_2; Z_2|W_2 = w_2) \\
&\le \lambda\text{-}SR_{RTD}(\mathfrak{q}_1) + \lambda\text{-}SR_{RTD}(\mathfrak{q}_2).
\end{aligned}
$$

This implies that

$$
\lambda\text{-}SR_{2\text{-}RTD}(\mathfrak{q}_1 \times \mathfrak{q}_2) \le \lambda\text{-}SR_{RTD}(\mathfrak{q}_1) + \lambda\text{-}SR_{RTD}(\mathfrak{q}_2),
$$

and completes the proof of the Lemma. □

Remark 9. We wish to bring following unique feature to this proof to the attention of the readers: in identifying the auxiliaries W_1, W_2 in terms of W, past or future of Z, past or future of Y, we actually chose different terms depending on $w \in \mathcal{W}$. This is a freedom that has never been exploited before (to the best of the knowledge of the authors). A consistent choice does not seem to work here.

6 Conclusion

In this paper we show a variety of results related to Marton's inner bound and its optimality. We also show that the tightest known outer bound is strictly sub-optimal. An outer bound is presented for product broadcast channels which is then shown to coincide with Marton's inner bound for classes of channels whose capacity regions were previously unknown. This outer bound turns out to be a strict improvement over the previously known tightest outer bound for product broadcast channels. It would be very interesting to extend this outer bound to non-product channels in a natural way. Further a variety of other interesting results are also established which aid in the computation of Marton's inner bound.

References

[1] T. Cover, "Broadcast channels," *IEEE Trans. Info. Theory*, vol. IT-18, pp. 2–14, January, 1972.

[2] A. El Gamal and Y.-H. Kim, *Network Information Theory.* Cambridge University Press, 2012.

[3] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Info. Theory*, vol. IT-25, pp. 306–311, May, 1979.

[4] C. Nair and A. El Gamal, "An outer bound to the capacity region of the broadcast channel," *IEEE Trans. Info. Theory*, vol. IT-53, pp. 350–355, January, 2007.

[5] C. Nair and Z. V. Wang, "On the inner and outer bounds for 2-receiver discrete memoryless broadcast channels," *Proceedings of the ITA Workshop*, 2008, cs.IT/0804.3825.

[6] A. Gohari and V. Anantharam, "Evaluation of Marton's inner bound for the general broadcast channel," *International Symposium on Information Theory*, pp. 2462–2466, 2009.

[7] V. Jog and C. Nair, "An information inequality for the bssc channel," *Proceedings of the ITA Workshop*, 2010.

[8] C. Nair, "A note on outer bounds for broadcast channel," *Presented at International Zurich Seminar*, 2010, http://arXiv.org/abs/1101.0640.

[9] A. Gohari, A. El Gamal, and V. Anantharam, "On an outer bound and an inner bound for the general broadcast channel," *International Symposium on Information Theory*, 2010, arxiv.org/abs/1006.5166.

[10] F. Terkelsen, "Some minimax theorems," *Mathematica Scandinavica*, vol. 31, pp. 405–413, 1972.

[11] A. El Gamal and Y.-H. Kim, *Network Information Theory.* Cambridge University Pres, 2011.

[12] C. Nair, "Capacity regions of two new classes of two-receiver broadcast channels," *Information Theory, IEEE Transactions on*, vol. 56, pp. 4207–4214, sep. 2010.

[13] C. Nair, Z. V. Wang, and Y. Geng, "An information inequality and evaluation of Marton's inner bound for binary input broadcast channels," *International Symposium on Information Theory*, 2010.

A A Min-Max Theorem

Theorem 6 (Theorem 3 of [10])**.** *Let X be a compact connected space, let Y be a set, and let $f : X \times Y \mapsto \mathbb{R}$ be a function satisfying:*

(i) For any $y_1, y_2 \in Y$ there exists $y_0 \in Y$ such that

$$f(x, y_0) \geq \frac{1}{2}\left(f(x, y_1) + f(x, y_2)\right), \forall x \in X.$$

(*ii*) *Every finite intersection of sets of the form* $\{x \in X : f(x, y) \leq \alpha)\}$ *with* $(y, \alpha) \in Y \times R$ *is closed and connected.*

Then

$$\sup_{y \in Y} \min_{x \in X} f(x, y) = \min_{x \in X} \sup_{y \in Y} f(x, y).$$

We now present a Corollary of the above theorem that can be potentially used in many information theory scenarios.

Corollary 4. *Let* Λ_d *be the d-dimensional simplex, i.e.* $\lambda_i \geq 0$ *and* $\sum_{i=1}^{d} \lambda_i = 1$*. Let* $\mathcal{P}$ *be a set of probability distributions* $p(u)$*. Let* $T_i(p(u)), i = 1, .., d$ *be a set of functions such that the set* $\mathcal{A}$*, defined by*

$$\mathcal{A} = \{(a_1, a_2, ..., a_d) \in \mathbb{R}^d : a_i \leq T_i(p(u)) \text{ for some } \ p(u) \in \mathcal{P}\},$$

is a convex set.

