

This is a postprint version of the following published document:

Lancho, A.; Koch, T.; Durisi, G. On Single-Antenna Rayleigh Block-Fading Channels at Finite Blocklength, *in IEEE Transactions on Information Theory*, 66(1), Jan. 2020, Pp. 496-519

DOI: <https://doi.org/10.1109/TIT.2019.2945782>

© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

# On Single-Antenna Rayleigh Block-Fading Channels at Finite Blocklength

Alejandro Lancho, *Member, IEEE*, Tobias Koch, *Senior Member, IEEE*, and Giuseppe Durisi, *Senior Member, IEEE*

**Abstract**—This paper concerns the maximum coding rate at which data can be transmitted over a noncoherent, 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. A high-SNR normal approximation of the maximum coding 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 at SNR values above 15 dB and when the number of coherence intervals is 10 or more.

## 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 next generations of cellular systems 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 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 a determined length  $n$  with a block-error probability no larger than  $\epsilon$ . Building upon Dobrushin's and Strassen's asymptotic results, Hayashi [2] and Polyanskiy, Poor and Verdú [3] showed that for various channels with a positive capacity  $C$ , the maximum coding rate can be tightly approximated by

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

where  $V$  denotes the channel dispersion [3, Def. 1],  $Q^{-1}(\epsilon)$  denotes the inverse of the Gaussian  $Q$ -function

$$Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (2)$$

and  $\mathcal{O}(\log n/n)$  comprises terms that decay no slower than  $\log n/n$ . The approximation that follows from (1) by ignoring the  $\mathcal{O}(\log n/n)$  term is sometimes referred to as *normal approximation*.

The work by Polyanskiy *et al.* [3] has been generalized to some wireless communication channels. For instance, the channel dispersion of coherent fading channels—where the receiver has perfect knowledge of the realizations of the fading coefficients—was studied by Polyanskiy and Verdú for the single-antenna case [4], and by Collins and Polyanskiy for the multiple-input single-output (MISO) [5] and the multiple-input multiple-output (MIMO) case [6], [7]. The channel dispersion of single-antenna quasistatic fading channels when both transmitter and receiver have perfect knowledge of the realization of the fading coefficients and the transmitter satisfies a long-term power constraint was obtained by Yang *et al.* [8]. In the noncoherent setting—where neither the transmitter nor the receiver have *a priori* knowledge of the realizations of the fading coefficients—the channel dispersion is only known in the quasistatic case, where it is zero [9], [10]. Upper and lower bounds on the second-order coding rate of quasistatic MIMO Rayleigh-fading channels have further been reported in [11] for the asymptotically-ergodic setup where the number of antennas grows linearly with the blocklength. For noncoherent Rayleigh block-fading channels, nonasymptotic bounds on the maximum coding rate were presented by Yang *et al.* for the

A. Lancho and T. Koch have received funding from the Spanish Ministerio de Economía y Competitividad under Grants TEC2013-41718-R and TEC2016-78434-C3-3-R (AEI/FEDER, EU), from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 714161), and from the Comunidad de Madrid under Grant S2103/ICE-2845. A. Lancho has also received funding from an FPU fellowship from the Spanish Ministerio de Educación, Cultura y Deporte under Grant FPU14/01274. T. Koch has also received funding from the Spanish Ministerio de Economía y Competitividad under Grant RYC-2014-16332 and from the 7th European Union Framework Programme under Grant 333680. G. Durisi has been supported by the Swedish Research Council under Grants 2012-4571 and 2016-03293. The material in this paper was presented in part at the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, June 2017, and at the 52nd Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, March 2018.

A. Lancho was with the Signal Theory and Communications Department, Universidad Carlos III de Madrid, 28911, Leganés, Spain and also with the Gregorio Marañón Health Research Institute, 28007, Madrid, Spain. He is now with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, 41296, Sweden (e-mail: lancho@ieee.org).

T. Koch is with the Signal Theory and Communications Department, Universidad Carlos III de Madrid, 28911, Leganés, Spain and also with the Gregorio Marañón Health Research Institute, 28007, Madrid, Spain (e-mail: koch@tsc.uc3m.es).

G. Durisi is with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg 41296, Sweden (e-mail: durisi@chalmers.se).

Copyright (c) 2019 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

single-antenna case [12] and by Östman *et al.* for the MIMO case [13], [14]. For further references see [1].

In a nutshell, in the noncoherent setting the channel dispersion is only known in the quasistatic case. For general block-fading channels, the maximum coding rate needs to be assessed by means of nonasymptotic bounds, whose evaluation is often computationally demanding. Obtaining an expression for the channel dispersion of noncoherent block-fading channels is difficult because for such channels the capacity-achieving input distribution is in general unknown. Thus, the standard approach of obtaining expressions of the form (1), which consists of first evaluating nonasymptotic upper and lower bounds on  $R^*(n, \epsilon)$  for the capacity-achieving input and output distributions and then analyzing these bounds in the limit as  $n \rightarrow \infty$ , cannot be followed. However, the behavior of capacity at high signal-to-noise ratio (SNR) is well understood. Indeed, it was demonstrated that an input distribution called *unitary space-time modulation* (USTM) achieves a lower bound on the capacity that is asymptotically tight [15], [16], [17]. Thus, a characterization of the channel dispersion at high SNR may be feasible.

In this paper, we present an expression similar to (1) of the maximum coding rate  $R^*(L, T, \epsilon, \rho)$  achievable over noncoherent, single-antenna, Rayleigh block-fading channels using error-correcting codes that span  $L$  coherence intervals of length  $T$ , have a block-error probability no larger than  $\epsilon$ , and satisfy the power constraint  $\rho$ . By replacing the capacity and channel dispersion by asymptotically tight approximations, we obtain a high-SNR normal approximation of  $R^*(L, T, \epsilon, \rho)$ . The obtained normal approximation is useful in two ways. On the one hand, it complements the nonasymptotic bounds provided in [12], [13], [14]. On the other hand, it allows for a mathematical analysis of  $R^*(L, T, \epsilon, \rho)$ .

The rest of this paper is organized as follows. Section II introduces the notation used in this paper. Section III presents the system model. Section IV introduces the most important quantities used in this paper. Section V is divided into three subsections. The first subsection presents the main result of the paper: a high-SNR normal approximation of  $R^*(L, T, \epsilon, \rho)$ . The second subsection discusses the accuracy of the normal approximation by means of numerical evaluations. The third subsection discusses some applications of our normal approximation. Section VI contains the proof of the main result. Section VII concludes the paper with a discussion of the presented results. Some of the proofs are deferred to the appendices.

## II. NOTATION

We denote scalar random variables by upper case letters such as  $X$ , and their realizations by lower case letters such as  $x$ . Likewise, we use boldface upper case letters to denote random vectors, *i.e.*,  $\mathbf{X}$ , and we use boldface lower case letters such as  $\mathbf{x}$  to denote their realizations. We use upper case letters with the standard font to denote distributions, and lower case letters with the standard font to denote probability density functions (pdfs). The serif font is used to denote constants independent of  $L$  and  $\rho$ , except  $E[\cdot]$ , which denotes the expectation operator, and  $P[\cdot]$ , which is used for probabilities. The

superscripts  $(\cdot)^T$  and  $(\cdot)^H$  denote transposition and Hermitian transposition, respectively. The complement of a set  $\mathcal{A}$  is denoted as  $\mathcal{A}^c$ . We use “ $\stackrel{\mathcal{L}}{=}$ ” to denote equality in distribution.

We denote by  $\log(\cdot)$  the natural logarithm, by  $I\{\cdot\}$  the indicator function, by  $\Gamma(\cdot)$  the Gamma function [18, Sec. 6.1.1], by  $\tilde{\gamma}(\cdot, \cdot)$  the regularized lower incomplete gamma function [18, Sec. 6.5], by  $\psi(\cdot)$  the digamma function [18, Sec. 6.3.2], by  ${}_2F_1(\cdot, \cdot; \cdot; \cdot)$  the Gauss hypergeometric function [19, Sec. 9.1], by  $E_1(\cdot)$  the exponential integral function [18, Sec. 5.1.1], and by  $\zeta(z, q)$  Riemann's zeta function [19, Sec. 9.511]. The gamma distribution with parameters  $z$  and  $q$  is denoted by  $\text{Gamma}(z, q)$ . We denote by  $\gamma \approx 0.5772$  Euler's constant.

Double limits such as

$$\lim_{\substack{L \rightarrow \infty, \\ \rho \rightarrow \infty}} f(L, \rho) = K \quad (3)$$

indicate that for every  $\epsilon > 0$  there exists a pair  $(L_0, \rho_0)$  independent of  $(L, \rho)$  such that for every  $L \geq L_0$  and  $\rho \geq \rho_0$  we have  $|f(L, \rho) - K| \leq \epsilon$ . We denote by  $\underline{\lim}$  the *limit inferior* and by  $\overline{\lim}$  the *limit superior*. *Double limit inferiors* and *double limit superiors* are defined accordingly using the above definition of a double limit. For example,

$$\underline{\lim}_{\substack{L \rightarrow \infty, \\ \rho \rightarrow \infty}} f(L, \rho) = \lim_{L_0 \rightarrow \infty} \inf_{L \geq L_0} \inf_{\rho \geq \rho_0} f(L, \rho). \quad (4)$$

## III. SYSTEM MODEL

We consider a single-antenna Rayleigh block-fading channel with coherence interval  $T > 2$ . For this channel model, the input-output relation within the  $\ell$ -th coherence interval is given by

$$\mathbf{Y}_\ell = H_\ell \mathbf{X}_\ell + \mathbf{W}_\ell \quad (5)$$

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 the additive noise, which is assumed to be a random vector with independent and identically distributed (i.i.d.), zero-mean, unit-variance, circularly-symmetric, complex Gaussian entries; and  $H_\ell$  is Rayleigh fading, *i.e.*, it is a zero-mean, unit-variance, circularly-symmetric, complex Gaussian random variable. We assume that  $H_\ell$  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 noncoherent setting where transmitter and receiver are aware of the distribution of  $H_\ell$  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 coherence intervals of length  $T$  needed to transmit the whole code. An  $(M, L, T, \epsilon, \rho)$  code for the channel (5) consists of the following:

- 1) An encoder  $f: \{1, \dots, M\} \rightarrow \mathbb{C}^{LT}$  that maps the message  $A$ , taking value in  $\{1, \dots, M\}$ , to a codeword

$\mathbf{X}^L = [\mathbf{X}_1, \dots, \mathbf{X}_L]$ . The codewords are assumed to satisfy the power constraint<sup>1</sup>

$$\|\mathbf{X}_\ell\|^2 \leq T\rho, \quad \ell = 1, \dots, L. \quad (6)$$

Since the variance of  $H_\ell$  and of the entries of  $\mathbf{W}_\ell$  are normalized to one,  $\rho$  in (6) can be interpreted as the average SNR at the receiver.

- 2) A decoder  $g: \mathbb{C}^{LT} \rightarrow \{1, \dots, M\}$  satisfying the maximum error probability constraint

$$\max_{1 \leq a \leq M} \mathbb{P}[g(\mathbf{Y}^L) \neq a | A = a] \leq \epsilon \quad (7)$$

where  $\mathbf{Y}^L = [\mathbf{Y}_1, \dots, \mathbf{Y}_L]$  is the channel output induced by the transmitted codeword  $\mathbf{X}^L = f(a)$  according to (5).

The *maximum coding rate*  $R^*(L, T, \epsilon, \rho)$  is defined as the largest rate  $\log M/(LT)$  (in nats per channel use) for which there exists an  $(M, L, T, \epsilon, \rho)$  code, i.e.,

$$R^*(L, T, \epsilon, \rho) \triangleq \sup \left\{ \frac{\log M}{LT} : \exists (M, L, T, \epsilon, \rho) \text{ code} \right\}. \quad (8)$$

#### IV. PRELIMINARIES

We next introduce some preliminary results that will be helpful in the remainder of the paper.

Conditioned on  $\mathbf{X}^L = \mathbf{x}^L$ , the output vector  $\mathbf{Y}^L$  is blockwise i.i.d. Gaussian. Thus, the conditional pdf of  $\mathbf{Y}_\ell$  given  $\mathbf{X}_\ell = \mathbf{x}$  is independent of  $\ell$  and satisfies

$$p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{1}{\pi^T(1 + \|\mathbf{x}\|^2)} \exp \left\{ -\|\mathbf{y}\|^2 + \frac{|\mathbf{y}^H \mathbf{x}|^2}{1 + \|\mathbf{x}\|^2} \right\} \quad (9)$$

for  $\mathbf{x}, \mathbf{y} \in \mathbb{C}^T$ . Here and throughout the paper, we omit the subscript  $\ell$  when immaterial. We shall refer to the distribution  $\mathbb{P}_{\mathbf{X}^L}^{(U)}$ , according to which  $\mathbf{X}^L = \sqrt{T\rho} \mathbf{U}^L$  (where  $\mathbf{U}^L = [\mathbf{U}_1, \dots, \mathbf{U}_L]$  and  $\mathbf{U}_1, \dots, \mathbf{U}_L$  are i.i.d. and uniformly distributed on the unit sphere in  $\mathbb{C}^T$ ), as USTM [15]. This distribution is relevant because it gives rise to a lower bound on capacity that is asymptotically tight at high SNR [16], [17]. In fact, it can be shown that this lower bound accurately approximates capacity already for intermediate SNR values. For example, [12, Fig. 1] illustrates that the lower bound is indistinguishable from the upper bound on capacity given in [12, Eq. (17)] for  $\rho \geq 10$  dB.

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

$$q_{\mathbf{Y}^L}^{(U)}(\mathbf{y}^L) = \prod_{\ell=1}^L q_{\mathbf{Y}}^{(U)}(\mathbf{y}_\ell), \quad \mathbf{y}^L = [\mathbf{y}_1, \dots, \mathbf{y}_L] \in \mathbb{C}^{LT} \quad (10)$$

where [12, Eq. (18)]

$$q_{\mathbf{Y}}^{(U)}(\mathbf{y}) = \frac{e^{-\|\mathbf{y}\|^2/(1+T\rho)} \|\mathbf{y}\|^{2(1-T)} \Gamma(T)}{\pi^T(1+T\rho)} \times \tilde{\gamma} \left( T-1, \frac{T\rho\|\mathbf{y}\|^2}{1+T\rho} \right) \left( 1 + \frac{1}{T\rho} \right)^{T-1} \quad (11)$$

<sup>1</sup>In the information theory literature, it is more common to impose a power constraint per codeword  $\mathbf{X}^L$ . However, practical systems typically require a per-coherence-interval constraint. Note that, in contrast to [14], where the power constraint (6) is assumed to hold with equality, here we consider the more general case where the power constraint may also be satisfied with strict inequality.

for  $\mathbf{y} \in \mathbb{C}^T$ . Observe that the expression of  $q_{\mathbf{Y}}^{(U)}$  contains a regularized lower incomplete gamma function. The following lemma presents an upper and lower bound on the logarithm of this function, which we shall use throughout the paper.

*Lemma 1:* The logarithm of the regularized lower incomplete gamma function  $\tilde{\gamma}(T-1, x)$ ,  $x > 0$  can be bounded as

$$0 \leq \log \frac{1}{\tilde{\gamma}(T-1, x)} \leq (T-1) \log \left( 1 + \frac{\Gamma(T)^{\frac{1}{T-1}}}{x} \right). \quad (12)$$

*Proof:* See Appendix A. ■

Throughout this paper, we shall denote by  $\mathbf{Y}^L$  a blockwise i.i.d. random vector whose conditional pdf, conditioned on  $\mathbf{X}^L = \mathbf{x}^L$ , is given by  $\prod_{\ell=1}^L p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}_\ell|\mathbf{x}_\ell)$  with  $p_{\mathbf{Y}|\mathbf{X}}$  as in (9). We shall denote by  $\tilde{\mathbf{Y}}^L$  a blockwise i.i.d. random vector that is independent of  $\mathbf{X}^L$  and has pdf  $q_{\mathbf{Y}^L}^{(U)}$ .

Conditioned on  $\|\mathbf{X}_\ell\|^2 = T\alpha_\ell$ , the random variables  $|\mathbf{Y}_\ell^H \mathbf{X}_\ell|^2$  and  $\|\mathbf{Y}_\ell\|^2$  can be written as

$$|\mathbf{Y}_\ell^H \mathbf{X}_\ell|^2 \stackrel{\mathcal{L}}{=} |H_\ell^* T\alpha_\ell + \mathbf{W}_\ell^*(1)\sqrt{T\alpha_\ell}|^2 \stackrel{\mathcal{L}}{=} T\alpha_\ell(1 + T\alpha_\ell)Z_{1,\ell} \quad (13)$$

$$\|\mathbf{Y}_\ell\|^2 \stackrel{\mathcal{L}}{=} \|H_\ell\sqrt{T\alpha_\ell}\mathbf{e}_1 + \mathbf{W}_\ell\|^2 \stackrel{\mathcal{L}}{=} (1 + T\alpha_\ell)Z_{1,\ell} + Z_{2,\ell} \quad (14)$$

where  $\mathbf{e}_1$  is the length- $T$  unitary vector  $[1, 0, \dots, 0]^T$ ,  $\{Z_{1,\ell}, \ell \in \mathbb{Z}\}$  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.

Conditioned on  $\|\mathbf{X}_\ell\|^2 = T\alpha_\ell$ , the random variables  $|\tilde{\mathbf{Y}}_\ell^H \mathbf{X}_\ell|^2$  and  $\|\tilde{\mathbf{Y}}_\ell\|^2$  can be written as

$$|\tilde{\mathbf{Y}}_\ell^H \mathbf{X}_\ell|^2 \stackrel{\mathcal{L}}{=} |(H_\ell^* \sqrt{T\rho} \mathbf{U}_\ell(1) + \mathbf{W}_\ell^*(1))\sqrt{T\alpha_\ell}|^2 \quad (15)$$

$$\|\tilde{\mathbf{Y}}_\ell\|^2 \stackrel{\mathcal{L}}{=} \|H_\ell\sqrt{T\rho}\mathbf{U}_\ell + \mathbf{W}_\ell\|^2. \quad (16)$$

In (13)–(16), the parameter  $\alpha_\ell$  lies in the interval  $[0, \rho]$  and can be thought of as the power allocated over the coherence interval  $\ell$ .

The *information density* between the random vectors  $\mathbf{X}^L$  and  $\mathbf{Y}^L$  is defined as

$$i(\mathbf{X}^L; \mathbf{Y}^L) \triangleq \log \left( \frac{p_{\mathbf{Y}^L|\mathbf{X}^L}(\mathbf{Y}^L | \mathbf{X}^L)}{p_{\mathbf{Y}^L}(\mathbf{Y}^L)} \right) \quad (17)$$

where  $p_{\mathbf{Y}^L}$  is the output pdf induced by the input distribution.<sup>2</sup> When the input distribution is USTM, the information density  $i(\mathbf{X}^L; \mathbf{Y}^L)$  can be expressed as

$$i(\mathbf{X}^L; \mathbf{Y}^L) = \sum_{\ell=1}^L i_\ell(T, \rho) \quad (18)$$

where

$$i_\ell(T, \rho) \triangleq (T-1) \log(T\rho) - \log \Gamma(T) - \frac{T\rho Z_{2,\ell}}{1+T\rho} + (T-1) \log \left( \frac{(1+T\rho)Z_{1,\ell} + Z_{2,\ell}}{1+T\rho} \right) - \log \tilde{\gamma} \left( T-1, \frac{T\rho((1+T\rho)Z_{1,\ell} + Z_{2,\ell})}{1+T\rho} \right). \quad (19)$$

<sup>2</sup>The existence of the conditional pdf  $p_{\mathbf{Y}^L|\mathbf{X}^L}$  implies that the output pdf  $p_{\mathbf{Y}^L}$  exists for every input distribution.

Using the left-most inequality in Lemma 1, we can lower-bound (19) by

$$\begin{aligned} \underline{i}_\ell(T, \rho) &\triangleq (T-1) \log(T\rho) - \log \Gamma(T) - \frac{T\rho Z_{2,\ell}}{1+T\rho} \\ &\quad + (T-1) \log\left(\frac{(1+T\rho)Z_{1,\ell} + Z_{2,\ell}}{1+T\rho}\right). \end{aligned} \quad (20)$$

The expected value of (19), denoted by  $I(T, \rho)$ , is given by

$$\begin{aligned} I(T, \rho) &\triangleq \mathbb{E}[i_\ell(T, \rho)] \\ &= (T-1) \log(T\rho) - \log \Gamma(T) - \frac{(T-1)T\rho}{1+T\rho} \\ &\quad (T-1) \mathbb{E}\left[\log\left(\frac{(1+T\rho)Z_1 + Z_2}{1+T\rho}\right)\right] \\ &\quad + \mathbb{E}\left[\log \tilde{\gamma}\left(T-1, \frac{T\rho((1+T\rho)Z_1 + Z_2)}{1+T\rho}\right)\right]. \end{aligned} \quad (21)$$

Likewise, the expected value of (20), denoted by  $\underline{I}(T, \rho)$ , can be computed as

$$\begin{aligned} \underline{I}(T, \rho) &\triangleq \mathbb{E}[\underline{i}_\ell(T, \rho)] \\ &= (T-1) \log(T\rho) - \log \Gamma(T) - \frac{(T-1)T\rho}{1+T\rho} \\ &\quad - (T-1) \log(1+T\rho) \\ &\quad + (T-1) \mathbb{E}[\log((1+T\rho)Z_1 + Z_2)] \\ &= (T-1) \log(T\rho) - \log \Gamma(T) \\ &\quad - (T-1) \left[ \log(1+T\rho) + \frac{T\rho}{1+T\rho} - \psi(T-1) \right] \\ &\quad + {}_2F_1\left(1, T-1; T; \frac{T\rho}{1+T\rho}\right) \end{aligned} \quad (22)$$

where the expected value in the fourth line has been solved using [19, Sec. 4.337-1] to integrate with respect to  $Z_1$  and [19, Sec. 4.352-1], [19, Sec. 3.381-4], and [20, Sec. 4.2.20] to integrate with respect to  $Z_2$ . Clearly,

$$I(T, \rho) \geq \underline{I}(T, \rho). \quad (23)$$

We define the *mismatched information density*<sup>3</sup> between the random vectors  $\mathbf{X}^L$  and  $\mathbf{Y}^L$  as

$$j(\mathbf{X}^L; \mathbf{Y}^L) \triangleq \log\left(\frac{p_{\mathbf{Y}^L|\mathbf{X}^L}(\mathbf{Y}^L | \mathbf{X}^L)}{q_{\mathbf{Y}^L}^{(U)}(\mathbf{Y}^L)}\right). \quad (24)$$

Using this definition together with (9), (10), and (11), the mismatched information density  $j(\mathbf{X}^L; \mathbf{Y}^L)$  can be written as

$$j(\mathbf{X}^L; \mathbf{Y}^L) = \sum_{\ell=1}^L j_\ell(\mathbf{X}_\ell; \mathbf{Y}_\ell) \quad (25)$$

<sup>3</sup>We use the word ‘‘mismatched’’ to indicate that the output distribution  $q_{\mathbf{Y}^L}^{(U)}$  in the denominator in (24) is not the one induced by the input distribution and the channel.

where

$$\begin{aligned} j_\ell(\mathbf{X}_\ell; \mathbf{Y}_\ell) &\triangleq \log\left(\frac{1+T\rho}{\Gamma(T)}\right) + \frac{|\mathbf{Y}_\ell^H \mathbf{X}_\ell|^2}{1+\|\mathbf{X}_\ell\|^2} - \frac{T\rho\|\mathbf{Y}_\ell\|^2}{1+T\rho} \\ &\quad + (T-1) \log\left(\frac{T\rho\|\mathbf{Y}_\ell\|^2}{1+T\rho}\right) - \log(1+\|\mathbf{X}_\ell\|^2) \\ &\quad - \log \tilde{\gamma}\left(T-1, \frac{T\rho\|\mathbf{Y}_\ell\|^2}{1+T\rho}\right). \end{aligned} \quad (26)$$

By (13) and (14),  $j(\mathbf{X}_\ell; \mathbf{Y}_\ell)$  depends on  $\mathbf{X}_\ell$  only via its magnitude. We can thus express  $j(\mathbf{X}_\ell; \mathbf{Y}_\ell)$ , conditioned on  $\|\mathbf{X}_\ell\|^2 = T\alpha_\ell$ , as

$$\begin{aligned} j_\ell(T, \alpha_\ell) &\triangleq (T-1) \log(T\rho) - \log \Gamma(T) - \frac{(T\rho - T\alpha_\ell)Z_{1,\ell}}{1+T\rho} \\ &\quad - \frac{T\rho Z_{2,\ell}}{1+T\rho} + \log\left(\frac{1+T\rho}{1+T\alpha_\ell}\right) \\ &\quad + (T-1) \log\left(\frac{(1+T\alpha_\ell)Z_{1,\ell} + Z_{2,\ell}}{1+T\rho}\right) \\ &\quad - \log \tilde{\gamma}\left(T-1, \frac{T\rho((1+T\alpha_\ell)Z_{1,\ell} + Z_{2,\ell})}{1+T\rho}\right). \end{aligned} \quad (27)$$

Note that the information densities  $i(\mathbf{X}^L; \mathbf{Y}^L)$  and  $j(\mathbf{X}^L; \mathbf{Y}^L)$  only differ in the output densities  $p_{\mathbf{Y}^L}$  and  $q_{\mathbf{Y}^L}^{(U)}$  in the denominators of (17) and (24), respectively. Consequently, for USTM inputs, where  $p_{\mathbf{Y}^L} = q_{\mathbf{Y}^L}^{(U)}$ , we have  $i_\ell(T, \rho) = j_\ell(T, \rho)$ .

