# Strengthened Cutset Upper Bounds on the Capacity of the Relay Channel and Applications

Abbas El Gamal, Amin Gohari and Chandra Nair

**Abstract**

We establish new upper bounds on the capacity of the relay channel which are tighter than all previous bounds. The upper bounds use traditional weak converse techniques involving mutual information inequalities and identification of auxiliary random variables via past and future channel random variable sequences. For the relay channel without self-interference, we show that the new bound is strictly tighter than all previous bounds for the Gaussian relay channel for every set of non-zero channel gains. When specialized to the class of relay channels 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 extends it to the non-symmetric case, 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.

## 1. Introduction

The relay channel, first introduced by van-der Meulen in 1971 [VDM71], is a canonical model of multi-hop communication networks in which a sender $X$ wishes to communicate to a receiver $Y$ with the help of a relay $(X_{\rm r}, Y_{\rm r})$ over a memoryless channel of the form $p(y, y_{\rm r}|x, x_{\rm r})$. The capacity of this channel, which is the highest achievable rate from $X$ to $Y$, is not known in general. In [CEG79], lower bounds on the capacity, later termed decode-forward, partial decode-forward, and compress-forward, and the cutset upper bound were established. These bounds were shown to coincide for several special classes of channels, including degraded [CEG79], semi-deterministic [EGA82], and 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 [EK12]. In a series of recent papers [WOX17], [WBO19], [LO19], motivated by Cover's problem concerning a relay channel with orthogonal receiver components [Cov87], new highly specialized upper bounds, which are also tighter than the cutset bound, are developed. While the bounds in [CEG79], [TU08], [ARY09a] use standard weak converse techniques involving basic mutual information bounds 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 has been used in several converse proofs, e.g., see [CK78], [EG79], and the new idea of auxiliary receiver. The upper bounds we present in this paper are natural extensions of the upper bound in [GN20]. Our bounds and their applications do not include an auxiliary receiver because we are not able to find an example with an auxiliary receiver that can strictly improve over the bound without it. We will focus our attention on the class of relay channels *without self-interference* $p(y_{\rm r}|x)p(y|x, x_{\rm r}, y_{\rm r})$ because they include and generalize several interesting relay channel settings that have been receiving significant attention in recent years.

Although the techniques used to establish the upper bounds in this paper have been employed in many previous works, the contributions of this paper are in the judicious manner in which these techniques are applied to obtain the tightest known upper bounds on the capacity of the relay channel and the rather nontrivial evaluations of these complex bounds to obtain tighter and more general bounds for several classes of relay channels.

**Organization of the paper and summary of the results**. In the following section we formally 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 upper bounds and applications with references to the corresponding sections.

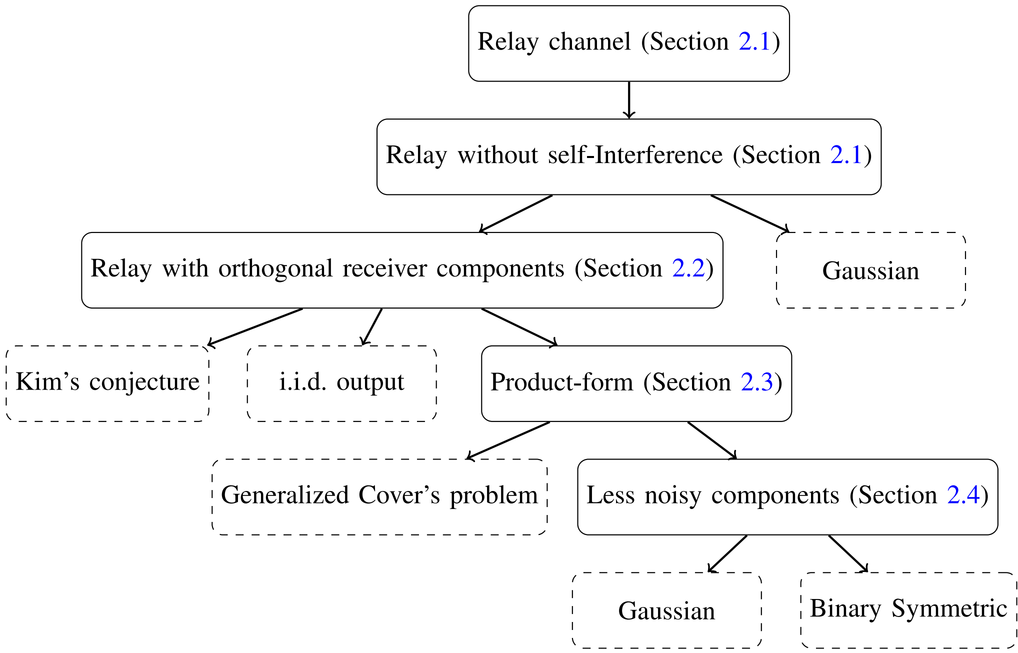


Fig. 1. Classes of relay channels for which upper bounds are established. An arrow from box A to box B indicates that class A includes class B. Dashed boxes indicate applications for which we compute the bounds.

The following summarizes the results in each section.

**Section 2.1**. Theorem 1 states our most general upper bound. Although the statement and subsequent bounds and applications of this upper 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 version of Theorem 1 which is easier to compute. This corollary is used to show the suboptimality of the cut-set upper bound and the improved bound in [GN20] for the Gaussian relay channel in Theorem 2 (also 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 called primitive). This proposition is used to prove a conjecture by Kim [Kim07, Question 2] in Theorem 3. Proposition 2 provides a new upper bound for 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 can improve upon the upper bounds in [TU08].

**Section 2.3**. Theorem 5 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].

**Section 2.4**. Proposition 4 specializes the bound in Proposition 1 to less-noisy product-form relay channels, which is used to: (i) derive the upper bound for the Gaussian case in Proposition 5 (also see Figure 5) which recovers as a special case an equivalent bound for the symmetric case in [WBO19] (see Theorem 7); and (ii) establish a tighter upper bound for a class of symmetric binary relay channels than the bounds in [WOX17], [BWÖ17] (see Theorem 8 and Figure 6).

## 2. Definitions and statement of the results

We adopt most of our notation from [EK12]. 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 $Y$.

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 $y_{\mathrm{r}}^{i-1}$ for each time $i \in [1:n]$, and a decoder that assigns an estimate $\hat{M}$ or an error message $\varepsilon$ 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 [EK12].

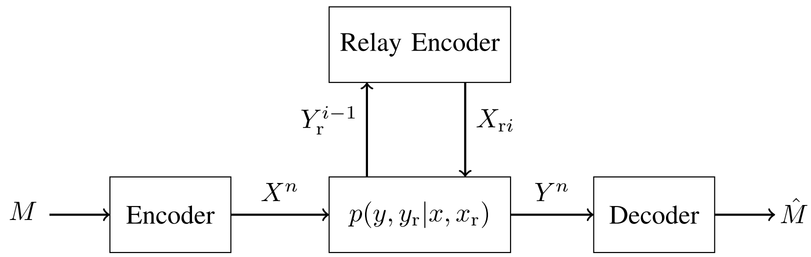


Fig. 2. Transmission of a message $M$ over a memoryless relay channel with $n$ channel uses.

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

We first consider the following class of relay channels.

**Definition 1.** 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 now present an upper bound for this class of relay channels.

**Theorem 1.** *Any achievable rate $R$ for a general discrete memoryless relay channel $p(y_{\mathrm{r}}|x)p(y|x, x_{\mathrm{r}}, y_{\mathrm{r}})$ must satisfy the following inequalities*

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

*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*

$$I(V, X_{\mathrm{r}}; Y_{\mathrm{r}}) - I(V, X_{\mathrm{r}}; Y) = I(U; Y_{\mathrm{r}}) - I(U; Y). \tag{6}$$

*Further 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.

*Remark* 1. As the proof of Theorem 1 indicates, the without self-interference assumption is used only to establish the Markov chain $U \to X, X_{\mathrm{r}} \to Y, Y_{\mathrm{r}}$. Hence, by relaxing 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]

$$C \le \max_{p(x, x_{\mathrm{r}})} \{I(X_{\mathrm{r}}, X; Y), I(X; Y, Y_{\mathrm{r}}|X_{\mathrm{r}})\}. \tag{7}$$

*Remark* 3. From (4) and (5), we deduce that

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

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 [EK12].

*Remark* 4. 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 capacity achieves 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 capacity is achievable by partial-decode-forward, since from (5) 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. 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 $Y_\mathrm{r} = (Y_{\mathrm{r}a}, Y_{\mathrm{r}b})$, where $p(y, y_\mathrm{r}|x, x_\mathrm{r}) = p(y_{\mathrm{r}a}|x_a)p(y|x_a, x_\mathrm{r})p(y_{\mathrm{r}b}|x_b)$, and $p(y_{\mathrm{r}b}|x_b)$ is a noiseless link of capacity $C_1$ while $p(y_{\mathrm{r}a}|x_a)$ and $p(y|x_a, x_\mathrm{r})$ 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 $p(y_{\mathrm{r}a}|x_a)p(y|x_a, x_\mathrm{r})$. Let $\mathcal{C}(C_1)$ be the capacity of this relay channel in terms of $C_1$ for a fixed channel $p(y_{\mathrm{r}a}|x_a)p(y|x_a, x_\mathrm{r})$. For $C_1 = \infty$, we have $\mathcal{C}(\infty) = \max_{p(x)} I(X, X_\mathrm{r}; Y)$. One can then ask what is the critical value $C_1^*$ such that $C_1^* = \inf\{C_1 : \mathcal{C}(C_1) = \mathcal{C}(\infty) = \max_{p(x)} I(X, X_\mathrm{r}; Y)\}$? It is immediate from the above discussion that $C_1^*$ can be characterized as the minimum value $C_1$ such that $R_{\mathsf{PDF}}(C_1) = \mathcal{C}(\infty) = \max_{p(x)} I(X, X_\mathrm{r}; Y)$ where $R_{\mathsf{PDF}}(C_1)$ 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. It is obtained by removing the constraint in (1), and relaxing the condition in (6).

