On Marton's inner bound for two receiver broadcast channels

Yanlin Geng, Amin Gohari, Chandra Nair, Yuanming Yu

Abstract

In this paper we establish some properties concerning the sum-rate evaluation of the 2-letter characterization of Marton's inner bound.

1 Introduction

The broadcast channel refers to a communication scenario where a single sender wishes to communicate (possibly different messages) with multiple receivers. We consider a simple setting of the problem where the 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 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. Usually we also require that the inner bound is evaluable or at least in a single-letter form. The best known inner bound to the capacity region of the two receiver broadcast channel is due to Marton[3]. Until recently[2], this bound was not evaluable[1] even though it was in a single letter form. It is not known if Marton's inner bound is optimal or not.

It was known (folk-lore) that by looking at multi-letter characterizations of the bound, one would be able to deduce whether Marton's inner bound is optimal or not. However since Marton's inner bound was not evaluable, there was very little attempt to look at the multi-letter characterization of the bound. In this paper, we undertake this long overdue approach and obtain some very interesting results in this direction.

Similar to this inner bound, there is also an outer bound called the UV outer bound that was the best known outer bound[5] to the capacity region. However our current approach on the inner bound allowed us to deduce that the UV outer bound is strictly sub-optimal. Since this result is of independent interest we prepared this in a separate write-up [1]. In the following sections we focus solely on the inner bound and its variations.

[1] We say that a region is evaluable if there is a finite dimensional characterization of this region.

2 Main

2.1 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

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

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

$$\begin{aligned} SR(\mathfrak{q}\times\mathfrak{q}) = \max_{p(u,v,w,x_1,x_2)} & \min\{I(W;Y_1,Y_2), I(W;Z_1,Z_2)\} + I(U;Y_1,Y_2|W) \\ & + I(V;Z_1,Z_2|W) - I(U;V|W) \end{aligned} \tag{2}$$

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 $SR_2(\mathfrak{q}) := \frac{1}{2}SR(\mathfrak{q}\times\mathfrak{q})$ 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 $SR_2(\mathfrak{q})$ and $SR(\mathfrak{q})$.

Lemma 1. *(Folklore) The following two statements are equivalent:*

1. *Marton's achievable strategy achieves the optimal sum rate, $SR^*(\mathfrak{q})$, for all broadcast channels, i.e. $SR(\mathfrak{q}) = SR^*(\mathfrak{q})$.*
2. *$SR_2(\mathfrak{q}) = SR(\mathfrak{q})$ for all $\mathfrak{q}(y,z|x)$.*

Proof. ($1 \implies 2$) This follows from two facts: first, $SR_2(\mathfrak{q})$ yields an achievable sum-rate for the broadcast channel $\mathfrak{q}(y,z|x)$, i.e. $SR_2(\mathfrak{q}) \leq SR^*(\mathfrak{q})$; and second, $SR_2(\mathfrak{q}) \geq SR(\mathfrak{q})$ 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 (1). 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

$$\begin{aligned}
&2SR_2(\mathfrak{q}) \\
&\geq \min\{I(\tilde{W};Y_1,Y_2), I(\tilde{W};Z_1,Z_2)\} \\
&\quad + I(\tilde{U};Y_1,Y_2|\tilde{W}) + I(\tilde{V};Z_1,Z_2|\tilde{W}) - I(\tilde{U};\tilde{V}|\tilde{W}) \\
&= \min\{I(W_1;Y_1), I(W_1;Z_1)\} + I(U_1;Y_1|W_1) \\
&\quad + I(V_1;Z_1|W_1) - I(U_1;V_1|W_1) \\
&\quad + \min\{I(W_2;Y_2), I(W_2;Z_2)\} + I(U_2;Y_2|W_2) \\
&\quad + I(V_2;Z_2|W_2) - I(U_2;V_2|W_2) \\
&= 2SR(\mathfrak{q}).
\end{aligned}$$