Then

$$\sup_{p(u) \in \mathcal{P}} \min_{\lambda \in \Lambda_d} \sum_{i=1}^{d} \lambda_i T_i(p(u)) = \min_{\lambda \in \Lambda_d} \sup_{p(u) \in \mathcal{P}} \sum_{i=1}^{d} \lambda_i T_i(p(u)).$$

Proof. Let $f(\lambda, p(u)) = \sum_{i=1}^{d} \lambda_i T_i(p(u))$. It suffices to verify that $f(\lambda, p(u))$ satisfies the conditions of Theorem 6. Since the set $\mathcal{A}$ is convex, we know that for any $p_1(u), p_2(u) \in \mathcal{P}$ we have a distribution $p_c(u) \in \mathcal{P}$ such that

$$T_i(p_c(u)) \geq \frac{1}{2}\big(T_i(p_1(u)) + T_i(p_2(u))\big), i = 1, ..., d.$$

Hence (using linearity in λ and non-negativity of λ_i) we have

$$f(\lambda, p_c(u)) \geq \frac{1}{2}\big(f(\lambda, p_1(u)) + f(\lambda, p_2(u))\big), \forall \lambda \in \Lambda_d.$$

Since $f(\lambda, p(u))$ is a linear function of λ, it is immediate that the set

$$\mathcal{B}(p(u), \alpha) = \{\lambda \in \Lambda_d : f(\lambda, p(u)) \leq \alpha\}$$

is closed for every pair $(p(u), \alpha) \in \mathcal{P} \times \mathbb{R}$. Further, due to the linearity in λ, if $\lambda_1, \lambda_2 \in \mathcal{B}(p(u), \alpha)$, then the line segment joining λ_1 and λ_2 belongs to $\mathcal{B}(p(u), \alpha)$. This implies that a finite intersection of sets, each containing λ_1 and λ_2 will also contain the line segment joining λ_1 and λ_2, showing that the finite intersection will be connected. Therefore finite intersections of the sets of the form $\mathcal{B}(p(u), \alpha)$ are closed and connected. Thus the Corollary 4 follows from Theorem 6. $\square$

We will now show how one can use the Corollary 4 to establish Lemma 1.

Proof. (Proof of Lemma 1) It is clear that

$$\begin{aligned}\max_{p(u,v,w,x)} \min_{\lambda \in [0,1]} \lambda\text{-}SR_M(\mathfrak{q}, p(u, v, w, x)) &\leq \max_{p(x)} \min_{\lambda \in [0,1]} \max_{p(u,v,w|x)} \lambda\text{-}SR_M(\mathfrak{q}, p(u, v, w, x)) \\ &\leq \min_{\lambda \in [0,1]} \max_{p(u,v,w,x)} \lambda\text{-}SR_M(\mathfrak{q}, p(u, v, w, x)).\end{aligned}$$

Therefore suffices to show that

$$\max_{p(u,v,w,x)} \min_{\lambda \in [0,1]} \lambda\text{-}SR_M(\mathfrak{q}, p(u, v, w, x)) = \min_{\lambda \in [0,1]} \max_{p(u,v,w,x)} \lambda\text{-}SR_M(\mathfrak{q}, p(u, v, w, x)).$$

Here we take $d = 2$ and set

$$T_1(p(u,v,w,x)) = I(W;Y) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$$
$$T_2(p(u,v,w,x)) = I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$$

It is clear that the set

$$\mathcal{A} = \{(a_1, a_2) : a_1 \le T_1(p(u,v,w,x)), a_2 \le T_2(p(u,v,w,x))\}$$

is a convex set. (In the standard manner, choose $\tilde{W} = (W, Q)$, and when $Q = 0$ choose $(U, V, W, X) \sim p_1(u,v,w,x)$ and $Q = 1$ choose $(U, V, W, X) \sim p_2(u,v,w,x)$). Hence from Corollary 4, we have the proof of Lemma 1. □

Remark 10. The proof of this claim in section 3.1.1 of [9] is very similar in flavor and uses the convexity of the set $\mathcal{A}$. However here we recover it from an application of some general theorems, and this technique and Corollary 4 may be helpful in other situations as well.

B Computing λ-SR_M for the semi-deterministic channel in Fig. 2

B.1 Maximum of λ-SR_M is obtained at the uniform input distribution

Consider the semi-deterministic channel of Figure 2. In this appendix we show that for any $\lambda \in [0, 1]$, λ-$SR_M(\mathfrak{q}, p(x))$ is less than or equal to λ-$SR_M(\mathfrak{q}, u(x))$ where u is the uniform distribution on $\mathcal{X}$.