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

$$\begin{aligned} R &\leq I(X; Y, Y_\mathrm{r}|X_\mathrm{r}) - I(V; Y|X_\mathrm{r}, Y_\mathrm{r}) - I(X; Y_\mathrm{r}|V, X_\mathrm{r}, Y) && (8)\\ &= I(X; Y_\mathrm{r}|X_\mathrm{r}) - I(X; Y_\mathrm{r}|V, X_\mathrm{r}) + I(X; Y|V, X_\mathrm{r}), && (9)\\ R &\leq I(X, X_\mathrm{r}; Y) - I(V; Y_\mathrm{r}|X_\mathrm{r}, X, Y) && (10) \end{aligned}$$

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

$$I(V, X_\mathrm{r}; Y_\mathrm{r}) - I(V, X_\mathrm{r}; Y) \leq \max_{p(u|x, x_\mathrm{r})} [I(U; Y_\mathrm{r}) - I(U; Y)].$$

*Further it suffices to consider $|\mathcal{V}| \leq |\mathcal{X}||\mathcal{X}_\mathrm{r}||\mathcal{Y}_\mathrm{r}| + 2$.*

*Remark* 5. The above bound is a strengthening of the bound given in [GN20, Theorem 1] for the choice of $J = Y_\mathrm{r}$. To see this, observe that

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

**Gaussian relay channel.** The Gaussian relay channel is defined by

$$\begin{aligned} Y_\mathrm{r} &= g_{12}X + Z_\mathrm{r},\\ Y &= g_{13}X + g_{23}X_\mathrm{r} + Z, \end{aligned} \qquad (11)$$

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

*Remark* 6. Note that as defined, the Gaussian relay channel belongs to the class of relay channels without self-interference.

Now, let $S_{12} = g_{12}^2 P$, $S_{13} = g_{13}^2 P$ and $S_{23} = g_{23}^2 P$. For this channel, the cut-set bound reduces to the following [EK12, Eq. 16.4]

$$R \leq \max_{0 \leq \rho \leq 1} \min \left\{ \mathsf{C}(S_{13} + S_{23} + 2\rho\sqrt{S_{13}S_{23}}), \mathsf{C}((1 - \rho^2)(S_{13} + S_{12})) \right\}$$

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

**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:*

$$\begin{aligned} R \leq &\frac{1}{2}\log\left((1-\rho^2)S_{21}+1\right) - \frac{1}{2}\log\left(\beta + S_{21}(1-\rho^2)\alpha + 2\sigma\sqrt{S_{21}(1-\rho^2)\alpha\beta}\right) + \\ &+ \frac{1}{2}\log\left(\beta(1-\sigma^2)\right) + \frac{1}{2}\log\left((1-\rho^2)\alpha S_{31}+1\right), \end{aligned} \tag{13}$$

$$R \leq \frac{1}{2}\log\left(1+S_{31}+S_{32}+2\rho\sqrt{S_{31}S_{32}}\right) + \frac{1}{2}\log(\beta(1-\sigma^2)) \tag{14}$$

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

$$\sigma = \frac{(1-\rho^2)\alpha S_{31}+1}{2T\sqrt{S_{21}(1-\rho^2)\alpha\beta}} - \frac{(1-\rho^2)\alpha S_{21}+\beta}{2\sqrt{S_{21}(1-\rho^2)\alpha\beta}},$$

*and*

$$T = \min\left[\frac{1+S_{31}+S_{32}+2\rho\sqrt{S_{31}S_{32}}}{(1-\rho^2)S_{21}+1}, \quad \lambda_{\max}\right],$$

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

$$2\rho\sqrt{S_{31}S_{32}} + S_{31} + S_{32} + 1 - \lambda\big(S_{32}S_{21}(1-\rho^2) + S_{31} + S_{32} + S_{21} + 2 + 2\rho\sqrt{S_{31}S_{32}}\big) + \lambda^2(S_{21}+1) = 0. \tag{15}$$

The proof of this theorem is given in Section 3.2.

Figure 3 compares the bound in Corollary 1 with the bound in [GN20, Proposition 1] for a scalar Gaussian relay channel.

### 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 2.** A relay channel is said to be *with orthogonal receiver components* (also referred to as *primitive*) (see Section 16.7.3 in [EK12]) 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 only on the capacity of the 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 inequalities*

$$R \leq I(X;Y_1,Y_{\mathrm{r}}) - I(U;Y_1|Y_{\mathrm{r}}), \tag{16}$$

$$R \leq I(X;Y_1,Y_{\mathrm{r}}) - I(V;Y_1|Y_{\mathrm{r}}) - I(X;Y_{\mathrm{r}}|V,Y_1) \tag{17}$$

$$= I(X;Y_1,V) - I(V;X|Y_{\mathrm{r}}), \tag{18}$$

$$R \leq I(X;Y_1) + C_0 - I(V;Y_{\mathrm{r}}|X,Y_1) \tag{19}$$

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

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

*Further it suffices to consider $|\mathcal{V}| \leq |\mathcal{X}||\mathcal{Y}_{\mathrm{r}}| + 2$ and $|\mathcal{U}| \leq |\mathcal{X}| + 1$.*

The proof of this proposition is given in Section 3.3.

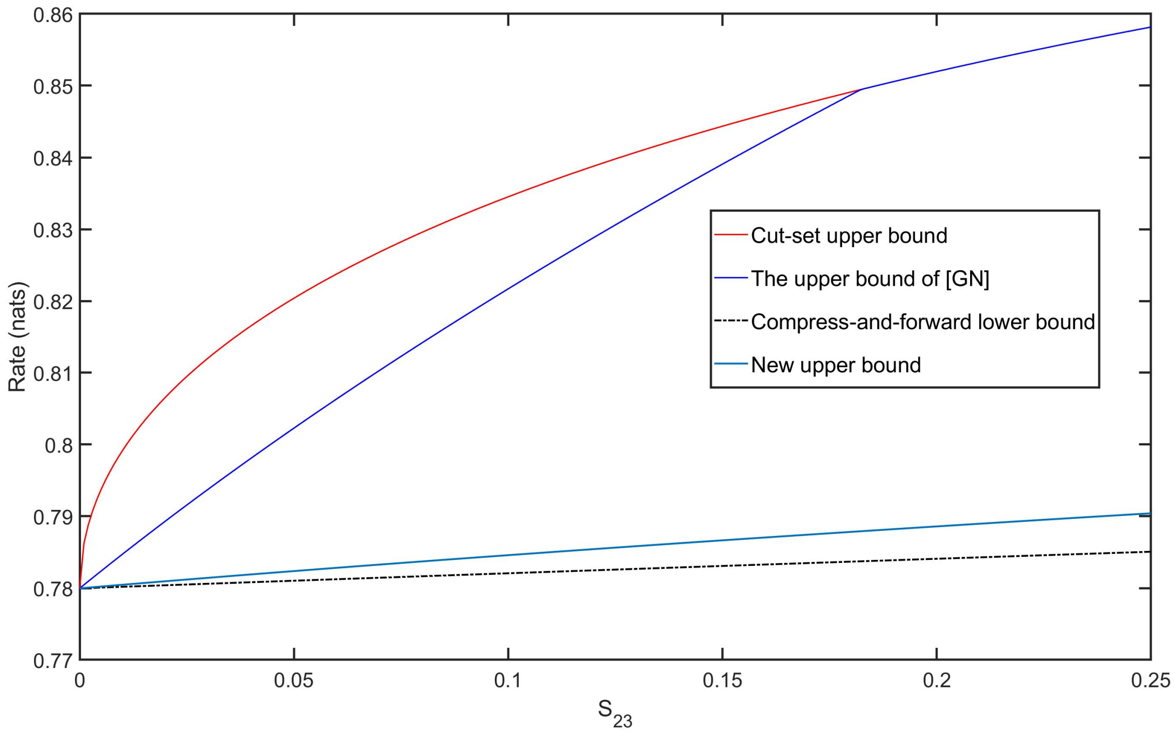


Fig. 3. Plots of the bounds for the Gaussian relay channel with $S_{13} = 3.7585$, $S_{12} = 1.2139$. The new upper bound is the evaluation of the bound in Corollary 1.

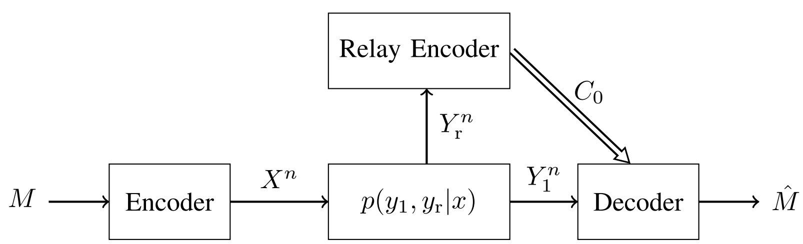


Fig. 4. Relay channel with orthogonal receiver components.

**Kim's conjecture.** We use Proposition 1 to prove a conjecture posed by Kim in [Kim07, Question 2] for a class of 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 $(Y_{\mathrm{r}}, Y_1)$ and the function $f$ (as defined in [OR95]).*

The proof of this theorem is given in Section 3.4.

