Saddlepoint Approximations for Short-Packet Wireless Communications

N.B. When citing this work, cite the original published paper.

(article starts on next page)

Saddlepoint Approximations for Short-Packet Wireless Communications

Alejandro Lancho, *Member, IEEE,* Johan Östman, *Student Member, IEEE,* Giuseppe Durisi, *Senior Member, IEEE,* Tobias Koch, *Senior Member, IEEE,* and Gonzalo Vazquez-Vilar, *Member, IEEE*

Abstract—In recent years, the derivation of nonasymptotic converse and achievability bounds on the maximum coding rate as a function of the error probability and blocklength has gained attention in the information theory literature. While these bounds are accurate for many scenarios of interest, they need to be evaluated numerically for most wireless channels of practical interest, and their evaluation is computationally demanding. This paper presents saddlepoint approximations of state-of-the-art converse and achievability bounds for noncoherent, single-antenna, Rayleigh block-fading channels. These approximations can be calculated efficiently and are shown to be accurate for SNR values as small as 0 dB and blocklengths of 168 channel uses or more.

Index Terms—Fading channels, finite-blocklength information theory, saddlepoint approximations, short packets, ultra-reliable low-latency communications.

I. INTRODUCTION

ONE of the main goals in the design of communication systems is to maximize the transmission rate (also referred to as *coding rate*) under given constraints on system parameters, such as power, bandwidth, latency, or reliability. In the absence of stringent latency constraints, which is the case for traditional wireless communication systems, the transmission rate can be improved by increasing the length of the packets sent to convey information from a transmitter to a receiver. We shall refer to the packet length as *blocklength*. This implies that asymptotic information-theoretical metrics,

such as *capacity* or *outage capacity*, are good benchmarks. Indeed, capacity is defined as the maximum coding rate at which data can be transmitted with vanishing error probability when the blocklength tends to infinity. Similarly, outage capacity is defined as the maximum coding rate at which data can be transmitted for a given error probability when the blocklength tends to infinity. These metrics have been widely used as performance benchmarks in the design of coding schemes for traditional wireless communication systems.

Next generation wireless communication systems will support services and applications requiring *ultra-reliable low-latency communications* (URLLC) [1], [2]. In URLLC, the devices exchange packets at low rates aiming for probabilities of error less than or equal to 10^{-5} and satisfying stringent latency constraints [2]. This is typically achieved by transmitting short packets. Recalling that capacity and outage capacity are defined under the assumption that the blocklength tends to infinity, it follows that for latency-constrained communications a more refined characterization of the maximum coding rate as a function of the blocklength is needed.

A. State of the Art

For most channel models of interest, obtaining a closed-form expression of the maximum coding rate is out of reach. Hence, there are two main directions to characterize the maximum coding rate as a function of the blocklength:

i) **Nonasymptotic bounds**: By obtaining upper and lower bounds on the maximum coding rate, the area in which this rate lies for a specific error probability and blocklength can be characterized. Often such bounds are expressed in terms of tail probabilities of sums of independent and identically distributed (i.i.d.) random variables and need to be evaluated numerically through computationally demanding procedures. Nonasymptotic bounds for several channel models can be found, e.g., in [3], [4]. The nonasymptotic bounds considered in this paper are the *meta-converse (MC) bound* [3, Th. 31] and the *random coding union bound with parameter s* (RCU$_s$) [5, Th. 1]. The MC bound and the RCU$_s$ bound yield an upper and a lower bound on the maximum coding rate for a fixed error probability and blocklength, respectively.

ii) **Refined asymptotic expansions**:[1] Perform asymptotic expansions of the maximum coding rate or the error probability that become accurate as the blocklength grows.

---

[1]We refer to these expansion as *refined* asymptotic expansions to better differentiate them from capacity or outage capacity, which are first-order asymptotic expansions of the maximum coding rate.

Typically, such expansions are available in closed form and describe how the maximum coding rate converges to capacity or how the error probability vanishes as the blocklength increases for a fixed coding rate.

The refined asymptotic expansions are usually obtained by expanding the tail probabilities appearing in the aforementioned nonasymptotic bounds. One possibility is to fix a reliability constraint and study the maximum coding rate as a function of the blocklength in the limit as the blocklength tends to infinity. This approach was followed *inter alia* by Polyanskiy *et al.* [3] who showed, for various channels with positive finite capacity C, that the maximum coding rate $R^*(n, \epsilon)$ at which data can be transmitted using an error-correcting code of a fixed blocklength n, with a block-error probability not larger than ϵ, can be expanded as

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

where V denotes the channel dispersion, $Q^{-1}(\cdot)$ denotes the inverse of the Gaussian Q-function, and $\mathcal{O}((\log n)/n)$ comprises terms that decay no slower than $(\log n)/n$. The approximation that follows from (1) by omitting the $\mathcal{O}((\log n)/n)$ term is sometimes referred to as *normal approximation*. The normal approximation has been established as a benchmark for short error-correcting codes. For example, [6] compares the performance of codes of blocklength $n = 128$ against the normal approximation of a Gaussian channel. The normal approximation further serves as a proxy for the maximum coding rate in the analysis and optimization of communication systems that exchange short packets. For example, the normal approximation has appeared in numerous papers on short-packet wireless communications, including [7]–[15].

The work by Polyanskiy *et al.* [3] has been generalized to several wireless communication channels; see, e.g., [16]–[20]. Particularly relevant for this work is the high-SNR normal approximation for noncoherent single-antenna Rayleigh block-fading channels derived in [20]. By means of numerical examples, it was shown in [20] that this normal approximation is accurate for probabilities of error above 10^{-3} and SNR values larger than or equal to 15 dB.

A second option to obtain refined asymptotic expansions is to fix the coding rate and to study the exponential decay of the error probability as the blocklength grows to infinity. Error-exponent results for block-fading channels can be found, e.g., in [21]–[25].

In general, normal approximations are accurate for moderate error probabilities and when the rate is close to capacity. In contrast, error exponents are accurate for moderate coding rates and when the probability of error is close to zero. URLLC services operate at error probabilities of around 10^{-5} and SNR values of around 0 dB [2]. For these values, both normal approximations and error exponents may become inaccurate, in which case they constitute poor benchmarks for short error-correcting codes, and using them as a proxy for the maximum coding rate or the error probability in the analysis and optimization of short-packet communication systems may give rise to misleading results. Consequently, there is a need

for refined approximations that characterize the coding rate for error probabilities below 10^{-5} and SNR values close to 0 dB.

B. Contributions

To provide approximations that are accurate in regimes where neither normal approximations nor error exponents are, this paper makes use of the classical *saddlepoint method* [26, Ch. XVI], [27]. The saddlepoint method has been applied in [28] to obtain approximations of the RCU$_{\mathrm{s}}$ bound and the *random coding union (RCU) bound* [3, Th. 16] for channels with finite alphabets. It has been further applied in [29], [30] to obtain approximations of the RCU bound and MC bound for symmetric memoryless channels. Similar approximations based on Laplace integration were used in [31], [32] to approximate the RCU bound and MC bound for Gaussian channels.

Saddlepoint expansions of nonasymptotic bounds can be obtained by applying the saddlepoint method to expand the tail probabilities of sums of i.i.d. random variables appearing in these bounds. These random variables typically depend on system parameters such as the SNR. Since existing saddlepoint methods ignore such dependencies, the error terms in the saddlepoint expansions also depend on these parameters and may even be unbounded in them. This is particularly problematic if the saddlepoint expansions are the starting point for asymptotic analyses.

