The Dispersion of Joint Source-Channel Coding

Da Wang, Amir Ingber, Yuval Kochman

Abstract

In this work we investigate the behavior of the distortion threshold that can be guaranteed in joint source-channel coding, to within a prescribed excess-distortion probability. We show that the gap between this threshold and the optimal average distortion is governed by a constant that we call the joint source-channel dispersion. This constant can be easily computed, since it is the sum of the source and channel dispersions, previously derived. The resulting performance is shown to be better than that of any separation-based scheme. For the proof, we use unequal error protection channel coding, thus we also evaluate the dispersion of that setting.

I. Introduction

One of the most basic results of Information Theory, joint source-channel coding, due to Shannon [1], states that in the limit of large block-length n, a discrete memoryless source with distribution P can be sent through a discrete memoryless channel with transition distribution W and reconstructed with some expected average distortion D, as long as

$$R(P, D) < \rho C(W), \tag{1}$$

where $R(P, D)$ is the rate-distortion function of the source, $C(W)$ is the channel capacity and the bandwidth expansion ratio ρ is the number of channel uses per source sample. We denote by $D^* = D^*(P, W, \rho)$ the distortion satisfying (1) with equality, known as the optimal performance theoretically attainable (OPTA). Beyond the *expected* distortion, one may be interested in ensuring that the distortion for one source block is below some threshold. To that end, we see an *excess distortion* event $\mathcal{E}(D)$ as

$$\mathcal{E}(D) \triangleq \{d(\mathbf{S}, \hat{\mathbf{S}}) > D\}, \tag{2}$$

where

$$d(\mathbf{s}, \hat{\mathbf{s}}) \triangleq \frac{1}{n}\sum_{i=1}^{n} d(s_i, \hat{s}_i) \tag{3}$$

is the distortion between the source and reproduction words $\mathbf{s}$ and $\hat{\mathbf{s}}$.

We are interested in the probability of this event as a function of the block length. We note that two different approaches can be taken. In the first, the distortion threshold is fixed to some $D \geq D^*$ and one considers how the excess-distortion probability ε approaches zero as the block length n grows. This leads to the joint source-channel excess-distortion exponent: [2], [3]

$$\varepsilon(n) \cong \exp\{-n \cdot E(P, W, \rho, D)\}. \tag{4}$$

One may ask an alternative question: for given excess distortion probability ε, let D_n be the optimal (minimal) distortion threshold that can be achieved at blocklength n. How does the sequence D_n approach D^*? In this work we show, that the sequence behaves as:

$$R(P, D_n) \cong \rho C(W) - \sqrt{\frac{V_J(P, W, \rho)}{n}} Q^{-1}(\varepsilon), \tag{5}$$

where $Q^{-1}(\cdot)$ is the inverse of the Gaussian cdf. We coin $V_J(P, W, \rho)$ the joint source-channel coding (JSCC) dispersion.

Similar problems have been stated and solved in the context of channel coding and lossless source coding in [4]. In [5] the channel dispersion result is tightened and extended, while in [6] (see also [7]) the parallel lossy source coding result is derived. In source coding, the rate redundancy above the rate-distortion function (or entropy in the lossless case) is measured, for a given excess-distortion probability ε:

$$R_n \cong R(P,D) + \sqrt{\frac{V_S(P,D)}{n}} Q^{-1}(\varepsilon), \tag{6}$$

where $V_S(P,D)$ is the source-coding dispersion. In channel coding, it is the rate gap below capacity, for a given error probability ε:

$$R_n \cong C(W) - \sqrt{\frac{V_C(W)}{n}} Q^{-1}(\varepsilon), \tag{7}$$

where $V_C(W)$ is the channel-coding dispersion. We show that the JSCC dispersion is related to the source and channel dispersions by the following simple formula (subject to certain regularity conditions):

$$V_J(P,W,\rho) = V_S(P,D^*) + \rho \cdot V_C(W). \tag{8}$$

The achievability proof of (8) is closely related to that of Csiszár for the exponent [3]. Namely, multiple source codebooks are mapped into an unequal error protection channel coding scheme. The converse proof combines the strong channel coding converse [8] with the D-covering of a type class (e.g., [9]).

The rest of the paper is organized as follows. Section II defines the notations. Section III revisits the channel coding problem, and extend the dispersion result (7) to the unequal error protection (UEP) setting. Section IV uses this framework to prove our main JSCC dispersion result. Then Section V shows the dispersion loss of separation-based schemes. Finally in Section VI we consider a formulation where the distortion ratios are fixed but the bandwidth expansion ratio ρ varies with n, and apply it to the lossless JSCC dispersion problem.

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 $\mathbf{X}$). caligraphic fonts (e.g. $\mathcal{X}$) represent a set and $\mathcal{P}(\mathcal{X})$ for 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 [10]. 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

$$\mathcal{P}_n(\mathcal{X}) \triangleq \{P \in \mathcal{P}(\mathcal{X}) : nP(x) \in \mathbb{Z}_+, \, \forall x \in \mathcal{X}\} \tag{9}$$

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 addiiton, 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 $I(P, \Lambda)$ 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 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 (3).

For our analysis, we also define the following information quantities [5]: given input distribution Φ and channel W, we define the information density of a channel as

$$i(x, y) \triangleq \log \frac{d\, W\,(y\,|\,x)}{d\, \Phi W(y)} = \frac{d\, I\,(\Phi, W)}{d\, W} = \frac{\partial\, I\,(\Phi, W)}{\partial\, W},$$

divergence variance as

$$V\,(\Phi \,\|\, \Psi) = \sum_{x \in \mathcal{X}} \Phi(x) \left[\log \frac{\Phi(x)}{\Psi(x)}\right]^2 - [D\,(\Phi \,\|\, \Psi)]^2,$$

unconditional information variance as

$$U\,(\Phi, W) \triangleq \mathrm{Var}\,[i(X, Y)] = V\,(\,\Phi \times W \,\|\, \Phi \times \Phi W\,)\,,$$

where $X \times Y$ has joint distribution $[\Phi \times W]$, conditional information variance as

$$\begin{aligned} V\,(\Phi, W) &\triangleq \mathbb{E}\,[\mathrm{Var}\,[i(X, Y)|X]] \\ &= V\,(\,\Phi \,\|\, \Phi W | \Phi) \\ &= \sum_{x \in \mathcal{X}} \Phi(x) \Big\{ \sum_{y \in \mathcal{Y}} W\,(\,y\,|\,x) \left[\log \frac{W\,(\,y\,|\,x)}{\Phi W(y)}\right]^2 \\ &\quad - [D\,(W_x \,\|\, \Phi W)]^2 \Big\}, \end{aligned}$$

and maximal/minimal conditional information variance as

$$V_{\max}(W) \triangleq \max_{\Phi \in \Pi(W)} V\,(\Phi, W)\,,$$
$$V_{\min}(W) \triangleq \min_{\Phi \in \Pi(W)} V\,(\Phi, W)\,.$$

For simplicity, we assume all channels in this paper satisify $V_{\min} > 0$, which holds for most channels (see [5, Appendix H] for detailed discussion).

In this paper, we use the notation $O\,(\cdot)$, $\Omega\,(\cdot)$ and $\Theta\,(\cdot)$, where $f(n) = O\,(g(n))$ if and only if $\limsup_{n \to \infty} \left|\frac{f(n)}{g(n)}\right| < \infty$, $f(n) = \Omega\,(g(n))$ if and only if $\liminf_{n \to \infty} \left|\frac{f(n)}{g(n)}\right| \ge 1$, and $f(n) = \Theta\,(g(n))$ if and only if $f(n) = O\,(g(n))$ and $f(n) = \Omega\,(g(n))$. In addition, $f(n) \le O(g(n))$ means that $f(n) \le cg(n)$ for some $c > 0$ and sufficiently large n. And we use the notation $\mathrm{poly}(n)$ to denote a sequence of numbers that is polynomial in n, i.e., $\mathrm{poly}(n) = \Theta\left(n^d\right)$ if the polynomial has degree d.

III. The Dispersion of UEP Channel Coding

In this section we introduce the dispersion of unequal error protection (UEP) coding. We use this framework in the next section to prove our main JSCC result, though we directly use one lemma proven here instead of the UEP dispersion theorem[1].

Given k classes of messages $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_k$, where $|\mathcal{M}_i| = N_i$, we can represent a message $m \in \mathcal{M} \triangleq \cup_i \mathcal{M}_i$ by its *class* i and *content* j, i.e., $m = (i, j)$, where $i \in \{1, 2, \ldots, k\}$ and $j \in \{1, 2, \ldots, N_i\}$. A scheme is comprised of an encoding function $f_{C;n} : \mathcal{M} \to \mathcal{X}^n$ and a decoder mapping $g_{C;n} : \mathcal{Y}^n \to \mathcal{M}$. The error probability for message m is $P_{\mathrm{e}}(m) \triangleq \mathbb{P}[\hat{m} \neq m]$, where $\hat{m}$ is the decoder output. We say that a scheme $(f_{C;n}, g_{C;n})$ is a *UEP scheme* with error probabilities $e_1, e_2, \ldots, e_k$ and rates $R_1, R_2, \ldots, R_k$ if

$$P_{\mathrm{e}}(m = (i, j)) \le e_i$$

for all messages, and

$$R_i = \frac{1}{n} \log N_i \quad \text{for all } i \in \{1, 2, \ldots, k\},$$

where n is the block length. We denote the codewords for message set M_i by $\mathcal{A}_i$, i.e.,

$$\mathcal{A}_i \triangleq \{f_{C;n}(m = (i, j)), j = 1, 2, \cdots, N_i\}.$$

As discussed in [5], dispersion gives a meaningful characterization on the rate loss at a certain block length and error probability. Here, we show that similar results hold for UEP channel codes.

Theorem 1 (UEP Dispersion, Achievability)**.** *Given a DMC $(\mathcal{X}, \mathcal{Y}, W)$, a sequence of integers $k_n = \mathrm{poly}(n)$, an infinite sequence of real numbers $\{\varepsilon_i \in (0, 1), i \in \mathbb{Z}^+\}$ and an infinite sequence of (not necessarily distinct) distributions $\left\{\Phi^{(i)} \in \mathcal{P}(\mathcal{X}), i \in \mathbb{Z}^+\right\}$, if $V\left(\Phi^{(i)}, W\right) > 0 \; \forall \, i$, then there exists a sequence of UEP schemes with k_n classes of messages and error probabilities $e_i \le \varepsilon_i$ such that for all $1 \le i \le k_n$,*

$$R_i = I\left(\Phi^{(i)}, W\right) - \sqrt{\frac{V_i}{n}} Q^{-1}(\varepsilon_i) + O\left(\frac{\log n}{n}\right), \tag{10}$$

where $V_i \triangleq V\left(\Phi^{(i)}, W\right)$ is the conditional information variance in (9).

The following corollary is immediate, substituting types $\{\Phi_i \in \Pi(W)\}$.

Corollary 2. *In the setting of Theorem 1, there exists a sequence of UEP codes with error probabilities $e_i \le \varepsilon_i$ such that*

$$R_i = C(W) - \sqrt{\frac{V_{C_i}}{n}} Q^{-1}(\varepsilon_i) + O\left(\frac{\log n}{n}\right),$$

where

$$V_{C_i} = \begin{cases} V_{\min}(W) & \varepsilon_i \le \frac{1}{2} \\ V_{\max}(W) & \varepsilon_i > \frac{1}{2} \end{cases}.$$

Remark 1. *In the theorem, the coefficient of the correction term $O(\log n/n)$ is unbounded for error probabilities that approach zero or one.*

Remark 2. *In the theorem, the message classes are cumulative, i.e., for each codeword length n, k_n message classes are used, which include the k_{n-1} classes used for $n-1$. Trivially, at least the same performance is achievable where only the message classes $k_{n-1}+1, \ldots, k_n$ are used. Thus, the theorem also applies to disjoint message sets, as long as their size is polynomial in n.*

Remark 3. *The rates of Corollary 2 are also necessary (up to the correction term). That is, any UEP code with error probabilities $e_1, e_2, \ldots, e_{k_n}$ such that $e_i \le \varepsilon_i$ must satisfy*

$$R_i \le C(W) - \sqrt{\frac{V_{C_i}}{n}} Q^{-1}(\varepsilon_i) + O\left(\frac{\log n}{n}\right).$$

[1]In this section we use n to denote the channel code block length, while in Sections IV to VI we use $m = \lfloor \rho n \rfloor$ as the channel code block length in the JSCC setting.

This is straightforward to see, as Theorem 48 of [5] shows that this is a bound in the single-codebook case.

Remark 4. *When taking a single codebook, i.e. $k_n = 1$ for all n, Corollary 2 reduces to the achievability part of the channel dispersion result [5, Theorem 49]. However, we have taken a slightly different path: we use constant-composition codebooks, resulting in the conditional information variance $V(\Phi, W)$, rather than i.i.d. codebooks which result in the generally higher (worse) unconditional information variance. As discussed in [5], these quantities are equal when a capacity-achieving distribution is used, but a scheme achieving $V(\Phi, W)$ may have an advantage under a cost constraint. Furthermore, we feel that our approach is more insightful, since it demonstrates that the stochastic effect that governs the dispersion is in the channel realization only, and not in the channel input (dual to the source dispersion being set by the source type only).*

The proof of Theorem 1 is based on the same construction used for the UEP exponent in [2]. A decoder that operates based on empirical mutual information (with varying threshold according to the codebook) is used, and if there is a unique codeword that has high enough empirical mutual information, it is declared; otherwise an error will be reported. This decoding rule may introduce two types of errors: the empirical mutual information for the actual codeword is not high enough, or the empirical mutual information for a wrong codeword is too high.

The following two lemmas address the effect of these error events. Lemma 3 shows that the empirical mutual information of the correct codeword is approximately normal distributed via the Central Limit Theorem, hence the probability of the first type of error (the empirical mutual information falls below the expected mutual information) is governed by the Q-function, from which we can obtain expression for the rate redundancy w.r.t. empirical mutual information. Lemma 4 shows that if we choose the codebook properly, the probability of the second type of error can be made negligible, relative to the probability of the first type of error.

