Qi, Ch.; Koch, T. A High-SNR Normal Approximation for MIMO Rayleigh Block-Fading Channels, *2020 IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 21-26 June 2020. IEEE, 2020, Pp. 3025-3030*
DOI: https://doi.org/10.1109/ISIT44484.2020.9174409

A High-SNR Normal Approximation for MIMO Rayleigh Block-Fading Channels

Chao Qi and Tobias Koch
Universidad Carlos III de Madrid, Leganés, Spain and Gregorio Marañón Health Research Institute, Madrid, Spain.
Emails: {chaoqi,koch}@tsc.uc3m.es

***Abstract*—This paper concerns the maximum coding rate at which a code of given blocklength can be transmitted with a given block-error probability over a non-coherent Rayleigh block-fading channel with multiple transmit and receive antennas (MIMO). In particular, a high-SNR normal approximation of the maximum coding rate is presented, which is proved to become accurate as the signal-to-noise ratio (SNR) and the number of coherence intervals L tend to infinity.**

I. Introduction

The next-generations of cellular systems are expected to support the traffic generated by sensors and devices involved in Internet of Things or machine-to-machine communications. Such systems require the transmission of short packets with low latency and ultra-high reliability [1]. While *capacity* and *outage capacity* provide accurate benchmarks for the throughput achievable in wireless communication systems when the package length is not restricted, for short-package wireless communications, a more refined analysis of the maximum coding rate as a function of the blocklength is needed. Such an analysis is provided in this paper.

Let $R^*(n,\epsilon)$ denote the maximum coding rate at which data can be transmitted using an error-correcting code of length n with a block-error probability no larger than ϵ. Building upon Dobrushin's and Strassen's asymptotic results, Polyanskiy *et al.* showed that for various channels with a positive capacity C, $R^*(n,\epsilon)$ can be tightly approximated as [2]

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

where V is the so-called channel dispersion; $Q^{-1}(\epsilon)$ denotes the inverse of the Q-function $Q(x) \triangleq \int_x^\infty (1/\sqrt{2\pi})e^{-t^2/2}\mathrm{d}t$, and $\mathcal{O}(\log n/n)$ comprises terms that decay no slower than $\log n/n$. The approximation that follows by ignoring the $\mathcal{O}(\log n/n)$ term is sometimes referred to as *normal approximation*. The work of Polyanskiy *et al.* has been generalized to several wireless communication channels [3]–[9]. For instance, the channel dispersion of coherent fading channels was obtained in [3], [5], [6]. In the non-coherent setting, the channel dispersion is known in the quasi-static case, where it is zero [4]. For general non-coherent Rayleigh block-fading channels, non-asymptotic bounds on the maximum coding rate were presented in [8], [9]. Saddlepoint approximations that accurately approximate these bounds in the single-antenna case with a negligible computational cost were given in [10], [11]. However, a closed-form expression of the channel dispersion for general non-coherent Rayleigh block-fading channels is still missing. Obtaining such an expression for non-coherent block-fading channels is difficult because their capacity-achieving input distribution is in general unknown. Fortunately, the high-SNR asymptotic behavior of the capacity of such channels is well understood [12], [13]. This fact was exploited in [7] to derive a high-SNR normal approximation of $R^*(n,\epsilon)$ for non-coherent, single-antenna, Rayleigh block-fading channels.

In this paper, we generalize [7] to the MIMO case. In particular, we present an expression of $R^*(n,\epsilon)$ similar to (1) for non-coherent MIMO Rayleigh block-fading channels. By deriving asymptotically-tight approximations on the capacity and the channel dispersion at high SNR, we obtain a high-SNR normal approximation of $R^*(n,\epsilon)$, which complements the existing non-asymptotic bounds.

II. System Model

We consider a Rayleigh block-fading channel with n_t transmit antennas, n_r receive antennas, and coherence interval T. Within the l-th coherence interval, the channel input-output relation is given by

$$\mathbf{Y}_l = \mathbf{X}_l\mathbf{H}_l + \mathbf{W}_l \tag{2}$$

where $\mathbf{X}_l \in \mathbb{C}^{T\times n_t}$ and $\mathbf{Y}_l \in \mathbb{C}^{T\times n_r}$ are the complex-valued transmitted and received matrices, respectively; $\mathbf{H}_l \in \mathbb{C}^{n_t\times n_r}$ is the complex-valued fading matrix with independent and identically distributed (i.i.d.) $\mathcal{CN}(0,1)$ entries; $\mathbf{W}_l \in \mathbb{C}^{T\times n_r}$ is the additive noise at the receiver with i.i.d. $\mathcal{CN}(0,1)$ entries. We assume that $\mathbf{H}_l$ and $\mathbf{W}_l$ are independent and take on independent realizations over successive coherence intervals, and the joint law of $\mathbf{H}_l$ and $\mathbf{W}_l$ does not depend on $\mathbf{X}_l$. We consider a non-coherent setting where neither the transmitter nor the receiver has a priori knowledge of the realizations of $(\mathbf{H}_l,\mathbf{W}_l)$, but both know their statistics perfectly.

We assume that $T \geq n_t + n_r$ and $n_r \geq n_t$. The latter assumption does not reduce capacity at high SNR [12], and we believe it is also reasonable in the finite-blocklength regime.