This shows that if $SR(\mathfrak{q})$ is the maximum achievable sum-rate then $SR_2(\mathfrak{q}) = SR(\mathfrak{q})$ 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

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

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

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

where $SR(\mathfrak{q}\otimes_n)$ 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

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

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

The capacity region of the broadcast channel is a very important open question in the area of multiuser information theory. Marton's inner bound represents the best known achievable region to this problem. Unfortunately, we do not know whether Marton's inner bound is optimal or not. Recently there has been a lot of effort on the outer bounds, and the best known outer bound have been established to be strictly sub-optimal[1]. Lemma 1 is an attempt at answering the same question regarding the inner bound. If one can find a channel for which $SR_2(\mathfrak{q}) > SR(\mathfrak{q})$ then Marton's inner bound is strictly sub-optimal, otherwise Marton's inner bound is optimal and would yield the capacity region. Rephrasing this argument, let $p^*(u,v,w,x)$ achieve the maximum sum-rate in (1) for some channel $\mathfrak{q}(y,z|x)$. 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. If it turns out that $p(\tilde{u},\tilde{v},\tilde{w},x_1,x_2)$ yields a global maximum, then $SR_2(\mathfrak{q}) = SR(\mathfrak{q})$.

2.2 λ-sum rate

Motivated by this argument about the factorization of the 2-letter Marton's inner bound, we consider a related expression which turns out to be quite useful. Given a broadcast channel $\mathfrak{q}(y,z|x)$ we define λ-sum rate (for $\lambda \in [0,1]$) to be

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

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

Here we state some results about the λ-sum rate. The proofs can be found in [1].

Lemma 2. *For a given channel $\mathfrak{q}(y,z|x)$, λ-$SR(\mathfrak{q})$ is convex in λ for $\lambda \in [0,1]$.*

Lemma 3. *λ-$SR(\mathfrak{q})$ is related to the Marton's inner bound as follows:*

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

i.e. the minimum value of λ-SR, for $\lambda \in [0,1]$, yields the Marton's inner bound.

Note: The proof is a consequence of Lemma 11 stated in the Appendix. Details can be found in [1].

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

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

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

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

Lemma 5. *To compute the maximum sum-rate in (3), it suffices to consider auxiliary random variables that satisfy $|\mathcal{U}|, |\mathcal{V}|, |\mathcal{W}| \leq |\mathcal{X}|$.*

We now prove an important property regarding any global maximizer of (3).

Lemma 6. *Let $p^*_\lambda(u,v,w,x)$ be any maximizer of the expression in (3), then the following holds:*
For all (u,v,w,x) such that $p^*_\lambda(u,v,w,x) > 0$,

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

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

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

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

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

where the notation λ-$SR(\mathfrak{q},p_1)$ 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

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

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

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 $\Psi(u_0,v_0,w_0,x_1) \leq \lambda$-$SR(\mathfrak{q})$. 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

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

Since for every $\epsilon > 0$ we have λ-$SR(\mathfrak{q},p_1) \leq \lambda$-$SR(\mathfrak{q},p_\lambda^*)$ it follows that

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

as desired. □

Definition 1. 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 $p_\mu^\epsilon(u,v,w,x)$ be a distribution defined according to

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

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

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

Remark 1. 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

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

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

$$\lambda\text{-}SR(\mathfrak{q},p)\ge\sum_{\mu(u,v,x)}\mu(u,v,x)\Psi_{1,\mu}(u,v,x)+\sum_{u,v,x,y,z}\mu(u,v,x)\mathfrak{q}(y,z|x)\log\frac{\mu(z)^{(1-\lambda)}\mu(y)^\lambda}{p(z)^{(1-\lambda)}p(y)^\lambda}.\tag{4}$$

We now state one of the main results of this paper.

Theorem 1. *Any enhanced local maximum of the λ-$SR(\mathfrak{q})$ 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 $\lambda\text{-}SR(\mathfrak{q},p)<\lambda\text{-}SR(\mathfrak{q},p^*_\lambda)=\lambda\text{-}SR(\mathfrak{q})$.

