An earlier version of this paper had a different title. This is the revised full version of ISIT 2020 paper on outer bounds for two receiver broadcast channels. The revised paper has been reorganized and there are some new results.

OUTER BOUNDS FOR MULTIUSER SETTINGS: THE AUXILIARY RECEIVER APPROACH

AMIN GOHARI AND CHANDRA NAIR

Abstract. This paper employs auxiliary receivers as a mathematical tool to identify Gallager-type auxiliary random variables and write outer bounds for some basic multiuser settings. This approach is then applied to the relay, interference, and broadcast channel settings, yielding new outer bounds that improve on existing outer bounds and strictly outperform classical outer bounds at least in some regimes. For instance, we strictly improve on: the cutset outer bound for the scalar Gaussian relay channel, the outer bounds for the Gaussian Z-interference channel, and the outer bounds for the two receiver broadcast channel.

1. Introduction

A number of techniques for proving infeasiblity results are known in the literature. The generic and classical approach is based on identification of the auxiliary random variables as the past and/or future of the underlying random variables to write single-letter converse bounds. We call such an identification to be a *Gallager-type* auxiliary identification [Gal74]. However, non-standard techniques have also been used in some specialized settings. Establishing the continuity of differential entropy [PW15] with respect to Wasserstein metric to develop an outer bound for the Gaussian Z-interference channel is an example of a non-classical approach. For the primitive relay channel, converse bounds based on concentration of Gaussian measure or reverse hypercontractivity are known [ARY09, WBO19, LO19, WO18, LO18], which are again non-classical approaches. It is also known that in distributed source/channel coding problems with dependent sources [KU10], less common measures of correlation based on maximal correlation or hypercontractivity can provide better converse bounds (see also [DMN18, GKS16] which studies the fundamental limits of this approach).

Cover, [Cov72], employed auxiliary random variables so that one can write achievable regions that captured the idea of superposition coding (clusters) for broadcast channels. Subsequently the use of auxiliary random variables at the sender side to develop achievable rate regions have been a useful tool in the information-theorists' tool box. In this paper, we propose auxiliary channels (or auxiliary receivers) to write outer bounds for basic multi-terminal settings and show that this can be used to develop bounds that outperform state-of-the-art bounds in some basic settings. We call the new family of outer bounds developed in this paper to be the *J-bounds*, with J being a generic pseudonym for an auxiliary receiver.

One can identify special instances of utilizing auxiliary receivers in prior works, notably in the genie-aided outer bound proofs. Our converse bounds based on auxiliary receiver generalize genie-aided outer bounds since the dummy receiver does not necessarily provide any extra information to the existing receivers. One may also interpret some of the existing bounds in the literature as special instances of J-bounds: See for instance, the auxiliary J in [GA10, Corollary 2] for the secret key agreement, auxiliary variable X in [WA08, Definition 3] for the multiterminal source coding problem, and the remote source and channels in [YLL18, Eq. 12, 13] for a joint source-channel coding over a broadcast channel.

The following well-known result will be used repeatedly in this paper.

Lemma 1 (Körner-Márton Lemma, (4.14) in [KM77]). *For any tuple of random variables (U, Y^n, Z^n) the following equality holds:*

$$\begin{aligned} H(Y^n|U) - H(Z^n|U) &= \sum_{i=1}^{n} H(Y_i|U, Y^{i-1}, Z_{i+1}^n) - H(Z_i|U, Y^{i-1}, Z_{i+1}^n) \\ &= \sum_{i=1}^{n} H(Y_i|U, Z^{i-1}, Y_{i+1}^n) - H(Z_i|U, Z^{i-1}, Y_{i+1}^n) \end{aligned}$$

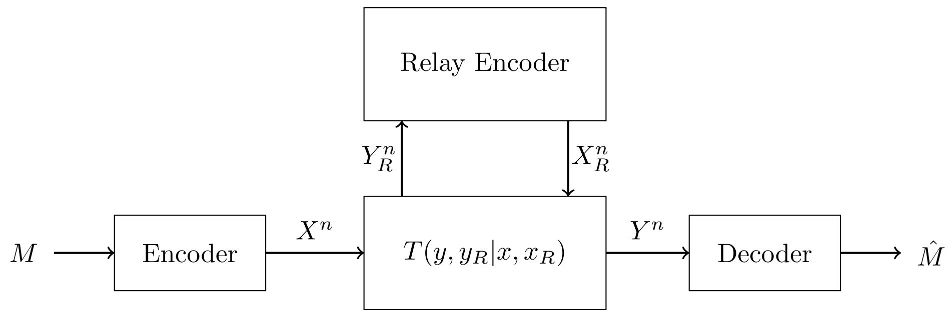

FIGURE 1. Transmission of a message M over a memoryless relay channel with n uses of the channel.

Remark 1. This equality has been repeatedly used in the literature to provide outer bounds or converses to capacity regions and in this paper we will continue to employ this frequently. Some generic ways of using this inequality has been illustrated in Lemma 6 in the Appendix. The authors would also like to remark that this lemma was called *Csiszar-sum-lemma* in the literature, based on its perceived first appearance as Lemma 7 in [CK78]. However a private communication to the authors by Körner revealed that this equality was first identified by Katalin Márton and used in [KM77, (4.14)] in her joint work with Janos Körner. Hence the authors find it appropriate to rechristen it as Körner-Márton Lemma.

1.1. **Organization.** This paper is organized as follows: in Sections 2 and 3 we give our new outer bounds for the relay and interference channels respectively. Finally, in Section 4 we give our outer bounds for the broadcast channel.

Notation: We adopt most of our notation from [EK12]. In particular, we use Y^{i-1} to denote the sequence $(Y_1, Y_2, \cdots, Y_i)$, and Y_i^j to denote $(Y_i, Y_{i+1}, \cdots, Y_j)$. For discrete settings, logarithms and in base two, and for continuous channels, the logarithms are in base e. Conditional distributions representing channels are denoted by $T(\cdot|\cdot)$.

2. Relay channel

A relay channel models transmission of a message from a sender X to a receiver Y in the presence of a helper relay node. A relay channel is described by a conditional distribution $T(y, y_R|x, x_R)$ where x_R and y_R are the input and output to the channel by the relay (see Fig. 1). A non-negative rate R is said to be achievable if the transmitter is able to send a message at rate R such that the probability of error tends to zero as n, the blocklength, tends to infinity. The supremum of all achievable rates is called the capacity region for the relay channel $T(y, y_R|x, x_R)$.

There are different achievability schemes in the literature for a single-relay channel such as decode-and-forward, compress-and-forward, etc (e.g. see [CG79, MHU19]). On the other hand, the best known upper bound for a general relay channel is the cut-set bound. While better bounds are known for the special class of relay channels with orthogonal receiver components (e.g. see [ARY09, WBO19, LO19, WO18]), these bounds are not applicable to a general relay channel. For more details on relay channels and a collection of known results please refer to Chapters 16 in [EK12].

The following theorem proves an upper bound to the capacity of a general relay channel $T(y, y_R|x, x_R)$. It is a J-version of the cut-set bound.

Theorem 1. *Consider a relay channel $T(y, y_R|x, x_R)$. Let $\mathcal{F}$ be the class of $T_{J|Y,Y_R,X,X_R}$ be a conditional distribution defined on arbitrary alphabet set $\mathcal{J}$ satisfying*

$$T_{J|X,X_R}(j|x, x_R) = T_{J|X,X_R}(j|x, x_R'), \quad \forall j, x, x_R, x_R' \tag{1}$$

where

$$T_{J|X,X_R}(j|x, x_R) = \sum_{y,y_R} T_{J|Y,Y_R,X,X_R}(j|y, y_R, x, x_R) T_{Y,Y_R|X,X_R}(y, y_R|x, x_R).$$

In other words $X_R \to X \to J$ is Markov. Then, there is some $p(x, x_R)$ such that a rate R is achievable only if for any $T_{J|Y,Y_R,X,X_R} \in \mathcal{F}$ we have

$$R \leq \min\big(I(X, X_R; Y),\ I(X; Y, Y_R|X_R),\ I(X; J, Y_R|X_R) + I(X, X_R; Y|W) - I(X, X_R; J|W)\big) \tag{2}$$

for some $p_{W,X,X_R,Y,Y_R,J} = p_{X,X_R} p_{W|X,X_R} T_{Y,Y_R|X,X_R} T_{J|Y,Y_R,X,X_R}$. Further it suffices to consider $|\mathcal{W}| \leq |\mathcal{X}||\mathcal{X}_R|$.

Proof. The cardinality bound on $|\mathcal{W}|$ comes from the standard Caratheodory-Bunt [Bun34] arguments and is omitted. Take an arbitrary code. The code defines a joint distribution $p_{M,X^n,X_R^n,Y^n,Y_R^n}$. Since randomization at the sender or relay does not improve the capacity region, w.l.o.g. we assume that X^n is a function of the message M and X_{Ri} is a function of Y_R^{i-1} $i = 1, .., n$, the past symbols received by the relay. Note that the first two constraints are the cut-set outer bound constraints. Therefore, it only remains to prove the third. Let

$$p_{J^n|M,X^n,X_R^n,Y^n,Y_R^n} = \prod_{i=1}^{n} T_{J_i|X_i,X_{Ri},Y_i,Y_{Ri}}.$$

We can think of $p_{J,Y,Y_R|X,X_R}$ as an extended memoryless relay channel. Observe that by Fano's inequality we have $n(R - \epsilon_n) \leq I(M; Y^n)$ for some ϵ_n that tends to zero as n tends to infinity. Next, we write

$$I(M; Y^n) = I(M; J^n) + I(M; Y^n) - I(M; J^n) \tag{3}$$

by adding and subtracting $I(M; J^n)$. The terms $I(M; J^n)$ and $I(M; Y^n) - I(M; J^n)$ are single-letterized separately as follows: starting with the latter, we have

$$\begin{aligned}
I(M; Y^n) - I(M; J^n) &\overset{(a)}{=} \sum_i I(M; Y_i|J_{i+1}^n, Y^{i-1}) - \sum_i I(M; J_i|J_{i+1}^n, Y^{i-1}) \\
&\overset{(b)}{=} \sum_i I(M, X_i; Y_i|J_{i+1}^n, Y^{i-1}) - \sum_i I(M, X_i; J_i|J_{i+1}^n, Y^{i-1}) \\
&\overset{(c)}{=} \sum_i I(M, X_i; Y_i|J_{i+1}^n, Y^{i-1}) - \sum_i I(X_i; J_i|J_{i+1}^n, Y^{i-1}) \\
&\overset{(d)}{\leq} \sum_i I(X_{Ri}, X_i; Y_i|J_{i+1}^n, Y^{i-1}) - \sum_i I(X_i; J_i|J_{i+1}^n, Y^{i-1}) \\
&= \sum_i I(X_{Ri}, X_i; Y_i|W_i) - \sum_i I(X_i; J_i|W_i) \\
&= \sum_i I(X_{Ri}, X_i; Y_i|W_i) - \sum_i I(X_{Ri}, X_i; J_i|W_i)
\end{aligned}$$

where (a) follows from Lemma 1,6, (b) follows from the fact that X_i is a function of M and (c), (d) follow from the Markov chains $(J_{i+1}^n, Y^{i-1}, M) \to X_i \to J_i$ and $(J_{i+1}^n, Y^{i-1}, M) \to (X_i, X_{Ri}) \to Y_i$ (established in Lemma 2), respectively. Finally we set

$$W_i = (J_{i+1}^n, Y^{i-1}). \tag{4}$$

Next, observe that

$$\begin{aligned}
I(M; J^n) &\leq I(M; J^n, Y_R^n) \\
&= \sum_{i=1}^{n} I(M; J_i, Y_{Ri}|J^{i-1}, Y_R^{i-1}, X_{R_i}) && (\because X_{Ri} = f_i(Y_R^{i-1})) \\
&\leq \sum_{i=1}^{n} I(M, J^{i-1}, Y_R^{i-1}; J_i, Y_{Ri}|X_{R_i}) \\
&\leq \sum_{i=1}^{n} I(X_i; J_i, Y_{Ri}|X_{R_i}) && \text{(Lemma 2).}
\end{aligned}$$

Thus we get

$$n(R-\epsilon_n) \leq I(M;Y^n) \leq \sum_{i=1}^{n} \left(I(X_i;J_i,Y_{Ri}|X_{R_i}) + I(X_{Ri},X_i;Y_i|W_i) - I(X_{Ri},X_i;J_i|W_i)\right).$$

This equation, along with the two cut-set constraints give the desired result. □

Lemma 2. *For the joint distribution on $(M,X^n,X_R^n,Y^n,Y_R^n,J^n)$ as induced by the code, the memoryless relay channel, and the auxiliary channel ($T_{J|Y,Y_R,X,X_R} \in \mathcal{F}$) as described by $p_{J^n|M,X^n,X_R^n,Y^n,Y_R^n} = \prod_{i=1}^n T_{J_i|X_i,X_{Ri},Y_i,Y_{Ri}}$ (thus defining an extended memoryless relay channel), the following Markov chains hold for $i \in [1:n]$:*

(1) $(X_i,X_{Ri},J_{i+1}^n) \to (X_i,X_{Ri}) \to (Y_i,Y_{Ri},J_i)$
(2) $(M,J^{i-1},Y_R^{i-1},Y^{i-1},J_{i+1}^n,X_{Ri}) \to X_i \to J_i$.

Proof. The second Markov chain follows from the first since $T_{J|Y,Y_R,X,X_R} \in \mathcal{F}$ implies that $(X_i,X_{Ri}) \to X_i \to J_i$ is Markov. Hence it remains to prove the first Markov chain. The memorylessness of the extended relay channel is equivalent to

$$(M,X^{i-1},X_R^{i-1},J^{i-1},Y_R^{i-1},Y^{i-1}) \to (X_i,X_{Ri}) \to (Y_i,Y_{Ri},J_i)$$

is Markov. Therefore it suffices to show that $I(J_{i+1}^n;Y_i,Y_{Ri},J_i|X_i,X_{Ri},M,J^{i-1},Y_R^{i-1},Y^{i-1}) = 0$. However note that

$$\begin{aligned}
&I(J_{i+1}^n;Y_i,Y_{Ri},J_i|X_i,X_{Ri},M,J^{i-1},Y_R^{i-1},Y^{i-1})\\
&\quad= \sum_{k=i+1}^{n} I(J_k;Y_i,Y_{Ri},J_i|X_i,X_{Ri},M,J^{i-1},Y_R^{i-1},Y^{i-1},J_{i+1}^{k-1})\\
&\quad= \sum_{k=i+1}^{n} I(J_k;Y_i,Y_{Ri},J_i|X_k,X_i,X_{Ri},M,J^{i-1},Y_R^{i-1},Y^{i-1},J_{i+1}^{k-1}) \qquad (\because X_k = g_k(M)).
\end{aligned}$$

Again since the extended relay channel is memoryless we know that $(M,X^{k-1},X_R^{k-1},J^{k-1},Y_R^{k-1},Y^{k-1}) \to (X_k,X_{Rk}) \to J_k$ is Markov and since in-addition we also have $X_{Rk} \to X_k \to J_k$ is Markov (by assumption on $T_{J|Y,Y_R,X,X_R}$), we obtain $(M,X^{k-1},X_R^{k-1},J^{k-1},Y_R^{k-1},Y^{k-1},X_{Rk}) \to X_k \to J_k$ is Markov. Hence

$$I(J_k;Y_i,Y_{Ri},J_i|X_k,X_i,X_{Ri},M,J^{i-1},Y_R^{i-1},Y^{i-1},J_{i+1}^{k-1}) = 0$$

for all $k \in [i+1:n]$, completing the proof. □

Corollary 1. *Consider the special case $T(y,y_R|x,x_R) = T_a(y_R|x)T_b(y|x,x_R,y_R)$ which includes, for instance, the Gaussian relay channels. For this class, any achievable rate R must satisfy*

$$R \leq \min\big(I(X,X_R;Y),\ I(X;Y,Y_R|X_R),\ I(X;Y_R|X_R) + \sup_{W \to X,X_R \to Y,Y_R} I(X,X_R;Y|W) - I(X,X_R;Y_R|W)\big), \tag{5}$$

for some $p(x,x_R)$.

Proof. For this class, since $X_R - X - Y_R$ is Markov, we can set $J = Y_R$, and the bound in (31) implies the result. □

Remark 2. This outer bound can be compared with the partial decode-and-forward lower bound which states that a rate R is achievable if

$$R \leq \min\big(I(X,X_R;Y),\ I(X;Y_R|X_R) + \sup_{W \to X,X_R \to Y,Y_R} I(X;Y|W,X_R) - I(X;Y_R|W,X_R)\big), \tag{6}$$

for some $p(x,x_R)$. More specifically, the last term in (6) is the same as that in (5) except that we have replaced W by (W,X_R).

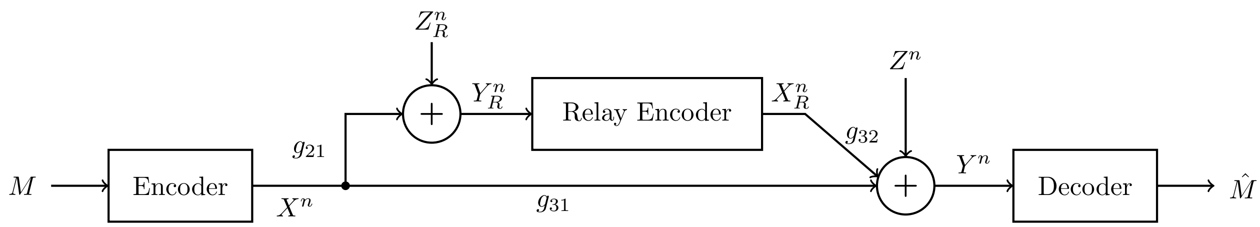

FIGURE 2. Depiction of a Gaussian relay channel.

Remark 3. A weaker form of the bound in (31) which does not involve auxiliary random variables is

$$R \leq \min\left(I(X, X_R; Y),\ I(X; Y, Y_R|X_R),\ I(X; J, Y_R|X_R) + I(X, X_R; Y|J)\ \right) \tag{7}$$

for any arbitrary $T_{J|Y,Y_R,X,X_R} \in \mathcal{F}$. This follows from the following:

$$I(X, X_R; Y|W) - I(X, X_R; J|W) \leq I(X, X_R; Y|JW) \leq I(X, X_R; Y|J).$$

In the next section we will show that the upper bound given in Theorem 1 can be strictly better than the cut-set bound for the Scalar Gaussian relay channel.

2.1. **Scalar Gaussian Relay Channel.** Consider a scalar Gaussian relay channel described by

$$Y_R = g_{21}X + Z_1, \quad Y = g_{31}X + g_{32}X_R + Z_2$$

where non-negative reals g_{21}, g_{31}, g_{32} are channel gains and Z_1 and Z_2 are independent standard Gaussian random variables. We assume that the power constraints on X_1 and X_2 are both given by P. This is depicted in Fig. 2.1. Let $S_{21} = g_{21}^2 P$, $S_{31} = g_{31}^2 P$ and $S_{32} = g_{32}^2 P$. Finally, let $\mathsf{C}(S) = \frac{1}{2}\ln(1+S)$. Using Theorem 1, we obtain the following the following upper bound:

Proposition 1. *The capacity of the relay channel is bounded from above by*

$$\min\Bigg(\mathsf{C}(S_{31}+S_{32}+2\rho\sqrt{S_{31}S_{32}}),\ \mathsf{C}((1-\rho^2)(S_{31}+S_{21})),$$
$$\min_{S_J \geq S_{21}}\left\{\mathsf{C}(S_J(1-\rho^2)) + \mathsf{C}\left(x_*^2 + \left(x_*\left(\sqrt{S_{31}}+\sqrt{\rho^2 S_{32}}\right)+\sqrt{(1-\rho^2)S_{32}}\right)^2\right) - \mathsf{C}\left(x_*^2(1+S_J)\right)\right\}\Bigg), \tag{8}$$

for some $\rho \in [-1,1]$, where x_* is the unique non-negative root of the quadratic equation:

$$x^2(\sqrt{S_{31}}+\sqrt{\rho^2 S_{32}})(\sqrt{(1-\rho^2)S_{32}})(1+S_J) + x\left((1-\rho^2)S_{32}(1+S_J) - (\sqrt{S_{31}}+\sqrt{\rho^2 S_{32}})^2 + S_J\right)$$
$$-(\sqrt{S_{31}}+\sqrt{\rho^2 S_{32}})(\sqrt{(1-\rho^2)S_{32}}) = 0. \tag{9}$$

The proof of the above proposition is given in Appendix B.1. The choice of J used in the proof is an enhancement of Y_R, *i.e.,* $X \to J \to Y_R$ forms a Markov chain.[1]

Observe that the first two terms in (8) correspond to the cut-set bound [EK12, Eq. 16.4]:

$$C \leq \max_{-1\leq\rho\leq 1}\min\left(\mathsf{C}(S_{31}+S_{32}+2\rho\sqrt{S_{31}S_{32}}),\ \mathsf{C}((1-\rho^2)(S_{31}+S_{21})),\right)$$
$$= \begin{cases} \mathsf{C}\left(\frac{\left(\sqrt{S_{21}S_{32}}+\sqrt{S_{31}(S_{31}+S_{21}-S_{32})}\right)^2}{S_{31}+S_{21}}\right) & \text{if } S_{21}\geq S_{32} \\ \mathsf{C}(S_{31}+S_{21}) & \text{otherwise.}\end{cases} \tag{10}$$

[1]The more restrictive choice of $J = Y_R$ is insufficient to obtain the stated bound in Proposition 2 on the slope. Moreover, in the conference version of this work, we provide a different example of a Gaussian MIMO relay channel where J is not an enhancement of Y_R, but is taken as part of Y.

Example 1. As an specific example, set $S_{21} = 1.2139$, $S_{31} = 3.7585$, $S_{32} = 0.032519$ and $S_J = S_{21}$. Then, the cut-set bound evaluates to 0.81327 (with maximizing $\rho = 0.4221655$), while the new upper bound is 0.79488 (with maximizing $\rho = 0.159498$). On the other hand, from the compress-and-forward lower bound, we know that the capacity is greater than or equal to [EK12, Eq. 16.12]

$$\mathsf{C}\left(S_{31} + \frac{S_{21}S_{32}}{S_{31}+S_{21}+S_{32}+1}\right).$$

This expression evaluates to 0.78066 for this example. The decode-and-forward evaluates to 0.39737.

2.2. **On the derivative of the capacity at $S_{32} = 0$.** Let $C(S_{21}, S_{31}, S_{32})$ denote the capacity of the scalar Gaussian relay channel with the given parameters. If $S_{32} = 0$, the link from the relay to the receiver is disabled and it becomes a point-to-point channel. Therefore $C(S_{21}, S_{31}, 0) = \mathsf{C}(S_{31})$ (achieved via the direct-transmission). We are interested in the derivative of $C(S_{21}, S_{31}, S_{32})$ with respect to S_{32} at $S_{32} = 0$, at some $S_{31} > S_{21} > 0$.