C. Qi and T. Koch have received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 714161). T. Koch has further received funding from the Spanish Ministerio de Economía y Competitividad under Grants RYC-2014-16332 and TEC2016-78434-C3-3-R (AEI/FEDER, EU).

The former assumption ensures that a unitary space-time modulation (USTM) input distribution achieves a lower bound on the capacity that is asymptotically tight in the sense that the difference between the lower bound and the capacity vanishes as the SNR tends to infinity [12], [13]. Such an input can be written as $\mathbf{X}_l = \sqrt{T\rho/n_t}\mathbf{U}_l$, where $\mathbf{U}_l \in \mathbb{C}^{T\times n_t}$ satisfies $\mathbf{U}_l^H\mathbf{U}_l = \mathbf{I}_{n_t}$ and is isotropically distributed.

For simplicity, we shall restrict ourselves to codes whose blocklength is an integer multiple of T, i.e., $n = LT$ for some L. An $(L, T, M, \epsilon, \rho)$ code consists of:

(1) An encoder f: $\{1, \cdots, M\} \to \mathbb{C}^{T\times n_t L}$ that maps a message A, which is uniformly distributed on $\{1, \cdots, M\}$, to a codeword $\mathbf{X}^L = [\mathbf{X}_1, \cdots, \mathbf{X}_L]$. The codewords are required to satisfy the power constraint

$$\|\mathbf{X}_l\|^2 \le T\rho, \quad l = 1, \cdots, L$$

where $\|\cdot\|$ denotes the Frobenius norm [14, Sec. 5.2]. Since the variances of $\mathbf{H}_l$ and $\mathbf{W}_l$ are normalized, ρ can be interpreted as the average SNR at the receiver.

(2) A decoder g: $\mathbb{C}^{T\times n_t L} \to \{1, \cdots, M\}$ satisfying a maximum error probability constraint

$$\max_{1\le a\le M} \Pr\left[g(\mathbf{Y}^L) \neq A | A = a\right] \le \epsilon$$

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

The maximum coding rate is defined as

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

where we denote by $\log(\cdot)$ the natural logarithm.

III. Main Result

A. High-SNR Normal Approximation

Theorem 1: Assume that $T \ge n_r + n_t$, $n_r \ge n_t$, and $0 < \epsilon < 1/2$. Then, at high SNR,

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

where $\mathcal{O}_L(\log L/L)$ comprises terms that are uniform in ρ and decay no slower than $\log L/L$; $o_\rho(1)$ comprises terms that are independent of L and vanish as $\rho \to \infty$; and

$$I(\rho) = -n_t n_r \log\left(1 + \frac{T\rho}{n_t}\right) + n_r n_t \log\frac{T\rho}{n_t} + \log\frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} + (T - n_t)\left(\sum_{i=0}^{n_t-1} \Psi(n_r - i) + n_t \log\frac{T\rho}{n_t}\right) - (T - n_t)n_t$$

$$\tilde{V} = n_t(T - n_t) + (T - n_t)^2 \sum_{i=0}^{n_t-1} \Psi'(n_r - i)$$

where $\Psi(\cdot)$ denotes Euler's digamma function, $\Psi'(\cdot)$ denotes its derivative [15, Sec. 8.36], and $\Gamma_{n_t}(\cdot)$ denotes the complex multivariate Gamma function [13, Eq. (3)].

Proof: See Sec. IV. ■

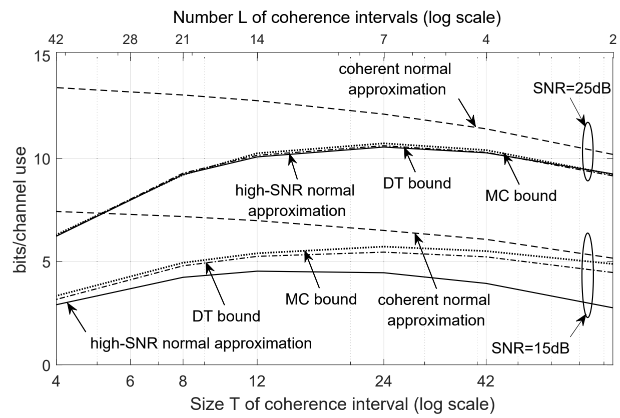

Fig. 1. Bounds on $R^*(L, T, \epsilon, \rho)$ for $n_t = n_r = 2$, $TL = 168$, $\epsilon = 10^{-3}$.

The quantity $I(\rho)/T$ is an asymptotically-tight lower bound on the capacity $C(\rho)$ of the non-coherent MIMO Rayleigh block-fading channel [13]. The ratio $\tilde{V}/T^2$ can be viewed as a high-SNR approximation of the channel dispersion. Observe that, at high SNR, $I(\rho)/T$ increases with n_r and $n_t \le T/2$, while $\tilde{V}/T^2$ is independent of the SNR and decreases with n_t and n_r.