Lemma 3 (Rate redundancy)**.** *For a DMC $(\mathcal{X}, \mathcal{Y}, W)$, given a an arbitrary distribution $\Phi \in \mathcal{P}(\mathcal{X})$ with $V(\Phi, V) > 0$, and a fixed probability ε, let $\Phi_n \in \mathcal{P}_n(\mathcal{X})$ be an n-type that approximates Φ as*

$$\|\Phi - \Phi_n\|_\infty \le \frac{1}{n}. \tag{11}$$

Let the rate redundancy *ΔR be the infimal value such that for $\mathbf{x} \in \mathcal{T}^n_{\Phi_n}$,*

$$\mathbb{P}\left[I\left(\Phi_n, P_{\mathbf{Y}|\mathbf{x}}\right) \le I(\Phi, W) - \Delta R, \mathbf{Y} \sim W^n(\cdot|\mathbf{x})\right] = \varepsilon, \tag{12}$$

then

$$\Delta R = \sqrt{\frac{V(\Phi, W)}{n}} Q^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{13}$$

Furthermore, the result holds if we replace (12) with

$$\mathbb{P}\left[I\left(\Phi_n, P_{\mathbf{Y}|\mathbf{x}}\right) \le I(\Phi, W) - \Delta R, \mathbf{Y} \sim W^n(\cdot|\mathbf{x})\right] = \varepsilon + \delta_n, \tag{14}$$

as long as $\delta_n = O\left(\frac{\log n}{\sqrt{n}}\right)$.

Proof sketch for Lemma 3: Applying Taylor expansion to the empirical mutual information $I(\Phi_n, P_{\mathbf{Y}|\mathbf{x}})$, where $\mathbf{Y}$ is the channel output corresponding to channel input $\mathbf{x}$, we have

$$\begin{aligned} I(\Phi_n, P_{\mathbf{Y}|\mathbf{x}}) \approx{}& I(\Phi_n, W) \\ &+ \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} (P_{\mathbf{Y}|\mathbf{x}}(y|x) - W(y|x)) I'_W(y|x), \end{aligned}$$

where the higher order terms only contribute to the correction term in the desired result, and

$$I'_W(y|x) \triangleq \left.\frac{\partial I(\Phi_n, V)}{\partial V(y|x)}\right|_{V=W}.$$

These first order terms can be represetned by sum of independent random variables with total variance $V(\Phi_n, W)/n$ and finite third moment, which faciliates the application of Berry-Esseen theorem (see, e.g., [11, Ch. XVI.5]) and gives

$$\mathbb{P}\left[I\left(\Phi_n, P_{\mathbf{Y}|\mathbf{x}}\right) \le I\left(\Phi_n, W\right) - \Delta R\right]$$
$$\approx Q\left(\left(\Delta_n + \Delta R\right)\sqrt{\frac{n}{V}}\right),$$

where $\Delta_n = O\left(\log n/n\right)$. Finally, we can show that given (11), $\left|V\left(\Phi, W\right) - V\left(\Phi_n, W\right)\right|$ and $\left|I\left(\Phi, W\right) - I\left(\Phi_n, W\right)\right|$ are small enough for (13) to hold. ■

Lemma 4. *For a DMC $(\mathcal{X}, \mathcal{Y}, W)$, there exists a sequence of UEP codes with $k_n = \mathrm{poly}(n)$ classes of messages, $\mathcal{A}_i \in \mathcal{T}^n_{\Phi_n^{(i)}}$, and rates $R_1, R_2, \ldots, R_{k_n}$, where $R_i \le H\left(\Phi_n^{(i)}\right) - \eta_n$,*

$$\eta_n \triangleq \frac{2}{n}\left(|\mathcal{X}|^2 + \log(n+1) + \log k_n + 1\right), \tag{15}$$

such that for any given $\mathbf{x} \in \mathcal{A}_i, i \in \{1, 2, \ldots, k_n\}$, any $\mathbf{x}' \neq \mathbf{x}$ and $\mathbf{x}' \in \mathcal{A}_{i'}$, $i' \in \{1, 2, \ldots, k_n\}$, and any $\gamma \in \mathbb{R}$,

$$\mathbb{P}\left[I\left(\Phi_n^{(i')}, P_{\mathbf{Y}|\mathbf{x}'}\right) - R_{i'} \ge \gamma, \mathbf{Y} \sim W^n\left(\cdot|\mathbf{x}\right)\right] \le$$
$$(n+1)^{|\mathcal{X}|^2|\mathcal{Y}|} \exp\left\{-n\left[\left|R_{i'} + \gamma - \eta_n\right|^+ - R_{i'}\right]\right\}.$$

Proof sketch for Lemma 4: This proof is based on the coding scheme in Lemma 6 of [2]. In that construction, given channel conditional type V, the fraction of the output sequences correspond to $\mathcal{A}_{i'}$ that overlaps with the output sequences of another codeword $\mathbf{x}$ in a message set $\mathcal{A}_i$ decays exponentially with the empirical mutual information $I\left(\Phi_n^{(i)}, V\right)$. Then by using a decoder based on empirical mutual information and by bounding the size of the output sequences that cause errors for the empirical mutual information decoder, we can show the desired result. ■

The detailed proofs of Lemmas 3 and 4 are given in Appendix A-B. Below we present the proof for Theorem 1.

Proof of Theorem 1: Fix some codeword length n. Without loss of generality, assume that the message is $m = (i, j)$ in class i, which is mapped to a channel input $\mathbf{x}(i, j) \in \mathcal{A}_i$. Each codebook $\mathcal{A}_i$ is drawn uniformly over the type class of $\Phi_n^{(i)} \in \mathcal{P}_n\left(\mathcal{X}\right)$, where $\Phi_n^{(i)}$ relates to $\Phi^{(i)}$ (which is a general probability distribution that is not necessarily in $\mathcal{P}_n\left(\mathcal{X}\right)$) by

$$|\Phi_n^{(i)}(x) - \Phi^{(i)}(x)|_\infty \le \frac{1}{n}.$$

For any $\mathbf{y} \in \mathcal{Y}^n$, define the measure for message m:

$$a_m(\mathbf{y}) \triangleq I(\Phi_n^{(i)}, P_{\mathbf{y}|\mathbf{x}(i,j)}) - R_i,$$

and let the decoder mapping $g_{C;n} : \mathcal{Y}^n \to \mathcal{M}$ be defined as follows, using thresholds γ_n to be specified:

$$g_{C;n}(\mathbf{y}) = \begin{cases} m & a_m(\mathbf{y}) \ge \gamma_n > \max_{m' \neq m} a_{m'}(\mathbf{y}) \\ \emptyset & \text{o.w. (declares a decoding failure)} \end{cases}$$

The error event is the union of the following two events:

$$\mathcal{E}_1 = \{a_m(\mathbf{y}) < \gamma_n\} \tag{16}$$
$$\mathcal{E}_2 = \left\{\exists m' \neq m \in \mathcal{M} \text{ s.t. } a_{m'}(\mathbf{y}) \ge \gamma_n\right\}. \tag{17}$$

Let $m' = (i', j')$ be a generic codeword different from m. For simplicity, we denote $\mathbf{x}(i, j)$ and $\mathbf{x}(i', j')$ by $\mathbf{x}$ and $\mathbf{x}'$ respectively in the rest of the proof. Note that i' may be equal to i.

We now choose

$$\gamma_n = 2\eta_n + \frac{1}{2n}\log k_n + \frac{a}{n}\log n, \tag{18}$$

where η_n is defined by (15) in Lemma 4 and $a = {}^{(d+1)}/2$, where d is the degree of the polynomial k_n. Note that $\gamma_n = O\left({}^{\log n}/n\right)$. Lemma 4 shows

$$\begin{aligned}\mathbb{P}\left[\mathcal{E}_2\right] &= \sum_j \mathbb{P}\left[I\left(\Phi_n^{(i')}, P_{\mathbf{Y}|\mathbf{x}'}\right) - R_{i'} \geq \gamma_n, \mathbf{x}' \in \mathcal{A}_{i'}, \mathbf{Y} \sim W\left(\cdot \mid \mathbf{x}\right)\right] \\ &\leq k_n (n+1)^{|\mathcal{X}|^2|\mathcal{Y}|} \\ &\quad \exp\left\{-n \min_{i'}\left[|R_{i'} + \gamma_n - \eta_n|^+ - R_{i'}\right]\right\} \\ &\leq k_n (n+1)^{|\mathcal{X}|^2|\mathcal{Y}|} \\ &\quad \exp\left\{-n\left[\eta_n + \frac{1}{2n}\log k_n + \frac{\log n^a}{n}\right]\right\} \\ &= \frac{\sqrt{k_n}}{n^a} = O\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

To analyze $\mathcal{E}_1$, let

$$\Delta R_i = I\left(\Phi^{(i)}, W\right) - R_i - \gamma_n. \tag{19}$$

Note that $\mathbb{P}\left[\mathcal{E}_1\right]$ may be written as

$$\mathbb{P}\left[I\left(\Phi_n^{(i)}, P_{\mathbf{Y}|\mathbf{x}}\right) - I\left(\Phi^{(i)}, W\right) \leq -\Delta R_i, \mathbf{Y} \sim W^n\left(\cdot|\mathbf{x}\right)\right].$$

Now employing (14) in Lemma 3 with $\varepsilon = \varepsilon_i$ and

$$\delta_n = -\mathbb{P}\left[\mathcal{E}_2\right] = O\left(\frac{1}{\sqrt{n}}\right),$$

we have

$$\Delta R_i = \sqrt{\frac{V(\Phi^{(i)}, W)}{n}} Q^{-1}\left(\varepsilon\right) + O\left(\frac{\log n}{n}\right)$$

is achievable. By the union bound, the error probabilities are no more than ε_i, as required. Finally, (19) leads to

$$R_i = I\left(\Phi^{(i)}, W\right) - \sqrt{\frac{V_i}{n}} Q^{-1}(\varepsilon_i) + O\left(\frac{\log n}{n}\right).$$

■

IV. Main Result: JSCC Dispersion

We now utilize the UEP framework in Section III to arrive at our main result.

For the sake of investigating the finite block-length behavior, we consider the *excess distortion* event $\mathcal{E}(D)$ defined in (2). When the distortion level is held fixed, Csiszár gives lower and upper bounds on the exponential decay of the excess distortion probability [3]. In this work, we fix the excess distortion probability to be constant with the blocklength n

$$\mathbb{P}\left[\mathcal{E}(D)\right] = \varepsilon \tag{20}$$

and examine how the distortion thresholds D_n approach the OPTA D^* (the distortion achieving equality in (1)), or equivalently, how $R(P, D_n)$ approaches $R(P, D^*) = \rho C(W)$. We find that it is governed by the joint source-channel dispersion (8). In this formula, the source dispersion is given by [6]:

$$V_S(P, D) = \mathrm{Var}\left[\frac{\partial}{\partial Q_i} R(Q, D)\bigg|_{Q=P}\right], \tag{21}$$

and the channel dispersion $V_C(W)$ is given by $V_{\min}(W)$, which is assumed to be equal to $V_{\max}(W)$.

Theorem 5. *Consider a JSCC problem with a DMS $(\mathcal{S}, \hat{\mathcal{S}}, P, d)$, a DMC $(\mathcal{X}, \mathcal{Y}, W)$ and bandwidth expansion factor ρ. Let the corresponding OPTA be D^*. Assume that $R(Q, D)$ is differentiable w.r.t. D and twice differentiable w.r.t.*

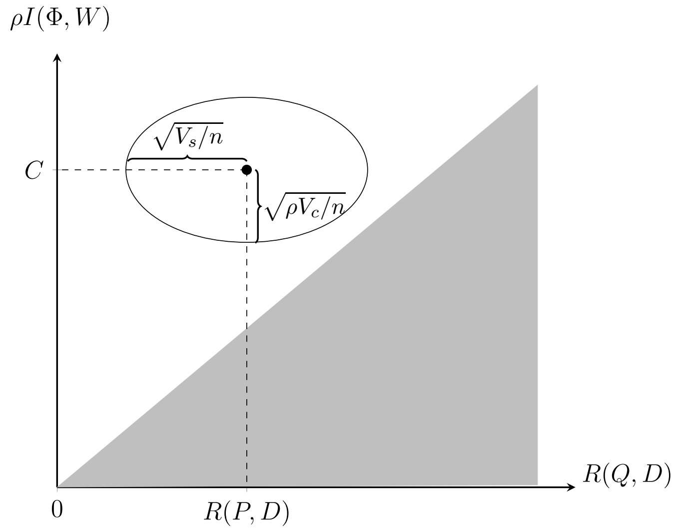

Fig. 1. Heuristic view of the main JSCC excess distortion event. The ellipse denotes the approximate one-standard-deviation region of the source-channel pair, while the gray area denotes the set of source-channel realizations leading to excess distortion.

Q in some neighborhood of (P, D^*). Also assume that the channel dispersion $V_{\min}(W) = V_{\max}(W) > 0$. Then for a fixed excess distortion probability $0 < \varepsilon < 1$, the optimal distortion thresholds D_n satisfy:

$$R(P, D_n) = \rho \cdot C(W) - \sqrt{\frac{V_J(P, W, \rho)}{n}} Q^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right),$$

where $V_J(P, W, \rho)$ is the JSCC dispersion (8).

We can give a heuristic explanation to this result, graphically depicted in Fig. 1. We know that the rate needed for describing the source is approximately Guassian, with mean $R(P, D_n)$ and variance $V_S(P, D_n)/n$. Similarly, the mutual information supplied by the channel is approximately Gaussian, with mean $\rho C(W)$ and variance $\rho V_C(W)/n$. We can now construct a codebook per source type, and map this set of codebooks to a channel UEP code. According to Section III, the dispersion of UEP given the rate of the chosen codebook is the same as only having that codebook. Consequently, an error occurs if the source and channel empirical behavior $(P_{\mathbf{s}}, P_{\mathbf{y}|\mathbf{x}})$ is such that

$$R(P_{\mathbf{s}}, D_n) > \rho \cdot I(\Phi, P_{\mathbf{y}|\mathbf{x}}).$$

The difference between the left and right hand sides is the difference of two independent approximately-Gaussian random variables, thus is approximately Gaussian with mean $R(P, D_n) - \rho C$ and variance $V_J(P, W, D)$, yielding (22) up to the correction term. However, for the proof we need to carefully consider the deviations from Gaussianity of both source and channel behaviors.

