A Strengthened Cutset Upper Bound on the Capacity of the Relay Channel and Applications

Abbas El Gamal, Amin Gohari and Chandra Nair

Abstract

We develop a new upper bound on the capacity of the relay channel that is tighter than all previous bounds. This upper bound is proved using traditional weak converse techniques involving mutual information inequalities and identification of auxiliary random variables via past and future channel random variable sequences. We show that it is strictly tighter than all previous bounds for the Gaussian relay channel with non-zero channel gains. When specialized to the relay channel with orthogonal receiver components, the bound resolves a conjecture by Kim on a class of deterministic relay channels. When further specialized to the class of product-form relay channels with orthogonal receiver components, the bound resolves a generalized version of Cover's relay channel problem, recovers the recent upper bound for the Gaussian case by Wu et al., and improves upon the recent bounds for the binary symmetric case by Wu et al. and Barnes et al., which are all obtained using non-traditional geometric proof techniques. We then develop an upper bound on the capacity of the relay channel with orthogonal receiver components which utilizes an auxiliary receiver and show that it is tighter than the bound by Tandon and Ulukus on the capacity of the relay channel with i.i.d. relay output sequence. Finally, we show through the Gaussian relay channel with i.i.d. relay output sequence that the bound with the auxiliary receiver can be strictly tighter than our main bound.[1]

1. Introduction

The relay channel was first introduced by van-der Meulen in 1971 [VDM71]. It is a basic model of a multi-hop communication network in which a sender X wishes to communicate to a receiver Y with the help of a relay $(X_\mathrm{r}, Y_\mathrm{r})$ over a memoryless channel of the form $p(y, y_\mathrm{r}|x, x_\mathrm{r})$. The capacity of this channel, which is the highest achievable rate from X to Y, is not known in general. In [CEG79], the cutset upper bound and several lower bounds on the capacity, later termed decode-forward, partial decode-forward, and compress-forward, were established. The cutset bound was shown to coincide with one or more of these lower bounds for several classes of relay channels, including degraded [CEG79], semi-deterministic [EGA82], and channels with orthogonal sender components [EGZ05] relay channels. In [ARY09a], the cutset bound was shown not to be tight in general via an example relay channel with orthogonal receiver components. These results and others are detailed in Chapter 19 of [EK11].

Motivated by Cover's problem concerning a relay channel with orthogonal receiver components [Cov87], a series of recent papers [WOX17], [WBO19], [LO19] developed specialized upper bounds that are also tighter than the cutset bound. While the older bounds in [CEG79], [TU08], [ARY09a] use standard weak converse techniques involving basic mutual information inequalities and identities, and Gallager-type auxiliary random variable identification, the recent bounds in [WOX17], [WBO19], [LO19] for symmetric Gaussian and binary symmetric relay channels with orthogonal receiver components use more sophisticated arguments from convex geometry and functional analysis. More recently, Gohari-Nair [GN20] developed a new upper bound on the capacity of the general relay channel and showed that it can be strictly tighter than the cutset bound. Their bound uses traditional converse techniques, including identification of auxiliary random variables using past and future channel variable sequences which have been used in previous converse proofs, for example, in [CK78], [EG79], and the new idea of introducing auxiliary receivers.

In this paper, which is a more complete and extended version of [EGGN21], we establish a new upper bound on the capacity of the relay channel. We show through several applications and examples that this bound is tighter than all previous upper bounds, including the cutset bound and the specialized bounds in [WOX17], [WBO19], [LO19], [GN20]. While our upper bound is motivated by the arguments in [GN20] and uses standard general converse techniques, it does not include auxiliary receivers because they make the bound computationally intractable in general and, as we will see, are not needed in the applications we consider. Nonetheless, we provide an upper bound for the relay channel with orthogonal receiver components involving an auxiliary receiver and show through an example that it is strictly tighter than our main bound.

[1]This paper was presented in part at the IEEE Symposium on Information Theory (ISIT) 2021.

Although the techniques used to establish the upper bounds in this paper have been employed in many previous works, our contributions are in (i) the judicious manner in which we minimize the discarded terms in the derivations of the constraints, which led to the tightest known upper bound on the capacity of the relay channel, and (ii) the rather nontrivial computation of the resulting bound to demonstrate strict improvements over previous bounds for several classes and examples of relay channels.

Organization of the paper and summary of the results. In the following section, we introduce the relay channel capacity problem and state and discuss our results. The proofs of these results are all given in Section 3. To help navigate the paper, Figure 1 depicts the classes of relay channels for which we provide new upper bounds and applications with references to the corresponding sections. Note that all our applications are for classes of relay channels *without self-interference* $p(y_{\mathrm{r}}|x)p(y|x, x_{\mathrm{r}}, y_{\mathrm{r}})$ because they include and generalize several interesting relay channel settings that have been receiving significant attention in recent years.

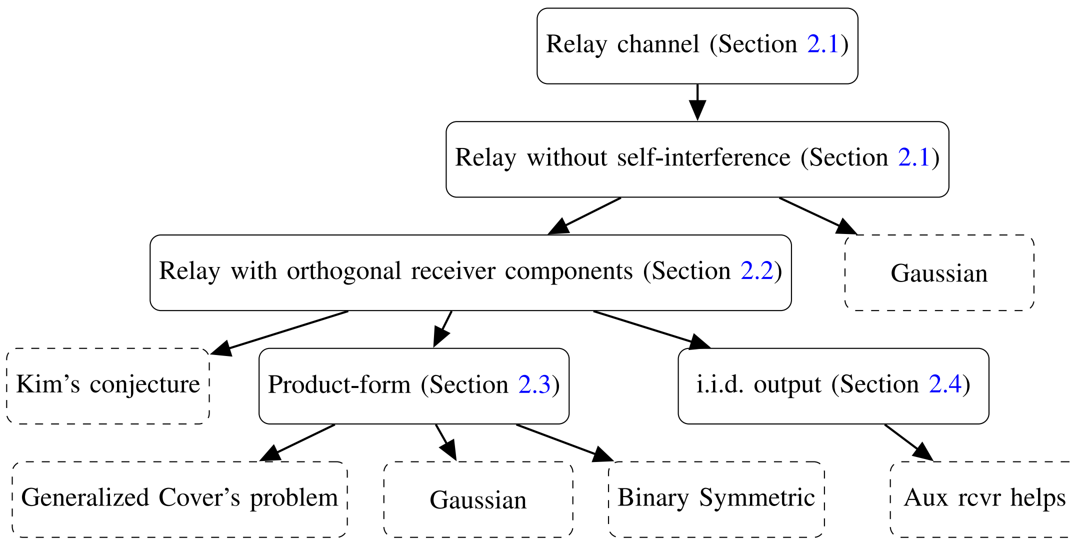

Fig. 1. Classes of relay channels for which upper bounds are established. An arrow from box A to box B indicates that class A includes class B. Dashed boxes indicate applications for which we compute the bounds.

The following is a summary of our results.

Section 2.1. Theorem 1 states our main upper bound. Although the statement and subsequent bounds and applications of this bound are for relay channels without self-interference, in Remark 1, we point out that with a minor relaxation, the upper bound holds for the general relay channel. Corollary 1 is a weakened but easier to evaluate version of Theorem 1 which is used in Theorem 2 to show the suboptimality of the cut-set upper bound and the improved bound in [GN20] for the Gaussian relay channel; see Figure 3.

Section 2.2. Proposition 1 provides an equivalent characterization of the upper bound in Theorem 1 for relay channels with orthogonal receiver components (also referred to as primitive relay). This proposition is used in Theorem 3 to prove a conjecture by Kim [Kim07, Question 2].

Section 2.3. Theorem 4 uses Proposition 1 to answer a generalized version of Cover's relay channel problem in [Cov87], which extends and improves upon results in [WOX17], [WBO19]. Proposition 2 uses Proposition 1 to derive an upper bound for the Gaussian case; see Figure 5. As shown in Lemma 1 it indirectly recovers previous results in [WBO19], [WBÖ17] by showing that the two seemingly different optimization problems evaluate to the same expression. Theorem 6 uses Proposition 1 to establish a tighter upper bound for a class of symmetric binary relay channels than the bounds in [WOX17], [BWÖ17]; see Figure 6.

Section 2.4. Theorem 7 provides a new upper bound on the capacity of the relay channel with orthogonal receiver components using the auxiliary receiver approach in [GN20]. Corollary 2 specializes Theorem 7 to relay channels with orthogonal receiver components in which the relay output is an i.i.d. sequence, independent of the sender and relay transmissions, which was previously studied in [ARY09b], [TU08], [AH83]. This bound is shown to be tight for the class of channels considered in [ARY09b] and strictly tighter than the upper bound by Tandon and Ulukus in [TU08]. We then show through a Gaussian relay channel with i.i.d. relay output sequence that the bound in Corollary 2 is strictly tighter than the upper bound in Proposition 1. Hence, including an auxiliary receiver into the upper bound can strictly improve upon our main upper bound in Theorem 1.

2. Definitions and statement of the results

We adopt most of our notation from [EK11]. In particular, we use Y^i to denote the sequence $(Y_1, Y_2, \cdots, Y_i)$, and Y_i^j to denote $(Y_i, Y_{i+1}, \cdots, Y_j)$. Unless stated otherwise, logarithms are to the base 2. We use $p(x)$ to indicate the probability mass function of a discrete random variable X and P_Y to indicate the probability distribution of an arbitrary random variable Y. We define $\mathsf{C}(x) = (1/2)\log(1+x)$ for $x \geq 0$.

The discrete memoryless relay channel depicted in Figure 2 consists of four alphabets $\mathcal{X}$, $\mathcal{X}_\mathrm{r}$, $\mathcal{Y}_\mathrm{r}$, $\mathcal{Y}$, and a collection of conditional pmfs $p(y_\mathrm{r}, y|x, x_\mathrm{r})$ on $\mathcal{Y}_\mathrm{r} \times \mathcal{Y}$. A $(2^{nR}, n)$ code for the discrete memoryless relay channel $p(y, y_\mathrm{r}|x, x_\mathrm{r})$ consists of a message set $[1 : 2^{nR}]$, an encoder that assigns a codeword $x^n(m)$ to each message $m \in [1 : 2^{nR}]$, a relay encoder that assigns a symbol $x_{\mathrm{r}i}(y_\mathrm{r}^{i-1})$ to each past received sequence y_r^{i-1} for each time $i \in [1 : n]$, and a decoder that assigns an estimate $\hat{M}$ or an error message ε to each received sequence y^n. We assume that the message M is uniformly distributed over $[1 : 2^{nR}]$. The definitions of the average probability of error, achievability and capacity follow those in [EK11].

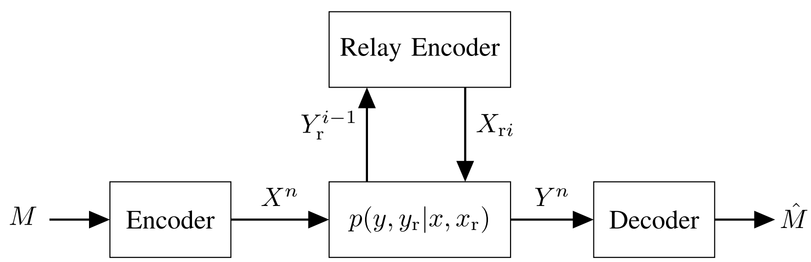

Fig. 2. Transmission of a message M over a memoryless relay channel with n channel uses.

2.1. Upper bound for the relay channel without self-interference

A relay channel is said to be *without self-interference* if $p(y, y_\mathrm{r}|x, x_\mathrm{r}) = p(y_\mathrm{r}|x)p(y|x, x_\mathrm{r}, y_\mathrm{r})$. We establish the following upper bound on the capacity of this class of relay channels.

Theorem 1. *Any achievable rate R for a discrete memoryless relay channel without self-interference $p(y_\mathrm{r}|x)p(y|x, x_\mathrm{r}, y_\mathrm{r})$ must satisfy the following inequalities*

$$\begin{aligned}
R &\leq I(X; Y, Y_\mathrm{r}|X_\mathrm{r}) - I(U; Y|X_\mathrm{r}, Y_\mathrm{r}), && (1)\\
&= I(X; Y|X_\mathrm{r}, U) + I(U; Y_\mathrm{r}|X_\mathrm{r}) + I(X; Y_\mathrm{r}|X_\mathrm{r}, U, Y) && (2)\\
R &\leq I(X; Y, Y_\mathrm{r}|X_\mathrm{r}) - I(V; Y|X_\mathrm{r}, Y_\mathrm{r}) - I(X; Y_\mathrm{r}|V, X_\mathrm{r}, Y) && (3)\\
&= I(X; Y_\mathrm{r}|X_\mathrm{r}) + I(X; Y|V, X_\mathrm{r}) - I(X; Y_\mathrm{r}|V, X_\mathrm{r}) && (4)\\
&= I(X; Y, V|X_\mathrm{r}) - I(V; X|X_\mathrm{r}, Y_\mathrm{r}), && (5)\\
R &\leq I(X, X_\mathrm{r}; Y) - I(V; Y_\mathrm{r}|X_\mathrm{r}, X, Y), && (6)
\end{aligned}$$

for some $p(u, x, x_\mathrm{r})p(y, y_\mathrm{r}|x, x_\mathrm{r})p(v|x, x_\mathrm{r}, y_\mathrm{r})$ satisfying

$$I(V, X_\mathrm{r}; Y_\mathrm{r}) - I(V, X_\mathrm{r}; Y) = I(U; Y_\mathrm{r}) - I(U; Y). \quad (7)$$

Moreover, we obtain an equivalent characterization of the bound if we drop the constraint on R in (6) and strengthen (7) to

$$I(V, X_\mathrm{r}; Y_\mathrm{r}) - I(V, X_\mathrm{r}; Y) = I(U; Y_\mathrm{r}) - I(U; Y) \leq I(X_\mathrm{r}; Y_\mathrm{r}). \quad (8)$$

In both characterizations of the bound, it suffices to consider $|\mathcal{V}| \le |\mathcal{X}||\mathcal{X}_\mathrm{r}||\mathcal{Y}_\mathrm{r}| + 2$ and $|\mathcal{U}| \le |\mathcal{X}||\mathcal{X}_\mathrm{r}| + 1$.

The proof of this theorem is given in Section 3.1. As in every weak converse, the starting point of the proof is to consider the joint distribution $(M, X^n, X_\mathrm{r}^n, Y^n, Y_\mathrm{r}^n)$ induced by any given code. Since by assumption, M can be recovered from Y^n, then by Fano's inequality $H(M) \approx I(M; Y^n)$. Note that to obtain the broadcast bound of the cut-set bound we begin from the following expansion

$$nR \le I(M; Y^n, Y_\mathrm{r}^n) \le \cdots \le \sum_{i=1}^{n} I(X_i; Y_i, Y_{\mathrm{r}i} | X_{\mathrm{r},i}),$$

which yields the single-letter term $I(X; Y, Y_\mathrm{r}|X_\mathrm{r})$ in which Y and Y_r are lumped together. In our proof, we instead begin use the tighter bound

$$nR \le I(M; Y_\mathrm{r}^n) + I(M; Y^n) - I(M; Y_\mathrm{r}^n),$$

and single-letterize $I(M; Y_\mathrm{r}^n)$ and $I(M; Y^n) - I(M; Y_\mathrm{r}^n)$ separately. We expand the second term, which represents the information *not* recovered by the relay, using the Csiszár-Körner-Marton sum lemma in two ways

$$\begin{aligned} I(M; Y^n) - I(M; Y_\mathrm{r}^n) &= \sum_i I(M; Y_i | Y_{\mathrm{r}i+1}^n, Y^{i-1}) - \sum_i I(M; Y_{\mathrm{r}i} | Y_{\mathrm{r}i+1}^n, Y^{i-1}) \\ &= \sum_i I(M; Y_i | Y_{i+1}^n, Y_\mathrm{r}^{i-1}) - \sum_i I(M; Y_{\mathrm{r}i} | Y_{i+1}^n, Y_\mathrm{r}^{i-1}). \end{aligned}$$

This naturally suggests the identifications of the auxiliary random variables V and U as

$$V_i = (Y_{i+1}^n, Y_\mathrm{r}^{i-1}), \quad U_i = (Y_{\mathrm{r}i+1}^n, Y^{i-1}),$$

The rest of the proof involves bounding similar carefully selected terms and using standard mutual information inequalities and identities to establish the stated bounds on the rate R.

Remark 1. As the proof of the theorem in Section 3.1 shows, the without self-interference assumption is used only to establish the Markov chain relationship $U \to X, X_\mathrm{r} \to Y, Y_\mathrm{r}$. Hence, by removing this condition, the above theorem readily extends to the general relay channel.

Remark 2. It is immediate that the above bound is at least as tight as the cut-set bound in [CEG79]

$$C \le \max_{p(x,x_\mathrm{r})} \min\{I(X_\mathrm{r}, X; Y), I(X; Y, Y_\mathrm{r}|X_\mathrm{r})\}. \tag{9}$$

Remark 3. From (5) and (6), we deduce that

$$R \le \max_{p(x,x_\mathrm{r})p(v|x,x_\mathrm{r},y_\mathrm{r})} \min\left\{I(X; Y, V|X_\mathrm{r}) - I(V; X|X_\mathrm{r}, Y_\mathrm{r}), I(X, X_\mathrm{r}; Y) - I(V; Y_\mathrm{r}|X_\mathrm{r}, X, Y)\right\}.$$

If we replace the maximum over $p(x, x_\mathrm{r})p(v|x, x_\mathrm{r}, y_\mathrm{r})$ with maximum over $p(x)p(x_\mathrm{r})p(v|x_\mathrm{r}, y_\mathrm{r})$, we obtain the equivalent form of the compress-forward lower bound without time-sharing random variable Q in [EK11]. Thus, the auxiliary random variable V can be interpreted as corresponding to the auxiliary random variable in the compress-forward lower bound.

Remark 4. The auxiliary random variable U in the upper bound can be interpreted as corresponding to the auxiliary random variable in the partial-decode-forward lower bound. This can be seen from (2) which has the terms $I(X; Y|X_\mathrm{r}, U) + I(U; Y_\mathrm{r}|X_\mathrm{r})$ of the partial-decode-forward lower bound plus an excess term $I(X; Y_\mathrm{r}|X_\mathrm{r}, U, Y)$.

Remark 5. If the cut-set bound is tight for a relay channel of the form $p(y, y_\mathrm{r}|x, x_\mathrm{r}) = p(y|x, x_\mathrm{r})p(y_\mathrm{r}|x)$ and the capacity coincides with the MAC bound in the cut-set bound, i.e., $C = I(X_\mathrm{r}, X; Y)$ for the maximizing $p(x, x_\mathrm{r})$, then the capacity is achievable by partial-decode-forward, since from (6) we obtain that $I(V; Y_\mathrm{r}|X_\mathrm{r}, X, Y) = 0$. The assumption $p(y, y_\mathrm{r}|x, x_\mathrm{r}) = p(y|x, x_\mathrm{r})p(y_\mathrm{r}|x)$ implies that $I(V; Y_\mathrm{r}|X_\mathrm{r}, X) = 0$. The constraint $V \to X, X_\mathrm{r} \to Y_\mathrm{r}$, then implies that

$$I(X; Y, Y_\mathrm{r}|X_\mathrm{r}) - I(V; Y|X_\mathrm{r}, Y_\mathrm{r}) - I(X; Y_\mathrm{r}|V, X_\mathrm{r}, Y) = I(V; Y_\mathrm{r}|X_\mathrm{r}) + I(X; Y|V, X_\mathrm{r}),$$

which is the partial-decode-forward lower bound (with auxiliary random variable V). As a concrete application, consider the following example which has the same flavor as the one introduced by Cover in [Cov87]: consider a relay channel of the form $X = (X_a, X_b)$ and $Y_{\rm r} = (Y_{{\rm r}a}, Y_{{\rm r}b})$, where $p(y, y_{\rm r}|x, x_{\rm r}) = p(y_{{\rm r}a}|x_a)p(y|x_a, x_{\rm r})p(y_{{\rm r}b}|x_b)$, and $p(y_{{\rm r}b}|x_b)$ is a noiseless link of capacity C_1 while $p(y_{{\rm r}a}|x_a)$ and $p(y|x_a, x_{\rm r})$ are arbitrary. In other words, we have a noiseless link of capacity C_1 from the sender to the relay in parallel to the channel $p(y_{{\rm r}a}|x_a)p(y|x_a, x_{\rm r})$. Let $\mathcal{C}(C_1)$ be the capacity of this relay channel in terms of C_1 for a fixed channel $p(y_{{\rm r}a}|x_a)p(y|x_a, x_{\rm r})$. For $C_1 = \infty$, we have $\mathcal{C}(\infty) = \max_{p(x)} I(X, X_{\rm r}; Y)$. One can then ask what is the critical value C_1^* such that $C_1^* = \inf\{C_1 : \mathcal{C}(C_1) = \mathcal{C}(\infty) = \max_{p(x)} I(X, X_{\rm r}; Y)\}$? It is immediate from the above discussion that C_1^* can be characterized as the minimum value C_1 such that $R_{\sf PDF}(C_1) = \mathcal{C}(\infty) = \max_{p(x)} I(X, X_{\rm r}; Y)$ where $R_{\sf PDF}(C_1)$ is the rate achieved by partial-decode-forward.

The following corollary is an immediate weakening of the upper bound in Theorem 1, which is easier to evaluate for the Gaussian relay channel.

Corollary 1. *Any achievable rate R for a discrete memoryless relay channel without interference $p(y_{\rm r}|x)p(y|x, x_{\rm r}, y_{\rm r})$ must satisfy the following conditions*

$$
\begin{aligned}
R &\le I(X; Y, Y_{\rm r}|X_{\rm r}) - I(V; Y|X_{\rm r}, Y_{\rm r}) - I(X; Y_{\rm r}|V, X_{\rm r}, Y) && (10)\\
&= I(X; Y_{\rm r}|X_{\rm r}) - I(X; Y_{\rm r}|V, X_{\rm r}) + I(X; Y|V, X_{\rm r}), && (11)
\end{aligned}
$$

for some $p(x, x_{\rm r})p(y, y_{\rm r}|x, x_{\rm r})p(v|x, x_{\rm r}, y_{\rm r})$ satisfying

$$
I(V, X_{\rm r}; Y_{\rm r}) - I(V, X_{\rm r}; Y) \le \min\left[I(X_{\rm r}; Y_{\rm r}), \ \max_{p(u|x,x_{\rm r})} (I(U; Y_{\rm r}) - I(U; Y))\right].
$$

Further it suffices to consider $|\mathcal{V}| \le |\mathcal{X}||\mathcal{X}_{\rm r}||\mathcal{Y}_{\rm r}| + 2$.

The corollary is established by weakening the alternate form of the upper bound in Theorem 1 simply by removing the constraint in (1) and relaxing the condition in (8).