For comparison, at high SNR, the capacity $C_c(\rho)$ of the coherent MIMO Rayleigh block-fading channel roughly behaves as $n_t \log(\rho)$ [16], and the channel dispersion converges to $n_t/T + \mathrm{Var}\big(\log\det(\mathbf{H}\mathbf{H}^H)\big)$ as $\rho \to \infty$ [6]. Noting that $I(\rho)/T$ roughly behaves as $(1 - n_t/T)n_t \log(\rho)$ and that $\tilde{V}/T^2$ can be written as $n_t(T - n_t)/T^2 + \big((T - n_t)/T\big)^2 \mathrm{Var}\big(\log\det(\mathbf{H}\mathbf{H}^H)\big)$, we observe that the high-SNR normal approximation presented in Theorem 1 corresponds to the normal approximation one obtains by transmitting one pilot symbol per coherence block for each transmit antenna to estimate the fading coefficient, and by then transmitting $T - n_t$ symbols over a coherent fading channel.

B. Numerical Results and Discussion

In Figs. 1–3, we depict the high-SNR normal approximation given in Theorem 1, the normal approximation of the coherent MIMO Rayleigh block-fading channel obtained in [6], a non-asymptotic (in ρ and L) lower bound on $R^*(L, T, \epsilon, \rho)$ that is based on the dependence testing (DT) bound, and an upper bound that is based on the meta converse (MC) bound obtained in [9] (both evaluated by the communication toolbox SPECTRE [17]).

In Fig. 1, we show $R^*(L, T, \epsilon, \rho)$ as a function of T with a fixed blocklength $n = LT = 168$ for $\rho = 15$dB and $\rho = 25$dB. Observe that the accuracy of the high-SNR normal approximation increases as the SNR increases. Moreover, the high-SNR normal approximation becomes more accurate as L increases. As expected, the normal approximation of the coherent setting is strictly larger than that of the non-coherent setting. The gap between two normal approximations becomes smaller as T increases, indicating that the cost for estimating the channel vanishes as T tends to infinity.

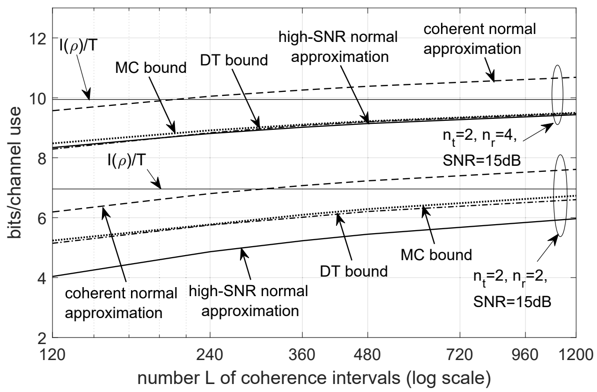

Fig. 2. Bounds on $R^*(L,T,\epsilon,\rho)$ for $T=24$, $\epsilon=10^{-3}$.

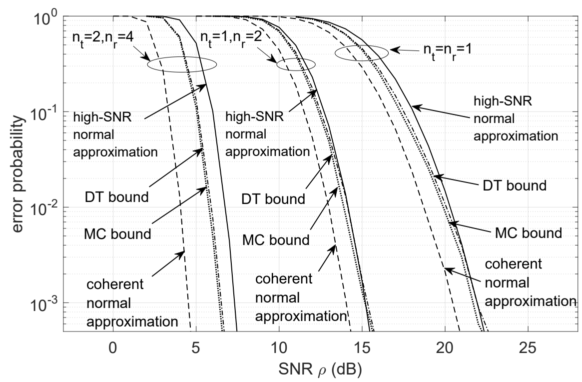

Fig. 3. Bounds on ϵ^* for $R=4$, $T=24$, $L=7$.

In Fig. 2, we show $R^*(L,T,\epsilon,\rho)$ as a function of the blocklength $n = LT$ for $T = 24$, $\rho = 15$dB, and $(n_t,n_r) \in \{(2,2),(2,4)\}$. Observe that the accuracy of the high-SNR normal approximation increases with n_r. Moreover, the gap between the high-SNR normal approximation and the coherent normal approximation appears to be independent of L. This agrees with the intuition that the cost for estimating the channel depends only on the coherence interval T.

In Fig. 3, we plot the minimum error probability ϵ^* as a function of the SNR for $R = 4$, $n = 168$ ($T = 24$ and $L = 7$), and $(n_t,n_r) \in \{(1,1),(1,2),(2,4)\}$. When the number of antennas increases, the diversity gain becomes larger, which is reflected in the slope of the plots. Observe that the error probability decreases significantly as the number of antennas increases. Further observe that the high-SNR normal approximation is accurate for $(n_t,n_r) \in \{(1,1),(1,2)\}$ and is overly pessimistic for $(n_t,n_r) = (2,4)$. Intuitively, for $(n_t,n_r) = (2,4)$, the SNR required to achieve a given error probability at a given rate is smaller than that in the other two cases, so the $o_\rho(1)$ terms in the high-SNR normal approximation are larger. In contrast, the coherent normal approximation is overly optimistic for all parameters considered in this figure.

IV. Proof of Theorem 1

The proof of Theorem 1 is based on a lower bound and an upper bound on $R^*(L,T,\epsilon,\rho)$. Due to space limitations, we only present an outline of the proof and defer the details to the longer version of the paper [18].