Remark 5. *In the (rather pathological) case where $V_{\min}(W) \neq V_{\max}(W)$, we cannot draw anymore the ellipse of Fig. 1. This is since the variance of the channel mutual information will be different between codebooks that have error probability smaller or larger than $1/2$. We can use $V_{\min}$ and $V_{\max}$ for upper and lower bounds on the JSCC dispersion. Also, when ε is close to zero or one, the dispersion of the channel part is very well approximated by $V_{\min}$ or $V_{\max}$, respectively.*

Remark 6. *The source and channel dispersions are known to be the second derivatives (with respect to the rate) of the source exponent at rate $R(P, D)$ and of the channel exponent at rate $C(W)$, respectively. Interestingly, the JSCC dispersion (8) is also connected to the second derivative of the JSCC exponent [3]:*

$$E(P, W, D, \rho) = \min_{R(P,D) \leq R \leq C} [E_S(P, D) + \rho E_C(W)]$$

(where E_S and E_C are the lossy source coding and sphere-packing exponents[2]*, respectively) via*

$$V_J(P,W,\rho) = \left[\frac{\partial^2 E(P,W,D,\rho)}{\partial R(P,D)^2} \bigg|_{D=D^*(P,W,\rho)} \right]^{-1},$$

where in the derivative P is held fixed.

The achievability part of Theorem 5 relies on the following lemma.

Lemma 6 (JSCC Distortion Redundancy)**.** *Consider a JSCC problem with a DMS $(\mathcal{S},\hat{\mathcal{S}},P,d)$, a DMC $(\mathcal{X},\mathcal{Y},W)$ and bandwidth expansion factor ρ. Let n be the length of the source block length, and let $m \triangleq \lfloor \rho n \rfloor$ be the length of the channel block length. Let Φ be an arbitrary distribution on $\mathcal{X}$, and let $\Phi_m \in \mathcal{P}_m(\mathcal{X})$ be an m-type that approximates Φ as*

$$\|\Phi - \Phi_m\|_\infty \leq \frac{1}{m}.$$

Let the channel input $\mathbf{x} \in \mathcal{X}^m$ have type Φ_m. Further, let $D^*(\Phi)$ be the solution to $R(P,D(\Phi)) = \rho I(\Phi,W)$. Assume that $R(Q,D)$, the RDF of a source Q with the same distortion measure, is twice differentiable w.r.t. D and the elements of Q at some neighborhood of $(P,D^*(\Phi))$. Let ε be a given probability and let $D_n > 0$ be the infimal value s.t.

$$\mathbb{P}\left[R(P_{\mathbf{S}},D_n) > \rho I(\Phi_m, P_{\mathbf{Y}|\mathbf{x}})\right] = \varepsilon. \tag{22}$$

Then, as n grows,

$$R(P,D_n) = \rho I(\Phi,W) - \sqrt{\frac{V_S(P)+\rho V_C(\Phi,W)}{n}}Q^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{23}$$

In addition, for any channel input (i.e., Φ_m is not restricted and may also depend upon the source sequence),

$$R(P,D_n) \leq \rho C(W) - \sqrt{\frac{V_J(P,W,\rho)}{n}}Q^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right), \tag{24}$$

where $V_J(P,W,\rho)$ is given by (8). Furthermore, all the above holds even if replace (22) with

$$\mathbb{P}\left[R(P_{\mathbf{S}},D_n) > \rho I(\Phi_m, P_{\mathbf{Y}|\mathbf{x}}) + \xi_n\right] = \varepsilon + \zeta_n, \tag{25}$$

for any given (vanishing) sequences ξ_n, ζ_n, as long as $\xi_n = O\left(\frac{\log n}{n}\right)$ and $\zeta_n = O\left(\frac{\log n}{\sqrt{n}}\right)$.

Proof sketch for Lemma 6: Similar to Lemma 3, we apply Taylor expansion to $R(P_{\mathbf{S}},D_n)$ and show that the first order term again can be expressed as sum of n independent random variables, and neglecting higher order terms does not affect the statement. Then $R(P_{\mathbf{S}},D_n) - \rho I(\Phi_m, P_{\mathbf{Y}|\mathbf{x}})$ can be shown to be the sum of $n+m$ indenpdent random variables, with total variance essentially $(V_S + \rho V_C(\Phi,W))/n$. Finally, similar to the derivation in Lemma 3, we apply the Berry-Esseen theorem and show (23) and (25) are true. ■

The detailed proof of Lemma 6 is given in Appendix B-B.

The converse part of Theorem 5 builds upon the following result, which states that for any JSCC scheme, the excess-distortion probability must be very high if the empirical mutual information over the channel is higher than the empirical source RDF.

Lemma 7 (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 $\Phi \in \mathcal{P}_n(\mathcal{X})$, let $G(Q,\Phi)$ be the set of source seqeuences in $\mathcal{T}_Q^n$ that are mapped (via JSCC encoder $f_{J;n}$) to channel codewords with type Φ, i.e.,*

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

Define all the channel outputs that covers $\mathbf{s}$ with distortion D as $\hat{B}(\mathbf{s},D)$, i.e.,

$$\hat{B}(\mathbf{s},D) = \{\mathbf{y} \in \mathcal{Y}^m : d(\mathbf{s}, g_{J;n}(\mathbf{y})) \leq D\} \tag{26}$$

[2]Sphere-packing exponent is only achievable when R is close to C, but this is sufficient for the derivative at $R = C$.

where $m = \lfloor \rho n \rfloor$ and $g_{J;n}$ is the JSCC decoder. If

$$|G(Q,\Phi)| \geq \frac{1}{(n+1)^{|\mathcal{X}|+1}} \left|\mathcal{T}_Q^n\right|, \tag{27}$$

then for a given distortion D and a channel with constant composition conditional distribution $V \in \mathcal{P}_m(\mathcal{Y}|\Phi)$, we have

$$\frac{1}{|G(Q,\Phi)|} \sum_{\mathbf{s}_i \in G(Q,\Phi)} \frac{\left|\mathcal{T}_V^m\left(f(\mathbf{s}_i)\right) \cap \hat{B}(\mathbf{s}_i, D)\right|}{\left|\mathcal{T}_V^m\left(f(\mathbf{s}_i)\right)\right|} \leq p(n) \exp^{-n[R(Q,D)-\rho I(\Phi,V)]^+} \tag{28}$$

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

The detailed proof of Lemma 7 is given in Appendix B-B. The proof uses an approach similar to that in the strong channel coding converse [8].

Below we present the proof for Theorem 5. The achievability proof is based on Lemmas 4 and 6, where we do not use directly Lemma 3 or Theorem 1, thus we do not suffer from the non-uniformity problem (see Remark 1). In other words, rather than evaluating the error probability per UEP codebook, we directly evaluate the average over all codebooks.

Proof: **Achievability:** Let

$$k_n = (n+1)^{|\mathcal{S}|+1} = \text{poly}(n). \tag{29}$$

At each block length n, we construct a source code $\mathcal{C} = \{\mathcal{C}_i\}$ as follows (the index n is omitted for notational simplicity). Each code $\mathcal{C}_i$ corresponds to one type $Q_i \in (\mathcal{P}_n(\mathcal{X}) \bigcap \Omega_n)$, where

$$\Omega_n = \left\{Q : \|P-Q\|_2^2 \leq |\mathcal{S}| \frac{\log n}{n}\right\}.$$

According to the refined type-covering Lemma [12], there exists codes $\mathcal{C}_i$ of rates

$$R_i \leq R(Q_i, D_n) + O\left(\frac{\log n}{n}\right), \tag{30}$$

that completely D_n-cover the corresponding types (where the redundancy term is uniform). We choose these to be the rates of the source code. The chosen codebook and codeword indices are then communicated using a dispersion-optimal UEP scheme as described in Section III with a capacity-achieving channel input distribution $\Phi \in \Pi(W)$. Specifically, each source codebook is mapped into a channel codebook of block length $\lfloor \rho n \rfloor$ and rate

$$\tilde{R}_i = \frac{R_i}{\rho}, \tag{31}$$

as long as

$$\tilde{R}_i \leq H(\Phi) - \eta_n, \tag{32}$$

where η_n is defined in Lemma 3. Otherwise, the mapping is arbitrary and we assume that an error will occur. The UEP scheme is thus used with different message classes at each n; such a scheme can only perform better than a scheme where the message classes accumulate, see Remark 2, thus we can use the results of Section III with number of codebooks:

$$\sum_{n'=1}^{n} |\mathcal{P}_n(\mathcal{X}) \bigcup \Omega_{n'}| \leq n \cdot |\mathcal{P}_n(\mathcal{X})| \leq (n+1)^{|\mathcal{S}|+1} = k_n,$$

where $\mathcal{P}_n(\mathcal{X})$ is defined in (9).

Error analysis: an excess-distortion event can occur only if one of the following events happened:

1) $P_{\mathbf{s}} \notin \Omega_n$, where $P_{\mathbf{s}}$ is the type of $\mathbf{s}$.
2) $\tilde{R}_i \geq H(\Phi) - \eta_n$.
3) $\mathcal{E}_2$ (17): an unrelated channel codeword had high empirical mutual information.
4) $\mathcal{E}_1$ (16): the true channel codeword had low empirical mutual information.

We show that the first three events only contribute to the correction term. According to [6, Lemma 2],

$$\mathbb{P}\left[P_{\mathbf{s}} \notin \Omega_n\right] \leq \frac{2|\mathcal{S}|}{n^2}.$$

By our assumption on the differentiability of $R(P, D)$, for large enough n the second event does not happen for any type in Ω_n. By Lemma 4, the probability of the third event is at most $O(1/\sqrt{n})$, uniformly. Thus, by the union bound, we need the probability of the last event to be at most $\varepsilon_n = \varepsilon - O(1/\sqrt{n})$.

Now following the analysis of $\mathcal{E}_1$ in the proof of Theorem 1, (19) indicates that event $\mathcal{E}_1$ is equivalent to

$$I\left(\Phi, P_{\mathbf{Y}|\mathbf{x}}\right) \le \tilde{R}_i + \gamma_n$$

where γ_n is defined in (18). (30) and (31) indicates this is equivalent to

$$\rho I\left(\Phi, P_{\mathbf{Y}|\mathbf{x}}\right) \le R(P_{\mathbf{s}}, D_n) + O\left(\frac{\log n}{n}\right).$$

On account of Lemma 6, this can indeed be satisfied with ε_n as required.

Converse: At the first stage of the proof we suppress the dependence on the block length n for conciseness. We first lower-bound the excess-distortion probability given that the source type is some $Q \in \mathcal{P}_n(S)$.

Let $\alpha(Q, \Phi) \triangleq \mathbb{P}\left[P_{\mathbf{X}} = \Phi | P_{\mathbf{S}} = Q\right]$ be the probability of having input type Φ giving that the source type is Q. Noting that given a source type, all strings within a type class are equally likely, we have

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

Now we have

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

Define the class of "frequent types" based on $\alpha(Q, \Phi)$:

$$A(Q) \triangleq \left\{\Phi \in \mathcal{P}_n(\mathcal{X}) : \alpha(Q, \Phi) \ge \frac{1}{(n+1)^{|\mathcal{X}|+1}}\right\}.$$

Note that

$$\mathbb{P}\left[P_{\mathbf{X}} \notin A(Q) | P_{\mathbf{S}} = Q\right] \le |\mathcal{P}_n(\mathcal{X})| \frac{1}{(n+1)^{|\mathcal{X}|+1}} \le \frac{1}{n+1},$$

thus $A(Q)$ is nonempty. Trivially, we have:

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

Next we use Lemma 7 to assert, for all $\Phi \in A(Q)$:

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

where $p(n)$ is given in (B.77). Since $\sum_{\Phi \in A(Q)} \alpha(Q, \Phi) \le 1$, we further have:

$$\begin{aligned}
\mathbb{P}\left[\mathcal{E}(D) | P_{\mathbf{S}} = Q\right] \ge 1 &- \frac{1}{n+1} \\
&+ p(n) \sum_{V \in \mathcal{P}_n(\mathcal{Y}|\Phi^*(Q))} \mathbb{P}\left[P_{\mathbf{y}|\mathbf{x}} = V | P_{\mathbf{x}} = \Phi^*(Q)\right] \exp\{-n[R(Q, D) - \rho I(\Phi^*(Q), V)]\},
\end{aligned}$$

where $\Phi^*(Q)$ minimizes the expression over all $\Phi \in A(Q)$ (if there are multiple maximizers, it is chosen arbitrarily). Collecting all source types, we have:

$$\begin{aligned}
\mathbb{P}\left[\mathcal{E}(D)\right] \ge \frac{n}{n+1} - p(n) \sum_{Q \in \mathcal{P}_n(\mathcal{S})} \sum_{V \in \mathcal{P}_n(\mathcal{Y}|\Phi^*(Q))} &\mathbb{P}\left[P_{\mathbf{S}} = Q\right] \mathbb{P}\left[P_{\mathbf{Y}|\mathbf{X}} = V | P_{\mathbf{X}} = \Phi^*(Q)\right] \cdot \\
&\exp\{-n[R(Q, D) - \rho I(\Phi^*(Q), V)]\}.
\end{aligned}$$

At this point we return the block length index n. Let Δ_n be some vanishing sequence to be specified later. Define the set

$$B(\Delta_n) \triangleq \{Q \in \mathcal{P}_n(\mathcal{S}), V \in \mathcal{P}_n(\mathcal{Y}|\Phi_n^*(Q)) : R(Q,D) - I(\Phi_n^*(Q), V) > \Delta_n\}.$$

For any sequence Δ_n we can write:

$$\mathbb{P}[\mathcal{E}(D)] \geq \mathbb{P}[B(\Delta_n)]\left[\frac{n}{n+1} - p(n)\exp\{-n\Delta_n\}\right],$$

where $\mathbb{P}[B(\Delta_n)] = \mathbb{P}\left[\mathbf{S} : P_{f_{J;n}(\mathbf{S})} \in B(\Delta_n)\right]$. Now choose $n\Delta_n = (1+p(n))\log(n+1)$ to obtain:

$$\mathbb{P}[\mathcal{E}(D)] \geq \left(1 - \frac{2}{n+1}\right)\mathbb{P}[B(\Delta_n)] \geq \frac{\mathbb{P}[B(\Delta_n)]}{1+\frac{2}{n-1}}.$$

Since we demand that $\mathbb{P}[\mathcal{E}(D)] \leq \varepsilon$ for all n, and inserting the definition of $B(\Delta_n)$ it must be that

$$\mathbb{P}[R(P_{\mathbf{S}}, D) - \rho I(\Phi_n^*(P_{\mathbf{S}}), V) > \Delta_n] \leq \varepsilon\left(1 + \frac{2}{n-1}\right).$$

Seeing that $\Delta_n = O(\log n/n)$, the desired result follows on account of (24) in Lemma 6. ■

V. THE LOSS OF SEPARATION

In this section we quantify the dispersion loss of a separation-based scheme with respect to the JSCC one. Using the separation approach, the interface between the source and channel parts is a fixed-rate message, as opposed to the variable-rate interface used in conjunction with multiple quantizers and UEP, shown in this work to achieve the JSCC dispersion.

