On Lossy Multi-Connectivity: Finite Blocklength Performance and Second-Order Asymptotics

Lin Zhou, Albrecht Wolf and Mehul Motani

Abstract

We consider the lossy transmission of a single source over parallel additive white Gaussian noise channels with independent quasi-static fading, which we term as the lossy multi-connectivity problem. We assume that only the decoder has access to the channel state information. Motivated by ultra-reliable and low latency communication requirements, we are interested in the finite blocklength performance of the problem, i.e., the minimal excess-distortion probability of transmitting \( k \) source symbols over \( n \) channel uses. By generalizing non-asymptotic bounds by Kostina and Verdú for the lossy joint source-channel coding problem, we derive non-asymptotic achievability and converse bounds for the lossy multi-connectivity problem. Using these non-asymptotic bounds and under mild conditions on the fading distribution, we derive approximations for the finite blocklength performance in the spirit of second-order asymptotics for any discrete memoryless source under an arbitrary bounded distortion measure. Furthermore, in the achievability part, we analyze the performance of a universal coding scheme by modifying the universal joint source-channel coding scheme by Csiszár and using a generalized minimum distance decoder. Our results demonstrate that the asymptotic notions of outage probability and outage capacity are in fact reasonable criteria even in the finite blocklength regime. Finally, we illustrate our results via numerical examples.

Index Terms

Parallel channels, Dispersion, 5G, Ultra-reliable and low latency, Finite blocklength analysis, Quasi-static fading, Joint source-channel coding

Lin Zhou and Mehul Motani are with the Department of Electrical and Computer Engineering, National University of Singapore (Emails: {ljzhou, motani}@nus.edu.sg). Albrecht Wolf is with Vodafone Chair Mobile Communication Systems, Technical University Dresden (Email: albrecht.wolf@tu-dresden.de).

The material in this work has been accepted in part to the IEEE GLOBECOM, Abu Dhabi, UAE, Dec. 2018 [1].
I. Introduction

In cellular communication for 5G systems, one is interested in ultra-reliable and low latency communication (URLLC). To gain insight for the wireless communication system for 5G, in this paper, using tools from finite blocklength information theory, we characterize the fundamental limits of lossy data transmission over parallel additive Gaussian white noise (AWGN) channels with independent quasi-static fading. This problem is also known as the multi-connectivity problem in its lossless form [2]. Thus, in this paper, we term our problem as the lossy multi-connectivity problem.

The main reason for considering quasi-static fading is that for low-latency communication systems, the length of the data packet is rather small and usually smaller than the channel coherence time of fading channels. This observation was made previously by Yang et al. [3]. The motivation for multi-connectivity is mainly for the reliable communication over fading channels, especially for the quasi-static fading channels suitable for low-latency communication. If only one channel is available and unfortunately, the channel is in deep fade condition, then it is clear that nothing can be transmitted reliably over the channel. However, when multiple channels are available, if at least one channel is not in deep fade condition, then some information can always be transmitted in a reliable manner. Hence, multi-connectivity provides a flexible communication framework that can trade diversity for multiplexing via the multiple routes to the destination. Furthermore, multi-connectivity can use different carrier frequencies, such that the multiple copies of the same information can, in the best case, be delivered in a single time slot.

To the best of our knowledge, existing works on quasi-static fading channels mainly focus on outage analysis with infinite signal to noise ratio (SNR) via the diversity-multiplexing tradeoff (DMT) ([4], [5], [6]). For the lossless multi-connectivity problem, the outage analysis was established by Wolf et al. [2]. However, as pointed by Yang et al. [3], the outage analysis is not necessarily a valid criterion for low-latency communication without careful verification. The main reason why one adopts the notion of outage analysis for quasi-static fading channel is that the classical capacity with vanishing error probability proposed by Shannon [7] is zero for most commonly encountered quasi-static fading channels [8, Chapter
However, capacity itself is an asymptotic notion requiring infinite blocklength and thus contradicts the need for low-latency communication. To fully understand the finite blocklength performance of channel coding over quasi-static fading channels, Yang et al. [3] adopted tools from finite block length information theory for point-to-point channels [9], [10], [11] and adapted the ideas to quasi-static MIMO channels.

In this paper, in the spirit of [3], we provide a detailed analysis for the lossy multi-connectivity problem using tools from finite blocklength information theory [10], [12], [13] and thus provide insights for the designers of future 5G URLLC systems. Our lossy multi-connectivity problem is essentially a lossy joint source channel coding (JSCC) problem where the channel consists of parallel AWGN channels with independent quasi-static fading.

A. Related Works

For the AWGN channel, the finite blocklength performance was characterized by Polyanskiy, Poor and Verdú [10], by Hayashi [9] and by Tan and Tomamichel [14]. Subsequently, the result was generalized to block fading channels by Polyanskiy and Verdú in [15]. Furthermore, for MIMO channel with quasi-static fading, under mild conditions on the fading distributions, the finite blocklength analysis was performed by Yang et al. in [3]. In general, it was shown in [3] that for quasi-static fading channels, the maximum number of messages per channel use which can transmitted over \( n \) uses of the quasi-static channel allowing average error probability \( \varepsilon \) is given by the outage capacity [8, Chapter 5] plus a remainder term which scales in the order of \( O(\log n/n) \). Yang et al. proved the interesting result by generalizing the \( \kappa/\beta \)-bound [10, Theorem 25] in the achievability part and the meta-converse theorem [10, Theorem 30] in the converse part. For the single input multiple output (SIMO) channel with quasi-static Rayleigh fading, MolavianJazi and Laneman [16] derived a similar result using information spectrum method for both directions. For other works on finite blocklength analysis for quasi-static fading channels, see [17] and the references therein.

We also review existing works on the lossy source coding (rate-distortion) problem dating back to Shannon [18]. The error exponent, which characterizes the speed of exponential decay of the excess-
distortion probability for the rate-distortion problem, was characterized by Marton [19] for any discrete memoryless source (DMS) under any bounded distortion measure and by Ihara and Kubo [20] for any Gaussian memoryless source (GMS) under the quadratic distortion measure. In terms of second-order asymptotics, for any DMS under any bounded distortion measure, the result was derived by Ingber and Kochman [21] using method of types [22] and a refined version of the type covering lemma [19]. Similar results were also derived by Kostina and Verdú for any DMS under any bounded distortion measure and any GMS under the quadratic distortion measure. In particular, Kostina and Verdú [12] derived non-asymptotic achievability and converse bounds using the so-called distortion-tilted information density. Subsequently, Kostina and Verdú [13] also generalized the results in [12] to derive the finite blocklength performance for the lossy joint source-channel coding problem [18].

For the lossy multi-connectivity problem with two parallel AWGN channels and quasi-static fading, Laneman et al. [23] considered the source and channel diversity for multiple settings of separate source-channel coding and joint source-channel coding under the average quadratic distortion criterion in the limit of high SNR. Note that, however, in this paper, we are interested in the finite blocklength performance for the lossy multi-connectivity problem for any value of SNR and any finite number of parallel AWGN channels with quasi-static fading under the excess-distortion probability criterion.

B. Main Contributions

First, we derive non-asymptotic converse and achievability bounds for the lossy multi-connectivity problem by generalizing the corresponding results for the lossy joint source-channel coding problem by Kostina and Verdú [13]. The non-asymptotic bounds hold for any value of SNR and any source distribution, discrete or continuous. In the expression of the non-asymptotic bounds, we make use of the distortion-tilted information density [12] and the fading information density [24].

Second, we derive the second-order asymptotics of the lossy multi-connectivity problem for discrete memoryless sources under bounded distortion measures. These results provide approximation for the finite blocklength performance. Both the achievability and converse parts follow by applying the Berry-Esseen
theorem to our non-asymptotic bounds and analyzing the remainder term judiciously. We remark that the JSCC scheme used to prove our non-asymptotic achievability bound is non-universal in the sense that the decoder has access to the source distribution and the channel law, as well as channel state information (CSI). Motivated by practical universal communication systems, we adapt a universal joint source-channel coding scheme dating back to Csiszár and used also by Wang, Ingber and Kochman [25] to our lossy multi-connectivity setting. Inspired by [26], our JSCC scheme combines unequal error protection [27] and a modified minimum distance decoder [28]. We show that the universal scheme achieves the same second-order asymptotics as the non-universal one and thus demonstrates that the knowledge of the source distribution and the channel law at the decoder are not necessary to achieve the optimal second-order performance.

Finally, we demonstrate the benefit of multi-connectivity numerically by comparing our results as the number of parallel channels varies. We show that with multiple parallel channels, the achievable excess-distortion probability $P_{e,k,n}(P, D) \ (\text{see } (7))$ is significantly decreased.

C. Organization of the Rest of the Paper

The rest of the paper is organized as follows. In Section II, we set up the notation and formulate our lossy multi-connectivity problem. Subsequently, in Section III, we present our main results: non-asymptotic bounds and second-order asymptotics. The proofs of second-order asymptotics are given in Section IV. For the smooth presentation of our main results, we defer all other proofs to the appendices.

