RESEARCH　　　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

The capacity of a class of state-dependent relay channel with orthogonal components and side information at the source and the relay

Zhixiang Deng[1,4], Baoyun Wang[1,2,3*] and Fei Lang[1]

Abstract

In this paper, a class of state-dependent relay channel with orthogonal channels from the source to the relay and from the source and the relay to the destination is studied. The two orthogonal channels are corrupted by two independent channel states S_R and S_D, respectively, which are known to both the source and the relay. The lower bounds on the capacity are established for the channel either with non-causal channel state information or with causal channel state information. Further, we show that the lower bound with non-causal channel state information is tight if the receiver output Y is a deterministic function of the relay input X_r, the channel state S_D, and one of the source inputs X_D, i.e., $Y = f(X_D, X_r, S_D)$, and the relay output Y_r is restricted to be controlled by only the source input X_R and the channel state S_R, i.e., the channel from the source to the relay is governed by the conditional probability distribution $P_{Y_r|X_R,S_R}$. The capacity for this class of semi-deterministic orthogonal relay channel is characterized exactly. The results are then extended to the Gaussian cases, modeling the channel states as additive Gaussian interferences. The capacity is characterized for the channel when the channel state information is known non-causally. However, when the channel state information is known causally, the capacity cannot be characterized in general. In this case, the lower bound on the capacity is established, and the capacity is characterized when the power of the relay is sufficiently large. Some numerical examples of the causal state information case are provided to illustrate the impact of the channel state and the role of the relay in information transmission and in cleaning the channel state.

Keywords: State-dependent relay channel with orthogonal components; Non-causal channel state information; Causal channel state information; Dirty paper coding

1 Introduction

We consider a state-dependent relay channel with orthogonal components as shown in Figure 1. The channel from the source to the relay and the channel from the source and the relay to the destination are assumed orthogonal. The source wants to send a message W to the destination with the help of the relay in n channel uses. Through a given memoryless probability law $P_{Y_r|X_R,X_r,S_R}$ $P_{Y|X_D,X_r,S_D}$, the channel outputs $\mathbf{Y}_r^n$ for the relay is controlled by the source inputs $\mathbf{X}_R^n$, the relay inputs $\mathbf{X}_r^n$, and
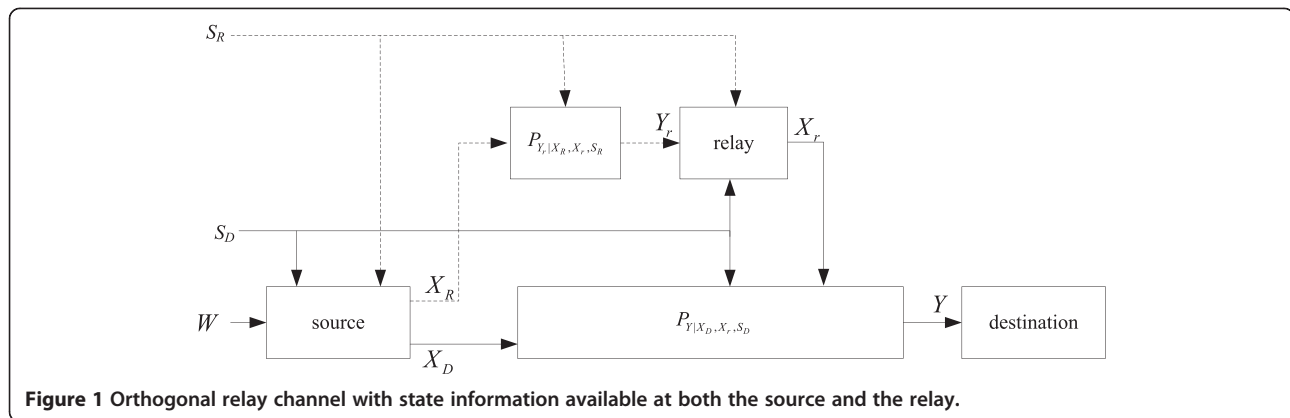
the channel state $\mathbf{S}_R^n$, while the channel outputs $\mathbf{Y}^n$ for the destination is controlled by the source input $\mathbf{X}_D^n$, the relay inputs $\mathbf{X}_r^n$, and the channel state $\mathbf{S}_D^n$. The state sequences $\mathbf{S}_R^n = (S_{R,1}, S_{R,2}, ..., S_{R,n})$ and $\mathbf{S}_D^n = (S_{D,1}, S_{D,2}, ..., S_{D,n})$ are independent and identically distributed (i.i.d.) with $S_{R,i} \sim Q_{S_R}(s_{R,i})$ and $S_{D,i} \sim Q_{S_D}(s_{D,i})$, respectively. We assume S_R and S_D are independent. The channel state information about S_R and S_D is known to the source and the relay causally (that is, only $\mathbf{S}_R^i$ and $\mathbf{S}_D^i$ are known before transmission i takes place) or non-causally (that is, entire $\mathbf{S}_R^n$ and $\mathbf{S}_D^n$ are known before communication commences). The destination estimates the message sent by the source from its received channel output $\mathbf{Y}^n$. In this paper, we study the capacity of this model.

---

* Correspondence: bywang@njupt.edu.cn
[1]College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 21003, China
[2]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
Full list of author information is available at the end of the article

Figure 1 Orthogonal relay channel with state information available at both the source and the relay.

1.1 Background

In many communication models, the communicating parties typically have some knowledge on the communicating channel or attempt to learn about the channel. State-dependent channels have brought wide attention in recent years [1]. Shannon first considered a single-user channel, wherein the channel state information was causally known to the transmitter [2]; the capacity of this channel was characterized. Gel'fand and Pinsker [3] derived a method to determine the capacity of a channel when the channel state information was non-causally known to the transmitter; this method was later called the Gel'fand-Pinsker (GP) coding scheme. In [4], Costa studied a Gaussian channel with additive white Gaussian noise (AWGN) and additive Gaussian interference known non-causally to the transmitter; it was demonstrated that with dirty paper coding (DPC), capacity can be achieved as if no interference existed in the channel.

Extensions to multiple user channels were performed by Gel'fand and Pinsker in [5], where it was shown that interference cancellation was possible in the Gaussian broadcast channel (BC) and Gaussian multiple-access channel (MAC). In multiple-user state-dependent channels, the channel state information may be known to all the users or only some of them. In [6], Sigurjonsson and Kim characterized the capacity of a degraded broadcast channel and the capacity of a physically degraded relay channel where the channel state information was causally known to the transmitters. Inner bounds for the two-user BC with non-causal side information at the transmitter were derived by extending Marto's achievable scheme to the state-dependent channels in [7]. In [8], Steinberg derived the inner and outer bounds for a degraded BC with non-causal side information and characterized the capacity region when the side information was provided to the transmitter in a causal manner. In [9], information theoretic performance limits for three classes of two-user state-dependent discrete memoryless BCs with non-causal side information at the encoder were derived.

The state-dependent two-user MAC with state information non-causally known to one of the encoders was considered in [10] and [11]. For the MAC with asymmetric channel state information at the transmitters and full channel state information at the decoder, a single-letter capacity region was characterized when the channel state available at one of the encoders was a subset of the channel state available at the other encoder [12]. However, for the general case, only the inner and outer bonds were derived. It is not easy to characterize the explicit capacity region for general state-dependent MACs even when the channel state information is known to all transmitters. Capacity regions are only characterized in some special cases, e.g., Gaussian MAC with additive interference known to both encoders [13]. In some cases where the cooperation between the transmitters is allowed, capacity regions are also characterized, e.g., in [14], explicit capacity region was characterized for the MAC with one informed encoder which transmitted a common message and a private message, while the uninformed encoder only transmitted a common message; in [15], the capacity region was derived for a two-user dirty paper Gaussian MAC with conferencing encoders.

The relay channels capture both the MAC and BC characteristics. The state-dependent relay channels were studied in [16-21]. Zaidi et al. [16] studied the relay channel with non-causal channel state information known to only the relay. The lower and upper bounds were derived by a coding scheme at the relay that used a combination of codeword splitting, Gel'f and-Pinsker binning, and decode-and-forward (DF) relaying. When the channel state information was known only at the source node, the lower and upper bounds were obtained in [17-20]. In [17], the coding scheme for the lower bound used techniques of rate splitting at the source, partial decode-and-forward (PDF) relaying, and a GP-like binning scheme. In order to derive the lower bound of the capacity, [18] proposed two achievable schemes: (i) state description by which the source described the channel state to the relay and the destination and (ii)

analog input description by which the source firstly computed the appropriate input that the relay would send had the relay known the channel state and then transmitted this appropriate input to the relay. With the same achievable schemes proposed in [18], the authors [19] obtained two corresponding lower bounds for the state-dependent relay channel with orthogonal components and channel state information known non-causally to the source. A similar orthogonal relay channel that was corrupted by an interference which was known non-causally to the source was considered in [20], in which several transmission strategies were proposed, assuming the interference had structure. Akhbari et al. [21] considered a state-dependent relay channel in three different cases: only the relay or only the source or both the source and the relay knew the channel state information non-causally. Lower bounds of the capacity were established based on using GP coding and compress-and-forward (CF) relaying for the three cases.

1.2 Motivation

State-dependent channels with state information available at the encoders can be used to model many systems, such as information embedding [22-24] which enables encoding a message into a host signal, computer memories with defective cells [25], communication systems with cognitive radios, etc. For the above examples, we are more interested in the communication systems with cognitive radios. In order to improve the frequency spectrum efficiency in wireless systems, some secondary users which are capable of acquiring some knowledge about the primary communication are introduced into an existing primary communication system [26]. Obtaining the knowledge of the primary communication, the secondary users can adapt their coding schemes to mitigate the interference caused by the primary communication. In such models, the channel state can be viewed as the signals of the primary communication and the informed encoders can be viewed as cognitive users [11].

For the state-dependent relay channel with orthogonal components considered in this paper, the channel states S_R and S_D are viewed as the signals of corresponding primary communication; the source and the relay are viewed as the secondary users which are capable of acquiring the channel state information. Thus, the model studied in this paper can be viewed as a secondary relay communication with cognitive source and cognitive relay. We are interested in studying the capacity of this model.

However, it is tedious to characterize the explicit capacity of the relay channels even if the channel is state-independent. The capacity for the state-independent relay channel is only characterized in some special channels, e.g., physically degraded/reversely degraded relay

channel [27], a class of deterministic relay channels [28] and a class of relay channels with orthogonal components [29]. To the best of our knowledge, explicit capacity results for the state-dependent relay channels with channel state information known to a part of the transmitters or all the transmitters were derived mainly in two cases: (i) physically degraded relay channels with state information *causally* known to both the source and the relay and (ii) *Gaussian* physically degraded relay channels with the channel state information *non-causally* known to the source and the relay. However, when the relay channel is corrupted by the channel state, explicit capacity has not been characterized yet, even if the channel state has structure and the structure is known to the source. In this paper, we try to find some capacity results of the state-dependent relay channel with orthogonal components.

1.3 Main contributions and organization of the paper

We investigate a state-dependent relay channel with orthogonal components, where the source communicates with the relay through a channel (say channel 1) orthogonal to another channel (say channel 2) through which the source and the relay communicate with the destination. We assume that channel 1 and channel 2 are affected by two independent channel states S_R and S_D, respectively. The channel state information about S_R and S_D is known to both the source and the relay non-causally or causally. In this setup, the main results of this paper are summarized as follows:

(1) A lower bound on the capacity of the channel is established when the channel state information is known to the source and the relay non-causally. The achievability is based on superposition coding at the source, PDF relaying at the relay, and cooperative GP coding at the source and the relay.

(2) When the channel state information is known to the source and the relay causally, an achievable rate of this channel is derived in a similar way as in the non-causal channel state information case, except that the auxiliary random variables U and U_r are independent of the channel state S_R and S_D.

(3) We show that the exact capacity for the channel with non-causal channel state information at the source and the relay can be characterized if the receiver output Y is a deterministic function of the relay input X_r, the channel state S_D, and one of the source inputs X_D, i.e., $Y = f(X_D, X_r, S)$, and the relay output Y_r is restricted to be controlled by only the source input X_R and the channel state S_R, i.e., the channel from the source to the relay is governed by the conditional probability distribution $P_{Y_r|X_R,S_R}$.

(4) Explicit capacity is also characterized for the Gaussian orthogonal relay channel with additive Gaussian interferences known to the source and the relay non-causally.

(5) For the Gaussian orthogonal relay channel with additive interferences known to the source and the relay causally, the capacity is derived when the power of the relay is sufficiently large.

The rest of the paper is organized as follows. In Section 2, we present the system model and some definitions as well as notations that will be used throughout the paper. Section 3 is devoted to establishing single-letter expressions for the lower bounds on the capacity of the discrete memoryless state-dependent orthogonal relay channel with channel state information known to the source and the relay either non-causally or causally. In Section 4, we show that when the channel state information is known non-causally, $Y = f(X_D, X_r, S_D)$ and the channel from the source to the relay is governed by the conditional probability distribution $P_{Y_r|X_R,S_R}$, the lower bound derived in Section 3 is tight; thus, the capacity is characterized exactly. In Section 5, the results are extended to the Gaussian cases. In Section 6, some numerical results are provided to illustrate the impact of the additive interferences and the role of the relay in information transmission and in cleaning the interference. In Section 7, we conclude this paper.

2 Notations and problem setup

Throughout this paper, random variables will be denoted by capital letters, while deterministic realizations thereof will be denoted by lower case letters. Vectors will be denoted by the boldface letters. The shorthand notation $\mathbf{x}_i^j$ is used to abbreviate $(x_i, x_{i+1}, ..., x_j)$, $\mathbf{x}^i$ is used to abbreviate $(x_1, x_2, ..., x_i)$, and x_i is used to denote the ith element of $\mathbf{x}^n$, where $1 \le i \le j \le n$. The probability law of a random variable X will be denoted by P_X, and the conditional probability distribution of Y given X will be denoted by $P_{Y|X}$. The alphabet of a scalar random variable X will be designated by the corresponding calligraphic letter $\mathcal{X}$. The cardinality of a set $\mathcal{J}$ will be denoted by $|\mathcal{J}|$. $\mathcal{T}_\varepsilon^n(X)$ denotes a set of strongly ε-typical sequences $\mathbf{x}^n \in \mathcal{X}^n$, while $\mathcal{A}_\varepsilon^n(X)$ denotes a set of weakly ε-typical sequences $\mathbf{x}^n \in \mathcal{X}^n$, where $\varepsilon > 0$. $E(\bullet)$ denotes expectation; $I(\bullet;\bullet)$ denotes the mutual information between two random variables. $\mathcal{N}(0, \sigma^2)$ denotes a Gaussian distribution with zero mean and variance σ^2.

