

## Article

# Channel Coding and Source Coding With Increased Partial Side Information

Avihay Sadeh-Shirazi \*, Uria Basher and Haim Permuter \*

Department of Electrical and Computer Engineering at the Ben Gurion University of the Negev, Beer Sheva 84105, Israel; uria.basher@gmail.com

\* Correspondence: avihays@gmail.com (A.S.); haimp@bgu.ac.il (H.P.); Tel.: +972-8-6461-558 (H.P.)

Received: 30 June 2017; Accepted: 30 August 2017; Published: 2 September 2017

**Abstract:** Let  $(S_{1,i}, S_{2,i}) \sim \text{i.i.d } p(s_1, s_2)$ ,  $i = 1, 2, \dots$  be a memoryless, correlated partial side information sequence. In this work, we study channel coding and source coding problems where the partial side information  $(S_1, S_2)$  is available at the encoder and the decoder, respectively, and, additionally, either the encoder's or the decoder's side information is increased by a limited-rate description of the other's partial side information. We derive six special cases of channel coding and source coding problems and we characterize the capacity and the rate-distortion functions for the different cases. We present a duality between the channel capacity and the rate-distortion cases we study. In order to find numerical solutions for our channel capacity and rate-distortion problems, we use the Blahut-Arimoto algorithm and convex optimization tools. Finally, we provide several examples corresponding to the channel capacity and the rate-distortion cases we presented.

**Keywords:** Blahut–Arimoto algorithm; channel capacity; channel coding; duality; Gelfand–Pinsker channel coding; partial side information; rate-distortion; source coding; Wyner–Ziv source coding

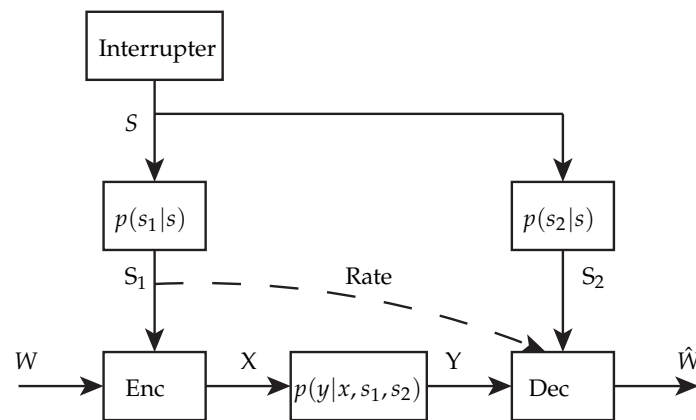
## 1. Introduction

In this paper, we investigate point-to-point channel models and rate-distortion problem models where both users have different and correlated partial side information and where, in addition, a rate-limited description of one of the user's side information is delivered to the other user. We then show the duality between the channel models and the rate-distortion models we investigate.

For the convenience of the reader, we refer to the state information as the side information, to the partial side information that is available to the encoder as the encoder's side information (ESI) and to the partial side information that is available to the decoder as the decoder's side information (DSI). We refer to the rate-limited description of the other user's side information as the increase in the side information. For example, if the decoder is informed with its DSI and, in addition, with a rate-limited description of the ESI, then we would say that the decoder is informed with increased DSI.

To make the motivation for this paper clear, let us look at the simple example depicted in Figure 1. In this setup, the communication between the Tx-Rx pair (the encoder-decoder) is interrupted by an undesired signal,  $S$ . The encoder and the decoder do not know  $S$  perfectly, but they each possess a version of  $S$ ; the encoder knows  $S_1$  (the ESI) and the decoder knows  $S_2$  (the DSI). For this example, let us assume that the source of the interruption is physically located in close proximity to the encoder (potentially, both signal sources are co-located). Thus, we assume that the encoder “knows more on  $S$ ” than the decoder; i.e.,  $H(S|S_1) < H(S|S_2)$ . We assume also that the transmitter can provide a rate-limited description of the ESI,  $S_1$ , to the decoder, thus increasing his DSI. In these circumstances, we pose the question; what is the capacity of the channel between the encoder and the decoder? This question is of practical importance. Knowing the channel capacity allows one to analyze its communication system better, answering questions such as “how close is the communication

system's performance to the capacity?" and "how important is the quality of the side information to the throughput?". Moreover, it allows one to design better practical codes, like polar codes and LDPC codes.



**Figure 1.** Increased partial side information example. The encoder wants to send a message to the decoder over an interrupted channel in the presence of side information. The encoder is provided with the ESI and the decoder is provided with increased DSI. i.e., the decoder is informed with a rate-limited description of the ESI, in addition to the DSI.

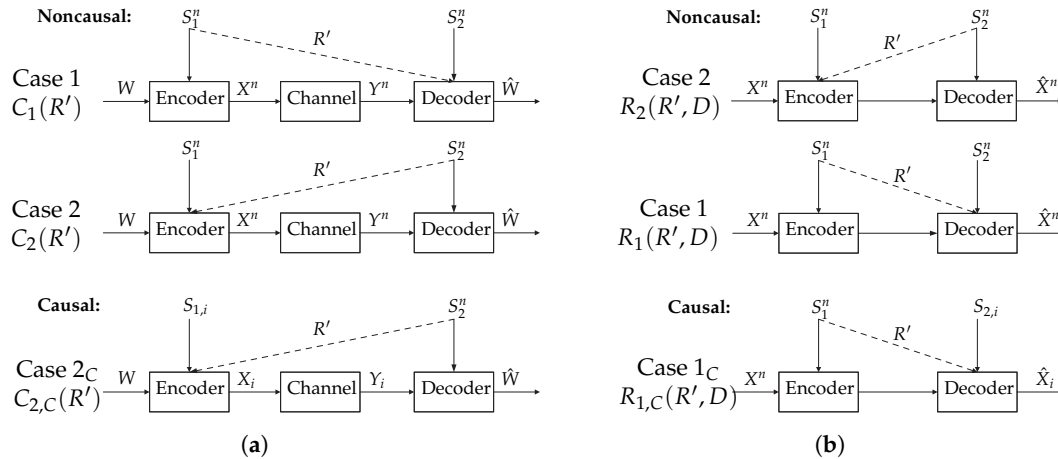
### 1.1. Channel Capacity in the Presence of State Information

The three problems of channel capacity in the presence of state information that we address in this paper are presented in Figure 2a. We make the assumption that the encoder is informed with partial state information, the ESI ( $S_1$ ), and the decoder is informed with different, but correlated, partial state information, which is the DSI ( $S_2$ ). The channel capacity problem cases are:

- Case 1: The decoder is provided with increased DSI; i.e., in addition to the DSI, the decoder is also informed with a rate-limited description of the ESI.
- Case 2: The encoder is informed with increased ESI.
- Case 2<sub>C</sub>: Similar to Case 2, with the exception that the ESI is known to the encoder in a causal manner. Notice that the rate-limited description of the DSI is still known to the encoder noncausally.

We will subsequently provide the capacity of Case 1 and Case 2<sub>C</sub> and characterize the lower and the upper bounds on Case 2, which differ only by a Markov relation. The results for the first case under discussion, Case 1, can be concluded from Steinberg's problem [1]. In [1], Steinberg introduced and solved the case in which the encoder is fully informed with the ESI and the decoder is informed with a rate-limited description of the ESI. Therefore, the innovation in Case 1 is that the decoder is also informed with the DSI. The solution for this problem can be derived by considering the DSI to be a part of the channel's output in Steinberg's solution. In the proof of the converse in his paper, Steinberg uses a new technique that involves using the Csiszár sum twice in order to get to a single-letter bound on the rate. We shall use this technique to present a duality in the converse of the Gelfand–Pinsker [2] and the Wyner–Ziv [3] problems, which, by themselves, constitute the basis for most of the results in this paper. In [3], Wyner and Ziv presented the rate-distortion function for data compression problems with side information at the decoder. We make use of their coding scheme in the achievability proof of the lower bound of Case 2 for describing the ESI with a limited rate at the decoder. In [2], Gelfand and Pinsker presented the capacity for a channel with noncausal channel state information (CSI) at the encoder. We use their coding scheme in the achievability proof of Case 1 and the lower bound of Case 2 for transmitting information over a channel where the ESI is the state information at the encoder. Therefore, we combine in our problems the Gelfand–Pinsker

and the Wyner-Ziv problems. Another related paper is [4], in which Shannon presented the capacity of a channel with causal CSI at the transmitter. We make use of Shannon's result in the achievability proof of Case 2<sub>C</sub> for communicating over a channel with causal ESI at the encoder. We also use Shannon's strategies [4], for developing an iterative algorithm to calculate the capacity of the cases we present in this paper.



**Figure 2.** Channel coding and source coding cases. **(a)** Channel coding with state information. Case 1: Rate-limited ESI at the decoder. Case 2: Rate-limited DSI at the encoder. Case 2<sub>C</sub>: Causal ESI and rate-limited DSI at the encoder; **(b)** Source coding with side information. Case 2: Rate-limited DSI at the encoder. Case 1: Rate-limited ESI at the decoder. Case 1<sub>C</sub>: Causal DSI and rate-limited ESI at the decoder. The cases are presented in this order to allow each source coding case to be parallel to the dual channel coding case.

Some related papers that can be found in the literature are mentioned herein. Heegard and El Gamal [5] presented a model of a state-dependent channel, where the transmitter is informed with the CSI at a rate limited to  $R_e$  and the receiver is informed with the CSI at a rate limited to  $R_d$ . This result relates to Case 1, Case 2 and Case 2<sub>C</sub> since we consider the rate-limited description of the ESI or the DSI as side information known at both the encoder and the decoder. Cover and Chiang [6] extended the Gelfand–Pinsker problem and the Wyner-Ziv problem to the case where both the encoder and the decoder are provided with different, but correlated, partial side information. They also showed a duality between the two cases, which is a topic that will be discussed later in this paper. Rozenzweig et al. [7] and Cemal and Steinberg [8] studied channels with partial state information at the transmitter. A detailed subject review on channel coding with state information was given by Keshet et al. in [9].

In addition to these three cases, we also present a more general case, where both the encoder and the decoder are informed with increased partial side information. i.e., the encoder and the decoder are each informed with partial side information, and, in addition, with a rate-limited description of the other's side information. We provide a lower bound on the capacity for this case; however, this bound does not necessarily coincide with the capacity and, therefore, this problem remains open.

### 1.2. Rate-Distortion with Side Information

In this paper, we address three problems of rate-distortion with side information, as presented in Figure 2b. In common with the channel capacity problems, we assume that the encoder is informed with the ESI ( $S_1$ ) and the decoder is informed with the DSI ( $S_2$ ), where the source,  $X$ , the ESI and the DSI are correlated. The rate-distortion problem cases we investigate in this paper are:

- Case 1: The decoder is provided with increased DSI.

- Case 1<sub>C</sub>: Similar to Case 1, with the exception that the ESI is known to the encoder in a causal manner. The rate-limited description of the ESI is still known to the decoder noncausally.
- Case 2: The encoder is informed with increased ESI.

Case 2 is a special case of Kaspi's [10] two-way source coding for  $K = 1$ . In [10], Kaspi introduced a model of multistage communication between two users, where each user may transmit up to  $K$  messages to the other user, dependent on the source and the previous received messages. For Case 2, we can consider sending the rate-limited description of the DSI as the first transmission and then, sending a function of the source, the ESI and the rate-limited description of the DSI as the second transmission. This fits into Kaspi's problem for  $K = 1$  and thus Kaspi's theorem also applies to Case 2. Kaspi's problem was later extended by Permuter et al. [11] to the case where a common rate-limited side information message is being conveyed to both users. Another strongly related paper is Wyner and Ziv's paper [3]. In the achievability of Case 1, we use the Wyner-Ziv coding scheme twice; once for describing the ESI at the decoder where the DSI is the side information and once for the main source and the ESI where the DSI is the side information. The rate-limited description of the ESI is the side information provided to both the encoder and the decoder. In [6] there is an extension to the Wyner-Ziv problem to the case where both the encoder and the decoder are provided with correlated partial side information. Weissman and El Gamal [12] and Weissman and Merhav [13] presented source coding with causal side information at the decoder, which relates to Case 1<sub>C</sub>. In addition, we present a generalized case of rate-distortion with two-sided increased partial side information. In this problem setup the encoder and the decoder are each informed with partial side information, and, in addition, with a rate-limited description of the other's side information. We present an upper bound on the optimal rate; however, this bound does not necessarily coincide with the optimal rate and, therefore, this problem remains open.

### 1.3. Duality

Within the scope of this work, we point out a duality relation between the channel capacity and the rate-distortion cases we discuss. The operational duality between channel coding and source coding was first mentioned by Shannon [14]. Pradhan et al. [15] and Pradhan and Ramchandran [16] studied the functional duality between some cases of channel coding and source coding, including the duality between the Gelfand–Pinsker problem and the Wyner-Ziv problem. This duality was also described by Cover and Chiang in [6], where they provided a transformation that makes duality between channel coding and source coding with two-sided state information apparent. Zamir et al. [17] and Su et al. [18] utilized the duality between channel coding and source coding with side information to develop coding schemes for the dual problems. Goldfeld, Permuter and Kramer [19] studied the duality between a two-encoder source coding with one-sided, rate-limited coordination and a semi-deterministic broadcast channel with one-sided decoder cooperation. More related works on the topic of duality can be found in the papers of Asnani et al. [20] and Gupta and Verdu [21].

In our paper, we show that the channel capacity cases and the rate-distortion cases we discuss are operational duals in a way that strongly relates to the Wyner-Ziv and Gelfand–Pinsker duality. We also provide a transformation scheme that shows this duality in a clear way. Moreover, we show a duality relation between Kaspi's problem and Steinberg's [1] problem by showing a duality relation between Case 2 source coding and Case 1 channel coding. Also, we show duality in the converse parts of the Gelfand–Pinsker and the Wyner-Ziv problems. We show that both converse parts can be proven in a perfectly dual way by using the Csiszár sum twice.

### 1.4. Computational Algorithms

Calculating channel capacity and rate-distortion problems, in general, and the Gelfand–Pinsker and the Wyner-Ziv problems, in particular, is not straightforward. Blahut [22] and Arimoto [23] suggested an iterative algorithm (to be referred to as the B-A algorithm) for numerically computing

the channel capacity and the rate-distortion problems. Willems [24] and Dupuis et al. [25] presented iterative algorithms based on the B-A algorithm for computing the Gelfand–Pinsker and the Wyner–Ziv functions. We use principles from Willems’ algorithms to develop an algorithm to numerically calculate the capacity for the cases we presented. More B-A based iterative algorithms for computing channel capacity and rate-distortion with side information can be found in [26,27]. A Blahut–Arimoto based algorithm for maximizing the directed-information can be found in [28].

### 1.5. Organization of the Paper and Main Contributions

To summarize, the main contributions of this paper are:

- We characterize the capacity and the rate-distortion functions of new channel and source coding problems with increased partial side information. We quantify the gain in the rate that can be achieved by having the parties involved share their partial side information with each other over a rate-limited secondary channel.
- We show a duality relationship between the channel capacity cases and the rate-distortion cases that we discuss.
- We provide a B-A based algorithm to solve the channel capacity problems we describe.
- We show a duality between the Gelfand–Pinsker capacity converse and the Wyner–Ziv rate-distortion converse.

The remainder of this paper is organized as follows. In Section 2 we introduce some notations for this paper and provide the settings of three channel coding and three source coding cases with increased partial side information. In Section 3 we present the main results for coding with increased partial side information; we provide the capacity and the rate-distortion for the cases we introduced in Section 2 and we point out the duality between the cases we examined. Section 4 contains illuminating examples for the cases discussed in the paper. In Section 5 we describe the B-A based algorithm we used in order to solve the capacity examples. We conclude the paper in Section 6 and we highlight two open problems; channel capacity and rate-distortion with two-sided rate-limited partial side information. Appendix A contains the duality derivation for the converse proofs of the Gelfand–Pinsker and the Wyner–Ziv problems and Appendices B, C, D and E contain the proofs for our theorems and lemmas.

## 2. Problem Setting and Definitions

In this section, we describe and formally define three cases of channel coding problems and three cases of source coding problems. All six cases are presented in Figure 2a,b.

**Notations.** We use subscripts and superscripts to denote vectors in the following ways:  $x^j = (x_1, \dots, x_j)$  and  $x_i^j = (x_i, \dots, x_j)$  for  $i \leq j$ . Moreover, we use the lower case  $x$  to denote sample value, the upper case  $X$  to denote a random variable, the calligraphic letter  $\mathcal{X}$  to denote the alphabet of  $X$ ,  $|\mathcal{X}|$  to denote the cardinality of the alphabet of  $X$  and  $p(x)$  to denote the probability  $\Pr\{X = x\}$ . We use the notation  $\mathcal{T}_\epsilon^{(n)}(X)$  to denote the strongly typical set of the random variable  $X$ , as defined in [29] (Chapter 11).

### 2.1. Definitions and Problem Formulation—Channel Coding with State Information

**Definition 1.** A discrete channel is defined by the set  $\{\mathcal{X}, \mathcal{S}_1, \mathcal{S}_2, p(s_1, s_2), p(y|x, s_1, s_2), \mathcal{Y}\}$ . The channel’s input sequence,  $\{X_i \in \mathcal{X}, i = 1, 2, \dots\}$ , the ESI sequence,  $\{S_{1,i} \in \mathcal{S}_1, i = 1, 2, \dots\}$ , the DSI sequence,  $\{S_{2,i} \in \mathcal{S}_2, i = 1, 2, \dots\}$ , and the channel’s output sequence,  $\{Y_i \in \mathcal{Y}, i = 1, 2, \dots\}$ , are discrete random variables drawn from the finite alphabets  $\mathcal{X}, \mathcal{S}_1, \mathcal{S}_2, \mathcal{Y}$ , respectively. Denote the message and the message space as  $W \in \{1, 2, \dots, 2^{nR}\}$  and let  $\hat{W}$  be the reconstruction of the message  $W$ . The random variables  $(S_{1,i}, S_{2,i})$  are i.i.d.  $\sim p(s_1, s_2)$  and the channel is memoryless, i.e., at time  $i$ , the output,  $Y_i$ , has a conditional distribution of

$$p(y_i|x^i, s_1^i, s_2^i, y^{i-1}) = p(y_i|x_i, s_{1,i}, s_{2,i}). \quad (1)$$

In the remainder of the paper, unless specifically mentioned otherwise, we refer to the ESI and the DSI as if they are known to the encoder and the decoder, respectively, in a noncausal manner. Also, as noted before, we use the term *increased* side information to indicate that the user's side information also includes a rate-limited description of the other user's partial side information. For example, when the decoder is informed with the DSI and with a rate-limited description of the ESI we would say that the decoder is informed with increased DSI.

**Problem Formulation.** For the channel  $p(y|x, s_1, s_2)$ , consider the following channel coding problem cases:

- Case 1: The encoder is informed with ESI and the decoder is informed with increased DSI.
- Case 2: The encoder is informed with increased ESI and the decoder is informed with DSI.
- Case  $2_C$ : The encoder is informed with increased causal ESI ( $S_1^i$  at time  $i$ ) and the decoder is informed with DSI. This case is the same as Case 2, except for the causal ESI.

All cases are presented in Figure 2a.

**Definition 2.** A  $(n, 2^{nR}, 2^{nR'})$  code,  $\{j \in 1, 2\}$ , for a channel with increased partial side information, as illustrated in Figure 2a, consists of two encoders and one decoder. The encoders are  $f$  and  $f_v$ , where  $f$  is the encoder for the channel's input and  $f_v$  is the encoder for the side information, and the decoder is  $g$ , as described for each case:

Case 1: Two encoders

$$\begin{aligned} f_v &: S_1^n \rightarrow \{1, 2, \dots, 2^{nR'_1}\}, \\ f &: \{1, 2, \dots, 2^{nR}\} \times S_1^n \times \{1, 2, \dots, 2^{nR'_1}\} \rightarrow \mathcal{X}^n, \end{aligned}$$

and a decoder

$$g: \mathcal{Y}^n \times S_2^n \times \{1, 2, \dots, 2^{nR'_1}\} \rightarrow \{1, 2, \dots, 2^{nR}\}. \quad (2)$$

Case 2: Two encoders

$$\begin{aligned} f_v &: S_2^n \rightarrow \{1, 2, \dots, 2^{nR'_2}\}, \\ f &: \{1, 2, \dots, 2^{nR}\} \times S_1^n \times \{1, 2, \dots, 2^{nR'_2}\} \rightarrow \mathcal{X}^n, \end{aligned}$$

and a decoder

$$g: \mathcal{Y}^n \times S_2^n \times \{1, 2, \dots, 2^{nR'_2}\} \rightarrow \{1, 2, \dots, 2^{nR}\}. \quad (3)$$

Case  $2_C$ : Two encoders

$$\begin{aligned} f_v &: S_2^n \rightarrow \{1, 2, \dots, 2^{nR'_2}\}, \\ f_i &: \{1, 2, \dots, 2^{nR}\} \times S_1^i \times \{1, 2, \dots, 2^{nR'_2}\} \rightarrow \mathcal{X}_i, \end{aligned}$$

and a decoder

$$g: \mathcal{Y}^n \times S_2^n \times \{1, 2, \dots, 2^{nR'_2}\} \rightarrow \{1, 2, \dots, 2^{nR}\}. \quad (4)$$

The average probability of error,  $P_e^{(n)}$ , for a  $(n, 2^{nR}, 2^{nR'})$  code is defined as

$$P_e^{(n)} = \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \Pr \{ \hat{W} \neq W \mid W = w \}, \quad (5)$$



where the index  $W$  is chosen according to a uniform distribution over the set  $\{1, 2, \dots, 2^{nR}\}$ . A rate pair  $(R, R')$  is said to be achievable if there exists a sequence of  $(n, 2^{nR}, 2^{nR'})$  codes such that the average probability of error  $P_e^{(n)} \rightarrow 0$  as  $n \rightarrow \infty$ .

**Definition 3.** The capacity of the channel,  $C(R')$ , is the supremum of all  $R$  such that the rate pair  $(R, R')$  is achievable.

## 2.2. Definitions and Problem Formulation—Source Coding with Side Information

Throughout this article we use the common definitions of rate-distortion as presented in [29].

