# A Strong Converse for Joint Source-Channel Coding

Da Wang
EECS Dept., MIT
Cambridge, MA 02139, USA
Email: dawang@mit.edu

Amir Ingber
Dept. of EE-Systems, TAU
Tel Aviv 69978, Israel
Email: ingber@eng.tau.ac.il

Yuval Kochman
EECS Dept., MIT
Cambridge, MA 02139, USA
Email: yuvalko@mit.edu

***Abstract*—We consider a discrete memoryless joint source-channel setting. Let $D$ be some distortion level, lower than the optimum performance theoretically attainable according to the separation theorem. We prove that for any joint source-channel scheme, the probability that the distortion is at most $D$ approaches zero as the block length increases. Furthermore, we show that the probability has an exponential behavior, and evaluate the optimal exponent.**

## 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 source (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]

$$\bar{E}_{\mathrm{sp}}(R,W) = \max_{\Phi} \min_{V} \left[ D(V \,\|\, W|\Phi) + |R - I(\Phi,V)|^{+} \right]. \tag{1}$$

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]):

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

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

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

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

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

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

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

$$E_{\mathrm{S}}(R,D,P) = \inf_{Q: R(Q,D) > R} D(Q \,\|\, P).$$

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 $\rho$ 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. However, to the best of our knowledge, a strong converse has not been presented, except for [8] , where Zhong et al. use very specific arguments to prove it for the Quadratic-Gaussian setting. Indeed, such a strong converse may be derived using previously known results. One way to do so, is to use equivalence to channel coding [9] and the strong channel converse. Alternatively, JSCC dispersion [10] implies a strong converse. Information spectrum methods, such as used by Han [11] to derive a strong converse to lossless JSCC.

In this work we take a direct path that allows us not only to prove a strong converse, but also derive the exponential behavior of the probability of success. Specifically, we show that whenever $R(D) > \rho C$, the probability of not having excess distortion for the optimal JSCC scheme decays with exponent

$$\begin{aligned} &\bar{E}_{\mathrm{JSCC}}(P,D,W,\rho) \\ &= \min_{R} [\bar{E}_{\mathrm{S}}(R,D,P) + \rho \bar{E}_{\mathrm{sp}}(R/\rho, W)], \end{aligned} \tag{3}$$

where the channel and source exponents are given by (1) and (2), respectively. This is analogous to the exponent of the error probability: the JSCC excess-distortion exponent for $R(D) < \rho C$ is upper-bounded by [12], [13]

$$\min_{R} \left[ E_{\mathrm{S}}(R,D,P) + \rho E_{\mathrm{sp}}(R,W) \right]. \tag{4}$$

## 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* $\mathcal{T}^n_{P_{\mathbf{x}}}$ 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 $N(a,b|\mathbf{x},\mathbf{y}) = P_{\mathbf{y}|\mathbf{x}}(a|b)N(a|\mathbf{x})$. 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

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

A discrete memoryless channel (DMC) $W : \mathcal{X} \to \mathcal{Y}$ is defined with its input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and conditional distribution $W(\cdot\,|\,x)$ of output letter $Y$ when the channel input letter $X$ equals $x \in \mathcal{X}$. Also, we abbreviate $W(\cdot\,|\,x)$ as $W_x(\cdot)$ for notational simplicity. We define mutual information as

$$I(\Phi, W) = \sum_{x,y} \Phi(x)W(y|x) \log \frac{\Phi(x)W(y|x)}{\Phi W(y)},$$

and the channel capacity is given by

$$C(W) = \max_{\Phi} I(\Phi, W),$$

and the set of capacity-achieving distributions is $\Pi(W) \triangleq \{\Phi : I(\Phi, W) = C(W)\}$.