II. PROBLEM FORMULATION

A. Notation

Random variables and their realizations are denoted in capital letters (e.g., $X$) and lower case letters (e.g., $x$), respectively. All sets (e.g., alphabets of random variables) are denoted in calligraphic font (e.g., $\mathcal{X}$). Let $X^n := (X_1, \ldots, X_n)$ be a random vector of length $n$ and $x^n$ a particular realization. We use $\|x^n\|$ to denote the $\ell_2$ norm $\sqrt{\sum_i x_i^2}$. The set of all probability distribution on $\mathcal{X}$ is denoted as $\mathcal{P}(\mathcal{X})$ and the
The set of all conditional probability distributions from $\mathcal{X}$ to $\mathcal{Y}$ is denoted as $\mathcal{P}(\mathcal{Y}|\mathcal{X})$. We use $\mathcal{R}$, $\mathcal{R}_+$ and $\mathcal{N}$ to denote the sets of real numbers, non-negative real numbers, and natural numbers, respectively. Given any two integers $a, b \in \mathcal{N}$, we use $[a : b]$ to denote the inclusive collection of all integers between $a$ and $b$, i.e., $[a : b] := \{c : c \in \mathcal{N}, a \leq c \leq b\}$. Furthermore, we use $[a]$ to denote the set $[1 : a]$ for any integer $a$. We use $\mathbf{I}_n$ to denote the $n \times n$ identity matrix. Given a real number $a \in \mathcal{R}$, we use $|a|^+$ to denote $\max\{a, 0\}$. Finally, all logarithms are natural logarithms and we use standard asymptotic notation such as $O(\cdot)$, $o(\cdot)$ and $\Theta(\cdot)$ [29].

### B. System Model

In this paper, as shown in Figure 1, we consider the lossy transmission of a single memoryless source over parallel AWGN channels with independent quasi-static fading. We assume that the memoryless source is generated i.i.d. according to a distribution $P_S$ defined on the alphabet $\mathcal{S}$, i.e., for any integer $k \in \mathcal{N}$ and any $s^k \in \mathcal{S}^k$, $P_{S^k}(s^k) = \prod_{i \in [k]} P_S(s_i)$.

Furthermore, we assume that there are in total $\Psi$ parallel AWGN channels with independent quasi-static fading. For each channel indexed by $t \in [\Psi]$, the additive noise $Z^n_t$ is generated i.i.d. according to the normal distribution, i.e., $\mathcal{N}(0, 1)$. Furthermore, the fading coefficient $A_t \in \mathcal{A}$ for each channel ($t \in [\Psi]$) remains unchanged in $n$ channel uses and the fading parameters $\{A_t\}_{t \in [\Psi]}$ are distributed i.i.d. according to a fading distribution $P_A$. Thus, the channel law for each independent channel indexed by $t \in [\Psi]$ is given by

$$Y^n_t = A_t X^n_t + Z^n_t, \; t \in [\Psi].$$  \hspace{1cm} (1)

Throughout the paper, we assume a maximal power constraint $P$, i.e., for any codewords $\{X^n_t\}_{t \in [\Psi]}$ for parallel channels, $\frac{1}{n} \sum_{t \in [\Psi]} \|X^n_t\|^2 \leq P$. Furthermore, we assume that only the receiver has access to CSI, i.e., $\{A_t\}_{t \in [\Psi]}$. Given the channel outputs $\{Y^n(t)\}_{t \in [\Psi]}$ and CSI $\{A_t\}_{t \in [\Psi]}$, the decoder estimates the source sequence as $\hat{S}^k$, which takes values in the alphabet $\hat{\mathcal{S}}^k$. To measure the performance, we define the
one-shot distortion measure $d : S \times S \to [0, \infty)$ and its multi-letter version $d(s^k, \hat{s}^k) := \frac{1}{k} \sum_{i \in [k]} d(s_i, \hat{s}_i)$ for any pair of source sequence $s^k \in S^k$ and the reproduced sequence $\hat{s}^k \in \hat{S}^k$.

To formally define a code for our problem, for each $t \in [\Psi]$, given any $P_t \in \mathbb{R}_+$, let $\mathbb{R}(n, P_t)$ be the collection of all length-$n$ sequences of real numbers with power $P_t$, i.e.,

$$\mathbb{R}(n, P_t) := \{x^n \in \mathbb{R}^n : \|x^n\|^2 = nP_t \}. \quad (2)$$

**Definition 1.** An $(k, n, P, D)$-code consists of

- **$\Psi$ encoders**

  $$f_t : S^k \to \mathbb{R}(n, P_t), \ \forall \ t \in [\Psi], \ \text{and} \quad (3)$$

- **a decoder**

  $$\phi : \prod_{t \in [\Psi]} \mathbb{R}(n, P_t) \times A^\Psi \to \hat{S}^k, \quad (4)$$

  for some power allocation vector $\{P_t\}_{t \in [\Psi]}$ satisfying

  $$\sum_{t \in [\Psi]} P_t \leq P. \quad (5)$$

Given an $(k, n, P, D)$-code, the source estimate can be expressed as $\hat{S}^k = \phi(\{Y^\infty_t, A^\Psi_t\}_{t \in [\Psi]})$ and the excess-distortion probability is defined as follows:

$$P_{e,k,n}(P, D) := \Pr\{d(s^k, \hat{s}^k) > D\}. \quad (6)$$

Given $(k, n, P, D)$, the optimal performance is evaluated by the minimal excess-distortion probability of any $(k, n, P, D)$-code, i.e.,

$$P_{e,k,n}^*(P, D) := \inf\{\varepsilon : \exists \ \text{an} \ (k, n, P, D)\text{-code s.t.} \ P_{e,k,n}(P, D) \leq \varepsilon\}. \quad (7)$$

Symmetrically, we let $k^*(n, \varepsilon, P, D)$ denote the maximum number of source symbols that can be transmitted over $n$ uses of parallel AWGN channels with maximum power constraint $P$ and multiplicative quasi-static fading so that the excess-distortion probability $P_{e,k,n}(P, D)$ is bounded above by $\varepsilon$, i.e.,

$$k^*(n, \varepsilon, P, D) := \sup\{k : \exists \ \text{an} \ (k, n, P, D)\text{-code s.t.} \ P_{e,k,n}(P, D) \leq \varepsilon\}. \quad (8)$$

In this paper, we will present asymptotic and non-asymptotic bounds on $P_{e,k,n}^*(P, D)$ and $k^*(n, \varepsilon, P, D)$. 
Fig. 1: System model for the lossy multi-connectivity problem with two parallel channels (i.e., $\Psi = 2$). We are interested in the finite blocklength performance of lossy transmission over parallel AWGN channels with independent quasi-static fading. We assume that only the decoder has access to the channel state information, i.e., the fading parameters $\{A_t\}_{t \in [\Psi]}$.

III. MAIN RESULTS

A. Preliminaries

Recall the definitions of the rate-distortion function $R(P_S, D)$ [18], [19] and the distortion-tilted information density [30], [12], i.e.,

$$R(P_S, D) := \inf_{P_{S|\hat{S}}} \mathbb{E}[d(S, \hat{S}) \leq D],$$

$$j(s, D) := -\log \mathbb{E}[\exp(\lambda^* D - d(s, \hat{S}))],$$

$$\lambda^* := -\frac{\partial R(P_S, D)}{\partial D},$$

where the expectation in (10) is with respect to the unconditional distribution $P_{\hat{S}}^*$, which is induced by the optimal test channel $P_{\hat{S}|S}^*$ for (9) and the source distribution $P_S$.

The definition of the distortion-tilted information density in (10) can also be extended to multi-letter case and thus we can define $j(s^k, D)$ for any length-$k$ source sequence $s^k$. It was shown in [12] that for any memoryless source, we have

$$j(s^k, D) = \sum_{i \in [k]} j(s_i, D).$$
Recall the channel law in (1). Thus, for each \( t \in [\Psi] \), the conditional distribution of the channel output \( Y^n_t \) given the input \( X^n_t \) and the fading parameter \( A_t \) is a product of normal distributions, i.e., for any \((x^n_t, a_t, y^n_t)\),

\[
P_{Y^n_t|X^n_t, A_t}(y^n_t|x^n_t, a_t) = \prod_{i\in[n]} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{\|y_{t,i} - a_t x_{t,i}\|^2}{2} \right).
\]

(13)

Now, let \( \{Q_{Y^n_t|A_t}\}_{t\in[\Psi]} \) be arbitrary conditional distributions and define the fading information densities [16, Eq. (9)] as

\[
\tilde{i}(x^n_t; y^n_t|a_t) := \log \frac{P_{Y^n_t|X^n_t, A_t}(y^n_t|x^n_t, a_t)}{Q_{Y^n_t|A_t}(y^n_t|a_t)}.
\]

(14)

For simplicity, throughout the paper, we let \( P := (P_1, \ldots, P_\Psi) \) to denote the power allocation vector and use \( A := (A_1, \ldots, A_\Psi) \) to denote fading parameters. Furthermore, we let \( \mathcal{P}_{\max}(P) := \{P : \sum_{t\in[\Psi]} P_t \leq P\} \) and \( \mathcal{P}_{\text{eq}}(P) := \{P : \sum_{t\in[\Psi]} P_t = P\} \) be the set of power allocation vectors satisfying the maximum and equal power constraint respectively.

### B. Non-Asymptotic Bounds

Our first result is a lower bound on the excess-distortion probability for any \((k, n, P, D)\)-code.

**Theorem 1.** Any \((k, n, P, D)\)-code satisfies that

\[
P_{e,k,n}(P, D) \geq \inf_{(P, \{P_{X^n_t|S^k}\}_{t\in[\Psi]})} \sup_{\gamma > 0} \left( -\exp(-n\gamma) + \Pr \left\{ \sum_{t\in[\Psi]} \tilde{i}(X^n_t, Y^n_t|A_t) \leq \beta(S^k, D) - n\gamma \right\} \right).
\]

(15)

We remark that the proof of Theorem 1 is similar to [13, Theorem 1] and omitted for simplicity. Note that in the right hand side of (15), the infimum over all stochastic encoders \( P_{X^n_t|S^k}, t \in [\Psi] \) makes the lower bound difficult to calculate. However, the following lemma states that for any AWGN channel with quasi-static fading, the right hand side of (15) does not depend on encoders under particular choices of auxiliary distributions \( \{Q_{Y^n_t|A_t}\}_{t\in[\Psi]} \).
To present the lemma, for each \( t \in [\Psi] \), given the power \( P_t \) and the fading coefficient \( A_t \), define

\[
L^n_t(P_t, A_t, Z^n_t) := \frac{n}{2} \log(1 + P_t A_t^2) + \sum_{j \in [n]} \frac{P_t A_t^2 (1 - Z_{t,j}^2)}{2(1 + P_t A_t^2)}.
\]  

(16)

**Lemma 2.** For each \( t \in [\Psi] \), given \( P_t \) and \( a_t \), choose the distribution \( Q_{Y^n_t | A_t} \) such that

\[
Q_{Y^n_t | A_t}(\cdot | a_t) \sim \mathcal{N}(0, (1 + P_t a_t^2)I_n).
\]  

(17)

Then, for any \( t \in [\Psi] \), channel input \( x^n_t \) and fading parameter \( a_t \), under the distribution of \( P_{Y^n_t | X^n_t A_t}(\cdot | x^n_t, a_t) \) (see (13)), the distribution of the fading information density \( \tilde{i}(x^n_t; Y^n_t | a_t) \) depends on \( x^n_t \) only through its power \( P_t = \frac{\|x^n_t\|^2}{n} \). Furthermore, \( \tilde{i}(x^n_t; Y^n_t | a_t) \) has the same distribution as \( L^n_t(P_t, A_t, Z^n_t) \).

