The Capacity of Several New Classes of Semi-Deterministic Relay Channels

Hon-Fah Chong and Mehul Motani
Electrical and Computer Engineering Department,
National University of Singapore, Singapore 117576.
✉: {chong.hon.fah,motani}@nus.edu.sg

Abstract

The relay channel consists of a transmitter input x_1, a relay input x_2, a relay output y_2 and a receiver output y_3. In this paper, we establish the capacity of three new classes of semi-deterministic relay channels: 1) A class of degraded semi-deterministic relay channels, 2) a class of semi-deterministic orthogonal relay channels and 3) a class of semi-deterministic relay channels with relay-transmitter feedback.

For the first class of relay channels, the output of the relay y_2 depends on a deterministic function of the transmitter's input x_1, i.e., on $s = f_1(x_1)$, rather than on x_1 directly. In addition, the relay channels satisfy the condition that $S \to (X_2, Y_2) \to Y_3$ forms a Markov chain for all input probability distributions $p(x_1, x_2)$. Hence, the first class of relay channels includes, but is strictly not limited to, the class of degraded relay channels previously considered by Cover and El Gamal. The partial decode-and-forward strategy achieves the capacity of the class of degraded semi-deterministic relay channels.

Next, we consider the class of semi-deterministic orthogonal relay channels where there are orthogonal channels from the relay to the receiver and from the transmitter to the receiver. In addition, the output of the relay y_2 is a deterministic function of x_1, x_2 and y_3, i.e., $y_2 = f_4(x_1, x_2, y_3)$. The class of semi-deterministic orthogonal relay channels is a generalization of the class of deterministic relay channels considered by Kim. The compress-and-forward strategy achieves the capacity of the class of semi-deterministic orthogonal relay channels.

For the third class of relay channels, there is a causal and noiseless feedback from the relay to the transmitter. In addition, similar to the second class of relay channels, the output of the relay y_2 is a deterministic function of x_1, x_2 and y_3. Both the generalized strategy of Gabbai and Bross and the hash-and-forward strategy of Kim achieve the capacity of the class of semi-deterministic relay channels with relay-transmitter feedback.

I. Introduction

The discrete-memoryless relay channel consists of four sets—$\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{Y}_2$, $\mathcal{Y}_3$—and a collection of conditional probability mass functions $p(.,.|x_1, x_2)$ on $\mathcal{Y}_2 \times \mathcal{Y}_3$, one for each $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. The transmitter input is denoted by $x_1 \in \mathcal{X}_1$, the relay input by $x_2 \in \mathcal{X}_2$, the relay output by $y_2 \in \mathcal{Y}_2$ and the receiver output by $y_3 \in \mathcal{Y}_3$.

A $\left(2^{NR}, N\right)$ code for a relay channel without feedback consists of a set of integers $\mathcal{W} = \left\{1, 2, ..., \left\lfloor 2^{NR} \right\rfloor\right\}$, an encoding function

$$e : \left\{1, 2, ..., \left\lfloor 2^{NR} \right\rfloor\right\} \to \mathcal{X}_1^N$$

a set of relay functions $\{\Psi_n\}_{n=1}^{n=N}$ such that

$$\Psi_n : \mathcal{Y}_2^{n-1} \to \mathcal{X}_2, \ \ 1 \le n \le N$$

and a decoding function

$$d : \mathcal{Y}_3^N \to \left\{1, 2, ..., \left\lfloor 2^{NR} \right\rfloor\right\}.$$

The relay is causal in nature. Hence, the input of the relay x_{2n} is allowed to depend only on the past outputs of the relay $y_{21}, y_{22}, ..., y_{2n-1}$. If the message $w \in \mathcal{W}$ is sent, let

$$\lambda(w) = \Pr\left\{d\left(Y_3^N\right) \neq w | w \text{ sent}\right\}$$

denote the conditional probability of error. The average probability of error is defined by

$$P_e^{(N)} = \frac{1}{\left\lfloor 2^{NR} \right\rfloor} \sum_w \lambda(w).$$

The probability of error is calculated under the uniform distribution over the codewords $w \in \mathcal{W}$. The rate R is said to be achievable by the relay channel if there exists a sequence of $\left(2^{NR}, N\right)$ codes with $P_e^{(N)} \to 0$ as $N \to \infty$. The capacity C of a relay channel is the supremum of the set of achievable rates.

For a relay channel with causal and noiseless relay-transmitter feedback, the only difference is that the transmitter consists of a set of encoding functions $\{\Xi_n\}_{n=1}^{n=N}$ such that

$$\Xi_n : \mathcal{W} \times \mathcal{Y}_2^{n-1} \to \mathcal{X}_1, \quad 1 \le n \le N.$$

The relay channel was first introduced by van der Meulen in [1], [2]. Cover & El Gamal established two fundamental coding theorems for the relay channel in an important paper [3]. In addition, these two coding theorems were combined in the same paper to give the best lower bound for the capacity of a general relay channel [3, Theorem 7]. Recently, Chong et al. determined a potentially larger achievable rate in [4, Theorem 2]. In particular, they determined that the following rate is achievable for any relay channel:

$$R_{\text{CMG}} = \sup \min \left\{ \begin{array}{c} I\left(X_1; \hat{Y}_2 Y_3 | U X_2\right) + I\left(U; Y_2 | V X_2\right) \\ I\left(X_1 X_2; Y_3\right) - I\left(Y_2; \hat{Y}_2 | U X_1 X_2 Y_3\right) \end{array} \right\} \tag{1}$$

where the supremum is taken over all joint probability mass functions of the form

$$p(v, u, x_1, x_2, y_2, \hat{y}_2, y_3) = p(v)\, p(u|v)\, p(x_1|u)\, p(x_2|v)\, p(y_2, y_3|x_1, x_2)\, p(\hat{y}_2|x_2, y_2, u) \tag{2}$$

and subject to the constraint

$$I(X_2; Y_3|UV) \ge I\left(\hat{Y}_2; Y_2 | U X_2 Y_3\right). \tag{3}$$

The capacity of the relay channel has been determined for the following special cases:

1) the degraded relay channel, the reversely degraded relay channel and the relay channel with causal noiseless feedback from the receiver to the relay [3];
2) the semideterministic relay channel [5];
3) a class of relay channels with orthogonal components [6];
4) a class of modulo-sum relay channels [7];
5) a class of deterministic relay channels [8].

