A High-SNR Normal Approximation for Single-Antenna Rayleigh Block-Fading Channels

(article starts on next page)

A High-SNR Normal Approximation
for Single-Antenna Rayleigh Block-Fading Channels

Alejandro Lancho[†], Tobias Koch[†], and Giuseppe Durisi[*]
[†]Universidad Carlos III de Madrid, Leganés, Spain, and Gregorio Marañón Health Research Institute, Madrid, Spain.
[*]Chalmers University of Technology, Gothenburg, Sweden
Emails: {alancho, koch}@tsc.uc3m.es, durisi@chalmers.se

Abstract—**This paper concerns the maximal achievable rate at which data can be transmitted over a non-coherent, single-antenna, Rayleigh block-fading channel using an error-correcting code of a given blocklength with a block-error probability not exceeding a given value. In particular, a high-SNR normal approximation of the maximal achievable rate is presented that becomes accurate as the signal-to-noise ratio (SNR) and the number of coherence intervals L over which we code tend to infinity. Numerical analyses suggest that the approximation is accurate already at SNR values of $15\,$dB.**

I. INTRODUCTION

There exists an increasing interest in the problem of transmitting short packets in wireless communications. For example, the vast majority of wireless connections in the fifth generation of cellular systems (5G) will most likely be originated by autonomous machines and devices, which predominantly exchange short packets. It is also expected that enhanced mobile-broadband services will be complemented by new services that target systems requiring reliable real-time communication with stringent requirements on latency and reliability. For more details see [1] and references therein. While in the absence of latency constraints, *capacity* and *outage capacity* provide accurate benchmarks for the throughput achievable in wireless communication systems, for low-latency wireless communications a more refined analysis of the maximal achievable rate as a function of the blocklength is needed. Such an analysis is provided in this paper.

Let $R^*(n, \epsilon)$ denote the maximal achievable rate at which data can be transmitted using an error-correcting code of a determined length n with a block-error probability not larger than ϵ. Building upon Dobrushin's and Strassen's asymptotic results, Polyanskiy, Poor and Verdú showed that for various

channels with a positive capacity C, the maximal achievable rate can be tightly approximated by [2]

$$R^*(n, \epsilon) = C - \sqrt{\frac{V}{n}}Q^{-1}(\epsilon) + \mathcal{O}\left(\log n/n\right) \qquad (1)$$

where V denotes the channel dispersion, $Q^{-1}(\cdot)$ denotes the inverse Q-function, and $\mathcal{O}\left(\log n/n\right)$ comprises terms that decay no slower than $\log n/n$. The approximation that follows from (1) by ignoring the $\mathcal{O}\left(\log n/n\right)$ terms is sometimes referred to as *normal approximation*.

The work by Polyanskiy *et al.* [2] has been generalized to some wireless communication channels; see, e.g., [3]–[10]. However, the channel dispersion in the non-coherent setting—where neither transmitter nor receiver have *a priori* knowledge of the realizations of the fading coefficients—is only known in the quasi-static case [5], where it is zero. For general non-coherent block-fading channels, non-asymptotic bounds on the maximal achievable rate that can be evaluated numerically were presented, e.g., in [3], [7]. Obtaining an expression for the channel dispersion of non-coherent block-fading channels is difficult because for such channels the capacity-achieving input distribution is in general unknown. Thus, the standard approach, which consists of first evaluating non-asymptotic upper and lower bounds on $R^*(n, \epsilon)$ for the capacity-achieving input distribution and then analyzing these bounds in the limit as $n \to \infty$, cannot be followed.

In this paper, we present an expression similar to (1) of the maximal rate $R^*(L, \epsilon, \rho)$ achievable over non-coherent, single-antenna, Rayleigh block-fading channels using error-correcting codes that span L coherence intervals, have a block-error probability not larger than ϵ, and satisfy the power constraint ρ. By replacing in (1) the capacity and channel dispersion by asymptotically tight approximations, we obtain a high-SNR normal approximation of $R^*(L, \epsilon, \rho)$.

II. SYSTEM MODEL

We consider a single-antenna Rayleigh block-fading channel with coherence interval T. For this channel model, the input-output relation within the ℓ-th coherence interval is given by

$$\mathbf{Y}_\ell = H_\ell \mathbf{X}_\ell + \mathbf{W}_\ell \qquad (2)$$

where $\mathbf{X}_\ell$ and $\mathbf{Y}_\ell$ are T-dimensional, complex-valued, random vectors containing the input and output signals, respectively;

$\mathbf{W}_\ell$ is a random vector with independent and identically distributed (i.i.d.) zero-mean, unit-variance, circularly-symmetric, complex Gaussian entries; and H_ℓ is Rayleigh fading, i.e., it is a zero-mean, unit-variance, circularly-symmetric, complex Gaussian random variable. We assume that H_ℓ and $\mathbf{W}_\ell$ are independent and take on independent realizations over successive coherence intervals. We further assume that the joint law of $(H_\ell, \mathbf{W}_\ell)$ does not depend on the channel inputs. We consider a non-coherent setting where transmitter and receiver are aware of the distribution of H_ℓ but not of its realization.