The proof of Lemma 2 follows from spherical symmetry and is given in Appendix A.

Invoking Theorem 1 and Lemma 2, we can now obtain the following non-asymptotic converse bound.

**Corollary 3.** Any \((k, n, P, D)\)-code satisfies that

\[
P_{e,k,n}(P, D) \geq \inf_{P \in \mathcal{P}_{\max}(P)} \sup_{\gamma > 0} \left\{ - \exp(-n\gamma) + \Pr \left\{ \sum_{t \in [\Psi]} L^n_t(P_t, A_t, Z^n_t) \leq j(S^n, D) - n\gamma \right\} \right\}.
\]  

(18)

In the following, we present a non-asymptotic achievability bound by generalizing [13, Theorems 7 and 8] to our lossy multi-connectivity setting. Recall the definition of \( L^n_t(\cdot) \) in (16) and that \( P^*_{S} \) is the distribution induced by the optimal test channel for the rate-distortion function \( R(P_S, D) \) (see (9)). Given any source sequence \( s^k \), define the non-excess-distortion probability

\[
\Phi(s^k, D) := \Pr_{P^*_{S}}\{d(s^k, \hat{s}^k) \leq D\}.
\]  

(19)

Finally, for simplicity, let

\[
\tilde{L}^n_t(P_t, A_t, Z^n_t) := L^n_t(P_t, A_t, Z^n_t) - \log \frac{1 + P_t A_t^2}{\sqrt{1 + 2P_t A_t^2}}.
\]  

(20)

**Theorem 4.** There exists an \((k, n, P, D)\)-code such that

\[
P_{e,k,n}(P, D)
\leq \inf_{P \in \mathcal{P}_{\max}(P)} \inf_{\gamma > 0} \left\{ \exp(-n\gamma + 1) + \mathbb{E} \left[ \exp \left( - \sum_{t \in [\Psi]} \tilde{L}^n_t(P_t, A_t, Z^n_t) - \log \frac{n\gamma}{\Phi(S^n, D)} \right) \right] \right\}.
\]  

(21)
The proof of Theorem 4 is given in Appendix B. We adapt the joint source-channel coding scheme in [13] which consists of the concatenation of source and channel codes. We remark that the coding scheme to prove Theorem 4 is non-universal since the decoder needs to know the exact channel law and the source distribution.

We remark that the result in Theorem 4 holds for arbitrary source distributions (discrete and continuous). As we will show later, using non-asymptotic bounds in Corollary 3 and Theorem 4, we can derive the second-order asymptotics for our lossy multi-connectivity problem, which provides tight approximation for \( k^*(n, \varepsilon, P, D) \) (see (8)). For the case of lossless transmission, the results in Corollary 3 and Theorem 4 holds with \( j(S^k, D) \) and \( \Phi(S^k, D) \) replaced by \( \log P_S^k(S^k) \).

C. Second-Order Asymptotics for a DMS

In this subsection, we consider any discrete memoryless source under arbitrary bounded distortion measure, i.e., \( S \) and \( \hat{S} \) are finite and \( \max_{(s, \hat{s}) \in S \times \hat{S}} d(s, \hat{s}) < \infty \). Before presenting the main theorem, we first recall and define necessary quantities. Recall the definitions of the distortion-dispersion function and the third absolute moment of the distortion-tilted information density [12], i.e.,

\[
V_s(P_S, D) := \text{Var}[j(S, D)], \quad (22)
\]
\[
T(P_S, D) := \mathbb{E}[|j(S, D)^3|]. \quad (23)
\]

Furthermore, given power allocation vector \( \mathbf{P} \) and fading parameters \( \mathbf{A} \), define

\[
U_1(\mathbf{P}, \mathbf{A}) := \sum_{t \in [\Psi]} \frac{1}{2} \log(1 + P_t A_t^2), \quad (24)
\]
\[
U_2(\mathbf{P}, \mathbf{A}) := \sum_{t \in [\Psi]} \frac{P_t A_t^2 (P_t A_t^2 + 2)}{2(1 + P_t A_t^2)}, \quad (25)
\]
\[
U_3(\mathbf{P}, \mathbf{A}) := \sum_{t \in [\Psi]} \frac{\sqrt{2} P_t^2 A_t^4 (7P_t A_t^2 + 12)}{(\sqrt{2}(1 + P_t A_t^2))^3}. \quad (26)
\]

For our result to hold, we need the following assumptions:

(i) Given source distribution \( P_S \) and distortion level \( D \),

\[
V_s(P_S, D) > 0, \quad T(P_S, D) < \infty; \quad (27)
\]
(ii) Given any $P \in \mathcal{P}_{eq}(P)$,

$$ \mathbb{E}_A \left[ \frac{U_3(P, A)}{U_2(P, A)} \right] < \infty, \quad \mathbb{E}_A \left[ \left( \frac{U_3(P, A)}{U_2(P, A)} \right)^2 \right] < \infty. \quad (28) $$

**Theorem 5.** Suppose $k = \Theta(n)$. Under the assumptions in (27) and (28), the optimal excess-distortion probability satisfies that

$$ P^*_{c,k,n}(P, D) = \inf_{P \in \mathcal{P}_{eq}(P)} \mathbb{E}_A \left[ Q \left( \frac{nU_1(P, A) - kR(P_S, D)}{\sqrt{kV_s(P_S, D) + nU_2(P, A)}} \right) \right] + O \left( \frac{\log n}{\sqrt{n}} \right). \quad (29) $$

The proof of Theorem 5 is given in Section IV-A. Several remarks are in order.

First, the result in Theorem 5 follows by applying the Berry-Esseen theorem to our non-asymptotic bounds in Corollary 3 and Theorem 4. Since the result in Theorem 4 is based on a non-universal JSCC scheme, we also prove the achievability part using a universal JSCC scheme which uses unequal error protection [27] and modified minimum distance decoding. We remark that the additional requirement

$$ \mathbb{E}_P \left[ \frac{1(\mathcal{A} = 0)}{\mathcal{A}} \right] < \infty \quad \text{and} \quad \mathbb{E}_P [\mathcal{A}] < \infty $$

is needed in order for our universal JSCC scheme to achieve the performance in Theorem 5.

Second, as can be seen in Theorem 5, the excess-distortion probability is dominated by the first term, which can be understood as the outage probability in the second-order sense. Actually, using [3, Lemma 17], we can show that under mild conditions on the fading distribution $P_A$,

$$ \inf_{P \in \mathcal{P}_{eq}(P)} \mathbb{E}_A \left[ Q \left( \frac{nU_1(P, A) - kR(P_S, D)}{\sqrt{kV_s(P_S, D) + nU_2(P, A)}} \right) \right] = P_{out}(k, n, D) + O \left( \frac{1}{n} \right), \quad (30) $$

where the outage probability is defined as

$$ P_{out}(k, n, D) := \inf_{P \in \mathcal{P}_{eq}(P)} \Pr \left\{ nU_1(P, A) \leq kR(P_S, D) \right\}. \quad (31) $$

Some interesting observations can be made. Theorem 5 suggests that in order to achieve the optimal excess-distortion probability, a joint source-channel coding scheme is required, as indicated by the denominator $kV_s(P_S, D) + nU_2(P, A)$. However, (30) indicates that with negligible loss of performance, a separate source-channel coding scheme can also achieve the optimal excess-distortion probability. Furthermore,
from [8, Exercise 5.17], we know that the outage probability $P_{\text{out}}(k, n, D)$ for parallel fading channels with i.i.d. fading parameters is achieved by equal power allocation, i.e., $P_t = \frac{P}{n}$ for all $t \in [\Psi]$.

Finally, we remark a bit on the assumptions for Theorem 5. The assumption in (28) is mild since it is satisfied by many common fading distributions, such as the Rayleigh fading [24]. Furthermore, the assumption that $k = \Theta(n)$ is valid since we can show that by allowing any non-vanishing excess-distortion probability, the asymptotic ratio of $\frac{k}{n}$ is finite and positive.

