Joint source-channel with side information coding error Exponents

Cheng Chang

HP Laboratories
HPL- 2009-18

Keyword(s):
joint source channel coding, side information, error exponent

Abstract:
In this paper, we study the upper and the lower bounds on the joint source-channel coding error exponent with decoder side-information. The results in the paper are non-trivial extensions of the Csiszár's classical paper [5]. Unlike the joint source-channel coding result in [5], it is not obvious whether the lower bound and the upper bound are equivalent even if the channel coding error exponent is known. For a class of channels, including the symmetric channels, we apply a game-theoretic result to establish the existence of a saddle point and hence prove that the lower and upper bounds are the same if the channel coding error exponent is known. More interestingly, we show that encoder side-information does not increase the error exponents in this case.

External Posting Date: February 6, 2009 [Fulltext] Approved for External Publication
Internal Posting Date: February 6, 2009 [Fulltext]

Submitted to IEEE Transactions on Information Theory, Allerton Conference 2009

Joint source-channel with side information coding error exponents

Cheng Chang

Abstract

In this paper, we study the upper and the lower bounds on the joint source-channel coding error exponent with decoder side-information. The results in the paper are non-trivial extensions of the Csiszár's classical paper [5]. Unlike the joint source-channel coding result in [5], it is not obvious whether the lower bound and the upper bound are equivalent even if the channel coding error exponent is known. For a class of channels, including the symmetric channels, we apply a game-theoretic result to establish the existence of a saddle point and hence prove that the lower and upper bounds are the same if the channel coding error exponent is known. More interestingly, we show that encoder side-information does not increase the error exponents in this case.

I. Introduction

In Shannon's very first paper on information theory [11], it is established that separate coding is optimal for memoryless source channel pairs. Reliable communication is possible if and only if the entropy of the source is lower than the capacity of the channel. However, the story is different when error exponent is considered. It is shown that joint source-channel coding achieves strictly better error exponent than separate[1] coding [5]. The key technical component of [5] is a channel coding scheme to protect different message sets with different channel coding error exponents. In this paper, we are concerned with the joint source-channel coding with side information problem as shown in Figure 1. For a special setup of Figure 1, where the discrete memoryless channel (DMC) is a noiseless channel with capacity[2] R, i.e. the source coding with side-information problem, the reliable reconstruction of a^n at the decoder is possible if and only if R is larger than the conditional entropy $H(P_{A|B})$ [13]. The error exponents of this problem is also studied in [8], [6] and more importantly in [1].

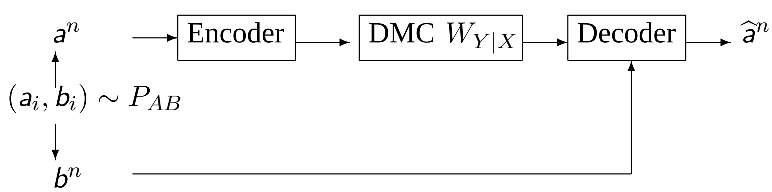


Fig. 1. Source coding with decoder side-information

The duality between source coding with decoder side-information and channel coding is established in the 80's [1]. This is an important result that all the channel coding error exponent bounds can be easily applied to source coding with side-information error exponent. The result is a consequence of the type covering lemma [6], also known as the Johnson-Stein-Lovász theorem [4]. With this duality result, we know that the error exponent of channel coding of channel $W_{Y|X}$ with channel code composition Q_X is

Cheng Chang is with Hewlett-Packard Labs, Palo Alto. Email: cheng.chang@hp.com

[1]In [5], Csiszár hand-wavily shows that the obvious separate coding scheme is suboptimal in terms achieving the best error exponent. The rather obvious result is rigidly proved in [14].

[2]In this paper, we use bits and $\log_2$, and R is always non-negative.

essentially the same problem as the error exponent of source coding with decoder side-information where the joint distribution is $Q_X \times W_{Y|X}$. Hence a natural question is what if we put these two dual problems together, what is the error exponent of joint source-channel coding with decoder side-information?

The more general case, where $W_{Y|X}$ is a noisy channel, is recently studied [15], [14]. It is shown that, not surprisingly, the reliable reconstruction of a^n is possible if and only if the channel capacity of the channel is larger than the conditional entropy of the source. A suboptimal error exponent based on a mixture scheme of separate coding and the joint source channel coding first developed in [5] is achieved. In this paper, we follow Csiszár's idea in [5] and develop a new coding scheme for joint source channel coding with decoder side-information. For a class of channels, including the symmetric channels, the resulted lower and upper bounds have the same property as the joint source-channel coding error exponent *without* side-information in [5]: they match if the channel coding error exponent is known at a critical rate. We use a game theoretic approach to interpret this result.

The outline of the paper is as follows. We review the problem setup and classical error exponent results in Section II. Then in Section III, we present the error exponent result for joint source-channel coding with both decoder and encoder side information which provides a simple upper bound to the error exponent investigated in the paper. This is a simple corollary of Theorem 5 in [5]. The main result of this paper is presented in Section IV. Some implications of these bounds are given in Section V.

II. Review of source and channel doing error exponents

In this paper random variables are denoted by a and b, the realizations of the random variables are denoted by a and b.

A. System model of joint source-channel coding with decoder side-information

As shown in Figure 1, the source and side-information, a^n and b^n respectively, are random variables i.i.d from distribution P_{AB} on a finite alphabet $\mathcal{A}\times\mathcal{B}$. The channel is memoryless with input/output probability transition $W_{Y|X}$, where the input/output alphabets $\mathcal{X}$ and $\mathcal{Y}$ are finite. Without loss of generality, we assume that the number of source symbols and the number of channel uses are equal, i.e. the encoder observes a^n and sends a codeword $x^n(a^n)$ of length n to the channel, the decoder observes the channel output y^n and side-information b^n which is not available to the encoder, the estimate is $\widehat{a}^n(b^n, y^n)$.

The error probability is the expectation of the decoding error average over all channel and source behaviors.

$$\Pr(\mathsf{a}^n \neq \widehat{\mathsf{a}}^n(\mathsf{b}^n, \mathsf{y}^n)) = \sum_{a^n, b^n} P_{AB}(a^n, b^n) \sum_{y^n} W_{Y|X}(y^n|x^n(a^n)) 1(a^n \neq \widehat{a}^n(b^n, y^n)). \tag{1}$$

The error exponent, for the optimal coding scheme, is defined as

$$E(P_{AB}, W_{Y|X}) = \lim \inf_{n\to\infty} -\frac{1}{n} \log \Pr(\mathsf{a}^n \neq \widehat{\mathsf{a}}^n(\mathsf{b}^n, \mathsf{y}^n)). \tag{2}$$

The main result of this paper is to establish both upper and lower bounds on $E(P_{AB}, W_{Y|X})$ and show the tightness of these bounds.

B. Classical error exponent results

[3] We review some classical results on channel coding error exponents and source coding with side-information error exponents. These bounds are investigated in [9], [6], [8] and [7].

[3]In this paper, we write the error exponents (both channel coding and source coding) in the style of Csiszár's method of types, equivalent Gallager style error exponents can be derived through the Fenchel duality.

1) Channel coding error exponents $E_c(R, W_{Y|X})$: Channel coding is a special case of joint source-channel coding with side-information: the source a and the side-information b are independent, i.e. $P_{AB} = P_A \times P_B$, and a is a uniform distributed random variable on $\{1, 2, ..., 2^R\}$. For the sake of simplicity, we assume that 2^R is an integer. This is not a problem if 2^R is not an integer since we can lump K symbols together and approximate 2^{KR} by an integer for some K, this is not a problem because $\lim_{K\to\infty} \frac{1}{K}\log_2(\lfloor 2^{KR}\rfloor) = R$. With this interpretation of channel coding, the definitions of error probability in (1) and error exponent in (2) still holds.

The channel coding error exponent $E_c(R, W_{Y|X})$ is lower bounded by the random coding error exponent and upper bounded by the sphere packing error exponent.

$$E_r(R, W_{Y|X}) \leq E_c(R, W_{Y|X}) \leq E_{sp}(R, W_{Y|X}) \tag{3}$$

$$\begin{aligned}
\text{where } E_r(R, W_{Y|X}) &= \max_{S_X} \inf_{V_{Y|X}} D(V_{Y|X}\|W_{Y|X}|S_X) + |I(V_{Y|X}; S_X) - R|^+ && (4)\\
&= \max_{S_X} E_r(R, S_X, W_{Y|X})\\
\text{and } E_{sp}(R, W_{Y|X}) &= \max_{S_X} \inf_{V_{Y|X}: I(V_{Y|X}; S_X) < R} D(V_{Y|X}\|W_{Y|X}|S_X) && (5)\\
&= \max_{S_X} E_{sp}(R, S_X, W_{Y|X})
\end{aligned}$$

Here S_X is the input composition (type) of the code words. $E_r(R, W_{Y|X}) = E_{sp}(R, W_{Y|X})$ in the high rate regime that $R > R_{cr}$ where R_{cr} is defined in [9] as the minimum rate for which the sphere packing $E_{sp}(R, W_{Y|X})$ and random coding error exponents $E_r(R, W_{Y|X})$ match for channel $W_{Y|X}$. There are tighter bounds on the channel coding error exponents $E_c(R, W_{Y|X})$ in the low rate regime for $R < R_{cr}$, known as straight-line lower bounds and expurgation upper bounds [9]. However, in this paper, we focus on the basic random coding and sphere packing bounds, as the main message can be effectively carried out.

It is well known [9] that both the random coding and the sphere-packing bounds are decreasing with R and are convex in R. And they are both positive if and only if $R < C(W_{Y|X})$, where $C(W_{Y|X})$ is the capacity of the channel $W_{Y|X}$.

2) Source coding with decoder side-information error exponents: This is also a special case of the general setup in Figure 1. This time the channel $W_{Y|X}$ is a noiseless channel with input-output alphabet $\mathcal{X} = \mathcal{Y}$ and $|\mathcal{X}| = 2^R$. Again, we can reasonably assume that 2^R is an integer.