Assume that $S_{31} > S_{21}$ while S_{32} is small. Then, the decode-and-forward lower bound equals $\mathsf{C}(S_{21})$ [EK12, Eq. 16.6] which is weaker than the direct-transmission lower bound. On the other hand, the bound from the compress-and-forward [EK12, Eq. 16.12] equals

$$\mathsf{C}\left(S_{31} + \frac{S_{21}S_{32}}{S_{31}+S_{21}+S_{32}+1}\right) = \mathsf{C}\left(S_{31} + \frac{S_{21}}{S_{31}+S_{21}+1}S_{32} + \mathcal{O}(S_{32}^2)\right).$$

This implies that

$$\frac{\partial}{S_{32}}C(S_{21}, S_{31}, S_{32})\big|_{S_{32}=0} \geq \frac{1}{2}\frac{S_{21}}{(1+S_{31})(S_{31}+S_{21}+1)}.$$

On the other hand the cut-set bound, see (10), is given by:

$$\mathsf{C}\left(\frac{\left(\sqrt{S_{21}S_{32}} + \sqrt{S_{31}(S_{31}+S_{21}-S_{32})}\right)^2}{S_{31}+S_{21}}\right) = \mathsf{C}\left(S_{31} + \frac{2\sqrt{S_{21}S_{31}}}{\sqrt{S_{31}+S_{21}}}\sqrt{S_{32}} + \mathcal{O}(S_{32})\right).$$

It has an infinite slope with respect to S_{32}, at $S_{32} = 0$ since the first order term is $\sqrt{S_{32}}$. The intuitive reason for the appearance of $\sqrt{S_{32}}$ is that the cut-set bound allows for cooperation between the relay and the transmitter.

The cut-set bound fails to provide any finite bound on the derivative of the capacity with respect to S_{32} at $S_{32} = 0$. However, the new bound gives a finite slope result:

Proposition 2. *For $S_{31} > S_{21}$, the derivative of the new upper bound with respect to S_{32} at $S_{32} = 0$ is less than or equal to*

$$\frac{1}{2}\frac{S_{31}^2(1+S_J)^2}{(1+S_{31})^2 S_J(S_{31}-S_J)}$$

where

$$S_J = \max\left(S_{21}, \frac{S_{31}}{S_{31}+2}\right).$$

The proof of the above proposition is given in Appendix B.2. See Fig. 3.1 for an illustration.

3. Interference Channel

A two-user interference channel models transmission of messages from two senders X_1 and X_2 to two receivers Y_1 and Y_2. An interference channel is described by a conditional distribution $T(y_1, y_2|x_1, x_2)$. A non-negative rate pair (R_1, R_2) is said to be achievable if the transmitter X_i is able to send a message at rate R_i to receiver Y_i such that the probability of error tends to zero as n, the blocklength, tends to infinity (see Fig. 3). The closure of the union of all achievable rate pairs is called the capacity region for the interference channel $T(y_1, y_2|x_1, x_2)$. A number of different outer bounds are known in the literature on the capacity of a general interference channel $T(y_1, y_2|x_1, x_2)$ (see Chapters 6 in [EK12] for an overview of interference channels). One can attempt to write the J-version of each of these bounds. As we aim to simply illustrate the use of J bounds, we only report one such bound here even though we were also able to write J versions of the outer bound given in [EO11] as well.

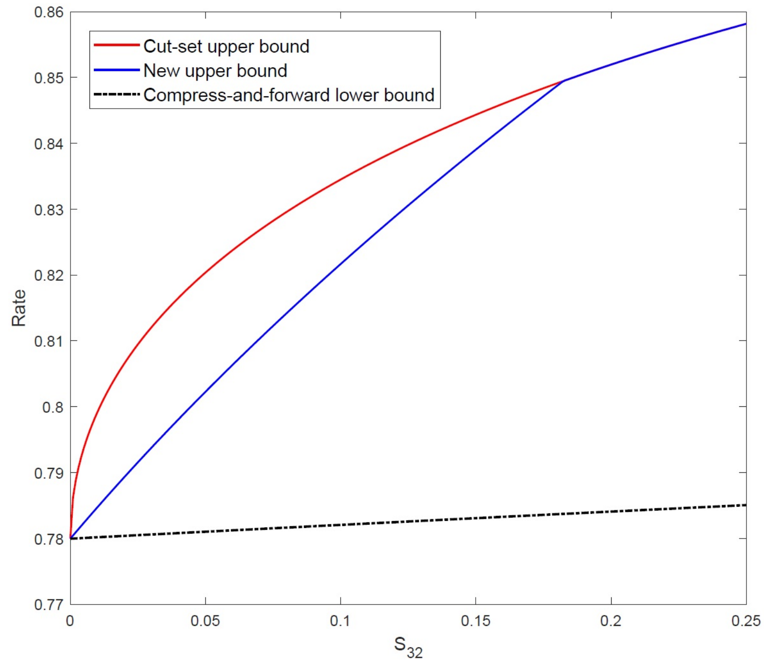

FIGURE 3. Illustration of the bounds for a Gaussian relay channel. Parameters are $S_{31} = 3.7585$, $S_{21} = 1.2139$.

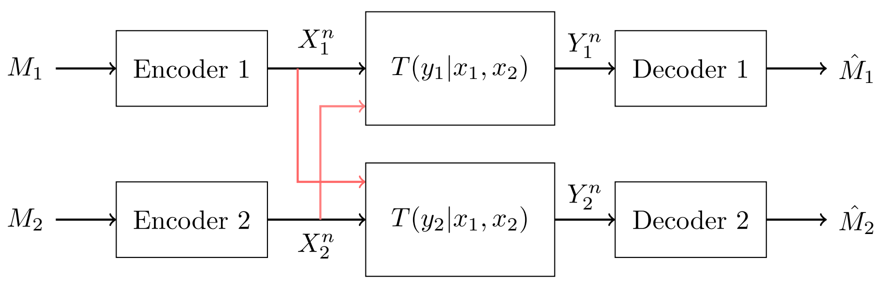

FIGURE 4. The setup for an interference channel.

Theorem 2. *Take an arbitrary interference channel* $T(y_1, y_2|x_1, x_2)$*. If* (R_1, R_2) *is achievable, then for any* $T_{J|X_1X_2Y_1Y_2}$ *such that* $p_{JY_1Y_2|X_1X_2} = T_{Y_1Y_2|X_1X_2}T_{J|X_1X_2Y_1Y_2}$ *satisfies*

$$p_{JY_1Y_2|X_1X_2} = p_{J|X_1}p_{Y_2|JX_2}p_{Y_1|JX_1X_2Y_2}, \tag{11}$$

we have

$$\begin{aligned}
R_1 &\leq \min(I(X_1; Y_1|X_2, Q), I(W; Y_2|Q) + I(X_1; J|W, Q) + I(X_1, X_2; Y_1|\hat{W}, Q) - I(X_1; J|\hat{W}, Q)), \\
R_2 &\leq \min(I(X_2; Y_2|W, X_1, Q), I(X_2; Y_2|W, Q) - I(X_2; J|W, Q)),
\end{aligned}$$

for some $p(q)p(x_1|q)p(x_2|q)p(w, \hat{w}|x_1, x_2, q)$ *satisfying*

$$I(X_1; J|W, Q) \geq I(X_1; Y_2|W, Q).$$

Further it suffices to consider $|\mathcal{Q}| \leq 4$, $|\mathcal{W}| \leq |\mathcal{X}_1||\mathcal{X}_2| + 2, |\hat{\mathcal{W}}| \leq |\mathcal{X}_1||\mathcal{X}_2|$.

Proof. The cardinality bounds on the auxiliary variables $W, \hat{W}$ follow from standard arguments and is omitted. One can prove the above theorem by identifying the auxiliary variable

$$W_i = (Y_2^{i-1}, J_{i+1}^n), \hat{W}_i = (Y_1^{i-1}, J_{i+1}^n)$$

for a given code (n, M_1, M_2). Let $Q \in [n]$ be a time-sharing random variable and let $W = (W_Q, Q)$ and $\hat{W} = (\hat{W}_Q, Q)$. The first bound $R_1 \leq I(X_1; Y_1|X_2, Q)$ is standard.

Observe that by Fano's inequality we have $n(R_1 - \epsilon_n) \leq I(M_1; Y_1^n) \leq I(X_1^n; Y_1^n)$ for some ϵ_n that tends to zero as n tends to infinity. Now, observe that

$$\begin{aligned}
I(X_1^n; Y_1^n) &= I(X_1^n; J^n) + I(X_1^n; Y_1^n) - I(X_1^n; J^n) \\
&\leq I(X_1^n; J^n) + I(X_1^n, X_2^n; Y_1^n) - I(X_1^n, X_2^n; J^n) && \because X_2^n \perp (X_1^n, J^n) \\
&= \sum_{i=1}^n \Big(I(X_1^n; J_i|J_{i+1}^n) + I(X_1^n, X_2^n; Y_{1i}|\hat{W}_i) - I(X_1^n, X_2^n; J_i|\hat{W}_i) \Big) && \text{(Lemma 1, 6)} \\
&= \sum_{i=1}^n \Big(I(X_{1i}; J_i|J_{i+1}^n) + I(X_{1i}, X_{2i}; Y_{1i}|\hat{W}_i) - I(X_{1i}; J_i|\hat{W}_i) \Big) && \text{(memoryless channels)} \\
&\leq \sum_{i=1}^n \Big(I(X_{1i}; J_i|W_i) + I(W_i; Y_{2i}) + I(X_{1i}, X_{2i}; Y_{1i}|\hat{W}_i) - I(X_{1i}; J_i|\hat{W}_i) \Big) && \text{(Lemma 1, 6)}
\end{aligned}$$

establishing the second bound on R_1.

Next, consider the bounds on R_2. Observe that by Fano's inequality we have $n(R_2 - \epsilon_n) \leq I(M_2; Y_2^n) \leq I(X_2^n; Y_2^n)$ for some ϵ_n that tends to zero as n tends to infinity. Observe that

$$\begin{aligned}
I(X_2^n; Y_2^n) &\leq I(X_2^n; Y_2^n|X_1^n) && \because X_2^n \perp X_1^n \\
&= \sum_{i=1}^n I(X_2^n; Y_{2i}|X_1^n, Y_2^{i-1}, J_{i+1}^n) && \because J_{i+1}^n \to X_{1i+1}^n \to (X_2^n, X_1^i) \to (Y_1^i, Y_2^i) \\
&\leq \sum_{i=1}^n I(X_{2i}; Y_{2i}|X_{1i}, W_i) && \text{(memoryless channels)}
\end{aligned}$$

establishing the first bound on R_2.

To get the second bound, observe that

$$\begin{aligned}
I(M_2; Y_2^n) &= I(M_2; Y_2^n) - I(M_2; J^n) && \because M_2 \perp J^n \\
&= \sum_{i=1}^n I(M_2; Y_{2i}|Y_2^{i-1}, J_{i+1}^n) - I(M_2; J_i|Y_2^{i-1}, J_{i+1}^n) && \text{(Lemma 1)} \\
&= \sum_{i=1}^n I(M_2, X_{2i}; Y_{2i}|Y_2^{i-1}, J_{i+1}^n) - I(M_2, X_{2i}; J_i|Y_2^{i-1}, J_{i+1}^n) && \because X_{2i} = f_{2i}(M_2) \\
&\stackrel{(a)}{=} \sum_{i=1}^n I(X_{2i}; Y_{2i}|Y_2^{i-1}, J_{i+1}^n) - I(X_{2i}; J_i|Y_2^{i-1}, J_{i+1}^n) + I(M_2; Y_{2i}|Y_2^{i-1}, J_{i+1}^n, X_{2i}) \\
&\qquad - I(M_2; J_i, Y_{2i}|Y_2^{i-1}, J_{i+1}^n, X_{2i}) \\
&\leq \sum_{i=1}^n I(X_{2i}; Y_{2i}|W_i) - I(X_{2i}; J_i|W_i),
\end{aligned}$$

where (a) follows since $(Y_2^{i-1}, J_{i+1}^n, M_2) \to (X_{2i}, J_i) \to Y_{2i}$ is Markov from (11). This established the second bound on R_2.

Finally, to show that $I(X_1; J|W, Q) \geq I(X_1; Y_2|W, Q)$, observe

$$\begin{aligned}
0 &\leq I(X_1^n; J^n X_2^n) - I(X_1^n; Y_2^n) && \because X_1^n \to J^n \to (X_2^n, Y_2^n) \\
&= I(X_1^n; J^n) - I(X_1^n; Y_2^n) && \because (X_1^n, J^n) \perp X_2^n \\
&= \sum_i I(X_1^n; J_i|W_i) - I(X_1^n; Y_{2i}|W_i) && \text{(Lemma 1)}
\end{aligned}$$

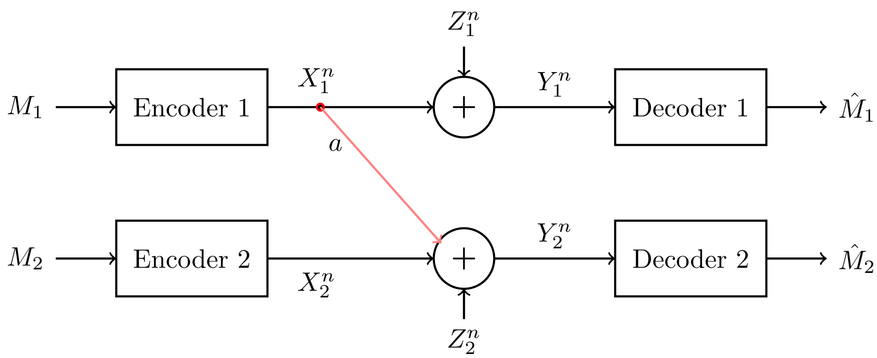

FIGURE 5. Illustration of a Gaussian Z-interference channel.

$$\begin{aligned} &= \sum_i I(X_{1i}; J_i|W_i) - I(X_1^n; Y_{2i}|W_i) && \text{(memoryless channel)} \\ &\leq \sum_i I(X_{1,i}; J_i|W_i) - I(X_{1i}; Y_{2i}|W_i), \end{aligned}$$

completing the proof of the constraint. □

Consider a Z-interference channel, i.e. $T(y_1, y_2|x_1, x_2) = T_1(y_1|x_1)T_2(y_2|x_1, x_2)$. For such a channel, we can simplify the outer bound in Theorem 2 as follows:

Corollary 2. *Let J be an auxiliary receiver defined by the channel $T_{J|X_1,X_2,Y_1,Y_2}$ such that $p_{JY_1Y_2|X_1X_2} = T_{Y_1Y_2|X_1X_2}T_{J|X_1X_2Y_1Y_2}$ satisfies*

$$p_{JY_1Y_2|X_1X_2} = p_{J|X_1}p_{Y_2|JX_2}p_{Y_1|JX_1},$$

and further let J be more-capable (as in [KM75]) than Y_1. Then, any rate pair (R_1, R_2) in the capacity of the Z-interference channel must satisfy the following constraints,

$$\begin{aligned} R_1 &\leq \min\{I(X_1; Y_1|Q), I(W; Y_2|Q) + I(X_1; J|W, Q)\} \\ R_2 &\leq \min\{I(X_2; Y_2|W, X_1, Q), I(X_2; Y_2|W, Q) - I(X_2; J|W, Q)\} \end{aligned}$$

for some $p(q)p(x_1|q)p(x_2|q)p(w|x_1, x_2, q)$ satisfying

$$I(X_1; J|W, Q) \geq I(X_1; Y_2|W, Q).$$

Proof. Due to the Z-nature of the interference channel observe that

$$I(X_1, X_2; Y_1|\hat{W}, Q) - I(X_1; J|\hat{W}, Q) = I(X_1; Y_1|\hat{W}, Q) - I(X_1; J|\hat{W}, Q) \leq 0,$$

where the last inequality is a consequence of J being more-capable than Y_1. □

3.1. **Gaussian Z-interference Channel.** Consider the two-user Z-Gaussian interference channel (GIC):

$$\begin{aligned} Y_1 &= X_1 + Z_1 \\ Y_2 &= aX_1 + X_2 + Z_2\,, \end{aligned} \tag{12}$$

with $a \geq 0$, $Z_i \sim \mathcal{N}(0, 1)$ and a power constraint on the n-letter codebooks:

$$\|X_1\|^2 \leq nP_1, \quad \|X_2\|^2 \leq nP_2. \tag{13}$$

See Fig. 3.1 for an illustration.

Theorem 3. *Take some arbitrary $\lambda \geq 1$ and $u, \alpha, \beta \in [0,1]$. Then, any achievable rate pair (R_1, R_2) for the scalar Gaussian interference channel with power constraints P_1, P_2 respectively must satisfy*

$$R_1 + \lambda R_2 \leq \frac{\lambda\alpha}{2}\log(K_2(1-\rho^2)+1) + \frac{\beta}{2}\log(1+P_1) + \frac{1-\beta}{2}\log\left(\frac{1+a^2P_1+P_2}{u^2}\right)$$
$$+\left(\frac{\lambda(1-\alpha)}{2} - \frac{(1-\beta)}{2}\right)\left(\log\frac{1+a^2K_1+K_2+2a\rho\sqrt{K_1K_2}}{K_1+u^2}\right) + \frac{\lambda(1-\alpha)}{2}\log\frac{K_1(1-\rho^2)+u^2}{a^2K_1(1-\rho^2)+1},$$

for some $K_1 \leq P_1$ and $K_2 \leq P_2$ and $\rho \in [-1,1]$ such that

$$(P_1-K_1)(P_2-K_2) \geq \rho^2 K_1 K_2,$$

and

$$\frac{K_1+u^2}{u^2} \geq \frac{1+a^2K_1+K_2+2a\rho\sqrt{K_1K_2}}{1+K_2(1-\rho^2)}. \tag{14}$$

Remark 4. The proof follows by showing the Gaussian optimality for the outer bound in Corollary 2. The proof technique used to show the Gaussian optimality is the one employed in [GN14], where a "Gallager-type" proof of sub-additivity (or single-letterization) of information-theoretic functionals is used to deduce a certain independence between two orthogonally rotated independent copies of the maximizing distribution, thereby establishing Gaussianity by the Skitovic-Darmois characterization. Sometimes the "Gallager-type" proof for the functional may not directly yield the requisite independence and so one has to consider perturbed functionals that have Gaussian maximizers and then one is able to use continuity to argue that Gaussians are a maximizer for the given functional, as in the case below.

Proof. The proof is given in Appendix C.1. ☐

Figure 3.1 plots the outer bound for $a = 0.8, P_1 = P_2 = 1$ for fixed choice of $\beta = 0$ and $u = 1$ in Theorem 3. Note that the curve passes through both the non-trivial corner points of the capacity region of the Gaussian Z-interference channel. This is formally established Lemma 3 below.

3.1.1. *On the slope of the capacity region at Costa's corner point.* Let $C_2 = \mathsf{C}(P_2)$, the maximum achievable rate at receiver Y_2. Costa [Cos85] "determined" the maximum value of R_1 such that (R_1, C_2) is achievable an $R_1^* = \mathsf{C}\left(\frac{a^2P_1}{1+P_2}\right)$; however Sason (see [Sas15] for a detailed discussion) observed that the proof of a certain lemma, regarding the continuity of differential entropy, had an error. Polyanskiy and Wu, [PW15] using the HWI inequality as the central piece, completed the continuity of entropy argument and established that $R_1^* = \mathsf{C}\left(\frac{a^2P_1}{1+P_2}\right)$ is indeed the maximum value of R_1 such that (R_1, C_2) is achievable. However, similar to the cut-set bound situation in Section 2.2, the outer bound derived by Polyanskiy-Wu bound does not show if the corner point is an exposed point or an extreme point of the capacity region. In Theorem 4 we show that Corollary 2 not only recovers the corner point, but also establishes that it is an exposed point[2] of the capacity region, thereby improving on the Polyanskiy-Wu bound. Further, we also show that it is better than Sato's outer bound for the interference channel (which is optimal at the other corner point).

Lemma 3. *Let $\mathcal{R}_{OB}$ denote the outer bound given in Theorem 3. The following hold:*

(i) *If $(R_1, C_2) \in \mathcal{R}_{OB}$, then*
$$R_1 \leq R_1^* = \frac{1}{2}\log\left(1+\frac{a^2P_1}{1+P_2}\right).$$

(ii) *Further, the outer bound given in Theorem 3 lies inside the outer bound by Sato ([Sat78]; see also Theorem 2 in [Kra04]). Consequently if $(C_1, R_2) \in \mathcal{R}_{OB}$, then*
$$R_2 \leq R_2^* = \frac{1}{2}\log\left(1+\frac{P_2}{1+a^2P_1}\right).$$

[2] For the Han and Kobayashi achievable region [HK81] with Gaussian signaling (whose optimality or sub-optimality is not yet determined for the Gaussian interference channel), it is known that the above corner point is an exposed point and in [CN16] the (non-trivial) slope of the above region at the corner point was computed.

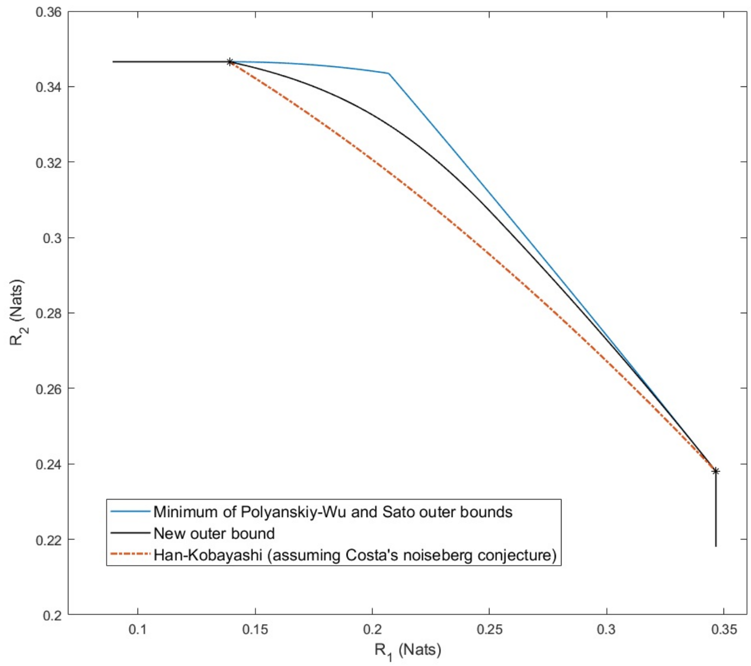

FIGURE 6. Illustration for $a = 0.8$, $P_1 = P_2 = 1$. The new outer bound is plotted for the fixed choice of $\beta = 0$ and $u = 1$ in Theorem 3 and optimized over α. The Polyanskiy-Wu bound is flat at Costa's corner point. However, Theorem 4 establishes that the capacity region has a kink at Costa's corner point, with the slope of the capacity region at Costa's corner point being less than or equal to -0.1323. The slope of the Han-Kobayashi region with Gaussian signaling [CN16] equals -0.3839 for this example.