Using Theorem 5 and (30), we obtain the following approximation for $k^*(n, \varepsilon, P, D)$ in (8). To present the result, given any maximal power constraint $P \in \mathbb{R}_+$ and any excess-distortion probability $\varepsilon \in (0, 1)$, define the outage capacity for our lossy multi-connectivity problem as

$$C_{\text{out}}^{\text{MC}}(P, \varepsilon) := \sup \left\{ r \in \mathbb{R}_+ : P_{\text{out}}(r, 1, D) \leq \varepsilon \right\}.$$  

**Corollary 6.** Assume that the outage probability satisfies that

$$\frac{\partial P_{\text{out}}(r, 1, D)}{\partial r} \bigg|_{r = C_{\text{out}}^{\text{MC}}(P, \varepsilon)} > 0.$$  

Then, under the conditions of Theorem 6, for any maximal power constraint $P \in \mathbb{R}_+$ and any excess-distortion probability $\varepsilon \in (0, 1)$, we have that

$$k^*(n, \varepsilon, P, D) = nC_{\text{out}}^{\text{MC}}(P, \varepsilon) + O(\log n).$$  

The proof of Corollary 6 is omitted since it follows similarly to the proof of [3, Theorem 3]. We remark that Corollary 6 implies that the dispersion for our lossy multi-connectivity problem is zero, which is consistent with existing literatures [3], [16]. Furthermore, our result in Corollary 6 justifies that the asymptotic notion of outage capacity, used extensively in wireless communication systems in case of quasi-static fading, is actually an accurate performance criterion even in the finite blocklength setting.

We remark that the results in Theorem 5 and Corollary 6 can also be established for Gaussian memoryless sources under quadratic distortion measures by using our non-asymptotic converse bounds. The converse proof remains the same and the achievability part can be done by adapting the universal coding scheme in [26] to the lossy multi-connectivity setting considered here.
Fig. 2: Power allocation: excess-distortion probability for lossy transmission over $\Psi = 2$ parallel AWGN channels with independent quasi-static Rayleigh fading for different power allocations $\alpha = [\alpha_1 P, \alpha_2 P]$ with power constraint $P = 5$.

D. Numerical Simulation

In this section, we illustrate our results in Corollary 3 and Theorem 4 via numerical simulations. We consider Rayleigh fading with scale parameter one, i.e., each fading parameter follows the same Rayleigh distribution $P_\mathcal{A}(a_t) = a_t \exp \left( -\frac{a_t^2}{2} \right)$ for any $a_t \in \mathbb{R}_+$. We are interested in the finite blocklength performance of the lossy transmission of the following two memoryless sources over parallel AWGN channels with independent quasi-static Rayleigh fading:

- Binary memoryless source (BMS) with bias $p$ under Hamming distortion measure. The source alphabet is $\mathcal{S} = \{0, 1\}$ and the source distribution is $P_S(0) = p$ and $P_S(1) = 1 - p$. For any two sequences $s^k$ and $\hat{s}^k$, the Hamming distortion is defined as $d(s^k, \hat{s}^k) = \frac{1}{k} \sum_{i \in [k]} 1\{s_i \neq \hat{s}_i\}$.

- GMS under the quadratic distortion measure. The source distribution is $\mathcal{N}(0, \sigma_S^2)$ and the quadratic distortion measure for any two sequences $s^k$ and $\hat{s}^k$ is defined as $d(s^k, \hat{s}^k) = \frac{1}{k} \sum_{i \in [k]} (s_i - \hat{s}_i)^2$.

In the following, we assume a maximal power constraint of $P = 5$, a fixed rate of source symbols per channel use of $k/n = 1$ and up to three parallel AWGN channels, i.e., $\Psi \leq 3$. 
In Figure 2, we plot the non-asymptotic bounds in Corollary 3 (dashed lines) and Theorem 4 (solid lines) via Monte-Carlo simulations for the case of $\Psi = 2$ and different power allocations $P = [P_1, P_2] = [\alpha_1, \alpha_2]P$ where $(\alpha_1, \alpha_2) \in \mathbb{R}_+$ satisfying $\alpha_1 + \alpha_2 \leq 1$. It is shown that the equal power allocation (i.e., $\alpha_1 = \alpha_2 = 0.5$) achieves the smallest excess-distortion probabilities in both the achievability and converse parts. We have verified that this is true for any possible values of $(\alpha_1, \alpha_2) \in \mathbb{R}_+$ such that $\alpha_1 + \alpha_2 \leq 1$ and thus conclude that the equal power allocation is optimal in the finite block length regime. Hence, in further numerical simulations, we plot our results for the equal power allocation only.

In Figure 3, for the case of $\Psi \in [3]$, we plot the non-asymptotic achievability bound in Theorem 4 (solid lines), the second-order asymptotics in Theorem 5 (dotted lines), and outage probability in (31) (dash-dotted lines) versus the blocklength. It can be observed that the achievable excess-distortion probability decreases significantly as the number of parallel channels increases and thus demonstrates the benefits of multi-connectivity in the finite blocklength setting. The same result is true for the converse result (see Corollary 3).

1The simulations included a great number of power allocations, which are omitted in Fig. 2 for clarity.
IV. PROOF OF SECOND-ORDER ASYMPTOTICS

A. Preliminaries

Using the definitions in (22), and (24) to (26), define the average dispersion function and the third-
absolute moment as

\[ V^{(P,A)}_{k+n\Psi} := \frac{kV_s(P_S,D) + nU_2(P,A)}{k + n\Psi}, \]  
\[ T^{(P,A)}_{k+n\Psi} := \frac{kT(P_S,D) + nU_3(P,A)}{k + n\Psi}. \]  

Furthermore, define the following three quantities

\[ \Omega_1(k, n, P) := \sup_{P \in P_{eq}(P)} \mathbb{E}_A \left[ 6T^{(P,A)}_{k+n\Psi} \left( \sqrt{V^{(P,A)}_{k+n\Psi}} \right)^{-3} \right], \]  
\[ \Omega_2(k, n, P) := \sup_{P \in P_{eq}(P)} \mathbb{E}_A \left[ \left( \sqrt{V^{(P,A)}_{k+n\Psi}} \right)^{-1} \right], \]  
\[ \Omega_3(k, n, P) := \sup_{P \in P_{eq}(P)} \mathbb{E}_A \left[ U_1(P,A) \left( \sqrt{V^{(P,A)}_{k+n\Psi}} \right)^{-1} \right]. \]

B. Converse Proof

In this section, we present the converse proof of our main results. Similar to \( P^{*,e,k,n}_{e,k,n}(P,D) \) in (7), we let \( P^{*,eq}_{e,k,n}(P,D) \) denote the corresponding minimal excess-distortion probability for any \((k, n, P, D)\)-code with equal power constraint. It can be easily shown that

\[ P^{*,e,k,n}_{e,k,n}(P,D) \geq P^{*,eq}_{e,k,n+1}(P,D) \]  

since for any \((k, n, P, D)\)-code with strict inequality in (5), we can always construct a \((k, n+1, P, D)\)-code with equality in (5). In this section, we will first derive a lower bound on \( P^{*,eq}_{e,k,n}(P,D) \) and then use (40) to establish a lower bound for \( P^{*,e,k,n}_{e,k,n}(P,D) \). Furthermore, the results in Theorem 1 and Corollary 3 hold also with equal power constraint as can be gleaned in their proofs.

For ease of notation, let

\[ L(P_t, A_t, Z^n_t) := \frac{\sum_{j \in [n]} P_t A^2_t (1 - Z^2_{t,j}) + 2A_t \sqrt{P_t} Z_{t,j}}{2(1 + P_t A^2_t)}. \]
In the following, we first bound the dominant term in Corollary 3. Using the definitions of $L^n_t(P_t, A_t, Z^n_t)$ in (16) and $j(S^k, D)$ in (12), for any power allocation vector $P \in \mathcal{P}_{eq}(P)$ and $\gamma > 0$, we have that

$$\Pr \left\{ \sum_{t \in [\Psi]} L^n_t(P_t, A_t, Z^n_t) \leq j(S^k, D) - n\gamma \right\}$$

$$= \Pr \left\{ \sum_{t \in [\Psi]} \left( \frac{n}{2} \log(1 + P_t A^2_t) + L(P_t, A_t, Z^n_t) \right) \leq \sum_{t \in [\Psi]} j(S_t, D) - n\gamma \right\}$$

$$= \Pr \left\{ \sum_{t \in [\Psi]} (j(S_t, D) - R(P_S, D)) - \sum_{t \in [\Psi]} L(P_t, A_t, Z^n_t) \geq \sum_{t \in [\Psi]} \frac{n}{2} \log(1 + P_t A^2_t) - kR(P_S, D) + n\gamma \right\}. \quad (42)$$

Using the fact that $\mathbb{E}[j(S_t, D)] = R(P_S, D)$ [12, Property 1], we conclude that $\{j(S_t, D) - R(P_S, D)\}_{t \in [k]}$ and $\{P_t A^2_t(1 - Z^2_{t,i} + 2A_t \sqrt{P_t Z_{t,i}})\}_{t \in [\Psi], i \in [n]}$ forms a sequence of $(k + n \Psi)$ independent random variables with zero mean. Applying the Berry-Esseen theorem [31], [32] (see also [33, Chapter 1]) to the result in (43) similarly to [16], we obtain that

$$\mathbb{E}_A \left[ \Pr \left\{ \sum_{t \in [\Psi]} L^n_t(P_t, A_t, Z^n_t) \leq j(S^k, D) - n\gamma \right\} - Q \left( \frac{\sum_{t \in [\Psi]} \frac{n}{2} \log(1 + P_t A^2_t) - kR(P_S, D) + n\gamma}{\sqrt{(k + n \Psi) V^{(P,A)}_{k+n\Psi}}} \right) \right]$$

$$\leq \mathbb{E}_A \left[ \frac{6T_{k+n\Psi}^{(P,A)} (V^{(P,A)}_{k+n\Psi})^{-3/2}}{\sqrt{(k + n \Psi)}} \right]. \quad (44)$$

