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. 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 for achievable rate regions in multiuser information theory settings are known in the literature. The generic and classical approach is based on the identification of the auxiliary random variables as the past and/or future of the underlying random variables to write single-letter converse (outer) 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 the Wasserstein metric to develop an outer bound for the Gaussian Z-interference channel is an example of a non-classical approach. For a class of relay channels called the primitive relay channel, converse bounds based on the blowing-up lemma, concentration of Gaussian measure, or reverse hypercontractivity are known [Zha88,Xue14,WBO19,LO19,WO18], 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 has been a useful tool in the information-theorists' toolbox. 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 receivers generalize genie-aided outer bounds since the auxiliary receiver's message may not be available 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, imaginary channel $\hat{V}_2$ in [LG09, Eq. 16] and the remote source and channels in [YLL18, Eq. 12, 13] for a joint source-channel coding over a broadcast channel.

The basic idea of our outer bounds is to expand the space of possibilities of (Gallager-type) identification of auxiliary variables in outer bounds: we consider one or multiple "auxiliary" receivers and use their past and/or future when identifying the auxiliary random variables. These auxiliary random variables are then used to derive new constraints on the achievable rates. For instance, one can use an existing upper bound to obtain a bound on the flow of information towards the introduced auxiliary receiver. Separately, one can bound the difference of the flow of information towards the legitimate receiver and the auxiliary receiver using the newly identified

This paper was presented in part at the IEEE International Symposium on Information Theory, 2020 [GN20].

auxiliary random variables. Adding up these two bounds yields an upper bound on the flow of information to the legitimate receiver. The idea of adding and subtracting the flow of information to an auxiliary receiver can be noted in our upper bounds on the relay and interference channels.

Another important benefit of introducing the auxiliary receivers is the possibility of inducing certain Markov chains and constraints on the auxiliary random variables. For instance, our upper bound for the relay channel involves an auxiliary random variable W that satisfies a certain Markov structure. This Markov chain follows from the restriction on the auxiliary receiver J, in that it is allowed to depend only on the transmitter's input signal and not on the relay's input signal (for another example, see the Markov chain structure induced by the choice of the auxiliary receiver in the follow-up work [EGGN21a, Theorem 7]).

The original motivation for the authors' introduction of auxiliary receivers came from the following observation in the context of broadcast channels: Suppose one erases the output of every receiver with probability ϵ, then the traditional single-letter outer bounds scale by $(1-\epsilon)$; however the achievable region does not (see Section 4.1). This motivated the authors, thanks also to a question asked by Young-Han Kim, to investigate whether the true capacity region also scaled by $(1-\epsilon)$. It was here that the auxiliary receiver idea originated as a tool to show that the true capacity region did not have the $(1-\epsilon)$ scaling property, as the outputs of the auxiliary channels need not undergo any erasure. Our outer bound in Theorem 7 addresses the above question. In this bound, new auxiliary random variables (defined using past and/or future of the auxiliary receiver symbols) are used to minimize the discarded terms in the various routine manipulations. In particular, the UV outer bound (a previously known outer bound on the capacity of a general broadcast channel) and the terms that are discarded in its derivation are considered. The bound is modified to minimize the discarded terms using the new auxiliary random variables. This then led to a (strict) tightening of the rate constraints, whose strictness is then demonstrated using a concrete example.

We also give a second outer bound for the general broadcast channel with two auxiliary receivers. Here, instead of just considering the information flow of the messages to one auxiliary receiver, we also modify the content of the message itself using another auxiliary receiver (see Remark 16).

One major difference between the approach in this paper and most of the earlier papers by the authors (and perhaps others) is that the outer bounds were not "guided" by achievable regions; that is, here we were not trying to develop a matching converse to an achievable region or even attempting to come close to one. The fact that a small change of perspective of the standard techniques do give us these improvements suggests that there is an entire unexplored landscape motivated by similar observations. It is also worthwhile to note that the new upper bounds developed here and in the follow-up works in [EGGN21a, GNN21] can recover and improve upon results obtained using (novel to the field) geometric techniques [PW15, WBO19, LO19, WO18].

Remark 1. There are some instances where there is, in the words of the wise reviewer, "a mathematical miasma" induced by routine and unending manipulations of information measures. In such instances, a reader may alternately use an automated software (https://github.com/cheuktingli/psitip, see [Li21]) to check the veracity of the manipulations.

Remark 2. The authors believe that there is lots of room to explore with respect to the auxiliary receiver approach for developing outer bounds even in the three basic settings considered in this paper. The main goal of this paper is to illustrate that this approach is fruitful and leads to strict improvements in relay, interference, and broadcast channels.

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. In Section 4 we give our outer bounds for the broadcast channel.

Notation: We adopt most of our notation from [EK12]. The set $\{1, 2, \cdots, n\}$ is denoted by $[n]$. We use Y^i to denote the sequence $(Y_1, Y_2, \cdots, Y_i)$, and Y_i^j to denote $(Y_i, Y_{i+1}, \cdots, Y_j)$. We also use $Y^{n\setminus i}$ to denote (Y^{i-1}, Y_{i+1}^n). For discrete settings, logarithms are in base two, and for continuous channels, the logarithms are in base e. Conditional distributions representing channels are denoted by $T(\cdot|\cdot)$. Given two random variables X and Y, we use $X \perp Y$ to denote X being independent of Y. We say that $X \multimap Y \multimap Z$ forms a Markov chain if $I(X; Z|Y) = 0$. For square matrices A and B of the same size, we write $A \preceq B$ if $B - A$ is a positive semi-definite matrix.

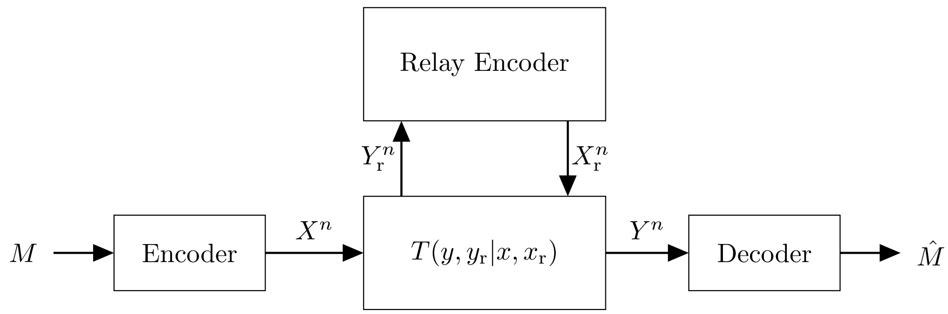

FIGURE 1. Transmission of a message M over a memoryless relay channel with n uses of the channel.

2. Relay channel

A relay channel models the transmission of a message from a sender to a receiver in the presence of a helper relay node. A relay channel is described by a conditional distribution $T(y, y_{\mathrm{r}}|x, x_{\mathrm{r}})$ where $x \in \mathcal{X}$ is the transmitter's input symbol, $x_{\mathrm{r}} \in \mathcal{X}_{\mathrm{r}}$ is the input to the channel by the relay, $y \in \mathcal{Y}$ is the output symbol at the receiver and $y_{\mathrm{r}} \in \mathcal{Y}_{\mathrm{r}}$ is the output symbol at the relay. An (n, R) code for a memoryless relay channel $T(y, y_{\mathrm{r}}|x, x_{\mathrm{r}})$, depicted in Fig. 1, consists of an encoder $\mathcal{E}$ that maps a message M (uniform over $[2^{nR}]$) to an input sequence $X^n \in \mathcal{X}^n$, *i.e.*, $X^n = \mathcal{E}(M)$, a relay encoder $\mathcal{E}_{\mathrm{r}i}$ that assigns a symbol $X_{\mathrm{r}i}$ to each past received sequence Y_{r}^{i-1} for $i \in [n]$, *i.e.*, $X_{\mathrm{r}i} = \mathcal{E}_{\mathrm{r}i}(Y_{\mathrm{r}}^{i-1})$, and a decoder $\mathcal{D}$ that produces an estimate $\hat{M}$ from Y^n. The following joint distribution is induced by the code over a memoryless relay channel $T(y, y_{\mathrm{r}}|x, x_{\mathrm{r}})$:

$$p(m)p(x^n|m)\left(\prod_{i=1}^{n} p(x_{\mathrm{r}i}|y_{\mathrm{r}}^{i-1}) \prod_{i=1}^{n} T(y_i, y_{\mathrm{r}i}|x_i, x_{\mathrm{r}i})\right) p(\hat{m}|y^n)$$

meaning that the current output symbols $(Y_i, Y_{\mathrm{r}i})$ are conditionally independent of $(M, X^{i-1}, X_{\mathrm{r}}^{i-1}, Y_{\mathrm{r}}^{i-1}, Y^{i-1})$ given $(X_i, X_{\mathrm{r}i})$. The error probability of the code is $\mathrm{P}[M \neq \hat{M}]$. 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 for the relay channel $T(y, y_{\mathrm{r}}|x, x_{\mathrm{r}})$.

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, CM11, BSC14]). Outer bounds are discussed in [TU08, ARY09, Xue14, WOX17, WBO19, LO19, WO18, EGGN21b] for the special class of relay channels with orthogonal receiver components (where $Y = (Y_1, Y_2)$ has two components and $p(y, y_{\mathrm{r}}|x, x_{\mathrm{r}}) = p(y_1, y_{\mathrm{r}}|x)p(y_2|x_{\mathrm{r}})$.) However these bounds are not applicable to a general relay channel and the best known upper bound for a general relay channel is the cutset bound [CG79]. For more details on relay channels and a collection of known results please refer to [EK12, Chapter 16]. Interested readers can also look into [EGGN21a] where we do explore the relay settings, using ideas developed in this paper, in much greater detail.

Definition 1. An auxiliary receiver is described by a conditional distribution $T_{J|Y,Y_{\mathrm{r}},X,X_{\mathrm{r}}}$ defined on arbitrary alphabet set $\mathcal{J}$. Let $\mathcal{F}$ be the class of $T_{J|Y,Y_{\mathrm{r}},X,X_{\mathrm{r}}}$ such that

$$T_{J,Y,Y_{\mathrm{r}}|X,X_{\mathrm{r}}} = T_{J|Y,Y_{\mathrm{r}},X,X_{\mathrm{r}}} T_{Y,Y_{\mathrm{r}}|X,X_{\mathrm{r}}}$$

also factorizes as

$$T_{J,Y,Y_{\mathrm{r}}|X,X_{\mathrm{r}}} = T_{J|X} T_{Y,Y_{\mathrm{r}}|J,X,X_{\mathrm{r}}}.$$

Equivalently, $\mathcal{F}$ is the class of $T_{J|Y,Y_{\mathrm{r}},X,X_{\mathrm{r}}}$, a conditional distribution defined on arbitrary alphabet set $\mathcal{J}$, satisfying

$$T_{J|X,X_{\mathrm{r}}}(j|x, x_{\mathrm{r}}) = T_{J|X,X_{\mathrm{r}}}(j|x, x'_{\mathrm{r}}), \quad \forall j, x, x_{\mathrm{r}}, x'_{\mathrm{r}} \tag{1}$$

where

$$T_{J|X,X_{\mathrm{r}}}(j|x, x_{\mathrm{r}}) = \sum_{y, y_{\mathrm{r}}} T_{J|Y,Y_{\mathrm{r}},X,X_{\mathrm{r}}}(j|y, y_{\mathrm{r}}, x, x_{\mathrm{r}}) T_{Y,Y_{\mathrm{r}}|X,X_{\mathrm{r}}}(y, y_{\mathrm{r}}|x, x_{\mathrm{r}}).$$

In other words $X_{\mathrm{r}} \multimap X \multimap J$ is Markov.

The following theorem proves an upper bound to the capacity of a general relay channel $T(y, y_{\mathrm{r}}|x, x_{\mathrm{r}})$. It is a J-version of the cutset bound.

Theorem 1. *Consider a relay channel $T(y, y_{\mathrm{r}}|x, x_{\mathrm{r}})$. Then, a rate R is achievable only if*

$$R \leq \max_{p_{X,X_{\mathrm{r}}}} \min_{T\in\mathcal{F}} \max_{p_{W|X,X_{\mathrm{r}}}} \min\big(I(X, X_{\mathrm{r}}; Y),\ I(X; Y, Y_{\mathrm{r}}|X_{\mathrm{r}}),\ I(X; J, Y_{\mathrm{r}}|X_{\mathrm{r}}) + I(X, X_{\mathrm{r}}; Y|W) - I(X, X_{\mathrm{r}}; J|W)\big) \tag{2}$$

where the first minimum is over $T_{J|Y,Y_{\mathrm{r}},X,X_{\mathrm{r}}} \in \mathcal{F}$ and the second maximum is over auxiliary random variables W satisfying $W \multimap (X, X_{\mathrm{r}}) \multimap (Y, Y_{\mathrm{r}}, J)$. Further it suffices to consider $|\mathcal{W}| \leq |\mathcal{X}||\mathcal{X}_{\mathrm{r}}|$.

Remark 3. Other upper bounds on the capacity of the relay channels are given in a follow-up work [EGGN21a]. The upper bound given in Theorem 1 is not subsumed by the upper bounds in [EGGN21a] because it applies to any arbitrary relay channel and imposes the Markov structure $W \multimap (X, X_{\mathrm{r}}) \multimap (J, Y_{\mathrm{r}}, Y)$ on the auxiliary random variable W.

Remark 4. The first two terms in the statement of the theorem are the cutset bound terms. The third term is the summation of $I(X; J, Y_{\mathrm{r}}|X_{\mathrm{r}})$ and $I(X, X_{\mathrm{r}}; Y|W) - I(X, X_{\mathrm{r}}; J|W)$. As the proof indicates, the term $I(X; J, Y_{\mathrm{r}}|X_{\mathrm{r}})$ is an upper bound on the flow of information to the auxiliary receiver J, while $I(X, X_{\mathrm{r}}; Y|W) - I(X, X_{\mathrm{r}}; J|W)$ is an upper bound on the flow of information to receiver Y minus the flow of information to auxiliary receiver J.

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 $p_{M,X^n,X_{\mathrm{r}}^n,Y^n,Y_{\mathrm{r}}^n}$. Since randomization at the sender or relay does not improve the capacity region, without loss of generality we assume that X^n is a function of the message M and $X_{\mathrm{r}i}$ 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 cutset upper bound constraints (see [EK12, Theorem 16.1]). Therefore, it only remains to prove the third.

The proof uses Fano's inequality and identifies auxiliary variables from the code-book induced distributions. We use several data-processing inequalities inferred from Markov chains in the course of our proof.

Let

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

We can think of $p_{J,Y,Y_{\mathrm{r}}|X,X_{\mathrm{r}}}$ 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

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

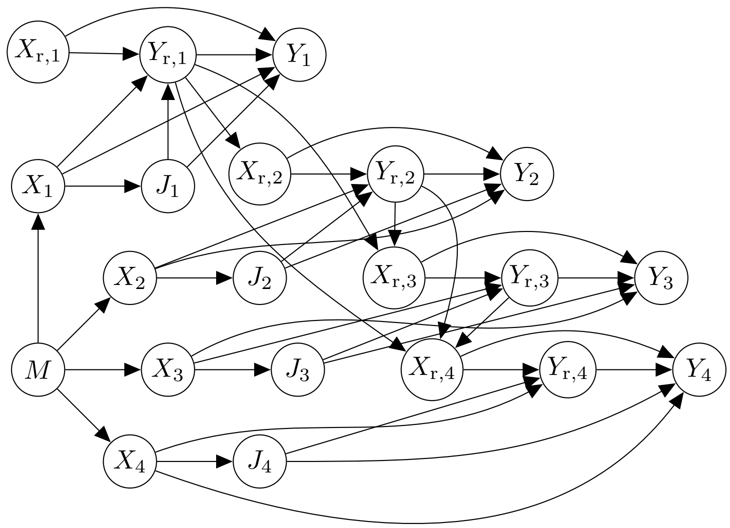

FIGURE 2. Bayesian network for probability distribution induced by codes for $T \in \mathcal{F}$, when $n = 4$. In drawing this Bayesian network, we utilize the fact that $T_{J,Y,Y_{\mathrm{r}}|X,X_{\mathrm{r}}} = T_{J|X} T_{Y,Y_{\mathrm{r}}|J,X,X_{\mathrm{r}}}$.

$$= \sum_i I(X_{\mathrm{r}i}, X_i; Y_i | W_i) - \sum_i I(X_{\mathrm{r}i}, X_i; J_i | W_i)$$

where (a) follows from Lemma 5 in Appendix A, (b) follows from the fact that X_i is a function of M and (c), (d), (e) follow from the Markov chains $(J_{i+1}^n, Y^{i-1}, M, X_{\mathrm{r}i}) -\!\!\circ\!\!- X_i -\!\!\circ\!\!- J_i$ and $(J_{i+1}^n, Y^{i-1}, M) -\!\!\circ\!\!- (X_i, X_{\mathrm{r}i}) -\!\!\circ\!\!- Y_i$ (deducible using the d-separation theorem and the Bayesian network in Figure 2)[1], respectively. Finally we set

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

Next, observe that

$$\begin{aligned}
I(M; J^n) &\le I(M; J^n, Y_{\mathrm{r}}^n) && (5)\\
&= \sum_{i=1}^n I(M; J_i, Y_{\mathrm{r}i} | J^{i-1}, Y_{\mathrm{r}}^{i-1}, X_{\mathrm{r}_i}) && \left(\because X_{\mathrm{r}i} = \mathcal{E}_{\mathrm{r}i}(Y_{\mathrm{r}}^{i-1})\right)\\
&\le \sum_{i=1}^n I(M, J^{i-1}, Y_{\mathrm{r}}^{i-1}; J_i, Y_{\mathrm{r}i} | X_{\mathrm{r}_i})\\
&\le \sum_{i=1}^n I(X_i; J_i, Y_{\mathrm{r}i} | X_{\mathrm{r}_i}) && \left(\because (J^{i-1}, Y_{\mathrm{r}}^{i-1}, M) -\!\!\circ\!\!- (X_i, X_{\mathrm{r}i}) -\!\!\circ\!\!- (J_i, Y_{\mathrm{r},i})\right).
\end{aligned}$$

Thus we get

$$n(R - \epsilon_n) \le I(M; Y^n) \le \sum_{i=1}^n \left(I(X_i; J_i, Y_{\mathrm{r}i} | X_{\mathrm{r}_i}) + I(X_{\mathrm{r}i}, X_i; Y_i | W_i) - I(X_{\mathrm{r}i}, X_i; J_i | W_i) \right). \tag{6}$$

[1]Consider the Bayesian network representation for $n = 4$ in Figure 2. Time index 3 can be thought of as the present index, while indices $1, 2$ can be considered past and index 4 can be considered future. We plot for $n = 4$ rather than $n = 3$ since $X_{\mathrm{r},1}$ is independent of M but every other $X_{\mathrm{r},i}$ is a function of Y_{r}^{i-1}.