We next introduce the notion of a channel code. For simplicity, we shall restrict ourselves to codes whose blocklength n satisfies $n = LT$, where L denotes the number of blocks of length T needed to transmit the whole code. An (M, L, ϵ, ρ) code for the channel (2) consists of the following:

1) An encoder $f\colon \{1, \ldots, M\} \to \mathbb{C}^{LT}$ that maps the message $A \in \{1, \ldots, M\}$ to a codeword $\mathbf{X}^L = [\mathbf{X}_1, \ldots, \mathbf{X}_L]$. The codewords are assumed to satisfy the power constraint[1]

$$\|\mathbf{X}_\ell\|^2 \le T\rho, \quad \ell = 1, \ldots, L. \tag{3}$$

Since the variances of H_ℓ and $\mathbf{W}_\ell$ are normalized, ρ in (3) can be interpreted as the average SNR at the receiver.

2) A decoder $g\colon \mathbb{C}^{LT} \to \{1, \ldots, M\}$ satisfying a maximum error probability constraint

$$\max_{1 \le a \le M} \mathbf{P}\left[g(\mathbf{Y}^L) \ne A \,\middle|\, A = a\right] \le \epsilon \tag{4}$$

where $\mathbf{Y}^L = [\mathbf{Y}_1, \ldots, \mathbf{Y}_L]$ is the channel output induced by the transmitted codeword $\mathbf{X}^L = f(a)$ according to (2).

The *maximal coding rate* is defined as

$$R^*(L, \epsilon, \rho) \triangleq \sup\left\{\frac{\log M}{LT} : \exists (M, L, \epsilon, \rho) \text{ code}\right\}. \tag{5}$$

III. PRELIMINARIES AND NOTATION

We shall refer to the distribution $\mathrm{P}_{\mathbf{X}^L}^{(\mathrm{U})}$, according to which $\mathbf{X}^L = \sqrt{T\rho}\mathbf{U}^L$ (where $\mathbf{U}^L = [\mathbf{U}_1, \ldots, \mathbf{U}_L]$ are i.i.d. and uniformly distributed on the unit sphere in $\mathbb{C}^T$) as unitary space-time modulation (USTM) [11]. This distribution is relevant because it yields a lower bound on capacity that is asymptotically tight at high SNR [12], [13].

The outputs $\mathbf{Y}^L$ induced by the USTM input distribution have pdf

$$\mathrm{q}_{\mathbf{Y}^L}^{(\mathrm{U})}(\mathbf{y}^L) = \prod_{\ell=1}^L \mathrm{q}_{\mathbf{Y}}^{(\mathrm{U})}(\mathbf{y}_\ell), \quad \mathbf{y}^L = [\mathbf{y}_1, \ldots, \mathbf{y}_L] \in \mathbb{C}^{LT} \tag{6}$$

where [3, Eq. (18)]

$$\mathrm{q}_{\mathbf{Y}}^{(\mathrm{U})}(\mathbf{y}) = \frac{e^{-\|\mathbf{y}\|^2/(1+T\rho)}\|\mathbf{y}\|^{2(1-T)}\Gamma(T)}{\pi^T(1+T\rho)}$$
$$\times \tilde{\gamma}\left(T-1, \frac{T\rho\|\mathbf{y}\|^2}{1+T\rho}\right)\left(1 + \frac{1}{T\rho}\right)^{T-1}, \quad \mathbf{y} \in \mathbb{C}^T. \tag{7}$$

Here, $\tilde{\gamma}(\cdot, \cdot)$ denotes the regularized lower incomplete gamma function and $\Gamma(\cdot)$ denotes the gamma function.

Conditioned on $\|\mathbf{X}_\ell\|^2 = T\alpha_\ell$, $\alpha_\ell \in [0, \rho]$, the distributions of $|\mathbf{Y}_\ell^H \mathbf{X}_\ell|^2$ and $\|\mathbf{Y}_\ell\|^2$ are as follows:

$$|\mathbf{Y}_\ell^H \mathbf{X}_\ell|^2 \overset{\mathscr{L}}{=} T\alpha_\ell(1 + T\alpha_\ell)Z_{1,\ell} \tag{8a}$$

$$\|\mathbf{Y}_\ell\|^2 \overset{\mathscr{L}}{=} (1 + T\alpha_\ell)Z_{1,\ell} + Z_{2,\ell} \tag{8b}$$

where $\overset{\mathscr{L}}{=}$ denotes equality in distribution, $\{Z_{1,\ell}, \ell \in \mathbb{Z}\}$ is a sequence of i.i.d. Gamma$(1,1)$-distributed random variables, and $\{Z_{2,\ell}, \ell \in \mathbb{Z}\}$ is a sequence of i.i.d. Gamma$(T-1, 1)$-distributed random variables (with Gamma(z, q) denoting the gamma distribution with parameters z and q). We will omit the subscript ℓ when immaterial.