However, the capacity of the general relay channel remains unknown. The achievability of the above classes of relay channels follows directly from appropriate substitutions for the auxiliary random variables in [3, Th. 7]. Moreover, except for the class of modulo-sum relay channels [7], the capacity of all the other classes of relay channels meet the cut-set upper bound. The question remains as to whether there exists other classes of relay channels where the lower bound given by [3, Th. 7] meets the cut-set upper bound.

In this paper, we answer this question affirmatively and establish the capacity of three new classes of semi-deterministic relay channels. The paper is organized as follows:

- In Section II, we consider the class of degraded semi-deterministic relay channels which strictly includes the class of degraded relay channels.
- In Section III, we consider the class of semi-deterministic orthogonal relay channels where there are orthogonal channels from the transmitter to the receiver and from the relay to the receiver. Furthermore, the class of semi-deterministic orthogonal relay channels satisfy the condition that the output of the relay y_2 is a deterministic function of x_1, x_2 and y_3.
- In Section IV, we consider the class of relay channels with causal and noiseless relay-transmitter feedback. Similar to the second class of relay channels, the output of the relay y_2 is a deterministic function of x_1, x_2 and y_3.

II. DEGRADED SEMI-DETERMINISTIC RELAY CHANNELS

We first describe the class of degraded semi-deterministic relay channels as shown in Fig. 1.

***Definition** 1:* Let $s \in \mathcal{S}$ be a deterministic function of $x_1 \in \mathcal{X}_1$, i.e.,

$$s = f_1(x_1). \tag{4}$$

Hence, we have $\|\mathcal{S}\| \leq \|\mathcal{X}_1\|$.

We define the class of degraded semi-deterministic relay channels as those channels which satisfy the following conditions:

- The conditional probability mass function of the channel can be expressed as

$$p(y_2, y_3|x_1, x_2) = p(y_3|x_1, x_2)\, p(y_2|s, x_2, y_3). \tag{5}$$

  We emphasize that $s = f_1(x_1)$ and use the notation s for the sake of brevity.
- In addition, we require that the following Markov chain:

$$S \to (X_2, Y_2) \to Y_3 \tag{6}$$

  holds true for all input probability distributions $p(x_1, x_2)$.

***Remark** 1:* A particular instance when condition (6) holds true for all input probability distributions is when s is a deterministic function of x_2 and y_2, i.e., $s = f_2(x_2, y_2)$. We can also see by inspection from (5) that the following Markov chain

$$X_1 \to (S, X_2, Y_3) \to Y_2 \tag{7}$$

holds true for all input probability distributions. More specifically, this follows from the equalities:

$$\begin{aligned} p\left(y_2|s,x_1,x_2,y_3\right) &\overset{(a)}{=} p\left(y_2|x_1,x_2,y_3\right) \\ &= \frac{p\left(y_2|x_1,x_2,y_3\right)p\left(y_3|x_1,x_2\right)}{p\left(y_3|x_1,x_2\right)} \\ &= \frac{p\left(y_2,y_3|x_1,x_2\right)}{p\left(y_3|x_1,x_2\right)} \\ &= \frac{p\left(y_3|x_1,x_2\right)p\left(y_2|s,x_2,y_3\right)}{p\left(y_3|x_1,x_2\right)} \\ &= p\left(y_2|s,x_2,y_3\right). \end{aligned} \tag{8}$$

where

(a) follows from the fact that s is a deterministic function of x_1.

Hence, the output of the relay y_2 depends (probabilistically) on s, x_2 and y_3 as shown in Fig. 1.

The main contribution of this paper for the class of degraded semi-deterministic relay channels described in Definition 1 is given by the following theorem:

Theorem *1:* The capacity of the degraded semi-deterministic relay channel is given by

$$C_1 = \sup_{p(x_1,x_2)} \min\left\{ I\left(S;Y_2|X_2\right) + I\left(X_1;Y_3|SX_2\right), I\left(X_1X_2;Y_3\right) \right\}. \tag{9}$$

Proof: 1) Achievability: This follows directly from substituting $V \triangleq X_2$, $U \triangleq (S,X_2)$ and $\hat{Y}_2 \triangleq \emptyset$ into [3, Th. 7].