**Relay channel with orthogonal receiver components and i.i.d. output**. Consider the following class of relay channels.

**Definition 3.** 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* 7. 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* 8. 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 $p(x)p(y_{\rm r})p(y_1|x,y_{\rm r})p(w|y_{\rm r})$ such that $I(W;Y_{\rm r}|Y_1) \le C_0$.

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

**Theorem 4** ( [TU08])**.** *Any achievable rate $R$ for the relay with orthogonal receiver components channel with i.i.d. output must satisfy the following*

$$R \le \min\{I(W,X;Y_1|Q), I(X;Y_1|Y_{\rm r},Q)\}$$

*for some $p(x,q)p(y_1|x,y_{\rm r})p(y_{\rm r})p(w|y_{\rm r},q)$ such that $I(W,Q;Y_{\rm r}) \le C_0$.*

*Remark* 9. The time-sharing random variable $Q$ was not included in the statement of the above theorem in [TU08], but we believe it is necessary to include it since the Markov chain $(W_Q,Q) \to Y_{{\rm r}Q} \to X_Q$ does not hold with the identification of auxiliary random variables in [TU08]. The upper bound in Theorem 4 with $Q$ still provides a converse for the example considered in [ARY09b].

Combining ideas from the proof of Theorem 4 with the proof techniques in this paper, we can obtain the following improved bound.

**Proposition 2.** *Any achievable rate $R$ satisfies*

$$\begin{aligned}R &\le I(X;Y_1|Q,W,V) - I(X;Y_{\rm r}|Q,W,V),\\ R &\le I(W,X;Y_1|Q)\end{aligned}$$

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

$$\begin{aligned}I(W;Y_{\rm r}|Q,Y_1) + I(V;Y_{\rm r}|Q,W) &\le C_0 + I(V;Y_1|Q,W),\\ I(W,Q;Y_{\rm r}) &\le C_0.\end{aligned}$$

The proof of this proposition is given in Section 3.5.

**Proposition 3.** *Consider a relay channel with i.i.d. output $p(y_{\rm r})p(y_1|x,y_{\rm r})$ such that the channel $p(y_1|x,Y_{\rm r}=y_{\rm r})$ is generic for every $y_{\rm r}$, and there is no random variable $K$ such that $H(K|X,Y_1)=H(K|X,Y_{\rm r})=0$ but $H(K|X)>0$. If the compress-forward lower bound with time-sharing does not match the bound in Theorem 4, then the bound in Proposition 2 strictly improves upon the bound in Theorem 4.*

The proof of this proposition is given in Section 3.6.

*Remark* 10. 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 4. 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 4. 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 $\epsilon$.

### *2.3. Product-form relay channels*

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

**Definition 4.** A relay channel with orthogonal receiver components is said to be *product-form* if $p(y_1,y_{\rm r}|x) = p(y_1|x)p(y_{\rm r}|x)$.

*Remark* 11. 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 4 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.

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

**Definition 5.** 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* 12. 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 6.** 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 5.** *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 3.7.

*Remark* 13. 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 5. 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.

### *2.4. Product-form relay channels with less noisy components*

Consider the following class of product-form relay channels.

**Definition 7.** A product-form relay channel described by $p(y_{\mathrm{r}}|x)p(y_1|x)$ and a relay-to-receiver link of capacity $C_0$ is said to have less-noisy components if it satisfies the conditions:

$$I(U;Y_{\mathrm{r}}) \leq I(U;Y_1) \text{ for every } p(u,x). \tag{20}$$

For this class of relay channels, we can specialize Proposition 1 to obtain the following bound.

**Proposition 4.** *Any achievable rate $R$ for product-form relay channel with less-noisy components must satisfy the following conditions*

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

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

$$I(V;Y_{\mathrm{r}}) - I(V;Y_1) \leq C_0. \tag{22}$$

*Further it suffices to consider $|\mathcal{V}| \leq |\mathcal{X}||\mathcal{Y}_{\mathrm{r}}| + 1$.*

The proof of this proposition is given in Section 3.8.

In the following we will also refer to the following class of relay channels.

**Definition 8.** A product-form relay channel is said to be *symmetric* if $\mathcal{Y}_1 = \mathcal{Y}_{\mathrm{r}}$, and $p_{Y_1|X}(y|x) = p_{Y_{\mathrm{r}}|X}(y|x)$ for all $x, y$.

It immediately follows that if a product-form relay channel is symmetric, then it is less noisy, and Proposition 4 provides an upper bound for the symmetric class.

**Gaussian Product-form relay channel with less-noisy components**. Consider a Gaussian relay channel with orthogonal receiver components described by

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

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 $S_{23} = 2^{2C_0} - 1$.

For $S_{13} \geq S_{12}$, the relay channel has less-noisy components and the upper bound in Proposition 4 reduces to the following.

**Proposition 5.** *Any achievable rate $R$ for the Gaussian product-form relay channel with $S_{13} \geq S_{12}$ must satisfy the following*

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

The proof of this proposition is given in Section 3.9.

*Remark* 14. Note that the compress-forward lower bound for this relay channel as given in [EK12, Eq. 16.17] implies that

$$C \geq \frac{1}{2} \log \left( 1 + S_{13} + \frac{S_{12}(S_{13}+1)S_{23}}{S_{12} + (S_{13}+1)(S_{23}+1)} \right). \tag{24}$$

Furthermore, the above lower bound can be improved by time-sharing at the transmitter [EK12, Sec. 16.8] or at the relay [WBO19, Footnote 2].

Figure 5 depicts the upper bound in Proposition 5 along with the cut-set upper bound and the compress-forward lower bound.

In [WBO19], the following upper bound on the capacity of the Gaussian product-form symmetric relay channel with $S_{12} = S_{13} = S$ is established.

**Theorem 6** ( [WBO19]). *For the symmetric Gaussian relay channel with orthogonal receiver components, any achievable rate $R$ satisfies*

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

*where*

$$h_\theta(\omega) = \frac{1}{2} \log \left( \frac{4 \sin^2 \frac{\omega}{2} (S + 1 - \sin^2 \frac{\omega}{2}) \sin^2 \theta}{(S+1)(\sin^2 \theta - \cos^2 \omega)} \right). \tag{25}$$

Although the techniques used to prove this theorem are completely different from those used in this paper, it turns out quite surprisingly that bound (25) coincides with the bound in Proposition 5 for the symmetric special case.

**Theorem 7.** *The bound in Theorem 6 simplifies to*

$$R \leq \frac{1}{2} \log \left( 1 + S + \frac{S(S+1)S_{23}}{(S+1)(S_{23}+1) - 1} \right),$$

*which coincides with the bound in Proposition 5 for $S_{12} = S_{13}$.*

The proof of this theorem is given in Section 3.10.

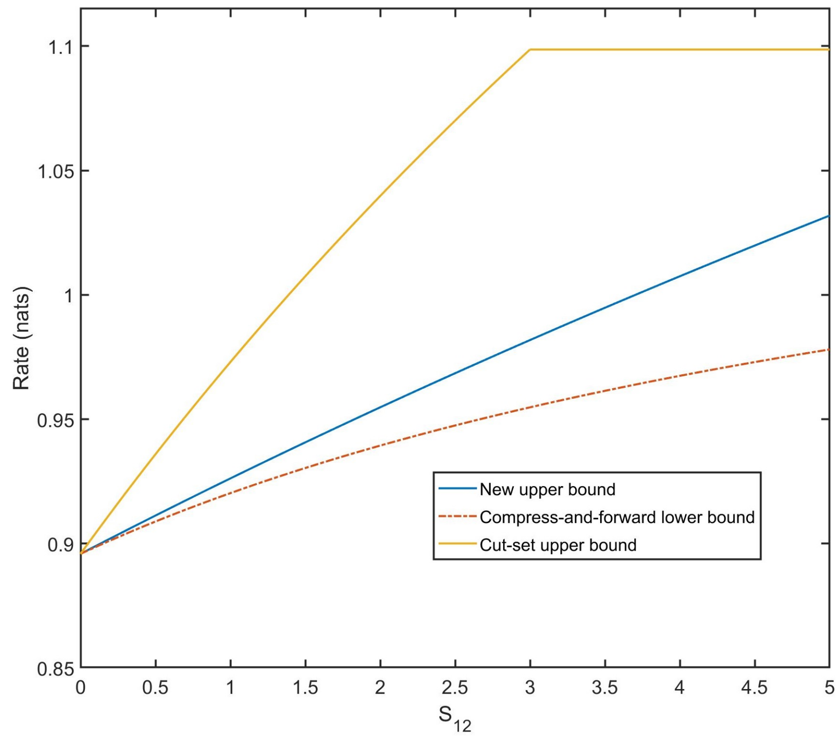


Fig. 5. Plots of the bounds for the Gaussian product-form relay channel with $S_{13} = 5$, $S_{23} = 0.5$.

**Symmetric binary relay channel with orthogonal receiver components**. Assume that the relay channel with orthogonal receiver components is 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 $x, y_1, y_\mathrm{r} \in \{0,1\}$. Moreover, 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 4 to this channel, we obtain the following bound.

**Theorem 8.** *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 $\rho$ must satisfy the following*

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

*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 3.11.

