Channel Coding and Source Coding With Increased Partial Side Information

Avihay Shirazi  
Ben Gurion University  
avihays@bgu.ac.il

Uriya Basher  
Ben Gurion University  
basher@bgu.ac.il

Haim H. Permuter  
Ben Gurion University  
haimp@bgu.ac.il

Abstract—Let \((S_{1,i}, S_{2,i}) \sim \text{i.i.d } p(s_{1,i}, s_{2,i}), i = 1, 2, \ldots\) be a memoryless correlated partial side information sequence. In this work we study channel coding and source coding problems, where the encoder is informed with the encoder’s side information (ESI, \(S_1\)) and the decoder is informed with the decoder’s side information (DSI, \(S_2\)), and additionally, either the encoder or the decoder is also informed with a version of the other’s side information with a rate limited to \(R_d\). We derive six special cases of channel coding and source coding problems and provide a single-letter characterization for the capacity or the rate-distortion function for the different cases. We then present a duality between the channel capacity and the rate distortion for the cases we study.

I. INTRODUCTION

Channel capacity with state information, source coding with side information and the duality between them were subjects of interest for many researchers throughout the years. Shannon [1] presented the capacity for a channel with causal CSI at the transmitter. Gelfand and Pinsker [2] presented the capacity for a channel with non-causal CSI at the transmitter. Heegard and El Gamal [3] presented a model of a state dependent channel, where the transmitter is informed with CSI at a rate limited to \(R_c\) and the receiver is informed with CSI at a rate limited to \(R_d\). They have presented a tight achievability for the capacity of the following cases: a) no CSI at the encoder and the decoder, b) fully informed encoder and decoder, c) full CSI at the encoder and no CSI at the decoder (the Gelfand-Pinsker problem), and d) full CSI at the decoder and \(R_c\) rate limited CSI at the encoder. Steinberg [4] presented the capacity for a channel class where the encoder is informed with full CSI and the decoder is informed with CSI at a rate limited to \(R_d\). Cover and Chiang [5] presented the capacity for a class of channel with two sided state information, where the encoder is informed with its local state information \(S_1\) and the decoder is informed with its local state information \(S_2\), and \(S_1, S_2\) are correlated. Keshet, Steinberg and Merhav [6] presented a detailed subject review on channel coding with state information. The papers [3], [4] and [5] are altogether strongly related to the current work and provide a basis for the forthcoming conclusions. With regard to source coding, in their landmark paper, Wyner and Ziv [7] presented the rate distortion for the case where the decoder is informed with full side information, and the encoder is informed with no side information. Cover and Chiang [5] presented a model of a rate distortion problem where the encoder is informed with \(S_1\) and the decoder is informed with \(S_2\), and \(S_1, S_2\) are correlated. Kaspi [8] presented the rate distortion for the case that there are two users (encoders) trying to communicate between them in a consecutive way. The problem mode presented in Kaspi’s paper is a general case of two of the source coding problems presented in this paper, and one can directly derive the rate regions for these two problems from [8, Theorem I]. Causal source coding was first introduced by Neuhoff and Gilbert [9] and was studied also by Weissman and Merhav [10], Weissman and El Gamal [11] and others. The duality between the channel coding and the source coding was first mentioned by Shannon [12]. Since that significant paper, this duality was studied by many, and included the duality between the Wyner-Ziv and Gelfand-Pinsker problems [13]. Cover and Chiang [5] presented in their work the duality for some fundamental channel capacity and rate distortion problems, and the duality for the two general cases mentioned above (the channel capacity or rate distortion with two sided state information). Let us denote the encoder side information as ESI and the decoder side information as DSI. For the channel coding problems, for the convenience of the reader, we will also refer to the transmitter as the encoder, the receiver as the decoder and state information as side information. Hence, the notations ESI and DSI also hold for channel coding problems. In this paper we will examine three cases of channel coding.
problems (see Figure 1) and three cases of source coding problems (see Figure 2). In all six cases the encoder and the decoder are each informed with their own local state (or side) information (ESI and DSI, respectively), where the ESI and DSI are correlated. In addition, one of the users’ side information is increased with a rate limited version of the other user’s side information (i.e., either rate limited ESI at the decoder or rate limited DSI at the encoder). We will present the information-theoretic duality between the channel coding and the source coding cases, and provide a single-letter solution for each case.