Proof. Suppose $R_2 = C_2 = \frac{1}{2}\log(1+P_2)$, then as

$$\frac{1}{2}\log(1+P_2) = C_2 \leq I(X_2;Y_2|W,X_1,Q) = I(X_2;Y_2|X_1,Q) - I(W;Y_2|X_1,Q) \leq \frac{1}{2}\log(1+P_2) - I(W;Y_2|X_1,Q),$$

it immediately follows that $X_2 \sim \mathcal{N}(0,P_2)$, is independent of (X_1,Q) and $I(W;Y_2|X_1,Q) = 0$. The last inequality further implies $I(W;X_2|X_1,Q) = 0$ (see Proposition 2 in [GN14]). Hence $X_2 \perp (W,X_1,Q)$ and $X_2 \perp (W,X_1,Q,J)$. Observe that

$$\begin{aligned} C_2 &\leq I(X_2;Y_2|W,Q) - I(X_2;J|W,Q) \\ &= I(X_2;Y_2|W,Q) \\ &= I(X_2;Y_2|W,X_1,Q) - I(X_2;X_1|W,Y_2,Q) \\ &\leq C_2 - I(X_2;X_1|W,Y_2,Q). \end{aligned}$$

implying $I(X_2;X_1|W,Y_2,Q) = 0$, which says that $I(X_1;aX_1+Z|W,Y_2,Q) = 0$. Since $I(Y_2;X_1|aX_1+Z,W,Q) = 0$, we have a double Markovity property [CK11]. From the double Markovity property and since the joint distribution of $(\tilde{Y}_1,Y_2)$ is indecomposable, we obtain that conditioned on (W,Q), $X_1 \perp (aX_1+Z,X_2)$. This implies, conditioned on (W,Q), that X_1 is independent of $X_1 + \frac{Z}{a}$. This implies that X_1 is a constant conditioned on (W,Q). Consequently $R_1 \leq I(W;Y_2|Q) = I(W,X_1;Y_2|Q) = I(X_1;Y_2|Q) \leq \frac{1}{2}\log\left(1+\frac{a^2P_1}{1+P_2}\right)$. This establishes the first part.

Sato established (see Theorem 2 in [Kra04]) that any achievable rate pair for the interference channel must satisfy for $\lambda \geq 1$

$$\lambda R_2 + R_1 \leq \frac{\lambda}{2}\log(1+a^2P_1+P_2) + \max_{K_1 \leq P_1}\left\{\frac{1}{2}\log(1+K_1) - \frac{\lambda}{2}\log(1+a^2K_1).\right\}$$

In particular if $1 \leq \lambda \leq \frac{1+a^2P_1}{a^2(1+P_1)}$ it is immediate that the above bound evaluates to

$$\frac{\lambda}{2}\log\left(1+\frac{P_2}{1+a^2P_1}\right) + \frac{1}{2}\log(1+P_1)$$

implying that it passes through (C_1, R_2^*).

From Theorem 3, putting $\beta = 0, \alpha = 0, u = 1$ we see that

$$R_1 + \lambda R_2 \leq \frac{1}{2}\log\left(\frac{1+a^2P_1+P_2}{1}\right) + \left(\frac{\lambda-1}{2}\right)\left(\log\frac{1+a^2K_1+K_2+2a\rho\sqrt{K_1K_2}}{K_1+1}\right) + \frac{\lambda}{2}\log\frac{K_1(1-\rho^2)+u^2}{a^2K_1(1-\rho^2)+1},$$

Therefore to compare the bounds it suffices to show that

$$\begin{aligned}
&\frac{1}{2}\log\left(\frac{1+a^2P_1+P_2}{1}\right) + \left(\frac{\lambda-1}{2}\right)\left(\log\frac{1+a^2K_1+K_2+2a\rho\sqrt{K_1K_2}}{K_1+1}\right) + \frac{\lambda}{2}\log\frac{K_1(1-\rho^2)+u^2}{a^2K_1(1-\rho^2)+1}\\
&\quad\leq \frac{\lambda}{2}\log(1+a^2P_1+P_2) + \frac{1}{2}\log(1+K_1) - \frac{\lambda}{2}\log(1+a^2K_1) \qquad \iff\\
&\left(\frac{\lambda-1}{2}\right)\left(\log(1+a^2K_1+K_2+2a\rho\sqrt{K_1K_2})\right) + \frac{\lambda}{2}\log\frac{K_1(1-\rho^2)+u^2}{a^2K_1(1-\rho^2)+1}\\
&\quad\leq \frac{\lambda-1}{2}\log(1+a^2P_1+P_2) + \frac{\lambda}{2}\log(1+K_1) - \frac{\lambda}{2}\log(1+a^2K_1).
\end{aligned}$$

This is immediate since

$$a^2(P_1-K_1) + (P_2-K_2) - 2a\rho\sqrt{K_1K_2} \geq a^2(P_1-K_1) + \frac{\rho^2K_1K_2}{P_1-K_1} - 2a\rho\sqrt{K_1K_2} \geq 0.$$

This completes the proof. □

We now establish a significantly stronger result regarding Costa's corner point by using Theorem 3.

Theorem 4. *Let $\mathcal{R}_{OB}$ denote the outer bound given in Theorem 3. Let $C_2 = \frac{1}{2}\log(1+P_2)$ and $R_1^* = \frac{1}{2}\log\left(1+\frac{a^2P_1}{1+P_2}\right)$. Then*

$$\max_{(R_1,R_2)\in\mathcal{R}_{OB}} \lambda R_2 + R_1 = \lambda C_2 + R_1^*$$

when

$$\lambda \geq 1 + \begin{cases} \frac{(1+P_2)(1-a^2)}{a^2P_2}\frac{\left(1+\sqrt{1+4a^2(1-a^2)P_2}\right)^2}{4a^2(1-a^2)P_2} & a^2 < \frac{1}{2} \\ \frac{(1+P_2)(1-a^2)}{a^2P_2}\frac{\left(1+\sqrt{1+P_2}\right)^2}{P_2} & a^2 \geq \frac{1}{2} \end{cases}.$$

Proof. The proof is presented in Appendix C.2. □

Remark 5. Exact computation of the outer bound in Theorem 3 is reasonably involved numerically. Even though it passes through both the corner points, the authors find no reason to believe that this matches with the Han–Kobayashi inner bound (with Gaussian signaling). If one considers Han-Kobayashi acheivable region $\mathcal{R}_{HK-GS}$ with Gaussian signaling, it was shown in [CN16] that $\max_{(R_1,R_2)\in\mathcal{R}_{HK-GS}} \lambda R_2 + R_1 = \lambda C_2 + R_1^*$ if and only if

$$\lambda \geq 1 + \max\left\{\frac{-\log a^2 - \frac{1-a^2}{(1+a^2P_1+P_2)}}{\log(1+P_2) - \frac{P_2}{1+P_2}}, \frac{(1+P_2)(1-a^2)}{a^2P_2}\right\}.$$

Note that the arguments in the Appendix can be seen to yields stronger result (that also involves P_1) for the outer bound but does not match the above value (except in the limiting case, when $P_1, P_2 \to \infty$). However, it is quite possible that a modification of the structure of the auxiliary receiver may close the gap.

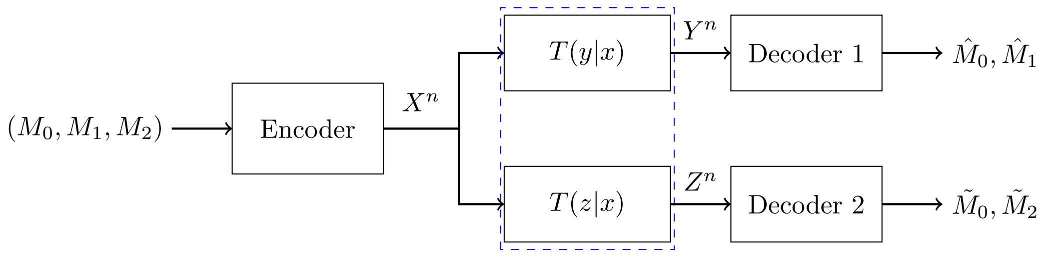

FIGURE 7. Illustration of a discrete memoryless broadcast channel.

4. BROADCAST CHANNEL

A two-receiver broadcast channel [Cov72] models transmission of messages from a single sender X to two receivers Y and Z. A discrete broadcast channel is described by a conditional distribution $T(y, z|x)$ with $|\mathcal{X}|, |\mathcal{Y}|, |\mathcal{Z}| < \infty$ (see Fig. 4). A non-negative rate triple (R_0, R_1, R_2) is said to be achievable if the transmitter is able to send a common message at rate R_0 and two private messages at rates R_1 and R_2 to the receivers Y and Z such that the probability of error tends to zero as n, the blocklength, tends to infinity. The closure of the union of all achievable rate pairs is called the capacity region for the broadcast channel $T(y, z|x)$. For more details on this model, the definition of the capacity region, and a collection of known results please refer to Chapters 5 and 8 in [EK12].

The best known achievable rate region for a two-receiver broadcast channel is the following inner bound [Mar79].

Theorem 5 (Marton '79)**.** *The union of non-negative rate triples (R_0, R_1, R_2) satisfying the constraints*

$$\begin{aligned} R_0 &\leq \min(I(W;Y), I(W;Z)) \\ R_0 + R_1 &\leq I(U, W; Y), \\ R_0 + R_2 &\leq I(V, W; Z), \\ R_0 + R_1 + R_2 &\leq \min\{I(W;Y), I(W;Z)\} + I(U;Y|W) \\ &\quad + I(V;Z|W) - I(U;V|W), \end{aligned}$$

for any triple of random variables (U, V, W) such that $(U, V, W) \to X \to (Y, Z)$ is achievable.

It is not known whether this region is the true capacity region or do there exist channels whose capacity region is strictly larger than the above region. The situation with respect to the outer bounds for the capacity region of the two-receiver broadcast channel is the following: among the various forms of the outer bounds proposed (e.g. see [Nai11]), the UV-outer bound noted below has been the best known computable outer bound for the general two-receiver broadcast channel with private messages. The UV outer bound [El 79, NE07, Nai11] for the capacity region of the broadcast channel is as follows:

Theorem 6 (UV outer bound)**.** *Any achievable rate (R_0, R_1, R_2) satisfies the constraints*

$$\begin{aligned} R_0 &\leq \min(I(W;Y), I(W;Z)) \\ R_0 + R_1 &\leq \min(I(W;Y), I(W;Z)) + I(U;Y|W), \\ R_0 + R_2 &\leq \min(I(W;Y), I(W;Z)) + I(V;Z|W), \\ R_0 + R_1 + R_2 &\leq \min(I(W;Y), I(W;Z)) + \min(I(U;Y|W) + I(X;Z|UW), I(V;Z|W) + I(X;Y|VW)) \end{aligned}$$

for some pair of random variables (U, V, W) such that $(U, V, W) \to X \to (Y, Z)$.

In [GGNY14], the authors showed a class of product broadcast channels whose capacity region is strictly contained in the region given by the UV outer bound. However the outer bound developed in [GGNY14] that (strictly) improved on the UV outer bound was valid only for the class of product broadcast channels. In this paper we provide some outer bounds for (general) broadcast channels that both (strictly) improve over the

UV outer bound, and the second one generalizes the one in [GGNY14] while remove the constraint that the broadcast channel must have a product structure.

4.1. **The J version of UV Outer Bound.** Our first outer bound was motivated by the following question: assume that we make the Y and Z receivers weaker by passing them through an erasure channel, *i.e.,* by considering $p(y',z'|x)=\sum_{y,z}p(y'|y)p(z'|z)T(y,z|x)$ where $p(y'|y)$ and $p(z'|z)$ are erasure channels with erasure probability ϵ. Then, for any p_{UVWX}, we have

$$I(W;Y')=(1-\epsilon)I(W;Y),\quad I(U;Y'|W)=(1-\epsilon)I(U;Y|W),\quad I(X;Y'|VW)=(1-\epsilon)I(X;Y|VW),$$
$$I(W;Z')=(1-\epsilon)I(W;Z),\quad I(V;Z'|W)=(1-\epsilon)I(V;Z|W),\quad I(X;Z'|UW)=(1-\epsilon)I(X;Z|UW).$$

Therefore, the UV outer bound scales by $1-\epsilon$ for an erased broadcast channel. However, Marton's inner bound involves a term $-I(U;V|W)$ in its sum-rate constraint which does not (immediately) scale by $1-\epsilon$. This raises the question of whether the capacity region scales by $1-\epsilon$ or not. Our first (new) outer bound below shows that the capacity region does not scale by $1-\epsilon$ for any $\epsilon\in(0,1)$ for the example of an erased Blackwell broadcast channel (see Lemma 4).

Theorem 7. *Given a broadcast channel characterized by $T(y,z|x)$, any achievable rate triple (R_0,R_1,R_2) must satisfy the following constraints:*

$$R_0\le\min\{I(W;Y),I(\hat{W};Y),I(W;Z),I(\tilde{W};Z)\}\tag{15a}$$
$$R_0+R_1\le\min\{I(W;Y),I(W;Z)\}+I(U;Y|W),\tag{15b}$$
$$\begin{aligned}R_0+R_1\le\min\Big\{&I(\tilde{W};Z)+\min\big[0,I(W;Y)-I(W;Z)\big],\ I(\tilde{W};J)+I(\hat{W};Y)-I(\hat{W};J)\Big\}\\&+I(\tilde{U};J|\tilde{W})+I(\hat{U};Y|\hat{W})-I(\hat{U};J|\hat{W}),\end{aligned}\tag{15c}$$
$$\begin{aligned}R_0+R_1\le\min\Big\{&I(\hat{W};Y)+\min\big[0,I(W;Z)-I(W;Y)\big],\ I(\hat{W};J)+I(\tilde{W};Z)-I(\tilde{W};J)\Big\}\\&+I(\hat{U};Y|\hat{W}),\end{aligned}\tag{15d}$$
$$R_0+R_2\le\min\{I(W;Y),I(W;Z)\}+I(V;Z|W),\tag{15e}$$
$$\begin{aligned}R_0+R_2\le\min\Big\{&I(\hat{W};Y)+\min\big[0,I(W;Z)-I(W;Y)\big],\ I(\hat{W};J)+I(\tilde{W};Z)-I(\tilde{W};J)\Big\}\\&+I(\hat{V};J|\hat{W})+I(\tilde{V};Z|\tilde{W})-I(\tilde{V};J|\tilde{W}),\end{aligned}\tag{15f}$$
$$\begin{aligned}R_0+R_2\le\min\Big\{&I(\tilde{W};Z)+\min\big[0,I(W;Y)-I(W;Z)\big],\ I(\tilde{W};J)+I(\hat{W};Y)-I(\hat{W};J)\Big\}\\&+I(\tilde{V};Z|\tilde{W}),\end{aligned}\tag{15g}$$
$$\begin{aligned}R_0+R_1+R_2\le\min\Big\{&I(\hat{W};Y)-I(\hat{W};J),\ I(\tilde{W};Z)-I(\tilde{W};J)\Big\}+I(X;J)\\&+I(\hat{U};Y|\hat{W})-I(\hat{U};J|\hat{W})+I(\tilde{V};Z|\tilde{W})-I(\tilde{V};J|\tilde{W}),\end{aligned}\tag{15h}$$
$$\begin{aligned}R_0+R_1+R_2\le\min\Big\{&I(W;Y),I(W;Z)\Big\}\\&+\min\Big\{I(V;Z|W)+I(X;Y|VW),I(U;Y|W)+I(X;Z|UW)\Big\},\end{aligned}\tag{15i}$$

for any auxiliary channel $p_{J|XYZ}$ and some choice of distribution over the variables

$$p(u,v,w,\tilde{u},\tilde{v},\tilde{w},\hat{u},\hat{v},\hat{w},x)=p(u,v,w,x)p(\tilde{w},\tilde{u},\tilde{v}|x)p(\hat{w},\hat{u},\hat{v}|x)$$

satisfying

$$I(\hat{W};J)-I(\hat{W};Y)+I(W;Y)-I(W;Z)+I(\tilde{W};Z)-I(\tilde{W};J)=0,\tag{16a}$$
$$I(\tilde{U};J|\tilde{W})-I(\tilde{U};Z|\tilde{W})+I(\hat{U};Y|\hat{W})-I(\hat{U};J|\hat{W})=I(U;Y|W)-I(U;Z|W),\tag{16b}$$

$$I(\tilde{V};Z|\tilde{W})-I(\tilde{V};J|\tilde{W})+I(\hat{V};J|\hat{W})-I(\hat{V};Y|\hat{W})=I(V;Z|W)-I(V;Y|W), \tag{16c}$$
$$0\le I(X;Z|\tilde{U}\tilde{W})-I(X;J|\tilde{U}\tilde{W})\le I(\tilde{V};Z|\tilde{W})-I(\tilde{V};J|\tilde{W}), \tag{16d}$$
$$0\le I(X;Y|\hat{V}\hat{W})-I(X;J|\hat{V}\hat{W})\le I(\hat{U};Y|\hat{W})-I(\hat{U};J|\hat{W}), \tag{16e}$$
$$I(V;Z|W)+I(X;Y|VW)=I(U;Y|W)+I(X;Z|UW). \tag{16f}$$

Moreover, in computing the bound it suffices to assume that $|\mathcal{W}|$, $|\hat{\mathcal{W}}|$ and $|\tilde{\mathcal{W}}|$ are less than or equal to $|\mathcal{X}|+6$, while $|\mathcal{U}|,|\mathcal{V}|,|\hat{\mathcal{U}}|,|\hat{\mathcal{V}}|,|\tilde{\mathcal{U}}|,|\tilde{\mathcal{V}}|\le|\mathcal{X}|+1$.

Proof. Take an arbitrary code (n,M_0,M_1,M_2) with error probability ϵ_n. Let Q be a time-sharing random variable, uniform over $[n]$, and independent of all previously defined random variables. Make the following identification

$$\hat{W}=(M_0,J^{Q-1},Y_{Q+1}^n,Q),\tilde{W}=(M_0,Z^{Q-1},J_{Q+1}^n,Q),W=(M_0,Z^{Q-1},Y_{Q+1}^n,Q),$$
$$U=\hat{U}=\tilde{U}=M_1,V=\hat{V}=\tilde{V}=M_2.$$

Then, the constraints given in the statement of the theorem can be directly verified to hold if we allow for a negligible violation of $g(\epsilon_n)$ where $g(\cdot)$ is a function that tends to zero as ϵ_n tends to zero. The constraints (15a),(15b),(15e) are standard and are essentially same (similar to UVW bound) but for completeness we present their starting points here. The following represent the n-letter starting points for the proof of the constraints, which can be obtained using Fano's inequality. They are then single-letterized using Lemma 6, guided by the identifications mentioned above.

$$nR_0\le\min\{I(M_0;Y^n),I(M_0;Y^n),I(M_0;Z^n),I(M_0;Z^n)\}+ng(\epsilon_n) \tag{17a}$$
$$n(R_0+R_1)\le\min\{I(M_0;Y^n),I(M_0;Z^n)\}+I(M_1;Y^n|M_0)+ng(\epsilon_n), \tag{17b}$$
$$\begin{aligned}n(R_0+R_1)\le\min\Big\{&I(M_0;Z^n)+\min\big[0,I(M_0;Y^n)-I(M_0;Z^n)\big],\ I(M_0;J^n)+I(M_0;Y^n)-I(M_0;J^n)\Big\}\\&+I(M_1;J^n|M_0)+I(M_1;Y^n|M_0)-I(M_1;J^n|M_0)+ng(\epsilon_n),\end{aligned} \tag{17c}$$
$$\begin{aligned}n(R_0+R_1)\le\min\Big\{&I(M_0;Y^n)+\min\big[0,I(M_0;Z^n)-I(M_0;Y^n)\big],\ I(M_0;J^n)+I(M_0;Z^n)-I(M_0;J^n)\Big\}\\&+I(M_1;Y^n|M_0)+ng(\epsilon_n),\end{aligned} \tag{17d}$$
$$n(R_0+R_2)\le\min\{I(M_0;Y^n),I(M_0;Z^n)\}+I(M_1;Z^n|M_0)+ng(\epsilon_n), \tag{17e}$$
$$\begin{aligned}n(R_0+R_2)\le\min\Big\{&I(M_0;Y^n)+\min\big[0,I(M_0;Z^n)-I(M_0;Y^n)\big],\ I(\hat{M}_0;J^n)+I(M_0;Z^n)-I(M_0;J^n)\Big\}\\&+I(M_2;J^n|M_0)+I(M_2;Z^n|M_0)-I(M_2;J^n|M_0)+ng(\epsilon_n),\end{aligned} \tag{17f}$$
$$\begin{aligned}n(R_0+R_2)\le\min\Big\{&I(M_0;Z^n)+\min\big[0,I(M_0;Y^n)-I(M_0;Z^n)\big],\ I(M_0;J^n)+I(M_0;Y^n)-I(M_0;J^n)\Big\}\\&+I(M_2;Z^n|M_0)+ng(\epsilon_n),\end{aligned} \tag{17g}$$
$$\begin{aligned}n(R_0+R_1+R_2)\le\min\Big\{&I(M_0;Y^n)-I(M_0;J^n),I(M_0;Z^n)-I(M_0;J^n)\Big\}+I(X^n;J^n)\\&+I(M_1;Y^n|M_0)-I(M_1;J^n|M_0)+I(M_2;Z^n|M_0)-I(M_2;J^n|M_0)+ng(\epsilon_n),\end{aligned} \tag{17h}$$
$$\begin{aligned}n(R_0+R_1+R_2)\le{}&\min\{I(M_0;Y^n),I(M_0;Z^n)\}\\&+\min\Big(I(M_2;Z^n|M_0)+I(X^n;Y^n|M_2M_0),I(M_1;Y^n|M_0)+I(X^n;Z^n|M_1,M_0)\Big)+ng(\epsilon_n),\end{aligned} \tag{17i}$$

Further the constraints can be established from the following starting points again using Lemma 6, and using Fano's inequality.

$$I(M_0;J^n)-I(M_0;Y^n)+I(M_0;Y^n)-I(M_0;Z^n)+I(M_0;Z^n)-I(M_0;J^n)=0, \tag{18a}$$
$$I(M_1;J^n|M_0)-I(M_1;Z^n|M_0)+I(M_1;Y^n|M_0)-I(M_1;J^n|M_0)=I(M_1;Y^n|M_0)-I(M_1;Z^n|M_0), \tag{18b}$$

$$I(M_2;Z^n|M_0) - I(M_2;J^n|M_0) + I(M_2;J^n|M_0) - I(M_2;Y^n|M_0) = I(M_2;Z^n|M_0) - I(M_2;Y^n|M_0), \tag{18c}$$
$$0 \leq I(X^n;Z^n|M_1,M_0) - I(X^n;J^n|M_1,M_0) + ng_1(\epsilon_n) \leq I(M_2;Z^n|M_0) - I(M_2;J^n|M_0) + ng_2(\epsilon_n), \tag{18d}$$
$$0 \leq I(X^n;Y^n|M_2M_0) - I(X^n;J^n|M_2M_0) + ng_3(\epsilon_n) \leq I(M_1;Y^n|M_0) - I(M_1;J^n|M_0) + ng_4(\epsilon_n), \tag{18e}$$
$$I(M_2;Z^n|M_0) + I(X^n;Y^n|M_2M_0) = I(M_1;Y^n|M_0) + I(X^n;Z^n|M_1M_0) + ng_5(\epsilon_n) \tag{18f}$$

Note that the bound depends only on the marginal distributions of $(\tilde{W},\tilde{U},X)$, $(\tilde{W},\tilde{V},X)$, $(\hat{W},\hat{U},X)$, $(\hat{W},\hat{V},X)$, (W,U,X) and (W,V,X). Therefore a consistent distributions of X is all that is needed to ensure the existence of a joint distribution. Then, similar to the original UV bound and using standard techniques, cardinality bounds on all of the auxiliary random variables can be imposed. Both of these are the primary reasons why we identified M_1 separately as $U,\tilde{V},\hat{U}$, and similarly for M_2. Therefore, for each $\epsilon_n > 0$, one can find a joint distributions $p_\epsilon(\tilde{u},\tilde{v},\tilde{w},\hat{u},\hat{v},\hat{w},x)$ with bounded alphabet sizes such that the constraints given in the statement of the theorem are violated by at most $g(\epsilon_n)$. Since the space of joint distributions with bounded alphabet sizes forms a compact set, by letting ϵ_n converge to zero, we can find a limit distribution $p(\tilde{u},\tilde{v},\tilde{w},\hat{u},\hat{v},\hat{w},x)$ for which all of the constraints in the theorem hold.