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

## 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*

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

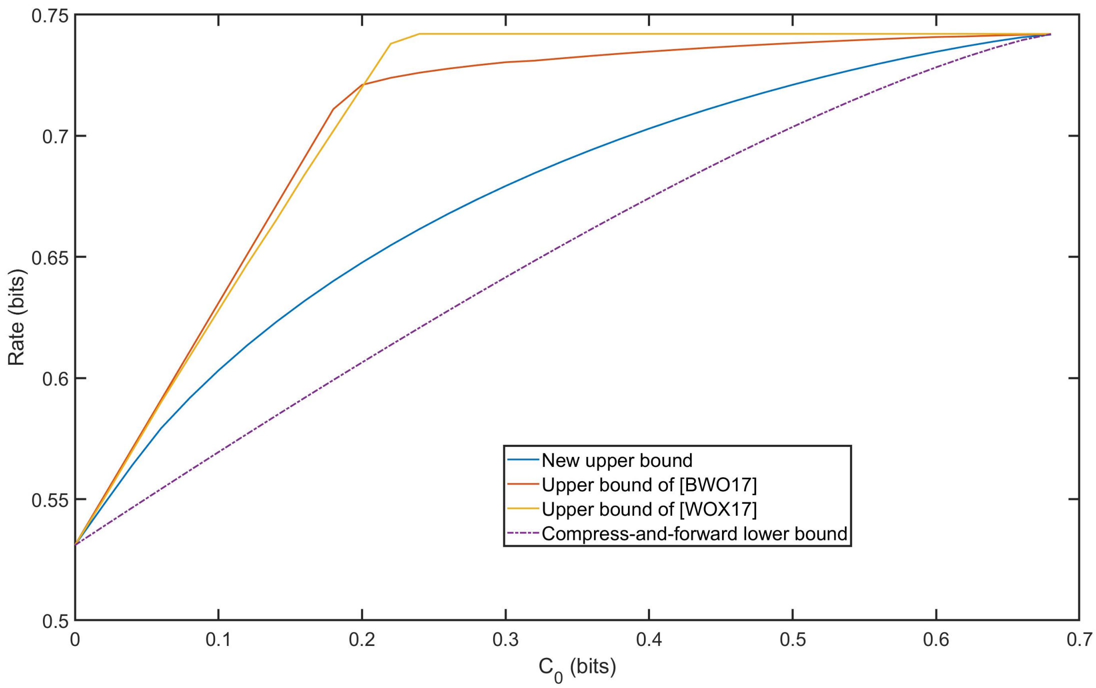


Fig. 6. Plots of the minimum of the two upper bounds give in [WOX17], the upper bound given in [BWÖ17], our new bound and the compress-forward lower bound for a symmetric binary relay channel with orthogonal receiver components with parameter $\rho = 0.1$.

Let

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

and $V = (Q, V_Q)$, $U = (Q, U_Q)$, $X = X_Q, X_{\mathrm{r}} = X_{\mathrm{r}Q}, Y = Y_Q, Y_{\mathrm{r}} = Y_{\mathrm{r}Q}$ for a time-sharing random variable $Q \overset{(d)}{=}$ Uniform$[1:n]$. 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. Lemma 1 implies the Markov chain $U \to (X, X_{\mathrm{r}}) \to (Y, Y_{\mathrm{r}})$ for relay channels of the form $p(y_{\mathrm{r}}|x)p(y|x, x_{\mathrm{r}}, y_{\mathrm{r}})$. 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

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

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

where in $(a)$ and $(b)$ we use the Markov chain $(Y_{\mathrm{r}}^{i-1}, Y_{\mathrm{r}i+1}^{n}, M, Y^{i-1}) \rightarrow (X_i, X_{\mathrm{r}i}, Y_{\mathrm{r}i}) \rightarrow Y_i$. 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 (3). Observe that by definition, $V_i$ implies that $X_{\mathrm{r}i}$. We have

$$
\begin{aligned}
\frac{1}{n} I(M; Y^n) &\leq \frac{1}{n} \left( I(M; Y_{\mathrm{r}}^n) + I(M; Y^n) - I(M; Y_{\mathrm{r}}^n) \right) \\
&\leq I(X; Y_{\mathrm{r}} | X_{\mathrm{r}}) + \frac{1}{n} \left( \sum_i I(M; Y_i | Y_{\mathrm{r}}^{i-1}, Y_{i+1}^n) - \sum_i I(M; Y_{\mathrm{r}i} | Y_{\mathrm{r}}^{i-1}, Y_{i+1}^n) \right) \\
&= I(X; Y_{\mathrm{r}} | X_{\mathrm{r}}) + \frac{1}{n} \left( \sum_i I(M, X_i; Y_i | Y_{\mathrm{r}}^{i-1}, X_{\mathrm{r}i}, Y_{i+1}^n) - \sum_i I(M, X_i; Y_{\mathrm{r}i} | Y_{\mathrm{r}}^{i-1}, X_{\mathrm{r}i}, Y_{i+1}^n) \right) \\
&= I(X; Y_{\mathrm{r}} | X_{\mathrm{r}}) + \frac{1}{n} \left( \sum_i I(X_i; Y_i | Y_{\mathrm{r}}^{i-1}, X_{\mathrm{r}i}, Y_{i+1}^n) + \sum_i I(M; Y_i | X_i, Y_{\mathrm{r}}^{i-1}, X_{\mathrm{r}i}, Y_{i+1}^n) \right. \\
&\qquad \left. - \sum_i I(X_i; Y_{\mathrm{r}i} | Y_{\mathrm{r}}^{i-1}, X_{\mathrm{r}i}, Y_{i+1}^n) - \sum_i I(M; Y_{\mathrm{r}i} | X_i, Y_{\mathrm{r}}^{i-1}, X_{\mathrm{r}i}, Y_{i+1}^n) \right) \\
&\leq I(X; Y_{\mathrm{r}} | X_{\mathrm{r}}) + \frac{1}{n} \left( \sum_i I(X_i; Y_i | Y_{\mathrm{r}}^{i-1}, X_{\mathrm{r}i}, Y_{i+1}^n) - \sum_i I(X_i; Y_{\mathrm{r}i} | Y_{\mathrm{r}}^{i-1}, X_{\mathrm{r}i}, Y_{i+1}^n) \right. \\
&\qquad \left. + \sum_i I(M; Y_i | X_i, Y_{\mathrm{r}}^{i-1}, X_{\mathrm{r}i}, Y_{i+1}^n, Y_{\mathrm{r}i}) \right) \\
&= I(X; Y_{\mathrm{r}} | X_{\mathrm{r}}) + I(X; Y | V, X_{\mathrm{r}}) - I(X; Y_{\mathrm{r}} | V, X_{\mathrm{r}}), \qquad (26)
\end{aligned}
$$

where in equation (26), 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 (5), note that

$$
\begin{aligned}
\frac{1}{n} I(M; Y^n) &\leq \frac{1}{n} I(X^n; Y^n) \\
&\leq \frac{1}{n} \sum_i I(X^n, Y_{i+1}^n; Y_i) \\
&= \frac{1}{n} \sum_i I(X_i; Y_i) + \frac{1}{n} \sum_i I(X^{n\setminus i}, Y_{i+1}^n; Y_i | X_i) \\
&\leq \frac{1}{n} \left( \sum_i I(X_i; Y_i) + I(V_i, X_{\mathrm{r}i}; Y_i | X_i) - I(V_i; Y_{\mathrm{r}i} | X_i, X_{\mathrm{r}i}) \right) \qquad (27)
\end{aligned}
$$

$$
\begin{aligned}
&\leq I(X;Y) + I(V, X_{\mathrm{r}}; Y|X) - I(V; Y_{\mathrm{r}}|X, X_{\mathrm{r}}) \\
&= I(X, X_{\mathrm{r}}; Y) + I(V; Y|X, X_{\mathrm{r}}) - I(V; Y_{\mathrm{r}}|X, X_{\mathrm{r}}) \\
&= I(X, X_{\mathrm{r}}; Y) - I(V; Y_{\mathrm{r}}|X_{\mathrm{r}}, X, Y),
\end{aligned} \tag{28}
$$

where (28) follows from the Markov chain relationship $V \to (X, X_{\mathrm{r}}, Y_{\mathrm{r}}) \to Y$ and (27) follows from

$$
\begin{aligned}
\sum_i I(Y_{\mathrm{r}}^{i-1}, Y_{i+1}^n; Y_{\mathrm{r}i}|X_i, X_{\mathrm{r}i}) &\overset{(a)}{=} \sum_i I(X^{n\setminus i}, Y_{\mathrm{r}}^{i-1}, Y_{i+1}^n; Y_{\mathrm{r}i}|X_i) - \sum_i I(X^{n\setminus i}; Y_{\mathrm{r}i}|X_i, Y_{\mathrm{r}}^{i-1}, Y_{i+1}^n) \\
&\quad - \sum_i I(X_{\mathrm{r}i}; Y_{\mathrm{r}i}|X_i) \qquad (29) \\
&\overset{(b)}{=} \sum_i I(Y_{i+1}^n; Y_{\mathrm{r}i}|X^n, Y_{\mathrm{r}}^{i-1}) - \sum_i I(X^{n\setminus i}; Y_{\mathrm{r}i}|X_i, Y_{\mathrm{r}}^{i-1}, Y_{i+1}^n) \\
&= \sum_i I(Y_{\mathrm{r}}^{i-1}; Y_i|X^n, Y_{i+1}^n) - \sum_i I(X^{n\setminus i}; Y_{\mathrm{r}i}|X_i, Y_{\mathrm{r}}^{i-1}, Y_{i+1}^n) \\
&\overset{(c)}{=} \sum_i I(Y_{\mathrm{r}}^{i-1}; Y_i|X^n, Y_{i+1}^n) - \sum_i I(X^{n\setminus i}; Y_{\mathrm{r}i}, Y_i|X_i, Y_{\mathrm{r}}^{i-1}, Y_{i+1}^n) \\
&= \sum_i I(X^{n\setminus i}, Y_{\mathrm{r}}^{i-1}; Y_i|X_i, Y_{i+1}^n) - \sum_i I(X^{n\setminus i}; Y_i|X_i, Y_{i+1}^n) \\
&\quad - \sum_i I(X^{n\setminus i}; Y_{\mathrm{r}i}, Y_i|X_i, Y_{\mathrm{r}}^{i-1}, Y_{i+1}^n) \\
&= \sum_i I(Y_{\mathrm{r}}^{i-1}, Y_{i+1}^n, X_{\mathrm{r}i}; Y_i|X_i) - \sum_i I(X^{n\setminus i}, Y_{i+1}^n; Y_i|X_i) \\
&\quad - \sum_i I(X^{n\setminus i}; Y_{\mathrm{r}i}|X_i, Y_{\mathrm{r}}^{i-1}, Y_{i+1}^n, Y_i) \\
&\leq \sum_i I(Y_{\mathrm{r}}^{i-1}, Y_{i+1}^n, X_{\mathrm{r}i}; Y_i|X_i) - \sum_i I(X^{n\setminus i}, Y_{i+1}^n; Y_i|X_i),
\end{aligned}
$$