To overcome this problem, we derive in Section IV saddlepoint expansions for random variables whose distribution depends on a system parameter θ, carefully analyze the error terms, and demonstrate that they are uniform in θ. The obtained expansions are then applied to the tail probabilities appearing in the nonasymptotic MC and RCU$_{\mathrm{s}}$ bounds, leading to saddlepoint expansions of these bounds with error terms that depend only on the blocklength and are uniform in the remaining parameters. By means of numerical analyses, we show that the *saddlepoint approximations* that follow by ignoring the error terms in the saddlepoint expansions are accurate, i.e., they are indistinguishable from the corresponding nonasymptotic bounds, if the SNR is greater than or equal to 0 dB and if the probability of error is larger than or equal to 10^{-8}. Furthermore, they require a much lower computational cost than the corresponding nonasymptotic bounds. Finally, we show that the normal approximation and the error exponent can be recovered from the saddlepoint expansions. They thus provide a unifying characterization of the two regimes, which are usually considered separately in the literature.

The rest of this paper is organized as follows. In Section II, we present the system model. In Section III, we introduce the most important quantities and the nonasymptotic bounds used in the paper. In Section IV, we derive saddlepoint expansions for random variables whose distribution depends on a parameter θ. In Section V, we apply those saddlepoint expansions to the nonasymptotic bounds presented in Section III. In Section VI, we briefly discuss how the normal approximation and error exponent can be obtained from the saddlepoint expansions. In Section VII, we present numerical examples that help to assess the accuracy of the presented
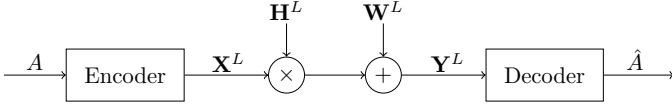
Fig. 1. Schema of the communication system with a single-antenna Rayleigh block-fading channel as channel model.

approximations. In Section VIII, we discuss the computational complexity and accuracy of the proposed saddlepoint approximations, and we compare them with the nonasymptotic bounds as well as the refined asymptotic expansions available in the literature. Section IX concludes the paper. Some of the proofs are deferred to the appendices.

II. SYSTEM MODEL

The system and channel model is illustrated in Fig. 1. We consider a single-antenna Rayleigh block-fading channel with coherence interval T—a channel model that is commonly used in the literature; see e.g., [33], [34]. For this channel model, the input-output relation within the ℓ-th coherence interval is given by

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

where $\mathbf{X}_\ell$ and $\mathbf{Y}_\ell$ are T-dimensional, complex-valued, random vectors containing the input and output signals, respectively; $\mathbf{W}_\ell$ is the additive noise with i.i.d., zero-mean, unit-variance, circularly-symmetric, complex Gaussian entries; and H_ℓ is a zero-mean, unit-variance, circularly-symmetric, complex Gaussian random variable. We assume that H_ℓ and $\mathbf{W}_\ell$ are independent and take on independent realizations over successive coherence intervals. Moreover, 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_ℓ but not of its realization.[2]

Note that, by considering a noncoherent setting, we do not preclude the possibility of estimating the fading coefficients, e.g., by transmitting pilot symbols. Instead, as in [35], we view the transmission of pilot symbols as a special case of coding and the estimation of the fading coefficients as part of the decoder. By studying the maximum coding rate of the noncoherent Rayleigh block-fading channel, we thus characterize the largest transmission rate at which data can be transmitted, irrespective of the employed channel-estimation method. See also [1].

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 entire codeword. An $(M, L, T, \epsilon, \rho)$-code for the channel (2) consists of:

1) An encoder $f: \{1, \ldots, M\} \to \mathbb{C}^{LT}$ that maps a message A, which is uniformly distributed on $\{1, \ldots, M\}$, to a

codeword $\mathbf{X}^L = [\mathbf{X}_1, \ldots, \mathbf{X}_L] = f(A)$. The codewords satisfy the power constraint[3]

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

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

2) A decoder $g: \mathbb{C}^{LT} \to \{1, \ldots, M\}$ satisfying the average error probability constraint

$$\mathsf{P}\big[g\big(\mathbf{Y}^L\big) \neq A\big] \leq \epsilon \qquad (4)$$

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

The *maximum coding rate* and the *minimum error probability* are respectively defined as[4]

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

$$\epsilon^*(L, T, R, \rho) \triangleq \inf\left\{ \epsilon : \exists (e^{LTR}, L, T, \epsilon, \rho)\text{-code} \right\}. \quad (5b)$$

In words, $R^*(L, T, \epsilon, \rho)$ denotes the largest transmission rate at which data can be transmitted with an error probability not exceeding ϵ using a channel code of blocklength LT. Likewise, $\epsilon^*(L, T, R, \rho)$ denotes the smallest error probability at which data can be transmitted at rate R using a channel code of blocklength LT. In this paper, we shall present our results in terms of error probabilities and use that upper (lower) bounds on $\epsilon^*(L, T, R, \rho)$ can be translated into lower (upper) bounds on $R^*(L, T, \epsilon, \rho)$ and *vice versa*.

III. PRELIMINARIES AND NONASYMPTOTIC BOUNDS

A. Preliminaries

According to (2), conditioned on $\mathbf{X}^L = \mathbf{x}^L$, the output vector $\mathbf{Y}^L$ is blockwise i.i.d. Gaussian. Thus, the conditional probability density function (pdf) of $\mathbf{Y}_\ell$ given $\mathbf{X}_\ell = \mathbf{x}$ is independent of ℓ and satisfies

$$\mathsf{p}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{1}{\pi^T(1 + \|\mathbf{x}\|^2)} \exp\left\{ -\|\mathbf{y}\|^2 + \frac{|\mathbf{y}^H \mathbf{x}|^2}{1 + \|\mathbf{x}\|^2} \right\}. \qquad (6)$$

Here and throughout the paper, we omit the subscript ℓ when immaterial.

We shall evaluate the achievability bounds for inputs of the form $\mathbf{X}^L = \sqrt{T\rho}\, \mathbf{U}^L$, where the components of $\mathbf{U}^L = [\mathbf{U}_1, \ldots, \mathbf{U}_L]$ are i.i.d. and uniformly distributed on the unit sphere in $\mathbb{C}^T$. This distribution of $\mathbf{X}^L$ can be viewed as a single-antenna particularization of *unitary space-time modulation* (USTM) [36] and is capacity achieving for the

---

[2]The assumption that H_ℓ and $\mathbf{W}_\ell$ take on independent realizations over successive coherence intervals is critical for the results obtained in this paper, namely, Proposition 1 in Section IV and the saddlepoint expansions presented in Section V that follow from it. In contrast, the assumption that H_ℓ is zero-mean Gaussian is not critical since Proposition 1 applies to any fading distribution for which (18) and (19) are satisfied.

[3]While in the information and communication theory literature, it is more common to impose a power constraint per codeword $\mathbf{X}^L$, practical systems typically require a per-coherence-interval constraint. Although it may be preferable to impose (3) with inequality, since it allows more freedom in optimizing the channel code, it seems plausible that using maximum power is optimal. For the high-SNR normal approximation presented in [20], this turns out to be the case.

[4]Here and throughout this paper, $\log(\cdot)$ denotes the natural logarithm. Thus, rates have the dimension nats per channel use.

power constraint (3).[5] In the following, we shall write $\bar{\mathsf{P}}_{\mathbf{X}}$ to denote the distribution of $\mathbf{X}_\ell = \sqrt{T\rho}\mathbf{U}_\ell$.