Note that λ-$SR_M(\mathfrak{q}, p(x))$ is concave in $p(x)$. To see this, take two marginal distributions $p_0(x)$ and $p_1(x)$, and assume that (U_0, V_0, W_0, X_0) and (U_1, V_1, W_1, X_1) are two set of random variables maximizing the expressions of λ-$SR_M(\mathfrak{q}, p_0(x))$ and λ-$SR_M(\mathfrak{q}, p_1(x))$ respectively. Take a uniform binary random variable Q, independent of all previously defined random variables. Let $U = U_Q$, $V = V_Q$, $W = (W_Q, Q)$, $X = X_Q$. Observe that X is distributed according to $\frac{p_0(x)}{2} + \frac{p_1(x)}{2}$, and furthermore if we compute the expression in λ-$SR_M(\mathfrak{q}, p(x))$ for (U, V, W, X), we get a value that is greater than or equal to the average of the corresponding values for (U_0, V_0, W_0, X_0) and (U_1, V_1, W_1, X_1). Therefore

$$\lambda\text{-}SR_M(\mathfrak{q}, p_0(x)) + \lambda\text{-}SR_M(\mathfrak{q}, p_1(x)) \le \lambda\text{-}SR_M(\mathfrak{q}, \frac{p_0(x)}{2} + \frac{p_1(x)}{2}).$$

Thus, λ-$SR_M(\mathfrak{q}, p(x))$ is concave in $p(x)$.

Take an arbitrary $p(x)$ of the form $\mathrm{P}(X) = (a, b, c, d)$. Because of the symmetries of the channel of Figure 2 with respect to the two receivers, we have

$$\begin{aligned}\lambda\text{-}SR_M(\mathfrak{q}, p(x) \sim (a,b,c,d)) &= \lambda\text{-}SR_M(\mathfrak{q}, p(x) \sim (b,a,d,c)) \\ &= \lambda\text{-}SR_M(\mathfrak{q}, p(x) \sim (c,d,a,b)) \\ &= \lambda\text{-}SR_M(\mathfrak{q}, p(x) \sim (d,c,b,a)).\end{aligned}$$

Here we have used the symmetry between inputs 1 and 2, and the symmetry between inputs 3 and

4, and the symmetry between the pair of inputs $(1,2)$ and $(3,4)$. Using the concavity of F, we have

$$\begin{aligned}
4\lambda\text{-}SR_M(\mathfrak{q}, p(x) \sim (a,b,c,d)) &= \lambda\text{-}SR_M(\mathfrak{q}, p(x) \sim (a,b,c,d)) + \lambda\text{-}SR_M(\mathfrak{q}, p(x) \sim (b,a,d,c)) + \\
&\quad \lambda\text{-}SR_M(\mathfrak{q}, p(x) \sim (c,d,a,b)) + \lambda\text{-}SR_M(\mathfrak{q}, p(x) \sim (d,c,b,a)) \\
&\le 4\lambda\text{-}SR_M(\mathfrak{q}, p(x) \sim (\frac{1}{4}(a+b+c+d), \frac{1}{4}(a+b+c+d), \\
&\qquad\qquad \frac{1}{4}(a+b+c+d), \frac{1}{4}(a+b+c+d))) \\
&= 4\lambda\text{-}SR_M(\mathfrak{q}, u(x)).
\end{aligned}$$

B.2 Computing the λ-sum-rate at the uniform input distribution

In this appendix we compute $\lambda\text{-}SR_M(\mathfrak{q}, u(x))$ at the uniform input distribution for the semi-deterministic channel given in Figure 2.

Claim 5. *The $\lambda \mapsto \lambda\text{-}SR_M(\mathfrak{q}, u(x))$ curve for the channel under consideration consists of two lines,*

$$\lambda\text{-}SR_M(\mathfrak{q}, u(x)) = \begin{cases} \frac{5}{3} - \frac{2}{3}\lambda & \lambda \in [0, \frac{1}{2}] \\ \frac{4}{3} & \lambda \in [\frac{1}{2}, 1] \end{cases} .$$

Proof. Note that

$$\begin{aligned}
\lambda\text{-}SR_M(\mathfrak{q}, u(x)) &= \max_{p(u,v,w|x)} \big\{\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)\big\} \\
&= \max_{p(u,w|x)} \big\{\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + H(Z|UW)\big\}.
\end{aligned}$$

In the last step we have used the inequality $I(V;Z|W) - I(U;V|W) \le H(Z|UW)$ together with the fact that $I(Z;Z|W) - I(U;Z|W) = H(Z|UW)$. Therefore $\lambda\text{-}SR_M(\mathfrak{q}, u(x))$ can be written as

$$\max_{p(u,w|x)} \big\{\lambda H(Y) + (1-\lambda)H(Z) + (1-\lambda)\big(H(Y|W) - H(Z|W)\big) + H(Z|UW) - H(Y|UW)\big\},$$

which is equal to

$$\begin{aligned}
&\lambda H(Y) + (1-\lambda)H(Z) + \max_{p(w|x)} \big\{(1-\lambda)\big(H(Y|W) - H(Z|W)\big) + \\
&\max_{p(u|w,x)} \big(H(Z|UW) - H(Y|UW)\big)\big\}. \qquad (19)
\end{aligned}$$