Finally, the cardinality bounds come from the standard Caratheodory-Bunt [Bun34] arguments and are omitted. □

Remark 6. From (15a),(15b), (15e), (15i), we can extract the following constraints:

$$\begin{aligned}
R_0 &\leq \min\{I(W;Y),I(W;Z)\}\\
R_0+R_1 &\leq \min\{I(W;Y),I(W;Z)\} + I(U;Y|W)\\
R_0+R_2 &\leq \min\{I(W;Y),I(W;Z)\} + I(V;Z|W)\\
R_0+R_1+R_2 &\leq \min\{I(W;Y),I(W;Z) + \min\{I(U;Y|W)+I(X;Z|UW), I(V;Z|W)+I(X;Y|VW)\}
\end{aligned}$$

and this implies that the outer bound in Theorem 7 is at least as good as the UV outer bound for all broadcast channels $T(y,z|x)$.

For generic broadcast channels the points $(C_1,0)$ and $(0,C_2)$ are the "corner" points of the capacity region and in [NKG16] the authors computed the slope of the capacity region at these points. However for some broadcast channels (that has zero Lebesgue measure in the space of parameters defining the broadcast channel given the input and output alphabet sizes), rate pairs of the form (C_1,R_2) for some $R_2>0$ and (R_1,C_2) for some $R_1>0$ are achievable. For such channels, we define (C_1,R_2^*) to be a *corner point* if no point of the form $(C_1,R_2^*+\epsilon)$ for any $\epsilon>0$ belongs to the capacity region. Similarly one can define an analogous corner point of the form (R_1^*,C_2). The results in [NKG16] are insufficient to determine these corner points.

The following corollary, which will be used later to show that Theorem 7 improves on the UV outer bound, relates to the study of corner points of the capacity region.

Corollary 3. *Consider a general broadcast channel $T(y,z|x)$ where $T(y|x)=\sum_{\hat{y}} T(y|\hat{y})\hat{T}(\hat{y}|x)$ for some $\hat{Y}$ which is an enhancement of Y. Furthermore, assume that*

- *$I(X;Y|U)=0$ implies $I(X;\hat{Y}|U)=0$.*

Then, the rate triple $(0,C_1,R_2)$ is achievable only if

$$R_2 \leq I(V;Z|W) - I(V;\hat{Y}|W) \tag{19}$$

for some $p(v,w,x)$ satisfying

$$C_1 = I(X;Y) \leq I(W;Z) + I(X;\hat{Y}|W). \tag{20}$$

Proof. Set $J=\hat{Y}$ in Theorem 7. We have

$$C_1 = R_0+R_1 \leq I(\hat{U},\hat{W};Y) = I(X;Y) - I(X;Y|\hat{U},\hat{W}) \leq I(X;Y) \leq C_1$$

implying that $I(X;Y)=C_1$ and $I(X;Y|\hat{U},\hat{W})=0$ and hence (from our assumption) that $I(X;\hat{Y}|\hat{U},\hat{W})=0$. From (15c), since $I(\hat{U};\hat{Y}|\hat{W}) \geq I(\hat{U};Y|\hat{W})$, we see that

$$C_1 = R_0+R_1 \leq I(\tilde{W};Z) + I(\tilde{U};\hat{Y}|\tilde{W}) \leq I(\tilde{W};Z) + I(X;\hat{Y}|\tilde{W}).$$

From (15h) we see that

$$C_1 + R_2 = R_0 + R_1 + R_2 \leq I(\hat{W}, \hat{U}; Y) + I(X; J|\hat{W}, \hat{U}) + I(\tilde{V}; Z|\tilde{W}) - I(\tilde{V}; J|\tilde{W}).$$

Since $I(X; J|\hat{W}, \hat{U}) = 0$ and $C_1 \geq I(X; Y) \geq I(\hat{W}, \hat{U}; Y)$, we have that $R_2 \leq I(\tilde{V}; Z|\tilde{W}) - I(\tilde{V}; J|\tilde{W})$. □

In the next section, we will demonstrate that Corollary 3 (and hence the outer bound of Theorem 7) outperforms the UV outer bound for a particular broadcast channel.

4.2. **Erasure Blackwell Channel.** The standard Blackwell channel is a deterministic broadcast channel $T(\hat{y}, \hat{z}|x)$ where $\mathcal{X} = \{0, 1, 2\}, \hat{\mathcal{Y}} = \{0, 1\}, \hat{\mathcal{Z}} = \{0, 1\}$, $\hat{Y} = \mathbf{1}[X = 2]$ and $\hat{Z} = \mathbf{1}[X = 1]$. The Erasure Blackwell channel is obtained when each of the outputs of the Blackwell broadcast channel are erased with probability ϵ. More specifically, we assume that $T(y, z|x) = \sum_{\hat{y},\hat{z}} T(\hat{y}, \hat{z}|x)p(y|\hat{y})p(z|\hat{z})$ where $p(y|\hat{y})$ and $p(z|\hat{z})$ are erasure channels with erasure probability ϵ. If $\epsilon = 0$, we get the Blackwell channel whose capacity is the union over all $p(x)$ of

$$\begin{aligned} R_1 &\leq H(\hat{Y}) \\ R_2 &\leq H(\hat{Z}) \\ R_1 + R_2 &\leq H(\hat{Y}, \hat{Z}). \end{aligned}$$

The UV-outer bound scales by $1 - \epsilon$ for the erased Blackwell channel. Thus the UV-outer bound reduces to the following for erased Blackwell:

$$\begin{aligned} R_1 &\leq (1 - \epsilon)H(\hat{Y}) \\ R_2 &\leq (1 - \epsilon)H(\hat{Z}) \\ R_1 + R_2 &\leq (1 - \epsilon)H(\hat{Y}, \hat{Z}). \end{aligned}$$

In particular, the corner point of the UV outer bound is $(R_1, R_2) = \big(1 - \epsilon, \frac{1}{2}(1 - \epsilon)\big)$. The outer bound developed for the corner point in Corollary 3 is used in Lemma 4 to show that the rate pair $(R_1, R_2) = \big(1 - \epsilon, \frac{1}{2}(1 - \epsilon)\big)$ is not achievable for any $\epsilon \in (0, 1)$. Therefore, the capacity region does not scale by $1 - \epsilon$.

Lemma 4. *The rate pair $(R_1, R_2) = \big(1 - \epsilon, \frac{1}{2}(1 - \epsilon)\big)$ is not achievable for any $\epsilon \in (0, 1)$ for the erasure Blackwell channel.*

Proof of Lemma 4 is given in Appendix D.1.

Remark 7. Even though Corollary 3 implies that the outer bound in Theorem 7 is strictly better than the UV outer bound for the erasure Blackwell channel, numerical results indicate that there is still a gap between the upper bound for the corner point in Corollary 3 and Marton's inner bound for the erasure Blackwell channel. For example, for $\epsilon = 0.1$, the UV outer bound has a corner point $(0.9, 0.45)$, while numerical simulations show that the new outer bound has a corner point $(0.9, 0.4265)$ and Marton's inner bound has a corner point $(0.9, 0.4205)$. Determining the true corner point for the erasure Blackwell channel remains an open problem.

The authors' original motivation for using auxiliary receivers for deriving converses came from the study of multi-letter extensions of Marton's Inner bound for the erasure Blackwell channel. We summarize some facts about the multi-letter Marton's inner bound (whose limit is the capacity region) for the erasure Blackwell channel. Our main result here is that we can identify one of the optimal auxiliaries for computing the weighted sum-rates of k-letter extensions of Marton's bound for all sufficiently large weights (independent of k).

Remark 8. The issue of determining the optimality (or sub-optimality) of Marton's inner bound stems from the inability to compute multi-letter extensions due to the dimensionality of the optimization problems and inability to identify the extremal auxiliaries. Thus the result here reduces the dimension as we determine the optimal U, leaving only V, W to be determined.

More generally, we consider a channel $p(y, z|x)$ such that

$$p(y, z|x) = \sum_{\hat{y}} p(y|\hat{y})p(z|x)p(\hat{y}|x)$$

where $\hat{Y} = f(X)$ is a function of X. For $\alpha \geq 1$, we can express the α-sum rate of Marton's inner bound as follows (this follows from a minimax theorem in [GGNY11]):

$$\max_{(R_1,R_2)\in\mathcal{R}_{\text{Marton}}} \alpha R_1+R_2 = \min_{\lambda\in[0,1]} \max_{p(u,v,w,x)} \Big\{(\alpha-\lambda)I(W;Y)+\lambda I(W;Z)+\alpha I(U;Y|W)+I(V;Z|W)-I(U;V|W)\Big\}.$$

Similarly, for the k-letter Marton we have

$$\max_{(R_1,R_2)\in\mathcal{R}_{\text{Marton}}} \alpha R_1 + R_2 =$$
$$\frac{1}{k}\min_{\lambda\in[0,1]} \max_{p(u,v,w,x^k)} \Big\{(\alpha-\lambda)I(W;Y^k)+\lambda I(W;Z^k)+\alpha I(U;Y^k|W)+I(V;Z^k|W)-I(U;V|W)\Big\}.$$

Proposition 3. *The following two statements hold:*

- *To evaluate $\alpha R_1 + R_2$ for k-letter Marton, it is optimal to set $U = \hat{Y}^k$ if $\alpha \geq \alpha^*(p_{\hat{Y}|X}p_{Y|\hat{Y}})$ where*
$$\alpha^*(p_{\hat{Y}|X}p_{Y|\hat{Y}}) = \sup_{p(x):I(\hat{Y};Y)\neq 0} \frac{I(X;\hat{Y})}{I(\hat{Y};Y)} = \sup_{p(x):I(\hat{Y};Y)\neq 0} \frac{H(\hat{Y})}{I(\hat{Y};Y)}. \tag{21}$$
In particular, if the channel from $\hat{Y}$ to Y is erasure with probability ϵ, we have $\alpha^* = \frac{1}{1-\epsilon}$. Consequently, for $k = 1$, α-sum rate of Marton's inner bound reduces to
$$\min_{\lambda\in[0,1]} \max_{p(v,w,x)} \Big\{(\alpha-\lambda)I(W;Y)+\lambda I(W;Z)+\alpha I(\hat{Y};Y|W)+I(V;Z|W)-I(\hat{Y};V|W)\Big\},$$
for $\alpha \geq \alpha^*(p_{\hat{Y}|X}p_{Y|\hat{Y}})$.
- *Take some $\alpha \geq \alpha^*(p_{\hat{Y}|X}p_{Y|\hat{Y}})$. Then, α-sum rate of Marton's inner bound equals*
$$\max_{p(v,x)} \Big\{\alpha I(\hat{Y};Y)+I(V;Z)-I(\hat{Y};V)\Big\},$$
if there exists some $\lambda^* \in [0,1]$ for which the function
$$p(x) \mapsto \max_{p(v|x)} \Big\{-(\alpha-\lambda^*)H(Y)-\lambda^* H(Z)+\alpha I(\hat{Y};Y)+I(V;Z)-I(V;\hat{Y})\Big\}. \tag{22}$$
is concave. In other words, concavity of the function in (22) is a sufficient condition for optimality of setting $W = \emptyset$ when computing Marton's inner bound.

Proof. We begin by proving the first part of the proposition. Consider the case of 1-letter Marton, *i.e.,* $k = 1$. We need to prove that

$$\alpha I(U;Y|W) - I(U;V|W) \leq \alpha I(\hat{Y};Y|W) - I(\hat{Y};V|W)$$

Equivalently, we should prove that

$$I(\hat{Y};V|W) - I(U;V|W) \leq \alpha I(\hat{Y};Y|W) - \alpha I(U;Y|W)$$

We have

$$I(\hat{Y};V|W) - I(U;V|W) \leq I(\hat{Y};V|UW) \leq H(\hat{Y}|UW) \leq \alpha I(\hat{Y};Y|UW) = \alpha I(\hat{Y};Y|W) - \alpha I(U;Y|W)$$

where we used the fact that $\alpha \geq \alpha^*$ to conclude that for any u, w we have

$$H(\hat{Y}|U=u,W=w) \leq \alpha I(\hat{Y};Y|U=u,W=w).$$

The result for the k-letter Marton follows from the tensorization property of α^* given in Lemma 5.

To show the second part of the proposition, first observe that

$$\min_{\lambda\in[0,1]} \max_{p(v,w,x)} \Big\{(\alpha-\lambda)I(W;Y)+\lambda I(W;Z)+\alpha I(\hat{Y};Y|W)+I(V;Z|W)-I(\hat{Y};V|W)\Big\} \tag{23}$$
$$\geq \max_{p(v,x)} \Big\{\alpha I(\hat{Y};Y)+I(V;Z)-I(\hat{Y};V)\Big\}, \tag{24}$$

as we can always choose $W = \emptyset$ in the inner maximization problem. On the other hand, assume that the concavity property holds for some λ^*. We have

$$\begin{aligned}
&\min_{\lambda\in[0,1]} \max_{p(v,w,x)} \Big\{(\alpha-\lambda)I(W;Y)+\lambda I(W;Z)+\alpha I(\hat{Y};Y|W)+I(V;Z|W)-I(\hat{Y};V|W)\Big\}\\
&\le \max_{p(v,w,x)} \Big\{(\alpha-\lambda^*)I(W;Y)+\lambda^* I(W;Z)+\alpha I(\hat{Y};Y|W)+I(V;Z|W)-I(\hat{Y};V|W)\Big\}\\
&= \max_{p(x)}(\alpha-\lambda^*)H(Y)+\lambda^* H(Z)+ \qquad (25)\\
&\quad \max_{p(w|x)} \Big\{-(\alpha-\lambda^*)H(Y|W)+\lambda^* H(Z|W)+\alpha I(\hat{Y};Y|W)+\max_{p(v|w,x)}\Big\{I(V;Z|W)-I(\hat{Y};V|W)\Big\}\Big\}.
\end{aligned}$$

The concavity property implies optimality of $W = \emptyset$ in (25).

□

Remark 9. Consider the erasure Blackwell channel with erasure probability ϵ. Let $\alpha = 1/(1-\epsilon)$. Simulations indicate that for small erasure probabilities $\epsilon \le 0.6$, the concavity of the function given in (22) holds if we choose $\lambda^* = 0.5$. On the other hand, if ϵ is larger, say larger than 0.63, then not only the function given in (22) is no longer concave, but simulations results also indicate that setting $W = \emptyset$ is indeed not optimal when computing Marton's inner bound.

Lemma 5. *For given channels $p_{J|X}$ and $p_{Y|J}$, define*

$$\alpha^*(p_{J|X}, p_{Y|J}) = \sup_{p(x):I(J;Y)\neq 0} \frac{I(X;J)}{I(J;Y)}. \qquad (26)$$

Take some natural number k and let $p_{J^k|X^k} = \prod_{i=1}^k p_{J_i|X_i}$ and $p_{Y^k|J^k} = \prod_{i=1}^k p_{Y_i|J_i}$ be memoryless channel extensions. Then

$$\alpha^*(p_{J^k|X^k}, p_{Y^k|J^k}) = \alpha^*(p_{J|X}, p_{Y|J}).$$

Proof. The direction

$$\alpha^*(p_{J^k|X^k}, p_{Y^k|J^k}) \ge \alpha^*(p_{J|X}, p_{Y|J})$$

follows by taking a product input distribution on X^k. For the other direction, take some arbitrary $p(x^k)$. Then,

$$\frac{I(X^k;J^k)}{I(J^k;Y^k)} = \frac{\sum_{i=1}^k I(J_i;X^k|J^{i-1})}{\sum_{i=1}^k I(J_i;Y^k|J^{i-1})} = \frac{\sum_{i=1}^k I(J_i;X_i|J^{i-1})}{\sum_{i=1}^k I(J_i;Y^k|J^{i-1})} \le \frac{\sum_{i=1}^k I(J_i;X_i|J^{i-1})}{\sum_{i=1}^k I(J_i;Y_i|J^{i-1})} \le \alpha^*_{p_{J|X}p_{Y|J}}$$

since for any i and every j^{i-1} where $p(J^{i-1} = j^{i-1}) > 0$ we have

$$\frac{I(J_i;X_i|J^{i-1}=j^{i-1})}{I(J_i;Y_i|J^{i-1}=j^{i-1})} \le \alpha^*(p_{J|X}, p_{Y|J}).$$

□

4.3. **The Second Outer Bound.** This outer bound was motivated directly from the outer bound for product broadcast channels developed in [GGNY14], that was used to demonstrate that the UV outer bound can be strictly improved. The outer bound in [GGNY14] critically used the product nature of the channel and if one perturbs the channel so as to lose the product nature, the outer bound became invalid. We now present an outer bound that varies continuously with respect to channel perturbations.

The outer bound of Theorem 7 uses a *single* auxiliary receiver J. In this section, we write another version of the UV bound with *two* auxiliary variables J and $\hat{J}$. This bound may be also interpreted as a Genie-aided bound with auxiliary receiver J provided to receiver Y, and auxiliary receiver $\hat{J}$ provided to receiver Z.

Theorem 8. *Given a broadcast channel $T(y,z|x)$ and any $T_{J,\hat{J}|X,Y,Z}$ any achievable non-negative rate triple (R_0,R_1,R_2) must satisfy the following constraints*

$$R_0 \le \min\{I(W_b;J)+I(W_a;Y|J), I(W_b;Z|\hat{J})+I(W_a;\hat{J})\} \qquad (27a)$$

$$R_0+R_1 \le I(U_b,W_b;J)+I(U_a,W_a;Y|J), \qquad (27b)$$

$$R_0 + R_1 \leq I(W_b; Z|\hat{J}) + I(W_a, J; \hat{J}) + I(U_b; J|W_b, \hat{J}) + I(U_a; Y|W_a, J), \tag{27c}$$

$$R_0 + R_2 \leq I(W_b, \hat{J}; J) + I(W_a; Y|J) + I(V_b; Z|W_b, \hat{J}) + I(V_a; \hat{J}|W_a, J), \tag{27d}$$

$$R_0 + R_2 \leq I(V_a, W_a; \hat{J}) + I(V_b, W_b; Z|\hat{J}) \tag{27e}$$

$$\begin{aligned} R_0 + R_1 + R_2 \leq \min\{&I(W_b, \hat{J}; J) + I(W_a; Y|J),\ I(W_b; Z|\hat{J}) + I(W_a, J; \hat{J})\} \\ &+ I(U_a; Y|W_a, J) + I(X; \hat{J}|U_a, W_a, J) \\ &+ \min\left\{I(U_b; J|W_b, \hat{J}) + I(X; Z|U_b, W_b, \hat{J}),\ I(V_b; Z|W_b, \hat{J}) + I(X; J|V_b, W_b, \hat{J})\right\}, \end{aligned} \tag{27f}$$

$$\begin{aligned} R_0 + R_1 + R_2 \leq \min\{&I(W_b, \hat{J}; J) + I(W_a; Y|J),\ I(W_b; Z|\hat{J}) + I(W_a, J; \hat{J})\} \\ &+ I(V_b; Z|W_b, \hat{J}) + I(X; J|V_b, W_b, \hat{J}) \\ &+ \min\left\{I(U_a; Y|W_a, J) + I(X; \hat{J}|U_a, W_a, J),\ I(V_a; \hat{J}|W_a, J) + I(X; Y|V_a, W_a, J)\right\}, \end{aligned} \tag{27g}$$

for some $p(w_a, v_a, u_a|x)p(w_b, v_b, u_b|x)p(x)$ *satisfying* $|\mathcal{W}_b|, |\mathcal{W}_a| \leq |\mathcal{X}|+7$, $|\mathcal{U}_b|, |\mathcal{V}_a| \leq |\mathcal{X}|+2$, $|\mathcal{V}_b|, |\mathcal{U}_a| \leq |\mathcal{X}|+1$.