As shown in Figure 1, we consider the state-dependent relay channel with orthogonal components denoted by $P_{Y,Y_r|X_R,X_D,X_r,S_D,S_R}$, where $Y \in \mathcal{Y}$ and $Y_r \in \mathcal{Y}_r$ are the channel outputs from the destination and the relay, respectively. $X_R \in \mathcal{X}_R$ and $X_D \in \mathcal{X}_D$ are the orthogonal channel

inputs from the source, while $X_r \in \mathcal{X}_r$ is the channel input from the relay. $S_R \in \mathcal{S}_R$ and $S_D \in \mathcal{S}_D$ denote the random channel states that corrupt channel 1 and channel 2, respectively. The channel states $S_{R,i}$ and $S_{D,i}$ at time instant i are independently drawn from the distribution Q_{S_R} and Q_{S_D}, respectively. The channel state information S_R and S_D is known to both the source and the relay non-causally or causally.

The message W is uniformly distributed over the set $\mathcal{W} = \{1, 2, ..., M\}$. The source transmits a message W to the destination with the help of a relay in n channel uses. Let $\mathbf{X}_R^n = (X_{R,1}, ..., X_{R,n})$, $\mathbf{X}_D^n = (X_{D,1}, ..., X_{D,n})$, and $\mathbf{X}_r^n = (X_{r,1}, ..., X_{r,n})$ be the channel inputs of the source and the relay, respectively. The relay channel is said to be memoryless and to have orthogonal components if

$$
\begin{aligned}
&P(\mathbf{y}_r^n, \mathbf{y}^n | \mathbf{x}_R^n, \mathbf{x}_D^n, \mathbf{x}_r^n, \mathbf{s}_D^n, \mathbf{s}_R^n) \\
&= \prod_{i=1}^n P(y_{r,i} | x_{r,i}, x_{R,i}, s_{R,i}) P(y_i | x_{r,i}, x_{D,i}, s_{D,i})
\end{aligned}
\tag{1}
$$

A (M, n) code for the state-dependent relay channel with channel state information non-causally known to the source and the relay consists of an encoding function at the source

$$
\varphi^n : \quad \{1, 2, ..., M\} \times \mathcal{S}_D^n \times \mathcal{S}_R^n \rightarrow \mathcal{X}_D^n \times \mathcal{X}_R^n
\tag{2}
$$

a sequence of encoding functions at the relay

$$
\varphi_{r,i} : \quad \mathcal{Y}_r^{i-1} \times \mathcal{S}_D^n \times \mathcal{S}_R^n \rightarrow \mathcal{X}_r
\tag{3}
$$

for $i = 1, 2, ..., n$, and a decoding function at the destination

$$
\phi^n : \quad \mathcal{Y}^n \rightarrow \{1, 2, ..., M\}.
$$

The information rate R is defined as $R = \frac{1}{n} \log_2 M$ bits per transmission

An (ε_n, n, R) code for the state-dependent relay channel with orthogonal components and non-causal state information is a code having average probability of error smaller than ε_n, i.e.,

$$
\Pr\left(W \ne \phi^n(y^n)\right) \le \varepsilon_n
$$

The rate R is said to be achievable if there exists a sequence of (ε_n, n, R) codes with $\lim_{n \to \infty} \varepsilon_n = 0$. The capacity of the channel is defined as the supremum of the set of achievable rates.

The definition of an (ε_n, n, R) code for the state-dependent relay channel with orthogonal components

and causal channel state information at the source and the relay is similar to that of the state-dependent relay channel with non-causal state information, except that the encoder consists of a sequence of maps $\{\varphi_i\}_{i=1}^n$, and $\{\varphi_{r,i}\}_{i=1}^n$, where i means the time index. Thus, the encoder mappings in (2) to (3) are replaced by

$$\varphi_i : \{1, 2, ..., M\} \times \mathcal{S}_D^i \times \mathcal{S}_R^i \to \mathcal{X}_D \times \mathcal{X}_R \qquad (4)$$

$$\varphi_{r,i} : \mathcal{Y}_r^{i-1} \times \mathcal{S}_D^i \times \mathcal{S}_R^i \to \mathcal{X}_r, \qquad (5)$$

respectively, where, $i = 1, 2, ..., n$. The definitions of achievable rate and capacity remain the same as in the non-causal state information case.

3 Discrete memoryless case

In this section, it is assumed that the alphabets $\mathcal{X}_D$, $\mathcal{X}_R$, $\mathcal{X}_r$, $\mathcal{Y}_r$, $\mathcal{Y}$, $\mathcal{S}_D$, and $\mathcal{S}_R$ are finite. Lower bounds on the capacity of the channel with non-causal channel state information or causal channel state information are established, respectively. In the proofs of the lower bounds in the discrete memoryless case, strong typicality is used.

3.1 Non-causal channel state information

The following theorem provides a lower bound on the capacity of the state-dependent orthogonal relay channel with channel state information non-causally known to the source and the relay.

Theorem 1 *For the orthogonal relay channel with channel state information non-causally known to both the source and the relay, the following rate is achievable*

$$R \leq \max \min\{I(X_R; Y_r | U_r, X_r, S_R) + I(U; Y | U_r)$$
$$- I(U; S_D | U_r), I(U, U_r; Y) - I(U, U_r; S_D)\}, \qquad (6)$$

where the maximization is over all measures on $\mathcal{S}_D \times \mathcal{S}_R \times \mathcal{X}_r \times \mathcal{U}_r \times \mathcal{X}_R, \times \mathcal{X}_D \times \mathcal{Y}_r \times \mathcal{Y}$ of the form

$$P_{S_R,S_D,X_r,U_r,U,X_R,X_D,Y_r,Y} = Q_{S_R} Q_{S_D} P_{U_r|S_D} P_{X_r|U_r,S_R,S_D}$$
$$\times P_{X_R|U_r,S_R} P_{U,X_D|U_r,S_D}$$
$$\times P_{Y_r|X_R,X_r,S_R} P_{Y|X_D,X_r,S_D} \qquad (7)$$

$U_r \in \mathcal{U}_r$ and $U \in \mathcal{U}$ are auxiliary random variables with

$$|\mathcal{U}_r| \leq |\mathcal{S}_D||\mathcal{S}_R||\mathcal{X}_D||\mathcal{X}_r| + 1 \qquad (8)$$

$$|\mathcal{U}| \leq |\mathcal{S}_D||\mathcal{S}_R||\mathcal{X}_r||\mathcal{X}_D|(|\mathcal{S}_D||\mathcal{S}_R||\mathcal{X}_r||\mathcal{X}_D| + 1) + 1 \qquad (9)$$

Remark 1 *Since the source and the relay know the channel state information non-causally, with PDF relaying, they can transmit the messages to the destination cooperatively with GP coding, namely, cooperative GP coding. The source communicates with the relay treating $\mathbf{s}_R^n$ as a time-sharing sequence for the same reason that the channel state information is known to both the source and the relay non-causally.*

3.1.1 Outline of the proof of Theorem 1

We now give a description of the coding scheme to derive the lower bound in Theorem 1. Detailed error analysis of Theorem 1 is given in Appendix 1. The achievable scheme is based on the combination of superposition coding at the source, PDF relaying at the relay, and cooperative GP coding at the source and the relay.

The message W is divided into two parts $W_D \in [1, 2^{nR_D}]$ and $W_D \in [1, 2^{nR_D}]$. Consider $B + 1$ blocks, each of n symbols. Let $\mathbf{s}_D^n(k)$ and $\mathbf{s}_R^n(k)$ be the state sequences in block k, $k = 1, 2, ..., B + 1$. A sequence of B messages $w(k) \in [1, 2^{nR}]$, with $w(k) = (w_D(k), w_R(k))$, $w_D(k) \in [1, 2^{nR_D}]$, $w_R(k) \in [1, 2^{nR_R}]$, and $R = R_D + R_R$, are sent over the channel in $n(B + 1)$ transmissions. During each of the first B blocks, the source encodes $w_D(k) \in [1, 2^{nR_D}]$ and sends it over the channel. Since both the source and the relay know the channel state sequence $s_R^n(k)$, the source encodes $w_R(k) \in [1, 2^{nR_R}]$ by treating $s_R^n(k)$ as a time-sharing sequence [30]. The message $w_R(k)$ is expressed as a unique set $\{m_{s_R}(k): s_R \in \mathcal{S}_R\}$ with $|\mathcal{S}_R|$ sub-messages. For each $s_R \in \mathcal{S}_R$, every sub-message $m_{s_R}(k)$ in the set is associated with a codeword $\mathbf{x}_R^n(m_{s_R}(k))$ from a corresponding sub-codebook $\mathcal{C}_{s_R}$. The set of codewords $\{\mathbf{x}_R^n(m_{s_R}(k)): s_R \in \mathcal{S}_R\}$ are sent over the channel multiplexed according to the state sequence $\mathbf{s}_R^n(k)$. The relay demultiplexes the received sequence $\mathbf{y}_r^n(k)$ into sub-sequences according to the state sequence $\mathbf{s}_R^n(k)$ and decode each sub-message $m_{s_R}(k)$. Consequently, $w_R(k)$ is decoded at the relay. The coding scheme is illustrated in Figure 2. With PDF relaying, the relay re-encodes $w_R(k)$ and sends it to the destination cooperatively with the source. In the last block $B + 1$, no new message is sent and let $w(B + 1) = (w_D(B + 1), w_R(B + 1)) = (1, 1)$. The average information rate $R(B/(B + 1))$ of the message over $B + 1$ blocks approaches R as $B \to \infty$.

3.1.2 Codebook generation

Fix a measure $P_{S_R,S_D,X_r,U_r,U,X_R,X_D}$ of the form (7).

(i) Generate $2^{n(R_R+R_{r,s})}$ i.i.d. codewords $\{\mathbf{u}_r^n(\tilde{w}_R, j_r)\}$ indexed by $\tilde{w}_R = 1, 2, ..., 2^{nR_R}$, $j_r = 1, 2, ..., 2^{nR_{r,s}}$, each with i.i.d. components drawn according to P_{U_r}.

(ii) For each $\mathbf{u}_r^n(\tilde{w}_R, j_r)$, generate $2^{n(R_D+R_{d,s})}$ i.i.d. codewords $\{\mathbf{u}^n(w_D, j_d | \tilde{w}_R, j_r)\}$ indexed by
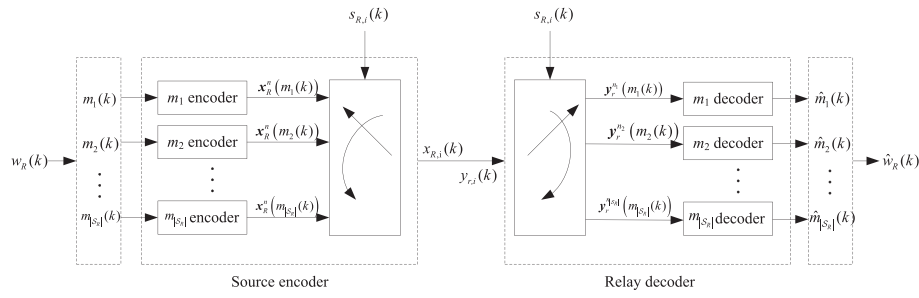
Figure 2 Multiplexed coding and decoding at the source and the relay [30].

$w_D = 1, 2, ..., 2^{nR_D}$, $j_d = 1, 2, ..., 2^{nR_{d,s}}$, each with i.i.d. components drawn according to $P_{U|U_r}$.

(iii) For each $u_r^n(\tilde{w}_R, j_r)$ and for each $s_R \in \mathcal{S}_R$, randomly and independently generate $2^{nR_{s_R}}$ sequences $\left\{ \mathbf{x}_R^n\left(m_{s_R} | s_R, \tilde{w}_R, j_r\right) \right\}$ indexed by $m_{s_R} \in \left[1, 2^{nR_{s_R}}\right]$, each with i.i.d. components according to $P_{X_R|U_r, S_R}$. These sequences constitute the sub-codebook $\mathcal{C}_{s_R}$ $(s_R \in \mathcal{S}_R)$. There are $|\mathcal{S}_R|$ such sub-codebooks for each $u_r^n(\tilde{w}_R, j_r)$. Set $R_R = \sum_{s_R \in \mathcal{S}_R} R_{s_R}$.

3.1.3 Encoding

We pick up the story in block k. Let $w(k) = (w_D(k), w_R(k)) \in \{1, 2, ..., 2^{nR}\}$, where $w_D(k) \in \left\{1, 2, ..., 2^{nR_D}\right\}$, $w_R(k) \in \left\{1, 2, ..., 2^{nR_R}\right\}$, be the new message to be sent from the source node at the beginning of block k. The encoding at the beginning of block k is as follows.

(i) The relay knows $w_R(k-1)$ (this will be justified below) and searches the smallest $j_r(k) \in \left\{1, 2, ..., 2^{nR_{r,s}}\right\}$, such that $u_r^n(w_R(k-1), j_r(k))$ is jointly typical with $s_D^n(k)$. If no such $j_r(k)$ exists, an error is declared and $j_r(k)$ is set to 1. By the covering lemma [31], this error probability tends to 0 as n approaches infinity, if $R_{r,s}$ satisfies

$$R_{r,s} \geq I(U_r; S_D) \tag{10}$$

Given $u_r^n(w_R(k-1), j_r(k))$, $s_R^n(k)$, and $s_D^n(k)$, the relay sends a vector $x_r^n(k)$ with i.i.d. components drawn according to the marginal $P_{X_r|U_r, S_D, S_R}$.

(ii) The source also knows $w_R(k-1)$ and $s_D^n(k)$, thereby knows $u_r^n(w_R(k-1), j_r(k))$. Then, the source searches $j_d(k) \in \{1, 2, ..., 2^{nR_{d,s}}\}$, such that $u^n(w_D(k), j_d(k) | w_R(k-1), j_r(k))$ is jointly typical with $s_D^n(k)$ given $u_r^n(w_R(k-1), j_r(k))$. If no such $j_d(k)$ exists, an error is declared and $j_d(k)$ is set to 1. By the covering lemma [31], this error

probability tends to 0 as n approaches infinity, if $R_{d,s}$ satisfies

$$R_{d,s} \geq I(U; S_D | U_r) \tag{11}$$

Given $u^n(w_D(k), j_d(k) | w_R(k-1), j_r(k))$, $u_r^n(w_R(k-1), j_r(k))$, and $s_D^n(k)$, the source then sends a vector $x_D^n(k)$ with i.i.d. components drawn according to the marginal $P_{X_D|U_r, U, S_D}$.