Let $\mathrm{P}(X|W=i) = (a_i, b_i, c_i, d_i)$, and $f(a_i, b_i, c_i, d_i) = \max_{p(u|x)} H(Z|U) - H(Y|U)$ conditioned on $\mathrm{P}(X) = (a_i, b_i, c_i, d_i)$. Observe that f is concave. The argument is similar to the one given above in the first part of this appendix and we will not repeat it here. Further, observe that $f(a_i, b_i, c_i, d_i) = f(b_i, a_i, d_i, c_i)$ because the symmetry between inputs 1 and 2, and the symmetry between inputs 3 and 4.

Consider the transformation $(a_i, b_i, c_i, d_i) \to (b_i, a_i, d_i, c_i)$, for all i while leaving $\mathrm{P}(W=i)$ unchanged. This preserves expression in equation (19) because of the symmetry between inputs 1 and 2, and the symmetry between inputs 3 and 4. Thus the transformation $(a_i, b_i, c_i, d_i) \to (\frac{a_i+b_i}{2}, \frac{a_i+b_i}{2}, \frac{c_i+d_i}{2}, \frac{c_i+d_i}{2})$, for all i while leaving $\mathrm{P}(W=i)$ unchanged, does not decrease the λ-sum-rate since $H(Y|W)$ and f are concave functions in (a_i, b_i, c_i, d_i), and $H(Z|W)$ that appears with a negative sign remains constant under this transformation. Therefore without loss of generality assume that $\mathrm{P}(X|W=i) = (\frac{x_i}{2}, \frac{x_i}{2}, \frac{1-x_i}{2}, \frac{1-x_i}{2})$ when optimizing the expression in equation (19). Let $\mathrm{P}(W=i) = w_i$. Then we require $\sum w_i x_i = \frac{1}{2}$.

Hence we can work out the expression as the maximum over w_i, x_i satisfying the above constraint of

$$\lambda \log 6 + (1-\lambda) + (1-\lambda) \sum_i w_i [\log 3 + \frac{2}{3} - \frac{2}{3} H(x_i, 1-x_i)] + \sum_i w_i f(\frac{x_i}{2}, \frac{x_i}{2}, \frac{1-x_i}{2}, \frac{1-x_i}{2}).$$

We now compute $f(\frac{x_i}{2}, \frac{x_i}{2}, \frac{1-x_i}{2}, \frac{1-x_i}{2})$. Observe that

$$\begin{aligned} H(Z) - H(Y) &= H(a+b, c+d) - H(\frac{a+b}{3}, \frac{a+c}{3}, \frac{a+d}{3}, \frac{b+c}{3}, \frac{b+d}{3}, \frac{c+d}{3}) \\ &\overset{(a)}{\leq} H(a+b, c+d) - H(\frac{a+b}{3}, \frac{a+c+d}{3}, \frac{a}{3}, \frac{b+c+d}{3}, \frac{b}{3}, \frac{c+d}{3}) \\ &\overset{(b)}{\leq} H(a+b, c+d) - H(\frac{a+b}{3}, \frac{a+b+c+d}{3}, \frac{a+b}{3}, \frac{c+d}{3}, \frac{0}{3}, \frac{c+d}{3}) \\ &= \frac{1}{3} H(a+b, c+d) - \log 3. \end{aligned}$$

The step (a) holds because the expression is convex in c and d once we fix $c+d$, therefore its maximum must occur at the boundaries. The step (b) holds because the expression is convex in a and b once we fix $a+b$, therefore its maximum must occur at the boundaries.

Therefore $H(Z) - H(Y) \leq \frac{1}{3} H(a+b, c+d) - \log 3$ for all permissible (a, b, c, d). Since the function $\frac{1}{3} H(a+b, c+d) - \log 3$ is concave, we conclude that $f(a_i, b_i, c_i, d_i) \leq \frac{1}{3} H(a+b, c+d) - \log 3$ for all permissible (a, b, c, d). Hence, at $(a, b, c, d) = (\frac{x_i}{2}, \frac{x_i}{2}, \frac{1-x_i}{2}, \frac{1-x_i}{2})$, we have

$$f(\frac{x_i}{2}, \frac{x_i}{2}, \frac{1-x_i}{2}, \frac{1-x_i}{2}) \leq \frac{1}{3} H(x_i, 1-x_i) - \log 3.$$

The equality can be indeed achieved by taking with probability half $(0, x_i, 0, 1-x_i)$ and with probability half $(x_i, 0, 1-x_i, 0)$. Thus, $f(\frac{x_i}{2}, \frac{x_i}{2}, \frac{1-x_i}{2}, \frac{1-x_i}{2}) = \frac{1}{3} H(x_i, 1-x_i) - \log 3$.