Define  $\beta(T, \rho) \triangleq \Gamma(T)^{\frac{1}{T-1}} \frac{1+T\rho}{T\rho}$ , and let

$$\begin{aligned} \bar{j}_\ell(T, \alpha_\ell) &\triangleq (T-1) \log(T\rho) - \log \Gamma(T) - \frac{(T\rho - T\alpha_\ell)Z_{1,\ell}}{1+T\rho} \\ &\quad - \frac{T\rho Z_{2,\ell}}{1+T\rho} + \log\left(\frac{1+T\rho}{1+T\alpha_\ell}\right) \\ &\quad + (T-1) \log\left(\frac{(1+T\alpha_\ell)Z_{1,\ell} + Z_{2,\ell}}{1+T\rho}\right) \\ &\quad + (T-1) \log\left(1 + \frac{\beta(T, \rho)}{(1+T\alpha_\ell)Z_{1,\ell} + Z_{2,\ell}}\right). \end{aligned} \quad (28)$$

By Lemma 1, we have that, with probability one,

$$j_\ell(T, \alpha_\ell) \leq \bar{j}_\ell(T, \alpha_\ell), \quad \alpha_\ell \in [0, \rho]. \quad (29)$$

Let  $J(T, \alpha_\ell) \triangleq \mathbb{E}[j_\ell(T, \alpha_\ell)]$  and  $\bar{J}(T, \alpha_\ell) \triangleq \mathbb{E}[\bar{j}_\ell(T, \alpha_\ell)]$  denote the conditional expected values of (27) and (28) given  $\|\mathbf{X}_\ell\|^2 = T\alpha_\ell$ . The latter expected value can be evaluated as

$$\begin{aligned} \bar{J}(T, \alpha_\ell) &= (T-1) \log(T\rho) - \log \Gamma(T) - \frac{(T-1)T\rho}{1+T\rho} \\ &\quad - (T-1) \log(1+T\rho) \\ &\quad + (T-1) \mathbb{E}\left[\log((1+T\alpha_\ell)Z_1 + Z_2 + \beta(T, \rho))\right] \\ &\quad - \frac{T\rho - T\alpha_\ell}{1+T\rho} + \log\left(\frac{1+T\rho}{1+T\alpha_\ell}\right). \end{aligned} \quad (30)$$

Clearly,

$$J(T, \alpha_\ell) \leq \bar{J}(T, \alpha_\ell), \quad \alpha_\ell \in [0, \rho]. \quad (31)$$

It can be shown that  $\bar{J}(\cdot)$  and  $\underline{I}(\cdot)$  bound the capacity [21]

$$C(T, \rho) = \sup_{\mathbf{P}_{\mathbf{X}^L}: \mathbb{E}[\|\mathbf{X}_\ell\|^2] \leq T\rho} \frac{\mathbb{E}[i(\mathbf{X}_\ell; \mathbf{Y}_\ell)]}{T}. \quad (32)$$

Indeed, on the one hand we have

$$C(T, \rho) \leq \sup_{0 \leq \alpha \leq \rho} \frac{J(T, \alpha)}{T} \leq \sup_{0 \leq \alpha \leq \rho} \frac{\bar{J}(T, \alpha)}{T} \quad (33)$$

where the first inequality follows from [22, Th. 5.1] and the second inequality follows from (31). On the other hand,

$$C(T, \rho) \geq \frac{I(T, \rho)}{T} \geq \frac{\underline{I}(T, \rho)}{T} \quad (34)$$

where the first inequality follows because USTM is a valid input distribution and the second inequality follows from (23). Note that  $J(T, \rho) = I(T, \rho)$  when the input distribution is USTM. It can be further shown that

$$\lim_{\rho \rightarrow \infty} \left\{ \sup_{0 \leq \alpha \leq \rho} \bar{J}(T, \alpha) - \underline{I}(T, \rho) \right\} = 0. \quad (35)$$

Thus, USTM yields an asymptotically tight lower bound on capacity, as already mentioned before.

Let

$$U(T, \rho) \triangleq \mathbb{E} \left[ (i_\ell(T, \rho) - I(T, \rho))^2 \right] \quad (36)$$

$$\bar{V}_\rho(T, \alpha) \triangleq \mathbb{E} \left[ (\bar{j}_\ell(T, \alpha) - \bar{J}(T, \alpha))^2 \right] \quad (37)$$

where the subscript  $\rho$  in  $\bar{V}_\rho(T, \alpha)$  is introduced to highlight that  $\bar{V}_\rho(T, \alpha)$  depends both on  $\alpha$  and  $\rho$ , but it is omitted when  $\alpha = \rho$ . In Lemma 9 (Appendix I) and Lemma 10 (Appendix J), we show that  $I(T, \rho)$ ,  $U(T, \rho)$ ,  $\bar{J}(T, \rho)$ , and  $\bar{V}_\rho(T, \rho)$  can be approximated as

$$I(T, \rho) = \underline{I}(T, \rho) + K_I(T, \rho) \quad (38)$$

$$U(T, \rho) = \tilde{U}(T) + K_U(T, \rho) \quad (39)$$

$$\bar{J}(T, \rho) = \underline{I}(T, \rho) + K_{\bar{J}}(T, \rho) \quad (40)$$

$$\bar{V}(T, \rho) = \tilde{U}(T) + K_{\bar{V}}(T, \rho) \quad (41)$$

where  $K_\xi(T, \rho)$ ,  $\xi = \{I, U, \bar{J}, \bar{V}\}$  are functions of  $T$  and  $\rho$  that satisfy

$$\lim_{\rho \rightarrow \infty} K_\xi(T, \rho) = 0, \quad T > 2. \quad (42)$$

A closed form expression of  $\underline{I}(T, \rho)$  is given in (22). Moreover,  $\tilde{U}(T)$  in (39) and (41) is defined as

$$\tilde{U}(T) \triangleq (T-1)^2 \frac{\pi^2}{6} + (T-1). \quad (43)$$

## V. MAIN RESULT

The main result of this paper is a high-SNR normal approximation on  $R^*(L, T, \epsilon, \rho)$  presented in Section V-A. In Section V-B, we assess the accuracy of this approximation by means of numerical examples. Possible applications are discussed in Section V-C.

### A. A High-SNR Normal Approximation

*Theorem 2:* Assume that  $T > 2$  and  $0 < \epsilon < \frac{1}{2}$ . Then, the maximum coding rate  $R^*(L, T, \epsilon, \rho)$  can be expanded as

$$R^*(L, T, \epsilon, \rho) = \frac{\underline{I}(T, \rho)}{T} - \sqrt{\frac{\tilde{U}(T) + K_{\tilde{U}}(T, \rho)}{LT^2}} Q^{-1}(\epsilon) + K_{\underline{I}}(T, \rho) + K_L(L, T, \rho) \quad (44)$$

where  $\underline{I}(T, \rho)$  and  $\tilde{U}(T)$  are defined in (22) and (43), respectively, and  $K_\xi(T, \rho)$ ,  $\xi = \{\underline{I}, \tilde{U}\}$  are functions of  $T$  and  $\rho$  that satisfy

$$\lim_{\rho \rightarrow \infty} K_\xi(T, \rho) = 0, \quad T > 2. \quad (45)$$

Similarly,  $K_L(L, T, \rho)$  is a function of  $T$ ,  $L$ , and  $\rho$  that satisfies

$$\sup_{\rho \geq \rho_0} |K_L(L, T, \rho)| \leq A \frac{\log L}{L}, \quad L \geq L_0 \quad (46)$$

for every  $T > 2$  and some  $A$ ,  $L_0$ , and  $\rho_0$  independent of  $L$  and  $\rho$ .

*Proof:* See Section VI. ■

*Remark 1:* The assumption that  $0 < \epsilon < 1/2$  is required to ensure that  $Q^{-1}(\epsilon)$  is nonnegative, which simplifies the manipulations of the channel dispersion. Treating the case  $1/2 < \epsilon < 1$  would require a separate analysis. For the sake of compactness, we decided to omit such an analysis, since we believe that  $0 < \epsilon < 1/2$  covers all cases of practical interest.

Ignoring the  $K_{\underline{I}}(T, \rho)$ ,  $K_{\tilde{U}}(T, \rho)$ , and  $K_L(L, T, \rho)$  terms in (44), we obtain the high-SNR normal approximation

$$R^*(L, T, \epsilon, \rho) \approx \frac{\underline{I}(T, \rho)}{T} - \sqrt{\frac{\tilde{U}(T)}{LT^2}} Q^{-1}(\epsilon). \quad (47)$$

The closed form expression for  $\underline{I}(T, \rho)$  in (22) contains a hypergeometric function, which is difficult to analyze mathematically. We therefore present also a simplified expression that is less accurate than (22) but easier to analyze. Specifically, it follows from Lemma 9 (Appendix I) that

$$\underline{I}(T, \rho) = (T-1) \log(T\rho) - \log \Gamma(T) - (T-1)(1+\gamma) + K'_{\underline{I}}(T, \rho) \quad (48)$$

where  $K'_{\underline{I}}(T, \rho)$  is a function of  $T$  and  $\rho$  that satisfies

$$\lim_{\rho \rightarrow \infty} K'_{\underline{I}}(T, \rho) = 0, \quad T > 2. \quad (49)$$

The quantity  $\underline{I}(T, \rho)/T$  is a high-SNR approximation of the information rate achievable with i.i.d. USTM inputs; cf. [23, Eq. (12)] (see also [12, Eq. (5)]). It is shown in [15, Th. 4] that  $\underline{I}(T, \rho)/T$  is also an asymptotically-tight lower bound on the capacity  $C(T, \rho)$  in the sense that

$$\lim_{\rho \rightarrow \infty} \left\{ C(T, \rho) - \frac{\underline{I}(T, \rho)}{T} \right\} = 0. \quad (50)$$

According to Theorem 2,

$$\frac{\tilde{U}(T)}{T^2} = \frac{(T-1)^2 \pi^2}{T^2 \cdot 6} + \frac{T-1}{T^2} \quad (51)$$

can be viewed as a high-SNR approximation of the channel dispersion.

For comparison, the capacity and dispersion of the coherent Rayleigh block-fading channel—where the receiver has perfect knowledge of the realizations of the fading coefficients—are given by [4], [7], [24]

$$C_c(\rho) = \mathbb{E}[\log(1 + \rho Z_1)] \quad (52a)$$

$$V_c(T, \rho) = \text{Var}[\log(1 + \rho Z_1)] + \frac{1}{T} - \frac{1}{T} \mathbb{E} \left[ \frac{1}{1 + \rho Z_1} \right]^2. \quad (52b)$$

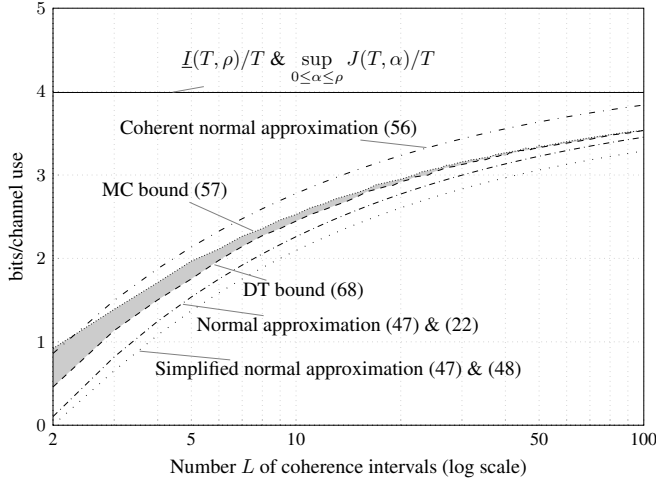


Fig. 1. Bounds on  $R^*(L, T, \epsilon, \rho)$  for  $\rho = 15$  dB,  $T = 20$ ,  $\epsilon = 10^{-3}$ . The shaded area indicates the area in which  $R^*(L, T, \epsilon, \rho)$  lies.

Note that

$$\lim_{\rho \rightarrow \infty} \{C_c(\rho) - (\log \rho - \gamma)\} = 0 \quad (53)$$

and, for every  $T$ ,

$$\lim_{\rho \rightarrow \infty} V_c(T, \rho) = \frac{\pi^2}{6} + \frac{1}{T}. \quad (54)$$

Furthermore, for the noncoherent channel the high-SNR capacity  $\underline{I}(T, \rho)/T$  satisfies (cf. (48))

$$\lim_{T \rightarrow \infty} \lim_{\rho \rightarrow \infty} \left\{ \frac{\underline{I}(T, \rho)}{T} - \frac{T-1}{T} [\log(\rho) - \gamma] \right\} = 0. \quad (55)$$

By comparing (53) and (55), we see that  $\underline{I}(T, \rho)/T$  is, up to terms that vanish as  $\rho \rightarrow \infty$  and  $T \rightarrow \infty$ , equal to  $(1 - 1/T)C_c(\rho)$ . Similarly, by comparing (51) and (54), we observe that  $\bar{U}(T)/T^2$  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 the heuristic that, at high SNR, one pilot symbol per coherence block should be transmitted to achieve both capacity and channel dispersion. However, this heuristic may be misleading since it is *prima facie* unclear whether one pilot symbol per coherence block suffices to obtain a fading estimate of sufficient accuracy. A more refined analysis of the maximum coding rate achievable with pilot-assisted transmission has been recently performed by Östman *et al.* [25].

### B. Numerical Examples

We illustrate the accuracy of the high-SNR normal approximation (47) by means of numerical examples. In Figs. 1 and 2, we show the approximation (47) as a function of  $L = n/T$  for a fixed coherence interval  $T$  and for different SNR values. In the normal approximation, we evaluate  $\underline{I}(T, \rho)$  using both the exact expression (22) as well as the approximation (48).

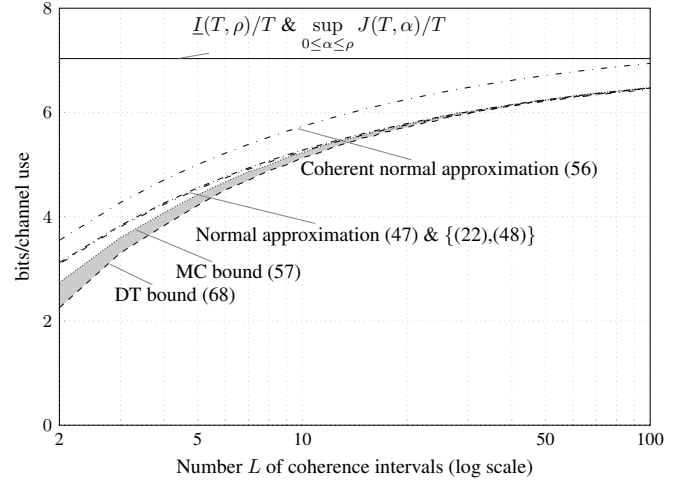


Fig. 2. Bounds on  $R^*(L, T, \epsilon, \rho)$  for  $\rho = 25$  dB,  $T = 20$ ,  $\epsilon = 10^{-3}$ . The shaded area indicates the area in which  $R^*(L, T, \epsilon, \rho)$  lies.

For comparison, we also plot the normal approximation of the coherent Rayleigh block-fading channel

$$R^*(L, T, \epsilon, \rho) \approx C_c(\rho) - \sqrt{\frac{V_c(T, \rho)}{L}} Q^{-1}(\epsilon) \quad (56)$$

where  $C_c(\rho)$  and  $V_c(T, \rho)$  are defined in (52a) and (52b), respectively. We further plot a nonasymptotic (in  $\rho$  and  $L$ ) lower bound on  $R^*(L, T, \epsilon, \rho)$  that is based on the dependence testing (DT) lower bound [3, Th. 22] with USTM channel inputs (see (68) below) and computed by Monte Carlo simulations. Similarly, we plot a nonasymptotic (in  $\rho$  and  $L$ ) upper bound on  $R^*(L, T, \epsilon, \rho)$  that is based on the meta converse (MC) upper bound [3, Th. 31] with auxiliary output pdf (10) (see (86) below).<sup>4</sup> More precisely, we plot the weakened version<sup>5</sup>

$$R^*(L, T, \epsilon, \rho) \leq \inf_{\xi > 0} \left\{ \frac{\log \xi}{LT} - \inf_{\alpha \in [\alpha, \rho]^L} \frac{\log \left( 1 - \epsilon - \mathbb{P} \left[ \sum_{\ell=1}^L j_\ell(\alpha_\ell) \geq \log \xi \right] \right)}{LT} \right\} \quad (57)$$

which is obtained by using [3, Eq. (102)] and was evaluated by Monte Carlo simulations. In (57),  $\alpha = (\alpha_1, \dots, \alpha_L)$  denotes the vector of power allocations. We finally plot  $\underline{I}(T, \rho)/T$  given in (22) and  $\sup_{0 \leq \alpha \leq \rho} J(T, \alpha)/T$  defined right before (30), which in both figures are indistinguishable from each other. By (33) and (34), we have that

$$\frac{\underline{I}(T, \rho)}{T} \leq C(T, \rho) \leq \sup_{0 \leq \alpha \leq \rho} \frac{J(T, \alpha)}{T}. \quad (58)$$

<sup>4</sup>The MC bound appeared in the form used in this paper (cf. (86)) in [3, Th. 31]. It can also be obtained by particularizing the quantum result of Nagaoka [26] to the classical case.

<sup>5</sup>The relaxation (57) of the MC bound coincides with the Verdú-Han bound [27, Th. 4] replacing the true output distribution  $\mathbb{P}_{\mathcal{Y}}$  by an arbitrary output distribution  $\mathbb{Q}_{\mathcal{Y}}$ . This bound for an arbitrary output distribution  $\mathbb{Q}_{\mathcal{Y}}$  is a particularization of the Hayashi-Nagaoka lemma for classical quantum channels [28, Lemma 4].

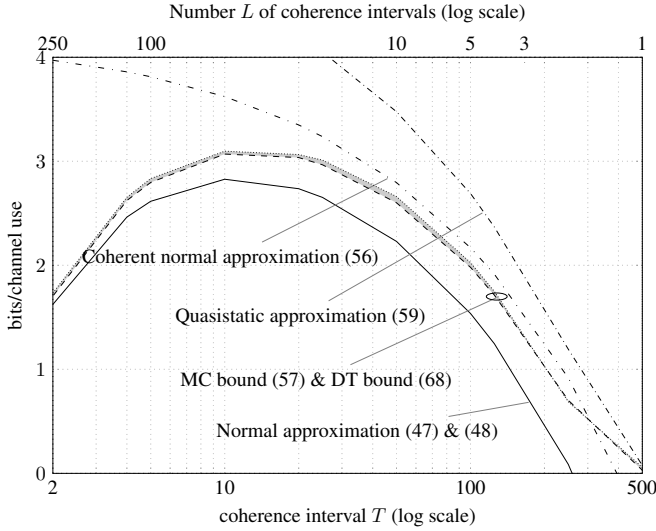


Fig. 3. Bounds on  $R^*(L, T, \epsilon, \rho)$  for  $LT = 500$ ,  $\epsilon = 10^{-3}$ ,  $\rho = 15$  dB. The MC bound and the DT bound are almost indistinguishable. The shaded area indicates the area in which  $R^*(L, T, \epsilon, \rho)$  lies.

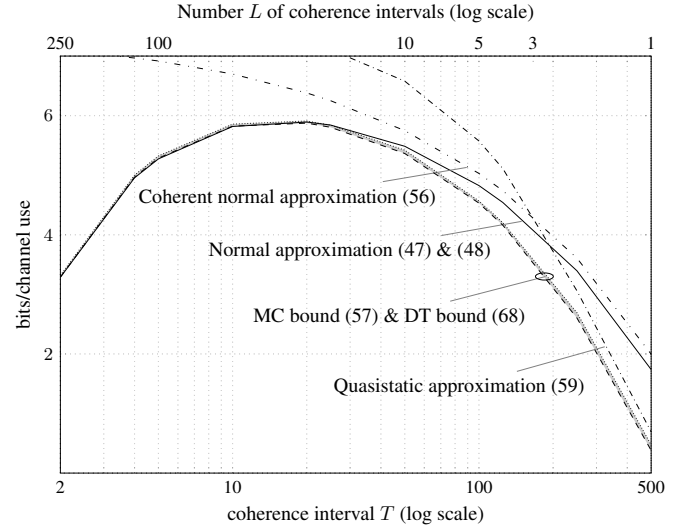


Fig. 4. Bounds on  $R^*(L, T, \epsilon, \rho)$  for  $LT = 500$ ,  $\epsilon = 10^{-3}$ ,  $\rho = 25$  dB. The MC bound and the DT bound are almost indistinguishable. The shaded area indicates the area in which  $R^*(L, T, \epsilon, \rho)$  lies.

