

Compress-and-forward strategy for relay channel with causal and non-causal channel state information

B. Akhbari M. Mirmohseni M.R. Aref

Information Systems and Security Lab (ISSL), Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran

E-mail: b_akhbari@ee.sharif.edu

Abstract: The discrete memoryless state-dependent relay channel (SD-RC) is considered in this study. Two main cases are investigated: SD-RC with non-causal channel state information (CSI) and SD-RC with causal CSI. In each case, the SD-RC with partial CSI at the source and the relay is considered. As special cases it includes three different situations in which perfect CSI is available: (i) only at the source, (ii) only at the relay and (iii) both at the source and the relay. For the non-causal situation, the authors establish lower bound on capacity (achievable rate) of the SD-RC, using Gel'fand-Pinsker coding at the nodes informed of CSI and compress-and-forward (CF) strategy at the relay. Using the Shannon's strategy and CF relaying, the authors derive lower bound on capacity of SD-RC in the causal case. Furthermore, in order to compare their derived bounds with the previously obtained results, which are based on the decode-and-forward (DF) strategy, the authors consider general Gaussian relay channel (RC) with additive independent and identically distributed Gaussian state and noise, and obtain lower bounds on capacity for the cases in which perfect CSI is available non-causally at the source or at the relay. They also present cases in which their lower bounds outperform DF-based bounds, and can achieve rates close to the upper bound. For causal case, a numerical example of the binary fading Gaussian RC with additive noise is provided.

1 Introduction

Owing to wide range of applications, recently investigation of communication strategies for state-dependent channels has received considerable attention [1]. In state-dependent channels, channel state information (CSI) can be available at the nodes causally or non-causally. When the transmitter knows CSI causally, the input to the channel at each time instant depends only on the past and the present CSI, whereas, in the non-causal case, the transmitter knows in advance the realisation of the entire state process from the beginning to the end of the block.

Causal channel state information at the transmitter (CSIT) can be assumed for example in wireless fading channels where the receiver estimates instantaneous state of the channel and sends it back to the transmitter via a

feedback link [1]. Shannon [2], for the first time, established the capacity of a state-dependent memoryless single-user channel whose states are independent and identically distributed (i.i.d) and CSI is causally known to the transmitter. Shannon showed that the capacity of this channel is equal to the capacity of an ordinary discrete memoryless channel with the same output alphabet and an extended input alphabet (i.e. $\mathcal{X}^{|\mathcal{S}|}$: \mathcal{S} and \mathcal{X} are state and input alphabets, respectively). In fact, optimal code for this channel is constructed over the alphabet of all mappings from \mathcal{S} to \mathcal{X} defined as Shannon's strategies, and as Shannon showed in [2], by adding a physical device in front of the channel depends on the message to be sent and the current state, capacity of this channel is achieved.

Non-causal CSIT can be considered in the context of computer memory with defects as the location of defective

cells. Gel'fand and Pinsker [3], characterised the capacity of the state-dependent single-user channel with non-causal CSIT by a coding scheme, which is called Gel'fand-Pinsker (GP) coding. Costa [4] applied GP results to the Gaussian channel with additive interference, known as non-causal CSIT, and showed that known interference does not reduce capacity [Costa's coding is also called dirty paper coding (DPC)].

Much research interests has recently been devoted to multiuser models with both causal and non-causal CSI [1, 5–7]. Among state-dependent multiuser models [8–10], some results have also been obtained for the relay channel (RC), recently [6, 7, 11–17].

The classical RC (state-independent) was first introduced in [18], and then extensively studied in [19]. Two basic coding strategies for RC were proposed in [19]: decode and forward (DF), and compress and forward (CF) strategy. The capacity of some special classes of the RC such as degraded relay channels (DRC) [19], semi-deterministic RCs [20, 21] and RCs with orthogonal components [22] have been established. Moreover, RC with delay [23], relay-broadcast channels [24, 25] and RC with private messages [26] have been investigated. Recent applications of relaying in wireless networks have revived the interest in studying relay networks [27–29].

In [7], the capacity of the discrete memoryless degraded state-dependent relay channel (SD-RC) has been derived, where the source and relay have access to identical causal CSI. The capacity achieving strategy in [7], consists of the extension of Shannon's strategy [2] and using DF scheme [19] at the relay. Based on the DF strategy for the RC, in [30] a lower bound on the capacity of T -node relay network ($(T-2)$ relay nodes, one transmitter and one destination) in the case where the source and the relays have access to non-causal and identical CSI has been established.

One of the most important issues in the state-dependent multiuser models is whether CSI is known symmetrically (to all the nodes) or asymmetrically (to only some of the nodes). Asymmetric knowledge of CSI may happen when only some of terminals equipped with cognition capability permits them to know the state of the channel. Asymmetric non-causal CSI in RC has been considered in [11, 13], where CSI has been known non-causally to the source or to the relay. Hence, only source or relay can combine the GP coding and the DF scheme. In [12] we have considered state-dependent cooperative relay-broadcast channel with asymmetric causal CSI and have established the capacity region of the degraded version of the channel. In [15], lower and upper bounds on the capacity of SD-RC where source and relay do not have identical CSI have been derived for four different classes (instantaneous, causal, delay-less, relaying with unlimited look ahead). Recently, achievable rates for a class of Gaussian RCs with

interference, which is known non-causally only to the source, have been derived using different coding strategies [16, 17].

In most of the previous works on SD-RC, only the DF strategy (or Partial DF (PDF)) has been considered. Conventional DF (or PDF) strategies are based on decoding the whole message (or part of it) by the relay. So, the relay node cooperates with the transmitter to send the decoded part to the destination. In the case that the channel from the source to the relay is remarkably better than the direct link between the source and the destination, the DF strategy is beneficial [22]. But, when the channel from the source to the relay is worse than the direct link or even when these two links are similar in average, DF scheme cannot be helpful and another scheme is needed. In a completely different approach known as CF strategy [19], relay compresses the received sequence (without decoding any part of the transmitted message), re-encodes and sends it to the destination.

In this paper, we focus on using the CF strategy in SD-RC for two reasons. (i) We expect that in SD-RC, analogous to the classical case, CF strategy outperforms the DF scheme when the link between the source and relay is worse than direct link, and can achieve rates close to the upper bound. (ii) In the conventional DF (or PDF), the source must know the relay input in order to have cooperation with it. But, when CSI is asymmetric (e.g. for the case where relay is only informed), the source cannot exploit the CSI. Hence, it does not know what the relay exactly sends and this introduces some loss in the coherence gain, which we would expect to achieve in the DF strategy, and as it has been shown in [13], although codeword splitting has been used in their scheme, their DF-based lower bound cannot be tight in the degraded Gaussian case. On the other hand, in the CF scheme independent codebooks are used at the source and the relay. Hence, using the CF approach in asymmetric scenario seems to be reasonable and in this paper in order to analyse this approach, we will consider different situations.

We investigate SD-RC with both causal and non-causal CSI in order to have a unified view. Furthermore, in order to consider asymmetry in CSI, for both causal and non-causal cases, we assume that the source and relay have partial CSI (not necessarily identical). This assumption includes three different situations in both causal and non-causal cases: perfect CSI is known (i) only to the source, (ii) only to the relay, and (iii) both to the source and the relay. In the non-causal case we establish lower bound on channel capacity (achievable rate) based on using GP coding at the informed nodes and CF strategy at the relay. In the causal case, using Shannon's strategies at the informed nodes and CF scheme at the relay, we establish lower bound on capacity. We show that our results for the causal case can be considered as a special case of non-causal CSI and this is congruent with the relation between the

expression for the capacity of the state-dependent single-user channel with causal CSI [2], and its non-causal counterpart, that is, the GP channel [3].

In order to compare our derived lower bounds with the established DF-based bounds for the state-dependent Gaussian RC with non-causal CSI, we consider those obtained in [11] and [13], and we present cases in which the CF approach outperforms the DF strategy. We also provide a numerical example of the binary fading additive white Gaussian noise (AWGN) RC, for causal case.

This paper is organised as follows. In Section 2, we present the state-dependent RC and its related definitions. In Section 3, we investigate the non-causal case. The causal case is addressed in Section 4. Section 5 specialises the results to the Gaussian case. The paper is concluded in Section 6.

2 Preliminaries and definitions

To specify the discrete memoryless state-dependent relay channel (DMSD-RC) depicted in Fig. 1, we define five finite sets: $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_2, \mathcal{Y}_3, \mathcal{S})$. A probability transition matrix $p(y_2, y_3 | x_1, x_2, s)$ is also defined for all $(x_1, x_2, y_2, y_3, s) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}_2 \times \mathcal{Y}_3 \times \mathcal{S}$. In this model, X_1 and X_2 are the source and the relay inputs, respectively. Y_3 and Y_2 are outputs of the destination and the relay, respectively. We also assume that the source and relay know an i.i.d noisy observation of states (partial CSI) drawn according to a known probability distribution $p(s, s_1, s_2)$, where $s \in \mathcal{S}, s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2$. This model is illustrated in Fig. 1. The CSI at the source (respectively at the relay) is perfect if $s_{1,n}$ (respectively $s_{2,n}$) is equal to s_n for each n . The source wants to transmit W to the destination with the help of relay in n channel uses.