For the USTM input distribution, we define the *generalized information density* as

$$i_s(\mathbf{x}_\ell; \mathbf{x}_\ell) \triangleq \log \frac{\mathsf{p}_{\mathbf{Y}_\ell|\mathbf{X}_\ell}(\mathbf{y}_\ell|\mathbf{x}_\ell)^s}{\int \mathsf{p}_{\mathbf{Y}_\ell|\mathbf{X}_\ell}(\mathbf{y}_\ell|\tilde{\mathbf{x}})^s \mathrm{d}\bar{\mathsf{P}}_{\mathbf{X}}(\tilde{\mathbf{x}})}, \quad s > 0. \quad (7)$$

Introducing the generalized information density is convenient because its cumulant generating function (CGF) is closely related to Gallager's $E_0(\rho, s)$ function, which plays an important role in the analysis of error exponents; see [5] for more details. It can be shown that $i_s(\mathbf{X}_\ell; \mathbf{Y}_\ell)$ depends on $\mathbf{X}_\ell$ only via $\|\mathbf{X}_\ell\|^2 = T\rho$. Conditioned on $\|\mathbf{X}_\ell\|^2 = T\rho$, we obtain that $i_s(\mathbf{X}_\ell; \mathbf{Y}_\ell)$ can be written as

$$i_s(\mathbf{X}_\ell; \mathbf{Y}_\ell) \overset{\mathscr{L}}{=} (T-1)\log(sT\rho) - \log\Gamma(T) - s\frac{T\rho\Upsilon_{2,\ell}}{1+T\rho}$$
$$+ (T-1)\log\left(\frac{(1+T\rho)\Upsilon_{1,\ell} + \Upsilon_{2,\ell}}{1+T\rho}\right)$$
$$- \log\tilde{\gamma}\left(T-1, s\frac{T\rho((1+T\rho)\Upsilon_{1,\ell} + \Upsilon_{2,\ell})}{1+T\rho}\right)$$
$$(8)$$

where we use "$\overset{\mathscr{L}}{=}$" to denote equality in distribution, and where $\Gamma(\cdot)$ and $\tilde{\gamma}(\cdot)$ denote the gamma function and the regularized lower incomplete gamma function, respectively. In (8), $\{\Upsilon_{1,\ell}\}_{\ell=1}^L$ are i.i.d. Gamma$(1,1)$-distributed random variables, and $\{\Upsilon_{2,\ell}\}_{\ell=1}^L$ are i.i.d. Gamma$(T-1,1)$-distributed random variables. For brevity, we let $i_{\ell,s}(\rho) \triangleq i_s(\mathbf{X}_\ell; \mathbf{Y}_\ell)$ and denote the expectation and variance of $i_{\ell,s}(\rho)$ by $I_s(\rho) \triangleq \mathsf{E}[i_{\ell,s}(\rho)]$ and $V_s(\rho) \triangleq \mathsf{Var}[i_{\ell,s}(\rho)]$, respectively.

B. Nonasymptotic Bounds

We next present the nonasymptotic bounds that we shall use in this paper. As upper bound on $\epsilon^*(L, T, R, \rho)$, we use the RCU$_s$ bound [5, Th. 1], which states that, for every $s > 0$, there exists a channel code of blocklength LT and rate R satisfying

$$\epsilon^*(L, T, R, \rho)$$
$$\leq \mathsf{P}\left[\sum_{\ell=1}^L (I_s(\rho) - i_{\ell,s}(\rho)) \geq LI_s(\rho) + \log U - LTR\right] \quad (9)$$

where U is uniformly distributed on the interval $[0, 1]$.

To present a lower bound on $\epsilon^*(L, T, R, \rho)$, consider the auxiliary output pdf

$$\mathsf{q}_{\mathbf{Y}_\ell}^s(\mathbf{y}_\ell) \triangleq \frac{1}{\mu(s)}\left(\int \mathsf{p}_{\mathbf{Y}_\ell|\mathbf{X}_\ell}(\mathbf{y}_\ell|\mathbf{x}_\ell)^s \mathrm{d}\bar{\mathsf{P}}_{\mathbf{X}}(\mathbf{x}_\ell)\right)^{1/s}, \quad s > 0 \quad (10)$$

where $\mu(s)$ is a normalizing factor. We define the *generalized mismatched information density*[6] as

$$j_s(\mathbf{X}_\ell; \mathbf{Y}_\ell) \triangleq \log\frac{\mathsf{p}_{\mathbf{Y}_\ell|\mathbf{X}_\ell}(\mathbf{Y}_\ell|\mathbf{X}_\ell)}{\mathsf{q}_{\mathbf{Y}_\ell}^s(\mathbf{y}_\ell)}$$
$$= \log\mu(s) + \frac{1}{s}i_s(\mathbf{X}_\ell; \mathbf{Y}_\ell), \quad s > 0. \quad (11)$$

For the sake of brevity, we define $j_{\ell,s}(\rho) \triangleq j_s(\mathbf{X}_\ell; \mathbf{Y}_\ell)$ and $J_s(\rho) \triangleq \mathsf{E}[j_{\ell,s}(\rho)]$. When $s = 1$, we have $j_{\ell,1}(\rho) = i_{\ell,1}(\rho)$ and $J_1(\rho) = I_1(\rho)$, in which case we omit the subscript and write $i_\ell(\rho) \triangleq i_{\ell,1}(\rho)$, $V(\rho) = V_1(\rho)$, and $C(\rho) \triangleq I_1(\rho)$.[7]

A lower bound on $\epsilon^*(L, T, R, \rho)$ follows by evaluating the MC bound [3, Th. 31] for the auxiliary pdf $\mathsf{q}_{\mathbf{Y}_\ell}^s$ and using [3, Eq. (102)]. This yields that, for every channel code of blocklength LT and rate R, we have

$$\epsilon^*(L, T, R, \rho) \geq -e^{(\log\xi - LTR)}$$
$$+ \mathsf{P}\left[\sum_{\ell=1}^L (I_s(\rho) - i_{\ell,s}(\rho)) \geq sLJ_s(\rho) - s\log\xi\right] \quad (12)$$

for every $\xi > 0$ and $s > 0$.

IV. SADDLEPOINT EXPANSION

Let $Z_{1,\theta}, \ldots, Z_{n,\theta}$ be a sequence of i.i.d., real-valued, zero-mean random variables. The distribution of $Z_{\ell,\theta}$ depends on $\theta \in \Theta$, where Θ denotes the set of possible values of θ.

For future reference, the moment generating function (MGF) of $Z_{\ell,\theta}$ is defined as

$$m_\theta(\zeta) \triangleq \mathsf{E}[e^{\zeta Z_{\ell,\theta}}] \quad (13)$$

and the CGF is defined as

$$\psi_\theta(\zeta) \triangleq \log m_\theta(\zeta). \quad (14)$$

Furthermore, the characteristic function is defined as

$$\varphi_\theta(\zeta) \triangleq \mathsf{E}[e^{\imath\zeta Z_{\ell,\theta}}] \quad (15)$$

where $\imath \triangleq \sqrt{-1}$. We denote by $m_\theta^{(k)}$ and $\psi_\theta^{(k)}$ the k-th derivative of $\zeta \mapsto m_\theta(\zeta)$ and $\zeta \mapsto \psi_\theta(\zeta)$, respectively. For the first three derivatives we also use the notation m_θ', m_θ'', m_θ''', ψ_θ', ψ_θ'', and ψ_θ'''.

A random variable $Z_{\ell,\theta}$ is said to be *lattice* if it is supported on the points b, $b \pm h$, $b \pm 2h\ldots$ for some b and h. A random variable that is not lattice is said to be *nonlattice*. It can be shown that a random variable is nonlattice if, and only if, [26, Ch. XV.1, Lemma 4]

$$|\varphi_\theta(\zeta)| < 1, \quad \text{for every } \zeta \neq 0. \quad (16)$$

We shall say that a family of random variables $Z_{\ell,\theta}$ (parametrized by θ) is nonlattice if

$$\sup_{\theta\in\Theta}|\varphi_\theta(\zeta)| < 1, \quad \text{for every } \zeta \neq 0. \quad (17)$$

---

[5]It is unknown whether USTM is also optimal at finite blocklength. However, for the blocklengths considered in the numerical examples of Section VII, the RCU$_s$ bound, computed for USTM inputs, is close to the MC bound, which applies to all input distributions satisfying (3). This suggests that USTM is at least close to optimal at finite blocklength.