Take a time-sharing random variable Q uniform over $\{1, 2, \cdots, n\}$ and independent of previously defined random variables. Observe that

$$\begin{aligned}
&\sum_{i=1}^{n} \left(I(X_i; J_i, Y_{\mathrm{r}i}|X_{\mathrm{r}_i}) + I(X_{\mathrm{r}i}, X_i; Y_i|W_i) - I(X_{\mathrm{r}i}, X_i; J_i|W_i)\right) \\
&= n\left(I(X_Q; J_Q, Y_{\mathrm{r}Q}|X_{\mathrm{r},Q}, Q) + I(X_{\mathrm{r}Q}, X_Q; Y_Q|W_Q, Q) - I(X_{\mathrm{r}Q}, X_Q; J_Q|W_Q, Q)\right) \\
&\leq n\left(I(Q, X_Q; J_Q, Y_{\mathrm{r}Q}|X_{\mathrm{r},Q}) + I(X_{\mathrm{r}Q}, X_Q; Y_Q|W_Q, Q) - I(X_{\mathrm{r}Q}, X_Q; J_Q|W_Q, Q)\right) \\
&= n\left(I(X_Q; J_Q, Y_{\mathrm{r}Q}|X_{\mathrm{r},Q}) + I(X_{\mathrm{r}Q}, X_Q; Y_Q|W_Q, Q) - I(X_{\mathrm{r}Q}, X_Q; J_Q|W_Q, Q)\right)
\end{aligned}$$

Let $X = X_Q$, $X_\mathrm{r} = X_{\mathrm{r}Q}$, $Y = Y_Q$, $Y_\mathrm{r} = Y_{\mathrm{r}Q}$, $J = J_Q$ and $W = (Q, W_Q)$. With this identification, (6) implies that

$$R \leq I(X; J, Y_\mathrm{r}|X_\mathrm{r}) + I(X, X_\mathrm{r}; Y|W) - I(X, X_\mathrm{r}; J|W).$$

From the proof of the cutset bound we also know that with the same identification of random variables we have

$$\begin{aligned}
R - \epsilon_n &\leq I(X, X_\mathrm{r}; Y), \\
R - \epsilon_n &\leq I(X; Y, Y_\mathrm{r}|X_\mathrm{r}),
\end{aligned}$$

where ϵ_n tends to zero as blocklength n tends to infinity.

The code defines a joint distribution on (X^n, X_r^n). Observe that the joint distribution induced on (X, X_r) does not depend on the choice of the auxiliary receiver J; it only depends on the original code. However, the conditional distribution $p_{W|X,X_\mathrm{r}}$ does depend on our choice of the auxiliary receiver J since W_i involves J_{i+1}^n. Therefore, we conclude that there exists some p_{X,X_r} such that for every auxiliary receiver J there exists some $p_{W|X,X_\mathrm{r}}$ such that

$$R \leq \min\big(I(X, X_\mathrm{r}; Y),\ I(X; Y, Y_\mathrm{r}|X_\mathrm{r}),\ I(X; J, Y_\mathrm{r}|X_\mathrm{r}) + I(X, X_\mathrm{r}; Y|W) - I(X, X_\mathrm{r}; J|W)\big).$$

Equivalently,

$$R \leq \max_{p_{X,X_\mathrm{r}}} \min_{T\in\mathcal{F}} \max_{p_{W|X,X_\mathrm{r}}} \min\big(I(X, X_\mathrm{r}; Y),\ I(X; Y, Y_\mathrm{r}|X_\mathrm{r}),\ I(X; J, Y_\mathrm{r}|X_\mathrm{r}) + I(X, X_\mathrm{r}; Y|W) - I(X, X_\mathrm{r}; J|W)\big).$$

This completes the proof. □

Remark 5. One could see that the term $I(M; J^n)$ in (5) could be also bounded from above by $nI(X, X_\mathrm{r}; J)$, yielding the apparently stronger constraint

$$R \leq \min\left(I(X, X_\mathrm{r}; J), I(X; J, Y_\mathrm{r}|X_\mathrm{r})\right) + I(X, X_\mathrm{r}; Y|W) - I(X, X_\mathrm{r}; J|W).$$

However, this apparent strengthening of the bound is not useful. To see this, observe that for the optimizing W we have

$$I(X, X_\mathrm{r}; J) + I(X, X_\mathrm{r}; Y|W) - I(X, X_\mathrm{r}; J|W) \geq I(X, X_\mathrm{r}; J) + I(X, X_\mathrm{r}; Y) - I(X, X_\mathrm{r}; J) = I(X, X_\mathrm{r}; Y)$$

since W being constant is always a valid choice.

Corollary 1. *Consider the special case $T(y, y_\mathrm{r}|x, x_\mathrm{r}) = T_a(y_\mathrm{r}|x)T_b(y|x, x_\mathrm{r}, y_\mathrm{r})$ which includes, for instance, the Gaussian relay channels. For this class, any achievable rate R must satisfy*

$$R \leq \min\big(I(X, X_\mathrm{r}; Y),\ I(X; Y, Y_\mathrm{r}|X_\mathrm{r}),\ I(X; Y_\mathrm{r}|X_\mathrm{r}) + \sup_{W \multimap (X,X_\mathrm{r}) \multimap (Y,Y_\mathrm{r})} I(X, X_\mathrm{r}; Y|W) - I(X, X_\mathrm{r}; Y_\mathrm{r}|W)\big), \quad (7)$$

for some $p(x, x_\mathrm{r})$.

Proof. For this class, since $X_\mathrm{r} \multimap X \multimap Y_\mathrm{r}$ is Markov, we can set $J = Y_\mathrm{r}$, and the bound in (2) implies the result. □

Remark 6. The upper bound in Corollary 1 can be compared with the partial decode-and-forward lower bound which states that a rate R is achievable if

$$R \leq \min\big(I(X, X_\mathrm{r}; Y),\ I(X; Y_\mathrm{r}|X_\mathrm{r}) + \sup_{W \multimap (X,X_\mathrm{r}) \multimap (Y,Y_\mathrm{r})} I(X; Y|W, X_\mathrm{r}) - I(X; Y_\mathrm{r}|W, X_\mathrm{r})\big), \quad (8)$$

for some $p(x, x_\mathrm{r})$. More specifically, the last term in (8) is the same as that in (7) except that we have replaced W by (W, X_r).

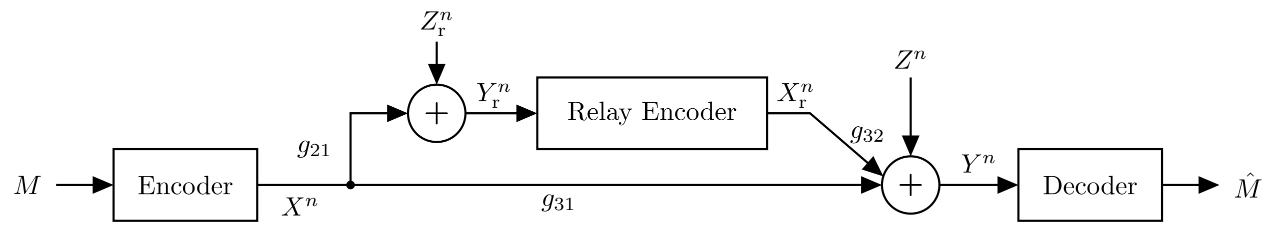

FIGURE 3. Depiction of a Gaussian relay channel.

Remark 7. A weaker form of the bound in (2) which does not involve auxiliary random variables is

$$R \leq \min\left(I(X, X_{\mathrm{r}}; Y),\ I(X; Y, Y_{\mathrm{r}}|X_{\mathrm{r}}),\ I(X; J, Y_{\mathrm{r}}|X_{\mathrm{r}}) + I(X, X_{\mathrm{r}}; Y|J)\ \right) \tag{9}$$

for any arbitrary $T_{J|Y,Y_{\mathrm{r}},X,X_{\mathrm{r}}} \in \mathcal{F}$. This follows from the following:

$$I(X, X_{\mathrm{r}}; Y|W) - I(X, X_{\mathrm{r}}; J|W) \leq I(X, X_{\mathrm{r}}; Y|J, W) \leq I(X, X_{\mathrm{r}}; Y|J).$$

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

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

$$Y_{\mathrm{r}} = g_{21}X + Z_1, \quad Y = g_{31}X + g_{32}X_{\mathrm{r}} + Z_2$$

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 and X_{r} are both given by P. This is depicted in Fig. 3. 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 upper bound:

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

$$\begin{aligned}\min\Bigg(&\mathsf{C}\left(S_{31} + S_{32} + 2\rho\sqrt{S_{31}S_{32}}\right),\ \mathsf{C}\left((1-\rho^2)(S_{31} + S_{21})\right),\\ &\min_{S_J \geq S_{21}}\left\{\mathsf{C}\left(S_J(1-\rho^2)\right) + \mathsf{C}\left(x_*^2 + \left(x_*\left(\sqrt{S_{31}} + \sqrt{\rho^2 S_{32}}\right) + \sqrt{(1-\rho^2)S_{32}}\right)^2\right) - \mathsf{C}\left(x_*^2(1 + S_J)\right)\right\}\Bigg),\end{aligned} \tag{10}$$

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

$$\begin{aligned}&x^2\left(\sqrt{S_{31}} + \sqrt{\rho^2 S_{32}}\right)(1 + S_J)\sqrt{(1-\rho^2)S_{32}} + x\left((1-\rho^2)S_{32}(1 + S_J) - \left(\sqrt{S_{31}} + \sqrt{\rho^2 S_{32}}\right)^2 + S_J\right)\\ &\quad - \left(\sqrt{S_{31}} + \sqrt{\rho^2 S_{32}}\right)\sqrt{(1-\rho^2)S_{32}} = 0.\end{aligned} \tag{11}$$

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.*, $X \multimap J \multimap Y_{\mathrm{r}}$ forms a Markov chain.[2]

Observe that the first two terms in (10) correspond to the cutset bound [EK12, Eq. 16.4]:

$$\begin{aligned}C &\leq \max_{-1\leq\rho\leq 1}\min\left(\mathsf{C}\left(S_{31} + S_{32} + 2\rho\sqrt{S_{31}S_{32}}\right),\ \mathsf{C}\left((1-\rho^2)(S_{31} + S_{21})\right)\right)\\ &= \begin{cases}\mathsf{C}\left(\frac{\left(\sqrt{S_{21}S_{32}} + \sqrt{S_{31}(S_{31} + S_{21} - S_{32})}\right)^2}{S_{31} + S_{21}}\right) & \text{if } S_{21} \geq S_{32}\\ \mathsf{C}(S_{31} + S_{21}) & \text{otherwise.}\end{cases}\end{aligned} \tag{12}$$

[2]The more restrictive choice of $J = Y_{\mathrm{r}}$ 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.

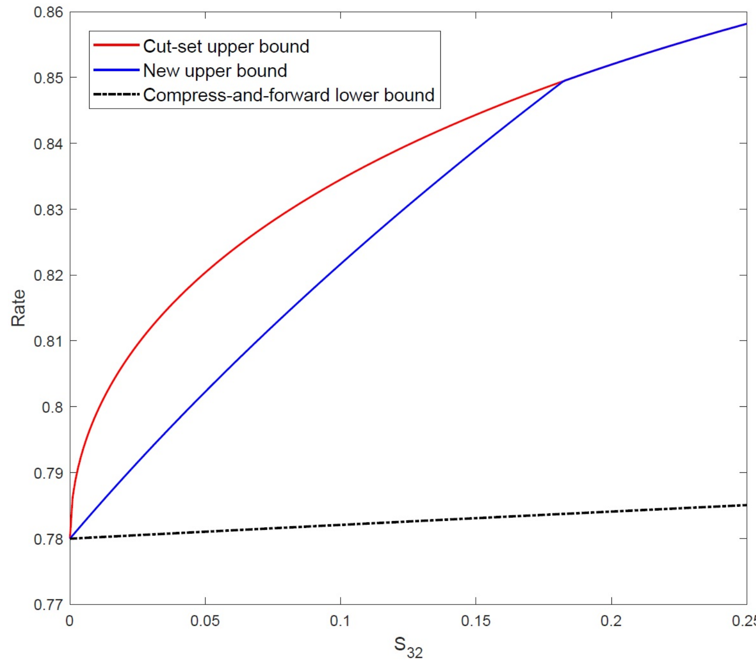

FIGURE 4. Illustration of the bounds for a Gaussian relay channel. Parameters are $S_{31} = 3.7585$, $S_{21} = 1.2139$.

Example 1. As a specific example, set $S_{21} = 1.2139$, $S_{31} = 3.7585$, $S_{32} = 0.032519$ and $S_J = S_{21}$. Then, the cutset 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 lower bound of [CM11] also evaluates to 0.78066. The decode-and-forward evaluates to 0.39737. See also Fig. 4 for a plot of the bounds.

Remark 8. The above bound is further improved in [EGGN21a]. In [EGGN21a], we also give a different upper bound with an auxiliary receiver J which is an enhancement of the relay's output variable Y_{r} (similar to the auxiliary receiver considered in this work). However, the identifications of the auxiliary variables in this work are different from that in [EGGN21a].

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

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

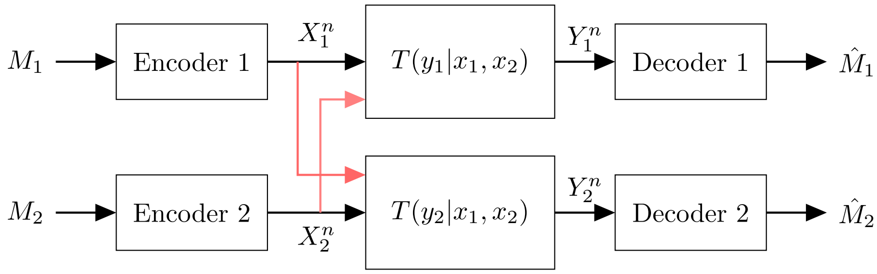

FIGURE 5. The setup for an interference channel.

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 cutset bound, see (12), 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 cutset bound allows for cooperation between the relay and the transmitter. To see this, observe that if we evaluate the cutset upper bound expression of $\min\left(I(X, X_\mathrm{r}; Y), I(X; Y, Y_\mathrm{r}|X_\mathrm{r})\right)$ assuming X and X_r to be independent Gaussians (no cooperation) we obtain a value $\min\left(\mathsf{C}(S_{31}+S_{21}), \mathsf{C}(S_{31}+S_{32})\right)$.

The cutset 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. Fig. 4 illustrates the finite slope of the new bound.