A discrete memoryless source (DMS) 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

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

where

$$d(\mathbf{s}, \hat{\mathbf{s}}) \triangleq \frac{1}{n} \sum_{i=1}^{n} d(s_i, \hat{s}_i) \tag{6}$$

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 (JSCC) 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 $\mathcal{C}^{(n)}_{\text{JSCC}}$ 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 (6). For a given JSCC scheme, we define the *error* event $\mathcal{E}(D)$ as

$$\mathcal{E}(D) \triangleq \mathcal{E}(D, f_{J;n}, g_{J;n}) \tag{7}$$

$$\triangleq \mathcal{E}\left(D, \mathcal{C}^{(n)}_{\text{JSCC}}\right) \tag{8}$$

$$\triangleq \{d(\mathbf{S}, \hat{\mathbf{S}}) > D\}, \tag{9}$$

and the *correct* event $\bar{\mathcal{E}}(D) \triangleq \mathcal{E}(D)^c = \{d(\mathbf{S}, \hat{\mathbf{S}}) \le D\}$. 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.,

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

## 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 $\rho$, when $\rho C(W) < R(P, D)$, 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 $\bar{E}(P, D, W, \rho)$ is given by $\bar{E}_{\text{JSCC}}(P, D, W, \rho)$ (3).*

**Remark 1** (Direct part of the theorem)**.** *The achievability of the exponent $\bar{E}_{\text{JSCC}}(P, D, W, \rho)$ may be proven by using unequal error protection (UEP), as done by Csiszár for the JSCC exponent where $R(D) < \rho C$ [13]. Loosely speaking, each source type-class $Q$ is quantized by a codebook of rate $R(Q, D)$, and then a UEP scheme is used to transmit these codebooks over the channel. The achievability proof [1] of the channel exponent $\bar{E}_{\text{sp}}(R, W)$ in (1) can be extended to the UEP setting, completing the proof. As this is a rather trivial extension of previous results, we concentrate in this work on proving the converse.*

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

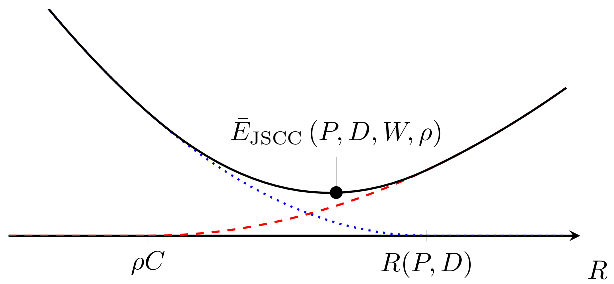


Fig. 1. When $R(P, D) > \rho C$, (3) is always minimized by a rate $R$ such that $R(P, D) > R > \rho C$, where in the plot the dashed curve is $\bar{E}_{\text{sp}}(R, W)$, the dotted curve is $\bar{E}_{\text{S}}(R, D, P)$ and the solid curve is $\bar{E}_{\text{S}}(R, D, P) + \rho\bar{E}_{\text{sp}}(R,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

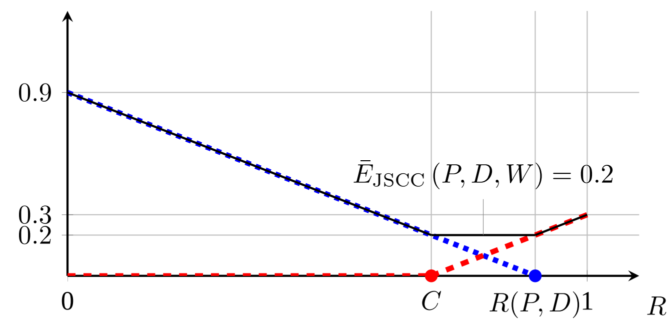


Fig. 2. $\bar{E}_{\text{JSCC}}(P, D, W, \rho)$ with $H_b(D) = 0.1$ and $H_b(\varepsilon) = 0.3$, where in the plot the dashed curve is $\bar{E}_{\text{sp}}(R, W)$, the dotted curve is $\bar{E}_{\text{S}}(R, D, P)$ and the solid curve is $\bar{E}_{\text{S}}(R, D, P) + \bar{E}_{\text{sp}}(R, W)$.