The source coding with side-information error exponent[4] $e(R, P_{AB})$ can be bounded as follows:

$$e_L(R, P_{AB}) \leq e(R, P_{AB}) \leq e_U(R, P_{AB}) \tag{6}$$

$$\begin{aligned}
\text{where} \qquad e_L(R, P_{AB}) &= \inf_{Q_{AB}} D(Q_{AB}\|P_{AB}) + |R - H(Q_{A|B})|^+\\
e_U(R, P_{AB}) &= \inf_{Q_{AB}: H(Q_{A|B}) > R} D(Q_{AB}\|P_{AB}).
\end{aligned}$$

The duality between channel coding and source coding with decoder side information had been well understood [1]. We give the following duality results on error exponents. .

$$\begin{aligned}
e(R, Q_A, P_{B|A}) &= E_c(H(Q_A) - R, Q_A, P_{B|A})\\
\text{or equivalently : } e(H(Q_A) - R, Q_A, P_{B|A}) &= E_c(R, Q_A, P_{B|A})
\end{aligned}$$

[4]In this paper, if $R \geq \log_2|\mathcal{A}|$ for source coding with side-information error exponents, we let the error exponent be ∞.

where $E_c(R, Q_A, P_{B|A})$ is the channel coding error exponent for channel $P_{B|A}$ at rate R and the codebook composition is Q_A. $e(R, Q_A, P_{B|A})$ is the source coding with side information error exponent at rate R with source sequences uniformly distributed in type Q_A and the side information is the output of channel $P_{B|A}$ with input sequence of type Q_A. So obviously, we have:

$$E_c(R, P_{B|A}) = \max_{Q_A}\{E_c(R, Q_A, P_{B|A})\}$$
$$e(R, P_{AB}) = \min_{Q_A}\{D(Q_A\|P_A) + e(R, Q_A, P_{B|A})\}$$

These results are established by the type covering lemma [5] on the operational level, i.e. a complete characterizations of the source coding with side information error exponent $e(R, Q_A, P_{B|A})$ implies a complete characterizations of the channel coding error exponent $E_c(H(Q_A) - R, Q_A, P_{B|A})$ and vice versa.

From these duality results, it is well known that both the lower and the upper bounds are increasing with R and are convex in R. And they are both positive if and only if $R > H(P_{A|B})$. The special case of the source coding with decoder side information problem is that the side information is independent of the source, i.e. $P_{AB} = P_A \times P_B$. In this case, the error exponent is completely characterized [6],

$$e(R, P_A) = \inf_{Q_A : H(Q_A) > R} D(Q_A\|P_A) \tag{7}$$

3) Joint source-channel coding error exponents [5]: In his seminal paper [5], the joint source-channel coding error exponents is studied. This is yet another special case of the general setup in Figure 1. When a and b are independent, i.e. $P_{AB} = P_A \times P_B$, we can drop all the b terms in (1). Hence the error probability is defined as:

$$\Pr(a^n \neq \widehat{a}^n(y^n)) = \sum_{a^n} P_A(a^n) \sum_{y^n} W_{Y|X}(y^n|x^n(a^n))1(a^n \neq \widehat{a}^n(y^n)). \tag{8}$$

Write the error exponent of (8) as $E(P_A, W_{Y|X})$. The lower and upper bounds of the error exponents are derived in [5]. It is shown that:

$$\min_R\{e(R, P_A) + E_{sp}(R, W_{Y|X})\} \leq E(P_A, W_{Y|X}) \leq \min_R\{e(R, P_A) + E_r(R, W_{Y|X})\} \tag{9}$$

The upper bound is derived by using standard method of types argument. The lower bound is a direct consequence of the channel coding Theorem 5 in [5].

The difference between the lower and upper bounds is in the channel coding error exponent. The joint source channel coding error exponent is "almost" completely characterized because the only possible improvement is to determine the channel coding error exponent which is still not completely characterized in the low rate regime where $R < R_{cr}$. However, let R^* be the rate that minimizes $\{e(R, P_A) + E_r(R, W_{Y|X})\}$, if $R^* \geq R_{cr}$ or equivalently $E_r(R^*, W_{Y|X}) = E_{sp}(R^*, W_{Y|X})$, then we have a complete characterization of the joint source channel coding error exponent:

$$E(P_A, W_{Y|X}) = e(R^*, P_A) + E_r(R^*, W_{Y|X}). \tag{10}$$

The goal of this paper is to derive a similar result for $E(P_{AB}, W_{Y|X})$ defined in (2) as that for the joint source channel coding in (9) and (10).

4) A recite of Theorem 5 in [5]: Given a sequence of positive integers $\{m_n\}$ with $\frac{1}{n}\log m_n \to 0$ and m_n message sets $\mathcal{A}_1, ...\mathcal{A}_{m_n}$ each with size $|\mathcal{A}_i| = 2^{nR_i}$. Then there exists a channel code (f_0, ϕ_0), where the encoder $f_0 : \bigcup_{i=0}^{m_n} \mathcal{A}_i \to \mathcal{X}^n$ where $f_0(a) = x^n(a) \in S_X^i$ for $a \in \mathcal{A}_i$ and the decoder $\phi_0 : \mathcal{Y}^n \to \bigcup_{i=0}^{m_n} \mathcal{A}_i$, write $\phi_0(y^n)$ as $\widehat{a}(y^n)$ s.t. for any message $a \in \mathcal{A}_i$, the decoding error

$$p_e(a) = \sum_{y^n} W_{Y|X}(y^n|x^n(a))1(a \neq \widehat{a}(y^n)) \leq 2^{n(E_r(R_i, S_X^i, W_{Y|X}) - \epsilon_n)}$$

for every channel $W_{Y|X}$, and $\epsilon_n \to 0$. In particular, if the channel $W_{Y|X}$ is known to the encoder, each S_X^i can be picked to maximize $E_r(R_i, S_X^i, W_{Y|X})$, hence for each $a \in \mathcal{A}_i$:

$$p_e(a) \leq 2^{n(E_r(R_i, W_{Y|X}) - \epsilon_n)}.$$

This channel coding theorem as Csiszár put it, the “main result of this paper” in [5]. We use this theorem directly in the proof of the lower bound in Proposition 1 and further modify it to show the lower bound in Theorem 1.

III. Joint source-channel coding error exponent with both decoder and encoder side-information

As a warmup to the more interesting scenario where the side-information is not known to the encoder, we present the upper/lower bounds when both the encoder and the decoder know the side-information. This setup is shown in Figure 2.

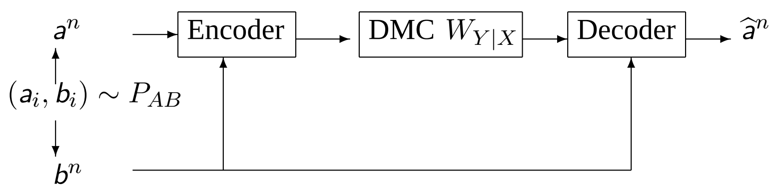


Fig. 2. Source coding with both decoder **and encoder** side-information

The error probability of the coding system is, similar to (1):

$$\Pr(a^n \neq \widehat{a}^n(b^n, y^n)) = \sum_{a^n, b^n} P_{AB}(a^n, b^n) \sum_{y^n} W_{Y|X}(y^n|x^n(a^n, b^n))1(a^n \neq \widehat{a}^n(b^n, y^n)). \tag{11}$$

The error exponent of this setup is denoted by $E_{both}(P_{AB}, W_{Y|X})$ which is defined in the same way as $E(P_{AB}, W_{Y|X})$ in (2). The difference is that the encoder observes both source a^n and the side-information b^n, hence the output of the encoder is a function of both: $x^n(a^n, b^n)$. So obviously, $E_{both}(P_{AB}, W_{Y|X})$ is not smaller than $E(P_{AB}, W_{Y|X})$.

Comparing (11) and (8), we can see the connections between joint source-channel coding with both decoder and encoder side information and joint source-channel coding. Knowing the side information b^n, the joint source channel coding with both encoder and decoder side information problem is essentially a channel coding problem with messages distributed on $\mathcal{A}^n$ with a distribution $P_{A|B}(a^n|b^n)$. Hence we can extend the results for joint source-channel coding error exponent [5]. We summarize the bounds on $E_{both}(P_{AB}, W_{Y|X})$ in the following proposition.

Proposition 1: Lower and upper bound on $E_{both}(P_{AB}, W_{Y|X})$

$$E_{both}(P_{AB}, W_{Y|X}) \leq \min_{R}\{e_U(R, P_{AB}) + E_{sp}(R, W_{Y|X})\}$$
$$E_{both}(P_{AB}, W_{Y|Z}) \geq \min_{R}\{e_U(R, P_{AB}) + E_r(R, W_{Y|X})\} \tag{12}$$

Not explicitly stated, but it should be clear that the range of R is $(0, \log_2 |\mathcal{A}|)$.

Proof: see Appendix A. Because $E_{both}(P_{AB}, W_{Y|X})$ is no smaller than $E(P_{AB}, W_{Y|X})$, so the lower bound of $E(P_{AB}, W_{Y|X})$ in Theorem 1 is also a lower bound for $E_{both}(P_{AB}, W_{Y|X})$. However, in the appendix, we give a simple proof of the lower bound on $E_{both}(P_{AB}, W_{Y|X})$ which is a corollary of Theorem 5 in [5]. □

Comparing the lower and the upper bounds for the case with both encoder and decoder side-information, we can easily see that if R^* minimizes $\{e_U(R, P_{AB}) + E_r(R, W_{Y|X})\}$ and $E_{sp}(R^*, W_{Y|X}) = E_r(R^*, W_{Y|X})$, then the upper bound and the lower bound match. Hence,

$$E_{both}(P_{AB}, W_{Y|X}) = e_U(R^*, P_{AB}) + E_r(R^*, W_{Y|X}). \tag{13}$$

In this case $E_{both}(P_{AB}, W_{Y|X})$ is completely characterized.

IV. Joint source-channel error exponents with only decoder side information

We study the more interesting problem where only decoder knows the side-information in this section. We first give a lower and an upper bound on the error exponent of joint source-channel coding with decoder only side-information. The result is summarized in the following Theorem.

Theorem 1: Lower and upper bound on the joint source channel coding with decoder side-information only, as setup in Figure 1, error exponent: For the error probability $\Pr(a^n \neq \widehat{a}^n(b^n, y^n))$ and error exponent $E(P_{AB}, W_{Y|X})$ defined in (1) and (2), we have the following lower and upper bounds:

$$E(P_{AB}, W_{Y|X}) \geq \tag{14}$$
$$\min_{Q_A} \max_{S_X(Q_A)} \min_{Q_{B|A}, V_{Y|X}} \{D(Q_{AB}\|P_{AB}) + D(V_{Y|X}\|W_{Y|X}|S_X(Q_A)) + |I(S_X(Q_A); V_{Y|X}) - H(Q_{A|B})|^+\}$$
$$E(P_{AB}, W_{Y|X}) \leq \tag{15}$$
$$\min_{Q_A} \max_{S_X(Q_A)} \min_{Q_{B|A}, V_{Y|X}: I(S_X(Q_A); V_{Y|X}) < H(Q_{A|B})} \{D(Q_{AB}\|P_{AB}) + D(V_{Y|X}\|W_{Y|X}|S_X(Q_A))\}$$

Proof: The main technical tool used here is the method of types. For the lower bound we propose a joint coding scheme for the joint source channel coding with side information problem. This scheme is a modification of the coding scheme first proposed in [5]. However, we cannot directly use the channel coding Theorem 5 in [5] because of the presence of the side information. In essence, we have to study a more complicated case using the method of types. Details see Appendix B. □

To simplify the expressions of the lower and upper bounds and later give a sufficient condition for these two bounds to match, we introduce the "digital interface" R and have the following corollary.