Now, using Corollary 3 and (44), by choosing $\gamma = \frac{\log n}{n}$, we conclude that

$$\mathbb{P}_{e,k,n}^{*,eq}(P, D)$$

$$\geq \inf_{P \in \mathcal{P}_{eq}(P)} \left\{ \mathbb{E}_A \left[ Q \left( \frac{nU_1(P,A) - kR(P_S, D) + \log n}{\sqrt{(k + n \Psi) V^{(P,A)}_{k+n\Psi}}} \right) - \frac{6T_{k+n\Psi}^{(P,A)} (V^{(P,A)}_{k+n\Psi})^{-3/2}}{\sqrt{(k + n \Psi)}} \right] - \frac{1}{n} \right\}. \quad (45)$$

$$\geq \inf_{P \in \mathcal{P}_{eq}(P)} \mathbb{E}_A \left[ Q \left( \frac{nU_1(P,A) - kR(P_S, D) + \log n}{\sqrt{(k + n \Psi) V^{(P,A)}_{k+n\Psi}}} \right) - \sup_{P \in \mathcal{P}_{eq}(P)} \mathbb{E}_A \left[ \frac{6T_{k+n\Psi}^{(P,A)} (V^{(P,A)}_{k+n\Psi})^{-3/2}}{\sqrt{(k + n \Psi)}} \right] - \frac{1}{n} \right]. \quad (46)$$

$$= \inf_{P \in \mathcal{P}_{eq}(P)} \mathbb{E}_A \left[ Q \left( \frac{nU_1(P,A) - kR(P_S, D) + \log n}{\sqrt{(k + n \Psi) V^{(P,A)}_{k+n\Psi}}} \right) - \frac{1}{n} \right] \quad (47)$$

$$\geq \inf_{P \in \mathcal{P}_{eq}(P)} \mathbb{E}_A \left[ Q \left( \frac{nU_1(P,A) - kR(P_S, D)}{\sqrt{(k + n \Psi) V^{(P,A)}_{k+n\Psi}}} \right) - \frac{\log n}{\sqrt{(k + n \Psi) V^{(P,A)}_{k+n\Psi}}} \right] - \frac{1}{n} \quad (48)$$

$$= \inf_{P \in \mathcal{P}_{eq}(P)} \mathbb{E}_A \left[ Q \left( \frac{nU_1(P,A) - kR(P_S, D)}{\sqrt{(k + n \Psi) V^{(P,A)}_{k+n\Psi}}} \right) - \frac{\Omega_2(k,n,P) \log n}{\sqrt{(k + n \Psi)}} - \frac{\Omega_1(k,n,P)}{\sqrt{(k + n \Psi)}} - \frac{1}{n} \right]. \quad (49)$$
\[
\inf_{P \in \mathcal{P}_{eq}(P)} \mathbb{E}_A \left[ Q \left( \frac{nU_1(P, A) - kR(P_S, D)}{\sqrt{(k + n\Psi) V^{(P, A)}_{k+n\Psi}}} \right) \right] + O\left( \frac{\log n}{\sqrt{n}} \right),
\]

(50)

where (47) follows from the definition of \( \Omega_1(k, n, P) \) in (37); (48) follows from [13, Eq. (255)] which states that \( Q(x + a) \geq Q(x) - \frac{n}{\sqrt{2\pi}} \geq Q(x) - a \) for any \( x \) and any \( a \geq 0 \); (49) follows from the definition of \( \Omega_2(k, n, P) \) in (38); and (50) follows from the conditions of Theorem 5.

Using the results in (40) and (50), we obtain that

\[
P^*_{e,k,n}(P, D) \geq P^*_{e,k,n+1}(P, D)
\]

(51)

\[
\geq \inf_{P \in \mathcal{P}_{eq}(P)} \mathbb{E}_A \left[ Q \left( \frac{(n+1)U_1(P, A) - kR(P_S, D)}{\sqrt{(k + (n+1)\Psi) V^{(P, A)}_{k+(n+1)\Psi}}} \right) \right] + O\left( \frac{\log n}{\sqrt{n}} \right)
\]

(52)

\[
\geq \inf_{P \in \mathcal{P}_{eq}(P)} \mathbb{E}_A \left[ Q \left( \frac{nU_1(P, A) - kR(P_S, D)}{\sqrt{(k + n\Psi) V^{(P, A)}_{k+n\Psi}}} \right) \right] - \frac{U_1(P, A)}{\sqrt{(k + n\Psi) V^{(P, A)}_{k+n\Psi}}} + O\left( \frac{\log n}{\sqrt{n}} \right)
\]

(53)

\[
\geq \inf_{P \in \mathcal{P}_{eq}(P)} \mathbb{E}_A \left[ Q \left( \frac{nU_1(P, A) - kR(P_S, D)}{\sqrt{(k + n\Psi) V^{(P, A)}_{k+n\Psi}}} \right) \right] - \frac{\Omega_3(k, n, P)}{\sqrt{k + n\Psi}} + O\left( \frac{\log n}{\sqrt{n}} \right)
\]

(54)

\[
= \inf_{P \in \mathcal{P}_{eq}(P)} \mathbb{E}_A \left[ Q \left( \frac{nU_1(P, A) - kR(P_S, D)}{\sqrt{(k + n\Psi) V^{(P, A)}_{k+n\Psi}}} \right) \right] + O\left( \frac{\log n}{\sqrt{n}} \right)
\]

(55)

where (53) follows from [13, Lemma 4]; (54) follows from the definition of \( \Omega_3(k, n, P) \) in (39); and (55) follows since \( \frac{\Omega_3(k, n, P)}{\sqrt{k+n\Psi}} = O\left( \frac{\log n}{\sqrt{n}} \right) \) similarly to (50).

C. Achievability Proof

In this subsection, we present the achievability proof of Theorem 5 for any DMS under a bounded distortion measure. For simplicity, in the following, we use \( Z \) to denote \( \{Z_i^n\}_{i \in [\Psi]} \) and similarly \( z \).

In the achievability proof, the following lemma [12, Lemma 2] is important. Recall the definitions of \( \Phi(s^k, D) \) in (19) and \( j(s, D) \) in (10).

**Lemma 7.** For any DMS under any bounded distortion measure, there exists constants \( c_1, c_2, c_3 \) such that

\[
\Pr \left\{ \log \frac{1}{\Phi(S^k, D)} \leq \sum_{i \in [k]} j(S_i, D) + (c_1 - 0.5) \log k + c_2 \right\} \geq 1 - \frac{c_3}{\sqrt{k}}.
\]

(56)
Recall the definition of $\tilde{L}_t^n(\cdot)$ in (20). For simplicity, let

$$\Upsilon_1(P, A, Z, S^k, D) := \sum_{i \in [\Psi]} \tilde{L}_t^n(P_i, A_i, Z_i^n) - \log n - \log \gamma - \log \frac{1}{\Phi(S^k, D)}, \quad (57)$$

$$\Upsilon_2(S^k, D) := \sum_{i \in [k]} j(S_i, D) + (c_1 - 0.5) \log k + c_2, \quad (58)$$

$$\Upsilon(P, A, Z, S^k, D) := \Upsilon_1(P, A, Z, S^k, D) + \log \frac{1}{\Phi(S^k, D)} - \Upsilon_2(S^k, D). \quad (59)$$

Furthermore, for any power vector $P$, define the set:

$$\mathcal{T}_{k,n}(P) := \left\{ (a,z,s^k) : \Upsilon(P, a, z, s^k, D) \geq \log n \right\}. \quad (60)$$

Using the result in Lemma 7, and weakening the result in Theorem 4 by letting $\gamma = \frac{\log n + 1}{n}$, we obtain that there exists an $(k, n, P, D)$-code such that for any $P \in \mathcal{P}_{\max}(P)$,

$$P_{e,k,n}(P, D) \leq \mathbb{E} \left[ \exp \left( - \left| \Upsilon_1(P, A, Z, S^k, D) \right|^+ \right) \right] + \frac{1}{n}$$

$$= \mathbb{E} \left[ \exp \left( - \left| \Upsilon_1(P, A, Z, S^k, D) \right|^+ \right) \times 1 \left\{ \log \frac{1}{\Phi(S^k, D)} \leq \Upsilon_2(S^k, D) \right\} \right]$$

$$+ \mathbb{E} \left[ \exp \left( - \left| \Upsilon_1(P, A, Z, S^k, D) \right|^+ \right) \times 1 \left\{ \log \frac{1}{\Phi(S^k, D)} > \Upsilon_2(S^k, D) \right\} \right] + \frac{1}{n} \quad (62)$$

$$\leq \mathbb{E} \left[ \exp \left( - \left| \Upsilon(P, A, Z, S^k, D) \right|^+ \right) \right] \times 1 \left\{ \log \frac{1}{\Phi(S^k, D)} \leq \Upsilon_2(S^k, D) \right\} \right]$$

$$+ \Pr \left\{ \log \frac{1}{\Phi(S^k, D)} > \Upsilon_2(S^k, D) \right\} + \frac{1}{n} \quad (63)$$

$$\leq \frac{1}{n} \Pr \{ (A, Z, S^k) \in \mathcal{T}_{k,n}(P) \} + \Pr \{ (A, Z, S^k) \not\in \mathcal{T}_{k,n}(P) \} + \frac{1}{n} + \frac{c_3}{\sqrt{k}} \quad (64)$$

$$\leq \frac{2}{n} + \frac{c_3}{\sqrt{k}} + \Pr \{ (A, Z, S^k) \not\in \mathcal{T}_{k,n}(P) \}, \quad (66)$$

where (63) follows since i) $\Upsilon(P, A, Z, S^k, D) \leq \Upsilon_1(P, A, Z, S^k, D)$ if $\log \frac{1}{\Phi(S^k, D)} \leq \Upsilon_2(S^k, D)$ and $\exp(-|x|^+)$ is non-increasing in $x$, and ii) $\exp(-|x|^+) \leq 1$ for any $x$ (recall that $|x|^+ = \max\{0, x\}$); (64) follows from Lemma 7; and (65) follows since $\mathbb{E}[\exp(-|X|^+)] \leq \frac{1}{a} \Pr \{ X \geq \log a \} + \Pr \{ X \leq \log a \}$ for any $a \geq 1$. 