**Definition 4.** The source sequence  $\{X_i \in \mathcal{X}, i = 1, 2, \dots\}$ , the ESI sequence  $\{S_{1,i} \in \mathcal{S}_1, i = 1, 2, \dots\}$  and the DSI sequence  $\{S_{2,i} \in \mathcal{S}_2, i = 1, 2, \dots\}$  are discrete random variables drawn from the finite alphabets  $\mathcal{X}, \mathcal{S}_1$  and  $\mathcal{S}_2$  respectively. The random variables  $(X_i, S_{1,i}, S_{2,i})$  are i.i.d  $\sim p(x, s_1, s_2)$ . Let  $\hat{\mathcal{X}}$  be the reconstruction alphabet and  $d_x : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow [0, \infty)$  be the distortion measure. The distortion between sequences is defined in the usual way:

$$d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^n d(x_i, \hat{x}_i). \quad (6)$$

**Problem Formulation.** For the source,  $X$ , the ESI,  $S_1$ , and the DSI,  $S_2$ , consider the following source coding problem cases:

- Case 1: The encoder is informed with ESI and the decoder is informed with increased DSI.
- Case 2: The encoder is informed with increased ESI and the decoder is informed with DSI.
- Case 1<sub>C</sub>: The encoder is informed with ESI and the decoder is informed with increased causal DSI ( $S_2^i$  at time  $i$ ). This case is the same as Case 1, except for the causal DSI.

All cases are presented in Figure 2b.

**Definition 5.** A  $(n, 2^{nR}, 2^{nR'}, D)$  code,  $\{j \in 1, 2\}$ , for the source  $X$  with increased partial side information, as illustrated in Figure 2b, consists of two encoders, one decoder and a distortion constraint. The encoders are  $f$  and  $f_v$ , where  $f$  is the encoder for the source and  $f_v$  is the encoder for the side information, and the decoder is  $g$ , as described for each case:

Case 1: Two encoders

$$\begin{aligned} f_v : \mathcal{S}_1^n &\rightarrow \{1, 2, \dots, 2^{nR'}\}, \\ f : \mathcal{X}^n \times \mathcal{S}_1^n &\times \{1, 2, \dots, 2^{nR'}\} \rightarrow \{1, 2, \dots, 2^{nR}\}, \end{aligned}$$

and a decoder

$$g : \{1, 2, \dots, 2^{nR}\} \times \mathcal{S}_2^n \times \{1, 2, \dots, 2^{nR'}\} \rightarrow \hat{\mathcal{X}}^n. \quad (7)$$

Case 2: Two encoders

$$\begin{aligned} f_v : \mathcal{S}_2^n &\rightarrow \{1, 2, \dots, 2^{nR'}\}, \\ f : \mathcal{X}^n \times \mathcal{S}_1^n &\times \{1, 2, \dots, 2^{nR'}\} \rightarrow \{1, 2, \dots, 2^{nR}\}, \end{aligned}$$

and a decoder

$$g : \{1, 2, \dots, 2^{nR}\} \times \mathcal{S}_2^n \times \{1, 2, \dots, 2^{nR'}\} \rightarrow \hat{\mathcal{X}}^n. \quad (8)$$

Case 1<sub>C</sub>: Two encoders

$$\begin{aligned} f_v &: \mathcal{S}_1^n \rightarrow \{1, 2, \dots, 2^{nR'_1}\}, \\ f &: \mathcal{X}^n \times \mathcal{S}_1^n \times \{1, 2, \dots, 2^{nR'_1}\} \rightarrow \{1, 2, \dots, 2^{nR}\}, \end{aligned}$$

and a decoder

$$g_i: \{1, 2, \dots, 2^{nR}\} \times \mathcal{S}_2^i \times \{1, 2, \dots, 2^{nR'_1}\} \rightarrow \hat{\mathcal{X}}_i. \quad (9)$$

The distortion constraint for all three cases is:

$$\mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n d(X_i, \hat{X}_i) \right] \leq D. \quad (10)$$

For a given distortion,  $D$ , and for any  $\epsilon > 0$ , the rate pair  $(R, R')$  is said to be achievable if there exists a  $(n, 2^{nR}, 2^{nR'}, D + \epsilon)$  code for the rate-distortion problem.

**Definition 6.** For a given  $R'$  and distortion  $D$ , the operational rate  $R^*(R', D)$  is the infimum of all  $R$ , such that the rate pair  $(R, R')$  is achievable.

### 3. Results

In this section, we present the main results of this paper. We will first present the results for the channel coding cases, then the main results for the source coding cases and, finally, we will present the duality between them.

#### 3.1. Channel Coding with Side Information

For a channel with two-sided state information as presented in Figure 2a, where  $(S_{1,i}, S_{2,i}) \sim p(s_1, s_2)$ , the capacity is as follows

**Theorem 1** (The capacity for the cases in Figure 2a). For the memoryless channel  $p(y|x, s_1, s_2)$ , where  $S_1$  is the ESI and  $S_2$  is the DSI and the side information  $(S_{1,i}, S_{2,i}) \sim p(s_1, s_2)$ , the channel capacity is

Case 1: The encoder is informed with ESI and the decoder is informed with increased DSI,

$$C_1^* = \max_{\substack{p(v_1|s_1)p(u|s_1,v_1)p(x|u,s_1,v_1) \\ \text{s.t. } R' \geq I(V_1;S_1) - I(V_1;Y,S_2)}} I(U;Y,S_2|V_1) - I(U;S_1|V_1). \quad (11)$$

Case 2: The encoder is informed with increased ESI and the decoder is informed with DSI;

Lower bounded by

$$C_2^{lb*} = \max_{\substack{p(v_2|s_2)p(u|s_1,v_2)p(x|u,s_1,v_2) \\ \text{s.t. } R' \geq I(V_2;S_2|S_1)}} I(U;Y,S_2|V_2) - I(U;S_1|V_2). \quad (12)$$

Upper bounded by

$$C_2^{ub1*} = \max_{\substack{p(v_2|s_1,s_2)p(u|s_1,v_2)p(x|u,s_1,v_2) \\ \text{s.t. } R' \geq I(V_2;S_2) - I(V_2;S_1)}} I(U;Y,S_2|V_2) - I(U;S_1|V_2) \quad (13)$$

and by

$$C_2^{ub2*} = \max_{\substack{p(v_2|s_2)p(u|s_1,s_2,v_2)p(x|u,s_1,v_2) \\ \text{s.t. } R' \geq I(V_2;S_2|S_1)}} I(U;Y,S_2|V_2) - I(U;S_1|V_2). \quad (14)$$



Case 2<sub>C</sub>: The encoder is informed with increased causal ESI ( $S_1^i$  at time  $i$ ) and the decoder is informed with DSI,

$$C_{2C}^* = \max_{\substack{p(v_2|s_2)p(u|v_2)p(x|u,s_1,v_2) \\ R' \geq I(V_2;S_2)}} I(U;Y,S_2|V_2). \quad (15)$$

For case  $j$ ,  $j \in \{1, 2\}$ , some joint distribution,  $p(s_1, s_2, v_j, u, x, y)$ , and  $(U, V_j)$  being some auxiliary random variables with bounded cardinality.

Appendix B contains the proof.

**Lemma 1.** For all three channel coding cases described in this section and for  $j \in \{1, 2\}$ , the following statements hold,

1. The function  $C_j(R')$  is a concave function of  $R'$ .
2. It is enough to take  $X$  to be a deterministic function of  $(U, S_1, V_j)$  to evaluate  $C_j$ .
3. The auxiliary alphabets  $\mathcal{U}$  and  $\mathcal{V}_j$  satisfy

for Case 1:	$ \mathcal{V}_1  \leq  \mathcal{X}  \mathcal{S}_1  \mathcal{S}_2  + 1$ and $ \mathcal{U}  \leq  \mathcal{X}  \mathcal{S}_1  \mathcal{S}_2 ( \mathcal{X}  \mathcal{S}_1  \mathcal{S}_2  + 1),$
for Case 2:	$ \mathcal{V}_2  \leq  \mathcal{S}_1  \mathcal{S}_2  + 1$ and $ \mathcal{U}  \leq  \mathcal{X}  \mathcal{S}_1  \mathcal{S}_2 ( \mathcal{S}_1  \mathcal{S}_2  + 1),$
for Case 2 <sub>C</sub> :	$ \mathcal{V}_2  \leq  \mathcal{S}_2  + 1$ and $ \mathcal{U}  \leq  \mathcal{X}  \mathcal{S}_2 ( \mathcal{S}_2  + 1).$

Appendix D contains the proof for the above lemma.

**Remark 1.** Please notice that in Equation (14), the rate of the side information,  $I(V_2; S_2|S_1)$ , can be written as  $I(V_2; S_2) - I(V_2|S_1)$ . This is true since the Markov relation  $V_2 - S_2 - S_1$  holds. Therefore, the only difference between the two upper bounds of Case 2,  $C_2^{ub1*}$  (13) and  $C_2^{ub2*}$  (14), is in the distribution over which we maximize. While for  $C_2^{ub1*}$  we restrict the maximization to distributions which maintain the Markov chain  $U - (S_1, V_2) - S_2$ , for the second upper bound,  $C_2^{ub2*}$ , we restrict the maximization to distributions which maintain  $V_2 - S_2 - S_1$ . We should note that we cannot state with certainty that one of the bounds is tighter than the other for all distributions  $p(y, x, s_1, s_2)$  and for all values of  $R'$ . Notwithstanding, one bound may be tighter than the other for all distributions.

### 3.2. Source Coding with Side Information

For the problem of source coding with side information as presented in Figure 2b, the rate-distortion function is as follows:

**Theorem 2** (The rate-distortion function for the cases in Figure 2b). For a bounded distortion measure  $d(x, \hat{x})$ , a source,  $X$ , and side information,  $S_1, S_2$ , where  $(X_i, S_{1,i}, S_{2,i}) \sim p(x, s_1, s_2)$ , the rate-distortion function is

Case 1: The encoder is informed with ESI and the decoder is informed with increased DSI,

$$R_1^*(D) = \min_{\substack{p(v_1|s_1)p(u|x,s_1,v_1)p(\hat{x}|\hat{u},s_2,v_1) \\ \text{s.t. } R' \geq I(V_1;S_1|S_2)}} I(U;X,S_1|V_1) - I(U;S_2|V_1). \quad (16)$$

Case 1<sub>C</sub>: The encoder is informed with ESI and the decoder is informed with increased causal DSI ( $S_2^i$  at time  $i$ ),

$$R_{1C}^*(D) = \min_{\substack{p(v_1|s_1)p(u|x,s_1,v_1)p(\hat{x}|\hat{u},s_2,v_1) \\ \text{s.t. } R' \geq I(V_1;S_1)}} I(U;X,S_1|V_1). \quad (17)$$

Case 2: The encoder is informed with increased ESI and the decoder is informed with DSI,

$$R_2^*(D) = \min_{\substack{p(v_2|s_2)p(u|x,s_1,v_2)p(\hat{x}|u,s_2,v_2) \\ \text{s.t. } R' \geq I(V_2;S_2) - I(V_2;X,S_1)}} I(U;X,S_1|V_2) - I(U;S_2|V_2). \quad (18)$$

For case  $j$ ,  $j \in \{1, 2\}$ , some joint distribution,  $p(x, s_1, s_2, v_j, u, \hat{x})$ , where  $\mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n d(X_i, \hat{X}_i) \right] \leq D$  and  $(U, V_j)$  being some auxiliary random variables with bounded cardinality.

Appendix C contains the proof.

**Lemma 2.** For all cases of rate-distortion problems in this section and for  $j \in \{1, 2\}$ , the following statements hold.

1. The function  $R_j(R', D)$  is a convex function of  $R'$  and  $D$ .
2. It is enough to take  $\hat{X}$  to be a deterministic function of  $(U, S_2, V_j)$  to evaluate  $R_j$ .
3. The auxiliary alphabets  $\mathcal{U}$  and  $\mathcal{V}_j$  satisfy

for Case 1:	$ \mathcal{V}_1  \leq  \mathcal{S}_1  \mathcal{S}_2  + 1$ and
	$ \mathcal{U}  \leq  \mathcal{X}  \mathcal{S}_1  \mathcal{S}_2 ( \mathcal{S}_1  \mathcal{S}_2  + 1),$
for Case 1 <sub>C</sub> :	$ \mathcal{V}_1  \leq  \mathcal{S}_1  + 1$ and
	$ \mathcal{U}  \leq  \mathcal{X}  \mathcal{S}_1 ( \mathcal{S}_1  + 1),$
for Case 2:	$ \mathcal{V}_2  \leq  \mathcal{X}  \mathcal{S}_1  \mathcal{S}_2  + 1$ and
	$ \mathcal{U}  \leq  \mathcal{X}  \mathcal{S}_1  \mathcal{S}_2 ( \mathcal{X}  \mathcal{S}_1  \mathcal{S}_2  + 1).$

Appendix D contains the proof for the above lemma.

### 3.3. Duality

We now investigate the duality between the channel coding and the source coding for the cases in Figure 2a,b. The following transformation makes the duality between the channel coding cases 1, 2, 2<sub>C</sub> and the source coding cases 2, 1, 1<sub>C</sub>, respectively, evident. The left column corresponds to channel coding and the right column to source coding. For cases  $j$  and  $\bar{j}$ , where  $j, \bar{j} \in \{1, 2\}$  and  $\bar{j} \neq j$ , consider the transformation:

$$\begin{aligned}
 &\text{channel coding} \longleftrightarrow \text{source coding} \\
 &C \longleftrightarrow R(D) \\
 &\text{maximization} \longleftrightarrow \text{minimization} \\
 &C_j \longleftrightarrow R_{\bar{j}}(D) \\
 &X \longleftrightarrow \hat{X} \\
 &Y \longleftrightarrow X \\
 &S_j \longleftrightarrow S_{\bar{j}} \\
 &V_j \longleftrightarrow V_{\bar{j}} \\
 &U \longleftrightarrow U \\
 &R' \longleftrightarrow R'.
 \end{aligned} \quad (19)$$

This transformation is an extension of the transformation provided in [6,15]. Note that while the channel capacity formula in Case  $j$  and the rate-distortion function in Case  $\bar{j}$  are dual to one another in the sense of maximization-minimization, the corresponding rates  $R'$  are not dual to each other in this sense; i.e., one would expect to see an opposite inequality ( $\geq \leftrightarrow \leq$ ) for dual cases, where we

have an inequality that is in the same direction ( $\leq \leftrightarrow \leq$ ) in the  $R'$  formulas. The duality in the side information rates,  $R'$ , is then in the sense that the arguments in the formulas for the dual  $R'$  are dual. This exception is due to the fact that while the Gelfand–Pinsker and the Wyner–Ziv problems for the main channel or the main rate-distortion problems are dual, the Wyner–Ziv problem for the side information stays the same; the only difference is the input and the output.

#### 4. Examples

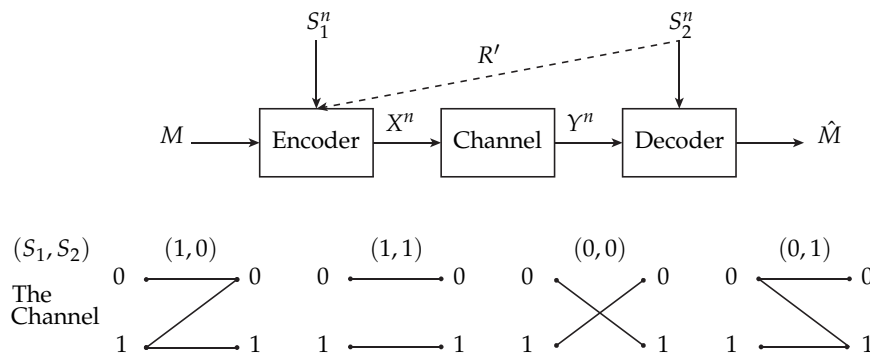
In this section, we provide examples for Case 2 of the channel coding theorem and for Case 1 of the source coding theorem. The numerical iterative algorithm, which we used to numerically calculate the lower bound,  $C_2^{lb}$ , is provided in the next section.

**Example 1** (Case 2 channel coding for a binary channel).

Consider the binary channel illustrated in Figure 3. The alphabet of the input, the output and the two states is binary  $\mathcal{X} = \mathcal{Y} = \mathcal{S}_1 = \mathcal{S}_2 = \{0, 1\}$  with  $(S_1, S_2) \sim \mathbf{P}_{S_1 S_2}$  being a joint probability mass function (PMF) matrix. The channel is dependent on the states  $S_1$  and  $S_2$ , where the encoder is fully informed with  $S_1$  and with  $S_2$  with a rate limited to  $R'$  and the decoder is fully informed with  $S_2$ . The dependence of the channel on the states is illustrated in Figure 3. If  $(S_1 = 1, S_2 = 0)$  then the channel is the *Z channel* with transition probability  $\epsilon$ , if  $(S_1 = 1, S_2 = 1)$  then the channel has no error, if  $(S_1 = 0, S_2 = 0)$  then the channel is the *X-channel* and if  $(S_1 = 0, S_2 = 1)$  then the channel is the *S-channel* with transition probability of  $\epsilon$ . The side information's joint PMF is

$$\mathbf{P}_{S_1 S_2} = \begin{pmatrix} 0.1 & 0.4 \\ 0.4 & 0.1 \end{pmatrix}.$$

The expressions for the lower bound on the capacity  $C_2^{lb}(R')$  and for  $R'$  are brought in Case 2 of Theorem 1.



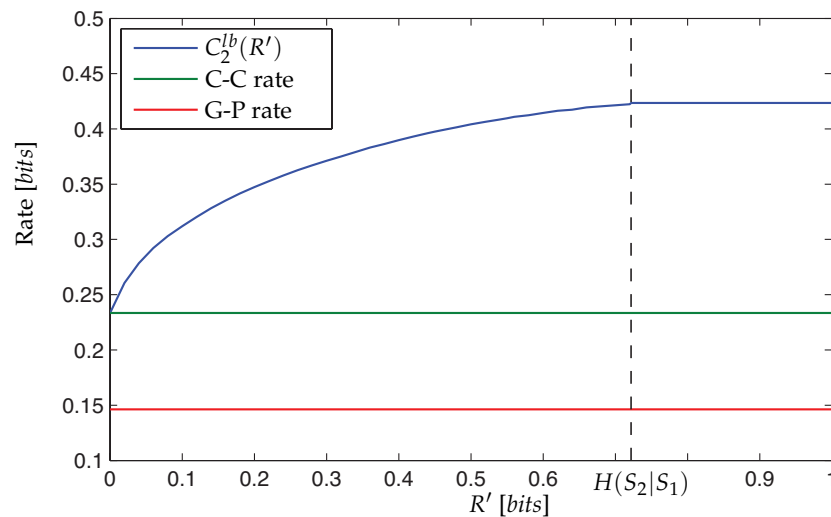
**Figure 3.** Example 1 Channel coding Case 2—channel topology.

In Figure 4, we provide the graph from of the computation of the lower bound on the capacity for the binary channel we are testing. In the graph, we present the lower bound,  $C_2^{lb}(R')$ , as a function of  $R'$ . We also provide the Cover & Chiang [6] capacity (where  $R' = 0$ ) and the Gelfand and Pinsker [2] capacity (where  $R' = 0$  and the decoder is not informed with  $S_2$ ).

Discussion:

1. The algorithm that we used to calculate  $C_2^{lb}(R')$  and  $R'$  combines a grid-search and a Blahut–Arimoto-like algorithms. We first construct a grid of probabilities of the random variable  $V_2$  given  $S_2$ , namely,  $w(v_2|s_2)$ . Then, for every probability  $w(v_2|s_2)$  such that  $I(V_2; S_2|S_1)$  is close enough to  $R'$  we calculate the maximum of  $I(U; Y, S_2|V_2) - I(U; S_1|V_2)$  using the iterative

- algorithm described in the next section. We then choose the maximum over those maximums and declare it to be  $C_2^{lb}$ . By taking a fine grid of the probabilities  $w(v_2|s_2)$  the operation's result can be arbitrarily close to  $C_2^{lb}$ .
2. For a given joint PMF matrix  $\mathbf{P}_{S_1 S_2}$ , we can see that  $C_2^{lb}(R')$  is non-decreasing in  $R'$ . Furthermore, since the expression  $I(V_2; S_2|S_1)$  is bounded by  $R_{\max} = \max_{p(v_2|s_2)} I(V_2; S_2|S_1) = H(S_2|S_1)$ , allowing  $R'$  to be greater than  $R_{\max}$  cannot improve  $C_2^{lb}$  any more. i.e.,  $C_2^{lb}(R' = R_{\max}) = C_2^{lb}(R' > R_{\max})$ . Therefore, it is enough to allow  $R' = R_{\max}$  to achieve  $C_2^{lb}$ , as if the encoder is fully informed with  $S_2$ .
  3. Although  $C_2^{lb}$  is a lower bound on the capacity, it can be significantly greater than the Cover-Chiang and the Gelfand-Pinsker rates for some channel models, as can be seen in this example. Moreover, we can actually state that  $C_2^{lb}$  is always greater than or equal to the Gelfand-Pinsker and the Cover-Chiang rates. This is due to the fact that when  $R' = 0$ ,  $C_2^{lb}$  coincides with the Cover-Chiang rate, which, in its turn, is always greater than or equal to the Gelfand-Pinsker rate; since  $C_2^{lb}$  is also non-decreasing in  $R'$ , it is obvious that our assertion holds.



**Figure 4.** Example 1. Channel coding Case 2 for the channel depicted in Figure 3, where the side information is distributed  $S_1 \sim \text{Bernoulli}(0.5)$ , and  $\Pr\{S_2 \neq S_1\} = 0.8$ .  $C_2^{lb}(R')$  is the lower bound on the capacity of this channel, C-C rate is the Cover-Chiang rate ( $R' = 0$ ) and G-P rate is the Gelfand-Pinsker rate ( $R' = 0$  and the decoder has no side information available at all). Notice that at the encoder the maximal uncertainty about  $S_2$  is  $H(S_2|S_1) = 0.7219$  bit. Therefore, for any  $R' \geq 0.7219$   $C_2^{lb}$  reaches its maximal value.

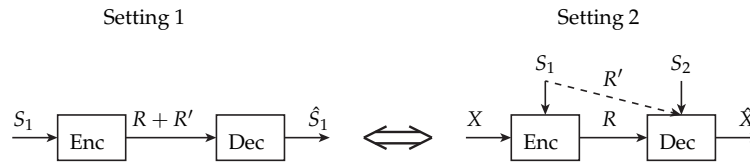
**Example 2** (Source coding Case 1 for a binary-symmetric source and Hamming distortion).