We thus conclude that the error term  $K_{\underline{I}}(T, \rho)$  in (44) is negligible for the SNR values considered in the figures. Observe that the high-SNR normal approximation of  $R^*(L, T, \epsilon, \rho)$  is accurate already for  $\rho = 15$  dB and  $L \geq 10$  when we use the exact expression (22) for  $\underline{I}(T, \rho)$ . For  $\rho = 25$  dB and  $L \geq 10$ , the normal approximation is accurate even when we approximate  $\underline{I}(T, \rho)$  using the simplified expression (48). Further observe that the normal approximation is pessimistic for  $\rho = 15$  dB and optimistic for  $\rho = 25$  dB. As expected, the normal approximation (56) of the coherent channel is strictly larger than the high-SNR normal approximation (47) and the gap between the two normal approximations appears to be independent of  $L$ . This agrees with the intuition that the cost for estimating the channel mainly depends on the length  $T$  of the coherence interval. Finally observe that the DT lower bound on  $R^*(L, T, \epsilon, \rho)$ , computed for USTM channel inputs, is close to the MC upper bound, which holds for any input distribution satisfying the power constraint (6), provided that  $\rho = 15$  dB and  $L \geq 5$  or  $\rho = 25$  dB and  $L \geq 2$ . Thus, USTM channel inputs, which achieve the capacity asymptotically as the SNR tends to infinity, also give rise to lower bounds on  $R^*(L, T, \epsilon, \rho)$  that are close to optimal for moderate SNR values and short blocklengths. A similar observation was also made in [12].

In Figs. 3 and 4, we show the high-SNR normal approximation (47) (with  $\underline{I}(T, \rho)/T$  evaluated using the approximation (48)) as a function of the coherence interval  $T$  for a fixed blocklength  $n$  (hence  $L$  is inversely proportional to  $T$ ). We further plot the normal approximation (56) of the coherent channel. For comparison, we also show the DT bound (see (68) below), evaluated for an USTM input distribution, and the weakened version of the MC bound (57) evaluated by Monte Carlo simulations. Finally, we present the normal approximation that was proposed in [9] for quasistatic multiple-input multiple-output (MIMO) block-fading channels. To adapt the

quasistatic MIMO block-fading channel to our system model, we replace  $\mathbb{H}$  in [9] by an  $L \times L$  diagonal matrix with diagonal entries  $H_1, \dots, H_L$ . Thus, specializing [9, Eq. (95)] to our case, we obtain

$$\epsilon \approx \mathbb{E} \left[ Q \left( \frac{C(\mathbb{H}) - LR^*(L, T, \epsilon, \rho)}{\sqrt{V(\mathbb{H})/T}} \right) \right] \quad (59)$$

where

$$C(\mathbb{H}) \triangleq \sum_{j=1}^L \log(1 + \rho |H_j|^2) \quad (60a)$$

$$V(\mathbb{H}) \triangleq L - \sum_{j=1}^L \frac{1}{\log(1 + \rho |H_j|^2)^2}. \quad (60b)$$

As already observed in Figs. 1 and 2, the high-SNR normal approximation is accurate for  $\rho = 15$  dB and  $L \geq 10$ , and it is indistinguishable from the DT and MC bounds for  $\rho = 25$  dB and  $L \geq 10$ . The high-SNR normal approximation becomes less accurate as  $L$  decreases. Observe that the normal approximation of the coherent channel provides a good approximation when  $T$  is large but becomes inaccurate when  $T \leq 100$ . Further observe that the normal approximation for the quasistatic case (59), which is tailored towards the case where  $L$  is small, becomes accurate only for  $L \leq 3$  in both figures. The figures show that, for a fixed blocklength  $n = LT$ , there is an optimal tradeoff between  $L$  and  $T$ . This may be relevant, e.g., for the design of orthogonal frequency-division multiplexing (OFDM) systems, when the duration of a codeword is smaller than the coherence time, hence only frequency diversity is available. The system designer can then determine the number of diversity branches  $L$  available to each user by assigning OFDM symbols from different time and frequency slots. Figs. 3 and 4 indicate the optimal value of  $L$  for  $n = 500$ ,  $\epsilon = 10^{-3}$ , and  $\rho = \{15, 25\}$  dB. We refer to [29] for a more detailed discussion.



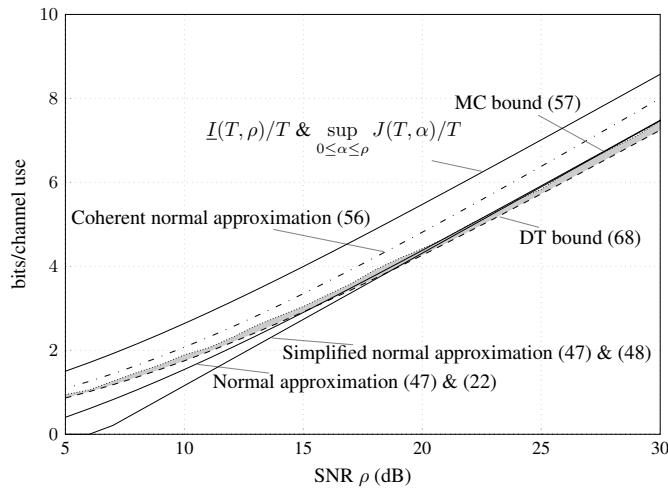


Fig. 5. Bounds on  $R^*(L, T, \epsilon, \rho)$  for  $T = 20$ ,  $L = 25$ ,  $\epsilon = 10^{-3}$ . The shaded area indicates the area in which  $R^*(L, T, \epsilon, \rho)$  lies.

In Fig. 5, we plot the high-SNR normal approximation (47), evaluating  $\underline{I}(T, \rho)$  using both (22) and (48), as a function of the SNR  $\rho$  for fixed  $T$  and  $L$ . We also plot the normal approximation (56) of the coherent channel. For comparison, we further plot the DT bound (see (68) below) evaluated for an USTM input distribution and the weakened version of the MC bound (57). Finally, we plot  $\underline{I}(T, \rho)/T$  given in (22) and  $\sup_{0 \leq \alpha \leq \rho} J(T, \alpha)/T$  defined right before (30), which in the figure are indistinguishable from each other. Recall that, by (33) and (34), these terms bound the capacity  $C(T, \rho)$  from below and from above, so we conclude that the error term  $K_{\underline{I}}(T, \rho)$  in (44) is negligible for the SNR values considered in the figures. Observe that the normal approximation that uses (22) becomes accurate already at SNR values of 15 dB, while the normal approximation that uses (48) is accurate for SNR values above 20 dB. Further observe that the normal approximation is pessimistic for  $\rho < 20$  dB and optimistic for  $\rho \geq 20$  dB. As expected, the normal approximation (56) of the coherent channel is strictly larger than the high-SNR normal approximation (47), but its gap to the nonasymptotic bounds decreases as  $\rho$  becomes small. Intuitively, this is because, as  $\rho$  decreases, knowledge of the fading coefficients becomes less important. Finally, we again observe that the DT lower bound on  $R^*(L, T, \epsilon, \rho)$  is close to the MC upper bound. Thus, USTM channel inputs, which achieve the capacity asymptotically as the SNR tends to infinity, are also close to optimal for all SNR values considered in the plot.

In Fig. 6, we plot the probability of error as a function of the SNR  $\rho$  for  $R = 4$ ,  $T = 20$ , and  $L = 25$ . Specifically, we show the high-SNR normal approximation (47) with  $\underline{I}(T, \rho)$  evaluated using (22), the normal approximation (56) of the coherent channel, the DT bound evaluated for an USTM input distribution, and the weakened version of the MC bound (57). For comparison, we further show the performance of an accumulate-repeat-jagged-accumulate (ARJA) low density parity check (LDPC) (3000,2000)-code combined with a 64-APSK modulation, pilot-assisted transmission (2

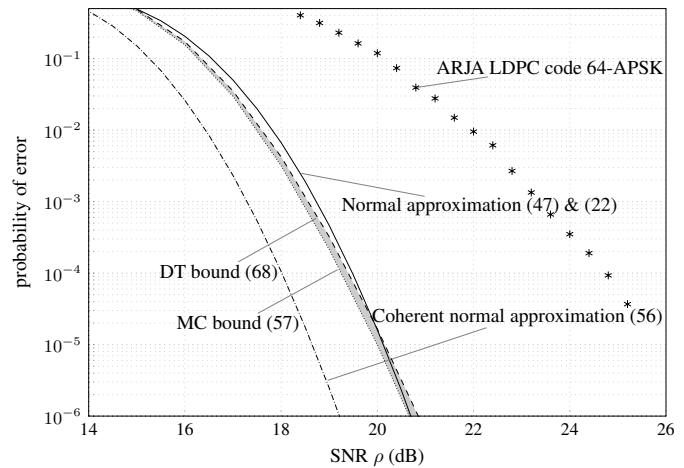


Fig. 6. Bounds on the probability of error  $\epsilon$  for  $R = 4$ ,  $T = 20$ ,  $L = 25$ . The shaded area indicates the area in which the true probability of error  $\epsilon$  lies.

pilot symbols per coherence block), and maximum likelihood channel estimation followed by mismatched nearest-neighbor decoding at the receiver (for details see [30, Sec. 4]). Observe that the high-SNR normal approximation is accurate for the whole range of SNR values evaluated. Further observe that the gap between the presented coding scheme and the rest of curves is substantial. This suggests that more sophisticated joint channel-estimation decoding procedures together with shaping techniques need to be adopted to close the gap; see also [31].

### C. Engineering Wisdom

As argued, e.g., in [1], the normal approximation can be used to analyze the performance of communication protocols. For example, let us consider the uplink scenario in [1, Sec. IV-C], where  $d$  devices intend to send  $k$  information bits to a base station within the time corresponding to  $n$  channel uses. The  $n$  channel uses are divided into  $s$  equally-sized slots of  $n_s \triangleq n/s$  channels uses and the devices apply a simple slotted-ALOHA protocol: each device picks randomly one of the  $s$  slots in the frame and sends its packet. If two or more devices pick the same slot, then a collision occurs and none of their packets is received correctly. If only one device picks a particular slot, then the error probability is calculated using the normal approximation. Specifically, in [1, Sec. IV-C] the normal approximation for the AWGN channel was considered, i.e.,<sup>6</sup>

$$R^*(n, \epsilon) \approx C_{\text{AWGN}}(\rho) - \sqrt{\frac{V_{\text{AWGN}}(\rho)}{n}} Q^{-1}(\epsilon) + \frac{1}{2} \frac{\log n}{n} \quad (61)$$

where

$$C_{\text{AWGN}}(\rho) = \log(1 + \rho) \quad (62a)$$

$$V_{\text{AWGN}}(\rho) = \rho \frac{2 + \rho}{(1 + \rho)^2}. \quad (62b)$$

<sup>6</sup>For the AWGN channel, the  $\mathcal{O}(\log n/n)$  in (1) can be replaced by  $(\log n)/(2n) + \mathcal{O}(1/n)$  [3], [32].

TABLE I  
 OPTIMAL NUMBER OF SLOTS FOR DIFFERENT CHANNEL MODELS AND  
 $n = LT = 480$ ,  $k = 256$ ,  $d = 12$ .

SNR	$T$	optimal number of slots $s$			
		noncoherent Rayleigh block- fading	coherent Rayleigh block- fading	AWGN	classic slotted- ALOHA
$\rho = 15$ dB	5	$s = 4$	$s = 6$	$s = 8$	$s = 12$
	20	$s = 6$	$s = 6$	$s = 8$	$s = 12$
$\rho = 25$ dB	5	$s = 8$	$s = 12$	$s = 12$	$s = 12$
	20	$s = 8$	$s = 8$	$s = 12$	$s = 12$

By solving (61) for  $\epsilon$ , we obtain an approximation of the packet error probability as a function of the packet length  $n$ , the number of information bits  $k = nR$  to be conveyed in a packet, and the SNR  $\rho$ , i.e.,

$$\epsilon^*(k, n, \rho) \approx Q\left(\frac{nC_{\text{AWGN}}(\rho) - k \log 2 + (\log n)/2}{\sqrt{nV_{\text{AWGN}}(\rho)}}\right). \quad (63)$$

By replacing (61) by our high-SNR normal approximation (47), we obtain the following approximation for the packet error probability when packets are transmitted over a noncoherent, single-antenna, Rayleigh block-fading channel of coherence interval  $T$ :

$$\epsilon^*(k, n, \rho) \approx Q\left(\frac{n\bar{I}(T, \rho) - kT \log 2}{\sqrt{nT\tilde{U}(T)}}\right). \quad (64)$$

Likewise, replacing (61) by the normal approximation for the coherent Rayleigh block-fading channel [4], [7], we obtain

$$\epsilon^*(k, n, \rho) \approx Q\left(\frac{nC_c(\rho) - k \log 2}{\sqrt{nTV_c(T, \rho)}}\right) \quad (65)$$

where

$$C_c(\rho) \triangleq \mathbb{E}[\log(1 + \rho Z_1)] \quad (66a)$$

$$V_c(T, \rho) \triangleq \text{Var}[\log(1 + \rho Z_1)] + \frac{1}{T} - \frac{1}{T} \mathbb{E}\left[\frac{1}{1 + \rho Z_1}\right]^2. \quad (66b)$$

The probability of successful transmission is given by [1, Eq. (24)]

$$P_{\text{success}} = \frac{d}{s} \left(1 - \frac{1}{s}\right)^{d-1} (1 - \epsilon^*(k, n_s, \rho)) \quad (67)$$

where  $(d/s)(1-1/s)^{d-1}$  is the probability that only one device transmits in a given slot [33, Sec. 5.3.2]. Our goal is to choose  $s$  such that the probability of successful transmission is maximized given  $d$ ,  $k$ ,  $n$ , and  $\rho$ . This problem entails a tradeoff between the probability of collision and the number of channel uses available for each packet, which affects the achievable error probability in a singleton slot.

As a concrete example, we consider the case where  $n = 480$ ,  $d = 12$ , and  $k = 256$ .<sup>7</sup> In Table I, we show the optimal number of slots  $s$  for the noncoherent Rayleigh block-fading channel (with  $\epsilon^*(k, n_s, \rho)$  approximated by (64)),

<sup>7</sup>The fact that  $n$  is fixed implies that the number of coherence intervals  $L$  changes inversely proportional to  $T$  for the block-fading cases.

the coherent Rayleigh block-fading channel (with  $\epsilon^*(k, n_s, \rho)$  approximated by (65)), the AWGN channel (with  $\epsilon^*(k, n_s, \rho)$  approximated by (63)), and the classic slotted-ALOHA protocol ( $\epsilon^*(k, n_s, \rho) = 0$ ) for the SNR values  $\rho = 15$  dB and  $\rho = 25$  dB and coherence intervals  $T = 5$  and  $T = 20$ . To be consistent with our system model, for the Rayleigh block-fading channel (both coherent and noncoherent) we only consider slot sizes  $n_s$  that are integer multiples of  $T$ . Observe that the optimal number of slots  $s$  depends critically on the SNR, the coherence interval, and the considered channel model. For example, for the classic slotted-ALOHA protocol, the optimal number of slots is  $s = 12$ , which coincides with the total number of devices  $d = 12$ . In contrast, for the AWGN channel, the optimal number of slots is  $s = 8$  for  $\rho = 15$  dB and coincides with the one of the classic slotted-ALOHA for  $\rho = 25$  dB. In most cases, the optimal number of slots  $s$  for the Rayleigh block-fading channel (both coherent and noncoherent) is yet again smaller and depends both on the SNR and the coherence interval  $T$ . When  $T = 20$ , the optimal number of slots  $s$  for the noncoherent Rayleigh block-fading channel coincides with that for the coherent channel. This agrees with the intuition that, when  $T$  is sufficiently large, the fading coefficients can be learned with little training overhead. In general, the optimal number of slots  $s$  decreases as the channel becomes less favorable. Intuitively, larger codes are required to combat the impairments due to AWGN and fading. Hence, the packet length  $n_s$  must be increased or, equivalently, the number of slots  $s = n/n_s$  must be reduced.

## VI. PROOF OF THEOREM 2

The proof of Theorem 2 is based on a lower bound on  $R^*(L, T, \epsilon, \rho)$ , given in Section VI-A, and on an upper bound on  $R^*(L, T, \epsilon, \rho)$ , given in Section VI-B. Since these bounds coincide up to the error terms  $K_{\text{DT}}(L, T, \rho)$  and  $K_{\text{MC}}(L, T, \rho)$  (whose difference is of order  $\log L/L$  uniform in  $\rho$ ) and up to the terms  $K_{\bar{J}}(T, \rho)$  and  $K_{\bar{V}}(T, \rho)$  given in (40) and (41) (which are independent of  $L$  and vanish as  $\rho \rightarrow \infty$ ), they prove (44).

### A. Dependence Testing Lower Bound

To obtain a lower bound on  $R^*(L, T, \epsilon, \rho)$ , we evaluate the DT bound [3, Th. 22] for the USTM input distribution defined in Section IV. Thus, assume that  $\mathbf{X}^L \sim P_{\mathbf{X}^L}^{(U)}$ , which implies  $\mathbf{Y}^L \sim q_{\mathbf{Y}^L}^{(U)}$ . One can show (see [14, App. A]) that the cumulative distribution function  $\mathbb{P}[i(\mathbf{x}^L; \tilde{\mathbf{Y}}^L) \leq \alpha]$  does not depend on  $\mathbf{x}^L$ . Furthermore, the USTM input distribution satisfies the power constraint (6) with probability one. A lower bound on  $R^*(L, T, \epsilon, \rho)$  follows therefore from the DT bound (maximum probability of error) [3, Th. 22], which, after a standard change of measure, can be stated as follows: there exists a code with  $M$  codewords, blocklength  $LT$ , and maximum probability of error  $\epsilon$  not exceeding

$$\epsilon \leq (M - 1) \mathbb{E} \left[ e^{-i(\mathbf{X}^L; \mathbf{Y}^L)} \mathbb{I}\{i(\mathbf{X}^L; \mathbf{Y}^L) > \log(M - 1)\} \right] + \mathbb{P}[i(\mathbf{X}^L; \mathbf{Y}^L) \leq \log(M - 1)]. \quad (68)$$

To show that (68) yields the lower bound

$$R^*(L, T, \epsilon, \rho) \geq \frac{I(T, \rho)}{T} - \sqrt{\frac{U(T, \rho)}{LT^2}} Q^{-1}(\epsilon) + \mathsf{K}_{\text{DT}}(L, T, \rho) \quad (69)$$

where for every  $T > 2$  and some  $A$ ,  $L_0$ , and  $\rho_0$  independent of  $L$ , and  $\rho$ ,

$$\sup_{\rho \geq \rho_0} |\mathsf{K}_{\text{DT}}(L, T, \rho)| \leq \frac{A}{L}, \quad L \geq L_0 \quad (70)$$

we follow almost verbatim the steps in [3, Eqs. (258)–(267)] (with  $\gamma$  in [3] replaced by  $M - 1$ ). The main difference is that, in our case,  $U(T, \rho)$  defined in (36) and  $B(T, \rho)$  defined as (cf. [3, Eq. (254)])

$$B(T, \rho) \triangleq \frac{6\mathbb{E}\left[|i_\ell(T, \rho) - I(T, \rho)|^3\right]}{U(T, \rho)^{3/2}} \quad (71)$$

depend on  $\rho$ . To ensure that the term  $\mathsf{K}_{\text{DT}}(L, T, \rho)$  in (69) is uniform in  $\rho$ , we will show that both  $U(T, \rho)$  and  $B(T, \rho)$  are bounded in  $\rho$ . We then apply the Berry-Esseen theorem [34, Ch. XVI.5] to obtain [3, Eq. (259)] with  $B(T, \rho)$  replaced by an upper bound  $B(T, \rho_0)$  that holds for all  $\rho \geq \rho_0$  and a sufficiently large  $\rho_0$ , followed by [3, Eqs. (261)–(265)], which gives

$$R^*(L, T, \epsilon, \rho) \geq \frac{I(T, \rho)}{T} - \sqrt{\frac{U(T, \rho)}{LT^2}} Q^{-1}(\tau) \quad (72)$$

where

$$\tau = \epsilon - \left(\frac{2\log 2}{\sqrt{2\pi}} + 5B(T, \rho_0)\right) \frac{1}{\sqrt{L}}. \quad (73)$$

A Taylor-series expansion of  $Q^{-1}(\tau)$  around  $\epsilon$  yields then that

$$\sup_{\rho \geq \rho_0} |Q^{-1}(\tau) - Q^{-1}(\epsilon)| \leq \frac{A}{\sqrt{L}}, \quad L \geq L_0 \quad (74)$$

for every  $T > 2$  and some  $A$ ,  $L_0$ , and  $\rho_0$  independent of  $L$  and  $\rho$ . Combining (74) with (72), we obtain (69).

To show that  $U(T, \rho)$  and  $B(T, \rho)$  are bounded in  $\rho$ , we resort to the following lemmas:

*Lemma 3:* Let  $0 \leq \delta \leq 1/2$  and let  $\bar{V}_\rho(T, \alpha)$  be defined in (37). For every  $\rho(1 - \delta) \leq \alpha \leq \rho$ , we have

$$\bar{V}_\rho(T, \alpha) \geq \left(\frac{T\rho}{1 + T\rho}\right)^2 (T - 1) - \Xi(T)\delta + \mathsf{K}_{\bar{V}}(T, \rho) \quad (75)$$

where  $\mathsf{K}_{\bar{V}}(T, \rho)$  is a function of  $T$  and  $\rho$  that satisfies

$$\lim_{\rho \rightarrow \infty} \mathsf{K}_{\bar{V}}(T, \rho) = 0, \quad T > 2 \quad (76)$$

and  $\Xi(T)$  is a positive constant that only depends on  $T$ .