We next introduce some notation and preliminary results that will be helpful in the remainder of the paper. The *information density* between $\mathbf{X}^L$ and $\mathbf{Y}^L$ is denoted by

$$i(\mathbf{X}^L; \mathbf{Y}^L) \triangleq \log\left(\frac{\mathrm{p}_{\mathbf{Y}^L|\mathbf{X}^L}(\mathbf{Y}^L \mid \mathbf{X}^L)}{\mathrm{p}_{\mathbf{Y}^L}(\mathbf{Y}^L)}\right) \tag{9}$$

where $\mathrm{p}_{\mathbf{Y}^L}$ is the output distribution induced by the input distribution. When the input distribution is USTM, the conditional information density, conditioned on $\|\mathbf{X}_\ell\|^2 = T\rho$, can be expressed as $\sum_{\ell=1}^L i_\ell(\rho)$, where

$$i_\ell(\rho) \triangleq (T-1)\log(T\rho) - \log\Gamma(T) - \frac{T\rho Z_{2,\ell}}{1+T\rho}$$
$$+ (T-1)\log\left(\frac{(1+T\rho)Z_{1,\ell} + Z_{2,\ell}}{1+T\rho}\right)$$
$$- \log\tilde{\gamma}\left(T-1, \frac{T\rho((1+T\rho)Z_{1,\ell} + Z_{2,\ell})}{1+T\rho}\right). \tag{10}$$

Since $\tilde{\gamma}(T-1, x) \le 1$, $x \ge 0$ we can lower-bound (10) by

$$\underline{i}_\ell(\rho) \triangleq (T-1)\log(T\rho) - \log\Gamma(T) - \frac{T\rho Z_{2,\ell}}{1+T\rho}$$
$$+ (T-1)\log\left(\frac{(1+T\rho)Z_{1,\ell} + Z_{2,\ell}}{1+T\rho}\right). \tag{11}$$

We define the *missmatched information density* as

$$j(\mathbf{X}^L; \mathbf{Y}^L) \triangleq \log\left(\frac{\mathrm{p}_{\mathbf{Y}^L|\mathbf{X}^L}(\mathbf{Y}^L \mid \mathbf{X}^L)}{\mathrm{q}_{\mathbf{Y}^L}^{(\mathrm{U})}(\mathbf{Y}^L)}\right). \tag{12}$$

Using this definition together with (2) and (7), conditioned on $\|\mathbf{X}_\ell\|^2 = T\alpha_\ell$, the missmatched information density can be expressed as $j(\mathbf{X}^L; \mathbf{Y}^L) = \sum_{\ell=1}^L j_\ell(\alpha_\ell)$, where

$$j_\ell(\alpha_\ell) = (T-1)\log(T\rho) - \log\Gamma(T) + \frac{(T\alpha_\ell - T\rho)Z_{1,\ell}}{1+T\rho}$$
$$+ (T-1)\log\left(\frac{(1+T\alpha_\ell)Z_{1,\ell} + Z_{2,\ell}}{1+T\rho}\right)$$
$$- \log\tilde{\gamma}\left(T-1, \frac{T\rho((1+T\alpha_\ell)Z_{1,\ell} + Z_{2,\ell})}{1+T\rho}\right)$$
$$- \frac{T\rho Z_{2,\ell}}{1+T\rho} + \log\left(\frac{1+T\rho}{1+T\alpha_\ell}\right). \tag{13}$$

Note that $j_\ell(\rho) = i_\ell(\rho)$ when the input distribution is USTM.

By lower-bounding $\tilde{\gamma}(\cdot,\cdot)$ using a result by Alzer [14, Th. 1] we obtain that $j_\ell(\alpha_\ell) \leq \bar{j}_\ell(\alpha_\ell)$, $\alpha_\ell \in [0,\rho]$, where

$$\bar{j}_\ell(\alpha_\ell) \triangleq (T-1)\log(T\rho) - \log\Gamma(T) + \frac{(T\alpha_\ell - T\rho)Z_{1,\ell}}{1+T\rho}$$
$$- \frac{T\rho Z_{2,\ell}}{1+T\rho} + \log\frac{1+T\rho}{1+T\alpha_\ell}$$
$$+ (T-1)\log\left(\frac{(1+T\alpha_\ell)Z_{1,\ell} + Z_{2,\ell}}{1+T\rho}\right)$$
$$+ (T-1)\log\left(1 + \frac{\Gamma(T)^{\frac{1}{T-1}}\frac{1+T\rho}{T\rho}}{(1+T\alpha_\ell)Z_{1,\ell} + Z_{2,\ell}}\right). \quad (14)$$