[6]We use the word "mismatched" to indicate that the auxiliary output pdf $\mathsf{q}_{\mathbf{Y}_\ell}^s$ is not necessarily the one induced by the input distribution and the channel.

[7]Recall that we choose the input distribution to be USTM, which is capacity achieving for the power constraint (3).

Similarly, we shall say that a family of distributions (parametrized by θ) is nonlattice if the corresponding family of random variables is nonlattice.

The following proposition presents saddlepoint expansions for families of random variables that are nonlattice. In Section V, these saddlepoint expansions will then be applied to the nonasymptotic bounds (9) and (12). While similar expansions could also be derived for families of lattice random variables, this would require a separate proof. Thus, for the sake of compactness, and because for many channel models of interest—including the Rayleigh block-fading channel (2)—the generalized information densities appearing in (9) and (12) are nonlattice, we do not consider families of lattice random variables in this paper.

Proposition 1: Let the family of i.i.d. random variables $\{Z_{\ell,\theta}\}_{\ell=1}^{n}$ (parametrized by θ) be nonlattice. Suppose that there exists a $\zeta_0 > 0$ such that

$$\sup_{\theta \in \Theta, |\zeta| < \zeta_0} \left| m_\theta^{(4)}(\zeta) \right| < \infty \tag{18}$$

and

$$\inf_{\theta \in \Theta, |\zeta| < \zeta_0} \psi_\theta''(\zeta) > 0. \tag{19}$$

Then, we have the following results:

Part 1: If for a given $\gamma \geq 0$ there exists a $\tau \in [0, \zeta_0)$ such that $n\psi_\theta'(\tau) = \gamma$,[8] then

$$\mathsf{P}\left[\sum_{\ell=1}^{n} Z_{\ell,\theta} \geq \gamma \right] = e^{n[\psi_\theta(\tau) - \tau\psi_\theta'(\tau)]} \Bigg[\Psi_\theta(\tau, \tau, n) \\ + \frac{K_\theta(\tau, \tau, n)}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}} \right) \Bigg] \tag{20}$$

where

$$\Psi_\theta(u, \tau, n) \triangleq e^{n\frac{u^2}{2}\psi_\theta''(\tau)} Q\left(u\sqrt{n\psi_\theta''(\tau)} \right) \tag{21}$$

$$K_\theta(u, \tau, n) \triangleq \frac{\psi_\theta'''(\tau)}{6\psi_\theta''(\tau)^{3/2}} \Bigg(-\frac{1}{\sqrt{2\pi}} + \frac{u^2 n\psi_\theta''(\tau)}{\sqrt{2\pi}} \\ - u^3 \psi_\theta''(\tau)^{3/2} n^{3/2} \Psi_\theta(u, \tau, n) \Bigg) \tag{22}$$

and $o(1/\sqrt{n})$ comprises terms that vanish faster than $1/\sqrt{n}$ and are uniform in τ and θ, i.e.,

$$\lim_{n \to \infty} \sup_{\tau \in [0,\zeta_0), \theta \in \Theta} \frac{o(1/\sqrt{n})}{1/\sqrt{n}} = 0. \tag{23}$$

Part 2: Let U be uniformly distributed on $[0,1]$. If for a given $\gamma \geq 0$ there exists a $\tau \in [0, \zeta_0)$ such that $n\psi_\theta'(\tau) = \gamma$, then

$$\mathsf{P}\left[\sum_{\ell=1}^{n} Z_{\ell,\theta} \geq \gamma + \log U \right] \\ = e^{n[\psi_\theta(\tau) - \tau\psi_\theta'(\tau)]} \Bigg[\Psi_\theta(\tau, \tau, n) + \Psi_\theta(1 - \tau, \tau, n) \\ + \frac{K_\theta(\tau, \tau, n) - K_\theta(1 - \tau, \tau, n)}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}} \right) \Bigg] \tag{24}$$

[8]In general, τ depends on n, θ, and γ. For the sake of compactness, we do not make this dependence explicit in the notation.

where $o(1/\sqrt{n})$ is uniform in τ and θ.

Proof: Part 1 is proved in Appendix I-A and Part 2 is proved in Appendix I-B. ∎

Remark 1: A Taylor series expansion of $\zeta \mapsto m_\theta^{(k)}(\zeta)$ around zero demonstrates that, if $m_\theta^{(k)}(\zeta)$ is bounded in $\theta \in \Theta$ and $|\zeta| < \zeta_0$, then the same is also true for $m_\theta^{(k-1)}(\zeta)$. Consequently, (18) implies that $\zeta \mapsto m_\theta(\zeta)$ and its first four derivatives are bounded in $\theta \in \Theta$ and $|\zeta| < \zeta_0$. This in turn implies that $\zeta \mapsto m_\theta(\zeta)$ is analytic on $\zeta \in (-\zeta_0, \zeta_0)$ for every $\theta \in \Theta$. Since Z_ℓ has zero mean by assumption, we further have that $m_\theta(\zeta) \geq 1$ by Jensen's inequality. We conclude that all derivatives of $\zeta \mapsto \psi_\theta(\zeta)$ exist on $\zeta \in (-\zeta_0, \zeta_0)$ for every $\theta \in \Theta$, and

$$\sup_{\theta \in \Theta, |\zeta| < \zeta_0} \left| \psi_\theta^{(k)}(\zeta) \right| < \infty, \quad k = 0, \dots, 4. \tag{25}$$

Remark 2: The asymptotic behaviors of $\Psi_\theta(u, \tau, n)$ and $K_\theta(u, \tau, n)$ depend critically on the asymptotic behavior of u. Indeed, the τ satisfying $n\psi_\theta'(\tau) = \gamma$, and hence also $u \in \{\tau, 1 - \tau\}$, depends in general on n. The term $\Psi_\theta(u, \tau, n)$ converges to a constant when u decays at least like $1/\sqrt{n}$, and it is of order $n^{a-1/2}$ when u is of order n^{-a}, $0 \leq a < 1/2$. Similarly, $K_\theta(u, \tau, n)$ converges to a constant when u decays at least like $1/\sqrt{n}$, and it vanishes as $n \to \infty$ when u is of order n^{-a}, $0 \leq a < 1/2$.

Since the difference $K_\theta(\tau, \tau, n) - K_\theta(1 - \tau, \tau, n)$ is difficult to evaluate, in the following corollary we present an upper bound on the saddlepoint expansion (24) that is easier to evaluate.

Corollary 2: Suppose that there exists a $\zeta_0 > 0$ satisfying (18) and (19). If for a given $\gamma \geq 0$ there exists a $\tau \in [0, \min\{\zeta_0, 1 - \delta\})$ (for some arbitrary $\delta > 0$ independent of n and θ) such that $n\psi_\theta'(\tau) = \gamma$, then the saddlepoint expansion (24) can be upper-bounded as

$$\mathsf{P}\left[\sum_{\ell=1}^{n} Z_{\ell,\theta} \geq \gamma + \log U \right] \leq e^{n[\psi_\theta(\tau) - \tau\psi_\theta'(\tau)]} \Bigg[\Psi_\theta(\tau, \tau, n) \\ + \Psi_\theta(1 - \tau, \tau, n) + \frac{\hat{K}_\theta(\tau)}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}} \right) \Bigg] \tag{26}$$

where

$$\hat{K}_\theta(\tau) \triangleq \frac{1}{\sqrt{2\pi}} \frac{\psi_\theta'''(\tau)}{6\psi_\theta''(\tau)^{3/2}} \tag{27}$$

and $o(1/\sqrt{n})$ is uniform in τ and θ.

Proof: See Appendix I-C. ∎

V. SADDLEPOINT EXPANSIONS OF RCU$_\mathrm{s}$ AND MC BOUNDS

We next apply Proposition 1 and Corollary 2 to the RCU$_\mathrm{s}$ bound (9) and MC bound (12) to obtain their saddlepoint expansions. To this end, we first express the MGF and the CGF in terms of the generalized information density (7) and discuss their regions of convergence.

