Error Exponents for Joint Source-Channel Coding With Side Information

Cheng Chang, *Member, IEEE*

***Abstract*—In this paper, we study upper and lower bounds on the error exponents for joint source-channel coding with decoder side-information. The results in the paper are nontrivial extensions of Csiszár's classical paper "Joint Source-Channel Error Exponent", *Problems of Control and Information Theory*, 1980. Unlike the joint source-channel coding result in Csiszár's paper, 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 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.**

***Index Terms*—Error exponent, error exponent game, joint source-channel coding, source coding with side-information.**

I. Introduction

In Shannon's very first paper on information theory [3], it is established that separation-based 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 the error exponent is considered. It is shown that joint source-channel coding achieves a strictly better error exponent than separation-based[1] coding [4]. The key technical component of [4] 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 Fig. 1. For a special setup of Fig. 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 $H(P_{A|B})$ [6]. The error exponent of this problem is also studied in [7], [8] and more importantly in [9].

The duality between source coding with decoder side-information and channel coding was established in the 80's [9]. 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 [8], also known as the Johnson-Stein-Lovász theorem [10]. With this duality result, we know that finding the error exponent of channel coding for channel $V_{Y|X}$ with channel code composition Q_X is essentially the same problem as that finding the error exponent of source coding with decoder side-information where the joint distribution is $Q_X \times V_{Y|X}$. 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?

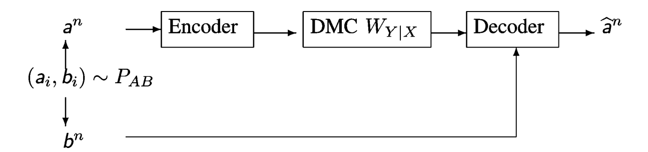

Fig. 1. Joint source-channel coding with decoder only side-information. Both the source and the channel are discrete memoryless.

This more general case, where $W_{Y|X}$ is a noisy channel in Fig. 1, is studied in [11] and recently in [12], [5]. It is shown in [11] that 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. In [12], [5], a suboptimal error exponent based on a mixture scheme of separation-based coding and the joint source-channel coding first developed in [4] is achieved. In this paper, we follow Csiszár's idea in [4] 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 resulting lower and upper bound have the same property as the joint source-channel coding error exponent *without* side-information in [4]: 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 III. Then in Section IV, 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 [4]. The main result of this paper is presented in Section V. Some implications of these bounds are given in Section VI.

II. Notation

We use serifed-fonts, e.g., a to indicate sample values, and sans-serif, e.g., a, to indicate random variables. We denote fi-

Manuscript received November 29, 2009; revised May 02, 2010; accepted January 25, 2011. Date of current version October 07, 2011. This work was conducted when the author was with Hewlett-Packard Laboratories, Palo Alto, CA. This material was presented in part at the 2009 Allerton Conference on Communication, Control, and Computing.

The author is with D. E. Shaw & Co., New York, NY 10036 USA (e-mail: chechang@ocf.berkeley.edu).

Communicated by E. Ordentlich, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2011.2165154

[1]In [4], Csiszár shows that the obvious separation-based coding scheme is suboptimal in terms of achieving the best error exponent. A more detailed treatment is given in [5].

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

nite sets by $\mathcal{A}$, $\mathcal{B}$, $\mathcal{X}$ and $\mathcal{Y}$. A probability distribution on finite set $\mathcal{A}$ is denoted by P_A or Q_A, a distribution on $\mathcal{A}\times\mathcal{B}$ is simply P_{AB}. A channel, or a probability transition matrix from finite set $\mathcal{X}$ to $\mathcal{Y}$ is denoted by $W_{Y|X}$ or $V_{Y|X}$. If a sequence a^n's empirical distribution is P_A, then we simply write $a^n \in P_A$. The empirical entropy of a^n, $H(a^n)$, is defined as the entropy of the empirical distribution, i.e., $H(a^n) \triangleq H(P_A)$. Similarly, for a sequence pair (a^n, b^n) with empirical distribution P_{AB}, the empirical conditional entropy $H(a^n|b^n)$ and the empirical mutual information $I(a^n, b^n)$ are defined respectively as the conditional entropy and the mutual information of the joint empirical distribution P_{AB}. The KL divergence between two distributions Q_A and P_A is denoted by $D(Q_A\|P_A)$.

There are numerous error exponents in this paper. The main topic of the paper, $E(P_{AB}, W_{Y|X})$, is the error exponent for joint source-channel coding with decoder side information, where the source/side-information channel pair is P_{AB} and $W_{Y|X}$. $E_{separate}(P_{AB}, W_{Y|X})$ is the error exponent if a separation-based coding scheme is applied. $E_{\rm both}(P_{AB}, W_{Y|X})$ is the error exponent while both encoder and decoder have access to the side-information. For channel coding, the error exponent for channel $W_{Y|X}$ at rate R is denoted by $E_c(R, W_{Y|X})$, the random coding lower bound is denoted by $E_r(R, W_{Y|X})$ and the sphere packing upper bound is denoted by $E_{sp}(R, W_{Y|X})$. For source coding with decoder side-information, the error exponent for source/side-information distribution P_{AB} at rate R is is denoted by $e(R, P_{AB})$, the lower bound and upper bound of this error exponent are denoted by $e_L(R, P_{AB})$ and $e_U(R, P_{AB})$ respectively. Other error exponents are defined later as they appear.

III. Review of Source and Channel Coding Error Exponents

A. System Model of Joint Source-Channel Coding With Decoder Side-Information

As shown in Fig. 1, the source and side-information, a^n and b^n respectively, are random variables drawn 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 matrix $W_{Y|X}$, 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 probability of error over all channel and source behaviors is

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

The error exponent, for optimal coding, is defined as $E(P_{AB}, W_{Y|X}) =$

$$\lim_{n\to\infty}\inf -\frac{1}{n}\log \Pr(\mathsf{a}^n \neq \widehat{\mathsf{a}}^n(\mathsf{b}^n, \mathsf{y}^n)). \quad (2)$$

The main goal 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

We review some classical results on channel coding error exponents and source coding with side-information error exponents.[3] These bounds are investigated in [13], [8], [7] and [14].

1) Channel Coding Error Exponent $E_c(R, W_{Y|X})$: 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 $\{1, 2, \ldots, 2^R\}$. For the sake of simplicity, we assume that 2^R is an integer. If 2^R is not an integer, we can lump K symbols together and approximate 2^{KR} by an integer for large K, this is not a problem because $\lim\limits_{K\to\infty}\frac{1}{K}\log_2(\lfloor 2^{KR}\rfloor) = R$. With this interpretation of channel coding, the definitions of error probability in (1) and error exponent in (2) still hold.

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

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

where $E_r(R, W_{Y|X}) = \max\limits_{S_X} E_r(R, S_X, W_{Y|X})$

$$E_r(R, S_X, W_{Y|X}) = \inf_{V_{Y|X}} D(V_{Y|X}\|W_{Y|X}|S_X) + |I(V_{Y|X}; S_X) - R|^+ \quad (4)$$

and $E_{sp}(R, W_{Y|X}) = \max\limits_{S_X} E_{sp}(R, S_X, W_{Y|X})$, where $E_{sp}(R, S_X, W_{Y|X}) =$

$$\max_{S_X}\inf_{V_{Y|X}: I(V_{Y|X}; S_X) < R} D(V_{Y|X}\|W_{Y|X}|S_X). \quad (5)$$

Here $|\cdot|^+ = \max\{0, \cdot\}$ and 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 [13] 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 [13]. 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 [13] 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) Error Exponents for Source Coding With Decoder Side-Information: This is also a special case of the general setup in Fig. 1. This time the channel $W_{Y|X}$ is a noiseless channel with

[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 Fenchel duality. This is discussed in detail in Problem 23 on pg. 192 [8].