Remark 6. The above bound is a strengthening of the bound given in [GN20, Theorem 1] for the choice of $J = Y_{\rm r}$. To see this, observe that

$$
\begin{aligned}
R &\le I(X; Y_{\rm r}|X_{\rm r}) + I(X; Y|V, X_{\rm r}) - I(X; Y_{\rm r}|V, X_{\rm r})\\
&= I(X; Y_{\rm r}|X_{\rm r}) + I(V, X_{\rm r}, X; Y) - I(V, X_{\rm r}, X; Y_{\rm r}) - I(V, X_{\rm r}; Y) + I(V, X_{\rm r}; Y_{\rm r})\\
&\le I(X; Y_{\rm r}|X_{\rm r}) + I(V, X_{\rm r}, X; Y) - I(V, X_{\rm r}, X; Y_{\rm r}) + \max_{p(u|x,x_{\rm r})} [I(U; Y_{\rm r}) - I(U; Y)]\\
&= I(X; Y_{\rm r}|X_{\rm r}) + \max_{p(u|x,x_{\rm r})} [I(X, X_{\rm r}; Y|U) - I(X, X_{\rm r}; Y_{\rm r}|U)] - I(V; Y_{\rm r}|X, X_{\rm r}, Y)\\
&\le I(X; Y_{\rm r}|X_{\rm r}) + \max_{p(u|x,x_{\rm r})} [I(X, X_{\rm r}; Y|U) - I(X, X_{\rm r}; Y_{\rm r}|U)]\,.
\end{aligned}
$$

Gaussian relay channel. The Gaussian relay channel is defined by

$$
\begin{aligned}
Y_{\rm r} &= g_{12}X + Z_{\rm r},\\
Y &= g_{13}X + g_{23}X_{\rm r} + Z,
\end{aligned} \tag{12}
$$

where g_{12}, g_{13}, and g_{23} are channel gains, and $Z \sim \mathcal{N}(0, 1)$ and $Z_{\rm r} \sim \mathcal{N}(0, 1)$ are independent noise components. We assume average power constraint P on each of X and $X_{\rm r}$.

Remark 7. Note that as defined, the Gaussian relay channel belongs to the class of relay channels without self-interference.

In the following discussion, we use the SNRs $S_{12} = g_{12}^2 P$, $S_{13} = g_{13}^2 P$ and $S_{23} = g_{23}^2 P$ to characterize the Gaussian relay channel. Recall that the cut-set bound for the Gaussian relay channel reduces to [CG79]

$$
\begin{aligned}
C &\le \max_{0\le\rho\le 1} \min\left\{{\sf C}(S_{13} + S_{23} + 2\rho\sqrt{S_{13}S_{23}}), {\sf C}((1-\rho^2)(S_{13} + S_{12}))\right\}\\
&= \begin{cases} {\sf C}\left(\frac{\left(\sqrt{S_{12}S_{23}} + \sqrt{S_{13}(S_{13}+S_{12}-S_{23})}\right)^2}{S_{13}+S_{12}}\right) & \text{if } S_{12} \ge S_{23},\\ {\sf C}(S_{13} + S_{12}) & \text{otherwise.}\end{cases}
\end{aligned} \tag{13}
$$

Also recall that the decode-forward and compress-forward (evaluated using Gaussian distributions) lower bounds for the Gaussian relay channel reduce to [EK11, Eqs. 16.6, 16.12]

$$C \geq \begin{cases} \mathsf{C}\left(\frac{\left(\sqrt{S_{13}(S_{12}-S_{23})}+\sqrt{S_{23}(S_{12}-S_{13})}\right)^2}{S_{12}}\right) & \text{if } S_{12} \geq S_{23}+S_{13}, \\ \mathsf{C}(S_{12}) & \text{otherwise.} \end{cases}$$

$$C \geq \mathsf{C}\left(S_{13} + \frac{S_{12}S_{23}}{S_{13}+S_{12}+S_{23}+1}\right),$$

respectively. The compress–forward outperforms decode–forward if and only if $S_{12}(1+S_{12}) \leq S_{13}(S_{13}+S_{23}+1)$. It is worth noting that when $S_{12}(1+S_{12}) = S_{13}(S_{13}+S_{23}+1)$, i.e., when compress–forward and decode–forward yield the same rate, the mixed strategy lower bound in [CG79, Theorem 7] evaluated using Gaussian distributions strictly improves upon both the compress–forward and decode–forward strategies [LGY13, Theorem 3]. It is also known that for small values of S_{12}, a simple two-letter amplify-forward strategy outperforms both compress–forward and decode–forward strategies [EGMZ06, Example 1]. Additionally, for certain values of channel gains, the lower bound of Chong, Motani and Garg [CMG07] improves upon the mixed strategy lower bound in [CG79] evaluated using Gaussian distributions [CMG07, Remark 6].

We are now ready to show that our upper bound is strictly tighter than the cutset bound.

Theorem 2. *The bound in Corollary 1 for the Gaussian relay channel is strictly tighter than the cut-set upper bound for every non-zero values of g_{12}, g_{13}, g_{23}. Furthermore, the bound reduces to the following. Any achievable rate R for the Gaussian relay channel must satisfy:*

$$\begin{aligned} R \leq\ & \frac{1}{2}\log\left((1-\rho^2)S_{12}+1\right) - \frac{1}{2}\log\left(\beta + S_{12}(1-\rho^2)\alpha + 2\sigma\sqrt{S_{12}(1-\rho^2)\alpha\beta}\right) + \\ & + \frac{1}{2}\log\left(\beta(1-\sigma^2)\right) + \frac{1}{2}\log\left((1-\rho^2)\alpha S_{13}+1\right), \end{aligned} \tag{14}$$

for some $0 \leq \alpha, \beta \leq 1$, $\rho \in [-1,1]$ such that $(1-\alpha)(1-\beta) \geq \sigma^2\alpha\beta$, where

$$\sigma = \frac{(1-\rho^2)\alpha S_{13}+1}{2T\sqrt{S_{12}(1-\rho^2)\alpha\beta}} - \frac{(1-\rho^2)\alpha S_{12}+\beta}{2\sqrt{S_{12}(1-\rho^2)\alpha\beta}},$$

and

$$T = \min\left[\frac{1+S_{13}+S_{23}+2\rho\sqrt{S_{13}S_{23}}}{(1-\rho^2)S_{12}+1}, \quad \lambda_{\max}\right],$$

where $\lambda_{\max}$ is the larger root of the quadratic equation

$$2\rho\sqrt{S_{13}S_{23}} + S_{13} + S_{23} + 1 - \lambda\big(S_{23}S_{12}(1-\rho^2) + S_{13} + S_{23} + S_{12} + 2 + 2\rho\sqrt{S_{13}S_{23}}\big) + \lambda^2(S_{12}+1) = 0. \tag{15}$$

The proof of this theorem is given in Section 3.2. As an outline, recall that the cutset bound in (9) is maximized by a jointly Gaussian X, X_{r} and that at the maximizing distribution, it coincides with the broadcast bound $I(X;Y,Y_{\mathrm{r}}|X_{\mathrm{r}})$ [EK11, Section 16.5]. Our bound involves the inequality

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

If the cutset bound is equal to the new upper bound, then it must be that $I(V;Y|X_{\mathrm{r}},Y_{\mathrm{r}}) = I(X;Y_{\mathrm{r}}|V,X_{\mathrm{r}},Y) = 0$. These two conditions are shown to imply that given $X_{\mathrm{r}} = x_{\mathrm{r}}$, $V \to Y_{\mathrm{r}} \to X$ and $Y_{\mathrm{r}} \to V \to X$. Given $X_{\mathrm{r}} = x_{\mathrm{r}}$, this "double Markovity condition" implies that X is independent of (V, Y_{r}). However, this is a contradiction since X is not independent of Y_{r} given $X_{\mathrm{r}} = x_{\mathrm{r}}$. This allows us to conclude that the new bound strictly improves upon the cut-set bound. The rest of the proof uses the sub-additivity, doubling, and rotation techniques from [GN14] to show that our upper bound is maximized by jointly Gaussian random variables.

Figure 3 compares the bound in Corollary 1 to the bound in [GN20, Proposition 1] for the scalar Gaussian relay channel, the cutset bound, and the compress-forward lower bound (evaluated using Gaussian distributions). Note that for the example in Figure 3, compress-forward outperforms decode-forward.

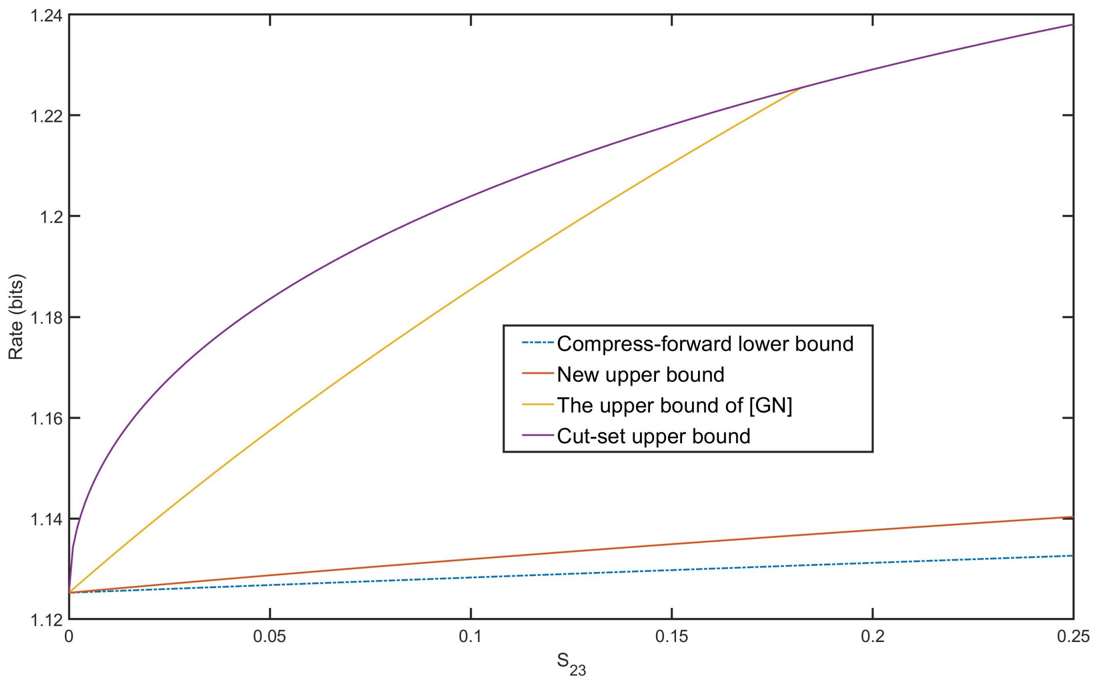

Fig. 3. Plots of the bounds for the Gaussian relay channel with $S_{13} = 3.7585$, $S_{12} = 1.2139$. The new upper bound is the evaluation of the bound in Corollary 1.

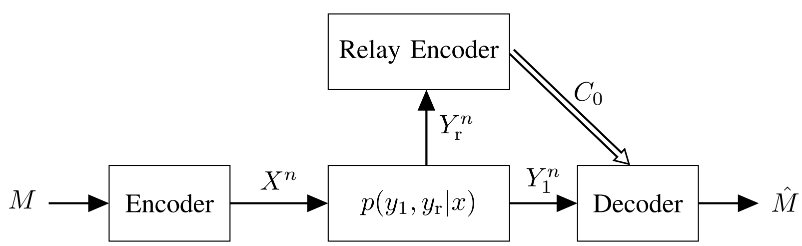

Fig. 4. Relay channel with orthogonal receiver components.

2.2. Relay channels with orthogonal receiver components

In this section we present results for the following sub-class of relay channels without self-interference.

Definition 1. A relay channel is said to be *with orthogonal receiver components* (also referred to as *primitive*) (see Section 16.7.3 in [EK11]) if $Y = (Y_1, Y_2)$, where $p(y_1, y_2, y_{\mathrm{r}}|x, x_{\mathrm{r}}) = p(y_1, y_{\mathrm{r}}|x)p(y_2|x_{\mathrm{r}})$. It is known that the capacity of the above relay channel depends on $p(y_2|x_{\mathrm{r}})$ only via the capacity of the point-to-point channel $p(y_2|x_{\mathrm{r}})$, hence we can substitute the relay-to-receiver channel $p(y_2|x_{\mathrm{r}})$ with a noiseless link of the same capacity C_0 [Kim07] as shown in Figure 4.

The following provides an equivalent characterization of the upper bound in Theorem 1 for relay channels with orthogonal receiver components.

Proposition 1. *Any achievable rate R for the relay channel with orthogonal receiver components $p(y_1, y_{\mathrm{r}}|x)$ with a relay-to-receiver link of capacity C_0 must satisfy the following conditions*

$$\begin{aligned} R &\leq I(X; Y_1, Y_{\mathrm{r}}) - I(V; Y_1|Y_{\mathrm{r}}) - I(X; Y_{\mathrm{r}}|V, Y_1) && (16) \\ &= I(X; Y_1, V) - I(V; X|Y_{\mathrm{r}}), && (17) \end{aligned}$$

for some $p(x)p(y_1, y_\mathrm{r}|x)p(v|x, y_\mathrm{r})$ *such that*

$$I(V; Y_\mathrm{r}) - I(V; Y_1) \leq C_0. \tag{18}$$

An equivalent form the bound without constraints is as follows: a rate R *is achievable if*

$$R \leq I(X; Y_1, Y_\mathrm{r}) - I(V; Y_1|Y_\mathrm{r}) - I(X; Y_\mathrm{r}|V, Y_1) = I(X; Y_1, V) - I(V; X|Y_\mathrm{r}) \tag{19}$$

$$R \leq I(X; Y_1) + C_0 - I(V; Y_\mathrm{r}|X, Y_1) \tag{20}$$

for some $p(x)p(y_1, y_\mathrm{r}|x)p(v|x, y_\mathrm{r})$*.*

Further it suffices to consider $|\mathcal{V}| \leq |\mathcal{X}||\mathcal{Y}_\mathrm{r}| + 1$ *when evaluating either form of the bound.*

The proof of this proposition is given in Section 3.3.

Remark 8. The only difference between the upper bound in Proposition 1 and the equivalent forms of the compress-forward lower bound in [EK11, Eq. 16.14] and [Kim07, Proposition 3] is that the upper bound takes the maximum over $p(x)p(v|x, y_\mathrm{r})$ while the compress-forward lower bound takes maximum of the same expression over $p(x)p(v|y_\mathrm{r})$.

Kim's conjecture. We use Proposition 1 to prove a conjecture posed by Kim in [Kim07, Question 2] for a class of deterministic relay channels with orthogonal receiver components described by $p(y_1, y_\mathrm{r}|x)$, where $X = f(Y_1, Y_\mathrm{r})$ for some function f.

Theorem 3. *Let* $\mathcal{C}(C_0)$ *be the supremum of achievable rates* R *for a given* C_0*. Let* C_0^* *be the minimum value of* C_0 *for which* $\mathcal{C}(C_0) = \mathcal{C}(\infty) = \log|\mathcal{X}|$*. Then* $C_0^* = H_G(Y_\mathrm{r}|Y_1)$ *and is achieved by a uniform distribution on* X*. Here* $H_G(Y_\mathrm{r}|Y_1)$ *denotes the conditional graph entropy of the characteristic graph of* (Y_r, Y_1) *and the function* f *(as defined in [OR95]).*

The proof of this theorem is given in Section 3.4. As an outline, consider the following constraint in the upper bound of Proposition 1:

$$R \leq I(X; Y_1, Y_\mathrm{r}) - I(V; Y_1|Y_\mathrm{r}) - I(X; Y_\mathrm{r}|V, Y_1). \tag{21}$$

If $\max_{p(x)} I(X; Y_1, Y_\mathrm{r})$ is achievable, it forces

$$I(V; Y_1|Y_\mathrm{r}) = I(X; Y_\mathrm{r}|V, Y_1) = 0. \tag{22}$$

Since X is a function of (Y_1, Y_r), $I(V; Y_1|Y_\mathrm{r}) = I(V; X, Y_1|Y_\mathrm{r}) \geq I(V; X|Y_\mathrm{r})$. This yields $I(V; X|Y_\mathrm{r}) = 0$, showing that the upper bound reduces to compress-forward lower bound (see Remark 8).

2.3. Product-form relay channels

Consider the following class of relay channels with orthogonal receiver components.

Definition 2. A relay channel with orthogonal receiver components is said to be *product-form* if $p(y_1, y_\mathrm{r}|x) = p(y_1|x)p(y_\mathrm{r}|x)$.

Remark 9. In [Zha88], Zhang provides a bound for a particular class of product-form relay channels for which the rate $\max_{p(x)} I(X; Y_1) + C_0$ is achievable. Proposition 1 recovers and generalizes his result. A similar argument as in Remark 5 shows that for any arbitrary product-form relay channel, the rate $\max_{p(x)} I(X; Y_1) + C_0$ is achievable if and only if this rate is achievable by partial-decode-forward.

We will also consider the following special case of the above class.

Definition 3. A product-form relay channel is said to be *symmetric* if $\mathcal{Y}_1 = \mathcal{Y}_\mathrm{r}$, $p(y_\mathrm{r}, y_1|x) = p(y_\mathrm{r}|x)p(y_1|x)$, and $p_{Y_1|X}(y|x) = p_{Y_\mathrm{r}|X}(y|x)$ for all x, y.

In the following we consider several applications for these classes of relay channels.

Generalized Cover Relay Channel Problem. We will need the following definitions.

Definition 4. A discrete-memoryless channel $p(y|x)$ is said to be *generic* if the channel matrix, P with entries $P_{x,y} = p(y|x)$, $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ is full row rank.

Remark 10. It is immediate that if $p(y_1|x_1)$ and $p(y_2|x_2)$ are generic, then so is $p(y_1|x_1) \otimes p(y_2|x_2)$. That is, the class of generic channels is closed under a product operation.

Definition 5. A product-form relay channel is said to be *generic* if the channel $p(y_1|x)$ is generic.

In [Cov87], Cover posed a special (symmetric) case of the following problem: Consider a generic product-form relay channel and let $\mathcal{C}(C_0)$ be the supremum of achievable rates R for a given C_0. What is the critical value C_0^* for which $C_0^* = \inf\{C_0 : \mathcal{C}(C_0) = \mathcal{C}(\infty) = \max_{p(x)} I(X;Y_1,Y_\mathrm{r})\}$?

This problem has recently attracted a fair amount of attention and non-traditional methods have been used to answer the question as well as obtain new upper bounds for $\mathcal{C}(C_0)$ for symmetric Gaussian channels and binary-symmetric channels. As we will show in the next subsections our new upper bound, which uses traditional converse techniques, recovers and (in the binary-symmetric case) improves on these recent results. In this section, we show that we can answer the generalized Cover relay channel problem.

We can answer the generalized Cover's open problem by evaluating the bound in Proposition 1.

Theorem 4. *Let C_0^* be the minimum value of C_0 such that $\mathcal{C}(C_0) = \mathcal{C}(\infty) = \max_{p(x)} I(X;Y_1,Y_\mathrm{r})$ for a generic product-form relay channel and R_0^* be the minimum value of C_0 such that $R_{\mathsf{CF}}(C_0) = \mathcal{C}(\infty) = \max_{p(x)} I(X;Y_1,Y_\mathrm{r})$ for the same relay channel. Then $C_0^* = R_0^*$.*

The proof of this theorem is given in Section 3.5. The proof is similar to that of Theorem 3. The generic assumption makes it possible to use the condition in (22) to deduce that $I(V;X|Y_\mathrm{r}) = 0$.

Remark 11. During the finalization of this manuscript the authors became aware of [Liu20] which uses results and techniques in convex geometry to arrive at a solution for Theorem 4. A mechanical glance at Theorem 1 of [Liu20] indicates that the outer bound was generated for the sole purpose of identifying C_0^* rather than providing an explicit bound on the capacity of the channel. In contrast, the result here follows from the upper bound established in Proposition 1.

Gaussian product-form relay channel. The Gaussian product-form relay channel is defined as

$$\begin{aligned} Y_1 &= X + W_1, \\ Y_\mathrm{r} &= X + W_\mathrm{r}, \end{aligned}$$

where $W_1 \sim \mathcal{N}(0, N_1)$ and $W_\mathrm{r} \sim \mathcal{N}(0, N_\mathrm{r})$ are independent of each other and of X, and a link of capacity C_0 from the relay to the destination. We assume average power constraint P on X and define $S_{12} = P/N_\mathrm{r}$, $S_{13} = P/N_1$ and $S_{23} = 2^{2C_0} - 1$.

The upper bound in Proposition 1 reduces to the following.

Proposition 2. *Any achievable rate R for the Gaussian product-form relay channel must satisfy the condition*

$$R \le \begin{cases} \frac{1}{2}\log\left(1 + S_{13} + \frac{S_{12}(S_{13}+1)S_{23}}{(S_{13}+1)(S_{23}+1)-1}\right), & \text{for } S_{12} \le S_{13} + S_{23} + S_{13}S_{23}, \\ \frac{1}{2}\log((1+S_{13})(1+S_{23})), & \text{otherwise.} \end{cases} \tag{23}$$

The proof of this proposition is given in Section 3.6. We use ideas from [GN14] to show the optimality of jointly Gaussian $(V, X, Y_1, Y_\mathrm{r})$ in computing the upper bound of Proposition 1, leading to an optimization problem with only three free variables. We reduce this to a one-dimensional optimization problem by observing two properties of any maximizer: (i) equation (18) holds with equality for the maximizing distribution and (ii) $K_{X,Y_\mathrm{r}} - K_{X,Y_\mathrm{r}|V}$ is a rank one matrix, where K_{X,Y_r} is the covariance matrix of (X, Y_r) and $K_{X,Y_\mathrm{r}|V}$ is the conditional covariance matrix of (X, Y_r) given V.

Remark 12. The decode-forward lower bound for the Gaussian product-form relay channel in [EK11, Eq. 16.16] is

$$C \ge \begin{cases} \frac{1}{2}\log(1 + S_{12}), & \text{for } S_{12} \le S_{13} + S_{23} + S_{13}S_{23}, \\ \frac{1}{2}\log((1+S_{13})(1+S_{23})), & \text{otherwise.} \end{cases} \tag{24}$$

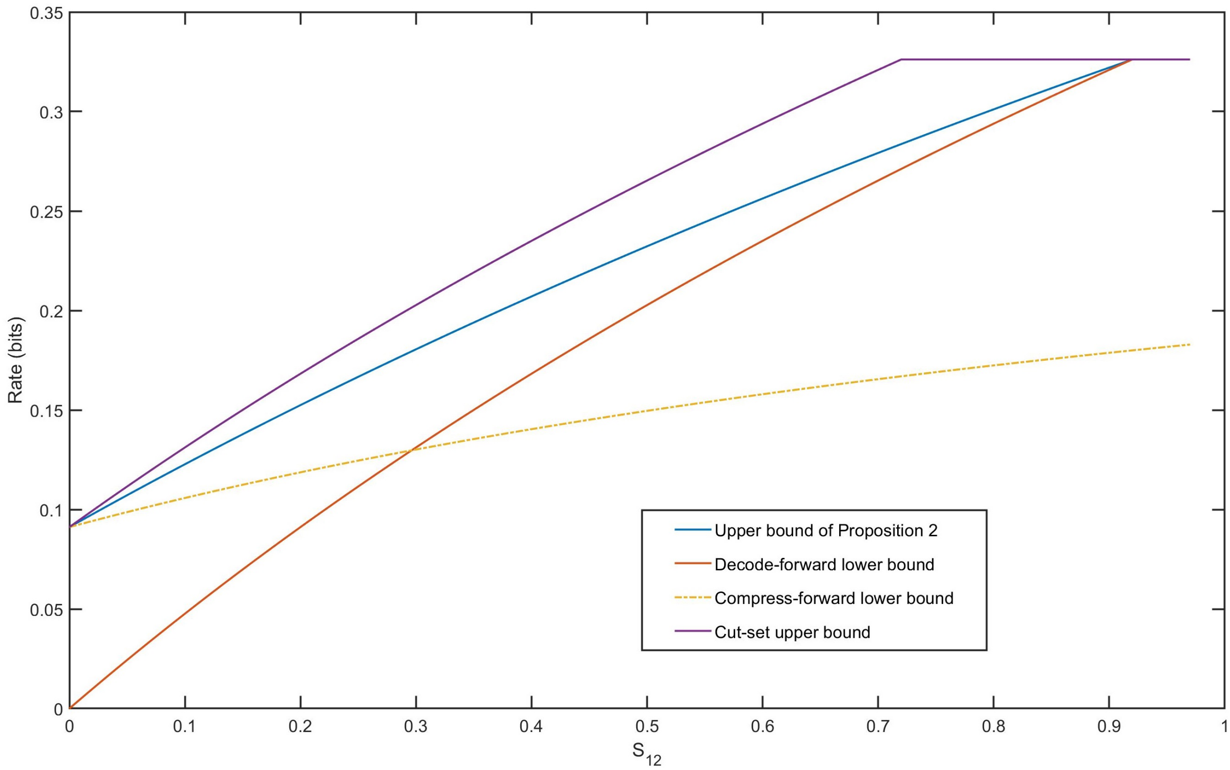

Fig. 5. Plots of the bounds for the Gaussian product-form relay channel with $S_{13} = 0.2$, $S_{23} = 0.6$.