*Proof:* See Appendix B. ■

*Lemma 4:* For every  $\rho_0 > 0$  and  $T > 2$ , we have

$$\sup_{\substack{\alpha \geq 0, \\ \rho \geq \rho_0}} \bar{V}_\rho(T, \alpha) < \infty \quad (77)$$

$$\sup_{\rho \geq \rho_0} U(T, \rho) < \infty. \quad (78)$$

*Proof:* See Appendix C. ■

*Lemma 5:* For every  $\rho_0 > 0$  and  $T > 2$ , we have

$$\sup_{\substack{\alpha \geq 0, \\ \rho \geq \rho_0}} \mathbb{E}\left[|\bar{j}_\ell(T, \alpha) - \bar{J}(T, \alpha)|^3\right] < \infty \quad (79)$$

$$\sup_{\rho \geq \rho_0} \mathbb{E}\left[|i_\ell(T, \rho) - I(T, \rho)|^3\right] < \infty. \quad (80)$$

*Proof:* See Appendix D. ■

For  $\delta = 0$ , Lemma 3 yields

$$\bar{V}(T, \rho) \geq \left(\frac{T\rho}{1 + T\rho}\right)^2 (T - 1) + \mathsf{K}_{\bar{V}}(T, \rho) \quad (81)$$

where  $\mathsf{K}_{\bar{V}}(T, \rho)$  satisfies (75). Together with (39), (41), and (42), this implies that

$$U(T, \rho) \geq \left(\frac{T\rho_0}{1 + T\rho_0}\right)^2 \frac{T - 1}{2}, \quad \rho \geq \rho_0 \quad (82)$$

for a sufficiently large  $\rho_0$ . Furthermore, Lemma 4 implies that, for every  $\rho_0 > 0$ , there exists an  $U_{\text{UB}}(T, \rho_0)$  that is independent of  $\rho$  and that satisfies

$$U(T, \rho) \leq U_{\text{UB}}(T, \rho_0), \quad \rho \geq \rho_0. \quad (83)$$

Finally, Lemma 5 implies that for every  $\rho_0 > 0$  there exists an  $S(T, \rho_0)$  that is independent of  $\rho$  and that satisfies

$$\mathbb{E}\left[|i_\ell(T, \rho) - I(T, \rho)|^3\right] \leq S(T, \rho_0), \quad \rho \geq \rho_0. \quad (84)$$

Combining (82) and (84), it follows that for a sufficiently large  $\rho_0 > 0$  there exists a  $B(T, \rho_0)$  that is independent of  $\rho$  and that satisfies

$$B(T, \rho) \leq \frac{6S(T, \rho_0)}{\left(\frac{T\rho_0}{1 + T\rho_0}\right)^3 \left(\frac{T-1}{2}\right)^{3/2}} \triangleq B(T, \rho_0), \quad \rho \geq \rho_0. \quad (85)$$

This concludes the proof of the lower bound (69).

## B. Meta Converse Upper Bound

An upper bound on  $R^*(L, T, \epsilon, \rho)$  follows from the MC bound [3, Th. 31] computed for the auxiliary pdf  $q_{\mathbf{Y}^L}^{(U)}$ , i.e.,

$$R^*(L, T, \epsilon, \rho) \leq \frac{1}{LT} \sup_{\alpha \in [0, \rho]^L} \log\left(\frac{1}{\beta(\alpha, q_{\mathbf{Y}^L}^{(U)})}\right). \quad (86)$$

Here,  $\alpha = (\alpha_1, \dots, \alpha_L)$  denotes the vector of power allocations, and  $\beta(\alpha, q_{\mathbf{Y}^L}^{(U)})$  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  $\epsilon$  [3, Eq. (100)]. Note that, by (13)–(16),  $\beta(\alpha, q_{\mathbf{Y}^L}^{(U)})$  depends on  $\mathbf{x}^L$  only via  $\alpha$  (recall that  $\|\mathbf{X}_\ell\|^2 = T\alpha_\ell$ ).

For  $0 < \delta < 1$ , let  $L_\delta(\alpha)$  denote the number of  $\alpha_\ell$ 's in  $\alpha$  that satisfy  $\rho(1 - \delta) \leq \alpha_\ell \leq \rho$ . The following lemma demonstrates that we can assume without loss of optimality that  $L_\delta(\alpha) \geq L/2$ , i.e., in at least half of the coherence intervals  $\alpha_\ell$  is larger than  $\rho(1 - \delta)$ .

*Lemma 6:* Let

$$\mathcal{A}_{\rho, \delta} \triangleq \{\alpha \in [0, \rho]^L : L_\delta(\alpha) \geq L/2\}. \quad (87)$$

For every  $0 < \delta < 1$ ,  $T > 2$ , and  $0 < \epsilon < 1/2$ , there exists a pair  $(L_0, \rho_0)$  independent of  $L$  and  $\rho$  such that, for  $L \geq L_0$

and  $\rho \geq \rho_0$ , the supremum in (86) can be replaced without loss of optimality by a supremum over  $\alpha \in \mathcal{A}_{\rho, \delta}$ .

*Proof:* See Appendix E. ■

In the following, we implicitly assume that  $L \geq L_0$  and  $\rho \geq \rho_0$  for some sufficiently large  $L_0$  and  $\rho_0$  so that Lemma 6 holds. Applying Lemma 6 to (86), and upper-bounding the right-hand side (RHS) of (86) using [3, Eq. (106)] and (29), we obtain

$$R^*(L, T, \epsilon, \rho) \leq \sup_{\alpha \in \mathcal{A}_{\rho, \delta}} \left\{ \frac{\log \xi(\alpha)}{LT} - \frac{\log(1 - \epsilon - \mathbb{P}[\sum_{\ell=1}^L \bar{j}_\ell(T, \alpha_\ell) \geq \log \xi(\alpha)])}{LT} \right\} \quad (88)$$

for every  $\xi: [0, \rho]^L \rightarrow (0, \infty)$ .

Let

$$\bar{B}(T, \alpha) \triangleq \frac{6 \sum_{\ell=1}^L \mathbb{E}[|\bar{j}_\ell(T, \alpha_\ell) - \bar{J}(T, \alpha)|^3]}{\left(\sum_{\ell=1}^L \bar{V}_\rho(T, \alpha_\ell)\right)^{3/2}}. \quad (89)$$

By Lemma 5, the expectation  $\mathbb{E}[|\bar{j}_\ell(T, \alpha) - \bar{J}(T, \alpha)|^3]$  can be upper-bounded by a constant  $\bar{S}(T, \rho_0)$  that is independent of  $\alpha$  and  $\rho$ . Furthermore, by the nonnegativity of  $\bar{V}_\rho(T, \alpha_\ell)$ ,

$$\sum_{\ell=1}^L \bar{V}_\rho(T, \alpha_\ell) \geq \sum_{\ell \in \mathcal{L}_\delta(\alpha)} \bar{V}_\rho(T, \alpha_\ell) \quad (90)$$

where  $\mathcal{L}_\delta(\alpha) \triangleq \{\ell = 1, \dots, L : \alpha_\ell \geq \rho(1 - \delta)\}$ . Lemma 3 demonstrates that, for  $\alpha \geq \rho(1 - \delta)$ ,

$$\bar{V}_\rho(T, \alpha) \geq \left(\frac{T\rho}{1+T\rho}\right)^2 (T-1) - \Xi(T)\delta + \mathsf{K}_{\bar{V}}(T, \rho). \quad (91)$$

Thus, for

$$\delta = \left(\frac{T\rho_0}{1+T\rho_0}\right)^2 \frac{T-1}{3\Xi(T)} \quad (92)$$

and  $\rho_0$  sufficiently large, we have

$$\sum_{\ell=1}^L \bar{V}_\rho(T, \alpha_\ell) \geq L_\delta(\alpha) \left(\frac{T\rho_0}{1+T\rho_0}\right)^2 \frac{T-1}{2}, \quad \rho \geq \rho_0. \quad (93)$$

It follows that, for every  $\alpha \in \mathcal{A}_{\rho, \delta}$  and  $\delta$  as chosen in (92),

$$\bar{B}(T, \alpha) \leq \frac{6L\bar{S}(T, \rho_0)}{\left(\frac{(T-1)L}{4}\right)^{3/2} \left(\frac{T\rho_0}{1+T\rho_0}\right)^3} \triangleq \frac{\bar{B}(T, \rho_0)}{\sqrt{L}}. \quad (94)$$

Let

$$\lambda = Q^{-1}\left(\epsilon + \frac{2\bar{B}(T, \rho_0)}{\sqrt{L}}\right) \quad (95)$$

and

$$\log \xi(\alpha) = \sum_{\ell=1}^L \bar{J}(T, \alpha_\ell) - \lambda \sqrt{\sum_{\ell=1}^L \bar{V}_\rho(T, \alpha_\ell)}. \quad (96)$$

With this choice, the Berry-Esseen theorem and (94) imply that, for every  $\alpha \in \mathcal{A}_{\rho, \delta}$ ,

$$\left| \mathbb{P}\left[\sum_{\ell=1}^L \bar{j}_\ell(T, \alpha_\ell) \leq \log \xi(\alpha)\right] - Q(\lambda) \right| \leq \frac{\bar{B}(T, \rho_0)}{\sqrt{L}}. \quad (97)$$

Thus, for such  $\alpha$ 's,

$$\mathbb{P}\left[\sum_{\ell=1}^L \bar{j}_\ell(T, \alpha_\ell) \leq \log \xi(\alpha)\right] \geq \epsilon + \frac{\bar{B}(T, \rho_0)}{\sqrt{L}}. \quad (98)$$

Substituting (98) into the upper bound (88), we obtain

$$\begin{aligned} R^*(L, T, \epsilon, \rho) &\leq \sup_{\alpha \in \mathcal{A}_{\rho, \delta}} \left\{ \frac{\sum_{\ell=1}^L \bar{J}(T, \alpha_\ell)}{LT} - \sqrt{\frac{\sum_{\ell=1}^L \bar{V}_\rho(T, \alpha_\ell)}{L^2 T^2}} Q^{-1}\left(\epsilon + \frac{2\bar{B}(T, \rho_0)}{\sqrt{L}}\right) \right\} \\ &\quad - \frac{\log \bar{B}(T, \rho_0)}{LT} + \frac{1}{2} \frac{\log L}{LT}. \end{aligned} \quad (99)$$

By the assumption  $0 < \epsilon < \frac{1}{2}$ , the inverse  $Q$ -function on the RHS of (99) is positive for sufficiently large  $L$ . It follows by the concavity of the square-root function and Jensen's inequality that (99) can be further upper-bounded as

$$\begin{aligned} R^*(L, T, \epsilon, \rho) &\leq \frac{1}{L} \sum_{\ell=1}^L \sup_{0 \leq \alpha_\ell \leq \rho} \left\{ \frac{\bar{J}(T, \alpha_\ell)}{T} - \sqrt{\frac{\bar{V}_\rho(T, \alpha_\ell)}{L T^2}} Q^{-1}\left(\epsilon + \frac{2\bar{B}(T, \rho_0)}{\sqrt{L}}\right) \right\} \\ &\quad - \frac{\log \bar{B}(T, \rho_0)}{LT} + \frac{1}{2} \frac{\log L}{LT} \\ &= \sup_{0 \leq \alpha \leq \rho} \left\{ \frac{\bar{J}(T, \alpha)}{T} - \sqrt{\frac{\bar{V}_\rho(T, \alpha)}{L T^2}} Q^{-1}\left(\epsilon + \frac{2\bar{B}(T, \rho_0)}{\sqrt{L}}\right) \right\} \\ &\quad - \frac{\log \bar{B}(T, \rho_0)}{LT} + \frac{1}{2} \frac{\log L}{LT} \end{aligned} \quad (100)$$

where the second step follows because the channel is block-wise i.i.d., so the terms inside the curly brackets do not depend on  $\ell$ .

Performing a Taylor-series expansion of the inverse  $Q$ -function around  $\epsilon$ , we obtain

$$\sup_{\rho \geq \rho_0} \left| Q^{-1}\left(\epsilon + \frac{2\bar{B}(T, \rho_0)}{\sqrt{L}}\right) - Q^{-1}(\epsilon) \right| \leq \frac{\mathsf{A}}{\sqrt{L}}, \quad L \geq L_0 \quad (101)$$

for every  $T > 2$  and some  $\mathsf{A}$ ,  $L_0$ , and  $\rho_0$  independent of  $L$  and  $\rho$ . Further using that, by Lemma 4,  $\bar{V}_\rho(T, \alpha)$  is bounded in  $\rho$  and  $\alpha$ , and collecting terms of order  $\log L/L$ , we can rewrite (100) as

$$R^*(L, T, \epsilon, \rho) \leq \sup_{0 \leq \alpha \leq \rho} \left\{ \frac{\bar{J}(T, \alpha)}{T} - \sqrt{\frac{\bar{V}_\rho(T, \alpha)}{L T^2}} Q^{-1}(\epsilon) \right\} + \mathsf{K}_1(L, T, \rho) \quad (102)$$

where  $\mathsf{K}_1(L, T, \rho)$  is a function of  $L$ ,  $T$ , and  $\rho$  that satisfies

$$\sup_{\rho \geq \rho_0} |\mathsf{K}_1(L, T, \rho)| \leq \mathsf{A} \frac{\log L}{L}, \quad L \geq L_0 \quad (103)$$

for every  $T > 2$  and some  $\mathsf{A}$ ,  $L_0$ , and  $\rho_0$  independent of  $L$  and  $\rho$ .

We next show that

$$\begin{aligned} & \sup_{0 \leq \alpha \leq \rho} \left\{ \frac{\bar{J}(T, \alpha)}{T} - \sqrt{\frac{\bar{V}_\rho(T, \alpha)}{LT^2}} Q^{-1}(\epsilon) \right\} \\ &= \frac{\bar{J}(T, \rho)}{T} - \sqrt{\frac{\bar{V}(T, \rho)}{LT^2}} Q^{-1}(\epsilon) + \mathsf{K}_2(L, T, \rho) \end{aligned} \quad (104)$$

where  $\mathsf{K}_2(L, T, \rho)$  is a function of  $L$ ,  $T$ , and  $\rho$  that satisfies

$$\sup_{\rho \geq \rho_0} |\mathsf{K}_2(L, T, \rho)| \leq \frac{\mathsf{A}}{L}, \quad L \geq L_0 \quad (105)$$

for every  $T > 2$  and some  $\mathsf{A}$ ,  $L_0$ , and  $\rho_0$  independent of  $L$  and  $\rho$ . By using (40) and (41), we finally obtain the desired upper bound

$$\begin{aligned} R^*(L, T, \epsilon, \rho) &\leq \frac{\bar{I}(T, \rho) + \mathsf{K}_{\bar{J}}(T, \rho)}{T} \\ &\quad - \sqrt{\frac{\bar{U}(T) + \mathsf{K}_{\bar{V}}(T, \rho)}{LT^2}} Q^{-1}(\epsilon) + \mathsf{K}_{\text{MC}}(L, T, \rho) \end{aligned} \quad (106)$$

where  $\mathsf{K}_{\text{MC}}(L, T, \rho)$  is a function of  $L$ ,  $T$ , and  $\rho$  that satisfies

$$\sup_{\rho \geq \rho_0} |\mathsf{K}_{\text{MC}}(L, T, \rho)| \leq \mathsf{A} \frac{\log L}{L}, \quad L \geq L_0 \quad (107)$$

for every  $T > 2$  and some  $\mathsf{A}$ ,  $L_0$ , and  $\rho_0$  independent of  $L$  and  $\rho$ ; and  $\mathsf{K}_\xi(T, \rho)$ ,  $\xi = \{\bar{J}, \bar{V}\}$  are functions of  $T$  and  $\rho$  that satisfy

$$\lim_{\rho \rightarrow \infty} \mathsf{K}_\xi(T, \rho) = 0, \quad T > 2. \quad (108)$$

To prove (104), we first present the following auxiliary results.

*Lemma 7:*

1) Assume that  $T > 2$ . For sufficiently large  $\rho$ , we have

$$\sup_{0 \leq \alpha \leq \rho} \bar{J}(T, \alpha) = \bar{J}(T, \rho). \quad (109)$$

2) Assume that  $T > 2$  and  $0 < \epsilon < \frac{1}{2}$ . Consider the supremum on the left-hand side (LHS) of (104). For sufficiently large  $L$  and  $\rho$ , we can assume without loss of optimality that  $\alpha \in [\rho(1 - \frac{\mathsf{K}(T)}{L}), \rho]$  for some nonnegative constant  $\mathsf{K}(T)$  that is independent of  $L$ ,  $\rho$ , and  $\alpha$ .

*Proof:* See Appendix G. ■

We next set out to prove (104). By Part 2) of Lemma 7, we can assume without loss of optimality that

$$\alpha \geq \rho \left( 1 - \frac{\mathsf{K}(T)}{L} \right). \quad (110)$$

Furthermore, we show in Appendix H that

$$\bar{V}_\rho(T, \alpha) \geq \bar{V}(T, \rho) - \Upsilon(T)\delta, \quad \rho(1 - \delta) \leq \alpha \leq \rho \quad (111)$$

where  $\Upsilon(T)$  is a positive constant that only depends on  $T$ . Particularizing this bound for  $\delta = \mathsf{K}(T)/L$ , we obtain

$$\bar{V}_\rho(T, \alpha) \geq \bar{V}(T, \rho) - \Upsilon(T) \frac{\mathsf{K}(T)}{L}, \quad \rho \left( 1 - \frac{\mathsf{K}(T)}{L} \right) \leq \alpha \leq \rho. \quad (112)$$

Combining (112) with Part 1) of Lemma 7, and using that, by the assumption  $0 < \epsilon < \frac{1}{2}$  we have  $Q^{-1}(\epsilon) > 0$ , we obtain

$$\begin{aligned} & \sup_{0 \leq \alpha \leq \rho} \left\{ \frac{\bar{J}(T, \alpha)}{LT} - \sqrt{\frac{\bar{V}_\rho(T, \alpha)}{LT^2}} Q^{-1}(\epsilon) \right\} \\ &\leq \frac{\bar{J}(T, \rho)}{T} - \sqrt{\frac{\bar{V}(T, \rho) - \frac{\Upsilon(T)\mathsf{K}(T)}{L}}{LT^2}} Q^{-1}(\epsilon) \\ &= \frac{\bar{J}(T, \rho)}{T} - \sqrt{\frac{\bar{V}(T, \rho)}{LT^2}} Q^{-1}(\epsilon) + \mathsf{K}_2(L, T, \rho) \end{aligned} \quad (113)$$

where  $\mathsf{K}_2(L, T, \rho)$  is as in (104). This proves (104) and concludes the proof of the upper bound.

## VII. CONCLUSION

We presented a high-SNR normal approximation for the maximum coding rate  $R^*(L, T, \epsilon, \rho)$  achievable over noncoherent, single-antenna, Rayleigh block-fading channels using an error-correcting code that spans  $L$  coherence intervals of length  $T$ , has a block-error probability no larger than  $\epsilon$ , and satisfies the power constraint  $\rho$ . The high-SNR normal approximation is roughly equal to the normal approximation 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 the heuristic that, at high SNR, one pilot symbol per coherence block should be transmitted to achieve both the capacity and the channel dispersion. While the approximation was derived under the assumption that the number of coherence intervals  $L$  and the SNR  $\rho$  tend to infinity, numerical analyses suggest that it becomes accurate already at SNR values of 15 dB and for 10 coherence intervals or more.

The obtained normal approximation is useful in two ways. First, it complements the nonasymptotic bounds provided in [12], [13], [14], whose evaluation is computationally demanding. Second, it lays the foundation for analytical studies that analyze the behavior of the maximum coding rate as a function of system parameters such as SNR, number of coherence intervals, or blocklength. An example of such a study was illustrated in Section V-C concerning the optimal design of a simple slotted-ALOHA protocol.

## APPENDIX A PROOF OF LEMMA 1

The left-most inequality in (12) follows because the regularized lower incomplete gamma function is no larger than 1. For the right-most inequality in (12), consider the following bound by Alzer [35, Th. 1] (see also [36, Eq. (5.4)])

$$\tilde{\gamma}(a, x) > (1 - e^{-s_a x})^a, \quad (x \geq 0, a > 0, a \neq 1) \quad (114)$$

where

$$s_a = \begin{cases} 1, & \text{if } 0 < a < 1 \\ \Gamma(a+1)^{-\frac{1}{a}}, & \text{if } a > 1. \end{cases} \quad (115)$$

In order to obtain the right-most inequality in (12), we first lower-bound  $\tilde{\gamma}(\cdot, \cdot)$  using (114)

$$\begin{aligned} \log \frac{1}{\tilde{\gamma}(T-1, x)} &\leq (T-1) \log \left( \frac{1}{1 - e^{-x\Gamma(T)^{-\frac{1}{T-1}}}} \right) \\ &= (T-1) \log \left( 1 + \frac{1}{e^{x\Gamma(T)^{-\frac{1}{T-1}}} - 1} \right) \end{aligned} \quad (116)$$

where the second step follows by simple algebraic manipulations. Since  $e^z \geq 1 + z$ , this can be further upper-bounded as

$$\log \frac{1}{\tilde{\gamma}(T-1, x)} \leq (T-1) \log \left( 1 + \frac{\Gamma(T)^{\frac{1}{T-1}}}{x} \right). \quad (117)$$

This proves Lemma 1.

### APPENDIX B PROOF OF LEMMA 3

To lower-bound  $\bar{V}_\rho(T, \alpha)$ , we begin by lower-bounding

$$\begin{aligned} &(\bar{j}_\ell(T, \alpha) - \bar{J}(T, \alpha))^2 \\ &= \left( -\frac{T\rho - T\alpha}{1 + T\rho}(Z_1 - 1) - \frac{T\rho}{1 + T\rho}(Z_2 - (T-1)) \right. \\ &\quad \left. + (T-1) \log((1 + T\alpha)Z_1 + Z_2 + \beta(T, \rho)) \right. \\ &\quad \left. - (T-1) \mathbb{E}[\log((1 + T\alpha)Z_1 + Z_2 + \beta(T, \rho))] \right)^2 \\ &\geq \left( \frac{T\rho - T\alpha}{1 + T\rho}(Z_1 - 1) + \frac{T\rho}{1 + T\rho}(Z_2 - (T-1)) \right)^2 \\ &\quad - 2 \left( \frac{T\rho - T\alpha}{1 + T\rho}(Z_1 - 1) + \frac{T\rho}{1 + T\rho}(Z_2 - (T-1)) \right) \\ &\quad \times \left( (T-1) \log((1 + T\alpha)Z_1 + Z_2) \right. \\ &\quad \left. - (T-1) \mathbb{E}[\log((1 + T\alpha)Z_1 + Z_2)] \right. \\ &\quad \left. + (T-1) \log \left( 1 + \frac{\beta(T, \rho)}{(1 + T\alpha)Z_1 + Z_2} \right) \right. \\ &\quad \left. - (T-1) \mathbb{E} \left[ \log \left( 1 + \frac{\beta(T, \rho)}{(1 + T\alpha)Z_1 + Z_2} \right) \right] \right). \end{aligned} \quad (118)$$

We next note that

$$\mathbb{E} \left[ (Z_1 - 1) \log \left( 1 + \frac{\beta(T, \rho)}{(1 + T\alpha)Z_1 + Z_2} \right) \right] \leq 0 \quad (119)$$

$$\mathbb{E} \left[ (Z_2 - (T-1)) \log \left( 1 + \frac{\beta(T, \rho)}{(1 + T\alpha)Z_1 + Z_2} \right) \right] \leq 0. \quad (120)$$

This follows from the inequalities

$$\begin{aligned} &(Z_1 - 1) \log \left( 1 + \frac{\beta(T, \rho)}{(1 + T\alpha)Z_1 + Z_2} \right) \\ &\leq (Z_1 - 1) \log \left( 1 + \frac{\beta(T, \rho)}{(1 + T\alpha) + Z_2} \right) \end{aligned} \quad (121)$$

and

$$\begin{aligned} &(Z_2 - (T-1)) \log \left( 1 + \frac{\beta(T, \rho)}{(1 + T\alpha)Z_1 + Z_2} \right) \\ &\leq (Z_2 - (T-1)) \log \left( 1 + \frac{\beta(T, \rho)}{(1 + T\alpha)Z_1 + (T-1)} \right) \end{aligned} \quad (122)$$

whose RHSs are zero mean because  $Z_1$  and  $Z_2$  are independent and have mean 1 and  $T-1$ , respectively.