In the remaining part of this subsection, we will upper bound the third term in (66). For simplicity, define the following function of power allocation vector $P$ and the fading vector $A$,

$$B_1(P, A) := \sum_{i \in [\Psi]} \log \frac{1 + P_i A_i^2}{\sqrt{1 + 2P_i A_i^2}} + (c_1 - 0.5) \log k + c_2 + 2 \log n + \log \gamma. \quad (67)$$

Then, using the definitions of $U_i(P, A)$ for $i \in [3]$ in (24) to (26) and applying the Berry-Esseen theorem similarly to (44), we have that

$$\Pr \left\{ (A, Z, S^k) \notin T_{k,n}(P) \right\} = \inf_{P \in \mathcal{P}_{\text{max}}(P)} \Pr \left\{ \sum_{i \in [k]} \frac{1}{\sqrt{2P_i A_i^2}} \right\} \leq \inf_{P \in \mathcal{P}_{\text{eq}}(P)} \Pr \left\{ \left( \sum_{i \in [k]} (S_i, D) - R(P_S, D) \right) \right\} \leq \inf_{P \in \mathcal{P}_{\text{eq}}(P)} \frac{\Pr \left\{ \sum_{i \in [k]} (S_i, D) - R(P_S, D) \right\} + \frac{1}{\sqrt{2P_i A_i^2}}}{\sqrt{k + n\Psi}} \leq \inf_{P \in \mathcal{P}_{\text{eq}}(P)} \frac{\Omega_1(k, n, P)}{\sqrt{k + n\Psi}} + \frac{\Omega_2(k, n, P) + \Omega_3(k, n, P) + c_1 \log k + c_2}{\sqrt{k + n\Psi}} \leq \inf_{P \in \mathcal{P}_{\text{eq}}(P)} \frac{\Omega_4(k, n, P) + \Omega_5(k, n, P) + c_1 \log k + c_2 + 2 \log n + \log \log(n)}{\sqrt{k + n\Psi}}, \quad (70)$$

where (72) follows from [13, Lemma 4] and the definition of $\Omega_4(k, n, P)$ in (37); and (73) follows form the definitions of $B_1(P, A)$ in (82) and $\Omega_5(k, n, P)$ in (39) and the fact that $\log \frac{1 + P_i A_i^2}{\sqrt{1 + 2P_i A_i^2}} \leq \frac{1}{2} \log(1 + P_t A_t^2)$ for any $t \in [\Psi]$.

For simplicity, let

$$\Omega_5(k, n, P) := \frac{2}{n} + \frac{c_3}{\sqrt{k}} + \frac{\Omega_1(k, n, P) + \Omega_3(k, n, P)}{\sqrt{k + n\Psi}} + \frac{c_1 \log k + c_2 + 2 \log n + \log \log(n)}{\sqrt{k + n\Psi}}. \quad (74)$$

Similarly to (50), we can verify that $\Omega_5(k, n, P) = O\left(\frac{\log n}{\sqrt{n}}\right)$.
D. Universal Achievability Coding Scheme

In this subsection, we present a universal JSCC scheme, which can achieve the same second-order asymptotics as our non-universal counterpart presented in Section IV-C. The JSCC scheme considered here dates back to Csiszár [34] who proposed the JSCC scheme consisting of unequal error protection and method of types to derive the error exponent for the JSCC problem. We adapted the idea to our setting by using a modified minimum distance decoder, which makes use of only the received codewords, the CSI and the knowledge of the source and channel codebooks.

1) Notation for Method of Types: Given a length-$k$ sequence $S^k$, we use $\hat{P}_{S^k}$ to denote the type (empirical distribution). The collection of all types on $S^k$ is denoted as $\mathcal{P}_k(S)$. For any type $Q \in \mathcal{P}_k(S)$, we use $\mathcal{T}_Q$ to denote the type class, i.e., all the length-$k$ sequences such that $\hat{P}_{S^k} = Q$. For simplicity, we assume that the types in $\mathcal{P}_k(S)$ are ordered in some sense, i.e., $\mathcal{P}_k(S) = \{Q_i\}_{i \in \mathcal{I}(\mathcal{P}_k(S))}$ and each index $i \in \mathcal{I}(\mathcal{P}_k(S))$ is associated with a type $Q_i$. It is known that $|\mathcal{P}_k(S)| \leq (k+1)^{|S|}$.

2) JSCC Scheme: Let $\{M_i\}_{i \in \mathcal{I}(\mathcal{P}_k(S))}$ be a sequence of integers to be determined. Furthermore, define the set

$$D := \{(r_1, r_2) \in \mathbb{R}_+^2 : r_1 \in \mathcal{I}(\mathcal{P}_k(S)), r_2 \in [M_{r_1}]\}.$$  

Codebook generation: For each $i \in \mathcal{I}(\mathcal{P}_k(S))$, generate $M_i$ independent source codewords $\hat{S}_i := \{\hat{S}^k(i, 1), \ldots, \hat{S}^k(i, M_i)\}$, each i.i.d. according to $P^*_S$. Furthermore, for each $i \in \mathcal{I}(\mathcal{P}_k(S))$ and $t \in [\Psi]$, generate $M_i$ independent channel codewords $\{X^n_t(i, 1), \ldots, X^n_t(i, M_i)\}$, each according to the uniform distribution over a sphere with radius $\sqrt{n}P_t$.

Encoding: Given $S^k$, for each $t \in [\Psi]$, the encoder $f_t$ transmits $X^n_t(I, J)$ if the type index of the source sequence is $I$ (i.e., $S^k \in \mathcal{T}_Q$) and the index minimizing the distortion between the source sequence and the source codewords in the $i$-th subcodebook $S_i$ is $J$, i.e.,

$$J = \arg\min_{j \in [1:M_i]} d(S^k, \hat{S}^k(I, j)).$$  

(76)
Decoding: Fix a threshold $\gamma_{k,n}$ to be specified. Given the channel outputs $\{Y^n_t\}_{t \in [\Psi]}$, the decoder $\phi$ outputs source estimate $\hat{S}^k(\hat{I}, \hat{J})$ if $(\hat{I}, \hat{J})$ are the unique pair such that

$$\sum_{t \in [\Psi]} \hat{i}(X^n_t(\hat{I}, \hat{J}); Y^n_t | A_t) - \log M_{\hat{i}} > \log \gamma_{k,n}. \quad (77)$$

Otherwise, the decoder declares an error.

Recalling the definition of $\hat{i}(X^n_t; Y^n_t | A_t)$ in (14), we obtain that

$$\sum_{t \in [\Psi]} \hat{i}(X^n_t(\hat{I}, \hat{J}); Y^n_t | A_t) = -\frac{\sum_{t \in [\Psi]} \|Y^n_t - A_t X^n_t(\hat{I}, \hat{J})\|^2}{2} + \sum_{t \in [\Psi]} \left( \frac{n}{2} \log(1 + P_t A_t^2) + \frac{\|Y^n_t\|^2}{2(1 + P_t A_t^2)} \right). \quad (78)$$

Hence, our decoder in (77) is a universal decoder which requires knowledge of only the channel outputs $\{Y^n_t\}_{t \in [\Psi]}$, the fading parameters $\{A_t\}_{t \in [\Psi]}$, the channel codebook $\{X^n_t(i,j)\}_{(i,j) \in \mathcal{D}}$ for each $t \in [\Psi]$ and the integers $\{M_i\}_{i \in [\mathcal{P}_k(S)]]}$.

3) Analysis of Excess-Distortion Probability: Given our coding scheme, following the analyses in [25], [26], we can upper bound the excess-distortion probability by

$$P_{e,k,n}(P, D) \leq \sum_{i \in [\mathcal{P}_k(S)]]} \Pr\{S^k \in \mathcal{T}_{Q_i}, \min_{\tilde{j} \in [M_i]} d(S^k, \hat{S}^k(i, \tilde{j})) > D\} \sum_{i \in [\mathcal{P}_k(S)]]} \Pr\{S^k \in \mathcal{T}_{Q_i}, (\hat{I}, \hat{J}) \neq (i, J)\}. \quad (79)$$

For each $i \in [\mathcal{P}_k(S)]]$, let

$$\log M_i := kR(Q_i, D) + C \log k, \quad (80)$$

where $C = 4|S||\hat{S}| + 9$ (see [35, Eq. (74)]). With this choice of $M_i$, invoking the type covering lemma for the lossy source coding problem [19], [21], [36], we conclude that for $n$ sufficiently large and any $i \in [\mathcal{P}_k(S)]]$,

$$\Pr\{S^k \in \mathcal{T}_{Q_i}, \min_{\tilde{j} \in [M_i]} d(S^k, \hat{S}^k(i, \tilde{j})) > D\} = 0. \quad (81)$$

For simplicity, given $s^k \in \mathcal{T}_{Q_i}$ and the $i$-th subcodebook $\hat{s}_i$, we use $j(s^k, \hat{s}_i)$ to denote the index of the codeword in $s_i$ which minimizes the distortion with respect to $s^k$. In the following lemma, we upper
bound the probability of decoding error in \((i, j(s^k, \hat{s}_i))\) wrongly conditioning on \(s^k\) and \(\hat{s}_i\). To present the lemma, define

\[
B_2(P, A) := \sum_{t \in [Q]} \frac{1 + P_i A_t^2}{\sqrt{1 + 2P_i A_t^2}} \left( \frac{2(1 + P_i A_t^2)}{\sqrt{\pi} \sqrt{P_i A_t^2 (P_i A_t^2 + 2)}} + \frac{12 \sqrt{2} P_i A_t^2 (7P_i A_t^2 + 12)}{(P_i A_t^2 + 2) \sqrt{P_i A_t^2 (P_i A_t^2 + 2)}} \right). \quad (82)
\]

It can be verified that under the conditions of Theorem 5, we have that there exists a constant \(c_6\) such that

\[
\sup_{P \in \mathcal{P}_{eq}(P)} \mathbb{E}[B_2(P, A)] \leq c_6. \quad (83)
\]

**Lemma 8.** For any \(i \in [|\mathcal{P}_k(S)|]\), given \(s^k \in T_Q\), and \(\hat{s}_i\), we obtain that

\[
\Pr\{(\hat{i}, \hat{j}) \neq (i, j(s^k, \hat{s}_i))|S^k = s^k, \hat{S}_i = \hat{s}_i\} \\ \leq \Pr\left\{ \sum_{t \in [Q]} \tilde{f}(X^n_t; Y^n_t|A_t) \leq \log \gamma_{k,n} + \log M_i \right\} + \frac{|\mathcal{P}_k(S)| \mathbb{E}_A[B_2(P, A)]}{\gamma_{k,n} \sqrt{n}}. \quad (84)
\]

The proof of Lemma 8 is inspired by the dependence testing bound \([10, \text{Theorem 18}]\) (see also \([16, \text{Theorem 1}]\)) and the analysis of its variant to quasi-static fading channel in \([24, \text{Chapter 4}]\). The proof of Lemma 8 is available in Appendix C for completeness.