Consider the source  $X = S_1 \oplus S_2$ , where  $S_1, S_2 \sim \text{i.i.d. Bernoulli}(0.5)$ , and consider the problem setting depicted in Case 1 of the source coding problems. It is sufficient for the decoder to reconstruct  $S_1$  with distortion  $\mathbb{E}[d(S_1, \hat{S}_1)] \leq D$  in order to reconstruct  $X$  with the same distortion. Furthermore, the two rate-distortion problem settings illustrated in Figure 5 are equivalent.

For every achievable rate in Setting 1,  $\mathbb{E}[d(S_1, \hat{S}_1)] \leq D$ . Denote  $\hat{X} \triangleq \hat{S}_1 \oplus S_2$ , then,  $d(S_1, \hat{S}_1) = S_1 \oplus \hat{S}_1 = (S_1 \oplus S_2) \oplus (\hat{S}_1 \oplus S_2) = X \oplus \hat{X} = d(X, \hat{X})$  and, therefore,  $\mathbb{E}[d(S_1, \hat{S}_1)] \leq D$  in Setting 1  $\Rightarrow \mathbb{E}[d(X, \hat{X})] \leq D$  in Setting 2. In the same way, for Setting 2, denote  $\hat{S}_1 \triangleq \hat{X} \oplus S_2$ . Then,  $d(X, \hat{X}) = X \oplus \hat{X} = S_1 \oplus \hat{S}_1$  and, therefore,  $\mathbb{E}[d(X, \hat{X})] \leq D$  in Setting 2  $\Rightarrow \mathbb{E}[d(S_1, \hat{S}_1)] \leq D$  in

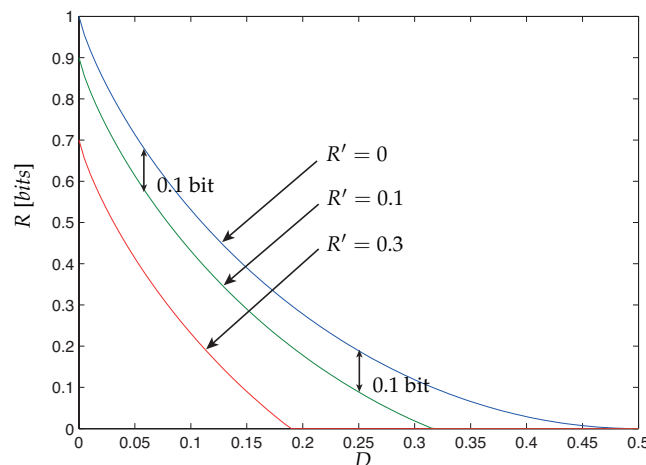
Setting 1. Hence, we can conclude that the two settings are equivalent and, for any given  $0 \leq D$  and  $0 \leq R'$ , the rate-distortion function is

$$R(D) = \begin{cases} 1 - H(D) - R' & 1 - H(D) - R' \geq 0 \\ 0 & 1 - H(D) - R' < 0 \end{cases} \quad (20)$$



**Figure 5.** The equivalent rate-distortion problem for Case 1 for the source  $X = S_1 \oplus S_2$  where  $S_1, S_2 \sim \text{i.i.d. Bernoulli}(0.5)$ .

In Figure 6 we present the plot resulting for this example. It is easy to verify that the Wyner and Ziv rate and the Cover and Chiang rate for this setting are  $R_{WZ}(D) = R_{CC}(D) = \max\{1 - H(D), 0\}$ .



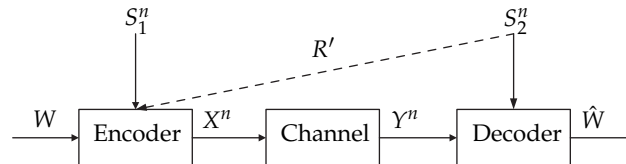
**Figure 6.** Example 2. Source coding Case 1 for binary-symmetric source and Hamming distortion. The source is given by  $X = S_1 \oplus S_2$ , where  $S_1, S_2 \sim \text{Bernoulli}(0.5)$ . The graph shows the rate-distortion function for different values of  $R'$ .

## 5. Semi-Iterative Algorithm

In this section, we provide algorithms that numerically calculate the lower bound on the capacity of Case 2 of the channel coding problems. The calculation of the Gelfand–Pinsker and the Wyner–Ziv problems has been addressed in many papers in the past, including [5,24–26]. All these algorithms are based on Arimoto’s [23] and Blahut’s [22] algorithms and on the fact that the Wyner–Ziv and the Gelfand–Pinsker problems can be presented as convex optimization problems. On the contrary, our problems are not convex in all of their optimization variables and, therefore, cannot be presented as convex optimization problems. In order to solve our problems we devised a different approach which combines a grid-search and a Blahut–Arimoto-like algorithm. In this section, we provide the mathematical justification for those two algorithms. Other algorithms to numerically compute the channel capacity or the rate-distortion of the rest of the cases presented in this paper can be derived using the principles that we describe in this section.

### 5.1. An Algorithm for Computing the Lower Bound on the Capacity of Case 2

Consider the channel in Figure 7 described by  $p(y|x, s_1, s_2)$  and consider the joint PMF  $p(s_1, s_2)$ . The capacity of this channel is lower bounded by  $\max I(U; Y, S_2|V_2) - I(U; S_1|V_2)$ , where the maximization is over all PMFs  $p(s_1, s_2)w(v_2|s_2)p(u|s_1, v_2)p(x|s_1, v_2, u)p(y|x, s_1, s_2)$  such that  $R' \geq I(V_2; S_2|S_1)$ . Notice that the lower bound expression is not concave in  $w(v_2|s_2)$ , which is the main difficulty with the computation of it. We first present an outline of the semi-iterative algorithm we developed, then we present the mathematical background and justification for the algorithm and, finally, we present the detailed algorithm.



**Figure 7.** Channel coding: Case 2.  $C_2^{lb} = \max I(U; Y, S_2|V_2) - I(U; S_1|V_2)$ , where the maximization is over all PMFs  $w(v_2|s_2)p(u|s_1, v_2)p(x|s_1, v_2, u)$  such that  $R' \geq I(V_2; S_2|S_1)$ .

For any fixed PMF  $w(v_2|s_2)$  denote

$$R_w \triangleq I(V_2; S_2|S_1), \quad (21)$$

$$C_{2,w}^{lb} \triangleq \max_{p(u|s_1, v_2)p(x|u, s_1, v_2)} I(U; Y, S_2|V_2) - I(U; S_1|V_2). \quad (22)$$

Then, the lower bound on the capacity,  $C_2^{lb}(R')$ , can be expressed as

$$C_2^{lb}(R') = \max_{\substack{w(v_2|s_2) \\ \text{s.t. } R' \geq R_w}} \max_{p(u|s_1, v_2)p(x|u, s_1, v_2)} [I(U; Y, S_2|V_2) - I(U; S_1|V_2)] \triangleq \max_{\substack{w(v_2|s_2) \\ \text{s.t. } R' \geq R_w}} C_{2,w}^{lb}. \quad (23)$$

The outline of the algorithm is as follows: for any given rate  $R' \leq H(S_2|S_1)$ ,  $\epsilon > 0$  and  $\delta > 0$ ,

1. Establish a fine and uniformly spaced grid of legal PMFs,  $w(v_2|s_2)$ , and denote the set of all of those PMFs as  $\mathcal{W}$ .
2. Establish the set  $\mathcal{W}^* := \{w(v_2|s_2) \mid w(v_2|s_2) \in \mathcal{W} \text{ and } R' - \epsilon \leq R_w \leq R'\}$ . This set is the set of all PMFs  $w(v_2|s_2)$  such that  $R_w$  is  $\epsilon$ -close to  $R'$  from below. If  $\mathcal{W}^*$  is empty, go back to step 1 and make the grid finer. Otherwise, continue.
3. For every  $w(v_2|s_2) \in \mathcal{W}^*$ , perform a Blahut-Arimoto-like optimization to find  $C_{2,w}^{lb}$  with accuracy of  $\delta$ .
4. Declare  $C_2^{lb}(R') = \max_{w(v_2|s_2) \in \mathcal{W}^*} C_2^{lb(\epsilon, \delta, \mathcal{W})}(R')$ .

**Remark 2. 1.** We considered only those  $R'$ 's such that  $R' \leq H(S_2|S_1)$  since  $H(S_2|S_1)$  is the maximal value that  $I(V_2; S_2|S_1)$  takes. The interpretation of this is that if the encoder is informed with  $S_1$ , we cannot increase its side information about  $S_2$  in more than  $H(S_2|S_1)$ . Therefore, for any  $H(S_2|S_1) \leq R'$ , we can limit  $R'$  to be equal to  $H(S_2|S_1)$  in order to compute the capacity;

2. Since  $C_{2,w}^{lb}(R')$  is continuous in  $w(v_2|s_2)$  and bounded (for example, by  $I(X; Y|S_1, S_2)$  from above and by  $I(X; Y)$  from below),  $C_2^{(\epsilon, \delta, \mathcal{W})}(R')$  can be arbitrarily close to  $C_2^{lb}(R')$  for  $\epsilon \rightarrow 0$ ,  $\delta \rightarrow 0$  and  $|\mathcal{W}| \rightarrow \infty$ .

#### 5.1.1. Mathematical background and justification

Here we focus on finding the lower bound on the capacity of the channel for a fixed distribution  $w(v_2|s_2)$ , i.e., finding  $C_{2,w}^{lb}$ . Note that the mutual information expression  $I(U; Y, S_2|V_2) - I(U; S_1|V_2)$  is concave in  $p(u|s_1, v_2)$  and convex in  $p(x|u, s_1, v_2)$ . Therefore, a standard convex maximization technique is not applicable for this problem. However, according to Dupuis, Yu and Willems [25],

we can write the expression for the lower bound as  $C_{2,w}^{lb} = \max_{q(t|s_1,v_2)} I(T; Y, S_2 | V_2) - I(T; S_1 | V_2)$ , where  $q(t|s_1, v_2)$  is a probability distribution over the set of all possible strategies  $t : \mathcal{S}_1 \times \mathcal{V}_2 \rightarrow \mathcal{X}$ , the input symbol  $X$  is selected using  $x = t(s_1, v_2)$  and  $p(y|x, s_1, s_2) = p(y|x, s_1, s_2, v_2) = p(y|t(s_1, v_2), s_1, s_2, v_2)$ . Now, since  $I(T; Y, S_2 | V_2) - I(T; S_1 | V_2)$  is concave in  $q(t|s_1, v_2)$ , we can use convex optimization methods to derive  $C_{2,w}^{lb}$ .

Denote the PMF

$$p(s_1, s_2, v_2, t, y) \triangleq p(s_1, s_2)w(v_2|s_2)q(t|s_1, v_2)p(y|t, s_1, s_2, v_2), \quad (24)$$

and denote also

$$J_w(q, Q) \triangleq \sum_{s_1, s_2, v_2, t, y} p(s_1, s_2, v_2, t, y) \log \frac{Q(t|y, s_2, v_2)}{q(t|s_1, v_2)}, \quad (25)$$

$$Q^*(t|y, s_2, v_2) \triangleq \frac{\sum_{s_1} p(s_1, s_2, v_2, t, y)}{\sum_{s_1, t'} p(s_1, s_2, v_2, t', y)}. \quad (26)$$

Notice that  $Q^*(t|y, s_2, v_2)$  is a marginal distribution of  $p(s_1, s_2, v_2, t, y)$  and that  $J_w(q, Q^*) = I(T; Y, S_2 | V_2) - I(T; S_1 | V_2)$  for the joint PMF  $p(s_1, s_2, v_2, t, y)$ .

The following lemma is the key for the iterative algorithm.

**Lemma 3.**

$$C_{2,w}^{lb} = \sup_{q'(t|s_1,v_2)} \max_{Q'(t|y,s_2,v_2)} J_w(q', Q'). \quad (27)$$

The proof for this is brought by Yeung in [30]. In addition, Yeung shows that the two-step alternating optimization procedure converges monotonically to the global optimum if the optimization function is concave. Hence, if we show that  $J_w(q, Q)$  is concave, we can maximize it using an alternating maximization algorithm over  $q$  and  $Q$ .

**Lemma 4.** The function  $J_w(q, Q)$  is concave in  $q$  and  $Q$  simultaneously.

We can now proceed to calculate the steps in the iterative algorithm.

**Lemma 5.** For a fixed  $q$ ,  $J_w(q, Q)$  is maximized for  $Q = Q^*$ .

**Proof.** The above follows from the fact that  $Q^*$  is a marginal distribution of  $p(s_1, s_2)w(v_2|s_2)q(t|s_1, v_2)p(y|t, s_1, s_2, v_2)$  and the property of the K-L divergence  $D(Q^*||Q') \geq 0$ .  $\square$

**Lemma 6.** For a fixed  $Q$ ,  $J_w(q, Q)$  is maximized for  $q = q^*$ , where  $q^*$  is defined by

$$q^*(t|s_1, v_2) = \frac{\prod_{s_2, y} Q(t|y, s_2, v_2)^{p(s_2|s_1, v_2)p(y|t, s_1, s_2, v_2)}}{\sum_{t'} \prod_{s_2, y} Q(t'|y, s_2, v_2)^{p(s_2|s_1, v_2)p(y|t', s_1, s_2, v_2)}}, \quad (28)$$

and

$$p(s_2|s_1, v_2) = \frac{p(s_1, s_2)w(v_2|s_2)}{\sum_{s_2'} p(s_1, s_2')w(v_2|s_2')}. \quad (29)$$

Define  $U_w(q)$  in the following way

$$U_w(q) = \sum_{s_1, v_2} p(s_1, v_2) \max_t \sum_{s_2, y} p(s_2|s_1, v_2)p(y|t, s_1, s_2, v_2) \log \frac{Q^*(t|y, s_2, v_2)}{q(t|s_1, v_2)}, \quad (30)$$



where  $Q^*$  is given in (26),  $p(s_1, v_2)$  and  $p(s_2|s_1, v_2)$  are marginal distributions of the joint PMF  $p(s_1, s_2, v_2, t, y) = p(s_1, s_2)w(v_2|s_2)q(t|s_1, v_2)p(y|t, s_1, s_2, v_2)$ . The following lemma will help us to define a termination condition for the algorithm.

**Lemma 7.** For every  $q(t|s_1, v_2)$  the function  $U_w(q)$  is an upper bound on  $C_{w,2}^{lb}$  and converges to  $C_{2,w}^{lb}$  for a large enough number of iterations.

## 5.2. Semi-Iterative Algorithm

The the algorithm for finding  $C_2^{lb}(R')$  is brought in Algorithm 1. Notice that the result of this algorithm,  $C_2^{(\epsilon, \delta, \mathcal{W})}(R')$ , can be arbitrarily close to  $C_2^{lb}(R')$  for  $\epsilon \rightarrow 0$ ,  $\delta \rightarrow 0$  and  $|\mathcal{W}| \rightarrow \infty$ .

---

### Algorithm 1 Numerically calculating $C_2^{lb}(R')$

---

```

1: Chose  $\epsilon > 0, \delta > 0$ 
2: Set  $R' \leftarrow \min\{R', H(S_2|S_1)\}$  {the amount of information needed for the encoder to know  $S_2$  given  $S_1$ }
3: Set  $C \leftarrow -\infty$ 
4: Establish a fine and uniformly spaced grid of legal PMFs  $w(v_2|s_2)$  and name it  $\mathcal{W}$ 
5: for all  $w$  in  $\mathcal{W}$  do
6:   Compute  $R_w$  using

$$R_w = I(V_2; S_2) - I(V_2; S_1)$$

7:   if  $R' - \epsilon \leq R_w \leq R'$  then
8:     Set  $Q(t|y, s_2, v_2)$  to be a uniform distribution over  $\{1, 2, \dots, |\mathcal{T}|\}$ , where  $\mathcal{T}$  is the alphabet of  $t$ .
     i.e.,  $Q(t|y, s_2, v_2) = \frac{1}{|\mathcal{T}|}, \forall t, y, s_2, v_2$ 
9:     repeat
10:      Set  $q(t|s_1, v_2) \leftarrow q^*(t|s_1, v_2)$  using

$$q^*(t|s_1, v_2) = \frac{\prod_{s_2, y} Q(t|y, s_2, v_2)^{p(s_2|s_1, v_2)p(y|t, s_1, s_2, v_2)}}{\sum_{t'} \prod_{s_2, y} Q(t'|y, s_2, v_2)^{p(s_2|s_1, v_2)p(y|t', s_1, s_2, v_2)}}$$

11:      Set  $(Q(t|y, s_2, v_2) \leftarrow Q^*(t|y, s_2, v_2)$  using

$$Q^*(t|y, s_2, v_2) = \frac{\sum_{s_1} p(s_1, s_2, v_2, t, y)}{\sum_{s_1, t'} p(s_1, s_2, v_2, t', y)}$$

12:      Compute  $J_w(q, Q)$  using

$$J_w(q, Q) = \sum_{s_1, s_2, v_2, t, y} p(s_1, s_2, v_2, t, y) \log \frac{Q(t|y, s_2, v_2)}{q(t|s_1, v_2)}$$

13:      Compute  $U_w(q)$  using

$$U_w(q) = \sum_{s_1, v_2} p(s_1, v_2) \max_t \sum_{s_2, y} p(s_2|s_1, v_2) p(y|t, s_1, s_2, v_2) \log \frac{Q^*(t|y, s_2, v_2)}{q(t|s_1, v_2)}$$

14:      until  $U_w(q) - J(q, Q) < \delta$ 
15:      if  $C \leq J_w(q, Q)$  then
16:        Set  $C \leftarrow J_w(q, Q)$ 
17:      end if
18:    end if
19:  end for
20: if  $C < 0$  then {there is no PMF  $w(v_2|s_2) \in \mathcal{W}$  such that  $R_w$  is  $\epsilon$ -close to  $R'$  from below}
21:   go to line 4 and make the grid finer
22: end if
23: Declare  $C_2^{lb(\epsilon, \delta, \mathcal{W})}(R') = C$ 

```

---

## 6. Open Problems

In this section, we discuss the generalization of the channel capacity and the rate-distortion problems that we presented in Section 3. We now consider the cases where the encoder and the decoder are simultaneously informed with a rate-limited description of both the ESI and the DSI, as illustrated in Figure 8. A lower bound on the capacity and an upper bound on the rate-distortion are suggested. Achievability schemes for the presented bounds can be easily derived using the same techniques that we used in the proofs for Theorems 1 and 2, and, hence, are omitted. We were unable to prove that the suggested bounds are tight, nor did we encounter any other such proofs in the published literature; therefore, we believe these problems to be open.

### 6.1. A Lower Bound on the Capacity of a Channel with Two-Sided Increased Partial Side Information

Consider the channel illustrated in Figure 8, where  $(S_{1,i}, S_{2,i})$  i.i.d.  $\sim p(s_1, s_2)$ . The encoder is informed with the ESI ( $S_1^n$ ) and rate-limited DSI and the decoder is informed with the DSI ( $S_2^n$ ) and rate-limited ESI. An  $(n, 2^{nR}, 2^{nR'_1}, 2^{nR'_2})$  code for the discussed channel consists of three encoding maps:

$$\begin{aligned} f_{v1} : S_1^n &\rightarrow \{1, 2, \dots, 2^{nR'_1}\}, \\ f_{v2} : S_2^n &\rightarrow \{1, 2, \dots, 2^{nR'_2}\}, \\ f : \{1, 2, \dots, 2^{nR}\} \times S_1^n \times \{1, 2, \dots, 2^{nR'_2}\} &\rightarrow \mathcal{X}^n, \end{aligned}$$

and a decoding map:

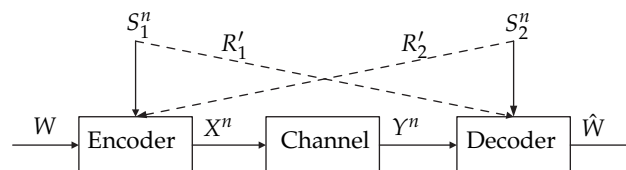
$$g : \mathcal{Y}^n \times S_2^n \times \{1, 2, \dots, 2^{nR'_1}\} \rightarrow \{1, 2, \dots, 2^{nR}\}.$$

*Fact 1:* The channel capacity,  $C_{12}^*$ , of this channel coding setup is bounded from below as follows:

$$C_{12}^* \geq \max_{\substack{p(v_1|s_1)p(v_2|s_2)p(u|s_1,v_1,v_2)p(x|u,s_1,v_1,v_2) \\ \text{s.t. } R'_1 \geq I(V_1;S_1) - I(V_1;Y,S_2,V_2) \\ R'_2 \geq I(V_2;S_2) - I(V_2;S_1)}} I(U;Y,S_2|V_1,V_2) - I(U;S_1|V_1,V_2), \quad (31)$$

for some joint distribution  $p(s_1, s_2, v_1, v_2, u, x, y)$  and  $U, V_1$  and  $V_2$  are some auxiliary random variables.

The proof for the achievability follows closely the proofs given in Appendix B and, therefore, is omitted.