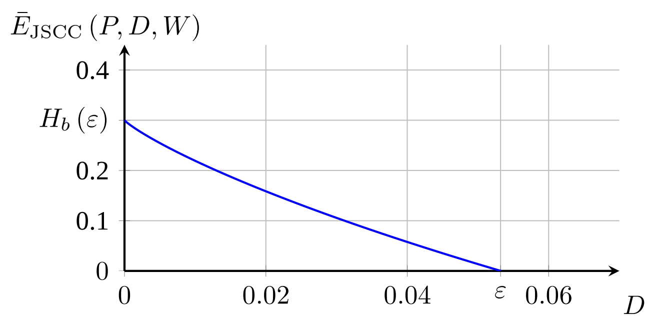


Fig. 3. $\bar{E}(P, D, W, \rho)$ with $H_b(\varepsilon) = 0.3$.

*the source and channel parameters, as follows.*

$$\begin{aligned}&\bar{E}_{\text{JSCC}}(P, D, W, \rho)\\ &= \min_{Q \in \mathcal{P}(\mathcal{S})} \Bigg[ D(Q \,\|\, P) + \max_{\Phi \in \mathcal{P}(\mathcal{X})} \min_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \Bigg( \rho D(V \,\|\, W|\Phi) \\ &\qquad + |R(Q, D) - \rho I(\Phi, V)|^+ \Bigg) \Bigg]. \end{aligned} \tag{10}$$

*We prove the equivalence of this form to (3) in Section IV-C. We use this form in the proof of Theorem 1, rather than the form (3).*

**Remark 3** (Minimizing rate)**.** *When $R(P, D) > \rho C$, the rate minimizing (3) satisfies $R(P, D) \geq R \geq \rho C$, 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.

**Lemma 2** (Joint source channel coding converse with fixed types)**.** *For a JSCC problem, given a source type $Q \in \mathcal{P}_n(\mathcal{S})$ and a channel input type $\Phi \in \mathcal{P}_n(\mathcal{X})$, define all the channel outputs that covers $\mathbf{s}$ with distortion $D$ as $\hat{B}(\mathbf{s}, D)$, i.e.,*

$$\hat{B}(\mathbf{s}, D) \triangleq \{\mathbf{y} \in \mathcal{Y}^m : d(\mathbf{s}, g_{J;n}(\mathbf{y})) \leq D\} \tag{11}$$

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

$$\begin{aligned}&\frac{1}{|G(Q, \Phi)|} \sum_{\mathbf{s}_i \in G(Q, \Phi)} \frac{\left|\mathcal{T}_V^m(f(\mathbf{s}_i)) \cap \hat{B}(\mathbf{s}_i, D)\right|}{|\mathcal{T}_V^m(f(\mathbf{s}_i))|} \\ &\leq \frac{p(n)}{\alpha(Q, \Phi)} \exp^{-n[R(Q,D) - \rho I(\Phi, V)]^+}, \end{aligned} \tag{12}$$

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

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 last lemma is similar to Lemma 3 in [14]. However, it provides a *uniform* bound over the size of the restricted $D$-ball. We prove this lemma, as well as Lemma 2 and the converse part of Theorem 1 in the next section. 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 $\varepsilon$, the capacity is given by $C(W) = 1 - H_b(\varepsilon)$. it can be shown that the optimizing $\Phi$ 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. $D < \varepsilon$,*

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

*For the case of $H_b(D) = 0.1$ and $H_b(\varepsilon) = 0.3$, we plot the $\bar{E}_{\text{S}}(R,D,P)$, $\bar{E}_{\text{sp}}(R,W)$ and $\bar{E}_{\text{JSCC}}(P,D,W)$ in Fig. 2. Finally, we show how $\bar{E}_{\text{JSCC}}(P,D,W)$ 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 $\hat{\mathbf{s}}$. Then the size of the set of source codewords with type $P$ that are $D$-covered by $\hat{\mathbf{s}}$ is

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

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

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

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

$$\begin{aligned}&|B(\hat{\mathbf{s}},P,D)| \\ \le & \sum_{\tilde{\Lambda}:\mathbb{E}_{Q,\tilde{\Lambda}}\left[d(\hat{S},S)\right]\le D,} \left|\mathcal{T}_{\tilde{\Lambda}}^n(\hat{\mathbf{s}})\right| \\ \le & (n+1)^{|\mathcal{S}||\hat{\mathcal{S}}|}\exp\left[n \max_{\tilde{\Lambda}:\mathbb{E}_{Q,\tilde{\Lambda}}\left[d(\hat{S},S)\right]\le D,} H\left(\tilde{\Lambda}|Q\right)\right].\end{aligned}$$

Note

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

hence

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

■

*Proof for Lemma 2:* In our proof, we first bound the denominator in (12) uniformly for all $\mathbf{s}_i$, and then bound the sum of the numerator over all $\mathbf{s}_i$, as done in [1] for the channel error exponent.