2) Converse: The converse follows from the cut-set upper bound [3, Th. 4]. Hence, C_1 is upper bounded by

$$C_1 \leq \sup_{p(x_1,x_2)} \min\left\{I\left(X_1;Y_2Y_3|X_2\right), I\left(X_1X_2;Y_3\right)\right\}. \tag{10}$$

The first term $I\left(X_1;Y_2Y_3|X_2\right)$ can be expressed as

$$\begin{aligned} I\left(X_1;Y_2Y_3|X_2\right) &\overset{(a)}{=} I\left(SX_1;Y_2Y_3|X_2\right) \\ &= I\left(S;Y_2Y_3|X_2\right) + I\left(X_1;Y_2Y_3|SX_2\right) \\ &= I\left(S;Y_2|X_2\right) + I\left(S;Y_3|X_2Y_2\right) + I\left(X_1;Y_2Y_3|SX_2\right) \\ &\overset{(b)}{=} I\left(S;Y_2|X_2\right) + I\left(X_1;Y_2Y_3|SX_2\right) \\ &= I\left(S;Y_2|X_2\right) + I\left(X_1;Y_3|SX_2\right) + I\left(X_1;Y_2|SX_2Y_3\right) \\ &\overset{(c)}{=} I\left(S;Y_2|X_2\right) + I\left(X_1;Y_3|SX_2\right). \end{aligned} \tag{11}$$

where

(a) follows from the fact that s is a deterministic function of x_1,

(b) follows from the fact that $S \to X_2Y_2 \to Y_3$ forms a Markov chain (condition (6)) and

(c) follows from the fact that $X_1 \to (S,X_2,Y_3) \to Y_2$ forms a Markov chain (condition (7)).

■

***Corollary** 1:* If s is a deterministic function of x_2 and y_2, the capacity is given by

$$C_1 = \sup_{p(x_1,x_2)} \min\left\{H\left(S|X_2\right) + I\left(X_1;Y_3|SX_2\right), I\left(X_1X_2;Y_3\right)\right\}. \tag{12}$$

Proof: The proof follows directly from Theorem 1 and the fact that s is a deterministic function of x_2 and y_2. ■

III. Semi-deterministic orthogonal relay channels

In [8], Kim considered a class of relay channels where there are orthogonal channels from the transmitter/relay to the receiver. The channel from the relay to the receiver is a noiseless one and the output of the relay y_2 is a deterministic function of x_1 and y_3. The cut-set upper bound is maximized by independent input probability distributions and is achievable by the compress-and-forward strategy. We can extend the result to a larger class of relay channels. We state this formally below.

***Theorem** 2:* If independent input probability distributions attain the cut-set upper bound and if y_2 is a deterministic function of x_1, x_2 and y_3, the capacity of the relay channel is given by the cut-set upper bound.

Proof: Refer to Appendix I. ■

In this section, we consider a class of relay channels which also satisfies the conditions of Theorem 2. The class of semi-deterministic orthogonal relay channels as shown in Fig. 2 is a generalization of the relay channels considered by Kim. There is a noisy link from the relay to the receiver and the output of the relay y_2 is a deterministic function of x_1, x_2 and y_3. Moreover, the output of the relay y_2 and the output of the receiver from the relay y_{32} depends (probabilistically) on an additional noise component which is independent of the inputs x_1 and x_2. The main part of the proof consists of showing that independent input probability distributions maximizes the cut-set upper bound for this class of relay channels.

We first define the class of semi-deterministic orthogonal relay channels below.

***Definition** 2:* The channel output of the receiver for the semi-deterministic orthogonal relay channel is given by $y_3 = (y_{31}, y_{32}) \in \mathcal{Y}_{31} \times \mathcal{Y}_{32}$. There is an orthogonal channel from the transmitter to the receiver (the output is denoted by $y_{31} \in \mathcal{Y}_{31}$) and from the relay to the receiver (the output is denoted by $y_{32} \in \mathcal{Y}_{32}$). Furthermore, the output of the relay Y_2 and the output of the receiver from the relay Y_{32} is dependent on a noise component T.

More specifically, we define the class of semi-deterministic orthogonal relay channels as those channels which satisfy the following conditions:

- There is a noise component T (with values $t \in \mathcal{T}$) independent of the inputs X_1 and X_2. We also require that t satisfies the following relationship

$$t = f_3\left(x_2, y_{32}\right) \tag{13}$$

  where f_3 is a deterministic function.

- The conditional probability mass function describing the channel can be expressed as

$$p\left(y_2,y_{31},y_{32}|x_1,x_2\right)=p\left(y_{31}|x_1\right)p\left(y_2|x_1,t,y_{31}\right)p\left(y_{32}|x_2,t\right)p\left(t\right). \tag{14}$$

  We emphasize that $t=f_3\left(x_2,y_{32}\right)$ and use the notation t for the sake of brevity.
- In addition, we require that y_2 satisfies the following relationship

$$y_2=f_4\left(x_1,x_2,y_{31},y_{32}\right) \tag{15}$$

  where f_4 is a deterministic function.

The main contribution of this paper for the class of semi-deterministic orthogonal relay channels described in Definition 2 is given by the following theorem:

Theorem *3:* The capacity of the semi-deterministic orthogonal relay channel is given by

$$C_2=\sup_{p(x_1)p(x_2)}\min\left\{I\left(X_1;Y_{31}\right)+I\left(X_2;Y_{32}\right),I\left(X_1;Y_2Y_{31}|T\right)\right\}. \tag{16}$$

Proof: Refer to Appendix II. ■

Remark *2:* We may verify that the class of relay channels considered by Kim satisfy the conditions of Definition 2 (set $T=\emptyset$ and $Y_{32}=X_2$, where we have $R_0=H\left(X_2\right)$). We can also combine the class of degraded semi-deterministic relay channels with the class of semi-deterministic orthogonal relay channels to obtain a new class of relay channels whose capacity can also be determined (see [9]).

Example *1:* We consider the discrete semi-deterministic modulo-sum relay channel shown in Fig. 3. This channel is specified by seven finite sets with the same cardinality, i.e.,

$$\mathcal{X}_1=\mathcal{X}_2=\mathcal{Y}_2=\mathcal{Y}_{31}=\mathcal{Y}_{32}=\mathcal{N}_1=\mathcal{N}_2=\{0,1,...,m-1\} \tag{17}$$

where m is an arbitrary positive integer greater than 1.