Formally, we define a separation-based encoder as the concatenation of the following elements.

1) A source encoder $f_{S;n} : \mathcal{S}^n \to \mathcal{M}_n$.
2) A source-channel mapping $\mathcal{M}_n \to \mathcal{M}_n$.
3) A channel encoder $f_{C;n} : \mathcal{M}_n \to \mathcal{S}^{\lfloor \rho n \rfloor}$.

The interface rate is $R_n = \log|\mathcal{M}_n|/n$. Finally, the source-channel mapping is randomized, in order to avoid "lucky" source-channel matching that leads to an effective "joint" scheme.[3] We assume that it is uniform over all permutations of $\mathcal{M}_n$, and that it is known at the decoder as well. Consequently, the decoder is the obvious concatenation of elements in reversed order. The excess distortion probability of the scheme is defined as the mean over all permutations.

In a separation-based scheme, an excess-distortion event occurs if one of the following: either the source coding results in excess distortion, or the channel coding results in a decoding error. Though it is possible that no excess distortion will occur when a channel error occurs (whether the source code has excess distortion or not), the probability of this event is exponentially small. Thus at every block-length n, the excess-distortion probability ε satisfies

$$\varepsilon = \varepsilon_{S;n} * \varepsilon_{C;n} - \delta_n \tag{33}$$

where $a * b = a + b - ab$, $\varepsilon_{S;n}$ and $\varepsilon_{C;n}$ are the source excess-distortion probability and channel error probability, repectively, at blocklength n, and δ_n is exponentially decaying with n. In this expression we take a fixed ε, in accordance with the dispersion setting; the system designer is still free to choose $\varepsilon_{S;n}$ and $\varepsilon_{C;n}$ by adjusting the rates R_n, as long as (33) is maintained.

We now employ the source and channel dispersion results (6), (7), which hold up to a correction term $O(log(n)/n)$,[4] to see that that for the optimal separation-based scheme:

$$R(D_n) = \rho C(W) - \min_{\varepsilon_{S;n} * \varepsilon_{C;n} \leq \varepsilon}\left[\sqrt{\frac{V_s(P,D^*)}{n}}Q^{-1}(\varepsilon_{S;n}) + \sqrt{\frac{\rho V_c(W)}{n}}Q^{-1}(\varepsilon_{C;n})\right] + O\left(\frac{\log n}{n}\right). \tag{34}$$

[3]For instance, the UEP scheme could be presented as a separation one if not for the randomized mapping.

[4]The redundancy terms are in general functions of the error probabilities, but for probabilities bounded away from zero and one they can be uniformly bounded; it will become evident that for positive and finite source and channel dispersions, this is indeed the case.

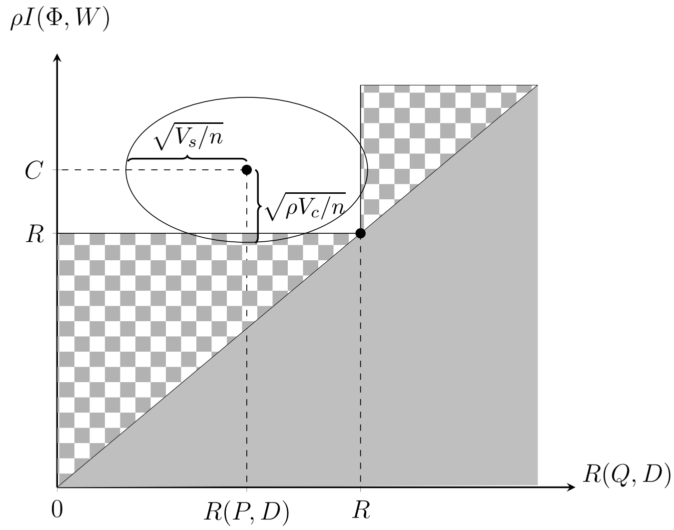

Fig. 2. Main JSCC excess distortion event: the loss of separation.

It follows, that up to the correction term it is optimal to choose *fixed* probabilities $\varepsilon_{S;n} = \varepsilon_S$ and $\varepsilon_{C;n} = \varepsilon_C$. Furthermore, the dependancy on n is the same as in the joint source-channel dispersion (23), but with different coefficient for the $1/\sqrt{n}$ term, i.e.,

$$R(D_n) = \rho C(W) - \sqrt{\frac{V_{\text{sep}}}{n}} Q^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{35}$$

Note that in the limit $\varepsilon \to 0$, $\sqrt{V_{\text{sep}}} = \sqrt{V_S} + \sqrt{\rho V_C}$.

In order to see why separation must have a loss, consider Fig. 1. The separation scheme designer is free to choose the digital interface rate R. Now whenever the random source-channel pair is either to the right of the point (R, R) due to a source type with $R(P_{\mathbf{S}}, D) > R$, or below it due to channel behavior $I(\Phi, P_{\mathbf{Y}|\mathbf{x}}) < R$, an excess distortion event will occur. Comparing to optimal JSCC, this adds the chessboard-pattern area on the plot. The designer may optimize R such that the probability of this area is minimized, but for any choice of R it will still have a strictly positive probability.

For quantifying the loss, it is tempting to look at the ratio between the coefficients of the $1/\sqrt{n}$ terms. However, this ratio may be in general infinite or negative, making the comparison difficult. We choose to define the equivalent probability $\tilde{\varepsilon}$ by rewriting (35) as

$$R(D_n) = \rho C - \sqrt{\frac{V(P, D, W, \rho)}{n}} Q^{-1}(\tilde{\varepsilon}) + O\left(\frac{\log n}{n}\right). \tag{36}$$

Thus, $\tilde{\varepsilon} < \varepsilon$ is the excess-distortion probability that a JSCC scheme could achieve under the same conditions, when the seperation scheme achieves ε. Substitution reveals that

$$\tilde{\varepsilon}(\varepsilon, \lambda) = Q\left(\frac{\min_{\varepsilon_s * \varepsilon_c \leq \varepsilon}\left[Q^{-1}(\varepsilon_s) + \sqrt{\lambda} Q^{-1}(\varepsilon_c)\right]}{\sqrt{1+\lambda}}\right), \tag{37}$$

where

$$\lambda \triangleq \frac{\rho V_c}{V_s}. \tag{38}$$

In general, numerical optimization is needed in order to obtain the equivalent probability. However, clearly $\tilde{\varepsilon}(\varepsilon, \lambda) = \tilde{\varepsilon}(\varepsilon, 1/\lambda)$. In the special symmetric case $\lambda = 1$ one may verify that the optimal probabilities are

$$\varepsilon_s = \varepsilon_c = 1 - \sqrt{1 - \varepsilon},$$

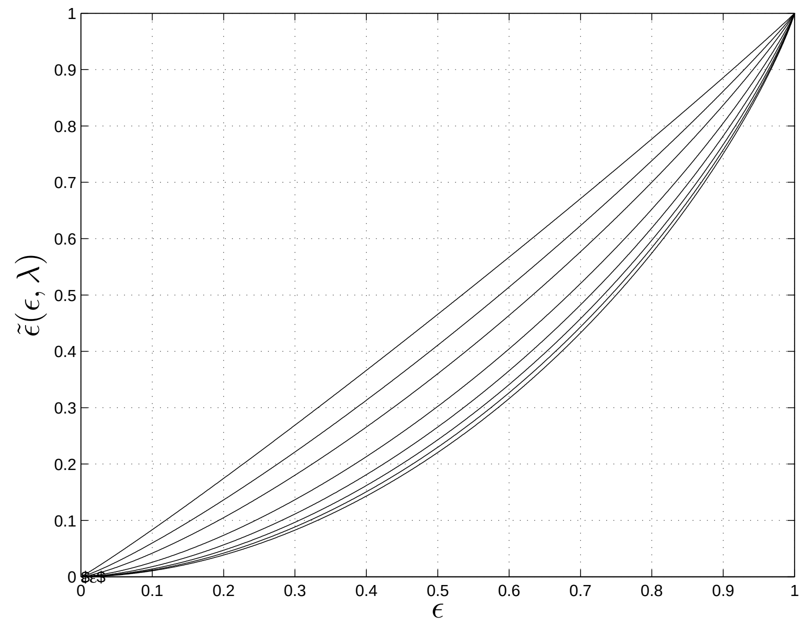

Fig. 3. $\tilde{\varepsilon}(\varepsilon, \lambda)$ as a function of ε for different values of λ. From bottom to top curve, $\lambda = \{1, 2, 3, 5, 10, 30, 100, 1000\}$.

thus

$$\tilde{\varepsilon}(\varepsilon, 1) = Q\left(\sqrt{2}Q^{-1}\left(1 - \sqrt{1-\varepsilon}\right)\right).$$

This reflects a large loss for low error probabilities. On the other hand,

$$\lim_{\lambda\to 0} \tilde{\varepsilon}(\varepsilon, \lambda) = \lim_{\lambda\to\infty} \tilde{\varepsilon}(\varepsilon, \lambda) = \varepsilon.$$

It seems that the symmetric case is the worst for separation, while when λ grows away from 1, either the source or the channel behave deterministically in the scale of interest, making the JSCC problem practically a digital one, i.e., either source coding over a clean channel or channel coding of equi-probable messages. This is somewhat similar to the loss of separation in terms of excess distortion exponent. This behavior is depicted in Fig. 3.

VI. BW Expansion and Lossless JSCC

We now wish to change the rules, by allowing the BW expansion ratio ρ, which was hitherto considered constant, to vary with the blocklength n. More specically, we takes some sequence ρ_n with $\lim_{n\to\infty} \rho_n = \rho$. It is not hard to verify that the results of Section IV remain valid[5], and ρ_n and D_n are related via:

$$R(P, D_n) = \rho_n C(W) - \sqrt{\frac{V_J(P, W, \rho)}{n}} Q^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right), \tag{39}$$

where for the calculation of the JSCC dispersion we use $D^*(P, W, \rho)$. In particular, one may choose to work with a *fixed* distortion threshold $D = D^*(P, W, \rho)$, and then (39) describes the convergence of the BW expansion ratio sequence to its limit ρ.

Equipped with this, we can now formulate a meaningful lossless JSCC dispersion problem. In (nearly) lossless coding we demand $\hat{\mathbf{S}} = \mathbf{S}$, otherwise we say that an *error event* $\mathcal{E}$ has occurred. We can see this as a special case of the lossy JSCC problem with Hamming distortion:

$$d(s_i, \hat{s}_i) = \begin{cases} 1 & \hat{s}_i = s_i \\ 0 & \text{otherwise}, \end{cases}$$

[5] Note that now the application of Berry-Esseen theorem is more involved, as we are now summing $\rho_n n + n$ independent random variables. However, its application still holds and results in Section IV can be proved by keeping track of ρ_n explicitly.

and with distortion threshold $D = 0$. While this setting does not allow for varying distortion thresholds, one may be interested in the number of channel uses needed to ensure a fixed error probability ε, as a function of the blocklength n. As an immediate corollary of (39), this is given by:

$$\rho_n = \frac{H(P)}{C(W)} + \sqrt{\frac{V_J(P,W,\rho)}{n}}\frac{Q^{-1}(\varepsilon)}{C(W)} + O\left(\frac{\log n}{n}\right). \tag{40}$$

In lossless JSCC dispersion, the source part of $V_J(P,W,\rho)$ simplifies to $\mathrm{Var}\left[\log P\right]$, in agreement with the lossless source coding dispersion of Strassen [4].

ACKNOWLEDGEMENT

The authors would like to thank Fadi Alajaji, Uri Erez, Meir Feder, Ligong Wang and Gregory W. Wornell for helpful discussions.

APPENDIX A
PROOFS FOR UEP CHANNEL CODING DISPERSION

In this appendix we provides proofs for results in Section III. We start by analyzing the Taylor expansion of empirical mutual information in Appendix A-A, which is crucial for proving Lemma 3, then we proceed to prove Lemmas 3 and 4 in Appendix A-B.

A. Analysis of the empirical mutual information

In this section, we investigate the Taylor expansion of the empirical mutual information at expected mutual information, i.e.,

$$I(\Phi, P_{\mathbf{Y}|\mathbf{x}}) = I(\Phi, W) + \sum_{x\in\mathcal{X}, y\in\mathcal{Y}} (P_{\mathbf{Y}|\mathbf{x}}(y|x) - W(y|x)) I'_W(y|x) \tag{A.41}$$

$$+ O\left(\sum_{x\in\mathcal{X}, y\in\mathcal{Y}} (P_{\mathbf{Y}|\mathbf{x}}(y|x) - W(y|x))^2\right), \tag{A.42}$$

where $I'_W(y|x) \triangleq \left.\frac{\partial I(\Phi,V)}{\partial V(y|x)}\right|_{V=W}$. Specifically, we characterize the first-order and higher-order correction terms of the Taylor expansion via Lemmas 8 and 10.