II. PROBLEM SETTING

In this section we will describe and formally define three cases of channel coding problems and three cases of source coding problems. All six cases are presented in Figures 1 and 2.

A. Definitions and Problem Formulation - Channel Coding with State Information

Definition 1. A discrete channel controlled by random parameters $K$ is the sextuple \( \{X, S_1, S_2, p(s_1, s_2), p(y|x, s_1, s_2), \mathcal{Y} \} \). The channel input sequence \( \{X_i \in X, i = 1, 2, \ldots \} \), the ESI sequence \( \{S_{1,i} \in S_1, i = 1, 2, \ldots \} \), the DSI sequence \( \{S_{2,i} \in S_2, i = 1, 2, \ldots \} \) and the channel output sequence \( \{Y_i \in \mathcal{Y}, i = 1, 2, \ldots \} \) are discrete random variables drawn from the finite alphabets \( X, S_1, S_2, \mathcal{Y} \), respectively. Denote the message and the message space as \( W \in \mathcal{W} := \{1, 2, \ldots, M\} \), and let \( W \) 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., the output \( Y^n \) has a conditional distribution

\[
P(y^n|x^n, s^n_1, s^n_2) = p(y^n|x^n, s^n_1, s^n_2),
\]

and since there is no feedback, it follows that

\[
P(y^n|x^n, s^n_1, s^n_2) = \prod_{i=1}^{n} p(y_i|x_i, s_{1,i}, s_{2,i}).
\]

Problem Formulation. For the channel $K$, consider the following channel coding problem cases:

- **Case 1** The encoder is informed with the ESI ($S^n_1$) and the decoder is informed with the DSI ($S^n_2$) at the decoder. Case 1: Rate limited ESI at the decoder.
- **Case 2** The encoder is informed with the ESI ($S^n_1$) and the decoder is informed with the DSI ($S^n_2$). Case 2: Rate limited DSI at the encoder.
- **Case 2C** The encoder is informed with the causal ESI ($S^n_1$) and rate limited DSI 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.

All cases are presented in Figure 1.

Definition 2. A \((n, M, \bar{M})\) code, \( j \in \{1, 2\} \), for the channel $K$ consists of two encoding maps and one decoding map, as described for each case:

**Case 1:** Encoding maps

\[
f_s : S^n_1 \rightarrow \{1, 2, \ldots, \bar{M}_j\},
\]

\[
f : \mathcal{W} \times S^n_1 \times \{1, 2, \ldots, \bar{M}_1\} \rightarrow X^n,
\]

and a decoding map

\[
g : \mathcal{Y} \times S^n_2 \times \{1, 2, \ldots, \bar{M}_1\} \rightarrow \mathcal{W}.
\]
Case 2: Encoding maps
\[ f_s : S_2^n \to \{1, 2, \ldots, \tilde{M}_2\}, \]
\[ f : W \times S_1^n \times \{1, 2, \ldots, \tilde{M}_2\} \to \mathcal{X}^n, \]
and a decoding map
\[ g : \mathcal{Y}^n \times S_2^n \times \{1, 2, \ldots, \tilde{M}_2\} \to \mathcal{W}. \]  

Case 2C: Encoding maps
\[ f_s : S_2^n \to \{1, 2, \ldots, \tilde{M}_2\}, \]
\[ f_i : \mathcal{W} \times S_i^n \times \{1, 2, \ldots, \tilde{M}_1\} \to \mathcal{X}_i, \]
and a decoding map
\[ g : \mathcal{Y}^n \times S_2^n \times \{1, 2, \ldots, \tilde{M}_2\} \to \mathcal{W}. \]

The rate pair \((R, \tilde{R}_a)\) of the code \((n, M, \tilde{M}_j)\) is defined as
\[ R = \frac{1}{n} \log M, \quad \tilde{R}_a = \frac{1}{n} \log \tilde{M}_j. \]

The average probability of error \(P_e(n)\) for a \((M, \tilde{M}_j, n)\) code is defined as
\[ P_e(n) = \frac{1}{M} \sum_{w=1}^{M} \Pr \{ \hat{W} \neq W \mid W = w \}. \]

where the index \(W\) is chosen according to a uniform distribution over the set \(\mathcal{W}\). A rate pair \((R, \tilde{R}_a)\) is said to be achievable if there exists a sequence of \((2^{-nR}, 2^{-nR_a}, n)\) codes such that the average probability of error \(P_e(n)\) goes to zero as \(n \to \infty\).