When $S_{12} \geq S_{13} + S_{23} + S_{13}S_{23}$, the upper bound in (23), the cut-set bound and the decode-forward lower bound all coincide. The condition $S_{12} \leq S_{13} + S_{23} + S_{13}S_{23}$ is equivalent to $I(X; Y_{\mathrm{r}}) \leq I(X; Y_1) + C_0$ for a Gaussian input $X \sim \mathcal{N}(0, P)$. The compress-forward lower bound for this relay channel in [EK11, Eq. 16.17] implies that

$$C \geq \frac{1}{2} \log\left(1 + S_{13} + \frac{S_{12}(S_{13}+1)S_{23}}{S_{12} + (S_{13}+1)(S_{23}+1)}\right). \tag{25}$$

Furthermore, this lower bound can be improved by time-sharing at the transmitter [EK11, Sec. 16.8] or at the relay [WBO19, Footnote 2].

Figure 5 plots the upper bound in Proposition 2 along with the cut-set upper bound and the compress-forward lower bound for an example Gaussian product-form relay channel.

In [WBÖ17], the following upper bound on the capacity of the Gaussian product-form relay channel is established.

Theorem 5 ([WBÖ17])**.** *For the Gaussian product-form relay channel, any achievable rate R must satisfy the condition*

$$R \leq \frac{1}{2}\log(1+S_{13}) + \sup_{\theta \in \left[\arcsin\left(\frac{1}{1+S_{23}}\right), \frac{\pi}{2}\right]} \min\left\{C_0 + \log\sin\theta, \min_{\omega \in \left(\frac{\pi}{2}-\theta, \frac{\pi}{2}\right]} h_\theta(\omega)\right\}, \tag{26}$$

where

$$h_\theta(\omega) = \frac{1}{2}\log\left(\frac{\left(S_{12} + S_{13} + \sin^2(\omega) - 2\cos(\omega)\sqrt{S_{12}S_{13}}\right)\sin^2\theta}{(S_{13}+1)(\sin^2\theta - \cos^2\omega)}\right). \tag{27}$$

Although the techniques used to prove this theorem are completely different from those used in this paper, it turns out quite surprisingly that bound (26) coincides with the bound in Proposition 2.

Lemma 1. *The bound in Theorem 5 simplifies to*

$$R \leq \begin{cases} \frac{1}{2}\log\left(1+S_{13}+\frac{S_{12}(S_{13}+1)S_{23}}{(S_{13}+1)(S_{23}+1)-1}\right), & \text{for } S_{12} \leq S_{13}+S_{23}+S_{13}S_{23}, \\ \frac{1}{2}\log((1+S_{13})(1+S_{23})), & \textit{otherwise}, \end{cases} \tag{28}$$

which coincides with the bound in Proposition 2.

The proof of this lemma is given in Section 3.7.

Symmetric binary relay channel with orthogonal receiver components. Consider a symmetric product-form relay channel described by $p(y_1, y_{\mathrm{r}}|x)$ and a link of rate C_0 from relay to the destination such that $p(y_1, y_{\mathrm{r}}|x) = p(y_1|x)p(y_{\mathrm{r}}|x)$, where $x, y_1, y_{\mathrm{r}} \in \{0, 1\}$. Further, assume that both the channels $p(y_1|x)$ and $p(y_{\mathrm{r}}|x)$ are binary symmetric channels with crossover probability $\rho \in [0, 1/2]$. By specializing Proposition 1 to this channel, we obtain the following bound.

Theorem 6. *Given arbitrary $\lambda \in [0,1]$ and $c \in [0,1]$, let $g_\lambda(c)$ be the maximum of $(1-\lambda)\left(H(Y_1) - H(Y_{\mathrm{r}})\right) + H(Y_{\mathrm{r}}|X)$ over all joint probability distributions $p(x, y_{\mathrm{r}})$ on $\{0,1\} \times \{0,1\}$ satisfying $p(x, y_{\mathrm{r}})(0,1) + p(x, y_{\mathrm{r}})(1,0) = c$. For any fixed $\lambda \in [0,1]$, let $\mathscr{C}[g_\lambda] : [0,1] \mapsto \mathbb{R}$ be the upper concave envelope of the function $g_\lambda(\cdot)$, i.e., the smallest concave function that dominates $g_\lambda(\cdot)$ from above. Any achievable rate R for a symmetric binary relay channel with orthogonal receiver components with parameter ρ must satisfy the following*

$$R \leq 1 - 2H_2(\rho) + \lambda C_0 + \mathscr{C}[g](\rho)$$

for any $\lambda \in [0,1]$, where $H_2(x) = -x\log(x) - (1-x)\log(1-x)$ is the binary entropy function.

The proof of this theorem is given in Section 3.8.

Figure 6 shows that our new upper bound strictly improves upon the bounds given in [WOX17] and [BWÖ17].

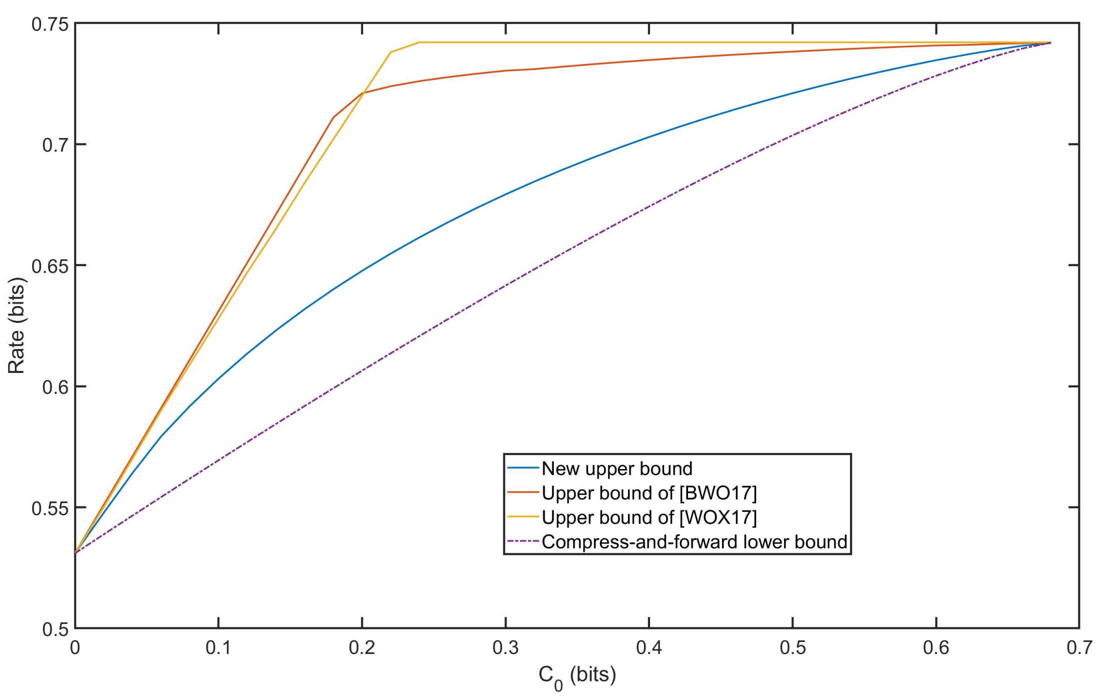

Fig. 6. Plots of the minimum of the two upper bounds give in [WOX17], the upper bound given in [BWÖ17], our new bound and the compress-forward lower bound for a symmetric binary relay channel with orthogonal receiver components with parameter $\rho = 0.1$.

2.4. Upper bound with auxiliary receiver and applications

We establish an upper bound on the capacity of the relay channel with orthogonal receiver components which uses an auxiliary receiver [GN20]. We use this bound to to obtain a tighter upper bound than the one given in [TU08] and to show the suboptimality of our main bound in Proposition 1.

As in [GN20], the auxiliary receiver J is an enhancement of the relay's output variable Y_{r}. However, the identification of the auxiliary variables in the following upper bound are different from those in [GN20].

Theorem 7. *Consider an arbitrary relay channel with orthogonal receiver components. Assume that for an auxiliary receiver J we have $p(j, y_{\mathrm{r}}, y_1|x) = p(j|x)p(y_{\mathrm{r}}|j)p(y_1|y_{\mathrm{r}}, j, x)$. Then any achievable rate R must satisfy the condition*

$$R \leq I(X; J) + I(X; Y_1|V, W, J, T) - I(X; Y_{\mathrm{r}}|V, W, J, T),$$

for some $p(t, x)p(j|x)p(y_{\mathrm{r}}|j)p(y_1|j, y_{\mathrm{r}}, x)p(w|t, y_{\mathrm{r}})p(v|t, x, y_{\mathrm{r}}, w)$ such that

$$I(V, W; Y_{\mathrm{r}}|J, T) - I(V, W; Y_1|J, T) \leq I(W; Y_{\mathrm{r}}|J, T) \leq C_0.$$

Further it suffices to consider $|\mathcal{T}| \leq |\mathcal{X}| + 3$, $|\mathcal{W}| \leq |\mathcal{Y}_{\mathrm{r}}| + 3$ and $|\mathcal{V}| \leq |\mathcal{X}||\mathcal{Y}_{\mathrm{r}}| + 1$.

The proof of this theorem is given in Section 3.9. As an outline, let $M_{\mathrm{r}} \in [1 : 2^{2C_0}]$ be a function of Y_{r}^n that can be viewed as the "message" from the relay to the receiver. While we derived the bound in Proposition 1 as a special case of the main bound in Theorem 1, it could also be directly established via the identification $V_i = (Y_{1i+1}^n, Y_{\mathrm{r}}^{i-1}, M_{\mathrm{r}})$. In Theorem 7, we instead use the following finer identification

$$V_i = Y_{1i+1}^n, \quad W_i = (M_{\mathrm{r}}, Y_{\mathrm{r}}^{i-1}), \quad T_i = (J^{i-1}, J_{i+1}^n),$$

Crucially, the introduction of the auxiliary receiver J and the auxiliary random variable T allows us to induce certain Markov chains and constraints on V and W that would otherwise not be possible to obtain.

Relay channels with i.i.d. output sequence. Consider the following class of relay channels.

Definition 6. A relay channel with orthogonal receiver components is said to be with i.i.d. output if its family of conditional probabilities have the form $p(y, y_{\mathrm{r}}|x, x_{\mathrm{r}}) = p(y_{\mathrm{r}})p(y_1|x, y_{\mathrm{r}})p(y_2|x_{\mathrm{r}})$.

Remark 13. In [ARY09b], a special case of this channel in which the channel from the transmitter to the receiver is a binary symmetric channel and the relay observes a corrupted version of the transmitter-receiver channel noise was used to establish the suboptimality of the cutset bound.

Remark 14. Communication over this relay channel is equivalent to communication over channels with rate-limited state information available at the receiver [ARY09b]. The Ahlswede-Han conjecture [AH83] for the channels with rate-limited state information is equivalent to the capacity for this relay channel with orthogonal receiver components being equal to $\max I(X; Y_1|W)$, where the maximum is over $p(x)p(y_{\mathrm{r}})p(y_1|x, y_{\mathrm{r}})p(w|y_{\mathrm{r}})$ such that $I(W; Y_{\mathrm{r}}|Y_1) \leq C_0$.

For the relay channel with orthogonal receiver components and i.i.d. output, we obtain the following corollary to Theorem 7.

Corollary 2. *Any achievable rate R for the relay channel with orthogonal receiver components and i.i.d. output must satisfy the condition*

$$R \leq I(X; Y_1|T, W, V) - I(X; Y_{\mathrm{r}}|T, W, V),$$

for some $p(t, x)p(y_1|x, y_{\mathrm{r}})p(y_{\mathrm{r}})p(w|t, y_{\mathrm{r}})p(v|t, x, y_{\mathrm{r}}, w)$ such that

$$I(V, W; Y_{\mathrm{r}}|T) - I(V, W; Y_1|T) \leq I(W; Y_{\mathrm{r}}|T) \leq C_0.$$

The proof of this corollary follows by setting the auxiliary receiver J to be a constant in the statement of Theorem 7. This choice is feasible because Y_{r} is independent of X.

In [TU08], the following upper bound on the capacity of the above relay channel is given.

Theorem 8 ([TU08]). *Any achievable rate R for the relay with orthogonal receiver components channel with i.i.d. output must satisfy the condition*

$$R \leq \min\{I(W, X; Y_1|T), I(X; Y_1|Y_{\mathrm{r}}, T)\}$$

for some $p(x,t)p(y_1|x,y_{\rm r})p(y_{\rm r})p(w|y_{\rm r},t)$ such that $I(W;Y_{\rm r}|T) \leq C_0$.

Remark 15. The time-sharing random variable T was not included in the statement of the above theorem in [TU08]. We believe it is necessary, however, since the Markov chain $(W_T, T) \to Y_{{\rm r},T} \to X_T$ does not hold with the identification of auxiliary random variables in [TU08]. The upper bound in Theorem 8 with T still provides a converse for the example considered in [ARY09b].

Next, we show that the bound in Corollary 2 strictly improves over the bound in Theorem 8.

Proposition 3. *The bound in Corollary 2 is less than or equal to the bound in Theorem 8. Moreover, the inclusion is strict for the following class of relay channels. Consider a relay channel with i.i.d. output sequence $p(y_{\rm r})p(y_1|x,y_{\rm r})$ such that the channel $p(y_1|x,Y_{\rm r}=y_{\rm r})$ is generic for every $y_{\rm r}$, and for which there is no random variable K such that $H(K|X,Y_1) = H(K|X,Y_{\rm r}) = 0$ but $H(K|X) > 0$. For such a channel, if the compress-forward lower bound with time-sharing does not match the bound in Theorem 8, then the bound in Corollary 2 strictly improves upon the bound in Theorem 8.*

The proof of this proposition is given in Section 3.10.

We can show that the class for which the inclusion is strict is non-empty. In [TU08, Section VII] it is shown that for a channel of the form $Y_1 = XY_{\rm r} + N$, where N and $Y_{\rm r}$ are Bernoulli random variables independent of the binary X, the compress-forward lower bound (without time-sharing) is significantly below the outer bound in Theorem 8. Further from Figure 4 in [TU08], it is evident that even with time-sharing the compress-forward lower bound continues to be below the bound in Theorem 8. Now consider the perturbed version of the channel $Y_1 = (XY_{\rm r}+N)Z+(X+N)(1-Z)$, where Z is independent of all other random variables and $\mathrm{P}(Z=1) = 1-\epsilon$. Observe that this perturbed channel satisfies the assumptions in Proposition 3; however, by the continuity of the various bounds in the channel parameters, the gap between the compress-forward rate (with time-sharing) and the upper bound continues to be non-zero for small enough values of ϵ.

Auxiliary receiver can help. Consider the Gaussian relay channel with orthogonal receiver components and i.i.d. output for which

$$Y_{\rm r} = Z_{\rm r}, \tag{29}$$
$$Y_1 = X + Y_{\rm r} + Z_1, \tag{30}$$

where $Z_1 \sim \mathcal{N}(0,N_1)$ and $Z_{\rm r} \sim \mathcal{N}(0,N_{\rm r})$ are independent of each other and of X. Assume average power constraint P on X. For this class of relay channels, we show that the bound with auxiliary receiver is strictly tighter than our main upper bound with no auxiliary receivers.

Theorem 9. *The bound in Corollary 2 strictly improves over the bound given in Proposition 1 for the channel given by (29)-(30) for any $N_1, N_{\rm r}, C_0 > 0$.*

The proof of this theorem is given in Section 3.11.

3. Proofs of the results

In the following sections we present the proofs of the results stated in the previous section in their order of appearance.

3.1. Proof of Theorem 1

We will first prove the theorem with the original constraints and then show the equivalence of the region to that obtained by dropping the constraint on R in (6) and instead strengthening (7) to (8).

Take an arbitrary code of length n inducing a joint distribution on $(M, X^n, X_{\rm r}^n, Y^n, Y_{\rm r}^n)$. The proof essentially follows from routine manipulations along with the following identifications:

$$V_i = (Y_{i+1}^n, Y_{\rm r}^{i-1}), \quad U_i = (Y_{{\rm r}i+1}^n, Y^{i-1}),$$

and $V = (Q, V_Q)$, $U = (Q, U_Q)$, $X = X_Q, X_{\rm r} = X_{{\rm r}Q}, Y = Y_Q, Y_{\rm r} = Y_{{\rm r}Q}$ for a time-sharing random variable $Q \stackrel{(d)}{=} \text{Uniform}[1:n]$. We employ the data-processing inequality, chain-rule, and the Csiszár-Körner-Marton identity repeatedly as is done usually. It may be useful for the readers to have the Bayesian Network diagram in Figure 7 and

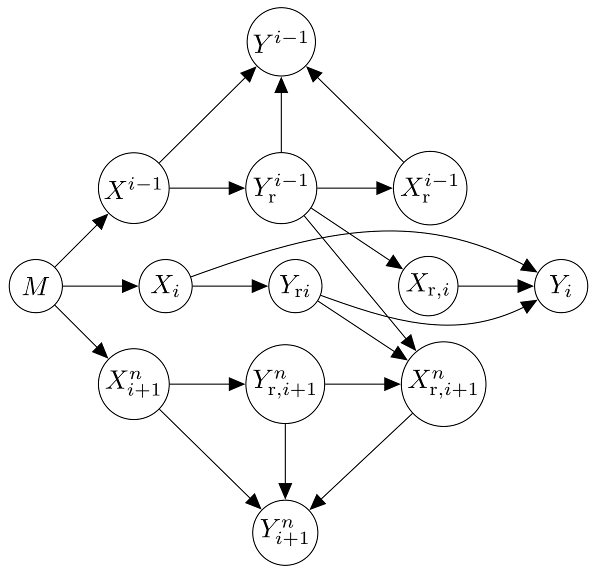

Fig. 7. A Bayesian network that contains causal relationships in the probability distribution induced by the codebooks for the relay channel without self-interference.

the d-separation theorem for deducing the various Markov chains employed. The Markov chain $V \to (X, X_{\mathrm{r}}, Y_{\mathrm{r}}) \to Y$ follow from the fact that the channel $p(y|x, x_{\mathrm{r}}, y_{\mathrm{r}})$ is memoryless. The Markov chain $U \to (X, X_{\mathrm{r}}) \to (Y, Y_{\mathrm{r}})$ for relay channels of the form $p(y_{\mathrm{r}}|x)p(y|x, x_{\mathrm{r}}, y_{\mathrm{r}})$ can be deduced from Fig. 7.

Since the expressions in the statement of the theorem depend only on the marginal distributions of $p(u, x, x_{\mathrm{r}})$ and $p(v|x, x_{\mathrm{r}}, y_{\mathrm{r}})$ the union, in the statement of the theorem, is taken over $p(u, x, x_{\mathrm{r}})p(y, y_{\mathrm{r}}|x, x_{\mathrm{r}})p(v|x, x_{\mathrm{r}}, y_{\mathrm{r}})$.

To show (1), we write

$$I(M; Y^n) = I(M; Y_{\mathrm{r}}^n) + I(M; Y^n) - I(M; Y_{\mathrm{r}}^n).$$

Following the steps in the proof of the cut-set bound, we obtain

$$I(M; Y_{\mathrm{r}}^n) \leq \sum_{i=1}^{n} I(X_i; Y_{\mathrm{r}i}|X_{\mathrm{r}i}).$$

On the other hand, we have

$$\begin{aligned}
I(M; Y^n) - I(M; Y_{\mathrm{r}}^n) &\leq I(M; Y^n|Y_{\mathrm{r}}^n) \\
&= \sum_{i=1}^{n} I(M; Y_i|Y^{i-1}, Y_{\mathrm{r}}^n) \\
&= \sum_{i=1}^{n} I(M, X_i; Y_i|X_{\mathrm{r}i}, Y^{i-1}, Y_{\mathrm{r}}^n) \\
&\overset{(a)}{=} \sum_{i=1}^{n} I(X_i; Y_i|X_{\mathrm{r}i}, Y^{i-1}, Y_{\mathrm{r}}^n) \\
&\overset{(b)}{\leq} \sum_{i=1}^{n} I(X_i; Y_i|X_{\mathrm{r}i}, Y^{i-1}, Y_{\mathrm{r}i}^n) \\
&= \sum_{i=1}^{n} I(X_i; Y_i|X_{\mathrm{r}i}, Y_{\mathrm{r}i}, U_i),
\end{aligned}$$

where in (a) and (b) we use the Markov chain $(Y_{\mathrm{r}}^{i-1}, Y_{\mathrm{r}i+1}^{n}, M, Y^{i-1}) \to (X_i, X_{\mathrm{r}i}, Y_{\mathrm{r}i}) \to Y_i$. This Markov chain follows from the fact that the channel $p(y|x, x_{\mathrm{r}}, y_{\mathrm{r}})$ is memoryless. Thus, we obtain

$$\begin{aligned}
I(M;Y^n) &= I(M;Y_{\mathrm{r}}^n) + I(M;Y^n) - I(M;Y_{\mathrm{r}}^n) \\
&\leq \sum_{i=1}^{n} I(X_i;Y_{\mathrm{r}i}|X_{\mathrm{r}i}) + I(X_i;Y_i|X_{\mathrm{r}i},Y_{\mathrm{r}i},U_i) \\
&\leq n\left(I(X;Y_{\mathrm{r}}|X_{\mathrm{r}}) + I(X;Y|X_{\mathrm{r}},Y_{\mathrm{r}},U)\right).
\end{aligned}$$