3. Interference Channel

A two-user interference channel models the 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 $T(y_1, y_2|x_1, x_2)$. An (n, R_1, R_2) code for a memoryless interference channel $T(y_1, y_2|x_1, x_2)$, depicted in Fig. 5, consists of two encoders $\mathcal{E}_1$ and $\mathcal{E}_2$ that map independent messages M_1 and M_2 (uniform over $[2^{nR_1}]$ and $[2^{nR_2}]$ respectively) to input sequences $X_1^n \in \mathcal{X}_1^n$ and $X_2^n \in \mathcal{X}_2^n$, *i.e.*, $X_1^n = \mathcal{E}_1(M_1)$ and $X_2^n = \mathcal{E}_2(M_2)$ and two decoders $\mathcal{D}_1$ and $\mathcal{D}_2$ that produce estimates $\hat{M}_1$ and $\hat{M}_2$ from Y_1^n and Y_2^n respectively. The following joint distribution is induced by the code over a memoryless interference channel $T(y_1, y_2|x_1, x_2)$:

$$p(m_1)p(m_2)p(x_1^n|m_1)p(x_2^n|m_2)\left(\prod_{i=1}^n T(y_{1i}, y_{2i}|x_{1i}, x_{2i})\right)p(\hat{m}_1|y_1^n)p(\hat{m}_2|y_2^n).$$

The error probability of the code is $\mathrm{P}[(M_1, M_2) \neq (\hat{M}_1, \hat{M}_2)]$. 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. 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 [EK12, Chapter 6] 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.

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 $T_{J|X_1,X_2,Y_1,Y_2}$ such that $p_{J,Y_1,Y_2|X_1,X_2} = T_{Y_1,Y_2|X_1,X_2}T_{J|X_1,X_2,Y_1,Y_2}$ satisfies*

$$p_{J,Y_1,Y_2|X_1,X_2} = p_{J|X_1}p_{Y_2|J,X_2}p_{Y_1|J,X_1,X_2,Y_2}, \tag{13}$$

we have

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

for some $p(q)p(x_1|q)p(x_2|q)p(w, \hat{w}|x_1, x_2, q)p(y_1, y_2, j|x_1, x_2)$ satisfying

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

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

Proof. The cardinality bounds on the auxiliary variables $W, \hat{W}$ follow from standard arguments and is omitted.

Take a code of length n with input sequences $X_1^n = \mathcal{E}_1(M_1)$ and $X_2^n = \mathcal{E}_2(M_2)$. Let

$$p_{J^n|M_1,M_2,X_1^n,X_2^n,Y_1^n,Y_2^n} = \prod_{i=1}^n T_{J_i|X_{1i},X_{2i},Y_{1i},Y_{2i}}.$$

We can think of $p_{J,Y_1,Y_2|X_1,X_2}$ as an extended interference channel. Observe from (13) that $p_{J_i|X_1^n,X_2^n} = p_{J_i|X_{1i}}$ and $p_{Y_{2i}|J^n,X_2^n} = p_{Y_{2i}|J_i,X_{2i}}$.

One can prove the above theorem by identifying the auxiliary variable

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

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 $R_1 \le I(X_1; Y_1|X_2, Q)$ is standard.

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

$$\begin{aligned} I(X_1^n; Y_1^n) &= I(X_1^n; J^n) + I(X_1^n; Y_1^n) - I(X_1^n; J^n)\\ &\le I(X_1^n; J^n) + I(X_1^n, X_2^n; Y_1^n) - I(X_1^n, X_2^n; J^n) && \because X_2^n \perp (X_1^n, J^n)\\ &= \sum_{i=1}^n \Big(I(X_1^n; J_i|J_{i+1}^n) + I(X_1^n, X_2^n; Y_{1i}|\hat{W}_i) - I(X_1^n, X_2^n; J_i|\hat{W}_i)\Big) && \text{(Lemma 4, Lemma 5)}\\ &= \sum_{i=1}^n \Big(I(X_{1i}; J_i|J_{i+1}^n) + I(X_{1i}, X_{2i}; Y_{1i}|\hat{W}_i) - I(X_{1i}; J_i|\hat{W}_i)\Big) && \text{(memoryless channels)}\\ &\le \sum_{i=1}^n \Big(I(X_{1i}; J_i|W_i) + I(W_i; Y_{2i}) + I(X_{1i}, X_{2i}; Y_{1i}|\hat{W}_i) - I(X_{1i}; J_i|\hat{W}_i)\Big) && \text{(Lemma 4, Lemma 5)} \end{aligned}$$

establishing the second bound on R_1.

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

$$\begin{aligned} I(X_2^n; Y_2^n) &\le I(X_2^n; Y_2^n|X_1^n) && \because X_2^n \perp X_1^n\\ &= \sum_{i=1}^n I(X_2^n; Y_{2i}|X_1^n, Y_2^{i-1}, J_{i+1}^n) && \because J_{i+1}^n -\!\!\circ\!\!- X_{1i+1}^n -\!\!\circ\!\!- (X_2^n, X_1^i) -\!\!\circ\!\!- (Y_1^i, Y_2^i) \end{aligned}$$

$$\leq \sum_{i=1}^{n} I(X_{2i};Y_{2i}|X_{1i},W_i) \qquad \text{(memoryless channels)}$$

establishing the first bound on R_2.

To get the second bound, observe that

$$\begin{aligned}
I(M_2;Y_2^n) &= I(M_2;Y_2^n) - I(M_2;J^n) && \because M_2 \perp J^n \\
&= \sum_{i=1}^{n} I(M_2;Y_{2i}|Y_2^{i-1},J_{i+1}^n) - I(M_2;J_i|Y_2^{i-1},J_{i+1}^n) && \text{(Lemma 4)} \\
&= \sum_{i=1}^{n} I(M_2,X_{2i};Y_{2i}|Y_2^{i-1},J_{i+1}^n) - I(M_2,X_{2i};J_i|Y_2^{i-1},J_{i+1}^n) && \because H(X_{2i}|M_2)=0 \\
&\overset{(a)}{=} \sum_{i=1}^{n} I(X_{2i};Y_{2i}|Y_2^{i-1},J_{i+1}^n) - I(X_{2i};J_i|Y_2^{i-1},J_{i+1}^n) + I(M_2;Y_{2i}|Y_2^{i-1},J_{i+1}^n,X_{2i}) \\
&\qquad - I(M_2;J_i,Y_{2i}|Y_2^{i-1},J_{i+1}^n,X_{2i}) \\
&\leq \sum_{i=1}^{n} I(X_{2i};Y_{2i}|W_i) - I(X_{2i};J_i|W_i),
\end{aligned}$$

where (a) follows since $(Y_2^{i-1},J_{i+1}^n,M_2) \multimap (X_{2i},J_i) \multimap Y_{2i}$ is Markov from (13). This establishes 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

$$\begin{aligned}
0 &\leq I(X_1^n;J^n,X_2^n) - I(X_1^n;Y_2^n) && \because X_1^n \multimap J^n \multimap (X_2^n,Y_2^n) \\
&= I(X_1^n;J^n) - I(X_1^n;Y_2^n) && \because (X_1^n,J^n) \perp X_2^n \\
&= \sum_i I(X_1^n;J_i|W_i) - I(X_1^n;Y_{2i}|W_i) && \text{(Lemma 4)} \\
&= \sum_i I(X_{1i};J_i|W_i) - I(X_1^n;Y_{2i}|W_i) && \text{(memoryless channel)} \\
&\leq \sum_i I(X_{1,i};J_i|W_i) - I(X_{1i};Y_{2i}|W_i),
\end{aligned}$$

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 $p_{J,Y_1,Y_2|X_1,X_2} = T_{Y_1,Y_2|X_1,X_2}T_{J|X_1,X_2,Y_1,Y_2}$ satisfies*

$$p_{J,Y_1,Y_2|X_1,X_2} = p_{J|X_1}p_{Y_2|J,X_2}p_{Y_1|J,X_1},$$

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. □

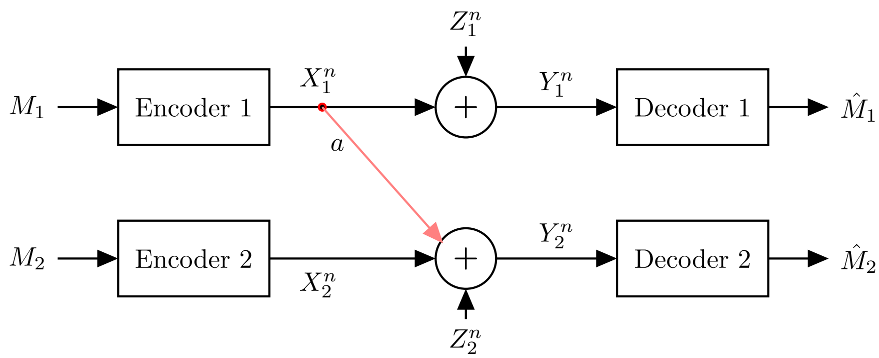

FIGURE 6. Illustration of a Gaussian Z-interference channel.

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

$$\begin{aligned} Y_1 &= X_1 + Z_1 \\ Y_2 &= aX_1 + X_2 + Z_2\,, \end{aligned} \tag{14}$$

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

$$\|X_1^n\|^2 \leq nP_1, \quad \|X_2^n\|^2 \leq nP_2. \tag{15}$$

See Fig. 6 for an illustration.

The assumption $0 < a < 1$ corresponds to the weak interference regime. The case of $a \geq 1$ corresponds to the strong interference regime and its capacity region is fully known [Sat81], [Cos85a].

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 Z-interference channel with power constraints P_1, P_2 respectively must satisfy*

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

for some $K_1 \leq P_1$ 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

$$\frac{K_1+u^2}{u^2} \geq \frac{1+a^2K_1+K_2+2a\rho\sqrt{K_1K_2}}{1+K_2(1-\rho^2)}. \tag{16}$$

Remark 9. 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 7 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 in Lemma 1 below.

Remark 10. Consider a Gaussian interference channel in the weak interference regime parameterized by the cross-gains $a, b \in (0,1)$:

$$\begin{aligned} Y_1 &= X_1 + bX_2 + Z_1 \\ Y_2 &= aX_1 + X_2 + Z_2. \end{aligned} \tag{17}$$

Then, Theorem 3 can be used to obtain two outer bounds on the capacity region of this channel, one by setting $b = 0$ and the other by setting $a = 0$. This follows from the simple observation (see [Cos85b, p. 612] or [PW15] for instance) that the capacity region of the Gaussian interference channel in the weak interference regime with cross-gains $a, b \in (0,1)$ is contained in the capacity region of either of the two Z-interference settings, one by setting $b = 0$ and the other by setting $a = 0$.

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 [Cos85b] aimed to determine the maximum value of R_1 such that (R_1, C_2) is achievable as $R_1^* = \mathsf{C}\left(\frac{a^2P_1}{1+P_2}\right)$. His argument involved two key steps. The first was to show the concavity of entropy power with added Gaussian noise [Cos85a]. The second involved approximating the empirical estimates of differential entropies seen at receivers by their distribution values. However, Sason (see [Sas04] and [Sas15] for a detailed discussion) observed that the approximation proof had a flaw and that Pinsker's inequality was insufficient to guarantee that the approximation error would grow linearly with block size n. The problem then rested open for eleven years until Polyanskiy and Wu [PW15], using Talagrand's inequality [Tal96] 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 cutset 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[3] 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 1. *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*

$$R_1 \le R_1^* = \frac{1}{2}\log\left(1 + \frac{a^2P_1}{1+P_2}\right).$$

(*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*

$$R_2 \le R_2^* = \frac{1}{2}\log\left(1 + \frac{P_2}{1+a^2P_1}\right).$$

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

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

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

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

[3] 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.

implying $I(X_2;X_1|W,Y_2,Q)=0$. From $I(X_2;X_1|W,Y_2,Q)=0$, we deduce that $I(X_1;Y_2-X_2|W,Y_2,Q)=I(X_1;aX_1+Z_2|W,Y_2,Q)=0$. On the other hand,

$$\begin{aligned} I(Y_2;X_1|aX_1+Z_2,W,Q) &= I(X_2;X_1|aX_1+Z_2,W,Q) \\ &\leq I(X_2;X_1,aX_1+Z_2,W,Q) \\ &= 0 \end{aligned}$$

where in the last step we used $X_2 \perp (W,X_1,Q)$. From $I(X_1;aX_1+Z_2|W,Y_2,Q)=0$ and $I(Y_2;X_1|aX_1+Z_2,W,Q)=0$ we have a double Markovity property (see Lemma 6). From the double Markovity property and since the joint distribution of (aX_1+Z_2,Y_2) is indecomposable (as defined in [CK11, Exercise 16.25]), we obtain that conditioned on (W,Q), $X_1 \perp (aX_1+Z_2,Y_2)$. This implies, conditioned on (W,Q), that X_1 is independent of $X_1+\frac{Z_2}{a}$. 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

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

Therefore to show that the outer bound given in Theorem 3 lies inside the outer bound by Sato, it suffices to show that

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

Equivalently, it suffices to show that

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

This is immediate since ($a\in(0,1)$) and

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

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*

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

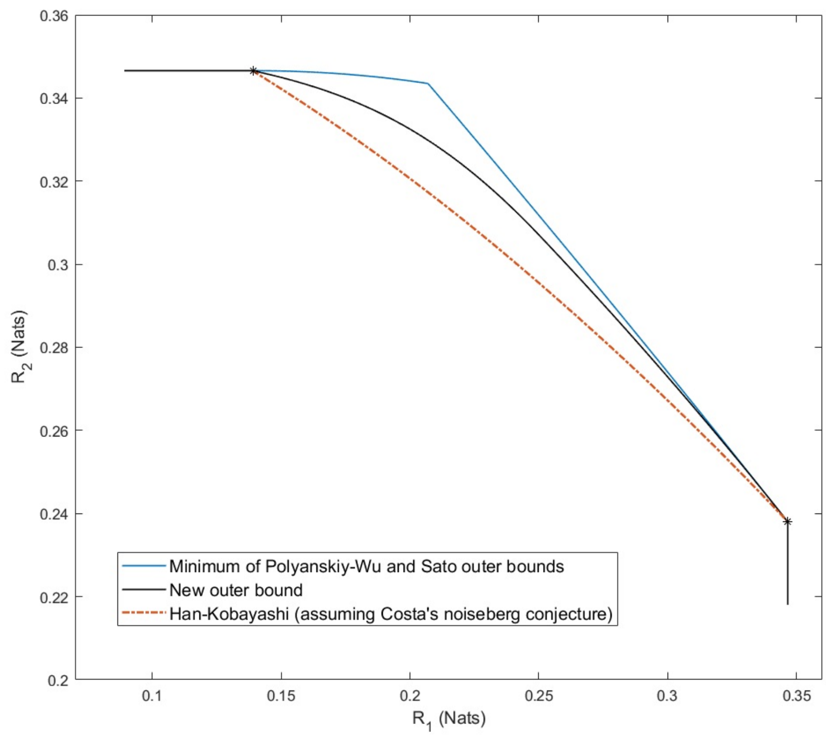

FIGURE 7. Illustration for $a = 0.8$, $P_1 = P_2 = 1$. The new outer bound is plotted for the fixed choice of $\beta = 0$ and $u = 1$ in Theorem 3 and optimized over α. The Polyanskiy-Wu bound is flat at Costa's corner point. However, Theorem 4 establishes that the capacity region has a kink at Costa's 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.

when

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

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

Remark 11. The 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 achievable 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 yield a 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.

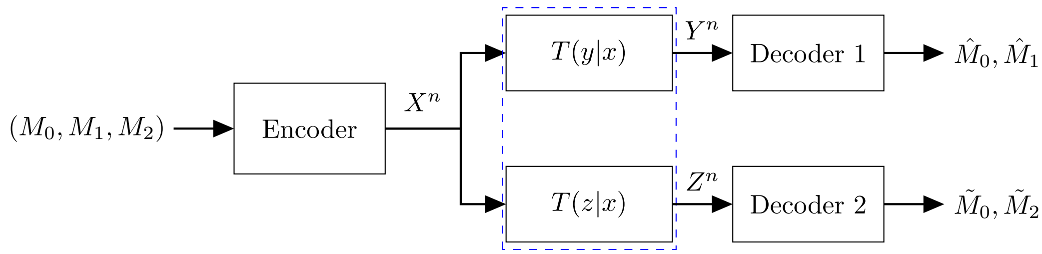

FIGURE 8. Illustration of a discrete memoryless broadcast channel.