Computing the expected value of (118), and using (119) and (120), we can lower-bound  $\bar{V}_\rho(T, \alpha)$  as

$$\begin{aligned} &\bar{V}_\rho(T, \alpha) \\ &\triangleq \mathbb{E}[(\bar{j}_\ell(T, \alpha) - \bar{J}(T, \alpha))^2] \\ &\geq \left( \frac{T\rho - T\alpha}{1 + T\rho} \right)^2 + \left( \frac{T\rho}{1 + T\rho} \right)^2 (T-1) \\ &\quad - \kappa_{T, \rho} \left( 1 - \frac{\alpha}{\rho} \right) \mathbb{E} \left[ (Z_1 - 1) \log \left( Z_1 + \frac{Z_2}{1 + T\alpha} \right) \right] \\ &\quad - \kappa_{T, \rho} \mathbb{E} \left[ (Z_2 - (T-1)) \log \left( Z_1 + \frac{Z_2}{1 + T\rho} \right) \right] \\ &\quad + \kappa_{T, \rho} \mathbb{E} \left[ (Z_2 - (T-1)) \log \left( \frac{(1 + T\rho)Z_1 + Z_2}{(1 + T\alpha)Z_1 + Z_2} \right) \right] \end{aligned} \quad (123)$$

where  $\kappa_{T, \rho} \triangleq 2(T-1)T\rho/(1 + T\rho)$ .

The first term on the RHS of (123) is nonnegative, so discarding it yields a lower bound. The third term on the RHS of (123) can be lower-bounded by upper-bounding

$$\begin{aligned} &\kappa_{T, \rho} \left( 1 - \frac{\alpha}{\rho} \right) \mathbb{E} \left[ (Z_1 - 1) \log \left( Z_1 + \frac{Z_2}{1 + T\alpha} \right) \right] \\ &\leq \kappa_{T, \rho} \left( 1 - \frac{\alpha}{\rho} \right) \sqrt{\mathbb{E}[(Z_1 - 1)^2] \mathbb{E} \left[ \log^2 \left( Z_1 + \frac{Z_2}{1 + T\alpha} \right) \right]} \\ &\leq 2(T-1) \delta \sqrt{\left( \frac{\pi^2}{6} + \gamma^2 + \psi^2(T) + \zeta(2, T) \right)} \end{aligned} \quad (124)$$

for  $\rho(1 - \delta) \leq \alpha \leq \rho$ . Here, the first inequality follows from the Cauchy-Schwarz inequality, and the last inequality follows because  $\kappa_{T, \rho} \leq 2(T-1)$ , because  $Z_1$  has variance 1, and because

$$\begin{aligned} &\mathbb{E} \left[ \log^2 \left( Z_1 + \frac{Z_2}{1 + T\alpha} \right) \right] \\ &\leq \mathbb{E}[\log^2(Z_1 + Z_2)] + \mathbb{E}[\log^2(Z_1)] \\ &= \frac{\pi^2}{6} + \gamma^2 + \zeta(2, T) + \psi^2(T) \end{aligned} \quad (125)$$

where we have evaluated the expected values using [19, Sec. 4.335-1] and [19, Sec. 4.358-2], respectively. The first inequality in (125) follows by treating the cases  $Z_1 + Z_2/(1 + T\alpha) \leq 1$  and  $Z_1 + Z_2/(1 + T\alpha) > 1$  separately, and by lower-bounding in the former case  $Z_1 + Z_2/(1 + T\alpha)$  by  $Z_1$  and upper-bounding in the latter case  $Z_1 + Z_2/(1 + T\alpha)$  by  $Z_1 + Z_2$ . Hence

$$\begin{aligned} \log^2 \left( Z_1 + \frac{Z_2}{1 + T\alpha} \right) &\leq \log^2(Z_1) \\ &\leq \log^2(Z_1) + \log^2(Z_1 + Z_2) \end{aligned} \quad (126)$$

if  $Z_1 + Z_2/(1 + T\alpha) \leq 1$ , and

$$\begin{aligned} \log^2 \left( Z_1 + \frac{Z_2}{1 + T\alpha} \right) &\leq \log^2(Z_1 + Z_2) \\ &\leq \log^2(Z_1) + \log^2(Z_1 + Z_2) \end{aligned} \quad (127)$$

if  $Z_1 + Z_2/(1 + T\alpha) > 1$ , which yields the desired bound.

The fifth term on the RHS (123) can be lower-bounded by upper-bounding

$$\begin{aligned} & \left| \mathbb{E} \left[ (Z_2 - (T-1)) \log \left( \frac{(1+T\rho)Z_1 + Z_2}{(1+T\alpha)Z_1 + Z_2} \right) \right] \right| \\ & \leq \mathbb{E} \left[ |Z_2 - (T-1)| \log \left( \frac{(1+T\rho)Z_1 + Z_2}{(1+T\alpha)Z_1 + Z_2} \right) \right] \\ & \leq \mathbb{E} [|Z_2 - (T-1)|] \log \left( \frac{\rho}{\alpha} \right) \\ & \leq \mathbb{E} [|Z_2 - (T-1)|] \log \left( \frac{1}{1-\delta} \right) \end{aligned} \quad (128)$$

for  $\rho(1-\delta) \leq \alpha \leq \rho$ . Combining (124)–(128) with (123), and upper-bounding  $\kappa_{T,\rho}$  by  $2(T-1)$ , we obtain the lower bound

$$\begin{aligned} & \bar{V}_\rho(T, \alpha) \\ & \geq \left( \frac{T\rho}{1+T\rho} \right)^2 (T-1) \\ & \quad - 2(T-1)\delta \sqrt{\left( \frac{\pi^2}{6} + \gamma^2 + \psi^2(T) + \zeta(2, T) \right)} \\ & \quad - \kappa_{T,\rho} \mathbb{E} \left[ (Z_2 - (T-1)) \log \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \right] \\ & \quad - 2(T-1) \mathbb{E} [|Z_2 - (T-1)|] \log \left( \frac{1}{1-\delta} \right). \end{aligned} \quad (129)$$

Only the second and fourth term on the RHS of (129) depend on  $\delta$ . The former term is linear in  $\delta$ , the latter term can be upper-bounded by a linear term by using that, for  $0 \leq \delta \leq 1/2$ ,

$$\log \left( \frac{1}{1-\delta} \right) \leq \frac{\delta}{1-\delta} \leq 2\delta. \quad (130)$$

Hence, there exists a positive constant  $\Xi(T)$  that only depends on  $T$  such that

$$\begin{aligned} & \bar{V}_\rho(T, \alpha) \geq \left( \frac{T\rho}{1+T\rho} \right)^2 (T-1) - \Xi(T)\delta \\ & \quad - \kappa_{T,\rho} \mathbb{E} \left[ (Z_2 - (T-1)) \log \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \right]. \end{aligned} \quad (131)$$

We conclude the proof of Lemma 3 by demonstrating that, for every  $T$ ,

$$\lim_{\rho \rightarrow \infty} \mathbb{E} \left[ (Z_2 - (T-1)) \log \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \right] = 0. \quad (132)$$

This is a direct consequence of the dominated convergence theorem [37, Section 1.26], which can be applied because

$$\begin{aligned} & \left| (Z_2 - (T-1)) \log \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \right| \\ & \leq |(Z_2 - (T-1))| \sqrt{\log^2(Z_1) + \log^2(Z_1 + Z_2)} \end{aligned} \quad (133)$$

which follows from the same steps as the first inequality in (125). Using the Cauchy-Schwarz inequality, the expected value of the RHS of (133) can be upper-bounded by

$$\sqrt{\mathbb{E} \left[ (Z_2 - (T-1))^2 \right] \mathbb{E} \left[ \log^2(Z_1) + \log^2(Z_1 + Z_2) \right]}$$

which is finite by (125) and because  $Z_2$  has finite variance.

## APPENDIX C PROOF OF LEMMA 4

We shall first prove (77). Using the definitions of  $\bar{j}_\ell(T, \alpha)$  and  $\bar{J}(T, \alpha)$  in (28) and (30), respectively, we upper-bound  $\bar{V}_\rho(T, \alpha)$  as

$$\begin{aligned} & \bar{V}_\rho(T, \alpha) \\ & = \mathbb{E} \left[ \left( \frac{T\rho - T\alpha}{1+T\rho} (1 - Z_1) + \frac{T\rho}{1+T\rho} (T-1 - Z_2) \right. \right. \\ & \quad \left. \left. + (T-1) \log \left( Z_1 + \frac{Z_2}{1+T\alpha} \right) \right. \right. \\ & \quad \left. \left. - (T-1) \mathbb{E} \left[ \log \left( Z_1 + \frac{Z_2}{1+T\alpha} \right) \right] \right. \right. \\ & \quad \left. \left. + (T-1) \log \left( 1 + \frac{\beta(T, \rho)}{(1+T\alpha)Z_1 + Z_2} \right) \right. \right. \\ & \quad \left. \left. - (T-1) \mathbb{E} \left[ \log \left( 1 + \frac{\beta(T, \rho)}{(1+T\alpha)Z_1 + Z_2} \right) \right] \right)^2 \right] \\ & \leq c_{4,2} \left( \frac{T\rho - T\alpha}{1+T\rho} \right)^2 \mathbb{E} [(Z_1 - 1)^2] \\ & \quad + c_{4,2} \left( \frac{T\rho}{1+T\rho} \right)^2 \mathbb{E} [(Z_2 - T + 1)^2] \\ & \quad + c_{4,2} (T-1)^2 \mathbb{E} \left[ \log^2 \left( Z_1 + \frac{Z_2}{1+T\alpha} \right) \right] \\ & \quad + c_{4,2} (T-1)^2 \mathbb{E} \left[ \log^2 \left( 1 + \frac{\beta(T, \rho)}{(1+T\alpha)Z_1 + Z_2} \right) \right] \end{aligned} \quad (134)$$

where we have used that

$$|a_1 + \dots + a_\eta|^\nu \leq c_{\eta,\nu} (|a_1|^\nu + \dots + |a_\eta|^\nu), \quad \eta, \nu \in \mathbb{Z}^+ \quad (135)$$

for some positive constant  $c_{\eta,\nu}$  that only depends on  $\eta$  and  $\nu$ , and that  $\mathbb{E}[(X - \mathbb{E}[X])^2] \leq \mathbb{E}[X^2]$  for every random variable  $X$ .

We next show that each term on the RHS of (134) is bounded in  $\rho$  and  $\alpha$ . Indeed, we have  $\mathbb{E}[(Z_1 - 1)^2] = 1$  and  $\mathbb{E}[(Z_2 - (T-1))^2] = (T-1)$ . Furthermore, since  $0 \leq (T\rho - T\alpha)/(1+T\rho) \leq 1$  and  $0 \leq T\rho/(1+T\rho) \leq 1$ , the first two terms on the RHS of (134) are bounded in  $\rho$  and  $\alpha$ . The third term on the RHS of (134) can be upper-bounded by (see (125))

$$\begin{aligned} & (T-1)^2 \mathbb{E} \left[ \log^2 \left( Z_1 + \frac{Z_2}{1+T\alpha} \right) \right] \\ & \leq (T-1)^2 \mathbb{E} [\log^2(Z_1 + Z_2)] + (T-1)^2 \mathbb{E} [\log^2(Z_1)]. \end{aligned} \quad (136)$$

Finally, for every  $\rho_0 > 0$  and  $\rho \geq \rho_0$ , the fourth term on the RHS of (134) can be upper-bounded by

$$\begin{aligned} & \mathbb{E} \left[ (T-1)^2 \log^2 \left( 1 + \frac{\beta(T, \rho)}{(1+T\alpha)Z_1 + Z_2} \right) \right] \\ & \leq (T-1)^2 \mathbb{E} \left[ \log^2 \left( 1 + \frac{\beta(T, \rho)}{Z_1 + Z_2} \right) \right] \\ & \leq (T-1)^2 \mathbb{E} \left[ \log^2 \left( 1 + \frac{\beta(T, \rho_0)}{Z_1 + Z_2} \right) \right] \end{aligned} \quad (137)$$

where the second inequality follows because  $\rho \mapsto \beta(T, \rho)$  is monotonically decreasing in  $\rho$ . Since the RHSs of (136) and (137) are finite, this proves (77).

The proof of (78) follows along similar lines. Indeed, using the definitions of  $i_\ell(T, \rho)$  and  $I(T, \rho)$  in (19) and (21), respectively, we can upper-bound  $U(T, \rho)$  as

$$\begin{aligned}
 U(T, \rho) &= \mathbb{E} \left[ \left( \frac{T\rho}{1+T\rho}(T-1-Z_2) \right. \right. \\
 &\quad + (T-1) \log \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \\
 &\quad - (T-1) \mathbb{E} \left[ \log \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \right] \\
 &\quad - \log \tilde{\gamma} \left( T-1, \frac{T\rho((1+T\rho)Z_1+Z_2)}{1+T\rho} \right) \\
 &\quad \left. \left. + \mathbb{E} \left[ \log \tilde{\gamma} \left( T-1, \frac{T\rho((1+T\rho)Z_1+Z_2)}{1+T\rho} \right) \right] \right)^2 \right] \\
 &\leq c_{3,2} \left( \frac{T\rho}{1+T\rho} \right)^2 \mathbb{E}[(Z_2 - T + 1)^2] \\
 &\quad + c_{3,2}(T-1)^2 \mathbb{E} \left[ \log^2 \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \right] \\
 &\quad + c_{3,2} \mathbb{E} \left[ \log^2 \tilde{\gamma} \left( T-1, \frac{T\rho((1+T\rho)Z_1+Z_2)}{1+T\rho} \right) \right]. \tag{138}
 \end{aligned}$$

We next show that each summand is bounded in  $\rho$ . Indeed, as shown before, the first and the second term on the RHS of (138) are bounded in  $\rho$ . As for the third term on the RHS of (138), we use Lemma 1 to obtain

$$\begin{aligned}
 &\mathbb{E} \left[ \log^2 \tilde{\gamma} \left( T-1, \frac{T\rho((1+T\rho)Z_1+Z_2)}{1+T\rho} \right) \right] \\
 &\leq (T-1)^2 \mathbb{E} \left[ \log^2 \left( 1 + \frac{\beta(T, \rho)}{(1+T\rho)Z_1+Z_2} \right) \right]. \tag{139}
 \end{aligned}$$

By the monotonicity of  $\rho \mapsto \beta(T, \rho)$ , it follows that for every  $\rho_0 > 0$  and  $\rho \geq \rho_0$ , the third term on the RHS of (138) is upper-bounded by

$$\begin{aligned}
 &\mathbb{E} \left[ \log^2 \tilde{\gamma} \left( T-1, \frac{T\rho((1+T\rho)Z_1+Z_2)}{1+T\rho} \right) \right] \\
 &\leq (T-1)^2 \mathbb{E} \left[ \log^2 \left( 1 + \frac{\beta(T, \rho_0)}{Z_1+Z_2} \right) \right]. \tag{140}
 \end{aligned}$$

Combining the above steps with (138) yields (78).

#### APPENDIX D PROOF OF LEMMA 5

We shall first prove (79). Using the definitions of  $\bar{j}_\ell(T, \alpha)$  and  $\bar{J}(T, \alpha)$  in (28) and (30), respectively, we can upper-bound

the third moment of  $\bar{j}_\ell(T, \alpha)$  as

$$\begin{aligned}
 &\mathbb{E} \left[ |\bar{j}_\ell(T, \alpha) - \bar{J}(T, \alpha)|^3 \right] \\
 &= \mathbb{E} \left[ \left| \frac{T\rho - T\alpha}{1+T\rho}(1-Z_1) + \frac{T\rho}{1+T\rho}(T-1-Z_2) \right. \right. \\
 &\quad + (T-1) \log \left( Z_1 + \frac{Z_2}{1+T\alpha} \right) \\
 &\quad - (T-1) \mathbb{E} \left[ \log \left( Z_1 + \frac{Z_2}{1+T\alpha} \right) \right] \\
 &\quad - (T-1) \log \left( 1 + \frac{\beta(T, \rho)}{(1+T\alpha)Z_1+Z_2} \right) \\
 &\quad \left. \left. + (T-1) \mathbb{E} \left[ \log \left( 1 + \frac{\beta(T, \rho)}{(1+T\alpha)Z_1+Z_2} \right) \right] \right|^3 \right] \\
 &\leq c_{6,3} \left| \frac{T\rho - T\alpha}{1+T\rho} \right|^3 \mathbb{E}[|Z_1 - 1|^3] \\
 &\quad + c_{6,3} \left| \frac{T\rho}{1+T\rho} \right|^3 \mathbb{E}[|Z_2 - T + 1|^3] \\
 &\quad + 2c_{6,3}(T-1)^3 \mathbb{E} \left[ \left| \log \left( Z_1 + \frac{Z_2}{1+T\alpha} \right) \right|^3 \right] \\
 &\quad + 2c_{6,3}(T-1)^3 \mathbb{E} \left[ \left| \log \left( 1 + \frac{\beta(T, \rho)}{(1+T\alpha)Z_1+Z_2} \right) \right|^3 \right] \tag{141}
 \end{aligned}$$

where we have used (135) and that  $\mathbb{E}[|X|^3] \geq |\mathbb{E}[X]|^3$  for every random variable  $X$ .

We next show that each term on the RHS of (141) is bounded in  $\rho$  and  $\alpha$ . Indeed, the first two terms on the RHS of (141) are bounded because the third central moments of the Gamma-distributed random variables  $Z_1$  and  $Z_2$  are bounded, and because  $0 \leq (T\rho - T\alpha)/(1+T\rho) \leq 1$  and  $0 \leq T\rho/(1+T\rho) \leq 1$ . The third term on the RHS of (141) can be upper-bounded by using that

$$\left| \log \left( Z_1 + \frac{Z_2}{1+T\alpha} \right) \right| \leq |\log Z_1| + |\log(Z_1 + Z_2)| \tag{142}$$

which follows from similar steps as the first inequality in (125). Hence, by (135)

$$\begin{aligned}
 &\mathbb{E} \left[ \left| \log \left( Z_1 + \frac{Z_2}{1+T\alpha} \right) \right|^3 \right] \\
 &\leq c_{2,3} (\mathbb{E}[|\log Z_1|^3] + \mathbb{E}[|\log(Z_1 + Z_2)|^3]). \tag{143}
 \end{aligned}$$

Finally, the fourth term on the RHS of (141) can be upper-bounded as

$$\begin{aligned}
 &(T-1)^3 \mathbb{E} \left[ \left| \log \left( 1 + \frac{\beta(T, \rho)}{(1+T\alpha)Z_1+Z_2} \right) \right|^3 \right] \\
 &\leq (T-1)^3 \mathbb{E} \left[ \left| \log \left( 1 + \frac{\beta(T, \rho)}{Z_1+Z_2} \right) \right|^3 \right] \\
 &\leq (T-1)^3 \mathbb{E} \left[ \left| \log \left( 1 + \frac{\beta(T, \rho_0)}{Z_1+Z_2} \right) \right|^3 \right] \tag{144}
 \end{aligned}$$



for every  $\rho_0 > 0$  and  $\rho \geq \rho_0$ . Here, the second inequality follows from the monotonicity of  $\rho \mapsto \beta(T, \rho)$ . Since the RHSs of (143) and (144) are finite, this proves (79).

We establish (80) along similar lines. Using the definitions of  $i_\ell(T, \rho)$  and  $I(T, \rho)$  in (19) and (21), respectively, we can upper-bound the third moment of  $i_\ell(T, \rho)$  as

$$\begin{aligned} & \mathbb{E} \left[ \left| i_\ell(T, \rho) - I(T, \rho) \right|^3 \right] \\ &= \mathbb{E} \left[ \left| \frac{T\rho}{1+T\rho}(T-1-Z_2) + (T-1) \log \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \right. \right. \\ & \quad \left. \left. - (T-1) \mathbb{E} \left[ \log \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \right] \right. \right. \\ & \quad \left. \left. - \log \tilde{\gamma} \left( T-1, \frac{T\rho((1+T\rho)Z_1+Z_2)}{1+T\rho} \right) \right. \right. \\ & \quad \left. \left. + \mathbb{E} \left[ \log \tilde{\gamma} \left( T-1, \frac{T\rho((1+T\rho)Z_1+Z_2)}{1+T\rho} \right) \right] \right|^3 \right] \\ &\leq c_{5,3} \left| \frac{T\rho}{1+T\rho} \right|^3 \mathbb{E} [|Z_2 - T + 1|^3] \\ & \quad + 2c_{5,3} (T-1)^3 \mathbb{E} \left[ \left| \log \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \right|^3 \right] \\ & \quad + 2c_{5,3} \mathbb{E} \left[ \left| \log \tilde{\gamma} \left( T-1, \frac{T\rho((1+T\rho)Z_1+Z_2)}{1+T\rho} \right) \right|^3 \right] \end{aligned} \quad (145)$$

where we have used (135) and that  $\mathbb{E}[|X|^3] \geq |\mathbb{E}[X]|^3$  for every random variable  $X$ .

As shown before, the first two terms on the RHS of (145) are bounded in  $\rho$ . As for the third term, we first use Lemma 1 to obtain

$$\begin{aligned} & \mathbb{E} \left[ \left| \log \tilde{\gamma} \left( T-1, \frac{T\rho((1+T\rho)Z_1+Z_2)}{1+T\rho} \right) \right|^3 \right] \\ & \leq (T-1)^3 \mathbb{E} \left[ \log^3 \left( 1 + \frac{\beta(T, \rho)}{(1+T\rho)Z_1+Z_2} \right) \right]. \end{aligned} \quad (146)$$

By the monotonicity of  $\rho \mapsto \beta(T, \rho)$ , it follows that for every  $\rho_0 > 0$  and  $\rho \geq \rho_0$ , the third term on the RHS of (145) is upper-bounded by

$$\begin{aligned} & \mathbb{E} \left[ \left| \log \tilde{\gamma} \left( T-1, \frac{T\rho((1+T\rho)Z_1+Z_2)}{1+T\rho} \right) \right|^3 \right] \\ & \leq (T-1)^3 \mathbb{E} \left[ \log^3 \left( 1 + \frac{\beta(T, \rho_0)}{(1+T\rho)Z_1+Z_2} \right) \right]. \end{aligned} \quad (147)$$

Since the RHS of (147) is finite, this proves (80).

