A Strong Converse for Joint Source-Channel Coding

Da Wang
EECS Dept., MIT
Cambridge, MA, USA
Email: dawang@mit.edu

Amir Ingber
EE Dept., Stanford University
Stanford, CA, USA
Email: ingber@stanford.edu

Yuval Kochman
School of CSE, HUJI
Jerusalem, Israel
Email: yuvalko@cs.huji.ac.il

Abstract—We consider a discrete memoryless joint source-channel setting. In this setting, if a source sequence is reconstructed with distortion below some threshold, we declare a success event. We prove that for any joint source-channel scheme, if this threshold lower (better) than the optimum average distortion, then the success probability approaches zero as the block length increases. Furthermore, we show that the probability has an exponential behavior, and evaluate the optimal exponent. Surprisingly, the best exponential behavior is attainable by a separation-based scheme.

I. Introduction

Information Theory produces sharp results: if the required performance is below a threshold, set by the problem parameters, then it may be achieved reliably, i.e., with probability that approaches one; otherwise, it can not. Converse results may be divided into categories, according to their strength, measured asymptotically as the block length goes to infinity. A *weak converse* indicates it is impossible for the probability of success to approach one. A *strong converse* further states that the success probability must approach zero. Beyond that, one may be interested in the rate in which this happens; specifically, an "exponentially-strong" converse states that the probability of success must approach zero exponentially. Indeed, for source and channel coding exponentially strong converses are known.

For a discrete memoryless channel (DMC) W, when the required channel rate R is above the channel capacity C, the probability of correct decoding decays with the following exponent [1] (Earlier in [2] it appears in a different form and proven to be a lower bound on the exponent):[1]

$$\bar{E}_{\mathrm{sp}}\left(R, W\right) = \max_{\Phi} \min_{V} \left[D\left(V \| W | \Phi\right) + |R - I\left(\Phi, V\right)|^{+} \right].$$

$$(1)$$

This exponent is closely related the sphere-packing exponent, which is an upper bound on the exponent of the error probability when $R < C$ [3] (in this form, [4]):

$$E_{\mathrm{sp}}\left(R, W\right) \triangleq \max_{\Phi} \min_{V : I(\Phi, V) \le R} D\left(V \| W | \Phi\right).$$

In lossy source coding, dual results can be obtained where the *excess-distortion probability* plays the role of error probability. Specifically, for some single-letter distortion measure

[1]For definition of of divergence and mutual information, see notation section in the sequel.

$d(\cdot, \cdot)$ and distortion threshold D, we say that the scheme is successful if

$$\frac{1}{n} \sum_{i=1}^{n} d(S_i, \hat{S}_i) \le D,$$

where S and $\hat{S}$ are the source and reconstruction sequences respectively.

For a discrete memoryless source (DMS) P, when the rate R is below the rate-distortion function (RDF) $R(P, D)$[2], the success probability has exponent [5, Problem 9.6]:

$$\bar{E}_{\mathrm{S}}\left(R, D, P\right) = \min_{Q} \left[D\left(Q \| P\right) + |R(Q, D) - R|^{+} \right]. \quad (2)$$

Similarly, this exponent is closely related to the lossy source coding excess-distortion exponent for $R > R(P, D)$ [6]:

$$E_{\mathrm{S}}\left(R, D, P\right) = \min_{Q : R(Q, D) \ge R} D\left(Q \| P\right).$$

In joint source-channel coding (JSCC), an average distortion D is achievable if $R(D) < \rho C$ and not achievable if the opposite holds, where ρ is the *bandwidth expansion factor* (number of channel uses per source sample). This result, due to Shannon [7], immediately implies a weak converse for the excess-distortion probability.

A (non-exponential) strong converse for JSCC has been shown for special cases: Quadratic-Gaussian [8], channels with additive noise [9], and lossless JSCC [10]. However, it has not been explicitly shown in general. We note that the strong converse for JSCC may be derived using previously known results. One approach is to use equivalence to channel coding [11] and the strong channel converse. Alternatively, JSCC dispersion [12] implies a strong converse. Another approach is the information spectrum method, which is used by Han in deriving the lossless JSCC case [10].