4. Broadcast Channel

A two-receiver broadcast channel [Cov72] models the transmission of messages from a single sender X to two receivers Y and Z. A discrete memoryless broadcast channel is described by a conditional distribution $T(y, z|x)$ with $|\mathcal{X}|, |\mathcal{Y}|, |\mathcal{Z}| < \infty$. An (n, R_0, R_1, R_2) code for a memoryless broadcast channel $T(y, z|x)$, depicted in Fig. 8, consists of an encoder $\mathcal{E}$ that maps three mutually independent messages M_0, M_1 and M_2 (uniform over $[2^{nR_0}]$, $[2^{nR_1}]$ and $[2^{nR_2}]$ respectively) to an input sequence $X^n \in \mathcal{X}^n$ *i.e.,* $X^n = \mathcal{E}(M_0, M_1, M_2)$ and two decoders $\mathcal{D}_1$ and $\mathcal{D}_2$ that produce estimates $(\hat{M}_0, \hat{M}_1)$ and $(\tilde{M}_0, \tilde{M}_2)$ from Y^n and Z^n respectively. Random variable M_0 represents the common message from transmitter to the two receivers while M_1 and M_2 are private messages to the receivers. The following joint distribution is induced by the code over a memoryless interference channel $T(y, z|x)$:

$$p(m_0)p(m_1)p(m_2)p(x^n|m_0, m_1, m_2)\left(\prod_{i=1}^{n} T(y_i, z_i|x_i)\right)p(\hat{m}_0, \hat{m}_1|y^n)p(\tilde{m}_0, \tilde{m}_2|z^n).$$

The error probability of the code is $\mathrm{P}[(M_0, M_1, M_0, M_2) \neq (\hat{M}_0, \hat{M}_1, \tilde{M}_0, \tilde{M}_2)]$. 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 [EK12, Chapters 5 and 8].

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* $(U, V, W) \multimap X \multimap (Y, Z)$ *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*

$$R_0 \leq \min(I(W;Y), I(W;Z)),$$

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

for some triple of random variables (U,V,W) such that $(U,V,W) -\!\!\circ\!\!- X -\!\!\circ\!\!- (Y,Z)$.

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 two outer bounds for (general) broadcast channels, both of which (strictly) improve over the UV outer bound. In particular, the second bound generalizes the one in [GGNY14] by relaxing 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 $p_{U,V,W,X}$, we have

$$
\begin{aligned}
&I(W;Y') = (1-\epsilon)I(W;Y), \quad I(U;Y'|W) = (1-\epsilon)I(U;Y|W), \quad I(X;Y'|V,W) = (1-\epsilon)I(X;Y|V,W), \\
&I(W;Z') = (1-\epsilon)I(W;Z), \quad I(V;Z'|W) = (1-\epsilon)I(V;Z|W), \quad I(X;Z'|U,W) = (1-\epsilon)I(X;Z|U,W).
\end{aligned}
$$

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 2).

Theorem 7. *Given a broadcast channel characterized by $T(y,z|x)$ and any achievable rate triple (R_0,R_1,R_2), one can find some input distribution $p(x)$ such that for any auxiliary channel $T_{J|X,Y,Z}$, the following constraints are satisfied:*

$$
R_0 \le \min\{I(W;Y), I(\hat{W};Y), I(W;Z), I(\tilde{W};Z)\}, \tag{18a}
$$

$$
R_0 + R_1 \le \min\{I(W;Y), I(W;Z)\} + I(U;Y|W), \tag{18b}
$$

$$
\begin{aligned}
R_0 + R_1 \le \min\Big\{ &I(\tilde{W};Z) + \min\big[0, I(W;Y) - I(W;Z)\big],\ I(\tilde{W};J) + I(\hat{W};Y) - I(\hat{W};J) \Big\} \\
&+ I(\tilde{U};J|\tilde{W}) + I(\hat{U};Y|\hat{W}) - I(\hat{U};J|\hat{W}),
\end{aligned} \tag{18c}
$$

$$
\begin{aligned}
R_0 + R_1 \le \min\Big\{ &I(\hat{W};Y) + \min\big[0, I(W;Z) - I(W;Y)\big],\ I(\hat{W};J) + I(\tilde{W};Z) - I(\tilde{W};J) \Big\} \\
&+ I(\hat{U};Y|\hat{W}),
\end{aligned} \tag{18d}
$$

$$
R_0 + R_2 \le \min\{I(W;Y), I(W;Z)\} + I(V;Z|W), \tag{18e}
$$

$$
\begin{aligned}
R_0 + R_2 \le \min\Big\{ &I(\hat{W};Y) + \min\big[0, I(W;Z) - I(W;Y)\big],\ I(\hat{W};J) + I(\tilde{W};Z) - I(\tilde{W};J) \Big\} \\
&+ I(\hat{V};J|\hat{W}) + I(\tilde{V};Z|\tilde{W}) - I(\tilde{V};J|\tilde{W}),
\end{aligned} \tag{18f}
$$

$$
\begin{aligned}
R_0 + R_2 \le \min\Big\{ &I(\tilde{W};Z) + \min\big[0, I(W;Y) - I(W;Z)\big],\ I(\tilde{W};J) + I(\hat{W};Y) - I(\hat{W};J) \Big\} \\
&+ I(\tilde{V};Z|\tilde{W}),
\end{aligned} \tag{18g}
$$

$$
\begin{aligned}
R_0 + R_1 + R_2 \le \min\Big\{ &I(\hat{W};Y) - I(\hat{W};J),\ I(\tilde{W};Z) - I(\tilde{W};J) \Big\} + I(X;J) \\
&+ I(\hat{U};Y|\hat{W}) - I(\hat{U};J|\hat{W}) + I(\tilde{V};Z|\tilde{W}) - I(\tilde{V};J|\tilde{W}),
\end{aligned} \tag{18h}
$$

$$R_0 + R_1 + R_2 \leq \min\Big\{I(W;Y), I(W;Z)\Big\} + \min\Big\{I(V;Z|W) + I(X;Y|V,W), I(U;Y|W) + I(X;Z|U,W)\Big\}, \tag{18i}$$

for some choice of distribution over the variables

$$p_{U,V,W,\tilde{U},\tilde{V},\tilde{W},\hat{U},\hat{V},\hat{W},X,Y,Z,J} = p_{U,V,W,X}\, p_{\tilde{W},\tilde{U},\tilde{V}|X}\, p_{\hat{W},\hat{U},\hat{V}|X}\, T_{Y,Z|X}\, T_{J|X,Y,Z}$$

satisfying

$$I(\tilde{W};Z) - I(\tilde{W};J) + I(\hat{W};J) - I(\hat{W};Y) = I(W;Z) - I(W;Y), \tag{19a}$$

$$I(\tilde{U};Z|\tilde{W}) - I(\tilde{U};J|\tilde{W}) + I(\hat{U};J|\hat{W}) - I(\hat{U};Y|\hat{W}) = I(U;Z|W) - I(U;Y|W), \tag{19b}$$

$$I(\tilde{V};Z|\tilde{W}) - I(\tilde{V};J|\tilde{W}) + I(\hat{V};J|\hat{W}) - I(\hat{V};Y|\hat{W}) = I(V;Z|W) - I(V;Y|W), \tag{19c}$$

and

$$0 \leq I(X;Z|\tilde{U},\tilde{W}) - I(X;J|\tilde{U},\tilde{W}) \leq I(\tilde{V};Z|\tilde{W}) - I(\tilde{V};J|\tilde{W}), \tag{20a}$$

$$0 \leq I(X;Y|\hat{V},\hat{W}) - I(X;J|\hat{V},\hat{W}) \leq I(\hat{U};Y|\hat{W}) - I(\hat{U};J|\hat{W}), \tag{20b}$$

$$I(V;Z|W) + I(X;Y|V,W) = I(U;Y|W) + I(X;Z|U,W). \tag{20c}$$

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* $|\mathcal{U}|, |\mathcal{V}|, |\hat{\mathcal{U}}|, |\hat{\mathcal{V}}|, |\tilde{\mathcal{U}}|, |\tilde{\mathcal{V}}| \leq |\mathcal{X}|+1$.

Proof. Take an arbitrary code (n, M_0, M_1, M_2) with error probability ϵ_n. Let

$$p_{J^n|M_0,M_1,M_2,X^n,Y^n,Z^n} = \prod_{i=1}^{n} T_{J_i|X_i,Y_i,Z_i}.$$

We can think of $p_{J,Y,Z|X}$ as an extended broadcast channel.

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 (18a), (18b), (18e) are standard and are essentially the same (similar to UVW bound) but for completeness we present their starting points here. The following represents 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 5, guided by the identifications mentioned above.

$$nR_0 \leq \min\{I(M_0;Y^n), I(M_0;Y^n), I(M_0;Z^n), I(M_0;Z^n)\} + ng(\epsilon_n) \tag{21a}$$

$$n(R_0+R_1) \leq \min\{I(M_0;Y^n), I(M_0;Z^n)\} + I(M_1;Y^n|M_0) + ng(\epsilon_n), \tag{21b}$$

$$n(R_0+R_1) \leq \min\Big\{I(M_0;Z^n) + \min\big[0, I(M_0;Y^n) - I(M_0;Z^n)\big],\ I(M_0;J^n) + I(M_0;Y^n) - I(M_0;J^n)\Big\} + I(M_1;J^n|M_0) + I(M_1;Y^n|M_0) - I(M_1;J^n|M_0) + ng(\epsilon_n), \tag{21c}$$

$$n(R_0+R_1) \leq \min\Big\{I(M_0;Y^n) + \min\big[0, I(M_0;Z^n) - I(M_0;Y^n)\big],\ I(M_0;J^n) + I(M_0;Z^n) - I(M_0;J^n)\Big\} + I(M_1;Y^n|M_0) + ng(\epsilon_n), \tag{21d}$$

$$n(R_0+R_2) \leq \min\{I(M_0;Y^n), I(M_0;Z^n)\} + I(M_1;Z^n|M_0) + ng(\epsilon_n), \tag{21e}$$

$$n(R_0+R_2) \leq \min\Big\{I(M_0;Y^n) + \min\big[0, I(M_0;Z^n) - I(M_0;Y^n)\big],\ I(M_0;J^n) + I(M_0;Z^n) - I(M_0;J^n)\Big\} + I(M_2;J^n|M_0) + I(M_2;Z^n|M_0) - I(M_2;J^n|M_0) + ng(\epsilon_n), \tag{21f}$$

$$
\begin{aligned}
n(R_0+R_2) &\le \min\Big\{I(M_0;Z^n)+\min\big[0,I(M_0;Y^n)-I(M_0;Z^n)\big],\ I(M_0;J^n)+I(M_0;Y^n)-I(M_0;J^n)\Big\}\\
&\quad + I(M_2;Z^n|M_0)+ng(\epsilon_n), \qquad (21g)\\
n(R_0+R_1+R_2) &\le \min\Big\{I(M_0;Y^n)-I(M_0;J^n), I(M_0;Z^n)-I(M_0;J^n)\Big\}+I(X^n;J^n)\\
&\quad + I(M_1;Y^n|M_0)-I(M_1;J^n|M_0)+I(M_2;Z^n|M_0)-I(M_2;J^n|M_0)+ng(\epsilon_n), \qquad (21h)\\
n(R_0+R_1+R_2) &\le \min\left\{I(M_0;Y^n), I(M_0;Z^n)\right\}\\
&\quad + \min\Big(I(M_2;Z^n|M_0)+I(X^n;Y^n|M_2,M_0), I(M_1;Y^n|M_0)+I(X^n;Z^n|M_1,M_0)\Big)+ng(\epsilon_n), \qquad (21i)
\end{aligned}
$$

Further the constraints can be established from the following starting points again using Lemma 5, and using Fano's inequality:

$$
\begin{array}{ccccc}
I(M_0;Z^n)-I(M_0;J^n) & + & I(M_0;J^n)-I(M_0;Y^n) & = & I(M_0;Z^n)-I(M_0;Y^n),\\
I(M_1;Z^n|M_0)-I(M_1;J^n|M_0) & + & I(M_1;J^n|M_0)-I(M_1;Y^n|M_0) & = & I(M_1;Z^n|M_0)-I(M_1;Y^n|M_0),\\
I(M_2;Z^n|M_0)-I(M_2;J^n|M_0) & + & I(M_2;J^n|M_0)-I(M_2;Y^n|M_0) & = & I(M_2;Z^n|M_0)-I(M_2;Y^n|M_0),
\end{array}
$$

and

$$
\begin{aligned}
&0 \le I(X^n;Z^n|M_1,M_0)-I(X^n;J^n|M_1,M_0)+ng_1(\epsilon_n) \le I(M_2;Z^n|M_0)-I(M_2;J^n|M_0)+ng_2(\epsilon_n),\\
&0 \le I(X^n;Y^n|M_2,M_0)-I(X^n;J^n|M_2,M_0)+ng_3(\epsilon_n) \le I(M_1;Y^n|M_0)-I(M_1;J^n|M_0)+ng_4(\epsilon_n),\\
&I(M_2;Z^n|M_0)+I(X^n;Y^n|M_2,M_0)=I(M_1;Y^n|M_0)+I(X^n;Z^n|M_1,M_0)+ng_5(\epsilon_n).
\end{aligned}
$$

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 consistent distributions on 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 12. From (18a), (18b), (18e), (18i), we can extract the following constraints:

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

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

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 $T(y|x)=\sum_{\hat{y}} T(y|\hat{y})T(\hat{y}|x)$ 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*

$$R_2 \leq I(V; Z|W) - I(V; \hat{Y}|W) \tag{23}$$

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

$$C_1 = I(X; Y) \leq I(W; Z) + I(X; \hat{Y}|W). \tag{24}$$

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 (18c), 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 (18h) we see that

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

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

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.** In this subsection we will focus on the private message case, i.e. the projection of the capacity region onto the plane $R_0 = 0$. 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. In this subsection, will demonstrate that Corollary 3 (and hence the outer bound of Theorem 7) leads to an improvement in the bound for the corner point for the erasure Blackwell channel described below.

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 2 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 2. *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 2 is given in Appendix D.1.

Remark 13. 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 14. 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 [GGNY14]):

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

Similarly, for the k-letter Marton we have

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

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*

$$\alpha^*(p_{\hat{Y}|X} p_{Y|\hat{Y}}) = \sup_{p(x):I(\hat{Y};Y)\neq 0} \frac{I(X;\hat{Y})}{I(\hat{Y};Y)} = \sup_{p(x):I(\hat{Y};Y)\neq 0} \frac{H(\hat{Y})}{I(\hat{Y};Y)}. \tag{25}$$

  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

$$p(x) \mapsto \max_{p(v|x)} \Big\{ -(\alpha-\lambda^*)H(Y) - \lambda^* H(Z) + \alpha I(\hat{Y};Y) + I(V;Z) - I(V;\hat{Y}) \Big\}. \tag{26}$$

is concave. In other words, concavity of the function in (26) 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

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

Equivalently, we should prove that

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

We have

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

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

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

The result for the k-letter Marton follows from the tensorization property of α^* given in Lemma 3.

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

$$\min_{\lambda\in[0,1]} \max_{p(v,w,x)} \Big\{ (\alpha-\lambda)I(W;Y) + \lambda I(W;Z) + \alpha I(\hat{Y};Y|W) + I(V;Z|W) - I(\hat{Y};V|W) \Big\} \tag{27}$$

$$\ge \max_{p(v,x)} \Big\{ \alpha I(\hat{Y};Y) + I(V;Z) - I(\hat{Y};V) \Big\}, \tag{28}$$

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

$$\begin{aligned}
&\min_{\lambda\in[0,1]} \max_{p(v,w,x)} \Big\{ (\alpha-\lambda)I(W;Y) + \lambda I(W;Z) + \alpha I(\hat{Y};Y|W) + I(V;Z|W) - I(\hat{Y};V|W) \Big\} \\
&\le \max_{p(v,w,x)} \Big\{ (\alpha-\lambda^*)I(W;Y) + \lambda^* I(W;Z) + \alpha I(\hat{Y};Y|W) + I(V;Z|W) - I(\hat{Y};V|W) \Big\} \\
&= \max_{p(x)} (\alpha-\lambda^*)H(Y) + \lambda^* H(Z) + \\
&\qquad \max_{p(w|x)} \Big\{ -(\alpha-\lambda^*)H(Y|W) - \lambda^* H(Z|W) + \alpha I(\hat{Y};Y|W) + \max_{p(v|w,x)} \Big\{ I(V;Z|W) - I(\hat{Y};V|W) \Big\} \Big\}.
\end{aligned} \tag{29}$$

The concavity property implies optimality of $W=\emptyset$ in (29).

□

Remark 15. 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 (26) holds if we choose $\lambda^* = 0.5$. On the other hand, if ϵ is larger, say larger than 0.631, then not only is the function given in (26) no longer concave, but simulation results also indicate that setting $W=\emptyset$ is not optimal when computing Marton's inner bound.