Lemma 8 (First order correction term for mutual information)**.** *If* $\mathbf{Y} \sim W^n(\cdot|\mathbf{x})$*, then*

$$\sum_{x\in\mathcal{X}, y\in\mathcal{Y}} (P_{\mathbf{Y}|\mathbf{x}}(y|x) - W(y|x)) I'_W(y|x) = \sum_x \sum_{j: j\in\mathcal{J}_x} Z_{x,j}$$

where $\mathcal{J}_x \triangleq \{j : x_j = x\}$, $\{Z_{x,j}, x\in\mathcal{X}, j\in\mathcal{J}_x\}$ *are independent random variables, and for a given* x, $\{Z_{x,j}, j\in\mathcal{J}_x\}$ *are identically distributed. Furthermore,*

$$\mathbb{E}\left[Z_{x,j}\right] = 0, \quad \forall\, x, j$$

$$\sum_x \sum_{j: j\in\mathcal{J}_x} \mathrm{Var}\left[Z_{x,j}\right] = \frac{V(\Phi, W)}{n},$$

$$\sum_x \sum_{j: j\in\mathcal{J}_x} \mathbb{E}\left[\left|Z_{x,j} - \mathbb{E}\left[Z_{x,j}\right]\right|^3\right] = O\left(\frac{1}{n^2}\right).$$

Proof of Lemma 8: Note

$$\begin{aligned}
\sum_{x\in\mathcal{X},y\in\mathcal{Y}}(P_{\mathbf{Y}|\mathbf{x}}(y|x)-W(y|x))I'_W(y|x) &= \sum_x\left[\sum_y(P_{\mathbf{Y}|\mathbf{x}}(y|x)-W(y|x))I'_W(y|x)\right]\\
&= \sum_x\left[\sum_y P_{\mathbf{Y}|\mathbf{x}}(y|x)I'_W(y|x)-\sum_y W(y|x)I'_W(y|x)\right]\\
&= \sum_x\left[\frac{1}{N(x|\mathbf{x})}\sum_y N_{x,y}(\mathbf{x},\mathbf{Y})I'_W(y|x)-E[I'_W(Y|x)]\right]\\
&= \sum_x\frac{1}{N(x|\mathbf{x})}\sum_{j:j\in\mathcal{J}_x}\left[I'_W(Y_j|x)-E[I'_W(Y|x)]\right].
\end{aligned}$$

Let $\tilde{Z}_{x,j}\triangleq I'_W(Y_j|x)-E[I'_W(Y|x)$ and $Z_{x,j}=\frac{1}{N(x|\mathbf{x})}\tilde{Z}_{x,j}$, then $\mathbb{E}\left[Z_{x,j}\right]=0$ and

$$\mathrm{Var}\left[\tilde{Z}_{x,j}\right]=\mathrm{Var}\left[I'_W(Y_j|x)\right]=\mathrm{Var}\left[I'_W(Y|x)\right].$$

By straightforward differentiation,

$$I'_W(y|x)=\left.\frac{\partial I(\Phi,V)}{\partial V(y|x)}\right|_{V=W}=\Phi(x)\log\frac{W(y|x)}{\Phi W(y)},$$

thus

$$\mathrm{Var}\left[\tilde{Z}_{x,j}\right]=\mathrm{Var}\left[I'_W(Y|x)\right]=\Phi^2(x)\mathrm{Var}\left[\log\frac{W(Y|x)}{\Phi W(Y)}\right].$$

Therefore

$$\begin{aligned}
\sum_x\sum_{j:j\in\mathcal{J}_x}\mathrm{Var}\left[Z_{x,j}\right] &= \sum_x\sum_{j:j\in\mathcal{J}_x}\frac{1}{N(x|\mathbf{x})^2}\mathrm{Var}\left[\tilde{Z}_{x,j}\right]\\
&= \sum_x\frac{1}{n\Phi(x)}\Phi(x)^2\mathrm{Var}\left[\log\frac{W(Y|x)}{\Phi W(Y)}\right]\\
&= \frac{1}{n}\sum_x\Phi(x)\mathrm{Var}\left[\log\frac{W(Y|x)}{\Phi W(Y)}\right]=\frac{V(\Phi,W)}{n}.
\end{aligned}$$

Finally, since any $Z_{x,j}$ is discrete and finite valued variables, the sum of the absolute third moment of these variables is bounded by some function $r_n=\Theta\left(\frac{1}{n^2}\right)$. ■

To investigate the higher order terms, we partition the channel realizations by its closeness to the true channel distribution W. Given input distribution Φ_n, we define

$$\Xi_n\triangleq\Xi_n\left(\Phi_n\right)\triangleq\left\{V\in\mathcal{P}_n\left(\mathcal{Y}|\Phi_n\right):\sum_{x,y}(V(y|x)-W(y|x))^2\leq|\mathcal{X}|\cdot|\mathcal{Y}|\cdot\frac{\log n}{n}\cdot\frac{1}{\Phi_n^{\min}}\right\}, \tag{A.43}$$

where $\Phi_n^{\min}\triangleq\min_{x\in\mathcal{X}}\Phi_n(x)$. As shown below in Lemma 9, Ξ_n is "typical" in the sense that it contains a channel realization with high probability.

Lemma 9. *If $\mathbf{x}\in\mathcal{X}^n$ has a type Φ_n and $\mathbf{Y}\in\mathcal{Y}^n$ is the output of the channel W^n with input $\mathbf{x}$, then*

$$\mathbb{P}\left[P_{\mathbf{Y}|\mathbf{x}}\notin\Xi_n\right]\leq\frac{2|\mathcal{X}|\cdot|\mathcal{Y}|}{n^2}.$$

Proof of Lemma 9: Let $\beta^2 = |\mathcal{X}|\cdot|\mathcal{Y}|\cdot\frac{\log n}{n}\cdot\frac{1}{\Phi_n^{\min}-\frac{1}{n}}$.

$$\begin{aligned}\mathbb{P}\left[P_{\mathbf{Y}|\mathbf{x}}\notin\Xi_n\right] &= \mathbb{P}\left[\sum_{a\in\mathcal{X},b\in\mathcal{Y}}\left(P_{\mathbf{Y}|\mathbf{x}}(b|a)-W(b|a)\right)^2>\beta^2\right]\\ &\overset{(a)}{\le}\mathbb{P}\left[\bigcup_{a\in\mathcal{X},b\in\mathcal{Y}}\left\{\left(P_{\mathbf{Y}|\mathbf{x}}(b|a)-W(b|a)\right)^2>\tfrac{\beta^2}{|\mathcal{X}||\mathcal{Y}|}\right\}\right]\\ &\overset{(b)}{\le}\sum_{a\in\mathcal{X},b\in\mathcal{Y}}\mathbb{P}\left[\left(P_{\mathbf{Y}|\mathbf{x}}(b|a)-W(b|a)\right)^2>\tfrac{\beta^2}{|\mathcal{X}||\mathcal{Y}|}\right]\\ &=\sum_{a\in\mathcal{X},b\in\mathcal{Y}}\mathbb{P}\left[\left|P_{\mathbf{Y}|\mathbf{x}}(b|a)-W(b|a)\right|>\tfrac{\beta}{\sqrt{|\mathcal{X}||\mathcal{Y}|}}\right],\end{aligned}\tag{A.44}$$

where (a) follows from the fact that in order for a sum of $|\mathcal{X}||\mathcal{Y}|$ elements to be above β^2, then at least one of the summands must be above $\beta^2/(|\mathcal{X}||\mathcal{Y}|)$. (b) follows from the union bound. For any $a\in\mathcal{X}, b\in\mathcal{Y}$, we have

$$\begin{aligned}&\mathbb{P}\left[\left|P_{\mathbf{Y}|\mathbf{x}}(b|a)-W(b|a)\right|>\tfrac{\beta}{\sqrt{|\mathcal{X}||\mathcal{Y}|}}\right]\\ =&\mathbb{P}\left[\left|\tfrac{1}{N_a(\mathbf{x})}\sum_{i:x_i=a}\left(\mathbb{1}_{Y_i=b}-W(b|a)\right)\right|>\tfrac{\beta}{\sqrt{|\mathcal{X}||\mathcal{Y}|}}\right]\\ \overset{(a)}{\le}&2\exp\left(-\frac{2\beta^2N_a(\mathbf{x})}{|\mathcal{X}|\cdot|\mathcal{Y}|}\right)\\ =&2\exp\left(-\frac{2\beta^2n\Phi_n(a)}{|\mathcal{X}|\cdot|\mathcal{Y}|}\right),\end{aligned}\tag{A.45}$$

where (a) follows from Hoeffding's inequality (see, e.g. [13, p. 191]). Applying (A.45) to each of the summands of (A.44) gives

$$\begin{aligned}\mathbb{P}\left[P_{\mathbf{Y}|\mathbf{x}}\notin\Xi_n\right]&\le\sum_{a\in\mathcal{X},b\in\mathcal{Y}}\mathbb{P}\left[\left|P_{\mathbf{Y}|\mathbf{x}}(b|a)-W(b|a)\right|>\tfrac{\beta}{\sqrt{|\mathcal{X}||\mathcal{Y}|}}\right]\\ &\le\sum_{a\in\mathcal{X},b\in\mathcal{Y}}2\exp\left(-\frac{2\beta^2n\Phi_n(a)}{|\mathcal{X}|\cdot|\mathcal{Y}|}\right)\\ &\le 2|\mathcal{Y}|\sum_{a\in\mathcal{X}}\exp\left(-\frac{2\beta^2n\Phi_n(a)}{|\mathcal{X}|\cdot|\mathcal{Y}|}\right)\\ &\le 2|\mathcal{X}|\cdot|\mathcal{Y}|\exp\left(-\frac{2\beta^2n\Phi_n^{\min}}{|\mathcal{X}|\cdot|\mathcal{Y}|}\right)\\ &=2|\mathcal{X}|\cdot|\mathcal{Y}|\frac{1}{n^2}.\end{aligned}\tag{A.46}$$

■

With Lemma 9, we can show that the higher order terms in (A.41) is in some sense negligible via Lemma 10.

Lemma 10 (Second order correction term for mutual information)**.** *If* $\mathbf{Y}\sim W^n(\cdot|\mathbf{x})$*, then exists* $J=J(|\mathcal{X}|,|\mathcal{Y}|,P_{\mathbf{x}})$ *such that*

$$\mathbb{P}\left[\sum_{x\in\mathcal{X},y\in\mathcal{Y}}\left(P_{\mathbf{Y}|\mathbf{x}}(y|x)-W(y|x)\right)^2\ge J\frac{\log n}{n}\right]\le\frac{2|\mathcal{X}||\mathcal{Y}|}{n^2}$$

Proof of Lemma 10: Let

$$J=|\mathcal{X}|\cdot|\mathcal{Y}|\cdot\frac{\log n}{n}\cdot\frac{2}{\Phi_n^{\min}}$$

then the lemma follows from the definition of Ξ_n and Lemma 9. ■

Finally, we show the following lemma that is useful for asymptotic analysis.

Lemma 11. *If $f_n = O(g_n)$, then there exist Γ_n and $\Gamma'_n = \Theta(\Gamma_n)$ such that*

$$\mathbb{P}\left[f_n \geq \Gamma'_n\right] \leq \mathbb{P}\left[g_n \geq \Gamma_n\right]$$
$$\mathbb{P}\left[f_n \leq -\Gamma'_n\right] \leq \mathbb{P}\left[g_n \geq \Gamma_n\right]$$

when n sufficiently large.

Proof of Lemma 10: By definition there exists $c > 0$ such that when n sufficiently large,

$$-cg_n \leq f_n \leq cg_n$$

Then letting $\Gamma'_n = c\Gamma_n$ completes the proof. ■

B. Proofs for UEP channel coding lemmas

In this section we provide proofs for Lemmas 3 and 4.

Proof for Lemma 3: We directly prove the stronger result where ΔR is defined according to (14). By Taylor expansion, we have

$$\begin{aligned} I(\Phi, P_{\mathbf{Y}|\mathbf{x}}) = I(\Phi, W) &+ \sum_{x\in\mathcal{X}, y\in\mathcal{Y}} (P_{\mathbf{Y}|\mathbf{x}}(y|x) - W(y|x)) I'_W(y|x) \\ &+ O\left(\sum_{x\in\mathcal{X}, y\in\mathcal{Y}} (P_{\mathbf{Y}|\mathbf{x}}(y|x) - W(y|x))^2\right), \end{aligned}$$

where $I'_W(y|x) \triangleq \left.\frac{\partial I(\Phi,V)}{\partial V(y|x)}\right|_{V=W}$. Let

$$A(\mathbf{Y}) = \sum_{x\in\mathcal{X}, y\in\mathcal{Y}} (P_{\mathbf{Y}|\mathbf{x}}(y|x) - W(y|x)) I'_W(y|x)$$

and

$$B(\mathbf{Y}) = O\left(\sum_{x\in\mathcal{X}, y\in\mathcal{Y}} (P_{\mathbf{Y}|\mathbf{x}}(y|x) - W(y|x))^2\right),$$

then

$$\begin{aligned} \varepsilon + \delta_n &= \mathbb{P}\left[I\left(\Phi_n, P_{\mathbf{Y}|\mathbf{x}}\right) \leq I(\Phi, W) - \Delta R, \mathbf{Y} \sim W^n(\cdot|\mathbf{x})\right] \\ &= \mathbb{P}\left[A(\mathbf{Y}) + B(\mathbf{Y}) \leq -\Delta R, \mathbf{Y} \sim W^n(\cdot|\mathbf{x})\right] \\ &\overset{(a)}{\geq} \mathbb{P}\left[A(\mathbf{Y}) + \Gamma_n \leq -\Delta R, \mathbf{Y} \sim W^n(\cdot|\mathbf{x})\right] - \mathbb{P}\left[B(\mathbf{Y}) \geq \Gamma_n, \mathbf{Y} \sim W^n(\cdot|\mathbf{x})\right] \end{aligned} \tag{A.47}$$

where $\Gamma_n > 0$ and (a) follows from (D.81). Similarly, (D.81) indicates

$$\begin{aligned} \varepsilon + \delta_n &= \mathbb{P}\left[A(\mathbf{Y}) + B(\mathbf{Y}) \leq -\Delta R, \mathbf{Y} \sim W^n(\cdot|\mathbf{x})\right] \\ &\leq \mathbb{P}\left[A(\mathbf{Y}) - \Gamma_n \leq -\Delta R, \mathbf{Y} \sim W^n(\cdot|\mathbf{x})\right] + \mathbb{P}\left[B(\mathbf{Y}) \leq -\Gamma_n, \mathbf{Y} \sim W^n(\cdot|\mathbf{x})\right] \end{aligned} \tag{A.48}$$