The conditional probability mass function $p\left(y_2,y_{31},y_{32}|x_1,x_2\right)$ is determined from the following equations:

$$Y_2=aX_1\oplus N_1\oplus N_2 \tag{18}$$

$$Y_{31}=bX_1\oplus N_1 \tag{19}$$

$$Y_{32}=X_2\oplus N_2 \tag{20}$$

where $\oplus$ denotes addition modulo m; a and b are positive integers; and N_1 and N_2 are independent noise variables defined over $\mathcal{N}_1$ and $\mathcal{N}_2$, respectively, with probability mass-functions given by

$$\mathbf{P}_{N_1}\triangleq\left(P_{N_1}\left(0\right),P_{N_1}\left(1\right),...,P_{N_1}\left(m-1\right)\right) \tag{21}$$

$$\mathbf{P}_{N_2}\triangleq\left(P_{N_2}\left(0\right),P_{N_2}\left(1\right),...,P_{N_2}\left(m-1\right)\right). \tag{22}$$

We note that when $a=b$, the discrete semi-deterministic modulo-sum relay channel is just a reversely degraded relay channel as $X_1\rightarrow\left(X_2,Y_{31},Y_{32}\right)\rightarrow Y_2$ forms a Markov chain. However, in general, when $a\neq b$, the discrete semi-deterministic modulo-sum relay channel does not fall under any classes of relay channels whose capacity has been previously proven.

Furthermore, we may easily verify that this class of modulo-sum relay channels satisfies the conditions of Definition 2 ($T = N_2$). Hence, the capacity of the class of semi-deterministic modulo-sum relay channels follows from Theorem 3.

IV. Semi-deterministic relay channels with causal noiseless relay-transmitter feedback

Finally, we consider a class of relay channels with causal noiseless relay-transmitter feedback as shown in Fig. 4.

***Definition** 3:* We define the class of semi-deterministic relay channels with causal noiseless relay-transmitter feedback as those channels which satisfy the condition that the output of the relay y_2 is a deterministic function of x_1, x_2 and y_3, i.e.,

$$y_2 = f_5\left(x_1, x_2, y_3\right). \tag{23}$$

The relay channel with causal noiseless relay-transmitter feedback was studied by Gabbai and Bross in [10]. In [10, Th. 3], Gabbai and Bross proved that the following rate is achievable for a discrete memoryless relay channel with causal noiseless relay-transmitter feedback:

$$R_{\text{GB}} = \sup \left\{ \begin{array}{c} I\left(X_1; \hat{Y}_2 Y_3 | V X_2\right) + I\left(V; Y_2 | X_2\right) \\ I\left(X_1; \hat{Y}_2 Y_3 | V X_2\right) + I\left(V X_2; Y_3\right) - I\left(\hat{Y}_2; Y_2 | X_2, V, Y_3\right) \end{array} \right\} \tag{24}$$

where the supremum is taken over all probability mass functions of the form

$$p\left(v, x_1, x_2, y_2, \hat{y}_2, y_3\right) = p\left(x_2\right) p\left(v|x_2\right) p\left(x_1|v, x_2\right) p\left(y_2, y_3|x_1, x_2\right) p\left(\hat{y}_2|x_2, y_2, v\right). \tag{25}$$

We can obtain the capacity of the class of relay channels in Definition 3 by appropriate substitutions for the auxiliary random variables in [10, Th. 3]. In particular, the capacity of the class of relay channels in Definition 3 is given by the following theorem:

***Theorem** 4:* The capacity of the semi-deterministic relay channel with causal noiseless relay-transmitter feedback as shown in Fig. 4, satisfying condition (23), is given by

$$C_3 = \sup_{p(x_1, x_2)} \min \left\{ I\left(X_1; Y_2 Y_3 | X_2\right), I\left(X_1 X_2; Y_3\right) \right\}. \tag{26}$$

Proof: 1) Achievability: We note that rate C_3 is achievable by substituting $V \triangleq X_2$ and $\hat{Y}_2 \triangleq Y_2$ in [10, Th. 3]. To see this, we note that the first term gives us $I\left(X_1; Y_2 Y_3 | X_2\right)$. For the second term, we obtain

$$\begin{aligned} & I\left(X_1; Y_2 Y_3 | X_2\right) + I\left(X_2; Y_3\right) - H\left(Y_2 | X_2, Y_3\right) \\ & = I\left(X_1 X_2; Y_3\right) + I\left(X_1; Y_2 | X_2 Y_3\right) - H\left(Y_2 | X_2, Y_3\right) \\ & = I\left(X_1 X_2; Y_3\right) + H\left(Y_2 | X_2 Y_3\right) - H\left(Y_2 | X_1 X_2 Y_3\right) - H\left(Y_2 | X_2, Y_3\right) \\ & = I\left(X_1 X_2; Y_3\right) - H\left(Y_2 | X_1 X_2 Y_3\right) \\ & \stackrel{(a)}{=} I\left(X_1 X_2; Y_3\right). \end{aligned} \tag{27}$$

where

(a) follows from condition (23).

Hence, the rate C_3 is achievable. An alternative proof of achievability may be shown using hash-and-forward [8] (see [11]).

2) Converse: The proof for the converse follows directly from the cut-set upper bound for the relay channel without feedback since it is also an upper bound for the relay channel with causal noiseless receiver-transmitter, relay-transmitter and receiver-relay feedback [3]. Therefore, the cut-set upper bound for the relay channel without feedback is also an upper bound for the relay channel with causal noiseless relay-transmitter feedback. ■

Remark *3:* This class of relay channels is also closely related to the class of deterministic relay channels solved by Kim in [8]. In [8], the output of the relay y_2 is a deterministic function of x_2 and y_3 but there is a noiseless link of rate R_0 from the relay to the receiver. In our case, the output of the relay y_2 is a deterministic function of x_1, x_2 and y_3 but there is a causal noiseless feedback from the relay to the transmitter. In both cases, the capacity may be achieved by both the hash-and-forward scheme or the compress-and-forward scheme.