A. Dependence Testing Lower Bound

To derive a lower bound on $R^*(L,T,\epsilon,\rho)$, we evaluate the DT lower bound [2, Th. 22] for a USTM input distribution. The DT bound states that there exists a code of blocklength $n = LT$ with M codewords and a maximum error probability satisfying

$$\begin{aligned}\epsilon \leq &(M-1)\mathbb{E}\Big[e^{-i(\mathbf{X}^L;\mathbf{Y}^L)} \cdot \mathbf{1}\big\{i(\mathbf{X}^L;\mathbf{Y}^L) > \log(M-1)\big\}\Big] \\ &+ P\big[i(\mathbf{X}^L;\mathbf{Y}^L) \leq \log(M-1)\big] \end{aligned} \tag{3}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function and $i(\mathbf{X}^L;\mathbf{Y}^L)$ is the information density between $\mathbf{X}^L$ and $\mathbf{Y}^L$, given by

$$i(\mathbf{X}^L;\mathbf{Y}^L) = \sum_{l=1}^{L} \log \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}_l|\mathbf{X}_l)}{f_{\mathbf{Y}}(\mathbf{Y}_l)}.$$

Here $f_{\mathbf{Y}|\mathbf{X}}$ denotes the conditional pdf of the channel output $\mathbf{Y}_l$ of (2) given the input $\mathbf{X}_l$ and $f_{\mathbf{Y}}$ denotes the output pdf induced by (2) with USTM channel inputs. In the following, we denote by $i(\rho)$ the random variable $\log\big(f_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}_l|\mathbf{X}_l)/f_{\mathbf{Y}}(\mathbf{Y}_l)\big)$ conditioned on $\|\mathbf{X}_l\|^2 = T\rho$. We further define $I(\rho) \triangleq \mathbb{E}[i(\rho)]$ and $U(\rho) \triangleq \mathbb{E}\big[\big(i(\rho) - I(\rho)\big)^2\big]$. Here and throughout the paper, we omit the subscript l where it is immaterial.

To bound the right-hand side (RHS) of (3), we follow [2, Eqs. (258)–(267)] (see also [7, Eqs. (69)–(85)]). In particular, to ensure that the Berry-Esseen ratio

$$B(\rho) \triangleq \frac{6\mathbb{E}\Big[\big|i(\rho) - I(\rho)\big|^3\Big]}{U(\rho)^{3/2}}$$

is uniformly bounded in L, we show in [18] that, for sufficiently large $\rho_0 > 0$,

$$\begin{aligned} U(\rho) \geq& \tilde{V}/2, \quad \rho \geq \rho_0 \\ \sup_{\rho \geq \rho_0} \mathbb{E}\Big[\big|i(\rho) - I(\rho)\big|^3\Big] <& \infty. \end{aligned}$$

Thus, for $\rho \geq \rho_0$ and sufficiently large ρ_0, $B(\rho)$ is upper-bounded by a finite constant that depends only on ρ_0 and T. It follows by the Berry-Esseen theorem and steps similar to [7, Eqs. (71)–(74)] that

$$R^*(L,T,\epsilon,\rho) \geq \frac{I(\rho)}{T} - \sqrt{\frac{U(\rho)}{LT^2}}Q^{-1}(\epsilon) + \mathcal{O}\Big(\frac{1}{L}\Big). \tag{4}$$

Since, as we show in [18], we have $U(\rho) = \tilde{V} + o_\rho(1)$, this demonstrates that the high-SNR normal approximation in Theorem 1 is achievable.

B. Meta Converse Upper Bound

To derive an upper bound on $R^*(L,T,\epsilon,\rho)$, we first introduce the mismatched information density

$$j(\mathbf{X}^L;\mathbf{Y}^L) \triangleq \sum_{l=1}^{L} \log \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}_l|\mathbf{X}_l)}{q_{\mathbf{Y}}(\mathbf{Y}_l)}$$

where $q_{\mathbf{Y}}$ is an auxiliary output pdf. In this paper, we choose the pdf $q_{\mathbf{Y}}$ that was chosen in [13, Eq. (25)] to derive a converse bound on the channel capacity of non-coherent MIMO Rayleigh block-fading channels.

Let $\mathbf{D}_l$ be a diagonal matrix with entries $(d_1, \ldots, d_{n_t})$, where $d_i^2 \triangleq \|(\mathbf{x}_l)_i\|^2$ denotes the power at transmit antenna i and where we use $(\mathbf{x}_l)_i$ to denote the i-th column of $\mathbf{x}_l$. Without loss of generality, we assume that $\text{tr}(\mathbf{D}_l^2) = T\alpha_l$, $\alpha_l \in [0, \rho]$, for each coherent block.

We next follow [6] and separate the codebook $\mathcal{C}$ into two sub-codebooks $\mathcal{C}_1$ and $\mathcal{C}_2$. Clearly, if the maximum error probabilities of $\mathcal{C}_1$ and $\mathcal{C}_2$ are ϵ, then the maximum error probability of $\mathcal{C}$ cannot be smaller than ϵ. Specifically, $\mathcal{C}_1$ contains all the codewords for which $d_i^2 > \bar{\delta}\rho$, $i = 1, \cdots, n_t$, (with $\bar{\delta}$ defined below) in at least half of the coherence intervals. It follows that the Berry-Esseen ratio $\bar{B}(\mathbf{D}_l)$ (see (7) below) can be upper-bounded by a positive value, so the Berry-Esseen theorem can be applied to derive an upper bound. For $\mathcal{C}_2$, which contains the remaining codewords, an upper bound is obtained using Chebyshev's inequality. We then prove that, as ρ and L tend to infinity, the cardinality of the entire codebook $\mathcal{C}$ can be approximated by the cardinality of $\mathcal{C}_1$. Consequently, $R^*(L, T, \epsilon, \rho)$ is asymptotically upper-bounded by the upper bound on the maximum coding rate of $\mathcal{C}_1$.