We denote the expected values of $\bar{j}_\ell(\alpha)$ conditioned on $\|\mathbf{X}_\ell\|^2 = T\alpha$, and of $i_\ell(\rho)$ and $\underline{i}_\ell(\rho)$, conditioned on $\|\mathbf{X}_\ell\|^2 = T\rho$, by $\bar{J}(\alpha)$, $I(\rho)$, and $\underline{I}(\rho)$, respectively. We further define

$$U(\rho) \triangleq \mathsf{E}\left[\left(i_\ell(\rho) - I(\rho)\right)^2\right] \quad (15a)$$

$$\bar{V}_\rho(\alpha) \triangleq \mathsf{E}\left[\left(\bar{j}_\ell(\alpha) - \bar{J}(\alpha)\right)^2\right] \quad (15b)$$

where the subscript ρ in $\bar{V}_\rho(\alpha)$ is introduced to highlight that $\bar{V}_\rho(\alpha)$ depends both on α and ρ, but it is omitted when $\alpha = \rho$.

In [15] we show that $I(\rho)$, $U(\rho)$, $\bar{J}(\rho)$ and $\bar{V}_\rho(\rho)$ can be approximated as

$$I(\rho) = \underline{I}(\rho) + o_\rho(1) \quad (16a)$$
$$U(\rho) = \tilde{U} + o_\rho(1) \quad (16b)$$
$$\bar{J}(\rho) = \underline{I}(\rho) + o_\rho(1) \quad (16c)$$
$$\bar{V}(\rho) = \tilde{U} + o_\rho(1) \quad (16d)$$

where $o_\rho(1)$ comprises terms that are independent of L and that vanish as $\rho \to \infty$. The term $\underline{I}(\rho)$ is given by

$$\underline{I}(\rho) = (T-1)\log(T\rho) - \log\Gamma(T) - (T-1)\left[\log(1+T\rho)\right.$$
$$\left. + \frac{T\rho}{1+T\rho} - \psi(T-1)\right] + {}_2F_1\left(1, T-1; T; \frac{T\rho}{1+T\rho}\right). \quad (17)$$

where $\psi(\cdot)$ and ${}_2F_1(\cdot,\cdot;\cdot;\cdot)$ denote the digamma function and the hypergeometric function, respectively. Furthermore,

$$\tilde{U} \triangleq (T-1)^2\frac{\pi^2}{6} + (T-1). \quad (18)$$

IV. MAIN RESULTS

The main result of this paper is a high-SNR normal approximation of $R^*(L,\epsilon,\rho)$ presented in Section IV-A. A discussion of this approximation is provided in Section IV-B.

A. High-SNR normal approximation

Theorem 1: Assume that $T > 2$ and $0 < \epsilon < 1/2$. Then,

$$R^*(L,\epsilon,\rho) = \frac{\underline{I}(\rho)}{T} + o_\rho(1) - \sqrt{\frac{\tilde{U} + o_\rho(1)}{LT^2}}Q^{-1}(\epsilon)$$
$$+ \mathcal{O}_L\left(\log L/L\right) \quad (19)$$

where $\mathcal{O}_L\left(\log L/L\right)$ comprises terms that are independent of ρ and that decay no slower than $\log L/L$.
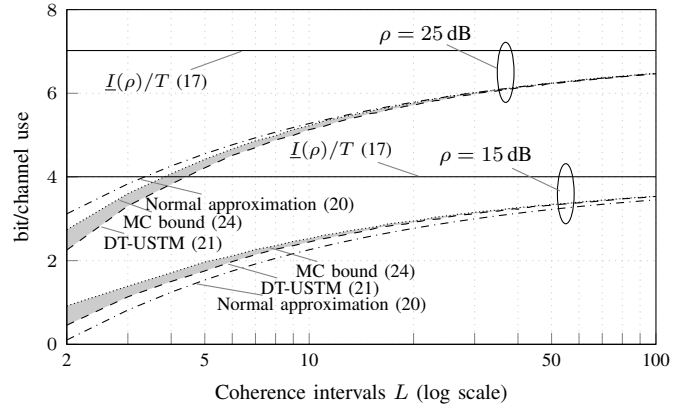
Proof: See Section V. ∎

Fig. 1: Bounds on $R^*(L,\epsilon,\rho)$ for $T = 20$, $\epsilon = 10^{-3}$, and SNR values $\rho = 15\,\text{dB}$ and $\rho = 25\,\text{dB}$.