Next, we show (4). Observe that by definition, V_i implies that $X_{\mathrm{r}i}$. We have

$$\begin{aligned}
\frac{1}{n}I(M;Y^n) &\leq \frac{1}{n}\left(I(M;Y_{\mathrm{r}}^n) + I(M;Y^n) - I(M;Y_{\mathrm{r}}^n)\right) \\
&\leq I(X;Y_{\mathrm{r}}|X_{\mathrm{r}}) + \frac{1}{n}\left(\sum_i I(M;Y_i|Y_{\mathrm{r}}^{i-1},Y_{i+1}^n) - \sum_i I(M;Y_{\mathrm{r}i}|Y_{\mathrm{r}}^{i-1},Y_{i+1}^n)\right) \\
&= I(X;Y_{\mathrm{r}}|X_{\mathrm{r}}) + \frac{1}{n}\left(\sum_i I(M,X_i;Y_i|Y_{\mathrm{r}}^{i-1},X_{\mathrm{r}i},Y_{i+1}^n) - \sum_i I(M,X_i;Y_{\mathrm{r}i}|Y_{\mathrm{r}}^{i-1},X_{\mathrm{r}i},Y_{i+1}^n)\right) \\
&= I(X;Y_{\mathrm{r}}|X_{\mathrm{r}}) + \frac{1}{n}\left(\sum_i I(X_i;Y_i|Y_{\mathrm{r}}^{i-1},X_{\mathrm{r}i},Y_{i+1}^n) + \sum_i I(M;Y_i|X_i,Y_{\mathrm{r}}^{i-1},X_{\mathrm{r}i},Y_{i+1}^n)\right. \\
&\qquad\qquad \left. - \sum_i I(X_i;Y_{\mathrm{r}i}|Y_{\mathrm{r}}^{i-1},X_{\mathrm{r}i},Y_{i+1}^n) - \sum_i I(M;Y_{\mathrm{r}i}|X_i,Y_{\mathrm{r}}^{i-1},X_{\mathrm{r}i},Y_{i+1}^n)\right) \\
&\leq I(X;Y_{\mathrm{r}}|X_{\mathrm{r}}) + \frac{1}{n}\left(\sum_i I(X_i;Y_i|Y_{\mathrm{r}}^{i-1},X_{\mathrm{r}i},Y_{i+1}^n) - \sum_i I(X_i;Y_{\mathrm{r}i}|Y_{\mathrm{r}}^{i-1},X_{\mathrm{r}i},Y_{i+1}^n)\right. \\
&\qquad\qquad \left. + \sum_i I(M;Y_i|X_i,Y_{\mathrm{r}}^{i-1},X_{\mathrm{r}i},Y_{i+1}^n,Y_{\mathrm{r}i})\right) \\
&= I(X;Y_{\mathrm{r}}|X_{\mathrm{r}}) + I(X;Y|V,X_{\mathrm{r}}) - I(X;Y_{\mathrm{r}}|V,X_{\mathrm{r}}),
\end{aligned} \tag{31}$$

where in equation (31), we used the fact that

$$I(M;Y_i|X_i,X_{\mathrm{r}i},Y_{i+1}^n,Y_{\mathrm{r}}^{i-1},Y_{\mathrm{r}i}) = 0.$$

To show (6), note that

$$\begin{aligned}
\frac{1}{n}I(M;Y^n) &\leq \frac{1}{n}I(X^n;Y^n) \\
&\leq \frac{1}{n}\sum_i I(X^n,Y_{i+1}^n;Y_i) \\
&= \frac{1}{n}\sum_i I(X_i;Y_i) + \frac{1}{n}\sum_i I(X^{n\setminus i},Y_{i+1}^n;Y_i|X_i) \\
&\leq \frac{1}{n}\left(\sum_i I(X_i;Y_i) + I(V_i,X_{\mathrm{r}i};Y_i|X_i) - I(V_i;Y_{\mathrm{r}i}|X_i,X_{\mathrm{r}i})\right) && (32) \\
&\leq I(X;Y) + I(V,X_{\mathrm{r}};Y|X) - I(V;Y_{\mathrm{r}}|X,X_{\mathrm{r}}) \\
&= I(X,X_{\mathrm{r}};Y) + I(V;Y|X,X_{\mathrm{r}}) - I(V;Y_{\mathrm{r}}|X,X_{\mathrm{r}}) \\
&= I(X,X_{\mathrm{r}};Y) - I(V;Y_{\mathrm{r}}|X_{\mathrm{r}},X,Y), && (33)
\end{aligned}$$

where (33) follows from the Markov chain relationship $V \to (X, X_{\mathrm{r}}, Y_{\mathrm{r}}) \to Y$ and (32) follows from

$$\sum_i I(Y_{\mathrm{r}}^{i-1},Y_{i+1}^n;Y_{\mathrm{r}i}|X_i,X_{\mathrm{r}i}) \stackrel{(a)}{=} \sum_i I(X^{n\setminus i},Y_{\mathrm{r}}^{i-1},Y_{i+1}^n;Y_{\mathrm{r}i}|X_i) - \sum_i I(X^{n\setminus i};Y_{\mathrm{r}i}|X_i,Y_{\mathrm{r}}^{i-1},Y_{i+1}^n)$$

$$
\begin{aligned}
&-\sum_i I(X_{\mathrm{r}i};Y_{\mathrm{r}i}|X_i) \qquad (34)\\
&\overset{(b)}{=}\sum_i I(Y_{i+1}^n;Y_{\mathrm{r}i}|X^n,Y_{\mathrm{r}}^{i-1})-\sum_i I(X^{n\backslash i};Y_{\mathrm{r}i}|X_i,Y_{\mathrm{r}}^{i-1},Y_{i+1}^n)\\
&\overset{(c)}{=}\sum_i I(Y_{\mathrm{r}}^{i-1};Y_i|X^n,Y_{i+1}^n)-\sum_i I(X^{n\backslash i};Y_{\mathrm{r}i}|X_i,Y_{\mathrm{r}}^{i-1},Y_{i+1}^n)\\
&\overset{(d)}{=}\sum_i I(Y_{\mathrm{r}}^{i-1};Y_i|X^n,Y_{i+1}^n)-\sum_i I(X^{n\backslash i};Y_{\mathrm{r}i},Y_i|X_i,Y_{\mathrm{r}}^{i-1},Y_{i+1}^n)\\
&=\sum_i I(X^{n\backslash i},Y_{\mathrm{r}}^{i-1};Y_i|X_i,Y_{i+1}^n)-\sum_i I(X^{n\backslash i};Y_i|X_i,Y_{i+1}^n)\\
&\quad-\sum_i I(X^{n\backslash i};Y_{\mathrm{r}i},Y_i|X_i,Y_{\mathrm{r}}^{i-1},Y_{i+1}^n)\\
&=\sum_i I(Y_{\mathrm{r}}^{i-1},Y_{i+1}^n,X_{\mathrm{r}i};Y_i|X_i)-\sum_i I(X^{n\backslash i},Y_{i+1}^n;Y_i|X_i)\\
&\quad-\sum_i I(X^{n\backslash i};Y_{\mathrm{r}i}|X_i,Y_{\mathrm{r}}^{i-1},Y_{i+1}^n,Y_i)\\
&\le\sum_i I(Y_{\mathrm{r}}^{i-1},Y_{i+1}^n,X_{\mathrm{r}i};Y_i|X_i)-\sum_i I(X^{n\backslash i},Y_{i+1}^n;Y_i|X_i),
\end{aligned}
$$

where (a) follows from the fact that $X_{\mathrm{r}i}$ is a function of Y_{r}^{i-1}, (b) follows because $(X^{n\backslash i},Y_{\mathrm{r}}^{i-1})\to(X_i,X_{\mathrm{r}i})\to Y_{\mathrm{r}i}$ form a Markov chain, (c) follows from Csiszár's sum lemma and (d) follows because $(X^{n\backslash i},Y_{i+1}^n)\to(X_i,Y_{\mathrm{r}}^{i-1},Y_{\mathrm{r}i})\to Y_i$ form a Markov chain. Next, observe that

$$I(V;Y_{\mathrm{r}}|Q)-I(V;Y|Q)=I(U;Y_{\mathrm{r}}|Q)-I(U;Y|Q)=\sum_i I(Y_{\mathrm{r}}^{i-1};Y_{\mathrm{r}i})-\sum_i I(Y^{i-1};Y_i)$$

Thus, $I(V;Y_{\mathrm{r}})-I(V;Y)=I(U;Y_{\mathrm{r}})-I(U;Y)$.

Proof of the equivalent form. from (8) and (3), we deduce that

$$
\begin{aligned}
R&\le I(X;Y,Y_{\mathrm{r}}|X_{\mathrm{r}})-I(V;Y|X_{\mathrm{r}},Y_{\mathrm{r}})-I(X;Y_{\mathrm{r}}|V,X_{\mathrm{r}},Y)+I(X_{\mathrm{r}};Y_{\mathrm{r}})-I(V,X_{\mathrm{r}};Y_{\mathrm{r}})+I(V,X_{\mathrm{r}};Y)\\
&=I(X,X_{\mathrm{r}};Y)-I(V;Y_{\mathrm{r}}|X_{\mathrm{r}},X,Y).
\end{aligned}
\tag{35}
$$

Therefore the bound described in the alternate form is tighter than the bound in Theorem 1. It remains to establish the other direction. Note that if (8) holds for a maximizing distribution $p(u,x,x_{\mathrm{r}})p(v|x,x_{\mathrm{r}},y_{\mathrm{r}})$, the proof is complete. Assume otherwise that

$$I(U;Y_{\mathrm{r}})-I(U;Y)=I(V,X_{\mathrm{r}};Y_{\mathrm{r}})-I(V,X_{\mathrm{r}};Y)>I(X_{\mathrm{r}};Y_{\mathrm{r}}).\tag{36}$$

Then, the inequality in (3) is strict since

$$
\begin{aligned}
R&\le I(X,X_{\mathrm{r}};Y)-I(V;Y_{\mathrm{r}}|X_{\mathrm{r}},X,Y)\\
&=I(X;Y,Y_{\mathrm{r}}|X_{\mathrm{r}})-I(V;Y|X_{\mathrm{r}},Y_{\mathrm{r}})-I(X;Y_{\mathrm{r}}|V,X_{\mathrm{r}},Y)+I(X_{\mathrm{r}};Y_{\mathrm{r}})-I(V,X_{\mathrm{r}};Y_{\mathrm{r}})+I(V,X_{\mathrm{r}};Y)\\
&<I(X;Y,Y_{\mathrm{r}}|X_{\mathrm{r}})-I(V;Y|X_{\mathrm{r}},Y_{\mathrm{r}})-I(X;Y_{\mathrm{r}}|V,X_{\mathrm{r}},Y).
\end{aligned}
\tag{37}
$$

Consider two independent Bernoulli time-sharing random variables $Q_1\sim\mathrm{B}(\theta)$ and $Q_2\sim\mathrm{B}(\theta)$ and independent of previously defined random variables. Set $\tilde{V}=(V,Q_1)$ if $Q_1=0$ and $\tilde{V}=Q_1$ if $Q_1=1$. Similarly, set $\tilde{U}=(U,Q_2)$ if $Q_2=0$ and $\tilde{U}=(X_{\mathrm{r}},Q_2)$ if $Q_2=1$. Observe that for any $\theta\in[0,1]$, the following equality

$$I(\tilde{U};Y_{\mathrm{r}})-I(\tilde{U};Y)=I(\tilde{V},X_{\mathrm{r}};Y_{\mathrm{r}})-I(\tilde{V},X_{\mathrm{r}};Y)$$

still holds. Moreover, inequalities (6) and (1) also continue to hold. For $\theta=0$, we have

$$I(\tilde{V},X_{\mathrm{r}};Y_{\mathrm{r}})-I(\tilde{V},X_{\mathrm{r}};Y)>I(X_{\mathrm{r}};Y_{\mathrm{r}}).\tag{38}$$

and for $\theta = 1$, we have

$$I(\tilde{V}, X_{\mathrm{r}}; Y_{\mathrm{r}}) - I(\tilde{V}, X_{\mathrm{r}}; Y) \leq I(X_{\mathrm{r}}; Y_{\mathrm{r}}) \tag{39}$$

one can find some $\theta^* \in [0,1]$ such that $\tilde{U}$ and $\tilde{V}$ satisfy

$$I(\tilde{V}, X_{\mathrm{r}}; Y_{\mathrm{r}}) - I(\tilde{V}, X_{\mathrm{r}}; Y) = I(\tilde{U}; Y_{\mathrm{r}}) - I(\tilde{U}; Y) = I(X_{\mathrm{r}}; Y_{\mathrm{r}}). \tag{40}$$

For this choice of θ^*, we claim that $p(\tilde{u}, x, x_{\mathrm{r}})p(\tilde{v}|x, x_{\mathrm{r}}, y_{\mathrm{r}})$ is a maximizing distribution for which (8) holds with equality. This follows from the fact that inequalities (6) and (1) hold for any $\theta \in [0,1]$. Using (40) and the expansion in (37), the constraint in (3) is identical with (6) and holds for θ^*.

The cardinality bounds on the auxiliary random variables come from the standard Caratheodory-Bunt [Bun34] arguments and is omitted.

This completes the proof.

3.2. Proof of Theorem 2

First, we show that the cut-set bound is not tight for any non-zero values of g_{12}, g_{13}, g_{23}. The proof is by contradiction. Hence let us assume that the cut-set bound is tight. That is,

$$C_{\mathrm{CS}} = \max_{P_{X,X_{\mathrm{r}}}:\, \mathrm{E}(X^2) \leq P,\, \mathrm{E}(X_{\mathrm{r}}^2) \leq P} \min\{I(X, X_{\mathrm{r}}; Y), I(X; Y, Y_{\mathrm{r}}|X_{\mathrm{r}})\}$$

is achievable. We know (see Section 16.5 of [EK11]) that the maximum is attained via the unique jointly Gaussian distribution

$$(X, X_{\mathrm{r}})_{opt} \sim \begin{cases} \mathcal{N}\left(0, \begin{bmatrix} P & \rho^* P \\ \rho^* P & P \end{bmatrix}\right) & \text{if } S_{12} > S_{23}, \\ \mathcal{N}\left(0, \begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}\right) & \text{if } S_{12} \leq S_{23}, \end{cases} \tag{41}$$

where $\rho^* \in (0,1)$ satisfies $I(X, X_{\mathrm{r}}; Y) = I(X; Y, Y_{\mathrm{r}}|X_{\mathrm{r}})$. Note that $C_{\mathrm{CS}} = I(X; Y, Y_{\mathrm{r}}|X_{\mathrm{r}})$ holds (is tight) in both of these cases.

From Corollary 1, we know that the capacity C satisfies

$$C \leq \max_{P_{X,X_{\mathrm{r}}}:\, \mathrm{E}(X^2) \leq P,\, \mathrm{E}(X_{\mathrm{r}}^2) \leq P} \min\{I(X, X_{\mathrm{r}}; Y) - I(V; Y_{\mathrm{r}}|X_{\mathrm{r}}, X, Y), I(X; Y, Y_{\mathrm{r}}|X_{\mathrm{r}}) - I(V; Y|X_{\mathrm{r}}, Y_{\mathrm{r}}) - I(X; Y_{\mathrm{r}}|V, X_{\mathrm{r}}, Y)\}$$

for some $P_{Y,Y_{\mathrm{r}}|X,X_{\mathrm{r}}} P_{V|X,X_{\mathrm{r}},Y_{\mathrm{r}}}$. If $C = C_{\mathrm{CS}}$ then it is necessary that when (X, X_{r}) is distributed as in (41), then from $C_{\mathrm{CS}} = I(X; Y, Y_{\mathrm{r}}|X_{\mathrm{r}})$, we deduce the existence of a $P_{V|X,X_{\mathrm{r}},Y_{\mathrm{r}}}$ such that

$$I(V; Y|X_{\mathrm{r}}, Y_{\mathrm{r}}) + I(X; Y_{\mathrm{r}}|V, X_{\mathrm{r}}, Y) = 0.$$

Our argument below shows that no such distribution exists.

Using Lemma 5, the first condition yields $I(X; V|X_{\mathrm{r}}, Y_{\mathrm{r}}) = 0$, $I(X; Y_{\mathrm{r}}|V, X_{\mathrm{r}}) = 0$. This gives the double Markovity conditions (Lemma 3): given $X_{\mathrm{r}} = x_{\mathrm{r}}$, $V \to Y_{\mathrm{r}} \to X$ and $Y_{\mathrm{r}} \to V \to X$. Thus there exists functions $g(X_{\mathrm{r}}, V)$ and $f(X_{\mathrm{r}}, Y_{\mathrm{r}})$ such that $f(X_{\mathrm{r}}, Y_{\mathrm{r}}) = g(X_{\mathrm{r}}, V)$ with probability 1, and

$$I(X; Y_{\mathrm{r}}, V|X_{\mathrm{r}}, f(X_{\mathrm{r}}, Y_{\mathrm{r}}), g(X_{\mathrm{r}}, V)) = 0.$$

Using an abuse of notation, let $B = f(X_{\mathrm{r}}, Y_{\mathrm{r}}) = g(X_{\mathrm{r}}, V)$ almost surely and so we have $I(X; V, Y_{\mathrm{r}}|B, X_{\mathrm{r}}) = 0$. This implies that $I(X; Y_{\mathrm{r}}|B, X_{\mathrm{r}}) = 0$. We also have the Markov chain $B \to (X_{\mathrm{r}}, Y_{\mathrm{r}}) \to X$. Observe that since $X, X_{\mathrm{r}}, Y_{\mathrm{r}}$ are jointly Gaussian, we can express $X = aX_{\mathrm{r}} + cY_{\mathrm{r}} + \hat{Z}$, where $\hat{Z}$ is independent of $Y_{\mathrm{r}}, X_{\mathrm{r}}$. Here

$$c = \frac{S_{12}(1 - \rho^*)}{g_{12}(1 + S_{12}(1 - \rho^*))} \neq 0.$$

Hence $\hat{Z}$ is also independent of B. Now we have $I(Y_{\mathrm{r}}; cY_{\mathrm{r}} + \hat{Z}|B, X_{\mathrm{r}}) = 0$. This holds only if Y_{r} is a function of (B, X_{r}) (see Lemma 5). Consequently Y_{r} is a function of (V, X_{r}) (since $B = g(X_{\mathrm{r}}, V)$) implying that $I(V; Y_{\mathrm{r}}|X_{\mathrm{r}}, X, Y) = \infty$. Thus, the constraint $R \leq I(X, X_{\mathrm{r}}; Y) - I(V; Y_{\mathrm{r}}|X_{\mathrm{r}}, X, Y)$ cannot hold. This establishes the requisite contradiction.

The optimality of Gaussian random variables for the evaluation of Corollary 1 is established in Lemma 6 in Appendix B. Therefore there exists some $\rho \in [-1,1]$ such that

$$K_{X,X_{\rm r},Z_1} = \begin{bmatrix} P & \rho P & 0 \\ \rho P & P & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{42}$$

and

$$K_{X,Z_1|VX_{\rm r}} = \begin{bmatrix} P(1-\rho^2)\alpha & \sigma\sqrt{P(1-\rho^2)\alpha\beta} \\ \sigma\sqrt{P(1-\rho^2)\alpha\beta} & \beta \end{bmatrix} \preceq \begin{bmatrix} P(1-\rho^2) & 0 \\ 0 & 1 \end{bmatrix}. \tag{43}$$

This gives us the conditions: $0 \le \alpha, \beta \le 1$ and $(1-\alpha)(1-\beta) \ge \sigma^2\alpha\beta$.

Now note that

$$\begin{aligned}
\max_{P_{U|X,X_{\rm r}}} (I(U;Y_{\rm r}) - I(U;Y)) &= I(X,X_{\rm r};Y_{\rm r}) - I(X,X_{\rm r};Y) + \max_{P_{U|X,X_{\rm r}}} (I(X,X_{\rm r};Y|U) - I(X,X_{\rm r};Y_{\rm r}|U)) \\
&\overset{(a)}{=} I(X,X_{\rm r};Y_{\rm r}) - I(X,X_{\rm r};Y) + \frac{1}{2}\log\lambda_{\max} \\
&= \frac{1}{2}\log(S_{12}+1) - \frac{1}{2}\log\left(1+S_{13}+S_{23}+2\rho\sqrt{S_{13}S_{23}}\right) + \frac{1}{2}\log\lambda_{\max},
\end{aligned}$$

where (a) follows from [GN20, Remark 13] and $\lambda_{\max}$ is the larger root of the quadratic polynomial

$$2\rho\sqrt{S_{13}S_{23}} + S_{13} + S_{23} + 1 - \lambda\big(S_{23}S_{12}(1-\rho^2) + S_{13} + S_{23} + S_{12} + 2 + 2\rho\sqrt{S_{13}S_{23}}\big) + \lambda^2(S_{12}+1) = 0. \tag{44}$$

Next, we use an observation about the maximizing variables from Lemma 2 below to determine the value of σ. The constraint in (46) is a linear equality in σ and yields

$$\sigma = \frac{(1-\rho^2)\alpha S_{13}+1}{2T\sqrt{S_{12}(1-\rho^2)\alpha\beta}} - \frac{(1-\rho^2)\alpha S_{12}+\beta}{2\sqrt{S_{12}(1-\rho^2)\alpha\beta}},$$

where

$$\frac{1}{2}\log(T) = \min\big\{-I(X;Y_{\rm r}|X_{\rm r}) + I(X,X_{\rm r};Y), \max_{P_{U|X,X_{\rm r}}}(I(X,X_{\rm r};Y|U) - I(X,X_{\rm r};Y_{\rm r}|U))\big\}.$$

Hence, we obtain that any achievable rate R must satisfy the conditions

$$\begin{aligned}
R \le \frac{1}{2}\log\big((1-\rho^2)S_{12}+1\big) &- \frac{1}{2}\log\left(\beta + S_{12}(1-\rho^2)\alpha + 2\sigma\sqrt{S_{12}(1-\rho^2)\alpha\beta}\right) \\
&+ \frac{1}{2}\log\big(\beta(1-\sigma^2)\big) + \frac{1}{2}\log\big((1-\rho^2)\alpha S_{13}+1\big)
\end{aligned} \tag{45}$$

for some $0 \le \alpha, \beta \le 1$, $\sigma, \rho \in [-1,1]$ such that $(1-\alpha)(1-\beta) \ge \sigma^2\alpha\beta$. This completes the proof.

Lemma 2. *Any maximizing distribution for the optimization problem of computing the maximum rate given by Theorem 2 must satisfy*

$$I(V,X_{\rm r};Y_{\rm r}) - I(V,X_{\rm r};Y) = \min\big\{I(X_{\rm r};Y_{\rm r}), \quad \max_{P_{U|X,X_{\rm r}}}(I(U;Y_{\rm r}) - I(U;Y))\big\}. \tag{46}$$

Proof. Assume that the maximizer does not satisfy (46). Then, from the alternate form of 1

$$I(V,X_{\rm r};Y_{\rm r}) - I(V,X_{\rm r};Y) < \min\big\{I(X_{\rm r};Y_{\rm r}), \quad \max_{P_{U|X,X_{\rm r}}}(I(U;Y_{\rm r}) - I(U;Y))\big\}.$$

Take a Bernoulli time-sharing random variable $Q \sim \mathrm{B}(\theta)$ and set $\tilde{V} = (V,Q)$ if $Q=0$ and $\tilde{V} = (Y_{\rm r}+\zeta W, Q)$ if $Q=1$, where W is a standard Gaussian noise, independent of previously defined random variables. When $\zeta = 0$, replacing V by $\tilde{V}$ would strictly increase the right hand side of (10) for any $\theta > 0$. On the other hand, for any arbitrary $\theta > 0$, we have

$$\lim_{\zeta\to 0}(I(\tilde{V},X_{\rm r};Y_{\rm r}) - I(\tilde{V},X_{\rm r};Y)) = \infty.$$

Also for any $\zeta > 0$,

$$\lim_{\theta\to 0} I(\tilde{V},X_{\rm r};Y_{\rm r}) - I(\tilde{V},X_{\rm r};Y) = I(V,X_{\rm r};Y_{\rm r}) - I(V,X_{\rm r};Y).$$

Therefore, one can find suitable $\theta, \zeta > 0$ such that for $\tilde{V}$ (11) has a larger value and still the constraints are satisfied for $\tilde{V}$. This is a contradiction.

$\square$

3.3. Proof of Proposition 1

Let R_1 be the upper bound of Theorem 1, R_2 be the upper bound of Proposition 1 in its form without any constraints, and R_3 be the the upper bound of Proposition 1 in its form with a constraint. It suffices to show that $R_1 \leq R_2$, $R_2 \leq R_3$ and $R_3 \leq R_1$.