**Figure 8.** Channel coding with two-sided increased partial side information.

### 6.2. An Upper Bound on the Rate-Distortion with Two-Sided Increased Partial Side Information

Consider the rate-distortion problem illustrated in Figure 9, where the source  $X$  and the side information  $S_1, S_2$  are distributed  $(X_i, S_{1,i}, S_{2,i}) \sim \text{i.i.d. } p(x, s_1, s_2)$ . The encoder is informed with the ESI ( $S_1^n$ ) and rate-limited DSI and the decoder is informed with the DSI ( $S_2^n$ ) and rate-limited ESI. An  $(n, 2^{nR}, 2^{nR'_1}, 2^{nR'_2}, D)$  code for the discussed rate-distortion problem consists of three encoding maps:

$$f_{v1} : S_1^n \rightarrow \{1, 2, \dots, 2^{nR'_1}\},$$

$$f_{v2} : \mathcal{S}_2^n \rightarrow \{1, 2, \dots, 2^{nR'_2}\},$$

$$f : \mathcal{X}^n \times \mathcal{S}_1^n \times \{1, 2, \dots, 2^{nR'_2}\} \rightarrow \{1, 2, \dots, 2^{nR}\},$$

and a decoding map:

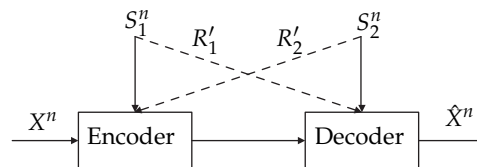
$$g : \{1, 2, \dots, 2^{nR}\} \times \mathcal{S}_2^n \times \{1, 2, \dots, 2^{nR'_1}\} \rightarrow \hat{\mathcal{X}}^n.$$

*Fact 2:* For a given distortion,  $D$ , and a given distortion measure,  $d(X, \hat{X}) : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathbb{R}^+$ , the rate-distortion function  $R_{12}^*(D)$  of this setup is bounded from above as follows:

$$R_{12}^*(D) \leq \min_{\substack{p(v_1|s_1)p(v_2|s_2)p(u|x,s_1,v_1,v_2)p(\hat{x}|u,s_2,v_1,v_2) \\ \text{s.t. } R'_1 \geq I(V_1;S_1) - I(V_1;S_2,V_2) \\ R'_2 \geq I(V_2;S_2) - I(V_2;X,S_1,V_1)}} I(U; X, S_1 | V_1, V_2) - I(U; S_2 | V_1, V_2), \quad (32)$$

for some joint distribution  $p(x, s_1, s_2, v_1, v_2, u, \hat{x})$  where  $\mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n d(X_i, \hat{X}_i) \right] \leq D$  and  $U, V_1$  and  $V_2$  are some auxiliary random variables.

The proof for the achievability follows closely the proofs given in Appendix C and, therefore, is omitted.



**Figure 9.** Source coding with two-sided increased partial side information.

**Acknowledgments:** Avihay Sadeh-Shirazi, Uria Basher and Haim Permuter were supported in part by the Israel Science Foundation and in part by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC under Grant 337752.

**Author Contributions:** Avihay Sadeh-Shirazi and Haim Permuter conceived and designed the study; Avihay Sadeh-Shirazi conducted the research with the aid and supervision of Haim Permuter. Avihay Sadeh-Shirazi wrote the paper with the contribution of Uria Basher. All authors have read and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Duality of the Converse of the Gelfand–Pinsker Theorem and the Wyner–Ziv Theorem

In this appendix, we provide proofs of the converse of the Gelfand–Pinsker capacity and the converse of the Wyner–Ziv rate in a dual way.

Channel capacity		Rate-distortion	
1	$nR = H(W)$	$nR = H(T)$	
2	$\stackrel{(a)}{\leq} I(W; Y^n) - I(W; S^n) + n\epsilon_n$	$\stackrel{(a)}{\geq} I(T; X^n) - I(T; S^n)$	
3	$= \sum_{i=1}^n \left[ I(W; Y_i   Y^{i-1}) - I(W; S_i   S_{i+1}^n) \right] + n\epsilon_n$	$= \sum_{i=1}^n \left[ I(T; X_i   X^{i-1}) - I(T; S_i   S_{i+1}^n) \right]$	
4	$= \sum_{i=1}^n \left[ I(W, S_{i+1}^n; Y_i   Y^{i-1}) - I(W, Y^{i-1}; S_i   S_{i+1}^n) \right] + \Delta - \Delta^* + n\epsilon_n$	$= \sum_{i=1}^n \left[ I(T, S_{i+1}^n; X_i   X^{i-1}) - I(W, X^{i-1}; S_i   S_{i+1}^n) \right] + \Delta - \Delta^*$	(A1)
5	$\stackrel{(b)}{\leq} \sum_{i=1}^n \left[ I(W, Y^{i-1}, S_{i+1}^n; Y_i) - I(W, Y^{i-1}, S_{i+1}^n; S_i) \right] + n\epsilon_n$	$\stackrel{(b)}{\geq} \sum_{i=1}^n \left[ I(T, X^{i-1}, S_{i+1}^n; X_i) - I(T, X^{i-1}, S_{i+1}^n; S_i) \right]$	
6	$= \sum_{i=1}^n \left[ I(U_i; Y_i) - I(U_i; S_i) \right] + n\epsilon_n,$	$= \sum_{i=1}^n \left[ I(U_i; X_i) - I(U_i; S_i) \right],$	

where

$$\begin{aligned}
\Delta &= \sum_{i=1}^n I(Y^{i-1}; S_i | W, S_{i+1}^n), & \Delta &= \sum_{i=1}^n I(X^{i-1}; S_i | T, S_{i+1}^n), \\
\Delta^* &= \sum_{i=1}^n I(S_{i+1}^n; Y_i | W, Y^{i-1}), & \Delta^* &= \sum_{i=1}^n I(S_{i+1}^n; X_i | T, X^{i-1}), \\
(a) & \text{ follows from Fano's inequality and from that fact that } W \text{ is independent of } S^n, & (a) & \text{ follows from Fano's inequality and from the fact that } T \text{ is independent of } S^n, \\
(b) & \text{ follows from the fact that } S_i \text{ is independent of } S_{i+1}^n. & (b) & \text{ follows from the fact that } S_i \text{ is independent of } S_{i+1}^n \text{ and that } X_i \text{ is independent of } X^{i-1}.
\end{aligned}
\tag{A2}$$

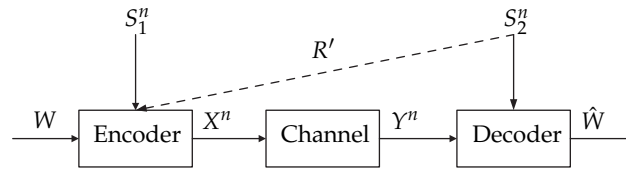
By substituting the output  $Y$  and the input  $X$  in the channel capacity theorem with the input  $X$  and the output  $\hat{X}$  in the rate-distortion theorem, respectively, we can observe duality in the converse proofs of the two theorems.

## Appendix B. Proof of Theorem 1

In this section, we provide the proofs for Theorem 1, Cases 2 and 2<sub>C</sub>. The results for Case 1, where the encoder is informed with ESI and the decoder is informed with increased DSI, can be derived directly from [1] (Section IV). In [1], Steinberg considered the case where the encoder is fully informed with the ESI and the decoder is informed with a rate-limited description of the ESI. Therefore, by considering the DSI,  $S_2^n$ , to be a part of the channel's output, we can apply Steinberg's result on the channel depicted in Case 1. For this reason, the proof for this case is omitted.

### Appendix B.1. Proof of Theorem 1, Case 2

Channel capacity Case 2 is presented in Figure A1. The proof of the lower bound,  $C_2^{lb}$ , is performed in the following way: for the description of the DSI,  $S_2$ , at a rate  $R'$  we use a Wyner-Ziv coding scheme where the source is  $S_2$  and the side information is  $S_1$ . Then, for the channel coding, we use a Gelfand-Pinsker coding scheme where the state information at the encoder is  $S_1$ ,  $S_2$  is a part of the channel's output and the rate-limited description of  $S_2$  is side information at both the encoder and the decoder. Notice that  $I(U; Y, S_2 | V_2) - I(U; S_1, V_2) = I(U; Y, S_2, V_2) - I(U; S_1, V_2)$  and that, since the Markov chain  $V_2 - S_2 - S_1$  holds, we can also write  $R' \geq I(V_2; S_2) - I(V_2; S_1)$ . We make use of these expressions in the following proof.



**Figure A1.** Channel capacity: Case 2. Lower bound:  $C_2^{lb} = \max I(U; Y, S_2 | V_2) - I(U; S_1 | V_2)$ , where the maximization is over all joint PMFs  $p(s_1, s_2, v_2, u, x, y)$  that maintain the Markov relations  $U - (S_1, V_2) - S_2$  and  $V_2 - S_2 - S_1$  and the constraint  $R' \geq I(V_2; S_2 | S_1)$ . Upper bounds:  $C_2^{ub1}$  is the result of the same expressions as for the lower bound, except that the maximization is taken over all PMFs that maintain the Markov chain  $U - (S_1, V_2) - S_2$ , and  $C_2^{ub2}$  is the result of the same expressions as for the lower bound, except that this time the maximization is taken over all PMFs that maintain  $V_2 - S_2 - S_1$ .

**Achievability:** (Channel capacity Case 2—Lower bound). Given  $(S_{1,i}, S_{2,i}) \sim$  i.i.d.  $p(s_1, s_2)$  and the memoryless channel  $p(y|x, s_1, s_2)$ , fix  $p(s_1, s_2, v_2, u, x, y) = p(s_1, s_2)p(v_2|s_2)p(u|s_1, v_2)p(x|u, s_1, v_2)p(y|x, s_1, s_2)$ , where  $x = f(u, s_1, v_2)$  (i.e.,  $p(x|u, s_1, v_2)$  can get the values 0 or 1).

#### Codebook generation and random binning

1. Generate a codebook  $\mathcal{C}_v$  of  $2^{n(I(V_2; S_2) + 2\epsilon)}$  sequences  $V_2^n$  independently using i.i.d.  $\sim p(v_2)$ . Label them  $v_2^n(k)$ , where  $k \in \{1, 2, \dots, 2^{n(I(V_2; S_2) + 2\epsilon)}\}$ , and randomly assign each sequence  $v_2^n(k)$  a bin number  $b_v(v_2^n(k))$  in the set  $\{1, 2, \dots, 2^{nR'}\}$ .
2. Generate a codebook  $\mathcal{C}_u$  of  $2^{n(I(U; Y, S_2, V_2) - 2\epsilon)}$  sequences  $U^n$  independently using i.i.d.  $\sim p(u)$ . Label them  $u^n(l)$ ,  $l \in \{1, 2, \dots, 2^{n(I(U; Y, S_2, V_2) - 2\epsilon)}\}$ , and randomly assign each sequence a bin number  $b_u(u^n(l))$  in the set  $\{1, 2, \dots, 2^{nR}\}$ .

Reveal the codebooks and the content of the bins to all encoders and decoders.

#### Encoding

1. *State Encoder:* Given the sequence  $S_2^n$ , search the codebook  $\mathcal{C}_v$  and identify an index  $k$  such that  $(v_2^n(k), S_2^n) \in \mathcal{T}_\epsilon^{(n)}(V_2, S_2)$ . If such a  $k$  is found, stop searching and send the bin number  $j = b_v(v_2^n(k))$ . If no such  $k$  is found, declare an error.
2. *Encoder:* Given the message  $W$ , the sequence  $S_1^n$  and the index  $j$ , search the codebook  $\mathcal{C}_v$  and identify an index  $k$  such that  $(v_2^n(k), S_1^n) \in \mathcal{T}_\epsilon^{(n)}(V_2, S_1)$ . If no such  $k$  is found or there is more than one such index, declare an error. If a unique  $k$ , as defined, is found, search the codebook  $\mathcal{C}_u$  and identify an index  $l$  such that  $(u^n(l), S_1^n, v_2^n(k)) \in \mathcal{T}_\epsilon^{(n)}(U, S_1, V_2)$  and  $b_u(u^n(l)) = W$ . If a unique  $l$ , as defined, is found, transmit  $x_i = f(u_i(l), S_{1,i}, v_{2,i}(k))$ ,  $i = 1, 2, \dots, n$ . Otherwise, if there is no such  $l$  or there is more than one, declare an error.

#### Decoding

Given the sequences  $Y^n, S_2^n$  and the index  $k$ , search the codebook  $\mathcal{C}_u$  and identify an index  $l$  such that  $(u^n(l), Y^n, S_2^n, v_2^n(k)) \in \mathcal{T}_\epsilon^{(n)}(U, Y, S_2, V_2)$ . If a unique  $l$ , as defined, is found, declare the message  $\hat{W}$  to be the bin index where  $u^n(l)$  is located, i.e.,  $\hat{W} = b_u(u^n(l))$ . Otherwise, if no such  $l$  is found or there is more than one, declare an error.

#### Analysis of the probability of error

Without loss of generality, let us assume that the message  $W = 1$  was sent and the indexes that correspond with the given  $W = 1, S_1^n, S_2^n$  are  $(k = 1, l = 1$  and  $j = 1)$ ; i.e.,  $v_2^n(1)$  corresponds with  $S_2^n$ ,  $b_v(v_2^n(1)) = 1$ ,  $u^n(1)$  is chosen according to  $(W = 1, S_1^n, v_2^n(1))$  and  $b_u(u^n(1)) = 1$ .

Define the following events:

$$\begin{aligned}
 E_1 &:= \left\{ \forall v_2^n(k) \in \mathcal{C}_v, (v_2^n(k), S_2^n) \notin \mathcal{T}_\epsilon^{(n)}(V_2, S_2) \right\} \\
 E_2 &:= \left\{ (v_2^n(1), S_1^n) \notin \mathcal{T}_\epsilon^{(n)}(V_2, S_1) \right\} \\
 E_3 &:= \left\{ \exists k' \neq 1 \text{ such that } b_v(v_2^n(k')) = 1 \text{ and } (v_2^n(k'), S_1^n) \in \mathcal{T}_\epsilon^{(n)}(V_2, S_1) \right\} \\
 E_4 &:= \left\{ \forall u^n(l) \in \mathcal{C}_u \text{ such that } b_u(u^n(l)) = 1, (u^n(l), S_1^n, v_2^n(1)) \notin \mathcal{T}_\epsilon^{(n)}(U, S_1, V_2) \right\} \\
 E_5 &:= \left\{ (u^n(1), Y^n, S_2^n, v_2^n(1)) \notin \mathcal{T}_\epsilon^{(n)}(U, Y, S_2, V_2) \right\} \\
 E_6 &:= \left\{ \exists l' \neq 1 \text{ such that } (u^n(l'), Y^n, S_2^n, v_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}(U, Y, S_2, V_2) \right\}
 \end{aligned}$$

The probability of error  $P_e^{(n)}$  is upper bounded by  $P_e^n \leq P(E_1) + P(E_2|E_1^c) + P(E_3|E_1^c, E_2^c) + P(E_4|E_1^c, E_2^c, E_3^c) + P(E_5|E_1^c, \dots, E_4^c) + P(E_6|E_1^c, \dots, E_5^c)$ . Using standard arguments, and assuming that  $(S_1^n, S_2^n) \in \mathcal{T}_\epsilon^{(n)}(S_1, S_2)$  and that  $n$  is large enough, we can state that

1.

$$\begin{aligned}
 P(E_1) &= \Pr \left\{ \bigcap_{v_2^n(k) \in \mathcal{C}_v} (v_2^n(k), S_2^n) \notin \mathcal{T}_\epsilon^{(n)}(V_2, S_2) \right\} \\
 &= \prod_{k=1}^{2^{n(I(V_2; S_2) + 2\epsilon)}} \Pr \left\{ (v_2^n(k), S_2^n) \notin \mathcal{T}_\epsilon^{(n)}(V_2, S_2) \right\} \\
 &= \prod_{k=1}^{2^{n(I(V_2; S_2) + 2\epsilon)}} \left( 1 - \Pr \left\{ (v_2^n(k), S_2^n) \in \mathcal{T}_\epsilon^{(n)}(V_2, S_2) \right\} \right) \\
 &\leq \left( 1 - 2^{-n(I(V_2; S_2) + \epsilon)} \right)^{2^{n(I(V_2; S_2) + 2\epsilon)}} \\
 &\leq e^{-2^{-n(I(V_2; S_2) + \epsilon)} 2^{n(I(V_2; S_2) + 2\epsilon)}} \\
 &= e^{-2^{n\epsilon}}.
 \end{aligned} \tag{A3}$$

The probability that there is no  $v_2^n(k)$  in  $\mathcal{C}_v$  such that  $(v_2^n(k), S_2^n)$  is strongly jointly typical is exponentially small provided that  $|\mathcal{C}_v| \geq 2^{n(I(V_2; S_2) + \epsilon)}$ . This follows from the standard rate-distortion argument that  $2^{nI(V_2; S_2)}$   $v_2^n$ 's "cover"  $S_2^n$ , therefore  $P(E_1) \rightarrow 0$ .

2. By the Markov lemma [31], since  $(S_1^n, S_2^n)$  are strongly jointly typical,  $(S_2^n, v_2^n(1))$  are strongly jointly typical and the Markov chain  $S_1 - S_2 - V_2$  holds, then  $(S_1^n, S_2^n, v_2^n(1))$  are strongly jointly typical with high probability. Therefore,  $P(E_2|E_1^c) \rightarrow 0$ .
- 3.

$$P(E_3|E_1^c, E_2^c) = \Pr \left\{ \bigcup_{\substack{v_2^n(k') \neq 1 \in \mathcal{C}_v \\ b_v(v_2^n(k')) = 1}} (v_2^n(k'), S_1^n) \in \mathcal{T}_\epsilon^{(n)}(V_2, S_1) \right\} \tag{A4}$$

$$\leq \sum_{\substack{v_2^n(k') \neq 1 \in \mathcal{C}_v \\ b_v(v_2^n(k')) = 1}} \Pr \left\{ (v_2^n(k'), S_1^n) \in \mathcal{T}_\epsilon^{(n)}(V_2, S_1) \right\} \tag{A5}$$

$$\leq \sum_{\substack{v_2^n(k') \neq 1 \in \mathcal{C}_v \\ b_v(v_2^n(k')) = 1}} 2^{-n(I(V_2; S_1) - \epsilon)} \tag{A6}$$

$$= 2^{n(I(V_2; S_2) + 2\epsilon - R')} 2^{-n(I(V_2; S_1) - \epsilon)} \tag{A7}$$

$$= 2^{n(I(V_2; S_2) - I(V_2; S_1) + 3\epsilon - R')}. \quad (\text{A8})$$

The probability that there is another index  $k'$ ,  $k' \neq 1$ , such that  $v_2^n(k')$  is in bin number 1 and that is strongly jointly typical with  $S_1^n$  is bounded by the number of  $v_2^n(k')$ 's in the bin times the probability of joint typicality. Therefore, if the number of bins  $R' > I(V_2; S_2) - I(V_2; S_1) + 3\epsilon$  then  $P(E_3|E_1^c, E_2^c) \rightarrow 0$ .

4. We use here the same argument we used for  $P(E_1)$ ; by the covering lemma, we can state that the probability that there is no  $u^n(l)$  in bin number 1 that is strongly jointly typical with  $(S_1^n, v_2^n(1))$  tends to zero for large enough  $n$  if the average number of  $u^n(l)$ 's in each bin is greater than  $2^{n(I(U; S_1, V_2) + \epsilon)}$ ; i.e.,  $|C_u|/2^{nR} > 2^{n(I(U; S_1, V_2) + \epsilon)}$ . This also implies that in order to avoid an error the number of words one should use is  $R < I(U; Y, S_2, V_2) - I(U; S_1, V_2) - 3\epsilon$ , where the last expression also equals  $I(U; Y, S_2|V_2) - I(U; S_1|V_2) - 3\epsilon$ .
5. As we argued for  $P(E_2|E_1^c)$ , since  $(X^n, u^n(1), S_1^n, v_2^n(1))$  is strongly jointly typical,  $(Y^n, X^n, S_1^n, S_2^n)$  is strongly jointly typical and the Markov chain  $(U, V_2) - (X, S_1, S_2) - Y$  holds, then, by the Markov lemma,  $(u^n(1), Y^n, S_2^n, v_2^n(1))$  is strongly jointly typical with high probability, i.e.,  $P(E_5|E_1^c, \dots, E_4^c) \rightarrow 0$ .
- 6.

$$\begin{aligned} P(E_6|E_1^c, \dots, E_5^c) &= \Pr \left\{ \bigcup_{u^n(l' \neq 1) \in C_u} (u^n(l'), Y^n, S_2^n, v_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}(U, Y, S_2, V_2) \right\} \\ &\leq \sum_{l'=2}^{2^{n(I(U; Y, S_2, V_2) - 2\epsilon)}} \Pr \left\{ (u^n(l'), Y^n, S_2^n, V_2^n) \in \mathcal{T}_\epsilon^{(n)}(U, Y, S_2, V_2) \right\} \\ &\leq \sum_{l'=2}^{2^{n(I(U; Y, S_2, V_2) - 2\epsilon)}} 2^{-n(I(U; Y, S_2, V_2) - \epsilon)} \\ &\leq 2^{n(I(U; Y, S_2, V_2) - 2\epsilon)} 2^{-n(I(U; Y, S_2, V_2) - \epsilon)} \\ &= 2^{-n\epsilon}. \end{aligned} \quad (\text{A9})$$

The probability that there is another index  $l'$ ,  $l' \neq 1$ , such that  $u^n(l')$  is strongly jointly typical with  $(Y^n, S_2^n, v_2^n(1))$  is bounded by the total number of  $u^n$ 's times the probability of joint typicality. Therefore, taking  $|C_u| \leq 2^{n(I(U; Y, S_2, V_2) - 2\epsilon)}$  assures us that  $P(E_6|E_1^c, \dots, E_5^c) \rightarrow 0$ . This follows the standard channel capacity argument that one can distinguish at most  $2^{nI(U; Y, S_2, V_2)}$  different  $u^n(l)$ 's given any typical member of  $\mathcal{Y}^n \times \mathcal{S}_2^n \times \mathcal{V}_2^n$ .

This shows that for rates  $R$  and  $R'$  as described and for large enough  $n$ , the error events are of arbitrarily small probability. This concludes the proof of the achievability and the lower bound on the capacity of Case 2.

**Converse:** (Channel capacity Case 2—Upper bound). We first prove that it is possible to bound the capacity from above by using two random variables,  $U$  and  $V$ , that maintain the Markov chain  $U - (S_1, V_2) - S_2$  (that is  $C_2^{ub1}$ ). Then, we prove that it is also possible to upper-bound the capacity by using  $U$  and  $V$  that maintain the Markov relation  $V_2 - S_2 - S_1$  (that is  $C_2^{ub2}$ ).