In this work we formally prove the strong converse for general lossy JSCC, and furthermore our converse is "exponentially-strong". We do that by showing that whenever $R(D) > \rho C$, the probability of not having excess distortion for the optimal JSCC scheme decays with exponent

$$\bar{E}_{\mathrm{JSCC}}\left(P, D, W, \rho\right)$$
$$= \min_{R}[\bar{E}_{\mathrm{S}}\left(R, D, P\right) + \rho \bar{E}_{\mathrm{sp}}\left(R/\rho, W\right)], \quad (3)$$

where the channel and source exponents are given by (1) and (2), respectively. This is analogous to the exponent of

[2]We write $R(P, D)$ as $R(D)$ when P is understood from context.

the excess distortion: the JSCC excess-distortion exponent for $R(D) < \rho C$ is upper-bounded by [13], [14]

$$\min_R \left[E_S \left(R, D, P \right) + \rho E_{\text{sp}} \left(R/\rho, W \right) \right]. \tag{4}$$

II. NOTATIONS

This paper uses lower case letters (e.g. x) to denote a particular value of the corresponding random variable denoted in capital letters (e.g. X). Vectors are denoted in bold (e.g. $\mathbf{x}$ or $\mathbf{X}$). Calligraphic fonts (e.g. $\mathcal{X}$) represent a set and $\mathcal{P}(\mathcal{X})$ denotes all the probability distributions on the alphabet $\mathcal{X}$. We use $\mathbb{Z}_+$ and $\mathbb{R}_+$ to denote the set of non-negative integer and real numbers respectively.

Our proofs make use of the method of types, and follow the notations in [5]. Specifically, the *type* of a sequence $\mathbf{x}$ with length n is denoted by $P_{\mathbf{x}}$, where the type is the empirical distribution of this sequence, i.e., $P_{\mathbf{x}}(a) = N(a|\mathbf{x})/n \, \forall a \in \mathcal{X}$, where $N(a|\mathbf{x})$ is the number of occurrences of a in sequence $\mathbf{x}$. The subset of the probability distributions $\mathcal{P}(\mathcal{X})$ that can be types of n-sequences is denoted as

$$\mathcal{P}_n(\mathcal{X}) \triangleq \{ P \in \mathcal{P}(\mathcal{X}) : nP(x) \in \mathbb{Z}_+, \forall x \in \mathcal{X} \} \tag{5}$$

and sometimes P_n is used to emphasize the fact that $P_n \in \mathcal{P}_n(\mathcal{X})$. A *type class* $\mathcal{T}_{P_{\mathbf{x}}}^n$ is defined as the set of sequences that have type $P_{\mathbf{x}}$. Given some sequence $\mathbf{x}$, a sequence $\mathbf{y}$ of the same length has *conditional type* $P_{\mathbf{y}|\mathbf{x}}$ if $N(a, b|\mathbf{x}, \mathbf{y}) = P_{\mathbf{y}|\mathbf{x}}(a|b)N(a|\mathbf{x})$, and we call the set of all $\mathbf{y}$ that have conditional type V given $\mathbf{x}$ the *V-shell* of $\mathbf{x}$, and denote it by $\mathcal{T}_V^n(\mathbf{x})$. Furthermore, the random variable corresponding to the conditional type of a random vector $\mathbf{Y}$ given $\mathbf{x}$ is denoted as $P_{\mathbf{Y}|\mathbf{x}}$. In addition, the possible conditional type given an input distribution $P_{\mathbf{x}}$ is denoted as

$$\mathcal{P}_n(\mathcal{Y}|P_{\mathbf{x}}) \triangleq \{ P_{\mathbf{y}|\mathbf{x}} : P_{\mathbf{x}} \times P_{\mathbf{y}|\mathbf{x}} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}) \}.$$

Given two distributions P and Q, their KL-divergence is defined as

$$D \left(P \parallel Q \right) \triangleq \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}. \tag{6}$$

A discrete memoryless channel $W : \mathcal{X} \to \mathcal{Y}$ is defined with its input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and conditional distribution $W \left(\cdot \mid x \right)$ of output letter Y when the channel input letter X equals $x \in \mathcal{X}$. Also, we abbreviate $W \left(\cdot \mid x \right)$ as $W_x(\cdot)$ for notational simplicity. Given a channel input distribution Φ, we denote the corresponding channel output distribution, and mutual information by

$$\Phi W(y) \triangleq \sum_{x \in \mathcal{X}} W \left(y \mid x \right) \Phi(x),$$

$$I \left(\Phi, W \right) \triangleq \sum_{x, y} \Phi(x) W(y|x) \log \frac{W(y|x)}{\Phi W(y)}.$$

And we denote channel capacity and the set of capacity-achieving distributions by $C(W) \triangleq \max_\Phi I \left(\Phi, W \right)$ and $\Pi(W) \triangleq \{ \Phi : I \left(\Phi, W \right) = C(W) \}$ respectively.

A discrete memoryless source is defined with source alphabet $\mathcal{S}$, reproduction alphabet $\hat{\mathcal{S}}$, source distribution P and a distortion measure $d : \mathcal{S} \times \hat{\mathcal{S}} \to \mathbb{R}_+$. Without loss of generality, we assume that for any $s \in \mathcal{S}$ there is $\hat{s} \in \hat{\mathcal{S}}$ such that

$d(s, \hat{s}) = 0$. The rate-distortion function (RDF) of a DMS $(\mathcal{S}, \hat{\mathcal{S}}, P, d)$ is given by

$$R(P, D) = \min_{\Lambda : E_{P, \Lambda} d(S, \hat{S}) \leq D} I(P, \Lambda), \tag{7}$$

where

$$d(\mathbf{s}, \hat{\mathbf{s}}) \triangleq \frac{1}{n} \sum_{i=1}^n d(s_i, \hat{s}_i) \tag{8}$$

is the distortion between the source and reproduction words $\mathbf{s}$ and $\hat{\mathbf{s}}$ and $I(P, \Lambda)$ is the mutual information over a channel with input distribution $P(S)$ and conditional distribution $\Lambda : \mathcal{S} \to \hat{\mathcal{S}}$.