Corollary 1: upper and lower bounds on $E(P_{AB}, W_{Y|X})$ with "digital interface" R

$$E(P_{AB}, W_{Y|X}) \leq \min_{Q_A} \max_{S_X(Q_A)} \min_{R}\{e_U(R, P_{AB}, Q_A) + E_{sp}(R, S_X(Q_A), W_{Y|X})\} \tag{16}$$
$$E(P_{AB}, W_{Y|Z}) \geq \min_{Q_A} \max_{S_X(Q_A)} \min_{R}\{e_U(R, P_{AB}, Q_A) + E_r(R, S_X(Q_A), W_{Y|X})\} \tag{17}$$

where $E_r(R, S_X(Q_A), W_{Y|X})$ is the standard random coding error exponent for channel $W_{Y|X}$ at rate R with input distribution $S_X(Q_A)$ defined in (4), while $e_U(R, P_{AB}, Q_A)$ is a peculiar source coding with side-information error exponent for source P_{AB} at rate R, where the empirical source distribution is fixed at Q_A. That is for Q_A

$$e_U(R, P_{AB}, Q_A) \triangleq \min_{Q_{B|A}:H(Q_{A|B})\geq R} D(Q_{AB}\|P_{AB}) \tag{18}$$

Proof: The proof is in Appendix C. □

With the simplified expression of the lower and upper bounds in Corollary 1, we can give a game theoretic interpretation of the bounds. And more importantly, we present some sufficient conditions for the two bounds to match.

A. A game theoretic interpretation of the bounds

The lower and upper bounds established in Corollary 1 clearly have a game theoretic interpretation. This is a two player zero sum game. The first player is "nature", the second player is the coding system, the payoff from "nature" to the coding system is the bounds on the error exponents in Corollary 1. "Nature" chooses the marginal of the source Q_A (observable to the coding system) and R which is essentially the side information $Q_{B|A}$ and the channel behavior $V_{Y|X}$ (non-observable to the coding system). The coding system choose $S_X(Q_A)$ after observing Q_A. Hence in this game, the "nature" has two moves, the first move on Q_A and the last move on R which is essentially $Q_{B|A}$ and $V_{Y|X}$, while the coding system has the middle move on $S_X(Q_A)$.

Comparing Corollary 1 for joint source-channel coding with decoder side information and the classical joint source-channel coding error exponent [5] in (9), it is desirable to have a sufficient condition that the lower bound and the upper bound match, i.e. the complete characterization as that in (10). It is simpler for the case in (9) since all is needed is that the sphere backing bound and the random coding bound to match at the critical rate R^* as discussed in Section II-B.3. However, for the two bounds in Corollary 1, it is not clear what the conditions are such that these two bounds match. Suppose that the solution of the game (16) is $(Q_A^u, S_X^u(Q_A), R^u)$ and solution of the game (17) is $(Q_A^l, S_X^l(Q_A), R^l)$. An obvious sufficient condition for the two bounds match is as follows:

$$(Q_A^l, S_X^l(Q_A), R^l) = (Q_A^u, S_X^u(Q_A), R^u) \text{ and } E_r(R^u, S_X^u(Q_A), W_{Y|X}) = E_{sp}(R^u, S_X^u(Q_A), W_{Y|X}) \tag{19}$$

This condition is hard to verify for *any* source channel pairs. In the next section, we try to simplify the condition under which these two bounds match for a class of channels.

B. A sufficient condition to reduce $\min\{\max\{\min\{\cdot\}\}\}$ to $\min\{\cdot\}$

The difficulty in studying the bounds in Corollary 1 is that the $\min$ and $\max$ operators are nested. The problem will be simplified if we can change the order of the $\min$ and $\max$ operators.

Corollary 2: For symmetric channels $W_{Y|X}$ defined on Page 94 in [9], this includes the binary symmetric and binary erasure channels, where the input distribution S_X to maximize the random coding error exponent $E_r(R, S_X, W_{Y|X})$ is uniform on $\mathcal{X}$, or for more general channels[5], where the input distribution S_X to maximize the random coding error exponent $E_r(R, S_X, W_{Y|X})$ is the same for all R, then the upper and lower bounds in Theorem 1 and Corollary 1 can be further simplified to the following forms:

$$E(P_{AB}, W_{Y|X}) \leq \min_R\{e_U(R, P_{AB}) + E_{sp}(R, W_{Y|X})\} \tag{20}$$

$$E(P_{AB}, W_{Y|Z}) \geq \min_R\{e_U(R, P_{AB}) + E_r(R, W_{Y|X})\} \tag{21}$$

[5]For example, a channel consisted of parallel symmetric channels.

Note: in this case, the upper and lower bounds for $E(P_{AB}, W_{Y|X})$ is the same as those for $E_{both}(P_{AB}, W_{Y|X})$ in Proposition 1. More discussions see Section V.

Proof: An important property for symmetric channels is that the input distribution that maximizes the random coding error exponent is constant for all rate R, hence the inner $\max\min\{\cdot\}$ is equal to $\min\max\{\cdot\}$, i.e.

$$\begin{aligned}
E(P_{AB}, W_{Y|X}) &\geq \min_{Q_A} \max_{S_X(Q_A)} \min_R \{e_U(R, P_{AB}, Q_A) + E_r(R, S_X(Q_A), W_{Y|X})\} \\
&= \min_{Q_A} \min_R \max_{S_X(Q_A)} \{e_U(R, P_{AB}, Q_A) + E_r(R, S_X(Q_A), W_{Y|X})\} \\
&= \min_{Q_A} \min_R \{e_U(R, P_{AB}, Q_A) + E_r(R, W_{Y|X})\} \qquad (22) \\
&= \min_R \{\min_{Q_A} \{e_U(R, P_{AB}, Q_A)\} + E_r(R, W_{Y|X})\} \\
&= \min_R \{e_U(R, P_{AB}) + E_r(R, W_{Y|X})\} \qquad (23)
\end{aligned}$$

where (22) follows the definition of random coding bound in (3) and (23) follows the obvious equality:

$$\min_{Q_A} e_U(R, P_{AB}, Q_A) = \min_{Q_{AB}: H(Q_{A|B}) \geq R} D(Q_{AB} \| P_{AB}) = e_U(R, P_{AB}).$$

The upper bound in 20 is trivial by noticing that $\max\min\{\cdot\} \leq \min\max\{\cdot\}$ [2], hence:

$$\begin{aligned}
E(P_{AB}, W_{Y|X}) &\leq \min_{Q_A} \max_{S_X(Q_A)} \min_R \{e_U(R, P_{AB}, Q_A) + E_{sp}(R, S_X(Q_A), W_{Y|X})\} \\
&\leq \min_{Q_A} \min_R \max_{S_X(Q_A)} \{e_U(R, P_{AB}, Q_A) + E_{sp}(R, S_X(Q_A), W_{Y|X})\} \\
&= \min_{Q_A} \min_R \{e_U(R, P_{AB}, Q_A) + E_{sp}(R, W_{Y|X})\} \\
&= \min_R \{\min_{Q_A} \{e_U(R, P_{AB}, Q_A)\} + E_{sp}(R, W_{Y|X})\} \\
&= \min_R \{e_U(R, P_{AB}) + E_{sp}(R, W_{Y|X})\} \qquad (24)
\end{aligned}$$

Corollary 2 is proved. □

With this corollary proved, we can give a sufficient condition under which the lower bound and upper bound match similar to that for the joint source-channel coding case in Section II-B.3. More discussions see Section V.

C. Why it is hard to generalize Corollary 2 to non-symmetric channels?

Whether $\max_{S_X(Q_A)} \min_R \{e_U(R, P_{AB}, Q_A) + E_r(R, S_X(Q_A), W_{Y|X})\}$ is equal to $\min_R \max_{S_X(Q_A)} \{e_U(R, P_{AB}, Q_A) + E_r(R, S_X(Q_A), W_{Y|X})\}$ is not obvious for general (non-symmetric) channels. A sufficient condition of the existence of a unique saddle point hence the equality is known as the Sion's Theorem [12] which states that:

$$\max_{\mu \in \mathcal{M}} \min_{\nu \in \mathcal{N}} f(\mu, \nu) = \min_{\nu \in \mathcal{N}} \max_{\mu \in \mathcal{M}} f(\mu, \nu) \qquad (25)$$

if $\mathcal{M}$ and $\mathcal{N}$ are convex, compact spaces and f a quasi-concave-convex (definitions see [2]) and continuous function on $\mathcal{M} \times \mathcal{N}$. For the function of interest,:

$$\max_{S_X(Q_A)} \min_R \{e_U(R, P_{AB}, Q_A) + E_r(R, S_X(Q_A), W_{Y|X})\}. \qquad (26)$$

We examine the sufficient condition under which a unique equilibrium exists, according to the Sion's Theorem. First, $e_U(R, P_{AB}, Q_A)+E_{sp}(R, S_X(Q_A), W_{Y|X})$ is quasi-convex in R because both $e_U(R, P_{AB}, Q_A)$ and $E_{sp}(R, S_X(Q_A), W_{Y|X})$ are convex, hence quasi-convex in R. However, (26) is not quasi concave on $S_X(Q_A)$:

$$E_r(R, S_X(Q_A), W_{Y|X}) = \inf_{V_{Y|X}} D(V_{Y|X} \| W_{Y|X} | S_X(Q_A)) + |I(V_{Y|X}; S_X(Q_A)) - R|^+,$$

notice that the first term is linear in $S_X(Q_A)$, the second term is quasi-concave but not concave. But the sum of a linear function and a quasi-concave function might not be quasi-concave. This shows that the $\min\max$ theorem cannot be established by using the Sion's Theorem. This does not mean that the $\min\max$ theorem cannot be proved. However for a non quasi-concave function that may have multiple peaks, $\min\max\{\cdot\}$ is not necessarily equal to $\max\min\{\cdot\}$.

V. "Almost" complete characterization of $E(P_{AB}, W_{Y|X})$ for symmetric channels

The sufficient condition in Corollary 2 is important, since binary symmetric and binary erasure channels are among the most well studied discrete memoryless channels. We further discuss the implications of the "almost" complete characterization of $E(P_{AB}, W_{Y|X})$ for symmetric channels.