Let $\Gamma'_n = J(\Phi_n, |\mathcal{X}|, |\mathcal{Y}|)$ in Lemma 10, then from Lemmas 10 and 11, there exists $\Gamma_n = \Theta(\Gamma'_n) = O\left(\frac{\log n}{n}\right)$ such that

$$\mathbb{P}\left[B(\mathbf{Y}) \geq \Gamma_n, \mathbf{Y} \sim W^n(\cdot|\mathbf{x})\right] \leq O\left(\frac{1}{n^2}\right), \tag{A.49}$$

$$\mathbb{P}\left[B(\mathbf{Y}) \leq -\Gamma_n, \mathbf{Y} \sim W^n(\cdot|\mathbf{x})\right] \leq O\left(\frac{1}{n^2}\right). \tag{A.50}$$

In addition, based on Lemma 8 and $Q(x) = 1 - Q(-x)$, we can apply Berry-Esseen theorem (see, e.g., [3, Ch. XVI.5]) and have for any $-\infty < \lambda < \infty$,

$$\left|\mathbb{P}\left[A(\mathbf{Y}) \geq \lambda\sigma, \mathbf{Y} \sim W^n\left(\cdot|x\right)\right] - Q\left(\lambda\right)\right| \leq \frac{T}{\sigma^3}, \tag{A.51}$$

$$\left|\mathbb{P}\left[A(\mathbf{Y}) \leq -\lambda\sigma, \mathbf{Y} \sim W^n\left(\cdot|x\right)\right] - Q\left(\lambda\right)\right| \leq \frac{T}{\sigma^3}, \tag{A.52}$$

where $\sigma^2 = V(\Phi, W)/n$ and T is bounded by c/n^2, where c is some constant. Denote $V\left(\Phi, W\right)$ as V, apply $\lambda_1 = (\Delta R + \Gamma_n)/\sigma$ and $\lambda_2 = (\Delta R - \Gamma_n)/\sigma$ to (A.51) and (A.52) respectively,

$$\left|\mathbb{P}\left[A(\mathbf{Y}) \geq \Delta R + \Gamma_n, \mathbf{Y} \sim W^n\left(\cdot|\mathbf{x}\right)\right] - Q\left(\left(\Delta R + \Gamma_n\right)\sqrt{\frac{n}{V}}\right)\right| \leq \frac{c}{\sqrt{nV^3}}, \tag{A.53}$$

$$\left|\mathbb{P}\left[A(\mathbf{Y}) \leq -(\Delta R - \Gamma_n), \mathbf{Y} \sim W^n\left(\cdot|\mathbf{x}\right)\right] - Q\left(\left(\Delta R - \Gamma_n\right)\sqrt{\frac{n}{V}}\right)\right| \leq \frac{c}{\sqrt{nV^3}}. \tag{A.54}$$

Therefore,

$$\begin{aligned} Q\left((\Delta R + \Gamma_n)\sqrt{\frac{n}{V}}\right) - \frac{c}{\sqrt{nV^3}} &\overset{(A.53)}{\leq} \mathbb{P}\left[A(\mathbf{Y}) \geq \Delta R + \Gamma_n, \mathbf{Y} \sim W^n\left(\cdot|\mathbf{x}\right)\right] \\ &\overset{(A.47)}{\leq} \varepsilon + \delta_n + \mathbb{P}\left[B(\mathbf{Y}) \geq \Gamma_n, \mathbf{Y} \sim W^n\left(\cdot|\mathbf{x}\right)\right] \\ &\overset{(A.49)}{=} \varepsilon + O\left(\frac{\log n}{\sqrt{n}}\right). \end{aligned}$$

Likewise,

$$\begin{aligned} Q\left((\Delta R - \Gamma_n)\sqrt{\frac{n}{V}}\right) + \frac{c}{\sqrt{nV^3}} &\overset{(A.54)}{\geq} \mathbb{P}\left[A(\mathbf{Y}) \leq -(\Delta R - \Gamma_n), \mathbf{Y} \sim W^n\left(\cdot|\mathbf{x}\right)\right] \\ &\overset{(A.48)}{\geq} \varepsilon + \delta_n - \mathbb{P}\left[B(\mathbf{Y}) \leq -\Gamma_n, \mathbf{Y} \sim W^n\left(\cdot|\mathbf{x}\right)\right] \\ &\overset{(A.50)}{=} \varepsilon + O\left(\frac{\log n}{\sqrt{n}}\right). \end{aligned}$$

From the smoothness of Q^{-1} around ε,

$$\begin{aligned} (\Delta R + \Gamma_n)\sqrt{\frac{n}{V}} &\geq Q^{-1}\left(\varepsilon + O\left(\frac{\log n}{\sqrt{n}}\right) + \frac{c}{\sqrt{nV^3}}\right) = Q^{-1}\left(\varepsilon\right) + O\left(\frac{\log n}{\sqrt{n}}\right), \\ (\Delta R - \Gamma_n)\sqrt{\frac{n}{V}} &\leq Q^{-1}\left(\varepsilon + O\left(\frac{\log n}{\sqrt{n}}\right) - \frac{c}{\sqrt{nV^3}}\right) = Q^{-1}\left(\varepsilon\right) + O\left(\frac{\log n}{\sqrt{n}}\right). \end{aligned}$$

Therefore,

$$\begin{aligned} \Delta R &\geq \sqrt{\frac{V}{n}}Q^{-1}\left(\varepsilon\right) + \sqrt{\frac{V}{n}}O\left(\frac{\log n}{\sqrt{n}}\right) - \Gamma_n = \sqrt{\frac{V}{n}}Q^{-1}\left(\varepsilon\right) + O\left(\frac{\log n}{n}\right), \\ \Delta R &\leq \sqrt{\frac{V}{n}}Q^{-1}\left(\varepsilon\right) + \sqrt{\frac{V}{n}}O\left(\frac{\log n}{\sqrt{n}}\right) + \Gamma_n = \sqrt{\frac{V}{n}}Q^{-1}\left(\varepsilon\right) + O\left(\frac{\log n}{n}\right), \end{aligned}$$

and finally

$$\Delta R = \sqrt{\frac{V}{n}}Q^{-1}\left(\varepsilon\right) + O\left(\frac{\log n}{n}\right).$$

■

Before proving Lemma 4, we include the following lemma [2] for completeness.

Lemma 12 ([2, Lemma 6]). *Given $\mathcal{X}$ and positive integers n, k_n, let*

$$\eta_n \triangleq \frac{2}{n}\left(|\mathcal{X}|^2 + \log(n+1) + \log k_n + 1\right).$$

Then for arbitrary (not necessarily distinct) distributions $\Phi_i \in \mathcal{P}_n(\mathcal{X})$ and positive integers N_i with

$$\frac{1}{n}\log N_i \le H(\Phi_i) - \eta_n, \quad i = 1, 2, \ldots, m,$$

there exist m disjoint sets $\mathcal{A}_i \subset \mathcal{X}^n$ such that

$$\mathcal{A}_i \subset \mathcal{T}^n_{\Phi_i}, |\mathcal{A}_i| = N_i, \quad i = 1, 2, \ldots, m,$$

and

$$|\mathcal{T}_{\bar{V}}(\mathbf{x})| \le N_j \exp\left\{-n\left[I\left(\Phi_i, \bar{V}\right) - \eta_n\right]\right\} \text{ if } \mathbf{x} \in \mathcal{A}_i$$

for every i, j and $\bar{V} : \mathcal{X}^n \to \mathcal{X}^n$, except for the case $i = j$ and $\bar{V}$ is the identity matrix.

Proof for Lemma 4: For $\mathbf{x}' \in \mathcal{A}_j$, $\mathbf{x}' \ne \mathbf{x}$, let the joint type for the triple $(\mathbf{x}, \mathbf{x}', \mathbf{y})$ be given as the joint distribution of RV's X, X', Y. Then from Lemma 12, we can find $\{\mathcal{A}_i\}$ such that $\mathcal{A}_i \subset \mathcal{T}^n_{\Phi_i}$ and $\frac{1}{n}\log N_i \le H(\Phi_i) - \eta_n$, thus X has distribution Φ_i and X' has distribution Φ_j. In addition, define

$$\mathcal{B}_V \triangleq \mathcal{B}_V(\mathbf{x}) \triangleq \left\{\mathbf{y} \in \mathcal{T}^n_V(\mathbf{x}) : \exists\, \mathbf{x}' \ne \mathbf{x} \text{ such that } \mathbf{x}' \in \mathcal{A}_j \text{ and } I\left(\mathbf{x}' \wedge \mathbf{y}\right) - R_j \ge \gamma\right\},$$

then the cardinality of $\cup_V \mathcal{B}_V$ is upper bounded by

$$\begin{aligned} |\cup_V \mathcal{B}_V| &\le N_j \exp\left\{-n\left[I\left(X, X'; Y\right) - H(Y|X) - \eta_n\right]\right\} \\ &\le N_j \exp\left\{nH(Y|X) - n\left|I\left(X, X'; Y\right) - \eta_n\right|^+\right\} \end{aligned}$$

Then for $\mathbf{y} \in \mathcal{B}_V$,

$$W(\mathbf{y}\,|\,\mathbf{x}) = \exp\{-n[D(V\,\|\,W|\Phi_i) + H(V|\Phi_i)]\}$$

Note that $I(\mathbf{x}' \wedge \mathbf{y}) - R_j \ge \gamma$ implies $I(X'; Y) - R_j \ge \gamma$, and $I(X, X'; Y) \ge I(X; Y)$,

$$I\left(X, X'; Y\right) - R_j \ge I\left(X'; Y\right) - R_j \ge \gamma$$

Hence,

$$\begin{aligned} W^n(\mathcal{B}_V|\,\mathbf{x}) &\le N_j \exp\left\{nH(Y|X) - n\left|I\left(X, X'; Y\right) - \eta_n\right|^+\right\} \exp\{-n[D(V\,\|\,W|\Phi_i) + H(V|\Phi_i)]\} \\ &= N_j \exp\left\{-n\left[D(V\,\|\,W|\Phi_i) + \left|I\left(X, X'; Y\right) - \eta_n\right|^+\right]\right\} \\ &\le N_j \exp\left\{-n\left[D(V\,\|\,W|\Phi_i) + |R_j + \gamma - \eta_n|^+\right]\right\} \end{aligned}$$

And

$$\begin{aligned} \mathbb{P}\left[I\left(\mathbf{x}' \wedge \mathbf{y}\right) - R_j \ge \gamma\right] &\le W^n\left(\bigcup_V \mathcal{B}_V \middle| \mathbf{x}\right) \\ &\le (n+1)^{|\mathcal{X}|^2|\mathcal{Y}|} N_j \exp\left\{-n\left[|R_j + \gamma - \eta_n|^+\right]\right\} \end{aligned}$$

■

Appendix B
Proofs for JSCC dispersion

This appendix contains proofs for results in Section IV. Similar to the development in Appendix A, we start by analyzing the Taylor expansion of the distortion-rate function in Appendix B-A, then prove the relevant key lemmas Appendix B-B.

A. Analysis of the distortion-rate function

In this section, we investigate that Taylor expansion of $R(P_{\mathbf{S}}, D_n)$. Denote the partial derivatives of $D(P,R)$ at $R = I(\Phi, W)$ and $Q = P$ as

$$D'_R \triangleq \left.\frac{\partial D(P,R)}{\partial R}\right|_{R=I(\Phi,W)},$$
$$D'_P(s) \triangleq \left.\frac{\partial D(Q,R)}{\partial Q(s)}\right|_{Q=P}.$$

Assuming $D(\cdot,\cdot)$ is smooth, Taylor expansion gives

$$\begin{aligned} D(P_{\mathbf{S}}, \rho I(\Phi, P_{\mathbf{Y}|\mathbf{x}}) + \xi'_n) &= D(P, \rho I(\Phi, W)) \\ &+ \sum_{s=1}^{|\mathcal{S}|} (P_{\mathbf{S}}(s) - P(s)) D'_P(s) \\ &+ (\rho I(\Phi, P_{\mathbf{Y}|\mathbf{x}}) + \xi'_n - \rho I(\Phi, W)) D'_R \\ &+ O\left(\sum_{s=1}^{|\mathcal{S}|} (P_{\mathbf{s}}(s) - P(s))^2 + (\rho I(\Phi, P_{\mathbf{Y}|\mathbf{x}}) + \xi'_n - \rho I(\Phi, W))^2 \right) \\ &= D(P, \rho I(\Phi, W)) \\ &+ \sum_{s=1}^{|\mathcal{S}|} \left(P_{\mathbf{S}}(s) - P(s)\right) D'_P(s) + \rho D'_R \sum_{x,y} \left(P_{\mathbf{Y}|\mathbf{x}}(y|x) - W(y|x) \right) I'_W(y|x) \qquad \text{(B.55)} \\ &+ B(\mathbf{S}, \mathbf{Y}, \xi'_n), \qquad \text{(B.56)} \end{aligned}$$

where $\xi'_n = O\left(\log n / n\right)$, and the correction term is

$$\begin{aligned} B(\mathbf{S}, \mathbf{Y}, \xi'_n) \triangleq& \xi'_n D'_R + O\left(\sum_{x,y} (P_{\mathbf{Y}|\mathbf{x}}(y|x) - W(y|x))^2 \right) \\ &+ O\left(\sum_{s=1}^{|\mathcal{S}|} (P_{\mathbf{s}}(s) - P(s))^2 + (\rho I(\Phi, P_{\mathbf{Y}|\mathbf{x}}) + \xi'_n - \rho I(\Phi, W))^2 \right). \qquad \text{(B.57)} \end{aligned}$$

For notational simplicity, we define

$$A(\mathbf{S}, \mathbf{Y}) \triangleq \sum_{s=1}^{|\mathcal{S}|} \left(P_{\mathbf{S}}(s) - P(s)\right) D'_P(s) + \rho D'_R \sum_{x,y} \left(P_{\mathbf{Y}|\mathbf{x}}(y|x) - W(y|x) \right) I'_W(y|x) \qquad \text{(B.58)}$$

The lemmas in this subsection is organized as follows. Lemma 13 shows that the first order terms of the Taylor expansion of $R(P_{\mathbf{S}}, D_n)$ with respect to P can be represented as the sum of n i.i.d. random variables. Then Lemma 14 shows that $A(\mathbf{S}, \mathbf{Y})$ can be represented represented as the sum of $n+m$ i.i.d. random variables. Finally, Lemmas 15 and 16 together with Lemmas 9 and 11 shows that the higher order terms in the Taylor expansion is negligible, as summarized in Lemma 17.