To show $R_1 \leq R_2$, assume that a rate R satisfies the inequalities given in Theorem 1 for a pair auxiliary variables (U, V). Theorem 1 implies that

$$\begin{aligned}
R &\leq I(X;Y,Y_{\mathrm{r}}|X_{\mathrm{r}}) - I(V;Y|X_{\mathrm{r}},Y_{\mathrm{r}}) - I(X;Y_{\mathrm{r}}|V,X_{\mathrm{r}},Y) \\
&\leq I(X;Y_1,Y_{\mathrm{r}}) - I(V,X_{\mathrm{r}};Y_1|Y_{\mathrm{r}}) - I(X;Y_{\mathrm{r}}|V,X_{\mathrm{r}},Y_1), \\
R &\leq I(X,X_{\mathrm{r}};Y) - I(V;Y_{\mathrm{r}}|X_{\mathrm{r}},X,Y) \\
&= I(X;Y_1) + I(X_{\mathrm{r}};Y_2|Y_1) - I(Y_{\mathrm{r}};X_{\mathrm{r}}|X) - I(V,X_{\mathrm{r}};Y_{\mathrm{r}}|X,Y_1) \qquad (47) \\
&\leq I(X;Y_1) + C_0 - I(V,X_{\mathrm{r}};Y_{\mathrm{r}}|X,Y_1),
\end{aligned}$$

where (47) follows from $I(Y_1,Y_{\mathrm{r}};X_{\mathrm{r}}|X) = 0$ and $I(Y_2;X,Y_1|X_{\mathrm{r}}) = 0$. This corresponds to the constraints in Proposition 1 for $\tilde{V} = (V, X_{\mathrm{r}})$.

Next, we show that $R_2 \leq R_3$. Let R be a rate satisfying

$$R \leq I(X;Y_1,Y_{\mathrm{r}}) - I(V;Y_1|Y_{\mathrm{r}}) - I(X;Y_{\mathrm{r}}|V,Y_1) \qquad (48)$$

$$R \leq I(X;Y_1) + C_0 - I(V;Y_{\mathrm{r}}|X,Y_1) \qquad (49)$$

for some $p(x)p(y_1,y_{\mathrm{r}}|x)p(v|x,y_{\mathrm{r}})$. It suffices to show that

$$I(V;Y_{\mathrm{r}}) - I(V;Y_1) \leq C_0. \qquad (50)$$

is satisfied by a maximizer of the minimum of (48) and (49). Assume that (50) is violated by a maximizing distribution and we have

$$I(V;Y_{\mathrm{r}}) - I(V;Y_1) > C_0.$$

In this case, the inequality (48) is strictly redundant since

$$\begin{aligned}
R &\leq I(X;Y_1) + C_0 - I(V;Y_{\mathrm{r}}|X,Y_1) \\
&< I(X;Y_1) + I(V;Y_{\mathrm{r}}) - I(V;Y_1) - I(V;Y_{\mathrm{r}}|X,Y_1) \\
&= I(X;Y_1,Y_{\mathrm{r}}) - I(V;Y_1|Y_{\mathrm{r}}) - I(X;Y_{\mathrm{r}}|V,Y_1).
\end{aligned}$$

Now, let $Q \sim B(\epsilon)$ be a time-sharing random variable and let $\tilde{V} = (Q, V)$ if $Q = 1$, and $\tilde{V} = Q$ if $Q = 0$. Then for an appropriate choice for ϵ, we have

$$I(\tilde{V};Y_{\mathrm{r}}) - I(\tilde{V};Y_1) = C_0.$$

For $\tilde{V}$, the constraint (48) is still redundant. Moreover, (48) and (49) have not decreased, and the constraint still holds. This shows that without loss of generality we can impose the constraint in (50) when evaluating the region.

Finally, to show that $R_3 \leq R_1$, assume that a rate R satisfies

$$R \leq I(X;Y_1,Y_{\mathrm{r}}) - I(V;Y_1|Y_{\mathrm{r}}) - I(X;Y_{\mathrm{r}}|V,Y_1) \qquad (51)$$

for some $p(x)p(y_1,y_{\mathrm{r}}|x)p(v|x,y_{\mathrm{r}})$ such that

$$0 \leq I(V;Y_{\mathrm{r}}) - I(V;Y_1) \leq C_0. \qquad (52)$$

The constraint $0 \leq I(V;Y_{\mathrm{r}}) - I(V;Y_1)$ is added above since it holds for any maximizer V. To see this, assume that

$$I(V;Y_{\mathrm{r}}) - I(V;Y_1) < 0,$$

then we have

$$\begin{aligned}R &\le I(X;Y_1,Y_\mathrm{r}) - I(V;Y_1|Y_\mathrm{r}) - I(X;Y_\mathrm{r}|V,Y_1)\\ &= I(X;Y_1) + I(V;Y_\mathrm{r}) - I(V;Y_1) - I(V;Y_\mathrm{r}|X,Y_1)\\ &< I(X;Y_1),\end{aligned}$$

which is less than the direct transmission bound. For such a joint distribution of (V,X,Y_1,Y_r), let (U,X_r,Y_2) be independent of (V,X,Y_1,Y_r) such that (i) $I(X_\mathrm{r};Y_2)=C_0$, (ii) $U\to X_\mathrm{r}\to Y_2$ forms a Markov chain, and (iii)

$$I(U;Y_2) = I(X_\mathrm{r};Y_2) - [I(V;Y_\mathrm{r}) - I(V;Y_1)].$$

This is feasible for some $U\to X_\mathrm{r}\to Y_2$ since

$$0\le I(V;Y_\mathrm{r}) - I(V;Y_1) \le C_0 = I(X_\mathrm{r};Y_2).$$

Now, consider the choice of (U,V) in Theorem 1. One can directly verify that the rate R satisfies the inequalities given in the alternate form of Theorem 1 for our choice of joint distribution of $(U,V,X,X_\mathrm{r},Y_1,Y_2,Y_\mathrm{r})$. Furthermore,

$$\begin{aligned}I(V,X_\mathrm{r};Y_\mathrm{r}) - I(V,X_\mathrm{r};Y) &= I(V;Y_\mathrm{r}) - I(V;Y_1) - I(X_\mathrm{r};Y_2)\\ &= I(U;Y_\mathrm{r}) - I(U;Y).\end{aligned}$$

This completes the proof.

3.4. Proof of Theorem 3

As argued in [Kim07], $C_0^*\le H_G(Y_\mathrm{r}|Y_1)$. It remains to show that $C_0^*\ge H_G(Y_\mathrm{r}|Y_1)$. Let $C_0=C_0^*$. From (16), we deduce that $I(V;Y_1|Y_\mathrm{r})=0$. Since X is a function of (Y_1,Y_r), we deduce that $I(V;X|Y_\mathrm{r})=0$. Consequently, (17) and (18) imply that the achievable rate must be achievable by compress-forward (see [Kim07, Proposition 3]). From $\log|\mathcal{X}| = I(X;V,Y_1)$, we deduce that X is uniformly distributed and $H(X|V,Y_1)=0$. From [OR95, Theorem 2], we deduce that $C_0^*\ge H_G(Y_\mathrm{r}|Y_1)$. This confirms Kim's conjecture in [Kim07].

3.5. Proof of Theorem 4

The result follows immediately from the definition that $R_0^*\ge C_0^*$. Therefore it suffices to show that $R_0^*\le C_0^*$. Let C_0 be such that $\mathcal{C}(C_0)=\mathcal{C}(\infty)=\max_{p(x)} I(X;Y_1,Y_\mathrm{r})$. From the constraint (16) of the upper bound in Proposition 1, it follows that if $\mathcal{C}(\infty)=\max_{p(x)} I(X;Y_1,Y_\mathrm{r})$ is achievable, for a maximizing distribution $p^*(x)$, there exists a distribution $p(v|x,y_\mathrm{r})$ such that $I(V;Y_1|Y_\mathrm{r})=0$. Since the channel $p_{Y_1|X}$ is generic, it follows from Lemma 4 that $I(V;Y_1|Y_\mathrm{r})=0$ implies that $I(V;X|Y_\mathrm{r})=0$. Therefore $V\to Y_\mathrm{r}\to X\to Y_1$ form a Markov chain. Then, the constraints in (17) and (18) imply that the rate R is achievable by compress-forward with the compression random variable V. Here, we utilize the characterization of the compress-forward for the relay channel with orthogonal receiver components given in [Kim07, Proposition 3]. Consequently, we have that $R_{\mathsf{CF}}(C_0)=\mathcal{C}(C_0)=\mathcal{C}(\infty)$. Since this holds for any C_0 such that $\mathcal{C}(C_0)=\mathcal{C}(\infty)=\max_{p(x)} I(X;Y_1,Y_\mathrm{r})$, we have that $R_0^*\le C_0^*$. This completes the proof.

3.6. Proof of Proposition 2

The sufficiency of considering Gaussian random variables for the evaluation of Proposition 1 in this context is established in Lemma 7 in the Appendix. Let the covariance of X,Y_r given V be

$$K_{X,Y_\mathrm{r}|V} = \begin{bmatrix} K_1 & \rho\sqrt{K_1K_2}\\ \rho\sqrt{K_1K_2} & K_2\end{bmatrix} \preceq \begin{bmatrix} P & P\\ P & P+N_\mathrm{r}\end{bmatrix}$$

for some $0\le K_1\le P$ and $0\le K_2\le N_\mathrm{r}+P$ and $\rho\in[-1,1]$ such that

$$(P-K_1)(P+N_\mathrm{r}-K_2)\ge (P-\rho\sqrt{K_1K_2})^2. \tag{53}$$

Then, the upper bound becomes the maximum of

$$\frac{1}{2}\log\left(\frac{P+N_\mathrm{r}}{N_\mathrm{r}}\right)+\frac{1}{2}\log\left(\frac{K_1+N_1}{N_1}\right)+\frac{1}{2}\log\left(1-\rho^2\right) \tag{54}$$

subject to

$$\frac{1}{2}\log(P+N_{\mathrm{r}})-\frac{1}{2}\log(K_2)-\frac{1}{2}\log(P+N_1)+\frac{1}{2}\log(K_1+N_1)\leq C_0.$$

Lemma 8 solves the above optimization problem, showing that the optimal K_1, K_2 and ρ are as follows: if $S_{12} \leq S_{13}+S_{23}+S_{13}S_{23}$, the optimizers are

$$K_1^* = \begin{cases} -N_1\frac{2^{2C_0}(P+N_{\mathrm{r}})-(P+N_{\mathrm{r}})}{2^{2C_0}(P+N_1)-(P+N_{\mathrm{r}})}+P\frac{2^{2C_0}(N_1-N_{\mathrm{r}})}{2^{2C_0}(P+N_1)-(P+N_{\mathrm{r}})} & S_{12}\geq S_{13}+S_{23}+S_{13}S_{23}, \\ P\left(1-\frac{P(N_1+P)^2(2^{2C_0}-1)}{(P+N_{\mathrm{r}})(2^{2C_0}-1)((N_1+P)^2-N_1^2 2^{-2C_0})+(N_{\mathrm{r}}-N_1)P^2}\right) & \text{otherwise.}\end{cases}$$

$$K_2^* = \frac{(K_1^*+N_1)(P+N_{\mathrm{r}})}{(P+N_1)2^{2C_0}},$$

$$\rho^* = \frac{P-\sqrt{(P-K_1^*)(P+N_{\mathrm{r}}-K_2^*)}}{\sqrt{K_1^*K_2^*}}.$$

Substituting these values in (54) and replacing $2^{2C_0}=(1+S_{23})$, $S_{12}=P/N_{\mathrm{r}}$ and $S_{13}=P/N_1$, the upper bound reduces to the form given in the statement of Proposition 2. This completes the proof.

3.7. Proof of Lemma 1

We seek to calculate

$$\frac{1}{2}\log\left(1+S_{13}\right)+\sup_{\theta\in\left[\arcsin\left(\frac{1}{1+S_{23}}\right),\frac{\pi}{2}\right]}\min\left\{C_0+\log\sin\theta,\min_{\omega\in\left(\frac{\pi}{2}-\theta,\frac{\pi}{2}\right]}h_\theta(\omega)\right\}, \tag{55}$$

where

$$h_\theta(\omega)=\frac{1}{2}\log\left(\frac{\left(S_{12}+S_{13}+\sin^2(\omega)-2\cos(\omega)\sqrt{S_{12}S_{13}}\right)\sin^2\theta}{(S_{13}+1)(\sin^2\theta-\cos^2\omega)}\right). \tag{56}$$

We make the following observations:

- Observation 1: The term $C_0+\log(\sin(\theta))$ is increasing in $\theta\in[0,\frac{\pi}{2}]$.
- Observation 2: The term $\min_{\omega\in\left(\frac{\pi}{2}-\theta,\frac{\pi}{2}\right]}h_\theta(\omega)$ is decreasing in θ. To see this, observe that for $0<\theta_1<\theta_2<\frac{\pi}{2}$ we have

$$\begin{aligned}&\min_{\omega\in(\frac{\pi}{2}-\theta_2,\frac{\pi}{2}]}\frac{1}{2}\log\left(\frac{\left(S_{12}+S_{13}+\sin^2(\omega)-2\cos(\omega)\sqrt{S_{12}S_{13}}\right)\sin^2\theta_2}{(S_{13}+1)(\sin^2\theta_2-\cos^2\omega)}\right)\\ &\leq\min_{\omega\in(\frac{\pi}{2}-\theta_1,\frac{\pi}{2}]}\frac{1}{2}\log\left(\frac{\left(S_{12}+S_{13}+\sin^2(\omega)-2\cos(\omega)\sqrt{S_{12}S_{13}}\right)\sin^2\theta_2}{(S_{13}+1)(\sin^2\theta_2-\cos^2\omega)}\right)\\ &\leq\min_{\omega\in(\frac{\pi}{2}-\theta_1,\frac{\pi}{2}]}\frac{1}{2}\log\left(\frac{\left(S_{12}+S_{13}+\sin^2(\omega)-2\cos(\omega)\sqrt{S_{12}S_{13}}\right)\sin^2\theta_1}{(S_{13}+1)(\sin^2\theta_1-\cos^2\omega)}\right).\end{aligned}$$

  Therefore $\min_{\omega\in\left(\frac{\pi}{2}-\theta,\frac{\pi}{2}\right]}h_\theta(\omega)$ is decreasing in θ.
- Observation 3: Given any fixed θ, the minimizer ω_* of $\min_{\omega\in\left(\frac{\pi}{2}-\theta,\frac{\pi}{2}\right]}h_\theta(\omega)$ can be explicitly computed (using routine algebra) as the smaller root of the quadratic polynomial

$$x^2-\frac{S_{12}+S_{13}+\cos^2\theta}{\sqrt{S_{12}S_{13}}}x+\sin^2(\theta)=0$$

  which is

$$\cos\omega_*=\frac{S_{13}+S_{12}+\cos^2\theta-\sqrt{(S_{12}-S_{13})^2+\cos^2\theta(4S_{12}S_{13}+2S_{12}+2S_{13}+\cos^2\theta)}}{2\sqrt{S_{12}S_{13}}}. \tag{57}$$

Consider two cases:

- *Case 1*: $S_{12} \geq S_{13}+S_{23}+S_{13}S_{23}$: we claim that the supremum in (55) is attained at $\theta_* = \pi/2$. Using observations 1 and 2 above, it suffices to verify that

$$C_0 + \log(\sin(\theta_*)) \leq \min_{\omega \in \left(\frac{\pi}{2}-\theta_*, \frac{\pi}{2}\right]} h_{\theta_*}(\omega).$$

For $\theta_* = \pi/2$, from (57) the minimizer ω_* satisfies

$$\cos\omega_* = \sqrt{\frac{S_{13}}{S_{12}}}.$$

Consequently,

$$h_\theta^*(\omega^*) = \frac{1}{2}\log\left(\frac{S_{12}+1}{S_{13}+1}\right).$$

Using the assumption $S_{12} \geq S_{13}+S_{23}+S_{13}S_{23}$ we note that

$$h_\theta(\omega_*) = \frac{1}{2}\log\left(\frac{S_{12}+1}{S_{13}+1}\right) \geq \frac{1}{2}\log(1+S_{23}) = C_0 = C_0 + \log(\sin(\theta_*)).$$

Substituting the maximizer $\theta_* = \pi/2$ in (55) shows that the bound in Theorem 5 is equivalent to

$$R \leq \frac{1}{2}\log((1+S_{13})(1+S_{23})).$$

- *Case* 2: $S_{12} < S_{13}+S_{23}+S_{13}S_{23}$: let $\theta_* \in [0, \frac{\pi}{2}]$ be defined according to

$$\sin^2\theta_* = \left(1 + \frac{S_{23}S_{12}}{(S_{13}+1)(S_{23}+1)-1}\right)\frac{1}{1+S_{23}}. \tag{58}$$

Observe that the above definition for $\sin^2\theta_*$ makes sense because the right hand side of (58) is less than one by the assumption that $S_{12} < S_{13}+S_{23}+S_{13}S_{23}$. Using Observations 1 and 2 above, in order to show that the supremum in (55) is attained at θ_*, it suffices to verify that

$$C_0 + \log(\sin(\theta_*)) = \min_{\omega \in \left(\frac{\pi}{2}-\theta_*, \frac{\pi}{2}\right]} h_{\theta_*}(\omega).$$

This equation can be verified by plugging in the value for $\cos(\omega_*)$ from (57). Substituting the maximizer θ_* in (55) shows that the bound in Theorem 5 is equivalent to

$$R \leq \frac{1}{2}\log\left(1 + S_{13} + \frac{S_{12}(S_{13}+1)S_{23}}{(S_{13}+1)(S_{23}+1)-1}\right). \tag{59}$$

This completes the proof.

3.8. *Proof of Theorem 6*

Using the symmetrization argument in [Nai13], without loss of generality, we can restrict X to be uniformly distribution when evaluating Proposition 1. To see this, given some arbitrary (V, X), take $Q \sim \mathrm{B}(0.5)$ independent of (V, X). Then, letting $V' = (V, Q)$, $X' = X + Q \mod 2)$, $Y_1' = Y_1 + Q \mod 2)$ and $Y_\mathrm{r}' = Y_\mathrm{r} + Q \mod 2)$, one can verify that $I(V'; Y_\mathrm{r}') - I(V'; Y_1') = I(V; Y_\mathrm{r}) - I(V; Y_1)$ and

$$I(X'; Y_1', Y_\mathrm{r}') - I(V'; Y_1'|Y_\mathrm{r}') - I(X'; Y_\mathrm{r}'|V', Y_1') \geq I(X; Y_1, Y_\mathrm{r}) - I(V; Y_1|Y_\mathrm{r}) - I(X; Y_\mathrm{r}|V, Y_1).$$

Moreover, X' is uniform.

From Proposition 1 for any $\lambda \geq 0$, any achievable rate R must satisfy

$$\begin{aligned}
R &\leq \max_{p(v|x,y_\mathrm{r})} \big(I(X; Y_1, Y_\mathrm{r}) - I(V; Y_1|Y_\mathrm{r}) - I(X; Y_\mathrm{r}|V, Y_1) + \lambda[C_0 - I(V; Y_\mathrm{r}) + I(V; Y_1)]\big) \\
&= \max_{p(v|x,y_\mathrm{r})} \big(I(X; Y_\mathrm{r}) + \lambda C_0 - H(Y_1|X) + H(Y_\mathrm{r}|X, V) + (1-\lambda)[H(Y_1|V) - H(Y_\mathrm{r}|V)]\big) \\
&= 1 - 2H_2(\rho) + \lambda C_0 + \max_{p(v|x,y_\mathrm{r})} \big(H(Y_\mathrm{r}|X, V) + (1-\lambda)[H(Y_1|V) - H(Y_\mathrm{r}|V)]\big).
\end{aligned}$$

Without loss of generality we can assume that $\lambda \in [0,1]$. To see this, observe that for $\lambda = 1$ the optimal choice for V is a constant and the upper bound becomes $1 - 2H_2(\rho) + C_0 + H(Y_{\rm r}|X)$. If $\lambda > 1$, the upper bound is greater than or equal to $1 - 2H_2(\rho) + C_0 + H(Y_{\rm r}|X)$ since V equal to a constant is one possible choice for V.

Let $p(x, y_{\rm r}) = (p_{00}, p_{01}, p_{10}, p_{11})$ and $f : p(x, y_{\rm r}) \mapsto (1-\lambda)\left(H(Y_1) - H(Y_{\rm r})\right) + H(Y_{\rm r}|X)$. The maximum of $(1-\lambda)[H(Y_1|V) - H(Y_{\rm r}|V)] + H(Y_{\rm r}|X, V)$ is equal to the upper concave envelope, $\mathscr{C}_{\mathbf{p}}(f)$, of the function f at

$$\left(\frac{1-\rho}{2}, \frac{\rho}{2}, \frac{\rho}{2}, \frac{1-\rho}{2}\right).$$

Observe that

$$f(p_{00}, p_{01}, p_{10}, p_{11}) = f(p_{11}, p_{10}, p_{01}, p_{00}). \tag{60}$$

Let

$$g(c) = \max_{\substack{p_{00}, p_{01}, p_{10}, p_{11} \\ p_{01}+p_{10}=c}} f(p_{00}, p_{01}, p_{10}, p_{11}).$$

Now define $\tilde{f}(p_{00}, p_{01}, p_{10}, p_{11}) = g(p_{01} + p_{10})$. Clearly $\tilde{f}(p_{00}, p_{01}, p_{10}, p_{11}) \geq f(p_{00}, p_{01}, p_{10}, p_{11})$ pointwise, and consequently the upper concave envelope of $\mathscr{C}_{\mathbf{p}}(\tilde{f}) \geq \mathscr{C}_{\mathbf{p}}(f)$. On the contrary, we can invoke the symmetry in (60) to immediately conclude that

$$g(p_{01} + p_{10}) = \tilde{f}\left(\frac{p_{00}+p_{11}}{2}, \frac{p_{01}+p_{10}}{2}, \frac{p_{01}+p_{10}}{2}, \frac{p_{00}+p_{11}}{2}\right) \leq \mathscr{C}_{\mathbf{p}}(f)\Big|_{\left(\frac{p_{00}+p_{11}}{2}, \frac{p_{01}+p_{10}}{2}, \frac{p_{01}+p_{10}}{2}, \frac{p_{00}+p_{11}}{2}\right)}.$$

This implies that

$$\mathscr{C}_c(g)\Big|_{\rho} = \mathscr{C}_{\mathbf{p}}(f)\Big|_{\left(\frac{1-\rho}{2}, \frac{\rho}{2}, \frac{\rho}{2}, \frac{1-\rho}{2}\right)},$$

which completes the proof.

3.9. Proof of Theorem 7

Let $M_{\rm r} \in [1 : 2^{2C_0}]$ be the message from relay to the receiver ($M_{\rm r}$ is a function of $Y_{\rm r}^n$), and define

$$V_i = Y_{1i+1}^n, \quad W_i = (M_{\rm r}, Y_{\rm r}^{i-1}), \quad T_i = (J^{i-1}, J_{i+1}^n),$$

and $V = (Q, V_Q)$, $W = (Q, W_Q)$, $T = (Q, T_Q)$ $X = X_Q, Y_1 = Y_{1Q}, Y_{\rm r} = Y_{{\rm r}Q}$ for a time-sharing random variable $Q \overset{(d)}{=} \text{Uniform}[1:n]$.