Substituting this in we get

$$1 + (1-\lambda)\frac{2}{3} + (\frac{1}{3} - \frac{2}{3}(1-\lambda)) \sum_i w_i H(x_i, 1-x_i).$$

We need to maximize this subject to $\sum w_i x_i = \frac{1}{2}$. Clearly when $(1-\lambda) \leq \frac{1}{2}$ the optimal choice is to set $x_i = \frac{1}{2}$. in the other interval, it is optimal to set $x_i = 0$ w.p. $\frac{1}{2}$ and $x_i = 1$ w.p. $\frac{1}{2}$. In this case we get $1 + (1-\lambda)\frac{2}{3}$. □

C Proof of outer bound (Claim 4) for product broadcast channels

Proof. Take a code of length n. Let Q be a random variable independent of the code book such that Q is uniform in $[1:n]$. Identify

$$\begin{aligned} W_1 &= (M_0, Z_2^{1:n}, Y_1^{1:Q-1}, Z_1^{Q+1:n}, Q), \\ W_2 &= (M_0, Y_1^{1:n}, Y_2^{1:Q-1}, Z_2^{Q+1:n}, Q), \\ U_1 &= U_2 = M_1, \\ V_1 &= V_2 = M_2, \\ X_1 &= X_{1Q}, \\ X_2 &= X_{2Q}. \end{aligned}$$

We need to verify that these choice of auxiliaries work. We begin with the sum rate. Using the Fano inequality and some manipulations we can write

$$
\begin{aligned}
&n(R_0+R_1+R_2)-nf_1(\epsilon_n)\\
&\leq \lambda I(M_0;Y_1^{1:n},Y_2^{1:n})+(1-\lambda)I(M_0;Z_1^{1:n},Z_2^{1:n})+I(M_1;Y_1^{1:n},Y_2^{1:n}|M_0)+I(M_2;Z_1^{1:n},Z_2^{1:n}|M_0)-I(M_1;M_2|M_0)\\
&=\lambda I(M_0;Y_1^{1:n},Y_2^{1:n})+(1-\lambda)I(M_0;Z_1^{1:n},Z_2^{1:n})+I(M_1;Y_1^{1:n},Y_2^{1:n}|M_0)+I(M_2;Z_1^{1:n},Z_2^{1:n}|M_1,M_0)\\
&\quad -I(M_1;M_2|M_0,Z_1^{1:n},Z_2^{1:n})\\
&=\lambda I(M_0;Y_2^{1:n}|Y_1^{1:n})+(1-\lambda)I(M_0;Z_2^{1:n})+I(M_1;Y_2^{1:n}|M_0,Y_1^{1:n})+I(M_2;Z_2^{1:n}|M_1,M_0)\\
&\quad +\lambda I(M_0;Y_1^{1:n})+(1-\lambda)I(M_0;Z_1^{1:n}|Z_2^{1:n})+I(M_1;Y_1^{1:n}|M_0)+I(M_2;Z_1^{1:n}|M_1,Z_2^{1:n},M_0)-I(M_1;M_2|M_0,Z_1^{1:n},Z_2^{1:n})\\
&=\lambda I(M_0;Y_2^{1:n}|Y_1^{1:n})+(1-\lambda)I(M_0;Z_2^{1:n})+I(M_1;Y_2^{1:n}|M_0,Y_1^{1:n})+I(M_2;Z_2^{1:n}|M_1,M_0)\\
&\quad +\lambda I(M_0;Y_1^{1:n})+(1-\lambda)I(M_0;Z_1^{1:n}|Z_2^{1:n})+I(M_1;Y_1^{1:n}|M_0)+I(M_2;Z_1^{1:n}|Z_2^{1:n},M_0)-I(M_1;M_2|M_0,Z_2^{1:n})\\
&\leq \lambda I(M_0;Y_2^{1:n}|Y_1^{1:n})+(1-\lambda)I(M_0,Y_1^{1:n};Z_2^{1:n})+I(M_1;Y_2^{1:n}|M_0,Y_1^{1:n})+I(M_2;Z_2^{1:n}|M_1,M_0,Y_1^{1:n})\\
&\quad +\lambda I(M_0,Z_2^{1:n};Y_1^{1:n})+(1-\lambda)I(M_0;Z_1^{1:n}|Z_2^{1:n})+I(M_1;Y_1^{1:n}|M_0,Z_2^{1:n})+I(M_2;Z_1^{1:n}|Z_2^{1:n},M_0)-I(M_1;M_2|M_0,Z_2^{1:n})\\
&\leq \lambda I(M_0;Y_2^{1:n}|Y_1^{1:n})+(1-\lambda)I(M_0,Y_1^{1:n};Z_2^{1:n})+I(M_1;Y_2^{1:n}|M_0,Y_1^{1:n})+I(X_2^{1:n};Z_2^{1:n}|M_1,M_0,Y_1^{1:n})\\
&\quad +\lambda I(M_0,Z_2^{1:n};Y_1^{1:n})+(1-\lambda)I(M_0;Z_1^{1:n}|Z_2^{1:n})+I(M_1;Y_1^{1:n}|M_0,Z_2^{1:n})+I(M_2;Z_1^{1:n}|Z_2^{1:n},M_0)-I(M_1;M_2|M_0,Z_2^{1:n})\\
&\leq \lambda I(M_0,Y_1^{1:n};Y_2^{1:n})+(1-\lambda)I(M_0,Y_1^{1:n};Z_2^{1:n})+I(M_1;Y_2^{1:n}|M_0,Y_1^{1:n})+I(X_2^{1:n};Z_2^{1:n}|M_1,M_0,Y_1^{1:n})\\
&\quad +\lambda I(M_0,Z_2^{1:n};Y_1^{1:n})+(1-\lambda)I(M_0,Z_2^{1:n};Z_1^{1:n})+I(M_1;Y_1^{1:n}|M_0,Z_2^{1:n})+I(M_2;Z_1^{1:n}|M_0,Z_2^{1:n})-I(M_1;M_2|M_0,Z_2^{1:n})
\end{aligned}
$$