Note that for every w,

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

Since $p(u,v,w,x)$ is an enhanced local maximum it follows from (4) that for every w such that $p^*_\lambda(w)>0$, we have

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

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

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

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

Consider the two-letter expression corresponding to the λ-sum rate in (3) given by

$$\lambda\text{-}SR_2(\mathfrak{q}) = \frac{1}{2}\lambda\text{-}SR(\mathfrak{q}\times\mathfrak{q}) \tag{5}$$

$$= \frac{1}{2}\max_{p(u,v,w,x_1,x_2)} \lambda I(W;Y_1,Y_2) + (1-\lambda)I(W;Z_1,Z_2) + I(U;Y_1,Y_2|W) + I(V;Z_1,Z_2|W) - I(U;V|W) \tag{6}$$

Lemma 5 implies that to compute the global maximum of the product channel $\mathfrak{q}\times\mathfrak{q}$ in (6) it suffices to consider $|\mathcal{U}|, |\mathcal{V}|, |\mathcal{W}| \leq |\mathcal{X}|^2$. If one wishes to verify $\lambda\text{-}SR_2(q) = \lambda\text{-}SR(q)$, then computations can be significantly reduced by using the following corollary to Theorem 1.

Corollary 2. *To verify that $\lambda\text{-}SR_2(\mathfrak{q}) = \lambda\text{-}SR(\mathfrak{q})$ 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 $\lambda\text{-}SR(\mathfrak{q})$, is an enhanced local maximum.*

We consider the two-letter factorization of the $\lambda\text{-}SR(\mathfrak{q})$ 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 $\lambda\text{-}SR(\mathfrak{q})$ factorizes then so does $SR(\mathfrak{q})$.

Corollary 3. *If $\lambda\text{-}SR_2(\mathfrak{q}) = \lambda\text{-}SR(\mathfrak{q})$ for all $\lambda \in [0,1]$ then $SR_2(\mathfrak{q}) = 2SR(\mathfrak{q})$*

Proof. The proof is immediate from Lemma 3. Observe that, under the assumption $\lambda\text{-}SR_2(\mathfrak{q}) = \lambda\text{-}SR(\mathfrak{q})$, we have

$$SR_2(\mathfrak{q}) = \min_{\lambda\in[0,1]} \lambda\text{-}SR_2(\mathfrak{q}) = \min_{\lambda\in[0,1]} \lambda\text{-}SR(\mathfrak{q}) = SR(\mathfrak{q}).$$

□

2.3 λ-SR for product channels

In this section we study the behavior of λ-SR($\mathfrak{q}$) for the product of two non-identical channels. The principal question we are concerned with is the following: Is it true that for all channels $\mathfrak{q}_1(y_1,z_1|x_1), \mathfrak{q}_2(y_2,z_2|x_2)$

$$\lambda\text{-}SR(\mathfrak{q}_1\times\mathfrak{q}_2) = \lambda\text{-}SR(\mathfrak{q}_1) + \lambda\text{-}SR(\mathfrak{q}_2). \tag{7}$$

We term this the *factorization of λ-SR.*

Remark 2. Equation (7) seems to be true for all the channels that have been considered, and a few others that we ventured to simulate. If one were brave, one could possibly conjecture that this is true. However at this point we just present this as a plausible line of attack to establish the optimality of Marton's coding strategy. Observe that if (7) holds, then by Corollary 3 it follows that Marton's coding strategy is optimal. The vice-versa need not hold, i.e. it may still be possible that Marton's coding strategy is optimal, and yet (7) does not hold. One might wonder why we did not consider the question

$$SR(\mathfrak{q}_1 \times \mathfrak{q}_2) \stackrel{?}{=} SR(\mathfrak{q}_1) + SR(\mathfrak{q}_2). \tag{8}$$

Indeed it turns out that we can show (in general)

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

There is a class of channels for which (7) holds but (8) does not (Lemma 8).