Proof. Take a code of length n with message triple (M_0, M_1, M_2) of rates (R_0, R_1, R_2) and with error probability of ϵ. Let Q be a random variable independent of the code book such that Q is uniform in $[n]$. Define

$$\begin{aligned} &W_{ai} = (M_0, Y^{i-1}, \hat{J}_{i+1}^n, J^{n\setminus i}), \quad W_{bi} = (M_0, J^{i-1}, Z_{i+1}^n, \hat{J}^{n\setminus i}) \\ &U_a = U_b = M_1, \qquad V_a = V_b = M_2. \end{aligned} \tag{28}$$

The outer bound follows from routine manipulations using Lemma 6, guided by the above identification, starting from each of the following n-letter expressions which are reasonably straightforward (please see E.2) to obtain using Fano's inequality:

$$nR_0 \leq \min\{I(M_0; J^n) + I(M_0; Y^n|J^n), I(M_0; Z^n|\hat{J}^n) + I(M_0; \hat{J}^n)\} + ng(\epsilon_n), \tag{29a}$$

$$n(R_0 + R_1) \leq I(M_1, M_0; J^n) + I(M_1, M_0; Y^n|J^n) + ng(\epsilon_n), \tag{29b}$$

$$n(R_0 + R_1) \leq I(M_0; Z^n|\hat{J}^n) + I(M_0, J^n; \hat{J}^n) + I(M_1; J^n|M_0, \hat{J}^n) + I(M_1; Y^n|M_0, J^n) + ng(\epsilon_n), \tag{29c}$$

$$n(R_0 + R_2) \leq I(M_0, \hat{J}^n; J^n) + I(M_0; Y^n|J^n) + I(M_2; Z^n|M_0, \hat{J}^n) + I(M_2; \hat{J}^n|M_0, J^n) + ng(\epsilon_n), \tag{29d}$$

$$n(R_0 + R_2) \leq I(M_2, M_0; \hat{J}^n) + I(M_2, M_0; Z^n|\hat{J}^n) + ng(\epsilon_n), \tag{29e}$$

$$\begin{aligned} n(R_0 + R_1 + R_2) \leq \min\{&I(M_0, \hat{J}^n; J^n) + I(M_0; Y^n|J^n), I(M_0; Z^n|\hat{J}^n) + I(M_0, J^n; \hat{J}^n)\} \\ &+ I(M_1; Y^n|M_0, J^n) + I(X^n; \hat{J}^n|M_1, M_0, J^n) \\ &+ \min\Bigg\{I(M_1; J^n|M_0, \hat{J}^n) + I(X^n; Z^n|M_1, M_0, \hat{J}^n), \\ &\qquad I(M_2; Z^n|M_0, \hat{J}^n) + I(X^n; J^n|M_2, M_0, \hat{J}^n)\Bigg\} + ng(\epsilon_n), \end{aligned} \tag{29f}$$

$$\begin{aligned} n(R_0 + R_1 + R_2) \leq \min\{&I(M_0, \hat{J}^n; J^n) + I(M_0; Y^n|J^n), I(M_0; Z^n|\hat{J}^n) + I(M_0, J^n; \hat{J}^n)\} \\ &+ I(M_2; Z^n|M_0, \hat{J}^n) + I(X^n; J^n|M_2, M_0, \hat{J}^n) \\ &+ \min\Bigg\{I(M_1; Y^n|M_0, J^n) + I(X^n; \hat{J}^n|M_1, M_0, J^n), \\ &\qquad I(M_2; \hat{J}^n|M_0, J^n) + I(X^n; Y^n|M_2, M_0, J^n)\Bigg\} + ng(\epsilon_n). \end{aligned} \tag{29g}$$

Finally, the cardinality bounds come from the standard Caratheodory-Bunt [Bun34] arguments and are omitted. ☐

Remark 10. An alternative approach to single-letterize (29a)-(29g) that skips using Lemma 6 is as follows: consider the UV bound in Theorem 6. Take for instance, the sum-rate constraint:

$$R_0 + R_1 + R_2 \leq I(W; Y) + I(U; Y|W) + I(X; Z|U, W).$$

This inequality is shown via the following expansion

$$I(M_0;Y^n)+I(M_1;Y^n|M_0)+I(M_2;Z^n|M_0,M_1)\leq\sum_i\left(I(W_i;Y_i)+I(U_i;Y_i|W_i)+I(X_i;Z_i|U_i,W_i)\right), \tag{30}$$

where $W_i=(M_0,Y^{i-1},Z_{i+1}^n),U_i=M_1,V_i=M_2$. The inequality (30) holds for any *arbitrary* joint distribution of p_{M_0,M_1,M_2,Y^n,Z^n}. Thus, it continues to hold if we formally replace M_0 and Z^n by $\tilde{M}_0=(M_0,J^n)$ and $\hat{J}^n$ respectively, while keeping all the other variables intact. With this replacement, the auxiliary variable W_i becomes $(M_0,J^n,Y^{i-1},\hat{J}_{i+1}^n)$ which is equal to (W_{ai},J_i) as defined in (28). This yields

$$\begin{aligned}&I(M_0,J^n;Y^n)+I(M_1;Y^n|M_0,J^n)+I(M_2;\hat{J}^n|M_0,J^n,M_1)\\&\leq\sum_i\Big(I(W_{ai},J_i;Y_i)+I(U_{ai};Y_i|W_{ai},J_i)+I(X_i;\hat{J}_i|U_{ai},W_{ai},J_i)\Big).\end{aligned}$$

Next, observe that the inequality (30) also continues to holds if we condition all the mutual information terms on J^n. This implies that

$$\begin{aligned}&I(M_0;Y^n|J^n)+I(M_1;Y^n|M_0,J^n)+I(M_2;\hat{J}^n|M_0,J^n,M_1)\\&\leq\sum_i\Big(I(W_{ai};Y_i|J_i)+I(U_{ai};Y_i|W_{ai},J_i)+I(X_i;\hat{J}_i|U_{ai},W_{ai},J_i)\Big).\end{aligned}$$

Similarly, one can obtain two sets of inequalities by replacing M_0 and Y^n by $(M_0,\hat{J}^n)$ and J^n respectively, or alternatively by conditioning all the terms on J^n. One can single-letterize (29a)-(29g) by writing the above four sets of inequalities for all the constraints in the UV bound, and mixing and matching appropriate equations from these four sets of inequalities.

As a special case of Theorem 4 assume that $H(J|Y)=H(\hat{J}|Z)=0$. More specifically, for a pair of bijective mappings $Y\leftrightarrow(Y_1,Y_2)$ and $Z\leftrightarrow(Z_1,Z_2)$, set $J=Y_1$ and $\hat{J}=Z_2$. Then, we obtain the following corollary:

Corollary 4. *Given a broadcast channel $T(y,z|x)$ any achievable non-negative rate triple (R_0,R_1,R_2) must satisfy the following constraints*

$$\begin{aligned}R_0&\leq\min\{I(W_b;Y_1)+I(W_a;Y_2|Y_1),I(W_b;Z_1|Z_2)+I(W_a;Z_2)\}\\R_0+R_1&\leq I(U_b,W_b;Y_1)+I(U_a,W_a;Y_2|Y_1),\\R_0+R_1&\leq I(W_b;Z_1|Z_2)+I(W_a,Y_1;Z_2)+I(U_b;Y_1|W_b,Z_2)+I(U_a;Y_2|W_a,Y_1),\\R_0+R_2&\leq I(W_b,Z_2;Y_1)+I(W_a;Y_2|Y_1)+I(V_b;Z_1|W_b,Z_2)+I(V_a;Z_2|W_a,Y_1),\\R_0+R_2&\leq I(V_a,W_a;Z_2)+I(V_b,W_b;Z_1|Z_2)\\R_0+R_1+R_2&\leq\min\{I(W_b,Z_2;Y_1)+I(W_a;Y_2|Y_1),I(W_b;Z_1|Z_2)+I(W_a,Y_1;Z_2)\}\\&\quad+I(U_a;Y_2|W_a,Y_1)+I(X;Z_2|U_a,W_a,Y_1)\\&\quad+\min\big\{I(U_b;Y_1|W_b,Z_2)+I(X;Z_1|U_b,W_b,Z_2),I(V_b;Z_1|W_b,Z_2)+I(X;Y_1|V_b,W_b,Z_2)\big\},\\R_0+R_1+R_2&\leq\min\{I(W_b,Z_2;Y_1)+I(W_a;Y_2|Y_1),I(W_b;Z_1|Z_2)+I(W_a,Y_1;Z_2)\}\\&\quad+I(V_b;Z_1|W_b,Z_2)+I(X;Y_1|V_b,W_b,Z_2)\\&\quad+\min\big\{I(U_a;Y_2|W_a,Y_1)+I(X;Z_2|U_a,W_a,Y_1),I(V_a;Z_2|W_a,Y_1)+I(X;Y_2|V_a,W_a,Y_1)\big\},\end{aligned}$$

for any pair of bijective mappings $Y\leftrightarrow(Y_1,Y_2)$ and $Z\leftrightarrow(Z_1,Z_2)$ and for some $p(w_a,v_a,u_a|x)p(w_b,v_b,u_b|x)p(x)$.

Remark 11. The following remarks are worth noting.

(1) This outer bound generalizes the outer bound of [GGNY14] to non-product broadcast channels. Consider the special case of $X=(X_1,X_2)$ and $T(y_1y_2,z_1z_2|x)$ being of the form

$$T(y_1y_2,z_1z_2|x)=T(y_1,z_1|x_1)T(y_2,z_2|x_2).$$

Then, the above outer bound reduces to the one given in [GGNY14]. Since the outer bound in [GGNY14] has been shown to strictly improve on the UV outer bound for some product broadcast channels, our new outer bound is also a strict improvement on the UV outer bound.

(2) Setting $Y_1 = Y,\ Z_1 = Z$, and $Y_2 = Z_2 = 0$ (constant random variables) reduces the above outer bound to the UV outer bond in [Nai11]. Hence this bound is at least as good as the UV outer bound for any broadcast channel. Finally, since this is strictly better than the UV for some product broadcast channels by virtue of the previous remark, this bound is a *strict* improvement over the UV outer bound.

(3) An interesting feature of the above outer bound is expressions like $I(W_1, Z_2; Y_1)$ where W_1 comes with Z_2 on one side, and Y_1 on the other side of the mutual information expression. This differs from the UV outer bound (or Marton's inner bound) where channel output variables and the auxiliary random variables appear on the opposite sides of the mutual information expressions.

5. Conclusion and Future work

New outer bounds for relay, interference, and broadcast channels have been developed using the idea of auxiliary receivers. The bounds were then employed to demonstrate aspects of the capacity region that were not determined from previous outer bounds such as: kinks (discontinuous derivatives) at the capacity region around corner points for the relay and the interference channel, and that capacity regions can shrink by more that $1 - \epsilon$ if the received symbols were erased with probability ϵ (a phenomenon that does not happen in the presence of feedback if the erasures are synchronous). We aimed to give an illustration of the techniques that one could use to develop outer bounds using auxiliary receivers, and we are positive that we have not harnessed the full potential of the auxiliary receivers even in the basic settings considered here.

In particular, a number of immediate future research directions is listed here:

(*i*) We note that there are many different ways to introduce auxiliary receivers. For instance, we give two outer bounds for broadcast channels. The examples for which these two bounds strictly improve over the UV outer bound are different. Unification of these two outer bounds into a single bound is left as a future work.

(*ii*) Any choice of auxiliary receivers in the results of this paper yields a valid and computable upper bound to the capacity region in discrete settings. A natural question is the to determine the smallest possible outer bound using these techniques. An immediate question in this direction: can one determine cardinality bounds on the sizes of auxiliary receivers so as to obtain the best upper bound.

(*iii*) The bound for the Gaussian relay bound channel was obtained by choosing the auxiliary channel from X to J to be an additive Gaussian channel. However, any arbitrary choice of the channel from X to J yields a valid upper bound. As a future work, one can study if one gets the best possible upper bound by taking the channel from X to J to be additive Gaussian.

Acknowledgement

The authors wish to thank Young-Han Kim for raising the question of the whether the capacity of the ϵ-erasure Blackwell broadcast channel scales as $1 - \epsilon$ and for some initial discussions towards determining the validity of the statement. Subsequent investigations by the authors demonstrated that the capacity region does not scale by $(1 - \epsilon)$ and this result motivated the rest of the findings of this paper.

The authors would like to thank Mohamamd Hossein Yassaee for alerting them to the fact that the upper bound for the relay channel improves over the cut-set bound for the scalar Gaussian relay channel. Prior to that, the authors had a more complicated example to show that the upper bound strictly improves over the cut-set bound. The authors would also like to thank Yannan Wang for doing the initial simulations concerning the outer bound and Marton's inner bound for the erased Blackwell channel.

References

[ARY09] M. Aleksic, P. Razaghi, and W. Yu, *Capacity of a class of modulo-sum relay channels*, IEEE Transactions on Information Theory **55** (2009), no. 3, 921–930.

[Bun34] L.N.H. Bunt, *Bijdrage tot de theorie der convexe puntverzamelingen*, Ph.D. thesis, Univ. Groningne, Amsterdam, 1934.

[CG79] T. Cover and A. E. Gamal, *Capacity theorems for the relay channel*, IEEE Transactions on Information Theory **25** (1979), no. 5, 572–584.

[CK78] I Csizár and J Körner, *Broadcast channels with confidential messages*, IEEE Trans. Info. Theory **IT-24** (May, 1978), 339–348.

[CK11] Imre Csiszar and Janos Korner, *Information theory: Coding theorems for discrete memoryless systems*, Cambridge University Press, 1 2011.

[CN16] M H M Costa and C Nair, *Gaussian z-interference channel: around the corner*, Information Theory and Applications Workshop (ITA) (2016).

[Cos85] M. H M Costa, *On the Gaussian interference channel*, Information Theory, IEEE Transactions on **31** (1985), no. 5, 607–615.

[Cov72] T. Cover, *Broadcast channels*, IEEE Transactions on Information Theory **18** (1972), no. 1, 2–14.

[DMN18] Anindya De, Elchanan Mossel, and Joe Neeman, *Non interactive simulation of correlated distributions is decidable*, Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2018, pp. 2728–2746.

[EK12] Abbas El Gamal and Young-Han Kim, *Network Information Theory*, Cambridge University Press, 2012.

[El 79] A El Gamal, *The capacity of a class of broadcast channels*, IEEE Trans. Info. Theory **IT-25** (March, 1979), 166–169.

[EO11] R. H. Etkin and E. Ordentlich, *Analysis of deterministic binary interference channels via a general outer bound*, IEEE Transactions on Information Theory **57** (2011), no. 5, 2597–2604.

[GA10] A. Gohari and V. Anantharam, *Information-theoretic key agreement of multiple terminals—part i*, IEEE Transactions on Information Theory **56** (2010), no. 8, 3973–3996.

[Gal74] R G Gallager, *Capacity and coding for degraded broadcast channels*, Probl. Peredac. Inform. **10(3)** (1974), 3–14.

[GGNY11] Yanlin Geng, Amin Gohari, Chandra Nair, and Yuanming Yu, *The capacity region of classes of product broadcast channels*, 2011.

[GGNY14] Y. Geng, A. Gohari, C. Nair, and Y. Yu, *On Marton's inner bound and its optimality for classes of product broadcast channels*, IEEE Transactions on Information Theory **60** (2014), no. 1, 22–41.

[GKS16] Badih Ghazi, Pritish Kamath, and Madhu Sudan, *Decidability of non-interactive simulation of joint distributions*, 2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2016, pp. 545–554.

[GN14] Y. Geng and C. Nair, *The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages*, IEEE Transactions on Information Theory **60** (2014), no. 4, 2087–2104.

[HK81] Te Sun Han and Kingo Kobayashi, *A new achievable rate region for the interference channel*, Information Theory, IEEE Transactions on **27** (1981), no. 1, 49–60.

[KM75] J Körner and K Marton, *Comparison of two noisy channels*, Topics in Inform. Theory(ed. by I. Csiszar and P.Elias), Keszthely, Hungary (August, 1975), 411–423.

[KM77] ———, *Images of a set via two channels and their role in mult-user communication*, IEEE Trans. Info. Theory **IT-23** (Nov, 1977), 751–761.

[Kra04] G. Kramer, *Outer bounds on the capacity of gaussian interference channels*, Information Theory, IEEE Transactions on **50** (2004), no. 3, 581–586.

[KU10] Wei Kang and Sennur Ulukus, *A new data processing inequality and its applications in distributed source and channel coding*, IEEE Transactions on Information Theory **57** (2010), no. 1, 56–69.

[LO18] Jingbo Liu and Ayfer Ozgur, *Capacity upper bounds for the relay channel via reverse hypercontractivity*, arXiv preprint arXiv:1811.11303 (2018).

[LO19] J. Liu and A. Ozgur, *New converses for the relay channel via reverse hypercontractivity*, 2019 IEEE International Symposium on Information Theory (ISIT), July 2019, pp. 2878–2882.

[LV07] Tie Liu and P. Viswanath, *An extremal inequality motivated by multiterminal information-theoretic problems*, Information Theory, IEEE Transactions on **53** (2007), no. 5, 1839–1851.

[Mar79] K. Marton, *A coding theorem for the discrete memoryless broadcast channel*, IEEE Transactions on Information Theory **25** (1979), no. 3, 306–311.

[MHU19] Marco Mondelli, S Hamed Hassani, and Rüdiger Urbanke, *A new coding paradigm for the primitive relay channel*, Algorithms **12** (2019), no. 10, 218.

[Nai11] Chandra Nair, *A note on outer bounds for broadcast channel*, CoRR **abs/1101.0640** (2011).

[NE07] C Nair and A El Gamal, *An outer bound to the capacity region of the broadcast channel*, IEEE Trans. Info. Theory **IT-53** (January, 2007), 350–355.

[NKG16] C. Nair, H. Kim, and A. El Gamal, *On the optimality of randomized time division and superposition coding for the broadcast channel*, 2016 IEEE Information Theory Workshop (ITW), Sept 2016, pp. 131–135.

[PW15] Yury Polyanskiy and Yihong Wu, *Wasserstein continuity of entropy and outer bounds for interference channels*, CoRR **abs/1504.04419** (2015).

[Sas15] I. Sason, *On the corner points of the capacity region of a two-user gaussian interference channel*, IEEE Trans. Inf. Theory **IT-61** (2015), 3682–3697.

[Sat78] H. Sato, *On degraded gaussian two-user channels (corresp.)*, Information Theory, IEEE Transactions on **24** (1978), no. 5, 637–640.

[WA08] Aaron B Wagner and Venkat Anantharam, *An improved outer bound for multiterminal source coding*, IEEE Transactions on Information Theory **54** (2008), no. 5, 1919–1937.

[WBO19] X. Wu, L. P. Barnes, and A. Ozgur, *"the capacity of the relay channel": Solution to cover's problem in the gaussian case*, IEEE Transactions on Information Theory **65** (2019), no. 1, 255–275.

[WO11] Shun Watanabe and Yasutada Oohama, *Secret key agreement from vector gaussian sources by rate limited public communication*, IEEE transactions on information forensics and security **6** (2011), no. 3, 541–550.

[WO18] X. Wu and A. Ozgur, *Cut-set bound is loose for gaussian relay networks*, IEEE Transactions on Information Theory **64** (2018), no. 2, 1023–1037.

[YLL18] L. Yu, H. Li, and W. Li, *Distortion bounds for source broadcast problems*, IEEE Transactions on Information Theory **64** (2018), no. 9, 6034–6053.

Appendix A. Preliminaries

The generic manipulations that are being used in the converses are the following.

Lemma 6. *For any set of random variables (U, V, A^n, B^n) the following hold:*

$$
\begin{aligned}
I(U;A^n|V) - I(U;B^n|V) &= \sum_{i=1}^n \Big(I(U;A_i|V,B_{i+1}^n,A^{i-1}) - I(U;B_i|V,B_{i+1}^n,A^{i-1})\Big)\\
&= \sum_{i=1}^n \Big(I(U;A_i|V,A_{i+1}^n,B^{i-1}) - I(U;B_i|V,A_{i+1}^n,B^{i-1})\Big)\\
I(U;B^n|V) + I(U;A^n|V) - I(U;B^n|V) &\le \sum_{i=1}^n \Big(I(U,B_{i+1}^n;B_i|V) + I(U,B_{i+1}^n,A^{i-1};A_i|V) - I(U,B_{i+1}^n,A^{i-1};B_i|V)\Big)\\
&= \sum_{i=1}^n \Big(I(U,B^{i-1};B_i|V) + I(U,B_{i+1}^n,A^{i-1};A_i|V) - I(U,B_{i+1}^n,A^{i-1};B_i|V)\Big)\\
&= \sum_{i=1}^n \Big(I(U,B_{i+1}^n;B_i|V) + I(U,A_{i+1}^n,B^{i-1};A_i|V) - I(U,A_{i+1}^n,B^{i-1};B_i|V)\Big)\\
&= \sum_{i=1}^n \Big(I(U,B^{i-1};B_i|V) + I(U,A_{i+1}^n,B^{i-1};A_i|V) - I(U,A_{i+1}^n,B^{i-1};B_i|V)\Big)\\
I(U;A^n|V) + I(V;B^n) &\le \sum_{i=1}^n \Big(I(U;A_i|V,A_{i+1}^n,B^{i-1}) + I(V,A_{i+1}^n,B^{i-1};B_i)\Big)\\
I(U;A^n|V) + I(V;B^n) &\le \sum_{i=1}^n \Big(I(U;A_i|V,B_{i+1}^n,A^{i-1}) + I(V,B_{i+1}^n,A^{i-1};B_i)\Big)
\end{aligned}
$$

Proof. The proof follows immediately from repeated applications of Lemma 1, chain rule for mutual information, and non-negativity of mutual information. The details are omitted as they are standard in literature. □

Appendix B. Proofs of Propositions for scalar Gaussian Relay channels

B.1. Proof of Proposition 1.

Proof. Remember that

$$Y_R = g_{21}X + Z_1, \quad Y = g_{31}X + g_{32}X_R + Z_2.$$

Take some $\alpha \in (0,1]$ and assume that $Z_1 = \alpha Z_3 + \sqrt{1-\alpha^2}Z_4$ where Z_2, Z_3 and Z_4 are mutually independent standard Gaussian random variables. Then, let

$$J = (g_{21}/\alpha)X + Z_3.$$

We have $Y_R = \alpha J + \sqrt{1-\alpha^2}Z_4$. Thus, J is an enhanced version of Y_R and $X \to J \to Y_R$ forms a Markov chain. Restricting the upper bound in Theorem 1 to the above families of J we obtain:

$$
\begin{aligned}
R \le \max_{p(X,X_R)\in\mathcal{P}} \min\Big(&I(X,X_R;Y),\ I(X;Y,Y_R|X_R),\\
&\min_{\alpha\in(0,1]}\ \max_{W:W\to(X,X_R)\to(Y,Y_R,J)} I(X;J,Y_R|X_R) + I(X,X_R;Y|W) - I(X;J|W)\Big),
\end{aligned}
\tag{31}
$$

where $\mathcal{P}$ is the set of all p_{X,X_R} satisfying the power constraint, i.e., $\mathrm{E}(X^2) \le P, \mathrm{E}(X_R^2) \le P$. Hence we know that there exists some $\rho \in [-1,1]$ such that

$$K_{X,X_R} \preceq \begin{bmatrix} P & \rho P \\ \rho P & P \end{bmatrix} =: K_\rho. \tag{32}$$

Elementary facts about Gaussian extremality for (conditional) differential entropy with respect to a covariance constraint shows that if (32) holds then

$$I(X,X_R;Y) \le \mathsf{C}(S_{31} + S_{32} + 2\rho\sqrt{S_{31}S_{32}}), \tag{33}$$

$$I(X;Y,Y_R|X_R) \le \mathsf{C}((1-\rho^2)(S_{31}+S_{21})), \tag{34}$$

where $\mathsf{C}(x) = \frac{1}{2}\ln(1+x)$.