Definition 3. The capacity of the channel \(C(\tilde{R}_a)\) is the supremum of all \(R\) such that the rate pair \((R, \tilde{R}_a)\) is achievable.

B. Definitions and Problem Formulation - Source Coding with Side Information

Throughout this article we use the regular definitions of rate distortion as presented in [14].

Definition 4. The source sequence \(\{X_i \in \mathcal{X}, i = 1, 2, \ldots\}\), the ESI sequence \(\{S_{1,i} \in S_1, i = 1, 2, \ldots\}\), the DSI sequence \(\{S_{2,i} \in S_2, i = 1, 2, \ldots\}\) are discrete random variables drawn from the finite alphabets \(\mathcal{X}, S_1\) and \(S_2\) 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 \(d_x : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)\) be the distortion measure. The distortion between sequences is defined in the usual way
\[ d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i). \]

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 the ESI \((S_1^n)\) and the decoder is informed with the DSI \((S_2^n)\) and rate limited ESI.
- **Case 2** The encoder is informed with the ESI \((S_1^n)\) and rate limited DSI and the decoder is informed with the DSI \((S_2^n)\).
- **Case 1C** The encoder is informed with the ESI and the decoder is informed with the causal ESI \((S_1^n)\) at time \(i\) and rate limited ESI (this case is similar to case 1, except for the causal DSI).

All cases are presented in Figure 2.

Definition 5. Let \(\mathcal{I}_M\) be the set of integers \(\mathcal{I}_M = \{1, 2, \ldots, M\}\). A \((n, M, \tilde{M}_j, D)\) code, \(j \in [1, 2]\) for the source coding cases illustrated in Figure 2 contains two encoders, a single decoder and a distortion constraint as described for each case:

**Case 1** Encoding maps
\[ f_s : S_1^n \to \{1, 2, \ldots, \tilde{M}_1\}, \]
\[ f : \mathcal{X}^n \times S_1^n \times \{1, 2, \ldots, \tilde{M}_1\} \to \mathcal{I}_M, \]
and a decoding map
\[ g : \mathcal{I}_M \times S_2^n \times \{1, 2, \ldots, \tilde{M}_2\} \to \hat{\mathcal{X}}^n. \]

**Case 2** Encoding maps
\[ f_s : S_2^n \to \{1, 2, \ldots, \tilde{M}_2\}, \]
\[ f : \mathcal{X}^n \times S_1^n \times \{1, 2, \ldots, \tilde{M}_2\} \to \mathcal{I}_M, \]
and a decoding map
\[ g : \mathcal{I}_M \times S_2^n \times \{1, 2, \ldots, \tilde{M}_2\} \to \hat{\mathcal{X}}^n. \]

**Case 1C** Encoding maps
\[ f_s : S_1^n \to \{1, 2, \ldots, \tilde{M}_1\}, \]
\[ f : \mathcal{X}^n \times S_1^n \times \{1, 2, \ldots, \tilde{M}_1\} \to \mathcal{I}_M, \]
and a decoding map
\[ g : \mathcal{I}_M \times S_2^n \times \{1, 2, \ldots, \tilde{M}_2\} \to \hat{\mathcal{X}}^n. \]

The distortion constraint for all three cases is:
\[ \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{n} d(X_i, \hat{X}_i) \right] \leq D. \]

The rate pair \((R, \tilde{R}_a)\) of the \((n, M, \tilde{M}_j, D)\) code is defined as
\[ R = \frac{1}{n} \log M, \quad \tilde{R}_a = \frac{1}{n} \log \tilde{M}_j. \]

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

Definition 6. For a given \(\tilde{R}_a\) and distortion \(D\), the opera-
tional rate $R^*(\hat{R}_s, D)$ is the infimum of all $R$, such that the rate pair $(R, \hat{R}_s)$ is achievable.

III. MAIN RESULTS

Here 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.