where $f_1(\epsilon)$ is a function that converges to zero as ϵ converges to zero. Thus,

$$
\begin{aligned}
&n(R_0+R_1+R_2)-nf_1(\epsilon_n)\\
&\leq \lambda I(M_0,Y_1^{1:n};Y_2^{1:n})+(1-\lambda)I(M_0,Y_1^{1:n};Z_2^{1:n})+I(M_1;Y_2^{1:n}|M_0,Y_1^{1:n})+I(X_2^{1:n};Z_2^{1:n}|M_1,M_0,Y_1^{1:n})\\
&\quad +\lambda I(M_0,Z_2^{1:n};Y_1^{1:n})+(1-\lambda)I(M_0,Z_2^{1:n};Z_1^{1:n})+I(M_1;Y_1^{1:n}|M_0,Z_2^{1:n})+I(M_2;Z_1^{1:n}|M_0,Z_2^{1:n})-I(M_1;M_2|M_0,Z_2^{1:n}).
\end{aligned}
$$

Similarly

$$
\begin{aligned}
&n(R_0+R_1+R_2)-nf_2(\epsilon_n)\\
&\leq \lambda I(M_0,Y_1^{1:n};Y_2^{1:n})+(1-\lambda)I(M_0,Y_1^{1:n};Z_2^{1:n})+I(M_1;Y_2^{1:n}|M_0,Y_1^{1:n})+I(M_2;Z_2^{1:n}|M_0,Y_1^{1:n})-I(M_1;M_2|M_0,Y_1^{1:n})\\
&\quad +\lambda I(M_0,Z_2^{1:n};Y_1^{1:n})+(1-\lambda)I(M_0,Z_2^{1:n};Z_1^{1:n})+I(M_2;Z_1^{1:n}|M_0,Z_2^{1:n})+I(X_1^{1:n};Y_1^{1:n}|M_0,M_2,Z_2^{1:n}).
\end{aligned}
$$

These lead to the following single letter bounds:

$$
\begin{aligned}
R_0+R_1+R_2 \leq\ & \lambda I(W_2;Y_2)+(1-\lambda)I(W_2;Z_2)+I(U_2;Y_2|W_2)+I(X_2;Z_2|U_2,W_2)\\
&+\lambda I(W_1;Y_1)+(1-\lambda)I(W_1;Z_1)+\min\big\{I(U_1;Y_1|W_1)+I(X_1;Z_1|U_1,W_1),\\
&\quad I(V_1;Z_1|W_1)+I(X_1;Y_1|V_1,W_1)\big\},\\
R_0+R_1+R_2 \leq\ & \lambda I(W_2;Y_2)+(1-\lambda)I(W_2;Z_2)+\min\big\{I(U_2;Y_2|W_2)+I(X_2;Z_2|U_2,W_2),\\
&\quad I(V_2;Z_2|W_2)+I(X_2;Y_2|V_2,W_2)\big\}\\
&+\lambda I(W_1;Y_1)+(1-\lambda)I(W_1;Z_1)+I(V_1;Z_1|W_1)+I(X_1;Y_1|V_1,W_1).
\end{aligned}
$$