With K_ρ defined as in (32), note that

$$\begin{aligned}
&\max_{\rho\in[-1,1]} \max_{\substack{p_{X,X_R}:\\ K_{X,X_R}=K_\rho}} \min_{\alpha\in(0,1]} \max_{W:W\to(X,X_R)\to(Y,Y_R,J)} I(X;J,Y_R|X_R) + I(X,X_R;Y|W) - I(X;J|W)\\
&\le \max_{\rho\in[-1,1]} \min_{\alpha\in(0,1]} \max_{\substack{p_{X,X_R}:K_{X,X_R}\preceq K_\rho\\ W:W\to(X,X_R)\to(Y,Y_R,J)}} I(X;J,Y_R|X_R) + I(X,X_R;Y|W) - I(X;J|W)\\
&\overset{(a)}{=} \max_{\rho\in[-1,1]} \min_{S_J\ge S_{21}} \left(\mathsf{C}(S_J(1-\rho^2)) + \left(\max_{\substack{p_{XX_R}\sim\mathcal{N}(0,K')\\ K'\preceq K_\rho}} I(X,X_R;Y) - I(X;J)\right)\right)
\end{aligned}$$

where (a) follows from Lemma 7 below and from $S_J := \frac{S_{21}}{\alpha^2}$. Further from the second part of Lemma 7, we know that $\left(\frac{g_{21}}{\alpha} - f g_{31}\right) X + f g_{32} X_R = 0$ almost surely, or that the maximizing K' takes the form

$$K' = \begin{bmatrix} aP & \pm P\sqrt{ab} \\ \pm P\sqrt{ab} & bP \end{bmatrix} \preceq \begin{bmatrix} P & \rho P \\ \rho P & P \end{bmatrix}.$$

For the matrix ordering above it is necessary and sufficient that $0 \le a, b \le 1$ and

$$1 - a - b \ge \rho^2 \mp 2\rho\sqrt{ab} \tag{35}$$

Observe that, when K' is defined as above

$$I(X,X_R;Y) - I(X;J) \quad = \frac{1}{2}\log(1 + aS_{31} + bS_{32} \pm 2\sqrt{abS_{31}S_{32}}) - \frac{1}{2}\log(1+aS_J).$$

Optimizing the term above with respect to b for a fixed a subject to (35) we obtain that

$$\begin{aligned}
&\max_{\substack{p_{XX_R}\sim\mathcal{N}(0,K')\\ K'\preceq K_\rho}} I(X,X_R;Y) - I(X;J)\\
&\quad = \max_{0\le a\le 1} \frac{1}{2}\log\left(1 + \left(\sqrt{aS_{31}} + \sqrt{\rho^2 a S_{32}} + \sqrt{(1-\rho^2)(1-a)S_{32}}\right)^2\right) - \frac{1}{2}\log(1+aS_J).
\end{aligned} \tag{36}$$

Setting $a = \frac{x^2}{1+x^2}$, we see that the optimal x for the above maximization problem is the unique non-negative root of the quadratic equation:

$$\begin{aligned}
&x^2(\sqrt{S_{31}} + \sqrt{\rho^2 S_{32}})(\sqrt{(1-\rho^2)S_{32}})(1+S_J) + x\left((1-\rho^2)S_{32}(1+S_J) - (\sqrt{S_{31}} + \sqrt{\rho^2 S_{32}})^2 + S_J\right)\\
&\quad - (\sqrt{S_{31}} + \sqrt{\rho^2 S_{32}})(\sqrt{(1-\rho^2)S_{32}}) = 0.
\end{aligned} \tag{37}$$

This completes the proof. □

Remark 12. From the above argument, we know that

$$\max_{\substack{p_{XX_R}\sim\mathcal{N}(0,K')\\ K'\preceq K_\rho}} I(X,X_R;Y) - I(X;J)$$

$$
\begin{aligned}
&= \max_{\substack{p_{XX_R}\sim\mathcal{N}(0,K_\rho)\\ W\to X,X_R\to Y,J}} I(X,X_R;Y|W) - I(XX_R;J|W)\\
&\le \max_{U,W\to X,X_R\to YJ} I(U;Y|W) - I(U;J|W) \qquad (38)\\
&\overset{(a)}{=} \frac{1}{2}\sum_i [\log\phi_i]_+,
\end{aligned}
$$

where (a) follows from [WO11, Eq. 7]. Here $[x]_+$ is zero if x is negative and x otherwise, and ϕ_i are the set of generalized eigenvalues for the pencil

$$\left(K^{\frac{1}{2}}_{X,X_R}\begin{bmatrix} g_{31}^2 & g_{31}g_{32}\\ g_{31}g_{32} & g_{32}^2\end{bmatrix}K^{\frac{1}{2}}_{X,X_R} + I_2,\quad K^{\frac{1}{2}}_{X,X_R}\begin{bmatrix}(g_{21}/\alpha)^2 & 0\\ 0 & 0\end{bmatrix}K^{\frac{1}{2}}_{X,X_R} + I_2\right).$$

Direct calculation shows that ϕ_i are the roots of the quadratic polynomial

$$2\rho\sqrt{S_{31}S_{32}} + S_{31} + S_{32} + 1 - \lambda\big(S_{32}S_J(1-\rho^2) + S_{31} + S_{32} + S_J + 2 + 2\rho\sqrt{S_{31}S_{32}}\big) + \lambda^2(S_J+1) = 0. \qquad (39)$$

Only one of the roots of this polynomial is larger than one. Denoting the larger root by $\lambda_{\max}$, we obtain

$$\max_{U,W\to X,X_R\to YJ} I(U;Y|W) - I(U;J|W) = \frac{1}{2}\ln\lambda_{\max}.$$

On the other hand, routine calculation shows that after substituting the unique non-negative root of (37) in (36) and simplifying the expression, we have

$$\max_{\substack{p_{XX_R}\sim\mathcal{N}(0,K')\\ K'\preceq K_\rho}} I(X,X_R;Y) - I(X,X_R;J) = \frac{1}{2}\ln\lambda_{\max}.$$

Thus, for this setting, the inequality in (38) is indeed an equality.

Theorem 9 (Theorem 8 in [LV07], see also Theorem 1 in [GN14])**.** *Let $\mathbf{Z}_1,\mathbf{Z}_2$ be two Gaussian vectors with with strictly positive definite covariance matrices K_{Z_1} and K_{Z_2}, respectively. Let $\mu\ge 1$ be a real number, S be a positive semidefinite matrix, and W be a random variable independent of $\mathbf{Z}_1,\mathbf{Z}_2$. Consider the optimization problem*

$$
\begin{aligned}
&\max_{p(\mathbf{x}|w)} h(\mathbf{X}+\mathbf{Z}_1|W) - \mu h(\mathbf{X}+\mathbf{Z}_2|W)\\
&\textit{subject to } Cov(\mathbf{X}) \preceq S
\end{aligned}
$$

where the maximization is over $p(w,\mathbf{x}) : Cov(\mathbf{X})\preceq S$, and $W,\mathbf{X}$ are independent of $\mathbf{Z}_1,\mathbf{Z}_2$. A Gaussian $p(\mathbf{x}|w)$ with the same covariance matrix for each w is an optimal solution of this optimization problem.

Remark 13. In [LV07], the constraint for the optimization problem is listed as $\mathrm{Cov}(\mathbf{X}|U)\preceq S$, which is seemingly a weaker statement. However what the authors proved and intended to prove is indeed the statement mentioned above. There is also an alternate proof of the result in [GN14].

Lemma 7. *Let A,B,C be matrices such that there exists $\tilde{A}$, $\tilde{B}$, $\tilde{C}$, and $\tilde{D}$ such that the extended matrices*

$$A_e := \begin{bmatrix} A & B\\ \tilde{A} & \tilde{B}\end{bmatrix},\quad C_e := \begin{bmatrix} C & 0\\ \tilde{C} & \tilde{D}\end{bmatrix}$$

are invertible. Assume that $\mathbf{Y} = A\mathbf{X} + B\mathbf{X_R} + \mathbf{Z}$ and $\mathbf{J} = C\mathbf{X}+\mathbf{Z}$ for some vectors $\mathbf{X},\mathbf{X_R},\mathbf{Y}$ and $\mathbf{J}$, and a Gaussian random vector $\mathbf{Z}$ independent of $(\mathbf{X},\mathbf{X_R})$ and distributed as $\mathcal{N}(0,I)$.

- *(i) To compute the maximum over $p(w,\mathbf{x},\mathbf{x_R})$, subject to $K_{\mathbf{XX_R}}\preceq K_0$ and $W-(\mathbf{X},\mathbf{X_R})-(\mathbf{Y},\mathbf{J})$, of the expression*

$$I(\mathbf{X};\mathbf{J}|\mathbf{X_R}) + I(\mathbf{XX_R};\mathbf{Y}|W) - I(\mathbf{XX_R};\mathbf{J}|W)$$

  it suffices to assume that $U\sim\mathcal{N}(0,K')$, $W\sim\mathcal{N}(0,K_0-K')$, where U and W are independent; and $[\mathbf{X}\ \mathbf{X_R}]^T = U+W$, for some $K'\preceq K_0$.
- *(ii) Further, if $\mathbf{X}^{(a)},\mathbf{X}^{(a)}_{\mathbf{R}}\sim\mathcal{N}(0,K')$, where K' is the maximizer mentioned above, then $(C-FA)\mathbf{X}^{(a)} - FB\mathbf{X}_R^{(a)} = 0$ for some F.*

Proof. Note that the matrices

$$A_\epsilon := \begin{bmatrix} A & B \\ \epsilon\tilde{A} & \epsilon\tilde{B} \end{bmatrix}, \quad C_\epsilon := \begin{bmatrix} C & 0 \\ \epsilon\tilde{C} & \epsilon\tilde{D} \end{bmatrix}$$

are invertible for every $\epsilon \neq 0$. Let $\tilde{\mathbf{Z}} \sim \mathcal{N}(0, I)$ be independent of $\mathbf{X}, \mathbf{X_R}$ and $\mathbf{Z}$. Define $\hat{\mathbf{Z}} = [\mathbf{Z}\ \tilde{\mathbf{Z}}]^T$. Define $\hat{\mathbf{X}} := [\mathbf{X}\ \mathbf{X_R}]^T$, $\hat{\mathbf{Y}}_\epsilon := A_\epsilon\tilde{\mathbf{X}} + \hat{\mathbf{Z}}$ and $\hat{\mathbf{J}}_\epsilon := C_\epsilon\tilde{\mathbf{X}} + \hat{\mathbf{Z}}$. Consider the expression

$$I(\mathbf{X}; \hat{\mathbf{J}}_\epsilon|\mathbf{X_R}) + I(\hat{\mathbf{X}}; \hat{\mathbf{Y}}_\epsilon|W) - I(\hat{\mathbf{X}}; \hat{\mathbf{J}}_\epsilon|W). \tag{40}$$

From Theorem 9 we know that there exists a matrix $K'_\epsilon \preceq K_0$, such that $\hat{\mathbf{X}}|\{W = w\} \sim \mathcal{N}(0, K'_\epsilon)$ for every w, maximizes the term $I(\hat{\mathbf{X}}; \hat{\mathbf{Y}}_\epsilon|W) - I(\hat{\mathbf{X}}; \hat{\mathbf{J}}_\epsilon|W)$. Thus the choice $U_\epsilon \sim \mathcal{N}(0, K'_\epsilon)$, $W_\epsilon \sim \mathcal{N}(0, K_0 - K'_\epsilon)$, where U_ϵ and W_ϵ are independent; and $[\mathbf{X}\ \mathbf{X_R}]^T = U_\epsilon + W_\epsilon$, maximizes the expression in (40) over $p(w, \mathbf{x}, \mathbf{x_R})$, subject to $K_{\mathbf{XX_R}} \preceq K_0$. For any $\mathbf{X}, \mathbf{X_R}$ such that $K_{\mathbf{XX_R}} \preceq K_0$, observe the following:

$$\begin{aligned} I(\mathbf{X}; \mathbf{J}|\mathbf{X_R}) &\le I(\mathbf{X}; \hat{\mathbf{J}}_\epsilon|\mathbf{X_R}) \\ &= I(\mathbf{X}; \mathbf{J}|\mathbf{X_R}) + I(\mathbf{X}; \epsilon\tilde{C}\mathbf{X} + \epsilon\tilde{D}\mathbf{X_R} + \tilde{\mathbf{Z}}|\mathbf{J}, \mathbf{X_R}) \\ &\le I(\mathbf{X}; \mathbf{J}|\mathbf{X_R}) + I(\mathbf{X}, \mathbf{X_R}; \epsilon\tilde{C}\mathbf{X} + \epsilon\tilde{D}\mathbf{X_R} + \tilde{\mathbf{Z}}) \\ &\le I(\mathbf{X}; \mathbf{J}|\mathbf{X_R}) + \frac{1}{2}\log\left|I + \epsilon^2[\tilde{C}\ \tilde{D}]K_0[\tilde{C}\ \tilde{D}]^T\right|. \end{aligned}$$

Similarly, mimicking the steps as above,

$$I(\mathbf{XX_R}; \mathbf{Y}|W) \le I(\hat{\mathbf{X}}; \hat{\mathbf{Y}}_\epsilon|W) \le I(\mathbf{XX_R}; \mathbf{Y}|W) + \frac{1}{2}\log\left|I + \epsilon^2[\tilde{A}\ \tilde{B}]K_0[\tilde{A}\ \tilde{B}]^T\right|,$$
$$I(\mathbf{XX_R}; \mathbf{J}|W) \le I(\hat{\mathbf{X}}; \hat{\mathbf{J}}_\epsilon|W) \le I(\mathbf{XX_R}; \mathbf{J}|W) + \frac{1}{2}\log\left|I + \epsilon^2[\tilde{C}\ \tilde{D}]K_0[\tilde{C}\ \tilde{D}]^T\right|.$$

Therefore as $\epsilon \to 0$, the maximum of the expression in (40) tends to maximum value of the optimization problem in the theorem. Since $K : K \preceq K_0$ is a compact set, there is a K' that is a subsequential limit of K'_ϵ. Hence from the continuity of $\log|K|$, we see that there is a maximizer for the optimization problem in the theorem of the form: $U \sim \mathcal{N}(0, K')$, $W \sim \mathcal{N}(0, K_0 - K')$, where U and W are independent; and $[\mathbf{X}\ \mathbf{X_R}]^T = U + W$, for some $K' \preceq K_0$. The completes the proof of (i).

Let $\mathbf{X}^{(a)}, \mathbf{X}_{\mathbf{R}}^{(a)} \sim \mathcal{N}(0, K')$, then from the optimality above we have that

$$I(\mathbf{X}^{(a)}\mathbf{X_R}^{(a)}; \mathbf{Y}^{(a)}) - I(\mathbf{X}^{(a)}\mathbf{X_R}^{(a)}; \mathbf{J}^{(a)}) \ge I(\mathbf{X}^{(a)}\mathbf{X_R}^{(a)}; \mathbf{Y}^{(a)}|W) - I(\mathbf{X}^{(a)}\mathbf{X_R}^{(a)}; \mathbf{J}^{(a)}|W) \tag{41}$$

for all W such that $(W, \mathbf{X}^{(a)}, \mathbf{X_R}^{(a)})$ is independent of $\mathbf{Z}$; where $\mathbf{Y}^{(a)} := A\mathbf{X}^{(a)} + B\mathbf{X_R}^{(a)} + \mathbf{Z}$ and $\mathbf{J}^{(a)} := C\mathbf{X}^{(a)} + \mathbf{Z}$. Express $C\mathbf{X}^{(a)} = F(A\mathbf{X}^{(a)} + B\mathbf{X_R}^{(a)}) + \mathbf{X}^{(b)}$, where $\mathbf{X}^{(b)}$ is independent of $A\mathbf{X}^{(a)} + B\mathbf{X_R}^{(a)}$ (such a decomposition is feasible for jointly Gaussian vectors). Let $W_b = \mathbf{X}^{(b)}$ to obtain that $W_b, \mathbf{Y}^{(a)}$ are independent. Consequently $I(\mathbf{X}^{(a)}\mathbf{X_R}^{(a)}; \mathbf{Y}^{(a)}) = I(\mathbf{X}^{(a)}\mathbf{X_R}^{(a)}; \mathbf{Y}^{(a)}|W_b)$ and using (41) we must have $W_b, \mathbf{J}^{(a)}$ to be independent. From a simple application of characteristic functions (see Proposition 2 in [GN14] for instance) it follows that $\mathrm{E}\left[\mathbf{X}^{(b)}(C\mathbf{X}^{(a)})^T\right] = 0$. This further implies that $\mathrm{E}(\mathbf{X}^{(b)}(\mathbf{X}^{(b)})^T) = 0$, showing that $\mathbf{X}^{(b)} = 0$ almost surely. Thus have $(C - FA)\mathbf{X}^{(a)} - FB\mathbf{X}_R^{(a)} = 0$ almost surely. □

B.2. **Proof of Proposition 2.**

Proof. When $S_{21} < S_J < S_{31}$, the third bound (the new one) for the capacity of the relay channel in Proposition 1, can be bounded from above (using routine calculations given in Appendix E) by

$$\begin{aligned} &\max_{\rho\in[-1,1]} \min_{S_J : S_{21}\le S_J < S_{31}} \frac{1}{2}\log(1 + S_J(1-\rho^2)) + \frac{1}{2}\log\left(\frac{1 + S_{31} + \rho^2 S_{32} + 2\sqrt{\rho^2 S_{31}S_{32}}}{1 + S_J} + \frac{S_{31}S_{32}(1-\rho^2)}{S_{31} - S_J}\right) \\ &\quad\le \min_{S_J : S_{21}\le S_J < S_{31}} \max_{\rho\in[-1,1]} \frac{1}{2}\log(1 + S_J(1-\rho^2)) + \frac{1}{2}\log\left(\frac{1 + S_{31} + \rho^2 S_{32} + 2\sqrt{\rho^2 S_{31}S_{32}}}{1 + S_J} + \frac{S_{31}S_{32}(1-\rho^2)}{S_{31} - S_J}\right) \\ &\quad\overset{(a)}{\le} \min_{S_J : S_{21}\le S_J < S_{31}} \frac{1}{2}\log\left(1 + S_{31} + S_{32}\left(\frac{(1+S_J)^2 S_{31}^2}{S_J(1+S_{31})(S_{31} - S_J)}\right)\right). \end{aligned}$$

where the bound in (a) again follows from routine algebra. Considering the derivative with respect to S_J, we get the optimum value for S_J being equal to

$$S_J = \max\left(S_{21}, \frac{S_{31}}{S_{31}+2}\right).$$

□

Appendix C. Proofs of results for Gaussian Z-interference

C.1. **Proof of Theorem 3.**

Proof. Let $J = X_1 + uZ$ where Z is a standard normal random variable such that $Z_2 = auZ + \sqrt{1-a^2u^2}Z'$ for some standard normal random variable Z' that is independent of Z. From Corollary 2 and standard arguments, it follows that for any $\alpha, \beta \in [0,1]$,

$$\begin{aligned} R_1 + \lambda R_2 \leq \sup_{p_Q p_{X_1|Q} p_{X_2|Q} p_{W|X_1,X_2,Q}} & \lambda\alpha I(X_2; Y_2|W, X_1, Q) + \lambda(1-\alpha)(I(X_2; Y_2|W, Q) - I(X_2; J|W, Q)) \\ & + \beta I(X_1; Y_1|Q) + (1-\beta)(I(W; Y_2|Q) + I(X_1; J|W, Q)), \end{aligned}$$

where X_1 and X_2 are assumed to satisfy the power constraints, and also satisfying that $I(X_1; J|W, Q) \geq I(X_1; Y_2|W, Q)$.

Let us make the following definitions

$$\begin{aligned} \mathbf{Y}_2^{(\epsilon)} &:= \begin{bmatrix} a & 1 \\ \epsilon & \epsilon \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} auZ_1 + \sqrt{1-a^2u^2}Z_2 \\ Z_3 \end{bmatrix} \\ \mathbf{J}^{(\epsilon)} &:= \begin{bmatrix} 1 & 0 \\ \epsilon & \epsilon \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} uZ_1 \\ Z_3 \end{bmatrix}, \\ \hat{\mathbf{J}}^{(\epsilon)} &:= \begin{bmatrix} \epsilon a & \epsilon \\ \epsilon & \epsilon \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} Z_4 \\ Z_3 \end{bmatrix} \\ \bar{\mathbf{J}}^{(\epsilon)} &:= \begin{bmatrix} \epsilon & 0 \\ \epsilon & \epsilon \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} Z_5 \\ Z_3 \end{bmatrix} \end{aligned}$$

where Z_i's are mutually independent standard Gaussian random variables. The construction ensures that $\hat{\mathbf{J}}^{(\epsilon)}$ is a stochastically degraded version of $\mathbf{Y}_2^{(\epsilon)}$ for any $\epsilon < 1$, and that $\mathbf{J}^{(\epsilon)} \to (X_1, \hat{\mathbf{J}}^{(\epsilon)}) \to X_2$ is Markov.

For any $\epsilon > 0, \gamma > 1$ consider the (perturbed) expression

$$\begin{aligned} & \lambda\alpha I(X_2; \mathbf{Y}_2|W, X_1, Q) - \lambda\alpha I(X_2; \hat{\mathbf{J}}|W, X_1, Q) + \lambda(1-\alpha)(I(X_2; \mathbf{Y}_2|W, Q) - I(X_2; \mathbf{J}|W, Q)) \\ & \quad + \beta I(X_1; Y_1|Q) + (1-\beta)(I(W; \mathbf{Y}_2|Q) + I(X_1; \mathbf{J}|W, Q) + I(X_2; \bar{\mathbf{J}}|W, X_1, Q)) \\ & \quad + \epsilon I(X_1, X_2; \mathbf{Y}_2|W, Q) - \gamma\epsilon I(X_1, X_2; \mathbf{J}|W, Q). \end{aligned}$$

This expression is essentially the same as the expression on the right-hand-side, save for perturbation terms in red. Note that, given power constraints, $\epsilon I(X_1, X_2; \mathbf{Y}_2|W, Q), \gamma\epsilon I(X_1, X_2; \mathbf{J}|W, Q), I(X_2; \hat{\mathbf{J}}|W, X_1, Q)$ are non-negative and bounded from above by some $g(\epsilon)$ that tends to zero as $\epsilon \to 0$. Let V be the maximum value of the expression among all the distributions satisfying the power constraints, the structure of the form $p_Q p_{X_1|Q} p_{X_2|Q} p_{W|X_1,X_2,Q}$, and $I(X_1; \mathbf{J}|W, Q) \geq I(X_1; \mathbf{Y}_2|W, Q)$. Let $p^*_Q p^*_{X_1|Q} p^*_{X_2|Q} p^*_{W|X_1,X_2,Q}$ be a maximizer[3].

Take two i.i.d. copies of the maximizer and denote them using subscripts a, b respectively. Let $(\cdot)_+ = \frac{(\cdot)_a + (\cdot)_b}{\sqrt{2}}$ and let $(\cdot)_- = \frac{(\cdot)_a - (\cdot)_b}{\sqrt{2}}$. Then, observe that for any $\epsilon < 1$, $\hat{\mathbf{J}}_-$ is a stochastically degraded version of $\mathbf{Y}_{2-}$ and $\hat{\mathbf{J}}_+$ is a stochastically degraded version of $\mathbf{Y}_{2+}$. Moreover, $\mathbf{J}_- \to (X_{1-}, \hat{\mathbf{J}}_-) \to X_{2-}$ is Markov.