A $(2^{nR}, n)$ code for the SD-RC consists of a message set $\mathcal{W} = \{1, \dots, 2^{nR}\}$, where the message W is uniformly distributed over the set \mathcal{W} , a decoding function at the destination: $d_3: \mathcal{Y}_3^n \rightarrow \mathcal{W}$ and a sequence of encoding functions at the relay where for non-causal CSI at the relay defined as

$$\phi_{2,i} : \mathcal{Y}_{2,1}^{i-1} \times \mathcal{S}_{2,1}^n \rightarrow \mathcal{X}_2 \quad \text{for } i = 1, \dots, n$$

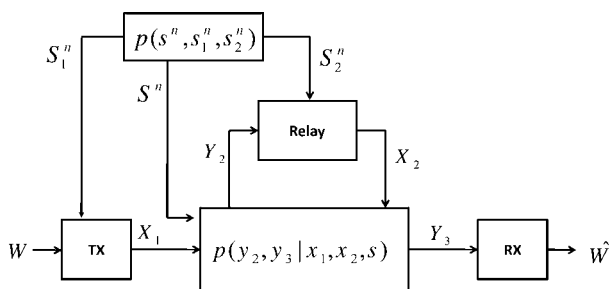


Figure 1 State-dependent RC

and for causal CSI as

$$\phi_{2,i} : \mathcal{Y}_{2,1}^{i-1} \times \mathcal{S}_{2,1}^i \rightarrow \mathcal{X}_2 \quad \text{for } i = 1, \dots, n$$

Definition 1: A rate R is said to be achievable for SD-RC, if there exists a sequence of codes $(2^{nR}, n)$ with average probability of error $P_e^{(n)} = \Pr(\hat{W} \neq W) \rightarrow 0$ as $n \rightarrow \infty$.

In this paper, upper case letters (e.g. X) are used to denote random variables, and their realisation are shown by lower case (e.g. x). We use $\|\mathcal{X}\|$ to denote the cardinality of finite discrete set \mathcal{X} . X_i^j indicates a sequence of random variables (X_i, \dots, X_j) . $p_X(x)$ denotes the probability mass function (p.m.f.) of X on \mathcal{X} , where occasionally subscript X is omitted. $\mathcal{A}_\epsilon^n(X, Y)$ denotes the set of ϵ -strongly, jointly typical n -length sequences based on $p(x, y)$ which may be indicated as \mathcal{A}_ϵ^n in the sequel, when it is obvious from the context.

3 SD-RC with non-causal CSI

In this section we consider DM-SDRC whose non-causal CSI is available partially at the source and the relay. In the following, achievable rate for this channel is derived in Theorem 1 based on combining the CF strategy at the relay and GP coding at the source and relay. As special cases, this theorem includes three different situations: The perfect CSI is known non-causally (i) only to the source, (ii) only to the relay and (iii) both to the source and the relay. Hence, as by-products of the following theorem, achievable rates for these three situations are provided.

Theorem 1: The capacity of the DMSD-RC with non-causal CSI \mathcal{S}_1 and \mathcal{S}_2 available at the source and the relay, respectively, is lower bounded by

$$C \geq R = \sup(I(U_1; \hat{Y}_2, Y_3 | U_2) - I(U_1; \mathcal{S}_1)) \quad (1)$$

$$\text{s.t. } I(U_2; Y_3) - I(U_2; \mathcal{S}_2) \geq I(\hat{Y}_2; Y_2, \mathcal{S}_2 | U_2, Y_3) \quad (2)$$

where the supremum is taken over all joint p.m.f. on $\mathcal{S} \times \mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{X}_1 \times \mathcal{X}_2 \times \hat{\mathcal{Y}}_2 \times \mathcal{Y}_2 \times \mathcal{Y}_3$ of the form

$$p(s, s_1, s_2, u_1, u_2, x_1, x_2, y_2, y_3, \hat{y}_2) = p(s, s_1, s_2)p(x_1, u_1 | s_1)p(x_2, u_2 | s_2) \times p(y_2, y_3 | x_1, x_2, s)p(\hat{y}_2 | y_2, u_2, s_2) \quad (3)$$

Proof: The proof of this theorem is based on using GP coding both at the source and relay and using the CF scheme at the relay. The proof appears in Appendix 1. □

Remark 1: Since the relay is informed of CSI \mathcal{S}_2 , it tries to use its knowledge for two different goals: (i) to cancel the effect of channel's state on its received signal (y_2) based on its knowledge of CSI, through acting as a decoder of the source-relay link and (ii) to compress the CSI (along with

its received signal, y_2) and send the result to the destination and let it to utilise partial CSI, where needed. As a simple example, if $y_2 = \emptyset$, the relay can compress only the CSI and send it to the destination. So the destination can utilise this partial CSI. Hence, in general, in order to achieve the two above-mentioned goals, we assume that in (3) \hat{y}_2 is conditioned on s_2 , besides u_2 and y_2 .

Remark 2: By setting $S_1 = S_2 = \emptyset$ and $U_1 = X_1$, $U_2 = X_2$ in (1)–(3), Theorem 1 reduces to the rate of classical RC in [19, Theorem 6].

Now, we specialise Theorem 1 to the cases where perfect CSI is available non-causally only at the source, only at the relay, and both at the source and relay.

Corollary 1: The capacity of the DMSD-RC with non-causal perfect CSI only at the source is lower bounded by

$$C \geq R = \sup(I(U_1; \hat{Y}_2, Y_3 | X_2) - I(U_1; S)) \quad (4)$$

$$\text{s.t. } I(X_2; Y_3) \geq I(\hat{Y}_2; Y_2 | X_2, Y_3) \quad (5)$$

where the supremum is taken over $p(s)p(x_1, u_1|s)p(x_2)p(y_2, y_3|x_1, x_2, s)p(\hat{y}_2|y_2, x_2)$.

Corollary 1 follows directly from Theorem 1 by setting $S_1 = S$, $S_2 = \emptyset$ and $U_2 = X_2$ in (1)–(3), since the perfect CSI is available only at the source, and only the source uses GP coding. By setting $S_2 = \emptyset$ it can be seen that \hat{y}_2 and s , given (y_2, x_2) become independent, which happens because the relay is not informed of CSI in Corollary 1.

Corollary 2: The capacity of the DMSD-RC with non-causal perfect CSI only at the relay is lower bounded by

$$C \geq R = \sup I(X_1; \hat{Y}_2, Y_3 | U_2) \quad (6)$$

$$\text{s.t. } I(U_2; Y_3) - I(U_2; S) \geq I(\hat{Y}_2; Y_2, S | U_2, Y_3) \quad (7)$$

where the supremum is taken over $p(s)p(x_1)p(x_2, u_2 | s)p(y_2, y_3 | x_1, x_2, s)p(\hat{y}_2 | y_2, u_2, s)$.

Corollary 2 follows directly from Theorem 1 by setting $S_1 = \emptyset$, $S_2 = S$ and $U_1 = X_1$ in (1)–(3), since the perfect CSI is available only at the relay, and only the relay uses GP coding.

Remark 3: By setting $S_1 = S_2 = S$ in (1)–(3), a lower bound on the capacity of DMSD-RC with non-causal perfect CSI at both the source and relay is derived.

4 SD-RC with causal CSI

In many practical applications, the state sequence is not known in advance, and can be known in a causal manner. Causal asymmetric CSI model may fit to the wireless

networks where some of the nodes could estimate the states of the channel (e.g. fading coefficients) with high accuracy. In this section, we consider DM-SDRC whose causal CSI is available partially at the source and the relay. In the following, achievable rate for this channel is derived in Theorem 2 based on using the Shannon's strategies at the source and relay by defining extended alphabet sets and the CF scheme. As special cases, this theorem includes three different situations: the perfect CSI is known causally (i) only to the source, (ii) only to the relay and (iii) both to the source and the relay. Hence, as by-products of the following theorem, achievable rates for these three situations are also provided.

Theorem 2: The capacity of the DMSD-RC with causal CSI S_1 and S_2 available at the source and the relay, respectively, is lower bounded by

$$C \geq R = \sup I(U_1; \hat{Y}_2, Y_3 | U_2) \quad (8)$$

$$\text{s.t. } I(U_2; Y_3) \geq I(\hat{Y}_2; Y_2, S_2 | U_2, Y_3) \quad (9)$$

where the supremum is taken over all joint p.m.f. on $S \times S_1 \times S_2 \times U_1 \times U_2 \times X_1 \times X_2 \times \hat{Y}_2 \times Y_2 \times Y_3$ of the form

$$\begin{aligned} & p(s, s_1, s_2, u_1, u_2, x_1, x_2, y_2, y_3, \hat{y}_2) \\ & = p(s, s_1, s_2)p(u_1)p(x_1|u_1, s_1)p(u_2)p(x_2|u_2, s_2) \\ & \quad \times p(y_2, y_3|x_1, x_2, s)p(\hat{y}_2|y_2, u_2, s_2) \end{aligned} \quad (10)$$

and $X_1 = f_1(U_1, S_1)$, $X_2 = f_2(U_2, S_2)$, where $f_1(\cdot)$ and $f_2(\cdot)$ are two arbitrary deterministic functions.