Observe that $T \to X \to (J, Y_{\rm r}, Y_1)$ forms a Markov chain. The assumption $X_i \to J_i \to Y_{{\rm r}i}$ implies that

$$I(Y_{\rm r}^n; X_i, J_i, Y_{1i}|J^{i-1}, J_{i+1}^n, Y_{{\rm r}i}) = 0,$$

implying that

$$I(M_{\rm r}, Y_{\rm r}^{i-1}; X_i, J_i, Y_{1i}|J^{i-1}, J_{i+1}^n, Y_{{\rm r}i}) = 0.$$

This yields the Markov chain $W \to (T, Y_{\rm r}) \to (X, J, Y_1)$.

Now consider

$$\begin{aligned}
I(X^n; M_{\rm r}, Y_1^n) &\leq I(X^n; J^n) + I(X^n; Y_{\rm r}^n|J^n) + I(X^n; Y_1^n|M_{\rm r}, J^n) - I(X^n; Y_{\rm r}^n|M_{\rm r}, J^n) \\
&\leq \sum_i I(X_i; J_i) + \sum_i I(X^n; Y_{1i}|Y_{\rm r}^{i-1}, Y_{1i+1}^n, M_{\rm r}, J^n) - \sum_i I(X^n; Y_{{\rm r}i}|Y_{\rm r}^{i-1}, Y_{1i+1}^n, M_{\rm r}, J^n) \\
&\leq \sum_i I(X_i; J_i) + \sum_i I(X_i; Y_{1i}|Y_{\rm r}^{i-1}, Y_{1i+1}^n, M_{\rm r}, J^n) - \sum_i I(X_i; Y_{{\rm r}i}|Y_{\rm r}^{i-1}, Y_{1i+1}^n, M_{\rm r}, J^n) \\
&\qquad + \sum_i I(X^n; Y_{1i}|X_i, Y_{{\rm r}i}, Y_{\rm r}^{i-1}, Y_{1i+1}^n, M_{\rm r}, J^n) \\
&\leq \sum_i I(X_i; J_i) + \sum_i I(X_i; Y_{1i}|Y_{\rm r}^{i-1}, Y_{1i+1}^n, M_{\rm r}, J^n) - \sum_i I(X_i; Y_{{\rm r}i}|Y_{\rm r}^{i-1}, Y_{1i+1}^n, M_{\rm r}, J^n) \\
&\leq n\left(I(X; J) + I(X; Y_1|V, W, J, T) - I(X; Y_{\rm r}|V, W, J, T)\right).
\end{aligned}$$

Next, observe that

$$
\begin{aligned}
n\left[I(V,W;Y_{\mathrm{r}}|J,T)-I(V,W;Y_1|J,T)\right] &= \sum_i I(M_{\mathrm{r}},Y_{\mathrm{r}}^{i-1},Y_{1i+1}^n;Y_{\mathrm{r}i}|J^n)-\sum_i I(M_{\mathrm{r}},Y_{\mathrm{r}}^{i-1},Y_{1i+1}^n;Y_{1i}|J^n)\\
&= \sum_i I(M_{\mathrm{r}};Y_{\mathrm{r}i}|Y_{\mathrm{r}}^{i-1},Y_{1i+1}^n,J^n)-\sum_i I(M_{\mathrm{r}},Y_{\mathrm{r}}^{i-1},Y_{1i+1}^n;Y_{1i}|J^n)\\
&\le \sum_i I(M_{\mathrm{r}};Y_{\mathrm{r}i}|Y_{\mathrm{r}}^{i-1},Y_{1i+1}^n,J^n)-\sum_i I(M_{\mathrm{r}};Y_{1i}|Y_{\mathrm{r}}^{i-1},Y_{1i+1}^n,J^n)\\
&= \sum_i I(M_{\mathrm{r}};Y_{\mathrm{r}i}|Y_{\mathrm{r}}^{i-1},Y_{1i+1}^n,J^n)-\sum_i I(M_{\mathrm{r}};Y_{1i}|Y_{\mathrm{r}}^{i-1},Y_{1i+1}^n,J^n)\\
&= I(M_{\mathrm{r}};Y_{\mathrm{r}}^n|J^n)-I(M_{\mathrm{r}};Y_1^n|J^n)\\
&\le \sum_i I(M_{\mathrm{r}},Y_{\mathrm{r}}^{i-1};Y_{\mathrm{r}i}|J^n)\\
&= nI(W;Y_{\mathrm{r}}|J,T).
\end{aligned}
$$

On the other hand,

$$
nI(W;Y_{\mathrm{r}}|J,T)=\sum_i I(M_{\mathrm{r}},Y_{\mathrm{r}}^{i-1};Y_{\mathrm{r}i}|J^n)=I(M_{\mathrm{r}};Y_{\mathrm{r}}^n|J^n)=H(M_{\mathrm{r}})-I(M_{\mathrm{r}};J^n)\le nC_0-I(M_{\mathrm{r}};J^n).
$$

Thus,

$$
I(V,W;Y_{\mathrm{r}}|J,T)-I(V,W;Y_1|J,T)\le I(W;Y_{\mathrm{r}}|J,T)\le C_0.
$$

The cardinality bounds on the auxiliary random variables come from the standard Caratheodory-Bunt [Bun34] arguments and is omitted.

This completes the proof.

3.10. Proof of Proposition 3

Let rate R be the bound given in Corollary 2. Observe that

$$
\begin{aligned}
R &\le I(X;Y_1|T,W,V)-I(X;Y_{\mathrm{r}}|T,W,V)\\
&\le I(X;Y_1|T,W,V,Y_{\mathrm{r}})\\
&\le I(X;Y_1|Y_{\mathrm{r}},T).
\end{aligned}
\tag{61}
$$

Next, using the constraint in Corollary 2

$$
\begin{aligned}
R &\le I(X;Y_1|T,W,V)-I(X;Y_{\mathrm{r}}|T,W,V)\\
&\le I(X;Y_1|T,W,V)-I(X;Y_{\mathrm{r}}|T,W,V)+I(W;Y_{\mathrm{r}}|T)-I(V,W;Y_{\mathrm{r}}|T)+I(V,W;Y_1|T)\\
&= I(W,X;Y_1|T)+I(V;Y_1|W,X,T)-I(X,V;Y_{\mathrm{r}}|W,T)\\
&\le I(W,X;Y_1|T)+I(V;Y_1|W,X,T)-I(V;Y_{\mathrm{r}}|W,X,T)\\
&= I(W,X;Y_1|T)+I(V;Y_1|W,X,T)-I(V;Y_1,Y_{\mathrm{r}}|W,X,T) &&(62)\\
&\le I(W,X;Y_1|T), &&(63)
\end{aligned}
$$

where (62) follows from the Markov structure on the auxiliary random variable V. Therefore the bound given in Corollary 2 is less than or equal to the bound in Theorem 8.

Next, let R be the bound given in Theorem 8. We have

$$
R=\min\{I(W,X;Y_1|T),I(X;Y_1|Y_{\mathrm{r}},T)\}
$$

for some $p(x,t)p(w|t,y_{\mathrm{r}})$ satisfying $I(W,T;Y_{\mathrm{r}})\le C_0$. Assume that rate R is achieved by the bound in Corollary 2. Consider the following two cases:

Case 1: $R=I(X;Y_1|Y_{\mathrm{r}},T)$. Then the chain of inequalities leading to (61) must be all equality, implying that

$$
I(X;Y_{\mathrm{r}}|W,V,Y_1,T)=I(V,W;Y_1|Y_{\mathrm{r}},T)=0.
$$

Since the channel $p(y_1|x)$ is generic, it follows from Lemma 4 that $I(V,W;Y_1|Y_{\mathrm{r}},T)=0$ implies that $I(V,W;X|Y_{\mathrm{r}},T)=0$. Therefore, $(V,W)\to(Y_{\mathrm{r}},T)\to X$ form a Markov chain. Since Y_{r} is independent of (X,T), we deduce that $I(V,W,Y_{\mathrm{r}};X|T)=0$. Therefore, $I(X;Y_{\mathrm{r}}|W,V,T)=0$. Hence,

$$R\le I(X;Y_1|W,V,T)-I(X;Y_{\mathrm{r}}|W,V,T)=I(X;Y_1|W,V,T). \tag{64}$$

Since $(V,W)\to(Y_{\mathrm{r}},T)\to X$ form a Markov chain, from the constraint

$$I(V,W;Y_{\mathrm{r}}|T)-I(V,W;Y_1|T)\le C_0,$$

we obtain

$$I(V,W;Y_{\mathrm{r}}|Y_1,T)\le C_0.$$

From this equation and (64), it follows that compress-forward with time-sharing is optimal (using the characterization in Proposition 3 of [Kim07]).

Case 2: $R=I(W,X;Y_1|T)<I(X;Y_1|Y_{\mathrm{r}},T)$. In this case we must have $I(W,T;Y_{\mathrm{r}})=C_0$ since if $I(W,T;Y_{\mathrm{r}})<C_0$ in Theorem 8, time-sharing between W and Y_{r} would strictly increase $I(W,X;Y_1|T)$.

Then, from the assumption that the rate R is achieved by the bound in Corollary 2 we obtain that the chain of inequalities leading to (63) must be all equality. Thus, $I(X;Y_{\mathrm{r}}|T,W)=I(V;Y_{\mathrm{r}}|X,T,W,Y_1)=0$ and

$$I(V,W;Y_{\mathrm{r}}|T)-I(V,W;Y_1|T)=I(W;Y_{\mathrm{r}}|T). \tag{65}$$

We have $I(V;Y_{\mathrm{r}}|X,T,W,Y_1)=0$ and $I(V;Y_1|X,T,W,Y_{\mathrm{r}})=0$. From the assumption of Proposition 3 and Lemma 3, we obtain that $I(V;Y_{\mathrm{r}},Y_1|X,T,W)=0$. From $I(X;Y_{\mathrm{r}}|W,T)=0$, we obtain

$$I(V,X;Y_{\mathrm{r}}|T,W)=0$$

Therefore, $I(V;Y_{\mathrm{r}}|W,T)=0$. Hence, from (65) it follows that $I(V,W;Y_1|T)=0$. Hence,

$$R\le I(X;Y_1|W,V,T)-I(X;Y_{\mathrm{r}}|W,V,T)\le I(X;Y_1|W,T).$$

From the Markov structure on W, $I(W;Y_1|Y_{\mathrm{r}},T)=0$ and therefore $I(W;Y_{\mathrm{r}}|Y_1,T)\le I(W;Y_{\mathrm{r}}|T)\le C_0$. This implies that the rate R is achieved by compress-forward with time-sharing (using the characterization in Proposition 3 of [Kim07]). This completes the proof.

3.11. Proof of Theorem 9

We first compute the bound given in Proposition 1 which is

$$R\le I(X;Y_1|V)-I(X;Y_{\mathrm{r}}|V)$$

for some $p(x)p(v|x,y_{\mathrm{r}})$ satisfying

$$I(V;Y_{\mathrm{r}})-I(V;Y_1)\le C_0. \tag{66}$$

It is shown in Lemma 9 in Appendix D that we can restrict to jointly Gaussian random variables to evaluate the bound. Let the covariance of X,Y_{r} given V be

$$K_{X,Y_{\mathrm{r}}|V}=\begin{bmatrix}K_1 & \rho\sqrt{K_1K_2}\\ \rho\sqrt{K_1K_2} & K_2\end{bmatrix}\preceq\begin{bmatrix}P & 0\\ 0 & N_{\mathrm{r}}\end{bmatrix}$$

for some $0\le K_1\le P$ and $0\le K_2\le N_{\mathrm{r}}$ and $\rho\in[-1,1]$ such that

$$(P-K_1)(N_{\mathrm{r}}-K_2)\ge\rho^2K_1K_2. \tag{67}$$

Then, the upper bound becomes the maximum of

$$\frac{1}{2}\log\left(1+\frac{(\sqrt{K_1}+\sqrt{K_2}\rho)^2}{K_2(1-\rho^2)+N_1}\right)+\frac{1}{2}\log(1-\rho^2) \tag{68}$$

subject to

$$\frac{1}{2}\log(N_{\mathrm{r}})-\frac{1}{2}\log(K_2)-\frac{1}{2}\log(P+N_{\mathrm{r}}+N_1)+\frac{1}{2}\log(K_1+K_2+2\rho\sqrt{K_1K_2}+N_1)\le C_0 \tag{69}$$

and $0 \le K_1 \le P$ and $0 \le K_2 \le N_{\mathrm{r}}$, $\rho \in [-1, 1]$ and

$$(P - K_1)(N_{\mathrm{r}} - K_2) \ge \rho^2 K_1 K_2. \tag{70}$$

We argue that for any maximizing K_1, K_2, ρ, equations (69) and (70) must hold with equality. Regarding equality in (69) any maximizing auxiliary random variable V in (66) must satisfy $I(V; Y_{\mathrm{r}}) - I(V; Y_1) = C_0$, otherwise time-sharing between V and Y_{r} would strictly improve the upper bound on the rate R. We next show that (70) must hold with equality.

First consider the case of $\rho \ge 0$. Since constraint (69) holds with equality, we seek to maximize

$$C_0 - \frac{1}{2}\log(N_{\mathrm{r}}) + \frac{1}{2}\log\left(\frac{K_2(1-\rho^2)}{K_2(1-\rho^2) + N_1}\right) + \frac{1}{2}\log(P + N_{\mathrm{r}} + N_1)$$

subject to $0 \le K_1 \le P$ and $0 \le K_2 \le N_{\mathrm{r}}$, $\rho \in [0, 1]$ and

$$\begin{aligned}
&\frac{1}{2}\log(N_{\mathrm{r}}) - \frac{1}{2}\log(K_2) - \frac{1}{2}\log(P + N_{\mathrm{r}} + N_1) + \frac{1}{2}\log(K_1 + K_2 + 2\rho\sqrt{K_1K_2} + N_1) = C_0,\\
&(P - K_1)(N_{\mathrm{r}} - K_2) \ge \rho^2 K_1 K_2.
\end{aligned}$$

Setting $K_3 = K_2(1-\rho^2)$ and $K_4 = K_2\rho^2$, this can be rephrased as maximizing

$$C_0 - \frac{1}{2}\log(N_{\mathrm{r}}) + \frac{1}{2}\log\left(\frac{K_3}{K_3 + N_1}\right) + \frac{1}{2}\log(P + N_{\mathrm{r}} + N_1)$$

subject to $0 \le K_1 \le P$ and $K_3, K_4 \ge 0$, $0 \le K_3 + K_4 \le N_{\mathrm{r}}$ and

$$\frac{1}{2}\log(N_{\mathrm{r}}) - \frac{1}{2}\log(K_3 + K_4) - \frac{1}{2}\log(P + N_{\mathrm{r}} + N_1) + \frac{1}{2}\log((\sqrt{K_1} + \sqrt{K_4})^2 + K_3 + N_1) = C_0, \tag{71}$$

$$(P - K_1)(N_{\mathrm{r}} - K_3 - K_4) \ge K_1K_4. \tag{72}$$

The constraint $(P - K_1)(N_{\mathrm{r}} - K_3 - K_4) \ge K_1K_4$ gives an upper bound on K_1:

$$K_1 \le P\left(1 - \frac{K_4}{N_{\mathrm{r}} - K_3}\right).$$

Assume that the above bound is strict. Let us fix K_4, and increase K_3 and K_1 simultaneously such that

$$\frac{(\sqrt{K_1} + \sqrt{K_4})^2 + K_3 + N_1}{K_3 + K_4}$$

is preserved. This will strictly increase the overall expression because $K_3/(K_3 + N_1)$ is strictly increasing in K_3, and (71) is preserved. This shows that (72) must hold with equality for any maximizer and completes the argument in the case of $\rho \ge 0$.

Next, consider the case of $\rho < 0$. We show that no maximizer can have $\rho < 0$. Assume otherwise and take an optimizer (K_1, K_2, ρ) where $\rho < 0$. If $\sqrt{K_1} < 2|\rho|\sqrt{K_2}$, the expression in (68) is non-positive and cannot be the maximizer. To see this, observe that if $\sqrt{K_1} < 2|\rho|\sqrt{K_2}$ then (68) is less than or equal to

$$\frac{1}{2}\log\left(1 + \frac{K_2\rho^2}{K_2(1-\rho^2) + N_1}\right) + \frac{1}{2}\log(1-\rho^2) = \frac{1}{2}\log\left(\frac{K_2(1-\rho^2) + N_1(1-\rho^2)}{K_2(1-\rho^2) + N_1}\right) \le 0. \tag{73}$$

Therefore we can assume that $\sqrt{K_1} \ge 2|\rho|\sqrt{K_2}$. Replacing ρ by $-\rho$ and $\sqrt{K_1}$ by the smaller $\sqrt{K_1} + 2\rho\sqrt{K_2}$ does not change (68) while still keeps all the constraints satisfied. However, since we have strictly decreased K_1, (70) will not hold with equality by this transformation. However, the argument for the case of $\rho \ge 0$ shows that (70) must hold with equality. Therefore, $\rho < 0$ cannot hold for any maximizer.

To sum this up, for any maximizing K_1, K_2, ρ, equations (69) and (70) must hold with equality.

Next, consider the bound given in Corollary 2:

$$R \le I(X; Y_1|T, W, V) - I(X; Y_{\mathrm{r}}|T, W, V)$$

for some $p(t, x)p(y_1|x, y_{\mathrm{r}})p(y_{\mathrm{r}})p(w|t, y_{\mathrm{r}})p(v|t, x, y_{\mathrm{r}}, w)$ such that

$$I(V, W, T; Y_{\mathrm{r}}) - I(V, W, T; Y_1) + I(T; Y_1) \le I(W, T; Y_{\mathrm{r}}) \le C_0. \tag{74}$$

It is shown in Lemma 9 in Appendix D that we can restrict to jointly Gaussian random variables to evaluate the above bound. We use the proof by contradiction. Setting $\tilde{V} = (V, W, T)$, it follows that the above bound is less than or equal to the bound in Proposition 1. Assume that the bound in Corollary 2 matches the one in Proposition 1. Since (66) holds with equality for any maximizer of Proposition 1, we must have $I(T; Y_1) = 0$. This implies $I(T; X) = 0$. Thus, $I(W, T; X) = 0$. Let

$$K_{X,Y_{\mathrm{r}}|V,W,T} = \begin{bmatrix} K_1 & \rho\sqrt{K_1K_2} \\ \rho\sqrt{K_1K_2} & K_2 \end{bmatrix} \preceq K_{X,Y_{\mathrm{r}}|W,T} = \begin{bmatrix} P & 0 \\ 0 & N_{\mathrm{r}}' \end{bmatrix} \preceq K_{X,Y_{\mathrm{r}}} = \begin{bmatrix} P & 0 \\ 0 & N_{\mathrm{r}} \end{bmatrix}$$

for some $0 \le K_1 \le P$ and $0 \le K_2 \le N_{\mathrm{r}}' \le N_{\mathrm{r}}$ and $\rho \in [-1, 1]$ such that

$$(P - K_1)(N_{\mathrm{r}}' - K_2) \ge \rho^2 K_1 K_2. \tag{75}$$

Since (70) holds with equality, we must have $N_{\mathrm{r}}' = N_{\mathrm{r}}$. Therefore, (W, T) is independent of Y_{r}. This contradicts (74).

4. Conclusion and Final remarks

We presented new upper bounds on the capacity of several classes of relay channels and showed through several applications that our bounds can strictly improve upon previous bounds.

One key insight that leads to strict improvements over previous bounds is the appearance of the same auxiliary variables in multiple constraints with incompatible choice of optimizing auxiliary for each constraint. In particular, in Theorem 1, bounds (1) and (3) that strengthen the term $I(X; Y, Y_{\mathrm{r}}|X_{\mathrm{r}})$ in the cut-set bound are maximized when $I(U; Y|X_{\mathrm{r}}, Y_{\mathrm{r}}) = I(V; Y|X_{\mathrm{r}}, Y_{\mathrm{r}}) = I(X; Y_{\mathrm{r}}|V, X_{\mathrm{r}}, Y) = 0$. On the other hand, bound (6) that strengthens the term $I(X, X_{\mathrm{r}}; Y)$ is maximized when $I(V; Y_{\mathrm{r}}|X_{\mathrm{r}}, Y, X) = 0$. For generic channels, these constraints cannot hold simultaneously, leading to strict improvement over the cut-set bound. This leads to a trade-off in the realization of the auxiliary which is clearly demonstrated in the statement of Proposition 1. Similarly, a maximizer for bound (16) in Proposition 1 is $V = Y_{\mathrm{r}}$ while constraint (18) prevents V from becoming close to Y_{r}. Observe that (18) is deduced from the constraint (7) on auxiliary random variables and can tighten the bounds.

Other key insights that led to the new results are the techniques used in the evaluation of the bounds: either by carefully relaxing the constraints as in Corollary 1 or by identifying the optimal auxiliaries in the associated non-convex optimization problems as in Proposition 2 and Theorem 1. There are also instances in which properties of the optimal auxiliaries have been identified that make the optimization problem amenable to numerical evaluation by reducing the search dimension as in Theorem 2 and Theorem 6.

Our results have focused on the class of relay channels without self-interference and subclasses and applications therein. The same techniques can be used to establish upper bounds on the capacity of the general relay channel (e.g., see comment after Theorem 1). We do not have concrete examples, however, that motivate such extensions.

We gave a specific instance for which the addition of an auxiliary receiver helped improve our main bound. However, there are many other ways that auxiliary receivers could be introduced in these bounds. The study of using auxiliary receivers to strictly improve existing bounds is still rather nascent. It behooves to highlight the following evaluation of auxiliary-receiver bound related observation: A bound with an auxiliary receiver is computed as an infimum over all auxiliary receiver realizations; hence every fixed choice of auxiliary receiver yields a valid and computable bound. However, to obtain the best possible bound, the resultant optimization problem takes a max-min-max-min formulation in which the innermost minimum is over the various rate constraints, the next maximum is over the choice of auxiliary random variables, the subsequent minimum is over the choice of auxiliary receivers, and the outermost maximum is over the choice of the input distributions. Making optimization problems of the above form tractable would be an interesting problem to investigate.

References

[AH83] Rudolf Ahlswede and Te Han, *On source coding with side information via a multiple-access channel and related problems in multi-user information theory*, IEEE Transactions on Information Theory **29** (1983), no. 3, 396–412.

[AK74] Rudolf Ahlswede and János Körner, *On common information and related characteristics of correlated information sources*, Presented at the 7th Prague Conference on Information Theory, 1974.

[ARY09a] Marko Aleksic, Peyman Razaghi, and Wei Yu, *Capacity of a class of modulo-sum relay channels*, IEEE Transactions on Information Theory **55** (2009), no. 3, 921–930.

[ARY09b] ———, *Capacity of a class of modulo-sum relay channels*, IEEE Transactions on Information Theory **55** (2009), no. 3, 921–930.

[Bun34] Lucas Nicolaas Hendrik Bunt, *Bijdrage tot de theorie der convexe puntverzamelingen*, Ph.D. thesis, Univ. Groningne, Amsterdam, 1934.

[BWÖ17] Leighton Pate Barnes, Xiugang Wu, and Ayfer Özgür, *A solution to cover's problem for the binary symmetric relay channel: Geometry of sets on the hamming sphere*, 2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton), IEEE, 2017, pp. 844–851.

[CEG79] Thomas Cover and Abbas El Gamal, *Capacity theorems for the relay channel*, IEEE Transactions on information theory **25** (1979), no. 5, 572–584.

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

[CK78] Imre Csiszár and Janos Korner, *Broadcast channels with confidential messages*, IEEE transactions on information theory **24** (1978), no. 3, 339–348.

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

[CMG07] Hon-Fah Chong, Mehul Motani, and Hari Krishna Garg, *Generalized backward decoding strategies for the relay channel*, IEEE Transactions on Information Theory **53** (2007), no. 1, 394–401.