Lemma 3. *For given channels $p_{J|X}$ and $p_{Y|J}$, define*

$$\alpha^*(p_{J|X}, p_{Y|J}) = \sup_{p(x): I(J;Y)\neq 0} \frac{I(X;J)}{I(J;Y)}. \tag{30}$$

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

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

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

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

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

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

□

4.3. **The Second Outer Bound.** This outer bound was motivated directly from the outer bound for product broadcast channels developed in [GGNY14], which 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 becomes 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*

$$R_0 \leq \min\{I(W_b;J)+I(W_a;Y|J), I(W_b;Z|\hat{J})+I(W_a;\hat{J})\}, \tag{31a}$$

$$R_0+R_1 \leq I(U_b,W_b;J)+I(U_a,W_a;Y|J), \tag{31b}$$

$$R_0+R_1 \leq I(W_b;Z|\hat{J})+I(W_a,J;\hat{J})+I(U_b;J|W_b,\hat{J})+I(U_a;Y|W_a,J), \tag{31c}$$

$$R_0+R_2 \leq I(W_b,\hat{J};J)+I(W_a;Y|J)+I(V_b;Z|W_b,\hat{J})+I(V_a;\hat{J}|W_a,J), \tag{31d}$$

$$R_0+R_2 \leq I(V_a,W_a;\hat{J})+I(V_b,W_b;Z|\hat{J}), \tag{31e}$$

$$\begin{aligned} R_0+R_1+R_2 \leq{}& \min\{I(W_b,\hat{J};J)+I(W_a;Y|J),\ I(W_b;Z|\hat{J})+I(W_a,J;\hat{J})\} \\ &+ I(U_a;Y|W_a,J)+I(X;\hat{J}|U_a,W_a,J) \\ &+ \min\big\{I(U_b;J|W_b,\hat{J})+I(X;Z|U_b,W_b,\hat{J}),\ I(V_b;Z|W_b,\hat{J})+I(X;J|V_b,W_b,\hat{J})\big\}, \end{aligned} \tag{31f}$$

$$\begin{aligned} R_0+R_1+R_2 \leq{}& \min\{I(W_b,\hat{J};J)+I(W_a;Y|J),\ I(W_b;Z|\hat{J})+I(W_a,J;\hat{J})\} \\ &+ I(V_b;Z|W_b,\hat{J})+I(X;J|V_b,W_b,\hat{J}) \\ &+ \min\big\{I(U_a;Y|W_a,J)+I(X;\hat{J}|U_a,W_a,J),\ I(V_a;\hat{J}|W_a,J)+I(X;Y|V_a,W_a,J)\big\}, \end{aligned} \tag{31g}$$

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

$$\begin{aligned} &W_{ai} = (M_0, Y^{i-1}, \hat{J}_{i+1}^n, J^{n\setminus i}), \quad W_{bi} = (M_0, J^{i-1}, Z_{i+1}^n, \hat{J}^{n\setminus i}) \\ &U_a = U_b = M_1, \qquad V_a = V_b = M_2, \end{aligned} \tag{32}$$

where $J^{n\setminus i} = (J^{i-1}, J_{i+1}^n)$ and $\hat{J}^{n\setminus i} = (\hat{J}^{i-1}, \hat{J}_{i+1}^n)$.

The outer bound follows from routine manipulations using Lemma 5, guided by the above identification, starting from each of the following n-letter expressions which are reasonably straightforward to obtain using Fano's inequality (please see Appendix E.3):

$$nR_0 \leq \min\{I(M_0;J^n)+I(M_0;Y^n|J^n), I(M_0;Z^n|\hat{J}^n)+I(M_0;\hat{J}^n)\}+ng(\epsilon_n), \tag{33a}$$

$$n(R_0+R_1) \leq I(M_1,M_0;J^n)+I(M_1,M_0;Y^n|J^n)+ng(\epsilon_n), \tag{33b}$$

$$n(R_0+R_1) \leq I(M_0;Z^n|\hat{J}^n) + I(M_0,J^n;\hat{J}^n) + I(M_1;J^n|M_0,\hat{J}^n) + I(M_1;Y^n|M_0,J^n) + ng(\epsilon_n), \quad (33c)$$

$$n(R_0+R_2) \leq I(M_0,\hat{J}^n;J^n) + I(M_0;Y^n|J^n) + I(M_2;Z^n|M_0,\hat{J}^n) + I(M_2;\hat{J}^n|M_0,J^n) + ng(\epsilon_n), \quad (33d)$$

$$n(R_0+R_2) \leq I(M_2,M_0;\hat{J}^n) + I(M_2,M_0;Z^n|\hat{J}^n) + ng(\epsilon_n), \quad (33e)$$

$$\begin{aligned} n(R_0+R_1+R_2) \leq\ & \min\{I(M_0,\hat{J}^n;J^n) + I(M_0;Y^n|J^n), I(M_0;Z^n|\hat{J}^n) + I(M_0,J^n;\hat{J}^n)\} \\ & + I(M_1;Y^n|M_0,J^n) + I(X^n;\hat{J}^n|M_1,M_0,J^n) \\ & + \min\Big\{I(M_1;J^n|M_0,\hat{J}^n) + I(X^n;Z^n|M_1,M_0,\hat{J}^n), \\ & \qquad I(M_2;Z^n|M_0,\hat{J}^n) + I(X^n;J^n|M_2,M_0,\hat{J}^n)\Big\} + ng(\epsilon_n), \end{aligned} \quad (33f)$$

$$\begin{aligned} n(R_0+R_1+R_2) \leq\ & \min\{I(M_0,\hat{J}^n;J^n) + I(M_0;Y^n|J^n), I(M_0;Z^n|\hat{J}^n) + I(M_0,J^n;\hat{J}^n)\} \\ & + I(M_2;Z^n|M_0,\hat{J}^n) + I(X^n;J^n|M_2,M_0,\hat{J}^n) \\ & + \min\Big\{I(M_1;Y^n|M_0,J^n) + I(X^n;\hat{J}^n|M_1,M_0,J^n), \\ & \qquad I(M_2;\hat{J}^n|M_0,J^n) + I(X^n;Y^n|M_2,M_0,J^n)\Big\} + ng(\epsilon_n). \end{aligned} \quad (33g)$$

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

Remark 16. An alternative approach to single-letterize (33a)-(33g) that skips using Lemma 5 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

$$I(M_0;Y^n) + I(M_1;Y^n|M_0) + I(M_2;Z^n|M_0,M_1) \leq \sum_i \left(I(W_i;Y_i) + I(U_i;Y_i|W_i) + I(X_i;Z_i|U_i,W_i)\right), \quad (34)$$

where $W_i = (M_0, Y^{i-1}, Z_{i+1}^n)$ and $U_i = M_1$. The inequality (34) 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 (32). 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 (34) 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 (33a)-(33g) 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 8 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 17. 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 than $1-\epsilon$ 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). As mentioned in the introduction, 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 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 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. 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 cutset 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 cutset bound. The authors would also like to thank Yan Nan Wang for doing the initial simulations concerning the outer bound and Marton's inner bound for the erased Blackwell channel.

The authors wish to thank Max Costa, Abbas El Gamal, Cheuk Ting Li, Yan Nan Wang, and Mohamamd Hossein Yassaee for comments on various aspects of the manuscript. The authors also wish to thank the anonymous reviewers for making several suggestions that greatly enhanced the presentation of this paper.

References

[AGS05] L. Ambrosio, N. Gigli, and G. Savare, *Gradient flows: In metric spaces and in the space of probability measures*, Lectures in Mathematics. ETH Zürich, Birkhäuser Basel, 2005.

[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.

[BSC14] A. Bennatan, S. Shamai, and A. R. Calderbank, *Soft-decoding-based strategies for relay and interference channels: Analysis and achievable rates using ldpc codes*, IEEE Transactions on Information Theory **60** (2014), no. 4, 1977–2009.

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

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

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

[CK11] ———, *Information theory: Coding theorems for discrete memoryless systems*, Cambridge University Press, 1 2011.

[CM11] H. Chong and M. Motani, *On achievable rates for the general relay channel*, IEEE Transactions on Information Theory **57** (2011), no. 3, 1249–1266.

[CN16] M. Costa and C. Nair, *Gaussian z-interference channel: around the corner*, Information Theory and Applications Workshop (ITA) (2016), 1–6.

[Cos85a] M. Costa, *A new entropy power inequality*, IEEE Transactions on Information Theory **31** (1985), no. 6, 751–760.

[Cos85b] ———, *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] A. De, E. Mossel, and J. 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.

[EGGN21a] A. El Gamal, A. Gohari, and C. Nair, *A strengthened cutset upper bound on the capacity of the relay channel and applications*, arXiv preprint arxiv:2101.11139 (2021).

[EGGN21b] Abbas El Gamal, Amin Gohari, and Chandra Nair, *Achievable rates for the relay channel with orthogonal receiver components*, IEEE Information Theory Workshop (2021).

[EK12] A. El Gamal and Y.-H. 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.

[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] B. Ghazi, P. Kamath, and M. 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.
[GN20] A. Gohari and C. Nair, *New outer bounds for the two-receiver broadcast channel*, 2020 IEEE International Symposium on Information Theory (ISIT), 2020, pp. 1492–1497.
[GNN21] A. Gohari, C. Nair, and D. Ng, *An information inequality motivated by the gaussian z-interference channel*, Proc. of IEEE ISIT, 2021.
[HK81] T. S. Han and K. 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] W. Kang and S. 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.
[KW10] A. Khisti and G. W. Wornell, *Secure transmission with multiple antennas i: The misome wiretap channel*, IEEE Transactions on Information Theory **56** (2010), no. 7, 3088–3104.
[LG09] N. Liu and A. Goldsmith, *Capacity regions and bounds for a class of z-interference channels*, IEEE transactions on information theory **55** (2009), no. 11, 4986–4994.
[Li21] Cheuk Ting Li, *An automated theorem proving framework for information-theoretic results*, 2021 IEEE International Symposium on Information Theory (ISIT), 2021, pp. 2750–2755.
[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] T. 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.
[Mat] MathOverflow, *Double markovity*, https://mathoverflow.net/questions/177312/double-markovity.
[Nai11] C. 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] Y. Polyanskiy and Y. Wu, *Wasserstein continuity of entropy and outer bounds for interference channels*, CoRR **abs/1504.04419** (2015), 3992–4002.
[Sas04] I. Sason, *On achievable rate regions for the gaussian interference channel*, IEEE transactions on information theory **50** (2004), no. 6, 1345–1356.
[Sas15] ———, *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.
[Sat81] H. Sato, *The capacity of the gaussian interference channel under strong interference (corresp.)*, IEEE Transactions on Information Theory **27** (1981), no. 6, 786–788.
[Tal96] M. Talagrand, *Transportation cost for gaussian and other product measures*, Geometric & Functional Analysis GAFA **6** (1996), no. 3, 587–600.
[TU08] R. Tandon and S. Ulukus, *A new upper bound on the capacity of a class of primitive relay channels*, 2008 46th Annual Allerton Conference on Communication, Control, and Computing, IEEE, 2008, pp. 1562–1569.
[WA08] A. B. Wagner and V. 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] S. Watanabe and Y. 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.
[WOX17] X. Wu, A. Ozgur, and L. Xie, *Improving on the cut-set bound via geometric analysis of typical sets*, IEEE Transactions on Information Theory **63** (2017), no. 4, 2254–2277.

[WSS06] H. Weingarten, Y. Steinberg, and S. Shamai, *The capacity region of the gaussian multiple-input multiple-output broadcast channel*, IEEE transactions on information theory **52** (2006), no. 9, 3936–3964.

[Xue14] F. Xue, *A new upper bound on the capacity of a primitive relay channel based on channel simulation*, IEEE transactions on information theory **60** (2014), no. 8, 4786–4798.

[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.

[Zha88] Z. Zhang, *Partial converse for a relay channel*, IEEE Transactions on Information Theory **34** (1988), no. 5, 1106–1110.

Appendix A. Preliminaries

The following well-known result has be used repeatedly in the proofs in this paper.

Lemma 4 (Körner-Márton Lemma, (4.14) in [KM77])**.** *For any tuple of random variables (U, Y^n, Z^n) the following equality holds:*

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

Remark 18*.* 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 5 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.

Motivated by the above lemma, the generic manipulations that are being used in the converses are the following.

Lemma 5. *For any set of random variables (U, V, A^n, B^n) the following hold:*

$$I(U; A^n|V) - I(U; B^n|V) = \sum_{i=1}^n \Big(I(U; A_i|V, B_{i+1}^n, A^{i-1}) - I(U; B_i|V, B_{i+1}^n, A^{i-1}) \Big) \tag{35}$$

$$= \sum_{i=1}^n \Big(I(U; A_i|V, A_{i+1}^n, B^{i-1}) - I(U; B_i|V, A_{i+1}^n, B^{i-1}) \Big),$$

$$I(U; B^n|V) + I(U; A^n|V) - I(U; B^n|V) \leq \sum_{i=1}^n \Big(I(U, B_{i+1}^n; B_i|V) + I(U, B_{i+1}^n, A^{i-1}; A_i|V) - I(U, B_{i+1}^n, A^{i-1}; B_i|V) \Big) \tag{36}$$

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

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

$$= \sum_{i=1}^n \Big(I(U, B^{i-1}; B_i|V) + I(U, A_{i+1}^n, B^{i-1}; A_i|V) - I(U, A_{i+1}^n, B^{i-1}; B_i|V) \Big),$$

$$I(U; A^n|V) + I(V; B^n) \leq \sum_{i=1}^n \Big(I(U; A_i|V, A_{i+1}^n, B^{i-1}) + I(V, A_{i+1}^n, B^{i-1}; B_i) \Big), \tag{37}$$

$$I(U; A^n|V) + I(V; B^n) \leq \sum_{i=1}^n \Big(I(U; A_i|V, B_{i+1}^n, A^{i-1}) + I(V, B_{i+1}^n, A^{i-1}; B_i) \Big).$$

Proof. The proof follows immediately from repeated applications of Lemma 4, chain rule for mutual information, and non-negativity of mutual information. The details are omitted as they are standard in literature. For instance, (35) can be shown by writing the telescopic sum

$$I(U;A^n|V) - I(U;B^n|V) = \sum_{i=1}^{n} \left(I(U;B_{i+1}^n, A^i|V) - I(U;B_i^n, A^{i-1}|V) \right).$$

To get (36), we use (35) as follows

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

The proof for (37) can be found at [EK12, p. 226] in the proof of the UV outer bound for a general broadcast channel.

□

Lemma 6. *Double Markovity [CK11, Exercise 16.25] Let U, X, Y be three (real-valued) random variables defined on the same probability space, such that both the Markov chains $U \to X \to Y$ and $U \to Y \to X$ hold. Then*

(1) There exists functions $f(X)$ and $g(Y)$ such that $f(X) = g(Y)$ with probability one.
(2) $U \to f(X) \to (X, Y)$ is Markov.

Proof. Disclaimer: We are only providing an outline of an argument to solve the Exercise for completeness. This argument is a generalization of an argument in the discrete case and follows along the lines of [Mat].

Let $F_{U|X}, F_{U|Y}$ denote the (regular) conditional distributions of U conditioned on X and Y respectively. We define an equivalence class according to $x_1 \equiv x_2$, if the conditional distributions satisfy $F_{U|X=x_1} = F_{U|X=x_2}$. This defines a partition of X. Let us similarly define an equivalence class and a partition of the values of Y.

Now define $f(X)$, a measurable function, such that it takes the same value in a partition and differs across partitions. Thus there is a bijection between $f(X)$ and the conditional distributions $F_{U|X}$. This yields $F_{U|X} = F_{U|f(X)}$.

The Markov chains imply that $F_{U|X,Y} = F_{U|X}$ with probability one, and $F_{U|X,Y} = F_{U|Y}$ with probability one. Therefore we have $F_{U|X} = F_{U|Y} = F_{U|X,Y}$ with probability one. On this set define $g(Y)$ to take the same value as $f(X)$. Clearly, by construction $g(Y) = f(X)$ with probability one. Further it is also clear by construction that $F_{U|X,Y,f(X)} = F_{U|X,Y} = F_{U|X} = F_{U|f(X)}$ with probability one. Thus we also have $U \to f(X) \to (X, Y)$ is Markov.

Remark 19. It is worth noting that if support of (X, Y) is the product of the support of X and the support of Y, then $f(X)$ and $g(Y)$ have to be constants, implying that U is independent of X, Y.

□

Appendix B. Proofs of Propositions for scalar Gaussian Relay channels

B.1. Proof of Proposition 1.