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 [15]:

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

where

$$e_L(R, P_{AB}) = \inf_{Q_{AB}} D(Q_{AB} \| P_{AB}) + |R - H(Q_{A|B})|^+$$
$$e_U(R, P_{AB}) = \inf_{Q_{AB}: H(Q_{A|B}) > R} D(Q_{AB} \| P_{AB}).$$

The duality between channel coding and source coding with decoder side information is well understood [9]. We give the following duality results on error exponents

$$e(R, Q_A, P_{B|A}) = E_c(H(Q_A) - R, Q_A, P_{B|A})$$

or equivalently

$$e(H(Q_A) - R, Q_A, P_{B|A}) = E_c(R, Q_A, P_{B|A})$$

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 [4] on the operational level, i.e., a complete characterization of the source coding with side information error exponent $e(R, Q_A, P_{B|A})$ implies a complete characterization 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 $R > H(P_{A|B})$. A special case of the source coding with decoder side information problem is when 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 [8]

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

3) Joint Source-Channel Coding Error Exponents [4]: In Csiszár's seminal paper [4], the joint source-channel coding error exponents is studied. This is yet another special case of the general setup in Fig. 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

$$\Pr(\mathsf{a}^n \ne \hat{\mathsf{a}}^n(\mathsf{y}^n)) = \sum_{a^n} \{P_A(a^n)$$
$$\sum_{y^n} W_{Y|X}(y^n | x^n(a^n)) 1(a^n \ne \hat{a}^n(y^n))\}. \tag{7}$$

We denote the error exponent of (7) by $E(P_A, W_{Y|X})$. The lower and upper bounds of the error exponents are derived in [4]. It is shown that:

$$\min_R \{e(R, P_A) + E_r(R, W_{Y|X})\} \le E(P_A, W_{Y|X})$$
$$\le \min_R \{e(R, P_A) + E_{sp}(R, W_{Y|X})\}. \tag{8}$$

The upper bound is derived by using standard method of types arguments. The lower bound is a direct consequence of the channel coding Theorem 5 in [4].

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 $R^* \ge R_{cr}$ 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

$$E(P_A, W_{Y|X}) = e(R^*, P_A) + E_r(R^*, W_{Y|X}). \tag{9}$$

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 (8) and (9).

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

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

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 $a \in \mathcal{A}_i$

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

This channel coding theorem as Csiszár put it, is the "main result of this paper" in [4]. We use this theorem directly in the proof of the lower bound in Proposition 1 and modify it to show the lower bound in Theorem 1.

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

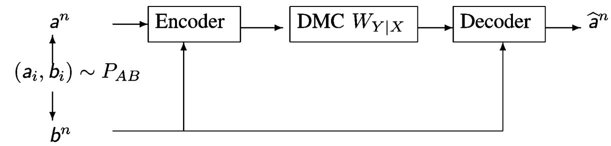

Fig. 2. Joint source-channel coding with both decoder and encoder side-information. Both the source and the channel are discrete memoryless.

IV. 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 Fig. 2.

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

$$\Pr(\mathsf{a}^n \neq \widehat{\mathsf{a}}^n(\mathsf{b}^n, \mathsf{y}^n)) = \sum_{a^n, b^n} \Big\{ P_{AB}(a^n, b^n) \sum_{y^n} W_{Y|X}(y^n | x^n(a^n, b^n)) 1(a^n \neq \widehat{a}^n(b^n, y^n)) \Big\}. \quad (10)$$

The error exponent of this setup is denoted by $E_{\mathrm{both}}(P_{AB}, W_{Y|X})$, 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, $E_{\mathrm{both}}(P_{AB}, W_{Y|X})$ is not smaller than $E(P_{AB}, W_{Y|X})$.

Comparing (10) and (7), 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 $P_{A|B}(\mathsf{a}^n | b^n)$. Hence, we can extend the results for joint source-channel coding error exponent [4]. We summarize the bounds on $E_{\mathrm{both}}(P_{AB}, W_{Y|X})$ in the following proposition.

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

$$E_{\mathrm{both}}(P_{AB}, W_{Y|X}) \leq \min_R \{ e_U(R, P_{AB}) + E_{sp}(R, W_{Y|X}) \}$$
$$E_{\mathrm{both}}(P_{AB}, W_{Y|X}) \geq \min_R \{ e_U(R, P_{AB}) + E_r(R, W_{Y|X}) \}.$$

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

Proof: see Appendix A. Because $E_{\mathrm{both}}(P_{AB}, W_{Y|X})$ is no smaller than $E(P_{AB}, W_{Y|X})$, the lower bound of $E(P_{AB}, W_{Y|X})$ in Theorem 1 is also a lower bound for $E_{\mathrm{both}}(P_{AB}, W_{Y|X})$. It is later shown that the lower bound in Theorem 1 is the same as the lower bound in Proposition 1 for symmetric channels, but otherwise not larger than the lower bound in Proposition 1. In the Appendix, we give a simple direct proof of the lower bound on $E_{\mathrm{both}}(P_{AB}, W_{Y|X})$ which is a corollary of Theorem 5 in [4]. □

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

$$E_{\mathrm{both}}(P_{AB}, W_{Y|X}) = e_U(R^*, P_{AB}) + E_r(R^*, W_{Y|X}). \quad (11)$$

In this case $E_{\mathrm{both}}(P_{AB}, W_{Y|X})$ is completely characterized.[5]

V. Joint Source-Channel Error Exponents With Only Decoder Side Information

We study the problem where only the 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 Fig. 1, error exponent: For the error probability $\Pr(\mathsf{a}^n \neq \widehat{\mathsf{a}}^n(\mathsf{b}^n, \mathsf{y}^n))$ and error exponent $E(P_{AB}, W_{Y|X})$ defined in (1) and (2), we have the following lower and upper bounds:

$$E(P_{AB}, W_{Y|X}) \geq \min_{Q_A} \max_{S_X(Q_A)} \min_{Q_{B|A}, V_{Y|X}} \{ D(Q_{AB} \| P_{AB}) + D(V_{Y|X} \| W_{Y|X} | S_X(Q_A)) + |I(S_X(Q_A); V_{Y|X}) - H(Q_{A|B})|^+ \} \quad (12)$$

$$E(P_{AB}, W_{Y|X}) \leq \min_{Q_A} \max_{S_X(Q_A)} \min_{Q_{B|A}, V_{Y|X} : I(S_X(Q_A); V_{Y|X}) < H(Q_{A|B})} \{ D(Q_{AB} \| P_{AB}) + D(V_{Y|X} \| W_{Y|X} | S_X(Q_A)) \}. \quad (13)$$

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 [4]. However, we cannot directly use the channel coding of Theorem 5 in [4] because of the presence of the side information. In essence, we have to study a more complicated case using the method of types. For 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 so that the channel and side-information only interact through R.