To define $\mathcal{C}_1$ and $\mathcal{C}_2$ mathematically, we first define the sets

$$\mathcal{D}_1 \triangleq \left\{\mathbf{D} \in \mathbb{D}_{n_t} : d_{i,i}^2 > \bar{\delta}\rho,\ i \in [1, n_t]\right\}$$
$$\mathcal{D}_2 \triangleq \mathbb{D}_{n_t} \setminus \mathcal{D}_1$$

where $\mathbb{D}_{n_t}$ denotes the set of $(n_t \times n_t)$-dimensional diagonal matrices with non-negative, real-valued entries; and $d_{i,i}$ denotes the i-th diagonal element of $\mathbf{D}$. Moreover,

$$\bar{\delta} = \frac{T}{n_t} - \frac{T}{2n_r\sqrt{n_t}\sqrt{\mathbb{E}\left[\left(\log\det(\mathbf{H}\mathbf{H}^H)\right)^2\right] + 1}}.$$

Let $L_{\mathcal{D}_1}(\mathbf{D}^L) \triangleq \sum_{l=1}^{L} \mathbf{1}\{\mathbf{D}_l \in \mathcal{D}_1\}$. Then, we define

$$\mathcal{C}_1 \triangleq \left\{\mathbf{x}^L \in \mathcal{C} : L_{\mathcal{D}_1}(\mathbf{D}^L) \geq L/2\right\}$$
$$\mathcal{C}_2 \triangleq \left\{\mathbf{x}^L \in \mathcal{C} : L_{\mathcal{D}_1}(\mathbf{D}^L) < L/2\right\}.$$

1) The cardinality of $\mathcal{C}_1$: An upper bound on $\log|\mathcal{C}_1|$ follows from the MC bound [2, Th. 31] and by an upper bound on $\log\left(f_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}_l|\mathbf{X}_l)/q_{\mathbf{Y}}(\mathbf{Y}_l)\right)$ which, conditioned on $\text{diag}\left\{\|(\mathbf{X}_l)_1\|, \cdots, \|(\mathbf{X}_l)_{n_t}\|\right\} = \mathbf{D}_l$, is denoted as $\bar{j}(\mathbf{D}_l)$ [18]:

$$\log|\mathcal{C}_1| \leq \sup_{\mathbf{x}^L \in \mathcal{C}_1} \Bigg\{ \log\xi(\alpha) - \log\left(1 - \epsilon - \Pr\left[\sum_{l=1}^{L} \bar{j}(\mathbf{D}_l) \geq \log\xi(\alpha)\right]\right)\Bigg\} \quad (5)$$

for every $\xi : [0, \rho]^L \to (0, \infty)$. Let $\bar{J}(\mathbf{D}_l) \triangleq \mathbb{E}[\bar{j}(\mathbf{D}_l)]$ and $\bar{V}(\mathbf{D}_l) \triangleq \mathbb{E}[(\bar{j}(\mathbf{D}_l) - \bar{J}(\mathbf{D}_l))^2]$. It can be shown that, for sufficiently large ρ_0, $\sup_{\rho \geq \rho_0} \mathbb{E}\left[\left|\bar{j}(\mathbf{D}_l) - \bar{J}(\mathbf{D}_l)\right|^3\right]$ is finite and depends only on ρ_0 and T, and that

$$\sum_{l:\ \mathbf{D}_l \in \mathcal{D}_1} \bar{V}(\mathbf{D}_l) \geq \frac{L}{2}\left[\frac{(T - n_t)n_t}{2}\right] \quad (6)$$

for $\rho \geq \rho_0$ and sufficiently large ρ_0 [18]. It follows that, for such ρ and ρ_0, the Berry-Esseen ratio satisfies

$$\bar{B}(\mathbf{D}_l) \triangleq \frac{6\sum_{l=1}^{L} \mathbb{E}\left[\left|\bar{j}(\mathbf{D}_l) - \bar{J}(\mathbf{D}_l)\right|^3\right]}{\left(\sum_{l=1}^{L} \bar{V}(\mathbf{D}_l)\right)^{3/2}} \leq \frac{\bar{B}(\rho_0)}{\sqrt{L}} \quad (7)$$

for some finite $\bar{B}(\rho_0)$ that depends only on ρ_0 and T.