---

[3]In the case of the Gaussian Z-interference channel, routine arguments will show that there is a maximizer - the power constraints yield tightness, and the additive Gaussian noise yields the continuity for the various terms with respect to weak convergence. (For one illustration, the readers are advised to look at the Appendix of [GN14]; these kind of arguments are rather standard in functional analysis.)

Now we have, by mimicking the single-letterization manipulations used in the proof of Theorem 2 after the equality in (a) below,

$$\begin{aligned}
2V &= \lambda\alpha I(X_{2a},X_{2b};\mathbf{Y}_{2a},\mathbf{Y}_{2b}|W_a,W_b,X_{1a},X_{1b},Q_a,Q_b) - \lambda\alpha I(X_{2a},X_{2b};\hat{\mathbf{J}}_a,\hat{\mathbf{J}}_b|W_a,W_b,X_{1a},X_{1b},Q_a,Q_b)\\
&\quad + \lambda(1-\alpha)I(X_{2a},X_{2b};\mathbf{Y}_{2a},\mathbf{Y}_{2b}|W_a,W_b,Q_a,Q_b)\\
&\quad - \lambda(1-\alpha)I(X_{2a},X_{2b};\mathbf{J}_a,\mathbf{J}_b|W_a,W_b,Q_a,Q_b) + \beta I(X_{1a},X_{1b};\mathbf{Y}_{1a},\mathbf{Y}_{1b}|Q_a,Q_b)\\
&\quad + (1-\beta)(I(W_a,W_b;\mathbf{Y}_{2a},\mathbf{Y}_{2b}|Q_a,Q_b) + I(X_{1a},X_{1b};\mathbf{J}_a,\mathbf{J}_b|W_a,W_b,Q_a,Q_b)\\
&\quad + I(X_{2a},X_{2b};\bar{\mathbf{J}}_a,\bar{\mathbf{J}}_b|W_a,W_b,X_{1a},X_{1b},Q_a,Q_b))\\
&\quad + \epsilon I(X_{1a},X_{1b},X_{2a},X_{2b};\mathbf{Y}_{2a},\mathbf{Y}_{2b}|W_a,W_b,Q_a,Q_b) - \gamma\epsilon I(X_{1a},X_{1b},X_{2a},X_{2b};\mathbf{J}_a,\mathbf{J}_b|W_a,W_b,Q_a,Q_b)\\
&\overset{(a)}{=} \lambda\alpha I(X_{2+},X_{2-};\mathbf{Y}_{2+},\mathbf{Y}_{2-}|W_a,W_b,X_{1+},X_{1-},Q_a,Q_b) - \lambda\alpha I(X_{2+},X_{2-};\hat{\mathbf{J}}_+,\hat{\mathbf{J}}_-|W_a,W_b,X_{1+},X_{1-},Q_a,Q_b)\\
&\quad + \lambda(1-\alpha)I(X_{2+},X_{2-};\mathbf{Y}_{2+},\mathbf{Y}_{2-}|W_a,W_b,Q_a,Q_b)\\
&\quad - \lambda(1-\alpha)I(X_{2+},X_{2-};\mathbf{J}_+,\mathbf{J}_-|W_a,W_b,Q_a,Q_b) + \beta I(X_{1+},X_{1-};\mathbf{Y}_{1+},\mathbf{Y}_{1-}|Q_a,Q_b)\\
&\quad + (1-\beta)(I(W_a,W_b;\mathbf{Y}_{2+},\mathbf{Y}_{2-}|Q_a,Q_b) + I(X_{1+},X_{1-};\mathbf{J}_+,\mathbf{J}_-|W_a,W_b,Q_a,Q_b)\\
&\quad + I(X_{2+},X_{2-};\bar{\mathbf{J}}_+,\bar{\mathbf{J}}_-|W_a,W_b,X_{1+},X_{1-},Q_a,Q_b))\\
&\quad + \epsilon I(X_{1+},X_{1-},X_{2+},X_{2-};\mathbf{Y}_{2+},\mathbf{Y}_{2-}|W_a,W_b,Q_a,Q_b) - \gamma\epsilon I(X_{1+},X_{1-},X_{2+},X_{2-};\mathbf{J}_+,\mathbf{J}_-|W_a,W_b,Q_a,Q_b)\\
&\overset{(b)}{=} \lambda\alpha I(X_{2+};\mathbf{Y}_{2+}|W_a,W_b,X_{1+},\mathbf{J}_-,Q_a,Q_b) - \lambda\alpha I(X_{2+};\hat{\mathbf{J}}_+|W_a,W_b,X_{1+},\mathbf{J}_-,Q_a,Q_b)\\
&\quad + \lambda(1-\alpha)I(X_{2+};\mathbf{Y}_{2+}|W_a,W_b,\mathbf{J}_-,Q_a,Q_b)\\
&\quad - \lambda(1-\alpha)I(X_{2+};\mathbf{J}_+|W_a,W_b,\mathbf{J}_-,Q_a,Q_b) + \beta I(X_{1+};\mathbf{Y}_{1+}|Q_a,Q_b)\\
&\quad + (1-\beta)(I(W_a,W_b,\mathbf{J}_-;\mathbf{Y}_{2+}|Q_a,Q_b) + I(X_{1+};\mathbf{J}_+|W_a,W_b,\mathbf{J}_-,Q_a,Q_b)\\
&\quad + I(X_{2+};\bar{\mathbf{J}}_+|W_a,W_b,X_{1+},\mathbf{J}_-,Q_a,Q_b))\\
&\quad + \epsilon I(X_{1+},X_{2+};\mathbf{Y}_{2+}|W_a,W_b,\mathbf{J}_-,Q_a,Q_b) - \gamma\epsilon I(X_{1+},X_{2+};\mathbf{J}_+|W_a,W_b,\mathbf{J}_-,Q_a,Q_b)\\
&\quad + \lambda\alpha I(X_{2-};\mathbf{Y}_{2-}|W_a,W_b,X_{1-},Q_a,Q_b,\mathbf{Y}_{2+}) - \lambda\alpha I(X_{2-};\hat{\mathbf{J}}_-|W_a,W_b,X_{1-},Q_a,Q_b,\mathbf{Y}_{2+})\\
&\quad + \lambda(1-\alpha)I(X_{2-};\mathbf{Y}_{2-}|W_a,W_b,\mathbf{Y}_{2+},Q_a,Q_b)\\
&\quad - \lambda(1-\alpha)I(X_{2-};\mathbf{J}_-|W_a,W_b,\mathbf{Y}_{2+},Q_a,Q_b) + \beta I(X_{1-};\mathbf{Y}_{1-}|Q_a,Q_b)\\
&\quad + (1-\beta)(I(W_a,W_b,\mathbf{Y}_{2+};\mathbf{Y}_{2-}|Q_a,Q_b) + I(X_{1-};\mathbf{J}_-|W_a,W_b,Q_a,Q_b,\mathbf{Y}_{2+})\\
&\quad + I(X_{2-};\bar{\mathbf{J}}_-|W_a,W_b,X_{1-},Q_a,Q_b,\mathbf{Y}_{2+}))\\
&\quad + \epsilon I(X_{1-},X_{2-};\mathbf{Y}_{2-}|W_a,W_b,Q_a,Q_b,\mathbf{Y}_{2_+}) - \gamma\epsilon I(X_{1-},X_{2-};\mathbf{J}_-|W_a,W_b,Q_a,Q_b,\mathbf{Y}_{2+})\\
&\quad - \big(\lambda\alpha I(X_{1-},\hat{\mathbf{J}}_-;\mathbf{Y}_{2+}|W_a,W_b,X_{1+},\mathbf{J}_-,Q_a,Q_b) - \lambda\alpha I(X_{1-},\hat{\mathbf{J}}_-;\hat{\mathbf{J}}_+|W_a,W_b,X_{1+},\mathbf{J}_-,Q_a,Q_b)\big)\\
&\quad - \big(\lambda\alpha I(X_{1+};\mathbf{Y}_{2-}|W_a,W_b,X_{1-},Q_a,Q_b,\mathbf{Y}_{2+}) - \lambda\alpha I(X_{1+};\hat{\mathbf{J}}_-|W_a,W_b,X_{1-},Q_a,Q_b,\mathbf{Y}_{2+})\big)\\
&\quad - \lambda(1-\alpha)I(X_{2-};\mathbf{J}_+|W_a,W_b,X_{2+},\mathbf{Y}_{2+},\mathbf{J}_-,Q_a,Q_b) - \lambda(1-\alpha)I(X_{2+};\mathbf{J}_-|W_a,W_b,X_{2-},\mathbf{Y}_{2-},Q_a,Q_b)\\
&\quad - \beta I(\mathbf{Y}_{1+};\mathbf{Y}_{1-}|Q_a,Q_b) - (1-\beta)I(\mathbf{Y}_{2+};\mathbf{Y}_{2-}|Q_a,Q_b) - (\gamma-1)\epsilon I(\mathbf{Y}_{2+};\mathbf{J}_-|W_a,W_b,Q_a,Q_b)
\end{aligned}$$

where (a) follows since bijections preserve mutual information and (b) from chain rule and data-processing equality from the Markov structure relating the various variables. In particular, we used that

$$\mathbf{J}_- \to (X_{1-},\hat{\mathbf{J}}_-,W_a,W_b,Q_a,Q_b) \to (X_{1+},X_{2+}) \to (\mathbf{Y}_{2+},\hat{\mathbf{J}}_+)$$

forms a Markov chain and the following expansion

$$\begin{aligned}
&I(X_{2+},X_{2-};\mathbf{Y}_{2+},\mathbf{Y}_{2-}|W_a,W_b,X_{1+},X_{1-},Q_a,Q_b) - I(X_{2+},X_{2-};\hat{\mathbf{J}}_+,\hat{\mathbf{J}}_-|W_a,W_b,X_{1+},X_{1-},Q_a,Q_b)\\
&= I(X_{2+};\mathbf{Y}_{2+}|W_a,W_b,X_{1+},X_{1-},\hat{\mathbf{J}}_-,Q_a,Q_b) - I(X_{2+};\hat{\mathbf{J}}_+|W_a,W_b,X_{1+},X_{1-},\hat{\mathbf{J}}_-,Q_a,Q_b)\\
&\quad + I(X_{2-};\mathbf{Y}_{2-}|W_a,W_b,X_{1+},X_{1-},\mathbf{Y}_{2+},Q_a,Q_b) - I(X_{2-};\hat{\mathbf{J}}_-|W_a,W_b,X_{1+},X_{1-},\mathbf{Y}_{2+},Q_a,Q_b)\\
&= I(X_{2+};\mathbf{Y}_{2+}|W_a,W_b,X_{1+},X_{1-},\hat{\mathbf{J}}_-,\mathbf{J}_-,Q_a,Q_b) - I(X_{2+};\hat{\mathbf{J}}_+|W_a,W_b,X_{1+},X_{1-},\hat{\mathbf{J}}_-,\mathbf{J}_-,Q_a,Q_b)
\end{aligned}$$

$$
\begin{aligned}
&+ I(X_{2-};\mathbf{Y}_{2-}|W_a,W_b,X_{1+},X_{1-},\mathbf{Y}_{2+},Q_a,Q_b) - I(X_{2-};\hat{\mathbf{J}}_{-}|W_a,W_b,X_{1+},X_{1-},\mathbf{Y}_{2+},Q_a,Q_b)\\
=\ & I(X_{2+};\mathbf{Y}_{2+}|W_a,W_b,X_{1+},\mathbf{J}_{-},Q_a,Q_b) - I(X_{2+};\hat{\mathbf{J}}_{+}|W_a,W_b,X_{1+},\mathbf{J}_{-},Q_a,Q_b)\\
&+ I(X_{2-};\mathbf{Y}_{2-}|W_a,W_b,X_{1-},\mathbf{Y}_{2+},Q_a,Q_b) - I(X_{2-};\hat{\mathbf{J}}_{-}|W_a,W_b,X_{1-},\mathbf{Y}_{2+},Q_a,Q_b)\\
&- \big(I(X_{1-},\hat{\mathbf{J}}_{-};\mathbf{Y}_{2+}|W_a,W_b,X_{1+},\mathbf{J}_{-},Q_a,Q_b) - I(X_{1-},\hat{\mathbf{J}}_{-};\hat{\mathbf{J}}_{+}|W_a,W_b,X_{1+},\mathbf{J}_{-},Q_a,Q_b)\big)\\
&- \big(I(X_{1+};\mathbf{Y}_{2-}|W_a,W_b,X_{1-},Q_a,Q_b,\mathbf{Y}_{2+}) - I(X_{1+};\hat{\mathbf{J}}_{2-}|W_a,W_b,X_{1-},Q_a,Q_b,\mathbf{Y}_{2+})\big).
\end{aligned}
$$

Since $\hat{\mathbf{J}}_{+}$ ($\hat{\mathbf{J}}_{-}$) is a stochastically degraded version of $\mathbf{Y}_{2+}$ ($\mathbf{Y}_{2-}$), we have that the red-colored expressions above satisfy

$$
\begin{aligned}
&\big(I(X_{1-},\hat{\mathbf{J}}_{-};\mathbf{Y}_{2+}|W_a,W_b,X_{1+},\mathbf{J}_{-},Q_a,Q_b) - I(X_{1-},\hat{\mathbf{J}}_{-};\hat{\mathbf{J}}_{+}|W_a,W_b,X_{1+},\mathbf{J}_{-},Q_a,Q_b) \geq 0\\
&\quad I(X_{1+};\mathbf{Y}_{2-}|W_a,W_b,X_{1-},Q_a,Q_b,\mathbf{Y}_{2+}) - I(X_{1+};\hat{\mathbf{J}}_{2-}|W_a,W_b,X_{1-},Q_a,Q_b,\mathbf{Y}_{2+}) \geq 0. \qquad (42)
\end{aligned}
$$

Further note that (this is to ensure the constraint $I(X_1;J|W,Q) \geq I(X_1;Y_2|W,Q)$ remains true after rotation)

$$
\begin{aligned}
0 &\leq I(X_{1a},X_{1b};\mathbf{J}_a,\mathbf{J}_b|W_a,W_b,Q_a,Q_b) - I(X_{1a},X_{1b};\mathbf{Y}_{2a},\mathbf{Y}_{2b}|W_a,W_b,Q_a,Q_b)\\
&= I(X_{1+},X_{1-};\mathbf{J}_{+},\mathbf{J}_{-}|W_a,W_b,Q_a,Q_b) - I(X_{1+},X_{1-};\mathbf{Y}_{2+},\mathbf{Y}_{2-}|W_a,W_b,Q_a,Q_b)\\
&= I(X_{1+},X_{1-};\mathbf{J}_{+}|W_a,W_b,Q_a,Q_b,\mathbf{J}_{-}) - I(X_{1+},X_{1-};\mathbf{Y}_{2+}|W_a,W_b,Q_a,Q_b,\mathbf{J}_{-})\\
&\quad + I(X_{1+},X_{1-};\mathbf{J}_{-}|W_a,W_b,Q_a,Q_b,\mathbf{Y}_{2+}) - I(X_{1+},X_{1-};\mathbf{Y}_{2-}|W_a,W_b,Q_a,Q_b,\mathbf{Y}_{2+})\\
&= I(X_{1+};\mathbf{J}_{+}|W_a,W_b,Q_a,Q_b,\mathbf{J}_{-}) - I(X_{1+};\mathbf{Y}_{2+}|W_a,W_b,Q_a,Q_b,\mathbf{J}_{-})\\
&\quad + I(X_{1-};\mathbf{J}_{-}|W_a,W_b,Q_a,Q_b,\mathbf{Y}_{2+}) - I(X_{1-};\mathbf{Y}_{2-}|W_a,W_b,Q_a,Q_b,\mathbf{Y}_{2+})\\
&\quad + I(X_{1-};\mathbf{J}_{+}|W_a,W_b,Q_a,Q_b,\mathbf{J}_{-},X_{1+}) - I(X_{1-};\mathbf{Y}_{2+}|W_a,W_b,Q_a,Q_b,\mathbf{J}_{-},X_{1+})\\
&\quad + I(X_{1+};\mathbf{J}_{-}|W_a,W_b,Q_a,Q_b,\mathbf{Y}_{2+},X_{1-}) - I(X_{1+};\mathbf{Y}_{2-}|W_a,W_b,Q_a,Q_b,\mathbf{Y}_{2+},X_{1-}).
\end{aligned}
$$

Now, observe that

$$
\begin{aligned}
I(X_{1-};\mathbf{J}_{+}|W_a,W_b,Q_a,Q_b,\mathbf{J}_{-},X_{1+}) &\leq I(X_{1-};\mathbf{J}_{+},\hat{\mathbf{J}}_{+}|W_a,W_b,Q_a,Q_b,\mathbf{J}_{-},X_{1+})\\
&\overset{(c)}{=} I(X_{1-};\hat{\mathbf{J}}_{+}|W_a,W_b,Q_a,Q_b,\mathbf{J}_{-},X_{1+})\\
&\overset{(d)}{\leq} I(X_{1-};\mathbf{Y}_{2+}|W_a,W_b,Q_a,Q_b,\mathbf{J}_{-},X_{1+})
\end{aligned}
$$

where (c) used $\mathbf{J}_{+} \to (\hat{\mathbf{J}}_{+},X_{1+}) \to (W_a,W_b,Q_a,Q_b,\mathbf{J}_{-},X_{1-})$ is Markov and (d) uses that $\hat{\mathbf{J}}_{+}$ is a stochastically degraded version of $\mathbf{Y}_{2+}$. Similarly, one obtains that

$$
I(X_{1+};\mathbf{J}_{-}|W_a,W_b,Q_a,Q_b,\mathbf{Y}_{2+},X_{1-}) \leq I(X_{1+};\mathbf{Y}_{2-}|W_a,W_b,Q_a,Q_b,\mathbf{Y}_{2+},X_{1-}),
$$

implying

$$
\begin{aligned}
&I(X_{1+};\mathbf{J}_{+}|W_a,W_b,Q_a,Q_b,\mathbf{J}_{-}) - I(X_{1+};\mathbf{Y}_{2+}|W_a,W_b,Q_a,Q_b,\mathbf{J}_{-})\\
&\quad + I(X_{1-};\mathbf{J}_{-}|W_a,W_b,Q_a,Q_b,\mathbf{Y}_{2+}) - I(X_{1-};\mathbf{Y}_{2-}|W_a,W_b,Q_a,Q_b,\mathbf{Y}_{2+}) \geq 0.
\end{aligned}
$$

Now set $\tilde{Q}$ be be uniform Bernoulli variable and when $\tilde{Q}=0$ we set $Q=(Q_a,Q_b), W=(W_a,W_b,\mathbf{J}_{-}), (X_1,X_2)=(X_{1+},X_{2+})$ and when $\tilde{Q}=1$ we set $Q=(Q_a,Q_b), W=(W_a,W_b,\mathbf{Y}_{2+}), (X_1,X_2)=(X_{1-},X_{2-})$. Notice that this distribution is a candidate maximizer of the expression and hence must induce a value of at most V. Therefore from (b) and (42) above we obtain that

$$
\begin{aligned}
2V \leq 2V &- \lambda(1-\alpha)I(X_{2-};\mathbf{J}_{+}|W_a,W_b,X_{2+},\mathbf{Y}_{2+},\mathbf{J}_{-},Q_a,Q_b) - \lambda(1-\alpha)I(X_{2+};\mathbf{J}_{-}|W_a,W_b,X_{2-},\mathbf{Y}_{2-},Q_a,Q_b)\\
&- \beta I(\mathbf{Y}_{1+};\mathbf{Y}_{1-}|Q_a,Q_b) - (1-\beta)I(\mathbf{Y}_{2+};\mathbf{Y}_{2-}|Q_a,Q_b) - (\gamma-1)\epsilon I(\mathbf{Y}_{2+};\mathbf{J}_{-}|W_a,W_b,Q_a,Q_b).
\end{aligned}
$$

This implies that the distribution generated above is a maximizer as well, but more importantly that the blue-colored term, $I(\mathbf{Y}_{2+};\mathbf{J}_{-}|W_a,W_b,Q_a,Q_b)=0$, implying that $\begin{bmatrix} X_{1+} & X_{2+}\end{bmatrix}$ is independent of $\begin{bmatrix} X_{1-} & X_{2-}\end{bmatrix}$, which further yields, by the Skitovic-Darmois characterization of Gaussians, that conditioned on W,Q, X_1,X_2 are jointly Gaussians and that they have the same covariance matrix (independent of W,Q). Since the arguments mimic Propositions 2, 8, and Corollary 3 of [GN14] we omit the details. Similarly from $I(\mathbf{Y}_{1+};\mathbf{Y}_{1-}|Q_a,Q_b)=0$

and $I(\mathbf{Y}_{2+};\mathbf{Y}_{2-}|Q_a,Q_b)=0$ we have that X_1 is a Gaussian with variance that does not depend on Q and so is X_2.