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

$$E(P_{AB}, W_{Y|X}) \leq \min_{Q_A} \max_{S_X(Q_A)} \min_R \{ e_U(R, P_{AB}, Q_A) + E_{sp}(R, S_X(Q_A), W_{Y|X}) \}, \quad (14)$$

$$E(P_{AB}, W_{Y|X}) \geq \min_{Q_A} \max_{S_X(Q_A)} \min_R \{ e_U(R, P_{AB}, Q_A) + E_r(R, S_X(Q_A), W_{Y|X}) \} \quad (15)$$

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 distribu-

[5] A special case is when the channel is noiseless with capacity R. This is the source coding with both encoder and decoder side-information problem. In this special case, we see that for any R, both the lower bound and the upper bound are $e_U(R, P_{AB})$. This is similar to the source coding error exponent studied in [8].

tion $S_X(Q_A)$ defined in (4), while $e_U(R, P_{AB}, Q_A)$ is the error exponent of the peculiar source coding with side-information problem for source P_{AB} at rate R, where the empirical source distribution is fixed at Q_A. That is, for fixed Q_A

$$e_U(R, P_{AB}, Q_A) \triangleq \min_{Q_{B|A}: H(Q_{A|B}) \geq R} D(Q_{AB} \| P_{AB}). \quad (16)$$

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. *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 $V_{Y|X}$ (nonobservable to the coding system). The coding system chooses $S_X(Q_A)$ 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 [4] in (8), it is desirable to have a sufficient condition for the lower bound and the upper bound to match, i.e., the complete characterization as in (9). It is simpler for the case in (8) since all that is needed is that the sphere packing bound and the random coding bound to match at the critical rate R^* as discussed in Section III-B-III. 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 in (14) is $(Q_A^u, S_X^u(Q_A), R^u)$ and solution of the game in (15) is $\left(Q_A^l, S_X^l(Q_A), R^l\right)$. An obvious sufficient condition for the two bounds to match is as follows:

$$\left(Q_A^l, S_X^l(Q_A), R^l\right) = (Q_A^u, S_X^u(Q_A), R^u) \text{ and}$$
$$E_r\left(R^u, S_X^u(Q_A), W_{Y|X}\right) = E_{sp}\left(R^u, S_X^u(Q_A), W_{Y|X}\right). \quad (17)$$

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

B. *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 a symmetric channel[6] $W_{Y|X}$, as defined on Page 94 in [13], for which the input distribution S_X to maximize the random coding error exponent $E_r(R, S_X, W_{Y|X})$ is uniform on $\mathcal{X}$ for all R, the upper and lower bounds in Theorem 1 and Corollary 1 can be further simplified to the following forms:

$$E(P_{AB}, W_{Y|X}) \leq \min_R \{e_U(R, P_{AB}) + E_{sp}(R, W_{Y|X})\} \quad (18)$$
$$E(P_{AB}, W_{Y|X}) \geq \min_R \{e_U(R, P_{AB}) + E_r(R, W_{Y|X})\}. \quad (19)$$

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

Proof: An important property for symmetric channels is that the input distribution that maximizes the random coding error exponent is constant for all rates R [13]; hence, the inner $\max\min\{\cdot\}$ is equal to $\min\max\{\cdot\}$, i.e.,

$$\begin{aligned}
&E(P_{AB}, W_{Y|X}) \\
&\quad \geq \min_{Q_A} \max_{S_X(Q_A)} \min_R \{e_U(R, P_{AB}, Q_A) \\
&\qquad + E_r(R, S_X(Q_A), W_{Y|X})\} \\
&\quad = \min_{Q_A} \min_R \max_{S_X(Q_A)} \{e_U(R, P_{AB}, Q_A) \\
&\qquad + E_r(R, S_X(Q_A), W_{Y|X})\} \\
&\quad = \min_{Q_A} \min_R \{e_U(R, P_{AB}, Q_A) + E_r(R, W_{Y|X})\} \quad (20) \\
&\quad = \min_R \{\min_{Q_A} \{e_U(R, P_{AB}, Q_A)\} + E_r(R, W_{Y|X})\} \\
&\quad = \min_R \{e_U(R, P_{AB}) + E_r(R, W_{Y|X})\} \quad (21)
\end{aligned}$$

where (20) follows the definition of random coding bound in (3) and (21) follows the obvious equality

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

The upper bound in (18) is trivial by noticing that $E(P_{AB}, W_{Y|X}) \leq E_{\text{both}}(P_{AB}, W_{Y|X})$ and the upper bound for $E_{\text{both}}(P_{AB}, W_{Y|X})$ in Proposition 1. However, we can also prove it by directly applying $\max\min\{\cdot\} \leq \min\max\{\cdot\}$ [16] to (14)

$$\begin{aligned}
&E(P_{AB}, W_{Y|X}) \\
&\quad \leq \min_{Q_A} \max_{S_X(Q_A)} \min_R \{e_U(R, P_{AB}, Q_A) \\
&\quad + E_{sp}(R, S_X(Q_A), W_{Y|X})\} \\
&\quad \leq \min_{Q_A} \min_R \max_{S_X(Q_A)} \{e_U(R, P_{AB}, Q_A) \\
&\qquad + E_{sp}(R, S_X(Q_A), W_{Y|X})\} \\
&\quad = \min_{Q_A} \min_R \{e_U(R, P_{AB}, Q_A) + E_{sp}(R, W_{Y|X})\} \\
&\quad = \min_R \{\min_{Q_A} \{e_U(R, P_{AB}, Q_A)\} + E_{sp}(R, W_{Y|X})\} \\
&\quad = \min_R \{e_U(R, P_{AB}) + E_{sp}(R, W_{Y|X})\}. \quad (22)
\end{aligned}$$

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 III-B-III. For more discussions, see Section VI.

[6] We quote the definition in [13]: "a DMC is defined to be symmetric if the set of outputs can be partitioned into subsets in such a way that for each subset the matrix of transition probabilities (using inputs as rows and outputs of the subset as columns) has the property that each row is a permutation of each other row and each column (if more than 1) is a permutation of each other column". This includes the binary symmetric and binary erasure channels.

C. Why is it Hard to Generalize Corollary 2 to Nonsymmetric 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 (nonsymmetric) channels. A sufficient condition of the existence of a unique saddle point for the equality to hold is known as Sion's Theorem [17], which states that:

$$\max_{\mu \in \mathcal{M}} \min_{\nu \in \mathcal{N}} f(\mu, \nu) = \min_{\nu \in \mathcal{N}} \max_{\mu \in \mathcal{M}} f(\mu, \nu) \tag{23}$$

if $\mathcal{M}$ and $\mathcal{N}$ are convex, compact spaces and f is quasi-convex[7] on $\mathcal{N}$ for all μ, quasi-concave on $\mathcal{M}$ for all ν and continuous on $\mathcal{M} \times N$. For the function of interest

$$\max_{S_X(Q_A)} \min_R \{e_U(R, P_{AB}, Q_A) + E_r(R, S_X(Q_A), W_{Y|X})\} \tag{24}$$

we examine the sufficient condition under which a unique equilibrium exists, according to Sion's Theorem. First, $e_U(R, P_{AB}, Q_A) + E_r(R, S_X(Q_A), W_{Y|X})$ is quasi-convex in R because both $e_U(R, P_{AB}, Q_A)$ and $E_r(R, S_X(Q_A), W_{Y|X})$ are convex, hence, quasi-convex in R. However, (24) is not necessarily 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 easily established by using 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 $\max\min\{\cdot\}$.

VI. "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 Figs. 3 and 4. The source a is a Bernoulli 0.5 random variable and the joint source has the distribution

$$P_{AB} = \begin{Bmatrix} 0.50 & 0.00 \\ 0.05 & 0.45 \end{Bmatrix}. \tag{25}$$

The channel $W_{Y|X}$ is a binary symmetric channel with crossover probability 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

[7]As defined in [16], a function $f : \mathcal{R}^n \to \mathcal{R}$ is called quasi-convex if its domain and all its sublevel sets: $\{x \in \text{domain of } f | f(x) \leq \alpha\}$ for all α are convex. A function f is quasi-concave if $-f$ is quasi-convex.