Ignoring the $\mathcal{O}_L(\log L/L)$ and the $o_\rho(1)$ terms in (19), we obtain the following high-SNR normal approximation:

$$R^*(L,\epsilon,\rho) \approx \frac{\underline{I}(\rho)}{T} - \sqrt{\frac{\tilde{U}}{LT^2}}Q^{-1}(\epsilon). \quad (20)$$

It is shown in [11, Th. 4] that $\underline{I}(\rho)/T$ is an asymptotically-tight lower bound on capacity. According to Theorem 1, $\tilde{U}/T^2$ can be viewed as a high-SNR approximation of the dispersion. For comparison, the dispersion of the coherent block-fading channel is $\pi^2/6 + 1/T + o_\rho(1)$ [4, Th. 2], and the dispersion of the (non-fading) complex AWGN channel is $1 + o_\rho(1)$ [1, Eq. (8)]. Observe that $\tilde{U}/T^2$ corresponds to the dispersion one obtains by transmitting one pilot symbol per coherence block to estimate the fading coefficient, and by then transmitting $T - 1$ symbols per coherence block over a coherent fading channel. This suggests that, at high SNR, one pilot symbol per coherence block suffices to estimate the fading coefficient.

It is perhaps surprising that the dispersion in the non-coherent case is smaller than the one of the coherent case. However, the capacity of the coherent fading channel is larger than $\underline{I}(\rho)/T$, so the rate $R^*(L,\epsilon,\rho)$ of the coherent channel is larger than that of the non-coherent channel.

B. Numerical Examples and Discussion

We illustrate the accuracy of the high-SNR normal approximation (20) by means of numerical examples. In Fig. 1 we show (20) as a function of $L = n/T$ for a fixed coherence interval T and for both $\rho = 15\,\text{dB}$ and $\rho = 25\,\text{dB}$. We further present a non-asymptotic (in ρ and L) lower bound on $R^*(L,\epsilon,\rho)$ that is based on the Dependence Testing (DT) bound [2, Th. 22] (see (21) below) as well as a non-asymptotic (in ρ and L) upper bound on $R^*(L,\epsilon,\rho)$ that is based on the Meta Converse (MC) bound [2, Th. 31] (see (24) below). We finally plot $\underline{I}(\rho)/T$. Observe that (20) is accurate already for $\rho = 15\,\text{dB}$. Further observe that (20) is pessimistic for $\rho = 15\,\text{dB}$ and optimistic for $\rho = 25\,\text{dB}$. This suggests that $o_\rho(1)$ is negative and that $\mathcal{O}_L\left(\log L/L\right)$ is positive.

In Fig. 2, we show the high-SNR normal approximation (20) as a function of the coherence interval T for a fixed block-length n (hence L is inversely proportional to T). For compari-
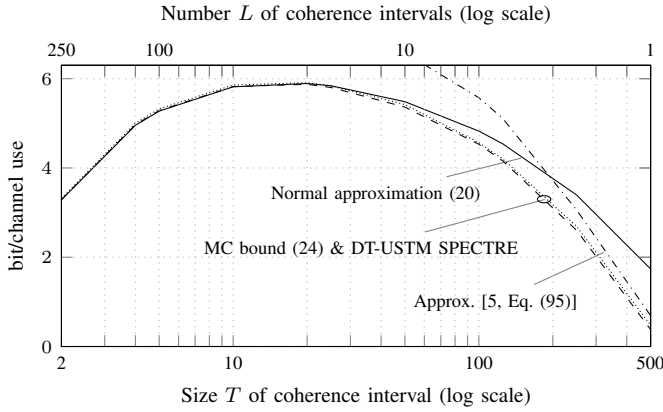
Fig. 2: Bounds on $R^*(L, \epsilon, \rho)$ for $LT = 500$, $\epsilon = 10^{-3}$, and $\rho = 25\,\mathrm{dB}$.

son, we also show the DT bound, evaluated for an USTM input distribution and implemented in the communication toolbox SPECTRE [16, Sec. 3.1], and the MC bound. Finally, we present the normal approximation that was given in [5, Eq. (95)] for quasi-static multiple-input multiple-output (MIMO) block-fading channels. To adapt the quasi-static MIMO block-fading channel to our system model, we replace $\mathbb{H}$ in [5] by an $L \times L$ diagonal matrix with diagonal entries $H_1, \ldots, H_L$. Observe that the high-SNR normal approximation (20) is accurate for $L \geq 10$ and then becomes less accurate as L decreases. Further observe that the normal approximation for the quasi-static case [5, Eq. (95)], which is tailored towards the case where L is small, becomes accurate only for $L \leq 2$.