First we give an example shown in Figure 3 and Figure 4. The source a is a Bernoulli 0.5 random variable and the joint distribution has the distribution

$$P_{AB} = \left\{ \begin{array}{cc} 0.50 & 0.00 \\ 0.05 & 0.45 \end{array} \right\} \tag{27}$$

The channel $W_{Y|X}$ is a binary symmetric channel with cross rate 0.025. The channel coding error exponent bounds $E_r(R, W_{Y|X})$ and $E_{sp}(R, W_{Y|X})$ and the source coding with decoder side-information upper bound $e_U(R, P_{AB})$ are plotted in Figure 3. The channel coding bound match while $R \geq R_{cr}$, where R_{cr} is defined in [9].

Note: the lower bound of the source coding with side information error exponent $e_L(R, P_{AB})$ is not plotted in the figure.

In Figure 4, we add both the lower and upper bounds on the joint source channel coding with decoder side information to the plot in Figure 3. For this source channel pair P_{AB} and $W_{Y|X}$, we have a complete characterization of $E_{both}(P_{AB}, W_{Y|X})$ because the channel is symmetric and the two bounds match at the minimal point, i.e. the two curves: $e_U(R, P_{AB}) + E_{sp}(R, W_{Y|X})$ and $e_U(R, P_{AB}) + E_r(R, W_{Y|X})$ match at the minimal point as shown in Figure 4. The value of the minimum is E_j shown in Figure 4.

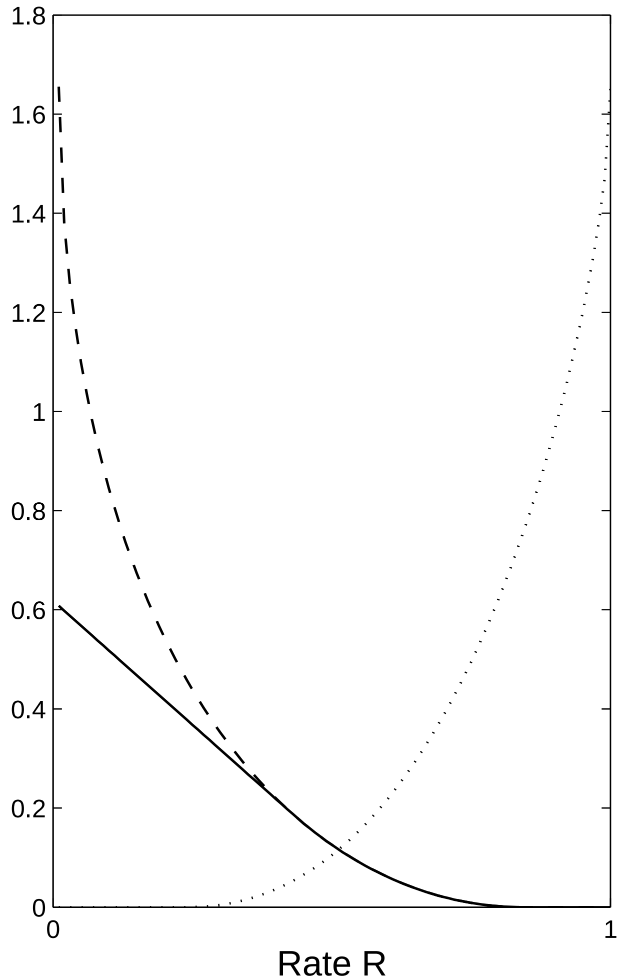
A. Encoder side information often does not help

Similar to Proposition 1, we can see the conditions under which we can give a complete characterization of the joint source channel coding with decoder only side information error exponent $E(P_{AB}, W_{Y|X})$. If R^* minimizes $\{e_U(R, P_{AB}) + E_r(R, W_{Y|X})\}$ and $E_{sp}(R^*, W_{Y|X}) = E_r(R^*, W_{Y|X})$, then the upper bound and the lower bound match. Hence:

$$E(P_{AB}, W_{Y|X}) = e_U(R^*, P_{AB}) + E_r(R^*, W_{Y|X}). \tag{28}$$

Comparing Corollary 2 and Proposition 1, we bound the error exponent with or without decoding side-information by the same lower and upper bounds. This does not mean that $E(P_{AB}, W_{Y|Z}) = E_{both}(P_{AB}, W_{Y|Z})$ always holds. But if the lower bound and upper bound match, which is shown in Figure 4, then we have:

$$E(P_{AB}, W_{Y|Z}) = E_{both}(P_{AB}, W_{Y|Z}) = e_U(R^*, P_{AB}) + E_r(R^*, W_{Y|X}). \tag{29}$$



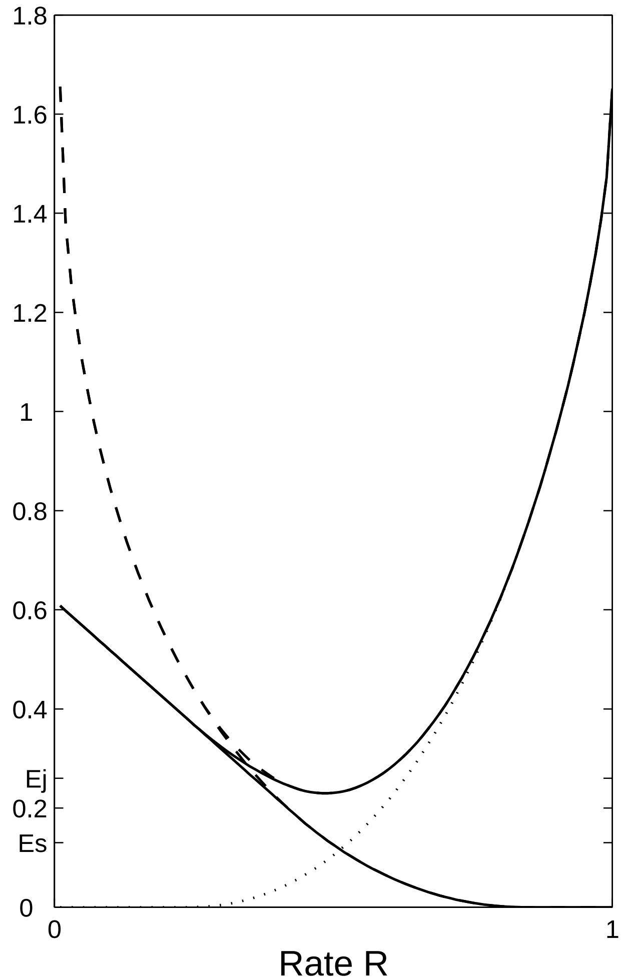


Fig. 3. The upper bound on source coding with side-information error exponent $e_U(R, P_{AB})$ is the dotted line. The random coding bound $E_r(R, W_{Y|X})$ and sphere packing bound $E_{sp}(R, W_{Y|X})$ for channel coding error exponents are the solid line and the dashed line respectively.

Fig. 4. $e_U(R, P_{AB})+E_{sp}(R, W_{Y|X})$ and $e_U(R, P_{AB})+E_r(R, W_{Y|X})$ are added to Figure 3 in dashed line and solid line respectively, they match at the minimal point hence the joint source-channel coding with decoder side-information error exponent is completely determined as $E(P_{AB}, W_{Y|X}) = E_j$ And E_s is the separate coding error exponent $E_{separate}(P_{AB}, W_{Y|X})$ defined in (33).

where R^* minimizes $e_U(R, P_{AB}) + E_r(R, W_{Y|X})$ and $R^* > R_{cr}$. This is another example for block coding where knowing side-information does not help increase the error exponent. In the contrary, as discussed in [3], in the delay constrained setup, there is a penalty for not knowing the side-information even if the channel is noiseless.

B. Separate coding is strictly sub-optimal

An obvious coding scheme for the problem in Figure 1 is to implement a separate coding scheme. A source encoder first encodes the source sequence a^n into a rate R, where R is determined later, bit stream $c^{nR}(a^n)$ then an independent channel encoder encodes the bits c^{nR} into channel inputs x^n. The channel decoder first decodes the channel output y^n into bits $\widehat{c}^{nR}$ and then the independent source decoder reconstructs $\widehat{a}^n$ from $\widehat{c}^{nR}$ and side information b^n. This is a separate coding scheme with outer source with side information coding and inner channel coding, both at rate R. If both coding are random coding that achieves the random coding error exponents for both source coding and channel coding respectively. The union bound of the error probability is as follows:

$$\begin{aligned}
\Pr(a^n \neq \widehat{a}^n(b^n, y^n)) &= \Pr(c^{nR} \neq \widehat{c}^{nR}(y^n)) + \Pr(a^n \neq \widehat{a}(\widehat{c}^{nR}(y^n), b^n), c^{nR} = c^{nR}(y^n)) && (30)\\
&\leq \Pr(c^{nR} \neq \widehat{c}^{nR}(y^n)) + \Pr(a^n \neq \widehat{a}(\widehat{c}^{nR}(y^n), b^n) | c^{nR} = c^{nR}(y^n)) && (31)\\
&\leq 2^{-n(E_r(R, W_{Y|X}) - \epsilon_n^1)} + 2^{-n(e_L(R, P_{AB}) - \epsilon_n^2)} && (32)
\end{aligned}$$

where ϵ_n^1 and ϵ_n^2 converges to zero as n goes to infinity. (30) follows the union bound argument that a decoding error occurs if either the inner channel coding fails or the outer source coding fails. (31) is true because conditional probability is large or equal to joint probability. Finally (32) is true because both the outer source coding and inner channel coding achieve the random coding error exponents. From (32) and that we can optimize the digital interface rate R between the channel coder and source coder, we know that the separate coding error exponent is

$$\max_R \{\min\{E_r(R, W_{Y|X}), e_L(R, P_{AB}\}\} \triangleq E_{separate}(P_{AB}, W_{Y|X}) \quad (33)$$

This separate coding scheme is also discussed for joint source channel coding in [5]. A similar bound is drawn. We next show why the separate coding error exponent $E_{separate}(P_{AB}, W_{Y|X})$ is in general strictly smaller than the lower bound of $E(P_{AB}, W_{Y|X})$ in (21).