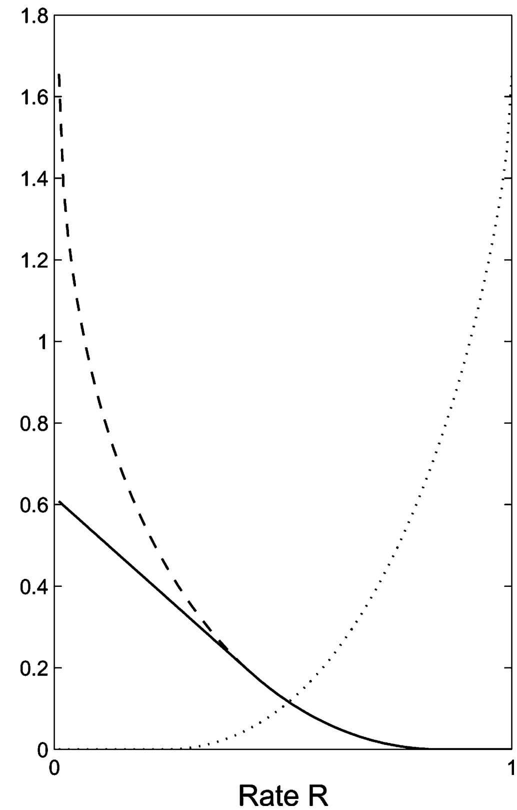

Fig. 3. Upper bound on source coding with side-information error exponent $e_U(R, P_{AB})$ is the dotted line. The random coding bound $E_r(R, W_{Y|X})$ and sphere packing bound $E_{sp}(R, W_{Y|X})$ for channel coding error exponents are the solid line and the dashed line respectively.

$e_U(R, P_{AB})$ are plotted in Fig. 3. The channel coding bounds match while $R \geq R_{cr}$, where R_{cr} is defined in [13]. 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 Fig. 4, we add both the lower and upper bounds on the joint source-channel coding with decoder side information to the plot in Fig. 3. For this source-channel pair P_{AB} and $W_{Y|X}$, we have a complete characterization of $E_{\text{both}}(P_{AB}, W_{Y|X})$ 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 Fig. 4. The value of the minimum is E_j shown in Fig. 4.

A. Encoder Side Information Does Not Always 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

$$E(P_{AB}, W_{Y|X}) = e_U(R^*, P_{AB}) + E_r(R^*, W_{Y|X}). \tag{26}$$

Comparing Corollary 2 and Proposition 1, we bound the error exponent with or without encoding side-information by the same lower and upper bounds. This does not mean that

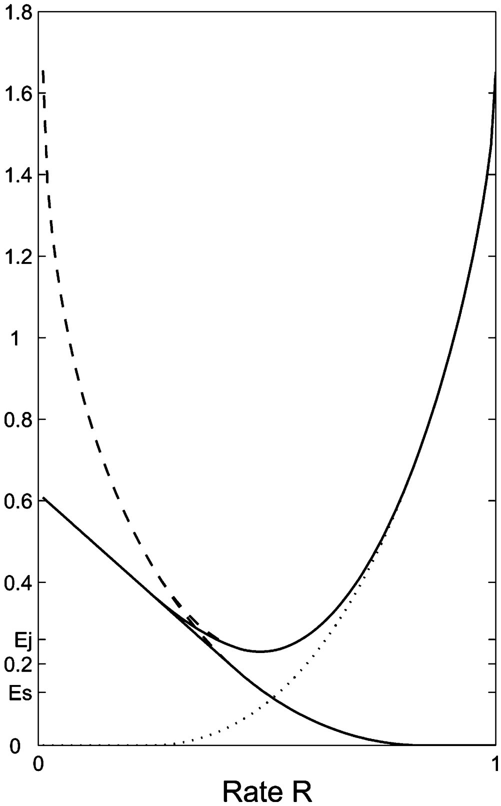

Fig. 4. $e_U(R, P_{AB}) + E_{sp}(R, W_{Y|X})$ and $e_U(R, P_{AB}) + E_r(R, W_{Y|X})$ are added to Fig. 3 in dashed line and solid line respectively. They match at the minimal point; hence, the joint source-channel coding with decoder side-information error exponent is completely determined as $E(P_{AB}, W_{Y|X}) = E_j$. And E_s is the separation-based coding error exponent $E_{separate}(P_{AB}, W_{Y|X})$ defined in (30).

$E(P_{AB}, W_{Y|X}) = E_{\text{both}}(P_{AB}, W_{Y|X})$ always holds. But if the lower bound and upper bound match, which is shown in Fig. 4, then we have

$$\begin{aligned} E(P_{AB}, W_{Y|X}) &= E_{\text{both}}(P_{AB}, W_{Y|X}) \\ &= e_U(R^*, P_{AB}) + E_r(R^*, W_{Y|X}) \end{aligned}$$

where R^* minimizes $e_U(R, P_{AB}) + E_r(R, W_{Y|X})$ and $R^* > R_{cr}$. This is another example[8] for block coding where knowing side-information at the encoder does not help increase the error exponent. In contrast, as discussed in [18], in the delay constrained setup, there is a penalty for not knowing the side-information even if the channel is noiseless.

B. Separation-Based Coding is Strictly Sub-Optimal

An obvious coding scheme for the problem in Fig. 1 is to implement a separation-based 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 separation-based coding scheme with outer source with side information coding and inner channel coding, both at rate R. If both coding schemes 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:

$$\begin{aligned} &\Pr(\mathsf{a}^n \neq \widehat{\mathsf{a}}^n(\mathsf{b}^n, \mathsf{y}^n)) \\ &\quad = \Pr(\mathsf{c}^{nR} \neq \widehat{\mathsf{c}}^{nR}(\mathsf{y}^n)) \\ &\qquad + \Pr(\mathsf{a}^n \neq \widehat{\mathsf{a}}(\widehat{\mathsf{c}}^{nR}(\mathsf{y}^n), \mathsf{b}^n), \mathsf{c}^{nR} = \mathsf{c}^{nR}(\mathsf{y}^n)) \quad (27) \\ &\quad \leq \Pr(\mathsf{c}^{nR} \neq \widehat{\mathsf{c}}^{nR}(\mathsf{y}^n)) \\ &\qquad + \Pr(\mathsf{a}^n \neq \widehat{\mathsf{a}}(\widehat{\mathsf{c}}^{nR}(\mathsf{y}^n), \mathsf{b}^n) | \mathsf{c}^{nR} = \mathsf{c}^{nR}(\mathsf{y}^n)) \quad (28) \\ &\quad \leq 2^{-n(E_r(R, W_{Y|X}) - \epsilon_n^1)} + 2^{-n(e_L(R, P_{AB}) - \epsilon_n^2)} \quad (29) \end{aligned}$$

where ϵ_n^1 and ϵ_n^2 converges to zero as n goes to infinity. Equation (27) follows the union bound argument that a decoding error occurs if either the inner channel coding fails or the outer source coding fails. Equation (28) is true because conditional probability is at least as large or equal to joint probability. Finally (29) is true because both the outer source coding and inner channel coding achieve the random coding error exponents. From (29) and that the optimization of the digital interface rate R between the channel coder and source coder, we know that a lower bound of the separation-based coding error exponent is

$$\begin{aligned} &\max_R \{\min\{E_r(R, W_{Y|X}), e_L(R, P_{AB}\}\} \\ &\qquad \triangleq E_{separate}(P_{AB}, W_{Y|X}). \end{aligned} \quad (30)$$

This separation-based coding scheme is also discussed for joint source-channel coding in [4], in which a similar bound is given. We next show why the separation-based 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 (19).