Example *2:* We consider a semi-deterministic Gaussian relay channel with causal noiseless relay-transmitter feedback as shown in Fig. 5. The outputs of the channel are given by

$$Y_2 = aX_1 + Z \tag{28}$$

$$Y_3 = (BX_1 + X_2 + Z, B) \tag{29}$$

where Z is Gaussian noise of variance σ^2, B is a random variable which takes on the value b_1 with probability α and the value b_2 with probability $1 - \alpha$. If $a \neq b_1$ or $a \neq b_2$, we readily see that this is neither a degraded Gaussian relay channel ($X_1 \to (X_2, Y_2) \to Y_3$ does not form a Markov chain) nor a reversely degraded Gaussian relay channel ($X_1 \to (X_2, Y_3) \to Y_2$ does not form a Markov chain). Moreover, we note that the output of the receiver y_3 is not a deterministic function of x_1, x_2 and y_2. Therefore, the transmitter only has feedback from the relay but not from the receiver. On the other hand, we note that the output of the relay y_2 is a deterministic function of x_1, x_2 and y_3. Its capacity is thus given by Theorem 4.

APPENDIX I
PROOF OF THEOREM 2

1) Achievability: Substituting $U \triangleq \emptyset$ and $V \triangleq \emptyset$ in [3, Th. 7] gives us the achievable rate

$$R = \sup I\left(X_1; \hat{Y}_2 Y_3 | X_2\right) \tag{30}$$

subject to the constraint

$$I\left(X_2; Y_3\right) \geq I\left(Y_2; \hat{Y}_2 | X_2 Y_3\right) \tag{31}$$

where the supremum is taken over all joint probability mass functions of the form

$$p\left(x_1\right) p\left(x_2\right) p\left(y_2, y_3 | x_1, x_2\right) p\left(\hat{y}_2 | x_2, y_2\right). \tag{32}$$

If $H(Y_2|X_2Y_3) \leq I(X_2;Y_3)$, we immediately obtain $R = I(X_1;Y_2Y_3|X_2)$. If $H(Y_2|X_2Y_3) > I(X_2;Y_3)$, we have

$$\begin{aligned}
R = I\left(X_1;\hat{Y}_2Y_3|X_2\right) &= I(X_1;Y_3|X_2) + I\left(X_1;\hat{Y}_2|X_2Y_3\right) \\
&\overset{(a)}{=} I(X_1;Y_3|X_2) + H\left(\hat{Y}_2|X_2Y_3\right) - H\left(\hat{Y}_2|X_1X_2Y_2Y_3\right) \\
&= I(X_1;Y_3|X_2) + H\left(\hat{Y}_2|X_2Y_3\right) - H\left(\hat{Y}_2|X_2Y_2Y_3\right) \\
&= I(X_1;Y_3|X_2) + I\left(Y_2;\hat{Y}_2|X_2Y_3\right) \\
&\overset{(b)}{=} I(X_1;Y_3|X_2) + I(X_2;Y_3) \\
&= I(X_1X_2;Y_3). \qquad (33)
\end{aligned}$$

where

(a) follows from the fact that y_2 is a deterministic function of x_1, x_2, y_3,

(b) follows from the fact that if $H(Y_2|X_2Y_3) > I(X_2;Y_3)$, there exists probability mass functions such that $I(X_2;Y_3) = I\left(Y_2;\hat{Y}_2|X_2Y_3\right)$.

2) Converse: The converse follows from the cut-set upper bound and the assumption that the cut-set upper bound is maximized by independent input probability distributions.

APPENDIX II
PROOF OF THEOREM 3

We first make some observations on the properties of the semi-deterministic orthogonal relay channel that will be useful later on.

***Property** 1:* For all input probability distributions $p(x_1,x_2)$, $(T,X_2,Y_{32}) \to X_1 \to Y_{31}$ forms a Markov chain.

***Property** 2:* For all input probability distributions $p(x_1,x_2)$, $(X_1,Y_{31}) \to X_2 \to Y_{32}$ forms a Markov chain.

***Property** 3:* For all input probability distributions $p(x_1,x_2)$, $(X_2,Y_{32}) \to (X_1,T) \to (Y_2,Y_{31})$ forms a Markov chain.

***Property** 4:* For independent input probability distributions, i.e., $p(x_1,x_2) = p(x_1)p(x_2)$, $(X_2,Y_{32}) \to T \to (Y_2,Y_{31})$ forms a Markov chain.

***Property** 5:* For independent input probability distributions, i.e., $p(x_1,x_2) = p(x_1)p(x_2)$, Y_{31} and Y_{32} are independent.

***Property** 6:* The terms $I(X_1;Y_2Y_{31}|T)$ and $I(X_1;Y_{31})$ for the semi-deterministic orthogonal relay channel is maximized by the marginal input probability distribution $p(x_1)$. The term $I(X_2;Y_{32})$ for the semi-deterministic orthogonal relay channel is maximized by the marginal input probability distribution $p(x_2)$.