Proof: The proof of this theorem is based on combining Shannon's strategies and CF scheme. Since in this theorem the source and the relay are informed of CSI causally, we use Shannon's strategies at both the source and the relay with defining extended alphabet sets. Shannon's strategies transform the original channel into one, with auxiliary inputs U_1 and U_2 . The proof appears in Appendix 2. \square

Remark 4: The expression for the achievable rate of Theorem 2 can be interpreted as a special case of Theorem 1, where U_1 and U_2 are independent of S_1 and S_2 . This is similar to the relation between the expression for the capacity of the state-dependent single-user channel with causal CSI [2], and its non-causal counterpart, that is the GP channel [3]. Moreover, based on the mentioned goals in Remark 1, we consider that in (10) \hat{y}_2 is conditioned on s_2 , in addition to u_2, y_2 .

Now, we specialise Theorem 2 to the cases where perfect CSI is available causally only at the source, only at the relay and both at the source and relay.

Corollary 3: The capacity of the DMSD-RC with causal perfect CSI only at the source is lower bounded by

$$C \geq R = \sup I(U_1; \hat{Y}_2, Y_3 | X_2) \quad (11)$$

$$\text{s.t. } I(X_2; Y_3) \geq I(\hat{Y}_2; Y_2 | X_2, Y_3) \quad (12)$$

where the supremum is taken over $p(s)p(u_1)p(x_1|u_1, s)p(x_2|p(y_2, y_3 | x_1, x_2, s)p(\hat{y}_2 | y_2, x_2))$ and $X_1 = f_1(U_1, S)$, where $f_1(\cdot)$ is an arbitrary deterministic function.

Corollary 3 follows directly from Theorem 2 by setting $S_1 = S, S_2 = \emptyset$ and $U_2 = X_2$ in (8)–(10), since the perfect CSI is available causally only at the source, and Shannon’s strategy is used only at the source.

Corollary 4: The capacity of the DMSD-RC with causal perfect CSI only at the relay is lower bounded by

$$C \geq R = \sup I(X_1; \hat{Y}_2, Y_3 | U_2) \quad (13)$$

$$\text{s.t. } I(U_2; Y_3) \geq I(\hat{Y}_2; Y_2, S | U_2, Y_3) \quad (14)$$

where the supremum is taken over $p(s)p(x_1)p(u_2)p(x_2 | u_2, s)p(y_2, y_3 | x_1, x_2, s)p(\hat{y}_2 | y_2, u_2, s)$ and $X_2 = f_2(U_2, S)$, where $f_2(\cdot)$ is an arbitrary deterministic function.

Corollary 4 follows directly from Theorem 2 by setting $S_1 = \emptyset, S_2 = S$ and $U_1 = X_1$ in (8)–(10), since the perfect CSI is available causally only at the relay, and Shannon’s strategy is used only at the relay.

Remark 5: By setting $S_1 = S_2 = S$ in (8)–(10), a lower bound on the capacity of DMSD-RC with causal perfect CSI at both the source and relay is derived.

5 State-dependent Gaussian RC

In this section in order to compare our derived lower bounds with the established DF-based bounds, we provide some examples for both non-causal and causal case.

5.1 Non-causal case

In this part we consider a general full-duplex Gaussian RC with i.i.d. and additive Gaussian state and noise processes. We consider the following cases: perfect Gaussian channel state is known non-causally: (i) only to the source (ii) only to the relay. We use the results obtained in Section 3 to provide achievable rate in each case.

The received signals at the relay and the destination at time $j = 1, \dots, n$ for Gaussian SD-RC are given by

$$Y_{2j} = aX_{1j} + Z_{2j} + S_j \quad (15)$$

$$Y_{3j} = X_{1j} + bX_{2j} + Z_{3j} + S_j \quad (16)$$

where X_{1j} and X_{2j} are the signals transmitted by the transmitter and the relay with individual average power constraints P_1 and P_2 , respectively. Positive constants a and b are known and denote the channel gains. Z_{2j} and Z_{3j} , with powers N_2 and N_3 , are i.i.d. and independent white Gaussian noise components at the relay and the destination, respectively. The components of S are also i.i.d. and zero-mean Gaussian with $E[S_j^2] = Q$.

Theorem 3: For the state-dependent general Gaussian RC with non-causal perfect CSI at the source, the rate (see (17))

$$\text{s.t. } \hat{N} = \frac{\beta^2}{b^2 P_2} [N_2 A + N_3 D + (a - 1)^2 C Q] \quad (18)$$

is achievable, where $A \triangleq P_1 + Q + 2\sigma_{1s} + N_3, B \triangleq P_1 + \alpha^2 Q + 2\alpha\sigma_{1s}, C \triangleq P_1(1 - \rho_{1s}^2), D \triangleq a^2 P_1 + Q + 2a\sigma_{1s}, -1 \leq \rho_{1s}, \alpha \leq 1$ and $0 \leq \beta \leq 1$.

Proof: The detailed proof is given in Appendix 3. \square

Theorem 4: For the state-dependent general Gaussian RC with non-causal perfect CSI at the relay, the rate (see (19))

$$\text{s.t. } \hat{N} = \frac{\beta^2 N_2 G + E}{F - G} \quad (20)$$

is achievable, where $A \triangleq P_2 + \alpha^2 Q + 2\alpha\sigma_{2s}, B \triangleq (\alpha b - 1)^2 (1 - \rho_{2s}^2), C \triangleq \alpha^2 (1 - \rho_{2s}^2), D \triangleq (\gamma + \beta)/\alpha, E \triangleq P_2 Q C [P_1 (D - a\beta((1/\alpha) - b))^2 + D^2 N_3] + a^2 \beta^2 P_1 N_3 A, F \triangleq P_2 (1 - \rho_{2s}^2) (P_1 + N_3 + b^2 P_2 + Q + 2b\sigma_{2s}), G \triangleq P_2 Q B + (P_1 + N_3) A$ and $-1 \leq \rho_{2s}, \alpha, \gamma \leq 1$.

Remark 6: For $Q = 0$ (state-independent channel), Theorems 3 and 4 reduce to the achievable rate for CF scheme in the standard interference-free Gaussian RC in [22].

Proof: The proof is similar to the proof of Theorem 3, except that the relay is informed in this case. So, $E[X_1 S] = 0, U_2 = X_2 + \alpha S$ (DPC), $E[X_2 S] = \sigma_{2s} = \rho_{2s} \sqrt{P_2 Q}$, where $-1 \leq \rho_{2s} \leq 1$ (GDPC) and $\hat{Y}_2 = \beta Y_2 + \gamma S + \hat{Z}$ (CSI is known to the relay, and it can compress its knowledge and send it

$$\max_{\rho_{1s}, \alpha, \beta} \frac{1}{2} \log \frac{\hat{N} (b^2 P_2 + A)}{(\beta^2 N_2 + \hat{N}) [(\alpha - 1)^2 Q + (BN_3/C)] + \beta^2 (a\alpha - 1)^2 N_3 Q} \quad (17)$$

$$\max_{\rho_{2s}, \alpha, \beta, \gamma} \frac{1}{2} \log \left(1 + P_1 \frac{[\beta^2 (a^2 N_3 + N_2) + \hat{N}] A + P_2 Q C (D - a\beta((1/\alpha) - b))^2}{(\beta^2 N_2 + \hat{N}) (G - P_1 A) + D^2 P_2 Q C N_3} \right) \quad (19)$$

to the destination to partially cancel the CSI). β is an arbitrary number and we also let γ to be negative in order to consider the role of relay as a channel state canceller of the source-relay link. Therefore we should maximise the rate in (19) subject to the condition (20), over parameters $-1 \leq \alpha, \rho_2, \gamma \leq 1$. \square

Now, we plot the derived achievable rates in (17) and (19), for some examples to compare with the corresponding DF-based lower and upper bounds for the state-dependent general Gaussian case in [11, 13].

In Fig. 2, the lower bound in (17), and for comparison, the lower and upper bounds in [11] are plotted for $a = 2, b = 1, P_1 = Q = N_3 = N_2 = 10$ dB against (P_2/N_3) . This figure shows that, although the source-relay link has a good condition ($a = 2, b = 1$), but at high (P_2/N_3) , CF-based lower bound outperforms DF. If we set $a = b = 1$ in the mentioned example, DF-based lower bound performs worse than what has been shown in Fig. 2.

Then, in Fig. 3, we consider $a = b = 1, Q = N_3 = N_2 = 10$ dB and $(P_2/N_3 = P_1/N_3 = \text{SNR})$ varies. We can see that in this case, CF outperforms DF. However, if we increase b (improving multiple-access side of the RC) the CF-based lower bound becomes very close to the upper bound at high SNR (e.g. for $b = 4$ the gap between the CF bound and upper bound is 0.0413 bits at 30 dB, and for $b = 8$, gap = 0.0110 bits at 26 dB), and can achieve almost tight rates for the general Gaussian case. As illustrated in [11], at high SNR, PDF-based lower bounds achieve the same rates as DF-based bounds. So at high SNR our derived bounds for these cases outperform PDF-based bounds, too.