V. PROOF OF THEOREM 1

The proof is based on a lower bound on $R^*(L, \epsilon, \rho)$, given in Section V-A, and on an upper bound on $R^*(L, \epsilon, \rho)$, given in Section V-B. Due to space limitations, some proofs are deferred to the longer version of our paper [15].

A. DT Lower Bound

To obtain a lower bound on $R^*(L, \rho, \epsilon)$, we evaluate the DT bound for USTM inputs. Thus, we assume that $\mathbf{X}^L \sim P_{\mathbf{X}^L}^{(U)}$, which implies $\mathbf{Y}^L \sim q_{\mathbf{Y}^L}^{(U)}$. For this choice, the cumulative distribution function $P[i(\mathbf{x}^L; \tilde{\mathbf{Y}}^L) \leq \alpha]$ (where $\tilde{\mathbf{Y}}^L \sim q_{\mathbf{Y}^L}^{(U)}$) does not depend on $\mathbf{x}^L$. A lower bound on $R^*(L, \epsilon, \rho)$ follows therefore from the DT bound (maximum probability of error) [2, Th. 22], which after a standard change of measure can be stated as follows: there exists a code with M codewords and maximal probability of error ϵ not exceeding

$$\epsilon \leq (M-1)\mathsf{E}\left[e^{-i(\mathbf{X}^L;\mathbf{Y}^L)}\mathrm{I}\{i(\mathbf{X}^L;\mathbf{Y}^L) > \log(M-1)\}\right] + \mathsf{P}\left[i\left(\mathbf{X}^L;\mathbf{Y}^L\right) \leq \log(M-1)\right] \quad (21)$$

where $\mathrm{I}\{\cdot\}$ denotes the indicator function. To show that (21) yields the lower bound

$$R^*(L, \epsilon, \rho) \geq \frac{I(\rho)}{T} - \sqrt{\frac{U(\rho)}{LT^2}}Q^{-1}(\epsilon) + \mathcal{O}_L\left(1/L\right) \quad (22)$$

(where $\mathcal{O}_L(1/L)$ comprises terms that are independent of ρ and that are of order $1/L$) we follow almost verbatim the steps

[2, Eqs. (258)–(267)] (with γ in [2] replaced by $M - 1$). The main difference is that in our case $U(\rho)$ and $B(\rho)$ (cf. [2, Eq. (254)]), defined as

$$B(\rho) \triangleq \frac{6\mathsf{E}\left[|i_\ell(\rho) - I(\rho)|^3\right]}{U(\rho)^{3/2}} \quad (23)$$

depend on ρ. To ensure that $\mathcal{O}_L(1/L)$ is independent of ρ, we thus need to ensure that both $U(\rho)$ and $B(\rho)$ are bounded in ρ. We then apply the Berry-Esseen theorem to obtain [2, Eq. (259)] with $B(\rho)$ replaced by an upper bound $B(T)$, followed by [2, Eqs. (261)–(266)]. This yields (22).

The proof that shows that $U(\rho)$ and $B(\rho)$ are bounded in ρ can be found in [15].

B. Meta Converse (MC) Upper Bound

An upper bound on $R^*(L, \epsilon, \rho)$ follows from the MC bound [2, Th. 31] computed for the auxiliary pdf $q_{\mathbf{Y}^L}^{(U)}$

$$R^*(L, \epsilon, \rho) \leq \frac{1}{LT}\sup_{\boldsymbol{\alpha} \in [0, \rho]^L} \log\left(\frac{1}{\beta(\boldsymbol{\alpha}, q_{\mathbf{Y}^L}^{(U)})}\right). \quad (24)$$

Here, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_L)$ denotes the vector of power allocations, and $\beta(\boldsymbol{\alpha}, q_{\mathbf{Y}^L}^{(U)})$ denotes the minimum probability of error under hypothesis $q_{\mathbf{Y}^L}^{(U)}$ if the probability of error under hypothesis $p_{\mathbf{Y}^L|\mathbf{X}^L=\mathbf{x}^L}$ does not exceed ϵ [2, Eq. (100)]. Note that $\beta(\boldsymbol{\alpha}, q_{\mathbf{Y}^L}^{(U)})$ depends on $\mathbf{x}^L$ only via $\boldsymbol{\alpha}$.

Fix an arbitrary $0 < \delta < 1$. We show in [15] that for sufficiently large L and ρ, the supremum in (24) can be replaced by a supremum over $\boldsymbol{\alpha} \in \mathcal{A}_{\rho, \delta}$, where

$$\mathcal{A}_{\rho, \delta} \triangleq \{\boldsymbol{\alpha} \in [0, \rho]^L : L_{\boldsymbol{\alpha}}(\delta) \geq L/2\} \quad (25)$$

and $L_{\boldsymbol{\alpha}}(\delta)$ denotes the number of α_ℓ's in $\boldsymbol{\alpha}$ that satisfy $\rho(1 - \delta) \leq \alpha_\ell \leq \rho$.