where $(a)$ follows from the fact that $X_{\mathrm{r}i}$ is a function of $Y_{\mathrm{r}}^{i-1}$, $(b)$ follows because $(X^{n\setminus i}, Y_{\mathrm{r}}^{i-1}) \to (X_i, X_{\mathrm{r}i}) \to Y_{\mathrm{r}i}$ form a Markov chain, and $(c)$ follows because $(X^{n\setminus i}, Y_{i+1}^n) \to (X_i, Y_{\mathrm{r}}^{i-1}, Y_{\mathrm{r}i}) \to Y_i$ 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, $I(V; Y_{\mathrm{r}}) - I(V; Y) = I(U; Y_{\mathrm{r}}) - I(U; Y)$. This completes the proof.

**Lemma 1** (Lemma 2 in [GN20], set $J = Y_{\mathrm{r}}$)**.** *For the joint distribution on $(M, X^n, X_{\mathrm{r}}^n, Y^n, Y_{\mathrm{r}}^n)$ as induced by a code for a memoryless relay channel satisfying $p(y, y_{\mathrm{r}}|x, x_{\mathrm{r}}) = p(y_{\mathrm{r}}|x_{\mathrm{r}})p(y|x, x_{\mathrm{r}}, y_{\mathrm{r}})$, the following Markov chains hold for $i \in [1:n]$:*

*1) $Y_{\mathrm{r}i+1}^n \to (X_i, X_{\mathrm{r}i}) \to (Y_i, Y_{\mathrm{r}i})$*
*2) $(M, Y_{\mathrm{r}}^{i-1}, Y^{i-1}, Y_{\mathrm{r}i+1}^n, X_{\mathrm{r}i}) \to X_i \to Y_{\mathrm{r}i}$.*

### *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 [EK12]) that the maximum is attained via the unique jointly Gaussian distribution

$$(X, X_{\mathrm{r}})_{opt} \sim \begin{cases} \mathcal{N}\left(0, \begin{bmatrix} P & \rho^* P \\ \rho^* P & P \end{bmatrix}\right) & \text{if } S_{12} > S_{23}, \\ \mathcal{N}\left(0, \begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}\right) & \text{if } S_{12} \le S_{23}, \end{cases} \tag{30}$$

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

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

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 $(X, X_{\mathrm{r}})$ is distributed as in (30), 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 $Y_{\mathrm{r}}$ is a function of $(B, X_{\mathrm{r}})$ (see Lemma 5). Consequently $Y_{\mathrm{r}}$ is a function of $(V, X_{\mathrm{r}})$ (since $B = g(X_{\mathrm{r}}, V)$) implying that $I(V; Y_{\mathrm{r}}|X_{\mathrm{r}}, X, Y) = \infty$. Thus, the constraint $R \le I(X, X_{\mathrm{r}}; Y) - I(V; Y_{\mathrm{r}}|X_{\mathrm{r}}, X, Y)$ 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

$$K_{X,X_{\mathrm{r}},Z_1} = \begin{bmatrix} P & \rho P & 0 \\ \rho P & P & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{31}$$

and

$$K_{X,Z_1|VX_{\mathrm{r}}} = \begin{bmatrix} P(1-\rho^2)\alpha & \sigma\sqrt{P(1-\rho^2)\alpha\beta} \\ \sigma\sqrt{P(1-\rho^2)\alpha\beta} & \beta \end{bmatrix} \preceq \begin{bmatrix} P(1-\rho^2) & 0 \\ 0 & 1 \end{bmatrix}. \tag{32}$$

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

Now note that

$$\begin{aligned} \max_{P_{U|X,X_{\mathrm{r}}}} (I(U; Y_{\mathrm{r}}) - I(U; Y)) &= I(X, X_{\mathrm{r}}; Y_{\mathrm{r}}) - I(X, X_{\mathrm{r}}; Y) + \max_{P_{U|X,X_{\mathrm{r}}}} (I(X, X_{\mathrm{r}}; Y|U) - I(X, X_{\mathrm{r}}; Y_{\mathrm{r}}|U)) \\ &\overset{(a)}{=} I(X, X_{\mathrm{r}}; Y_{\mathrm{r}}) - I(X, X_{\mathrm{r}}; Y) + \frac{1}{2}\log\lambda_{\max} \\ &= \frac{1}{2}\log(S_{21} + 1) - \frac{1}{2}\log\left(1 + S_{31} + S_{32} + 2\rho\sqrt{S_{31}S_{32}}\right) + \frac{1}{2}\log\lambda_{\max}, \end{aligned}$$

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

$$2\rho\sqrt{S_{31}S_{32}} + S_{31} + S_{32} + 1 - \lambda\big(S_{32}S_{21}(1-\rho^2) + S_{31} + S_{32} + S_{21} + 2 + 2\rho\sqrt{S_{31}S_{32}}\big) + \lambda^2(S_{21}+1) = 0. \quad (33)$$

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

$$\begin{aligned} R &\leq \frac{1}{2}\log\left((1-\rho^2)S_{21}+1\right) - \frac{1}{2}\log\left(\beta + S_{21}(1-\rho^2)\alpha + 2\sigma\sqrt{S_{21}(1-\rho^2)\alpha\beta}\right) \\ &\quad + \frac{1}{2}\log\left(\beta(1-\sigma^2)\right) + \frac{1}{2}\log\left((1-\rho^2)\alpha S_{31}+1\right) \quad (34) \\ R &\leq \frac{1}{2}\log\left(1 + S_{31} + S_{32} + 2\rho\sqrt{S_{31}S_{32}}\right) + \frac{1}{2}\log(\beta(1-\sigma^2)) \quad (35) \end{aligned}$$

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

$$\frac{(1-\rho^2)\alpha S_{31}+1}{(1-\rho^2)\alpha S_{21} + \beta + 2\sigma\sqrt{S_{21}(1-\rho^2)\alpha\beta}} \leq \lambda_{\max}.$$

To complete the proof, we use an observation about the maximizing variables from Lemma 2 below to determine the value of $\sigma$. The constraint in (36) is a linear equality in $\sigma$ and yields

$$\sigma = \frac{(1-\rho^2)\alpha S_{31}+1}{2T\sqrt{S_{21}(1-\rho^2)\alpha\beta}} - \frac{(1-\rho^2)\alpha S_{21}+\beta}{2\sqrt{S_{21}(1-\rho^2)\alpha\beta}},$$

where

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

This completes the proof.

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

$$I(V, X_{\mathrm{r}}; Y_{\mathrm{r}}) - I(V, X_{\mathrm{r}}; Y) = \min\big\{I(X_{\mathrm{r}}; Y_{\mathrm{r}}), \quad \max_{P_{U|X,X_{\mathrm{r}}}}(I(U; Y_{\mathrm{r}}) - I(U;Y))\big\}. \quad (36)$$

*Proof.* Assume that the maximizer does not satisfy (36). Then, three cases are possible:

$$I(V, X_{\mathrm{r}}; Y_{\mathrm{r}}) - I(V, X_{\mathrm{r}}; Y) < I(X_{\mathrm{r}}; Y_{\mathrm{r}}) \leq \max_{P_{U|X,X_{\mathrm{r}}}}\left[I(U; Y_{\mathrm{r}}) - I(U;Y)\right], \quad (37)$$

or

$$I(X_{\mathrm{r}}; Y_{\mathrm{r}}) < I(V, X_{\mathrm{r}}; Y_{\mathrm{r}}) - I(V, X_{\mathrm{r}}; Y) \leq \max_{P_{U|X,X_{\mathrm{r}}}}\left[I(U; Y_{\mathrm{r}}) - I(U;Y)\right], \quad (38)$$

or

$$I(V, X_{\mathrm{r}}; Y_{\mathrm{r}}) - I(V, X_{\mathrm{r}}; Y) < \max_{P_{U|X,X_{\mathrm{r}}}}\left[I(U; Y_{\mathrm{r}}) - I(U;Y)\right] \leq I(X_{\mathrm{r}}; Y_{\mathrm{r}}) \quad (39)$$

Consider cases (37) and (39). Note that inequality (9) implies (10) if

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

Therefore, the right hand side of (9) is strictly less than (10), and (10) is redundant. Take a Bernoulli time-sharing random variable $Q \sim \mathrm{B}(\theta)$ and set $\tilde{V} = (V, Q)$ if $Q = 0$ and $\tilde{V} = (Y_{\mathrm{r}} + \zeta W, Q)$ 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 (8) for any $\theta > 0$. On the other hand, for any arbitrary $\theta > 0$, we have

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

Also for any $\zeta > 0$,

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

Therefore, one can find suitable $\theta, \zeta > 0$ such that for $\tilde{V}$, (10) is still redundant, (9) has a larger value and still the constraints are satisfied for $\tilde{V}$. This is a contradiction.