In Fig. 4, the lower bound in (19), and also the lower and the upper bounds in [13] are plotted for $a = b = 1, P_1 = P_2 = 20$ dB and $Q = N_2 = 10$ dB, as

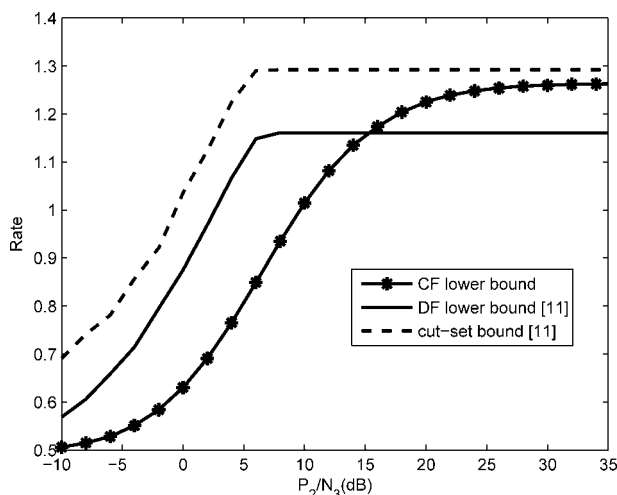


Figure 2 Lower and upper bounds obtained by different methods for state-dependent general Gaussian RC with non-causal CSI known only to the source, $a = 2, b = 1, P_1 = Q = N_3 = N_2 = 10$ dB

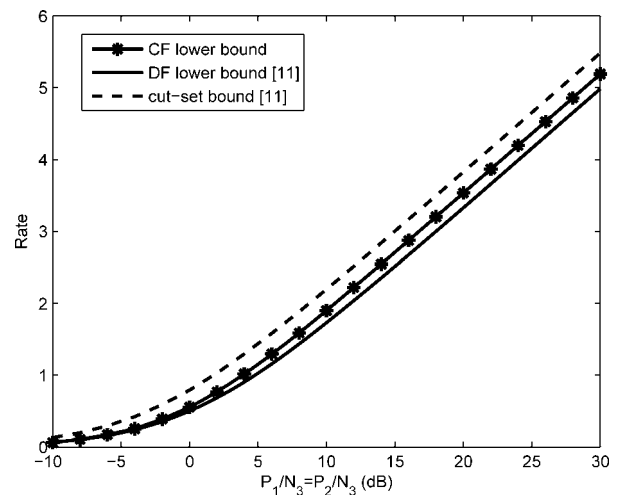


Figure 3 Lower and upper bounds obtained by different methods for state-dependent general Gaussian RC with non-causal CSI known only to the source, $a = b = 1, Q = N_3 = N_2 = 10$ dB

functions of SNR (P_1/N_3) . It is observed that for SNR values higher than 10 dB, CF outperforms the DF lower bound and can achieve the rates nearly close to the upper bound. In this case γ is approximately equal to $-\beta$, which shows that the relay tries to achieve near optimality, through cancelling the known state and then compresses the results (i.e. $\hat{Y}_2 = \beta(Y_2 - S) + \hat{Z}$), and re-encodes it to send to the destination.

In Fig. 5, the lower and upper bounds are plotted as function of the interference power Q , for fixed value of the power at the source and the relay when CSI is only available at the relay. The curves are depicted for $a = b = 1, P_1 = 25$ dB, $P_2 = N_2 = 20$ dB and $N_3 = 10$ dB. This figure shows that CF-based lower bound is a decreasing function of Q . We can see that CF-based

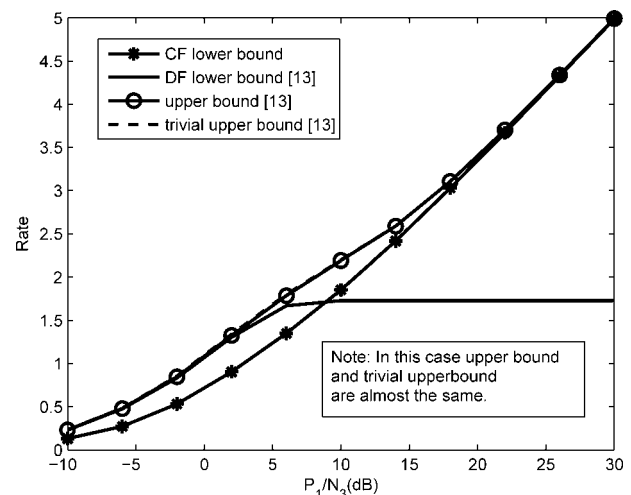


Figure 4 Lower and upper bounds obtained by different methods for state-dependent general Gaussian RC with non-causal CSI known only to the relay, $a = b = 1, P_1 = P_2 = 20$ dB, $Q = N_2 = 10$ dB

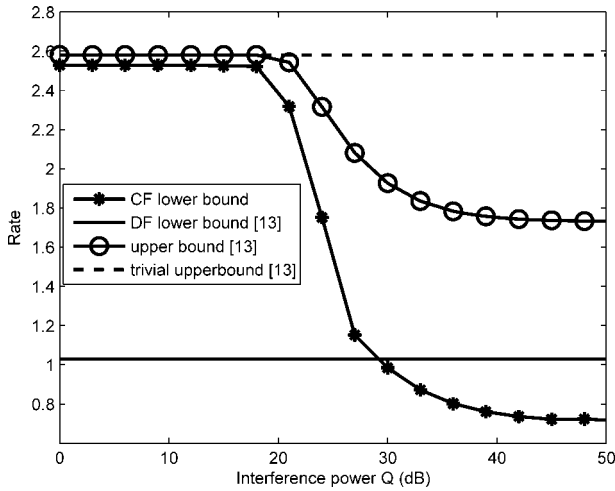


Figure 5 Lower and upper bounds obtained by different methods for state-dependent general Gaussian RC with non-causal CSI known only to the relay, $a = b = 1$, $P_1 = 25$ dB, $P_2 = N_2 = 20$ dB, $N_3 = 10$ dB

achievable rate for very large values of Q is strictly positive which illustrates that, even in presence of strong interference, the transmission from the uninformed source to the uninformed destination is possible. Furthermore, we can see that for values of Q between 0 to 30 dB, DF performs worse than the CF-based lower bound.

5.2 Causal case

As mentioned before, for the causal case optimal codes are constructed over an extended input alphabets $\mathcal{X}^{\|\mathcal{S}\|}$ (or, equivalently over the alphabet of mappings from $\mathcal{S} \rightarrow \mathcal{X}$). So, if $\|\mathcal{S}\|$ is infinite in limit, it may cause some practical problems for code construction. Hence, for simplicity in evaluating the derived achievable rates in Section 4, we consider binary channel states. In fact, we consider a binary flat fading AWGN RC. So, the received signals at the relay and destination at time $j = 1, \dots, n$ are given by

$$Y_{2j} = X_{1j} + Z_{2j} \tag{21}$$

$$Y_{3j} = S_j X_{1j} + X_{2j} + Z_{3j} \tag{22}$$

where the parameters in (21)–(22) defined similar to the definitions for (15)–(16). The only difference is in the definition of S_j , where we assume i.i.d. on/off fading coefficients with $p(s_j = 1) = 1 - p(s_j = 0) = \alpha$, here. Note that S_j is a binary fading coefficient which influences on the source–destination link. So if $s_j = 1$, the considered model reduces to classic RC and if $s_j = 0$, it reduces to an RC without direct link between source and destination. As an example, we consider the case where causal CSI is available only at the source. Hence, for the mentioned scenario, we evaluate the achievable rate and its condition in (11) and (12) by choosing input probability distribution $p(u_1), p(x_2), p(\hat{y}_2 | y_2, x_2)$ and deterministic function $X_1 = f_1(U_1, S)$. Similar to Section 5.1, we evaluate the related

terms for a Gaussian distribution. We take deterministic function $X_1 = f_1(U_1, S) = SU_1$ and the following probability distribution

$$p(u_1) = \frac{1}{\sqrt{2\pi P_1/\alpha}} \exp\left(\frac{-u_1^2}{2P_1/\alpha}\right) \tag{23}$$

Note that the chosen deterministic function f_1 is not necessarily optimal, but it results in an achievable rate for the mentioned scenario. Let $X_2 \sim \mathcal{N}(0, P_2)$, which is independent of U_1 (because of CF scheme), and let $\hat{Y}_2 = Y_2 + \hat{Z}$, where compression noise $\hat{Z} \sim \mathcal{N}(0, \hat{N})$ is independent of S, U_1, X_2, Z_2, Z_3 . Hence, we have

$$p(\hat{y}_2 | y_2) = p_{\hat{Z}}(\hat{y}_2 - y_2) = \frac{1}{\sqrt{2\pi\hat{N}}} \exp\left(\frac{-(\hat{y}_2 - y_2)^2}{2\hat{N}}\right)$$

For the above specified parameters, we have

$$p(y_3) = (1 - \alpha) \frac{1}{\sqrt{2\pi(N_3 + P_2)}} \exp\left(\frac{-y_3^2}{2(N_3 + P_2)}\right) + \alpha \frac{1}{\sqrt{2\pi((P_1/\alpha) + N_3 + P_2)}} \exp\left(\frac{-y_3^2}{2((P_1/\alpha) + N_3 + P_2)}\right)$$

By calculating other probability distributions in (11) and (12), we numerically evaluate the rate and its condition in Corollary 3.