We may readily prove Property 1-Property 4 from Definition 2. Property 5 can be shown from the following equalities:

$$\begin{aligned}
p(y_{31},y_{32}) &= \sum_{x_1,x_2\in\mathcal{X}_1\times\mathcal{X}_2} p(y_{31},y_{32}|x_1,x_2)\,p(x_1)\,p(x_2) \\
&= \sum_{x_1,x_2\in\mathcal{X}_1\times\mathcal{X}_2}\sum_{y_2\times\mathcal{Y}_2} p(y_2,y_{31},y_{32}|x_1,x_2)\,p(x_1)\,p(x_2)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{x_1,x_2\in\mathcal{X}_1\times\mathcal{X}_2}\sum_{y_2\times\mathcal{Y}_2} p\left(y_{31}|x_1\right)p\left(y_2|x_1,t,y_{31}\right)p\left(y_{32}|x_2,t\right)p\left(t\right)p\left(x_1\right)p\left(x_2\right)\\
&= \sum_{x_1,x_2\in\mathcal{X}_1\times\mathcal{X}_2} p\left(y_{31}|x_1\right)p\left(y_{32}|x_2,t\right)p\left(t\right)p\left(x_1\right)p\left(x_2\right)\sum_{y_2\times\mathcal{Y}_2} p\left(y_2|x_1,t,y_{31}\right)\\
&= \sum_{x_1,x_2\in\mathcal{X}_1\times\mathcal{X}_2} p\left(y_{31}|x_1\right)p\left(y_{32}|x_2,t\right)p\left(t\right)p\left(x_1\right)p\left(x_2\right)\\
&\overset{(a)}{=} \sum_{x_1,x_2\in\mathcal{X}_1\times\mathcal{X}_2} p\left(y_{31}|x_1\right)p\left(y_{32},t|x_2\right)p\left(x_1\right)p\left(x_2\right)\\
&\overset{(b)}{=} \sum_{x_1,x_2\in\mathcal{X}_1\times\mathcal{X}_2} p\left(y_{31}|x_1\right)p\left(y_{32}|x_2\right)p\left(x_1\right)p\left(x_2\right)\\
&= p\left(y_{31}\right)p\left(y_{32}\right)
\end{aligned}
\tag{34}
$$

where

(a) follows from the fact that T is independent of X_2 and

(b) follows from the fact that t is a deterministic function of x_2 and y_{32} (condition (13)).

To prove Property 6, we first consider the term $H\left(Y_2Y_{31}|T\right)$ which depends only on the probability distribution $p\left(t\right)$ (which is independent of X_1 and X_2) and the conditional probability distribution $p\left(y_2,y_{31}|t\right)$. We note that the conditional probability distribution $p\left(y_2,y_{31}|t\right)$ depends on the marginal probability distribution $p\left(x_1\right)$ and not on the joint probability distribution $p\left(x_1,x_2\right)$ from the following equalities:

$$
\begin{aligned}
p\left(y_2,y_{31}|t\right) &\overset{(a)}{=} \sum_{x_1\in\mathcal{X}_1}\sum_{x_2\in\mathcal{X}_2} p\left(y_2,y_{31}|t,x_1,x_2\right)p\left(x_1,x_2\right)\\
&\overset{(b)}{=} \sum_{x_1\in\mathcal{X}_1}\sum_{x_2\in\mathcal{X}_2} p\left(y_2,y_{31}|t,x_1\right)p\left(x_1,x_2\right)\\
&= \sum_{x_1\in\mathcal{X}_1} p\left(y_2,y_{31}|t,x_1\right)p\left(x_1\right)\sum_{x_2\in\mathcal{X}_2} p\left(x_2|x_1\right)\\
&= \sum_{x_1\in\mathcal{X}_1} p\left(y_2,y_{31}|t,x_1\right)p\left(x_1\right).
\end{aligned}
\tag{35}
$$

where

(a) follows from the fact that T is independent of X_1 and X_2 and

(b) follows from the fact that $X_2 \to \left(X_1,T\right) \to \left(Y_2,Y_{31}\right)$ forms a Markov chain (Property 3).

Hence, the term $H\left(Y_2Y_{31}|T\right)$ is maximized by the marginal input probability distribution $p\left(x_1\right)$. Similarly, the term $H\left(Y_2Y_{31}|TX_1\right)$ is maximized by the marginal input probability distribution $p\left(x_1\right)$. Therefore, $I\left(X_1;Y_2Y_{31}|T\right)$ is maximized by the marginal input probability distribution $p\left(x_1\right)$. We can likewise prove that $I\left(X_1;Y_{31}\right)$ is maximized by the marginal input probability distribution $p\left(x_1\right)$ from the fact that $X_2 \to X_1 \to Y_{31}$ forms a Markov chain (Property 1). We can also prove that $I\left(X_2;Y_{32}\right)$ is maximized by the marginal input probability distribution $p\left(x_2\right)$ from the fact that $X_1 \to X_2 \to Y_{32}$ forms a Markov chain (Property 2).

1) Achievability: This follows from the compress-and-forward strategy. If $H\left(Y_2|X_2Y_{31}Y_{32}\right) \le I\left(X_2;Y_{31}Y_{32}\right)$, following the proof of Theorem 2, we immediately obtain

$$
\begin{aligned}
R &= I\left(X_1;Y_2Y_{31}Y_{32}|X_2\right)\\
&= I\left(X_1;Y_{32}|X_2\right) + I\left(X_1;Y_2Y_{31}|X_2Y_{32}\right)
\end{aligned}
$$

$$\stackrel{(a)}{=} I\left(X_1;Y_2Y_{31}|X_2Y_{32}\right)$$
$$\stackrel{(b)}{=} I\left(X_1;Y_2Y_{31}|X_2Y_{32}T\right)$$
$$\stackrel{(c)}{=} I\left(X_1;Y_2Y_{31}|T\right) \tag{36}$$

where

(a) follows from the fact that $X_1 \to X_2 \to Y_{32}$ forms a Markov chain (Property 2),

(b) follows from the fact that t is a deterministic function of x_2 and y_{32} (condition (13)) and

(c) follows from the fact that $(X_2, Y_{32}) \to (X_1, T) \to (Y_2, Y_{31})$ forms a Markov chain for all input probability distributions (Property 3) and that $(X_2, Y_{32}) \to T \to (Y_2, Y_{31})$ forms a Markov chain for independent input probability distributions (Property 4).