Fix the rates  $R$  and  $R'$  and a sequence of codes  $(n, 2^{nR}, 2^{nR'})$  that achieve the capacity. By Fano's inequality,  $H(W|Y^n, S_2^n) \leq n\epsilon_n$ , where  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . Let  $T_2 = f_v(S_2^n)$ , and define  $V_{2,i} = (T_2, Y^{i-1}, S_{1,i+1}^n, S_2^{i-1})$ ,  $U_i = W$ ; hence, the Markov chain  $U_i - (S_{1,i}, V_{2,i}) - S_{2,i}$  is maintained. The proof for this follows.

$$\begin{aligned} p(u_i|s_{1,i}, v_{2,i}, s_{2,i}) &= p(w|s_{1,i}, t_2, y^{i-1}, s_{1,i+1}^n, s_2^{i-1}, s_{2,i}) \\ &= \sum_{x^{i-1}, s_1^{i-1}} p(w, x^{i-1}, s_1^{i-1}|s_{1,i}, t_2, y^{i-1}, s_{1,i+1}^n, s_2^{i-1}, s_{2,i}) \\ &= \sum_{x^{i-1}, s_1^{i-1}} p(s_1^{i-1}|t_2, y^{i-1}, s_{1,i}, s_2^{i-1}) p(x^{i-1}|t_2, y^{i-1}, s_1^n, s_2^{i-1}) p(w|x^{i-1}, t_2, y^{i-1}, s_1^n, s_2^{i-1}) \end{aligned}$$



$$=p(w|t_2, y^{i-1}, s_{1,i+1}^n, s_2^{i-1}, s_{1,i}). \quad (\text{A10})$$

Next, consider

$$\begin{aligned} nR' &\geq H(T_2) \\ &\geq H(T_2|S_1^n) - H(T_2|S_1^n, S_2^n) \\ &= I(T_2; S_2^n|S_1^n) \\ &= H(S_2^n|S_1^n) - H(S_2^n|T_2, S_1^n) \\ &= \sum_{i=1}^n [H(S_{2,i}|S_1^n, S_2^{i-1}) - H(S_{2,i}|T_2, S_1^n, S_2^{i-1})] \\ &\stackrel{(a)}{=} \sum_{i=1}^n [H(S_{2,i}|S_{1,i}) - H(S_{2,i}|T_2, S_1^n, S_2^{i-1}, Y^{i-1})] \\ &\stackrel{(b)}{=} \sum_{i=1}^n [H(S_{2,i}|S_{1,i}) - H(S_{2,i}|T_2, S_{1,i+1}^n, S_2^{i-1}, Y^{i-1}, S_{1,i})] \\ &= \sum_{i=1}^n [H(S_{2,i}|S_{1,i}) - H(S_{2,i}|V_{2,i}, S_{1,i})] \\ &= \sum_{i=1}^n I(S_{2,i}; V_{2,i}|S_{1,i}), \end{aligned} \quad (\text{A11})$$

where (a) follows from the fact that  $S_{2,i}$  is independent of  $(S_1^{i-1}, S_{1,i+1}^n, S_2^{i-1})$  given  $S_{1,i}$ , and the fact that  $Y^{i-1}$  is independent of  $S_{2,i}$  given  $(T_2, S_1^n, S_2^{i-1})$  (the proof for this follows) and (b) follows from the fact that conditioning reduces entropy.

$$\begin{aligned} p(y^{i-1}|t_2, s_1^n, s_2^{i-1}, s_{2,i}) &= \sum_{x^n, w} p(y^{i-1}, x^n, w|t_2, s_1^n, s_2^{i-1}, s_{2,i}) \\ &= \sum_{x^n, w} p(w)p(x^n|w, t_2, s_1^n)p(y^{i-1}|x^{i-1}, s_1^{i-1}, s_2^{i-1}) \\ &= p(y^{i-1}|t_2, s_1^n, s_2^{i-1}), \end{aligned} \quad (\text{A12})$$

where we used the facts that  $W$  is independent of  $(T_2, S_1^n, S_{2,i}^n)$ ,  $X^n$  is a function of  $(W, T_2, S_1^n)$  and that the channel is memoryless; i.e.,  $Y^{i-1}$  is independent of  $(W, T_2, S_{1,i}^n, S_{2,i}^n)$  given  $(X^{i-1}, S_1^{i-1}, S_2^{i-1})$ . We continue the proof of the converse by considering the following set of inequalities:

$$\begin{aligned} nR &= H(W) \\ &\leq H(W|T_2) - H(W|T_2, Y^n, S_2^n) + n\epsilon_n \\ &= I(W; Y^n, S_2^n|T_2) + n\epsilon_n \\ &= \sum_{i=1}^n I(W; Y_i, S_{2,i}|T_2, Y^{i-1}, S_2^{i-1}) + n\epsilon_n \\ &\stackrel{(b)}{=} \sum_{i=1}^n [I(W, S_{1,i+1}^n; Y_i, S_{2,i}|T_2, Y^{i-1}, S_2^{i-1}) \\ &\quad - I(S_{1,i+1}^n; Y_i, S_{2,i}|W, T_2, Y^{i-1}, S_2^{i-1})] + n\epsilon_n \\ &\stackrel{(c)}{=} \sum_{i=1}^n [I(W, S_{1,i+1}^n; Y_i, S_{2,i}|T_2, Y^{i-1}, S_2^{i-1}) \\ &\quad - I(S_{1,i}; Y^{i-1}, S_2^{i-1}|W, T_2, S_{1,i+1}^n)] + n\epsilon_n \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \left[ I(W; Y_i, S_{2,i} | T_2, Y^{i-1}, S_{1,i+1}^n, S_2^{i-1}) \right. \\
&\quad \left. - I(S_{1,i}; W | T_2, Y^{i-1}, S_{1,i+1}^n, S_2^{i-1}) \right] \\
&\quad + \Delta - \Delta^* + n\epsilon_n,
\end{aligned} \tag{A13}$$

where

$$\Delta = \sum_{i=1}^n I(S_{1,i+1}^n; Y_i, S_{2,i} | T_2, Y^{i-1}, S_2^{i-1}), \tag{A14}$$

$$\Delta^* = \sum_{i=1}^n I(S_{1,i}; Y^{i-1}, S_2^{i-1} | T_2, S_{1,i+1}^n), \tag{A15}$$

(b) follows from the mutual information properties and (c) follows from the Csiszár sum identity.

By using the Csiszár sum on (A14) and (A15), we get

$$\Delta = \Delta^*, \tag{A16}$$

and, therefore, from (A11) and (A13)

$$R' \geq \frac{1}{n} \sum_{i=1}^n I(S_{2,i}; V_{2,i} | S_{1,i}) \tag{A17}$$

$$R - \epsilon_n \leq \frac{1}{n} \sum_{i=1}^n \left[ I(U_i; Y_i, S_{2,i} | V_{2,i}) - I(U_i; S_{1,i} | V_{2,i}) \right]. \tag{A18}$$

Using the convexity of  $R'$  and Jensen's inequality, the standard time sharing argument for  $R$  and the fact that  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ , we can conclude that

$$R' \geq I(V_2; S_2 | S_1), \tag{A19}$$

$$R \leq I(U; Y, S_2 | V_2) - I(U; S_1 | V_2), \tag{A20}$$

where  $U$  and  $V$  maintain the Markov chain  $U - (S_1, V_2) - S_2$ .

We now proceed to prove that it is possible to upper-bound the capacity of Case 2 by using two random variables,  $U$  and  $V$ , that maintain the Markov chain  $V_2 - S_2 - S_1$ . Fix the rates  $R$  and  $R'$  and a sequence of codes  $(n, 2^{nR}, 2^{nR'})$  that achieve the capacity. By Fano's inequality,  $H(W | Y^n, S_2^n) \leq n\epsilon_n$ , where  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . Let  $T_2 = f_v(S_2^n)$  and define  $V_{2,i} = (T_2, S_2^{i-1})$ ,  $U_i = (W, Y^{i-1}, S_{1,i+1}^n)$ . The Markov chain  $V_{2,i} - S_{2,i} - S_{1,i}$  is maintained. Then,

$$\begin{aligned}
nR' &\geq H(T_2) \\
&\stackrel{(a)}{=} \sum_{i=1}^n \left[ H(S_{2,i} | S_1^n, S_2^{i-1}) - H(S_{2,i} | T_2, S_1^n, S_2^{i-1}) \right] \\
&\stackrel{(b)}{=} \sum_{i=1}^n \left[ H(S_{2,i} | S_{1,i}) - H(S_{2,i} | T_2, S_{1,i}, S_{1,i+1}^n, S_2^{i-1}) \right] \\
&\geq \sum_{i=1}^n \left[ H(S_{2,i} | S_{1,i}) - H(S_{2,i} | T_2, S_{1,i}, S_2^{i-1}) \right]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \left[ H(S_{2,i}|S_{1,i}) - H(S_{2,i}|V_{2,i}, S_{1,i}) \right] \\
&= \sum_{i=1}^n I(S_{2,i}; V_{2,i}|S_{1,i}),
\end{aligned} \tag{A21}$$

where (a) follows from the same reasoning as in (A11), and (b) follows from the fact that  $S_{2,i}$  is independent of  $(S_1^{i-1}, S_{1,i+1}^n, S_2^{i-1})$  given  $S_{1,i}$ , and the fact that  $(Y^{i-1}, S_1^{i-1})$  is independent of  $S_{2,i}$  given  $(T_2, S_{1,i}^n, S_2^{i-1})$ ; the proof for this follows.

$$\begin{aligned}
p(y^{i-1}, s_1^{i-1}|t_2, s_{1,i}^n, s_2^{i-1}, s_{2,i}) &= \sum_{x^n, w} p(y^{i-1}, s_1^{i-1}, x^n, w|t_2, s_{1,i}^n, s_2^{i-1}, s_{2,i}) \\
&= \sum_{x^n, w} p(w)p(s_1^{i-1}|s_2^{i-1})p(x^n|w, t_2, s_1^n)p(y^{i-1}|x^{i-1}, s_1^{i-1}, s_2^{i-1}) \\
&= p(y^{i-1}, s_1^{i-1}|t_2, s_{1,i}^n, s_2^{i-1}),
\end{aligned} \tag{A22}$$

where we used the facts that  $W$  is independent of  $(T_2, S_{1,i}^n, S_{2,i}^n)$ ,  $S_1^{i-1}$  is independent of  $(T_2, S_{1,i}^n, S_{2,i}^n)$  given  $S_2^{i-1}$ ,  $X^n$  is a function of  $(W, T_2, S_1^n)$  and that the channel is memoryless; i.e.,  $Y^{i-1}$  is independent of  $(W, T_2, S_{1,i}^n, S_{2,i}^n)$  given  $(X^{i-1}, S_1^{i-1}, S_2^{i-1})$ .

In order to complete our proof, we need the following lemma.

**Lemma A1.** *The following inequality holds:*

$$\sum_{i=1}^n I(S_{1,i}; W, Y^{i-1}, S_{1,i+1}^n | T_2, S_2^{i-1}) \leq \sum_{i=1}^n I(S_{1,i}; W, Y^{i-1}, S_2^{i-1} | T_2, S_{1,i+1}^n). \tag{A23}$$

**Proof.** Notice that

$$\sum_{i=1}^n I(S_{1,i}; W, Y^{i-1}, S_{1,i+1}^n | T_2, S_2^{i-1}) = \sum_{i=1}^n I(S_{1,i}; W, Y^{i-1}, S_{1,i+1}^n, S_2^{i-1} | T_2) - I(S_{1,i}; S_2^{i-1} | T_2) \tag{A24}$$

and that

$$\sum_{i=1}^n I(S_{1,i}; W, Y^{i-1}, S_2^{i-1} | T_2, S_{1,i+1}^n) = \sum_{i=1}^n I(S_{1,i}; W, Y^{i-1}, S_{1,i+1}^n, S_2^{i-1} | T_2) - I(S_{1,i}; S_{1,i+1}^n | T_2). \tag{A25}$$

Therefore, it is enough to show that  $\sum_{i=1}^n -I(S_{1,i}; S_2^{i-1} | T_2) \leq \sum_{i=1}^n -I(S_{1,i}; S_{1,i+1}^n | T_2)$  holds in order to prove the lemma. Consider

$$\begin{aligned}
\sum_{i=1}^n -I(S_{1,i}; S_{1,i+1}^n | T_2) - \left( \sum_{i=1}^n -I(S_{1,i}; S_2^{i-1} | T_2) \right) &= \sum_{i=1}^n H(S_{1,i} | T_2, S_{1,i+1}^n) - H(S_{1,i} | T_2, S_2^{i-1}) \\
&= \sum_{i=1}^n H(S_1^n | T_2) - H(S_{1,i} | T_2, S_2^{i-1}) \\
&= \sum_{i=1}^n H(S_{1,i} | T_2, S_1^{i-1}) - H(S_{1,i} | T_2, S_2^{i-1}) \\
&\stackrel{(a)}{\geq} 0,
\end{aligned} \tag{A26}$$

where (a) follows from the fact that the Markov chain  $S_{1,i} - (T_2, S_2^{i-1}) - (T_2, S_1^{i-1})$  holds and from the data processing inequality. This completes the proof of the lemma.  $\square$

We continue the proof of the converse by considering the following set of inequalities:

$$\begin{aligned}
nR &= H(W) \\
&\leq H(W|T_2) - H(W|T_2, Y^n, S_2^n) + n\epsilon_n \\
&= I(W; Y^n, S_2^n | T_2) + n\epsilon_n \\
&= \sum_{i=1}^n I(W; Y_i, S_{2,i} | T_2, Y^{i-1}, S_2^{i-1}) + n\epsilon_n \\
&\stackrel{(a)}{=} \sum_{i=1}^n \left[ I(W, S_{1,i+1}^n; Y_i, S_{2,i} | T_2, Y^{i-1}, S_2^{i-1}) \right. \\
&\quad \left. - I(S_{1,i+1}^n; Y_i, S_{2,i} | W, T_2, Y^{i-1}, S_2^{i-1}) \right] + n\epsilon_n \\
&\stackrel{(b)}{=} \sum_{i=1}^n \left[ I(W, S_{1,i+1}^n; Y_i, S_{2,i} | T_2, Y^{i-1}, S_2^{i-1}) \right. \\
&\quad \left. - I(S_{1,i}; Y^{i-1}, S_2^{i-1} | W, T_2, S_{1,i+1}^n) \right] + n\epsilon_n \\
&= \sum_{i=1}^n \left[ I(W, S_{1,i+1}^n; Y_i, S_{2,i} | T_2, Y^{i-1}, S_2^{i-1}) \right. \\
&\quad \left. - I(S_{1,i}; W, Y^{i-1}, S_2^{i-1} | T_2, S_{1,i+1}^n) \right] + n\epsilon_n \\
&\stackrel{(c)}{\leq} \sum_{i=1}^n \left[ I(W, S_{1,i+1}^n; Y_i, S_{2,i} | T_2, Y^{i-1}, S_2^{i-1}) \right. \\
&\quad \left. - I(S_{1,i}; W, Y^{i-1}, S_{1,i+1}^n | T_2, S_2^{i-1}) \right] + n\epsilon_n \\
&\leq \sum_{i=1}^n \left[ I(W, Y^{i-1}, S_{1,i+1}^n; Y_i, S_{2,i} | T_2, S_2^{i-1}) \right. \\
&\quad \left. - I(S_{1,i}; W, Y^{i-1}, S_{1,i+1}^n | T_2, S_2^{i-1}) \right] + n\epsilon_n \\
&= \sum_{i=1}^n I(U_i; Y_i, S_{1,i+1}^n | V_{2,i}) - I(U_i; S_{1,i} | V_{2,i}), \tag{A27}
\end{aligned}$$

where (a) follows from the mutual information properties, (b) follows from the Csiszár sum identity and (c) follows from Lemma A1. Therefore,

$$R' \geq \frac{1}{n} \sum_{i=1}^n I(S_{2,i}; V_{2,i} | S_{1,i}) \tag{A28}$$

$$R - \epsilon_n \leq \frac{1}{n} \sum_{i=1}^n \left[ I(U_i; Y_i, S_{2,i} | V_{2,i}) - I(U_i; S_{1,i} | V_{2,i}) \right]. \tag{A29}$$

Using the convexity of  $R'$  and Jensen's inequality, the standard time sharing argument for  $R$  and the fact that  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ , we can conclude that

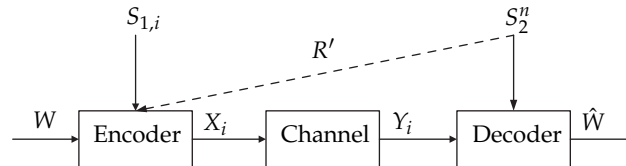
$$R' \geq I(V_2; S_2 | S_1), \tag{A30}$$

$$R \leq I(U; Y, S_2 | V_2) - I(U; S_1 | V_2), \tag{A31}$$

where the Markov chain  $V_2 - S_2 - S_1$  holds. Therefore, we can conclude that the expression given in (12) is an upper-bound to any achievable rate. This concludes the proof of the upper-bound and the proof of Theorem 1 Case 2.

### Appendix B.2. Proof of Theorem 1, Case 2<sub>C</sub>

Channel capacity Case 2<sub>C</sub> is illustrated in Figure A2. For describing the DSI,  $S_2$ , with a rate  $R'$  we use the standard rate-distortion coding scheme. Then, for the channel coding we use the Shannon strategy [4] coding scheme where the channel's causal state information at the encoder is  $S_1$ ,  $S_2$  is a part of the channel's output and the rate-limited description of  $S_2$  is the side information at both the encoder and the decoder.



**Figure A2.** Channel capacity: Case 2 with causal ESI.  $C_{2C} = \max I(U; Y, S_2 | V_2)$ , where the maximization is over all PMFs  $p(v_2 | s_2) p(u | v_2) p(x | u, s_1, v_2)$  such that  $R' \geq I(V_2; S_2)$ .

**Achievability:** (Channel capacity Case 2<sub>C</sub>). Given  $(S_{1,i}, S_{2,i}) \sim \text{i.i.d. } p(s_1, s_2)$ , where the ESI is known in a causal way ( $S_1^i$  at time  $i$ ), and the memoryless channel  $p(y | x, s_1, s_2)$ , fix  $p(s_1, s_2, v_2, u, x, y) = p(s_1, s_2) p(v_2 | s_2) p(u | v_2) p(x | u, s_1, v_2) p(y | x, s_1, s_2)$ , where  $x = f(u, s_1, v_2)$  (i.e.,  $p(x | u, s_1, v_2)$  can get the values 0 or 1).

*Codebook generation and random binning*

1. Generate a codebook  $\mathcal{C}_v$  of  $2^{n(I(V_2; S_2) + 2\epsilon)}$  sequences  $V_2^n$  independently using i.i.d.  $\sim p(v_2)$ . Label them  $v_2^n(k)$  where  $k \in \{1, 2, \dots, 2^{n(I(V_2; S_2) + 2\epsilon)}\}$ .
2. For each  $v_2^n(k)$  generate a codebook  $\mathcal{C}_u(k)$  of  $2^{n(I(U; Y, S_2 | V_2) - 2\epsilon)}$  sequences  $U^n$  distributed independently according to i.i.d.  $\sim p(u | v_2)$ . Label them  $u^n(w, k)$ , where  $w \in \{1, 2, \dots, 2^{n(I(U; Y, S_2 | V_2) - 2\epsilon)}\}$ , and associate the sequences  $u^n(w, \cdot)$  with the message  $W = w$ .

Reveal the codebooks and the content of the bins to all encoders and decoders.

*Encoding*

1. *State Encoder:* Given the sequence  $S_2^n$ , search the codebook  $\mathcal{C}_v$  and identify an index  $k$  such that  $(v_2^n(k), S_2^n) \in \mathcal{T}_\epsilon^{(n)}(V_2, S_2)$ . If such a  $k$  is found, stop searching and send it. Otherwise, if no such  $k$  is found, declare an error.
2. *Encoder:* Given the message  $W \in \{1, 2, \dots, 2^{n(I(U; Y, S_2 | V_2) - 2\epsilon)}\}$ , the index  $k$  and  $S_1^i$  at time  $i$ , identify  $u^n(W, k)$  in the codebook  $\mathcal{C}_u(k)$  and transmit  $x_i = f(u_i(W, k), S_{1,i}, v_{2,i}(k))$  at any time  $i \in \{1, 2, \dots, n\}$ . The element  $x_i$  is the result of a multiplexer with an input signal  $(u_i(W, k), v_{2,i}(k))$  and a control signal  $S_{1,i}$ .

*Decoding*

Given  $Y^n, S_2^n$  and  $k$ , look for a unique index  $\hat{W}$ , associated with the sequence  $u^n(\hat{W}, k) \in \mathcal{C}_u(k)$ , such that  $(Y^n, S_2^n, u^n(\hat{W}, k)) \in \mathcal{T}_\epsilon^{(n)}(Y, U, S_2 | v_2^n(k))$ . If a unique such  $\hat{W}$  is found, declare that the sent message was  $\hat{W}$ . Otherwise, if no unique index  $\hat{W}$  exists, declare an error.

*Analysis of the probability of error*

Without loss of generality, let us assume that the message  $W = 1$  was sent and the index  $k$  that correspond with  $S_2^n$  is  $k = 1$ ; i.e.,  $v_2^n(1)$  corresponds to  $S_2^n$  and  $u^n(1, 1)$  is chosen according to  $(W = 1, v_2^n(1))$ .

Define the following events:

$$E_1 := \left\{ \forall v_2^n(k) \in \mathcal{C}_v, (S_2^n, v_2^n(k)) \notin \mathcal{T}_\epsilon^{(n)}(S_2, V_2) \right\}$$

$$E_2 := \left\{ (u^n(1,1), Y^n, S_2^n) \notin \mathcal{T}_\epsilon^{(n)}(U, Y, S_2 | v_2^n(1)) \right\}$$

$$E_3 := \left\{ \exists w' \neq 1 : u^n(w', 1) \in \mathcal{C}_u(1) \text{ and } (u^n(w', 1), Y^n, S_2^n) \in \mathcal{T}_\epsilon^{(n)}(U, Y, S_2 | v_2^n(1)) \right\}.$$

The probability of error  $P_e^{(n)}$  is upper bounded by  $P_e^n \leq P(E_1) + P(E_2|E_1^c) + P(E_3|E_1^c, E_2^c)$ . Using standard arguments and assuming that  $(S_1^n, S_2^n) \in \mathcal{T}_\epsilon^{(n)}(S_1, S_2)$  and that  $n$  is large enough, we can state that

1. For each sequence  $v_2^n \in \mathcal{C}_v$ , the probability that  $v_2^n$  is not jointly typical with  $S_2^n$  is at most  $(1 - 2^{-n(I(V_2; S_2) + \epsilon)})$ . Therefore, having  $2^{n(I(V_2; S_2) + 2\epsilon)}$  i.i.d. sequences in  $\mathcal{C}_v$ , the probability that none of those sequences is jointly typical with  $S_2^n$  is bounded by

$$\begin{aligned} P(E_1) &\leq (1 - 2^{-n(I(V_2; S_2) + \epsilon)})^{2^{n(I(V_2; S_2) + 2\epsilon)}} \\ &\leq e^{-2^{n(I(V_2; S_2) + 2\epsilon)} 2^{-n(I(V_2; S_2) + \epsilon)}} \\ &= e^{-2^{n\epsilon}}, \end{aligned} \quad (\text{A32})$$

where, for every  $\epsilon > 0$ , the last line goes to zero as  $n$  goes to infinity.