(iii) Meanwhile, to send a message $w_R(k) \in [1, 2^{nR}]$, express it as a unique set of messages $\{m_{s_R}(k) : s_R \in \mathcal{S}_R\}$. The source, knowing the codeword $u_r^n(w_R(k-1), j_r(k))$, considers the set of codewords $\left\{ x_R^n\left(m_{s_R}(k) | s_R, w_R(k-1), j_r(k)\right) : s_R \in \mathcal{S}_R \right\}$. Store each codeword in a first-in-first-out (FIFO) buffer of length n. A multiplexer is used to choose a symbol at each transmission time $i \in [1, n]$ from one of the FIFO buffers according to the state $s_{R,i}(k)$. Then, the chosen symbol is transmitted.

3.1.4 Decoding

At the end of block k, the relay and the destination observe $\mathbf{y}_r^n(k)$ and $\mathbf{y}^n(k)$, respectively.

(i) Having successfully decoded $w_R(k-1)$ in block $k-1$ and knowing $u_r^n(w_R(k-1), j_r(k))$, $x_r^n(k)$ and $s_R^n(k)$, the relay node estimates $\hat{w}_R(k)$ from $y_r^n(k)$. According to the state sequence $s_R^n(k)$, the relay demultiplexes $y_r^n(k)$, $u_r^n(w_R(k-1), j_r(k))$ and $x_r^n(k)$ into subsequences $\{y_{r,s_R}^{n_{s_R}(k)}(k), s_R \in \mathcal{S}_R\}$, $\{u_{r,s_R}^{n_{s_R}(k)}(w_R(k-1), j_r(k)), s_R \in \mathcal{S}_R\}$ and $\{x_{r,s_R}^{n_{s_R}(k)}(k), s_R \in \mathcal{S}_R\}$, respectively, where $\sum_{s_R \in \mathcal{S}_R} n_{s_R}(k) = n$. Assuming $s_R^n(k) \in T_\varepsilon^n(S_R)$, and thus $n_{s_R}(k) \geq n(1-\varepsilon)p(s_R)$ for all $s_R \in \mathcal{S}_R$, it finds for each $s_R \in \mathcal{S}_R$ a unique $\hat{m}_{s_R}(k)$ such that the codeword sub-sequence $x_R^{n(1-\varepsilon)p(s_R)}\left(\hat{m}_{s_R}(k) | s_R, w_R(k-1), j_r(k)\right)$ is jointly typical with $\mathbf{y}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(k)$, given $u_{r,s_R}^{n(1-\varepsilon)p(s_R)}(w_R(k-1), j_r(k))$ and

$\mathbf{x}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(k)$. By the law of large numbers (LLN) and the packing lemma [31], the probability error of each decoding step approaches 0 as $n \to \infty$ if $R_{s_R} \le p(s_R)I(X_R; Y_r|U_r, X_r, S_R = s_R)$. Therefore, the total probability error in decoding $\hat{w}_R(k)$ approaches 0 for sufficiently large n if the following condition is satisfied:

$$R_R = \sum_{s_R \in \mathcal{S}_R} R_{s_R}$$
$$\le I(X_R; Y_r|U_r, X_r, S_R) \tag{12}$$

(ii) Observing $y^n(k)$, the destination finds a pair $(\hat{w}_R(k-1), \hat{w}_D(k))$ such that

$$(\mathbf{u}^n(\hat{w}_D(k), \hat{j}_d(k)|\hat{w}_R(k-1), \hat{j}_r(k)), \mathbf{u}_r^n(\hat{w}_R(k-1), \hat{j}_r(k)),$$
$$\mathbf{y}^n(k)) \in \mathcal{T}_\varepsilon^n(U, U_r, Y)$$

for some $\hat{j}_d(k) \in \{1, 2, \ldots, 2^{nR_{d,s}}\}$ and $\hat{j}_r(k) \in \{1, 2, \ldots, 2^{nR_{r,s}}\}$. If there is no such pair or it is not unique, an error is declared. By the packing lemma [31], it can be shown that for sufficiently large n, decoding is correct with high probability if

$$R_D + R_{d,s} \le I(U; Y|U_r)$$
$$R_D + R_{d,s} + R_R + R_{r,s} \le I(U, U_r; Y) \tag{13}$$

Combining (11) to (13), $w(k-1) = (w_D(k-1), w_R(k-1))$ is decoded correctly with high probability at the end of block k, if

$$R \le I(X_R; Y_r|U_r, X_r, S_R) + I(U; Y|U_r) - I(U; S_D|U_r)$$
$$R \le I(U, U_r; Y) - I(U, U_r; S_D) \tag{14}$$

The detailed analysis of error probability is shown in Appendix 1.

3.2 Causal channel state information
In many practical communication systems, the state sequences are not known to the encoders in advance. For the case that the channel state information is provided to the source and the relay causally, the capacity is lower bounded as the following theorem.

Theorem 2 *The capacity of the orthogonal relay channel with channel state information causally known to both the source and relay is lower bounded by*

where $U_r \in \mathcal{U}_r$ and $U \in \mathcal{U}$ are auxiliary random variables with

$$|\mathcal{U}_r| \le |\mathcal{S}_D||\mathcal{S}_R||\mathcal{X}_D||\mathcal{X}_r| + 1 \tag{16}$$

$$|\mathcal{U}| \le |\mathcal{S}_D||\mathcal{S}_R||\mathcal{X}_r||\mathcal{X}_D|(|\mathcal{S}_D||\mathcal{S}_R||\mathcal{X}_r||\mathcal{X}_D| + 1) + 1 \tag{17}$$

Remark 2 *The achievable rate region in Theorem 2 is obtained by specializing the expression for the region in Theorem 1 to the case where the auxiliary random variables U and U_r are independent of S_D and S_R. This is similar to the relation between the expression for the capacity of the state-dependent channel with causal channel state information introduced by Shannon [2] and its non-causal counterpart, the Gel'fand-Pinsker channel [3].*

Proof The achievability poof is derived in a similar way as in the non-causal channel state information case except that the auxiliary random variables U and U_r are independent of the channel states S_D and S_R, and the channel inputs of the source and the relay are restricted to the mappings $x_D = f_D(u, s_D)$ and $x_r = f_r(u_r, s_D, s_R)$, respectively, where $f_D(\cdot)$ and $f_r(\cdot)$ are deterministic functions. The details are omitted for brevity.

4 Semi-deterministic orthogonal relay channel with non-causal channel state information
In this section, we show that the lower bound derived in Theorem 1 is tight for a class of semi-deterministic orthogonal relay channel, where, the output Y of the destination is a deterministic function of X_D, X_r and S_D, i.e., $Y = f(X_D, X_r, S_D)$, and the output Y_r of the relay node is controlled only by X_R and S_R, i.e., the channel from the source to the relay is governed by the conditional distribution $P_{Y_r|X_R,S_R}$. This assumption is reasonable in many cases, e.g., when the two orthogonal channels use two different frequency bands, the received signal Y_r at the relay node will not be affected by its input signal X_r. The channel can be expressed as

$$P(y_r, y|x_R, x_D, x_r, s_D, s_R) = P(y_r|x_R, s_R)\mathbf{1}\{y = f(x_D, x_r, s_D)\} \tag{18}$$

where, $f(\cdot)$ is a deterministic function and $\mathbf{1}\{\cdot\}$ denotes the indicator function. The channel state information on S_R and S_D is known to both the source and the relay non-causally. The capacity of this class of semi-deterministic orthogonal relay channel is characterized as shown in the following theorem.

$$C_{CS} \ge \max_{\substack{p(s_D)p(s_R)p(u_r)\,p(x_R|u_r, s_R)\,p(u|u_r, s_D) \\ x_r = f_r(u_r, s_D, s_R), x_D = f_D(u, s_D)}} \min\{I(X_R; Y_r|U_r, X_r, S_R) + I(U; Y|U_r), I(U, U_r; Y)\}, \tag{15}$$

Theorem 3 *The capacity of the channel (18) with the channel state information known non-causally to the source and the relay is characterized as*

$$C = \max \min \{I(X_R; Y_r|S_R) + H(Y|U_r, S_D), H(Y) - I(U_r, Y; S_D)\}, \quad (19)$$

where the maximization is over all measures on $\mathcal{S}_D \times \mathcal{S}_R \times \mathcal{X}_r \times \mathcal{U}_r \times \mathcal{X}_R \times \mathcal{X}_D \times \mathcal{Y}_r \times \mathcal{Y}$ of the form

$$\begin{aligned}
P(s_D, s_R, x_r, u_r, x_R, x_D, y_r, y) \\
= Q(s_D)Q(s_R)P(x_R|s_R)P(u_r, x_r, x_D|s_D) \\
\times P(y_r|x_R, s_R)\mathbf{1}\{y = f(x_D, x_r, s_D)\}
\end{aligned} \quad (20)$$

$U_r \in \mathcal{U}_r$ is an auxiliary random variables with

$$|\mathcal{U}_r| \le |\mathcal{S}_D||\mathcal{S}_R||\mathcal{X}_D||\mathcal{X}_r| + 1 \quad (21)$$

and $\mathbf{1}\{\cdot\}$ denotes the indicator function.

Proof The achievability follows from Theorem 1. First note that the joint distribution of (20) can also be written as

$$\begin{aligned}
P(s_D, s_R, x_r, u_r, x_R, x_D, y_r, y) \\
= Q(s_D)Q(s_R)P(x_R|s_R)P(u_r, y|s_D) \\
\times P(x_r, x_D|s_D, u_r, y)P(y_r|x_R, s_R)
\end{aligned} \quad (22)$$

with additional requirement that

$$y = f(x_D, x_r, s_D). \quad (23)$$

Note that, when $P_{U_r, Y, S_D}(u_r, y, s_D)$ is fixed, all the items on the right-hand side (RHS) of (19) are fixed except for $I(X_R; Y_r|S_R)$, which is independent of $P_{X_r, X_D \mid S_D, U_r, Y}(x_r, x_D|s_D, u_r, y)$. Therefore, the maximization over all joint distributions of the form (20) can be replaced by the maximization only over those distributions, where x_r and x_D are two deterministic functions of (s_D, u_r, y), i.e., of the form

$$\begin{aligned}
P(s_D, s_R, x_r, u_r, x_R, x_D, y_r, y) \\
= Q(s_D)Q(s_R)P(x_R|s_R)P(u_r, y|s_D)\mathbf{1}\{x_r = g_r(u_r, s_D)\} \\
\times \mathbf{1}\{x_D = g_d(y, u_r, s_D)\}P(y_r|x_R, s_R)
\end{aligned} \quad (24)$$

for some mappings g_r: $(u_r, s_D) \to x_r$, g_d: $(y, u_r, s_D) \to x_D$ and subject to (23). Thus, we only have to prove the achievability of the rate that satisfies (19) for some distribution of the form (24).

The achievability follows directly from Theorem 1 by taking $U = Y$ since $Y = f(X_D, X_r, S_D)$, letting X_R be

independent of U_r and X_r considering the fact that Y_r is only determined by X_R and S_R, and by setting $x_r = g_r(u_r, s_D)$, $x_D = g_d(y, u_r, s_D)$. Note that with these choices of the random variables, if we chose stochastic kernels $P_{X_R|S_R}$ and $P_{U_r, Y|S_D}$, two deterministic mappings g_r:$(u_r, s_D) \to x_r$ and g_d:$(y, u_r, s_D) \to x_D$, combined with $Q_{S_D}Q_{S_R}$ and the channel law, the joint distribution (24) for which (23) is satisfied will be determined.

The proof of the converse is as follows.

Consider an (ε_n, n, R) code with an average error probability $P_e^{(n)} \le \varepsilon_n$. By Fano's inequality, we have

$$H(W|Y^n) \le nRP_e^{(n)} + 1 = n\delta_n \quad (25)$$

where $\delta_n \to 0$ as $n \to +\infty$. Thus,

$$nR = H(W) \le I(W; \mathbf{Y}^n) + n\delta_n \quad (26)$$

Defining the auxiliary random variable $\bar{U}_{r,i} = (\mathbf{Y}^{i-1}, \mathbf{S}_{D,i+1}^n)$, we have

$$\begin{aligned}
I(W; \mathbf{Y}^n) &\le I(W; \mathbf{Y}^n, \mathbf{Y}_r^n) \\
&\le I(W; \mathbf{Y}^n, \mathbf{Y}_r^n | \mathbf{S}_D^n, \mathbf{S}_R^n) \\
&= \sum_i I(W; Y_i, Y_{r,i} | \mathbf{Y}^{i-1}, \mathbf{Y}_r^{i-1}, \mathbf{S}_D^n, \mathbf{S}_R^n) \\
&= \sum_i I(W; Y_{r,i} | \mathbf{Y}^{i-1}, \mathbf{Y}_r^{i-1}, \mathbf{S}_D^n, \mathbf{S}_R^n) \\
&\quad + \sum_i I(W; Y_i | \mathbf{Y}^{i-1}, \mathbf{Y}_r^{i-1}, Y_{r,i}, \mathbf{S}_D^n, \mathbf{S}_R^n),
\end{aligned} \quad (27)$$

where the second inequality follows from the fact that $\mathbf{S}_D^n$ and $\mathbf{S}_R^n$ are independent of W.