Moreover, for the considered model in (21)–(22), following similar steps and based on CF scheme we also calculate achievable rate for the case where the source ignores its knowledge about CSI, and we denote it as ‘state-ignorant CF scheme’. Furthermore, to provide DF-based achievable rate for the considered model in (21)–(22) with causal CSI at the source, we exploit the derived

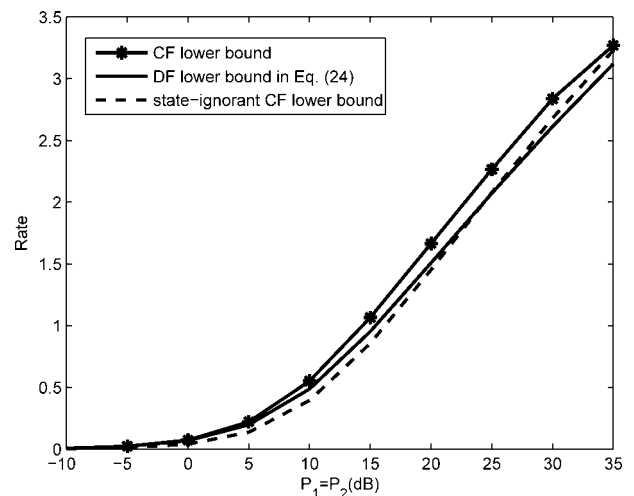


Figure 6 Lower bounds obtained by different methods for the binary fading AWGN RC with causal CSI known only to the source, $\alpha = 0.5$, $N_2 = N_3 = 10$ dB

Table 1 Lower bounds (in bps/Hz) obtained by different methods for the binary fading AWGN RC, $P_1 = 20$ dB, $P_2 = 10$ dB, $N_2 = N_3 = 10$ dB

α	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.8$	$\alpha = 1$
Coding strategies					
DF scheme with causal CSI at the source in (24)	0.65	1.22	1.50	1.69	1.72
CF scheme with causal CSI at the source	0.68	1.27	1.56	1.73	1.76
state-ignorant CF scheme	0.50	0.92	1.25	1.60	1.76

DF-based lower bound in [11, Corollary 2]. Note that this lower bound is originally for SD-RC with non-causal CSI available only at the source, but by setting U_1 and S given X_2 to be independent in [11, Corollary 2] (i.e. setting $I(U_1; S|X_2) = 0$) and setting $X_1 = f_1(U_1, S)$, this bound reduces to the DF-based lower bound for SD-RC with causal CSI only at the source, which becomes as follows

$$R_{\text{DF}}^{\text{causal}} = \max \min \{I(U_1; Y_2 | X_2), I(U_1, X_2; Y_3)\} \quad (24)$$

where the maximisation is over $p(s)p(x_2)p(u_1|x_2)$. $X_1 = f_1(U_1, S)$ where $f_1(\cdot)$ is an arbitrary deterministic function. (Note that the bound in (24) is the same as the derived DF-based lower bound in [15, Theorem 2], by setting $S_s = S$, $S_r = S_d = \emptyset$, $V = X_2$ in [15, Theorem 2] (since here the case in which causal CSI known only to the source is considered).)

Hence, to numerically evaluate the DF-based lower bound in (24), we choose f_1 as in the CF scheme (i.e. $X_1 = f_1(U_1, S) = U_1 S$), and choose $p(u_1)$ as in (23) and $X_2 \sim \mathcal{N}(0, P_2)$, and maximising the rate over jointly Gaussian probability distribution for $p(u_1, x_2)$. Note that in the DF scheme U_1 and X_2 are dependent. These computed lower bounds (CF lower bound, DF lower bound in (24) and state-ignorant CF lower bound) are plotted for $\alpha = 0.5$, $N_2 = N_3 = 10$ dB against $P_1 = P_2$ in Fig. 6. Moreover, the results for different values of α and for $P_1 = 20$ dB, $P_2 = 10$ dB, $N_2 = N_3 = 10$ dB are illustrated in Table 1.

The results in Fig. 6 show that for these parameters, CF-based bound outperforms DF-based bound for causal CSI. For the assumed parameters in Table 1, DF and CF lower bound works nearly the same. Also as it was expected, by knowing causal CSI at the source higher rates are achieved with respect to the case where the source ignores its knowledge about CSI in state-dependent channel.

6 Conclusion

We investigated SD-RC with causal and non-causal CSI to gain a unified view. We considered SD-RC whose non-causal CSI is available partially at the source and the relay. We established achievable rate for this channel based on the combination of the CF strategy with GP coding at the

source and the relay. As special cases, this set-up includes three distinct scenarios in which the perfect non-causal CSI is available only at the source, only at the relay and both at the source and relay. We also investigated similar set-up for the SD-RC with causal CSI, and based on using Shannon's strategy and the CF scheme, a lower bound on the capacity of the SD-RC with causal CSI for this set-up was established. We showed that, similar to the relation between the expression for the capacity of single-user state-dependent channel with causal CSI, and its non-causal counterpart, the expression for the achievable rate of SD-RC with causal CSI is a special case of the non-causal situation. We further illustrated our results for non-causal case via Gaussian examples and presented some cases for the general Gaussian RC with states, in which our derived lower bounds outperform the DF-based lower bounds derived in [11, 13], and can achieve rates nearly close to the derived upper bounds in [11, 13]. For causal case, a numerical example of the binary fading Gaussian RC with additive Gaussian noise was provided. In future work we intend to study the SD-RC with private messages.

7 Acknowledgments

This work was partially supported by Iranian National Science Foundation (INSF) under contract No. 84,5193-2006 and by Iran Telecommunication Research Center (ITRC) under contract No. T500/20958. The materials in this paper have been presented in part at the IEEE International Symposium on Information Theory (ISIT), Korea, 28 June–3 July 2009, and IEEE Information Theory Workshop (ITW), Italy, 11–16 October 2009.

8 References

- [1] KESHET G., STEINBERG Y., MERHAV N.: 'Channel coding in the presence of side information: subject review', *Found. Trends Commun. Inf. Theory*, 2007, **4**, (6), pp. 445–586
- [2] SHANNON C.E.: 'Channels with the side information at the transmitter', *IBM J. Res. Dev.*, 1958, **2**, pp. 289–293
- [3] GEL'FAND S., PINSKER M.: 'Coding for channels with random parameters', *Probab. Control Inf. Theory*, 1980, **9**, (1), pp. 19–31