If $H\left(Y_2|X_2Y_{31}Y_{32}\right) > I\left(X_2;Y_{31}Y_{32}\right)$, following the proof of Theorem 2, we have

$$R = I\left(X_1X_2;Y_{31}Y_{32}\right)$$
$$= I\left(X_1;Y_{31}\right) + I\left(X_2;Y_{31}|X_1\right) + I\left(X_1X_2;Y_{32}|Y_{31}\right)$$
$$\stackrel{(a)}{=} I\left(X_1;Y_{31}\right) + I\left(X_1X_2;Y_{32}|Y_{31}\right)$$
$$\stackrel{(b)}{=} I\left(X_1;Y_{31}\right) + H\left(Y_{32}\right) - H\left(Y_{32}|Y_{31},X_1,X_2\right)$$
$$\stackrel{(c)}{=} I\left(X_1;Y_{31}\right) + H\left(Y_{32}\right) - H\left(Y_{32}|X_2\right)$$
$$= I\left(X_1;Y_{31}\right) + I\left(X_2;Y_{32}\right) \tag{37}$$

where

(a) follows from the fact that $X_2 \to X_1 \to Y_{31}$ forms a Markov chain (Property 1),

(b) follows from the fact that Y_{31} and Y_{32} are independent if the input probability distributions are independent (Property 5) and

(c) follows from the fact that $(X_1, Y_{31}) \to X_2 \to Y_{32}$ forms a Markov chain (Property 2).

Hence, the compress-and-forward strategy achieves the rate

$$R = \sup_{p(x_1)p(x_2)} \min\left\{I\left(X_1;Y_{31}\right) + I\left(X_2;Y_{32}\right), I\left(X_1;Y_2Y_{31}|T\right)\right\} \tag{38}$$

for the semi-deterministic orthogonal relay channel.

2) Converse: The converse follows from the cut-set upper bound. We will show that the cut-set upper bound is maximized by independent input probability distributions. For the first term in (10), we obtain

$$I\left(X_1;Y_2Y_{31}Y_{32}|X_2\right)$$
$$= H\left(Y_{32}|X_2\right) + H\left(Y_2Y_{31}|Y_{32}X_2\right) - H\left(Y_2Y_{31}Y_{32}|X_1X_2\right)$$
$$\stackrel{(a)}{=} H\left(Y_{32}|X_2\right) + H\left(Y_2Y_{31}|TY_{32}X_2\right) - H\left(Y_2Y_{31}Y_{32}|X_1X_2\right)$$
$$\stackrel{(b)}{\leq} H\left(Y_{32}|X_2\right) + H\left(Y_2Y_{31}|T\right) - H\left(Y_2Y_{31}Y_{32}|X_1X_2\right)$$
$$\stackrel{(c)}{=} H\left(Y_{32}|X_2\right) + H\left(Y_2Y_{31}|T\right) - H\left(Y_2Y_{31}|X_1X_2Y_{32}\right) - H\left(Y_{32}|X_2\right)$$
$$\stackrel{(a)}{=} H\left(Y_2Y_{31}|T\right) - H\left(Y_2Y_{31}|X_1X_2Y_{32}T\right)$$
$$\stackrel{(d)}{=} H\left(Y_2Y_{31}|T\right) - H\left(Y_2Y_{31}|X_1T\right)$$

$$= I\left(X_1;Y_2Y_{31}|T\right) \tag{39}$$

(a) follows from the fact that t is a deterministic function of x_2 and y_{32} (condition (13)),

(b) follows from the fact that conditioning reduces entropy,

(c) follows from the fact that $X_1 \rightarrow X_2 \rightarrow Y_{32}$ forms a Markov chain (Property 2) and

(d) follows from the fact that $(X_2, Y_{32}) \rightarrow (X_1, T) \rightarrow (Y_2, Y_{31})$ forms a Markov chain (Property 3).

We note that $I\left(X_1;Y_2Y_{31}|T\right)$ is maximized by the marginal input probability distribution $p\left(x_1\right)$ (Property 6). Furthermore, we note that (b) can be replaced by an equality as long as the input probability distributions are independent (Property 4). Hence, independent input probability distributions maximize the first term in (10). For the second term in (10), we obtain

$$\begin{aligned}
I\left(X_1X_2;Y_{31}Y_{32}\right) &= H\left(Y_{31}\right) + H\left(Y_{32}|Y_{31}\right) - H\left(Y_{31},Y_{32}|X_1,X_2\right) \\
&\overset{(a)}{\leq} H\left(Y_{31}\right) + H\left(Y_{32}\right) - H\left(Y_{31},Y_{32}|X_1,X_2\right) \\
&\overset{(b)}{=} H\left(Y_{31}\right) + H\left(Y_{32}\right) - H\left(Y_{31}|X_1\right) - H\left(Y_{32}|X_2\right) \\
&= I\left(X_1;Y_{31}\right) + I\left(X_2;Y_{32}\right).
\end{aligned} \tag{40}$$

(a) follows from the fact that conditioning reduces entropy and

(b) follows from the fact that $(X_2, Y_{32}) \rightarrow X_1 \rightarrow Y_{31}$ forms a Markov chain (Property 1) and from the fact that $(X_1, Y_{31}) \rightarrow X_2 \rightarrow Y_{32}$ forms a Markov chain (Property 2).