First, obviously, $E_{separate}(P_{AB}, W_{Y|X}) \leq \max_R \{\min\{E_r(R, W_{Y|X}), e_U(R, P_{AB})\}\}$. Secondly $\{E_r(R, W_{Y|X})$ is monotonically decreasing, $e_U(R, P_{AB})$ is monotonically increasing, and both are continuous and convex as shown in Figure 4. This means that for rate $\bar{R}$ such that $E_r(\bar{R}, W_{Y|X}) = e_U(\bar{R}, P_{AB})$:

$$E_{separate}(P_{AB}, W_{Y|X}) = E_r(\bar{R}, W_{Y|X}) = e_U(\bar{R}, P_{AB})$$

Now let R^* be the rate to minimize $\{e_U(R, P_{AB}) + E_r(R, W_{Y|X})\}$, i.e.

$$E(P_{AB}, W_{Y|X}) \geq e_U(R^*, P_{AB}) + E_r(R^*, W_{Y|X}).$$

There are three scenarios. First if $R^* = \bar{R}$, then

$$E(P_{AB}, W_{Y|X}) \geq e_U(R^*, P_{AB}) + E_r(R^*, W_{Y|X}) = 2E_r(\bar{R}, W_{Y|X}) = 2E_{separate}(P_{AB}, W_{Y|X}).$$

Secondly, if $R^* < \bar{R}$,

$$E(P_{AB}, W_{Y|X}) \geq E_r(R^*, W_{Y|X}) > E_r(\bar{R}, W_{Y|X}) = E_{separate}(P_{AB}, W_{Y|X}).$$

Finally if $R^* > \bar{R}$,

$$E(P_{AB}, W_{Y|X}) \geq e_U(R^*, P_{AB}) > e_U(\bar{R}, P_{AB}) = E_{separate}(P_{AB}, W_{Y|X}).$$

So in all cases, the joint source channel coding error exponent $E(P_{AB}, W_{Y|X})$ is strictly larger than the separate coding error exponent $E_{separate}(P_{AB}, W_{Y|X})$. This is clearly illustrated in Figure 4.

Note: $E_{separate}(P_{AB}, W_{Y|X})$ is an achievable separate coding error exponent from the obvious separate coding scheme. What we prove is that this obvious one is strictly smaller than the joint source-channel coding error exponent. This is similar to the claim Csiszár makes in [5]. It should be clear that the upper bound of any separate source channel coding error exponent is $\max_R\{\min\{E_{sp}(R, W_{Y|X}), e_U(R, P_{AB}\}\}$ which is comparable to (33). The proof hinges on the complete transparency between the source coding and channel coding, otherwise we have a joint coding schemes. A detailed discussion is in [14].

ACKNOWLEDGEMENT

The author would like to thank Neri Merhav and Filippo Balestrieri for pointing out the Sion's theorem in [12], and Fady Alajaji for many helpful discussions on the technical results of this paper.

VI. CONCLUSIONS

We study the joint source channel coding with decoder side-information problem, with or without encoder side-information. This is an extension of Csiszár's joint source channel coding error exponent problem in [5]. To derive the lower bound, we use a novel joint source channel with decoder side-information decoding scheme. We further investigate the conditions under which the lower bounds and upper bounds match. A game theoretic approach is applied to show the equivalence of the lower and upper bound. This approach might be useful in simplifying other error exponents with a cascade of min-max operators , for example, the Wyner-Ziv coding error exponent recently studied in [10].

APPENDIX

A. Proof of upper and lower bounds on $E_{both}(P_{AB}, W_{Y|X})$

We prove Proposition 1 in this section. The upper bound and lower bounds are simple corollaries of the method of types and Theorem 5 in [5] respectively.

1) Upper bound: Consider a distribution Q_{AB}, the joint source channel encoder observes the realization of the source (a^n, b^n) with type Q_{AB}, for the case where the decoder knows the side-information b^n. There are[6] $2^{n(H(Q_{A|B})-\epsilon_n^1)}$ many equally likely sequences $\in \mathcal{A}^n$ conditional on b^n. These are the sequences with the same joint probability with b^n as the sequence a^n. Even knowing the joint type Q_{AB} (given by a genie) and the side-information b^n, the decoder needs to guess the correct one from the channel output y^n. This is a channel coding problem with rate $H(Q_{A|B}) - \epsilon_n^1$.

Now consider the channel input $x^n(a^n, b^n)$ where b^n is the side-information, notice that there are at most $(n+1)^{|\mathcal{X}|}$ many different input types, there is a type $S_X(Q_{AB})$, such that more than $(n+1)^{-|\mathcal{X}|} = 2^{-n\epsilon_n^2}$ fraction of the channel inputs given side-information b^n and the joint type of (a^n, b^n) being Q_{AB} have type $S_X(Q_{AB})$. For a channel $V_{Y|X}$, such that the channel capacity of the channel given the input distribution S_X is smaller than $H(Q_{A|B})$, i.e.

$$I(S_X(Q_{AB}); V_{Y|X}) < H(Q_{A|B}),$$

then if the channel $W_{Y|X}$ behaves like $V_{Y|X}$ with the code book with type $S_X(Q_{AB})$, with high probably, the decoder cannot correctly decide from one of the $2^{nH(Q_{A|B})}$ sequences. This is guaranteed by the Blowing up Lemma [6] or see a detailed proof in [7].

The probability that both the source behaves like $Q_{A|B}$ and the channel behaves like $V_{Y|X}$ is

$$2^{-n(D(Q_{AB}\|P_{AB})+D(V_{Y|X}\|W_{Y|X}|S_X(Q_{AB}))-\epsilon_n^3)}. \quad (34)$$

Notice that the source behavior Q_{AB} and the channel behavior $V_{Y|X}$ are arbitrary, as long as $H(Q_{A|B}) >$

[6] Here ϵ_n^i goes to zero as n goes to infinity, $i = 1, 2, 3$.

$I(S_X(Q_{AB}); V_{Y|X})$, we can upper bound the error exponent as follows:

$$
\begin{aligned}
& E_{both}(P_{AB}, W_{Y|Z}) \\
\leq\ & \min_{Q_{AB}, V_{Y|Z}: H(Q_{A|B}) > I(S_X(Q_{AB}); V_{Y|X})} \{D(Q_{AB}\|P_{AB}) + D(V_{Y|X}\|W_{Y|X}|S_X(Q_{AB}))\} && (35)\\
=\ & \min_R \{ \min_{Q_{AB}, V_{Y|Z}: H(Q_{A|B}) > R > I(S_X(Q_{AB}); V_{Y|X})} D(Q_{AB}\|P_{AB}) + D(V_{Y|X}\|W_{Y|X}|S_X(Q_{AB}))\} && (36)\\
=\ & \min_R \{ \min_{Q_{AB}: H(Q_{A|B}) > R} \{D(Q_{AB}\|P_{AB}) + \min_{V_{Y|Z}: R > I(S_X(Q_{AB}); V_{Y|X})} D(V_{Y|X}\|W_{Y|X}|S_X(Q_{AB}))\}\} && (37)\\
\leq\ & \min_R \{ \min_{Q_{AB}: H(Q_{A|B}) > R} \{D(Q_{AB}\|P_{AB}) + E_{sp}(R, W_{Y|X})\}\} && (38)\\
=\ & \min_R \{e_U(R, P_{AB}) + E_{sp}(R, W_{Y|X})\} && (39)
\end{aligned}
$$

(35) is a direct consequence of (34). In (36), we introduce the "digital interface" R, the equivalence in (36) and (37) should be obvious. (38) and (39) are by definitions of the channel coding and source coding error exponents. □

2) Lower bound: Given a side-information sequence b^n which is known to both the encoder and the decoder. We partition the source sequence set $\mathcal{A}^n$ based on their joint type with b^n. The number of joint types $m_n \leq (n+1)^{|\mathcal{A}||\mathcal{B}|}$ and denote by Q^i_{AB}, $i = 1, 2, ...m_n$ the joint types. It should be clear that the Q^i_{AB}'s here all have the same marginal distribution as b^n.

$$\text{Let } \mathcal{A}_i(b^n) = \{a^n : (a^n, b^n) \in Q^i_{AB}\}, \quad i = 1, 2, ...m_n.$$

Obviously, $\mathcal{A}_i$'s form a partition of $\mathcal{A}^n$. And each set has size $|A_i(b^n)| \leq 2^{nH(Q^i_{A|B})}$. Now we can apply Theorem 5 of [5] as recited earlier: there exists a channel code f_0, ϕ_0, such that for each $a^n \in \mathcal{A}_i(b^n)$, i.e. $(a^n, b^n) \in Q^i_{AB}$:

$$p_{e,b^n}(a^n) = \sum_{y^n} W_{Y|X}(y^n|x^n(a^n, b^n)) 1(a^n \neq \widehat{a}^n(b^n, y^n)) \leq 2^{-n(E_r(H(Q^i_{A|B}), W_{Y|X}) - \epsilon_n)}. \qquad (40)$$

The joint source channel coding error probability is hence:

$$
\begin{aligned}
\Pr(\mathsf{a}^n \neq \widehat{\mathsf{a}}^n(\mathsf{b}^n, \mathsf{y}^n)) &= \sum_{a^n, b^n} P_{AB}(a^n, b^n) \sum_{y^n} W_{Y|X}(y^n|x^n(a^n)) 1(a^n \neq \widehat{a}^n(b^n, y^n)) \\
&= \sum_{Q_{AB}} \sum_{(a^n, b^n) \in Q_{AB}} P_{AB}(a^n, b^n) \sum_{y^n} W_{Y|X}(y^n|x^n(a^n, b^n)) 1(a^n \neq \widehat{a}^n(b^n, y^n)) \\
&\leq \sum_{Q_{AB}} \sum_{(a^n, b^n) \in Q_{AB}} P_{AB}(a^n, b^n) 2^{-n(E_r(H(Q_{A|B}), W_{Y|X}) - \epsilon_n)} && (41)\\
&\leq \sum_{Q_{AB}} 2^{-nD(Q_{AB}\|P_{AB})} 2^{-n(E_r(H(Q_{A|B}), W_{Y|X}) - \epsilon_n)} \\
&\leq (n+1)^{|\mathcal{A}||\mathcal{B}|} \max_{Q_{AB}} \{2^{-nD(Q_{AB}\|P_{AB})} 2^{-n(E_r(H(Q_{A|B}), W_{Y|X}) - \epsilon_n)}\} \\
&\leq 2^{-n(\min_{Q_{AB}} \{D(Q_{AB}\|P_{AB}) + E_r(H(Q_{A|B}), W_{Y|X})\} - \epsilon'_n)} && (42)
\end{aligned}
$$

(41) follows by substituting in (40) and the rest inequalities are by method of types. $\epsilon'_n \to 0$, so we can

lower bound the error exponent as

$$\begin{aligned}
E_{both}(P_{AB}, W_{Y|Z}) &\geq \min_{Q_{AB}} \{D(Q_{AB}\|P_{AB}) + E_r(H(Q_{A|B}), W_{Y|X})\} & (43)\\
&= \min_{R}\{\min_{Q_{AB}:H(Q_{A|B})=R}\{D(Q_{AB}\|P_{AB}) + E_r(H(Q_{A|B}), W_{Y|X})\}\} & (44)\\
&= \min_{R}\{\min_{Q_{AB}:H(Q_{A|B})=R}\{D(Q_{AB}\|P_{AB})\} + E_r(R, W_{Y|X})\} & (45)\\
&= \min_{R\geq H(P_{A|B})}\{\min_{Q_{AB}:H(Q_{A|B})=R}\{D(Q_{AB}\|P_{AB}) + E_r(R, W_{Y|X})\}\} & (46)\\
&= \min_{R\geq H(P_{A|B})}\{\min_{Q_{AB}:H(Q_{A|B})\geq R}\{D(Q_{AB}\|P_{AB}) + E_r(R, W_{Y|X})\}\} & (47)\\
&= \min_{R\geq H(P_{A|B})}\{e_U(R, P_{AB}) + E_r(R, W_{Y|X})\} & (48)\\
&= \min_{R}\{e_U(R, P_{AB}) + E_r(R, W_{Y|X})\} & (49)
\end{aligned}$$