Finally, consider (38). In this case,

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

which implies that the right hand side of equation (9) is strictly larger than that of (10), and hence (9) is redundant. Take a time-sharing random variable $Q \sim \mathrm{B}(\theta)$, and set $\tilde{V} = (V, Q)$ if $Q = 0$, and $\tilde{V} = Q$ if $Q = 1$. This time sharing would decrease the expression $I(V, X_{\mathrm{r}}; Y_{\mathrm{r}}) - I(V, X_{\mathrm{r}}; Y)$, and would increase (10). This is a contradiction. $\square$

*3.3. Proof of Proposition 1*

We first show that if a rate $R$ satisfies the inequalities given in Theorem 1 for a pair auxiliary variables $(U, V)$, it also satisfies the inequalities given in Proposition 1 for the choice of auxiliary variables $(U, \tilde{V})$ where $\tilde{V} = (V, X_{\mathrm{r}})$. For a relay channel with orthogonal receiver components, Theorem 1 implies that

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

This corresponds to the constraints in Proposition 1 for $U$ and $\tilde{V} = (V, X_{\mathrm{r}})$. To verify the constraints of Proposition 1 for $U$ and $\tilde{V}$, note that the channel structure and the Markov chain relationships $V \to (X, X_{\mathrm{r}}, Y_{\mathrm{r}}) \to Y$ and $U \to (X, X_{\mathrm{r}}) \to (Y, Y_{\mathrm{r}})$ imply that $(V, X_{\mathrm{r}}) \to (X, Y_{\mathrm{r}}) \to Y_1$ and $U \to X \to (Y_1, Y_{\mathrm{r}})$.

Observe that

$$\begin{aligned}
I(U; Y_{\mathrm{r}}) - I(U; Y_1) - I(X_{\mathrm{r}}; Y_2|Y_1) &\leq I(U; Y_{\mathrm{r}}) - I(U; Y_1, Y_2) \\
&\overset{(a)}{=} I(V, X_{\mathrm{r}}; Y_{\mathrm{r}}) - I(V, X_{\mathrm{r}}; Y_1, Y_2) \\
&= I(V, X_{\mathrm{r}}; Y_{\mathrm{r}}) - I(V, X_{\mathrm{r}}; Y_1) - I(X_{\mathrm{r}}; Y_2|Y_1) \\
&= I(\tilde{V}; Y_{\mathrm{r}}) - I(\tilde{V}; Y_1) - I(X_{\mathrm{r}}; Y_2|Y_1),
\end{aligned}$$

where $(a)$ follows from Theorem 1. Thus $I(U; Y_{\mathrm{r}}) - I(U; Y_1) \leq I(\tilde{V}; Y_{\mathrm{r}}) - I(\tilde{V}; Y_1)$. Further,

$$\begin{aligned}
I(\tilde{V}; Y_{\mathrm{r}}) - I(\tilde{V}; Y_1) - C_0 &\leq I(\tilde{V}; Y_{\mathrm{r}}) - I(\tilde{V}; Y_1) - I(X_{\mathrm{r}}; Y_2|Y_1) \\
&= I(U; Y_{\mathrm{r}}) - I(U; Y_1, Y_2) \\
&\leq I(U; Y_{\mathrm{r}}) - I(U; Y_1),
\end{aligned}$$

which establishes the constraint.

Next, conversely assume that a rate $R$ satisfies the inequalities given in Proposition 1 for some pair of auxiliary variables $(U, V)$. This defines a joint distribution over $(U, V, X, Y_1, Y_{\mathrm{r}})$. Let $(U', X_{\mathrm{r}}, Y_2)$ be independent of $(U, V, X, Y_1, Y_{\mathrm{r}})$ 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)

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

This is feasible for some $U' \to X_{\mathrm{r}} \to Y_2$ since from the constraint on $U$ and $V$ in Proposition 1 we have

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

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

$$I(V, X_{\mathrm{r}}; Y_{\mathrm{r}}) - I(V, X_{\mathrm{r}}; Y) = I(V; Y_{\mathrm{r}}) - I(V; Y_1) - I(X_{\mathrm{r}}; Y_2)$$

$$= I(U; Y_{\mathrm{r}}) - I(U; Y_1) - I(U'; Y_2)$$
$$= I(\tilde{U}; Y_{\mathrm{r}}) - I(\tilde{U}; Y).$$

Moreover,

$$\begin{aligned} I(V, X_{\mathrm{r}}; Y_{\mathrm{r}}) - I(V, X_{\mathrm{r}}; Y) &= I(V; Y_{\mathrm{r}}) - I(V; Y_1) - C_0 \\ &\le I(U; Y_{\mathrm{r}}) - I(U; Y_1) \\ &\le I(U; Y_{\mathrm{r}}) \le I(X; Y_{\mathrm{r}}) = I(X, X_{\mathrm{r}}; Y_{\mathrm{r}}). \end{aligned}$$

The second constraint in Theorem 1 can be verified directly too. 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 (17), we deduce that $I(V; Y_1|Y_{\mathrm{r}}) = 0$. Since $X$ is a function of $(Y_1, Y_{\mathrm{r}})$, we deduce that $I(V; X|Y_{\mathrm{r}}) = 0$. Consequently, (18) and (19) imply that the achievable rate $R = \log|\mathcal{X}|$ must satisfy the condition

$$R \le \max_{p(x)p(v|y_{\mathrm{r}})} \min\big\{I(X; V, Y_1), I(X; Y_1) + C_0 - I(V; Y_{\mathrm{r}}|X, Y_1)\big\}.$$

This matches the rate achieved by compress-forward (see [EK12, Eq. 16.14]). Thus, compress-forward must achieve the rate $\log|\mathcal{X}|$ when $C_0 = C_0^*$. Using the characterization of the compress-forward lower bound in Proposition 3 of [Kim07], we obtain that $\log|\mathcal{X}| = I(X; \tilde{V}, Y_1)$ for some $p(\tilde{v}|y_{\mathrm{r}})$ such that $I(\tilde{V}; Y_{\mathrm{r}}|Y_1) \le C_0^*$. Therefore, $X$ is uniformly distributed and $H(X|\tilde{V}, Y_1) = 0$. 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 Proposition 2*

Let $V_i = (Y_{1i+1}^n, Y_{2i+1}^n, Y_{\mathrm{r}}^{i-1}), W_i = (Y_{2i+1}^n, Y_{\mathrm{r}}^{i-1})$ and $V = (Q, V_Q)$, $W = (Q, W_Q)$, $X = X_Q, X_{\mathrm{r}} = X_{\mathrm{r}Q}, Y_1 = Y_{1Q}, Y_2 = Y_{2Q}$ and $Y_{\mathrm{r}} = Y_{\mathrm{r}Q}$ for a time-sharing random variable $Q \sim$ Uniform$[1:n]$. Note that $(W_i, Y_{\mathrm{r}i})$ is independent of $X_i$ and that $H(W_i|V_i) = 0$ and from Theorem 1, we can deduce the Markov chain $(V, W, Q) \to (X, X_{\mathrm{r}}, Y_{\mathrm{r}}) \to Y_1$. Since $p(y_1|x, x_{\mathrm{r}}, y_{\mathrm{r}}) = p(y_1|x, y_{\mathrm{r}})$, we obtain the Markov chain $(V, W, Q) \to (X, Y_{\mathrm{r}}) \to Y_1$.

Using the proof of Theorem 1 with this identification of auxiliary $V$, we have

$$R \le I(X; Y_{\mathrm{r}}|X_{\mathrm{r}}) + \frac{1}{n}\sum_i \big[I(X_i; Y_{1i}, Y_{2i}|V_i, X_{\mathrm{r}i}) - I(X_i; Y_{\mathrm{r}i}|V_i, X_{\mathrm{r}i})\big] \tag{40}$$
$$= \frac{1}{n}\sum_i \big[I(X_i; Y_{1i}|V_i, X_{\mathrm{r}i}) - I(X_i; Y_{\mathrm{r}i}|V_i, X_{\mathrm{r}i})\big] \tag{41}$$
$$= I(X; Y_1|W, V, Q) - I(X; Y_{\mathrm{r}}|W, V, Q), \tag{42}$$

where the last step follows from the fact that $V_i$ implies $(X_{\mathrm{r}i}, W_i)$.

Next,

$$\begin{aligned} I(V; Y_{\mathrm{r}}|Q) - I(V; Y|Q) &= \sum_i I(Y_{\mathrm{r}}^{i-1}; Y_{\mathrm{r}i}) - \sum_i I(Y^{i-1}; Y_i) \\ &= -\sum_i I(Y^{i-1}; Y_i) \le 0. \end{aligned}$$

On the other hand,

$$I(V; Y_{\mathrm{r}}|Q) - I(V; Y|Q) \ge I(V; Y_{\mathrm{r}}|Q) - I(V; Y_1|Q) - C_0.$$

Therefore,

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

This inequality can be written as

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

Next, observe that

$$\begin{aligned}
nR &\le I(X^n; Y_1^n, Y_2^n)\\
&= I(X^n; Y_1^n | Y_2^n)\\
&= H(Y_1^n|Y_2^n) - H(Y_1^n|X^n, Y_2^n)\\
&\le \sum H(Y_{1i}) - \sum_i H(Y_{1i}|X^n, Y_2^n, Y_1^{i-1})\\
&\le \sum H(Y_{1i}) - \sum_i H(Y_{1i}|X^n, Y_2^n, Y_1^{i-1}, Y_{\mathrm{r}}^{i-1}, X_{\mathrm{r}}^{i-1})\\
&= \sum H(Y_{1i}) - \sum_i H(Y_{1i}|X_i, Y_2^n, Y_{\mathrm{r}}^{i-1})\\
&\overset{(a)}{=} \sum H(Y_{1i}) - \sum_i H(Y_{1i}|X_i, Y_{2i+1}^n, Y_{\mathrm{r}}^{i-1})\\
&= \sum I(Y_{1i}; X_i, W_i),
\end{aligned}$$

where $(a)$ follows since $H(X_{\mathrm{r}}^i|Y_{\mathrm{r}}^{i-1}) = 0$.