Since the choice of the auxiliaries do not depend on λ, we conclude that

$$
\begin{aligned}
R_0+R_1+R_2 \leq\ & \min\{I(W_1;Y_1)+I(W_2;Y_2),I(W_1;Z_1)+I(W_2;Z_2)\}+I(U_2;Y_2|W_2)+I(X_2;Z_2|U_2,W_2)\\
&+\min\big\{I(U_1;Y_1|W_1)+I(X_1;Z_1|U_1,W_1),I(V_1;Z_1|W_1)+I(X_1;Y_1|V_1,W_1)\big\},\\
R_0+R_1+R_2 \leq\ & \min\{I(W_1;Y_1)+I(W_2;Y_2),I(W_1;Z_1)+I(W_2;Z_2)\}+I(V_1;Z_1|W_1)+I(X_1;Y_1|V_1,W_1)\\
&+\min\big\{I(U_2;Y_2|W_2)+I(X_2;Z_2|U_2,W_2),I(V_2;Z_2|W_2)+I(X_2;Y_2|V_2,W_2)\big\}.
\end{aligned}
$$

It remains to verify the following inequalities

$$\begin{aligned}
R_0 &\leq I(W_1;Y_1) + I(W_2;Y_2), \\
R_0 &\leq I(W_1;Z_1) + I(W_2;Z_2), \\
R_0 + R_1 &\leq I(W_1;Y_1) + I(W_2;Y_2) + I(U_1;Y_1|W_1) + I(U_2;Y_2|W_2), \\
R_0 + R_1 &\leq I(W_1;Z_1) + I(W_2;Z_2) + I(U_1;Y_1|W_1) + I(U_2;Y_2|W_2), \\
R_0 + R_2 &\leq I(W_1;Y_1) + I(W_2;Y_2) + I(V_1;Z_1|W_1) + I(V_2;Z_2|W_2), \\
R_0 + R_2 &\leq I(W_1;Z_1) + I(W_2;Z_2) + I(V_1;Z_1|W_1) + I(V_2;Z_2|W_2).
\end{aligned}$$

The first single-letter formula holds because one can verify that $I(W_1;Y_1) \geq \frac{1}{n}I(M_0;Y_1^{1:n})$ and $I(W_2;Y_2) \geq \frac{1}{n}I(M_0;Y_2^{1:n}|Y_1^{1:n})$. These imply that $I(W_1;Y_1) + I(W_2;Y_2) \geq I(M_0;Y_2^{1:n},Y_1^{1:n})$. One can finish the proof using the Fano inequality. The second inequality on R_0 can be proved similarly. The third inequality holds because $I(U_1W_1;Y_1) \geq \frac{1}{n}I(M_0M_1;Y_1^{1:n})$, $I(U_2W_2;Y_2) \geq \frac{1}{n}I(M_0M_1;Y_2^{1:n}|Y_1^{1:n})$ and (M_0, M_1) can be recovered from $(Y_1^{1:n}, Y_2^{1:n})$ with high probability. The fourth inequality holds because:

$$\begin{aligned}
&n(R_0 + R_1) - nf_3(\epsilon) \\
&\leq I(M_0;Z_1^{1:n},Z_2^{1:n}) + I(M_1;Y_1^{1:n},Y_2^{1:n}|M_0) \\
&\leq I(M_0;Z_1^{1:n},Z_2^{1:n}) + I(M_1,Z_2^{1:n};Y_1^{1:n}|M_0) + I(M_1;Y_2^{1:n}|M_0,Y_1^{1:n}) \\
&= I(M_0;Z_1^{1:n}|Z_2^{1:n}) + I(M_0;Z_2^{1:n}) + I(Y_1^{1:n};Z_2^{1:n}|M_0) + I(M_1;Y_1^{1:n}|M_0,Z_2^{1:n}) + I(M_1;Y_2^{1:n}|M_0,Y_1^{1:n}) \\
&\leq I(M_0,Z_2^{1:n};Z_1^{1:n}) + I(M_0,Y_1^{1:n};Z_2^{1:n}) + I(M_1;Y_1^{1:n}|M_0,Z_2^{1:n}) + I(M_1;Y_2^{1:n}|M_0,Y_1^{1:n}) \\
&= \sum_{i=1}^{n} \big(I(M_0,Z_2^{1:n};Z_1^i|Z_1^{i+1:n}) + I(M_0,Y_1^{1:n};Z_2^i|Z_2^{i+1:n}) + I(M_1;Y_1^i|M_0,Z_2^{1:n},Y_1^{1:i-1}) + I(M_1;Y_2^i|M_0,Y_1^{1:n},Y_2^{1:i-1})\big) \\
&\leq \sum_{i=1}^{n} \big(I(M_0,Z_2^{1:n},Z_1^{i+1:n};Z_1^i) + I(M_0,Y_1^{1:n},Z_2^{i+1:n};Z_2^i) + I(M_1;Y_1^i|M_0,Z_2^{1:n},Y_1^{1:i-1}) + I(M_1;Y_2^i|M_0,Y_1^{1:n},Y_2^{1:i-1})\big) \\
&= \sum_{i=1}^{n} \big(I(M_0,Z_2^{1:n},Y_1^{1:i-1},Z_1^{i+1:n};Z_1^i) + I(M_0,Y_1^{1:n},Y_2^{1:i-1},Z_2^{i+1:n};Z_2^i) + I(M_1;Y_1^i|M_0,Z_2^{1:n},Y_1^{1:i-1}) \\
&\qquad - I(Y_1^{1:i-1};Z_1^i|M_0,Z_2^{1:n},Z_1^{i+1:n}) + I(M_1;Y_2^i|M_0,Y_1^{1:n},Y_2^{1:i-1}) - I(Y_2^{1:i-1};Z_2^i|M_0,Y_1^{1:n},Z_2^{i+1:n})\big)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{n} \big(I(M_0,Z_2^{1:n},Y_1^{1:i-1},Z_1^{i+1:n};Z_1^i) + I(M_0,Y_1^{1:n},Y_2^{1:i-1},Z_2^{i+1:n};Z_2^i) + I(M_1;Y_1^i|M_0,Z_2^{1:n},Y_1^{1:i-1}) \\
&\qquad - I(Z_1^{i+1:n};Y_1^i|M_0,Z_2^{1:n},Y_1^{1:i-1}) + I(M_1;Y_2^i|M_0,Y_1^{1:n},Y_2^{1:i-1}) - I(Z_2^{i+1:n};Y_2^i|M_0,Y_1^{1:n},Y_2^{1:i-1})\big) \\
&= \sum_{i=1}^{n} \big(I(M_0,Z_2^{1:n},Y_1^{1:i-1},Z_1^{i+1:n};Z_1^i) + I(M_0,Y_1^{1:n},Y_2^{1:i-1},Z_2^{i+1:n};Z_2^i) + I(M_1;Y_1^i|M_0,Z_2^{1:n},Y_1^{1:i-1},Z_1^{i+1:n}) \\
&\qquad + I(M_1;Y_2^i|M_0,Y_1^{1:n},Y_2^{1:i-1},Z_2^{i+1:n})\big) \\
&= n(I(W_1;Z_1) + I(W_2;Z_2) + I(U_1;Y_1|W_1) + I(U_2;Y_2|W_2)).
\end{aligned}$$