Calculate the two terms in (27) separately as follows:

$$\begin{aligned}
\sum_i & I(W; Y_{r,i} | \mathbf{Y}^{i-1}, \mathbf{Y}_r^{i-1}, \mathbf{S}_D^n, \mathbf{S}_R^n) \\
&= \sum_i H(Y_{r,i} | \mathbf{Y}^{i-1}, \mathbf{Y}_r^{i-1}, \mathbf{S}_D^n, \mathbf{S}_R^n) \\
&\quad - H(Y_{r,i} | \mathbf{Y}^{i-1}, \mathbf{Y}_r^{i-1}, \mathbf{S}_D^n, \mathbf{S}_R^n, W) \\
&\overset{(a)}{=} \sum_i H(Y_{r,i} | \mathbf{Y}^{i-1}, \mathbf{Y}_r^{i-1}, \mathbf{S}_D^n, \mathbf{S}_R^n) \\
&\quad - H(Y_{r,i} | \mathbf{Y}^{i-1}, \mathbf{Y}_r^{i-1}, \mathbf{S}_D^n, \mathbf{S}_R^n, W, X_{R,i}) \\
&\overset{(b)}{\le} \sum_i H(Y_{r,i} | S_{R,i}) - H(Y_{r,i} | S_{R,i}, X_{R,i}) \\
&= \sum_i I(X_{R,i}; Y_{r,i} | S_{R,i}),
\end{aligned} \quad (28)$$

where (a) holds since $X_{R,i}$ is a function of $(W, \mathbf{S}_D^n, \mathbf{S}_R^n)$; (b) follows from the fact that conditioning reduces

entropy and the Markov chain $(\mathbf{Y}^{i-1}, \mathbf{Y}_r^{i-1}, \mathbf{S}_D^n, \mathbf{S}_R^{i-1}, \mathbf{S}_{R,i+1}^n, W) \leftrightarrow (X_{R,i}, S_{R,i}) \leftrightarrow Y_{r,i}$.

$$
\begin{aligned}
\sum_i & I(W; \ Y_i | \ \mathbf{Y}^{i-1}, \mathbf{Y}_r^{i-1}, Y_{r,i}, \mathbf{S}_D^n, \mathbf{S}_R^n) \\
&= \sum_i H(Y_i | \mathbf{Y}^{i-1}, \mathbf{Y}_r^{i-1}, Y_{r,i}, \mathbf{S}_D^n, \mathbf{S}_R^n) \\
&\quad - H(Y_i | \mathbf{Y}^{i-1}, \mathbf{Y}_r^{i-1}, Y_{r,i}, \mathbf{S}_D^n, \mathbf{S}_R^n, W) \\
&= \sum_i H(Y_i | \mathbf{Y}^{i-1}, \mathbf{Y}_r^{i-1}, Y_{r,i}, \mathbf{S}_D^n, \mathbf{S}_R^n) \\
&\quad - H(Y_i | \mathbf{Y}^{i-1}, \mathbf{Y}_r^{i-1}, Y_{r,i}, \mathbf{S}_D^n, \mathbf{S}_R^n, W, X_{D,i}, X_{r,i}) \\
&\overset{(a)}{=} \sum_i H(Y_i | \mathbf{Y}^{i-1}, \mathbf{Y}_r^{i-1}, Y_{r,i}, \mathbf{S}_D^n, \mathbf{S}_R^n) \\
&\overset{(b)}{\leq} \sum_i H(Y_i | \mathbf{Y}^{i-1}, \mathbf{S}_{D,i+1}^n, S_{D,i}) \\
&= \sum_i H(Y_i | \ \bar{U}_{r,i}, S_{D,i}),
\end{aligned}
$$
$$(29)$$

where (a) holds since $X_{r,i}$ is a function of $(\mathbf{Y}_r^{i-1}, \mathbf{S}_D^n, \mathbf{S}_R^n)$; (b) follows from the fact that conditioning reduces entropy.

From (26) to (29), we have

$$
R \leq \frac{1}{n} \left(\sum_i I(X_{R,i}; Y_{r,i} | \ S_{R,i}) + H(Y_i | \ U_{r,i}, S_{R,i}) \right) + \delta_n
$$
$$(30)$$

The proof of the bound $I(W; \mathbf{Y}^n)$ given in the second term in (19) is as follows:

$$
\begin{aligned}
I(W; \mathbf{Y}^n) &= \sum_i I(W; Y_i | \mathbf{Y}^{i-1}) \\
&\leq \sum_i I(W, \mathbf{Y}^{i-1}; Y_i) \\
&= \sum_i I\left(W, \mathbf{Y}^{i-1}, \mathbf{S}_{D,i+1}^n; Y_i\right) - I(\mathbf{S}_{D,i+1}^n; Y_i | W, \mathbf{Y}^{i-1}) \\
&\overset{(a)}{=} \sum_i I\left(W, \mathbf{Y}^{i-1}, \mathbf{S}_{D,i+1}^n; Y_i\right) - I(\mathbf{Y}^{i-1}; S_{D,i} | W, \mathbf{S}_{D,i+1}^n) \\
&\overset{(b)}{=} \sum_i H(Y_i) - H(Y_i | W, \mathbf{Y}^{i-1}, \mathbf{S}_{D,i+1}^n) \\
&\quad - I\left(W, \mathbf{Y}^{i-1}, \mathbf{S}_{D,i+1}^n; S_{D,i}\right) \\
&= \sum_i H(Y_i) - H(Y_i | W, \mathbf{Y}^{i-1}, \mathbf{S}_{D,i+1}^n) \\
&\quad - \sum_i \left(I\left(W, \mathbf{Y}^{i-1}, \mathbf{S}_{D,i+1}^n, Y_i; S_{D,i}\right) \right. \\
&\quad \left. - I(Y_i; S_{D,i} | W, \mathbf{Y}^{i-1}, \mathbf{S}_{D,i+1}^n) \right) \\
&\overset{(c)}{\leq} \sum_i H(Y_i) - I\left(W, \mathbf{Y}^{i-1}, \mathbf{S}_{D,i+1}^n, Y_i; S_{D,i}\right) \\
&\leq \sum_i H(Y_i) - I\left(\mathbf{Y}^{i-1}, \mathbf{S}_{D,i+1}^n, Y_i; S_{D,i}\right) \\
&= \sum_i H(Y_i) - I\left(\bar{U}_{r,i}, Y_i; S_{D,i}\right),
\end{aligned}
$$
$$(31)$$

where (a) holds due to Csiszar and Korner's sum identity; (b) follows since $S_{D,i}$ is independent of $(W, \mathbf{S}_{D,i+1}^n)$, and (c) follows from the fact that $H(Y_i | W, \mathbf{Y}^{i-1}, \mathbf{S}_{D,i+1}^n) \geq I(Y_i; S_{D,i} | W, \mathbf{Y}^{i-1}, \mathbf{S}_{D,i+1}^n)$.

By (26) and (31),

$$
R \leq \frac{1}{n} \sum_i \left(H(Y_i) - I(\bar{U}_{r,i}, Y_i; S_{D,i}) \right) + \delta_n
$$
$$(32)$$

From the above, we have

$$
\begin{aligned}
R &\leq \frac{1}{n} \left(\sum_i I(X_{R,i}; Y_{r,i} | S_{R,i}) + \sum_i H(Y_i | \bar{U}_{r,i}, S_{D,i}) \right) + \delta_n \\
R &\leq \frac{1}{n} \sum_i \left(H(Y_i) - I(\bar{U}_{r,i}, Y_i; S_{D,i}) \right) + \delta_n
\end{aligned}
$$
$$(33)$$

Introduce a time-sharing random variable T, which is uniformly distributed over $\{1, 2, \ldots, n\}$ and denote the collection of random variables

$$
(X_R, X_r, Y_r, Y, \bar{U}_r, S_D, S_R) = (X_{R,T}, X_{r,T}, Y_{r,T}, Y_T, \bar{U}_{r,T}, S_{D,T}, S_{R,T}).
$$

Considering the first bound in (33), we have

$$
\begin{aligned}
\frac{1}{n} &\left(\sum_i I(X_{R,i}; Y_{r,i} | S_{R,i}) + \sum_i H(Y_i | \bar{U}_{r,i}, S_{D,i}) \right) \\
&= I(X_R; Y_r | S_R, T) + H(Y | \bar{U}_r, S_D, T) \\
&= H(Y_r | S_R, T) - H(Y_r | X_R, S_R, T) + H(Y | \bar{U}_r, S_D, T) \\
&\leq I(X_R; Y_r | X_r, S_R) + H(Y | \bar{U}_r, S_D, T),
\end{aligned}
$$
$$(34)$$

where the last step follows from the fact that T is independent of all the other variables and the Markov chain $T \leftrightarrow (X_R, S_R) \leftrightarrow Y_r$.

Similarly, considering the second bound in (33), we have

$$
\begin{aligned}
\frac{1}{n} \sum_i &\left(H(Y_i) - I(\bar{U}_{r,i}, Y_i; S_{D,i}) \right) \\
&= H(Y|T) - I(\bar{U}_r, Y; S_D | T) \\
&\leq H(Y) - I(\bar{U}_r, T, Y; S_D) + I(T; S_D) \\
&= H(Y) - I(\bar{U}_r, T, Y; S_D)
\end{aligned}
$$
$$(35)$$

Defining $U_r = (\bar{U}_r, T)$, we get

$$
\begin{aligned}
R &\leq I(X_R; Y_r | S_R) + H(Y | U_r, S_D) + \delta_n \\
R &\leq H(Y) - I(U_r, Y; S_D) + \delta_n
\end{aligned}
$$
$$(36)$$

Therefore, for a given sequence of (ε_n, n, R) code with ε_n going to zero as n goes to infinity, there exists a measure of the form $P_{S_D S_R, X_r, X_R, X_D} = Q_{S_D} Q_{S_R} P_{X_r | S_D S_R} P_{X_R, X_D | X_r, S_D, S_R}$, such that the rate R essentially satisfies (19).

Considering the facts that $I(X_R; Y_r | S_R)$ is determined by the joint distribution P_{X_R, S_R, Y_r} and the other three

items on the RHS of (19) is independent of P_{X_R,S_R,Y_r}, the maximum in (19) taken over all joint probability mass functions $P_{S_D,S_R,X_r,U_r,X_R,X_D,Y_r,Y}$ is equivalent to that taken over all joint probability mass functions of the form

$$P(s_D, s_R, x_r, u_r, x_R, x_D, y_r, y)$$
$$= Q(s_D)Q(s_R)P(x_R|s_R)P(u_r, x_r, x_D|s_D)$$
$$\times P(y_r|x_R, s_R)\mathbf{1}\{y = f(x_D, x_r, s_D)\}$$

The bound of the cardinality of $\mathcal{U}_r$ can be proven in a similar way as that proven in Theorem 1. It is omitted here for brevity.

This concludes the proof.

5 Memoryless Gaussian case

In this section, we study a state-dependent Gaussian relay channel with orthogonal components in which the channel states and the noise are additive and Gaussian. As shown in Figure 3, we consider the state-dependent Gaussian orthogonal relay channel, where channel 1 (dashed line) uses a different frequency band as compared to that used by channel 2 (solid line). The two orthogonal channels, channel 1 and channel 2, are corrupted by two independent additive Gaussian interferences S_R and S_D, respectively, which are known to the source and the relay. The channel can be described as

$$Y_r = X_R + S_R + Z_r \tag{37}$$

$$Y = X_D + X_r + S_D + Z_d \tag{38}$$

where Y_r and Y are the channel outputs of the relay and the destination, respectively; the Gaussian i.i.d. random variables (X_R, X_D) and X_r are channel inputs from the source and the relay with the average power constraints $E(X_R^2) + E(X_D^2) \leq P$ and $E(X_r^2) \leq \gamma P$. The additive interferences S_R, S_D and the noises Z_r, Z_d are assumed to be zero-mean Gaussian i.i.d. with $E(S_R^2) = Q_R$, $E(S_D^2) = Q_D$

and $E(Z_r^2) = E(Z_d^2) = N$. Further, we assume that S_R, S_D, Z_r, and Z_d are independent mutually. As in the discrete memoryless case, we will discuss the capacity of the channel when the additive interference sequences are known to the source and the relay non-causally and causally, respectively.

5.1 Channel state information non-causally known to the source and the relay

For the channel shown in Figure 3, when the channel state information is known non-causally to the source and the relay, using cooperative DPC, the capacity is characterized as in the following theorem.

Theorem 4 *The capacity of the Gaussian orthogonal relay channel with the channel state information non-causally known to both the source and the relay is given by*

$$C(P, \gamma P) = \max_{0 \leq \beta, \ \rho \leq 1} \min$$
$$\times \left\{ \mathcal{C}\left(\frac{\bar{\beta}P}{N}\right) + \mathcal{C}\left(\frac{\beta(1-\rho^2)P}{N}\right), \mathcal{C}\left(\frac{(\beta+\gamma+2\rho\sqrt{\beta\gamma})P}{N}\right) \right\}, \tag{39}$$

where $\mathcal{C}(x) = \frac{1}{2}\log_2(1+x)$ and $\bar{\beta} = 1-\beta$.

Remark 3 *As in many other dirty paper channels with channel state information known non-causally at the encoders, with dirty paper coding, the capacity of the channel considered here is as same as that of the state-independent relay channel with orthogonal components. In fact, (39) also characterizes the capacity of the state-independent Gaussian orthogonal relay channel. Therefore, no matter the channel state information is either causally or non-causally known to the source and the relay, (39) serves as an upper bound on the capacity of the channel shown in Figure 3.*

Proof We only need to prove the achievability of (39) since the expression in (39) characterizes the capacity of the state-independent orthogonal relay channel [29] which obviously serves as an upper bound of the channel in this paper.
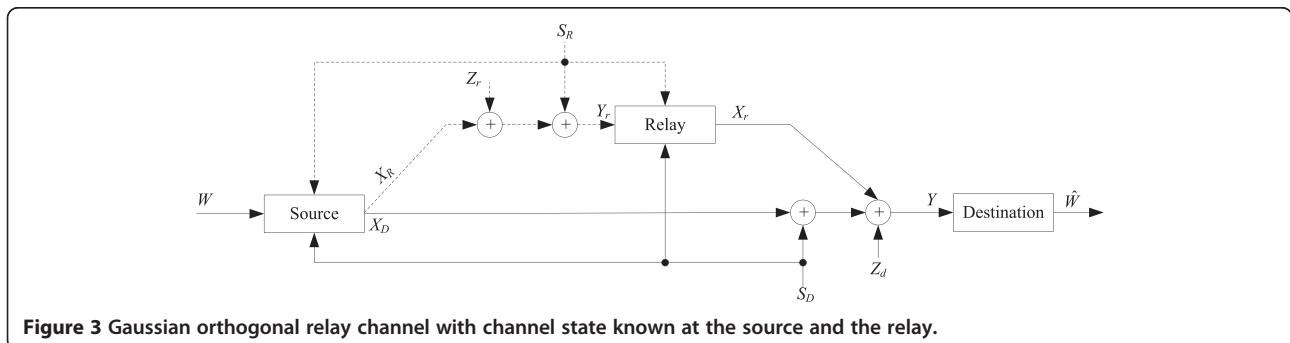

Figure 3 Gaussian orthogonal relay channel with channel state known at the source and the relay.