Finally,

$$\sum_i I(W_i; Y_{\mathrm{r}i}) \le \sum_i I(Y_2^n, Y_{\mathrm{r}}^{i-1}; Y_{\mathrm{r}i}) = I(Y_2^n; Y_{\mathrm{r}}^n) \le I(Y_2^n; X_{\mathrm{r}}^n) \le nC_0.$$

Thus, $I(W, Q; Y_{\mathrm{r}}) = I(W; Y_{\mathrm{r}}|Q) \le C_0$. This completes the proof.

### *3.6. Proof of Proposition 3*

From Theorem 4, we have

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

for some $p(x, q)p(w|q, y_{\mathrm{r}})$ satisfying $I(W, Q; Y_{\mathrm{r}}) \le C_0$.

Assume that rate $R$ is achieved by the bound in Proposition 2. Consider the following two cases:

*Case 1*: $R = I(X; Y_1|Y_{\mathrm{r}}, Q)$. Since

$$\begin{aligned}
I(X; Y_1|Q, W, V) - I(X; Y_{\mathrm{r}}|Q, W, V) &\le I(X; Y_1|Q, W, V, Y_{\mathrm{r}})\\
&\le I(X; Y_1|Y_{\mathrm{r}}, Q),
\end{aligned}$$

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

$$R \le I(X; Y_1|W, V, Q) - I(X; Y_{\mathrm{r}}|W, V, Q) = I(X; Y_1|W, V, Q) \tag{43}$$

The constraint

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

is equivalent to

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

Since $(V, W) \to (Y_{\mathrm{r}}, Q) \to X$ form a Markov chain, we obtain

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

From this equation and (43), we deduce that compress-forward with time-sharing is optimal (using the characterization in Proposition 3 of [Kim07]).

*Case 2*: $R = I(W, X; Y_1|Q) < I(X; Y_1|Y_{\mathrm{r}}, Q)$. In this case we must have $I(W, Q; Y_{\mathrm{r}}) = C_0$ since if $I(W, Q; Y_{\mathrm{r}}) < C_0$ in Theorem 4, time-sharing between $W$ and $Y_{\mathrm{r}}$ would strictly increase $I(W, X; Y_1|Q)$.

Then, from the assumption that the rate $R$ is achieved by the bound in Proposition 2 we obtain

$$I(X;Y_1|W,V,Q) - I(X;Y_\mathrm{r}|W,V,Q) \geq I(W,X;Y_1|Q) \tag{44}$$

and

$$I(W,Q;Y_\mathrm{r}) + I(V;Y_1|Q,W) \geq I(W;Y_\mathrm{r}|Q,Y_1) + I(V;Y_\mathrm{r}|Q,W). \tag{45}$$

Adding up the above two inequalities, we obtain

$$I(V;Y_1|W,X,Q) \geq I(V;Y_\mathrm{r}|X,Q,W),$$

which is equivalent to $-I(V;Y_\mathrm{r}|X,Q,W,Y_1) \geq 0$. Thus, we have $I(V;Y_\mathrm{r}|X,Q,W,Y_1) = 0$. Moreover, inequalities (44) and (45) must be equalities. We had $I(V;Y_\mathrm{r}|X,Q,W,Y_1) = 0$ and $I(V;Y_1|X,Q,W,Y_\mathrm{r}) = 0$. From the assumption of the proposition and Lemma 3, we obtain that $I(V;Y_\mathrm{r},Y_1|X,W,Q) = 0$. Therefore,

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

Hence, since (45) holds with equality, it follows that $I(W,Q;Y_\mathrm{r}) + I(V;Y_1|Q,W) = I(W;Y_\mathrm{r}|Q,Y_1)$. Equivalently, $I(W;Y_1|Q) + I(V;Y_1|Q,W) = I(W;Y_1|Q,Y_\mathrm{r}) = 0$. Thus, $I(W;Y_1|Q) = I(V;Y_1|Q,W) = 0$. Hence,

$$R \leq I(X;Y_1|W,Q)$$

and $I(W;Y_\mathrm{r}|Y_1,Q) \leq I(W;Y_\mathrm{r}|Q) \leq C_0$. 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.

### 3.7. Proof of Theorem 5

The result follows immediately from the definition that $R_0^* \geq C_0^*$. Therefore it suffices to show that $R_0^* \leq C_0^*$. 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 (17) 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 (18) and (19) 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 [EK12, Eq. 16.14]

$$R_{\mathsf{CF}}(C_0) = \max_{p(x)p(v|y_\mathrm{r})} \min\big\{I(X;V,Y_1), I(X;Y_1) + C_0 - I(V;Y_\mathrm{r}|X,Y_1)\big\}.$$

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 $R_0^* \leq C_0^*$. This completes the proof.

### 3.8. Proof of Proposition 4

Consider arbitrary auxiliaries $U$ and $V$ for the evaluation of the upper bound in Proposition 1. If

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

then since $I(U;Y_\mathrm{r}) \leq I(U;Y_1)$, the same rate $R$ is achievable by setting $U = \emptyset$ and keeping the same $V$. If

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

then utilizing $p(y_1,y_\mathrm{r}|x) = p(y_1|x)p(y_\mathrm{r}|x)$, we have

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

which is less than the direct transmission bound. Thus, setting both $U = \emptyset$ and $V = \emptyset$ yields a larger rate. This shows that without loss of generality, we can assume that $I(V;Y_\mathrm{r}) - I(V;Y_1) \geq 0$, and therefore, can set $U = \emptyset$.

Next, the constraint $R \le I(X;Y_1) + C_0 - I(V;Y_{\mathrm{r}}|X,Y_1)$ is redundant because

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

where the last step follows from $p(y_1,y_{\mathrm{r}}|x) = p(y_1|x)p(y_{\mathrm{r}}|x)$. This completes the reduction of the upper bound in Proposition 1 to the form in Proposition 4.

### *3.9. Proof of Proposition 5*

The sufficiency of considering Gaussian random variables for the evaluation of Proposition 4 in this context is established in Lemma 7 in the Appendix. Let the covariance of $X, Y_{\mathrm{r}}$ 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

$$(P-K_1)(P+N_{\mathrm{r}}-K_2) \ge (P-\rho\sqrt{K_1K_2})^2. \tag{46}$$

Then, the upper bound becomes the maximum of

$$\frac{1}{2}\log\left(\frac{P+N_{\mathrm{r}}}{N_{\mathrm{r}}}\right) + \frac{1}{2}\log\left(\frac{K_1+N_1}{N_1}\right) + \frac{1}{2}\log\left(1-\rho^2\right) \tag{47}$$

subject to

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

Lemma 8 solves the above optimization problem, showing that the optimal $K_1, K_2$ and $\rho$ are

$$K_1^* = P\left(1 - \frac{P(N_1+P)^2(2^{2C_0}-1)}{(P+N_{\mathrm{r}})(2^{2C_0}-1)((N_1+P)^2 - N_1^2 2^{-2C_0}) + (N_{\mathrm{r}}-N_1)P^2}\right) \tag{48}$$

$$K_2^* = \frac{(K_1^*+N_1)(P+N_{\mathrm{r}})}{(P+N_1)2^{2C_0}} \tag{49}$$

$$\rho^* = \frac{P - \sqrt{(P-K_1^*)(P+N_{\mathrm{r}}-K_2^*)}}{\sqrt{K_1^*K_2^*}}. \tag{50}$$

Substituting these values in (47) 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

$$\frac{1}{2}\log\left(1 + S_{13} + \frac{S_{12}(S_{13}+1)S_{23}}{(S_{13}+1)(S_{23}+1)-1}\right)$$

as desired. This completes the proof.

### *3.10. Proof of Theorem 7*

We begin by simplifying the bound in Theorem 6. We seek to calculate

$$\frac{1}{2}\log(1+S) + \sup_{\theta\in\left[\arcsin(\frac{1}{1+S_{23}}),\frac{\pi}{2}\right]} \min\left\{C_0 + \log\sin(\theta), \min_{\omega\in\left(\frac{\pi}{2}-\theta,\frac{\pi}{2}\right]} h_\theta(\omega)\right\}, \tag{51}$$

where

$$h_\theta(\omega) = \frac{1}{2}\log\left(\frac{4\sin^2\left(\frac{\omega}{2}\right)\left(S+1-\sin^2\left(\frac{\omega}{2}\right)\right)\sin^2(\theta)}{(S+1)(\sin^2(\theta)-\cos^2(\omega))}\right). \tag{52}$$

Observe that $C_0+\log(\sin(\theta))$ is increasing in $\theta\in(0,\frac{\pi}{2})$. We will show that $\min_{\omega\in\left(\frac{\pi}{2}-\theta,\frac{\pi}{2}\right]} h_\theta(\omega)$ is decreasing in $\theta$ and that the supremum in (51) is attained at a $\theta_*$ such that

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

First, consider the problem of computing $\min_{\omega\in\left(\frac{\pi}{2}-\theta,\frac{\pi}{2}\right]} h_\theta(\omega)$, i.e.

$$\min_{\omega\in(\frac{\pi}{2}-\theta,\frac{\pi}{2}]}\frac{1}{2}\log\left(\frac{4\sin^2\left(\frac{\omega}{2}\right)\left(S+1-\sin^2\left(\frac{\omega}{2}\right)\right)\sin^2(\theta)}{(S+1)(\sin^2(\theta)-\cos^2(\omega))}\right).$$