A discrete memoryless joint source-channel coding problem consists of a DMS $(\mathcal{S}, \hat{\mathcal{S}}, P, d)$, a DMC $W : \mathcal{X} \to \mathcal{Y}$ and a *bandwidth expansion factor* $\rho \in \mathbb{R}_+$. A JSCC scheme $\mathcal{C}_{\text{JSCC}}^{(n)}$ is comprised of an encoder mapping $f_{J;n} : \mathcal{S}^n \to \mathcal{X}^{\lfloor \rho n \rfloor}$ and decoder mapping $g_{J;n} : \mathcal{Y}^{\lfloor \rho n \rfloor} \to \hat{\mathcal{S}}^n$. Given a source block $\mathbf{s}$, the encoder maps it to a sequence $\mathbf{x} = f_{J;n}(\mathbf{s}) \in \mathcal{X}^{\lfloor \rho n \rfloor}$ and transmits this sequence through the channel. The decoder receives a sequence $\mathbf{y} \in \mathcal{Y}^{\lfloor \rho n \rfloor}$ distributed according to $W(\cdot|\mathbf{x})$, and maps it to a source reconstruction $\hat{\mathbf{s}}$. The corresponding distortion is given by (8). For a given JSCC scheme, we define the *error* event $\mathcal{E}(D)$ as

$$\mathcal{E}(D) \triangleq \mathcal{E}(D, f_{J;n}, g_{J;n}) \triangleq \mathcal{E}\left(D, \mathcal{C}_{\text{JSCC}}^{(n)} \right)$$

$$\triangleq \{ d(\mathbf{S}, \hat{\mathbf{S}}) > D \},$$

and the *correct* event $\bar{\mathcal{E}}(D) \triangleq \mathcal{E}(D)^c = \{ d(\mathbf{S}, \hat{\mathbf{S}}) \leq D \}$. Finally, for block length n, we define the *best* correct event $\bar{\mathcal{E}}_n(D)$ as an event that corresponds to the JSCC scheme that produces the minimum error probability, i.e.,

$$\bar{\mathcal{E}}_n(D) \in \underset{\left\{ \mathcal{C}_{\text{JSCC}}^{(n)} \right\}}{\arg \min} \mathbb{P} \left[\bar{\mathcal{E}}(D) \right].$$

III. MAIN RESULT

The following formally states the exponential decay rate of the probability of success at distortion thresholds $R(D) > \rho C$, thus also serves as a strong converse for JSCC coding.

Theorem 1 (Strong Converse for JSCC). *Given a discrete memoryless JSCC problem with DMS $(\mathcal{S}, \hat{\mathcal{S}}, P, d)$, DMC $(\mathcal{X}, \mathcal{Y}, W)$ and bandwidth expansion factor ρ, let $\bar{E}(P, D, W, \rho)$ be the exponent of the success probability for the best sequence of JSCC schemes:*

$$\bar{E}(P, D, W, \rho) \triangleq \lim_{n \to \infty} -\frac{1}{n} \log \mathbb{P} \left[\bar{\mathcal{E}}_n(D) \right].$$

Then $\bar{E}(P, D, W, \rho)$ exists and is given by $\bar{E}_{\text{JSCC}}(P, D, W, \rho)$ (3).

Remark 1 (Direct part of the theorem). *The achievability of the exponent $\bar{E}_{\text{JSCC}}(P, D, W, \rho)$ can be proven by a separation scheme, noting that the probability of no excess distortion is lower-bounded by the product of the probability of no excess distortion in source coding and the probability of no channel error, for any chosen digital rate R. These probabilities have exponents $\bar{E}_S(R, D, P)$ and $\rho \bar{E}_{\text{sp}}(R/\rho, W)$, respectively (see Section I). Thus this proves separation theorem for the (rather strange from a practical point of view) case*

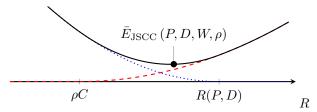Fig. 1. When $R(P,D) > \rho C$, (3) is always minimized by a rate R such that $R(P,D) > R > \rho C$, where in the plot the dashed curve is $\rho \bar{E}_{\text{sp}}(R/\rho, W)$, the dotted curve is $\bar{E}_{\text{S}}(R,D,P)$ and the solid curve is $\bar{E}_{\text{S}}(R,D,P) + \rho \bar{E}_{\text{sp}}(R/\rho, W)$.

where one is interested in achieving a distortion threshold with exponentially small probability.