2.3.1 Sufficient conditions for factorization of λ-$SR(\mathfrak{q}_1 \times \mathfrak{q}_2)$

In this section, we derive sufficient conditions under which (7) 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:*

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

Proof.

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

□

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 2. *The λ-$SR(\mathfrak{q}_1 \times \mathfrak{q}_2)$ factorizes (as in (7)) if any one of the conditions 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 $X_2 \to Z_2$ be deterministic. Then we will show that

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

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

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

Observe that

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

Similar reasoning will allow one to deal with the case $X_1 \to Y_1$ 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 proceed to show that

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

Again, a similar argument will work when $X_1 \to Z_1$ is deterministic.

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 $p(x_2), I(X_2;Y_2) \ge I(X_2;Z_2)$. Then observe that

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

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

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. □

Remark 3. The above Theorem can be used to prove the optimality of Marton's inner bound for a variety of channels, including: more-capable, semi-deterministic, product of reversely semi-deterministic, product of reversely more-capable, etc. Proof can also be found in [1]. Here we outline the proof for reversely more-capable channel for the sum-rate.

Lemma 7. *Consider a reversely more capable channel, i.e.* $X = (X_1, X_2), Y = (Y_1, Y_2), Z = (Z_1, Z_2)$ *and* $\mathfrak{q}(y_1, y_2, z_1, z_2|x_1, x_2) = \mathfrak{q}_1(y_1, z_1|x_1)\mathfrak{q}_2(y_2, z_2|x_2)$*. Further* Y_1 *is more capable than* Z_1 *(i.e.* $\forall p(x_1), I(X_1; Y_1) \geq I(X_1; Z_1)$*) and* Z_2 *is more capable than* Y_2 *(i.e.* $\forall p(x_2), I(X_2; Z_2) \geq I(X_2; Y_2)$*). Then* $SR^*(\mathfrak{q}) = SR(\mathfrak{q})$*, i.e. Marton's sum-rate is optimal.*

Proof. The proof follows from the following fact: Consider the n-fold product $(\mathfrak{q}_1 \times \mathfrak{q}_2)\otimes_n$. Since in each of the $2n$ components, one receiver is more capable than the other, it follows by repeated use of Theorem 2 that

$$\lambda\text{-}SR((\mathfrak{q}_1 \times \mathfrak{q}_2)\otimes_n) = n\lambda\text{-}SR(\mathfrak{q}_1) + n\lambda\text{-}SR(\mathfrak{q}_2) = n\lambda\text{-}SR(\mathfrak{q}_1 \times \mathfrak{q}_2).$$

Now

$$\begin{aligned} SR^*(\mathfrak{q}_1 \times \mathfrak{q}_2) &= \liminf_n \min_{\lambda\in[0,1]} \frac{1}{n}\lambda\text{-}SR((\mathfrak{q}_1 \times \mathfrak{q}_2)\otimes_n) \\ &= \min_{\lambda\in[0,1]} \lambda\text{-}SR(\mathfrak{q}_1) + \lambda\text{-}SR(\mathfrak{q}_2) \\ &= \min_{\lambda\in[0,1]} \lambda\text{-}SR(\mathfrak{q}_1 \times \mathfrak{q}_2) \\ &= SR(\mathfrak{q}_1 \times \mathfrak{q}_2). \end{aligned}$$

Thus Marton's strategy achieves the optimal sum-rate for the product of reversely more capable broadcast channels. □

Lemma 8. *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*

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

Proof. From [4], 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. Hence from Lemma 7 we have that

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

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

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

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 $SR(\mathfrak{q}_1) = C$, the common capacity.

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

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

□

2.4 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 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

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

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

Lemma 9. *It suffices to consider $|W| \leq |2|$ to compute the $SR_{RTD}(\mathfrak{q})$.*

Proof. We prove this by showing that for every $p(x)$, to compute the maximum of the expression (over $p(w|x)$)

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

it suffices to consider $|W| \leq 2$.