For $0<\theta_1<\theta_2<\frac{\pi}{2}$, observe that

$$\begin{aligned}\min_{\omega\in(\frac{\pi}{2}-\theta_2,\frac{\pi}{2}]}\frac{1}{2}\log\left(\frac{4\sin^2\left(\frac{\omega}{2}\right)\left(S+1-\sin^2\left(\frac{\omega}{2}\right)\right)\sin^2(\theta_2)}{(S+1)(\sin^2(\theta_2)-\cos^2(\omega))}\right)&\le\min_{\omega\in(\frac{\pi}{2}-\theta_1,\frac{\pi}{2}]}\frac{1}{2}\log\left(\frac{4\sin^2\left(\frac{\omega}{2}\right)\left(S+1-\sin^2\left(\frac{\omega}{2}\right)\right)\sin^2(\theta_2)}{(S+1)(\sin^2(\theta_2)-\cos^2(\omega))}\right)\\&\le\min_{\omega\in(\frac{\pi}{2}-\theta_1,\frac{\pi}{2}]}\frac{1}{2}\log\left(\frac{4\sin^2\left(\frac{\omega}{2}\right)\left(S+1-\sin^2\left(\frac{\omega}{2}\right)\right)\sin^2(\theta_1)}{(S+1)(\sin^2(\theta_1)-\cos^2(\omega))}\right).\end{aligned}$$

Therefore $\min_{\omega\in\left(\frac{\pi}{2}-\theta,\frac{\pi}{2}\right]} h_\theta(\omega)$ is decreasing in $\theta$. Now, let $\theta_*\in[0,\frac{\pi}{2}]$ be defined according to

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

The minimizing $\omega$ for the computation of $h_{\theta_*}(\omega)$ can be seen to be $\omega_*$ satisfying

$$\cos\omega_*=\frac{2S+\cos^2\theta_*-\cos\theta_*\cdot\sqrt{4S^2+4S+\cos^2\theta_*}}{2S}.$$

Plugging these choices one can verify that

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

Substituting the maximizer $\theta_*$ in (51) shows that the bound in Theorem 6 is equivalent to

$$R\le\frac{1}{2}\log\left(1+S+\frac{S(S+1)S_{23}}{(S+1)(S_{23}+1)-1}\right).$$

This completes the proof.

### *3.11. Proof of Theorem 8*

Using the symmetrization argument in [Nai13], without loss of generality, we can restrict $X$ to be uniformly distribution when evaluating Proposition 4. 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

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

Moreover, $X'$ is uniform.

From Proposition 4 for any $\lambda\ge0$, any achievable rate $R$ must satisfy

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

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 $1 - 2H_2(\rho) + C_0 + H(Y_\mathrm{r}|X)$. If $\lambda > 1$, the upper bound is greater than or equal to $1 - 2H_2(\rho) + C_0 + H(Y_\mathrm{r}|X)$ since $V$ equal to a constant is one possible choice for $V$.

Let $p(x, y_\mathrm{r}) = (p_{00}, p_{01}, p_{10}, p_{11})$ and $f : p(x, y_\mathrm{r}) \mapsto (1-\lambda)\,(H(Y_1) - H(Y_\mathrm{r})) + H(Y_\mathrm{r}|X)$. The maximum of $(1-\lambda)[H(Y_1|V) - H(Y_\mathrm{r}|V)] + H(Y_\mathrm{r}|X, V)$ 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

$$f(p_{00}, p_{01}, p_{10}, p_{11}) = f(p_{11}, p_{10}, p_{01}, p_{00}). \tag{53}$$

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 (53) 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

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

completing the proof of the theorem.

## 4. Conclusion and Final remarks

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

Our upper bounds use standard converse techniques but take their application beyond what has been previously done. One key insight that lead 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 (2) 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 (5) 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 4. Similarly, a maximizer for bound (21) in Proposition 4 is $V = Y_\mathrm{r}$ while constraint (22) prevents $V$ from becoming close to $Y_\mathrm{r}$. Observe that (22) is deduced from the constraint (6) 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 5 and Theorem 7. There are also instances in which properties of the optimal auxiliaries have been identified that makes the optimization problem amenable to numerical evaluation by reducing the search dimension as in Theorem 2 and Theorem 8.

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.

The question of whether the addition of an auxiliary receiver can help improve the bounds remains open. Observe that a bound with 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 inner most 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 outer-most 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.

[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.

[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.

[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.

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

[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.

[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.

[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* 15. While this is not explicitly stated in the references, the above statement holds true for random variables defined on Polish spaces under the Borel $\sigma$-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 5) 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 $Y_\epsilon$, 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 $R_\epsilon$ 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 $P^\epsilon$ achieves the supremum of $R_\epsilon$ (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 $P^\epsilon$ 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

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

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

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

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

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

A similar inequality for B can also be written. Further observe that the dagger variables satisfy the constraint as well. This implies that for $P^\epsilon$ 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 4 for the Gaussian product-form relay channel*

**Lemma 7.** *To evaluate Proposition 4 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 $Y_\mathrm{r}$ 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 $R_\epsilon$ satisfying the following:

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

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

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

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

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

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

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

Further observe that the dagger variables satisfy

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

as well. Since the dagger variables are also a feasible choice, for $P^\epsilon$ 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 4 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*

$$K_1^*=P\left(1-\frac{P(N_1+P)^2(2^{2C_0}-1)}{(P+N_\mathrm{r})(2^{2C_0}-1)((N_1+P)^2-N_1^2 2^{-2C_0})+(N_\mathrm{r}-N_1)P^2}\right) \tag{54}$$

$$K_2^*=\frac{(K_1^*+N_1)(P+N_\mathrm{r})}{(P+N_1)2^{2C_0}} \tag{55}$$

$$\rho^*=\frac{P-\sqrt{(P-K_1^*)(P+N_\mathrm{r}-K_2^*)}}{\sqrt{K_1^*K_2^*}}. \tag{56}$$

*Proof.* Let

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

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

$$(P-K_1)(P+N_\mathrm{r}-K_2)\ge(P-\rho\sqrt{K_1K_2})^2. \tag{57}$$

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

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

will still hold. The expression we wish to maximize

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

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

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

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

$$(P-K_1)(P+N_\mathrm{r}-K_2)\ge(P-\rho\sqrt{K_1K_2})^2 \tag{58}$$

must hold with equality. Thus, any optimal choice for $\rho$ must satisfy $\rho\ge0$ and

$$\rho=\frac{P-\sqrt{(P-K_1)(P+N_\mathrm{r}-K_2)}}{\sqrt{K_1K_2}}. \tag{59}$$

Note that the other solution for $\rho$ in (58) is

$$\rho=\frac{P+\sqrt{(P-K_1)(P+N_\mathrm{r}-K_2)}}{\sqrt{K_1K_2}} \tag{60}$$

However, the optimizer $\rho$ must satisfy (59). If instead (60) holds, reducing $\rho$ to $P/\sqrt{K_1K_2}$ would strictly increase the objective function while continuing to satisfy (57).

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 $Y_{\mathrm{r}}$ would strictly improve the upper bound on the rate $R$. In other words, the maximizing distribution is such that (22) 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

$$\begin{aligned}&K_1\zeta(-2^{-2C_0}N_1+2^{-2C_0}P+N_1+P)+2\zeta P(2^{-2C_0}N_1-N_1-P)\\ &=(2^{-2C_0}N_1\zeta-N_1\zeta-P\zeta-P)\sqrt{(P-K_1)((P+N_1)\zeta-(K_1+N_1)\zeta 2^{-2C_0})}\end{aligned} \tag{61}$$

If we raise both sides to power two, we get a quadratic equation. This quadratic equation has two roots. One root is $K_1 = -N_1\frac{\zeta-2^{-2C_0}\zeta}{1-2^{-2C_0}\zeta} + P\frac{1-\zeta}{1-2^{-2C_0}\zeta}$, and the other is

$$K_1^* = P\left(1-\frac{P(N_1+P)(2^{-2C_0}-1)}{\zeta(2^{-2C_0}-1)((N_1+P)^2-N_1^2 2^{-2C_0})+(1-\zeta)2^{-2C_0}P^2}\right).$$

The root $K_1 = -N_1\frac{\zeta-2^{-2C_0}\zeta}{1-2^{-2C_0}\zeta} + P\frac{1-\zeta}{1-2^{-2C_0}\zeta}$ is never in $[0,P]$ and hence not acceptable. The reason is as follows: assume that the root is in $[0,P]$. Then, we obtain that

$$N_1\frac{\zeta-2^{-2C_0}\zeta}{1-2^{-2C_0}\zeta} \leq P\frac{1-\zeta}{1-2^{-2C_0}\zeta} \leq P+N_1\frac{\zeta-2^{-2C_0}\zeta}{1-2^{-2C_0}\zeta}$$

We must have $1-2^{-2C_0}\zeta \leq 0$ otherwise from $\zeta \geq 1$,

$$N_1\frac{\zeta-2^{-2C_0}\zeta}{1-2^{-2C_0}\zeta} \leq P\frac{1-\zeta}{1-2^{-2C_0}\zeta}$$

cannot hold. Next, from

$$P\frac{1-\zeta}{1-2^{-2C_0}\zeta} \leq P+N_1\frac{\zeta-2^{-2C_0}\zeta}{1-2^{-2C_0}\zeta}$$

we deduce

$$P\frac{2^{-2C_0}\zeta-\zeta}{1-2^{-2C_0}\zeta} \leq N_1\frac{\zeta-2^{-2C_0}\zeta}{1-2^{-2C_0}\zeta}$$

which cannot hold when $1-2^{-2C_0}\zeta \leq 0$.

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 $K_1 = K_1^*$.

$\square$