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 Fig. 4. This means that for rate $\bar{R}$ such that $E_r(\bar{R}, W_{Y|X}) = e_U(\bar{R}, P_{AB})$

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

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

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

[8]For the classical source coding with decoder side-information problem, knowing side-information at the encoder does not help increase the error exponent if $R \leq R_{cr}$, where R_{cr} is the minimum rate that the sphere packing bound and random coding bound agree [9].

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

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

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

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

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

Note: $E_{separate}(P_{AB}, W_{Y|X})$ is an achievable error exponent from the obvious separation-based 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 [4]. It should be clear that the upper bound of any separation-based source-channel coding error exponent is $\max_R\{\min\{E_{sp}(R, W_{Y|X}), e_U(R, P_{AB}\}\}$ which is comparable to (30). The proof hinges on the complete transparency between the source coding and channel coding, otherwise we have a joint coding scheme. A detailed discussion is in [5].

VII. Conclusions

We studied 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 [4]. To derive the lower bound, we used a novel joint source-channel with decoder side-information decoding scheme. We further investigated the conditions under which the lower bounds and upper bounds match. A game theoretic approach was 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 [19].

Appendix

A) Proof of Upper and Lower Bounds on $E_{\text{both}}(P_{AB}, W_{Y|X})$: 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 [4] respectively.

2) 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[9] $2^{n(H(Q_{A|B})-\epsilon_n^1)}$ many equally likely sequences $\in \mathcal{A}^n$ conditioned 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.,

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

if the channel $W_{Y|X}$ behaves like $V_{Y|X}$ with the code book with type $S_X(Q_{AB})$, with high probability, the decoder cannot correctly decide from one of the $2^{nH(Q_{A|B})}$ sequences. This is guaranteed by the Blowing up Lemma [8] or see a detailed proof in [14].

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

$$2^{-n\left(D(Q_{AB}\|P_{AB})+D(V_{Y|X}\|W_{Y|X}|S_X(Q_{AB}))-\epsilon_n^3\right)}. \tag{31}$$

Notice that the source behavior Q_{AB} and the channel behavior $V_{Y|X}$ are arbitrary. As long as $H(Q_{A|B}) > I(S_X(Q_{AB}); V_{Y|X})$, we can upper bound the error exponent as follows:

$$\begin{aligned} &E_{\text{both}}(P_{AB}, W_{Y|X}) \\ &\leq \min_{Q_{AB}, V_{Y|X}: H(Q_{A|B}) > I(S_X(Q_{AB}); V_{Y|X})} \{D(Q_{AB}\|P_{AB}) \\ &\quad + D(V_{Y|X}\|W_{Y|X}|S_X(Q_{AB}))\} \qquad (32) \\ &= \min_R\{\min_{Q_{AB}, V_{Y|X}: H(Q_{A|B}) > R > I(S_X(Q_{AB}); V_{Y|X})} \\ &\qquad D(Q_{AB}\|P_{AB}) + D(V_{Y|X}\|W_{Y|X}|S_X(Q_{AB}))\} \qquad (33) \\ &= \min_R\{\min_{Q_{AB}: H(Q_{A|B}) > R}\{D(Q_{AB}\|P_{AB}) + \\ &\qquad \min_{V_{Y|X}: R > I(S_X(Q_{AB}); V_{Y|X})} D(V_{Y|X}\|W_{Y|X}|S_X(Q_{AB}))\}\} \\ &\qquad (34) \\ &\leq \min_R\{\min_{Q_{AB}: H(Q_{A|B}) > R}\{ \\ &\qquad D(Q_{AB}\|P_{AB}) + E_{sp}(R, W_{Y|X})\}\} \qquad (35) \\ &= \min_R\{e_U(R, P_{AB}) + E_{sp}(R, W_{Y|X}).\} \qquad (36) \end{aligned}$$

(32) is a direct consequence of (31). In (33), we introduce the "digital interface" R, the equivalence in (33) and (34) should be obvious. Equation (35) and (36) are by definitions of the channel coding and source coding error exponents. □

3) Lower Bound: Fix 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}, $i = 1, 2, \ldots m_n$ the joint types. It should be clear that the Q^i_{AB}'s here all have the same marginal distribution as b^n

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

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

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 [4] 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., $(a^n, b^n) \in Q^i_{AB}$

$$\begin{aligned} p_{e,b^n}(a^n) &= \sum_{y^n} W_{Y|X}(y^n|x^n(a^n,b^n))1(a^n \neq \widehat{a}^n(b^n,y^n)) \\ &\leq 2^{-n(E_r(H(Q^i_{A|B}),W_{Y|X})-\epsilon_n)}. \end{aligned} \tag{37}$$

The joint source-channel coding error probability is hence

$$\begin{aligned} &\Pr(\mathsf{a}^n \neq \widehat{\mathsf{a}}^n(\mathsf{b}^n,\mathsf{y}^n)) \\ &= \sum_{a^n,b^n} \{P_{AB}(a^n,b^n) \\ &\quad \sum_{y^n} W_{Y|X}(y^n|x^n(a^n,b^n))1(a^n \neq \widehat{a}^n(b^n,y^n))\} \\ &= \sum_{Q_{AB}} \sum_{(a^n,b^n)\in Q_{AB}} \{P_{AB}(a^n,b^n) \\ &\quad \sum_{y^n} W_{Y|X}(y^n|x^n(a^n,b^n))1(a^n \neq \widehat{a}^n(b^n,y^n))\} \\ &\leq \sum_{Q_{AB}} \sum_{(a^n,b^n)\in Q_{AB}} \{P_{AB}(a^n,b^n) \\ &\qquad \times 2^{-n(E_r(H(Q_{A|B}),W_{Y|X})-\epsilon_n)}\} \qquad (38) \\ &\leq \sum_{Q_{AB}} 2^{-nD(Q_{AB}\|P_{AB})} \times 2^{-n(E_r(H(Q_{A|B}),W_{Y|X})-\epsilon_n)} \\ &\leq (n+1)^{|\mathcal{A}||\mathcal{B}|} \max_{Q_{AB}} \{2^{-nD(Q_{AB}\|P_{AB})} \\ &\qquad \times 2^{-n(E_r(H(Q_{A|B}),W_{Y|X})-\epsilon_n)}\} \\ &\leq 2^{-n\left(\min_{Q_{AB}}\{D(Q_{AB}\|P_{AB})+E_r(H(Q_{A|B}),W_{Y|X})\}-\epsilon'_n\right)}. \qquad (39) \end{aligned}$$