2. The random variable  $Y^n$  is distributed according to  $p(y|x, s_1, s_2) = p(y|x, s_1, s_2, v_2)$ , therefore, having  $(S_2^n, v_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}(S_2, V_2)$  implies that  $(Y^n, S_2^n, v_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}(Y, S_2, V_2)$ . Recall that  $x_i = f(u_i(1,1), S_{1,i}, v_2(1))$  and that  $U^n$  is generated according to  $p(u|v_2)$ ; therefore,  $(X^n, S_1^n, u^n(1,1), v_2^n(1))$  is jointly typical. Thus, by the Markov lemma [31], we can state that  $(Y^n, S_2^n, u^n(1,1), v_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}(Y, S_2, U, V_2)$  with high probability for a large enough  $n$ .
3. Now, the probability for a random  $U^n$ , such that  $(U^n, v_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}(U, V_2)$ , to be also jointly typical with  $(Y^n, S_2^n, v_2^n(1))$  is upper bounded by  $2^{-n(I(U, Y, S_2 | V_2) - \epsilon)}$ , hence

$$\begin{aligned} P(E_3|E_1^c, E_2^c) &\leq \sum_{1 < w'}^{|\mathcal{C}_u(1)|} \Pr \left\{ (u^n(w', 1), Y^n, S_2^n) \in \mathcal{T}_\epsilon^{(n)}(U, Y, S_2 | v_2^n(1)) \right\} \\ &\leq \sum_{1 < w'}^{|\mathcal{C}_u(1)|} 2^{-n(I(U, Y, S_2 | V_2) - \epsilon)} \\ &\leq 2^{n(I(U, Y, S_2 | V_2) - 2\epsilon)} 2^{-n(I(U, Y, S_2 | V_2) - \epsilon)} \\ &= 2^{-n\epsilon}, \end{aligned} \quad (\text{A33})$$

which goes to zero exponentially fast with  $n$  for every  $\epsilon > 0$ .

Therefore,  $P_e^{(n)} = P(\hat{W} \neq W)$  goes to zero as  $n \rightarrow \infty$ .

**Converse:** (Channel capacity case  $2_c$ ). Fix the rates  $R$  and  $R'$  and a sequence of codes  $(n, 2^{nR}, 2^{nR'})$  that achieve capacity. By Fano's inequality,  $H(W|Y^n, S_2^n) \leq n\epsilon_n$ , where  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . Let  $T_2 = f_v(S_2^n)$ , and define  $V_{2,i} = (T_2, Y^{i-1}, S_2^{i-1})$ ,  $U_i = W$ . Then,

$$\begin{aligned} nR' &\geq H(T_2) \\ &\geq H(T_2) - H(T_2 | S_2^n) \\ &= I(T_2; S_2^n) \\ &= H(S_2^n) - H(S_2^n | T_2) \\ &= \sum_{i=1}^n \left[ H(S_{2,i} | S_2^{i-1}) - H(S_{2,i} | T_2, S_2^{i-1}) \right] \\ &\stackrel{(a)}{=} \sum_{i=1}^n \left[ H(S_{2,i}) - H(S_{2,i} | T_2, S_2^{i-1}, Y^{i-1}) \right] \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n I(S_{2,i}; T_2, Y^{i-1}, S_2^{i-1}) \\
&= \sum_{i=1}^n I(S_{2,i}; V_{2,i}),
\end{aligned} \tag{A34}$$

where (a) follows from the fact that  $S_{2,i}$  is independent of  $S_2^{i-1}$  and the fact that  $S_{2,i}$  is independent of  $Y^{i-1}$  given  $(T_2, S_2^{i-1})$ . The proof for this follows.

$$\begin{aligned}
p(y^{i-1}|t_2, s_2^{i-1}, s_{2,i}) &= \sum_{w, x^{i-1}, s_1^{i-1}} p(y^{i-1}, w, x^{i-1}, s_1^{i-1}|t_2, s_2^{i-1}, s_{2,i}) \\
&= \sum_{w, x^{i-1}, s_1^{i-1}} p(w)p(s_1^{i-1}|s_2^{i-1})p(x^{i-1}|w, t_2, s_1^{i-1})p(y^{i-1}|x^{i-1}, s_1^{i-1}, s_2^{i-1}) \\
&= p(y^{i-1}|t_2, s_2^{i-1}),
\end{aligned} \tag{A35}$$

where we used the fact that  $W$  is independent of  $(T_2, S_2^{i-1}, S_{2,i})$ ,  $S_1^{i-1}$  is independent of  $(T_2, S_{2,i})$  given  $S_2^{i-1}$ ,  $X^{i-1}$  is a function of  $(W, T_2, S_1^{i-1})$  and that  $Y^{i-1}$  is independent of  $(W, T_2, S_{2,i})$  given  $(X^{i-1}, S_1^{i-1}, S_2^{i-1})$ . We now continue with the proof of the converse.

$$\begin{aligned}
nR &\leq H(W) \\
&\leq H(W|T_2) - H(W|T_2, Y^n, S_2^n) + n\epsilon_n \\
&= I(W; Y^n, S_2^n|T_2) + n\epsilon_n \\
&= \sum_{i=1}^n I(W; Y_i, S_{2,i}|T_2, Y^{i-1}, S_2^{i-1}) + n\epsilon_n \\
&= \sum_{i=1}^n I(U_i; Y_i, S_{2,i}|V_{2,i}) + n\epsilon_n
\end{aligned} \tag{A36}$$

and therefore, from (A34) and (A36)

$$R' \geq \frac{1}{n} \sum_{i=1}^n I(S_{2,i}; V_{2,i}) \tag{A37}$$

$$R - \epsilon_n \leq \frac{1}{n} \sum_{i=1}^n I(U_i; Y_i, S_{2,i}|V_{2,i}). \tag{A38}$$

Using the convexity of  $R'$  and Jensen's inequality, the standard time-sharing argument for  $R$  and the fact that  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ , we can conclude that

$$R' \geq I(V_2; S_2), \tag{A39}$$

$$R \leq I(U; Y, S_2|V_2). \tag{A40}$$

Notice that the Markov chain  $V_{2,i} - S_{2,i} - S_{1,i}$  holds since  $(Y^{i-1}, S_2^{i-1})$  is independent of  $S_{1,i}$  and  $T_2(S_2^n)$  is dependent on  $S_{1,i}$  only through  $S_{2,i}$ . Notice also that the Markov chain  $U_i - V_{2,i} - (S_{1,i}, S_{2,i})$  holds since

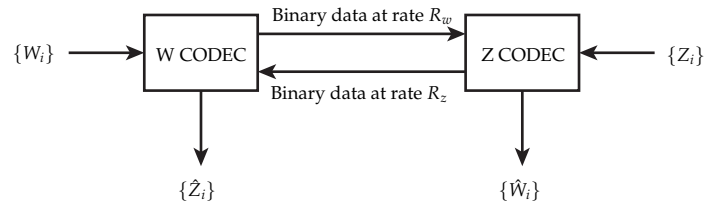
$$\begin{aligned}
p(w|t_2, y^{i-1}, s_2^{i-1}, s_{1,i}, s_{2,i}) &= \sum_{x^{i-1}, s_1^{i-1}} p(w, x^{i-1}, s_1^{i-1}|t_2, y^{i-1}, s_2^{i-1}, s_{1,i}, s_{2,i}) \\
&= \sum_{x^{i-1}, s_1^{i-1}} p(s_1^{i-1}|t_2, y^{i-1}, s_2^{i-1})p(x^{i-1}|t_2, y^{i-1}, s_1^{i-1}, s_2^{i-1})p(w|t_2, x^{i-1}, s_1^{i-1}) \\
&= p(w|t_2, y^{i-1}, s_2^{i-1}).
\end{aligned} \tag{A41}$$



This concludes the converse, and the proof of Theorem 1 Case 2<sub>C</sub>.

### Appendix C. Proof of Theorem 2

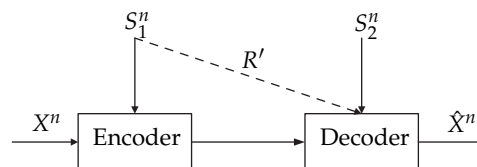
In this section, we provide the proof of Theorem 2, Cases 1 and 1<sub>C</sub>. Case 2, where the encoder is informed with increased ESI and the decoder is informed with DSI is a special case of [10] for  $K = 1$  and, therefore, the proof for this case is omitted. Following Kaspi's scheme (Figure A3) for  $K = 1$ , at the first stage, node  $W$  sends a description of  $W$  with a rate limited to  $R_w$ , then, after reconstructing  $\hat{W}$  at the  $Z$  node, it sends a function of  $Z$  and  $\hat{W}$  over to node  $W$  with a rate limited to  $R_z$ . Let  $S_2$  be  $W$  in Kaspi's scheme and  $(X, S_1)$  be  $Z$  in Kaspi's scheme. Consider  $D_z = d(Z_i, \hat{Z}_i) = d((X, S_{1,i}), (\hat{X}_i, \hat{S}_{1,i})) = d(X_i, \hat{X}_i) = D$ . Then, it is apparent that Case 2 of the rate-distortion problems is a special case of Kaspi's two-way problem for  $K = 1$ .



**Figure A3.** Kaspi's two-way source coding scheme. The total rates are  $R_w = \sum_{k=1}^K R_w^k$  and  $R_z = \sum_{k=1}^K R_z^k$  and the expected per-letter distortions are  $D_w = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n d(W_i, \hat{W}_i)\right]$  and  $D_z = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n d(Z_i, \hat{Z}_i)\right]$ .

#### Appendix C.1. Proof of Theorem 2, Case 1

Rate-distortion Case 1 is presented in Figure A4. We use the Wyner-Ziv coding scheme for the description of the ESI,  $S_1$ , at a rate  $R'$ , where the source is  $S_1$  and the side information at the decoder is  $S_2$ . Then, to describe the main source,  $X$ , with distortion less than or equal to  $D$  we use the Wyner-Ziv coding scheme again, where this time,  $S_2$  is the side information at the decoder,  $S_1$  is a part of the source and the rate-limited description of  $S_1$  is the side information at both the encoder and the decoder. Notice that  $I(U; X, S_1 | V_1) - I(U; S_2 | V_1) = I(U; X, S_1, V_1) - I(U; S_1, V_1)$  and that since the Markov chain  $V_1 - S_1 - S_2$  holds, it is also possible to write  $R' \geq I(V_1; S_1) - I(V_1; S_2)$ ; we use these expressions in the following proof.



**Figure A4.** Rate-distortion: Case 1.  $R_1(D) = \min I(U; X, S_1 | V_1) - I(U; S_2 | V_1)$ , where the minimization is over all PMFs  $p(v_1 | s_1)p(u | x, s_1, v_1)p(\hat{x} | u, s_2, v_1)$  such that  $R' \geq I(V_1; S_1 | S_2)$  and  $\mathbb{E}[d(X, \hat{X})] \leq D$ .

**Achievability:** (Rate-distortion Case 1). Given  $(X_i, S_{1,i}, S_{2,i})$  i.i.d.  $\sim p(x, s_1, s_2)$  and the distortion measure  $D$ , fix  $p(x, s_1, s_2, v_1, u, \hat{x}) = p(x, s_1, s_2)p(v_1 | s_1)p(u | x, s_1, v_1)p(\hat{x} | u, s_2, v_1)$  that satisfies  $\mathbb{E}[d(X, \hat{X})] = D$  and  $\hat{x} = f(u, s_2, v_1)$ .

*Codebook generation and random binning*

1. Generate a codebook,  $\mathcal{C}_v$ , of  $2^{n(I(V_1; S_1) + 2\epsilon)}$  sequences,  $V_1^n$ , independently using i.i.d.  $\sim p(v_1)$ . Label them  $v_1^n(k)$ , where  $k \in \{1, 2, \dots, 2^{n(I(V_1; S_1) + 2\epsilon)}\}$  and randomly assign each sequence  $v_1^n(k)$  a bin number  $b_v(v_1^n(k))$  in the set  $\{1, 2, \dots, 2^{nR'}\}$ .

2. Generate a codebook  $\mathcal{C}_u$  of  $2^{n(I(U;X,S_1,V_1)+2\epsilon)}$  sequences  $U^n$  independently using i.i.d.  $\sim p(u)$ . Label them  $u^n(l)$ , where  $l \in \{1, 2, \dots, 2^{n(I(U;X,S_1,V_1)+2\epsilon)}\}$ , and randomly and assign each  $u^n(l)$  a bin number  $b_u(u^n(l))$  in the set  $\{1, 2, \dots, 2^{nR}\}$ .

Reveal the codebooks and the content of the bins to all encoders and decoders.

#### Encoding

1. *State Encoder*: Given the sequence  $S_1^n$ , search the codebook  $\mathcal{C}_v$  and identify an index  $k$  such that  $(S_1^n, v_1^n(k)) \in \mathcal{T}_\epsilon^{(n)}(S, V_1)$ . If such a  $k$  is found, stop searching and send the bin number  $j = b_v(v_1^n(k))$ . If no such  $k$  is found, declare an error.
2. *Encoder*: Given the sequences  $X^n, S_1^n$  and  $v_1^n(k)$ , search the codebook  $\mathcal{C}_u$  and identify an index  $l$  such that  $(X^n, S_1^n, v_1^n(k), u^n(l)) \in \mathcal{T}_\epsilon^{(n)}(X, S_1, V_1, U)$ . If such an  $l$  is found, stop searching and send the bin number  $w = b_u(u^n(l))$ . If no such  $l$  is found, declare an error.

#### Decoding

Given the bins indices  $w$  and  $j$  and the sequence  $S_2^n$ , search the codebook  $\mathcal{C}_v$  and identify an index  $k$  such that  $(S_2^n, v_1^n(k)) \in \mathcal{T}_\epsilon^{(n)}(S_2, V_1)$  and  $b_v(v_1^n(k)) = j$ . If no such  $k$  is found or there is more than one such index, declare an error. If a unique  $k$ , as defined, is found, search the codebook  $\mathcal{C}_u$  and identify an index  $l$  such that  $(S_2^n, v_1^n(k), u^n(l)) \in \mathcal{T}_\epsilon^{(n)}(S_2, V_1, U)$  and  $b_u(u^n(l)) = w$ . If a unique  $l$ , as defined, is found, declare  $\hat{X}_i = f_i(u_i^n(l), S_{2,i}, v_{1,i}(k))$ ,  $i = 1, 2, \dots, n$ . Otherwise, if there is no such  $l$  or there is more than one, declare an error.

#### Analysis of the probability of error

Without loss of generality, for the following events  $E_2, E_3, E_4, E_5$  and  $E_6$ , assume that  $v_1^n(k = 1)$  and  $b_v(v_1^n(k = 1)) = 1$  correspond to the sequences  $(X^n, S_1^n, S_2^n)$  and for the events  $E_5$  and  $E_6$  assume that  $u^n(l = 1)$  and  $b_u(u^n(l = 1)) = 1$  correspond to the same given sequences. Define the following events:

$$\begin{aligned}
 E_1 &:= \left\{ \forall v_1^n(k) \in \mathcal{C}_v, (S_1^n, v_1^n(k)) \notin \mathcal{T}_\epsilon^{(n)}(S_1, V_1) \right\} \\
 E_2 &:= \left\{ (S_1^n, v_1^n(1)) \in \mathcal{T}_\epsilon^{(n)}(S_1, V_1) \text{ but } (S_2^n, v_1^n(1)) \notin \mathcal{T}_\epsilon^{(n)}(S_2, V_1) \right\} \\
 E_3 &:= \left\{ \exists k' \neq 1 \text{ such that } b_v(v_1^n(k')) = 1 \text{ and } (S_2^n, v_1^n(k')) \in \mathcal{T}_\epsilon^{(n)}(S_2, V_1) \right\} \\
 E_4 &:= \left\{ \forall u^n(l) \in \mathcal{C}_u, (X^n, S_1^n, v_1^n(1), u^n(l)) \notin \mathcal{T}_\epsilon^{(n)}(X, S_1, V_1, U) \right\} \\
 E_5 &:= \left\{ (X^n, S_1^n, v_1^n(1), u^n(1)) \in \mathcal{T}_\epsilon^{(n)}(X, S_1, V_1, U) \text{ but } (S_2^n, v_1^n(1), u^n(1)) \notin \mathcal{T}_\epsilon^{(n)}(S_2, V_1, U) \right\} \\
 E_6 &:= \left\{ \exists l' \neq 1 \text{ such that } b_u(u^n(l')) = 1 \text{ and } (S_2^n, v_1^n(1), u^n(l')) \in \mathcal{T}_\epsilon^{(n)}(S_2, V_1, U) \right\}.
 \end{aligned}$$

The probability of error  $P_e^{(n)}$  is upper bounded by  $P_e^{(n)} \leq P(E_1) + P(E_2|E_1^c) + P(E_3|E_1^c, E_2^c) + P(E_4|E_1^c, E_2^c, E_3^c) + P(E_5|E_1^c, \dots, E_4^c) + P(E_6|E_1^c, \dots, E_5^c)$ . Using standard arguments and assuming that  $(X^n, S_1^n, S_2^n) \in \mathcal{T}_\epsilon^{(n)}(X, S_1, S_2)$  and that  $n$  is large enough, we can state that

1.

$$\begin{aligned}
 P(E_1) &= \Pr \left\{ \bigcap_{v_1^n(k) \in \mathcal{C}_v} (S_1^n, v_1^n(k)) \notin \mathcal{T}_\epsilon^{(n)}(S_1, V_1) \right\} \\
 &\leq \prod_{k=1}^{2^{n(I(V_1;S_1)+\epsilon)}} \Pr \{ (S_1^n, v_1^n(k)) \notin \mathcal{T}_\epsilon^{(n)}(S_1, V_1) \}
 \end{aligned}$$

$$\begin{aligned}
&\leq e^{-2^n(I(V_1;S_1)+2\epsilon)} 2^{-nI(S_1;V_1)-n\epsilon} \\
&= e^{-2^{n\epsilon}}.
\end{aligned} \tag{A42}$$

The probability that there is no  $v_1^n(k)$  in  $\mathcal{C}_v$  such that  $(S_1^n, v_1^n(k))$  is strongly jointly typical is exponentially small provided that  $|\mathcal{C}_v| > 2^{n(I(S_1;V_1)+\epsilon)}$ . This follows from the standard rate-distortion argument that  $2^{nI(S_1;V_1)}$   $v_1^n(k)$ s “cover”  $S_1^n$ , therefore  $P(E_1) \rightarrow 0$ .

2. By the Markov lemma, since  $(S_1^n, S_2^n)$  are strongly jointly typical and  $(S_1^n, v_1^n(1))$  are strongly jointly typical and the Markov chain  $V_1 - S_1 - S_2$  holds, then  $(S_1^n, S_2^n, v_1^n(1))$  are also strongly jointly typical. Thus,  $P(E_2|E_1^c) \rightarrow 0$ .
- 3.

$$\begin{aligned}
P(E_3) &= \Pr \left\{ \bigcup_{\substack{v_1^n(k' \neq 1) \\ b_v(v_1(k'))=1}} (S_2^n, v_1^n(k')) \in \mathcal{T}_\epsilon^{(n)}(S_1, V_1) \right\} \\
&\leq \sum_{\substack{v_1^n(k' \neq 1) \\ b_v(v_1(k'))=1}} \Pr \left\{ (S_1^n, v_1^n(k')) \in \mathcal{T}_\epsilon^{(n)}(S_1, V_1) \right\} \\
&\leq 2^{n(I(V_1;S_1)+2\epsilon-R')} 2^{-n(I(S_2;V_1)-\epsilon)}.
\end{aligned} \tag{A43}$$

The probability that there is another index  $k', k' \neq 1$ , such that  $v_1^n(k')$  is in bin number 1 and that it is strongly jointly typical with  $S_2^n$  is bounded by the number of  $v_1^n(k')$ 's in the bin times the probability of joint typicality. Therefore, if  $R' > I(V_1;S_1) - I(V_1;S_2) + 3\epsilon$  then  $P(E_3|E_1^c, E_2^c) \rightarrow 0$ . Furthermore, using the Markov chain  $V_1 - S_1 - S_2$ , we can see that the inequality can be presented as  $R' > I(V_1;S_1|S_2) + 3\epsilon$ .

4. We use here the same argument we used for  $P(E_1)$ . By the covering lemma we can state that the probability that there is no  $u^n(l)$  in  $\mathcal{C}_u$  that is strongly jointly typical with  $(X^n, S_1^n, v_1^n(k))$  tends to 0 as  $n \rightarrow \infty$  if  $R'_u > I(U;X,S_1,V_1) + \epsilon$ . Hence,  $P(E_4|E_1^c, E_2^c, E_3^c) \rightarrow 0$ .
5. Using the same argument we used for  $P(E_2|E_1^c)$ , we conclude that  $P(E_4|E_1^c, E_2^c, E_3^c) \rightarrow 0$ .
6. We use here the same argument we used for  $P(E_2|E_1^c)$ . Since  $(U, X, S_1, V_1)$  are strongly jointly typical,  $(X, S_1, S_2)$  are strongly jointly typical and the Markov chain  $(U, V_1) - (X, S_1) - S_2$  holds, then  $(U, X, S_1, S_2, V_1)$  are also strongly jointly typical.
7. The probability that there is another index  $l', l' \neq 1$  such that  $u^n(l')$  is in bin number 1 and that it is strongly jointly typical with  $(S_2^n, v_1^n(1))$  is exponentially small provided that  $R \geq I(U;X,S_1,V_1) - I(U;S_2,V_1) + 3\epsilon = I(U;X,S_1|V_1) - I(U;S_2|V_1) + 3\epsilon$ . Notice that  $2^{n(I(U;X,S_1,V_1)-R)}$  stands for the average number of sequences  $u^n(l)$ 's in each bin indexed  $w$  for  $w \in \{1, 2, \dots, 2^{nR}\}$ .

This shows that for rates  $R$  and  $R'$  as described, and for large enough  $n$ , the error events are of arbitrarily small probability. This concludes the proof of the achievability for the source coding Case 1.