(43) is a direct consequence of (42), in (44) we again introduce the "digital interface" variable R. (45) and (48) are by definitions of $E_r(R, W_{Y|X})$ and $e_U(R, P_{AB})$ respectively. (46) is true because $E_r(R, W_{Y|X})$ is monotonically increasing with R and for $R < H(P_{B|A})$,

$$\min_{Q_{AB}:H(Q_{A|B})=R} D(Q_{AB}\|P_{AB}) \geq 0 = \min_{Q_{AB}:H(Q_{A|B})=H(P_{A|B})} D(Q_{AB}\|P_{AB}).$$

(47) is true because $D(Q_{AB}\|P_{AB})$ is convex in Q_{AB} and the global minimum is $Q^*_{AB} = P_{AB}$, but $H(Q^*_{A|B}) = H(P_{A|B}) \geq R$ which means the minimum point is on the boundary. Lastly (49) is because for $R < H(P_{A|B})$, $e_U(R, P_{AB})$ is constant at 0, while $E_r(R, W_{Y|X})$ is monotonically increasing with R. □

B. Lower and upper bounds on $E(P_{AB}, W_{Y|X})$

We give the proof of Theorem 1 here.

1) Lower bound: From the definition of the error exponent, we need to find a encoding rule $x : \mathcal{A}^n \to \mathcal{X}^n$ and decoding rule $\widehat{a} : \mathcal{B}^n \times \mathcal{Y}^n \to \mathcal{X}^n$such chat the error probability :

$$\Pr(a^n \neq \widehat{a}^n(b^n, y^n)) = \sum_{a^n,b^n} P_{AB}(a^n, b^n) \sum_{y^n} W_{Y|X}(y^n|x^n(a^n))1(a^n \neq \widehat{a}^n(b^n, y^n)) \tag{50}$$

is upper bounded by $2^{n(E-\epsilon_n)}$ where $\epsilon_n \to 0$, where E is the right hand side of (14).

We first describe the encoder and decoder, then prove that this coding system achieves the lower bound.

The encoder only observes the source sequence a^n. For all those sequences a^n with type Q_A, the channel input is $x^n(a^n)$ that has type $S_X(Q_A)$, i.e. the channel input type only depends on the type of the source, where $S_X(Q_A)$ is the distribution to maximize the following exponent:

$$\min_{Q_{B|A}, V_{Y|X}} \{D(Q_{AB}\|P_{AB}) + D(V_{Y|X}\|W_{Y|X}|S_X(Q_A)) + |I(S_X(Q_A); V_{Y|X}) - H(Q_{A|B})|^+\}.$$

The decoder observes both the side-information b^n and the channel output y^n, the decoder takes both the conditional entropy and mutual information across the channel into account:

$$\widehat{a}^n(b^n, y^n) = \arg\max_{a^n} I(x^n(a^n); y^n) - H(a^n|b^n) \tag{51}$$

We next need to show that there exists such a encoder/decoder pair that achieve the error exponent in (14). We also use the method of random selection of codebooks. We denote by $\mathcal{C}$ the set of the codebooks such that the codewords for $a^n \in Q_A$ all have composition $S_X(Q_A)$. Obviously $\mathcal{C}$ is finite, we let ζ be the random variable uniformly distributed on $\mathcal{C}$. We use codebook c if $\zeta = c$, i.e. we use the codebooks with equal probability. The most important property of this codebook distribution is the point-wise independence of the codewords, for all $a^n \in Q_A$ and $\tilde{a}^n \in \tilde{Q}_A$, for any two valid codewords $s^n \in S_X(Q_A)$ and $\tilde{s}^n \in S_X(\tilde{Q}_A)$:

$$\overset{\zeta}{\Pr}(x^n(a^n) = s^n, x^n(\tilde{a}^n) = \tilde{s}^n) = \overset{\zeta}{\Pr}(x^n(a^n) = s^n) \overset{\zeta}{\Pr}(x^n(\tilde{a}^n) = \tilde{s}^n) = \frac{1}{|S_X(Q_A)|} \frac{1}{|S_X(\tilde{Q}_A)|} \tag{52}$$

We calculate the average error probability on the whole codebook set $\mathcal{C}$. Write the average error probability as p_e^n, then first we have:

$$p_e^n = E(\overset{\zeta}{\Pr}(a^n \neq \widehat{a}^n(b^n, y^n)) = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \overset{c}{\Pr}(a^n \neq \widehat{a}^n(b^n, y^n)), \tag{53}$$

where $E(\overset{\zeta}{\Pr}(a^n \neq \widehat{a}^n(b^n, y^n))$ is the expected error probability over all codebooks under the codebook distribution ζ.

For a fixed codebook $c \in \mathcal{C}$

$$\begin{aligned}
& \overset{c}{\Pr}(a^n \neq \widehat{a}^n(b^n, y^n)) \\
= & \sum_{a^n, b^n} P_{AB}(a^n, b^n) \overset{c}{\Pr}(a^n \neq \widehat{a}^n(b^n, y^n)) \\
= & \sum_{Q_{AB}} \sum_{(a^n, b^n) \in Q_{AB}} \left(P_{AB}(a^n, b^n) \overset{c}{\Pr}(a^n \neq \widehat{a}^n(b^n, y^n)) \right) \\
= & \sum_{Q_{AB}} \sum_{(a^n, b^n) \in Q_{AB}} \left(P_{AB}(a^n, b^n) \sum_{y^n} W_{Y|X}(y^n | x^n(a^n)) 1^c(a^n \neq \widehat{a}^n(b^n, y^n)) \right) \\
= & \sum_{Q_{AB}} \sum_{(a^n, b^n) \in Q_{AB}} \\
& \left(P_{AB}(a^n, b^n) \sum_{V_{Y|X}} \sum_{y^n : (x^n(a^n), y^n) \in S_X(Q_A) \times V_{Y|X}} W_{Y|X}(y^n | x^n(a^n)) 1^c(a^n \neq \widehat{a}^n(b^n, y^n)) \right)
\end{aligned} \tag{54}$$

For $(a^n, b^n) \in Q_{AB}$, so the source sequence a^n has marginal distribution Q_A, from the codebook generation we know that the codeword $x^n(a^n) \in S_X(Q_A)$. For side-information $b^n \in \mathcal{B}^n$, we partition $\mathcal{A}^n$ according to the joint type with b^n:

$$Q_{\tilde{A}B}(b^n) = \{\tilde{a}^n \in \mathcal{A}^n : (\tilde{a}^n, b^n) \in Q_{\tilde{A}B}\}.$$

We partition $S_X(Q_{\tilde{A}})$ according to the joint distribution with y^n. For a joint distribution U_{XY} s.t. $U_X = S_X(Q_{\tilde{A}})$ and $y^n \in U_Y$:

$$U_{XY}(Q_{\tilde{A}}, y^n) = \{x^n \in S_X(Q_{\tilde{A}}) : (x^n, y^n) \in U_{XY}\}.$$

For $(a^n, b^n) \in Q_{AB}$ and channel output $y^n \in \mathcal{Y}^n$, s.t. $(x^n(a^n), y^n) \in V_{Y|X}$, a decoding error is made if there exists a source sequence $\tilde{a}^n \neq a^n$,s.t. $\tilde{a}^n \in Q_{\tilde{A}B}(b^n)$ where $Q_{\tilde{A}B}$ may or may not be Q_{AB} and the code word $x^n(\tilde{a}^n) \in U_{XY}(y^n, Q_{\tilde{A}})$, where $U_X = S_X(Q_{\tilde{A}})$ and $y^n \in U_Y$:

$$\begin{aligned}
& I(x^n(\tilde{a}^n); y^n) - H(\tilde{a}^n | b^n) \geq I(x^n(a^n); y^n) - H(a^n | b^n) \\
i.e. \quad & I_{U_{XY}}(X; Y) - H(Q_{\tilde{A}|B}) \geq I(S_X(Q_A); V_{Y|X}) - H(Q_{A|B})
\end{aligned} \tag{55}$$

Now we can expand the indicator function in (54) as follows, for a codebook c:

$$
\begin{aligned}
&1^c(a^n \neq \widehat{a}^n(b^n, y^n)) \\
&= 1^c\left(\exists \tilde{a}^n \neq a^n, s.t. I(x^n(\tilde{a}^n), y^n) - H(Q_{\tilde{A}|B}) \geq I(S_X(Q_A); V_{Y|X}) - H(Q_{A|B})\right) \\
&\leq \min\{1, \sum_{Q_{\tilde{A}B},\ U_{XY}: I_{U_{XY}}(X,Y) - H(Q_{\tilde{A}|B}) \geq I(S_X(Q_A); V_{Y|X}) - H(Q_{A|B})} 1^c(\exists \tilde{a}^n \neq a^n \text{ and } \tilde{a}^n \in Q_{\tilde{A}B}(b^n), \\
&\quad s.t.\ x^n(\tilde{a}^n) \in U_{XY}(Q_{\tilde{A}}, y^n))\} \qquad (56)
\end{aligned}
$$