[Cov87] Thomas Cover, *The capacity of the relay channel*, Open Problems in Communication and Computation, Springer, 1987, pp. 72–73.

[EG79] Abbas El Gamal, *The capacity of a class of broadcast channels*, IEEE Transactions on Information Theory **25** (1979), no. 2, 166–169.

[EGA82] Abbas El Gamal and Mohammad-Reza Aref, *The capacity of the semideterministic relay channel (corresp.)*, IEEE Transactions on Information Theory **28** (1982), no. 3, 536–536.

[EGGN21] Abbas El Gamal, Amin Gohari, and Chandra Nair, *Strengthened cutset upper bounds on the capacity of the relay channel and applications*, IEEE International Symposium on Information Theory, IEEE, 2021.

[EGMZ06] Abbas El Gamal, Mehdi Mohseni, and Sina Zahedi, *Bounds on capacity and minimum energy-per-bit for awgn relay channels*, IEEE Transactions on information theory **52** (2006), no. 4, 1545–1561.

[EGZ05] Abbas El Gamal and Sina Zahedi, *Capacity of a class of relay channels with orthogonal components*, IEEE Transactions on Information Theory **51** (2005), no. 5, 1815–1817.

[EK11] Abbas El Gamal and Young-Han Kim, *Network information theory*, Cambridge University Press, 2011.

[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] Amin Gohari and Chandra Nair, *Outer bounds for multiuser settings: The auxiliary receiver approach*, submitted to the IEEE Transactions on Information Theory (2020).

[Kim07] Young-Han Kim, *Coding techniques for primitive relay channels*, Proc. Forty-Fifth Annual Allerton Conf. Commun., Contr. Comput, 2007, p. 2007.

[LGY13] K. Luo, R. H. Gohary, and H. Yanikomeroglu, *Analysis of the generalized df-cf for gaussian relay channels: Decode or compress?*, IEEE Transactions on Communications **61** (2013), no. 5, 1810–1821.

[Liu20] Jingbo Liu, *Minorization via mixed volumes and cover's problem for general channels*, arXiv preprint arXiv:2012.14521 (2020).

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

[Nai13] Chandra Nair, *Upper concave envelopes and auxiliary random variables*, International Journal of Advances in Engineering Sciences and Applied Mathematics **5** (2013), no. 1, 12–20.

[OR95] Alon Orlitsky and James R Roche, *Coding for computing*, Proceedings of IEEE 36th Annual Foundations of Computer Science, IEEE, 1995, pp. 502–511.

[TU08] Ravi Tandon and Sennur 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.

[VDM71] Edward C Van Der Meulen, *Three-terminal communication channels*, Advances in applied Probability (1971), 120–154.

[WBÖ17] Xiugang Wu, Leighton Pate Barnes, and Ayfer Özgür, *The geometry of the relay channel*, 2017 IEEE International Symposium on Information Theory (ISIT), IEEE, 2017, pp. 2233–2237.

[WBO19] Xiugang Wu, Leighton Pate Barnes, and Ayfer 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.

[WOX17] Xiugang Wu, Ayfer Ozgur, and Liang-Liang Xie, *Improving on the cut-set bound via geometric analysis of typical sets*, IEEE Transactions on Information Theory **63** (2017), no. 4, 2254–2277.

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

APPENDIX

A. Some Mathematical Preliminaries

Lemma 3 (Double-Markovity, Exercise 16.25 in [CK11], also see [AK74])**.** *Let (X, Y, Z) be random variables such that $X \to Y \to Z$ and $X \to Z \to Y$ are Markov chains. Then there exists functions $f(Y)$ and $g(Z)$ such that $\mathrm{P}(f(Y) = g(Z)) = 1$ and X is conditionally independent of (Y, Z) given $(f(X), g(Y))$.*

Remark 16. While this is not explicitly stated in the references, the above statement holds true for random variables defined on Polish spaces under the Borel σ-algebra, which guarantees the existence of regular conditional probabilities. In this paper, we work with finite valued random variables; with an exception being additive Gaussian noise settings.

Lemma 4. *Let $p(y|x)$ be a* generic *channel (see Definition 4) and assume that $W \to X \to Y$ form a Markov chain. Then, $I(W; Y) = 0$ implies $I(W; X) = 0$.*

Proof. Given $I(W;Y)=0$. Let w_1, w_2 be such that $p(w_i)>0$ for $i=1,2$. Since $I(W;Y)=0$ we have $P(Y=y|W=w_1)=P(Y=y|W=w_2)$. This implies that $\sum_x P(X=x|W=w_1)\vec{v}_x=\sum_x P(X=x|W=w_2)\vec{v}_x$, where $\vec{v}_x := p(y|x)$. From full row-rank, i.e., the linear independence of $\{\vec{v}_x\}$, it follows that $P(X=x|W=w_1)=P(X=x|W=w_2)$ for every x. Since this holds for any pair w_1, w_2 such that $p(w_i)>0$, we obtain $I(W;X)=0$, i.e., W is independent of X. □

Lemma 5. *Let $Y=X+Z$ be an AWGN channel. Assume that $W\to X\to Y$ form a Markov chain. Then, we have*

(i) $I(W;Y)=0$ *implies that* $I(W;X)=0$.

(ii) $I(W;X|Y)=0$ *(or equivalently* $I(W;Y)=I(W;X)$*) implies that* $I(W;X)=0$.

(iii) $I(X;Y|W)=0$ *implies that* X *is a function of* W.

Proof. (i): This is standard and follows from the non-vanishing property of the characteristic function of a Gaussian distribution.

(ii): This follows from Lemma 3 along with the observation that if $f(X)=g(X+Z)$ with probability one, then both functions $f(X), g(X+Z)$ have to be constant with probability one.

(iii): Note that $I(X;X+Z)=0$ implies that the characteristic function of X satisfies the equation $\Phi_X(t_1+t_2)=\Phi_X(t_1)\Phi_X(t_2)$, whose only solution (in the space of characteristic functions) is $\Phi_X(t)=e^{itc}$, implying that $X=c$ with probability one. Returning to $I(X;X+Z|W)=0$, we have that conditioned on W, X is a constant; implying that X is a function of W as required. □

B. Evaluation of Corollary 1 for the Gaussian relay channel

Lemma 6. *For the evaluation of Corollary 1 for the Gaussian relay channel, we can assume that the random variables are jointly Gaussian satisfying the requisite Markov Chains.*

Proof. We follow the ideas in [GN14] and further focus only on the key steps that are unique here. Let us consider a class of problems in which we replace $Y_{\rm r}$ with $\begin{bmatrix} Y_{\rm r} \\ \epsilon Z_{\rm r}+W\end{bmatrix}$ and Y by $\begin{bmatrix} Y \\ \epsilon Z_{\rm r}+W\end{bmatrix}$, where $W\sim\mathcal{N}(0,1)$ is independent of previously defined random variables. Let us call these random variables $Y_{{\rm r},\epsilon}$ and Y_ϵ, respectively. The main proof step is to show that for the optimization problem described below, Gaussian distributions are the maximizers for any $\epsilon>0$. Following this step, one can move to the limit $\epsilon=0$ (the original problem) by using the power constraints, the additive Gaussian noise model, and other arguments found in the Appendix of [GN14]. These are rather standard in analysis and hence the details are omitted.

Consider the following optimization problem: for $\epsilon>0$, we wish to compute the supremum of R_ϵ satisfying the following:

$$
\begin{aligned}
R_\epsilon &\le I(X;Y_{\rm r}|X_{\rm r}) - I(X;Y_{{\rm r},\epsilon}|V,X_{\rm r}) + I(X;Y_\epsilon|V,X_{\rm r}) - \epsilon h(Y_\epsilon|V,X_{\rm r}),\\
R_\epsilon &\le I(X,X_{\rm r};Y) - h(Y_{\rm r}|X_{\rm r},X,Y) + h(Y_{{\rm r},\epsilon}|X_{\rm r},X,Y,V)\\
&\quad + \epsilon\big(I(X;Y_{\rm r}|X_{\rm r}) - I(X;Y_{{\rm r},\epsilon}|V,X_{\rm r}) + I(X;Y_\epsilon|V,X_{\rm r}) - \epsilon h(Y_\epsilon|V,X_{\rm r})\big)
\end{aligned}
$$

for some $P_{X,X_{\rm r}}P_{Y,Y_{\rm r}|X,X_{\rm r}}P_{V|X,X_{\rm r},Z_{\rm r}}$ such that $\mathrm{E}(X^2)\le P$, $\mathrm{E}(X_{\rm r}^2)\le P$,

$$
h(Y_\epsilon|V,X_{\rm r}) - h(Y_{{\rm r},\epsilon}|V,X_{\rm r}) \le \max_{P_{U|X,X_{\rm r}}}\left[h(Y_\epsilon|U) - h(Y_{{\rm r},\epsilon}|U)\right].
$$

Suppose P^ϵ achieves the supremum of R_ϵ (the existence of maximizer follows from reasonably standard arguments - please refer to the Appendix of [GN14] for an illustration of the ideas involved), and under P^ϵ let the values of the right-hand-sides of the constraints be A, B. Then by taking the usual doubling followed by rotation we obtain

$$
\begin{aligned}
A = &\frac{1}{2}\big(I(X_+;Y_{{\rm r},+}|X_{{\rm r},+}) - I(X_+;Y_{{\rm r},\epsilon,+}|V_+,X_{{\rm r},+}) + I(X_+;Y_{\epsilon,+}|V_+,X_{{\rm r},+}) - \epsilon h(Y_{\epsilon,+}|V_+,X_{{\rm r},+}),\\
&+ I(X_-;Y_{{\rm r},-}|X_{{\rm r},-}) - I(X_-;Y_{{\rm r},\epsilon,-}|V_-,X_{{\rm r},-}) + I(X_-;Y_{\epsilon,-}|V_-,X_{{\rm r},-}) - \epsilon h(Y_{\epsilon,-}|V_-,X_{{\rm r},-})\big)\\
&- \frac{1}{2}\big(I(X_{{\rm r},-};Y_{{\rm r},+}|X_{{\rm r},+}) + I(X_-;Y_{{\rm r},\epsilon,+}|V_+,X_{{\rm r},+},X_+) - I(X_-;Y_{\epsilon,+}|V_+,X_{{\rm r}+},X_+) + I(X_{{\rm r},+},Y_{{\rm r},+};Y_{{\rm r},-}|X_{{\rm r},-})\\
&\quad + I(X_+;Y_{{\rm r},\epsilon,-}|V_-,X_{{\rm r},-},X_-) - I(X_+;Y_{\epsilon,-}|V_-,X_{{\rm r}-},X_-) + \epsilon I(Y_{{\rm r},\epsilon,-};Y_{\epsilon,+}|V_1,V_2,X_{{\rm r},-},X_{{\rm r},+})\big),
\end{aligned}
$$

where $V_+ = (V_1, V_2, X_{\mathrm{r},-}, Y_{\mathrm{r},\epsilon,-}), V_- = (V_1, V_2, X_{\mathrm{r},+}, Y_{\epsilon,+})$. Observe that

$$I(X_-; Y_{\mathrm{r},\epsilon,+}|V_+, X_{\mathrm{r},+}, X_+) - I(X_-; Y_{\epsilon,+}|V_+, X_{\mathrm{r}+}, X_+) = I(X_-; Y_{r,+}|V_+, X_{\mathrm{r},+}, X_+, \epsilon Z_{\mathrm{r}_+} + W_+).$$

A similar equality also holds for the second pair of terms colored in olive and blue.

Now let Q be a uniformly distributed binary random variable taking values $+$ and $-$ each with probability 0.5, and define $V_\dagger = (V_Q, Q)$, $X_\dagger = X_Q$ and other variables similar to that of X. Then we obtain

$$\begin{aligned} A \le\ & I(X_\dagger; Y_{\mathrm{r},\dagger}|X_{\mathrm{r},\dagger}) - I(X_\dagger; Y_{\mathrm{r},\epsilon,\dagger}|V_\dagger, X_{\mathrm{r},\dagger}) + I(X_\dagger; Y_{\epsilon,\dagger}|V_+, X_{\mathrm{r},\dagger}) - \epsilon h(Y_{\epsilon,\dagger}|V_\dagger, X_{\mathrm{r},\dagger}), \\ & - \frac{1}{2}\big(I(X_{\mathrm{r},-}; Y_{\mathrm{r},+}|X_{\mathrm{r},+}) + I(X_{\mathrm{r},+}, Y_{\mathrm{r},+}; Y_{\mathrm{r},-}|X_{\mathrm{r},-}) + \epsilon I(Y_{\mathrm{r},\epsilon,-}; Y_{\epsilon,+}|V_1, V_2, X_{\mathrm{r},-}, X_{\mathrm{r},+})\big) \end{aligned}$$

A similar inequality for B can also be written. Further observe that the dagger variables satisfy the constraint as well. This implies that for P^ϵ to be optimal, it is necessary (from the Skitovic-Darmois characterization, see [GN14] for details) that

$$I(X_{\mathrm{r},+}, Y_{\mathrm{r},+}; Y_{\mathrm{r},-}|X_{\mathrm{r},-}) = 0, \quad I(Y_{\mathrm{r},\epsilon,-}; Y_{\epsilon,+}|V_1, V_2, X_{\mathrm{r},-}, X_{\mathrm{r},+}) = 0,$$

implying Gaussianity of the conditional distributions $P_{X|X_\mathrm{r}}$ and $P_{X,Y_\mathrm{r}|X_\mathrm{r},V}$; and further that the covariance of $P_{X,Y_\mathrm{r}|X_\mathrm{r},V}$ does not depend on the conditioned variables. ☐

C. Evaluation of Proposition 1 for the Gaussian product-form relay channel

Lemma 7. *To evaluate Proposition 1 for the Gaussian product-form relay channel, we can assume that the random variables are jointly Gaussian satisfying the requisite Markov Chains.*

Proof. As in the proof of Lemma 6 we follow the ideas in [GN14] and the steps are very similar to those of the proof of the previous lemma. Consider a class of problems in which we replace Y_r with $\begin{bmatrix} Y_\mathrm{r} \\ \epsilon Z_\mathrm{r} + W \end{bmatrix}$ and Y_1 by $\begin{bmatrix} Y_1 \\ \epsilon Z_\mathrm{r} + W \end{bmatrix}$, where $W \sim \mathcal{N}(0,1)$ is independent of previously defined random variables. Let us call these random variables $Y_{\mathrm{r},\epsilon}$ and $Y_{1,\epsilon}$, respectively. For $\epsilon > 0$, we wish to compute the supremum of R_ϵ satisfying the following:

$$R_\epsilon \le I(X; Y_\mathrm{r}) + I(X; Y_{1,\epsilon}|V) - I(X; Y_{\mathrm{r},\epsilon}|V) - \epsilon h(Y_{1,\epsilon}|V)$$

for some $P_X P_{Y_1,Y_\mathrm{r}|X} P_{V|X,Z_\mathrm{r}}$ such that

$$I(V; Y_{\mathrm{r},\epsilon}) - I(V; Y_{1,\epsilon}) \le C_0,$$

and identical power constraints. Suppose P^ϵ achieves the supremeum of R_ϵ (existence justified using routine arguments), and under P^ϵ let the value of the right-hand-side be A. Then by taking the usual doubling followed by rotation we obtain

$$\begin{aligned} A =\ & \frac{1}{2}\Big(I(X_+; Y_{\mathrm{r},+}) - I(X_+; Y_{\mathrm{r},\epsilon,+}|V_+) + I(X_+; Y_{1,\epsilon,+}|V_+) - \epsilon h(Y_{1,\epsilon,+}|V_+), \\ & + I(X_-; Y_{\mathrm{r},-}) - I(X_-; Y_{\mathrm{r},\epsilon,-}|V_-) + I(X_-; Y_{1,\epsilon,-}|V_-) - \epsilon h(Y_{1,\epsilon,-}|V_-)\Big) \\ & - \frac{1}{2}\Big(I(X_-; Y_{\mathrm{r},\epsilon,+}|V_+, X_+) - I(X_-; Y_{1,\epsilon,+}|V_+, X_+) + I(Y_{\mathrm{r},+}; Y_{\mathrm{r},-}) \\ & + I(X_+; Y_{\mathrm{r},\epsilon,-}|V_-, X_-) - I(X_+; Y_{1,\epsilon,-}|V_-, X_-) + \epsilon I(Y_{\mathrm{r},\epsilon,-}; Y_{1,\epsilon,+}|V_1, V_2)\Big) \end{aligned}$$

where $V_+ = (V_1, V_2, Y_{\mathrm{r},\epsilon,-}), V_- = (V_1, V_2, Y_{1,\epsilon,+})$. Observe that

$$I(X_-; Y_{\mathrm{r},\epsilon,+}|V_+, X_+) - I(X_-; Y_{1,\epsilon,+}|V_+, X_+) = I(X_-; Y_{r,+}|V_+, X_+, \epsilon Z_{\mathrm{r}_+} + W_+).$$

As before, let Q be a uniformly distributed binary random variable taking values $+$ and $-$ each with probability 0.5, and define $V_\dagger = (V_Q, Q)$, $X_\dagger = X_Q$ and other variables similar to that of X. Using this we obtain that

$$\begin{aligned} A \le\ & I(X_\dagger; Y_{\mathrm{r},\dagger}) - I(X_\dagger; Y_{\mathrm{r},\epsilon,\dagger}|V_\dagger) + I(X_\dagger; Y_{1,\epsilon,\dagger}|V_+) - \epsilon h(Y_{\epsilon,\dagger}|V_\dagger), \\ & - \frac{1}{2}\Big(I(Y_{\mathrm{r},+}; Y_{\mathrm{r},-}) + \epsilon I(Y_{\mathrm{r},\epsilon,-}; Y_{\epsilon,+}|V_1, V_2)\Big). \end{aligned}$$

Further observe that the dagger variables satisfy

$$I(V_\dagger; Y_{\mathrm{r},\epsilon,\dagger}) - I(V_\dagger; Y_{1,\epsilon,\dagger}) \le C_0$$

as well. Since the dagger variables are also a feasible choice, for P^ϵ to be optimal, it is necessary (from the Skitovic-Darmois characterization, see [GN14] for details) that $I(Y_{\mathrm{r},+}; Y_{\mathrm{r},-}) = 0$ and $I(Y_{\mathrm{r},\epsilon,-}; Y_{\epsilon,+}|V_1, V_2) = 0$, implying that P_X and $P_{X,Y_\mathrm{r}|V}$ are both Gaussians and that the covariance of the latter does not depend on V. □

Lemma 8. *In evaluating Proposition 1 in the space of jointly Gaussian distributions, the covariance of $K_{X,Y_\mathrm{r}|V}$ is unique for the maximizing distribution and equals*

$$K_{X,Y_\mathrm{r}|V} = \begin{bmatrix} K_1^* & \rho^* \sqrt{K_1^* K_2^*} \\ \rho^* \sqrt{K_1^* K_2^*} & K_2^* \end{bmatrix}$$

where

$$K_1^* = \begin{cases} -N_1 \frac{2^{2C_0}(P+N_\mathrm{r}) - (P+N_\mathrm{r})}{2^{2C_0}(P+N_1) - (P+N_\mathrm{r})} + P \frac{2^{2C_0}(N_1 - N_\mathrm{r})}{2^{2C_0}(P+N_1) - (P+N_\mathrm{r})} & S_{12} \ge S_{13} + S_{23} + S_{13}S_{23}, \\ P\left(1 - \frac{P(N_1+P)^2(2^{2C_0}-1)}{(P+N_\mathrm{r})(2^{2C_0}-1)((N_1+P)^2 - N_1^2 2^{-2C_0}) + (N_\mathrm{r} - N_1)P^2}\right) & \textit{otherwise}. \end{cases} \tag{76}$$

$$K_2^* = \frac{(K_1^* + N_1)(P + N_\mathrm{r})}{(P+N_1)2^{2C_0}} \tag{77}$$

$$\rho^* = \frac{P - \sqrt{(P - K_1^*)(P + N_\mathrm{r} - K_2^*)}}{\sqrt{K_1^* K_2^*}}. \tag{78}$$

Proof. Let

$$K_{X,Y_\mathrm{r}|V} = \begin{bmatrix} K_1 & \rho\sqrt{K_1 K_2} \\ \rho\sqrt{K_1 K_2} & K_2 \end{bmatrix} \preceq \begin{bmatrix} P & P \\ P & P + N_\mathrm{r} \end{bmatrix}$$

for some $0 \le K_1 \le P$ and $0 \le K_2 \le N_\mathrm{r} + P$ and $\rho \in [-1, 1]$ satisfying

$$(P - K_1)(P + N_\mathrm{r} - K_2) \ge (P - \rho\sqrt{K_1 K_2})^2. \tag{79}$$

Then, the bound becomes

$$R \le \frac{1}{2}\log\left(\frac{P + N_\mathrm{r}}{N_\mathrm{r}}\right) + \frac{1}{2}\log\left(\frac{K_1 + N_1}{N_1}\right) + \frac{1}{2}\log\left(1 - \rho^2\right)$$

subject to

$$\frac{1}{2}\log(P + N_\mathrm{r}) - \frac{1}{2}\log(K_2) - \frac{1}{2}\log(P + N_1) + \frac{1}{2}\log(K_1 + N_1) \le C_0.$$

The optimizer ρ is non-negative; otherwise if the optimizer ρ is negative, moving to $\rho = 0$ strictly increases the expression. This is because if we decrease ρ^2 while fixing K_1 and K_2, the constraint

$$(P - K_1)(P + N_\mathrm{r} - K_2) \ge (P - \rho\sqrt{K_1 K_2})^2$$

will still hold. The expression we wish to maximize

$$(K_1 + N_1)(1 - \rho^2)$$

will also increase as we decrease ρ^2. Increasing K_2 while decreasing ρ such that $\rho\sqrt{K_2}$ is preserved, shows that either $K_2 = P + N_\mathrm{r}$ or else

$$(P - K_1)(P + N_\mathrm{r} - K_2) = (P - \rho\sqrt{K_1 K_2})^2.$$

Even in the case of $K_2 = P + N_\mathrm{r}$, the inequality

$$(P - K_1)(P + N_\mathrm{r} - K_2) \ge (P - \rho\sqrt{K_1 K_2})^2 \tag{80}$$

must hold with equality. Thus, any optimal choice for ρ must satisfy $\rho \ge 0$ and

$$\rho = \frac{P - \sqrt{(P - K_1)(P + N_\mathrm{r} - K_2)}}{\sqrt{K_1 K_2}}. \tag{81}$$

Note that the other solution for ρ in (80) is

$$\rho = \frac{P + \sqrt{(P-K_1)(P+N_\mathrm{r}-K_2)}}{\sqrt{K_1K_2}} \tag{82}$$

However, the optimizer ρ must satisfy (81). If instead (82) holds, reducing ρ to $P/\sqrt{K_1K_2}$ would strictly increase the objective function while continuing to satisfy (79).