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

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

Hence

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

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

$$\begin{aligned}&\mathbf{y} \in \mathcal{T}_V(f(\mathbf{s})) \cap \hat{B}(\mathbf{s},D) \\ \Rightarrow &\mathbf{s} \in B(g_{J;n}(\mathbf{y}),Q,D) \cap G(Q,\Phi).\end{aligned} \tag{13}$$

Therefore, any $\mathbf{y}$ will be counted at most $|B(g_{J;n}(\mathbf{y}),Q,D) \cap G(Q,\Phi)|$ 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

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

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

$$\begin{aligned}&\frac{\alpha(Q,\Phi)}{|G(Q,\Phi)|}\sum_{\mathbf{s}_i \in G(Q,\Phi)} \left|\mathcal{T}_V(f(\mathbf{s}_i)) \cap \hat{B}(\mathbf{s}_i,D)\right| \\ \le & \frac{1}{\left|\mathcal{T}_Q^n\right|} B_u \left|\bigcup_{\mathbf{s}_i \in \mathcal{T}_Q^n} \mathcal{T}_V(f(\mathbf{s}_i)) \cap \hat{B}(\mathbf{s}_i,D)\right| \\ \le & \frac{1}{\left|\mathcal{T}_Q^n\right|} B_u |\mathcal{T}_\Psi^n|.\end{aligned}$$

Noting

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

we have

$$\begin{aligned}&\frac{1}{n}\log\left[\frac{1}{\left|\mathcal{T}_Q^n\right|} B_u |\mathcal{T}_\Psi^n|\right] \\ \le & \frac{|\mathcal{S}|}{n}\log(n+1) - H(Q) \\ &+ \frac{|\mathcal{S}|\left|\hat{\mathcal{S}}\right|}{n}\log(n+1) + H(Q) - R(Q,D) \\ &+ \rho H(\Psi) \\ \le & \rho H(\Psi) - R(Q,D) \\ &+ \frac{|\mathcal{S}|\left|\hat{\mathcal{S}}\right|}{n}\log(n+1) + \frac{|\mathcal{S}|}{n}\log(n+1).\end{aligned}$$

Combining the bounds for both numerator and denominator, we have

$$\begin{aligned}&\frac{1}{n}\log\left[\frac{\alpha(Q,\Phi)}{|G(Q,\Phi)|}\sum_{\mathbf{s}_i \in G(Q,\Phi)} \frac{\left|\mathcal{T}_V(f(\mathbf{s}_i)) \cap \hat{B}(\mathbf{s}_i,D)\right|}{\mathcal{T}_V(f(\mathbf{s}_i))}\right] \\ \le & \rho H(\Psi) - \rho H(V|\Phi) - R(Q,D) + \frac{|\mathcal{X}||\mathcal{Y}|}{m}\log(m+1) \\ &+ \frac{|\mathcal{S}|\left|\hat{\mathcal{S}}\right|}{n}\log(n+1) + \frac{|\mathcal{S}|}{n}\log(n+1).\end{aligned}$$

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

$$p(n) = (\rho n+1)^{\rho|\mathcal{X}||\mathcal{Y}|}(n+1)^{\left(|\mathcal{S}||\hat{\mathcal{S}}|+|\mathcal{S}|\right)}, \tag{14}$$

and the proof is completed. ■

### B. Proof of Main Result

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

$$\begin{aligned}&\mathbb{P}\left[\bar{\mathcal{E}}(D)\right]\\ \leq&(n+1)^{|\mathcal{S}|}\max_{Q\in\mathcal{P}_n(\mathcal{S})}\mathbb{P}\left[\bar{\mathcal{E}}(D)\,\middle|\,P_{\mathbf{S}}=Q\right]e^{-nD(Q\,\|\,P)}.\end{aligned}\quad(15)$$