For the channel given by (37) and (38), we evaluate the achievable rate in (6) with the choice of the jointly Gaussian random variables U, U_r, S_R, S_D, X_R, X_D, and X_R given by

$$U_0 = X_{D,0} + \alpha(1-\alpha_r)S_D \tag{40}$$

$$U_r = \left(1 + \rho\sqrt{\beta/\gamma}\right)X_r + \alpha_r S_D \tag{41}$$

$$U = U_0 + \frac{\rho\sqrt{\beta/\gamma}}{\left(1 + \rho\sqrt{\beta/\gamma}\right)}U_r \tag{42}$$

$$X_D = X_{D0} + \rho\sqrt{\beta/\gamma}X_r, \tag{43}$$

where $E(X_D^2) = \beta P$, $E(X_R^2) = \bar{\beta}P$, $E(X_r^2) = \gamma P$, $E(X_r X_D) = \rho\sqrt{\beta\gamma}P$ and $X_{D,0} \sim \mathcal{N}(0, (1-\rho^2)\beta P)$ is independent of X_r. The parameter β is the ratio of the source power allocated to X_D, while $\bar{\beta} = 1-\beta$ is the ratio of the source power allocated to X_R. The parameter ρ is the correlation coefficient between X_r and X_D. With the above definitions of the random variables, it is straightforward to show the achievable rate in (39). The calculation is straightforward, thus omitted for brevity.

However, the calculation above is somewhat algebraic. Proceeding similarly to Costa's dirty paper coding, we extend the result in Theorem 1 for the discrete memoryless (DM) case to memoryless channels with discrete time and continuous alphabets by standard arguments [32]. An alternative proof is outlined in Appendix 2.

5.2 Channel state information known at the source and the relay causally

When the channel state information is known to the source and the relay causally, the capacity is not characterized in general. The following theorem gives a lower bound on the capacity.

Theorem 5 *For the Gaussian orthogonal relay channel with the channel state information causally known to the source and the relay, the following rate is achievable:*

where $\mathcal{C}(x) = \frac{1}{2}\log_2(1+x)$ *and* $\bar{\beta} = 1-\beta$.

Remark 4 *Since the interference S_R is additive and known to both the source and the relay, the relay can remove S_R completely before decoding the message from the source. Actually, the interference S_R does not affect the achievable rate.*

Remark 5 *The source and the relay expend parts $\rho_{d,s}^2 \beta P$ and $\rho_{r,s}^2 \gamma P$ of their power respectively to clean S_D from the channel and use the remaining power for cooperative information transmission. It is different from many other dirty paper channels with non-causal channel state information at the transmitters where the channel states can be completely cleaned by choosing appropriate auxiliary random variables, e.g., by dirty paper coding. If $Q_D = 0$, the entire power of the source and the relay will be used for information transmission, i.e., $\rho_{r,s} = \rho_{d,s} = 0$. This reduces to the capacity of the state-independent relay channel with orthogonal components as shown in [29] since S_R does not affect the achievable rate.*

Proof The result in Theorem 2 for the discrete memoryless case can be extended to memoryless channels with discrete time and continuous alphabets using standard techniques [32]. The proof follows through evaluation of the lower bound of Theorem 2 using the following jointly Gaussian input distribution. Fix $0 \le \beta \le 1$, $-1 \le \rho_{d,s}, \rho_{r,s}, \rho_{d,r} \le 1$ and $\bar{\beta} = 1-\beta$. Let $X_R \sim \mathcal{N}(0, \bar{\beta}P)$, $U_r \sim \mathcal{N}(0, (1-\rho_{r,s}^2)\gamma P)$, $U' \sim \mathcal{N}(0, (1-\rho_{d,s}^2)(1-\rho_{d,r}^2)\beta P)$, where U_r and U' are independent. Let $U = \rho_{d,r}\sqrt{\frac{(1-\rho_{d,s}^2)\beta P}{(1-\rho_{r,s}^2)\gamma P}}U_r + U'$. We define $X_r = U_r + \rho_{r,s}\sqrt{\gamma P/Q_D}S_D$ and $X_D = U + \rho_{d,s}\sqrt{\beta P/Q_D}S_D$. With these definitions, it can be easily verified that $U \sim \mathcal{N}(0, (1-\rho_{d,s}^2)\beta P)$, $X_r \sim \mathcal{N}(0, \gamma P)$, and $X_D \sim \mathcal{N}(0, \beta P)$. Note that U, U_r, and U' are independent of S_D. Obviously, from these definitions, it is evident that $E(X_R^2) + E(X_D^2) \le P$ and $E(X_r^2) \le \gamma P$. Through straightforward algebra, it can be shown that the evaluation of the lower bound in Theorem 2 using the above choices gives the lower bound in Theorem 5. The computation details are omitted here for brevity.

We next characterize the capacity of the state-dependent Gaussian orthogonal relay channel with causal

$$R(P, \gamma P) \le \max_{\substack{0 \le \beta \le 1 \\ -1 \le \rho_{d,s}, \rho_{r,s}, \rho_{d,r} \le 1}} \min\left\{ \mathcal{C}\left(\frac{\bar{\beta}P}{N}\right) + \mathcal{C}\left(\frac{\left(1-\rho_{d,s}^2\right)\left(1-\rho_{d,r}^2\right)\beta P}{\left(\sqrt{Q_D} + \rho_{d,s}\sqrt{\beta P} + \rho_{r,s}\sqrt{\gamma P}\right)^2 + N}\right), \right.$$
$$\left. \mathcal{C}\left(\frac{\left(1-\rho_{d,s}^2\right)\beta P + \left(1-\rho_{r,s}^2\right)\gamma P + 2\rho_{d,r}\sqrt{\left(1-\rho_{d,s}^2\right)\left(1-\rho_{r,s}^2\right)\beta\gamma P}}{\left(\sqrt{Q_D} + \rho_{d,s}\sqrt{\beta P} + \rho_{r,s}\sqrt{\gamma P}\right)^2 + N}\right)\right\} \tag{44}$$

channel state information when the power of the relay is sufficiently large. As shown in Theorem 5, a part of the relay's power is used to clean the interference S_D. When the power of the relay is sufficiently large, the interference S_D can be cleaned completely and the capacity of the channel can be determined as shown in the following theorem.

Theorem 6 *For the Gaussian orthogonal relay channel with the additive interference sequences known at the source and the relay causally, when the power of the relay satisfies*

$$\gamma P \geq \left(\frac{P}{4N} + \frac{N}{4P} + \frac{Q_\mathrm{D}}{P} + \frac{1}{2}\right)P \tag{45}$$

the capacity can be characterized as

$$C(P, \gamma P) = \max_{0 \leq \beta \leq 1} C\left(\frac{\bar{\beta}P}{N}\right) + C\left(\frac{\beta P}{N}\right) \tag{46}$$

Remark 6 *When the power of the relay is sufficiently large such that the interference S_D is completely cleaned by the relay using part of its power and its remaining power is sufficiently large such that the relay-destination link does not constrain the achievable rate, the message sent from the source is split into two parts: one part is sent directly to the destination through a point-to-point source-destination channel and the other is sent to the destination through a two-hop source-relay-destination channel with DF relaying. The two parts are sent independently, and the rate can be expressed as the sum of the rates of the source-destination channel and the two-hop source-relay-destination channel (the rate of the later is constrained by the source-relay link).*

Proof Define $\boldsymbol{\rho} = (\rho_\mathrm{d,s}, \rho_\mathrm{d,r}, \rho_\mathrm{r,s})$. We denote the two terms on the RHS in (44) as

$$R_1(\beta, \boldsymbol{\rho}) = C\left(\frac{\bar{\beta}P}{N}\right) + C\left(\frac{\left(1-\rho_\mathrm{d,s}^2\right)\left(1-\rho_\mathrm{d,r}^2\right)\beta P}{\left(\sqrt{Q_\mathrm{D}} + \rho_\mathrm{d,s}\sqrt{\beta P} + \rho_\mathrm{r,s}\sqrt{\gamma P}\right)^2 + N}\right) \tag{47}$$

$$R_2(\beta, \boldsymbol{\rho}) = C\left(\frac{\left(1-\rho_\mathrm{d,s}^2\right)\beta P + \left(1-\rho_\mathrm{r,s}^2\right)\gamma P + 2\rho_\mathrm{d,r}\sqrt{\left(1-\rho_\mathrm{d,s}^2\right)\left(1-\rho_\mathrm{r,s}^2\right)\beta \gamma P}}{\left(\sqrt{Q_D} + \rho_\mathrm{d,s}\sqrt{\beta P} + \rho_\mathrm{r,s}\sqrt{\gamma P}\right)^2 + N}\right) \tag{48}$$

Let $R(\beta, \boldsymbol{\rho}) = \min\{R_1(\beta, \boldsymbol{\rho}), R_2(\beta, \boldsymbol{\rho})\}$. Then, $R(P, \gamma P) \leq$

$$\max_{\substack{0 \leq \beta \leq 1 \\ -1 \leq \rho_{d,s}, \rho_{r,s}, \rho_{d,r} \leq 1}} R(\beta, \boldsymbol{\rho}) \text{ is achievable.}$$

It is easy to verify that if $\gamma P \geq Q_\mathrm{D}$, for any fixed β, $R_1(\beta, \boldsymbol{\rho})$ is the maximal when $\boldsymbol{\rho} = \boldsymbol{\rho}_1^* = \left(0, 0, -\sqrt{Q_\mathrm{D}/(\gamma P)}\right)$.

Denote the maximum of $R_1(\beta, \boldsymbol{\rho})$ as $R_1^*(\beta)$. Therefore, we have

$$R_1^*(\beta) = R_1(\beta, \boldsymbol{\rho}_1^*) = C\left(\frac{\bar{\beta}P}{N}\right) + C\left(\frac{\beta P}{N}\right) \tag{49}$$

$$R_2(\beta, \boldsymbol{\rho}_1^*) = C\left(\frac{\beta P + \gamma P - Q_\mathrm{D}}{N}\right) \tag{50}$$

Next, we will show the condition under which $R_2(\beta, \boldsymbol{\rho}_1^*)$ is always larger than $R_1^*(\beta)$ for any β. Let

$$1 + \frac{\beta P + \gamma P - Q_\mathrm{D}}{N} \geq \left(1 + \frac{\bar{\beta}P}{N}\right)\left(1 + \frac{\beta P}{N}\right) \tag{51}$$

The inequality in (51) is equivalent to

$$P^2\beta^2 - \left(P^2 - PN\right)\beta + \gamma PN - Q_\mathrm{D}N - PN \geq 0 \tag{52}$$

It is easy to show that if

$$\gamma \geq \frac{P}{4N} + \frac{N}{4P} + \frac{Q_\mathrm{D}}{P} + \frac{1}{2} \tag{53}$$

the inequality in (52) holds for any β. Thus, if $\gamma \geq \frac{P}{4N} + \frac{N}{4P} + \frac{Q_\mathrm{D}}{P} + \frac{1}{2}$, the following inequality is always satisfied for any β

$$R_2(\beta, \boldsymbol{\rho}_1^*) \geq R_1^*(\beta) \tag{54}$$

For any β, we have

$$R(\beta, \boldsymbol{\rho}_1^*) = \min\left\{R_1(\beta, \boldsymbol{\rho}_1^*), R_2(\beta, \boldsymbol{\rho}_1^*)\right\} = R_1^*(\beta) \tag{55}$$

Therefore,

$$R(P, \gamma P) \leq \max_{0 \leq \beta \leq 1} R_1^*(\beta) = \max_{0 \leq \beta \leq 1} C\left(\frac{\bar{\beta}P}{N}\right) + C\left(\frac{\beta P}{N}\right) \tag{56}$$

is achievable.

As mentioned in Remark 3, (39) serves as an upper bound on the capacity of the channel considered here. The converse proof follows by proving that (46) matches the upper bound in (39) if the condition in (45) is satisfied. We denote the two terms on the RHS in (39) as

$$C_1(\beta, \rho) = C\left(\frac{\bar{\beta}P}{N}\right) + C\left(\frac{\beta(1-\rho^2)P}{N}\right) \tag{57}$$

$$C_2(\beta, \rho) = C\left(\frac{(\beta + \gamma + 2\rho\sqrt{\beta\gamma})P}{N}\right) \tag{58}$$

Let $C(\beta, \rho) = \min\{C_1(\beta, \rho), C_2(\beta, \rho)\}$. Similar to steps from (47) to (55), it is easy to prove that for any β if $\gamma \geq \frac{P}{4N} + \frac{N}{4P} + \frac{1}{2}$

$$C(\beta, 0) = \min\{C_1(\beta, 0), C_2(\beta, 0)\} = C_1(\beta, 0)$$
$$= \mathcal{C}\left(\frac{\bar{\beta}P}{N}\right) + \mathcal{C}\left(\frac{\beta P}{N}\right) \tag{59}$$

Next, we have to prove that for any β, under the condition $\gamma \geq \frac{P}{4N} + \frac{N}{4P} + \frac{1}{2}$, $C(\beta, \rho)$ is maximized when $\rho = 0$. Denote the maximal of $C(\beta, \rho)$ as $C^*(\beta)$, i.e., $C^*(\beta) = \max_{-1 \leq \rho \leq 1} C(\beta, \rho)$. This can be proven by contradiction.

Assume that $C(\beta, \rho)$ is maximized when $\rho = \rho'$ $(\rho' \neq 0)$. By (59), we get

$$C(\beta, \rho') \geq C(\beta, 0) = C_1(\beta, 0) \tag{60}$$

However, we have

$$C(\beta, \rho') = \min\{C_1(\beta, \rho'), C_2(\beta, \rho')\} \leq C_1(\beta, \rho') \tag{61}$$

From (57), it is easy to verify that for any β, $C_1(\beta, \rho)$ is maximized when $\rho = 0$. Thus, (60) and (61) are contradictory. This proves that for any β, $C^*(\beta) = C_1(\beta, 0) = \mathcal{C}$

$\left(\frac{\bar{\beta}P}{N}\right) + \mathcal{C}\left(\frac{\beta P}{N}\right)$. Thus, the maximization problem in (39) is equivalent to the following maximization problem

$$C(P, \gamma P) \leq \max_{0 \leq \beta \leq 1} \mathcal{C}\left(\frac{\bar{\beta}P}{N}\right) + \mathcal{C}\left(\frac{\beta P}{N}\right) \tag{62}$$

This completes the proof.

6 Numerical examples

In this section, we provide some numerical examples for the achievable rate in Theorem 5. With these examples, we will show the impact of the channel state and the role of the relay in information transmission and in cleaning the channel state.