We next choose $\lambda = Q^{-1}\left(\epsilon + 2\bar{B}(\rho_0)/\sqrt{L}\right)$ and

$$\log\xi(\alpha) = \sum_{l=1}^{L} \bar{J}(\mathbf{D}_l) - \lambda\sqrt{\sum_{l=1}^{L} \bar{V}(\mathbf{D}_l)}$$

and apply the Berry-Esseen theorem to (5) to obtain

$$\begin{aligned}
&\frac{\log|\mathcal{C}_1|}{L} \\
&\leq \frac{1}{L}\sup_{\mathbf{x}^L \in \mathcal{C}_1}\left\{\sum_{l=1}^{L} \bar{J}(\mathbf{D}_l) - \sqrt{\sum_{l=1}^{L} \bar{V}(\mathbf{D}_l)}Q^{-1}\left(\epsilon + \frac{2\bar{B}(\rho_0)}{\sqrt{L}}\right)\right\} \\
&\quad - \frac{\log\bar{B}(\rho_0)}{L} + \frac{\log L}{2L} \\
&\leq \frac{1}{L}\sup_{\mathbf{x}^L \in \mathcal{C}_1}\sum_{l=1}^{L}\left\{\bar{J}(\mathbf{D}_l) - \sqrt{\frac{\bar{V}(\mathbf{D}_l)}{L}}Q^{-1}(\epsilon)\right\} + \mathcal{O}_L\left(\frac{\log L}{L}\right)
\end{aligned} \quad (8)$$

where the $\mathcal{O}_L(\log L/L)$ term is uniform in ρ and decays no slower than $\log L/L$. Here, the second inequality follows by performing a Taylor series expansion of the inverse Q-function around ϵ and from Jensen's inequality applied to the square-root function. Since, in each block, $\mathbf{x}_l$ independently satisfies the power constraint $\|\mathbf{x}_l\|^2 \leq T\rho$, the supremum on the RHS of (8) can be written as

$$\sum_{l:\ \mathbf{D}_l \in \mathcal{D}_1} \sup_{\mathbf{D} \in \mathcal{D}_1}\left\{\bar{J}(\mathbf{D}) - \sqrt{\frac{\bar{V}(\mathbf{D})}{L}}Q^{-1}(\epsilon)\right\} + \sum_{l:\ \mathbf{D}_l \in \mathcal{D}_2} \sup_{\mathbf{D} \in \mathcal{D}_2}\left\{\bar{J}(\mathbf{D}) - \sqrt{\frac{\bar{V}(\mathbf{D})}{L}}Q^{-1}(\epsilon)\right\} \quad (9)$$

maximized over all sequences $\mathbf{D}^L$ satisfying $L_{\mathcal{D}_1}(\mathbf{D}^L) \geq L/2$. It can be shown that, for every $\mathbf{D} \in \mathcal{D}_1$ with $\text{tr}(\mathbf{D}^2) = T\alpha$, [18]

$$\bar{J}(\mathbf{D}) \leq \bar{J}\left(\sqrt{\frac{\rho}{\alpha}}\mathbf{D}\right), \quad \bar{V}(\mathbf{D}) \geq \bar{V}\left(\sqrt{\frac{\rho}{\alpha}}\mathbf{D}\right) - \frac{\Upsilon(T)}{L}$$

where $\Upsilon(T)$ is a non-negative constant that depends only on T. Defining $\tilde{\mathcal{D}}_1 \triangleq \{\mathbf{D} \in \mathcal{D}_1 \colon \text{tr}(\mathbf{D}^2) = T\rho\}$, it follows that

$$\begin{aligned}
&\sup_{\mathbf{D} \in \mathcal{D}_1}\left\{\bar{J}(\mathbf{D}) - \sqrt{\frac{\bar{V}(\mathbf{D})}{L}}Q^{-1}(\epsilon)\right\} \\
&\leq \sup_{\mathbf{D} \in \tilde{\mathcal{D}}_1}\left\{\bar{J}(\mathbf{D}) - \sqrt{\frac{\bar{V}(\mathbf{D}) - \Upsilon(T)/L}{L}}Q^{-1}(\epsilon)\right\} \\
&= \sup_{\mathbf{D} \in \tilde{\mathcal{D}}_1}\left\{\bar{J}(\mathbf{D}) - \sqrt{\frac{\bar{V}(\mathbf{D})}{L}}Q^{-1}(\epsilon)\right\} + \mathcal{O}_L\left(\frac{1}{L}\right)
\end{aligned} \quad (10)$$

where the $\mathcal{O}_L(1/L)$ term is uniform in ρ and decays no slower than $1/L$. To optimize over $\mathbf{D}$, we prove in [18] that, for every $\mathbf{D} \in \tilde{\mathcal{D}}_1$, $\bar{V}(\mathbf{D})$ can be approximated as

$$\bar{V}(\mathbf{D}) = \tilde{V} + K_{\bar{V}}(T, \mathbf{D}) \tag{11}$$

for $\rho \geq \rho_0$ and sufficiently large ρ_0, where $\tilde{V}$ is given in Theorem 1 and $\lim_{\rho\to\infty} \sup_{\mathbf{D}\in\tilde{\mathcal{D}}_1} |K_{\bar{V}}(T, \mathbf{D})| = 0$. Consequently, defining $K_{\bar{V}}(T) \triangleq \sup_{\mathbf{D}\in\tilde{\mathcal{D}}_1} |K_{\bar{V}}(T, \mathbf{D})|$, the RHS of (10) can be upper-bounded by

$$\sup_{\mathbf{D}\in\tilde{\mathcal{D}}_1} \bar{J}(\mathbf{D}) - \sqrt{\frac{\tilde{V} - K_{\bar{V}}(T)}{L}} Q^{-1}(\epsilon) + \mathcal{O}_L\Big(\frac{1}{L}\Big). \tag{12}$$