Lemma 13. *Under the conditions of Lemma 6,*

$$\sum_{s \in \mathcal{S}} (P_{\mathbf{S}}(s) - P(s)) D'_P(s) = \sum_{i=1}^{n} \tilde{Z}_i$$

where $\{\tilde{Z}_i, i = 1, 2, \cdots, n\}$ *are i.i.d. random variables such that*

$$\mathbb{E}\left[\tilde{Z}_i\right] = 0$$
$$\mathrm{Var}\left[\tilde{Z}_i\right] = \frac{V_D}{n^2}$$

where $V_D = V_S \cdot (D'_R)^2$.

Proof:

$$\begin{aligned}\sum_{s\in\mathcal{S}}(P_{\mathbf{S}}(s)-P(s))D'_P(s) &= \frac{1}{n}\sum_{i=1}^{n}D'_P(S_i)-\sum_{s\in\mathcal{S}}P(s)D'_P(s)\\ &= \frac{1}{n}\sum_{i=1}^{n}D'_P(S_i)-E[D'_P(S)]\end{aligned}$$

Let $\tilde{Z}_i \triangleq D'_P(S_i) - E[D'_P(S)]$, then

$$\mathbb{E}\left[\tilde{Z}_i\right] = 0,$$

and

$$\mathrm{Var}\left[D'_P(S_i) - E[D'_P(S)]\right] = \mathrm{Var}\left[D'_P(S)\right].$$

By elementary calculus it can be shown that for all $s \in \mathcal{S}$,

$$D'_P(s) = \frac{\partial D(P,R)}{\partial P(s)} = -\frac{\partial R(P,D)}{\partial P(s)}\frac{\partial D(P,R)}{\partial R} = -R'(s)D'_R.$$

Therefore,

$$V_D = \mathrm{Var}\left[D'_P(S)\right] = \mathrm{Var}\left[R'(S)\right](D'_R)^2 = V_S \cdot (D'_R)^2.$$

∎

Lemma 14 (First order correction term for distortion-rate function)**.** *Under the conditions of Lemma 6, (B.58), i.e.,* $A(\mathbf{S},\mathbf{Y})$ *is the sum of* $n+m$ *independent random variables, whose sum of variance is*

$$\frac{1}{n}\left[\rho(D'_R)^2V_S + \rho(D'_R)^2V\left(\Phi,W\right) + O\left(\frac{\log n}{n}\right)\right]$$

and sum of the absolute third moment is bounded by some constant.

Proof for Lemma 14: According to Lemmas 8 and 13, (B.58) can be interpreted as the sum of $n+m$ independent random variables. Let σ_n^2 be the sum of the variance of these $n+m$ variables, then

$$\begin{aligned}\sigma_n^2 &= n\frac{1}{n^2}V_D + \sum_{x\in\mathcal{X}} m\Phi_m(x)\left(\frac{\rho D'_R}{m\Phi_m(x)}\right)^2 V_C(x)\\ &= \frac{1}{n}V_D + \sum_{x\in\mathcal{X}}\frac{(\rho D'_R)^2}{m\Phi_m(x)}V_C(x).\\ &= \frac{1}{n}\left[V_D + \rho(D'_R)^2V\left(\Phi_m,W\right)\right]\\ &= \frac{1}{n}\left[\rho(D'_R)^2V_S + \rho(D'_R)^2V\left(\Phi,W\right) + O\left(\frac{\log n}{n}\right)\right] \end{aligned} \tag{B.59}$$

Define r to be the sum of the absolute third moment of these variables. Since these are discrete and finite valued variables, r is bounded by $\frac{1}{n^2}J_3$, for some constant J_3.

∎

To investigate the higher order terms, we partition the source type by its closeness to the source distribution P. Given source distribution P, we define

$$\Omega_n \triangleq \Omega_n(P) \triangleq \left\{Q \in \mathcal{T}_n : \|P-Q\|_2^2 \le |\mathcal{S}|\frac{\log n}{n}\right\}. \tag{B.60}$$

In addition, we show the following property of set Ξ_n (defined in (A.43) in Appendix A-A):

Lemma 15. *If* $P_{\mathbf{y}|\mathbf{x}} \in \Xi_n$*, then*

$$\left(I(\Phi, P_{\mathbf{y}|\mathbf{x}}) - I(\Phi, W)\right)^2 = O\left(\frac{\log n}{n}\right). \tag{B.61}$$

Proof: By definition of Ξ_n,

$$\sum_{x,y}(P_{\mathbf{y}|\mathbf{x}}(y|x) - W(y|x))^2 = O\left(\frac{\log n}{n}\right), \tag{B.62}$$

and therefore

$$\max_{x,y}(P_{\mathbf{y}|\mathbf{x}}(y|x) - W(y|x))^2 = O\left(\frac{\log n}{n}\right). \tag{B.63}$$

The zero-th order Taylor approximation of $I(\Phi, P_{\mathbf{y}|\mathbf{x}})$ around $W = P_{\mathbf{y}|\mathbf{x}}$ is given by

$$\begin{aligned} I(\Phi, P_{\mathbf{y}|\mathbf{x}}) &= I(\Phi, W) + O\left(\sum_{x,y}\left|P_{\mathbf{y}|\mathbf{x}}(y|x) - W(y|x)\right|\right) &\text{(B.64)}\\ &= I(\Phi, W) + O\left(\max_{x,y}\left|P_{\mathbf{y}|\mathbf{x}}(y|x) - W(y|x)\right|\right), &\text{(B.65)}\end{aligned}$$

therefore

$$(I(\Phi, P_{\mathbf{y}|\mathbf{x}}) - I(\Phi, W))^2 = O\left(\max_{x,y}\left|P_{\mathbf{y}|\mathbf{x}}(y|x) - W(y|x)\right|^2\right), \tag{B.66}$$

and the required result follows from (B.63). ■

The bounding of $B(\mathbf{S}, \mathbf{Y}, \xi'_n)$ is mainly based on the following lemma.

Lemma 16. *There exists constant* $J > 0$ *such that*

$$\mathbb{P}\left[\sum_{x,y}(P_{\mathbf{Y}|\mathbf{x}}(y|x) - W(y|x))^2 + \sum_{s=1}^{|\mathcal{S}|}(P_{\mathbf{s}}(s) - P(s))^2 + (\rho I(\Phi, P_{\mathbf{Y}|\mathbf{x}}) + \xi'_n - \rho I(\Phi, W))^2 \geq J\frac{\log n}{n}\right] \leq O\left(\frac{1}{n^2}\right)$$

Proof: Based on Lemma 15, we have

$$\begin{aligned} &\mathbb{P}\left[\sum_{x,y}(P_{\mathbf{Y}|\mathbf{x}}(y|x) - W(y|x))^2 + \sum_{s=1}^{|\mathcal{S}|}(P_{\mathbf{s}}(s) - P(s))^2 + (\rho I(\Phi, P_{\mathbf{Y}|\mathbf{x}}) + \xi'_n - \rho I(\Phi, W))^2 \geq J\frac{\log n}{n}\right] \\ &\leq \mathbb{P}\left[P_{\mathbf{S}} \notin \Omega_n \text{ or } P_{\mathbf{Y}|\mathbf{x}} \notin \Xi_n\right] \\ &\leq \mathbb{P}\left[P_{\mathbf{S}} \notin \Omega_n\right] + \mathbb{P}\left[P_{\mathbf{Y}|\mathbf{x}} \notin \Xi_n\right] \\ &\overset{(a)}{\leq} \frac{2|\mathcal{S}|}{n^2} + \frac{2|\mathcal{X}| \cdot |\mathcal{Y}|}{m^2} \\ &= O\left(\frac{1}{n^2}\right). \end{aligned}$$

(a) follows from Lemma 9 and [6, Lemma 2]. ■

Lemma 17 (Second order correction term for distortion-rate function)**.** *For* $\xi'_n = O\left(\frac{\log n}{n}\right)$*, there exists* $\Gamma_{n,1} = O\left(\frac{\log n}{n}\right)$ *and* $\Gamma_{n,2} = O\left(\frac{\log n}{n}\right)$ *such that*

$$\mathbb{P}\left[B(\mathbf{S}, \mathbf{Y}, \xi'_n) > \Gamma_{n,1}\right] \leq O\left(\frac{1}{n^2}\right) \tag{B.67}$$

$$\mathbb{P}\left[B(\mathbf{S}, \mathbf{Y}, \xi'_n) < -\Gamma_{n,2}\right] \leq O\left(\frac{1}{n^2}\right) \tag{B.68}$$

Proof: Let $\Gamma_{n,1} = \xi'_n D'_R + (J + |D'_R|)\log n/n$ and $\Gamma_{n,2} = -\xi'_n D'_R + (J + |D'_R|)\log n/n$, where the J is given by Lemma 16, then the proof follows from Lemma 16 and Lemma 11. ■

B. Proofs for JSCC lemmas

This section first shows Lemma 6 (JSCC Distortion Redundancy Lemma), upon which proofs for both the achievability and converse of the main theorem builds. Then it shows the proof for Lemma 7, which is essential for establishing the converse result.

Proof for Lemma 6: We directly prove the stronger result where D_n is defined according to (25).

We first note that for D_n,

$$\mathbb{P}\left[R(P_{\mathbf{S}}, D_n) \geq \rho I(\Phi_m, P_{\mathbf{Y}|\mathbf{x}}) + \xi_n\right] \geq \varepsilon + \zeta_n. \tag{B.69}$$

By Lemma 19, for any conditional type V, there is a constant $J_1 = J_1(|\mathcal{X}|, |\mathcal{Y}|)$ such that

$$|I(\Phi_m, V) - I(\Phi, V)| \leq J_1 \frac{\log m}{m},$$

Therefore,

$$\begin{aligned}
\varepsilon + \zeta_n &\leq \mathbb{P}\left[R(P_{\mathbf{S}}, D_n) \geq \rho I(\Phi_m, P_{\mathbf{Y}|\mathbf{x}}) + \xi_n\right] \\
&\leq \mathbb{P}\left[R(P_{\mathbf{S}}, D_n) \geq \rho I(\Phi, P_{\mathbf{Y}|\mathbf{x}}) - J_1 \tfrac{\log m}{m} + \xi_n\right] \\
&= \mathbb{P}\left[R(P_{\mathbf{S}}, D_n) \geq \rho I(\Phi, P_{\mathbf{Y}|\mathbf{x}}) + \xi'_n\right] \\
&= \mathbb{P}\left[D_n \leq D\left(P_{\mathbf{S}}, \rho I(\Phi, P_{\mathbf{Y}|\mathbf{x}}) + \xi'_n\right)\right],
\end{aligned} \tag{B.70}$$

where $xi'_n = O\left(\log n/n\right)$. Let $\Delta D_n \triangleq D_n - D^*$, (B.70) now becomes

$$\begin{aligned}
\varepsilon + \zeta_n &= \mathbb{P}\left[D_n \leq D(P_{\mathbf{S}}, \rho I(\Phi, P_{\mathbf{Y}|\mathbf{x}}) + \xi'_n)\right] \\
&= \mathbb{P}\left[\Delta D_n \leq A(\mathbf{S}, \mathbf{Y}) + B(\mathbf{S}, \mathbf{Y}, \xi'_n)\right].
\end{aligned}$$

Applying (D.78) and (D.79) gives

$$\begin{aligned}
\varepsilon + \zeta_n &\leq \mathbb{P}\left[A(\mathbf{S}, \mathbf{Y}) + \Gamma_{n,1} \geq \Delta D_n\right] + \mathbb{P}\left[B(\mathbf{S}, \mathbf{Y}, \xi'_n) > \Gamma_{n,1}\right] \\
\varepsilon + \zeta_n &\geq \mathbb{P}\left[A(\mathbf{S}, \mathbf{Y}) - \Gamma_{n,2} \geq \Delta D_n\right] - \mathbb{P}\left[B(\mathbf{S}, \mathbf{Y}, \xi'_n) < -\Gamma_{n,2}\right]
\end{aligned}$$

From Lemmas 11 and 17 we have

$$\begin{aligned}
\mathbb{P}\left[B(\mathbf{S}, \mathbf{Y}, \xi'_n) < -\Gamma_{n,2}\right] &\leq O\left(\frac{1}{n^2}\right) \\
\mathbb{P}\left[B(\mathbf{S}, \mathbf{Y}, \xi'_n) > \Gamma_{n,1}\right] &\leq O\left(\frac{1}{n^2}\right)
\end{aligned}$$

Since $\zeta_n = O\left(\frac{\log n}{\sqrt{n}}\right)$, we absorb the $O\left(1/n^2\right)$ terms and have:

$$\begin{aligned}
\varepsilon + O\left(\frac{\log n}{\sqrt{n}}\right) &\geq \mathbb{P}\left[A(\mathbf{S}, \mathbf{Y}) \geq \Delta D_n - \Gamma_{n,1}\right] \\
\varepsilon + O\left(\frac{\log n}{\sqrt{n}}\right) &\leq \mathbb{P}\left[A(\mathbf{S}, \mathbf{Y}) \geq \Delta D_n + \Gamma_{n,2}\right],
\end{aligned}$$