The fifth inequality follows in a similar fashion, and the sixth one is similar to the third one. Hence the outer bound is valid. $\square$

D Proof of Lemma 10

Proof of Lemma 10: This is a consequence of Corollary 4. Let $d = 2$, let

$$\begin{aligned} &T_1(p(w, x_1, x_2)) \\ &\quad = I(W; Y_1, Y_2) + \sum_{w \in \mathcal{A}_1} \mathrm{P}(W = w) I(X_1, X_2; Y_1, Y_2 | W = w) \\ &\qquad + \sum_{w \in \mathcal{A}_2} \mathrm{P}(W = w) \big(I(X_1; Y_1, Y_2 | W = w) + I(X_2; Z_1, Z_2 | W = w) - I(X_1; X_2 | W = w) \big) \\ &\qquad + \sum_{w \in \mathcal{A}_3} \mathrm{P}(W = w) \big(I(X_1; Y_1, Y_2 | W = w) + I(X_2; Z_1, Z_2 | W = w) - I(X_1; X_2 | W = w) \big) \\ &\qquad + \sum_{w \in \mathcal{A}_4} \mathrm{P}(W = w) I(X_1, X_2; Z_1, Z_2 | W = w). \end{aligned}$$

$$\begin{aligned} &T_2(p(w, x_1, x_2)) \\ &\quad = I(W; Z_1, Z_2) + \sum_{w \in \mathcal{A}_1} \mathrm{P}(W = w) I(X_1, X_2; Y_1, Y_2 | W = w) \\ &\qquad + \sum_{w \in \mathcal{A}_2} \mathrm{P}(W = w) \big(I(X_1; Y_1, Y_2 | W = w) + I(X_2; Z_1, Z_2 | W = w) - I(X_1; X_2 | W = w) \big) \\ &\qquad + \sum_{w \in \mathcal{A}_3} \mathrm{P}(W = w) \big(I(X_1; Y_1, Y_2 | W = w) + I(X_2; Z_1, Z_2 | W = w) - I(X_1; X_2 | W = w) \big) \\ &\qquad + \sum_{w \in \mathcal{A}_4} \mathrm{P}(W = w) I(X_1, X_2; Z_1, Z_2 | W = w). \end{aligned}$$

It is clear that the set

$$\mathcal{G} = \{(g_1, g_2) : g_1 \leq T_1(p(w, x_1, x_2)), g_2 \leq T_2(p(w, x_1, x_2))\}$$

is a convex set. (In the standard manner, choose $\tilde{W} = (W, Q)$; When $Q = 0$ choose $(W, X_1, X_2) \sim p_1(w, x_1, x_2)$ and when $Q = 1$ choose $(W, X_1, X_2) \sim p_2(w, x_1, x_2)$). Hence from Corollary 4, we have the proof of Lemma 10.