It can be shown that the supremum over $\tilde{\mathcal{D}}_1$ is achieved when $d_i^2 = T\rho/n_t, i \in [1, n_t]$, and we denote the resulting value of $\bar{J}(\mathbf{D})$ by $\tilde{J}(\rho)$ [18]. Combining (10) and (12), we thus obtain that

$$\sup_{\mathbf{D}\in\mathcal{D}_1} \left\{ \bar{J}(\mathbf{D}) - \sqrt{\frac{\bar{V}(\mathbf{D})}{L}} Q^{-1}(\epsilon) \right\} \leq \tilde{J}(\rho) - \sqrt{\frac{\tilde{V} - K_{\bar{V}}(T)}{L}} Q^{-1}(\epsilon) + \mathcal{O}_L\Big(\frac{1}{L}\Big). \tag{13}$$

Regarding the second sum in (9), it can be shown that, for $L \geq L_0$, $\rho \geq \rho_0$, and sufficiently large L_0 and ρ_0,

$$\sup_{\mathbf{D}\in\mathcal{D}_2} \left\{ \bar{J}(\mathbf{D}) - \sqrt{\frac{\bar{V}(\mathbf{D})}{L}} Q^{-1}(\epsilon) \right\} \leq \tilde{J}(\rho) - \sqrt{\frac{\tilde{V} - K_{\bar{V}}(T)}{L}} Q^{-1}(\epsilon) + \mathcal{O}_L\Big(\frac{1}{L}\Big). \tag{14}$$

To obtain (14), we first prove that [18]

$$\sup_{\mathbf{D}\in\mathcal{D}_2} \bar{J}(\mathbf{D}) \leq \sup_{\mathbf{D}\in\tilde{\mathcal{D}}_2} \bar{J}(\mathbf{D})$$

where $\tilde{\mathcal{D}}_2 \triangleq \{\mathbf{D} \in \mathcal{D}_2 \colon \mathrm{tr}(\mathbf{D}^2) = T\rho\}$. We then show that, for $L \geq L_0$ and sufficiently large L_0, [18]

$$\tilde{J}(\rho) - \sup_{\mathbf{D}\in\tilde{\mathcal{D}}_2} \bar{J}(\mathbf{D}) \geq \tau$$

for some positive constant τ that is independent of ρ and L. Lower-bounding $\bar{V}(\mathbf{D}) \geq 0$, it thus follows that

$$\begin{aligned}
&\tilde{J}(\rho) - \sqrt{\frac{\tilde{V} - K_{\bar{V}}(T)}{L}} Q^{-1}(\epsilon) + \mathcal{O}_L\Big(\frac{1}{L}\Big) \\
&\quad - \sup_{\mathbf{D}\in\tilde{\mathcal{D}}_2} \left\{ \bar{J}(\mathbf{D}) - \sqrt{\frac{\bar{V}(\mathbf{D})}{L}} Q^{-1}(\epsilon) \right\} \\
&\geq \tilde{J}(\rho) - \sqrt{\frac{\tilde{V} - K_{\bar{V}}(T)}{L}} Q^{-1}(\epsilon) + \mathcal{O}_L\Big(\frac{1}{L}\Big) - \sup_{\mathbf{D}\in\tilde{\mathcal{D}}_2} \bar{J}(\mathbf{D}) \\
&\geq \tau - \sqrt{\frac{\tilde{V} - K_{\bar{V}}(T)}{L}} Q^{-1}(\epsilon) + \mathcal{O}_L\Big(\frac{1}{L}\Big)
\end{aligned}$$

which is strictly positive for $L \geq L_0$, $\rho \geq \rho_0$, and sufficiently large L_0 and ρ_0, so (14) follows. Thus, combining (13) and (14) with (9) yields for $L \geq L_0$ and $\rho \geq \rho_0$

$$\frac{\log|\mathcal{C}_1|}{L} \leq \tilde{J}(\rho) - \sqrt{\frac{\tilde{V} - K_{\bar{V}}(T)}{L}} Q^{-1}(\epsilon) + \mathcal{O}_L\Big(\frac{\log L}{L}\Big). \tag{15}$$

2) The cardinality of $\mathcal{C}_2$: To bound the cardinality of $\mathcal{C}_2$, we show in [18] that $\bar{V}(\mathbf{D}) \leq \bar{K}(T)$ for every $\mathbf{D}$, where $\bar{K}(T)$ is a positive constant that depends only on T. Chebyshev's inequality then yields that

$$\log|\mathcal{C}_2| \leq \sup_{\mathbf{x}^L\in\mathcal{C}_2} \sum_{l=1}^{L} \bar{J}(\mathbf{D}_l) + \sqrt{\frac{L\bar{K}(T)}{2(1-\epsilon)}} + \log(1-\epsilon).$$