A. Main Results - Channel Coding with Side Information

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

\[ C_1^* = \max_{p(x|s_1)p(u|x_1,s_1)p(x|u,x_1,\hat{s}_1)} \left[ I(U;Y,S_1|\hat{S}_1) - I(U;S_1|\hat{S}_1) \right]. \]

Using the transformation $R\to \hat{R}$, the capacity is as follows

\[ C_2^* = \max_{p(x|s_2)p(u|x_2,s_2)} \left[ I(U;Y,S_2|\hat{S}_2) - I(U;S_2|\hat{S}_2) \right]. \]

Case 2C: The encoder is informed with causal ESI $(S_1^t)$ at time $i$ and rate limited DSI and the decoder is informed with the DSI $(S_2^t)$.

\[ C_{2C}^* = \max_{p(x|s_2)p(u|x_2,s_2)} I(U;Y,S_2|\hat{S}_2). \]

For $j \in \{1,2\}$, some joint distribution $p(x_1, s_1, s_2, \hat{s}_j, u, x, y)$ and $(U, \hat{S}_j)$ - some auxiliary random variables with bounded cardinality.

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

(i) The function $C_j(\hat{R}_s)$ is a concave function of $\hat{R}_s$.

(ii) It is enough to take $X$ to be a deterministic function of $(U, S_1, \hat{S}_j)$ to evaluate $C_j$.

B. Main Results - Source Coding with Side Information

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

Theorem 2 (The rate distortion function for the cases in Figure 2). 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 the ESI $(S_1^t)$ and the decoder is informed with the DSI $(S_2^t)$ and rate limited ESI.

\[ R_1^*(D) = \min_{p(x|s_1)p(u|x_1,s_1)\text{ s.t. } R_2 \geq I(S_1|\hat{S}_1) - I(S_1;S_2)} \left[ I(U;X,S_1|\hat{S}_1) - I(U;S_2|\hat{S}_1) \right]. \]

Case 1C: The encoder is informed with the ESI and the decoder is informed with the causal ESI $(S_2^t)$ at time $i$ and rate limited ESI.

\[ R_{1C}^*(D) = \min_{p(x|s_1)p(u|x_1,s_1)\text{ s.t. } R_2 \geq I(S_1;S_1)} \left[ I(U;X,S_1|\hat{S}_1) - I(U;S_2|\hat{S}_1) \right]. \]

Case 2: The encoder is informed with the ESI $(S_1^t)$ and rate limited DSI and the decoder is informed with the DSI $(S_2^t)$.

\[ R_2^*(D) = \min_{p(x|s_2)p(u|x_2,s_2)} \left[ I(U;X,S_1|\hat{S}_2) - I(U;S_2|\hat{S}_2) \right]. \]

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

(i) The function $R_j(\hat{R}_s, D)$ is a convex function of $\hat{R}_s$ and $D$.

(ii) It is enough to take $\hat{X}$ to be a deterministic function of $(U, S_2, \hat{S}_j)$ to evaluate $R_j$.

C. Main Results - Duality

We now investigate the duality between the channel coding and the source coding for the cases in Figures 1 and 2. With the following transformation it is noticeable that channel coding cases 1, 2, 2C are dual to the source coding cases 2, 1, 1C, respectively. The left column correspond with channel coding and the left column with source coding. Let $j, \tilde{j} \in \{1,2\}, \tilde{j} \neq j$, and consider the transformation:

\[ \begin{align*}
\text{channel coding } &\leftrightarrow \text{source coding} \\
C &\leftrightarrow R(D) \\
\text{maximization } &\leftrightarrow \text{minimization} \\
C_j &\leftrightarrow R_j(D) \\
X &\leftrightarrow \hat{X} \\
Y &\leftrightarrow X \\
S_j &\leftrightarrow S_j \\
\hat{S}_j &\leftrightarrow \hat{S}_j
\end{align*} \]

756
then, the duality between the channel coding and the source coding cases is evident. This transformation is an extension of the transformation provided in [5]. Note that while the channel capacity formula in case \( j \) and the rate distortion function in case \( j \) are dual to each other in the sense of maximization-minimization, the corresponding rates \( \hat{R}_s \) 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, but instead, we have the inequality is in the same direction \((\leq \leftrightarrow \leq)\) in the \( \hat{R}_s \) formulas. The duality in the side information rates \( \hat{R}_s \), then, is in the sense that the arguments in the formulas for the dual \( \hat{R}_s \) 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.