The MGF and the CGF of $I_s(\rho) - i_{\ell,s}(\rho)$ are defined as

$$m_{\rho,s}(\tau) \triangleq \mathsf{E}\left[e^{\tau(I_s(\rho) - i_{\ell,s}(\rho))} \right] \tag{28}$$

$$\psi_{\rho,s}(\tau) \triangleq \log m_{\rho,s}(\tau) \tag{29}$$

and depend on the parameters $\theta = \{\rho, s\}$. For some arbitrary $0 < \underline{s} < \overline{s} < \infty$, $0 < \underline{\rho} < \overline{\rho} < \infty$, $0 < a < 1$, and $0 < b < \min\left\{\frac{T}{T-1}, \frac{1+T\overline{\rho}}{\overline{s}T\overline{\rho}}\right\}$, let

$$\mathcal{S} \triangleq \{(\tau, \rho, s) \in \mathbb{R}^3 : \tau \in [a, b], \rho \in [\underline{\rho}, \overline{\rho}], s \in [\underline{s}, \overline{s}]\}. \quad (30)$$

It can be shown that [37, Lemma 4.2]

$$\sup_{(\tau, \rho, s) \in \mathcal{S}} m_{\rho, s}^{(k)}(\tau) < \infty \quad (31)$$

for every nonnegative integer k. So $\mathcal{S}$ is in the region of convergence of $m_{\rho, s}$. We are now ready to present the following saddlepoint expansions.

Theorem 3 (Saddlepoint Expansion RCU$_s$): The coding rate R and minimum error probability ϵ^* can be parametrized by $(\tau, \rho, s) \in \mathcal{S}$ as

$$R(\tau, s) = \frac{1}{T}(I_s(\rho) - \psi'_{\rho, s}(\tau)) \quad (32a)$$

$$\epsilon^*(\tau, s) \leq e^{L[\psi_{\rho,s}(\tau) - \tau\psi'_{\rho,s}(\tau)]}\left[\Psi_{\rho,s}(\tau, \tau, n)\right.$$
$$\left. + \Psi_{\rho,s}(1 - \tau, \tau, n) + \frac{\hat{K}_{\rho,s}(\tau)}{\sqrt{L}} + o\left(\frac{1}{\sqrt{L}}\right)\right] \quad (32b)$$

where $o(1/\sqrt{L})$ is uniform in τ, s, and ρ.[9]

Proof: The desired result follows by applying Corollary 2 to (9). Indeed, it can be shown that the family of random variables $I_s(\rho) - i_{\ell,s}(\rho)$ (parametrized by (ρ, s)) is nonlattice [37, Lemma B.2]. Furthermore, (31) implies that the first condition (18) required for Proposition 1 and Corollary 2 is satisfied. Regarding the second condition (19), it can be observed that $V_s(\rho)$ is strictly increasing in ρ (for a fixed s) and strictly increasing in s (for a fixed ρ). Consequently, it is bounded away from zero for every $\rho \geq \underline{\rho}$ and $s \geq \underline{s}$ (for arbitrary $\underline{\rho} > 0$ and $\underline{s} > 0$). Since $\psi''_{\rho,s}(0) = V_s(\rho)$, it follows from Lemma 7 in Appendix IV that the second condition (19) required in Proposition 1 and Corollary 2 is satisfied, too. ∎

Remark 3: The set $\mathcal{S}$ with $\overline{s} = 1$ includes $0 \leq \tau < 1$. In this case, the identity (32a) characterizes all rates R between the critical rate, defined as [38, Eq. (5.6.30)]

$$R_s^{\text{cr}}(\rho) \triangleq \frac{1}{T}\left(I_s(\rho) - \psi'_{\rho,s}(1)\right) \quad (33)$$

and $I_s(\rho)$. Solving (32a) for τ, we obtain from Theorem 3 an upper bound on the minimum error probability $\epsilon^*(L, T, R, \rho)$ as a function of the rate $R \in (R_s^{\text{cr}}(\rho), I_s(\rho)]$, $s \in (0, 1]$.

Theorem 4 (Saddlepoint Expansion MC): For every rate R and $(\tau, \rho, s) \in \mathcal{S}$

$$\epsilon^*(L, T, R, \rho)$$
$$\geq e^{L[\psi_{\rho,s}(\tau) - \tau\psi'_{\rho,s}(\tau)]}\left[\Psi_{\rho,s}(\tau, \tau, n) + \frac{K_{\rho,s}(\tau, \tau, L)}{\sqrt{L}}\right.$$
$$\left. + o\left(\frac{1}{\sqrt{L}}\right)\right] - e^{L\left[J_s(\rho) - \frac{\psi'_{\rho,s}(\tau)}{s} - TR\right]} \quad (34)$$

where $o(1/\sqrt{L})$ is uniform in τ, s, and ρ.

Proof: The inequality (34) follows by applying Proposition 1 to (12) with $\log \xi = LJ_s(\rho) - L\psi'_{\rho,s}(\tau)/s$. The rest of the proof is similar to the proof of Theorem 3. ∎

VI. NORMAL APPROXIMATION AND ERROR EXPONENT

It is possible to recover the normal approximation and the error exponent of the channel from the saddlepoint expansions. Detailed derivations can be found in [37, Chs. 6.3–6.4].

Specifically, an asymptotic analysis of (32a) and (34) yields the normal approximation

$$R^*(L, T, \epsilon, \rho) = \frac{C(\rho)}{T} - \sqrt{\frac{V(\rho)}{LT^2}}Q^{-1}(\epsilon) + \mathcal{O}\left(\frac{\log L}{L}\right) \quad (35)$$

where $\mathcal{O}((\log L)/L)$ comprises terms that are of order of $(\log L)/L$ and are uniform in ρ.

Note that there are no closed-form expressions for $C(\rho)$ and $V(\rho)$, so these quantities must be evaluated numerically by computing the mean and variance of $i_\ell(\rho)$. A closed-form high-SNR normal approximation was presented in [20, Th. 2]. This expansion can be recovered from (35) by using that [20, Eqs. (38) and (39)]

$$C(\rho) = (T - 1)\log(T\rho) - \log \Gamma(T)$$
$$- (T - 1)\left[\log(1 + T\rho) + \frac{T\rho}{1 + T\rho} - \psi(T - 1)\right]$$
$$+ {}_2F_1\left(1, T - 1; T; \frac{T\rho}{1 + T\rho}\right) + o_\rho(1) \quad (36)$$

$$V(\rho) = (T - 1)^2\frac{\pi^2}{6} + (T - 1) + o_\rho(1) \quad (37)$$

where $o_\rho(1)$ comprises terms that are uniform in L and vanish as $\rho \to \infty$. In (36), $\psi(\cdot)$ denotes the digamma function and ${}_2F_1(\cdot, \cdot; \cdot; \cdot)$ denotes the Gaussian hypergeometric function.