By monotonicity of the terms in the outer bound it is immediate that variance of X_1 is P_1, the variance of X_2 is P_2, and the covariance of $X_1,X_2|W$ is given by some

$$\begin{bmatrix} K_1 & \rho\sqrt{K_1K_2} \\ \rho\sqrt{K_1K_2} & K_2 \end{bmatrix} \preceq \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix}.$$

Now we substitute this distribution and obtain the bound in the weighted-sum rate. □

C.2. **Proof of Theorem 4: Slope at Costa's corner point.** Let $\lambda_1=\lambda\alpha$ and $\lambda_2=\lambda(1-\alpha)$. Setting $\beta=0$ and ignoring the constraint in (14), we obtain that

$$\begin{aligned} R_1+(\lambda_1+\lambda_2)R_2 \le{}& \frac{\lambda_1}{2}\log(K_2(1-\rho^2)+1)+\frac{1}{2}\log\left(\frac{1+a^2P_1+P_2}{u^2}\right) \\ &+\frac{\lambda_2-1}{2}\log\left(\frac{1+a^2K_1+K_2+2a\rho\sqrt{K_1K_2}}{K_1+u^2}\right)+\frac{\lambda_2}{2}\log\left(\frac{K_1(1-\rho^2)+u^2}{a^2K_1(1-\rho^2)+1}\right), \end{aligned}$$

for some $K_1\le P_1$, $K_2\le P_2$ and $\rho\in[-1,1]$ satisfying

$$(P_1-K_1)(P_2-K_2)\ge\rho^2K_1K_2. \tag{43}$$

Note that the expression is increasing in K_2; and hence we fix K_1,ρ and substitute for the maximal K_2 satisfying (43) to obtain

$$\begin{aligned} &\lambda_1\log\left(\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}(1-\rho^2)+1\right)+\lambda_2\log\left(\frac{K_1(1-\rho^2)+1}{a^2K_1(1-\rho^2)+1}\right) \\ &\quad+(\lambda_2-1)\log\left(\frac{1+a^2K_1+\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}+2a\rho\sqrt{K_1\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}}}{K_1+1}\right) \\ &\le(\lambda_1+\lambda_2-1)\log(1+P_2). \end{aligned}$$

Rearrangement of terms yields

$$\begin{aligned} &\lambda_1\log\left(\frac{\frac{P_2(P_1-K_1)(1-\rho^2)}{P_1-K_1+\rho^2K_1}+1}{1+P_2}\right)+\log\left(\frac{K_1(1-\rho^2)+1}{a^2K_1(1-\rho^2)+1}\right) \\ &\quad+(\lambda_2-1)\log\left(\frac{(1+a^2K_1+\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}+2a\rho\sqrt{K_1\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}})(1+K_1(1-\rho^2))}{(K_1+1)(a^2K_1(1-\rho^2)+1)(1+P_2)}\right) \\ &\le 0. \end{aligned}$$

From the concavity of log it suffices that

$$\begin{aligned} &\lambda_1\left(\frac{\frac{P_2(P_1-K_1)(1-\rho^2)}{P_1-K_1+\rho^2K_1}+1}{1+P_2}\right)+\left(\frac{K_1(1-\rho^2)+1}{a^2K_1(1-\rho^2)+1}\right) \\ &\quad+(\lambda_2-1)\left(\frac{(1+a^2K_1+\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}+2a\rho\sqrt{K_1\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}})(1+K_1(1-\rho^2))}{(K_1+1)(a^2K_1(1-\rho^2)+1)(1+P_2)}\right) \\ &\le\lambda_1+\lambda_2. \end{aligned}$$

This is equivalent to

$$-\lambda_1\left(\frac{P_2P_1\rho^2}{(P_1-K_1+\rho^2K_1)(1+P_2)}\right)+\left(\frac{K_1(1-a^2)(1-\rho^2)}{a^2K_1(1-\rho^2)+1}\right)$$

$$+(\lambda_2-1)\left(\frac{(1+a^2K_1+\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}+2a\rho\sqrt{K_1\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}})(1+K_1(1-\rho^2))}{(K_1+1)(a^2K_1(1-\rho^2)+1)(1+P_2)}-1\right)$$
$$\leq 0.$$

Since

$$2a\rho\sqrt{K_1\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}}\leq\frac{\lambda_1}{\lambda_2-1}\left(\frac{P_2P_1\rho^2}{(P_1-K_1+\rho^2K_1)}\right)\frac{(K_1+1)(a^2K_1(1-\rho^2)+1)}{(1+K_1(1-\rho^2))}+\frac{\lambda_2-1}{\lambda_1}\frac{(1+K_1(1-\rho^2))}{(K_1+1)(a^2K_1(1-\rho^2)+1)}\frac{a^2K_1((P_1-K_1))}{P_1}$$

It suffices that

$$\left(\frac{K_1(1-a^2)(1-\rho^2)}{a^2K_1(1-\rho^2)+1}\right)+(\lambda_2-1)\left(\frac{(1+a^2K_1+\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}+\frac{\lambda_2-1}{\lambda_1}\frac{(1+K_1(1-\rho^2))}{(K_1+1)(a^2K_1(1-\rho^2)+1)}\frac{a^2K_1((P_1-K_1))}{P_1})(1+K_1(1-\rho^2))}{(K_1+1)(a^2K_1(1-\rho^2)+1)(1+P_2)}-1\right)\leq 0.$$

This is equivalent to

$$\begin{aligned}&\big(K_1(1-a^2)(1-\rho^2)\big)\\&\leq(\lambda_2-1)\left(a^2K_1(1-\rho^2)+1-\frac{(1+a^2K_1+\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}+\frac{\lambda_2-1}{\lambda_1}\frac{(1+K_1(1-\rho^2))}{(K_1+1)(a^2K_1(1-\rho^2)+1)}\frac{a^2K_1((P_1-K_1))}{P_1})(1+K_1(1-\rho^2))}{(K_1+1)(1+P_2)}\right).\end{aligned}\tag{44}$$

This can be rewritten as

$$\big((1-a^2)(1-\rho^2)\big)\leq(\lambda_2-1)\left(\frac{a^2P_2(1-\rho^2)(1+K_1)+\rho^2(1-a^2)}{(1+K_1)(1+P_2)}\right)+\frac{(\lambda_2-1)}{(1+K_1)(1+P_2)}\left(\left(\frac{\rho^2P_2(1+P_1))}{P_1-K_1+\rho^2K_1}\right)-\frac{\lambda_2-1}{\lambda_1}\left(\frac{(1+K_1(1-\rho^2))^2}{(K_1+1)(a^2K_1(1-\rho^2)+1)}\frac{a^2((P_1-K_1))}{P_1}\right)\right).$$

Note that

$$\frac{(1+K_1(1-\rho^2))^2}{(a^2K_1(1-\rho^2)+1)}\leq\left(\frac{K_1}{a^2}(1-\rho^2)+1\right).$$

Therefore it suffices that

$$\big((1-a^2)(1-\rho^2)\big)\leq(\lambda_2-1)\left(\frac{a^2P_2(1-\rho^2)(1+K_1)+\rho^2(1-a^2)}{(1+K_1)(1+P_2)}\right)+\frac{(\lambda_2-1)}{(1+K_1)(1+P_2)}\left(\left(\frac{\rho^2P_2(1+P_1))}{P_1}\right)-\frac{\lambda_2-1}{\lambda_1}\left(\frac{(1+\frac{K_1}{a^2}(1-\rho^2))}{(K_1+1)}\frac{a^2((P_1-K_1))}{P_1}\right)\right).$$

Since the expression in linear in ρ^2, we just need to ensure that this holds for $\rho^2=0$ and $\rho^2=1$. At $\rho^2=0$ we require

$$\big((1-a^2)\big)\leq(\lambda_2-1)\left(\frac{a^2P_2}{(1+P_2)}\right)-\frac{(\lambda_2-1)}{(1+K_1)(1+P_2)}\left(\frac{\lambda_2-1}{\lambda_1}\left(\frac{(a^2+K_1)}{(K_1+1)}\frac{(P_1-K_1)}{P_1}\right)\right)$$

Optimizing over $K_1:0\leq K_1\leq P_1$, it suffices that

$$\big((1-a^2)\big)\leq(\lambda_2-1)\left(\frac{a^2P_2}{(1+P_2)}\right)$$

$$- \frac{(\lambda_2-1)^2}{\lambda_1 P_1(1+P_2)} \times \begin{cases} \frac{(P_1+a^2)^2}{4(1+P_1)(1-a^2)} & P_1 > \frac{a^2}{1-2a^2}, a^2 < \frac{1}{2} \\ a^2 P_1 & o.w. \end{cases}.$$

Or one can even optimize over P_1 to get a bound, that works for all P_1,

$$1-a^2 \leq (\lambda_2-1)\left(\frac{a^2P_2}{(1+P_2)}\right) - \frac{(\lambda_2-1)^2}{\lambda_1(1+P_2)} \times \begin{cases} \frac{1}{4(1-a^2)} & a^2 < \frac{1}{2} \\ a^2 & o.w. \end{cases}.$$

Therefore we clearly need

$$\frac{\lambda_2-1}{\lambda_1} < \begin{cases} 4a^2(1-a^2)P_2 & a^2 < \frac{1}{2} \\ P_2 & o.w. \end{cases}.$$

At $\rho^2 = 1$, we require

$$0 \leq (\lambda_2-1)\left(\frac{(1-a^2)}{(1+K_1)(1+P_2)}\right) + \frac{(\lambda_2-1)}{(1+K_1)(1+P_2)}\left(\left(\frac{P_2(1+P_1))}{P_1}\right) - \frac{\lambda_2-1}{\lambda_1}\left(\frac{1}{(K_1+1)}\frac{a^2((P_1-K_1))}{P_1}\right)\right)$$

Or equivalently

$$0 \leq (1-a^2) + \left(\left(\frac{P_2(1+P_1))}{P_1}\right) - \frac{\lambda_2-1}{\lambda_1}\left(\frac{1}{(K_1+1)}\frac{a^2((P_1-K_1))}{P_1}\right)\right).$$

Optimizing over K_1 it suffices that

$$0 \leq (1-a^2) + \left(\left(\frac{P_2(1+P_1))}{P_1}\right) - \frac{\lambda_2-1}{\lambda_1}a^2\right).$$

Therefore, we need

$$\frac{\lambda_2-1}{\lambda_1} \leq \frac{\frac{P_2(1+P_1)}{P_1} + (1-a^2)}{a^2}.$$

Thus the constraint (from $\rho^2 = 0$) is the active one. We are now left with computing the minimum $\lambda = \lambda_1 + \lambda_2$ satisfying the above constraints.

Case (i): $a^2 < \frac{1}{2}$

We seek to minimize $\lambda_1 + \lambda_2$ subject to

$$1-a^2 \leq (\lambda_2-1)\left(\frac{a^2P_2}{1+P_2}\right) - \frac{(\lambda_2-1)^2}{\lambda_1(1+P_2)} \times \frac{1}{4(1-a^2)}.$$

Let $\frac{\lambda_1}{\lambda_2-1} := \gamma$. We seek to minimize

$$\frac{(1-a^2)(1+\gamma)}{\frac{a^2P_2}{1+P_2} - \frac{1}{4\gamma(1+P_2)(1-a^2)}} = \frac{(1+P_2)(1-a^2)}{a^2P_2}\frac{\gamma(1+\gamma)}{\gamma - \frac{1}{4a^2(1-a^2)P_2}}$$

The minimum value is

$$\frac{(1+P_2)(1-a^2)}{a^2P_2}\frac{\left(1+\sqrt{1+4a^2(1-a^2)P_2}\right)^2}{4a^2(1-a^2)P_2}$$

obtained when $\gamma = \frac{\left(1+\sqrt{1+4a^2(1-a^2)P_2}\right)}{4a^2(1-a^2)P_2}$.

Case (ii): $a^2 \geq \frac{1}{2}$

We seek to minimize $\lambda_1 + \lambda_2$ subject to

$$1-a^2 \leq (\lambda_2-1)\left(\frac{a^2P_2}{1+P_2}\right) - \frac{(\lambda_2-1)^2a^2}{\lambda_1(1+P_2)}.$$

As before, let $\frac{\lambda_1}{\lambda_2-1} := \gamma$. We seek to minimize

$$\frac{(1-a^2)(1+\gamma)}{\frac{a^2P_2}{1+P_2} - \frac{a^2}{\gamma(1+P_2)}} = \frac{(1+P_2)(1-a^2)}{a^2P_2}\frac{\gamma(1+\gamma)}{\gamma - \frac{1}{P_2}}$$

The minimum value is

$$\frac{(1+P_2)(1-a^2)}{a^2P_2}\frac{\left(1+\sqrt{1+P_2}\right)^2}{P_2}$$

obtained when $\gamma = \frac{\left(1+\sqrt{1+P_2}\right)}{P_2}$.

Appendix D. Proofs of Results for the Broadcast Channel

D.1. **Proof of Lemma 4.** We use the proof by contradiction. Observe that an erasure channel $p(\hat{y}|y)$ satisfies the assumptions of Corollary 3. From Corollary 3, there exists $p(v,w,x)$ such that

$$\frac{1}{2}(1-\epsilon) \leq I(V;Z|W) - I(V;\hat{Y}|W) \tag{45}$$

$$1-\epsilon = I(\hat{Y};Y) \leq I(W;Z) + I(X;\hat{Y}|W) = I(W;Z) + H(\hat{Y}|W). \tag{46}$$

Note that $I(\hat{Y};Y) = (1-\epsilon)H(\hat{Y})$, thus $1-\epsilon = I(\hat{Y};Y)$ implies that $\hat{Y}$ is uniform. Next note that

$$\begin{aligned}
\frac{1-\epsilon}{2} &\leq I(V;Z|W) - I(V;\hat{Y}|W) \\
&= (1-\epsilon)(I(V;\hat{Z}|W) - I(V;\hat{Y}|W)) - \epsilon I(V;\hat{Y}|W) \\
&= (1-\epsilon)(I(V;\hat{Z}|\hat{Y}W) - I(V;\hat{Y}|\hat{Z},W)) - \epsilon I(V;\hat{Y}|W) \\
&= (1-\epsilon)(H(\hat{Z}|\hat{Y}) - I(W;\hat{Z}|\hat{Y}) - H(\hat{Z}|V,W,\hat{Y}) - I(V;\hat{Y}|\hat{Z},W)) - \epsilon I(V;\hat{Y}|W) \\
&\leq (1-\epsilon)\,\mathrm{P}(\hat{Y}=0)H(\hat{Z}|\hat{Y}=0)
\end{aligned}$$

Since $\hat{Y}$ is uniform, we have $\frac{1}{2} \leq H(\hat{Z}|\hat{Y}) = \frac{1}{2}H(\hat{Z}|\hat{Y}=0) \leq \frac{1}{2}$, implying that $P(X=0) = P(X=1) = \frac{1}{4}$. The above sequence also implies that the following four terms $I(W;\hat{Z}|\hat{Y}), H(\hat{Z}|V,W,\hat{Y}), I(V;\hat{Y}|W), I(V;\hat{Y}|\hat{Z},W)$ are zero. Let $\mathcal{W}_0 := \{w : \mathrm{P}(\hat{Y}=1|W=w) \neq 0\}$. For all $w \in \mathcal{W}_0$, and for any v such that $\mathrm{P}(V=v|W=w) > 0$ observe that

$$\begin{aligned}
p(v|w)p(\hat{y}=1|w) &= p(v|w)p(\hat{y}=1|vw)p(\hat{z}=0|vw,\hat{y}=1) && \because I(V;\hat{Y}|W)=0, p(\hat{z}=0|vw,\hat{y}=1)=1 \\
&= p(v|w)p(\hat{z}=0|vw)p(\hat{y}=1|vw,\hat{z}=0) \\
&= p(v|w)p(\hat{z}=0|vw)p(\hat{y}=1|w,\hat{z}=0) && \because I(V;\hat{Y}|\hat{Z},W)=0.
\end{aligned}$$

Canceling $p(v|w)$ we see that $p(\hat{z}=0|vw)$ does not depend on v, implying that $I(V;\hat{Z}|W=w)=0$. Since we have $I(V;\hat{Z}|W) = \frac{1}{2}$, this implies that $\sum_{w\in\mathcal{W}_0}\mathrm{P}(W=w) \leq \frac{1}{2}$. Hence

$$\frac{1}{2} = \sum_w \mathrm{P}(W=w,\hat{Y}=1) = \sum_{w\in\mathcal{W}_0}\mathrm{P}(W=w,\hat{Y}=1) \leq \sum_{w\in\mathcal{W}_0}\mathrm{P}(W=w) \leq \frac{1}{2}.$$

The above implies that $\sum_{w\in\mathcal{W}_0}\mathrm{P}(W=w) = \frac{1}{2}$ and that $w \in \mathcal{W}_0$ implies $\hat{Y}=1$. On the other hand, by definition $w \notin \mathcal{W}_0$ implies $\hat{Y}=0$, or that $\hat{Y}$ is a function of W. Now, applying this to (46), we obtain that $1-\epsilon \leq I(W;Z) = (1-\epsilon)I(W;\hat{Z}) \leq (1-\epsilon)H(\hat{Z}) = (1-\epsilon)H_2(\frac{1}{4})$, a contradiction.

Appendix E. Routine Calculations: For the Referees verification purposes

E.1. **Steps in Proof of Proposition 2.** .

E.1.1. *Argument 1.* We first show the algebra manipulations for the first inequality in Appendix B.2.

$$
\begin{aligned}
&\mathsf{C}\left(x_*^2 + \left(x_*\left(\sqrt{S_{31}} + \sqrt{\rho^2 S_{32}}\right) + \sqrt{(1-\rho^2)S_{32}}\right)^2\right) - \mathsf{C}\left(x_*^2(1+S_J)\right) \\
&\quad \le \frac{1}{2}\log\left(\frac{1+S_{31}+\rho^2 S_{32} + 2\sqrt{\rho^2 S_{31}S_{32}}}{1+S_J} + \frac{S_{31}S_{32}(1-\rho^2)}{S_{31}-S_J}\right), && \Longleftrightarrow \\
&1 + x_*^2 + \left(x_*\left(\sqrt{S_{31}} + \sqrt{\rho^2 S_{32}}\right) + \sqrt{(1-\rho^2)S_{32}}\right)^2 \\
&\quad \le \left(\frac{1+S_{31}+\rho^2 S_{32} + 2\sqrt{\rho^2 S_{31}S_{32}}}{1+S_J} + \frac{S_{31}S_{32}(1-\rho^2)}{S_{31}-S_J}\right)(1+x_*^2(1+S_J)), && \Longleftrightarrow \\
&2x_*\left(\sqrt{S_{31}} + \sqrt{\rho^2 S_{32}}\right)\sqrt{(1-\rho^2)S_{32}} \\
&\quad \le \frac{S_{31}-S_J+\rho^2 S_{32} + 2\sqrt{\rho^2 S_{31}S_{32}}}{1+S_J} + \frac{S_J S_{32}(1-\rho^2)}{S_{31}-S_J} + x_*^2(1+S_J)\frac{S_{31}S_{32}(1-\rho^2)}{S_{31}-S_J}.
\end{aligned}
$$

Completing the square on x_*, it suffices to show that

$$
\begin{aligned}
&\frac{(S_{31}-S_J)\left(\sqrt{S_{31}} + \sqrt{\rho^2 S_{32}}\right)^2}{S_{31}(1+S_J)} \le \frac{S_{31}-S_J+\rho^2 S_{32} + 2\sqrt{\rho^2 S_{31}S_{32}}}{1+S_J} + \frac{S_J S_{32}(1-\rho^2)}{S_{31}-S_J} && \Longleftrightarrow \\
&\frac{S_J}{1+S_J} \le \frac{S_J}{1+S_J}\frac{\left(\sqrt{S_{31}} + \sqrt{\rho^2 S_{32}}\right)^2}{S_{31}} + \frac{S_J S_{32}(1-\rho^2)}{S_{31}-S_J},
\end{aligned}
$$

which is immediate.

E.1.2. *Argument 2, step (a).* Note that for $\rho^2 \le 1$ and $S_{21} \le S_J < S_{31}$

$$
\begin{aligned}
&\frac{1}{2}\log(1+S_J(1-\rho^2)) + \frac{1}{2}\log\left(\frac{1+S_{31}+\rho^2 S_{32} + 2\sqrt{\rho^2 S_{31}S_{32}}}{1+S_J} + \frac{S_{31}S_{32}(1-\rho^2)}{S_{31}-S_J}\right) \\
&\quad \overset{(a)}{\le} \frac{1}{2}\log\left(1+S_{31}+S_{32}\left(\frac{(1+S_J)^2 S_{31}^2}{S_J(1+S_{31})(S_{31}-S_J)}\right)\right) && \Longleftrightarrow \\
&1+S_{31}+\rho^2 S_{32} + 2\sqrt{\rho^2 S_{31}S_{32}} + \frac{S_{31}S_{32}(1-\rho^2)}{S_{31}-S_J}(1+S_J(1-\rho^2)) \\
&\quad \le 1+S_{31}+S_{32}\left(\frac{(1+S_J)^2 S_{31}^2}{S_J(1+S_{31})(S_{31}-S_J)}\right) + \rho^2 S_J\left(\frac{1+S_{31}+\rho^2 S_{32} + 2\sqrt{\rho^2 S_{31}S_{32}}}{1+S_J}\right).
\end{aligned}
$$

Since $\frac{(1+S_J)S_{31}}{S_{31}-S_J} + \frac{S_{31}(1+S_J)}{S_J(1+S_{31})} = \frac{(1+S_J)^2 S_{31}^2}{S_J(1+S_{31})(S_{31}-S_J)}$, the last inequality is equivalent to

$$
\begin{aligned}
&\rho^2 S_{32} + 2\sqrt{\rho^2 S_{31}S_{32}} + \frac{S_{31}S_{32}(1-\rho^2)}{S_{31}-S_J}(1+S_J(1-\rho^2)) \\
&\quad \le S_{32}\left(\frac{(1+S_J)S_{31}}{S_{31}-S_J} + \frac{S_{31}(1+S_J)}{S_J(1+S_{31})}\right) + \rho^2 S_J\left(\frac{1+S_{31}+\rho^2 S_{32} + 2\sqrt{\rho^2 S_{31}S_{32}}}{1+S_J}\right).
\end{aligned}
$$

The last inequality follows from the two immediate ones below

$$
\begin{aligned}
\rho^2 S_{32} + \frac{S_{31}S_{32}(1-\rho^2)}{S_{31}-S_J}(1+S_J(1-\rho^2)) &\le S_{32}\frac{(1+S_J)S_{31}}{S_{31}-S_J} \\
2\sqrt{\rho^2 S_{31}S_{32}} &\le \frac{S_{32}S_{31}(1+S_J)}{S_J(1+S_{31})} + \rho^2 S_J\left(\frac{1+S_{31}}{1+S_J}\right).
\end{aligned}
$$

E.2. **Steps in the proof of Theorem 4.** To see (29c) observe that

$$
\begin{aligned}
& I(M_0; Z^n|\hat{J}^n) + I(M_0, J^n; \hat{J}^n) + I(M_1; J^n|M_0, \hat{J}^n) + I(M_1; Y^n|M_0, J^n) \\
& = I(M_0; Z^n, \hat{J}^n) + I(M_1; Y^n, J^n|M_0) + I(\hat{J}^n; J^n|M_1, M_0) \\
& \geq n(R_0 + R_1) - ng(\epsilon_n)
\end{aligned}
$$

To see (29f) observe that

$$
\begin{aligned}
& \min\{I(M_0, \hat{J}^n; J^n) + I(M_0; Y^n|J^n), I(M_0; Z^n|\hat{J}^n) + I(M_0, J^n; \hat{J}^n)\} \\
& \quad + I(M_1; Y^n|M_0, J^n) + I(M_2; \hat{J}^n|M_1, M_0, J^n) \\
& \quad + \min\Big\{ I(M_1; J^n|M_0, \hat{J}^n) + I(M_2; Z^n|M_1, M_0, \hat{J}^n), \\
& \qquad I(M_2; Z^n|M_0, \hat{J}^n) + I(M_1; J^n|M_2, M_0, \hat{J}^n) \Big\} \\
& = \min\{I(M_0; Y^n, J^n), I(M_0; Z^n, \hat{J}^n)\} \\
& \quad + \min\Big\{ I(M_1; Y^n, J^n|M_0) + I(M_2; Z^n, \hat{J}^n|M_1, M_0) + I(J^n; \hat{J}^n|M_0, M_1, M_2), \\
& \qquad I(M_2; Z^n, \hat{J}^n|M_0) + I(M_1; Y^n, J^n|M_0) + I(J^n; \hat{J}^n|M_0, M_1, M_2) + I(M_1; M_2|M_0, J^n, \hat{J}^n) \Big\} \\
& \geq n(R_0 + R_1 + R_2) - ng(\epsilon_n)
\end{aligned}
$$