Equation (38) follows by substituting in (37) and the rest of the inequalities are by method of types. $\epsilon'_n \to 0$, so we can lower bound the error exponent as

$$\begin{aligned} &E_{\text{both}}(P_{AB}, W_{Y|X}) \\ &\quad \geq \min_{Q_{AB}} \{D(Q_{AB}\|P_{AB}) + E_r(H(Q_{A|B}), W_{Y|X})\} \qquad (40) \\ &\quad = \min_R \{\min_{Q_{AB}:H(Q_{A|B})=R} \{D(Q_{AB}\|P_{AB}) \\ &\qquad + E_r(H(Q_{A|B}), W_{Y|X})\}\} \qquad (41) \\ &\quad = \min_R \{\min_{Q_{AB}:H(Q_{A|B})=R} \{D(Q_{AB}\|P_{AB})\} \\ &\qquad + E_r(R, W_{Y|X})\} \qquad (42) \\ &\quad = \min_{R\geq H(P_{A|B})} \{\min_{Q_{AB}:H(Q_{A|B})=R} \{D(Q_{AB}\|P_{AB}) \\ &\qquad + E_r(R, W_{Y|X})\}\} \qquad (43) \\ &\quad = \min_{R\geq H(P_{A|B})} \{\min_{Q_{AB}:H(Q_{A|B})\geq R} \{D(Q_{AB}\|P_{AB}) \\ &\qquad + E_r(R, W_{Y|X})\}\} \qquad (44) \\ &\quad = \min_{R\geq H(P_{A|B})} \{e_U(R, P_{AB}) + E_r(R, W_{Y|X})\} \qquad (45) \\ &\quad = \min_R \{e_U(R, P_{AB}) + E_r(R, W_{Y|X})\}. \qquad (46) \end{aligned}$$

Equation (40) is a direct consequence of (39), in (41) we again introduce the "digital interface" variable R. Equations (42) and (45) are by definitions of $E_r(R, W_{Y|X})$ and $e_U(R, P_{AB})$ respectively. Equation (43) is true because $E_r(R, W_{Y|X})$ is monotonically decreasing with R and for $R < H(P_{B|A})$

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

Equation (44) 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 (46) 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 decreasing with R. □

D) Lower and Upper Bounds on $E(P_{AB}, W_{Y|X})$: We give the proof of Theorem 1 here.

5) Lower Bound: From the definition of the error exponent, we need to find an 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 that the error probability:

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

is upper bounded by $2^{n(E-\epsilon_n)}$ where $\epsilon_n \to 0$, where E is the right hand side of (12).

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:

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

The 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:

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

We next need to show that there exists such a encoder/decoder pair that achieve the error exponent in (12). 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.[10]

[10]In the rest of this section, we use $\overset{\zeta}{\Pr}$ to denote the probability under the codebook distribution ζ and use $\overset{c}{\Pr}$ to denote the probability when a particular codebook c is used.

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 $\tilde{s}^n \in S_X(\tilde{Q}_A)$

$$\begin{aligned}&\overset{\zeta}{\Pr}(x^n(a^n) = s^n, x^n(\tilde{a}^n) = \tilde{s}^n)\\&= \overset{\zeta}{\Pr}(x^n(a^n) = s^n)\overset{\zeta}{\Pr}(x^n(\tilde{a}^n) = \tilde{s}^n)\\&= \frac{1}{|S_X(Q_A)|}\frac{1}{|S_X(\tilde{Q}_A)|}.\end{aligned}$$

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

$$\begin{aligned}p_e^n &= E(\overset{\zeta}{\Pr}(\mathsf{a}^n \neq \widehat{\mathsf{a}}^n(\mathsf{b}^n, \mathsf{y}^n))\\&= \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}}\overset{c}{\Pr}(\mathsf{a}^n \neq \widehat{\mathsf{a}}^n(\mathsf{b}^n, \mathsf{y}^n))\end{aligned} \tag{47}$$

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

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

$$\begin{aligned}&\overset{c}{\Pr}(\mathsf{a}^n \neq \widehat{\mathsf{a}}^n(\mathsf{b}^n, \mathsf{y}^n))\\&= \sum_{a^n,b^n} P_{AB}(a^n,b^n)\overset{c}{\Pr}(a^n \neq \widehat{a}^n(b^n, \mathsf{y}^n))\\&= \sum_{Q_{AB}}\sum_{(a^n,b^n)\in Q_{AB}}\left(P_{AB}(a^n,b^n)\overset{c}{\Pr}(a^n \neq \widehat{a}^n(b^n, \mathsf{y}^n))\right)\\&= \sum_{Q_{AB}}\sum_{(a^n,b^n)\in Q_{AB}}\Bigg(P_{AB}(a^n,b^n)\sum_{y^n}W_{Y|X}(y^n|x^n(a^n))\\&\qquad 1^c(a^n \neq \widehat{a}^n(b^n,y^n))\Bigg)\\&= \sum_{Q_{AB}}\sum_{(a^n,b^n)\in Q_{AB}}\\&\qquad\Bigg(P_{AB}(a^n,b^n)\sum_{V_{Y|X}}\sum_{y^n:(x^n(a^n),y^n)\in S_X(Q_A)\times V_{Y|X}}\\&\qquad W_{Y|X}(y^n|x^n(a^n))1^c(a^n \neq \widehat{a}^n(b^n,y^n))\Bigg).\end{aligned} \tag{48}$$

For $(a^n,b^n) \in Q_{AB}$, 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 $y^n \in U_Y$

$$\begin{aligned}&I(x^n(\tilde{a}^n);y^n) - H(\tilde{a}^n|b^n) \geq I(x^n(a^n);y^n) - H(a^n|b^n), \text{ i.e.,}\\&I_{U_{XY}}(X;Y) - H(Q_{\tilde{A}|B}) \geq I(S_X(Q_A);V_{Y|X}) - H(Q_{A|B}).\end{aligned} \tag{49}$$

In (49), we rewrite the entropy and mutual information terms according to the empirical distributions of the sequence pairs. Specifically, $I_{U_{XY}}(X;Y)$ is the mutual information between X and Y where $(X,Y) \sim U_{XY}$, and $I(S_X(Q_A);V_{Y|X})$ is the mutual information between X and Y where $X \sim S_X(Q_A)$ and $(X,Y) \sim S_X(Q_A) \times V_{Y|X}$.

Now we can expand the indicator function in (48) as follows, for a codebook c

$$\begin{aligned}&1^c(a^n \neq \widehat{a}^n(b^n,y^n))\\&= 1^c\big(\exists \tilde{a}^n \neq a^n, s.t.\ I(x^n(\tilde{a}^n),y^n) - H(Q_{\tilde{A}|B})\\&\quad \geq I(S_X(Q_A);V_{Y|X}) - H(Q_{A|B})\big)\\&\leq \min\Big\{1\\&\sum_{Q_{\tilde{A}B},\, U_{XY}: I_{U_{XY}}(X,Y) - H(Q_{\tilde{A}|B}) \geq I(S_X(Q_A);V_{Y|X}) - H(Q_{A|B})}\\&\quad 1^c(\exists \tilde{a}^n \neq a^n \text{ and } \tilde{a}^n \in Q_{\tilde{A}B}(b^n)\\&\qquad s.t.\ x^n(\tilde{a}^n) \in U_{XY}(Q_{\tilde{A}}, y^n)\Big\}.\end{aligned} \tag{50}$$