Under the uniform codebook distribution ζ, for $\tilde{a}^n \neq a^n$, $x^n(\tilde{a}^n)$ is uniformly distributed in $S_X(Q_{\tilde{A}})$ independent of $x^n(a^n)$, so for all $Q_{\tilde{A}B}$ and U_{XY} with the proper marginals($b^n \in Q_B$, $U_X = S_X(Q_{\tilde{A}})$ and $y^n \in U_Y$) and satisfying (55):

$$
\begin{aligned}
& E(1(\exists \tilde{a}^n \neq a^n \text{ and } \tilde{a}^n \in Q_{\tilde{A}B}(b^n), s.t.\ x^n(\tilde{a}^n) \in U_{XY}(Q_{\tilde{A}}, y^n))) \\
=\ & \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} 1^c(\exists \tilde{a}^n \neq a^n \text{ and } \tilde{a}^n \in Q_{\tilde{A}B}(b^n), s.t.\ x^n(\tilde{a}^n) \in U_{XY}(Q_{\tilde{A}}, y^n)) \\
=\ & \overset{\zeta}{\Pr}(\exists \tilde{a}^n \neq a^n \text{ and } \tilde{a}^n \in Q_{\tilde{A}B}(b^n), s.t.\ x^n(\tilde{a}^n) \in U_{XY}(Q_{\tilde{A}}, y^n)) \\
\leq\ & |Q_{\tilde{A}B}(b^n)| \overset{\zeta}{\Pr}(\ x^n(\tilde{a}^n) \in U_{XY}(Q_{\tilde{A}}, y^n) | \tilde{a}^n \neq a^n \text{ and } \tilde{a}^n \in Q_{\tilde{A}B}(b^n)) \qquad (57) \\
=\ & |Q_{\tilde{A}B}(b^n)| \frac{|U_{XY}(Q_{\tilde{A}}, y^n)|}{|S_X(Q_{\tilde{A}})|} \qquad (58) \\
\leq\ & 2^{n\epsilon_n} 2^{nH(Q_{\tilde{A}|B})} \frac{2^{nH(U_{X|Y})}}{2^{nH(U_X)}} \qquad (59) \\
=\ & 2^{-n(I_{U_{XY}}(X,Y) - H(Q_{\tilde{A}|B}) - \epsilon_n)} \\
\leq\ & 2^{-n(I(S_X(Q_A); V_{Y|X}) - H(Q_{A|B}) - \epsilon_n)} \qquad (60)
\end{aligned}
$$

where $\epsilon_n \to 0$. (57) is by a union bound argument. (58) is true because the codeword $x^n(\tilde{a}^n)$ is uniformly distributed in $S_X(Q_{\tilde{A}})$. (59) is by the method of types. (60) is true because the condition in (55) is satisfied.

Combining (56) and (60) and noticing that the numbers of types of U_{XY} and $Q_{\tilde{A}B}$ are polynomials of n, hence sub-exponential, we have:

$$
\begin{aligned}
E(1(a^n \neq \widehat{a}^n(b^n, y^n))) &\leq \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} 1^c(a^n \neq \widehat{a}^n(b^n, y^n)) \\
&\leq \min\{1, 2^{-n(I(S_X(Q_A); V_{Y|X}) - H(Q_{A|B}) - \epsilon_n^1)}\} \\
&= 2^{-n|I(S_X(Q_A); V_{Y|X}) - H(Q_{A|B}) - \epsilon_n^1|^+} \qquad (61)
\end{aligned}
$$

Finally, we substitute (61) and (54) into (53). Notice that the number of types of $V_{Y|X}$ and Q_{AB} are polynomials in n and the usual method of types argument (upper bounding the probability of

$P_{AB}(a^n, b^n) \in Q_{AB})$ etc.), we have:

$$\begin{aligned} p_e^n &= \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}} \Pr^c(a^n \neq \widehat{a}^n(b^n, y^n)) \\ &\leq \sum_{Q_{AB}, V_{Y|X}} 2^{-n(D(Q_{AB}\|P_{AB})+D(V_{Y|X}\|W_{Y|X}|S_X(Q_A))+|I(S_X(Q_A);V_{Y|X})-H(Q_{A|B})-\epsilon_n^1|^+-\epsilon_n^2)} \\ &\leq \sum_{Q_{AB}, V_{Y|X}} 2^{-n(D(Q_{AB}\|P_{AB})+D(V_{Y|X}\|W_{Y|X}|S_X(Q_A))+|I(S_X(Q_A);V_{Y|X})-H(Q_{A|B})|^+-\epsilon_n^1-\epsilon_n^2)} \\ &\leq 2^{-n(\min_{Q_{AB},V_{Y|X}}\{D(Q_{AB}\|P_{AB})+D(V_{Y|X}\|W_{Y|X}|S_X(Q_A))+|I(S_X(Q_A);V_{Y|X})-H(Q_{A|B})|^+\}-\epsilon_n^3)} \end{aligned}$$

where $\epsilon_n^i \to 0$ for $i = 1, 2, 3$. Notice that p_e^n is the average error probability of the codebook set $\mathcal{C}$, so there exists at least a codebook c, such that the error probability is no bigger than p_e^n.

Now we lower bound the achievable error exponent by

$$\begin{aligned} &\min_{Q_{AB},V_{Y|X}} \{D(Q_{AB}\|P_{AB}) + D(V_{Y|X}\|W_{Y|X}|S_X(Q_A)) + |I(S_X(Q_A); V_{Y|X}) - H(Q_{A|B})|^+\} \\ =& \min_{Q_A} \min_{Q_{B|A},V_{Y|X}} \{D(Q_{AB}\|P_{AB}) + D(V_{Y|X}\|W_{Y|X}|S_X(Q_A)) + |I(S_X(Q_A); V_{Y|X}) - H(Q_{A|B})|^+\} \\ =& \min_{Q_A} \max_{S_X(Q_A)} \min_{Q_{B|A},V_{Y|X}} \{D(Q_{AB}\|P_{AB}) + D(V_{Y|X}\|W_{Y|X}|S_X(Q_A)) + |I(S_X(Q_A); V_{Y|X}) - H(Q_{A|B})|^+\} \end{aligned}$$

The last equality is true because that the codeword composition $S_X(Q_A)$ can be picked according to the source composition Q_A. And by our code book selection we always pick the composition to maximize the error exponent

$$\min_{Q_{B|A},V_{Y|X}} \{D(Q_{AB}\|P_{AB}) + D(V_{Y|X}\|W_{Y|X}|S_X(Q_A)) + |I(S_X(Q_A); V_{Y|X}) - H(Q_{A|B})|^+\}.$$

Here we slightly abuse the notations where $S_X(Q_A)$ is always the optimal distribution to maximize the above exponent given Q_A.

The lower bound on $E(P_{AB}, W_{Y|X})$ in Theorem (1) is just proved. ■

2) Upper bound: [7] First we fix the source composition Q_A, there are $2^{n(H(Q_A)-\epsilon_n^1)}$ sequences in $\mathcal{A}^n$ with type Q_A. When the encoder observes the source sequence a^n, it has to send a code word $x^n(a^n)$ to the channel $W_{Y|X}$. There are at most $(n+1)^{|X|}$ different types, so at least

$$\frac{2^{n(H(Q_A)-\epsilon_n^1)}}{(n+1)^{|X|}} = 2^{n(H(Q_A)-\epsilon_n^{2'})}$$

of the codewords for $a^n \in Q_A$ have the same composition, we write this composition $S_X(Q_A)$, and

$$A_1 = \{a^n \in Q_A : x^n(a^n) \in S_X(Q_A)\}, \text{ where } |A_1| = 2^{n(H(Q_A)-\epsilon_n^2)}.$$

Now we fix the conditional type $Q_{B|A}$, so we have the marginal Q_B and the joint distribution Q_{AB} determined by Q_A and $Q_{B|A}$. Write

$$Q_{A|B}(b^n) = \{a^n : (a^n, b^n) \in Q_{AB}\} \text{ and } Q_{B|A}(a^n) = \{b^n : (a^n, b^n) \in Q_{AB}\}.$$

Obviously $|Q_B| = 2^{n(H(Q_B)-\epsilon_n^3 n)}$ and for all b^n: $|Q_{A|B}(b^n)| = 2^{n(H(Q_{A|B})-\epsilon_n^4)}$, for all a^n: $|Q_{B|A}(a^n)| = 2^{n(H(Q_{B|A})-\epsilon_n^{4'})}$.

[7]In this proof, $\epsilon_n^i > 0$ and $\epsilon_n^i \to 0$, $i = 1, 2, 3, 4, 4', 5, 6$ and 7.

Let $B_1 = \{b^n \in Q_B : |Q_{A|B}(b^n) \bigcap A_1| \geq 2^{n(H(Q_{A|B})-\epsilon_n^5)}\}$, where $\epsilon_n^5 = \epsilon_n^2 + \epsilon_n^{4'} + \frac{1}{n}$. We show next that the size of B_1 is of the order $2^{nH(Q_B)}$.

Let $AB_1 = \{(a^n, b^n) : a^n \in A_1 \text{ and } (a^n, b^n) \in Q_{AB}\}$, we compute the size of AB_1 from two different ways.

First

$$|AB_1| = |A_1||Q_{B|A}(a^n)| = 2^{n(H(Q_{AB})-\epsilon_n^2-\epsilon_n^{4'})}. \tag{62}$$

Secondly

$$\begin{aligned}
|AB_1| &= |\{(a^n, b^n) : b^n \in B_1,\ a^n \in A_1 \text{ and } (a^n, b^n) \in Q_{AB}\} \\
&\quad \bigcup \{(a^n, b^n) : b^n \in Q_B - B_1,\ a^n \in A_1 \text{ and } (a^n, b^n) \in Q_{AB}\}| && (63)\\
&\leq |B_1||Q_{A|B}(b^n)| + |Q_B - B_1| 2^{n(H(Q_{A|B})-\epsilon_n^5)} && (64)\\
&= |B_1| 2^{n(H(Q_{A|B})-\epsilon_n^4)} + (2^{n(H(Q_B)-\epsilon_n^3)} - |B_1|) 2^{n(H(Q_{A|B})-\epsilon_n^5)} \\
&\leq |B_1| 2^{nH(Q_{A|B})} + 2^{nH(Q_B)} 2^{n(H(Q_{A|B})-\epsilon_n^5)} && (65)
\end{aligned}$$

(63) is by the definition of AB_1 and B_1, (64) is by the definition of B_1, (65) is true because all ϵ_n^i's are positive.

Combining (62) and (65) and use the fact that $\epsilon_n^5 = \epsilon_n^2 + \epsilon_n^{4'} + \frac{1}{n}$, we have:

$$\begin{aligned}
|B_1| 2^{nH(Q_{A|B})} &\geq 2^{n(H(Q_{AB})-\epsilon_n^2-\epsilon_n^{4'})} - 2^{nH(Q_B)} 2^{n(H(Q_{A|B})-\epsilon_n^5)} \\
&= 2^{n(H(Q_{AB})-\epsilon_n^2-\epsilon_n^{4'})} \times \frac{1}{2}.
\end{aligned}$$

Hence $|B_1| \geq 2^{n(H(Q_B)-\epsilon_n^2-\epsilon_n^{4'}-\frac{1}{n})} = 2^{n(H(Q_B)-\epsilon_n^5)}$.

Now we consider the decoding error of the following events and show that this error events gives us an upper bound on the error exponent stated in this theorem:

Source and side information pair $AB^* = \{(a^n, b^n) : a^n \in A_1, b^n \in B_1, (a^n, b^n) \in Q_{AB}\}$.