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

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

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

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

and let $\alpha(Q,\Phi)\triangleq\mathbb{P}\left[P_{\mathbf{X}}=\Phi|P_{\mathbf{S}}=Q\right]$ be the probability of having channel input type $\Phi$ given that the source type is $Q$. 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,

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

Next we use Lemma 2 to assert,

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

where $p(n)$ is a polynomial given in (14). Therefore,

$$\begin{aligned}&\mathbb{P}\left[\bar{\mathcal{E}}(D)|P_{\mathbf{S}}=Q\right]\\ \leq&\sum_{\Phi\in\mathcal{A}_m}\alpha(Q,\Phi)\sum_{V\in\mathcal{P}_m(\mathcal{Y}|\Phi)}\mathbb{P}\left[P_{\mathbf{y}|\mathbf{x}}=V|P_{\mathbf{x}}=\Phi\right]\Bigg(\frac{p(n)}{\alpha(Q,\Phi)}\\ &\quad\cdot\exp\{-n[R(Q,D)-\rho I(\Phi,V)]\}\Bigg)\\ \leq&\sum_{\Phi\in\mathcal{A}_m}p(n)\exp\left\{-m\left[\min_{V\in\mathcal{P}_m(\mathcal{Y}|\Phi)}D\left(V\,\|\,W|\Phi\right)\right]\right\}\\ &\quad\cdot\exp\left\{-n\left|R(Q,D)-\rho I\left(\Phi,V\right)\right|^+\right\}\\ \leq&\sum_{\Phi\in\mathcal{A}_m}p(n)\cdot\exp\Bigg\{-n\Bigg[\min_{V\in\mathcal{P}_m(\mathcal{Y}|\Phi)}\rho D\left(V\,\|\,W|\Phi\right)\\ &\quad+\left|R(Q,D)-\rho I(\Phi,V)\right|^+\Bigg]\Bigg\}\end{aligned}\quad(16)$$

Combining (15) and (16), we have

$$\begin{aligned}&\mathbb{P}\left[\bar{\mathcal{E}}(D)\right]\\ \leq&\text{poly}(n)\exp\Bigg\{-n\Bigg[\min_{Q\in\mathcal{P}_n(\mathcal{S})}\Bigg(D\left(Q\,\|\,P\right)\\ &\quad+\min_{\Phi\in\mathcal{A}_m}\min_{V\in\mathcal{P}_m(\mathcal{Y}|\Phi)}\rho D\left(V\,\|\,W|\Phi\right)\\ &\quad+\left|R(Q,D)-\rho I(\Phi,V)\right|^+\Bigg)\Bigg]\Bigg\}.\end{aligned}$$

Therefore,

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

On account of the exponent equivalence, the proof is completed. ■

### C. Proof of Exponent Equivalence

In this section we prove the equivalence between (10) and (3).

To that end, let the $\Phi,V$ and $Q$ that achieves $\bar{E}_{\text{JSCC}}\left(P,D,W,\rho\right)$ be $\Phi^*,V^*$ and $Q^*$. Let $R_{\text{S}}$ and $R_{\text{C}}$ be the source and channel rate respectively, where $R_{\text{C}}=R_{\text{S}}/\rho$. If $R(Q^*,D)-\rho I\left(\Phi^*,V^*\right)\leq 0$, then

$$\begin{aligned}&\bar{E}_{\text{JSCC}}\left(P,D,W,\rho\right)\\ =&D\left(Q^*\,\|\,P\right)+\rho D\left(V^*\,\|\,W|\Phi^*\right)\\ =&\min_Q D\left(Q\,\|\,P\right)+\max_\Phi\min_V\rho D\left(V\,\|\,W|\Phi\right),\end{aligned}$$

and for any $R(Q^*,D)\leq R'_{\text{S}}\leq\rho I\left(\Phi^*,V^*\right)$ and $R'_{\text{C}}=R'_{\text{S}}/\rho$, we have