In the following we implicitly assume that $L \geq L_0$ and $\rho \geq \rho_0$ for some sufficiently large L_0 and ρ_0. It follows from [15], [2, Eq. (106)], and because $j_\ell(\alpha_\ell) \leq \bar{j}_\ell(\alpha_\ell)$, $\alpha_\ell \in [0, \rho]$, that (24) can be upper-bounded as

$$R^*(L, \epsilon, \rho) \leq \sup_{\boldsymbol{\alpha} \in \mathcal{A}_{\rho, \delta}} \left\{ \frac{\log \xi(\boldsymbol{\alpha})}{LT} \right.$$
$$\left. - \frac{\log\left(1 - \epsilon - \mathsf{P}\left[\sum_{\ell=1}^L \bar{j}_\ell(\alpha_\ell) \geq \log \xi(\boldsymbol{\alpha})\right]\right)}{LT} \right\} \quad (26)$$

for an arbitrary $\xi \colon [0, \rho]^L \to (0, \infty)$.

Let

$$\bar{B}(\boldsymbol{\alpha}) \triangleq \frac{6\sum_{\ell=1}^L \mathsf{E}\left[|\bar{j}_\ell(\alpha_\ell) - \bar{J}(\alpha_\ell)|^3\right]}{\left(\sum_{\ell=1}^L \bar{V}_\rho(\alpha_\ell)\right)^{3/2}}. \quad (27)$$

We show in [15] that the expectation $\mathsf{E}\left[|\bar{j}_\ell(\alpha) - \bar{J}(\alpha)|^3\right]$ can be upper-bounded by a constant $\bar{S}(T)$ that only depends on T. We further show that, for $\rho(1 - \delta) \leq \alpha \leq \rho$,

$$\bar{V}_\rho(\alpha) \geq (T - 1) + o_\rho(1) + o_\delta(1) \quad (28)$$

where $o_\delta(1)$ comprises terms that are independent of L and ρ and that vanish as $\delta \downarrow 0$. It thus follow that, for ρ_0 sufficiently large and δ sufficiently small,

$$\sum_{\ell=1}^{L} \bar{V}_\rho(\alpha_\ell) \geq L_{\boldsymbol{\alpha}}(\delta) \frac{T-1}{2}. \tag{29}$$

Hence, for every $\boldsymbol{\alpha} \in \mathcal{A}_{\rho,\delta}$ and δ sufficiently small,

$$\bar{B}(\boldsymbol{\alpha}) \leq \frac{6L\bar{S}(T)}{\left(\frac{(T-1)L}{4}\right)^{3/2}} \triangleq \frac{\bar{B}(T)}{\sqrt{L}}. \tag{30}$$

Let

$$\lambda = Q^{-1}\left(\epsilon + \frac{2\bar{B}(T)}{\sqrt{L}}\right) \tag{31}$$

and

$$\log \xi(\boldsymbol{\alpha}) = \sum_{\ell=1}^{L} \bar{J}(\alpha_\ell) - \lambda \sqrt{\sum_{\ell=1}^{L} \bar{V}_\rho(\alpha_\ell)}. \tag{32}$$

With this choice, the Berry-Esseen theorem and (30) imply that, for every $\boldsymbol{\alpha} \in \mathcal{A}_{\rho,\delta}$,

$$\mathsf{P}\left[\sum_{\ell=1}^{L} \bar{j}_\ell(\alpha_\ell) \leq \log \xi(\boldsymbol{\alpha})\right] \geq \epsilon + \frac{\bar{B}(T)}{\sqrt{L}}. \tag{33}$$

Substituting (33) into (26), we obtain

$$R^*(L, \epsilon, \rho) \leq \frac{1}{2} \frac{\log L}{LT} - \frac{\log \bar{B}(T)}{LT}$$
$$+ \sup_{\boldsymbol{\alpha} \in [0,\rho]^L} \left\{ \frac{\sum_{\ell=1}^{L} \bar{J}(\alpha_\ell)}{LT} - \lambda \sqrt{\frac{\sum_{\ell=1}^{L} \bar{V}_\rho(\alpha_\ell)}{L^2 T^2}} \right\}. \tag{34}$$

By the assumption $0 < \epsilon < 1/2$, we have that λ is positive for L_0 sufficiently large. Thus, by the concavity of $x \mapsto \sqrt{x}$ and Jensen's inequality

$$R^*(L, \epsilon, \rho) \leq \sup_{0 \leq \alpha \leq \rho} \left\{ \frac{\bar{J}(\alpha)}{T} - Q^{-1}(\epsilon)\sqrt{\frac{\bar{V}_\rho(\alpha)}{LT^2}} \right\}$$
$$+ \mathcal{O}_L\left(\log L/L\right) \tag{35}$$

where we performed a Taylor series expansion to approximate λ by $Q^{-1}(\epsilon)$. Indeed, we show in [15] that $\bar{V}_\rho(\alpha)$ is bounded in ρ and α, so the remainder can be combined with the first two terms on the RHS of (34) in a $\mathcal{O}_L(\log L/L)$ term.