Using Lemma 8, choosing \(\gamma_{k,n} = |\mathcal{P}_k(S)| \leq (k + 1)^{|S|}\) and applying the Berry-Esseen theorem, we obtain that for any \(P \in \mathcal{P}_{eq}(P)\),

\[
\sum_{i \in [|\mathcal{P}_k(S)|]} \Pr\{S^k \in T_{Q_k}, (\hat{i}, \hat{j}) \neq (i, J)\} \\ = \sum_{i \in [|\mathcal{P}_k(S)|]} \sum_{s^k \in T_{Q_k}} \mathbb{E}_{S|S}(s^k|S^k) \Pr\{(\hat{i}, \hat{j}) \neq (i, j(s^k, \hat{s}_i))|S^k = s^k, \hat{S}_i = \hat{s}_i\} \quad (85)
\]

\[
\leq \sum_{i \in [|\mathcal{P}_k(S)|]} \sum_{s^k \in T_{Q_k}} \mathbb{E}_{S|S}(s^k) \Pr\left\{ \sum_{t \in [Q]} \tilde{f}(X^n_t; Y^n_t|A_t) \leq \log \gamma_{k,n} + \log M_i \right\} + \frac{|\mathcal{P}_k(S)| \mathbb{E}_A[B_2(P, A)]}{\gamma_{k,n} \sqrt{n}} \quad (86)
\]

\[
= \Pr\left\{ \sum_{t \in [Q]} \tilde{f}(X^n_t; Y^n_t|A_t) \leq \log \gamma_{k,n} + nR(\hat{P}_{\theta_k}, D) + C \log k \right\} + \frac{|\mathcal{P}_k(S)| \mathbb{E}_A[B_2(P, A)]}{\gamma_{k,n} \sqrt{n}} \quad (87)
\]

\[
\leq \Pr\left\{ \sum_{t \in [Q]} \tilde{f}(X^n_t; Y^n_t|A_t) \leq |S| \log(k + 1) + nR(\hat{P}_{\theta_k}, D) + C \log k \right\} \quad (88)
\]

\[
= \Pr\left\{ S^k \in T^k, \sum_{t \in [Q]} \tilde{f}(X^n_t; Y^n_t|A_t) \leq |S| \log(k + 1) + kR(\hat{P}_{\theta_k}, D) + C \log k \right\} + \Pr\{S^k \notin T^k\} \quad (89)
\]
where (87) follows from the choice of $M_k$ in (83), the assumption that $\Omega_4$ and the definitions of (43); (93) follows by applying the Berry-Esseen theorem as in (44); (94) follows from [13, Lemma P. A. Proof of Lemma 2

(see [35, Eq. (91)]) and the result in Lemma 2; (91) follows from [37, Lemma 22]; (92) is similar to (43); (93) follows by applying the Berry-Esseen theorem as in (44); (94) follows from [13, Lemma 4] and the definitions of $\Omega_1(k, n, P)$ in (37) and $\Omega_2(k, n, P)$ in (38); and (95) follows from the result in (83), the assumption that $k = \Theta(n)$ and similar arguments leading to (50).

APPENDIX

A. Proof of Lemma 2

Recall the distributions $P_{Y_i^n | X_i^n A_t}$ in (13) and $Q_{Y_i^n | A_t}$ in (17). Given any $t \in [\Psi]$, channel input $x_t^n \in \mathbb{R}(n, P_t)$ (see (2)) and fading parameter $a_t$, using the definition of $\bar{i}(X_i^n; Y_i^n | A_t)$ and the fact $Z_i^n =$
Due to the channel law (see (1)), under the distribution of $P_{Y^n_t|x^n_t,A_t}(\cdot|x^n_t,a_t)$, we have

$$
\tilde{i}(x^n_t;Y^n_t|a_t) = \log \left( (1 + P_t a_t^2)^{\frac{n}{2}} \exp \left\{ - \frac{||Y^n_t - a_t x^n_t||^2}{2} + \frac{||Y^n_t||^2}{2(1 + P_t a_t^2)} \right\} \right)
$$

(96)

$$
= \frac{n}{2} \log(1 + P_t a_t^2) - \frac{||Z^n_t||^2}{2} + \frac{||Z^n_t + a_t x^n_t||^2}{2(1 + P_t a_t^2)}.
$$

(97)

Hence, the distribution of $\tilde{i}(x^n_t;Y^n_t|a_t)$ depends on $x^n_t$ only through $||Z^n_t + a_t x^n_t||^2$. Noting that

$$
||Z^n_t + a_t x^n_t||^2 = ||Z^n_t||^2 + a_t^2 ||x^n_t||^2 + 2 \sum_{j \in [n]} a_t x_{t,j} Z_{t,j}
$$

(98)

$$
= ||Z^n_t||^2 + nP_t a_t^2 + 2 \sum_{j \in [n]} a_t x_{t,j} Z_{t,j},
$$

we conclude that the distribution of $\tilde{i}(x^n_t;Y^n_t|a_t)$ depends on $x^n_t$ only through $\sum_{j \in [n]} A_t x_{t,j} Z_{t,j}$. However, since $Z^n_t$ is i.i.d. according to $\mathcal{N}(0,1)$, we conclude that

$$
\sum_{j \in [n]} a_t x_{t,j} Z_{t,j} \sim \mathcal{N}(0, nP_t a_t^2),
$$

(100)

independent of the choice of $x^n_t \in \mathbb{R}(n,P_t)$. Therefore, we conclude that the distribution of $\tilde{i}(x^n_t;Y^n_t|a_t)$ depends on $x^n_t$ only through its power $P_t$ under $P_{Y^n_t|x^n_t,A_t}(\cdot|x^n_t,a_t)$. It remains to show that $\tilde{i}(x^n_t;Y^n_t|a_t)$ is distributed as $L^n_t(P_t, A_t, Z^n_t)$. We can now choose $x^n_t = x^n_0 = (\sqrt{P_t}, \ldots, \sqrt{P_t})$ for simplicity. With this choice, under the distribution of $P_{Y^n_t|x^n_t,A_t}(\cdot|x^n_t,a_t)$, using the result in (97) and the definition of $L^n_t(P_t, A_t, Z^n_t)$ in (16), we obtain that $\tilde{i}(x^n_t;Y^n_t|a_t)$ for any $x^n_t \in \mathbb{R}(n,P_t)$ and any fading parameter $a_t$ has the same distribution as

$$
\tilde{i}(x^n_0;Y^n_0|A_t) = \frac{n}{2} \log(1 + P_t a_t^2) - \frac{||Z^n_0||^2}{2} + \frac{||Z^n_0 + a_t x^n_0||^2}{2(1 + P_t a_t^2)}
$$

(101)

$$
= L^n_t(P_t, a_t, Z^n_t).
$$

(102)

**B. Proof of Theorem 4**

Since the proof of Theorem 4 is similar to [13], we only emphasize the differences here. Fix an arbitrary real number $\gamma > 0$. Given any source sequence $s^k$, recalling the definition of $\Phi(s^k,D)$ in (19), define

$$
W(s^k, D) := \left[ \frac{n \gamma}{\Phi(s^k,D)} \right].
$$

(103)
Codebook generation: For each \( i \in [M] \), generate a random sequence \( \hat{S}^k(i) \) i.i.d. according to \( P_{\hat{S}}^* \). The collection of these \( M \) recovery sequences form the random source codebook. Furthermore, for each \( t \in [\Psi] \) and each \( i \in [M] \), generate a channel codeword \( X_t^n(i) \) according to a uniform distribution over the sphere with radius \( \sqrt{nP_t} \), i.e.,

\[
P_{X_t^n}(x_t^n) = \frac{1\{x_t^n ||x_t^n||^2 - nP_t\}}{A_n(\sqrt{n}P_t)},
\]

where \( 1\{\cdot\} \) is indicator function and \( A_n(r) \) is the surface area of an \( n \)-dimensional sphere with radius \( r \). The collection of the \( n\Psi \) channel codewords form the channel codebook. We assume that the source and channel codebooks are known by both the encoder and the decoder. For simplicity, we will use \( \hat{S} \) to denote the random source codebook and \( \hat{s} \) a particular realization. Similarly we will use \( X \) and \( x \) to denote the channel codebooks. Recall that we use \( A \) to denote the fading parameters. We will also use \( a \) to denote a particular realization of \( A \).

Encoding: Fix an integer \( M \) s.t. \( W(s^k, D) \leq M \) for all \( s^k \). Given a source sequence \( S^k \), for each \( t \in [\Psi] \), the encoder \( f_t \) outputs the channel codeword \( X_t^n(i(j(S^k))) \) where \( j(S^k) := \min\{m, W(S^k, D)\} \) if \( d(S^k, \hat{S}^k(m)) \leq D < \min_{i \in [m-1]} d(S^k, \hat{S}^k(i)) \) and \( j(S^k) = M \) otherwise. The index \( j(S^k) \) can be essentially understood as the minimum of \( W(S^k, D) \) and the index of the source codeword which minimizes the distortion with respect to \( S^k \).