Based on Lemma 14, by the (non-i.i.d. version of the) Berry-Esseen theorem ([11, XVI.5, Theorem 2]) we have that for any a and n,

$$\left|\mathbb{P}\left[A(\mathbf{S}, \mathbf{Y}) \geq \lambda \cdot \sigma_n\right] - Q(\lambda)\right| \leq \frac{6T_n}{\sigma_n^3} = O\left(\frac{1}{\sqrt{n}}\right),$$

where T_n is bounded by c/n^2, with c being a constant. Let $\lambda_1 = (\Delta D_n - \Gamma_{n,1})/\sigma$ and $\lambda_2 = (\Delta D_n + \Gamma_{n,2})/\sigma$, then,

$$\begin{aligned}
\varepsilon + O\left(\frac{\log n}{\sqrt{n}}\right) &\geq Q((\Delta D_n - \Gamma_{n,1})/\sigma) + O\left(\frac{1}{\sqrt{n}}\right), \\
\varepsilon + O\left(\frac{\log n}{\sqrt{n}}\right) &\leq Q((\Delta D_n + \Gamma_{n,2})/\sigma) + O\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}$$

absorbing the $O\left(\frac{1}{\sqrt{n}}\right)$ on the right hand side, we have

$$\varepsilon + O\left(\frac{\log n}{\sqrt{n}}\right) \geq Q((\Delta D_n - \Gamma_{n,1})/\sigma_n),$$
$$\varepsilon + O\left(\frac{\log n}{\sqrt{n}}\right) \leq Q((\Delta D_n + \Gamma_{n,2})/\sigma_n).$$

From the smoothness of Q^{-1} around ε, noting $\Gamma_{n,i} = O\left(\log n/n\right), i = 1, 2$ and replace σ_n as in Lemma 14, we obtain

$$\Delta D_n \leq D'_R \sqrt{\frac{V_C + \rho V(\Phi, W)}{n}} Q^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right), \tag{B.71}$$

$$\Delta D_n \geq D'_R \sqrt{\frac{V_C + \rho V(\Phi, W)}{n}} Q^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{B.72}$$

Therefore,

$$\Delta D_n = D'_R \sqrt{\frac{V_C + \rho V(\Phi, W)}{n}} Q^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{B.73}$$

We add D^* and apply $R(P, D)$ to both sides of (B.71). With the Taylor approximation we have

$$R(P, D_n) = I(\Phi, W) + \sqrt{\frac{V_S + \rho V(\Phi, W)}{n}} Q^{-1}(\varepsilon)|D'_R|R'_D + O\left(\frac{\log n}{n}\right).$$

where $R'_D \triangleq \frac{\partial R(P,D)}{\partial D}$. Finally, note that D'_R is negative, and combined with the fact that $D'_R R'_D = 1$ we have the required

$$R(P, D_n) = I(\Phi, W) - \sqrt{\frac{V_S + \rho V(\Phi, W)}{n}} Q^{-1}\left(\varepsilon\right) + O\left(\frac{\log n}{n}\right).$$

In order to establish (24), write:

$$\varepsilon_n \triangleq \mathbb{P}\left[R(P_{\mathbf{S}}, D_n) > \rho I(\Phi_m(\mathbf{S}), P_{\mathbf{Y}|\mathbf{x}}) + \xi_n\right] = \sum_{\mathbf{s} \in \mathcal{S}^n} \mathbb{P}\left[\mathbf{S} = \mathbf{s}\right] \mathbb{P}\left[I(\Phi_m(\mathbf{s}), P_{\mathbf{Y}|\mathbf{x}}) < T_n(P_{\mathbf{s}})\right],$$

where

$$T_n(P_{\mathbf{s}}) \triangleq \frac{R(P_{\mathbf{s}}, D_n) - \xi_n}{\rho}.$$

Clearly, the optimal $\Phi_m(\mathbf{s})$ is only a function of $T_n(P_{\mathbf{s}})$. Thus,

$$\varepsilon_n \geq \sum_t \mathbb{P}\left[T_n(P_{\mathbf{s}}) = t\right] \mathbb{P}\left[I(\Phi_m(t), P_{\mathbf{Y}|\mathbf{x}}) < t\right]. \tag{B.74}$$

Without loss of generality we restrict the thresholds to those satisfying

$$t \geq C(W) - O\left(\frac{\log n}{n}\right), \tag{B.75}$$

since otherwise the theorem is satisfied trivially. Now define the set

$$\Pi(W, \delta) \triangleq \{\Phi \in \mathcal{P}\left(\mathcal{X}\right) : \exists \Phi^* \in \Pi(W) : \|\Phi - \Phi^*\| \leq \delta\}.$$

Since $I(\Phi, W)$ is concave in Φ, it follows that

$$\sup_{\Phi \notin \Pi(W,\delta)} I(\Phi, W) = C(W) - \epsilon(\delta)$$

where $\epsilon(\delta) > 0$ for any $\delta > 0$. Thus, for thresholds that satisfy (B.75) and for $\Phi \notin \Pi(W, \delta)$ (for any choice of $\delta > 0$):

$$\lim_{n \to \infty} \mathbb{P}\left[I(\Phi, P_{\mathbf{Y}|\mathbf{x}}) < t\right] = 1.$$

It follows that we may restrict $\Phi_m(t)$ in (B.74) to any set $\Pi(W,\delta)$ with $\delta > 0$. Since inside that set the Hessian of $I(P,W)$ (as a function of W) can be uniformly bounded (see [5, Appendix I]), we have that (A.53) and (A.54) holds uniformly (i.e. with the same constant A) for all $\Phi \in \Pi(W,\delta)$. Consequently,

$$\mathbb{P}\left[I(\Phi_m(t), P_{\mathbf{Y}|\mathbf{x}}) < t\right] \geq 1 - Q\left(\left(t - I(\Phi_m(t), W)\right)\sqrt{\frac{n}{V(\Phi_m(t), W)}}\right) + O\left(\frac{1}{\sqrt{n}}\right)$$

Since without the last correction term the probability is minimized by any $\Phi^*(W) \in \Pi(W)$ and that correction term is uniform, we have that

$$\mathbb{P}\left[I(\Phi_m(t), P_{\mathbf{Y}|\mathbf{x}}) < t\right] \geq \mathbb{P}\left[I(\Phi^*(W), P_{\mathbf{Y}|\mathbf{x}}) < t\right] - O\left(\frac{1}{\sqrt{n}}\right).$$

Then, (B.74) becomes:

$$\varepsilon_n + O\left(\frac{1}{\sqrt{n}}\right) \geq \sum_t \mathbb{P}\left[T_n(P_{\mathbf{s}}) = t\right] \mathbb{P}\left[I(\Phi^*(W), P_{\mathbf{Y}|\mathbf{x}}) < t\right] = \mathbb{P}\left[R(P_{\mathbf{S}}, D_n) > \rho I(\Phi^*(W), P_{\mathbf{Y}|\mathbf{x}}) + \xi_n\right]$$

Since the $O\left({}^1/\sqrt{n}\right)$ term may be included in a ξ_n sequence, it follows that one cannot do better, to the approximation required, then using a fixed input type $\Phi^*(W)$ for all source strings, resulting in (24).

■

To show the converse of the JSCC problem define in Section I, we first upper bound the fraction of source codeword that is D-covered by a given reconstruction sequence.

Lemma 18 (Restricted D-ball size)**.** *Given source type P and a reconstruction sequence $\hat{\mathbf{s}}$, define restricted D-ball as*

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

Then

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

Proof: 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

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

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

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

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

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

Note

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

hence

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

■

Remark 7. *Lemma 3 in [9], is similar to Lemma 18. However, it does not bound the size of the restricted D-ball uniformly, and we choose to prove Lemma 18, which is necessary for proving Lemma 7.*

Proof for Lemma 7: In our proof, we first bound the denominator in (28) uniformly for all $\mathbf{s}_i$, and then bound the sum of the numerator over all $\mathbf{s}_i$, as done in [8] for the channel error exponent.

a) Bounding the denominator: Based on standard results in method of types [14], 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

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

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

$$\mathbf{y} \in \mathcal{T}_V(f(\mathbf{s})) \cap \hat{B}(\mathbf{s}, D) \Rightarrow \mathbf{s} \in B(g_{J;n}(\mathbf{y}), Q, D) \cap G(Q,\Phi), \tag{B.76}$$

hence any $\mathbf{y}$ will be counted at most $|B(g_{J;n}(\mathbf{y}), Q, D) \cap G(Q,\Phi)|$ times. According to Lemma 18, 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 Φ. Therefore,

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

Noting

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

we have

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

Combining the bounds for both numerator and denominator, we have

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

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

$$p(n) = (\rho n + 1)^{\rho n |\mathcal{X}||\mathcal{Y}|} (n+1)^{n\left[(|\mathcal{S}||\hat{\mathcal{S}}|)(|\mathcal{X}|+1)(|\mathcal{S}|)\right]}, \tag{B.77}$$

and the proof is completed. ■

Appendix C
Continuity of the Mutual Information Function

In this section we show the continuity of the mutual information function, which shows that for investigation in dispersion, arguments based on types is essentially the same as arguments based on general probability distributions.

Lemma 19. *For $P, Q \in \mathcal{P}(\mathcal{X})$, if $\|P - Q\|_\infty \leq \delta \leq 1/(2|\mathcal{X}||\mathcal{Y}|)$, then*

$$|I(P,W) - I(Q,W)| \leq \delta |\mathcal{X}| \log |\mathcal{Y}| - |\mathcal{Y}||\mathcal{X}| \delta \log |\mathcal{X}| \delta.$$

Therefore, when $\delta = \Theta\left(\frac{1}{n}\right)$,

$$|I(P,W) - I(Q,W)| = O\left(\frac{\log n}{n}\right)$$

Proof: Let $P_Y = [P \times W]_Y$ and $Q_Y = [Q \times W]_Y$, note

$$\|P_Y - Q_Y\|_1 \leq \delta |\mathcal{X}| |\mathcal{Y}|.$$

Let $\delta' = |\mathcal{X}| \delta$, then Lemma 1.2.7 in [10] shows,

$$\begin{aligned}
|I(P,W) - I(Q,W)| &= |(H(P_Y) - H(W|P)) - (H(Q_Y) - H(W|Q))| \\
&\leq |(H(P_Y) - H(Q_Y))| + |(H(W|P) - H(W|Q))| \\
&\leq -|\mathcal{Y}| \delta' \log \delta' + \delta |\mathcal{X}| \log |\mathcal{Y}| \\
&= \delta |\mathcal{X}| \log |\mathcal{Y}| - |\mathcal{Y}||\mathcal{X}| \delta \log |\mathcal{X}| \delta.
\end{aligned}$$

■

Appendix D
Elementary Probability Inequalities

In this section we prove several simple probability inequalities used in our derivation.

Lemma 20. *Let A and B be two (generally dependent) random variables and let c be a constant. Then for any values $\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4$, the following holds:*

$$\mathbb{P}[A + B > c] \leq \mathbb{P}[A > c - \Gamma_1] + \mathbb{P}[B > \Gamma_1], \tag{D.78}$$
$$\mathbb{P}[A + B > c] \geq \mathbb{P}[A > c + \Gamma_2] - \mathbb{P}[B < -\Gamma_2], \tag{D.79}$$
$$\mathbb{P}[A + B < c] \leq \mathbb{P}[A < c + \Gamma_3] + \mathbb{P}[B < -\Gamma_3], \tag{D.80}$$
$$\mathbb{P}[A + B < c] \geq \mathbb{P}[A < c - \Gamma_4] - \mathbb{P}[B > \Gamma_4]. \tag{D.81}$$

Proof: To show (D.78), let $\mathcal{E}_A = \{A > c - \Gamma_1\}$, $\mathcal{E}_B = \{B > \Gamma_1\}$, and $\mathcal{E} = \{A + B > c\}$. Note that

$$\mathcal{E}_A{}^c \bigcap \mathcal{E}_B{}^c \subseteq \mathcal{E}^c,$$

hence by De Morgan's law,

$$\mathcal{E}_A \bigcup \mathcal{E}_B \supseteq \mathcal{E}.$$

We prove (D.78) by the union bound

$$\mathbb{P}[\mathcal{E}] \leq \mathbb{P}[\mathcal{E}_A] + \mathbb{P}[\mathcal{E}_B].$$

Apply (D.78) on $-A, -B, -c$ and Γ_2, we obtain (D.79) after rearrangement.
Subtract 1 from both sides of (D.79) and replace Γ_2 by Γ_3, we obtain (D.80) after rearrangement.
Apply (D.79) on $-A, -B, -c$ and Γ_4, we obtain obtain (D.81) after rearrangement.

■

References

[1] 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.

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

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

[4] V. Strassen. Asymptotische abschätzungen in shannon?s informationstheorie. In *Trans. Third Prague Conf. Information Theory*, pages 679–723, 1962.

[5] Y. Polyanskiy, H.V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. *IEEE Trans. Info. Theory*, 56(5):2307–2359, May 2010.

[6] A. Ingber and Y. Kochman. The dispersion of lossy source coding. In *Proc. of the Data Compression Conference*, Snowbird, Utah, March 2011.

[7] V. Kostina and S. Verdu. Fixed-length lossy compression in the finite blocklength regime: Discrete memoryless sources. In *IEEE International Symposium on Information Theory Proceedings (ISIT), St. Petersburg, Russia*, pages 41–45, Aug. 2011.

[8] G. Dueck and J. Korner. Reliability function of a discrete memoryless channel at rates above capacity. *infotheory,*, 25(1):82–85, jan 1979.

[9] 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.

[10] I. Csiszár and J. Korner. *Information Theory - Coding Theorems for Discrete Memoryless Systems*. Academic Press, New York, 1981.

[11] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, New York, 2nd edition, 1971.

[12] B. Yu and T.P. Speed. A rate of convergence result for a universal d-semifaithful code. *IEEE Trans. Info. Theory*, 39(3):813 –820, may. 1993.

[13] David Pollard. *Convergence of stochastic processes*. Springer, 1984.

[14] I. Csiszár and J. Korner. *Information Theory - Coding Theorems for Discrete Memoryless Systems*. Academic Press, New York, 1981.