To compute $SR_{RTD}(\mathfrak{q}, p(x))$, for every $w \in \mathcal{A}$ we can assume w.l.o.g. that $I(X;Y|W=w) \geq I(X;Z|W=w)$. Otherwise, say it does not hold for w_0, then move w_0 from $\mathcal{A}$ to $\mathcal{A}^c$ and the sum-rate increases. In other words, conditioned on $W = w_0$ set $U = \emptyset, V = X$.

Similarly for every $w \in \mathcal{A}^c$ we can assume w.l.o.g. that $I(X;Y|W=w) \leq I(X;Z|W=w)$. For every $p(x)$, define a function $J(X) = \max\{I(X;Y), I(X;Z)\}$, and let $J(X|W) = \sum_w \mathrm{P}(W=w) J(X|W=w)$. Using this definition, we can now write

$$\begin{aligned} SR_{RTD}(\mathfrak{q}, p(x)) &= \max_{p(w|x)} \lambda I(W;Y) + (1-\lambda) I(W;Z) + J(X|W) \\ &= \lambda I(X;Y) + (1-\lambda) I(X;Z) + \max_{p(w|x)} \big(J(X|W) - \lambda I(X;Y|W) - (1-\lambda) I(X;Z|W)\big) \\ &= \lambda I(X;Y) + (1-\lambda) I(X;Z) + \hat{R}(X), \end{aligned}$$

where for any $p(x), R(X) := J(X) - \lambda I(X;Y) - (1-\lambda) I(X;Z)$, and $\hat{R}(X)$ denotes the value of the upper concave envelope of $R(X)$ evaluated at $p(x)$. To compute the upper concave envelope[2] at any point one only needs to take the convex combination of some two points on the original surface, and hence $|W| = 2$ is sufficient. □

[2] Upper concave envelope of a function $f(x)$ is defined as the smallest concave function, $g(x)$, such that $g(x) \geq f(x)$. Such a function is well-defined and is the $\inf\{h(x) : h(x) \geq f(x), h(x) \text{ is concave }\}$

It was shown [6] 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. $SR(\mathfrak{q}) = SR_{RTD}(\mathfrak{q})$ when $|X| = 2$. This result is based on the inequality that whenever $|X| = 2$ we have

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

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

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 $2RTD$. We define

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

Similarly define

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

In a similar fashion to the proof of Lemma 3 one can show the following Lemma.

Lemma 10. *The following holds:*

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

The proof is given in the Appendix.

Remark 4. Suppose there is a binary input channel $\mathfrak{q}(y,z|x)$ such that $SR_{2RTD}(\mathfrak{q}\times\mathfrak{q}) > 2SR_{RTD}(\mathfrak{q})$ then it would immediately imply that

$$SR_2(\mathfrak{q}) \geq \frac{1}{2}SR_{2RTD}(\mathfrak{q}\times\mathfrak{q}) > SR_{RTD}(\mathfrak{q}) = SR(\mathfrak{q})$$

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 $\lambda\text{-}SR_{2RTD}(\mathfrak{q}_1\times\mathfrak{q}_2) = \lambda\text{-}SR_{RTD}(\mathfrak{q}_1) + \lambda\text{-}SR_{RTD}(\mathfrak{q}_2)$ for channels with arbitrary input cardinality. Hence from Lemma 10 it will immediately follow that $SR_{2RTD}(\mathfrak{q}\times\mathfrak{q}) = 2SR_{RTD}(\mathfrak{q})$.