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 (49)

$$\begin{aligned}&E\big(1(\exists \tilde{a}^n \neq a^n \text{ and } \tilde{a}^n \in Q_{\tilde{A}B}(b^n)\\&\quad s.t.\ x^n(\tilde{a}^n) \in U_{XY}(Q_{\tilde{A}}, y^n))\big)\\&= \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}} 1^c(\exists \tilde{a}^n \neq a^n \text{ and } \tilde{a}^n \in Q_{\tilde{A}B}(b^n)\\&\quad s.t.\ x^n(\tilde{a}^n) \in U_{XY}(Q_{\tilde{A}}, y^n))\\&= \overset{\zeta}{\Pr}(\exists \tilde{a}^n \neq a^n \text{ and } \tilde{a}^n \in Q_{\tilde{A}B}(b^n)\\&\quad s.t.\ x^n(\tilde{a}^n) \in U_{XY}(Q_{\tilde{A}}, y^n))\\&\leq |Q_{\tilde{A}B}(b^n)|\overset{\zeta}{\Pr}\left(x^n(\tilde{a}^n) \in U_{XY}(Q_{\tilde{A}}, y^n)|\tilde{a}^n \neq a^n\right.\\&\quad \left.\text{and } \tilde{a}^n \in Q_{\tilde{A}B}(b^n)\right) &(51)\\&= |Q_{\tilde{A}B}(b^n)|\frac{|U_{XY}(Q_{\tilde{A}}, y^n)|}{|S_X(Q_{\tilde{A}})|} &(52)\\&\leq 2^{n\epsilon_n}2^{nH(Q_{\tilde{A}|B})}\frac{2^{nH(U_{X|Y})}}{2^{nH(U_X)}} &(53)\\&= 2^{-n(I_{U_{XY}}(X,Y) - H(Q_{\tilde{A}|B}) - \epsilon_n)}\\&\leq 2^{-n(I(S_X(Q_A);V_{Y|X}) - H(Q_{A|B}) - \epsilon_n)} &(54)\end{aligned}$$

where $\epsilon_n \to 0$. Equation (51) is by a union bound argument. Equation (52) is true because the codeword $x^n(\tilde{a}^n)$ is uniformly distributed in $S_X(Q_{\tilde{A}})$. Equation (53) is by the method of types. Equation (54) is true because the condition in (49) is satisfied.

Combining (50) and (54) and noticing that the numbers of types, U_{XY} and $Q_{\tilde{A}B}$, is polynomials of n, hence sub-exponential, we have

$$\begin{aligned}&E(1(a^n \neq \widehat{a}^n(b^n, y^n)))\\ &\leq \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}} 1^c(a^n \neq \widehat{a}^n(b^n, y^n))\\ &\leq \min\left\{1, 2^{-n\left(I(S_X(Q_A);V_{Y|X})-H(Q_{A|B})-\epsilon_n^1\right)}\right\}\\ &= 2^{-n|I(S_X(Q_A);V_{Y|X})-H(Q_{A|B})-\epsilon_n^1|^+}. \qquad (55)\end{aligned}$$

Finally, we substitute (55) and (48) into (47). 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 $P_{AB}(\mathsf{a}^n, \mathsf{b}^n) \in Q_{AB}$) etc.), we have

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

where $\exp\{t\} = 2^t$ and $\epsilon_n^i \to 0$ 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

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

The last equality is true because 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

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

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. ■

6) Upper Bound: [11] 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

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

of the codewords for $a^n \in Q_A$ have the same composition, we write this composition $S_X(Q_A)$, and $A_1 = \{a^n \in Q_A : x^n(a^n) \in S_X(Q_A)\}$, 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 $Q_{A|B}(b^n) = \{a^n : (a^n, b^n) \in Q_{AB}\}$ 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: $|Q_{B|A}(a^n)| = 2^{n(H(Q_{B|A})-\epsilon_n^{4'})}$.

Let

$$B_1 = \left\{b^n \in Q_B : |Q_{A|B}(b^n)\bigcap A_1| \geq 2^{n\left(H(Q_{A|B})-\epsilon_n^5\right)}\right\}$$

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

$$|AB_1| = |A_1||Q_{B|A}(a^n)| = 2^{n\left(H(Q_{AB})-\epsilon_n^2-\epsilon_n^{4'}\right)}. \qquad (56)$$

Secondly

$$\begin{aligned}&|AB_1|\\ &= \Big|\{(a^n, b^n) : b^n \in B_1,\ a^n \in A_1 \text{ and } (a^n, b^n) \in Q_{AB}\}\\ &\qquad \bigcup\{(a^n, b^n) : b^n \in Q_B - B_1,\ a^n \in A_1\\ &\qquad\qquad \text{and } (a^n, b^n) \in Q_{AB}\}\Big| \qquad (57)\\ &\leq |B_1||Q_{A|B}(b^n)| + |Q_B - B_1|2^{n\left(H(Q_{A|B})-\epsilon_n^5\right)} \qquad (58)\\ &= |B_1|2^{n(H(Q_{A|B})-\epsilon_n^4)}\\ &\quad + \left(2^{n(H(Q_B)-\epsilon_n^3)} - |B_1|\right)2^{n\left(H(Q_{A|B})-\epsilon_n^5\right)}\\ &\leq |B_1|2^{nH(Q_{A|B})} + 2^{nH(Q_B)}2^{n\left(H(Q_{A|B})-\epsilon_n^5\right)}. \qquad (59)\end{aligned}$$

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

Equation (57) is by the definition of AB_1 and B_1, (58) is by the definition of B_1, (59) is true because all ϵ_n^i's are positive.

Combining (56) and (59) and use the fact that $\epsilon_n^5 = \epsilon_n^2 + \epsilon_n^{4'} + \frac{1}{n}$, we have

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

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 event gives us an upper bound on the error exponent stated in the 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 $|B_1|$

$$\begin{aligned}|AB^*| &\geq |B_1| \times 2^{n\left(H(Q_{A|B})-\epsilon_n^5\right)\}}\\ &\geq 2^{n\left(H(Q_B)-\epsilon_n^5\right)} \times 2^{n\left(H(Q_{A|B})-\epsilon_n^5\right)}\\ &\geq 2^{n\left(H(Q_{AB})-2\epsilon_n^5\right)}.\end{aligned} \tag{60}$$

So obviously the probability of AB^* is

$$\begin{aligned}P_{AB}(AB^*) &= |AB^*| 2^{-n(D(Q_{AB}\|P_{AB})+H(Q_{AB}))}\\ &\geq 2^{-n\left(D(Q_{AB}\|P_{AB})+2\epsilon_n^5\right)}.\end{aligned} \tag{61}$$

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 if 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 [14].

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

$$\begin{aligned}&\exp\left\{ -n\Big(\min_{V_{Y|X}: I(S_X(Q_A); V_{Y|X}) < H(Q_{A|B}) - \epsilon_n^5} D(V_{Y|X}\|W_{Y|X}|S_X(Q_A)) + \epsilon_n^6\Big)\right\}\\ &\geq \exp\left\{ -n\Big(\min_{V_{Y|X}: I(S_X(Q_A); V_{Y|X}) < H(Q_{A|B})} D(V_{Y|X}\|W_{Y|X}|S_X(Q_A)) + \epsilon_n^7\Big)\right\},\end{aligned} \tag{62}$$

where $\exp\{t\} = 2^t$, 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 (61) and (62), 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

$$\begin{aligned}&\exp\left\{ -n\Big(\min_{Q_A} \max_{S_X(Q_A)} \min_{Q_{B|A}} \{D(Q_{AB}\|P_{AB}) + \min_{V_{Y|X}: I(S_X(Q_A); V_{Y|X}) < H(Q_{A|B})} D(V_{Y|X}\|W_{Y|X}|S_X(Q_A))\} + 2\epsilon_n^5 + \epsilon_n^7\Big)\right\}\\ &= \exp\left\{ -n\Big(\min_{Q_A} \max_{S_X(Q_A)} \min_{[Q_{B|A}, V_{Y|X}: I(S_X(Q_A); V_{Y|X}) < H(Q_{A|B})]} \{D(Q_{AB}\|P_{AB}) + D(V_{Y|X}\|W_{Y|X}|S_X(Q_A))\} + 2\epsilon_n^5 + \epsilon_n^7\Big)\right\}.\end{aligned}$$

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. ■

G) Proof of Corollary 1: The proofs for both lower and upper bounds with the "digital interface" are similar.