IV. PROOF OF THEOREM 1 CASE 2

In order to solve this problem we have solve first a Wyner-Ziv problem for the side information, where the state encoder is the W-Z encoder, the encoder is the W-Z decoder and \( S_1 \) is the W-Z side information at the decoder. Then, we have to solve a Gelfand-Pinsker problem for the main channel, where \((S_1, \hat{S}_2)\) is the G-P side information at the encoder and we consider \((Y, S_2, \hat{S}_2)\) to be the output of the G-P channel.

Achievability: Given \((S_{1,i}, S_{2,i}) \sim \text{i.i.d.} \ p(s_1, s_2)\) and the memoryless channel \( p(y|x, s_1, s_2) \), fix \( p(s_1, s_2) p(\hat{s}_2|s_2)p(u|s_1, s_2) p(x|u, s_1, s_2)\) where \( x = f(u, s_1, s_2) \) (i.e., \( p(x|u, s_1, s_2) \) gets the values 0 or 1).

Codebook Generation and Random Binning

1) Generate at random \( 2^{nR_u^0} \) sequences \( \tilde{S}_2^n(k) \) drawn i.i.d. \( \sim \prod_{i=1}^{n} p(\tilde{s}_2) \), index them and construct the codebook \( C_s = \{ \tilde{S}_2(1), \tilde{S}_2(2), \ldots, \tilde{S}_2(2^{nR_u^0}) \} \). Randomly and uniformly distribute the sequences \( \tilde{S}_2^n(k) \), \( k \in \{1, 2, \ldots, 2^{nR_u^0} \} \) into \( 2^{nR_u} \) random bins. Let \( j \) denote the index assigned to the bin \( B_{s}(j) \).

2) Generate at random \( 2^{nR_u^0} \) sequences \( U^n(l) \) drawn i.i.d. \( \sim \prod_{i=1}^{n} p(u_i) \), index them and build the codebook \( C_u = \{ U^n(1), U^n(2), \ldots, U^n(2^{nR_u^0}) \} \). Randomly and uniformly distribute the sequences \( U^n(l) \), \( l \in \{1, 2, \ldots, 2^{nR_u^0} \} \) into \( 2^{nR_u} \) random bins. Let \( w \) denote the index assigned to bin \( B_u(w) \).

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

Encoding

1) State Encoder: Given the sequence \( S_2^n \), the state encoder looks for a \( \tilde{S}_2^n(k) \in C_s \) such that \( (\tilde{S}_2^n, \tilde{S}_2^n(k)) \) are jointly typical. Upon finding such a \( \tilde{S}_2^n(k) \), the state encoder stops searching and sends the index \( j \) of the bin \( B_s(j) \), where \( \tilde{S}_2(k) \) is located. If no such \( k \) exists, the state encoder declares an error.

2) Encoder: Given the message \( W = w \), sequence \( S_2^n \) and the index \( j \), the encoder will look for a \( \tilde{S}_2^n(k) \) inside the bin \( B_s(j) \) such that \( (S_2^n, \tilde{S}_2^n(k)) \in T_r^n(S_1, \hat{S}_2) \). If a unique such \( \tilde{S}_2^n(k) \) is found, the encoder will look inside the bin \( B_u(w) \) (the bin that corresponds with the message \( w \)) for a \( U^n(l) \in B_u(w) \) such that \( (U^n(l), S_2^n, \tilde{S}_2^n(k)) \in T_r^n(U, S_1, \hat{S}_2) \). If the encoder finds a unique such \( U^n(l) \), it transmits \( x_i = f(U_i, S_{1,i}, \hat{S}_{2,i}(k)) \) at any time \( i \in \{1, 2, \ldots, n\} \). Otherwise, if the encoder is not able to find either a unique \( U^n(l) \) or a unique \( \tilde{S}_2^n(k) \), it declares an error.

Decoding

Given \( Y^n, S_2^n \) and \( \tilde{S}_2^n(k) \), the decoder looks for a \( U^n(l) \in C_u \) such that \( (Y^n, U^n(l), S_2^n, \tilde{S}_2^n(k)) \in T_r^n(Y, U, S_1, \hat{S}_2) \). If it finds a unique such \( U^n(l) \), it declares the message to be the index \( \hat{w} = w \) of the bin \( B_u(w) \), where \( U^n(l) \) is located. Otherwise, if it cannot find a unique such sequence, it declares 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 \( W=1, S_2^n \) are \((k = K, l = L \) and \( j = J)\). i.e., \( \tilde{S}_2^n(K) \) corresponds with \( S_2^n, \tilde{S}_2^n(K) \in B_s(J) \) and \( U^n(L) \) is chosen according to \((W = 1, S_2^n, \tilde{S}_2^n(K))\).