Theorem 3. *The following holds:*

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

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

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

Hence it suffices to show that

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

Observe that

$$\begin{aligned}
&\lambda I(W;Y_1,Y_2) + (1-\lambda)I(W;Z_1,Z_2) + \sum_{w\in\mathcal{A}_1} \mathrm{P}(W=w)I(X_1,X_2;Y_1,Y_2|W=w)\\
&\quad + \sum_{w\in\mathcal{A}_2} \mathrm{P}(W=w)\big(I(X_1;Y_1,Y_2|W=w) + I(X_2;Z_1,Z_2|W=w) - I(X_1;X_2|W=w)\big)\\
&\quad + \sum_{w\in\mathcal{A}_3} \mathrm{P}(W=w)\big(I(X_1;Y_1,Y_2|W=w) + I(X_2;Z_1,Z_2|W=w) - I(X_1;X_2|W=w)\big)\\
&\quad + \sum_{w\in\mathcal{A}_4} \mathrm{P}(W=w)I(X_1,X_2;Z_1,Z_2|W=w)\\
&= \lambda H(Y_1,Y_2) + (1-\lambda)H(Z_1,Z_2)\\
&\quad + \sum_{w\in\mathcal{A}_1} \mathrm{P}(W=w)\big(I(X_1,X_2;Y_1,Y_2|W=w) - \lambda H(Y_1,Y_2|W=w) - (1-\lambda)H(Z_1,Z_2|W=w)\big)\\
&\quad + \sum_{w\in\mathcal{A}_2} \mathrm{P}(W=w)\big(I(X_1;Y_1,Y_2|W=w) + I(X_2;Z_1,Z_2|W=w) - I(X_1;X_2|W=w)\\
&\qquad\qquad - \lambda H(Y_1,Y_2|W=w) - (1-\lambda)H(Z_1,Z_2|W=w)\big) \qquad (9)\\
&\quad + \sum_{w\in\mathcal{A}_3} \mathrm{P}(W=w)\big(I(X_1;Y_1,Y_2|W=w) + I(X_2;Z_1,Z_2|W=w) - I(X_1;X_2|W=w)\\
&\qquad\qquad - \lambda H(Y_1,Y_2|W=w) - (1-\lambda)H(Z_1,Z_2|W=w)\big)\\
&\quad + \sum_{w\in\mathcal{A}_4} \mathrm{P}(W=w)\big(I(X_1,X_2;Z_1,Z_2|W=w) - \lambda H(Y_1,Y_2|W=w) - (1-\lambda)H(Z_1,Z_2|W=w)\big).
\end{aligned}$$

The idea of the proof is to factorize each of the four summation terms in (9) separately.

Consider the following manipulations of the terms.

$$\begin{aligned} &I(X_1,X_2;Y_1,Y_2|W=w) - \lambda H(Y_1,Y_2|W=w) - (1-\lambda)H(Z_1,Z_2|W=w) \\ &\quad = I(X_1;Y_1|W=w,Y_2) + I(X_2;Y_2|W=w,Z_1) - \lambda H(Y_1|W=w,Y_2) \\ &\qquad - \lambda H(Y_2|W=w,Z_1) - (1-\lambda)H(Z_1|W=w,Y_2) - (1-\lambda)H(Z_2|W=w,Z_1) \end{aligned} \tag{10}$$

$$\begin{aligned} &I(X_1;Y_1,Y_2|W=w) + I(X_2;Z_1,Z_2|W=w) - I(X_1;X_2|W=w) \\ &\qquad - \lambda H(Y_1,Y_2|W=w) - (1-\lambda)H(Z_1,Z_2|W=w) \\ &\quad = I(X_1;Y_1|W=w,Y_2) + I(X_2;Z_2|W=w,Z_1) - \lambda H(Y_1|W=w,Y_2) \\ &\qquad - \lambda H(Y_2|W=w,Z_1) - (1-\lambda)H(Z_1|W=w,Y_2) - (1-\lambda)H(Z_2|W=w,Z_1) \\ &\qquad + I(X_2;Z_1|W=w) + I(X_1;Y_2|W=w,Z_1) - I(X_1;X_2|W=w) \\ &\quad \le I(X_1;Y_1|W=w,Y_2) + I(X_2;Z_2|W=w,Z_1) - \lambda H(Y_1|W=w,Y_2) \\ &\qquad - \lambda H(Y_2|W=w,Z_1) - (1-\lambda)H(Z_1|W=w,Y_2) - (1-\lambda)H(Z_2|W=w,Z_1) \end{aligned} \tag{11}$$