8) Proof of (15), 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:

$$\begin{aligned}&E(P_{AB}, W_{Y|X})\\ &\geq \min_{Q_A} \max_{S_X(Q_A)} \min_{Q_{B|A}, V_{Y|X}} \big\{D(Q_{AB}\|P_{AB}) + D(V_{Y|X}\|W_{Y|X}|S_X(Q_A)) + |I(S_X(Q_A); V_{Y|X}) - H(Q_{A|B})|^+\big\}\\ &= \min_{Q_A} \max_{S_X(Q_A)} \min_{R} \big\{ \min_{Q_{B|A}, V_{Y|X}: H(Q_{A|B}) = R} D(Q_{AB}\|P_{AB}) + D(V_{Y|X}\|W_{Y|X}|S_X(Q_A)) + |I(S_X(Q_A); V_{Y|X}) - R|^+\big\}\\ &= \min_{Q_A} \max_{S_X(Q_A)} \min_{R} \big\{ \min_{Q_{B|A}: H(Q_{A|B}) = R} D(Q_{AB}\|P_{AB}) + E_r(R, S_X(Q_A), W_{Y|X})\big\} \quad (63)\\ &\geq \min_{Q_A} \max_{S_X(Q_A)} \min_{R} \big\{ \min_{Q_{B|A}: H(Q_{A|B}) \geq R} D(Q_{AB}\|P_{AB}) + E_r(R, S_X(Q_A), W_{Y|X})\big\} \quad (64)\\ &= \min_{Q_A} \max_{S_X(Q_A)} \min_{R} \big\{ e_U(R, P_{AB}, Q_A) + E_r(R, S_X(Q_A), W_{Y|X})\big\} \quad (65)\end{aligned}$$

where in (63), $E_r(R, S_X(Q_A), W_{Y|X})$ is the standard random coding error exponent for channel $W_{Y|X}$ at rate R and input distribution $S_X(Q_A)$, (64) is trivial and $e_U(R, P_{AB}, Q_A)$, defined in (16), is the upper bound of the error exponent of a peculiar source coding problem with side-information for source P_{AB} at rate R, where the empirical source distribution is fixed at Q_A. □

9) Proof of (14), the Upper Bound: Similar to the proof for (15), we have the following equalities:

$$
\begin{aligned}
&E(P_{AB}, W_{Y|X}) \\
&\le \min_{Q_A} \max_{S_X(Q_A)} \min_{Q_{B|A}, V_{Y|X}: I(S_X(Q_A); V_{Y|X}) < H(Q_{A|B})} \\
&\quad \{D(Q_{AB} \| P_{AB}) + D(V_{Y|X} \| W_{Y|X} | S_X(Q_A))\} \\
&= \min_{Q_A} \max_{S_X(Q_A)} \min_{R} \min_{Q_{B|A}, V_{Y|X}: I(S_X(Q_A); V_{Y|X}) < R < H(Q_{A|B})} \\
&\quad \{D(Q_{AB} \| P_{AB}) + D(V_{Y|X} \| W_{Y|X} | S_X(Q_A))\} \\
&= \min_{Q_A} \max_{S_X(Q_A)} \min_{R} \{ \min_{Q_{B|A}: H(Q_{A|B}) > R} D(Q_{AB} \| P_{AB}) \\
&\quad + \min_{V_{Y|X}: I(S_X(Q_A); V_{Y|X}) < R} D(V_{Y|X} \| W_{Y|X} | S_X(Q_A))\} \\
&= \min_{Q_A} \max_{S_X(Q_A)} \min_{R} \{ \min_{Q_{B|A}: H(Q_{A|B}) > R} D(Q_{AB} \| P_{AB}) \\
&\quad + E_{sp}(R, S_X(Q_A), W_{Y|X})\} \qquad (66) \\
&= \min_{Q_A} \max_{S_X(Q_A)} \min_{R} \{ e_U(R, P_{AB}, Q_A) \\
&\quad + E_{sp}(R, S_X(Q_A), W_{Y|X})\} \qquad (67)
\end{aligned}
$$

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 (16). □

ACKNOWLEDGMENT

The author would like to thank Neri Merhav and Filippo Balestrieri for pointing out Sion's theorem in [17], and Fady Alajaji for many helpful discussions on the technical results of this paper. The author would also like to thank the Associate Editor, anonymous reviewers and Hari Palaiyanur for their insightful comments that helped to greatly improve the presentation of the paper.

REFERENCES

[1] C. Chang, "Joint source-channel with side information coding error exponents," in *HP Labs Tech. Rep..* : , 2009, vol. HPL-2009-18.
[2] C. Chang, "Joint source-channel with side information coding error exponents," presented at the 48th Annu. Allerton Conf. Communications, Control, and Computing, Oct. 2009.
[3] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, 1948.
[4] I. Csiszár, "Joint source-channel error exponent," *Probl. Control Inf. Theory*, vol. 9, pp. 315–328, 1980.
[5] Y. Zhong, "Joint Source-Channel Coding Reliability Function for Single and Multi-Terminal Communication Systems," Ph.D. dissertation, Queen's University, London, U.K., 2008.
[6] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, 1973.
[7] R. Gallager, Source Coding With Side Information and Universal Coding, Mass. Instit. Technol., 1976, Tech. Rep. LIDS-P-937.
[8] I. Csiszár and J. Körner*, Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
[9] R. Ahlswede, "Coloring hypergraphs: A new approach to multi-user source coding II," *J. Combin., Inf. Syst. Sci.*, vol. 5, pp. 220–268, 1980.
[10] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein*, Covering Codes*. New York: Elsevier, 1997.
[11] S. Shamai(Shitz) and S. Verdú, "Capacity of channels with side information," *Eur. Trans. Telecommun.*, vol. 6, no. 5, pp. 587–600, Sep./Oct. 1995.
[12] Y. Zhong, F. Alajaji, and L. L. Campbell, "Source side information can increase the joint source-channel coding error exponent," presented at the 10th Canadian Workshop on Information Theory, Edmonton AB, Canada, Jun. 2007.
[13] R. Gallager*, Information Theory and Reliable Communication*. Hoboken, NJ: Wiley, 1971.
[14] R. Gallager, Fixed Composition Arguments and Lower Bounds to Error Probability [Online]. Available: http://web.mit.edu/gallager/www/notes/notes5.pdf
[15] I. Csiszár, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2505–2523, 1998.
[16] S. Boyd and L. Vandenberghe*, Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[17] M. Sion, "On general minmax theorems," *Pacific J. Math.*, vol. 8, pp. 171–176, 1958.
[18] C. Chang and A. Sahai, "The price of ignorance: The impact on side-information for delay in lossless source coding," *IEEE Trans. Inf. Theory*, submitted for publication.
[19] B. Kelly and A. B. Wagner, "Error exponents and test channel optimization for the Wyner-Ziv problem," presented at the 46th Annu. Allerton Conf. Communications, Control, and Computing, Oct. 2007.

Cheng Chang (M'08) received a B.E. degree from Tsinghua University, Beijing, in 2000, and a Ph.D. degree from University of California at Berkeley in 2007. He is currently a quantitative analyst with the D. E. Shaw Group in New York. In 2008, he spent a year in the information theory group at HP Labs as a postdoctoral researcher. His research interests are signal processing, control theory, information theory and machine learning.