- [4] COSTA M.H.M.: 'Writing on dirty paper', *IEEE Trans. Inf. Theory*, 1983, **29**, (3), pp. 439–441
- [5] JAFAR S.A.: 'Capacity with causal and non-causal side information: a unified view', *IEEE Trans. Inf. Theory*, 2006, **52**, (12), pp. 5468–5474
- [6] KIM Y.H., SUTIVONG A., SIGURJONSSON S.: 'Multiple user writing on dirty paper'. Proc. IEEE Int. Symp. on Information Theory (ISIT'04), Chicago, IL, June/July 2004, p. 534
- [7] SIGURJONSSON S., KIM Y.H.: 'On multiple user channels with state information at the transmitters'. Proc. IEEE Int. Symp. on Information Theory (ISIT'05), Adelaide, Australia, September 2005, pp. 72–76
- [8] STEINBERG Y.: 'Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information', *IEEE Trans. Inf. Theory*, 2005, **51**, (8), pp. 2867–2877
- [9] SOMEKH-BARUCH A., SHAMAI (SHITZ) S., VERDÚ S.: 'Cooperative multiple-access encoding with states available at one transmitter', *IEEE Trans. Inf. Theory*, 2008, **54**, (10), pp. 4448–4469
- [10] SOMEKH-BARUCH A., SHAMAI (SHITZ) S., VERDÚ S.: 'Cognitive interference channels with state information'. Proc. IEEE Int. Symp. on Information Theory (ISIT'08), Toronto, Canada, July 2008, pp. 1353–1357
- [11] ZAIDI A., VANDENDORPE L.: 'Lower bounds on the capacity of relay channel with states at the source', *EURASIP J. Wirel. Commun. Netw.*, 2009, Article ID: 634296; doi:10.1155/2009/634296
- [12] KHOSRAVI-FARSANI R., AKHBARI B., MIRMOHSENI M., AREF M.R.: 'Cooperative relay-broadcast channels with causal channel state information'. Proc. IEEE Int. Symp. on Information Theory (ISIT'09), Seoul, Korea, June/July 2009, pp. 1174–1178
- [13] ZAIDI A., KOTAGIRI S., LANEMAN J.N., VANDENDORPE L.: 'Cooperative relaying with state at the relay'. Proc. IEEE Information Theory Workshop (ITW'08), Porto, Portugal, May 2008, pp. 139–143
- [14] ZAIDI A., KOTAGIRI S., LANEMAN J.N., VANDENDORPE L.: 'Cooperative relaying with state available non-causally at the relay', *IEEE Trans. Inf. Theory*, To be published, available online at <http://arxiv.org/abs/0911.4704>
- [15] KHORMUJI M.N., SKOGLUND M.: 'The relay channel with partial causal state information'. Proc. IEEE Int. Symp. on Information Theory and its Application (ISITA'08), Auckland, New Zealand, December 2008, pp. 1–6
- [16] KHORMUJI M.N., SKOGLUND M.: 'On cooperative downlink transmission with frequency reuse'. Proc. IEEE Int. Symp. on Information Theory (ISIT'09), Seoul, Korea, June/July 2009, pp. 849–853
- [17] KHORMUJI M.N., SKOGLUND M.: 'Instantaneous forwarding strategies for relay channels with known interference'. Proc. IEEE Information Theory Workshop (ITW'09), Volos, Greece, June 2009, pp. 42–46
- [18] VAN DER MEULEN E.C.: 'Three-terminal communication channels', *Adv. Appl. Probab.*, 1971, **3**, pp. 120–154
- [19] COVER T.M., EL GAMAL A.: 'Capacity theorems for relay channels', *IEEE Trans. Inf. Theory*, 1979, **25**, (5), pp. 572–584
- [20] EL GAMAL A., AREF M.R.: 'The capacity of semi-deterministic relay channel', *IEEE Trans. Inf. Theory*, 1982, **28**, (3), p. 536
- [21] AREF M.R.: 'Information flow in relay networks'. PhD dissertation, Stanford University, Stanford, CA, October 1980
- [22] ZAHEDI S.: 'On reliable communication over relay networks'. PhD dissertation, Stanford University, Stanford, CA, August 2005
- [23] EL GAMAL A., HASSANPOUR N., MAMMEN J.: 'Relay networks with delays', *IEEE Trans. Inf. Theory*, 2007, **53**, (10), pp. 3413–3431
- [24] LIANG Y., VEERAVALLI V.V.: 'Cooperative relay broadcast channels', *IEEE Trans. Inf. Theory*, 2007, **53**, (3), pp. 900–928
- [25] LIANG Y., KRAMER G.: 'Rate region for relay broadcast channels', *IEEE Trans. Inf. Theory*, 2007, **53**, (10), pp. 3517–3535
- [26] TANNIOUS R., NOSRATINIA A.: 'Relay channel with private messages', *IEEE Trans. Inf. Theory*, 2007, **53**, (10), pp. 3777–3785
- [27] KRAMER G., GASTPAR M., GUPTA P.: 'Cooperative strategies and capacity theorems for relay networks', *IEEE Trans. Inf. Theory*, 2005, **51**, (9), pp. 3037–3063
- [28] GHABELI L., AREF M.R.: 'Symmetric relaying based on partial decoding and the capacity of a class of relay networks', *IET Commun.*, 2009, **3**, (1), pp. 151–159
- [29] YASSAEI M.H., AREF M.R.: 'Generalized compress-and-forward strategy for relay networks'. Proc. IEEE Int. Symp. on Information Theory (ISIT'08), Toronto, Canada, June 2008, pp. 2683–2687
- [30] ZAIDI A., VANDENDORPE L., DUHAMEL P.: 'Lower bounds on the capacity regions of the relay channel and the cooperative relay-broadcast channel with non-causal side

information'. Proc. IEEE Communication Conf. (ICC'07), Glasgow, Scotland, June 2007, pp. 6005–6011

[31] COVER T.M., THOMAS J.A.: 'Elements of information theory'. Wiley Series in Telecommunications (Wiley, 2006, 2nd edn.)

[32] KAILATH T.: 'Lectures on Wiener and Kalman filtering' (Springer Verlag, 1985, 2nd edn.)

9 Appendix 1

9.1 Proof of Theorem 1

Outline of the proof: The proof is based on random coding scheme which combines GP coding at the source and the relay where partial CSI is available, and CF strategy at the relay. We consider a block Markov encoding scheme where a sequence of $B - 1$ messages are transmitted in B blocks, each containing n symbols. As $B \rightarrow \infty$, for fixed n , the rate $R(B - 1)/B$ is arbitrarily close to R .

Random coding: For any joint p.m.f. defined in (3):

1. Generate $2^{n(R+R')}$ i.i.d. u_1^n sequences each with probability $p(u_1^n) = \prod_{j=1}^n p(u_{1j})$. Label these $u_1^n(w, m)$, where $w \in [1, 2^{nR}]$ and $m \in [1, 2^{nR'}]$.

2. Generate $2^{n(R_2+R'_2)}$ i.i.d. u_2^n sequences each with probability $p(u_2^n) = \prod_{j=1}^n p(u_{2j})$. Label these $u_2^n(t, k)$, where $t \in [1, 2^{nR_2}]$ and $k \in [1, 2^{nR'_2}]$.

3. For each $u_2^n(t, k)$, generate $2^{n\hat{R}_2}$ i.i.d. \hat{y}_2^n sequences according to $p(\hat{y}_2^n | u_2^n) = \prod_{j=1}^n p(\hat{y}_{2j} | u_{2j})$, where for every $u_2 \in \mathcal{U}_2$, we define

$$p(\hat{y}_2 | u_2) = \frac{\sum_{s, s_1, s_2, u_1, x_1, x_2, y_2, y_3} p(s, s_1, s_2, u_1, u_2, x_1, x_2, y_2, y_3, \hat{y}_2)}{\sum_{s, s_1, s_2, u_1, x_1, x_2, y_2, y_3, \hat{y}_2} p(s, s_1, s_2, u_1, u_2, x_1, x_2, y_2, y_3, \hat{y}_2)}$$

Label these $\hat{y}_2^n(z | t)$, $z \in [1, 2^{n\hat{R}_2}]$.

4. Randomly partition the set $\{1, \dots, 2^{n\hat{R}_2}\}$ into 2^{nR_2} bins, defined as $B(t)$ where $t \in [1, 2^{nR_2}]$.

Encoding (at the beginning of block i): We assume that the channel state S_1^n and S_2^n in each block is non-causally known to the source and the relay, respectively.

1. Let w_i be the new message to be sent in block i . The source looks for the smallest m such that $(u_1^n(w_i, m), s_1^n(i)) \in A_\epsilon^n(U_1, S_1)$. Denote this m with m_i . Based on the GP coding [3], there exists such an index m_i , if n is large enough and

$$R' \geq I(U_1; S_1) \quad (25)$$

Based on the chosen m_i , the source given $u_1^n(w_i, m_i)$ in block i transmits i.i.d. $x_1^n(w_i)$ according to $p(x_1^n | u_1^n, s_1^n)$.

2. At the relay, assume that $(\hat{y}_2^n(z_{i-1} | t_{i-1}), y_2^n(i-1), u_2^n(t_{i-1}, k_{i-1}), s_2^n(i-1)) \in A_\epsilon^n$ and $z_{i-1} \in B(t_i)$. Knowing t_i and $s_2^n(i)$, the relay seeks for the smallest k such that $(u_2^n(t_i, k), s_2^n(i)) \in A_\epsilon^n$. Denote this k as k_i . Based on the GP coding, for sufficiently large n , there exists such an index k_i , if

$$R'_2 \geq I(U_2; S_2) \quad (26)$$

Then the relay given $(u_2^n(t_i, k_i), s_2^n(i))$ in block i , transmits i.i.d. $x_2^n(t_i)$ sequence, drawn according to the marginal $p(x_2^n | u_2^n, s_2^n)$.

Decoding (at the end of block i):

1. The relay finds a unique index z such that $(\hat{y}_2^n(z | t_i), y_2^n(i), u_2^n(t_i, k_i), s_2^n(i)) \in A_\epsilon^n$. Using [19, Lemmas 1, 2] the probability P_z that there is no such z is bounded by

$$P_z \leq (P\{(\hat{y}_2^n(z | t_i), y_2^n(i), u_2^n(t_i, k_i), s_2^n(i)) \notin A_\epsilon^n\})^{2^{n\hat{R}_2}} \leq (1 - (1 - \epsilon)2^{-n[8\epsilon + I(\hat{Y}_2; Y_2, S_2 | U_2)]})^{2^{n\hat{R}_2}} \stackrel{(a)}{\leq} \exp(-(1 - \epsilon)2^{n\hat{R}_2} 2^{-n[8\epsilon + I(\hat{Y}_2; Y_2, S_2 | U_2)]})$$

where (a) follows from $(1 - \mu x)^F \leq \exp(-\mu F x)$ for $0 \leq x \leq 1$. Hence, we see that as long as

$$\hat{R}_2 > I(\hat{Y}_2; Y_2, S_2 | U_2) + 8\epsilon \quad (27)$$

and for sufficiently large n , there exists such an index z with arbitrarily high probability.