Define the following events:

\[
E_1 := \{ S_2^n \in T_r^n(S_2), (S_2^n, \tilde{S}_2^n(k)) \notin T_r^n(S_2, \tilde{S}_2) \} \\
\forall \tilde{S}_2^n(k) \in C_s \tag{30}
\]

\[
E_2 := \{ \exists k' \neq K : \tilde{S}_2^n(k') \in B_s(J) \text{ and} \} \\
(\tilde{S}_2^n, \tilde{S}_2^n(k')) \in T_r^n(S_1, \hat{S}_2) \tag{31}
\]

\[
E_3 := \{ \forall U^n(l) \in B_u(1), (U^n(l), S_2^n, \tilde{S}_2^n(K)) \} \\
\notin T_r^n(U, S_1, \hat{S}_2) \tag{32}
\]

\[
E_4 := \{ \exists l' \neq L : U^n(l') \in C_u \text{ and} \} \\
(U^n(l'), Y^n, S_2^n, \tilde{S}_2^n(K)) \notin T_r^n(U, Y, S_2, \tilde{S}_2) \} \tag{33}
\]

The probability of error \( P_u^n \) is upper bounded by \( P_u^n \leq P(E_1) + P(E_2) + P(E_3) + P(E_4) \). Using standard arguments, and assuming that \((S_2^n, \tilde{S}_2^n) \in T_r^n(S_1, \hat{S}_2) \) and that \( n \) is large enough, we can state that
The probability that there is no \( k \in \{1, 2, \ldots, 2^nR_t'\} \) such that \((S^n_2, \tilde{S}_2^n(k))\) is strongly jointly typical is exponentially small provided that \( R_t' > \bar{I}(S_2; \tilde{S}_2) + \epsilon. \) This follows from the standard rate distortion argument that \( 2^nR_t', \tilde{S}_2^n \) "cover" \( S_2^n \), therefore \( P(E_1) \to 0. \)

2) By the Markov lemma, since \((S^n_1, S^n_2)\) are strongly jointly typical and \((S^n_2, \tilde{S}_2^n(k))\) are strongly jointly typical, then \((S^n_1, S^n_2, \tilde{S}_2^n(k))\) are also strongly jointly typical. This follows that

\[
P(E_2) = \Pr \left\{ \bigcup_{\tilde{S}_2^n(k') \in C_u \atop k' \neq K} (S^n_2, \tilde{S}_2^n(k')) \in \mathcal{T}_e(n)(S_2, \tilde{S}_2) \right\}
\]

\[
\leq \sum_{k'=1 \atop k' \neq K} \Pr \left\{ (S^n_2, \tilde{S}_2^n(k')) \in \mathcal{T}_e(n)(S_2, \tilde{S}_2) \right\}
\]

\[
\leq 2^{n(R_t'-R_s)} \cdot 2^{-n(I(S_2; \tilde{S}_2) - \epsilon)}.
\]

the probability that there is another \( \tilde{S}_2^n(k^*) \), \( k^* \neq K \), in the bin \( B_s(J) \) that is strongly jointly typical with \( S_1^n \) is bounded by the number of \( \tilde{S}_2^n(k) \)'s in the bin times the probability of joint typicality. Therefore, if \( R_s > \bar{I}(S_2; \tilde{S}_2) - \bar{I}(S_1; \tilde{S}_2) + \epsilon \) then \( P(E_2) \to 0. \) Furthermore, using the Markov chain \( \tilde{S}_2 - S_2 - S_1 \) we can see that the inequality can be presented as \( R_s > \bar{I}(S_2; S_2, S_1). \)

3) 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 the bin \( B_u(l) \) that is strongly jointly typical with \((S^n_1, \tilde{S}_2^n(k))\), \( P(E_3) \to 0 \) as \( n \to \infty \) if \( R_u - R > \bar{I}(U; S_1, \tilde{S}_2) + \epsilon. \) Notice that \( 2^n(R_u - R) \) stands for the average number of \( U^n(l) \)'s in