The saddlepoint expansions (32b) and (34) can be also written as an exponential term times a subexponential factor. The exponential terms of both expansions coincide for rates $R_{1/2}^{\text{cr}}(\rho) < R < C(\rho)$, where $R_{1/2}^{\text{cr}}(\rho)$ is the critical rate (33) evaluated at $s = 1/2$. So they characterize the *reliability function*, defined as

$$E_r(T, R, \rho) \triangleq \lim_{L \to \infty} -\frac{1}{L}\log \epsilon^*(L, T, R, \rho). \quad (38)$$

Let now $\underline{\rho} \leq \rho \leq \overline{\rho}$ and $\underline{\tau} < \tau < \overline{\tau}$ for some arbitrary $0 < \underline{\rho} < \overline{\rho} < \infty$ and $0 < \underline{\tau} < \overline{\tau} < 1$, and set $s_\tau \triangleq 1/(1 + \tau)$. The coding rate R and the minimum error probability ϵ^* can be parametrized by $\tau \in (\underline{\tau}, \overline{\tau})$ as

$$R(\tau) = \frac{1}{T}\left(I_{s_\tau}(\rho) - \psi'_{\rho,s_\tau}(\tau)\right) \quad (39a)$$

$$\underline{\mathsf{A}}_\rho(\tau) \leq \epsilon^*(L, T, R, \rho)e^{-L[\psi_{\rho,s_\tau}(\tau) - \tau\psi'_{\rho,s_\tau}(\tau)]} \leq \overline{\mathsf{A}}_\rho(\tau) \quad (39b)$$

where

$$\overline{\mathsf{A}}_\rho(\tau) \triangleq \frac{1}{\sqrt{2\pi L\tau^2\psi''_{\rho,s_\tau}(\tau)}} + \frac{1}{\sqrt{2\pi L(1 - \tau)^2\psi''_{\rho,s_\tau}(\tau)}}$$
$$+ \frac{|\hat{K}_{\rho,s_\tau}(\tau)|}{\sqrt{L}} + o\left(\frac{1}{\sqrt{L}}\right) \quad (40)$$

$$\underline{\mathsf{A}}_\rho(\tau) \triangleq \frac{s_\tau^{\frac{1}{s_\tau}}}{\tau\left(2\pi L\psi''_{\rho,s_\tau}(\tau)\right)^{\frac{1}{2s_\tau}}} + o\left(\frac{1}{L^{\frac{1}{2s_\tau}}}\right). \quad (41)$$

---

[9]The error terms appearing in the asymptotic expansions depend, in general, also on T. However, we do not make this dependence explicit in the notation, since we view T as a fixed parameter.

The little-o term in (40) is uniform in ρ and τ. The little-o term in (41) is uniform in ρ (for every given τ).

The products of $\overline{A}_\rho(\tau)$ and $\underline{A}_\rho(\tau)$ with $e^{L[\psi_{\rho,s_\tau}(\tau)-\tau\psi'_{\rho,s_\tau}(\tau)]}$ yield approximations of the RCU$_s$ and MC bounds, respectively. Note that the first term of $\underline{A}_\rho(\tau)$ is positive and of order $L^{-\frac{1+\tau}{2}}$. This order coincides with that of the subexponential factor of the random-coding upper bound on the error probability for some memoryless channels and rates above the critical rate [39], [40].

It follows from (39a), (39b), and the behaviors of $\overline{A}_\rho(\tau)$ and $\underline{A}_\rho(\tau)$ that the reliability function $E_r(T,R,\rho)$ can be parametrized by $\tau \in (0,1)$ as

$$E_r(\tau) = \tau\psi'_{\rho,\frac{1}{1+\tau}}(\tau) - \psi_{\rho,\frac{1}{1+\tau}}(\tau) \tag{42a}$$

$$R(\tau) = \frac{1}{T}\left(I_{\frac{1}{1+\tau}}(\rho) - \psi'_{\rho,\frac{1}{1+\tau}}(\tau)\right). \tag{42b}$$

VII. NUMERICAL EXAMPLES

To show the accuracy of the saddlepoint approximations, we evaluate them for different channel parameters and compare them with the nonasymptotic bounds as well as the refined asymptotic expansions available in the literature.

In all figures, we plot approximations of the RCU$_s$ bound in red and approximations of the MC bound in blue, which are obtained by disregarding the $o(1/\sqrt{L})$ terms in the saddlepoint expansions and by optimizing numerically over the parameters s and τ. Solid lines ("saddlepoint") depict the saddlepoint approximations (32b) and (34), and dashed lines ("pref$\times$EE") depict (39b). We further plot the nonasymptotic bounds (9) and (12), evaluated using Monte-Carlo simulations, with dots and highlight in grey the area where the true maximum coding rate lies. Finally, we plot the normal approximation (35) ("NA") and the error-exponent approximations ("EE" and "$L^{-\frac{1+\tau}{2}}\times$EE") that follow by solving

$$\epsilon^*(L,T,R,\rho) = e^{-LE_r(T,R,\rho)} \tag{43a}$$

$$\epsilon^*(L,T,R,\rho) = \frac{1}{L^{\frac{1+\tau}{2}}}e^{-LE_r(T,R,\rho)} \tag{43b}$$

for R, where τ in (43b) is the parameter for which (42b) is equal to R.

In Figs. 2a and 2b, we study $R^*(L,T,\epsilon,\rho)$ as a function of L for $n = LT = 168$ (hence T is inversely proportional to L), $\epsilon = 10^{-5}$, and SNR values 0 dB and 6 dB, respectively. Observe that the approximations (32b), (34), and (39b) are almost indistinguishable from the nonasymptotic bounds. Further observe that, compared to the saddlepoint expansions, the normal approximation and the error-exponent approximations are loose, although their accuracy increases for larger SNR values (and larger values of L, for the normal approximation). Finally, the error-exponent approximation "$L^{-\frac{1+\tau}{2}}\times$EE" is accurate for the entire range of parameters.

In Figs. 3a and 3b, we study $R^*(L,T,\epsilon,\rho)$ as a function of ϵ for $n = 168$ ($T = 12$ and $L = 14$) and SNR values 0 dB and 6 dB, respectively. In addition to the aforementioned bounds and approximations, we plot $R^{\text{cr}}_{1/2}(0)$ to indicate the minimum rate that can be characterized by (32a). Observe that the approximations (32b), (34), and (39b) are almost indistinguishable from the nonasymptotic bounds. Further observe how
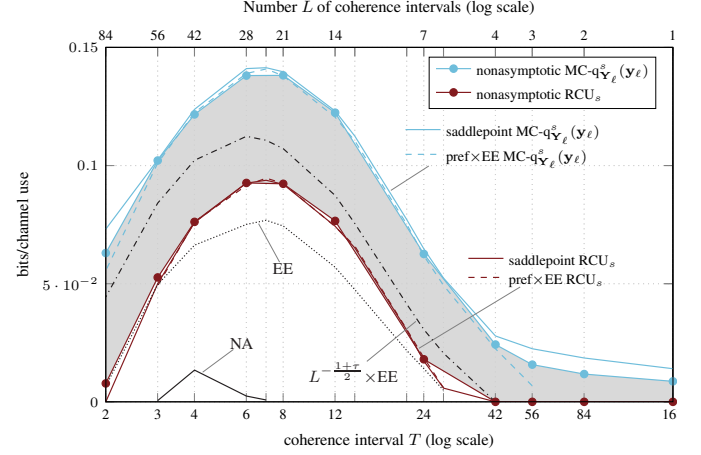
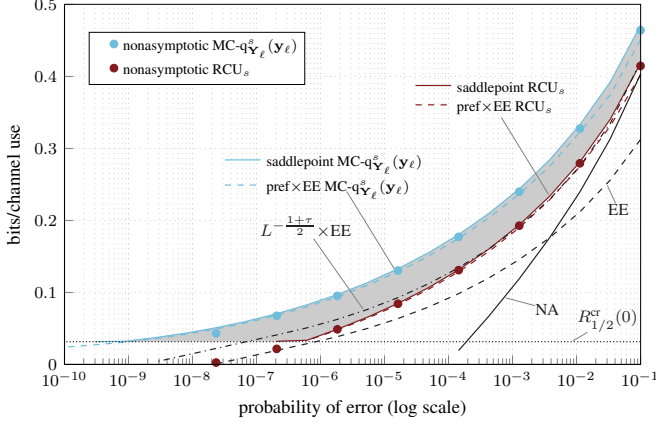
(a) $n = 168$, $\epsilon = 10^{-5}$, and $\rho = 0$ dB.
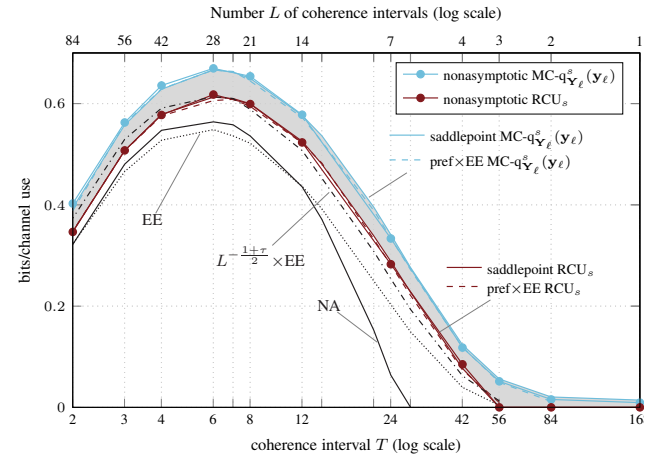
(b) $n = 168$, $\epsilon = 10^{-5}$, and $\rho = 6$ dB.