2. The destination finds unique (\hat{t}_i, \hat{k}_i) such that $(u_2^n(\hat{t}_i, \hat{k}_i), y_3^n(i)) \in A_\epsilon^n$. This step can be done with arbitrarily small probability of error (i.e. $\hat{t}_i = t_i, \hat{k}_i = k_i$), if n is sufficiently large and

$$R_2 + R'_2 < I(U_2; Y_3) \quad (28)$$

3. Knowing t_{i-1} and k_{i-1} (from the previous block), the destination calculates a set of indices z such that

$$\mathcal{L}(y_3^n(i-1)) \triangleq (\hat{y}_2^n(z | t_{i-1}), u_2^n(t_{i-1}, k_{i-1}), y_3^n(i-1)) \in A_\epsilon^n(\hat{Y}_2, U_2, Y_3)$$

Then the destination declares that \hat{z}_{i-1} has been sent in block $i-1$, if $\hat{z}_{i-1} \in B(t_i) \cap \mathcal{L}(y_3^n(i-1))$. We compute the probability that z_{i-1} was chosen incorrectly. We write F_{i-1}^c for the event 'all decisions in block $i-1$ were correct'.

Thus we first bound $E[\|\mathcal{L}(y_3^n(i-1))\| | F_{i-1}^c]$

$$E[\|\mathcal{L}(y_3^n(i-1))\| | F_{i-1}^c] = E[\psi(z_{i-1} | y_3^n(i-1)) | F_{i-1}^c] + \sum_{z \neq z_{i-1}} E[\psi(z | y_3^n(i-1)) | F_{i-1}^c]$$

where

$$\psi(z | y_3^n(i-1)) = \begin{cases} 1 & \text{if } (\hat{y}_2^n(z | t_{i-1}), u_2^n(t_{i-1}, k_{i-1}), y_3^n(i-1)) \in A_\epsilon^n \\ 0 & \text{otherwise} \end{cases}$$

Now, for $z \neq z_{i-1}$ in $\mathcal{L}(y_3^n(i-1))$, Y_3 and \hat{Y}_2 are jointly independent given U_2 . From [31, Chapter 15]

$$E[\psi(z | y_3^n(i-1)) | F_{i-1}^c] \leq 2^{-n[-6\epsilon + I(\hat{Y}_2; Y_3 | U_2)]} \quad z \neq z_{i-1}$$

Since there are $2^{n\hat{R}_2} - 1$ choices for $z \neq z_{i-1}$

$$E[\|\mathcal{L}(y_3^n(i-1))\| | F_{i-1}^c] = 1 + (2^{n\hat{R}_2} - 1)2^{-n[-6\epsilon + I(\hat{Y}_2; Y_3 | U_2)]} \leq 1 + 2^{n\hat{R}_2} 2^{-n[-6\epsilon + I(\hat{Y}_2; Y_3 | U_2)]}$$

As long as t_i has been decoded correctly, an error is made only if there is a $z \neq z_{i-1}$ in $\mathcal{L}(y_3^n(i-1))$ which its bin number is t_i . Thus, the error occurred in step 3, is

$$\begin{aligned} P(\text{error in step3}) &\leq P\{\exists z \neq z_{i-1} \text{ such that } z \in B(t_i) \\ &\quad \cap \mathcal{L}(y_3^n(i-1)) | F_{i-1}^c\} \\ &\leq E \left[\sum_{z \neq z_{i-1}, z \in \mathcal{L}(y_3^n(i-1))} P(z \in B(t_i) | F_{i-1}^c) \right] \\ &\leq E[\|\mathcal{L}(y_3^n(i-1))\| | F_{i-1}^c] 2^{-nR_2} \\ &\leq (1 + 2^{n\hat{R}_2} 2^{-n[-6\epsilon + I(\hat{Y}_2; Y_3 | U_2)]}) 2^{-nR_2} \end{aligned}$$

We see that as long as

$$R_2 > \hat{R}_2 - I(\hat{Y}_2; Y_3 | U_2) + 6\epsilon \quad (29)$$

the destination can determine z_{i-1} reliably.

4. Finally, the destination uses both $\hat{y}_2^n(z_{i-1} | t_{i-1})$ and $y_3^n(i-1)$, and declares that \hat{w}_{i-1} and \hat{m}_{i-1} have been sent in block $i-1$, if there are unique indices such that

$$(u_1^n(\hat{w}_{i-1}, \hat{m}_{i-1}), y_3^n(i-1), \hat{y}_2^n(z_{i-1} | t_{i-1}), u_2^n(t_{i-1}, k_{i-1})) \in A_\epsilon^n$$

Using [19, Lemma 2], with arbitrarily high probability $\hat{w}_{i-1} = w_{i-1}$ and $\hat{m}_{i-1} = m_{i-1}$, if n is sufficiently large and

$$R + R' < I(U_1; Y_3, \hat{Y}_2 | U_2) \quad (30)$$

Now combining (26)–(29) yields

$$\begin{aligned} I(U_2; Y_3) - I(U_2; S_2) &\geq I(\hat{Y}_2; Y_2, S_2 | U_2) - I(\hat{Y}_2; Y_3 | U_2) \\ &= H(\hat{Y}_2 | U_2, Y_3) - H(\hat{Y}_2 | Y_2, S_2, U_2) \\ &\stackrel{(a)}{=} H(\hat{Y}_2 | U_2, Y_3) - H(\hat{Y}_2 | Y_2, S_2, U_2, Y_3) \\ &= I(\hat{Y}_2; Y_2, S_2 | U_2, Y_3) \end{aligned}$$

where (a) follows from the fact that \hat{Y}_2 is independent of Y_3 given Y_2, S_2, U_2 .

Combining (25) and (30) also yields (1). Thus the rate in (1) s.t. (2) is achievable. \square

10 Appendix 2

10.1 Proof of Theorem 2

Outline of the proof: As already mentioned, we combine CF strategy at the relay with Shannon's strategies both at the source and relay where partial CSI is available with defining extended alphabet sets. Similar to the proof of the previous theorem, a block Markov encoding scheme is considered.

Random coding: For any joint p.m.f. defined in (10):

1. Generate 2^{nR} i.i.d. u_1^n sequences according to $p(u_1^n) = \prod_{j=1}^n p(u_{1j})$. Index them as $u_1^n(w)$ where $w \in [1, 2^{nR}]$.
2. Generate 2^{nR_2} i.i.d. u_2^n sequences each with probability $p(u_2^n) = \prod_{j=1}^n p(u_{2j})$. Index them as $u_2^n(t)$ where $t \in [1, 2^{nR_2}]$.
3. For each $u_2^n(t)$, generate $2^{n\hat{R}_2}$ i.i.d. \hat{y}_2^n sequences each with probability $p(\hat{y}_2^n | u_2^n) = \prod_{j=1}^n p(\hat{y}_{2j} | u_{2j})$, where for $u_2 \in \mathcal{U}_2$ and $\hat{y}_2 \in \hat{\mathcal{Y}}_2$ we define

$$\begin{aligned} p(\hat{y}_2 | u_2) &= \sum_{s, s_1, s_2, u_1, x_1, x_2, y_2, y_3} p(s, s_1, s_2) p(u_1) p(x_1 | u_1, s_1) \\ &\quad \times p(x_2 | u_2, s_2) p(y_2, y_3 | x_1, x_2, s) p(\hat{y}_2 | y_2, u_2, s_2) \end{aligned} \quad (31)$$

Index them as $\hat{y}_2^n(z | t)$ where $z \in [1, 2^{n\hat{R}_2}]$.

4. Randomly partition the set $\{1, \dots, 2^{n\hat{R}_2}\}$ into 2^{nR_2} bins defined as $B(t)$ where $t \in [1, 2^{nR_2}]$.

Encoding (at the beginning of block i):

1. Let w_i be the new message to be sent from the source in block i . Upon receiving $s_{1j}(i)$ (the value of the known state process at time $j(1 \leq j \leq n)$ in block i at the source), the source sends $x_{1j}(i) = f_1(u_{1j}(w_i), s_{1j}(i))$.

2. At the relay assume

$$\begin{aligned} &(\hat{y}_2^n(z_{i-1}|t_{i-1}), y_2^n(i-1), u_2^n(t_{i-1}), s_2^n(i-1)) \\ &\in \mathcal{A}_\epsilon^n(\hat{Y}_2, Y_2, U_2, S_2) \end{aligned}$$

and $z_{i-1} \in B(t_i)$. Upon receiving $s_{2j}(i)$ (the value of the known state process at time j ($1 \leq j \leq n$) in block i at the relay), the relay sends $x_{2j}(i) = f_2(u_{2j}(t_i), s_{2j}(i))$.

Note that, although the channel state S_2 in block i is causally known to the relay (at time j , knows only the CSI S_2 from time 1 to j), but the relay knows completely the value of the state process of block $i-1$ (i.e. $s_2^n(i-1)$).