## APPENDIX E

### PROOF OF LEMMA 6

Consider the upper bound (86), namely,

$$R^*(L, T, \epsilon, \rho) \leq \sup_{\alpha \in [0, \rho]^L} \log \left( \frac{1}{\beta(\alpha, \mathbf{q}_{\mathbf{Y}^L}^{(U)})} \right). \quad (148)$$

In the following, we show that, for sufficiently large  $L$  and  $\rho$ , we can assume without loss of optimality that  $\alpha \in \mathcal{A}_{\rho, \delta}$ . To

this end, we demonstrate that for all  $\alpha \notin \mathcal{A}_{\rho, \delta}$  and sufficiently large  $L$  and  $\rho$ , we can find a lower bound on  $R^*(L, T, \epsilon, \rho)$  that exceeds an upper bound on (148). Hence, such  $\alpha$  cannot be optimal.

A lower bound on  $R^*(L, T, \epsilon, \rho)$  follows from (69), and by bounding  $I(T, \rho)$  and  $U(T, \rho)$  using (23) and (83), respectively:

$$\begin{aligned} R^*(L, T, \epsilon, \rho) &\geq \frac{I(T, \rho)}{T} - \sqrt{\frac{U_{\text{UB}}(T, \rho_0)}{LT^2}} Q^{-1}(\tau) \\ &\triangleq \frac{R_{\text{LB}}(T, \rho)}{T}, \quad \rho \geq \rho_0 \end{aligned} \quad (149)$$

with  $\tau$  defined in (73). Recall that, by the assumption  $0 < \epsilon < \frac{1}{2}$ , we have  $Q^{-1}(\tau) > 0$  for  $L$  sufficiently large.

It follows from [3, Eq. (106)] and (29) that the RHS of (148) can be upper-bounded as

$$\begin{aligned} \sup_{\alpha \in [0, \rho]^L} \log \left( \frac{1}{\beta(\alpha, \mathbf{q}_{\mathbf{Y}^L}^{(U)})} \right) &\leq \sup_{\alpha \in [0, \rho]^L} \left\{ \frac{\log \xi(\alpha)}{LT} \right. \\ & \quad \left. - \frac{\log \left( 1 - \epsilon - \mathbb{P} \left[ \sum_{\ell=1}^L \bar{j}_\ell(T, \alpha_\ell) \geq \log \xi(\alpha) \right] \right)}{LT} \right\} \end{aligned} \quad (150)$$

for every  $\xi: [0, \rho]^L \rightarrow (0, \infty)$ . By Lemma 4, for every  $\rho_0 > 0$  there exists a  $\bar{V}_{\text{UB}}(T, \rho_0)$  that is independent of  $\alpha$  and  $\rho$  and that satisfies

$$\bar{V}_\rho(T, \alpha) \leq \bar{V}_{\text{UB}}(T, \rho_0), \quad \alpha \geq 0, \rho \geq \rho_0. \quad (151)$$

Let

$$\log \xi(\alpha) = \sum_{\ell=1}^L \bar{J}(T, \alpha_\ell) + \sqrt{\frac{L \bar{V}_{\text{UB}}(T, \rho_0)}{(1-\epsilon) - \frac{1}{\sqrt{L}}}}. \quad (152)$$

By Chebyshev's inequality [34, Ch. V.7] and (151), we obtain

$$\begin{aligned} & \mathbb{P} \left[ \sum_{\ell=1}^L \bar{j}_\ell(T, \alpha_\ell) \geq \log \xi(\alpha) \right] \\ & \leq \frac{\sum_{\ell=1}^L \bar{V}_\rho(T, \alpha_\ell)}{L \bar{V}_{\text{UB}}(T, \rho_0)} \left( 1 - \epsilon - \frac{1}{\sqrt{L}} \right) \\ & \leq 1 - \epsilon - \frac{1}{\sqrt{L}}, \quad \rho \geq \rho_0. \end{aligned} \quad (153)$$

Combining (153) with (150) then yields

$$\begin{aligned} R^*(L, \epsilon, \rho) &\leq \sup_{\alpha \in [0, \rho]^L} \frac{\sum_{\ell=1}^L \bar{J}(T, \alpha_\ell)}{LT} \\ & \quad + \sqrt{\frac{\bar{V}_{\text{UB}}(T, \rho_0)}{LT^2(1-\epsilon) - T^2\sqrt{L}}} + \frac{\log L}{2LT} \\ &\triangleq \sup_{\alpha \in [0, \rho]^L} \frac{1}{L} \sum_{\ell=1}^L \frac{R_{\text{UB}}(T, \alpha_\ell)}{T}, \quad \rho \geq \rho_0. \end{aligned} \quad (154)$$

The  $\alpha$ 's for which  $\frac{1}{L} \sum_{\ell=1}^L R_{\text{UB}}(T, \alpha_\ell)/T$  is smaller than (149) can be discarded without loss of optimality, since the upper bound can never be smaller than the lower bound. We next use this argument to show that the fraction of  $\alpha_\ell$ 's in

$\alpha$  that satisfy  $\alpha_\ell \geq \rho(1 - \delta)$  tends to 1 as  $L$  and  $\rho$  tend to infinity. Specifically, we consider the difference

$$\begin{aligned} & \frac{1}{L} \sum_{\ell=1}^L [R_{\text{LB}}(T, \rho) - R_{\text{UB}}(T, \alpha_\ell)] \\ &= \frac{1}{L} \sum_{\ell=1}^L \left[ \frac{T\rho - T\alpha_\ell}{1 + T\rho} + \log\left(\frac{1 + T\alpha_\ell}{1 + T\rho}\right) \right. \\ & \quad + (T - 1)\mathbb{E} \left[ \log \frac{(1 + T\rho)Z_{1,\ell} + Z_{2,\ell} + \beta(T, \rho)}{(1 + T\alpha_\ell)Z_{1,\ell} + Z_{2,\ell} + \beta(T, \rho)} \right] \\ & \quad - (T - 1)\mathbb{E} \left[ \log \left( 1 + \frac{\beta(T, \rho)}{(1 + T\rho)Z_{1,\ell} + Z_{2,\ell}} \right) \right] \\ & \quad \left. - \sqrt{\frac{U_{\text{UB}}(T, \rho_0)}{L}} Q^{-1}(\tau) - \sqrt{\frac{\bar{V}_{\text{UB}}(T, \rho_0)}{L(1 - \epsilon) - \sqrt{L}}} - \frac{\log L}{2L} \right] \end{aligned} \quad (155)$$

where we have evaluated  $R_{\text{LB}}(T, \rho)$  and  $R_{\text{UB}}(T, \alpha_\ell)$  using (22) and (30). We next fix a sufficiently large  $\rho_0$  and assume that  $\rho \geq \rho_0$ . Since  $\rho \mapsto \beta(T, \rho)$  is decreasing in  $\rho$ , we can lower-bound the third-term on the RHS of (155) by replacing  $\beta(T, \rho)$  by  $\beta(T, \rho_0)$ . We can further lower-bound (155) by omitting the first term on the RHS of (155), which is nonnegative since  $\alpha_\ell \leq \rho$ . This yields

$$\begin{aligned} & \frac{1}{L} \sum_{\ell=1}^L [R_{\text{LB}}(T, \rho) - R_{\text{UB}}(T, \alpha_\ell)] \\ & \geq \frac{1}{L} \sum_{\ell=1}^L \left[ \log\left(\frac{1 + T\alpha_\ell}{1 + T\rho}\right) \right. \\ & \quad + (T - 1)\mathbb{E} \left[ \log \frac{(1 + T\rho)Z_{1,\ell} + T - 1 + \beta(T, \rho_0)}{(1 + T\alpha_\ell)Z_{1,\ell} + T - 1 + \beta(T, \rho_0)} \right] \\ & \quad - (T - 1)\mathbb{E} \left[ \log \left( 1 + \frac{\beta(T, \rho)}{(1 + T\rho)Z_{1,\ell} + Z_{2,\ell}} \right) \right] \\ & \quad \left. - \sqrt{\frac{U_{\text{UB}}(T, \rho_0)}{L}} Q^{-1}(\tau) - \sqrt{\frac{\bar{V}_{\text{UB}}(T, \rho_0)}{L(1 - \epsilon) - \sqrt{L}}} - \frac{\log L}{2L} \right] \\ & \triangleq \frac{1}{L} \sum_{\ell=1}^L \Delta_{L,T,\rho}(\alpha_\ell), \quad \rho \geq \rho_0. \end{aligned} \quad (156)$$

In the following, we analyze the behaviour of the function  $\alpha_\ell \mapsto \Delta_{L,T,\rho}(\alpha_\ell)$ . Let

$$\begin{aligned} g_{T,\rho}(\alpha_\ell) & \triangleq \log\left(\frac{1 + T\alpha_\ell}{1 + T\rho}\right) \\ & \quad + (T - 1)\mathbb{E} \left[ \log \frac{(1 + T\rho)Z_{1,\ell} + T - 1 + \beta(T, \rho_0)}{(1 + T\alpha_\ell)Z_{1,\ell} + T - 1 + \beta(T, \rho_0)} \right] \end{aligned} \quad (157)$$

and

$$\begin{aligned} \omega_{L,T,\rho} & \triangleq (T - 1)\mathbb{E} \left[ \log \left( 1 + \frac{\beta(T, \rho)}{(1 + T\rho)Z_1 + Z_2} \right) \right] \\ & \quad + \sqrt{\frac{U_{\text{UB}}(T, \rho_0)}{L}} Q^{-1}(\tau) + \sqrt{\frac{\bar{V}_{\text{UB}}(T, \rho_0)}{L(1 - \epsilon) - \sqrt{L}}} + \frac{\log L}{2L}. \end{aligned} \quad (158)$$

Thus,  $\Delta_{L,T,\rho}(\alpha_\ell) = g_{T,\rho}(\alpha_\ell) - \omega_{L,T,\rho}$ . Note that  $\frac{\partial}{\partial \alpha_\ell} g_{T,\rho}(\alpha_\ell) = \frac{\partial}{\partial \alpha_\ell} \Delta_{L,T,\rho}(\alpha_\ell)$ , since  $\omega_{L,T,\rho}$  does not depend on  $\alpha_\ell$ . Further note that

$$\begin{aligned} & \lim_{\substack{L \rightarrow \infty, \\ \rho \rightarrow \infty}} \omega_{L,T,\rho} \\ &= \lim_{\rho \rightarrow \infty} (T - 1)\mathbb{E} \left[ \log \left( 1 + \frac{\beta(T, \rho)}{(1 + T\rho)Z_1 + Z_2} \right) \right] \\ & \quad + \lim_{L \rightarrow \infty} \sqrt{\frac{U_{\text{UB}}(T, \rho_0)}{L}} Q^{-1}(\tau) \\ & \quad + \lim_{L \rightarrow \infty} \sqrt{\frac{\bar{V}_{\text{UB}}(T, \rho_0)}{L(1 - \epsilon) - \sqrt{L}}} + \lim_{L \rightarrow \infty} \frac{\log L}{2L} \\ &= 0 \end{aligned} \quad (159)$$

where the second line in (159) is zero by the dominated convergence theorem. The following lemma discusses the behavior of  $\alpha_\ell \mapsto g_{T,\rho}(\alpha_\ell)$ .

**Lemma 8:** The function  $\alpha \mapsto g_{T,\rho}(\alpha)$  has the following properties:

- 1) The derivative of  $\alpha \mapsto g_{T,\rho}(\alpha)$  is either strictly positive, strictly negative, or changes its sign once from positive to negative. This implies that  $g_{T,\rho}(\alpha)$ ,  $0 \leq \alpha \leq \rho$  is minimized at the boundary of  $[0, \rho]$ , and it has a unique maximizer.
- 2) The derivative of  $\alpha \mapsto g_{T,\rho}(\alpha)$  with respect to  $\alpha$ , denoted by  $\alpha \mapsto g'_{T,\rho}(\alpha)$ , does not depend on  $\rho$ .
- 3) It holds that  $g_{T,\rho}(\rho) = 0$ . Furthermore,  $g_{T,\rho}(0) \rightarrow \infty$  as  $\rho \rightarrow \infty$  for  $T > 2$ .
- 4) Let  $\alpha^*$  denote the unique maximizer of  $\alpha \mapsto g_{T,\rho}(\alpha)$ , which by Part 2) does not depend on  $\rho$ . For  $T > 2$  and every  $\alpha' > \alpha^*$  independent of  $\rho$ , we have

$$\sup_{\rho \geq \alpha'} \sup_{\alpha' \leq \alpha \leq \rho} \rho g'_{T,\rho}(\alpha) < 0. \quad (160)$$

*Proof:* See Appendix F. ■

We next study those  $\alpha$ 's for which  $\sum_{\ell=1}^L \Delta_{L,T,\rho}(\alpha_\ell) \geq 0$ , since they can be discarded without loss of optimality. Let

$$\mathcal{L}_\delta(\alpha) \triangleq \{\ell = 1, \dots, L : \alpha_\ell \geq \rho(1 - \delta)\} \quad (161)$$

and let  $L_\delta(\alpha)$  denote the number of  $\alpha_\ell$ 's in  $\alpha$  that satisfy  $\rho(1 - \delta) \leq \alpha_\ell \leq \rho$ . (Thus,  $L_\delta(\alpha)$  is the cardinality of  $\mathcal{L}_\delta(\alpha)$ .) Further let

$$\Delta_{L,T,\rho}^*(\delta) \triangleq \inf_{0 \leq \alpha \leq \rho(1 - \delta)} \Delta_{L,T,\rho}(\alpha). \quad (162)$$

The sum of  $\Delta_{L,T,\rho}(\alpha_\ell)$  in (156) can be expressed as

$$\begin{aligned} & \sum_{\ell=1}^L \Delta_{L,T,\rho}(\alpha_\ell) \\ &= \sum_{\mathcal{L}_\delta(\alpha)} \Delta_{L,T,\rho}(\alpha_\ell) + \sum_{\mathcal{L}_\delta^c(\alpha)} \Delta_{L,T,\rho}(\alpha_\ell). \end{aligned} \quad (163)$$

By Parts 1) and 3) of Lemma 8,

$$\Delta_{L,T,\rho}(\alpha_\ell) \geq -\omega_{L,T,\rho}, \quad 0 \leq \alpha_\ell \leq \rho \quad (164)$$

for  $T > 2$  and  $\rho$  sufficiently large. Thus, we can lower-bound the first sum on the RHS of (163) by  $-L_\delta(\alpha)\omega_{L,T,\rho}$  and the

second sum on the RHS of (163) by  $(L - L_\delta(\alpha))\Delta_{L,T,\rho}^*(\delta)$ . This yields

$$\begin{aligned} & \sum_{\ell=1}^L \Delta_{L,T,\rho}(\alpha_\ell) \\ & \geq (L - L_\delta(\alpha))\Delta_{L,T,\rho}^*(\delta) - L_\delta(\alpha)\omega_{L,T,\rho}. \end{aligned} \quad (165)$$

It follows that we can discard without loss of optimality every  $\alpha$  for which

$$L\Delta_{L,T,\rho}^*(\delta) \geq L_\delta(\alpha)[\omega_{L,T,\rho} + \Delta_{L,T,\rho}^*(\delta)] \quad (166)$$

since for such  $\alpha$ 's the RHS of (165), and hence also (155), is nonnegative. We conclude that an  $\alpha$  maximizing (148) must satisfy

$$\frac{L_\delta(\alpha)}{L} > 1 - \frac{\omega_{L,T,\rho}}{\omega_{L,T,\rho} + \Delta_{L,T,\rho}^*(\delta)}. \quad (167)$$

As we shall show below, for every  $0 < \delta < 1$  we have

$$\omega_{L,T,\rho} + \Delta_{L,T,\rho}^*(\delta) \geq -\delta \sup_{\rho \geq \alpha'} \sup_{\alpha' \leq \alpha \leq \rho} \rho g'_{T,\rho}(\alpha) \quad (168)$$

for some  $0 < \alpha' < \rho(1 - \delta)$  that is independent of  $\rho$ . We further show that the RHS of (168) is independent of  $L$  and  $\rho$  and strictly positive. It follows that

$$\frac{L_\delta(\alpha)}{L} > 1 - \frac{\omega_{L,T,\rho}}{-\delta \sup_{\rho \geq \alpha'} \sup_{\alpha' \leq \alpha \leq \rho} \rho g'_{T,\rho}(\alpha)} \quad (169)$$

which, by (159), tends to one as  $\rho$  and  $L$  tend to infinity. Thus, for every  $0 < \delta < 1$ , there exist sufficiently large  $L_0$  and  $\rho_0$  such that

$$L_\delta(\alpha) \geq L/2, \quad L \geq L_0, \rho \geq \rho_0. \quad (170)$$

This proves Lemma 6.

It remains to show (168). Let  $\alpha_{\min} = \rho(1 - \delta)$ . By Part 1) of Lemma 8,  $\alpha \mapsto g_{T,\rho}(\alpha)$  has exactly one maximizer, which we shall denote by  $\alpha^*$ . Since  $\omega_{L,T,\rho}$  does not depend on  $\alpha$ , it follows that  $\alpha^*$  is also the maximizer of  $\alpha \mapsto \Delta_{L,T,\rho}(\alpha)$ . Furthermore, the infimum of  $\Delta_{L,T,\rho}(\alpha)$  over  $0 \leq \alpha \leq \alpha_{\min}$ , denoted by  $\Delta_{L,T,\rho}^*(\delta)$ , is either achieved at  $\alpha = 0$  or at  $\alpha_{\min}$ .

By Part 3) of Lemma 8 and (159), we have

$$\lim_{\substack{L \rightarrow \infty, \\ \rho \rightarrow \infty}} \Delta_{L,T,\rho}(0) = \infty. \quad (171)$$

We next show that

$$\Delta_{L,T,\rho}(\alpha_{\min}) + \omega_{L,T,\rho} \geq -\delta \sup_{\rho \geq \alpha'} \sup_{\alpha' \leq \alpha \leq \rho} \rho g'_{T,\rho}(\alpha). \quad (172)$$

If  $\alpha_{\min} \leq \alpha^*$ , then this is clearly satisfied, since in this case  $\Delta_{L,T,\rho}(\alpha_{\min}) \geq \Delta_{L,T,\rho}(0)$  and  $\Delta_{L,T,\rho}(0)$  tends to infinity as  $L \rightarrow \infty$  and  $\rho \rightarrow \infty$ . However, in general this case does not occur for large  $\rho$  and  $L$ , since  $\alpha_{\min}$  tends to infinity as  $\rho \rightarrow \infty$  and, by Part 2) of Lemma 8,  $\alpha^*$  does not depend on  $\rho$ , which implies that  $\alpha_{\min} > \alpha^*$  for  $\rho$  sufficiently large. We thus focus on the case where  $\alpha_{\min} > \alpha^*$ . Note that

$$\Delta_{L,T,\rho}(\rho) - \Delta_{L,T,\rho}(\alpha_{\min}) = -\omega_{L,T,\rho} - \Delta_{L,T,\rho}(\alpha_{\min}) \quad (173)$$

since  $g_{T,\rho}(\rho) = 0$ . Thus, by the mean value theorem [38, Th. 5.10], there exists an  $x_0 \in [\alpha_{\min}, \rho]$  such that

$$\begin{aligned} & -\omega_{L,T,\rho} - \Delta_{L,T,\rho}(\alpha_{\min}) \\ & = \int_{\alpha_{\min}}^{\rho} \Delta'_{L,T,\rho}(\alpha) d\alpha = \rho \delta \Delta'_{L,T,\rho}(x_0) \end{aligned} \quad (174)$$

where  $\Delta'_{L,T,\rho}(\cdot)$  denotes the derivative of  $\alpha \mapsto \Delta_{L,T,\rho}(\alpha)$ . We can therefore lower-bound

$$\begin{aligned} \Delta_{L,T,\rho}(\alpha_{\min}) + \omega_{L,T,\rho} & \geq -\delta \sup_{\alpha_{\min} \leq \alpha \leq \rho} \rho \Delta'_{L,T,\rho}(\alpha) \\ & \geq -\delta \sup_{\rho \geq \alpha'} \sup_{\alpha' \leq \alpha \leq \rho} \rho g'_{T,\rho}(\alpha) \end{aligned} \quad (175)$$

for every  $\alpha' \in (\alpha^*, \alpha_{\min})$  independent of  $L$  and  $\rho$ .<sup>8</sup> In (175), the second inequality follows by noting that  $\Delta'_{L,T,\rho} = g'_{T,\rho}$  and by further optimizing over  $\rho$ . It remains to show that the RHS of (175) is independent of  $L$  and  $\rho$  and strictly positive. To this end, we first note that  $\alpha \mapsto g_{T,\rho}(\alpha)$  is independent of  $L$ . Furthermore, by optimizing over  $\rho \geq \alpha'$ , the RHS of (175) becomes also independent of  $\rho$ . Finally, by Part 4) of Lemma 8,

$$\sup_{\rho \geq \alpha'} \sup_{\alpha' \leq \alpha \leq \rho} \rho g'_{T,\rho}(\alpha) < 0, \quad T > 2, \rho \geq \alpha' \quad (176)$$

for every  $\alpha' \in (\alpha^*, \alpha_{\min})$  independent of  $L$  and  $\rho$ . Thus, the claim (168) follows, which concludes the proof of Lemma 6.

## APPENDIX F PROOF OF LEMMA 8

The derivative of  $\alpha \mapsto g_{T,\rho}(\alpha)$  can be expressed as

$$\begin{aligned} g'_{T,\rho}(\alpha) & = \frac{T}{1 + T\alpha} \\ & - (T - 1) \mathbb{E} \left[ \frac{TZ_1}{(1 + T\alpha)Z_1 + (T - 1) + \beta(T, \rho_0)} \right] \\ & = T \left[ \frac{1}{1 + T\alpha} - \frac{T - 1}{1 + T\alpha} \right. \\ & \quad \left. + \frac{T - 1}{1 + T\alpha} \varphi \left( \frac{1 + T\alpha}{T - 1 + \beta(T, \rho_0)} \right) \right] \\ & = \frac{T}{1 + T\alpha} \left[ -(T - 2) \right. \\ & \quad \left. + (T - 1) \varphi \left( \frac{1 + T\alpha}{T - 1 + \beta(T, \rho_0)} \right) \right] \end{aligned} \quad (177)$$

where

$$\varphi(x) \triangleq \frac{1}{x} e^{\frac{1}{x}} \mathbb{E}_1 \left( \frac{1}{x} \right), \quad x > 0. \quad (178)$$

The first equality follows because, by [39, App. A.9], we can swap derivative and expected value; the second equality follows by solving the expected value using [19, Sec. 3.353-5.7]. Note that the RHS of (177) does not depend on  $\rho$ . Hence Part 2) of Lemma 8 follows immediately.