Remark 2 (Alternative form). *The exponent $\bar{E}_{\text{JSCC}}(P,D,W,\rho)$ may be written explicitly as a function of the source and channel parameters as follows:*

$$\bar{E}_{\text{JSCC}}(P,D,W,\rho)$$
$$= \min_{Q \in \mathcal{P}(\mathcal{S})} \left[D(Q \| P) + \max_{\Phi \in \mathcal{P}(\mathcal{X})} \min_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \left(\rho D(V \| W | \Phi) \right. \right.$$
$$\left. \left. + |R(Q,D) - \rho I(\Phi, V)|^+ \right) \right]. \quad (9)$$

We use this form in the proof of Theorem 1, rather than the form (3). The equivalence proof can be shown by letting Φ^*, V^* and Q^* be the optimizing distributions for (9), and noting that regardless of the $|\cdot|^+$ operation taking effect, (9) can be decomposed into two parts that correspond to $\bar{E}_{\text{S}}(R,D,P)$ and $\rho \bar{E}_{\text{sp}}(R/\rho, W)$. We omit the detailed proof due to space limits.

Remark 3 (Minimizing rate). *When $R(P,D) \leq \rho C$, proving Theorem 1 is trivial as $\bar{E}(P,D,W,\rho) = 0$. When $R(P,D) > \rho C$, the rate minimizing (3) satisfies $R(P,D) > R > \rho C$, as shown in Fig. 1. This is parallel to the excess-distortion exponent where $R(P,D) < \rho C$, where (4) is minimized by a rate $R(P,D) < R < \rho C$.*

The proof of our main result builds on the following key lemma, which shows for any JSCC scheme, when the source, channel and channel codebook are all constant-composition, the probability of a no-excess distortion event vanishes exponentially with the exponent in (11).

Lemma 2 (Joint source channel coding converse with fixed types). *For a JSCC problem, we define all the channel outputs that covers some source sequence $\mathbf{s}$ with distortion D as $\hat{B}(\mathbf{s}, D)$, i.e.,*

$$\hat{B}(\mathbf{s}, D) \triangleq \{\mathbf{y} \in \mathcal{Y}^m : d(\mathbf{s}, g_{J;n}(\mathbf{y})) \leq D\} \quad (10)$$

where $m = \lfloor \rho n \rfloor$ and $g_{J;n}$ is the JSCC decoder. Then given a source type $Q \in \mathcal{P}_n(\mathcal{S})$ and a channel input type $\Phi \in \mathcal{P}_n(\mathcal{X})$, for a distortion D and a channel with constant composition conditional distribution $V \in \mathcal{P}_m(\mathcal{Y}|\Phi)$, we have

$$\frac{1}{|G(Q,\Phi)|} \sum_{\mathbf{s}_i \in G(Q,\Phi)} \frac{\left| \mathcal{T}_V^m(f(\mathbf{s}_i)) \cap \hat{B}(\mathbf{s}_i, D) \right|}{\left| \mathcal{T}_V^m(f(\mathbf{s}_i)) \right|}$$
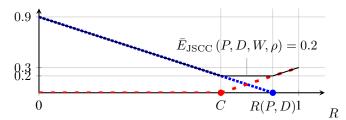$$\leq \frac{p(n)}{\alpha(Q,\Phi)} \exp^{-n[R(Q,D) - \rho I(\Phi, V)]^+}, \quad (11)$$

Fig. 2. $\bar{E}_{\text{JSCC}}(P,D,W,\rho)$ with $\rho = 1$, $H_b(D) = 0.1$ and $H_b(\varepsilon) = 0.3$, where in the plot the dashed curve is $\bar{E}_{\text{sp}}(R,W)$, the dotted curve is $\bar{E}_{\text{S}}(R,D,P)$ and the solid curve is $\bar{E}_{\text{S}}(R,D,P) + \bar{E}_{\text{sp}}(R,W)$.

where

$$\alpha(Q,\Phi) \triangleq \mathbb{P}[P_{\mathbf{X}} = \Phi | P_{\mathbf{S}} = Q] \quad (12)$$

is the probability of having channel input type Φ given that the source type is Q, $G(Q,\Phi)$ is the set of source sequences in $\mathcal{T}_Q^n$ that are mapped (via JSCC encoder $f_{J;n}$) to channel codewords with type Φ, i.e.,

$$G(Q,\Phi) \triangleq \{\mathbf{s} \in \mathcal{T}_Q^n : \mathbf{x} = f_{J;n}(\mathbf{s}) \in \mathcal{T}_\Phi^m\}, \quad (13)$$

and $p(n)$ is a polynomial that depends only on the source, channel and reconstruction alphabet sizes and ρ.