Proof. Remember that

$$Y_{\mathrm{r}} = g_{21}X + Z_1, \quad Y = g_{31}X + g_{32}X_{\mathrm{r}} + Z_2.$$

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

$$\begin{aligned} R \le \max_{p(X,X_{\mathrm{r}})\in\mathcal{P}} \min\Big(& I(X,X_{\mathrm{r}};Y),\ I(X;Y,Y_{\mathrm{r}}|X_{\mathrm{r}}), \\ & \min_{\alpha\in(0,1]} \max_{W:W \multimap (X,X_{\mathrm{r}}) \multimap (Y,Y_{\mathrm{r}},J)} I(X;J,Y_{\mathrm{r}}|X_{\mathrm{r}}) + I(X,X_{\mathrm{r}};Y|W) - I(X;J|W)\Big), \end{aligned} \tag{38}$$

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

$$K_{X,X_{\mathrm{r}}} \preceq \begin{bmatrix} P & \rho P \\ \rho P & P \end{bmatrix} =: K_\rho, \tag{39}$$

where $K_{X,X_{\mathrm{r}}} \preceq K_\rho$ stands for $K_\rho - K_{X,X_{\mathrm{r}}}$ being a positive semi-definite matrix.

Elementary facts about Gaussian extremality[4] for (conditional) differential entropy with respect to a covariance constraint shows that if (39) holds then

$$I(X,X_{\mathrm{r}};Y) \le \mathsf{C}(S_{31} + S_{32} + 2\rho\sqrt{S_{31}S_{32}}), \tag{40}$$

$$I(X;Y,Y_{\mathrm{r}}|X_{\mathrm{r}}) \le \mathsf{C}((1-\rho^2)(S_{31} + S_{21})), \tag{41}$$

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

With K_ρ defined as in (39), note that

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

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 $\left(\frac{g_{21}}{\alpha} - f g_{31}\right) X - f g_{32} X_{\mathrm{r}} = 0$ 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

$$1 - a - b \ge \rho^2 \mp 2\rho\sqrt{ab} \tag{42}$$

Observe that, when K' is defined as above

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

[4]If X, Y are random vectors with a block covariance matrix given by $K := \begin{bmatrix} K_X & K_{X,Y} \\ K_{X,Y}^T & K_Y \end{bmatrix}$; then $h(X,Y)$ and $h(X|Y)$ are maximized when $(X,Y) \sim \mathcal{N}(0,K)$. One way to observe the second part is that

$$h(X|Y) = h(X - K_{X,Y}K_Y^{-1}Y|Y) \le h(X - K_{X,Y}K_Y^{-1}Y) \le \frac{1}{2}\log\left((2\pi e)^n \left|K_X - K_{X,Y}K_Y^{-1}K_{X,Y}^T\right|\right).$$

Optimizing the term above with respect to b for a fixed a subject to (42) we obtain that

$$\begin{aligned}&\max_{\substack{p_{X,X_{\mathrm{r}}}\sim\mathcal{N}(0,K')\\ K'\preceq K_\rho}} I(X,X_{\mathrm{r}};Y)-I(X;J)\\ &\quad=\max_{0\le a\le 1}\frac{1}{2}\log\left(1+\left(\sqrt{aS_{31}}+\sqrt{\rho^2 aS_{32}}+\sqrt{(1-\rho^2)(1-a)S_{32}}\right)^2\right)-\frac{1}{2}\log(1+aS_J). \qquad (43)\end{aligned}$$

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:

$$\begin{aligned}&x^2(\sqrt{S_{31}}+\sqrt{\rho^2 S_{32}})(\sqrt{(1-\rho^2)S_{32}})(1+S_J)+x\left((1-\rho^2)S_{32}(1+S_J)-(\sqrt{S_{31}}+\sqrt{\rho^2 S_{32}})^2+S_J\right)\\ &\quad-(\sqrt{S_{31}}+\sqrt{\rho^2 S_{32}})(\sqrt{(1-\rho^2)S_{32}})=0. \qquad (44)\end{aligned}$$

This completes the proof. ☐

Remark 20. From the above argument, we know that

$$\begin{aligned}&\max_{\substack{p_{X,X_{\mathrm{r}}}\sim\mathcal{N}(0,K')\\ K'\preceq K_\rho}} I(X,X_{\mathrm{r}};Y)-I(X;J)\\ &=\max_{\substack{p_{X,X_{\mathrm{r}}}\sim\mathcal{N}(0,K_\rho)\\ W\multimap(X,X_{\mathrm{r}})\multimap(Y,J)}} I(X,X_{\mathrm{r}};Y|W)-I(X,X_{\mathrm{r}};J|W)\\ &\le\max_{(U,W)\multimap(X,X_{\mathrm{r}})\multimap(Y,J)} I(U;Y|W)-I(U;J|W) \qquad (45)\\ &\overset{(a)}{=}\frac{1}{2}\sum_i[\log\phi_i]_+,\end{aligned}$$

where (a) follows from, for instance, [WO11, Eq. 7] (this, in turn, utilizes the enhancement technique of [WSS06] for the MIMO Gaussian broadcast channel). Here $[x]_+$ is zero if x is negative and x otherwise, and ϕ_i are the set of generalized eigenvalues for the pencil

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

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

$$2\rho\sqrt{S_{31}S_{32}}+S_{31}+S_{32}+1-\lambda\big(S_{32}S_J(1-\rho^2)+S_{31}+S_{32}+S_J+2+2\rho\sqrt{S_{31}S_{32}}\big)+\lambda^2(S_J+1)=0. \qquad (46)$$

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

$$\max_{(U,W)\multimap(X,X_{\mathrm{r}})\multimap(Y,J)} I(U;Y|W)-I(U;J|W)=\frac{1}{2}\ln\lambda_{\max}.$$

On the other hand, routine calculation shows that after substituting the unique non-negative root of (44) in (43) and simplifying the expression, we have

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

Thus, for this setting, the inequality in (45) is indeed an equality. We remark that the optimality of $U=(X,X_{\mathrm{r}})$ in (45) is similar to the observation made in [KW10] for the problem of secure transmission over a Gaussian wiretap channel with multiple antennas.

Theorem 9 (Theorem 8 in [LV07], see also Theorem 1 in [GN14])**.** *Let $\mathbf{Z}_1,\mathbf{Z}_2$ be two Gaussian vectors 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*

$$\max_{p(\mathbf{x},w)} h(\mathbf{X}+\mathbf{Z}_1|W)-\mu h(\mathbf{X}+\mathbf{Z}_2|W)$$

$$\text{subject to } Cov(\mathbf{X}) \preceq S$$

where the maximization is over $p(w, \mathbf{x})$ *such that* $Cov(\mathbf{X}) \preceq S$*, 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 21. In [LV07], the constraint for the optimization problem is listed as $\mathrm{Cov}(\mathbf{X}|W) \preceq S$, 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 $\mathbf{Y} = A\mathbf{X} + B\mathbf{X}_{\mathrm{r}} + \mathbf{Z}$ *and* $\mathbf{J} = C\mathbf{X} + \mathbf{Z}$ *for some vectors* $\mathbf{X}, \mathbf{X}_{\mathrm{r}}, \mathbf{Y}$ *and* $\mathbf{J}$*, and a Gaussian random vector* $\mathbf{Z}$ *independent of* $(\mathbf{X}, \mathbf{X}_{\mathrm{r}})$ *and distributed as* $\mathcal{N}(0, I)$*. Let* K_0 *be an arbitrary symmetric positive semi-definite matrix.*

- (i) *To compute the maximum over* $p(w, \mathbf{x}, \mathbf{x}_{\mathrm{r}})$*, subject to* $K_{\mathbf{X},\mathbf{X}_{\mathrm{r}}} \preceq K_0$ *and* $W \multimap (\mathbf{X}, \mathbf{X}_{\mathrm{r}}) \multimap (\mathbf{Y}, \mathbf{J})$*, of the expression*
$$I(\mathbf{X}; \mathbf{J}|\mathbf{X}_{\mathrm{r}}) + I(\mathbf{X}, \mathbf{X}_{\mathrm{r}}; \mathbf{Y}|W) - I(\mathbf{X}, \mathbf{X}_{\mathrm{r}}; \mathbf{J}|W)$$
it suffices to assume that $[\mathbf{X}\ \mathbf{X}_{\mathrm{r}}]^T = U + W$ *where* U *and* W *are independent, and* $U \sim \mathcal{N}(0, K')$*,* $W \sim \mathcal{N}(0, K_0 - K')$ *for some* $K' \preceq K_0$*.*
- (ii) *Further, let* $[\mathbf{X}^*, \mathbf{X}_{\mathrm{r}}^*] \sim \mathcal{N}(0, K')$ *be the maximizer mentioned above. Then* $(C - FA)\mathbf{X}^* - FB\mathbf{X}_{\mathrm{r}}^* = 0$ *for some matrix* 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 $\epsilon \neq 0$. This follows from the assumption that A_e and C_e are invertible. Let $\tilde{\mathbf{Z}} \sim \mathcal{N}(0, I)$ be independent of $\mathbf{X}, \mathbf{X}_{\mathrm{r}}$ and $\mathbf{Z}$. Define $\hat{\mathbf{Z}} = [\mathbf{Z}\ \tilde{\mathbf{Z}}]^T$. Define $\hat{\mathbf{X}} := [\mathbf{X}\ \mathbf{X}_{\mathrm{r}}]^T$, $\hat{\mathbf{Y}}_\epsilon := A_\epsilon\hat{\mathbf{X}} + \hat{\mathbf{Z}}$ and $\hat{\mathbf{J}}_\epsilon := C_\epsilon\hat{\mathbf{X}} + \hat{\mathbf{Z}}$. Consider the expression

$$I(\mathbf{X}; \hat{\mathbf{J}}_\epsilon|\mathbf{X}_{\mathrm{r}}) + I(\hat{\mathbf{X}}; \hat{\mathbf{Y}}_\epsilon|W) - I(\hat{\mathbf{X}}; \hat{\mathbf{J}}_\epsilon|W). \tag{47}$$

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 $[\mathbf{X}\ \mathbf{X}_{\mathrm{r}}]^T = U_\epsilon + W_\epsilon$, maximizes the expression in (47) over $p(w, \mathbf{x}, \mathbf{x}_{\mathrm{r}})$, subject to $K_{\mathbf{X},\mathbf{X}_{\mathrm{r}}} \preceq K_0$. For any $\mathbf{X}, \mathbf{X}_{\mathrm{r}}$ such that $K_{\mathbf{X},\mathbf{X}_{\mathrm{r}}} \preceq K_0$, observe the following:

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

Similarly, mimicking the steps as above,

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

Therefore as $\epsilon \to 0$, the maximum of the expression in (47) 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 $[\mathbf{X}\ \mathbf{X}_{\mathrm{r}}]^T = U + W$, for some $K' \preceq K_0$. This completes the proof of (i).

Let $[\mathbf{X}^* \ \mathbf{X}_{\mathrm{r}}^*]^T \sim \mathcal{N}(0, K')$ be a maximizer. Let $\mathbf{Y}^* := A\mathbf{X}^* + B\mathbf{X}_{\mathrm{r}}{}^* + \mathbf{Z}$ and $\mathbf{J}^* := C\mathbf{X}^* + \mathbf{Z}$ where $[\mathbf{X}^*, \mathbf{X}_{\mathrm{r}}^*]$ is independent of $\mathbf{Z}$. Then from the optimality of $[\mathbf{X}^*, \mathbf{X}_{\mathrm{r}}^*]$ we have that

$$I(\mathbf{X}^*, \mathbf{X}_{\mathrm{r}}{}^*; \mathbf{Y}^*) - I(\mathbf{X}^*, \mathbf{X}_{\mathrm{r}}{}^*; \mathbf{J}^*) \geq I(\mathbf{X}^*, \mathbf{X}_{\mathrm{r}}{}^*; \mathbf{Y}^*|\tilde{W}) - I(\mathbf{X}^*, \mathbf{X}_{\mathrm{r}}{}^*; \mathbf{J}^*|\tilde{W}) \tag{48}$$

for any $\tilde{W}$ such that $(\tilde{W}, \mathbf{X}^*, \mathbf{X}_{\mathrm{r}}{}^*)$ is independent of $\mathbf{Z}$. Express $C\mathbf{X}^* = F(A\mathbf{X}^* + B\mathbf{X}_{\mathrm{r}}{}^*) + \tilde{\mathbf{X}}$, where $\tilde{\mathbf{X}}$ is independent of $A\mathbf{X}^* + B\mathbf{X}_{\mathrm{r}}{}^*$ (such a decomposition is feasible for jointly Gaussian vectors).[5] Since $\tilde{\mathbf{X}}$ is a function of $\mathbf{X}^*$ and $\mathbf{X}_{\mathrm{r}}{}^*$, it is independent of $\mathbf{Z}$. Let $\tilde{W} = \tilde{\mathbf{X}}$. We obtain that $\tilde{W}$ and $\mathbf{Y}^*$ are independent. Consequently, $I(\mathbf{X}^*, \mathbf{X}_{\mathrm{r}}{}^*; \mathbf{Y}^*) = I(\mathbf{X}^*, \mathbf{X}_{\mathrm{r}}{}^*; \mathbf{Y}^*|\tilde{W})$ and using (48), we deduce that $\tilde{W}$ $(= \tilde{\mathbf{X}})$ and $\mathbf{J}^*$ are independent. Consequently, $\tilde{\mathbf{X}}$ and $\mathbf{J}^* = C\mathbf{X}^* + \mathbf{Z}$ are uncorrelated, implying that $\tilde{W} = \tilde{\mathbf{X}}$ and $C\mathbf{X}^*$ are uncorrelated (hence independent). Since $C\mathbf{X}^* = F(A\mathbf{X}^* + B\mathbf{X}_{\mathrm{r}}{}^*) + \tilde{\mathbf{X}}$ and $A\mathbf{X}^* + B\mathbf{X}_{\mathrm{r}}{}^*$ are uncorrelated with $\tilde{\mathbf{X}}$, one obtains that $\mathrm{E}(\tilde{\mathbf{X}}(\tilde{\mathbf{X}})^T) = 0$, showing that $\tilde{\mathbf{X}} = 0$ almost surely. Thus, we have $(C - FA)\mathbf{X}^* - FB\mathbf{X}_{\mathrm{r}}^* = 0$ 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.2) by

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

where the bound in (a) again follows from routine algebra given in Appendix E.2. 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 $\alpha, \beta \in [0, 1]$,

$$\begin{aligned}
R_1 + \lambda R_2 \leq \sup_{p_Q p_{X_1|Q} p_{X_2|Q} p_{W|X_1,X_2,Q}} \Big[&\lambda\alpha I(X_2; Y_2|W, X_1, Q) + \lambda(1-\alpha)(I(X_2; Y_2|W, Q) - I(X_2; J|W, Q))\\
&+ \beta I(X_1; Y_1|Q) + (1-\beta)(I(W; Y_2|Q) + I(X_1; J|W, Q))\Big],
\end{aligned} \tag{49}$$

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

$$\mathbf{Y}_2^{(\epsilon)} := \begin{bmatrix} a & 1 \\ \epsilon & \epsilon \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} auZ_1 + \sqrt{1 - a^2u^2}Z_2 \\ Z_3 \end{bmatrix},$$

[5]For every two jointly Gaussian random vectors $\mathbf{S}$ and $\mathbf{V}$, one can decompose $\mathbf{S}$ as $\mathbf{S} = F\mathbf{V} + \mathbf{R}$ for some matrix F and random vector $\mathbf{R}$ such that $\mathbf{V}$ is independent of $\mathbf{R}$. Without loss of generality assume that $\mathbf{V}$ has an invertible covariance matrix. Then $F = \mathbb{E}(\mathbf{S}\mathbf{V}^T)\mathbb{E}(\mathbf{V}\mathbf{V}^T)^{-1}$ yields the desired decomposition. Please see Theorem 2.3 in https://web.mit.edu/gallager/www/papers/chap2.pdf

$$\mathbf{J}^{(\epsilon)} := \begin{bmatrix} 1 & 0 \\ \epsilon & \epsilon \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} uZ_1 \\ Z_3 \end{bmatrix},$$
$$\hat{\mathbf{J}}^{(\epsilon)} := \begin{bmatrix} \epsilon a & \epsilon \\ \epsilon & \epsilon \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} Z_4 \\ Z_3 \end{bmatrix},$$
$$\bar{\mathbf{J}}^{(\epsilon)} := \begin{bmatrix} \epsilon & 0 \\ \epsilon & \epsilon \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} Z_5 \\ Z_3 \end{bmatrix},$$

where Z_i's are mutually independent standard Gaussian random variables, independent of (X_1, X_2). The construction ensures that $\hat{\mathbf{J}}^{(\epsilon)}$ is a stochastically degraded version of $\mathbf{Y}_2^{(\epsilon)}$ for any $\epsilon < 1$, and that $\mathbf{J}^{(\epsilon)} \multimap (X_1, \hat{\mathbf{J}}^{(\epsilon)}) \multimap X_2$ is Markov. For simplicity, we sometimes drop the (ϵ) and simply write $\hat{\mathbf{J}}$ instead of $\hat{\mathbf{J}}^{(\epsilon)}$ (and similarly for other variables).