For $\gamma = 1$, Figure 4 shows a comparison of the capacity of the state-independent $(Q_R = Q_D = 0)$ relay channel with orthogonal components and the achievable rate derived in Theorem 5. Obviously, the larger the power of the additive interference, more power of the source and the relay will be used to clean the interference; this results in a lower achievable rate. As the power values (P) of the source and relay increase, a larger amount of interference can be cleaned, leaving more power for information transmission. Consequently, the achievable rate will approach the capacity of the state-independent relay channel with orthogonal components as P increases. This can also be verified from (44) such that if
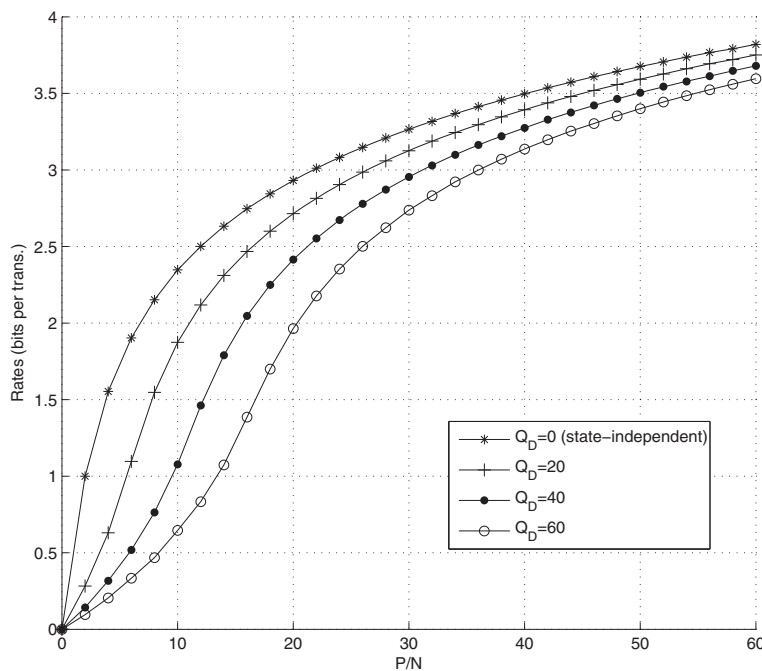

Figure 4 Achievable rates vs. SNR under different power values of the interference.

$P \gg Q_D$, the impact of the additive interference S_D will be negligible with respect to P. The maximization problem of (44) is approximate to that of (39) by taking $\rho_{r,s} \to 0$ and $\rho_{d,s} \to 0$.

For $P/N = 30$, Figure 5 shows the role of the relay in cleaning the channel state. As the power of the relay increases, the achievable rate increases. In particular, when the power of the relay is sufficiently large such that the channel state can be cleaned completely and the relay-destination link does not become the bottleneck for the achievable rate, the achievable rate matches the upper bound. This has been proven in Theorem 6. Figure 5 vividly illustrates this result.

7 Conclusions

In this paper, we consider a state-dependent relay channel with orthogonal channels from the source to the relay and from the source and the relay to the destination. The orthogonal channels are affected by two independent channel states, respectively, and the channel state information is known to both the source and the relay either non-causally or causally. In the non-causal state information case, the lower bound on the capacity of the channel is established with superposition coding at the source, PDF relaying at the relay and cooperative GP coding at the source and the relay. We further show that if the output of the destination Y is a deterministic function of the relay input X_r, the channel state S_D and one of the source inputs X_D, i.e., $Y = f(X_D, X_r, S_D)$ and the relay output Y_r is restricted to be controlled by only the source input X_R and the channel state S_R, the lower

bound is tight, and the capacity can be characterized exactly. As for the causal channel state information case, the lower bound on the capacity is also derived. The expression for the achievable rate in the causal state information case can be interpreted as a special case of that for the achievable rate in the non-causal state information case, where the auxiliary random variables U and U_r are independent of S_R and S_D. This is similar to the relation between the expression for the capacity of the state-dependent channel with causal channel state information introduced by Shannon [2] and its non-causal counterpart, the Gel'fand-Pinsker channel [3].

Further, we investigate the Gaussian state-dependent relay channel with orthogonal components, modeling the channel states as additive Gaussian interferences. Capacity is characterized when the additive interference sequences are known non-causally. The expression for the capacity is the same as that for the capacity of the state-independent relay channel with orthogonal components. This observation is similar to the results for the multiple user state-dependent channels shown in [6]. When the state information is known causally, however, the capacity is not characterized in general. In this case, with carefully chosen auxiliary random variables, achievable rate is derived. It is shown that when the power of the relay is sufficiently large, the capacity can be characterized exactly. Finally, two numerical examples are given to illustrate the impact of the channel state and the role of the relay in information transmission and in cleaning the state. The simulation results show that the larger the power of the additive interference, the more
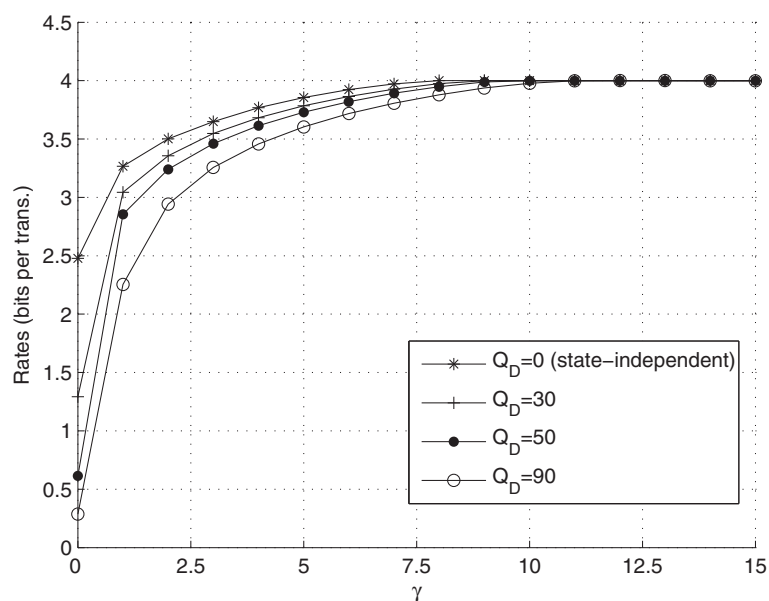

Figure 5 Achievable rates vs. γ under different power values of the interference.

power of the source and the relay will be spent to clean the interference; this results in a lower achievable rate. However, as the power P increases, the impact of the interference will be negligible if $P \gg Q_D$ and the achievable rate will approach the capacity of the state-independent relay channel with orthogonal components. The simulation results also illustrate that when the power of the relay satisfies $\gamma P \geq \left(\frac{P}{4N} + \frac{N}{4P} + \frac{Q_D}{P} + \frac{1}{2}\right) P$, the capacity of the channel can be characterized.

Appendices
Appendix 1
Proof of Theorem 1: Analysis of probability of error
The average probability of error is given by

$$P_e \leq \sum_{\substack{\mathbf{s}_D^n \notin \mathcal{T}_\varepsilon^n(S_D) \\ \mathbf{s}_R^n \notin \mathcal{T}_\varepsilon^n(S_R)}} \Pr(\mathbf{s}_D^n) \Pr(\mathbf{s}_R^n)$$
$$+ \sum_{\substack{\mathbf{s}_D^n \in \mathcal{T}_\varepsilon^n(S_D) \\ \mathbf{s}_R^n \in \mathcal{T}_\varepsilon^n(S_R)}} \Pr(\mathbf{s}_D^n) \Pr(\mathbf{s}_R^n) Pr(error | \mathbf{s}_D^n, \mathbf{s}_R^n) \quad (63)$$

By the AEP, the first term $\Pr(\mathbf{s}_D^n \notin \mathcal{T}_\varepsilon^n(S_D)) \Pr(\mathbf{s}_R^n \notin \mathcal{T}_\varepsilon^n(S_R))$ on the RHS of (63) goes to 0 as $n \to \infty$. It is sufficient to upper bound the second term of the RHS of (63). We now examine the probabilities of the error events associated with the encoding and decoding steps. The error event is contained in the union of the following error events, where E_{1k} and E_{2k} correspond to the encoding steps in block k, $E_{3s_R k}$ and $E_{4s_R k}$ correspond to decoding the message $\hat{m}_{s_R}(k)$ at the relay in block k given $S_R = s_R$, the events E_{5k} and E_{6k} correspond to decoding $w_D(k)$ at the destination in block k. The probability of error $Pr(error | \mathbf{s}_D^n, \mathbf{s}_R^n)$ is upper bounded as

$$Pr(error \mid \mathbf{s}^{Dn}, \mathbf{s}_R^n) \leq Pr(E_{1k}) + Pr(E_{2k})$$
$$+ \sum_{s_R \in \mathcal{S}_R} P(S_R = s_R) Pr(E_{3s_R k} | E_{1k}^c E_{2k}^c)$$
$$+ \sum_{s_R \in \mathcal{S}_R} P(S_R = s_R) Pr(E_{4s_R k} | E_{1k}^c E_{2k}^c E_{3s_R k}^c)$$
$$+ Pr(E_{5k} | E_{1k}^c E_{2k}^c) + Pr(E_{6k} | E_{1k}^c E_{2k}^c E_{5k}^c),$$

where $E_{mk}^c (m = 1, 2, 3s_R, 5)$ denotes the corresponding event complement of E_{mk}.

Let E_{1k} be the event that there is no sequence $\mathbf{u}_r^n (w_R(k-1), j_r(k))$ jointly typical with $\mathbf{s}_D^n(k)$, i.e.,

$$E_{1k} = \{\nexists j_r(k) \in \{1, 2, ..., 2^{nR_{r,s}}\} \ s.t. (\mathbf{u}_r^n(w_R(k-1),$$
$$j_r(k)), \mathbf{s}_D^n(k)) \in \mathcal{T}_\varepsilon^n(U_r, S_D)\}$$

For $\mathbf{u}_r^n(w_R(k-1), j_r(k))$ and $\mathbf{s}_D^n(k)$ generated independently with i.i.d. components according to P_{U_r} and Q_{S_D}, respectively, the probability that there exits at least one $j_r(k) \in \{1, 2, ..., 2^{nR_{r,s}}\}$ such that $\mathbf{u}_r^n(w_R(k-1), j_r(k))$ is

jointly typical with $\mathbf{s}_D^n(k)$ is greater than $(1-\varepsilon) 2^{-n(I(U_r; S_D)+\delta(\varepsilon))}$ for n sufficiently large. There are $2^{nR_{r,s}}$ such $\mathbf{u}_r^n$ s in each bin. Therefore, the probability of event E_{1k} is bounded by

$$\Pr(E_{1k}) \leq \left[1 - (1-\varepsilon) 2^{-n(I(U_r; S_D)+\delta(\varepsilon))}\right]^{2^{nR_{r,s}}} \quad (64)$$

Taking the logarithm on both sides of (64) and following from the inequality $\ln(x) \leq x - 1$, we have $ln(\Pr(E_{1k})) \leq -(1-\varepsilon) 2^{n(R_{r,s} - I(U_r; S_D) - \delta(\varepsilon))}$. Thus, if $R_{r,s} > I(U_r; S_D) + \delta(\varepsilon)$, $\Pr(E_{1k}) \to 0$ as $n \to \infty$, where $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$.

Let E_{2k} be the event that there is no sequence $\mathbf{u}^n(w_D(k), j_d(k) | w_R(k-1), j_r(k))$ jointly typical with $\mathbf{s}_D^n(k)$, given $\mathbf{u}_r^n(w_R(k-1), j_r(k))$, i.e.,

$$E_{2k} = \{\nexists j_d(k) \in \{1, 2, ..., 2^{nR_{d,s}}\} \ s.t.\}$$
$$(\mathbf{u}^n(w_D(k), j_d(k) | w_R(k-1), j_r(k)),$$
$$\mathbf{u}_r^n(w_R(k-1), j_r(k)), \mathbf{s}_D^n(k)) \in \mathcal{T}_\varepsilon^n(U, U_r, S_D)\}$$

Similar to the analysis on the probability of the event E_{1k}, if $R_{d,s} > I(U; S_D | U_r) + \delta(\varepsilon)$, $\Pr(E_{2k}) \to 0$ as $n \to \infty$.

For each $s_R \in \mathcal{S}_R$, let $E_{3s_R k}$ be the event that $\mathbf{x}_R^{n(1-\varepsilon)p(s_R)}(m_{s_R}(k) | s_R, w_R(k-1), j_r(k))$ is not jointly typical with $\mathbf{y}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(k)$, given $\mathbf{u}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(w_R(k-1), j_r(k))$, $\mathbf{x}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(k)$ and $S_R = s_R$, i.e.,

$$E_{3s_R k} = \{(\mathbf{x}_R^{n(1-\varepsilon)p(s_R)}(m_{s_R}(k) | s_R, w_R(k-1), j_r(k)),$$
$$\mathbf{u}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(w_R(k-1), j_r(k)), \mathbf{x}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(k),$$
$$\mathbf{y}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(k))$$
$$\notin \mathcal{T}_\varepsilon^n(X_R, U_r, X_r, S_R, Y_r) \ given \ S_R = s_R\}$$

By the LLN, for all $s_R \in \mathcal{S}_R$, $Pr(E_{3s_R k} | E_{1k}^c E_{2k}^c) \to 0$ as $n \to \infty$. Consequently, $\sum_{s_R \in \mathcal{S}_R} P(S_R = s_R) Pr(E_{3s_R k} | E_{1k}^c E_{2k}^c) \to 0$ as $n \to \infty$.