Similar to the analysis of $\log|\mathcal{C}_1|$, it can be shown that [18]

$$\begin{aligned}
&\sup_{\mathbf{x}^L\in\mathcal{C}_2} \sum_{l=1}^{L} \bar{J}(\mathbf{D}_l) \\
&\quad\leq \sup_{D^L\colon L_{\mathcal{D}_1}(\mathcal{D}^L)<L/2} \Bigg\{ \sum_{l\colon \mathbf{D}_l\in\mathcal{D}_1} \sup_{\mathbf{D}\in\tilde{\mathcal{D}}_1} \bar{J}(\mathbf{D}) \\
&\qquad\qquad + \sum_{l\colon \mathbf{D}_l\in\mathcal{D}_2} \sup_{\mathbf{D}\in\tilde{\mathcal{D}}_2} \bar{J}(\mathbf{D}) \Bigg\} \\
&\quad\leq \sup_{D^L\colon L_{\mathcal{D}_1}(\mathbf{D}^L)<L/2} \Big\{ L_{\mathcal{D}_1}(\mathbf{D}^L)\tilde{J}(\rho) \\
&\qquad\qquad + (L - L_{\mathcal{D}_1}(\mathbf{D}^L))(\tilde{J}(\rho) - \tau) \Big\} \\
&\quad< L\tilde{J}(\rho) - \frac{L}{2}\tau.
\end{aligned}$$

It follows that

$$\frac{\log|\mathcal{C}_2|}{L} < \tilde{J}(\rho) - \frac{1}{2}\tau + \sqrt{\frac{\bar{K}(T)}{2L(1-\epsilon)}} + \frac{\log(1-\epsilon)}{L}. \tag{16}$$

3) The cardinality of $\mathcal{C}$: Denote the RHSs of (15) and (16) by κ_1 and κ_2, respectively. It follows that

$$\log|\mathcal{C}| \leq \log e^{L\kappa_1} + \log\big(1 + e^{L(\kappa_2-\kappa_1)}\big). \tag{17}$$

By the behavior of $K_{\bar{V}}(T, \mathbf{D})$ in (11), we have that $\tilde{V} + K_{\bar{V}}(T) \leq 2\tilde{V}$ for $\rho \geq \rho_0$ and sufficiently large ρ_0. Thus,

$$\kappa_2 - \kappa_1 \leq -\frac{1}{2}\tau + \sqrt{\frac{\bar{K}(T)}{2L(1-\epsilon)}} + \sqrt{\frac{2\tilde{V}}{L}} Q^{-1}(\epsilon) + \mathcal{O}_L\Big(\frac{\log L}{L}\Big)$$

which tends to $-\tau/2$ as $L \to \infty$. Applied to (17), this implies that $\log|\mathcal{C}|/L \leq \kappa_1 + o_L(1/L)$, where the $o_L(1/L)$ term is uniform in $\rho \geq \rho_0$ and vanishes faster than $1/L$. Consequently,

$$R^*(L, T, \epsilon, \rho) \leq \frac{\tilde{J}(\rho)}{T} - \sqrt{\frac{\tilde{V} - K_{\bar{V}}(T)}{LT^2}} Q^{-1}(\epsilon) + \mathcal{O}_L\Big(\frac{\log L}{L}\Big). \tag{18}$$

Comparing the lower bound (4) with the converse bound (18), and using that, as shown in [18], $\tilde{J}(\rho) = I(\rho) + o_\rho(1)$, we obtain Theorem 1.

REFERENCES

[1] G. Durisi, T. Koch, and P. Popovski, “Toward massive, ultrareliable, and low-latency wireless communication with short packets,” *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.

[2] Y. Polyanskiy, H. V. 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] Y. Polyanskiy and S. Verdú, “Scalar coherent fading channel: Dispersion analysis,” in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, pp. 2959–2963, Jul. 2011.

[4] 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, Jul. 2014.

[5] A. Collins and Y. Polyanskiy, “Dispersion of the coherent MIMO block-fading channel,” in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, pp. 1068–1072, Jul. 2016.

[6] A. Collins and Y. Polyanskiy, “Coherent multiple-antenna block-fading channels at finite blocklength,” *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 380–405, Jan. 2019.

[7] A. Lancho, T. Koch, and G. Durisi, “On single-antenna Rayleigh block-fading channels at finite blocklength,” *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 496–519, Jan. 2020.

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

[9] 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.

[10] A. Lancho, J. Östman, G. Durisi, T. Koch, and G. Vazquez-Vilar, “Saddlepoint approximations for noncoherent single-antenna rayleigh block-fading channels,” in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, pp. 612–616, Jul. 2019.

[11] A. Lancho, J. Östman, G. Durisi, T. Koch, and G. Vazquez-Vilar, “Saddlepoint approximations for short-packet wireless communications,” *IEEE Trans. Wireless Commun.*, to be published. doi: 10.1109/TWC.2020.2987573.

[12] L. Zheng and D. N. C. 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, Jan. 2013.

[14] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 1990.

[15] A. Jeffrey and D. Zwillinger, *Table of integrals, series, and products*. Elsevier, 2007.

[16] E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Trans. Telecom.*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999.

[17] Spectre, *SPECTRE: short packet communication toolbox*, 2015, GitHub repository. [Online]. Available: https://github.com/yp-mit/spectre.

[18] C. Qi and T. Koch, “On MIMO Rayleigh blockfading channels at finite blocklengths,” *preprint*. [Online]. Available: http://www.tsc.uc3m.es/~koch/Publications.html.