<sup>8</sup>Since  $\alpha^*$  is independent of  $\rho$  and  $\alpha_{\min} \rightarrow \infty$  as  $\rho \rightarrow \infty$ , it follows that such an  $\alpha'$  exists.

We next prove Part 1) of Lemma 8. Because  $T/(1+T\alpha)$  in (177) is nonnegative, the sign of  $\alpha \mapsto g'_{T,\rho}(\alpha)$  is determined by the terms inside the square brackets. Note that  $x \mapsto \varphi(x)$  is strictly decreasing since, by [19, Sec. 3.353-3],

$$\frac{1}{x} e^{\frac{1}{x}} \mathbf{E}_1\left(\frac{1}{x}\right) = 1 - \int_0^1 e^{-\frac{t}{(1-t)x}} dt \quad (179)$$

and  $x \mapsto e^{-\frac{t}{(1-t)x}}$  is strictly positive and strictly increasing in  $x$ . Hence, the function inside the squared brackets is strictly decreasing in  $\alpha$ . This implies that  $\alpha \mapsto g'_{T,\rho}(\alpha)$  is either strictly positive, strictly negative, or changes its sign once from positive to negative.

We next prove Part 3) of Lemma 8 by showing that  $\lim_{\rho \rightarrow \infty} g_{T,\rho}(0) = \infty$  for  $T > 2$ . To this end, we express  $g_{T,\rho}(0)$  as

$$\begin{aligned} g_{T,\rho}(0) &= (T-2) \mathbf{E} \left[ \log \left( 1 + \frac{T\rho Z_1}{Z_1 + (T-1) + \beta(T, \rho_0)} \right) \right] \\ &\quad + \mathbf{E} \left[ \log \left( Z_1 + \frac{T-1 + \beta(T, \rho_0)}{1 + T\rho} \right) \right] \\ &\quad - \mathbf{E} \left[ \log(Z_1 + T - 1 + \beta(T, \rho_0)) \right]. \end{aligned} \quad (180)$$

The first expected value on the RHS of (180) tends to infinity as  $\rho \rightarrow \infty$ , whereas the other expected values are bounded in  $\rho$ . For  $T > 2$ , it follows that the RHS of (180) tends to infinity as  $\rho \rightarrow \infty$ . Hence the claim follows.

We finally prove Part 4) of Lemma 8 by analyzing  $\rho g'_{T,\rho}(\alpha)$ . It follows from (177) that

$$\begin{aligned} \rho g'_{T,\rho}(\alpha) &= \frac{T\rho}{1+T\alpha} \left[ -(T-2) \right. \\ &\quad \left. + (T-1)\varphi\left(\frac{1+T\alpha}{T-1+\beta(T, \rho_0)}\right) \right]. \end{aligned} \quad (181)$$

Observe that the function inside the square brackets is independent of  $L$  and  $\rho$ . Further note that, as argued above, it is strictly decreasing in  $\alpha$ . Hence, its supremum over  $\alpha' \leq \alpha \leq \rho$  is achieved for  $\alpha = \alpha'$ . Furthermore, this function is strictly negative for  $T > 2$  and  $\alpha' > \alpha^*$ . As for the term outside the curly brackets, we have for every  $\alpha' > \alpha^*$

$$\inf_{\rho \geq \alpha'} \inf_{\alpha' \leq \alpha \leq \rho} \frac{T\rho}{1+T\alpha} = \frac{T\alpha'}{1+T\alpha'} > 0. \quad (182)$$

Combining these two results, we conclude that

$$\sup_{\rho \geq \alpha'} \sup_{\alpha' \leq \alpha \leq \rho} \rho g'_{T,\rho}(\alpha) < 0, \quad T > 2, \alpha' > \alpha^*. \quad (183)$$

This proves Part 4) of Lemma 8 and concludes the proof of Lemma 8.

#### APPENDIX G PROOF OF LEMMA 7

1) *Part 1)*: The difference between  $\bar{J}(T, \alpha)$  and  $\bar{J}(T, \rho)$  can be lower-bounded by

$$\bar{J}(T, \rho) - \bar{J}(T, \alpha) \geq g_{T,\rho}(\alpha) \quad (184)$$

where the function  $\alpha \mapsto g_{T,\rho}(\alpha)$  was defined in (157). By Parts 1) and 3) of Lemma 8 (Appendix E),  $\alpha \mapsto g_{T,\rho}(\alpha)$  is nonnegative for sufficiently large  $\rho$ . It follows that, for such  $\rho$ ,

$$\sup_{0 \leq \alpha \leq \rho} \bar{J}(T, \alpha) = \bar{J}(T, \rho). \quad (185)$$

This proves Part 1) of Lemma 7.

2) *Part 2)*: To study

$$\sup_{0 \leq \alpha \leq \rho} \left\{ \frac{\bar{J}(T, \alpha)}{T} - \sqrt{\frac{\bar{V}_\rho(T, \alpha)}{LT^2}} Q^{-1}(\epsilon) \right\} \quad (186)$$

we consider the difference

$$\begin{aligned} \bar{J}(T, \rho) - \bar{J}(T, \alpha) &- \left( \sqrt{\frac{\bar{V}(T, \rho)}{L}} - \sqrt{\frac{\bar{V}_\rho(T, \alpha)}{L}} \right) Q^{-1}(\epsilon) \\ &\geq g_{T,\rho}(\alpha) - \left( \sqrt{\frac{\bar{V}(T, \rho)}{L}} - \sqrt{\frac{\bar{V}_\rho(T, \alpha)}{L}} \right) Q^{-1}(\epsilon). \end{aligned} \quad (187)$$

Clearly, every  $\alpha$  for which the RHS of (187) is nonnegative is suboptimal and can be discarded without loss of optimality. We continue by lower-bounding  $\bar{V}_\rho(T, \alpha) \geq 0$  and by using that  $\bar{V}(T, \rho) \leq \bar{V}_{\text{UB}}(T, \rho_0)$ ,  $\rho \geq \rho_0$  for sufficiently large  $\rho_0$  and for some constant  $\bar{V}_{\text{UB}}(T, \rho_0)$  that is independent of  $\rho$  (Lemma 4). Since by the assumption  $0 < \epsilon < \frac{1}{2}$  we have  $Q^{-1}(\epsilon) > 0$ , this yields

$$\begin{aligned} g_{T,\rho}(\alpha) &- \sqrt{\frac{\bar{V}(T, \rho)}{L}} Q^{-1}(\epsilon) + \sqrt{\frac{\bar{V}_\rho(T, \alpha)}{L}} Q^{-1}(\epsilon) \\ &\geq g_{T,\rho}(\alpha) - \sqrt{\frac{\bar{V}_{\text{UB}}(T, \rho_0)}{L}} Q^{-1}(\epsilon) \\ &\triangleq f_{L,T,\rho}(\alpha). \end{aligned} \quad (188)$$

Again, the values of  $\alpha$  for which  $f_{L,T,\rho}(\alpha) \geq 0$  are suboptimal and can be discarded without loss of optimality.

Let us write  $f_{L,T,\rho}$  as  $f_{L,T,\rho}(\alpha) \triangleq g_{T,\rho}(\alpha) - \omega_{L,T}$ , where

$$\omega_{L,T} \triangleq \sqrt{\frac{\bar{V}_{\text{UB}}(T, \rho_0)}{L}} Q^{-1}(\epsilon). \quad (189)$$

Note that  $\Delta_{L,T,\rho}$  defined in (156) and  $f_{L,T,\rho}$  only differ in terms that do not depend on  $\alpha$  (namely,  $\omega_{L,T,\rho}$  and  $\omega_{L,T}$ ), so they have the same behavior with respect to  $\alpha$  as summarized in Lemma 8. Let  $\delta_{L,T} \triangleq 1 - \alpha_0/\rho$ , where  $\alpha_0$  is the unique real root of  $\alpha \mapsto f_{L,T,\rho}(\alpha)$ . Indeed, we know that  $\alpha \mapsto f_{L,T,\rho}(\alpha)$  has at least one root because  $\omega_{L,T} \geq 0$  and  $\omega_{L,T} \rightarrow 0$  as  $L \rightarrow \infty$ , so  $f_{L,T,\rho}(\rho) = -\omega_{L,T} \leq 0$  and  $f_{L,T,\rho}(0) > 0$  for  $L$  and  $\rho$  sufficiently large. Furthermore, we have  $f'_{L,T,\rho} = g'_{T,\rho}$  and  $\alpha \mapsto g'_{T,\rho}(\alpha)$  is either strictly positive, strictly negative, or changes its sign once from positive to negative (Part 1) of Lemma 8). Consequently,  $f_{L,T,\rho}(\alpha)$ ,  $0 \leq \alpha \leq \rho$  is minimized at an endpoint of  $[0, \rho]$  and it has a unique maximizer, so its root is unique. By the same line of arguments, we also conclude that all  $\alpha$ 's between 0 and  $\rho(1 - \delta_{L,T})$  can be discarded without loss of optimality, since for such  $\alpha$ 's the function  $f_{L,T,\rho}(\alpha)$  is nonnegative.

To study the behavior of  $\delta_{L,T}$ , we next note that

$$\omega_{L,T} = -(f_{L,T,\rho}(\rho) - f_{L,\rho}(\alpha_0)). \quad (190)$$

It follows then by similar steps as in (174)–(175) that

$$\omega_{L,T} \geq -\delta_{L,T} \sup_{\alpha_0 \leq \alpha \leq \rho} \rho f'_{L,T,\rho}(\alpha). \quad (191)$$

Let  $\alpha^*$  denote the unique maximizer of  $\alpha \mapsto f_{L,T,\rho}(\alpha)$ . Recall that, by Part 2) of Lemma 8,  $\alpha^*$  does not depend on  $\rho$ . We next show that we can find an  $\tilde{\alpha}$  independent of  $L$  and  $\rho$  such that  $\alpha^* < \tilde{\alpha} < \alpha_0$ . Indeed, by Lemma 8, we have that  $g_{T,\rho}(\alpha^*) > 0$  for sufficiently large  $\rho$ . This in turn implies that

$$\lim_{\substack{L \rightarrow \infty, \\ \rho \rightarrow \infty}} f_{L,T,\rho}(\alpha^*) > 0 \quad (192)$$

since  $\omega_{L,T} \rightarrow 0$  as  $L \rightarrow \infty$ . We next note that, for every  $\tilde{\alpha}$ ,

$$\begin{aligned} & \lim_{\substack{L \rightarrow \infty, \\ \rho \rightarrow \infty}} f_{L,T,\rho}(\tilde{\alpha}) \\ & \geq \lim_{\substack{L \rightarrow \infty, \\ \rho \rightarrow \infty}} f_{L,T,\rho}(\alpha^*) - |f_{L,T,\rho}(\tilde{\alpha}) - f_{L,T,\rho}(\alpha^*)| \end{aligned} \quad (193)$$

where the difference

$$\begin{aligned} & f_{L,T,\rho}(\tilde{\alpha}) - f_{L,T,\rho}(\alpha^*) \\ & = g_{T,\rho}(\tilde{\alpha}) - g_{T,\rho}(\alpha^*) \\ & = \log\left(\frac{1 + T\tilde{\alpha}}{1 + T\alpha^*}\right) \\ & \quad + (T-1)\mathbb{E}\left[\log\frac{(1 + T\alpha^*)Z_1 + T - 1 + \beta(T, \rho_0)}{(1 + T\tilde{\alpha})Z_1 + T - 1 + \beta(T, \rho_0)}\right] \end{aligned} \quad (194)$$

is independent of  $L$  and  $\rho$ . By the continuity of  $\alpha \mapsto g_{T,\rho}(\alpha)$ , this difference can be made arbitrarily small by choosing  $\tilde{\alpha}$  sufficiently close to  $\alpha^*$ . It thus follows from (192)–(194) that there exists an  $\tilde{\alpha} \in (\alpha^*, \rho]$  that is independent of  $L$  and  $\rho$  and that satisfies

$$\lim_{\substack{L \rightarrow \infty, \\ \rho \rightarrow \infty}} f_{L,T,\rho}(\tilde{\alpha}) > 0. \quad (195)$$

In other words, if  $L$  and  $\rho$  are sufficiently large, then we can find an  $\tilde{\alpha} \in (\alpha^*, \alpha_0)$  that is independent of  $L$  and  $\rho$ . In this case the RHS of (191) can be further lower-bounded by

$$\begin{aligned} \omega_{L,T} & \geq -\delta_{L,T} \sup_{\tilde{\alpha} \leq \alpha \leq \rho} \rho f'_{L,T,\rho}(\alpha) \\ & \geq -\delta_{L,T} \sup_{\rho \geq \tilde{\alpha}} \sup_{\tilde{\alpha} \leq \alpha \leq \rho} \rho f'_{L,T,\rho}(\alpha). \end{aligned} \quad (196)$$

We next argue that the constant

$$F(T) \triangleq -\sup_{\rho \geq \tilde{\alpha}} \sup_{\tilde{\alpha} \leq \alpha \leq \rho} \rho f'_{L,T,\rho}(\alpha) \quad (197)$$

is independent of  $L$  and  $\rho$  and strictly positive. Indeed, we have that  $f'_{L,T,\rho} = g'_{T,\rho}$ , which is independent of  $L$ . Furthermore, by optimizing over  $\rho \geq \tilde{\alpha}$ , the RHS of (197) becomes independent of  $\rho$ . Finally, setting  $\alpha' = \tilde{\alpha}$  in (160) (Part 4) of Lemma 8) yields

$$\sup_{\rho \geq \tilde{\alpha}} \sup_{\tilde{\alpha} \leq \alpha \leq \rho} \rho g'_{T,\rho}(\alpha) < 0, \quad \rho \geq \tilde{\alpha} \quad (198)$$

hence the claim follows. Consequently, we obtain from (196) and the definition of  $\omega_L$  and  $F(T)$  that, for sufficiently large  $L_0$  and  $\rho_0$ ,

$$\delta_{L,T} \leq \frac{\sqrt{\bar{V}_{\text{UB}}(T, \rho_0)} Q^{-1}(\epsilon)}{F(T)} \frac{1}{\sqrt{L}}, \quad \rho \geq \rho_0, L \geq L_0. \quad (199)$$

We next tighten this bound on  $\delta_{L,T}$ . Indeed, using that without loss of optimality we can assume  $\rho(1 - \delta_{L,T}) \leq \alpha \leq \rho$ , we can derive a tighter lower bound on (187) by lower-bounding  $\bar{V}_\rho(T, \alpha)$  using the lower bound given in Appendix H instead of lower-bounding it by zero. Specifically, by (208) in Appendix H,

$$\begin{aligned} \sqrt{\frac{\bar{V}_\rho(T, \alpha)}{L}} & \geq \sqrt{\frac{\bar{V}(T, \rho) - \Upsilon(T)\delta_{L,T}}{L}} \\ & \geq \sqrt{\frac{\bar{V}(T, \rho)}{L}} - \sqrt{\frac{\Upsilon(T)\delta_{L,T}}{L}} \end{aligned} \quad (200)$$

for every  $\rho(1 - \delta_{L,T}) \leq \alpha \leq \rho$ . We can thus lower-bound (187) as

$$\begin{aligned} \bar{J}(T, \rho) - \bar{J}(T, \alpha) & - \left( \sqrt{\frac{\bar{V}(T, \rho)}{L}} - \sqrt{\frac{\bar{V}_\rho(T, \alpha)}{L}} \right) Q^{-1}(\epsilon) \\ & \geq g_{T,\rho}(\alpha) - \sqrt{\frac{\Upsilon(T)\delta_{L,T}}{L}} Q^{-1}(\epsilon) \\ & \triangleq \tilde{f}_{L,T,\rho}(\alpha), \quad \rho(1 - \delta_{L,T}) \leq \alpha \leq \rho. \end{aligned} \quad (201)$$

Again, the values of  $\alpha$  for which  $\tilde{f}_{L,T,\rho}(\alpha) \geq 0$  are suboptimal and can be discarded without loss of optimality.

Let us write  $\tilde{f}_{L,T,\rho}(\alpha) = g_{T,\rho}(\alpha) - \tilde{\omega}_{L,T}$ , where

$$\tilde{\omega}_{L,T} \triangleq \sqrt{\frac{\Upsilon(T)\delta_{L,T}}{L}} Q^{-1}(\epsilon). \quad (202)$$

Further let  $\tilde{\delta}_{L,T} \triangleq 1 - \tilde{\alpha}_0/\rho$ , where  $\tilde{\alpha}_0$  is the unique real root of  $\alpha \mapsto \tilde{f}_{L,T,\rho}(\alpha)$ . As above, it can be shown that all  $\alpha$ 's between 0 and  $\rho(1 - \tilde{\delta}_{L,T})$  can be discarded without loss of optimality, since for such  $\alpha$ 's the function  $\tilde{f}_{L,T,\rho}(\alpha)$  is nonnegative. By repeating the steps (191)–(199) with  $\omega_{L,T}$  replaced by  $\tilde{\omega}_{L,T}$ , we obtain that

$$\begin{aligned} \tilde{\delta}_{L,T} & \leq \frac{1}{F(T)} \sqrt{\frac{\Upsilon(T)\delta_{L,T}}{L}} Q^{-1}(\epsilon) \\ & \leq \left( \frac{Q^{-1}(\epsilon)}{F(T)} \right)^{3/2} \sqrt{\Upsilon(T)} \sqrt{\bar{V}_{\text{UB}}(T, \rho_0)} \frac{1}{L^{3/4}} \end{aligned} \quad (203)$$

for every  $\rho \geq \rho_0$ ,  $L \geq L_0$ , and sufficiently large  $L_0$  and  $\rho_0$ . The last inequality follows by upper-bounding  $\delta_{L,T}$  using (199).

If we perform the above steps  $N$  times, then we obtain that, without loss of optimality,

$$\alpha \geq \rho \left( 1 - \delta_{L,T}^{(N)} \right) \quad (204)$$

where  $\delta_{L,T}^{(N)}$  satisfies

$$\begin{aligned} 0 & \leq \delta_{L,T}^{(N)} \\ & \leq \frac{\left( \frac{Q^{-1}(\epsilon)\sqrt{\Upsilon(T)}}{F(T)} \right)^{2-2^{-N+1}}}{L^{1-2^{-N}}} \left( \frac{\bar{V}_{\text{UB}}(T, \rho_0)}{\Upsilon(T)} \right)^{2^{-N}} \end{aligned} \quad (205)$$

for every  $\rho \geq \rho_0$ ,  $L \geq L_0$ , and sufficiently large  $L_0$  and  $\rho_0$ .<sup>9</sup> Thus, by letting  $N$  tend to infinity, we conclude that we can assume without loss of optimality that

$$\alpha \geq \rho \left( 1 - \delta_{L,T}^{(\infty)} \right) \quad (206)$$

<sup>9</sup>While, in principle,  $L_0$  and  $\rho_0$  may depend on  $N$ , it can be shown that one can find pairs  $(L_0, \rho_0)$  that are independent of  $N$  and that satisfy (205) for every  $N$ .

where  $\delta_{L,T}^{(\infty)}$  satisfies

$$0 \leq \delta_{L,T}^{(\infty)} \leq \frac{\left(\frac{Q^{-1}(\epsilon)\sqrt{\Upsilon(T)}}{F(T)}\right)^2}{L}. \quad (207)$$

This concludes the proof of Part 2) of Lemma 7.

#### APPENDIX H LOWER BOUND ON $\bar{V}_\rho(T, \alpha)$

We show that for all  $\rho(1 - \delta) \leq \alpha \leq \rho$ ,  $0 \leq \delta \leq 1/2$ , and  $\rho \geq \rho_0$ , we have

$$\bar{V}_\rho(T, \alpha) \geq \bar{V}(T, \rho) - \Upsilon(T)\delta \quad (208)$$

where  $\Upsilon(T)$  is a positive constant that only depends on  $T$ . Let  $\Omega(T, \alpha) \triangleq \bar{j}_\ell(T, \alpha) - \bar{J}(T, \alpha)$ , i.e.,

$$\begin{aligned} \Omega(T, \alpha) = & -\frac{T\rho - T\alpha}{1 + T\rho}(Z_1 - 1) - \frac{T\rho}{1 + T\rho}(Z_2 - (T - 1)) \\ & + (T - 1) \log((1 + T\alpha)Z_1 + Z_2 + \beta(T, \rho)) \\ & - (T - 1) \mathbb{E}[\log((1 + T\alpha)Z_1 + Z_2 + \beta(T, \rho))]. \end{aligned} \quad (209)$$

It follows that  $\bar{V}_\rho(T, \alpha) = \mathbb{E}[\Omega^2(T, \alpha)]$ . We next analyze the difference

$$\begin{aligned} & \bar{V}(T, \rho) - \bar{V}_\rho(T, \alpha) \\ &= \mathbb{E}[(\Omega(T, \rho) - \Omega(T, \alpha))(\Omega(T, \rho) + \Omega(T, \alpha))] \\ &\leq \sqrt{\mathbb{E}[(\Omega(T, \rho) - \Omega(T, \alpha))^2] \mathbb{E}[(\Omega(T, \rho) + \Omega(T, \alpha))^2]} \end{aligned} \quad (210)$$

where the inequality follows from the Cauchy-Schwarz inequality. On the one hand, using (135), we have for every  $\rho_0 > 0$ ,

$$\begin{aligned} & \sup_{\substack{\alpha > 0, \\ \rho \geq \rho_0}} \mathbb{E}[(\Omega(T, \rho) + \Omega(T, \alpha))^2] \\ &\leq c_{2,2} \sup_{\rho \geq \rho_0} \mathbb{E}[\Omega^2(T, \rho)] + c_{2,2} \sup_{\substack{\alpha \geq 0, \\ \rho \geq \rho_0}} \mathbb{E}[\Omega^2(T, \alpha)] \\ &= c_{2,2} \sup_{\rho \geq \rho_0} \bar{V}(T, \rho) + c_{2,2} \sup_{\substack{\alpha \geq 0, \\ \rho \geq \rho_0}} \bar{V}_\rho(T, \alpha) \end{aligned} \quad (211)$$

which, by Lemma 4, is bounded. On the other hand, using (135) and that  $\mathbb{E}[(X - \mathbb{E}[X])^2] \leq \mathbb{E}[X^2]$  for every random variable  $X$ , we obtain

$$\begin{aligned} & \mathbb{E}[(\Omega(T, \rho) - \Omega(T, \alpha))^2] \\ &= \mathbb{E}\left[\left(\frac{T\rho - T\alpha}{1 + T\rho}(Z_1 - 1) + (T - 1) \log\left(\frac{(1 + T\rho)Z_1 + Z_2 + \beta(T, \rho)}{(1 + T\alpha)Z_1 + Z_2 + \beta(T, \rho)}\right) - (T - 1) \mathbb{E}\left[\log\left(\frac{(1 + T\rho)Z_1 + Z_2 + \beta(T, \rho)}{(1 + T\alpha)Z_1 + Z_2 + \beta(T, \rho)}\right)\right]\right)^2\right] \\ &\leq c_{2,2}(T - 1)^2 \mathbb{E}\left[\log^2\left(\frac{(1 + T\rho)Z_1 + Z_2 + \beta(T, \rho)}{(1 + T\alpha)Z_1 + Z_2 + \beta(T, \rho)}\right)\right] \\ &\quad + c_{2,2} \left(\frac{T\rho - T\alpha}{1 + T\rho}\right)^2. \end{aligned} \quad (212)$$

When  $\rho(1 - \delta) \leq \alpha \leq \rho$ , this can be further upper-bounded as

$$\begin{aligned} & \mathbb{E}\left[(\Omega(T, \rho) - \Omega(T, \alpha))^2\right] \\ &\leq c_{2,2}(T - 1)^2 \log^2\left(\frac{1}{1 - \delta}\right) + c_{2,2}\delta^2 \\ &\leq c_{2,2}(4(T - 1)^2 + 1)\delta^2 \end{aligned} \quad (213)$$

where the last inequality follows from (130). Combining (211) and (213) with (210) yields (208).