For each $s_R \in \mathcal{S}_R$, let $E_{4s_R k}$ be the event that $\mathbf{x}_R^{n(1-\varepsilon)p(s_R)}(\hat{m}_{s_R}(k) | s_R, w_R(k-1), j_r(k))$ is jointly typical with $\mathbf{y}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(k)$, given $\mathbf{u}_{r,s_R}^{n_{s_R}(k)}(w_R(k-1), j_r(k))$, $\mathbf{x}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(k)$ and $S_R = s_R$ for some $\hat{m}_{s_R}(k) \neq m_{s_R}(k)$, i.e.,

$$E_{4s_R k} = \{\exists \hat{m}_{s_R}(k) \in \{1, 2, ..., 2^{nR_{s_R}}\} \ s.t. \hat{m}_{s_R}(k) \neq m_{s_R}(k),$$
$$(\mathbf{x}_R^{n(1-\varepsilon)p(s_R)}(\hat{m}_{s_R}(k) | s_R, w_R(k-1), j_r(k)),$$
$$\mathbf{u}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(w_R(k-1), j_r(k)), \mathbf{x}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(k),$$
$$\mathbf{y}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(k))$$
$$\in \mathcal{T}_\varepsilon^n(X_R, U_r, X_r, S_R, Y_r) \ given \ S_R = s_R\}$$

Conditioned on the events E_{1k}^c, E_{2k}^c and $E_{3s_R k}^c$, for all $s_R \in \mathcal{S}_R$, by the joint typicality lemma [31], the probability that $(\mathbf{x}_R^{n(1-\varepsilon)p(s_R)}(\hat{m}_{s_R}(k) | s_R, w_R(k-1), j_r(k)), \mathbf{u}_{r,s_R}^{n(1-\varepsilon)p(s_R)}$

$(w_R(k-1), j_r(k)), \mathbf{x}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(k), \mathbf{y}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(k) \in \mathcal{T}_{\varepsilon}^{n(1-\varepsilon)p(s_R)}$ $(X_R, U_r, X_r, S_R, Y_r)$ given $\mathbf{u}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(w_R(k-1), j_r(k))$, $\mathbf{x}_{r,s_R}^{n(1-\varepsilon)p(s_R)}(k)$ and $S_R = s_R$ for $\hat{m}_{s_R}(k) \neq m_{s_R}(k)$ is less than $2^{-n(1-\varepsilon)p(s_R)(I(X_R;Y_r|U_r,X_r,S_R=s_R)-\delta(\varepsilon))}$ for sufficiently large n. There are $2^{nR_{s_R}}$ (exactly $2^{nR_{s_R}}-1$) such $\mathbf{x}_R^{n(1-\varepsilon)p(s_R)}$s. Thus, the conditional probability of event $E_{4s_R k}$ given E_{1k}^c, E_{2k}^c and $E_{3s_R k}^c$ is upper bounded by

$$Pr(E_{4s_R k}|E_{1k}^c E_{2k}^c E_{3s_R k}^c)$$
$$\leq 2^{-n((1-\varepsilon)p(s_R)I(X_R;Y_r|U_r,X_r,S_R=s_R)-(1-\varepsilon)p(s_R)\delta(\varepsilon)-R_{s_R})} \tag{65}$$

From (65), $Pr(E_{4s_R k}|E_{1k}^c E_{2k}^c E_{3s_R k}^c) \to 0$ as $n \to \infty$ if

$$R_{s_R} < (1-\varepsilon)p(s_R)(I(X_R;Y_r|U_r,X_r,S_R=s_R)-\delta(\varepsilon))$$

Since $R_R = \sum_{s_R \in S_R} R_{s_R}$, $\sum_{s_R \in S_R} P(S_R = s_R) Pr(E_{4s_R k}|E_{1k}^c E_{2k}^c E_{3s_R k}^c) \to 0$ as $n \to \infty$ if

$$R_R < (1-\varepsilon)(I(X_R;Y_r|U_r,X_r,S_R)-\delta(\varepsilon)),$$

where $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$.

Let E_{5k} be the event that $\mathbf{u}^n(w_D(k), j_d(k)|w_R(k-1), j_r(k))$, $\mathbf{u}_r^n(w_R(k-1), j_r(k))$ and $\mathbf{y}^n(k)$ are not jointly typical, i.e.,

$$E_{5k} = \{(\mathbf{u}^n(w_D(k), j_d(k)|w_R(k-1), j_r(k)), \mathbf{u}_r^n(w_R(k-1),$$
$$j_r(k)), \mathbf{y}^n(k)) \notin \mathcal{T}_{\varepsilon}^n(U, U_r, Y)\}$$

Conditioned on the events E_{1k}^c and E_{2k}^c, we have $Pr(E_{5k}|E_{1k}^c E_{2k}^c) \to 0$ as $n \to \infty$ by the Markov Lemma.

Let E_{6k} be the event that $\mathbf{u}^n(\hat{w}_D(k), \hat{j}_d(k)|\hat{w}_R(k-1), \hat{j}_r(k))$ and $\mathbf{u}_r^n(\hat{w}_R(k-1), \hat{j}_r(k))$ are jointly typical with $\mathbf{y}^n(k)$ for some $(\hat{w}_D(k), \hat{w}_R(k-1)) \in \{1, 2, ..., 2^{nR_D}\} \times \{1, 2, ..., 2^{nR_R}\}$, $\hat{j}_d(k) \in \{1, 2, ..., 2^{nR_{d,s}}\}$ and $\hat{j}_r(k) \in \{1, 2, ..., 2^{nR_{r,s}}\}$, with $(\hat{w}_D(k), \hat{w}_R(k-1)) \neq (w_D(k), w_R(k-1))$, i.e.,

$$E_{6k} = \{\exists (\hat{w}_D(k), \hat{w}_R(k-1)) \in \{1, 2, \cdots, 2^{nR_D}\}$$
$$\times \{1, 2, \cdots, 2^{nR_R}\}, \hat{j}_d(k) \in \{1, 2, \cdots, 2^{nR_{d,s}}\},$$
$$\hat{j}_r(k) \in \{1, 2, \cdots, 2^{nR_{r,s}}\}$$
$$s.t. (\hat{w}_D(k), \hat{w}_R(k-1)) \neq (w_D(k), w_R(k-1)),$$
$$(\mathbf{u}^n(\hat{w}_D(k), \hat{j}_d(k)|\hat{w}_R(k-1), \hat{j}_r(k)),$$
$$\mathbf{u}_r^n(\hat{w}_R(k-1), \hat{j}_r(k)), \mathbf{y}^n(k)) \in \mathcal{T}_{\varepsilon}^n(U, U_r, Y)\}$$

We split the potential event E_{6k} into three disjoint parts: first, $\hat{w}_D(k) = w_D(k)$ and $\hat{w}_R(k-1) \neq w_R(k-1)$;

second, $\hat{w}_D(k) \neq w_D(k)$ and $\hat{w}_R(k-1) = w_R(k-1)$; third, $\hat{w}_D(k) \neq w_D(k)$ and $\hat{w}_R(k-1) \neq w_R(k-1)$, i.e.,

$$E_{6k_1} = \{\exists \hat{w}_R(k-1) \in \{1, 2, \cdots, 2^{nR_R}\}, \hat{j}_r(k) \in \{1, 2, \cdots, 2^{nR_{r,s}}\}$$
$$s.t. \hat{w}_R(k-1) \neq w_R(k-1),$$
$$(\mathbf{u}^n(w_D(k), j_d(k)|\hat{w}_R(k-1), \hat{j}_r(k)),$$
$$\mathbf{u}_r^n(\hat{w}_R(k-1), \hat{j}_r(k)), \mathbf{y}^n(k)) \in \mathcal{T}_{\varepsilon}^n(U, U_r, Y)\}$$

$$E_{6k_2} = \{\exists \hat{w}_D(k) \in \{1, 2, \cdots, 2^{nR_D}\}, \hat{j}_d(k) \in \{1, 2, \cdots, 2^{nR_{d,s}}\},$$
$$s.t. \hat{w}_D(k) \neq w_D(k),$$
$$(\mathbf{u}^n(\hat{w}_D(k), \hat{j}_d(k)|w_R(k-1), j_r(k)),$$
$$\mathbf{u}_r^n(w_R(k-1), j_r(k)), \mathbf{y}^n(k)) \in \mathcal{T}_{\varepsilon}^n(U, U_r, Y)\}$$

$$E_{6k_3} = \{\exists (\hat{w}_D(k), \hat{w}_R(k-1)) \in \{1, 2, \cdots, 2^{nR_D}\}$$
$$\times \{1, 2, \cdots, 2^{nR_D}\}, \hat{j}_d(k) \in \{1, 2, \cdots, 2^{nR_{d,s}}\},$$
$$\hat{j}_r(k) \in \{1, 2, ..., 2^{nR_{r,s}}\} s.t. \hat{w}_D(k) \neq w_D(k),$$
$$\hat{w}_R(k-1) \neq w_R(k-1),$$
$$(\mathbf{u}^n(\hat{w}_D(k), \hat{j}_d(k)|\hat{w}_R(k-1), \hat{j}_r(k)),$$
$$\mathbf{u}_r^n(\hat{w}_R(k-1), \hat{j}_r(k)), \mathbf{y}^n(k)) \in \mathcal{T}_{\varepsilon}^n(U, U_r, Y)\}$$

Conditioned on the events E_{1k}^c, E_{2k}^c and E_{5k}^c, by the joint typicality lemma, the probability of the above three events are bounded by

$$Pr(E_{6k_1}|E_{1k}^c E_{2k}^c E_{5k}^c) \leq 2^{-n(I(U,U_r;Y)-\delta(\varepsilon)-R_R)} \tag{66}$$

$$Pr(E_{6k_2}|E_{1k}^c E_{2k}^c E_{5k}^c) \leq 2^{-n(I(U;Y|U_r)-\delta(\varepsilon)-R_D)} \tag{67}$$

$$Pr(E_{6k_3}|E_{1k}^c E_{2k}^c E_{5k}^c) \leq 2^{-n(I(U, U_r;Y)-\delta(\varepsilon)-(R_D+R_R))} \tag{68}$$

From (66) to (68), $Pr(E_{6k\_1}|E_{1k}^c E_{2k}^c E_{5k}^c) \to 0$, $Pr(E_{6k\_2}|E_{1k}^c E_{2k}^c E_{5k}^c) \to 0$ and $Pr(E_{6k\_3}|E_{1k}^c E_{2k}^c E_{5k}^c) \to 0$ as $n \to \infty$, if

$$R_D + R_{d,s} < I(U;Y|U_r)-\delta(\varepsilon)$$

$$R_D + R_{d,s} + R_R + R_{r,s} < I(U, U_r;Y)-\delta(\varepsilon),$$

where $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$. Note that (68) makes (66) unnecessary. Therefore, $Pr(E_{6k}|E_{1k}^c E_{2k}^c E_{5k}^c) \leq Pr(E_{6k\_1}|E_{1k}^c E_{2k}^c E_{5k}^c) + Pr(E_{6k\_2}|E_{1k}^c E_{2k}^c E_{5k}^c) + Pr(E_{6k\_3}|E_{1k}^c E_{2k}^c E_{5k}^c)$ approaches 0 as $n \to \infty$.

It remains to show that the alphabet sizes of the random variables U_r and U can be limited without loss of generality as stated in (8) and (9), respectively. This is done by the support lemma [33]. Fix a distribution μ of $(S_D, S_R, U_r, U, X_r, X_R, X_D, Y_r, Y_D)$ on the Borel σ-algebra of $\mathcal{P}(\mathcal{S}_D, \mathcal{S}_R, \mathcal{U}_r, \mathcal{U}, \mathcal{X}_r, \mathcal{X}_R, \mathcal{X}_D, \mathcal{Y}_r, \mathcal{Y}_D)$ that has the form (7).

To bound (8) on $|\mathcal{U}_r|$, note that we have

$$
\begin{aligned}
I_\mu\big(X_R; Y_r\big|U_r, X_r, S_R\big) &+ I_\mu(U; Y|U_r) - I_\mu(U; S_D|U_r) \\
= H_\mu(X_R, X_r, S_R|U_r) &+ H_\mu(X_r, S_R, Y_r|U_r) \\
- H_\mu(X_r, S_R|U_r) &- H_\mu(X_R, X_r, S_R, Y_r|U_r) \\
+ H_\mu(Y|U_r) &- H_\mu(U, Y|U_r) + H_\mu(U, S_D|U_r) \\
- H_\mu(S_D|U_r)
\end{aligned}
\tag{69}
$$

$$
\begin{aligned}
I_\mu(U, U_r; Y) &- I_\mu(U, U_r; S_D) \\
= H_\mu(Y) &- H_\mu(S_D) - H_\mu\big(U, Y|U_r\big) + H_\mu(U, S_D|U_r)
\end{aligned}
\tag{70}
$$

Hence, it suffices to show that the following functionals of $\mu(S_D, S_R, U_r, U, X_r, X_R, X_D, Y_r, Y_D)$:

$$
\begin{aligned}
f_{S_D, S_R, X_r, X_D} &= \mu(s_D, s_R, x_r, x_D) \quad \forall (s_D, s_R, x_r, x_D) \in \mathcal{S}_D \\
&\times \mathcal{S}_R \times \mathcal{X}_r \times \mathcal{X}_D
\end{aligned}
\tag{71}
$$

$$
\begin{aligned}
f_0(\mu) = \int_\mu d_\mu(u_r) \big(& H_\mu(X_R, X_r, S_R|U_r) + H_\mu(X_r, S_R, Y_r|U_r) \\
& - H_\mu(X_r, S_R|U_r) - H_\mu(X_R, X_r, S_R, Y_r|U_r) \\
& + H_\mu(Y|u_r) - H_\mu(U, Y|u_r) \\
& + H_\mu(U, S_D|u_r) - H_\mu(S_D|u_r)\big)
\end{aligned}
\tag{72}
$$

$$
f_1(\mu) = \int_\mu d_\mu(u_r)\big(H_\mu\big(U, S_D|u_r\big) - H_\mu(U, Y|u_r)\big)
\tag{73}
$$

can be preserved with another measure μ' that has the form (7). This condition can be satisfied by the support lemma. Observing that there are $|\mathcal{S}_D||\mathcal{S}_R||\mathcal{X}_r||\mathcal{X}_D| + 1$ functionals (in (71), there are $|\mathcal{S}_D||\mathcal{S}_R||\mathcal{X}_r||\mathcal{X}_D| - 1$ degrees of freedom) in (71) to (73), according to the support lemma, the cardinality of the alphabet of the auxiliary random variable U_r can be taken to $|\mathcal{S}_D||\mathcal{S}_R| |\mathcal{X}_r||\mathcal{X}_D| + 1$ without altering $I(X_R; Y_r|U_r, X_r, S_R) + I(U; Y|U_r) - I(U; S_D|U_r)$ and $I(U, U_r; Y) - I(U, U_r; S_D)$.

Once the alphabet of U_r is fixed, the alphabet of U is bounded in a similar way. $|\mathcal{S}_D||\mathcal{S}_R||\mathcal{X}_r||\mathcal{X}_D| \times (|\mathcal{S}_D||\mathcal{S}_R||\mathcal{X}_r||\mathcal{X}_D| + 1) + 1$ functionals must be satisfied to preserve the joint distribution of U_r, S_D, S_R, X_r, X_D and two more functionals to preserve

$$
\begin{aligned}
I_\mu(U; Y|U_r) &- I_\mu(U; S_D|U_r) \\
= H_\mu(Y|U_r) &- H_\mu(S_D|U_r) - H_\mu\big(U_r, Y|U\big) + H_\mu(S_D, U_r|U)
\end{aligned}
\tag{74}
$$

$$
\begin{aligned}
I_\mu(U, U_r; Y) &- I_\mu(U, U_r; S_D) \\
= H_\mu(Y) &- H_\mu(S_D) - H_\mu\big(U_r, Y|U\big) + H_\mu(U_r, S_D|U)
\end{aligned}
\tag{75}
$$

yielding the bound indicated in (9).