The proof of this lemma is based upon the exponential channel coding converse [1], combined with the following.

Lemma 3 (Restricted D-ball size). *Given source type P and a reconstruction sequence $\hat{\mathbf{s}}$, define restricted D-ball as*

$$B(\hat{\mathbf{s}}, P, D) \triangleq \{\mathbf{s} \in \mathcal{T}_P^n : d(\mathbf{s}, \hat{\mathbf{s}}) \leq D\}.$$

Then

$$|B(\hat{\mathbf{s}}, P, D)| \leq (n+1)^{|\mathcal{S}||\hat{\mathcal{S}}|} \exp\{n[H(P) - R(P,D)]\}.$$

This result is similar to the bound over the size of the restricted D-ball presented in Lemma 3 in [15]. However, This bound is *uniform* over source types and does not depend on the reconstruction alphabet. We prove this lemma, as well as Lemma 2 and the converse part of Theorem 1 in the next section[3]. We conclude this section by presenting the following example.

Example 1. *Transmitting a binary symmetric source (BSS) over a binary symmetric channel (BSC) subject to the Hamming distortion with the bandwidth expansion factor $\rho = 1$.*

For a BSS, the RDF is given by $R(P,D) = 1 - H_b(D)$, where $H_b(\cdot)$ is the binary entropy function. It can be shown that (1) is always minimized by a uniform distribution and:

$$\bar{E}_{\text{S}}(R,D,P) = |1 - H_b(D) - R|^+.$$

For a BSC with cross over probability ε, the capacity is given by $C(W) = 1 - H_b(\varepsilon)$. it can be shown that the optimizing Φ and V in (1) are always symmetric and:

$$\bar{E}_{\text{sp}}(R,W) = |R - 1 + H_b(\varepsilon)|^+.$$

Therefore, when $R(P,D) > C(W)$, i.e. $D < \varepsilon$,

$$\bar{E}_{\text{JSCC}}(P,D,W) = \inf_R [\bar{E}_{\text{S}}(R,D,P) + \bar{E}_{\text{sp}}(R,W)]$$
$$= R(P,D) - C(W) = H_b(\varepsilon) - H_b(D).$$

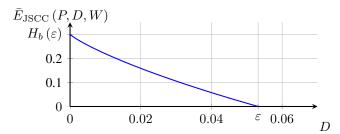[3]Lemma 2 and Lemma 3 are also used in [12] for JSCC dispersion, but the proofs were not included there.

Fig. 3. $\bar{E}(P, D, W, \rho)$ as a function of D for Example 1, where $\rho = 1$ and $H_b(\varepsilon) = 0.3$.

For the case of $H_b(D) = 0.1$ and $H_b(\varepsilon) = 0.3$, we plot the $\bar{E}_S(R, D, P)$, $\bar{E}_{sp}(R, W)$ and $\bar{E}_{JSCC}(P, D, W)$ in Fig. 2. Finally, we show $\bar{E}_{JSCC}(P, D, W)$ as a function of D when $H_b(\varepsilon) = 0.3$ in Fig. 3.

IV. PROOFS

A. Proof of Key Lemmas

Proof of Lemma 3: Let $P \in \mathcal{P}_n(\mathcal{S})$ be a given type and let Q be the type of $\hat{s}$. Then the size of the set of source codewords with type P that are D-covered by $\hat{s}$ is

$$|B(\hat{s}, P, D)| = \left| \bigcup_{\substack{\Lambda: \mathbb{E}[d(S,\hat{S})] \le D, \\ P\Lambda = Q}} \{s \in \mathcal{T}_P^n : P_{s,\hat{s}} = P \times \Lambda\} \right|.$$

Note there are at most $(n+1)^{|\mathcal{S}||\hat{\mathcal{S}}|}$ joint types, and

$$\{s \in \mathcal{T}_P^n : P_{s,\hat{s}} = P \times \Lambda\} = \mathcal{T}_{\tilde{\Lambda}}^n(\hat{s}),$$

where $\tilde{\Lambda}$ is the reverse channel from $\hat{S}$ to S such that $Q \times \tilde{\Lambda} = P \times \Lambda$. Therefore,

$$|B(\hat{s}, P, D)| \le \sum_{\substack{\tilde{\Lambda}: \mathbb{E}_{Q,\tilde{\Lambda}}[d(\hat{S},S)] \le D,}} \left| \mathcal{T}_{\tilde{\Lambda}}^n(\hat{s}) \right|$$

$$\le (n+1)^{|\mathcal{S}||\hat{\mathcal{S}}|} \exp\left[n \max_{\tilde{\Lambda}: \mathbb{E}_{Q,\tilde{\Lambda}}[d(\hat{S},S)] \le D,} H\left(\tilde{\Lambda}|Q\right) \right].$$