Fig. 2. $R^*(L,T,\epsilon,\rho)$ as a function of L for a fixed $n = LT$.

the normal approximation becomes accurate for large error probabilities, whereas the error-exponent approximation "EE" becomes accurate for small error probabilities. In contrast, the error-exponent approximation "$L^{-\frac{1+\tau}{2}}\times$EE" is accurate for the entire range of parameters.

In Figs. 4a and 4b, we study $\epsilon^*(L,T,R,\rho)$ for a fixed rate R as a function of the SNR ρ. Specifically, in Fig. 4a we show $\epsilon^*(L,T,R,\rho)$ for $n = 168$ ($T = 24$ and $L = 7$) and $R = 0.48$, and in Fig. 4b we show $\epsilon^*(L,T,R,\rho)$ for $n = 500$ ($T = 20$ and $L = 25$) and $R = 4$. In both figures, we also show the high-SNR normal approximation ("high-SNR-NA") presented in [20, Th. 2]. In addition to the aforementioned bounds and approximations, in Fig. 4a we show the simulated performance of a transmission scheme for which 4 channel uses per coherence interval are allocated to coded pilot symbols belonging to a quaternary phase shift keying (QPSK) constellation, and the remaining channel uses are allocated to QPSK symbols encoded by a binary quasi-cyclic code. Decoding is performed via ordered statistics decoding (OSD), which uses scaled nearest-neighbor decoding (SNN) as the metric to solve the likelihood ratios. This coding scheme is denoted as "OSD-SNN" and is depicted in orange. We also show the simulated performance of a transmission

(a) $L = 14$, $T = 12$, and $\rho = 0$ dB.
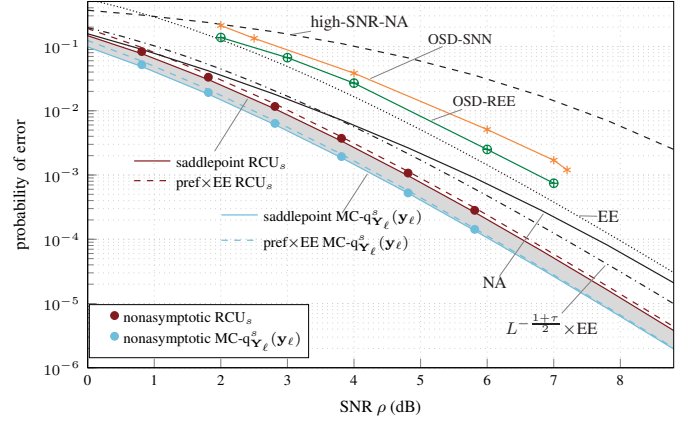

(a) $L = 7$, $T = 24$, and $R = 0.48$.

(b) $L = 14$, $T = 12$, and $\rho = 6$ dB.

(b) $L = 25$, $T = 20$, and $R = 4$.

Fig. 3. $R^*(L, T, \epsilon, \rho)$ as a function of ϵ.

Fig. 4. $\epsilon^*(L, T, R, \rho)$ as a function of ρ.
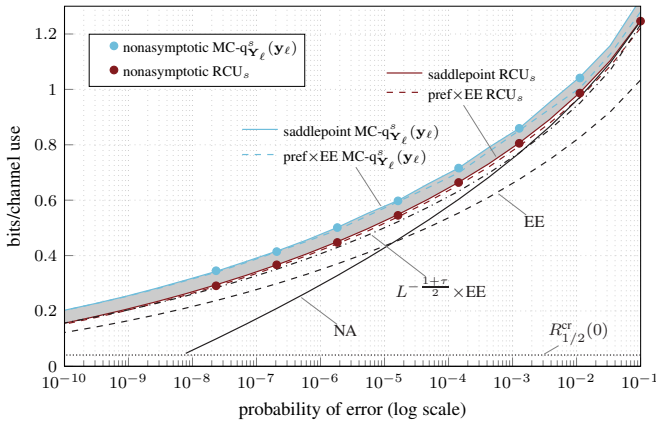
scheme based on OSD-SNN that additionally performs a re-estimation of the fading coefficients based on the initial OSD decision. This coding scheme is denoted as "OSD-REE" and is depicted in green. The derivation and design of these codes can be found in [25]. Observe that, for this scenario, the error-exponent approximation "$L^{-\frac{1+\tau}{2}} \times$EE" is not as accurate as the saddlepoint approximations. In Fig. 4b, we additionally show the performance of an accumulate-repeat-jagged-accumulate (ARJA) low-density parity-check (LDPC) code combined with 64-APSK modulation, the transmission of 2 pilot symbols per coherence block, and maximum-likelihood channel estimation followed by mismatched nearest-neighbor decoding at the receiver ("ARJA LDPC code 64-APSK"); for details see [6, Sec. 4]. Observe that the gap between the performance of the presented transmission schemes and the rest of the curves is substantial in both figures. This suggests that more sophisticated (possibly joint) channel-estimation and decoding procedures together with shaping techniques need to be adopted to close the gap.

VIII. DISCUSSION

In the following, we discuss the complexity of the numerical evaluation of the proposed saddlepoint approximations and compare it with that of the nonasymptotic bounds as well as
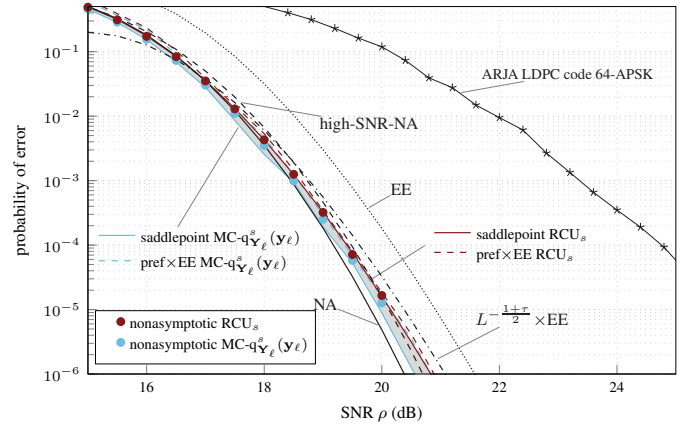
of the refined asymptotic expansions available in the literature. We then provide additional remarks on when one should use which approximation.