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

$$\begin{aligned} &I(X_2;Y_1,Y_2|W=w) + I(X_1;Z_1,Z_2|W=w) - I(X_1;X_2|W=w) \\ &\quad \le I(X_1;Z_1|W=w,Z_2) + I(X_2;Y_2|W=w,Y_1) - \lambda H(Y_1|W=w,Z_2) \\ &\qquad - \lambda H(Y_2|W=w,Y_1) - (1-\lambda)H(Z_1|W=w,Z_2) - (1-\lambda)H(Z_2|W=w,Y_1) \end{aligned} \tag{12}$$

Finally

$$\begin{aligned} &I(X_1,X_2;Z_1,Z_2|W=w) - \lambda H(Y_1,Y_2|W=w) - (1-\lambda)H(Z_1,Z_2|W=w) \\ &\quad = I(X_1;Z_1|W=w,Z_2) + I(X_2;Z_2|W=w,Y_1) - \lambda H(Y_1|W=w,Z_2) \\ &\qquad - \lambda H(Y_2|W=w,Y_1) - (1-\lambda)H(Z_1|W=w,Z_2) - (1-\lambda)H(Z_2|W=w,Y_1) \end{aligned} \tag{13}$$

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 (10), (11), (12), (13), and the definitions of $W_1, W_2, \mathcal{B}, \mathcal{C}$ we can bound the expression in (9) by

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

This implies that

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

and completes the proof of the Lemma. □

Remark 5. 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.

3 Conclusion

In this paper we considered the 2-letter sum rate of the Marton's inner bound and deduced some properties of this expression. We showed that Marton's sum rate for product of non-identical channels does not factorize and hence our focus was on a very related quantity called the λ-sum rate. We showed sufficient conditions for its factorization and used this to establish the sum capacity of the product of two reversely semi-deterministic channels. We also introduced the idea of an enhanced local maximum and tied it to the factorization of the λ-SR, and in turn the factorization of the 2-letter sum rate of Marton's inner bound. We also showed that a particular strategy called the randomized time division and its natural extension to a product channel, does factorize and in this process introduced new ideas in the identification of auxiliary random variables. We hope that the techniques developed here will eventually lead us to determining the optimality of Marton's inner bound.

Acknowledgement

The work of Amin Gohari and Chandra Nair was partially supported by a grant from the University Grants Committee of the Hong Kong Special Administrative Region, China (Project No. AoE/E-02/08). This work was also partially supported by the Institute of Theoretical Computer Science and Communication and the Institute of Network Coding at the The Chinese University of Hong Kong.

References

[1] Y Geng, A Gohari, C Nair, and Y Yu, *UV outer bound is strictly sub-optimal*, In preparation (2011).

[2] Amin Aminzadeh Gohari and Venkat Anantharam, *Evaluation of Marton's inner bound for the general broadcast channel*, CoRR **abs/0904.4541** (2009).

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

[4] C Nair, *Capacity regions of two new classes of two-receiver broadcast channels*, Information Theory, IEEE Transactions on **56** (2010), no. 9, 4207–4214.

[5] ______, *A note on outer bounds for broadcast channel*, Presented at International Zurich Seminar (2010).

[6] 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).

Lemma 11. *[1] 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 $\mathcal{G}$, defined by*

$$\mathcal{G} = \{(g_1, g_2, ..., g_d) \in \mathbb{R}^d : g_i \leq T_i(p(u)) \text{ for some } p(u) \in \mathcal{P}\},$$

Then

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

is a convex set.

Proof of Lemma 10: This is a consequence of Lemma 11. 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 Lemma 11, we have the proof of Lemma 10.