Note

$$R(P, D) = \min_{\Lambda: \mathbb{E}_{P,\Lambda}[d(S,\hat{S})] \le D} I(P, \Lambda)$$

$$= H(P) - \max_{\tilde{\Lambda}: \mathbb{E}_{Q,\tilde{\Lambda}}[d(\hat{S},S)] \le D} H\left(\tilde{\Lambda}|Q\right),$$

hence

$$|B(\hat{s}, P, D)| \le (n+1)^{|\mathcal{S}||\hat{\mathcal{S}}|} \exp\left\{ n[H(P) - R(P, D)] \right\}.$$ ∎

Proof for Lemma 2: In our proof, we first bound the denominator in (11) uniformly for all s_i, and then bound the sum of the numerator over all s_i, as done in [1] for the channel error exponent.

a) Bounding the denominator: Standard results in method of types [5] show for $f(s) \in \mathcal{T}_\Phi^n$,

$$(m+1)^{-|\mathcal{X}||\mathcal{Y}|} \exp\{mH(V|\Phi)\} \le |\mathcal{T}_V^m(f(s))|,$$

$$\frac{1}{|\mathcal{T}_V^m(f(s))|} \le (m+1)^{|\mathcal{X}||\mathcal{Y}|} \exp\{-mH(V|\Phi)\}.$$

b) Bounding the sum of numerator: Since $s \in G(Q, \Phi)$,

$$y \in \mathcal{T}_V(f(s)) \cap \hat{B}(s, D)$$
$$\Rightarrow s \in B(g_{J;n}(y), Q, D) \cap G(Q, \Phi). \quad (14)$$

Therefore, any y will be counted at most $|B(g_{J;n}(y), Q, D) \cap G(Q, \Phi)|$ times. According to Lemma 3, this is upper bounded by $B_u = (n+1)^{|\mathcal{S}||\hat{\mathcal{S}}|} \exp\{n[H(Q) - R(Q, D)]\}$. In addition, it is obvious that

$$\bigcup_{s_i \in G(Q, \Phi)} \mathcal{T}_V(f(s_i)) \cap \hat{B}(s_i, D) \subset \mathcal{T}_\Psi^n,$$

where $\Psi = \Phi V$ is the channel output distribution corresponding to Φ. Therefore,

$$\frac{\alpha(Q, \Phi)}{|G(Q, \Phi)|} \sum_{s_i \in G(Q, \Phi)} \left| \mathcal{T}_V(f(s_i)) \cap \hat{B}(s_i, D) \right|$$

$$\le \frac{1}{|\mathcal{T}_Q^n|} B_u \left| \bigcup_{s_i \in \mathcal{T}_Q^n} \mathcal{T}_V(f(s_i)) \cap \hat{B}(s_i, D) \right| \le \frac{1}{|\mathcal{T}_Q^n|} B_u |\mathcal{T}_\Psi^n|.$$

Noting

$$(n+1)^{-|\mathcal{S}|} \exp\{nH(Q)\} \le |\mathcal{T}_Q^n|$$
$$|\mathcal{T}_\Psi^m| \le \exp\{mH(\Psi)\},$$

$$\frac{1}{n} \log\left[\frac{1}{|\mathcal{T}_Q^n|} B_u |\mathcal{T}_\Psi^n| \right]$$

$$\le \frac{|\mathcal{S}|}{n} \log(n+1) - H(Q) + \rho H(\Psi)$$

$$+ \frac{|\mathcal{S}||\hat{\mathcal{S}}|}{n} \log(n+1) + H(Q) - R(Q, D)$$

$$\le \rho H(\Psi) - R(Q, D)$$

$$+ \frac{|\mathcal{S}||\hat{\mathcal{S}}|}{n} \log(n+1) + \frac{|\mathcal{S}|}{n} \log(n+1).$$

Combining the bounds for both numerator and denominator, we have

$$\frac{1}{n} \log\left[\frac{\alpha(Q, \Phi)}{|G(Q, \Phi)|} \sum_{s_i \in G(Q, \Phi)} \frac{\left| \mathcal{T}_V(f(s_i)) \cap \hat{B}(s_i, D) \right|}{\mathcal{T}_V(f(s_i))} \right]$$

$$\le \rho H(\Psi) - \rho H(V|\Phi) - R(Q, D) + \frac{|\mathcal{X}||\mathcal{Y}|}{m} \log(m+1)$$

$$+ \frac{|\mathcal{S}||\hat{\mathcal{S}}|}{n} \log(n+1) + \frac{|\mathcal{S}|}{n} \log(n+1).$$

Note $m = \lfloor \rho n \rfloor \le \rho n$, let

$$p(n) = (\rho n + 1)^{\lceil \rho |\mathcal{X}||\mathcal{Y}| \rceil} (n+1)^{(|\mathcal{S}||\hat{\mathcal{S}}|+|\mathcal{S}|)}, \quad (15)$$

and the proof is completed. ∎

B. Proof of Main Result

Proof for Theorem 1 (converse part): As discussed in Remark 3, we only need to consider the case that $\rho C(W) < R(P, D)$.