We also note that $I\left(X_1;Y_{31}\right)$ is maximized by the marginal input probability distribution $p\left(x_1\right)$ and similarly, $I\left(X_2;Y_{32}\right)$ is maximized by the marginal input probability distribution $p\left(x_2\right)$ (Property 6). Furthermore, (a) can be replaced by an equality as long as the input probability distributions are independent (Property 5). Hence, independent input probability distributions achieve capacity.

REFERENCES

[1] E. C. van der Meulen, "Transmission of information in a T-terminal discrete memoryless channel," Ph.D. dissertation, Univ. of California, Berkeley, CA, Jun. 1968.

[2] ——, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[3] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sept. 1979.

[4] H. F. Chong, M. Motani, and H. K. Garg, "Generalized backward decoding strategies for the relay channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 394–401, Jan 2007.

[5] A. E. Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Trans. Inf. Theory*, vol. 28, no. 3, p. 536, May 1982.

[6] A. E. Gamal and S. Zahedi, "Capacity of a class of relay channels with orthogonal components," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1815–1817, May 2005.

[7] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 921–930, Mar. 2009.

[8] Y.-H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1328–1329, Mar. 2008.

[9] H. F. Chong and M. Motani, "The capacity of a class of mixture semi-deterministic relay channels," National University of Singapore, Tech. Rep., 2010.

[10] Y. Gabbai and S. I. Bross, "Achievable rates for the discrete memoryless relay channel with partial feedback configurations," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 4989–5007, Nov. 2006.

[11] H. F. Chong and M. Motani, "The capacity regions of some classes of deterministic relay channels," in *IEEE Intern. Symp. on Inform. Theory, 2008*, Toronto, Ontario, Canada, Jul. 2008, pp. 344–348.

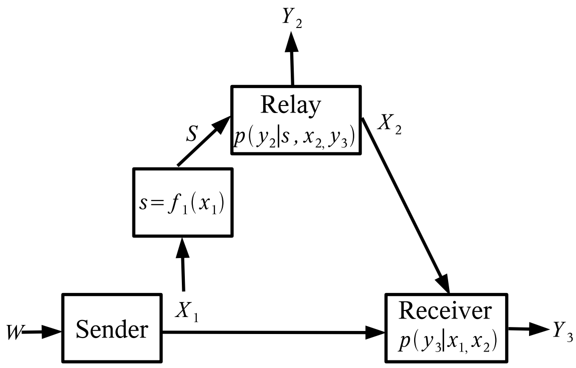

Fig. 1. Degraded semi-deterministic relay channel

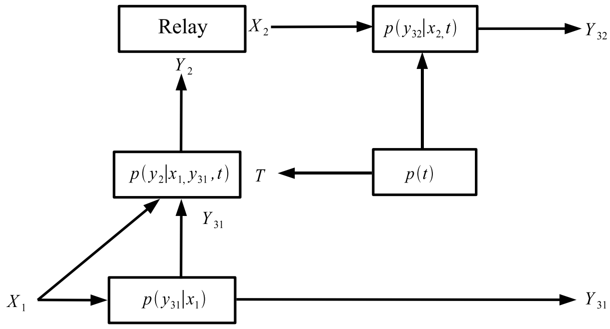

Fig. 2. Semi-deterministic orthogonal relay channel

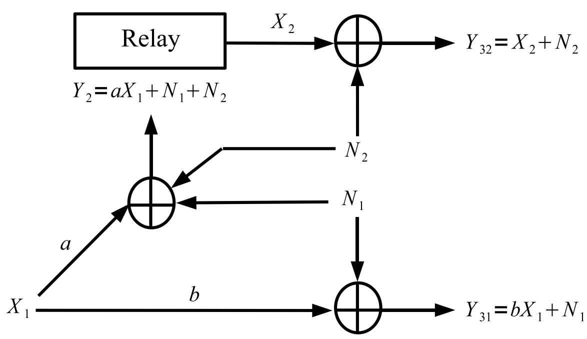

Fig. 3. Semi-deterministic modulo-sum relay channel

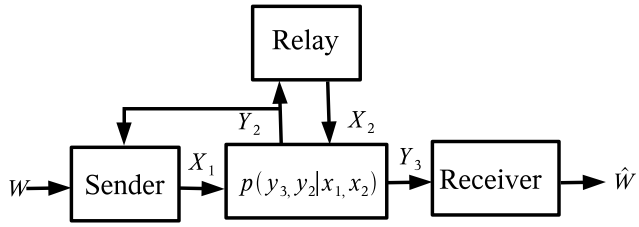

Fig. 4. Relay channels with causal noiseless relay-transmitter feedback

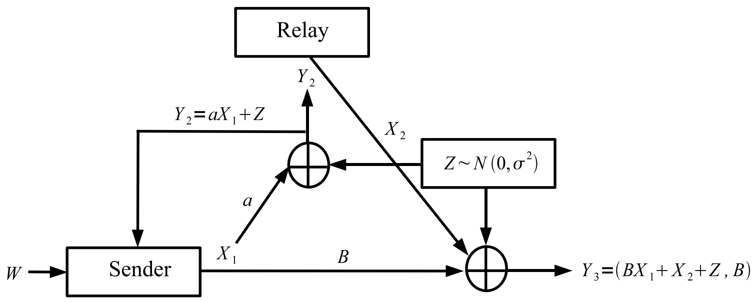

Fig. 5. A semi-deterministic Gaussian relay channel with causal noiseless relay-transmitter feedback