**Converse:** (Rate-distortion Case 1). Fix a distortion measure  $D$ , the rates  $R', R \geq R(D) = \min I(U;X,S_1|V_1) - I(U;S_2|V_1) = \min I(U;X,S_1|S_2,V_1)$  and a sequence of codes  $(n, 2^{nR}, 2^{nR'})$  such that  $\mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n d(X_i, \hat{X}_i) \right] = D$ . Let  $T_1 = f_v(S_1^n)$ ,  $T = f(X^n, S_1^n, T)$  and define  $V_{1,i} = (T_1, S_{1,i+1}^n, S_2^{i-1}, S_{2,i+1}^n)$  and  $U_i = T$ . Notice that  $\hat{X}_i = \hat{X}_i(T, T_1, S_2^n)$  and, therefore,  $\hat{X}_i$  is a function of  $(U_i, V_{1,i}, S_{2,i})$ .

$$\begin{aligned}
nR' &\geq H(T_1) \\
&\geq H(T_1|S_2^n) - H(T_1|S_1^n, S_2^n)
\end{aligned}$$

$$\begin{aligned}
&= I(T_1; S_1^n | S_2^n) \\
&= H(S_1^n | S_2^n) - H(S_1^n | T_1, S_2^n) \\
&= \sum_{i=1}^n \left[ H(S_{1,i} | S_{1,i+1}^n, S_2^n) - H(S_{1,i} | T_1, S_{1,i+1}^n, S_2^n) \right] \\
&\stackrel{(a)}{=} \sum_{i=1}^n \left[ H(S_{1,i} | S_{2,i}) - H(S_{1,i} | T_1, S_{1,i+1}^n, S_2^{i-1}, S_{2,i+1}^n, S_{2,i}) \right] \\
&= \sum_{i=1}^n \left[ H(S_{1,i} | S_{2,i}) - H(S_{1,i} | V_{1,i}, S_{2,i}) \right] \\
&= \sum_{i=1}^n I(S_{1,i}; V_{1,i} | S_{2,i}), \tag{A44}
\end{aligned}$$

where (a) follows from the fact that  $S_{1,i}$  is independent of  $(S_{1,i+1}^n, S_2^{i-1}, S_{2,i+1}^n)$  given  $S_{2,i}$ .

$$\begin{aligned}
nR &\geq H(T) \\
&\geq H(T | T_1, S_2^n) - H(T | T_1, X^n, S_1^n, S_2^n) \\
&= I(T; X^n, S_1^n | T_1, S_2^n) \\
&= H(X^n, S_1^n | T_1, S_2^n) - H(X^n, S_1^n | T, T_1, S_2^n) \\
&= \sum_{i=1}^n \left[ H(X_i, S_{1,i} | T_1, S_2^n, X_{i+1}^n, S_{1,i+1}^n) - H(X_i, S_{1,i} | T, T_1, S_2^n, X_{i+1}^n, S_{1,i+1}^n) \right] \\
&\stackrel{(b)}{=} \sum_{i=1}^n \left[ H(X_i, S_{1,i} | T_1, S_{1,i+1}^n, S_2^n) - H(X_i, S_{1,i} | T, T_1, S_2^n, X_{i+1}^n, S_{1,i+1}^n) \right] \\
&\stackrel{(c)}{\geq} \sum_{i=1}^n \left[ H(X_i, S_{1,i} | T_1, S_{1,i+1}^n, S_2^n) - H(X_i, S_{1,i} | T, T_1, S_{1,i+1}^n, S_2^n) \right] \\
&= \sum_{i=1}^n I(X_i, S_{1,i}; T | T_1, S_{1,i+1}^n, S_2^n) \\
&= \sum_{i=1}^n I(X_i, S_{1,i}; U_i | V_{1,i}, S_{2,i}) \\
&= \sum_{i=1}^n R\left(\mathbb{E}\left[d(X_i, \hat{X}_i)\right]\right) \\
&\stackrel{(d)}{\geq} nR\left(\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n d(X_i, \hat{X}_i)\right]\right) \\
&= nR(D), \tag{A45}
\end{aligned}$$

where (b) follows from the fact that  $(X_i, S_{1,i})$  is independent of  $X_{i+1}^n$  given  $(T_1, S_{1,i+1}^n, S_2^n)$ ; this is because  $X_{i+1}^n$  is independent of  $(T_1, X^i, S_1^i)$  given  $(S_{1,i+1}^n, S_{2,i+1}^n)$ , (c) follows from the fact that conditioning reduces entropy and (d) follows from the convexity of  $R(D)$  and Jensen's inequality.

Using also the convexity of  $R'$  and Jensen's inequality, we can conclude that

$$R' \geq I(V_1; S_1 | S_2), \tag{A46}$$

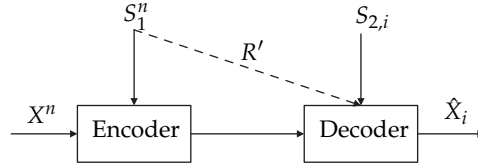
$$R \geq I(U; X, S_1 | V_1, S_2). \tag{A47}$$

It is easy to verify that  $(T_1, S_{1,i+1}^n, S_2^{i-1}, S_{2,i+1}^n) - S_{1,i} - S_{2,i}$  forms a Markov chain, since  $T_1(S_1^n)$  depends on  $S_{2,i}$  only through  $S_{1,i}$ . The structure  $T - (T_1, S_{1,i+1}^n, S_2^{i-1}, S_{2,i+1}^n, X_i, S_{1,i}) - S_{2,i}$  also forms a Markov chain since  $S_{2,i}$  contains no information about  $(S_1^{i-1}, X^{i-1}, X_{i+1}^n)$  given  $(T_1, S_{1,i}^n, S_2^{i-1}, S_{2,i+1}^n, X_i)$  and, therefore, contains no information about  $T(X^n, S_1^n, T_1)$ .

This concludes the converse, and the proof of Theorem 2 Case 1.

#### Appendix C.2. Proof of Theorem 2, Case 1<sub>C</sub>

Rate-distortion Case 1<sub>C</sub> is illustrated in Figure A5. For describing the ESI,  $S_1$ , with a rate  $R'$  we use the standard rate-distortion coding scheme. Then, for the main source,  $X$ , we use a Weissman-El Gamal [12] coding scheme where the DSI,  $S_2$ , is the causal side information at the decoder,  $S_1$  is a part of the source and the rate-limited description of  $S_1$  is the side information at both the encoder and decoder.



**Figure A5.** Rate-distortion: Case 1 with causal DSI.  $R_{1C}(D) = \min I(U; X, S_1 | V_1)$ , where the minimization is over all PMFs  $p(v_1 | s_1)p(u | x, s_1, v_1)p(\hat{x} | u, s_2, v_1)$  such that  $R' \geq I(V_1; S_1)$  and  $\mathbb{E}[d(X, \hat{X})] \leq D$ .

**Achievability:** (Rate-distortion Case 1<sub>C</sub>). Given  $(X_i, S_{1,i}, S_{2,i}) \sim \text{i.i.d. } p(x, s_1, s_2)$  where the DSI is known in a causal way ( $S_2^i$  in time  $i$ ) and the distortion measure is  $D$ , fix  $p(x, s_1, s_2, v_1, u, \hat{x}) = p(x, s_1, s_2)p(v_1 | s_1)p(u | x, s_1, v_1)p(\hat{x} | u, s_2, v_1)$  that satisfies  $\mathbb{E}[d(X, \hat{X})] = D$  and that  $\hat{x} = f(u, s_2, v_1)$ .

#### Codebook generation and random binning

1. Generate a codebook  $\mathcal{C}_v$  of  $2^{n(I(V_1; S_1) + 2\epsilon)}$  sequences  $V_1^n$  independently using i.i.d.  $\sim p(v_2)$ . Label them  $v_1^n(k)$  where  $k \in \{1, 2, \dots, 2^{n(I(V_1; S_1) + 2\epsilon)}\}$ .
2. For each  $v_1^n(k)$  generate a codebook  $\mathcal{C}_u(k)$  of  $2^{n(I(U; X, S_1 | V_1) + 2\epsilon)}$  sequences  $U^n$  distributed independently according to i.i.d.  $\sim p(u | v_1)$ . Label them  $u^n(w, k)$ , where  $w \in \{1, 2, \dots, 2^{n(I(U; X, S_1 | V_1) + 2\epsilon)}\}$ .

Reveal the codebooks to all encoders and decoders.

#### Encoding

1. *State Encoder:* Given the sequence  $S_1^n$ , search the codebook  $\mathcal{C}_v$  and identify an index  $k$  such that  $(v_1^n(k), S_1^n) \in \mathcal{T}_\epsilon^{(n)}(V_1, S_1)$ . If such a  $k$  is found, stop searching and send it. Otherwise, if no such  $k$  is found, declare an error.
2. *Encoder:* Given  $X^n, S_1^n$  and the index  $k$ , search the codebook  $\mathcal{C}_u(k)$  and identify an index  $w$  such that  $(u^n(w, k), X^n, S_1^n) \in \mathcal{T}_\epsilon^{(n)}(U, X, S_1 | v_1^n(k))$ . If such an index  $w$  is found, stop searching and send it. Otherwise, declare an error.

#### Decoding

Given the indices  $w, k$  and the sequence  $S_1^i$  at time  $i$ , declare  $\hat{x}_i = f(u_i(w, k), S_{2,i}, v_{1,i}(k))$ .

#### Analysis of the probability of error

Without loss of generality, let us assume that  $v_1^n(1)$  corresponds to  $S_1^n$  and that  $u^n(1, 1)$  corresponds to  $(X^n, S_1^n, v_1^n(1))$ .

Define the following events:

$$E_1 := \left\{ \forall v_1^n(k) \in \mathcal{C}_v, (v_1^n(k), S_1^n) \notin \mathcal{T}_\epsilon^{(n)}(S_1, V_1) \right\}$$

$$E_2 := \left\{ \forall u^n(w, 1) \in \mathcal{C}_u(1), (X^n, S_1^n, u^n(w, 1)) \notin \mathcal{T}_\epsilon^{(n)}(X, S_1, U) \right\}$$

The probability of error  $P_e^{(n)}$  is upper bounded by  $P_e^n \leq P(E_1) + P(E_2|E_1^c)$ . Assuming that  $(S_1^n, S_2^n) \in \mathcal{T}_\epsilon^{(n)}(S_1, S_2)$ , we can state that by the standard rate-distortion argument, having more than  $2^{n(I(V_1; S_1) + \epsilon)}$  sequences  $v_1^n(k)$  in  $\mathcal{C}_v$  and a large enough  $n$  assures us with probability arbitrarily close to 1 that we would find an index  $k$  such that  $(v_1^n(k), S_1^n) \in \mathcal{T}_\epsilon^{(n)}(V_1, S_1)$ . Therefore,  $P(E_1) \rightarrow 0$  as  $n \rightarrow \infty$ . Now, if  $(v_1^n(1), S_1^n) \in \mathcal{T}_\epsilon^{(n)}(V_1, S_1)$ , using the same argument, we can also state that having more than  $2^{n(I(U; X, S_1|V_1) + \epsilon)}$  sequences  $u^n(w, 1)$  in  $\mathcal{C}_u(1)$  assures us that  $P(E_2|E_1^c) \rightarrow 0$  as  $n \rightarrow \infty$ . This concludes the proof of the achievability.

**Converse:** (Rate-distortion Case 1C). Fix a distortion measure  $D$ , the rates  $R', R \geq R(D) = \min I(U; X, S_1|V_1)$  and a sequence of codes  $(n, 2^{nR}, 2^{nR'})$  such that  $\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n d(X_i, \hat{X}_i)\right] = D$ . Let  $T_1 = f_v(S_1^n)$ ,  $T = f(X^n, S_1^n, T_1)$  and define  $V_{1,i} = (T_1, S_{1,i+1}^n)$ ,  $U_i = T$ . Notice that  $\hat{X}_i = \hat{X}_i(T, T_1, S_2^i)$ , and, therefore,  $\hat{X}_i$  is a function of  $(U_i, V_{1,i}, S_2^i)$ .

$$\begin{aligned}
 nR' &\geq H(T_1) \\
 &\geq H(V) - H(T_1|S_1^n) \\
 &= I(T_1; S_1^n) \\
 &= H(S_1^n) - H(S_1^n|T_1) \\
 &= \sum_{i=1}^n \left[ H(S_{1,i}|S_{1,i+1}^n) - H(S_{1,i}|T_1, S_{1,i+1}^n) \right] \\
 &\stackrel{(a)}{=} \sum_{i=1}^n \left[ H(S_{1,i}) - H(S_{1,i}|T_1, S_{1,i+1}^n) \right] \\
 &= \sum_{i=1}^n \left[ H(S_{1,i}) - H(S_{1,i}|V_{1,i}) \right] \\
 &= \sum_{i=1}^n I(S_{1,i}; V_{1,i}), \tag{A48}
 \end{aligned}$$

where (a) follows the fact that  $S_{1,i}$  is independent of  $S_{1,i+1}^n$ .

$$\begin{aligned}
 nR &\geq H(T) \\
 &\geq H(T|T_1) - H(T|T_1, X^n, S_1^n) \\
 &= I(T; X^n, S_1^n|T_1) \\
 &= H(X^n, S_1^n|T_1) - H(X^n, S_1^n|T, T_1) \\
 &= \sum_{i=1}^n \left[ H(X_i, S_{1,i}|T_1, X_{i+1}^n, S_{1,i+1}^n) - H(X_i, S_{1,i}|T, T_1, X_{i+1}^n, S_{1,i+1}^n) \right] \\
 &\stackrel{(b)}{=} \sum_{i=1}^n \left[ H(X_i, S_{1,i}|T_1, S_{1,i+1}^n) - H(X_i, S_{1,i}|T, T_1, X_{i+1}^n, S_{1,i+1}^n) \right] \\
 &\stackrel{(c)}{\geq} \sum_{i=1}^n \left[ H(X_i, S_{1,i}|T_1, S_{1,i+1}^n) - H(X_i, S_{1,i}|T, T_1, S_{1,i+1}^n) \right] \\
 &= \sum_{i=1}^n I(X_i, S_{1,i}; T|T_1, S_{1,i+1}^n) \\
 &= \sum_{i=1}^n I(X_i, S_{1,i}; U_i|V_{1,i}) \\
 &= \sum_{i=1}^n R\left(\mathbb{E}\left[d(X_i, \hat{X}_i)\right]\right) \\
 &\stackrel{(d)}{\geq} nR\left(\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n d(X_i, \hat{X}_i)\right]\right)
 \end{aligned}$$

$$=nR(D) \quad (\text{A49})$$

where (b) follows from the fact that  $(X_i, S_{1,i})$  is independent of  $X_{i+1}^n$  given  $(T_1, S_{1,i+1}^n)$ , (c) follows from the fact that conditioning reduces entropy and (d) follows from the convexity of  $R(D)$  and Jensen's inequality.

Using also the convexity of  $R'$  and Jensen's inequality, we can conclude that

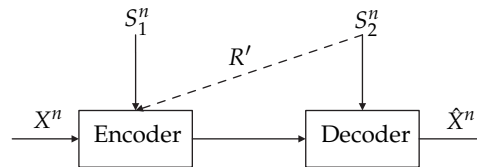
$$R' \geq I(V_1; S_1), \quad (\text{A50})$$

$$R \geq I(U; X, S_1 | V_1). \quad (\text{A51})$$

It is easy to verify that both Markov chains  $V_{1,i} - S_{1,i} - (X_i, S_{2,i})$  and  $U_i - (X_i, S_{1,i}, V_{1,i}) - S_{2,i}$  hold. This concludes the converse, and the proof of Theorem 2 Case 1<sub>C</sub>.

### Appendix C.3. Proof of Theorem 2, Case 2

Rate-distortion Case 2 (see Figure A6) is a special case of [10] for  $K = 1$ , and hence, the proof is omitted.



**Figure A6.** Rate distortion: Case 2.  $R_2(D) = \min I(U; X, S_1 | V_2) - I(U; S_2 | V_2)$ , where the minimization is over all PMFs  $p(v_2 | s_2)p(u | x, s_1, v_2)p(\hat{x} | u, s_2, v_2)$  such that  $R' \geq I(V_2; S_2) - I(V_2; X, S_1)$  and  $\mathbb{E}[d(X, \hat{X})] \leq D$ .

### Appendix D. Proof of Lemma 1

We provide here a partial proof of Lemma 1. In the first part we prove the concavity of  $C_2^{lb}(R')$  in  $R'$  for Case 2, the second part contains the proof that it is enough to take  $X$  to be a deterministic function of  $(S_1, V_1, U)$  in order to achieve the capacity  $C_1(R')$  for Case 1 and in the third part we prove the cardinality bound for Case 1. The proofs of these three parts for the rest of the cases can be derived using the same techniques and therefore are omitted. The proof of Lemma 2 can also be readily concluded using the techniques we use in this appendix and is omitted as well.

*Part 1:* We prove here that for Case 2 of the channel capacity problems, the lower bound on the capacity,  $C_2^{lb}(R')$ , is a concave function of the state information rate,  $R'$ . Recall that the expression for  $C_2^{lb}$  is  $C_2^{lb}(R') = \max I(U; Y, S_2 | V_2) - I(U; S_1 | V_2)$  where the maximization is over all probabilities  $p(s_1, s_2)p(v_2 | s_2)p(u | s_1, v_2)p(x | u, s_1, v_2)p(y | x, s_1, s_2)$  such that  $R' \geq I(V_2; S_2 | S_1)$ . This means that we want to prove that for any two rates,  $R'^{(1)}$  and  $R'^{(2)}$ , and for any  $0 \leq \alpha \leq 1$  and  $\bar{\alpha} = 1 - \alpha$  the capacity maintains  $C_2^{lb}(\alpha R'^{(1)} + \bar{\alpha} R'^{(2)}) \geq \alpha C_2^{lb}(R'^{(1)}) + \bar{\alpha} C_2^{lb}(R'^{(2)})$ . Let  $(U^{(1)}, V_2^{(1)}, X^{(1)}, Y^{(1)})$  and  $(U^{(2)}, V_2^{(2)}, X^{(2)}, Y^{(2)})$  be the random variables that meet the conditions on  $R'^{(1)}$  and on  $R'^{(2)}$  and also achieve  $C_2^{lb}(R'^{(1)})$  and  $C_2^{lb}(R'^{(2)})$ , respectively. Let us introduce the auxiliary random variable  $Q \in \{1, 2\}$ , independent of  $S_1, S_2, V_2, U, X$  and  $Y$ , and distributed according to  $\Pr\{Q = 1\} = \alpha$  and  $\Pr\{Q = 2\} = \bar{\alpha}$ . Then, consider

$$\begin{aligned} \alpha R'^{(1)} + \bar{\alpha} R'^{(2)} &= \alpha [I(V_2^{(1)}; S_2) - I(V_2^{(1)}; S_1)] + \bar{\alpha} [I(V_2^{(2)}; S_2) - I(V_2^{(2)}; S_1)] \\ &\stackrel{(a)}{=} \alpha [I(V_2^{(1)}; S_2 | Q = 1) - I(V_2^{(1)}; S_1 | Q = 1)] + \bar{\alpha} [I(V_2^{(2)}; S_2 | Q = 2) - I(V_2^{(2)}; S_1 | Q = 2)] \\ &\stackrel{(b)}{=} I(V_2^{(Q)}; S_2 | Q) - I(V_2^{(Q)}; S_1 | Q) \\ &\stackrel{(c)}{=} I(V_2^{(Q)}, Q; S_2) - I(V_2^{(Q)}, Q; S_1), \end{aligned} \quad (\text{A52})$$



and

$$\begin{aligned} \alpha C_2^{lb}(R'^{(1)}) + \bar{\alpha} C_2^{lb}(R'^{(2)}) &= \alpha [I(U^{(1)}; Y^{(1)}, S_2 | V_2^{(1)}) - I(U^{(1)}; S_1 | V_2^{(1)})] \\ &\quad + \bar{\alpha} [I(U^{(2)}; Y^{(2)}, S_2 | V_2^{(2)}) - I(U^{(2)}; S_1 | V_2^{(2)})] \\ &\stackrel{(d)}{=} I(U^{(Q)}; Y^{(Q)}, S_2 | V_2^{(Q)}, Q) - I(U^{(Q)}; S_1 | V_2^{(Q)}, Q), \end{aligned} \quad (A53)$$

where (a), (b), (c) and (d) all follow from the fact that  $Q$  is independent of  $(S_1, S_2, V_2, U, X, Y)$  and from  $Q$ 's probability distribution. Now, let  $V'_2 = (V_2^{(Q)}, Q)$ ,  $U' = U^{(Q)}$ ,  $Y' = Y^{(Q)}$  and  $X' = X^{(Q)}$ . Then, following from the equalities above, for any two rates  $R'^{(1)}$  and  $R'^{(2)}$  and for any  $0 \leq \alpha \leq 1$ , there exists a set of random variables  $(U', V'_2, X', Y')$  that maintains

$$\alpha R'^{(1)} + \bar{\alpha} R'^{(2)} = I(V'_2; S_2) - I(V'_2; S_1), \quad (A54)$$

and

$$\begin{aligned} C_2^{lb}(\alpha R'^{(1)} + \bar{\alpha} R'^{(2)}) &\geq I(U'; Y', S_2 | V'_2) - I(U'; S_1 | V'_2) \\ &= \alpha C_2^{lb}(R'^{(1)}) + \bar{\alpha} C_2^{lb}(R'^{(2)}). \end{aligned} \quad (A55)$$

This completes the proof of the concavity of  $C_2^{lb}(R')$  in  $R'$ .