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

$$\lambda\alpha I(X_2; \mathbf{Y}_2|W, X_1, Q) + \lambda(1-\alpha)\big[I(X_2; \mathbf{Y}_2|W, Q) - I(X_2; \mathbf{J}|W, Q)\big] \\ + \beta I(X_1; Y_1|Q) + (1-\beta)\big[I(W; \mathbf{Y}_2|Q) + I(X_1; \mathbf{J}|W, Q)\big] + T_{\epsilon,\gamma} \tag{50}$$

where $T_{\epsilon,\gamma}$ is a perturbation term, to the expression in (49), defined by

$$T_{\epsilon,\gamma} := (1-\beta)I(X_2; \bar{\mathbf{J}}|W, X_1, Q) - \lambda\alpha I(X_2; \hat{\mathbf{J}}|W, X_1, Q) + \epsilon I(X_1, X_2; \mathbf{Y}_2|W, Q) - \gamma\epsilon I(X_1, X_2; \mathbf{J}|W, Q).$$

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.[6]

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, $\mathbf{J}_- \multimap (X_{1-}, \hat{\mathbf{J}}_-) \multimap X_{2-}$ is Markov. Further as Z's are Gaussian random variables, under this transformation, $Z_{i,+}$ and $Z_{i,-}$, $1 \leq i \leq 5$, are independent of each other and satisfy the same independence relationship to other variables as in the original setting.

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

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

[6] In the case of the Gaussian Z-interference channel, routine arguments will show that there is a maximizer - the power constraints yield tightness (for instance, Proposition 17 and Theorem 4 in [GN14]), and the additive Gaussian noise yields the continuity (Proposition 18 in [GN14]) for the various terms with respect to weak convergence. The interested readers can also look into Section 5.2 of [AGS05].

$$
\begin{aligned}
&\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)\big[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)\\
&\qquad\qquad + I(X_{2+};\bar{\mathbf{J}}_+|W_a,W_b,X_{1+},\mathbf{J}_-,Q_a,Q_b)\big]\\
&\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)\big[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+})\\
&\qquad\qquad + I(X_{2-};\bar{\mathbf{J}}_-|W_a,W_b,X_{1-},Q_a,Q_b,\mathbf{Y}_{2+})\big]\\
&\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-},\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. The detailed justification of going from (a) to (b) is given in Appendix E.1. Since $\hat{\mathbf{J}}_+$ ($\hat{\mathbf{J}}_-$) is a stochastically degraded version of $\mathbf{Y}_{2+}$ ($\mathbf{Y}_{2-}$), we have that the expressions within the curly braces above satisfy

$$
\begin{aligned}
I(X_{1-},\hat{\mathbf{J}}_-;\mathbf{Y}_{2+}|W_a,W_b,X_{1+},\mathbf{J}_-,Q_a,Q_b) - I(X_{1-},\hat{\mathbf{J}}_-;\hat{\mathbf{J}}_+|W_a,W_b,X_{1+},\mathbf{J}_-,Q_a,Q_b) &\geq 0,\\
I(X_{1+};\mathbf{Y}_{2-}|W_a,W_b,X_{1-},Q_a,Q_b,\mathbf{Y}_{2+}) - I(X_{1+};\hat{\mathbf{J}}_{2-}|W_a,W_b,X_{1-},Q_a,Q_b,\mathbf{Y}_{2+}) &\geq 0. \qquad (51)
\end{aligned}
$$

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)

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

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) uses the fact that $\mathbf{J}_+ \multimap (\hat{\mathbf{J}}_+,X_{1+}) \multimap (W_a,W_b,Q_a,Q_b,\mathbf{J}_-,X_{1-})$ 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 (51) 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 last 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 (16), we obtain that

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

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

$$(P_1-K_1)(P_2-K_2)\geq\rho^2K_1K_2. \tag{52}$$

We choose $u=1$. Note that the expression is increasing in K_2; and hence we fix K_1,ρ and substitute for the maximal K_2 satisfying (52) to obtain

$$\begin{aligned}&\frac{\lambda_1}{2}\log(K_2(1-\rho^2)+1)+\frac{1}{2}\log\left(1+a^2P_1+P_2\right)\\&\quad+\frac{\lambda_2-1}{2}\log\left(\frac{1+a^2K_1+K_2+2a\rho\sqrt{K_1K_2}}{K_1+1}\right)+\frac{\lambda_2}{2}\log\left(\frac{K_1(1-\rho^2)+1}{a^2K_1(1-\rho^2)+1}\right)\\&\leq\frac{\lambda_1}{2}\log\left(\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}(1-\rho^2)+1\right)+\frac{1}{2}\log\left(1+a^2P_1+P_2\right)\\&\quad+\frac{\lambda_2-1}{2}\log\left(\frac{1+a^2K_1+\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}+2a\rho\sqrt{K_1\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}}}{K_1+1}\right)+\frac{\lambda_2}{2}\log\left(\frac{K_1(1-\rho^2)+1}{a^2K_1(1-\rho^2)+1}\right).\end{aligned}$$

To prove our result, note that it suffices to show that the maximum over K_1,K_2,ρ is upper bounded by $\frac{1}{2}\log(1+a^2P_1+P_2)+\frac{(\lambda_1+\lambda_2-1)}{2}\log(1+P_2)$. Equivalently, we desire to show:

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

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

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

$$
\begin{aligned}
&-\lambda_1 \left(\frac{P_2P_1\rho^2}{(P_1-K_1+\rho^2K_1)(1+P_2)}\right) + \left(\frac{K_1(1-a^2)(1-\rho^2)}{a^2K_1(1-\rho^2)+1}\right)\\
&\quad + (\lambda_2-1)\left(\frac{(1 + a^2K_1 + \frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1} + 2a\rho\sqrt{K_1\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}})(1+K_1(1-\rho^2))}{(K_1+1)(a^2K_1(1-\rho^2)+1)(1+P_2)} - 1\right)\\
&\le 0.
\end{aligned}
$$

Since

$$
\begin{aligned}
2a\rho\sqrt{K_1\frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1}} \le{}& \frac{\lambda_1}{\lambda_2-1}\left(\frac{P_2P_1\rho^2}{(P_1-K_1+\rho^2K_1)}\right)\frac{(K_1+1)(a^2K_1(1-\rho^2)+1)}{(1+K_1(1-\rho^2))}\\
&+ \frac{\lambda_2-1}{\lambda_1}\frac{(1+K_1(1-\rho^2))}{(K_1+1)(a^2K_1(1-\rho^2)+1)}\frac{a^2K_1((P_1-K_1))}{P_1}
\end{aligned}
$$

it suffices that

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

This is equivalent to

$$
\begin{aligned}
&\left(K_1(1-a^2)(1-\rho^2)\right) \qquad (53)\\
&\le (\lambda_2-1)\left(a^2K_1(1-\rho^2)+1 - \frac{(1 + a^2K_1 + \frac{P_2(P_1-K_1)}{P_1-K_1+\rho^2K_1} + \frac{\lambda_2-1}{\lambda_1}\frac{(1+K_1(1-\rho^2))}{(K_1+1)(a^2K_1(1-\rho^2)+1)}\frac{a^2K_1((P_1-K_1))}{P_1})(1+K_1(1-\rho^2))}{(K_1+1)(1+P_2)}\right).
\end{aligned}
$$

This can be rewritten as

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

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

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

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

$$\begin{aligned}\big((1-a^2)\big) &\leq (\lambda_2-1)\left(\frac{a^2P_2}{(1+P_2)}\right)\\ &\quad -\frac{(\lambda_2-1)^2}{\lambda_1P_1(1+P_2)}\times\begin{cases}\frac{(P_1+a^2)^2}{4(1+P_1)(1-a^2)} & \text{if } P_1>\frac{a^2}{1-2a^2}, a^2<\frac{1}{2}\\ a^2P_1 & \text{otherwise}\end{cases}.\end{aligned}$$

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

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

Therefore we clearly need

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

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

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

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

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

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

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

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 2.** 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

$$\frac{1}{2}(1-\epsilon) \leq I(V;Z|W) - I(V;\hat{Y}|W) \tag{54}$$

$$1 - \epsilon = I(\hat{Y};Y) \leq I(W;Z) + I(X;\hat{Y}|W) = I(W;Z) + H(\hat{Y}|W). \tag{55}$$

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 (55), we obtain that $1-\epsilon \leq I(W;Z) = (1-\epsilon)I(W;\hat{Z}) \leq (1-\epsilon)H(\hat{Z}) = (1-\epsilon)H_2\left(\frac{1}{4}\right)$, a contradiction.

Appendix E. Routine Calculations

E.1. **Justification of equalities in the proof of Theorem 3.** Consider the first equality:

$$\begin{aligned} 2V &= \lambda\alpha I(X_{2a},X_{2b};\mathbf{Y}_{2a},\mathbf{Y}_{2b}|W_a,W_b,X_{1a},X_{1b},Q_a,Q_b) - \lambda\alpha I(X_{2a},X_{2b};\hat{\mathbf{J}}_a,\hat{\mathbf{J}}_b|W_a,W_b,X_{1a},X_{1b},Q_a,Q_b) \\ &\quad + \lambda(1-\alpha)I(X_{2a},X_{2b};\mathbf{Y}_{2a},\mathbf{Y}_{2b}|W_a,W_b,Q_a,Q_b) \\ &\quad - \lambda(1-\alpha)I(X_{2a},X_{2b};\mathbf{J}_a,\mathbf{J}_b|W_a,W_b,Q_a,Q_b) + \beta I(X_{1a},X_{1b};\mathbf{Y}_{1a},\mathbf{Y}_{1b}|Q_a,Q_b) \\ &\quad + (1-\beta)\big[I(W_a,W_b;\mathbf{Y}_{2a},\mathbf{Y}_{2b}|Q_a,Q_b) + I(X_{1a},X_{1b};\mathbf{J}_a,\mathbf{J}_b|W_a,W_b,Q_a,Q_b) \\ &\qquad\qquad + I(X_{2a},X_{2b};\bar{\mathbf{J}}_a,\bar{\mathbf{J}}_b|W_a,W_b,X_{1a},X_{1b},Q_a,Q_b)\big] \\ &\quad + \epsilon I(X_{1a},X_{1b},X_{2a},X_{2b};\mathbf{Y}_{2a},\mathbf{Y}_{2b}|W_a,W_b,Q_a,Q_b) - \gamma\epsilon I(X_{1a},X_{1b},X_{2a},X_{2b};\mathbf{J}_a,\mathbf{J}_b|W_a,W_b,Q_a,Q_b) \\ &= \lambda\alpha I(X_{2+},X_{2-};\mathbf{Y}_{2+},\mathbf{Y}_{2-}|W_a,W_b,X_{1+},X_{1-},Q_a,Q_b) - \lambda\alpha I(X_{2+},X_{2-};\hat{\mathbf{J}}_+,\hat{\mathbf{J}}_-|W_a,W_b,X_{1+},X_{1-},Q_a,Q_b) \\ &\quad + \lambda(1-\alpha)I(X_{2+},X_{2-};\mathbf{Y}_{2+},\mathbf{Y}_{2-}|W_a,W_b,Q_a,Q_b) \\ &\quad - \lambda(1-\alpha)I(X_{2+},X_{2-};\mathbf{J}_+,\mathbf{J}_-|W_a,W_b,Q_a,Q_b) + \beta I(X_{1+},X_{1-};\mathbf{Y}_{1+},\mathbf{Y}_{1-}|Q_a,Q_b) \\ &\quad + (1-\beta)\big[I(W_a,W_b;\mathbf{Y}_{2+},\mathbf{Y}_{2-}|Q_a,Q_b) + I(X_{1+},X_{1-};\mathbf{J}_+,\mathbf{J}_-|W_a,W_b,Q_a,Q_b) \\ &\qquad\qquad + I(X_{2+},X_{2-};\bar{\mathbf{J}}_+,\bar{\mathbf{J}}_-|W_a,W_b,X_{1+},X_{1-},Q_a,Q_b)\big] \qquad (56) \\ &\quad + \epsilon I(X_{1+},X_{1-},X_{2+},X_{2-};\mathbf{Y}_{2+},\mathbf{Y}_{2-}|W_a,W_b,Q_a,Q_b) - \gamma\epsilon I(X_{1+},X_{1-},X_{2+},X_{2-};\mathbf{J}_+,\mathbf{J}_-|W_a,W_b,Q_a,Q_b) \end{aligned}$$

This equality replaces the a, b copies of random variables with their rotated $+$ and $-$ versions. It holds since the mapping $(X_{2a},X_{2b}) \mapsto (X_{2+},X_{2-})$ (and the similar mapping for other pairs of random variables) is a bijection and hence preserves mutual information terms. Indeed the determinant of the Jacobian of the transformation $(r_a,r_b) \mapsto (r_+,r_-)$ is one, hence it would even preserve differential entropies.

The next step is equating (a) and (b) in the proof of Theorem 3. To justify this step we will break things one by one. One of the repeated tricks that we will employ is the following

$$\begin{aligned} &I(X_+,X_-;Y_+,Y_-|U) - I(X_+,X_-;\hat{Y}_+,\hat{Y}_-|U) \\ &\quad = I(X_+,X_-;Y_+|U,\hat{Y}_-) - I(X_+,X_-;\hat{Y}_+|U,\hat{Y}_-) + I(X_+,X_-;Y_-|U,Y_+) - I(X_+,X_-;\hat{Y}_-|U,Y_+) \end{aligned}$$

This is just Lemma 6 with n=2.

- *Considering the terms that are multiplied by $\lambda\alpha$ in (56), we have:*

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

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

where in the second equality the (immediate) Markov chain

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

is used to introduce $\mathbf{J}_-$ in the conditioning. Next,

$$
\begin{aligned}
&I(X_{2+}, X_{2-}; \mathbf{Y}_{2+}|W_a, W_b, X_{1+}, X_{1-}, \mathbf{J}_-, \hat{\mathbf{J}}_-, Q_a, Q_b) - I(X_{2+}, X_{2-}; \hat{\mathbf{J}}_+|W_a, W_b, X_{1+}, X_{1-}, \mathbf{J}_-, \hat{\mathbf{J}}_-, Q_a, Q_b) \\
&\quad + I(X_{2+}, X_{2-}; \mathbf{Y}_{2-}|W_a, W_b, X_{1+}, X_{1-}, \mathbf{Y}_{2+}, Q_a, Q_b) - I(X_{2+}, X_{2-}; \hat{\mathbf{J}}_-|W_a, W_b, X_{1+}, X_{1-}, \mathbf{Y}_{2+}, Q_a, Q_b) \\
&= I(X_{2+}, X_{2-}, X_{1-}, \hat{\mathbf{J}}_-; \mathbf{Y}_{2+}|W_a, W_b, X_{1+}, \mathbf{J}_-, Q_a, Q_b) - I(X_{2+}, X_{2-}, X_{1-}, \hat{\mathbf{J}}_-; \hat{\mathbf{J}}_+|W_a, W_b, X_{1+}, \mathbf{J}_-, Q_a, Q_b) \\
&\quad - I(X_{1-}, \hat{\mathbf{J}}_-; \mathbf{Y}_{2+}|W_a, W_b, X_{1+}, \mathbf{J}_-, Q_a, Q_b) + I(X_{1-}, \hat{\mathbf{J}}_-; \hat{\mathbf{J}}_+|W_a, W_b, X_{1+}, \mathbf{J}_-, Q_a, Q_b) \\
&\quad + I(X_{2+}, X_{2-}, X_{1+}; \mathbf{Y}_{2-}|W_a, W_b, X_{1-}, \mathbf{Y}_{2+}, Q_a, Q_b) - I(X_{2+}, X_{2-}, X_{1+}; \hat{\mathbf{J}}_-|W_a, W_b, X_{1-}, \mathbf{Y}_{2+}, Q_a, Q_b) \\
&\quad - I(X_{1+}; \mathbf{Y}_{2-}|W_a, W_b, X_{1-}, \mathbf{Y}_{2+}, Q_a, Q_b) + I(X_{1+}; \hat{\mathbf{J}}_-|W_a, W_b, X_{1-}, \mathbf{Y}_{2+}, Q_a, Q_b) \\
&= I(X_{2+}; \mathbf{Y}_{2+}|W_a, W_b, X_{1+}, \mathbf{J}_-, Q_a, Q_b) - I(X_{2+}; \hat{\mathbf{J}}_+|W_a, W_b, X_{1+}, \mathbf{J}_-, Q_a, Q_b) \\
&\quad - I(X_{1-}, \hat{\mathbf{J}}_-; \mathbf{Y}_{2+}|W_a, W_b, X_{1+}, \mathbf{J}_-, Q_a, Q_b) + I(X_{1-}, \hat{\mathbf{J}}_-; \hat{\mathbf{J}}_+|W_a, W_b, X_{1+}, \mathbf{J}_-, Q_a, Q_b) \\
&\quad + I(X_{2-}; \mathbf{Y}_{2-}|W_a, W_b, X_{1-}, \mathbf{Y}_{2+}, Q_a, Q_b) - I(X_{2-}; \hat{\mathbf{J}}_-|W_a, W_b, X_{1-}, \mathbf{Y}_{2+}, Q_a, Q_b) \qquad (57) \\
&\quad - I(X_{1+}; \mathbf{Y}_{2-}|W_a, W_b, X_{1-}, \mathbf{Y}_{2+}, Q_a, Q_b) + I(X_{1+}; \hat{\mathbf{J}}_-|W_a, W_b, X_{1-}, \mathbf{Y}_{2+}, Q_a, Q_b)
\end{aligned}
$$

where in the last step we used the fact that the channels are additive Gaussian (and hence Z_+'s are independent of other variables), we have Y_{2+} depends only on the "inputs" (X_{1+}, X_{2+}), (and similarly for the negative terms).