Let $\mathbb{P}\left[\bar{\mathcal{E}}(D)\right] = 1 - \mathbb{P}[\mathcal{E}(D)]$. By following a similar argument in [1, Proof of Lemma 5], clearly,

$$\mathbb{P}\left[\bar{\mathcal{E}}(D)\right]$$
$$\leq (n+1)^{|\mathcal{S}|} \max_{Q \in \mathcal{P}_n(\mathcal{S})} \mathbb{P}\left[\bar{\mathcal{E}}(D) \,\middle|\, P_{\mathbf{S}} = Q\right] e^{-nD(Q \,\|\, P)}. \quad (16)$$

Let $m \triangleq \lfloor \rho n \rfloor$ and let $\mathcal{A}_m$ denote the channel codebook in this JSCC scheme. Then conditioning again, we have

$$\mathbb{P}\left[\bar{\mathcal{E}}(D)|P_{\mathbf{S}} = Q\right]$$
$$= \sum_{\Phi \in \mathcal{A}_m} \mathbb{P}[P_{\mathbf{X}} = \Phi|P_{\mathbf{S}} = Q] \, \mathbb{P}\left[\bar{\mathcal{E}}(D)|P_{\mathbf{S}} = Q, P_{\mathbf{x}} = \Phi\right].$$

Recall the definitions for $G(Q, \Phi)$ and $\alpha(Q, \Phi)$ in (12) and (13), noting that given a source type, all strings within a type class are equally likely, hence

$$\alpha(Q, \Phi) = \frac{|\{\mathbf{s} \in \mathcal{T}_Q^n : \mathbf{x} = f_{J;n}(\mathbf{s}) \in \mathcal{T}_\Phi^m\}|}{|\mathcal{T}_Q^n|} = \frac{|G(Q, \Phi)|}{|\mathcal{T}_Q^n|}.$$

Therefore,

$$\mathbb{P}\left[\bar{\mathcal{E}}(D)|P_{\mathbf{S}} = Q\right]$$
$$= \sum_{\Phi \in \mathcal{A}_m} \alpha(Q, \Phi) \mathbb{P}\left[\bar{\mathcal{E}}(D)|P_{\mathbf{S}} = Q, P_{\mathbf{x}} = \Phi\right]$$
$$= \sum_{\Phi \in \mathcal{A}_m} \alpha(Q, \Phi) \Bigg(\sum_{V \in \mathcal{P}_m(\mathcal{Y}|\Phi)} \mathbb{P}\left[P_{\mathbf{y}|\mathbf{x}} = V|P_{\mathbf{x}} = \Phi\right]$$
$$\cdot \mathbb{P}\left[\bar{\mathcal{E}}(D)|P_{\mathbf{S}} = Q, P_{\mathbf{x}} = \Phi, P_{\mathbf{y}|\mathbf{x}} = V\right] \Bigg).$$

Next we use Lemma 2 to assert,

$$\mathbb{P}\left[\bar{\mathcal{E}}(D)|P_{\mathbf{S}} = Q, P_{\mathbf{x}} = \Phi, P_{\mathbf{y}|\mathbf{x}} = V\right]$$
$$\leq \frac{1}{|G(Q, \Phi)|} \sum_{\mathbf{s}_i \in G(Q, \Phi)} \frac{\left|\mathcal{T}_V^m\left(f(\mathbf{s}_i)\right) \cap \hat{B}(\mathbf{s}_i, D)\right|}{|\mathcal{T}_V^m\left(f(\mathbf{s}_i)\right)|}$$
$$\leq \frac{p(n)}{\alpha(Q, \Phi)} \exp\left\{-n \left|R(Q, D) - \rho I\left(\Phi, V\right)\right|^+\right\}$$

where $p(n)$ is a polynomial given in (15). Therefore,

$$\mathbb{P}\left[\bar{\mathcal{E}}(D)|P_{\mathbf{S}} = Q\right]$$
$$\leq \sum_{\Phi \in \mathcal{A}_m} \alpha(Q, \Phi) \sum_{V \in \mathcal{P}_m(\mathcal{Y}|\Phi)} \mathbb{P}\left[P_{\mathbf{y}|\mathbf{x}} = V|P_{\mathbf{x}} = \Phi\right] \Bigg(\frac{p(n)}{\alpha(Q, \Phi)}$$
$$\cdot \exp\{-n[R(Q, D) - \rho I(\Phi, V)]\}\Bigg)$$
$$\leq \sum_{\Phi \in \mathcal{A}_m} p(n) \exp\left\{-m \left[\min_{V \in \mathcal{P}_m(\mathcal{Y}|\Phi)} D\left(V \,\|\, W|\Phi\right)\right]\right\}$$
$$\cdot \exp\left\{-n \left|R(Q, D) - \rho I\left(\Phi, V\right)\right|^+\right\}$$
$$\leq \sum_{\Phi \in \mathcal{A}_m} p(n) \cdot \exp\left\{-n \left[\min_{V \in \mathcal{P}_m(\mathcal{Y}|\Phi)} \rho D\left(V \,\|\, W|\Phi\right)\right.\right.$$
$$\left.\left. + \left|R(Q, D) - \rho I(\Phi, V)\right|^+\right]\right\} \quad (17)$$