First, for each $(a^n, b^n) \in AB^*$:

$$P_{AB}(a^n, b^n) = 2^{-n(D(Q_{AB}\|P_{AB})+H(Q_{AB}))}.$$

Secondly, the size of AB^* is lower bounded as follows from the definition of B_1 and the lower bound on $|B_1|$:

$$\begin{aligned}
|AB^*| &\geq |B_1| \times 2^{n(H(Q_{A|B})-\epsilon_n^5)}\} \\
&\geq 2^{n(H(Q_B)-\epsilon_n^5)} \times 2^{n(H(Q_{A|B})-\epsilon_n^5)} \\
&\geq 2^{n(H(Q_{AB})-2\epsilon_n^5)} && (66)
\end{aligned}$$

So obviously the probability of AB^* is

$$P_{AB}(AB^*) = |AB^*| 2^{-n(D(Q_{AB}\|P_{AB})+H(Q_{AB}))} \geq 2^{-n(D(Q_{AB}\|P_{AB})+2\epsilon_n^5)}. \tag{67}$$

Thirdly, if the side-information is $b^n \in B_1$ there are at least $2^{n(H(Q_{A|B})-\epsilon_n^5)}$ many a^n's such that $(a^n, b^n) \in Q_{AB}$, that is, there are at least $2^{n(H(Q_{A|B})-\epsilon_n^5)}$ many source sequences with the same likelihood given the side-information b^n (even there exists a "genie" that tells the decoder that the joint distribution of (a^n, b^n) is Q_{AB}). Furthermore, the channel input codeword $x^n(a^n)$ for these source sequences all have composition $S_X(Q_A)$. Hence we have a channel coding problem with rate $H(Q_{A|B}) - \epsilon_n^5$ and fixed input composition $S_X(Q_A)$. This is the standard channel coding sphere packing bound studied in [7].

So if $b^n \in B_1$, then **average** error probability for $(a^n, b^n) \in AB^*$ is at least:

$$2^{-n(\min_{V_{Y|X}:I(S_X(Q_A);V_{Y|X})<H(Q_{A|B})-\epsilon_n^5}\{D(V_{Y|X}\|W_{Y|X}|S_X(Q_A))\}+\epsilon_n^6)}$$
$$\geq 2^{-n(\min_{V_{Y|X}:I(S_X(Q_A);V_{Y|X})<H(Q_{A|B})}\{D(V_{Y|X}\|W_{Y|X}|S_X(Q_A))\}+\epsilon_n^7)}, \quad (68)$$

where ϵ_n^5 and ϵ_n^6 goes to zero as n goes to infinity, hence $\epsilon_n^7 \to 0$ because $I(S_X(Q_A); V_{Y|X})$ is continuous in $V_{Y|X}$ and $D(V_{Y|X}\|W_{Y|X}|S_X(Q_A))$ is convex in $V_{Y|X}$.

Finally we combine (67) and (68), and notice that the above analysis is true for any(adversary) distribution of the source Q_A, and any(optimal) channel codebook composition $S_X(Q_A)$, and any(adversary) $Q_{B|A}$ after Q_A and $S_X(Q_A)$ are chosen, the error probability is lower bounded by:

$$2^{-n(\min_{Q_A}\max_{S_X(Q_A)}\min_{Q_{B|A}}\{D(Q_{AB}\|P_{AB})+\min_{V_{Y|X}:I(S_X(Q_A);V_{Y|X})<H(Q_{A|B})}\{D(V_{Y|X}\|W_{Y|X}|S_X(Q_A))\}\}+2\epsilon_n^5+\epsilon_n^7)}$$
$$= 2^{-n(\min_{Q_A}\max_{S_X(Q_A)}\min_{Q_{B|A},V_{Y|X}:I(S_X(Q_A);V_{Y|X})<H(Q_{A|B})}\{D(Q_{AB}\|P_{AB})+D(V_{Y|X}\|W_{Y|X}|S_X(Q_A))\}+2\epsilon_n^5+\epsilon_n^7)}$$

Both ϵ_n^5 and ϵ_n^7 converges to zero as n goes to infinity, the upper bound in Theorem 1 is just proved. ■

C. Proof of Corollary 1

The proofs for both lower bounds and uppers with the "digital interface" are similar.

1) Proof of (17), the lower bound: By introducing the auxiliary variable R to separate the source coding and channel coding error exponents and the definition of error exponents, the following equalities should be obvious.

$$\begin{aligned}
& E(P_{AB}, W_{Y|X}) \\
\geq\ & \min_{Q_A}\max_{S_X(Q_A)}\min_{Q_{B|A},V_{Y|X}} D(Q_{AB}\|P_{AB}) + D(V_{Y|X}\|W_{Y|X}|S_X(Q_A)) + |I(S_X(Q_A);V_{Y|X}) - H(Q_{A|B})|^+ \\
=\ & \min_{Q_A}\max_{S_X(Q_A)}\min_{R} \\
& \{\min_{Q_{B|A},V_{Y|X}:H(Q_{A|B})=R} D(Q_{AB}\|P_{AB}) + D(V_{Y|X}\|W_{Y|X}|S_X(Q_A)) + |I(S_X(Q_A);V_{Y|X}) - R|^+\} \\
=\ & \min_{Q_A}\max_{S_X(Q_A)}\min_{R}\{\min_{Q_{B|A}:H(Q_{A|B})=R} D(Q_{AB}\|P_{AB}) + E_r(R, S_X(Q_A), W_{Y|X})\} && (69) \\
=\ & \min_{Q_A}\max_{S_X(Q_A)}\min_{R}\{e'_U(R, P_{AB}, Q_A) + E_r(R, S_X(Q_A), W_{Y|X})\} && (70)
\end{aligned}$$

where $E_r(R, S_X(Q_A), W_{Y|X})$ is the standard random coding error exponent for channel $W_{Y|X}$ at rate R and input distribution $S_X(Q_A)$, while $e'_U(R, P_{AB}, Q_A)$ is a peculiar source coding with side-information error exponent for source P_{AB} at rate R, where the empirical source distribution is fixed at Q_A. That is for Q_A

$$e'_U(R, P_{AB}, Q_A) \triangleq \min_{Q_{B|A}:H(Q_{A|B})=R} D(Q_{AB}\|P_{AB})$$

(70) needs more examination. It is obvious that

$$e'_U(R, P_{AB}, Q_A) \geq \min_{Q_{B|A}:H(Q_{A|B})\geq R} D(Q_{AB}\|P_{AB}) \triangleq e_U(R, P_{AB}, Q_A).$$

where $e_U(R, P_{AB}, Q_A)$ is defined in (18). Now (70) becomes

$$E(P_{AB}, W_{Y|X}) \geq \min_{Q_A}\max_{S_X(Q_A)}\min_{R}\{e_U(R, P_{AB}, Q_A) + E_r(R, S_X(Q_A), W_{Y|X})\}. \quad (71)$$

2) Proof of (16), the upper bound: Similar to the proof for (17), we have the following equalities:

$$
\begin{aligned}
& E(P_{AB}, W_{Y|X}) \\
\leq\ & \min_{Q_A} \max_{S_X(Q_A)} \min_{Q_{B|A}, V_{Y|X}: I(S_X(Q_A); V_{Y|X}) < H(Q_{A|B})} \{D(Q_{AB} \| P_{AB}) + D(V_{Y|X} \| W_{Y|X} | S_X(Q_A))\} \\
=\ & \min_{Q_A} \max_{S_X(Q_A)} \min_{R} \min_{Q_{B|A}, V_{Y|X}: I(S_X(Q_A); V_{Y|X}) < R < H(Q_{A|B})} \{D(Q_{AB} \| P_{AB}) + D(V_{Y|X} \| W_{Y|X} | S_X(Q_A))\} \\
=\ & \min_{Q_A} \max_{S_X(Q_A)} \min_{R} \{ \min_{Q_{B|A}: H(Q_{A|B}) > R} D(Q_{AB} \| P_{AB}) + \min_{V_{Y|X}: I(S_X(Q_A); V_{Y|X}) < R} D(V_{Y|X} \| W_{Y|X} | S_X(Q_A))\} \\
=\ & \min_{Q_A} \max_{S_X(Q_A)} \min_{R} \{ \min_{Q_{B|A}: H(Q_{A|B}) > R} D(Q_{AB} \| P_{AB}) + E_{sp}(R, S_X(Q_A), W_{Y|X})\} \qquad (72) \\
=\ & \min_{Q_A} \max_{S_X(Q_A)} \min_{R} \{ e_U(R, P_{AB}, Q_A) + E_{sp}(R, S_X(Q_A), W_{Y|X})\} \qquad (73)
\end{aligned}
$$

where $E_{sp}(R, S_X(Q_A), W_{Y|X})$ is the standard sphere packing bound defined in (5) and $e_U(R, P_{AB}, Q_A)$ is defined in (18). □

References

[1] Rudolf Ahlswede. Coloring hypergraphs: A new approach to multi-user source coding ii. *Journ. of Combinatorics, Information and System Sciences*, 5:220–268, 1980.

[2] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[3] Cheng Chang and Anant Sahai. The price of ignorance: the impact on side-information for delay in lossless source coding. *submitted to IEEE Transactions on Information Theory*, 2007. `http://arxiv.org/abs/0712.0873`.

[4] Gérard Cohen, Iiro Honkala, Simon Litsyn, and Antoine Lobstein. *Covering Codes*. Elsevier, 1997.

[5] Imre Csiszár. Joint source-channel error exponent. *Problem of Control and Information Theory*, 9:315–328, 1980.

[6] Imre Csiszár and János Körner. *Information Theory*. Akadémiai Kiadó, Budapest, 1986.

[7] Robert Gallager. Fixed composition arguments and lower bounds to error probability. `http://web.mit.edu/gallager/www/notes/notes5.pdf`.

[8] Robert Gallager. Source coding with side information and universal coding. Technical Report LIDS-P-937, Mass. Instit. Tech., 1976.

[9] Robert G. Gallager. *Information Theory and Reliable Communication*. John Wiley, New York, NY, 1971.

[10] Benjamin Kelly and Aaron B. Wagner. Error exponents and test channel optimization for the wyner-ziv problem. *Proceedings of the 46th Annual Allerton Conference on Communications, Control, and Computing*, Oct. 2007.

[11] Claude Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27, 1948.

[12] Maurice Sion. On general minmax theorems. *Pacific Journal of Mathematics*, 8:171– 176, 1958.

[13] David Slepian and Jack Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19, 1973.

[14] Yangfan Zhong. *Joint Source-Channel Coding Reliability Function for Single and Multi-Terminal Communication Systems*. Ph.D. Thesis, Queens University, 2008.

[15] Yangfan Zhong, Fady Alajaji, and L. Lorne Campbell. Source side information can increase the joint source-channel coding error exponent. *Proceedings of the Tenth Canadian Workshop on Information Theory, Edmonton*, June 2007.