- *Considering the terms that are multiplied by* $\lambda(1-\alpha)$ *in* (56)*, we have:*

$$
\begin{aligned}
&I(X_{2+}, X_{2-}; \mathbf{Y}_{2+}, \mathbf{Y}_{2-}|W_a, W_b, Q_a, Q_b) - I(X_{2+}, X_{2-}; \mathbf{J}_+, \mathbf{J}_-|W_a, W_b, Q_a, Q_b) \\
&\quad = I(X_{2+}, X_{2-}; \mathbf{Y}_{2+}|W_a, W_b, \mathbf{J}_-, Q_a, Q_b) - I(X_{2+}, X_{2-}; \mathbf{J}_+|W_a, W_b, \mathbf{J}_-, Q_a, Q_b) \\
&\quad\quad + I(X_{2+}, X_{2-}; \mathbf{Y}_{2-}|W_a, W_b, \mathbf{Y}_{2+}, Q_a, Q_b) - I(X_{2+}, X_{2-}; \mathbf{J}_-|W_a, W_b, \mathbf{Y}_{2+}, Q_a, Q_b) \\
&\quad = I(X_{2+}; \mathbf{Y}_{2+}|W_a, W_b, \mathbf{J}_-, Q_a, Q_b) - I(X_{2+}; \mathbf{J}_+|W_a, W_b, \mathbf{J}_-, Q_a, Q_b) \\
&\quad\quad + I(X_{2-}; \mathbf{Y}_{2-}|W_a, W_b, \mathbf{Y}_{2+}, Q_a, Q_b) - I(X_{2-}; \mathbf{J}_-|W_a, W_b, \mathbf{Y}_{2+}, Q_a, Q_b) \\
&\quad\quad + I(X_{2-}; \mathbf{Y}_{2+}|W_a, W_b, X_{2+}, \mathbf{J}_-, Q_a, Q_b) - I(X_{2-}; \mathbf{J}_+|W_a, W_b, X_{2+}, \mathbf{J}_-, Q_a, Q_b) \\
&\quad\quad + I(X_{2+}; \mathbf{Y}_{2-}|W_a, W_b, X_{2-}, \mathbf{Y}_{2+}, Q_a, Q_b) - I(X_{2+}; \mathbf{J}_-|W_a, W_b, X_{2-}, \mathbf{Y}_{2+}, Q_a, Q_b) \\
&\quad = I(X_{2+}; \mathbf{Y}_{2+}|W_a, W_b, \mathbf{J}_-, Q_a, Q_b) - I(X_{2+}; \mathbf{J}_+|W_a, W_b, \mathbf{J}_-, Q_a, Q_b) \\
&\quad\quad + I(X_{2-}; \mathbf{Y}_{2-}|W_a, W_b, \mathbf{Y}_{2+}, Q_a, Q_b) - I(X_{2-}; \mathbf{J}_-|W_a, W_b, \mathbf{Y}_{2+}, Q_a, Q_b) \qquad (58) \\
&\quad\quad - I(X_{2-}; \mathbf{J}_+|W_a, W_b, X_{2+}, \mathbf{J}_-, \mathbf{Y}_{2+}, Q_a, Q_b) - I(X_{2+}; \mathbf{J}_-|W_a, W_b, X_{2-}, \mathbf{Y}_{2-}, \mathbf{Y}_{2+}, Q_a, Q_b)
\end{aligned}
$$

Here the last inequality uses

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

The first equation follows from the structure of Y_{2+} which is a linear combination of X_{2+}, J_+ and some independent noise. The second equation follows similarly.

- *Considering the terms that are multiplied by* β *in* (56)*, we have:*

  Since Y_1 is X_1 with some additive Gaussian noise, the following decomposition is immediate

$$I(X_{1+}, X_{1-}; \mathbf{Y}_{1+}, \mathbf{Y}_{1-}|Q_a, Q_b) = I(X_{1+}; \mathbf{Y}_{1+}|Q_a, Q_b) + I(X_{1-}; \mathbf{Y}_{1-}|Q_a, Q_b) - I(\mathbf{Y}_{1+}; \mathbf{Y}_{1-}|Q_a, Q_b). \qquad (59)$$

- *Considering the terms that are multiplied by* $1-\beta$ *in* (56)*, we have:*
The key equality used here is (one can also swap $(A_2, B_1) \leftrightarrow (A_1, B_2)$)

$$h(A_1, A_2|C) - h(B_1, B_2|C) = h(A_1|C, A_2) - h(B_1|C, A_2) + h(A_2|C, B_1) - h(B_2|C, B_1).$$

$$\begin{aligned}
&I(W_a, W_b; \mathbf{Y}_{2+}, \mathbf{Y}_{2-}|Q_a, Q_b) + I(X_{1+}, X_{1-}; \mathbf{J}_+, \mathbf{J}_-|W_a, W_b, Q_a, Q_b) + I(X_{2+}, X_{2-}; \bar{\mathbf{J}}_+, \bar{\mathbf{J}}_-|W_a, W_b, X_{1+}, X_{1-}, Q_a, Q_b)\\
&\quad = h(\mathbf{Y}_{2+}, \mathbf{Y}_{2-}|Q_a, Q_b) + \left[h(\mathbf{J}_+, \mathbf{J}_-|W_a, W_b, Q_a, Q_b) - h(\mathbf{Y}_{2+}, \mathbf{Y}_{2-}|W_a, W_b, Q_a, Q_b)\right]\\
&\qquad + \left[h(\bar{\mathbf{J}}_+, \bar{\mathbf{J}}_-|W_a, W_b, X_{1+}, X_{1-}, Q_a, Q_b) - h(\mathbf{J}_+, \mathbf{J}_-|W_a, W_b, X_{1+}, X_{1-}, Q_a, Q_b)\right]\\
&\qquad - h(\bar{\mathbf{J}}_+, \bar{\mathbf{J}}_-|W_a, W_b, X_{2+}, X_{2-}, X_{1+}, X_{1-}, Q_a, Q_b)\\
&\quad = h(\mathbf{Y}_{2+}|Q_a, Q_b) + h(\mathbf{Y}_{2-}|Q_a, Q_b) - I(\mathbf{Y}_{2+}; \mathbf{Y}_{2-}|Q_a, Q_b)\\
&\qquad + \Big[h(\mathbf{J}_+|W_a, W_b, Q_a, Q_b, \mathbf{J}_-) - h(\mathbf{Y}_{2+}|W_a, W_b, Q_a, Q_b, \mathbf{J}_-)\\
&\qquad\quad + h(\mathbf{J}_-|W_a, W_b, Q_a, Q_b, \mathbf{Y}_{2+}) - h(\mathbf{Y}_{2-}|W_a, W_b, Q_a, Q_b, \mathbf{Y}_{2+})\Big]\\
&\qquad + \Big[h(\bar{\mathbf{J}}_+|W_a, W_b, X_{1+}, X_{1-}, Q_a, Q_b, \mathbf{J}_-) - h(\mathbf{J}_+|W_a, W_b, X_{1+}, X_{1-}, Q_a, Q_b, \mathbf{J}_-)\\
&\qquad\quad + h(\bar{\mathbf{J}}_-|W_a, W_b, X_{1+}, X_{1-}, Q_a, Q_b, \bar{\mathbf{J}}_+,) - h(\mathbf{J}_-|W_a, W_b, X_{1+}, X_{1-}, Q_a, Q_b, \bar{\mathbf{J}}_+,)\Big]\\
&\qquad - h(\bar{\mathbf{J}}_+|W_a, W_b, X_{2+}, X_{2-}, X_{1+}, X_{1-}, Q_a, Q_b, \mathbf{J}_-) - h(\bar{\mathbf{J}}_-|W_a, W_b, X_{2+}, X_{2-}, X_{1+}, X_{1-}, Q_a, Q_b, \bar{\mathbf{J}}_+)
\end{aligned}$$

In the last line we used the equality

$$h(\bar{\mathbf{J}}_+|W_a, W_b, X_{2+}, X_{2-}, X_{1+}, X_{1-}, Q_a, Q_b) = h(\bar{\mathbf{J}}_+|W_a, W_b, X_{2+}, X_{2-}, X_{1+}, X_{1-}, Q_a, Q_b, \mathbf{J}_-),$$

introducing an extra $\mathbf{J}_-$ for free because $\bar{\mathbf{J}}_+$ depends on the inputs (X_{2+}, X_{1+}) and some independent additive Gaussian noise. Now we can recombine them back into mutual information terms as follows

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

The key observation is that both the terms in the curly braces are zero. This is where the peculiar construction of the two variables $\mathbf{J}$ and $\bar{\mathbf{J}}$ are useful. Observe that the first components of both are X_1 with different independent additive Gaussians. When conditioned on say X_{1+} the additive Gaussians of both components are completely independent of the remaining terms and drop away from the mutual information. The second components of both $\mathbf{J}$ and $\bar{\mathbf{J}}$ are *identical*. Hence the terms in the curly braces are zero. This yields that the sum of the terms multiplied by $(1-\beta)$ in (56) is equivalent to

$$\begin{aligned}
&I(W_a, W_b, \mathbf{J}_-; \mathbf{Y}_{2+}|Q_a, Q_b) + I(X_{1+}; \mathbf{J}_+|W_a, W_b, \mathbf{J}_-, Q_a, Q_b) + I(X_{2+}; \bar{\mathbf{J}}_+|W_a, W_b, X_{1+}, \mathbf{J}_-, Q_a, Q_b)\\
&\quad + I(W_a, W_b, \mathbf{Y}_{2+}; \mathbf{Y}_{2-}|Q_a, Q_b) + I(X_{1+}, X_{1-}; \mathbf{J}_-|W_a, W_b, Q_a, Q_b, \mathbf{Y}_{2+}) + I(X_{2+}, X_{2-}; \bar{\mathbf{J}}_-|W_a, W_b, X_{1+}, X_{1-}, Q_a, Q_b, \mathbf{Y}_{2+})\\
&\quad - I(\mathbf{Y}_{2+}; \mathbf{Y}_{2-}|Q_a, Q_b)
\end{aligned} \tag{60}$$

- *Considering the terms that are multiplied by* ϵ *in* (56)*, we have:*

$$\begin{aligned}
&I(X_{1+}, X_{1-}, X_{2+}, X_{2-}; \mathbf{Y}_{2+}, \mathbf{Y}_{2-}|W_a, W_b, Q_a, Q_b) - \gamma I(X_{1+}, X_{1-}, X_{2+}, X_{2-}; \mathbf{J}_+, \mathbf{J}_-|W_a, W_b, Q_a, Q_b)\\
&\quad = I(X_{1+}, X_{2+}; \mathbf{Y}_{2+}|W_a, W_b, \mathbf{J}_-, Q_a, Q_b) - \gamma I(X_{1+}, X_{2+}; \mathbf{J}_+|W_a, W_b, \mathbf{J}_-, Q_a, Q_b)\\
&\qquad + I(X_{1-}, X_{2-}; \mathbf{Y}_{2-}|W_a, W_b, Q_a, Q_b, \mathbf{Y}_{2_+}) - \gamma I(X_{1-}, X_{2-}; \mathbf{J}_-|W_a, W_b, Q_a, Q_b, \mathbf{Y}_{2+})\\
&\qquad - (\gamma-1) I(\mathbf{Y}_{2+}; \mathbf{J}_-|W_a, W_b, Q_a, Q_b)
\end{aligned} \tag{61}$$

This is exactly as in Example 2 of [GN14] and follows from Lemma 4.

Using (57), (58), (59), (60), (61) to substitute the terms in (56) we see that $2V$ can be further written as

$$\begin{aligned}
2V = {} & \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)\\
& + \lambda(1-\alpha)I(X_{2+};\mathbf{Y}_{2+}|W_a,W_b,\mathbf{J}_-,Q_a,Q_b)\\
& - \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)\\
& + (1-\beta)\big[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)\\
& \qquad + I(X_{2+};\bar{\mathbf{J}}_+|W_a,W_b,X_{1+},\mathbf{J}_-,Q_a,Q_b)\big]\\
& + \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)\\
& + \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+})\\
& + \lambda(1-\alpha)I(X_{2-};\mathbf{Y}_{2-}|W_a,W_b,\mathbf{Y}_{2+},Q_a,Q_b)\\
& - \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)\\
& + (1-\beta)\big[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+})\\
& \qquad + I(X_{2-};\bar{\mathbf{J}}_-|W_a,W_b,X_{1-},Q_a,Q_b,\mathbf{Y}_{2+})\big]\\
& + \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+})\\
& - \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]\\
& - \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]\\
& - \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). \qquad (62)
\end{aligned}$$

This completes the justification of equating (a) and (b) in the proof of Theorem 3.

E.2. **Steps in the proof of Proposition 2.**

E.2.1. *Argument 1.* We first show the algebra manipulations for the first inequality in Appendix B.2.

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

or equivalently,

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

which is equivalent to showing,

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

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

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

or equivalently,

$$\frac{S_J}{1+S_J} \leq \frac{S_J}{1+S_J}\frac{\left(\sqrt{S_{31}}+\sqrt{\rho^2 S_{32}}\right)^2}{S_{31}} + \frac{S_J S_{32}(1-\rho^2)}{S_{31}-S_J},$$

which is immediate.

E.2.2. *Argument 2, step* (a). Note that for $\rho^2 \leq 1$ and $S_{21} \leq S_J < S_{31}$

$$\frac{1}{2}\log(1+S_J(1-\rho^2)) + \frac{1}{2}\log\left(\frac{1+S_{31}+\rho^2 S_{32}+2\sqrt{\rho^2 S_{31}S_{32}}}{1+S_J} + \frac{S_{31}S_{32}(1-\rho^2)}{S_{31}-S_J}\right)$$
$$\stackrel{(a)}{\leq} \frac{1}{2}\log\left(1+S_{31}+S_{32}\left(\frac{(1+S_J)^2 S_{31}^2}{S_J(1+S_{31})(S_{31}-S_J)}\right)\right)$$

or equivalently,

$$1+S_{31}+\rho^2 S_{32}+2\sqrt{\rho^2 S_{31}S_{32}} + \frac{S_{31}S_{32}(1-\rho^2)}{S_{31}-S_J}(1+S_J(1-\rho^2))$$
$$\leq 1+S_{31}+S_{32}\left(\frac{(1+S_J)^2 S_{31}^2}{S_J(1+S_{31})(S_{31}-S_J)}\right) + \rho^2 S_J\left(\frac{1+S_{31}+\rho^2 S_{32}+2\sqrt{\rho^2 S_{31}S_{32}}}{1+S_J}\right).$$

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

$$\rho^2 S_{32}+2\sqrt{\rho^2 S_{31}S_{32}} + \frac{S_{31}S_{32}(1-\rho^2)}{S_{31}-S_J}(1+S_J(1-\rho^2))$$
$$\leq S_{32}\left(\frac{(1+S_J)S_{31}}{S_{31}-S_J} + \frac{S_{31}(1+S_J)}{S_J(1+S_{31})}\right) + \rho^2 S_J\left(\frac{1+S_{31}+\rho^2 S_{32}+2\sqrt{\rho^2 S_{31}S_{32}}}{1+S_J}\right).$$

The last inequality follows from the two immediate ones below

$$\rho^2 S_{32} + \frac{S_{31}S_{32}(1-\rho^2)}{S_{31}-S_J}(1+S_J(1-\rho^2)) \leq S_{32}\frac{(1+S_J)S_{31}}{S_{31}-S_J}$$
$$2\sqrt{\rho^2 S_{31}S_{32}} \leq \frac{S_{32}S_{31}(1+S_J)}{S_J(1+S_{31})} + \rho^2 S_J\left(\frac{1+S_{31}}{1+S_J}\right).$$

E.3. **Steps in the proof of Theorem 8.** To see (33c) observe that

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

To see (33f) observe that

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