4) We can assume that \( X^n \) is a deterministic function of \((U^n(l), S^n_1, S^n_2, \tilde{S}_2^n(K))\) (see Lemma 1). Therefore

\[
P(E_4) = \Pr \left\{ \bigcup_{l' \in \mathcal{C}_u \atop l' \neq L} (U^n(l'), Y^n, S^n_2, \tilde{S}_2^n(K)) \in \mathcal{T}_e(n)(U, Y, S_2, \tilde{S}_2) \right\}
\]

\[
\leq \sum_{l'=1 \atop l' \neq L} \Pr \left\{ (U^n(l'), Y^n, S^n_2, \tilde{S}_2^n(K)) \in \mathcal{T}_e(n)(U, Y, S_2, \tilde{S}_2) \right\}
\]

\[
\leq 2^{nR_u} \cdot 2^{-n(I(U; Y, S_2, \tilde{S}_2) - \epsilon)}.
\]

This shows that for rates \( R \) and \( \tilde{R}_s \) as described, and for large enough \( n \), the error events are of arbitrarily small probability. This concludes the proof of the achievability for the channel coding Case 2.

Converse: Fix a rate \( \tilde{R}_s \) and a sequence of codes \((2^{nR_u}, 2^{n\tilde{R}_s}, n)\) that achieve capacity. By Fano's inequality, \( H(W|Y^n, S^n_2) \leq n\epsilon_n \), where \( \epsilon_n \to 0 \) as \( n \to \infty \). Let \( T = f_s(S^n_2) \), and define \( \tilde{S}_{2,i} = (T, Y_{i-1}, S_{1,i+1}, S_{2,i}^{-1}) \). Then,

\[
n\tilde{R}_s \geq H(T)
\]

\[
\geq H(T|S^n_1) - H(T|S^n_1, S^n_2)
\]

\[
= I(T; S^n_2|S^n_1) - H(S^n_2|T, S^n_1)
\]

\[
= \sum_{i=1}^{n} \left[ H(S_{2,i}|S^n_1, S_{2,i}^{-1}) - H(S_{2,i}|T, S^n_1, S_{2,i}^{-1}) \right]
\]

\[
= \sum_{i=1}^{n} \left[ H(S_{2,i}|S_{1,i}) - H(S_{2,i}|S_{1,i}, S_{2,i}^{-1}) - H(S_{2,i}|T, S_{1,i}, S^n_1, S_{2,i}^{-1}, Y_{i-1}) \right]
\]

\[
= \sum_{i=1}^{n} \left[ H(S_{2,i}|S_{1,i}) - H(S_{2,i}|S_{2,i}) \right]
\]

\[
= \sum_{i=1}^{n} I(S_{2,i}; \tilde{S}_{2,i}|S_{1,i})
\]

where (a) follows from the fact that \( S_{2,i} \) is independent of \((S_{1,i}^{i-1}, S_{1,i+1}^{i}, S_{2,i}^{i-1})\) given \( S_{1,i} \), and the fact that \( S_{2,i} \) is independent of \((Y_{i-1}, S_{1,i}^{i-1})\) given \((T, S_{1,i}^{i-1}, S_{2,i}^{i-1})\).
\[ nR = H(W) \]
\[ \leq H(W|T) - H(W|T, Y^n, S^n_2) + n\epsilon_n \]
\[ = I(W; Y^n, S^n_2|T) + n\epsilon_n \]
\[ = \sum_{i=1}^{n} I(W; Y_i, S_{2,i}|T, Y^{i-1}, S^{i-1}_2) + n\epsilon_n \]
\[ \stackrel{(b)}{=} \sum_{i=1}^{n} \left[ I(W, S^n_{1,i+1}; Y_i, S_{2,i}|T, Y^{i-1}, S^{i-1}_2) - I(S^n_{1,i+1}; Y_i)|W, T, Y^{i-1}, S^{i-1}_2|T, Y^{i-1}, S^{i-1}_2) \right] + n\epsilon_n \]
\[ \stackrel{(c)}{=} \sum_{i=1}^{n} \left[ I(W, S^n_{1,i+1}; Y_i, S_{2,i}|T, Y^{i-1}, S^{i-1}_2) - I(S^n_{1,i}; Y^{i-1}, S^{i-1}_2|W, T, S^n_{1,i+1}) \right] + n\epsilon_n \]
\[ = \sum_{i=1}^{n} \left[ I(W, Y_i, S_{2,i}|T, Y^{i-1}, S^n_{1,i+1}) - I(S^n_{1,i}; W|T, Y^{i-1}, S^n_{1,i+1}) \right] + \Delta - \Delta^* + n\epsilon_n, \tag{46} \]