$$\bar{E}_{\text{S}}\left(R'_{\text{S}},D,P\right)=\min_Q\left[D\left(Q\,\|\,P\right)\right],$$
$$\bar{E}_{\text{sp}}\left(R'_{\text{C}},W\right)=\max_\Phi\min_V D\left(V\,\|\,W|\Phi\right),$$

and thus

$$\begin{aligned}&\inf_{R_{\mathrm{S}}} \bar{E}_{\mathrm{S}}\left(R_{\mathrm{S}}, D, P\right)+\rho \bar{E}_{\mathrm{sp}}\left(R_{\mathrm{C}}, W\right) \\ \leq & \bar{E}_{\mathrm{S}}\left(R_{\mathrm{S}}^{\prime}, D, P\right)+\rho \bar{E}_{\mathrm{sp}}\left(R_{\mathrm{C}}^{\prime}, W\right) \\ = & \bar{E}_{\mathrm{JSCC}}(P, D, W, \rho) .\end{aligned} \tag{17}$$

Similarly, we can show that when $R(Q^*, D) - \rho I\left(\Phi^*, V^*\right) > 0$,

$$\begin{aligned}&\inf_{R_{\mathrm{S}}} \bar{E}_{\mathrm{S}}\left(R_{\mathrm{S}}, D, P\right)+\rho \bar{E}_{\mathrm{sp}}\left(R_{\mathrm{C}}, W\right) \\ \leq & \bar{E}_{\mathrm{JSCC}}(P, D, W, \rho) .\end{aligned} \tag{18}$$

However, we also know that for any $R$

$$\begin{aligned}&\bar{E}_{\mathrm{S}}\left(R_{\mathrm{S}}, D, P\right)+\rho \bar{E}_{\mathrm{sp}}\left(R_{\mathrm{C}}, W\right) \\ \geq & \min_{Q \in \mathcal{P}(\mathcal{S})}\left(D\left(Q \,\|\, P\right)+\max_{\Phi \in \mathcal{P}(\mathcal{X})} \min_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \rho D\left(V \,\|\, W | \Phi\right)\right. \\ &\left. \quad + \left|R(Q, D)-\rho I(\Phi, V)\right|^{+}\right),\end{aligned}$$

hence

$$\inf_{R_{\mathrm{S}}} \bar{E}_{\mathrm{S}}\left(R_{\mathrm{S}}, D, P\right)+\rho \bar{E}_{\mathrm{sp}}\left(R_{\mathrm{C}}, W\right) \geq \bar{E}_{\mathrm{JSCC}}(P, D, W, \rho) . \tag{19}$$

Therefore, (17) and (18) and (19) suggest that

$$\bar{E}_{\mathrm{JSCC}}(P, D, W, \rho)=\inf_{R_{\mathrm{S}}} \bar{E}_{\mathrm{S}}\left(R_{\mathrm{S}}, D, P\right)+\rho \bar{E}_{\mathrm{sp}}\left(R_{\mathrm{C}}, W\right) .$$

## REFERENCES


[1] G. Dueck and J. Korner. Reliability function of a discrete memoryless channel at rates above capacity. *IEEE Trans. Info. Theory*, 25(1):82–85, January 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. Korner. *Information Theory - Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, New York, 2011.

[6] K. Marton. Error exponent for source coding with a fidelity criterion. *IEEE Trans. Info. Theory*, IT-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, Vol. 7*, pages 142–163, 1959.

[8] Y. Zhong, F. Alajaji, and L.L. Campbell. On the excess distortion exponent for memoryless Gaussian source-channel pairs. In *ISIT-2006, Seattle, WA*, 2006.

[9] 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.

[10] 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.

[11] Te Sun Han. *Information-Spectrum Method in Information Theory*. Springer, 1 edition, November 2002.

[12] I Csiszár. Joint source-channel error exponent. *Prob. of Cont. and Info. Th.*, 9(5):315–328, 1980.

[13] I. Csiszár. On the error exponent of source-channel transmission with a distortion threshold. *IEEE Trans. Info. Theory*, IT-28:823–838, Nov. 1982.

[14] Z. Zhang, E.H. Yang, and V. Wei. The redundancy of source coding with a fidelity criterion - Part one: Known statistics. *IEEE Trans. Info. Theory*, IT-43:71–91, Jan. 1997.