Next, we can infer that any the maximizing auxiliary random variable V must satisfy $I(V;Y_\mathrm{r}) - I(V;Y_1) = C_0$, otherwise time-sharing between V and Y_r would strictly improve the upper bound on the rate R. In other words, the maximizing distribution is such that (18) holds with equality. This yields

$$K_2 = \frac{(K_1+N_1)(P+N_\mathrm{r})}{(P+N_1)2^{2C_0}}$$

and

$$\begin{aligned}\rho &= \frac{P - \sqrt{(P-K_1)(P+N_\mathrm{r}-K_2)}}{\sqrt{K_1K_2}}\\ &= \frac{P - \sqrt{(P-K_1)(P+N_\mathrm{r}-(K_1+N_1)\frac{P+N_\mathrm{r}}{P+N_1}2^{-2C_0})}}{\sqrt{K_1(K_1+N_1)\frac{P+N_\mathrm{r}}{P+N_1}2^{-2C_0}}}.\end{aligned}$$

We wish to maximize $(K_1+N_1)(1-\rho^2)$. Letting $\zeta = \frac{P+N_\mathrm{r}}{P+N_1} \geq 1$, we seek to maximize, subject to $K_1 \in [0,P]$, the following expression:

$$\begin{aligned}&\frac{K_1(K_1+N)\zeta 2^{-2C_0} - \left(P - \sqrt{(P-K_1)((P+N_1)\zeta - (K_1+N_1)\zeta 2^{-2C_0})}\right)^2}{K_1\zeta 2^{-2C_0}}\\ &= \frac{-P^2 - P(P+N_1)\zeta + K_1(P+N_1)\zeta + P(K_1+N_1)\zeta 2^{-2C_0} + 2P\sqrt{(P-K_1)((P+N_1)\zeta - (K_1+N_1)\zeta 2^{-2C_0})}}{K_1\zeta 2^{-2C_0}}.\end{aligned}$$

Taking the derivative with respect to K_1, and simplifying it we obtain

$$\begin{aligned}&K_1\zeta(-2^{-2C_0}N_1 + 2^{-2C_0}P + N_1 + P) + 2\zeta P(2^{-2C_0}N_1 - N_1 - P)\\ &= (2^{-2C_0}N_1\zeta - N_1\zeta - P\zeta - P)\sqrt{(P-K_1)((P+N_1)\zeta - (K_1+N_1)\zeta 2^{-2C_0})}\end{aligned} \tag{83}$$

If we raise both sides to power two, we get a quadratic equation. This quadratic equation has the following two roots:

$$\begin{aligned}K_{1a} &= -N_1\frac{\zeta - 2^{-2C_0}\zeta}{1-2^{-2C_0}\zeta} + P\frac{1-\zeta}{1-2^{-2C_0}\zeta},\\ K_{1b} &= P\left(1 - \frac{P(N_1+P)(2^{-2C_0}-1)}{\zeta(2^{-2C_0}-1)((N_1+P)^2 - N_1^2 2^{-2C_0}) + (1-\zeta)2^{-2C_0}P^2}\right).\end{aligned}$$

One can verify that in the case of

$$S_{13} + S_{23} + S_{13}S_{23} \leq S_{12}$$

only the root K_{1a} is in $[0,P]$ while when

$$S_{13} + S_{23} + S_{13}S_{23} > S_{12}$$

only the root K_{1b} is in $[0,P]$. Let K_1^* denote the unique root that lies in $[0,P]$. The function

$$\frac{-P^2 - P(P+N_1)\zeta + K_1(P+N_1)\zeta + P(K_1+N_1)\zeta 2^{-2C_0} + 2P\sqrt{(P-K_1)((P+N_1)\zeta - (K_1+N_1)\zeta 2^{-2C_0})}}{K_1\zeta 2^{-2C_0}}.$$

is increasing for $K_1 \in (0, K_1^*]$ and decreasing for $K_1 \in [K_1^*, P]$. Therefore, it reaches its maximum at $K_1 = K_1^*$. $\square$

D. Evaluation of Corollary 2 and Proposition 1 for the Gaussian relay channel with orthogonal receiver components and i.i.d.output

Lemma 9. *It suffices to consider jointly Gaussian distributions when evaluating the outer bounds in Corollary 2 or Proposition 1 for the channel described by (29), (30).*

Proof. As with the other arguments, we will first define a perturbed Y_1 defined according to

$$Y_{1,\epsilon} = \begin{bmatrix} X + Y_{\mathrm{r}} + Z_1 \\ \epsilon Y_{\mathrm{r}} + Z_2 \end{bmatrix}. \tag{84}$$

Here $Z_2 \sim \mathcal{N}(0,1)$ is mutually independent of all other random variables.

Let us consider the argument for Proposition 1 first. Consider the maximum of

$$I(X;Y_{1,\epsilon}|V) - I(X;Y_{\mathrm{r}}|V) - \delta I(Y_{1,\epsilon};\delta Y_{1,\epsilon} + Z_3|V) \tag{85}$$

subject to $p(x)p(y_{1,\epsilon},y_{\mathrm{r}}|x)p(v|x,y_{\mathrm{r}})p(z_3)$ and $I(V;Y_{\mathrm{r}}) - I(V;Y_{1,\epsilon}) \le C_0$. Here $Z_3 \sim \mathcal{N}(0, I_2)$ is mutually independent of all other random variables. The idea is to show that the maximum is attained by a jointly Gaussian distribution, and then take the limit and $\epsilon, \delta \to 0$.

Take two independent copies of the maximizer $(X_1, Y_{\mathrm{r},1}, Y_{1,\epsilon,1}, V_1, Z_{3,1})$ and $(X_2, Y_{\mathrm{r},2}, Y_{1,\epsilon,2}, V_2, Z_{3,2})$, and perform two orthogonal rotations to obtain the $+$ and $-$ random variables $(X_+, Y_{\mathrm{r},+}, Y_{1,\epsilon,+}, Z_{3,+})$ and $(X_-, Y_{\mathrm{r},-}, Y_{1,\epsilon,-}, Z_{3,-})$. Let $V = (V_1, V_2)$. Now observe that

$$\begin{aligned}
&I(X_+,X_-;Y_{1,\epsilon,+},Y_{1,\epsilon,-}|V) - I(X_+,X_-;Y_{\mathrm{r},+},Y_{\mathrm{r},-}|V)\\
&\quad = I(X_+;Y_{1,\epsilon,+}|V,Y_{1,\epsilon,-}) - I(X_+;Y_{\mathrm{r},+}|V,Y_{1,\epsilon,-})\\
&\qquad + I(X_-;Y_{1,\epsilon,-}|V,Y_{\mathrm{r},+}) - I(X_-;Y_{\mathrm{r},-}|V,Y_{\mathrm{r},+})\\
&\qquad + I(X_-;Y_{1,\epsilon,+}|V,Y_{1,\epsilon,-},X_+) - I(X_-;Y_{\mathrm{r},+}|V,Y_{1,\epsilon,-},X_+)\\
&\qquad + I(X_+;Y_{1,\epsilon,-}|V,Y_{\mathrm{r},+},X_-) - I(X_+;Y_{\mathrm{r},-}|V,Y_{\mathrm{r},+},X_-)\\
&\quad = I(X_+;Y_{1,\epsilon,+}|V,Y_{1,\epsilon,-}) - I(X_+;Y_{\mathrm{r},+}|V,Y_{1,\epsilon,-})\\
&\qquad + I(X_-;Y_{1,\epsilon,-}|V,Y_{\mathrm{r},+}) - I(X_-;Y_{\mathrm{r},-}|V,Y_{\mathrm{r},+})\\
&\qquad - I(X_-;Y_{\mathrm{r},+}|V,Y_{1,\epsilon,-},Y_{1,\epsilon,+},X_+) - I(X_+;Y_{\mathrm{r},-}|V,Y_{\mathrm{r},\epsilon,+},Y_{1,\epsilon,-},X_-).
\end{aligned}$$

The last step follows from

$$I(X_-;Y_{1,\epsilon,+}|Y_{\mathrm{r},+},V,Y_{1,\epsilon,-},X_+) = I(X_+;Y_{1,\epsilon,-}|Y_{\mathrm{r},-},V,Y_{\mathrm{r},+},X_-) = 0$$

which follows from the fact that $(Z_{1,+}, Z_{1,-})$ and $(Z_{2,+}, Z_{2,-})$ are pairs of independent random variables. Next, we also have

$$\begin{aligned}
&I(Y_{1,\epsilon,+},Y_{1,\epsilon,-};\delta Y_{1,\epsilon,+} + W_{3+},\delta Y_{1,\epsilon,-} + Z_{3,-}|V)\\
&\quad = I(Y_{1,\epsilon,+};\delta Y_{1,\epsilon,+} + Z_{3,+}|V,Y_{1,\epsilon,-}) + I(Y_{1,\epsilon,-};\delta Y_{1,\epsilon,-} + Z_{3,-}|V)\\
&\qquad + I(Y_{1,\epsilon,-};\delta Y_{1,\epsilon,+} + Z_{3,+}|V,\delta Y_{1,\epsilon,-} + Z_{3,-})\\
&\quad = I(Y_{1,\epsilon,+};\delta Y_{1,\epsilon,+} + Z_{3,+}|V,Y_{1,\epsilon,-}) + I(Y_{1,\epsilon,-};\delta Y_{1,\epsilon,-} + Z_{3,-}|V,Y_{\mathrm{r},+})\\
&\qquad + I(Y_{1,\epsilon,-};\delta Y_{1,\epsilon,+} + Z_{3,+}|V,\delta Y_{1,\epsilon,-} + Z_{3,-}) + I(Y_{\mathrm{r},+};\delta Y_{1,\epsilon,-} + Z_{3,-}|V).
\end{aligned} \tag{86}$$

Thus, we obtain

$$\begin{aligned}
&I(X_+,X_-;Y_{1,\epsilon,+},Y_{1,\epsilon,-}|V) - I(X_+,X_-;Y_{\mathrm{r},+},Y_{\mathrm{r},-}|V) - \delta I(Y_{1,\epsilon,+},Y_{1,\epsilon,-};\delta Y_{1,\epsilon,+} + Z_{3,+},\delta Y_{1,\epsilon,-} + Z_{3,-}|V)\\
&\quad = I(X_+;Y_{1,\epsilon,+}|V,Y_{1,\epsilon,-}) - I(X_+;Y_{\mathrm{r},+}|V,Y_{1,\epsilon,-}) - \delta I(Y_{1,\epsilon,+};\delta Y_{1,\epsilon,+} + Z_{3,+}|V,Y_{1,\epsilon,-})\\
&\qquad + I(X_-;Y_{1,\epsilon,-}|V,Y_{\mathrm{r},+}) - I(X_-;Y_{\mathrm{r},-}|V,Y_{\mathrm{r},+}) - \delta I(Y_{1,\epsilon,-};\delta Y_{1,\epsilon,-} + Z_{3,-}|V,Y_{\mathrm{r},+})\\
&\qquad - I(X_-;Y_{\mathrm{r},+}|V,Y_{1,\epsilon,-},Y_{1,\epsilon,+},X_+) - I(X_+;Y_{\mathrm{r},-}|V,Y_{\mathrm{r},\epsilon,+},Y_{1,\epsilon,-},X_-)\\
&\qquad - \delta I(Y_{1,\epsilon,-};\delta Y_{1,\epsilon,+} + Z_{3,+}|V,\delta Y_{1,\epsilon,-} + Z_{3,-}) - \delta I(Y_{\mathrm{r},+};\delta Y_{1,\epsilon,-} + Z_{3,-}|V)
\end{aligned} \tag{87}$$

Further observe that

$$I(V;Y_{\mathrm{r},+},Y_{\mathrm{r},-}) - I(V;Y_{1,\epsilon,+},Y_{1,\epsilon,-})$$

$$= I(V;Y_{\mathrm{r},+}) + I(V,Y_{\mathrm{r},+};Y_{\mathrm{r},-}) - I(V;Y_{1,\epsilon,-}) - I(V,Y_{1,\epsilon,-};Y_{1,\epsilon,+}) + I(Y_{1,\epsilon,-};Y_{1,\epsilon,+})$$
$$= I(V,Y_{1,\epsilon,-};Y_{\mathrm{r},+}) + I(V,Y_{\mathrm{r},+};Y_{\mathrm{r},-}) - I(V,Y_{\mathrm{r},+};Y_{1,\epsilon,-}) - I(V,Y_{1,\epsilon,-};Y_{1,\epsilon,+}) + I(Y_{1,\epsilon,-};Y_{1,\epsilon,+})$$

implying that

$$I(V,Y_{1,\epsilon,-};Y_{\mathrm{r},+}) + I(V,Y_{\mathrm{r},+};Y_{\mathrm{r},-}) - I(V,Y_{\mathrm{r},+};Y_{1,\epsilon,-}) - I(V,Y_{1,\epsilon,-};Y_{1,\epsilon,+}) \le 2C_0.$$

We identify $V_+ = (V, Y_{1,\epsilon,-})$ and $V_- = (V, Y_{\mathrm{r},+})$. Now by taking a uniform binary Q independent, and taking each of $+$ distribution and $-$ distribution with probability $\frac{1}{2}$, we get a new maximizer, which satisfies the constraint and yields a value at least as large as the original maximizer. For the maximum value to be equal we must have certain terms to be zero, in particular the blue term in (87) must be zero, *i.e.,*

$$I(Y_{1,\epsilon,-};\delta Y_{1,\epsilon,+} + Z_{3,+}|V,\delta Y_{1,\epsilon,-} + Z_{3,-}) = 0. \tag{88}$$

From (88) and $I(\delta Y_{1,\epsilon,-} + Z_{3,-};\delta Y_{1,\epsilon,+} + Z_{3,+}|V,Y_{1,\epsilon,-}) = 0$, we obtain a double-Markovity condition as defined in Lemma 3. We obtain that conditioned on V, we have $Y_{1,\epsilon,-}, Y_{1,\epsilon,+}$ are independent. This in particular implies that X_+, Y_{r_+} and X_-, Y_{r_-} are independent conditioned on V, yielding the desired Gaussianity of (X, Y_r) given V. Next we claim that taking X to be Gaussian minimizes $I(V;Y_\mathrm{r}) - I(V;Y_1)$. To see this, observe the constraint $h(Y_\mathrm{r}) - h(Y_{1,\epsilon}) - h(Y_\mathrm{r}|V) + h(Y_{1,\epsilon}|V) = I(V;Y_\mathrm{r}) - I(V;Y_{1,\epsilon}) \le C_0$. Note that, we have established that conditioned on V, (X, Y_r) are Gaussians. We are given that Y_r is Gaussian. Therefore taking X to be Gaussian maximizes the term $h(Y_{1,\epsilon})$ and thus keeps the constraint satisfied.

Now let us consider the argument for Corollary 2. Using the definition of $Y_{1,\epsilon}$ and Z_3 as before (see (84)), consider a maximizer of

$$R \le I(X;Y_{1,\epsilon}|V,W,T) - I(X;Y_\mathrm{r}|V,W,T) - \delta I(Y_{1,\epsilon};\delta Y_{1,\epsilon} + Z_3|V,W,T) - \delta I(Y_\mathrm{r};\delta Y_\mathrm{r} + Z_3|W,T) \tag{89}$$

over $p(t,x)p(y_{1,\epsilon}|x,y_\mathrm{r})p(y_\mathrm{r})p(w|t,y_\mathrm{r})p(v|t,x,y_\mathrm{r},w)p(z_3)$ subject to

$$I(V,W;Y_\mathrm{r}|T) - I(V,W;Y_{1,\epsilon}|T) \le I(W;Y_\mathrm{r}|T) \le C_0.$$

As before, take two copies of the maximizer and perform two orthogonal rotations to obtain the $+$ and $-$ random variables. Set $V = (V_1, V_2), W = (W_1, W_2), T = (T_1, T_2)$. We have

$$\begin{aligned} &I(Y_{\mathrm{r},+},Y_{\mathrm{r},-};\delta Y_{\mathrm{r},+} + Z_{3,+},\delta Y_{\mathrm{r},-} + Z_{3,-}|W,T) \\ &\quad = I(Y_{\mathrm{r},+};\delta Y_{\mathrm{r},+} + Z_{3,+}|W,T) + I(Y_{\mathrm{r},-};\delta Y_{\mathrm{r},-} + Z_{3,-}|W,T,\delta Y_{\mathrm{r},+} + Z_{3,+}) \\ &\quad = I(Y_{\mathrm{r},+};\delta Y_{\mathrm{r},+} + Z_{3,+}|W,T) + I(Y_{\mathrm{r},-};\delta Y_{\mathrm{r},-} + Z_{3,-}|W,T,Y_{\mathrm{r},+}) \\ &\qquad + I(Y_{\mathrm{r},+};\delta Y_{\mathrm{r},-} + Z_{3,-}|\delta Y_{\mathrm{r},+} + Z_{3,+},W,T) \end{aligned} \tag{90}$$

The exact manipulation for the first constraint above that we did in (87) goes through with (V,W,T) replacing V, and further using (90) we obtain the following

$$\begin{aligned} &I(X_+,X_-;Y_{1,\epsilon,+},Y_{1,\epsilon,-}|V,W,T) - I(X_+,X_-;Y_{\mathrm{r},+},Y_{\mathrm{r},-}|V,W,T) \\ &\quad - \delta I(Y_{1,\epsilon,+},Y_{1,\epsilon,-};\delta Y_{1,\epsilon,+} + Z_{3,+},\delta Y_{1,\epsilon,-} + Z_{3,-}|V,W,T) - \delta I(Y_{\mathrm{r},+},Y_{\mathrm{r},-};\delta Y_{\mathrm{r},+} + Z_{3,+},\delta Y_{\mathrm{r},-} + Z_{3,-}|W,T) \\ &\quad = I(X_+;Y_{1,\epsilon,+}|V,W,T,Y_{1,\epsilon,-}) - I(X_+;Y_{\mathrm{r},+}|V,W,T,Y_{1,\epsilon,-}) - \delta I(Y_{1,\epsilon,+};\delta Y_{1,\epsilon,+} + Z_{3,+}|V,W,T,Y_{1,\epsilon,-}) \\ &\qquad - \delta I(Y_{\mathrm{r},+};\delta Y_{\mathrm{r},+} + Z_{3,+}|W,T) \\ &\qquad + I(X_-;Y_{1,\epsilon,-}|V,W,T,Y_{\mathrm{r},+}) - I(X_-;Y_{\mathrm{r},-}|V,W,T,Y_{\mathrm{r},+}) - \delta I(Y_{1,\epsilon,-};\delta Y_{1,\epsilon,-} + Z_{3,-}|V,W,T,Y_{\mathrm{r},+}) \\ &\qquad - \delta I(Y_{\mathrm{r},-};\delta Y_{\mathrm{r},-} + Z_{3,-}|W,T,Y_{\mathrm{r},+}) \\ &\qquad - I(X_-;Y_{\mathrm{r},+}|V,W,T,Y_{1,\epsilon,-},Y_{1,\epsilon,+},X_+) - I(X_+;Y_{\mathrm{r},-}|V,W,T,Y_{\mathrm{r},\epsilon,+},Y_{1,\epsilon,-},X_-) \\ &\qquad - \delta I(Y_{1,\epsilon,-};\delta Y_{1,\epsilon,+} + Z_{3,+}|V,W,T,\delta Y_{1,\epsilon,-} + Z_{3,-}) - \delta I(Y_{\mathrm{r},+};\delta Y_{1,\epsilon,-} + Z_{3,-}|V,W,T) \\ &\qquad - \delta I(Y_{\mathrm{r},+};\delta Y_{\mathrm{r},-} + Z_{3,-}|\delta Y_{\mathrm{r},+} + Z_{3,+},W,T) \end{aligned} \tag{91}$$

Let us identify $V_+ = (V, Y_{1,\epsilon,-})$, $V_- = (V, Y_{\mathrm{r},+})$, $W_+ = W$, $W_- = (W, Y_{\mathrm{r},+})$, and $T_+ = T_- = T$. To verify the constraint, we need to show that

$$I(V_+,W_+;Y_{\mathrm{r},+}|T_+) - I(V_+,W_+;Y_{1,\epsilon,+}|T_+) + I(V_-,W_-;Y_{\mathrm{r},-}|T_-) - I(V_-,W_-;Y_{1,\epsilon,-}|T_-)$$

$$\leq I(W_+;Y_{\mathrm{r},+}|T_+)+I(W_-;Y_{\mathrm{r},-}|T_-).$$

Observe that

$$\begin{aligned}
&I(V_+,W_+;Y_{\mathrm{r},+}|T_+)-I(V_+,W_+;Y_{1,\epsilon,+}|T_+)+I(V_-,W_-;Y_{\mathrm{r},-}|T_-)-I(V_-,W_-;Y_{1,\epsilon,-}|T_-)\\
&=I(V,W,Y_{1,\epsilon,-};Y_{\mathrm{r},+}|T)-I(V,W,Y_{1,\epsilon,-};Y_{1,\epsilon,+}|T)+I(V,W,Y_{\mathrm{r},+};Y_{\mathrm{r},-}|T)-I(V,W,Y_{\mathrm{r},+};Y_{1,\epsilon,-}|T)\\
&=I(V,W;Y_{\mathrm{r},+},Y_{\mathrm{r},-}|T)-I(V,W;Y_{1,\epsilon,+},Y_{1,\epsilon,-}|T)-I(Y_{1,\epsilon,-};Y_{1,\epsilon,+}|T)\\
&\overset{(a)}{\leq} I(W;Y_{\mathrm{r},+},Y_{\mathrm{r},-}|T)-I(Y_{1,\epsilon,-};Y_{1,\epsilon,+}|T)\\
&\leq I(W;Y_{\mathrm{r},+}|T)+I(W,Y_{\mathrm{r},+};Y_{\mathrm{r},-}|T)\\
&=I(W_+;Y_{\mathrm{r},+}|T_+)+I(W_-;Y_{\mathrm{r},-}|T_-),
\end{aligned}$$

where (a) follows from the fact that random variables (W_1,T_1) and (W_2,T_2) of the two copies of the maximizer satisfy the constraints

$$\begin{aligned}
I(V_1,W_1;Y_{\mathrm{r},1}|T_1)-I(V_1,W_1;Y_{1,\epsilon,1}|T_1)\leq I(W_1;Y_{\mathrm{r},1}|T_1)\\
I(V_2,W_2;Y_{\mathrm{r},2}|T_2)-I(V_2,W_2;Y_{1,\epsilon,2}|T_2)\leq I(W_2;Y_{\mathrm{r},2}|T_2).
\end{aligned}$$

Thus the first part of the constraint is satisfied for $+$ and $-$ variables. Finally observe that

$$\begin{aligned}
&I(W_+;Y_{\mathrm{r},+}|T_+)+I(W_-;Y_{\mathrm{r},-}|T_-)\\
&\quad=I(W;Y_{\mathrm{r},+}|T)+I(W,Y_{\mathrm{r},+};Y_{\mathrm{r},-}|T)\\
&\quad=I(W;Y_{\mathrm{r},+},Y_{\mathrm{r},-}|T)\leq 2C_0.
\end{aligned}$$

This shows that the constraint is satisfied. Now we denote $T=(T,Q)$ where Q is uniform binary and conditioned on $Q=0$ we use the $+$ distribution and conditioned on $Q=-$ we use the - distribution. Thus, we get a new maximizer, which satisfies the constraint and yields a value at least as large as the original maximizer.

For the maximizing distribution (89) is tight and we must have the following

$$\begin{aligned}
I(Y_{\mathrm{r},+};\delta Y_{\mathrm{r},-}+Z_{3,-}|\delta Y_{\mathrm{r},+}+Z_{3,+},W,T)=0\\
I(Y_{1,\epsilon,-};\delta Y_{1,\epsilon,+}+Z_{3,+}|V,W,T,\delta Y_{1,\epsilon,-}+Z_{3,-})=0.
\end{aligned}$$

Now using double-Markovity, we conclude that conditioned on V,W,T, the random variables X,Y_{r} are jointly Gaussian for the maximizer and the covariance matrix of X,Y_{r} does not depend on the values of V,W,T. Moreover, and conditioned on W,T, the random variable Y_{r} is Gaussian for the maximizer. As observed earlier taking X to be Gaussian conditioned on T maximizes $h(Y_{1,\epsilon}|T)$ and hence keeps the C_0 constraint satisfied. This implies that we have a Gaussian maximizer. $\square$