Decoding: Prior to describing the decoding rule, define the random variable \( U \) taking values in \([M+1]\) as follows:

\[
U := \begin{cases} 
  j(S^k) & \text{if } d(S^k, \hat{S}^k(j(S^k))) \leq D \\
  M + 1 & \text{otherwise.}
\end{cases}
\]

(105)

Given channel outputs \( \{Y_t^n\}_{t \in [\Psi]} \) and fading parameters \( \{A_t\}_{t \in [\Psi]} \), the decoder declares \( \hat{S}^k(\hat{J}) \) as the source estimate if

\[
\hat{J} = \arg\max_{\hat{j} \in [M]} P_{U|S}(\hat{j}|\hat{S}) \prod_{t=1}^{\Psi} P_{Y_t^n|X_t^nA_t}(Y_t^n|X_t^n(\hat{j}), A_t).
\]

(106)

Note that the decoder in (106) is similar to a MAP decoder with the only difference that \( P_{U|S}(\hat{j}|\hat{S}) \) should be replaced by \( \prod_{t \in [\Psi]} P_{X_t^n}(X_t^n(\hat{j})) \).
Conditioning on the fading parameters $a$, the source codebook $\hat{s}$ and the channel codebook $x$, similarly to [13, Eq. (90)], we can upper bound the excess-distortion probability by

$$P_{e,k,n}(P, D | A = a, X = x, \hat{S} = \hat{s})$$

$$\leq \sum_{u \in [M]} P_{U|\hat{s}}(u|\hat{s}) \Pr \left\{ \bigcup_{j \in [M]: j \neq u} \left\{ \frac{P_{U|^\hat{s}}(j|\hat{s}) \prod_{t=1}^{n} P_{Y^n_t|X^n_t A_t}(Y^n_t|X^n_t, a_t)}{P_{U|\hat{s}}(u|\hat{s}) \prod_{t=1}^{n} P_{Y^n_t|X^n_t A_t}(Y^n_t|X^n_t, a_t)} \geq 1 \right\} \right\} + \Pr_{P_{U|\hat{s}}} \left\{ U > W(S^k, D)|\hat{S} = \hat{s} \right\}. \quad (107)$$

For each $t \in [\Psi]$, let $P_{Y^n_t|A_t}$ be induced by the channel law $P_{Y^n_t|X^n_t A_t}$, the input distribution $P_{X^n_t}$ and the fading distribution $f_A$. Using the definition of $Q_{Y^n_t|A_t}$ in (17) and [24, Eq. (4.91)], we obtain that for any $t \in [\Psi]$ and any fading parameter $a_t$,

$$\max_{y^n} \frac{P_{Y^n_t|A_t}(y^n_t|a_t)}{Q_{Y^n_t|A_t}(y^n_t|a_t)} \leq \frac{1 + P_t a_t^2}{\sqrt{1 + 2P_t a_t^2}}. \quad (108)$$

In the remaining part of this subsection, wherever we use $\mathbb{E}$, we mean the expectation with respect to the following distribution

$$\left( \prod_{i \in [M]} P_{S_i}^k(s^k) \right) \left( \prod_{t \in [\Psi]} \prod_{j \in [M]} P_{X^n_t}(X^n_t(j)) \right) \left( \prod_{t \in [\Psi]} P_{Y^n_t|X^n_t A_t} \right). \quad (109)$$

Using the definition of $W(S^k, D)$ in (103), combing (107), (108) and following the analyses in [13, Eq. (91)-(106)], we can obtain that

$$P_{e,k,n}(P, D)$$

$$\leq \mathbb{E} \left[ \min \left\{ 1, W(S^k, D) \prod_{t \in [\Psi]} \frac{P_{Y^n_t|A_t}(Y^n_t|A_t)}{Q_{Y^n_t|A_t}(Y^n_t|X^n_t, A_t)} \right\} \right] + \mathbb{E} \left[ (1 - \Phi(S^k, D))^{W(S^k, D)} \right] \quad (110)$$

$$\leq \mathbb{E} \left[ \min \left\{ 1, W(S^k, D) \prod_{t \in [\Psi]} \frac{1 + P_t A_t^2}{\sqrt{1 + 2P_t A_t^2}} \frac{Q_{Y^n_t|A_t}(Y^n_t|A_t)}{P_{Y^n_t|X^n_t A_t}(Y^n_t|X^n_t, A_t)} \right\} \right]$$

$$+ \mathbb{E} \left[ (1 - \Phi(S^k, D))^{\frac{2}{W(S^k, D) - 1}} \right] \quad (111)$$

$$\leq \mathbb{E} \left[ \exp \left( - \sum_{t \in [\Psi]} \left( \tilde{i}(X^n_t; Y^n_t|A_t) - \log \frac{1 + P_t A_t^2}{\sqrt{1 + 2P_t A_t^2}} - \log W(S^k, D) \right) \right) \right]$$

$$+ \mathbb{E} \left[ \exp \left\{ - \Phi(S^k, D) \left( \frac{n^{\gamma}}{\Phi(S^k, D)} - 1 \right) \right\} \right] \quad (112)$$
\[\begin{align*}
&= \mathbb{E}\left[ \exp\left( - \sum_{t \in [\Psi]} \left( \bar{i}(X^n_t; Y^n_t | A_t) - \log \frac{1 + P_t A_{t}^2}{\sqrt{1 + 2P_t A_{t}^2}} \right) - \log \frac{n \gamma^2}{\Phi(S^k, D)} \right) \right] \\
&\quad + \exp(-n \gamma + 1) \\
&= \mathbb{E}\left[ \exp\left( - \sum_{t \in [\Psi]} \left( L^n_t(P_t, A_t, Z^n_t) - \log \frac{1 + P_t A_{t}^2}{\sqrt{1 + 2P_t A_{t}^2}} \right) - \log \frac{n \gamma^2}{\Phi(S^k, D)} \right) \right] \\
&\quad + \exp(-n \gamma + 1),
\end{align*}\]

where (112) follows by using the inequality \((1 - a)^M \leq \exp(-Ma)\) for any \(a \in [0, 1]\); and (114) follows from \(\Phi(s^k, D) \leq 1\) for any \(s^k\) and the result in Lemma 2 which states that under the distribution \(P_{Y^n | X^n, A_t}\), the fading information density \(\bar{i}(X^n; Y^n | A_t)\) has the same distribution as \(L^n_t(P_t, A_t, Z^n_t)\).

The proof of Theorem 4 is now complete.

C. Proof of Lemma 8

Recall the definition of \(D\) in (75). Given \(s^k \in \mathcal{T}_Q\), and \(\hat{s}_i\), we obtain that

\[\begin{align*}
\Pr\{\hat{I}, \hat{J} \neq (i, j(S^k, \hat{s}_i)) | S^k = s^k, \hat{s}_i = \hat{s}_i\} \\
= \Pr\left\{ \sum_{t \in [\Psi]} \bar{i}(X^n_t(i, j(s^k, \hat{s}_i)); Y^n_t | A_t) \leq \log \gamma_{k,n} + \log M_t \right\} \\
+ \Pr\left\{ \exists (\bar{i}, \bar{j}) \in D : (\bar{i}, \bar{j}) \neq (i, j(s^k, \hat{s}_i)), \sum_{t \in [\Psi]} \bar{i}(X^n_t(\bar{i}, \bar{j}); Y^n_t | A_t) > \gamma_{k,n} + \log M_t \right\}, \\
\leq \Pr\left\{ \sum_{t \in [\Psi]} \bar{i}(X^n_t(i, j(s^k, \hat{s}_i)); Y^n_t | A_t) \leq \log \gamma_{k,n} + \log M_t \right\} \\
+ \sum_{(\bar{i}, \bar{j}) \in D: (\bar{i}, \bar{j}) \neq (i, j(s^k, \hat{s}_i))} \Pr\left\{ \sum_{t \in [\Psi]} \bar{i}(X^n_t(\bar{i}, \bar{j}); Y^n_t | A_t) > \log \gamma_{k,n} + \log M_t \right\} \\
\leq \Pr\left\{ \sum_{t \in [\Psi]} \bar{i}(X^n_t; Y^n_t | A_t) \leq \log \gamma_{k,n} + \log M_t \right\} \\
+ \sum_{i=1}^{\lfloor P_k(S) \rfloor} M_t \Pr\left\{ \sum_{t \in [\Psi]} \bar{i}(X^n_t; \tilde{Y}^n_t | A_t) > \log \gamma_{k,n} + \log M_t \right\} \\
\leq \Pr\left\{ \sum_{t \in [\Psi]} \bar{i}(X^n_t; Y^n_t | A_t) \leq \log \gamma_{k,n} + \log M_t \right\} + \frac{\|P_k(S)\| \mathbb{E}_A[B_2(P, A)]}{\gamma_{k,n} \sqrt{n}},
\end{align*}\]

where in (117), for each \(t \in [\Psi]\), \((X^n_t, A_t, Y^n_t, \tilde{Y}^n_t) \sim P_{X^n_t} P_{A_t} P_{Y^n_t | X^n_t, A_t} P_{\tilde{Y}^n_t | A_t}\) and (117) follows since the channel output \(Y^n_t = X^n_t(i, j(s^k, \hat{s}_i)) + Z^n_t\) is independent of all codewords \(X^n_t(\bar{i}, \bar{j})\) where \((\bar{i}, \bar{j}) \neq (i, j(s^k, \hat{s}_i))\).
(i, j(s^k, s_i)); and (118) follows from using the result in [24, Eq. (4.93)] and the definition of $B_2(P, A)$ in (82).

REFERENCES