We next show that

$$\sup_{0 \leq \alpha \leq \rho} \left\{ \frac{\bar{J}(\alpha)}{T} - \sqrt{\frac{\bar{V}_\rho(\alpha)}{LT^2}} Q^{-1}(\epsilon) \right\}$$
$$= \frac{\bar{J}(\rho)}{T} - \sqrt{\frac{\bar{V}(\rho)}{LT^2}} Q^{-1}(\epsilon) + \mathcal{O}_L\left(1/L\right). \tag{36}$$

It then follows by (16c) and (16d) that this upper bound coincides with (22) up to $o_\rho(1)$ terms.

Indeed, we demonstrate in [15] that for $T > 2$ and ρ_0 sufficiently large, we have

$$\sup_{0 \leq \alpha \leq \rho} \bar{J}(\alpha) = \bar{J}(\rho). \tag{37}$$

We further demonstrate that for $T > 2$, $0 < \epsilon < 1/2$, and sufficiently large L_0 and ρ_0, the supremum in (36) can be replaced by a supremum over $\alpha \in [\rho(1 - \mathsf{K}/L), \rho]$, where K is some non-negative constant. Together with the lower bound

$$\bar{V}_\rho(\alpha) \geq \bar{V}(\rho) - \Upsilon \frac{\mathsf{K}}{L}, \quad \rho\left(1 - \frac{\mathsf{K}}{L}\right) \leq \alpha \leq \rho \tag{38}$$

for some non-negative constant Υ (see [15]), this yields

$$\sup_{0 \leq \alpha \leq \rho} \left\{ \frac{\bar{J}(\alpha)}{LT} - \sqrt{\frac{\bar{V}_\rho(\alpha)}{LT^2}} Q^{-1}(\epsilon) \right\}$$
$$= \frac{\bar{J}(\rho)}{T} - \sqrt{\frac{\bar{V}(\rho) - \Upsilon\mathsf{K}/L}{LT^2}} Q^{-1}(\epsilon). \tag{39}$$

This proves (36) and concludes the proof of the upper bound.

REFERENCES

[1] G. Durisi, T. Koch, and P. Popovski, "Towards massive, ultra-reliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sept. 2016.

[2] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[3] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Diversity versus channel knowledge at finite block-length," in *Proc. IEEE ITW*, Sept. 2012, pp. 572–576.

[4] Y. Polyanskiy and S. Verdú, "Scalar coherent fading channel: Dispersion analysis," in *Proc. IEEE ISIT*, July 2011, pp. 2959–2963.

[5] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, July 2014.

[6] J. Östman, W. Yang, G. Durisi, and T. Koch, "Diversity versus multiplexing at finite blocklength," in *Proc. IEEE ISWCS*, Aug. 2014, pp. 702–706.

[7] G. Durisi, T. Koch, J. Östman, Y. Polyanskiy, and W. Yang, "Short-packet communications over multiple-antenna Rayleigh-fading channels," *IEEE Trans. on Commun.*, vol. 64, no. 2, pp. 618–629, Feb. 2016.

[8] A. Collins and Y. Polyanskiy, "Orthogonal designs optimize achievable dispersion for coherent MISO channels," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Jun. 2014, pp. 2524–2528.

[9] J. Hoydis, R. Couillet, and P. Piantanida, "The second-order coding rate of the MIMO quasi-static Rayleigh fading channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6591–6622, Dec. 2015.

[10] A. Collins and Y. Polyanskiy, "Dispersion of the coherent MIMO block-fading channel," in *Proc. IEEE ISIT*, July 2016, pp. 1068–1072.

[11] B. Hochwald and T. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.

[12] L. Zheng and D. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[13] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 117–132, Feb. 2013.

[14] H. Alzer, "On some inequalities for the incomplete gamma function," *Math Comp.*, vol. 66, no. 218, pp. 771–778, Apr. 1997.

[15] A. Lancho, T. Koch, and G. Durisi, "On single-antenna Rayleigh block-fading channels at finite blocklength," preprint. [Online]. Available: http://www.tsc.uc3m.es/~koch

[16] A. Collins, G. Durisi, T. Erseghe, V. Kostina, J. Östman, Y. Polyanskiy, I. Tal, and W. Yang. (2016) SPECTRE, short packet communication toolbox. [Online]. Available: "https://github.com/yp-mit/spectre"