This completes the proof.

Appendix 2
Outline of the proof of Theorem 4

With the random variables defined in (40) to (43), we rewrite (37) and (38) as follows:

$$
Y_r = X_R + S_R + Z_r
\tag{76}
$$

$$
\begin{aligned}
Y &= X_D + X_r + S_D + Z_d \\
&= X_{D0} + \big(1 + \rho\sqrt{\beta/\gamma}\big)X_r + S_D + Z_d
\end{aligned}
\tag{77}
$$

With these definitions, it is easy to verify that $U = X_D + \left(\alpha(1+\alpha_r) + \frac{\alpha_r\rho\sqrt{\beta/\gamma}}{(1+\rho\sqrt{\beta/\gamma})}\right)S_D$, U_0 and U_r are conditionally independent, given S_D, and α and α_r will be clear in the sequel.

Use the two random variables U_r and U to generate the auxiliary codewords $\mathbf{u}_r^n$ and $\mathbf{u}^n$. Under these choices of the random variables, codebook generation, encoding and decoding are the same as that in the DM case except that successive decoding is used to decode the quadruple $\big(\hat{w}_R(k-1), \hat{j}_r(k), \hat{w}_D(k), \hat{j}_d(k)\big)$ instead of joint decoding in the DM case.

Observing $\mathbf{y}^n(k)$, the destination firstly find a pair $\big(\hat{w}_R(k-1), \hat{j}_r(k)\big)$ such that

$$
\big(\mathbf{u}_r^n\big(\hat{w}_R(k-1), \hat{j}_r(k)\big), \mathbf{y}^n(k)\big) \in \mathcal{A}_\varepsilon^n(U_r, Y)
$$

If there is no such pair or it is not unique, an error is declared. By the packing lemma [31], it can be shown that for sufficiently large n, decoding is correct with high probability if

$$
R_R + R_{r,s} \le I(U_r; Y)
\tag{78}
$$

From (10) and (78), $R_R \le I(U_r; Y) - I(U_r; S_D)$.

Treating X_{D0} in (77) as noise, we define the auxiliary random variable U_r as

$$
U_r = \big(1 + \rho\sqrt{\beta/\gamma}\big)X_r + \alpha_r S_D
\tag{79}
$$

By Costa's dirty paper coding, let

$$
\alpha_r = \frac{\big(\sqrt{\gamma} + \rho\sqrt{\beta}\big)^2 P}{\big(\gamma + \beta + 2\rho\sqrt{\beta\gamma}\big)P + N}
\tag{80}
$$

then

$$
\begin{aligned}
R_R &\le I(U_r; Y) - I(U_r; S_D) \\
&= \mathcal{C}\left(\frac{\big(\sqrt{\gamma} + \rho\sqrt{\beta}\big)^2 P}{(1-\rho^2)\beta P + N}\right)
\end{aligned}
\tag{81}
$$

is achievable.

Since the destination has successfully decoded $\mathbf{u}_r^n$ $\left(\hat{w}_R(k-1),\hat{j}_r(k)\right)$, the destination can peel off U_r to make the channel to the destination equivalent to

$$
\begin{aligned}
Y^{'} &= Y - U_r \\
&= X_{D0} + \left(1 + \rho\sqrt{\frac{\beta}{\gamma}}\right)X_r + S_D + Z_d - \left[\left(1 + \rho\sqrt{\frac{\beta}{\gamma}}\right)X_r + \alpha_r S_D\right] \\
&= X_{D0} + (1 + \alpha_r)S_D + Z_d
\end{aligned}
\tag{82}
$$

The destination finds a unique pair $\left(\hat{w}_D(k),\hat{j}_d(k)\right)$ such that

$$
\begin{aligned}
(\mathbf{u}^n(\hat{w}_D(k),&\hat{j}_d(k)|\hat{w}_R(k-1),\hat{j}_r(k)),\mathbf{u}_r^n\left(\hat{w}_R(k-1),\hat{j}_r(k)\right),\mathbf{y}^{'n}(k)) \\
&\times \in \mathcal{A}_\varepsilon^n(U,U_r,Y^{'})
\end{aligned}
$$

If there is no such pair or it is not unique, an error is declared. By the packing lemma [31], it can be shown that for sufficiently large n, decoding is correct with high probability if

$$
R_D + R_{d,s} \leq I\left(U;Y^{'}|U_r\right) \tag{83}
$$

From (11) and (83), $R_D \leq I(U;\ Y^{'}|U_r) - I(U;\ S_D|U_r)$ is achievable. The proof will continue after the following lemma.

Lemma 1 *For the considered channel here, if $R_D \leq I(U;$ $Y^{'}|U_r) - I(U;\ S_D|U_r)$ is achievable, $R_D \leq I(U_0;\ Y^{'}) - I(U_0;$ $S_D)$ is achievable.*

Proof We only have to verify $I(U;\ Y^{'}|U_r) - I(U;\ S_D|U_r) \geq I(U_0;\ Y^{'}) - I(U_0;\ S_D)$.

$$
\begin{aligned}
I\left(U;Y^{'}|U_r\right) &- I(U;S_D|U_r) \\
&= H(U|U_r) - H\left(U|U_r,Y^{'}\right) - H(U|U_r) + H(U|U_r,S_D) \\
&= H(U|U_r,S_D) - H\left(U|U_r,Y^{'}\right) \\
&= H\left(U_0 + \frac{\rho\sqrt{\beta/\gamma}}{\left(1 + \rho\sqrt{\beta/\gamma}\right)}U_r|U_r,S_D\right) \\
&\quad - H\left(U_0 + \frac{\rho\sqrt{\beta/\gamma}}{\left(1 + \rho\sqrt{\beta/\gamma}\right)}U_r|U_r,Y^{'}\right) \\
&= H(U_0|U_r,S_D) - H(U_0|U_r,Y^{'}) \\
&\geq H(U_0|S_D) - H(U_0|Y^{'}) \\
&= I(U_0;Y^{'}) - I(U_0;S_D),
\end{aligned}
\tag{84}
$$

where the inequality holds due to the facts that U_0 and U_r are conditionally independent given S_D and conditioning reduces entropy.

Thus, if $R_D \leq I(U;\ Y^{'}|U_r) - I(U;\ S_D|U_r)$ is achievable, we have

$$
R_D \leq I\left(U_0;Y^{'}\right) - I(U_0;S_D) \tag{85}
$$

is achievable.

This completes the proof of Lemma 1.
By (82), we define

$$
U_0 = X_{D,0} + \alpha(1 - \alpha_r)S_D \tag{86}
$$

With Costa's dirty paper coding, let

$$
\alpha = \frac{(1 - \rho^2)\beta P}{(1 - \rho^2)\beta P + N} \tag{87}
$$

Then by Lemma 1, we have

$$
\begin{aligned}
R_D &\leq I\left(U_0;Y^{'}\right) - I(U_0;S_D) \\
&= \mathcal{C}\left(\frac{(1 - \rho^2)\beta P}{N}\right)
\end{aligned}
\tag{88}
$$

is achievable.

The decoding at the relay is the same as that in the DM case except that the interference sequence $\mathbf{s}_R^n(k)$ can be subtracted before decoding $w_R(k)$ since $\mathbf{s}_R^n(k)$ is additive and known to the relay. This makes the channel from the source to the relay equivalent to

$$
Y_r^{'} = X_R + Z_r \tag{89}
$$

Therefore,

$$
R_R \leq \mathcal{C}\left(\frac{\bar{\beta}P}{N}\right) \tag{90}
$$

is achievable.

Combining (81), (88) and (90), we show that $R = R_R + R_D$ is achievable when the following condition is satisfied

$$
\begin{aligned}
C(P,\gamma P) &= \max_{0 \leq \beta,\ \rho \leq 1} \min \\
&\times \left\{\mathcal{C}\left(\frac{(\beta + \gamma + 2\rho\sqrt{\beta\gamma})P}{N_2}\right),\mathcal{C}\left(\frac{\bar{\beta}P}{N_1}\right) + \mathcal{C}\left(\frac{\beta(1 - \rho^2)P}{N_2}\right)\right\}
\end{aligned}
\tag{91}
$$

This completes the proof.

Author details
[1]College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 21003, China. [2]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China. [3]College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210003, China. [4]College of Internet of Things Engineering, Hohai University, Changzhou 213022, China.

References

1. G. Keshet, Y. Steinberg, N. Merhav, Channel coding in the presence of side information: subject review. Found. Trends Commun. Inf. Theory **4**(6), 44–586 (2008)
2. C.E. Shannon, Channels with side information at the transmitter. IBM J. Res. Develop. **2**(4), 289–293 (1958)
3. S.I. Gel'fand, M.S. Pinsker, Coding for channel with random parameters. Prob. Contr. Inf. Theory **9**(1), 19–31 (1980)
4. M.H.M. Costa, Writing on dirty paper. IEEE Trans. Inf. Theory **29**(3), 439–441 (1983)
5. S.I. Gel'fand, M.S. Pinsker, On Gaussian channel with random parameters, in *Paper presented at the IEEE International Symposium on Information Theory, U.S.S.R Tashkent, 18–22 Sept 1984*, pp. 247–250
6. S. Sigurjonsson, Y.H. Kim, On multiple user channels with state information at the transmitters, in *Paper presented at the IEEE international symposium on information theory (ISIT), Adelaide, 4–9 Sept 2005*, pp. 72–76
7. Y. Steinberg, S. Shamai, (Shitz), Achievable rates for the broadcast channel with states known at the transmitter, in *Paper presented at the IEEE international symposium on information theory (ISIT), Adelaide, SA, 4–9 Sept 2005*, pp. 2184–2188
8. Y. Steinberg, Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information. IEEE Trans. Inf. Theory. **51**(8), 2867–2877 (2005)
9. K.G. Nagananda, C.R. Murthy, S. Kishore, Capacity bounds for state-dependent broadcast channels. (Cornell University Library, 2012). http://arxiv.org/abs/1110.0124. Accessed 12 Aug 2012
10. S. Kotagiri, J.N. Laneman, Achievable rates for multiple access channels with state information known at one encoder, in *Paper presented at the Allerton conference on communications, control, and computing, Monticello, IL, USA, 29 Sept–1 Oct 2004*
11. S. Kotagiri, J.N. Laneman, Multiple access channels with state information known to some encoders and independent messages. EURASIP J. Wireless Commun. Netw. (2008). doi:10.1155/2008/450680
12. Y. Cemal, Y. Steinberg, The multiple-access channel with partial state information at the encoders. IEEE Trans. Inf. Theory. **51**(11), 3392–4003 (2005)
13. Y. Kim, A. Sutivong, S. Sigurjonsson, Multiple user writing on dirty paper, in *Paper presented at the IEEE international symposium on information and theory (ISIT), Chicago, 27 June–2 July 2004*, p. 534
14. A. Somekh-Baruch, S. Shamai (Shitz), S. Verdú, Cooperative multiple-access encoding with states available at one transmitter. IEEE Trans. Inf. Theory **54**(10), 4448–4469 (2008)
15. S.I. Bross, A. Lapidoth, M. Wigger, Dirty-paper coding for the Gaussian multiaccess channel with conferencing. IEEE Trans. Inf. Theory **58**(9), 5640–5668 (2012)
16. A. Zaidi, S.P. Kotagiri, J. Nicholas Laneman, L. Vandendorpe, Cooperative relaying with state available noncausally at the relay. IEEE Trans. Inf. Theory **56**(5), 2272–2298 (2010)
17. A. Zaidi, L. Vandendorpe, Lower bounds on the capacity of the relay channel with states at the source. EURASIP J. Wireless Commun. Netw. (2009). doi:10.1155/2009/634296
18. A. Zaidi, S. Shamai (Shitz), P. Piantanida, L. Vandendorpe, Bounds on the capacity of the relay channel with noncausal state at source. IEEE Trans. Inf. Theory **59**(5), 2639–2672 (2013)
19. A. Zaidi, S. Shamai, P. Piantanida, L. Vandendorpe, On the capacity of a class of relay channels with orthogonal components and noncausal state information at source, in *Paper presented at the 5th international symposium on wireless pervasive computing (ISWPC), Modena, Italy, 5–7 May 2010*, pp. 528–533
20. K. Bakanoglu, E. Erkip, O. Simeone, S. Shamai (Shitz), Relay channel with orthogonal components and structured interference known at the source. IEEE Trans. Comm. **61**(4), 1277–289 (2012)
21. B. Akhbari, M. Mirmohseni, M.R. Aref, Compress-and-forward strategy for the relay channel with non-causal state information, in *Paper presented at the IEEE international symposium on information theory (ISIT), Seoul, Korea, 28 June–3 July 2009*, pp. 1169–1173
22. B. Chen, G.W. Wornell, Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE Trans. Inf. Theory **47**(4), 1423–1443 (2001)
23. A.S. Cohen, A. Lapidoth, The Gaussian watermarking game. IEEE Trans. Inf. Theory **48**(6), 1639–1667 (2002)
24. P. Moulin, J.A. O'Sullivan, Information-theoretic analysis of information hiding. IEEE Trans. Inf. Theory **49**(3), 563–593 (2003)
25. A.V. Kusnetsov, B.S. Tsybakov, Coding in a memory with defective cells. Probl. Predach. Inf. **10**, 52–60 (1974)
26. J. Mitola, *Cognitive radio: an integrated agent architecture for software defined radio* (Dissertation, KTH Royal Institute of Technology, 2000)
27. T.M. Cover, A.E.L. Gamal, Capacity theorems for the relay channel. IEEE Trans. Inf. Theory **IT-25**(5), 572–584 (1979)
28. Y.H. Kim, Capacity of a class of relay channels. IEEE Trans. Inf. Theory **54**(3), 1328–1329 (2008)
29. A. El Gamal, S. Zahedi, Capacity of a class of relay channels with orthogonal components. IEEE Trans. Inf. Theory **51**(5), 1815–1817 (2005)
30. A.J. Goldsmith, P.P. Varaiya, Capacity of fading channels with channel side information IEEE Trans. Inf. Theory **43**(6), 1986–1992 (1997)
31. A. El Gamal, Y.H. Kim, *Network Information Theory* (Cambridge University Press, Cambridge, 2011)
32. T.M. Cover, J.A. Thomas, *Elements of Information Theory*, 2nd edn. (Liley, New York, 2006)
33. I. Csiszár, J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Channels* (Academic, Waltham, 1981)