*Part 2:* We prove here that it is enough to take  $X$  to be a deterministic function of  $(U, S_1, V_1)$  in order to maximize  $I(U; Y, S_2, V_1) - I(U; S_1, V_1)$ . Fix  $p(u, v_1 | s_1)$ . Note that

$$\begin{aligned} p(y, s_2 | u, v_1) &= \sum_{x, s_1} p(s_1 | u, v_1) p(s_2 | s_1, v_1, u) p(x | s_1, s_2, v_1, u) p(y | x, s_1, s_2, v_1, u) \\ &= \sum_{x, s_1} p(s_1 | u, v_1) p(s_2 | s_1) p(x | s_1, v_1, u) p(y | x, s_1, s_2) \end{aligned} \quad (A56)$$

is linear in  $p(x | u, v_1, s_1)$ . This follows from the fact that fixing  $p(u, v_1 | s_1)$  also defines  $p(s_1 | u, v_1)$  and from the following Markov chains  $S_2 - S_1 - (V_1, U)$ ,  $X - (S_1, V_1, U) - S_2$  and  $Y - (X, S_1, S_2) - (V_1, U)$ . Hence, since  $I(U; Y, S_2 | V_1)$  is convex in  $p(y, s_2 | v_1)$  it is also convex in  $p(x | u, v_1, s_1)$ . Noting also that  $I(U; S_1 | V_1)$  is constant given a fixed  $p(u, v_1 | s_1)$ , we can conclude that  $I(U; Y, S_2 | V_1) - I(U; S_1 | V_1)$  is convex in  $p(x | u, v_1, s_1)$  and, hence, it gets its maximum at the boundaries of  $p(x | u, v_1, s_1)$ , i.e., when the last is equal 0 or 1. This implies that  $X$  can be expressed as a deterministic function of  $(U, V_1, S_1)$ .

*Part 3:* We prove now the cardinality bound for Theorem 1. First, let us recall the support lemma [32] (p. 310). Let  $\mathcal{P}(\mathcal{Z})$  be the set of PMFs on the set  $\mathcal{Z}$ , and let the set  $\mathcal{P}(\mathcal{Z} | \mathcal{Q}) \subseteq \mathcal{P}(\mathcal{Z})$  be a collection of PMFs  $p(z | q)$  on  $\mathcal{Z}$  indexed by  $q \in \mathcal{Q}$ . Let  $g_j$ ,  $j = 1, \dots, k$ , be continuous functions on  $\mathcal{P}(\mathcal{Z} | \mathcal{Q})$ . Then, for any  $Q \sim F_Q(q)$ , there exists a finite random variable  $Q' \sim p(q')$  taking at most  $k$  values in  $\mathcal{Q}$  such that

$$\begin{aligned} \mathbb{E}[g_j(p_{Z|Q}(z|Q))] &= \int_{\mathcal{Q}} g_j(p_{Z|Q}(z|q)) dF(q) \\ &= \sum_{q'} g_j(p_{Z|q}(z|q')) p(q'). \end{aligned} \quad (A57)$$

We first reduce the alphabet size of  $V_1$  while considering the alphabet size of  $U$  to be constant and then we calculate the cardinality of  $U$ . Consider the following continuous functions of  $p(x, s_1, s_2, u | v_1)$

$$g_j = \begin{cases} P_{XS_1S_2|V}(j|v_1), & j \in \{1, 2, \dots, |\mathcal{X}||\mathcal{S}_1||\mathcal{S}_2| - 1\}, \\ I(V_1; S_1) - I(V_1; Y, S_2) & j = |\mathcal{X}||\mathcal{S}_1||\mathcal{S}_2|, \\ I(U; Y, S_2 | V_1 = v_1) - I(U; S_1 | V_1 = v_1) & j = |\mathcal{X}||\mathcal{S}_1||\mathcal{S}_2| + 1. \end{cases} \quad (A58)$$

Then, by the support lemma, there exists a random variable  $V'_1$  with  $|\mathcal{V}'_1| \leq |\mathcal{X}||\mathcal{S}_1||\mathcal{S}_2| + 1$  such that  $p(x, s_1, s_2)$ ,  $I(V_1; S_1) - I(V_1; Y, S_2)$  and  $I(U; Y, S_2|V_1) - I(U; S_1|V_1)$  are preserved. Notice that the probability of  $U$  might have changed due to changing  $V_1$ ; we denote the corresponding  $U$  as  $U'$ . Next, for  $v'_1 \in \mathcal{V}'_1$  and the corresponding probability  $p(v'_1)$  that we found in the previous step, we consider  $|\mathcal{X}||\mathcal{S}_1||\mathcal{S}_2||\mathcal{V}'_1|$  continuous functions of  $p(x, s_1, s_2, v'_1|u')$

$$f_j = \begin{cases} P_{XS_1S_2V'_1|U'}(j|u') & j = \{1, 2, \dots, |\mathcal{X}||\mathcal{S}_1||\mathcal{S}_2||\mathcal{V}'_1| - 1\}, \\ I(U'; Y, S_2|V'_1) - I(U'; S_1|V'_1) & j = |\mathcal{X}||\mathcal{S}_1||\mathcal{S}_2||\mathcal{V}'_1|. \end{cases} \quad (\text{A59})$$

Thus, there exists a random variable  $U''$  with  $|\mathcal{U}''| \leq |\mathcal{X}||\mathcal{S}_1||\mathcal{S}_2||\mathcal{V}'_1|$  such that the mutual information expressions above and all the desired Markov conditions are preserved. Notice that the expression  $I(V_1; S_1) - I(V_1; Y, S_2)$  is being preserved since  $p(x, s_1, s_2, v'_1)$  is being preserved.

To conclude, we can bound the cardinality of the auxiliary random variables of Theorem 1 Case 1 by  $|\mathcal{V}_1| \leq |\mathcal{X}||\mathcal{S}_1||\mathcal{S}_2| + 1$  and  $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{S}_1||\mathcal{S}_2||\mathcal{V}_1| \leq |\mathcal{X}||\mathcal{S}_1||\mathcal{S}_2|(|\mathcal{X}||\mathcal{S}_1||\mathcal{S}_2| + 1)$  without limiting the generality of the solution.  $\square$

## Appendix E. Proofs for Section 5

### Appendix E.1. Proof of Lemma 4

**Proof.** For  $0 \leq \alpha \leq 1$  and  $\bar{\alpha} = 1 - \alpha$

$$\begin{aligned} J_w(\alpha q_1 + \bar{\alpha} q_2, \alpha Q_1 + \bar{\alpha} Q_2) &= \sum_{s_1, s_2, v_2, t, y} p(s_1, s_2) w(v_2|s_2) p(y|t, s_1, s_2, v_2) (\alpha q_1 + \bar{\alpha} q_2) \log \frac{\alpha Q_1 + \bar{\alpha} Q_2}{\alpha q_1 + \bar{\alpha} q_2} \\ &\stackrel{(a)}{\geq} \sum_{s_1, s_2, v_2, t, y} p(s_1, s_2) w(v_2|s_2) p(y|t, s_1, s_2, v_2) \left( \alpha q_1 \log \frac{Q_1}{q_1} + \bar{\alpha} q_2 \log \frac{Q_2}{q_2} \right) \\ &= \alpha J_w(q_1, Q_1) + \bar{\alpha} J_w(q_2, Q_2), \end{aligned} \quad (\text{A60})$$

where (a) follows from the log-sum inequality:

$$\sum_i a_i \log \frac{a_i}{b_i} \geq a \log \frac{a}{b}, \quad (\text{A61})$$

for  $\sum_i a_i = a$  and  $\sum_i b_i = b$ .  $\square$

### Appendix E.2. Proof of Lemma 6

**Proof.** Let us calculate  $q^*$  using the KKT conditions. We want to maximize  $J_w(q^*, Q)$  over  $q^*$ , where for all  $t, s_1$  and  $v_2$ ,  $0 \leq q^*(t|s_1, v_2) \leq 1$  and  $\sum_{t'} q^*(t'|s_1, v_2) = 1$ .

For fixed  $s_1$  and  $v_2$ ,

$$0 = \frac{\partial}{\partial q^*} \left( J_w(q^*, Q) + (1 - \sum_t q^*(t|s_1, v_2)) \nu_{s_1, v_2} \right) \quad (\text{A62})$$

$$= \sum_{s_2, y} p(s_1, s_2) w(v_2|s_2) p(y|t, s_1, s_2, v_2) \left( \log \frac{Q(t|y, s_2, v_2)}{q^*(t|s_1, v_2)} - 1 \right) - \nu_{s_1, v_2}, \quad (\text{A63})$$

divide by  $p(s_1, v_2)$ ,

$$0 = -\log q^*(t|s_1, v_2) + \frac{\sum_{s_2, y} p(s_1, s_2) w(v_2|s_2) p(y|t, s_1, s_2, v_2)}{p(s_1, v_2)} \log Q(t|y, s_2, v_2) - 1 + \frac{\nu_{s_1, v_2}}{p(s_1, v_2)}, \quad (\text{A64})$$

define  $-1 + \frac{\nu_{s_1, v_2}}{p(s_1, v_2)} = \log \nu'_{s_1, v_2}$ , hence

$$q^*(t|s_1, v_2) = \nu'_{s_1, v_2} \prod_{s_2, y} Q(t|y, s_2, v_2)^{p(s_2|s_1, v_2) p(y|t, s_1, s_2, v_2)}, \quad (\text{A65})$$

and from the constraint  $\sum_{t'} q^*(t'|s_1, v_2) = 1$  we get that

$$q^*(t|s_1, v_2) = \frac{\prod_{s_2, y} Q(t|y, s_2, v_2)^{p(s_2|s_1, v_2)p(y|t, s_1, s_2, v_2)}}{\sum_{t'} \prod_{s_2, y} Q(t'|y, s_2, v_2)^{p(s_2|s_1, v_2)p(y|t', s_1, s_2, v_2)}}. \quad (\text{A66})$$

□

### Appendix E.3. Proof of Lemma 7

The proof for this lemma is done in three steps: first, we prove that  $U_w(q_1)$  is greater than or equal to  $J_w(q_0, Q_0^*)$  for any two PMFs  $q_0(t|s_1, v_2)$  and  $q_1(t|s_1, v_2)$ , then, we use Lemmas 3 and 5 to state that for the optimal PMF,  $q_c(t|s_1, v_2)$ ,  $C_{2,w}^{lb} = J_w(q_c, Q_c^*)$ , and, therefore,  $U_w(q)$  is an upper bound of  $C_{2,w}^{lb}$  for every  $q(t|s_1, v_2)$ . Thirdly, we prove that  $U_w(q)$  converges to  $C_{2,w}^{lb}$ .

**Proof.** Consider any two PMFs,  $q_0(t|s_1, v_2)$  and  $q_1(t|s_1, v_2)$ , their corresponding  $\{p_0(s_1, s_2, v_2, t, y), Q_0^*(t|y, s_2, v_2)\}$  and  $\{p_1(s_1, s_2, v_2, t, y), Q_1^*(t|y, s_2, v_2)\}$ , respectively, according to (24) and (26) and consider also the following inequalities:

$$\begin{aligned} & \sum_{s_1, s_2, v_2, t, y} p_0(s_1, s_2, v_2, t, y) \log \frac{Q_1^*(t|y, s_2, v_2)}{q_1(t|s_1, v_2)} - J_w(q_0, Q_0^*) \\ &= \sum_{s_1, s_2, v_2, t, y} p_0(s_1, s_2, v_2, t, y) \left( \log \frac{Q_1^*(t|y, s_2, v_2)}{q_1(t|s_1, v_2)} - \log \frac{Q_0^*(t|y, s_2, v_2)}{q_0(t|s_1, v_2)} \right) \\ &= \sum_{s_1, s_2, v_2, t, y} p_0(s_1, s_2, v_2, t, y) \log \left( \frac{Q_1^*(t|y, s_2, v_2)}{Q_0^*(t|y, s_2, v_2)} \frac{q_0(t|s_1, v_2)}{q_1(t|s_1, v_2)} \right) \\ &= \mathbb{D}(q_0(t|s_1, v_2) \| q_1(t|s_1, v_2)) - \mathbb{D}(Q_0^*(t|y, s_2, v_2) \| Q_1^*(t|y, s_2, v_2)) \\ &\stackrel{(a)}{=} \mathbb{D}(q_0(t|s_1, s_2, v_2)p(y|t, s_1, s_2, v_2)p(s_1, s_2)w(v_2|s_2) \| q_1(t|s_1, s_2, v_2)p(y|t, s_1, s_2, v_2)p(s_1, s_2)w(v_2|s_2)) \\ &\quad - \mathbb{D}(Q_0^*(t|y, s_2, v_2) \| Q_1^*(t|y, s_2, v_2)) \\ &= \mathbb{D}(p_0(s_1, s_2, v_2, t, y) \| p_1(s_1, s_2, v_2, t, y)) - \mathbb{D}(Q_0^*(t|y, s_2, v_2) \| Q_1^*(t|y, s_2, v_2)) \\ &\stackrel{(b)}{=} \mathbb{D}(p_0(s_2, v_2, y)Q_0^*(t|y, s_2, v_2)p_0(s_1|s_2, v_2, t, y) \| p_1(s_2, v_2, y)Q_1^*(t|y, s_2, v_2)p_1(s_1|s_2, v_2, t, y)) \\ &\quad - \mathbb{D}(Q_0^*(t|y, s_2, v_2) \| Q_1^*(t|y, s_2, v_2)) \\ &= \mathbb{D}(p_0(s_2, v_2, y) \| p_1(s_2, v_2, y)) + \mathbb{D}(p_0(s_1|s_2, v_2, t, y) \| p_1(s_1|s_2, v_2, t, y)) \\ &\stackrel{(c)}{=} \geq 0, \end{aligned} \quad (\text{A67})$$

where  $\mathbb{D}(\cdot \| \cdot)$  is the K-L divergence,  $p_j(s_2, v_2, y)$  and  $p_j(s_1|s_2, v_2, t, y)$  are marginal distributions of  $p_j(s_1, s_2, v_2, t, y)$  for  $j = 0, 1$ , (a) follows from the fact that  $T$  is independent of  $S_2$  given  $(S_1, V_2)$  and from the K-L divergence properties, (b) follows from the fact that  $Q_j^*(t|y, s_2, v_2)$  is a marginal distribution of  $p_j(s_1, s_2, v_2, t, y)$  for  $j = 0, 1$  and (c) follows from the fact that  $\mathbb{D}(\cdot \| \cdot) \geq 0$  always.

Thus,

$$\begin{aligned} J(q_0, Q_0^*) &\leq \sum_{s_1, s_2, v_2, t, y} p_0(s_1, s_2, v_2, t, y) \log \frac{Q_1^*(t|y, s_2, v_2)}{q_1(t|s_1, v_2)} \\ &= \sum_{s_1, s_2, v_2, t, y} p(s_1, s_2)w(v_2|s_2)q_0(t|s_1, v_2)p(y|t, s_1, s_2, v_2) \log \frac{Q_1^*(t|y, s_2, v_2)}{q_1(t|s_1, v_2)} \\ &= \sum_{s_1, v_2} p(s_1, v_2) \sum_t q_0(t|s_1, v_2) \sum_{s_2} p(s_2|s_1, v_2) \sum_y p(y|t, s_1, s_2, v_2) \log \frac{Q_1^*(t|y, s_2, v_2)}{q_1(t|s_1, v_2)} \\ &\leq \sum_{s_1, v_2} p(s_1, v_2) \max_{t'} \sum_{s_2} p(s_2|s_1, v_2) \sum_y p(y|t', s_1, s_2, v_2) \log \frac{Q_1^*(t'|y, s_2, v_2)}{q_1(t'|s_1, v_2)} \\ &= U_w(q_1). \end{aligned} \quad (\text{A68})$$

We proved that  $U_w(q_1)$  is greater than or equal to  $J_w(q_0, Q_0^*)$  for any choice of  $q_0(t|s_2, v_2)$  and  $q_1(t|s_1, v_2)$ . Therefore, by taking  $q_0(t|s_1, v_2)$  to be the distribution that achieves  $C_{2,w}^{lb}$  and by considering Lemmas 3 and 5, we conclude that  $U_w(q) \geq C_{w,2}$  for any choice of  $q(t|s_1, v_2)$ .

In order to prove that  $U_w(q)$  converges to  $C_{2,w}^{lb}$  let us rewrite Equation (A63) as

$$\sum_{s_2, y} p(s_2|s_1, v_2) p(y|t, s_1, s_2, v_2) \log \frac{Q(t|y, s_2, v_2)}{q^*(t|s_1, v_2)} = v'_{s_1, v_2}. \quad (\text{A69})$$

We can see that for a fixed  $Q$ , the right hand side of the equation is independent of  $t$ . Considering also

$$\begin{aligned} J_w(q, Q) &= \sum_{s_1, s_2, v_2, t, y} p(s_1, s_2) w(v_2|s_2) q(t|s_1, v_2) p(y|t, s_1, s_2, v_2) \log \frac{Q(t|y, s_2, v_2)}{q(t|s_1, v_2)} \\ &\leq \sum_{s_1, v_2} p(s_1, v_2) \max_{t'} \sum_{s_2} p(s_2|s_1, v_2) \sum_y p(y|t', s_1, s_2, v_2) \log \frac{Q^*(t'|y, s_2, v_2)}{q(t'|s_1, v_2)}, \end{aligned} \quad (\text{A70})$$

we can conclude that the equation holds when the PMF  $q$  is the PMF that achieves  $C_{2,w}^{lb}$ .  $\square$

## References

1. Steinberg, Y. Coding for Channels With Rate-Limited Side Information at the Decoder, with Applications. *IEEE Trans. Inf. Theory* **2008**, *54*, 4283–4295.
2. Gel'fand, S.I.; Pinsker, M.S. Coding for Channel with Random Parameters. *Probl. Control Theory* **1980**, *9*, 19–31.
3. Wyner, A.; Ziv, J. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory* **1976**, *22*, 1–10.
4. Shannon, C.E. Channels with side information at the transmitter. *IBM J. Res. Dev.* **1958**, *2*, 289–293.
5. Heegard, C.; Gamal, A.A.E. On the capacity of computer memory with defects. *IEEE Trans. Inf. Theory* **1983**, *29*, 731–739.
6. Cover, T.M.; Chiang, M. Duality between channel capacity and rate distortion with two-sided state information. *IEEE Trans. Inf. Theory* **2006**, *48*, 1629–1638.
7. Rosenzweig, A.; Steinberg, Y.; Shamai, S. On channels with partial channel state information at the transmitter. *IEEE Trans. Inf. Theory* **2005**, *51*, 1817–1830.
8. Cemal, Y.; Steinberg, Y. Coding Problems for Channels With Partial State Information at the Transmitter. *IEEE Trans. Inf. Theory* **2007**, *53*, 4521–4536.
9. Keshet, G.; Steinberg, Y.; Merhav, N. Channel Coding in the Presence of Side Information. *Found. Trends Commun. Inf. Theory* **2007**, *4*, 445–586.
10. Kaspi, A.H. Two-way source coding with a fidelity criterion. *IEEE Trans. Inf. Theory* **1985**, *31*, 735–740.
11. Permuter, H.; Steinberg, Y.; Weissman, T. Two-Way Source Coding With a Helper. *IEEE Trans. Inf. Theory* **2010**, *56*, 2905–2919.
12. Weissman, T.; Gamal, A.E. Source Coding With Limited-Look-Ahead Side Information at the Decoder. *IEEE Trans. Inf. Theory* **2006**, *52*, 5218–5239.
13. Weissman, T.; Merhav, N. On causal source codes with side information. *IEEE Trans. Inf. Theory* **2005**, *51*, 4003–4013.
14. Shannon, C.E. Coding Theorems for a Discrete Source with a Fidelity Criterion. *IRE Nat. Conv. Rec.* **1959**, *4*, 142–163.
15. Pradhan, S.S.; Chou, J.; Ramchandran, K. Duality between source coding and channel coding and its extension to the side information case. *IEEE Trans. Inf. Theory* **2003**, *49*, 1181–1203.
16. Pradhan, S.S.; Ramchandran, K. On functional duality in multiuser source and channel coding problems with one-sided collaboration. *IEEE Trans. Inf. Theory* **2006**, *52*, 2986–3002.
17. Zamir, R.; Shamai, S.; Erez, U. Nested linear/lattice codes for structured multiterminal binning. *IEEE Trans. Inf. Theory* **2006**, *48*, 1250–1276.

18. Su, J.; Eggers, J.; Girod, B. Illustration of the duality between channel coding and rate distortion with side information. In Proceedings of the 2000 Conference Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 29 October–1 November 2000; Volume 2, pp. 1841–1845.
19. Goldfeld, Z.; Permuter, H.H.; Kramer, G. Duality of a Source Coding Problem and the Semi-Deterministic Broadcast Channel With Rate-Limited Cooperation. *IEEE Trans. Inf. Theory* **2016**, *62*, 2285–2307.
20. Asnani, H.; Permuter, H.H.; Weissman, T. Successive Refinement With Decoder Cooperation and Its Channel Coding Duals. *IEEE Trans. Inf. Theory* **2013**, *59*, 5511–5533.
21. Gupta, A.; Verdu, S. Operational Duality Between Lossy Compression and Channel Coding. *IEEE Trans. Inf. Theor.* **2011**, *57*, 3171–3179.
22. Blahut, R.E. Computation of channel capacity and rate-distortion functions. *IEEE Trans. Inform. Theory* **1972**, *18*, 460–473.
23. Arimoto, S. An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless Channels. *IEEE Trans. Inf. Theory* **1972**, *18*, 14–20.
24. Willems, F.M.J. *Computation of the Wyner-Ziv Rate-Distortion Function*; Research Report; University of Technology: Hong Kong, China, 1983.
25. Dupuis, F.; Yu, W.; Willems, F. Blahut-Arimoto algorithms for computing channel capacity and rate-distortion with side information. In Proceedings of the 2004 International Symposium on Information Theory, Chicago, IL, USA, 27 June–2 July 2004; p. 179.
26. Cheng, S.; Stankovic, V.; Xiong, Z. Computing the channel capacity and rate-distortion function with two-sided state information. *IEEE Trans. Inf. Theory* **2005**, *51*, 4418–4425.
27. Sumszyk, O.; Steinberg, Y. Information embedding with reversible stegotext. In Proceedings of the 2009 IEEE International Symposium on Information Theory, Seoul, Korea, 28 June–3 July 2009; pp. 2728–2732.
28. Naiss, I.; Permuter, H.H. Extension of the Blahut-Arimoto Algorithm for Maximizing Directed Information. *IEEE Trans. Inf. Theory* **2013**, *59*, 204–222.
29. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 1991.
30. Yeung, R.W. *Information Theory and Network Coding*, 1 ed.; Springer: Berlin, Germany, 2008.
31. Berger, T. Multiterminal source coding. In *Information Theory Approach to Communications*; Longo, G., Ed.; CSIM Course and Lectures; Springer: Berlin, Germany, 1978; pp. 171–231.
32. Csiszar, I.; Korner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*/Imre Csiszar and Janos Korner; Academic Press: Cambridge, MA, USA, 1981.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).