Decoding (at the end of block i):

1. The relay seeks a unique index z such that

$$(\hat{y}_2^n(z|t_i), y_2^n(i), u_2^n(t_i), s_2^n(i)) \in \mathcal{A}_\epsilon^n(\hat{Y}_2, Y_2, U_2, S_2) \quad (32)$$

For sufficiently large n , similar to the proof of Step 1 in Theorem 1, the probability that the relay can find such an index z is arbitrarily high, if

$$\hat{R}_2 > I(\hat{Y}_2; Y_2, S_2|U_2) \quad (33)$$

2. The destination finds a unique \hat{t}_i such that $(u_2^n(\hat{t}_i), y_3^n(i)) \in \mathcal{A}_\epsilon^n(U_2, Y_3)$. With arbitrarily high probability $\hat{t}_i = t_i$, if n is sufficiently large and

$$R_2 \leq I(U_2; Y_3) \quad (34)$$

3. Knowing t_{i-1} (from the previous block), the destination makes a list code of indices z such that

$$\begin{aligned} \mathcal{L}(y_3^n(i-1)) &\triangleq (\hat{y}_2^n(z|t_{i-1}), u_2^n(t_{i-1}), y_3^n(i-1)) \\ &\in \mathcal{A}_\epsilon^n(\hat{Y}_2, U_2, Y_3) \end{aligned}$$

Then the destination looks for a unique index \hat{z}_{i-1} which belongs to both the list code and the relevant bin, that is

$$\hat{z}_{i-1} \in B(t_i) \cap \mathcal{L}(y_3^n(i-1)) \quad (35)$$

Similar to the proof of Step 3 in Theorem 1, $\hat{z}_{i-1} = z_{i-1}$ with small enough probability of error, if n is sufficiently large and

$$R_2 > \hat{R}_2 - I(\hat{Y}_2; Y_3|U_2) \quad (36)$$

4. Then using both $\hat{y}_2^n(z_{i-1}|t_{i-1})$ and $y_3^n(i-1)$, the destination declares that \hat{w}_{i-1} was sent in block $i-1$ if there is a unique \hat{w}_{i-1} such that

$$(u_1^n(\hat{w}_{i-1}), y_3^n(i-1), \hat{y}_2^n(z_{i-1}|t_{i-1}), u_2^n(t_{i-1})) \in \mathcal{A}_\epsilon^n \quad (37)$$

This step can be done with small probability of error (i.e.

$\hat{w}_{i-1} = w_{i-1}$), if n is sufficiently large and

$$R < I(U_1; Y_3, \hat{Y}_2|U_2) \quad (38)$$

Now (33), (34) and (36) result in

$$\begin{aligned} I(U_2; Y_3) &\geq I(\hat{Y}_2; Y_2, S_2|U_2) - I(\hat{Y}_2; Y_3|U_2) \\ &= I(\hat{Y}_2; Y_2, S_2|U_2, Y_3) \end{aligned} \quad (39)$$

Thus we see from (38) and (39), that the rate in (8) s.t. (9) is achievable. \square

11 Appendix 3

11.1 Proof of Theorem 3

The proof is based on the evaluation of the achievable rate and its condition in (4) and (5) with an appropriate choice of input distribution. Note that for the CF scheme similar to the classical Gaussian RC [22], the best choice of the probability distribution is not known. So, we evaluate the related terms for a Gaussian distribution. If CF outperforms the DF scheme for Gaussian distribution, it is obvious that for the best choice of the probability distribution CF outperforms DF too. Subsequently, X_1 and X_2 are assumed to be zero-mean Gaussian with variances P_1 and P_2 and $E[X_1X_2] = 0$. Moreover, $E[S^2] = Q$, $E[X_2S] = 0$ (only the source is informed). Similar to Costa's initial DPC [4], the auxiliary random variable U_1 is defined as $U_1 = X_1 + \alpha S$, but similar to [9, 11, 13], arbitrary correlation is assumed between X_1 and S which is called generalised DPC to partially cancel the state. Hence, $E[X_1S] = \sigma_{1s} = \rho_{1s}\sqrt{P_1Q}$ where $-1 \leq \rho_{1s} \leq 1$. $\hat{Y}_2 = \beta Y_2 + \hat{Z}$ where β is a constant, and compression noise \hat{Z} is a zero-mean Gaussian random variable with variance \hat{N} , which is independent of S, X_1, X_2, Z_2, Z_3 . S is also independent of Z_2 and Z_3 . Evaluation of (4) with the specified parameters in the proof of Theorem 3 is as follows ($h(\cdot)$ denotes differential entropy)

$$\begin{aligned} R &< I(U_1; Y_3, \hat{Y}_2|X_2) - I(U_1; S) \\ &= I(U_1; Y_3|X_2) + I(U_1; \hat{Y}_2|X_2, Y_3) - I(U_1; S) \\ &= h(\hat{Y}_2|X_2, Y_3) - h(\hat{Y}_2|X_2, Y_3, U_1) + h(Y_3|X_2) \\ &\quad - h(Y_3|X_2, U_1) - I(U_1; S) \end{aligned} \quad (40)$$

Now, for the Gaussian distribution

$$h(\hat{Y}_2|X_2, Y_3) = \frac{1}{2} \log[(2\pi e)(E(\hat{Y}_2^2) - E(\hat{Y}_2 E(\hat{Y}_2|X_2, Y_3)))] \quad (41)$$

$$= \frac{1}{2} \log[(2\pi e)(E(\hat{Y}_2^2) - E(\hat{Y}_2(\gamma_1 X_2 + \gamma_2 Y_3)))] \quad (42)$$

$$= \frac{1}{2} \log[(2\pi e)((a^2\beta^2 P_1 + \beta^2 Q + 2a\beta^2 \sigma_{1s} + \beta^2 N_2 + \hat{N}) - \gamma_2 E(\hat{Y}_2 Y_3))] \quad (43)$$

where $E(\hat{Y}_2|X_2, Y_3)$ is the minimum mean-squared error estimator of \hat{Y}_2 given (X_2, Y_3) .

Equality in (42) follows from the fact that for the Gaussian distribution, linear estimator is optimal (i.e. $E(\hat{Y}_2|X_2, Y_3) = \gamma_1 X_2 + \gamma_2 Y_3$) [32]. Equality in (43) follows by substituting $E(\hat{Y}_2^2)$, $E(\hat{Y}_2 X_2) = 0$ and $E(\hat{Y}_2 Y_3) = \beta(aP_1 + (a+1)\sigma_{1s} + Q)$ into (42).

Since from the orthogonality principle [32], γ_1 and γ_2 can be found as

$$\begin{aligned} \gamma_1 &= -b\gamma_2 \\ \gamma_2 &= \frac{\beta(aP_1 + (a+1)\sigma_{1s} + Q)}{P_1 + Q + 2\sigma_{1s} + N_3} \end{aligned}$$

then, putting γ_1 and γ_2 into (43) yields (see equation at the bottom of the page)

Following the same procedure for evaluating other entropy functions, and defining $A \triangleq P_1 + Q + 2\sigma_{1s} + N_3$, $B \triangleq P_1 + \alpha^2 Q + 2\alpha\sigma_{1s}$, $C \triangleq P_1(1 - \rho_{1s}^2)$, $D \triangleq a^2 P_1 + Q + 2a\sigma_{1s}$, we can derive them as follows (see (44))

Substituting (44) into (40) yields (17). To obtain \hat{N} , we should compute the constraint in (5) with the above specified parameters similar to the above-mentioned procedure for computing the entropy functions. The derived entropy functions are

$$\begin{aligned} h(\hat{Y}_2|X_2, Y_2, Y_3) &= h(\hat{Y}_2|X_2, Y_2) = \frac{1}{2} \log[(2\pi e)\hat{N}] \\ I(X_2; Y_3) &= \frac{1}{2} \log \left[1 + \frac{b^2 P_2}{P_1 + Q + N_3 + 2\sigma_{1s}} \right] \end{aligned} \quad (45)$$

Putting (45) into (2), \hat{N} is derived as

$$\hat{N} = \frac{\beta^2}{b^2 P_2} [N_2 A + N_3 D + (a-1)^2 C Q]$$

which completes the proof of Theorem 3. □

$$h(\hat{Y}_2|X_2, Y_3) = \frac{1}{2} \log \left[(2\pi e) \left(\beta^2 N_2 + \hat{N} + \frac{\beta^2 [(a-1)^2 (Q P_1 - \sigma_{1s}^2) + N_3 (a^2 P_1 + Q + 2a\sigma_{1s})]}{P_1 + Q + 2\sigma_{1s} + N_3} \right) \right]$$

$$\begin{aligned} h(\hat{Y}_2|X_2, Y_3, U_1) &= \frac{1}{2} \log \left[(2\pi e) \left(\beta^2 N_2 + \hat{N} + \frac{\beta^2 (a\alpha - 1)^2 N_3 C Q}{(\alpha - 1)^2 C Q + B N_3} \right) \right] \\ h(\hat{Y}_2|X_2, Y_3) &= \frac{1}{2} \log \left[(2\pi e) \left(\beta^2 N_2 + \hat{N} + \frac{\beta^2 [(a-1)^2 (Q P_1 - \sigma_{1s}^2) + N_3 (a^2 P_1 + Q + 2a\sigma_{1s})]}{P_1 + Q + 2\sigma_{1s} + N_3} \right) \right] \\ h(Y_3|X_2, U_1) &= \frac{1}{2} \log \left[(2\pi e) \left((\alpha - 1)^2 \frac{Q C}{B} + N_3 \right) \right] \\ h(Y_3|X_2) &= \frac{1}{2} \log [(2\pi e) A] \\ I(U_1; S) &= \frac{1}{2} \log \left[\frac{B}{C} \right] \end{aligned} \quad (44)$$