#### APPENDIX I

##### HIGH-SNR APPROXIMATIONS OF INFORMATION RATES

*Lemma 9:* The quantities  $\bar{J}(T, \rho)$ ,  $I(T, \rho)$ , and  $\underline{I}(T, \rho)$  satisfy

$$\lim_{\rho \rightarrow \infty} \left\{ \bar{J}(T, \rho) - ((T - 1) \log(T\rho) - \log \Gamma(T) - (T - 1)(1 + \gamma)) \right\} = 0 \quad (214)$$

$$\lim_{\rho \rightarrow \infty} \left\{ I(T, \rho) - ((T - 1) \log(T\rho) - \log \Gamma(T) - (T - 1)(1 + \gamma)) \right\} = 0 \quad (215)$$

$$\lim_{\rho \rightarrow \infty} \left\{ \underline{I}(T, \rho) - ((T - 1) \log(T\rho) - \log \Gamma(T) - (T - 1)(1 + \gamma)) \right\} = 0 \quad (216)$$

for every  $T > 2$ .

*Proof:* We can express  $\bar{J}(T, \rho)$ ,  $I(T, \rho)$ , and  $\underline{I}(T, \rho)$  as (see (30), (21), and (22))

$$\begin{aligned} \bar{J}(T, \rho) = & (T - 1) \log(T\rho) - \log \Gamma(T) - (T - 1) \frac{T\rho}{1 + T\rho} \\ & + (T - 1) \mathbb{E}\left[\log\left(Z_1 + \frac{Z_2}{(1 + T\rho)}\right)\right] \\ & + (T - 1) \mathbb{E}\left[\log\left(1 + \frac{\beta(T, \rho)}{(1 + T\rho)Z_1 + Z_2}\right)\right] \end{aligned} \quad (217)$$

$$\begin{aligned} I(T, \rho) = & (T - 1) \log(T\rho) - \log \Gamma(T) - (T - 1) \frac{T\rho}{1 + T\rho} \\ & + (T - 1) \mathbb{E}\left[\log\left(Z_1 + \frac{Z_2}{(1 + T\rho)}\right)\right] \\ & - \mathbb{E}\left[\log \tilde{\gamma}\left(T - 1, \frac{T\rho((1 + T\rho)Z_1 + Z_2)}{1 + T\rho}\right)\right] \end{aligned} \quad (218)$$

$$\begin{aligned} \underline{I}(T, \rho) = & (T - 1) \log(T\rho) - \log \Gamma(T) - (T - 1) \frac{T\rho}{1 + T\rho} \\ & + (T - 1) \mathbb{E}\left[\log\left(Z_1 + \frac{Z_2}{(1 + T\rho)}\right)\right]. \end{aligned} \quad (219)$$

Note that these expressions differ only in terms that vanish as  $\rho \rightarrow \infty$ . Indeed, we have for every  $T > 2$

$$\lim_{\rho \rightarrow \infty} \left\{ \mathbb{E}\left[\log\left(Z_1 + \frac{Z_2}{(1 + T\rho)}\right)\right] - \gamma \right\} = 0 \quad (220)$$

$$\lim_{\rho \rightarrow \infty} \mathbb{E}\left[\log\left(1 + \frac{\beta(T, \rho)}{(1 + T\rho)Z_1 + Z_2}\right)\right] = 0 \quad (221)$$

$$\lim_{\rho \rightarrow \infty} \mathbb{E}\left[\log \tilde{\gamma}\left(T - 1, \frac{T\rho((1 + T\rho)Z_1 + Z_2)}{1 + T\rho}\right)\right] = 0. \quad (222)$$

We further have that

$$\lim_{\rho \rightarrow \infty} (T-1) \frac{T\rho}{1+T\rho} = (T-1). \quad (223)$$

Hence (214)–(216) follow.

It remains to prove (220)–(222). Equation (220) follows because, by the dominated convergence theorem,

$$\begin{aligned} & \lim_{\rho \rightarrow \infty} \mathbb{E} \left[ \log \left( Z_1 + \frac{Z_2}{(1+T\rho)} \right) \right] \\ &= \mathbb{E} \left[ \lim_{\rho \rightarrow \infty} \log \left( Z_1 + \frac{Z_2}{(1+T\rho)} \right) \right] \end{aligned} \quad (224)$$

and because  $\mathbb{E}[\log Z_1] = -\gamma$ . The dominated convergence theorem can be applied since (see (142))

$$\left| \log \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \right| \leq |\log(Z_1 + Z_2)| + |\log(Z_1)| \quad (225)$$

and  $\mathbb{E}[|\log(Z_1 + Z_2)| + |\log(Z_1)|] < \infty$ .

Similarly, (221) and (222) follow by the dominated convergence theorem and by noting that the terms inside the expected values on the LHS of (221) and (222) vanish as  $\rho \rightarrow \infty$ . The dominated convergence theorem can be applied because, for every  $\rho_0 > 0$  and  $\rho \geq \rho_0$ ,

$$\begin{aligned} & \left| \log \tilde{\gamma} \left( T-1, \frac{T\rho((1+T\rho)Z_1 + Z_2)}{1+T\rho} \right) \right| \\ & \leq (T-1) \log \left( 1 + \frac{\beta(T, \rho)}{(1+T\rho)Z_1 + Z_2} \right) \\ & \leq (T-1) \log \left( 1 + \frac{\beta(T, \rho_0)}{Z_1 + Z_2} \right) \end{aligned} \quad (226)$$

and because the expected value of the RHS of (226) is finite. Here, the first inequality follows from Lemma 1, and the second inequality follows because  $\rho \mapsto \beta(T, \rho)$  is monotonically decreasing in  $\rho$ . ■

## APPENDIX J

### HIGH-SNR APPROXIMATIONS OF DISPERSIONS

*Lemma 10:* The quantities  $\bar{V}(T, \rho)$  and  $U(T, \rho)$  defined in (41) and (39), respectively, satisfy

$$\lim_{\rho \rightarrow \infty} \bar{V}(T, \rho) = (T-1)^2 \frac{\pi^2}{6} + (T-1) \quad (227)$$

$$\lim_{\rho \rightarrow \infty} U(T, \rho) = (T-1)^2 \frac{\pi^2}{6} + (T-1) \quad (228)$$

for every  $T > 2$ .

*Proof:* We first prove (227) by analyzing  $\bar{V}(T, \rho)$  in the limit as  $\rho \rightarrow \infty$ . To this end, we first note that

$$\begin{aligned} & \bar{j}_\ell(T, \rho) - \bar{J}(T, \rho) \\ &= \frac{T\rho}{1+T\rho} (T-1 - Z_2) + (T-1) \log \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \\ & \quad - (T-1) \mathbb{E} \left[ \log \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \right] \\ & \quad + (T-1) \log \left( 1 + \frac{\beta(T, \rho)}{(1+T\rho)Z_1 + Z_2} \right) \\ & \quad - (T-1) \mathbb{E} \left[ \log \left( 1 + \frac{\beta(T, \rho)}{(1+T\rho)Z_1 + Z_2} \right) \right] \end{aligned} \quad (229)$$

tends to

$$T-1 - Z_2 + (T-1) \log(Z_1) - (T-1) \mathbb{E}[\log Z_1] \quad (230)$$

as  $\rho \rightarrow \infty$ . (To obtain  $\mathbb{E}[\log Z_1]$ , we interchange limit and expectation, which can be justified by the dominated convergence theorem.) Since  $Z_1$  and  $Z_2$  are independent, we have that

$$\begin{aligned} & \mathbb{E} \left[ (T-1 - Z_2 + (T-1) \log(Z_1) - (T-1) \mathbb{E}[\log Z_1])^2 \right] \\ &= \mathbb{E}[(T-1 - Z_2)^2] \\ & \quad + (T-1)^2 \left( \mathbb{E}[\log^2(Z_1)] - \mathbb{E}[\log Z_1]^2 \right) \\ &= (T-1) + (T-1)^2 \frac{\pi^2}{6}. \end{aligned} \quad (231)$$

It remains to show that we can swap limit and expectation. To this end, we next argue that the dominated convergence theorem applies. Indeed, proceeding similarly as in Appendix C, we note that for every  $\rho_0 > 0$  and  $\rho \geq \rho_0$

$$\begin{aligned} & (\bar{j}_\ell(T, \rho) - \bar{J}(T, \rho))^2 \\ & \leq c_{5,2} \left( \frac{T\rho}{1+T\rho} \right)^2 (Z_2 - T + 1)^2 \\ & \quad + c_{5,2} (T-1)^2 \log^2 \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \\ & \quad + c_{5,2} (T-1)^2 \mathbb{E} \left[ \log \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \right]^2 \\ & \quad + c_{5,2} (T-1)^2 \log^2 \left( 1 + \frac{\beta(T, \rho)}{(1+T\rho)Z_1 + Z_2} \right) \\ & \quad + c_{5,2} (T-1)^2 \mathbb{E} \left[ \log \left( 1 + \frac{\beta(T, \rho)}{(1+T\rho)Z_1 + Z_2} \right) \right]^2 \\ & \leq c_{5,2} (Z_2 - T + 1)^2 + (T-1)^2 \log^2(Z_1) \\ & \quad + c_{5,2} (T-1)^2 \log^2(Z_1 + Z_2) \\ & \quad + c_{5,2} (T-1)^2 \mathbb{E} [|\log(Z_1 + Z_2)| + |\log(Z_1)|]^2 \\ & \quad + c_{5,2} (T-1)^2 \log^2 \left( 1 + \frac{\beta(T, \rho_0)}{Z_1 + Z_2} \right) \\ & \quad + c_{5,2} (T-1)^2 \mathbb{E} \left[ \log \left( 1 + \frac{\beta(T, \rho_0)}{Z_1 + Z_2} \right) \right]^2. \end{aligned} \quad (232)$$

To obtain the second inequality in (232), we upper-bound the second term using that (see (125))

$$\log^2 \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \leq \log^2(Z_1 + Z_2) + \log^2(Z_1),$$

the third term using that (see (142))

$$\left| \log \left( Z_1 + \frac{Z_2}{1+T\rho} \right) \right| \leq |\log(Z_1 + Z_2)| + |\log(Z_1)|,$$

the fourth term using that, for every  $\rho_0 > 0$  and  $\rho \geq \rho_0$ ,

$$\begin{aligned} \log^2 \left( 1 + \frac{\beta(T, \rho)}{(1+T\rho)Z_1 + Z_2} \right) & \leq \log^2 \left( 1 + \frac{\beta(T, \rho)}{Z_1 + Z_2} \right) \\ & \leq \log^2 \left( 1 + \frac{\beta(T, \rho_0)}{Z_1 + Z_2} \right), \end{aligned}$$

and the fifth term using that, for every  $\rho_0 > 0$  and  $\rho \geq \rho_0$ ,

$$\log \left( 1 + \frac{\beta(T, \rho)}{(1+T\rho)Z_1 + Z_2} \right) \leq \log \left( 1 + \frac{\beta(T, \rho_0)}{Z_1 + Z_2} \right).$$

Since the expected value of the RHS of (232) is finite, the dominated convergence theorem applies and (227) follows.

To prove (228) we proceed similarly. Indeed, by Lemma 1,

$$\begin{aligned} & i_\ell(T, \rho) - I(T, \rho) \\ &= \frac{T\rho}{1+T\rho}(T-1-Z_2) + (T-1) \log\left(Z_1 + \frac{Z_2}{1+T\rho}\right) \\ & \quad - (T-1) \mathbb{E} \left[ \log\left(Z_1 + \frac{Z_2}{1+T\rho}\right) \right] \\ & \quad - \log \tilde{\gamma}\left(T-1, \frac{T\rho((1+T\rho)Z_1+Z_2)}{1+T\rho}\right) \\ & \quad + \mathbb{E} \left[ \log \tilde{\gamma}\left(T-1, \frac{T\rho((1+T\rho)Z_1+Z_2)}{1+T\rho}\right) \right] \quad (233) \end{aligned}$$

tends to (230) as  $\rho$  tends to infinity. It remains to show that limit and expectation can be swapped. We next argue that this follows from dominated convergence theorem. Indeed, using (135), we obtain for every  $\rho_0 > 0$  and  $\rho \geq \rho_0$  that

$$\begin{aligned} & (i_\ell(T, \rho) - I(T, \rho))^2 \\ & \leq c_{5,2} \left(\frac{T\rho}{1+T\rho}\right)^2 (Z_2 - T + 1)^2 \\ & \quad + c_{5,2} (T-1)^2 \log^2\left(Z_1 + \frac{Z_2}{1+T\rho}\right) \\ & \quad + c_{5,2} (T-1)^2 \mathbb{E} \left[ \log\left(Z_1 + \frac{Z_2}{1+T\rho}\right) \right]^2 \\ & \quad + c_{5,2} \log^2 \tilde{\gamma}\left(T-1, \frac{T\rho((1+T\rho)Z_1+Z_2)}{1+T\rho}\right) \\ & \quad + c_{5,2} \mathbb{E} \left[ \log \tilde{\gamma}\left(T-1, \frac{T\rho((1+T\rho)Z_1+Z_2)}{1+T\rho}\right) \right]^2 \\ & \leq c_{5,2} (Z_2 - T + 1)^2 \\ & \quad + c_{5,2} (T-1)^2 \log^2(Z_1 + Z_2) + (T-1)^2 \log^2(Z_1) \\ & \quad + c_{5,2} (T-1)^2 \mathbb{E} \left[ |\log(Z_1 + Z_2)| + |\log(Z_1)| \right]^2 \\ & \quad + c_{5,2} (T-1)^2 \log^2\left(1 + \frac{\beta(T, \rho_0)}{Z_1 + Z_2}\right) \\ & \quad + c_{5,2} (T-1)^2 \mathbb{E} \left[ \log\left(1 + \frac{\beta(T, \rho_0)}{Z_1 + Z_2}\right) \right]^2. \quad (234) \end{aligned}$$

Here, we upper-bound the first three terms as in (232), and the fourth and fifth term using Lemma 1 and the monotonicity of  $\rho \mapsto \beta(T, \rho)$ . Since the expected value of the RHS of (234) is finite, the dominated convergence theorem applies and (228) follows. ■

#### ACKNOWLEDGMENT

The authors gratefully acknowledge fruitful discussions with Gonzalo Vazquez-Vilar. They further thank Mustafa C. Coşkun for producing the performance curve of the ARJA LDPC code shown in Fig. 6.

#### 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, Sep. 2016.  
 [2] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.

[3] 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.  
 [4] Y. Polyanskiy and S. Verdú, "Scalar coherent fading channel: Dispersion analysis," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, St. Petersburg, Russia, Jul. 2011, pp. 2959–2963.  
 [5] A. Collins and Y. Polyanskiy, "Orthogonal designs optimize achievable dispersion for coherent MISO channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jun. 2014, pp. 2524–2528.  
 [6] —, "Dispersion of the coherent MIMO block-fading channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 1068–1072.  
 [7] —, "Coherent multiple-antenna block-fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 380–405, Jan. 2019.  
 [8] W. Yang, G. Caire, G. Durisi, and Y. Polyanskiy, "Optimum power control at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4598–4615, Sep. 2015.  
 [9] 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.  
 [10] E. Molavianjazi and J. N. Laneman, "On the second-order coding rate of non-ergodic fading channels," in *Proc. 51st Allerton Conf. Commun. Control Comput.*, Monticello, IL, USA, Oct. 2013, pp. 583–587.  
 [11] 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.  
 [12] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Diversity versus channel knowledge at finite block-length," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Lausanne, Switzerland, Sep. 2012, pp. 572–576.  
 [13] J. Östman, W. Yang, G. Durisi, and T. Koch, "Diversity versus multiplexing at finite blocklength," in *Proc. IEEE Int. Symp. on Wireless Commun. Syst. (ISWCS)*, Barcelona, Spain, Aug. 2014, pp. 702–706.  
 [14] G. Durisi, T. Koch, J. Östman, Y. Polyanskiy, and W. Yang, "Short-packet communications over multiple-antenna Rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 618–629, Feb. 2016.  
 [15] 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.  
 [16] 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.  
 [17] 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.  
 [18] M. Abramowitz, *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*. New York, NY, USA: Dover Publications, 1974.  
 [19] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. Amsterdam, The Netherlands: Elsevier/Academic Press, 2007.  
 [20] M. Geller and E. W. Ng, "A table of integrals of the exponential integral," *Journal of Research of the National Bureau of Standards - B, Mathematics and Mathematical Science*, vol. 738, no. 3, pp. 191–210, Jul.-Sep. 1969.  
 [21] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.  
 [22] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.  
 [23] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.  
 [24] T. Ericson, "A Gaussian channel with slow fading," *IEEE Trans. Inf. Theory*, vol. 16, no. 3, pp. 353–355, May 1970.  
 [25] J. Östman, G. Durisi, E. G. Ström, M. C. Coşkun, and G. Liva, "Short packets over block-memoryless fading channels: Pilot-assisted or noncoherent transmission?" *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1521–1536, Feb. 2019.  
 [26] H. Nagaoka, "Strong converse theorems in quantum information theory," in *Proc. ERATO Workshop on Quantum Information Science*, Tokyo, Japan, Sep. 2001, p. 6–8.  
 [27] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.  
 [28] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1753–1768, Jul. 2003.



- [29] J. Östman, R. Devassy, G. C. Ferrante, and G. Durisi, "Low-latency short-packet transmissions: Fixed length or HARQ?" in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.
- [30] M. C. Coşkun, G. Durisi, T. Jerkovits, G. Liva, W. Ryan, B. Stein, and F. Steiner, "Efficient error-correcting codes in the short blocklength regime," *Physical Communication*, vol. 34, pp. 66 – 79, Jun. 2019.
- [31] M. C. Coşkun, G. Liva, J. Östman, and G. Durisi, "Low-complexity joint channel estimation and list decoding of short codes," in *Proc. 12th Int. ITG Conf. Sys. Commun. Coding (SCC)*, Rostock, Germany, Feb. 2019.
- [32] V. Y. F. Tan and M. Tomamichel, "The third-order term in the normal approximation for the AWGN channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2430–2438, May 2015.
- [33] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach*, 6th ed. New Jersey, USA: Pearson, 2012.
- [34] W. Feller, *An Introduction To Probability Theory And Its Applications*, 2nd ed. New York, NY, USA: Wiley, 1971, vol. II.
- [35] H. Alzer, "On some inequalities for the incomplete gamma function," *Math Comp.*, vol. 66, no. 218, pp. 771–778, Apr. 1997.
- [36] W. Gautschi, "The incomplete gamma functions since Tricomi," in *Atti dei Convegni Lincei, 147*, Accademia Nazionale dei Lincei, Rome, 1998, pp. 203–237.
- [37] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY, USA: McGraw-Hill, 1987.
- [38] —, *Principles of Mathematical Analysis*, 3rd ed. New York, NY, USA: McGraw-Hill, 1976.
- [39] R. Durrett, *Probability: Theory and Examples*, 3rd ed. Belmont, USA: Duxbury Advances Series, 2005.

**Alejandro Lancho** (S'14–M'19) received the B.E., M.Sc. and Ph.D. degrees in Electrical Engineering from the Universidad Carlos III de Madrid, Spain, in 2013, 2014 and 2019, respectively. During the fall of 2017, he was a visiting researcher at Chalmers University of Technology. Since October, 2019, he works at Chalmers University of Technology as a postdoctoral researcher.

Alejandro Lancho was an FPU fellow (Spanish Ministerio de Educación Cultura y Deporte) during his Ph.D. studies. He was among the six finalists for the IEEE Jack Keil Wolf ISIT Student Paper Award at the 2017 IEEE International Symposium on Information Theory, Aachen, Germany, 2017. His research interests are in the areas of information theory and wireless communications.

**Tobias Koch** (S'02–M'09–SM'16) is a Visiting Professor and Ramón y Cajal Research Fellow with the Signal Theory and Communications Department of Universidad Carlos III de Madrid (UC3M). He received the M.Sc. degree in electrical engineering (with distinction) in 2004 and the Ph.D. degree in electrical engineering in 2009, both from ETH Zurich, Switzerland. From June 2010 until May 2012 he was a Marie Curie Intra-European Research Fellow with the University of Cambridge, UK. He was also a research intern at Bell Labs, Murray Hill, NJ, USA in 2004, and the Universitat Pompeu Fabra (UPF), Barcelona, Spain, in 2007. He joined the Signal Processing Group of UC3M in June 2012. His research interests are in digital communication theory and information theory.

Dr. Koch received a Starting Grant from the European Research Council (ERC), a Ramón y Cajal Research Fellowship, a Marie Curie Intra-European Fellowship, a Marie Curie Career Integration Grant, and a Fellowship for Prospective Researchers from the Swiss National Science Foundation. He further received a medal of the 2018 Young Researchers Award "Agustín de Betancourt y Molina" by the Real Academia de Ingeniería. In 2013–2016 he served as Vice Chair of the Spain Chapter of the IEEE Information Theory Society.

**Giuseppe Durisi** (S'02–M'06–SM'12) received the Laurea degree summa cum laude and the Doctor degree both from Politecnico di Torino, Italy, in 2001 and 2006, respectively. From 2002 to 2006, he was with Istituto Superiore Mario Boella, Torino, Italy. From 2006 to 2010 he was a postdoctoral researcher at ETH Zurich, Zurich, Switzerland. In 2010, he joined Chalmers University of Technology, Gothenburg, Sweden, where he is now professor with the Communication Systems Group. He is also co-director of Chalmers ICT Area of Advance, and of Chalmers AI Research Center.

Dr. Durisi is a senior member of the IEEE. He is the recipient of the 2013 IEEE ComSoc Best Young Researcher Award for the Europe, Middle East, and Africa Region, and is co-author of a paper that won a "student paper award" at the 2012 International Symposium on Information Theory, and of a paper that won the 2013 IEEE Sweden VT-COM-IT joint chapter best student conference paper award. In 2015, he joined the editorial board of the IEEE Transactions on Communications as associate editor. From 2011 to 2014, he served as publications editor for the IEEE Transactions on Information Theory. His research interests are in the areas of communication and information theory and machine learning.