Combining (16) and (17), we have

$$\mathbb{P}\left[\bar{\mathcal{E}}(D)\right]$$
$$\leq p'(n) \exp\left\{-n \left[\min_{Q \in \mathcal{P}_n(\mathcal{S})} \left(D\left(Q \,\|\, P\right)\right.\right.\right.$$
$$+ \min_{\Phi \in \mathcal{A}_m} \min_{V \in \mathcal{P}_m(\mathcal{Y}|\Phi)} \rho D\left(V \,\|\, W|\Phi\right)$$
$$\left.\left.\left. + \left|R(Q, D) - \rho I(\Phi, V)\right|^+\right)\right]\right\},$$

where $p'(n)$ is some polynomial in n. Therefore,

$$\bar{E}\left(P, D, W, \rho\right)$$
$$= \liminf_{n \to \infty} -\frac{1}{n} \log \mathbb{P}\left[\bar{\mathcal{E}}_n(D)\right]$$
$$\geq \min_{Q \in \mathcal{P}(\mathcal{S})} \left(D\left(Q \,\|\, P\right) + \max_{\Phi \in \mathcal{P}(\mathcal{X})} \min_{V \in \mathcal{P}_n(\mathcal{Y}|\Phi)} \rho D\left(V \,\|\, W|\Phi\right)\right.$$
$$\left. + \left|R(Q, D) - \rho I(\Phi, V)\right|^+\right).$$

Finally, it can be shown that the exponents in (9) and (3) are equivalent, whose proof is omitted due to space limit. This completes the proof. ∎

REFERENCES

[1] G. Dueck and J. Körner. Reliability function of a discrete memoryless channel at rates above capacity. *IEEE Trans. Info. Theory*, 25(1):82–85, Jan. 1979.
[2] S. Arimoto. On the converse to the coding theorem for discrete memoryless channels (corresp.). *IEEE Trans. Info. Theory*, 19(3):357–359, May 1973.
[3] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp. Lower bounds to error probability for coding on discrete memoryless channels, part I-II. *Inform. Contr.*, 10:65–103,522–552, 1967.
[4] E. A. Haroutunian. Estimates of the error exponent for the semi-continuous memoryless channel. *Problemy Pered. Inform. (Problems of Inform. Trans.)*, 4, No. 4:37–48, 1968.
[5] I. Csiszár and J. Körner. *Information Theory - Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, New York, second edition, 2011.
[6] K. Marton. Error exponent for source coding with a fidelity criterion. *IEEE Trans. Info. Theory*, 20:197–199, Mar. 1974.
[7] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. In *Institute of Radio Engineers, International Convention Record*, volume 7, pages 142–163, 1959.
[8] Y. Zhong, F. Alajaji, and L.L. Campbell. Joint source-channel coding excess distortion exponent for some memoryless continuous-alphabet systems. *IEEE Trans. Info. Theory*, 55(3):1296–1319, Mar. 2009.
[9] Y. Zhong, F. Alajaji, and L.L. Campbell. Joint source-channel coding error exponent for discrete communication systems with Markovian memory. *IEEE Trans. Info. Theory*, 53(12):4457–4472, Dec. 2007.
[10] T. S. Han. *Information-Spectrum Method in Information Theory*. Springer, November 2002.
[11] M. Agarwal, A. Sahai, and S. Mitter. Coding into a source: a direct inverse rate-distortion theorem. In *Proceedings of the 44th Annual Allerton Conference on Communication, Control and Computing*, 2003.
[12] D. Wang, A. Ingber, and Y. Kochman. The dispersion of joint source-channel coding. In *49th Annual Allerton Conference on Communication, Control and Computing*, Monticelo, IL, Sep. 2011.
[13] I Csiszár. Joint source-channel error exponent. *Prob. of Cont. and Info. Th.*, 9(5):315–328, 1980.
[14] I. Csiszár. On the error exponent of source-channel transmission with a distortion threshold. *IEEE Trans. Info. Theory*, 28:823–838, Nov. 1982.
[15] Z. Zhang, E.H. Yang, and V. Wei. The redundancy of source coding with a fidelity criterion. 1. known statistics. *IEEE Trans. Info. Theory*, 43:71–91, Jan. 1997.