where
\[ \Delta = \sum_{i=1}^{n} I(S^n_{1,i+1}; Y_i, S_{2,i}|T, Y^{i-1}, S^{i-1}_2), \tag{47} \]
\[ \Delta^* = \sum_{i=1}^{n} I(S^n_{1,i}; Y^{i-1}, S^{i-1}_2|T, S^n_{1,i+1}), \tag{48} \]

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

By using the Csiszar sum on (47) and (48), we get
\[ \Delta = \Delta^*, \tag{49} \]

and therefore, from (45) and (46)
\[ \tilde{R}_s \geq \frac{1}{n} \sum_{i=1}^{n} I(S_{2,i}; S_{\tilde{2},i}|S_{1,i}) \tag{50} \]
\[ R - \epsilon_n \leq \frac{1}{n} \sum_{i=1}^{n} \left[ I(U_i; Y_i, S_{2,i}|S_{\tilde{2},i}) - I(U_i; S_{1,i}|S_{\tilde{2},i}) \right]. \tag{51} \]

Using the convexity of \( \tilde{R}_s \) and Jansen’s inequality, the standard time sharing argument for \( R \) and the fact that \( \epsilon_n \to 0 \) as \( n \to \infty \), we can conclude that
\[ \tilde{R}_s \geq I(S_{\tilde{2}}; S_2|S_1), \tag{52} \]
\[ R \leq I(U; Y, S_2|S_{\tilde{2}}) - I(U; S_1|S_{\tilde{2}}). \tag{53} \]

Notice that the Markov chains \( S_{\tilde{2}} - S_2 - S_1 \) and \( U - (S_1, S_{\tilde{2}}) - S_2 \) hold. This concludes the converse, and the proof of Theorem 1 Case 2.

**References**

[1] C. E. Shannon, “Channels with side information at the transmitter,” *IBM J. Res. Dev.*, vol. 2, no. 4, pp. 289–293, 1958.

[2] S. I. Gel'fand and M. S. Pinsker, “Coding for channel with random parameters,” *Problems of Control Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[3] C. Heegard and A. A. E. Gamal, “On the capacity of computer memory with defects,” *IEEE Transactions on Information Theory*, vol. 29, no. 5, pp. 731–739, 1983.

[4] Y. Steinberg, “Coding for channels with rate-limited side information at the decoder, with applications,” *Information Theory, IEEE Transactions on*, vol. 54, no. 9, pp. 4283–4295, sept. 2008.

[5] T. M. Cover and M. Chiang, “Duality between channel capacity and rate distortion with two-sided state information,” *CoRR*, vol. abs/cs/0508050, 2005.

[6] G. Keshet, Y. Steinberg, and N. Merhav, “Channel coding in the presence of side information,” *Found. Trends Commun. Inf. Theory*, vol. 4, no. 6, pp. 445–586, 2007.

[7] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *Information Theory, IEEE Transactions on*, vol. 22, no. 1, pp. 1–10, jan 1976.

[8] A. H. Kaspi, “Two-way source coding with a fidelity criterion,” *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 735–740, 1985.

[9] D. L. Neuhoff and R. K. Gilbert, “Causal source codes,” *IEEE Transactions on Information Theory*, vol. 28, no. 5, pp. 701–713, 1982.

[10] T. Weissman and N. Merhav, “On causal source codes with side information,” *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 4003–4013, 2005.

[11] T. Weissman and A. E. Gamal, “Source coding with limited-lookahead side information at the decoder,” *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5218–5239, 2006.

[12] C. E. Shannon, “Coding theorems for a discrete source with a fidelity criterion,” vol. 7, no. 1, pp. 142–163, Mar. 1959.

[13] R. J. Barron, B. Chen, and G. W. Wornell, “The duality between information embedding and source coding with side information and some applications,” *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1159–1180, 2003.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & sons, 1991.