A. Complexity Analysis

We denote by N the cost of numerically evaluating a one-dimensional integral and by K the cost of performing an optimization over an auxiliary parameter. In the following, we provide a coarse estimate of the complexity of the presented bounds and approximations in terms of N and K. The nonasymptotic bounds (9) and (12) require the evaluation of the distribution function of $\sum_{\ell=1}^{L} i_{\ell,s}(\rho)$. By (8), the generalized information density $i_{\ell,s}(\rho)$ can be expressed in terms of the two random variables $\Upsilon_{1,\ell}$ and $\Upsilon_{2,\ell}$. It follows that the evaluation of the RCU$_s$ bound requires the computation of a $(2L + 1)$-dimensional integral (over $\{\Upsilon_{1,\ell}\}_{\ell=1}^{L}$, $\{\Upsilon_{2,\ell}\}_{\ell=1}^{L}$, and U) and an optimization over s. Similarly, the MC bound requires the computation of a $(2L)$-dimensional integral and an optimization over s and ξ. The numerical evaluation of an L-dimensional integral has a complexity of roughly N^L. Hence, the overall complexity of the RCU$_s$ bound is KN^{2L+1}, and the complexity of the MC bound is K^2N^{2L}. The saddlepoint approximations (32b) and (34), and the prefactor-and-error-exponent approximations (39b), depend on $I_s(\rho)$, $\psi_{\rho,s}$,

TABLE I
COMPLEXITY AND COMPUTATIONAL TIME OF THE PRESENTED BOUNDS AND APPROXIMATIONS FOR $L = 30, T = 10, \epsilon = 10^{-6}, \rho = 6$ dB.

Bounds / Approximations	Comp. Complexity	Comp. Time
nonasymptotic bounds	RCU_s: KN^{2L+1} (optimization over s)	RCU_s: 3720 s
	MC: $K^2 N^{2L}$ (optimization over s, ξ)	MC: 3280 s
saddlepoint approximations	$5K^2N^2$ (optimization over τ, s)	RCU_s: 162 s
		MC: 194 s
prefactor-and-error-exponent approximations	$5KN^2$ (optimization over τ)	RCU_s: 170 s
		MC: 207 s
error-exponent approximations	$3KN^2$ (optimization over τ)	140 s
normal approximation	$2N^2$ (no optimization)	0.7 s
high-SNR normal approximation	available in closed form	5 s

$\psi'_{\rho,s}$, $\psi''_{\rho,s}$, and $\psi'''_{\rho,s}$, so they can be obtained by solving 5 two-dimensional integrals and by optimizing over (τ, s) (saddlepoint approximations) or over τ (prefactor-and-error-exponent approximations). The error-exponent approximations that follow by solving (43a) and (43b) for R can be obtained by evaluating $I_{s_\tau}(\rho)$, ψ_{ρ,s_τ}, and ψ'_{ρ,s_τ}, which corresponds to the evaluation of 3 two-dimensional integrals, and by optimizing over τ. The normal approximation (35) can be obtained by evaluating $C(\rho)$ and $V(\rho)$, which corresponds to the evaluation of 2 two-dimensional integrals. The high-SNR normal approximation [20, Th. 2] is available in closed form.

In Table I, we summarize the computational complexities of the presented bounds and approximations. We further show the computational time (in seconds) required to numerically compute bounds and approximations on $R^*(L, T, \epsilon, \rho)$ for the parameters $L = 30$, $T = 10$, $\epsilon = 10^{-6}$, and $\rho = 6$ dB on a PC-cluster node with 96 GB of RAM, powered by an Intel Xeon Gold 6130 processor. Observe that the computational complexity of the nonasymptotic bounds grows exponentially in L, whereas the approximations proposed in this paper have a computational complexity that is independent of L. This is a significant reduction in computation cost, especially if L is large. As a consequence, the saddlepoint approximations have a computational time that is about a factor of 20 (RCU_s bound) or 17 (MC bound) smaller than that of the nonasymptotic bounds. The prefactor-and-error-exponent approximations have a similar complexity and computational time as the saddlepoint approximations, while the complexity and computational time of the error-exponent approximation is slightly smaller. The normal approximation has, by far, the lowest computational time, since no optimization over auxiliary parameters is required. Perhaps surprisingly, the high-SNR normal approximation requires more computational time than the normal approximation, despite the fact that the former is available in closed form and the latter must be evaluated numerically. The reason is that the high-SNR approximation (36) of capacity depends on a Gaussian hypergeometric function, whose evaluation is costly. Last but not least, it is worth pointing out that setting $s = 1/(1 + \tau)$ in the saddlepoint approximations and optimizing over τ, which reduces the computational complexity to $5KN^2$, yields accurate results.

B. Saddlepoint Approximations vs. Normal and Error-Exponent Approximations

Intuitively, the error-exponent approximation (43a) is accurate for small values of ρ and ϵ, whereas the normal approximation (35) is accurate for large values of ρ and ϵ. For example, we observe from Fig. 3a that, for $\rho = 0$ dB, the normal approximation is accurate for $\epsilon > 10^{-2}$. In contrast, for $\rho = 0$ dB, the error-exponent approximation "EE" is accurate for error probabilities below 10^{-4}. Similarly, we observe from Fig. 3b that, for $\rho = 6$ dB, the normal approximation is accurate for error probabilities above 10^{-3}, while the error-exponent approximation is accurate for $\epsilon < 10^{-7}$. As noted in [20], the high-SNR normal approximation is accurate for $\rho \geq 15$ dB, $L \geq 10$, and large values of ϵ. In contrast, the saddlepoint approximations (32b) and (34), and the approximations (39b), are accurate over the entire range of system parameters. Finally, the error-exponent approximation "$L^{-\frac{1+\tau}{2}} \times \text{EE}$" is accurate over the entire range of system parameters, albeit not as accurate as the saddlepoint approximations. That said, the normal approximations require a computational time that is two orders of magnitude lower than that of the remaining approximations. Furthermore, the high-SNR normal approximation is available in closed form, which makes it suitable for analytical studies.

In summary, the normal approximation is the go-to choice when the SNR and error probabilities are sufficiently large. For example, for an SNR of 6 dB, this is the case for error probabilities above 10^{-3}. Indeed, in this regime the normal approximation is reasonably accurate, and it can either be computed efficiently or it can be approximated by the high-SNR normal approximation, which is available in closed form. In contrast, the saddlepoint approximations arise as easy-to-compute alternatives to the nonasymptotic bounds when one wishes to characterize the maximum coding rate for a large range of system parameters or for small SNR values and error probabilities, where the normal approximations are inaccurate.

IX. SUMMARY AND CONCLUSIONS

In this paper, we applied the saddlepoint method to derive approximations of the MC and the RCU_s bounds for single-antenna Rayleigh block-fading channels. While these approximations must still be evaluated numerically, they only require

the evaluation of two-dimensional integrals. This is in contrast to the nonasymptotic MC and RCU$_s$ bounds, which require the evaluation of $(2L)$-dimensional integrals. Numerical evidence shows that the saddlepoint approximations are accurate for the entire range of system parameters for which the nonasymptotic bounds are computable. They thus arise as easy-to-compute alternatives to the nonasymptotic bounds.

APPENDIX I
PROOFS OF PROPOSITION 1 AND COROLLARY 2

A. Proof of Proposition 1, Part 1

The proof follows closely the steps by Feller [26, Ch. XVI]. Since we consider a random variable $Z_{\ell,\theta}$ that depends on a parameter θ, we provide here a self-contained proof proving uniformity in the extra parameter. Let F_θ denote the distribution of $Y_{\ell,\theta} \triangleq Z_{\ell,\theta} - \tilde\gamma$, where $\tilde\gamma \triangleq \gamma/n$. Then, the CGF of Y_k is given by $\tilde\psi_\theta(\zeta) \triangleq \psi_\theta(\zeta) - \zeta\tilde\gamma$. Consider the tilted random variable $V_{\ell,\theta}$ with distribution

$$\vartheta_{\theta,\tau}(x) = e^{-\tilde\psi_\theta(\tau)} \int_{-\infty}^{x} e^{\tau t} dF_\theta(t)$$
$$= e^{-\psi_\theta(\tau)+\tau\tilde\gamma} \int_{-\infty}^{x} e^{\tau t} dF_\theta(t) \qquad (44)$$

where the parameter τ lies in $(-\zeta_0, \zeta_0)$. Note that the exponential term $e^{-\psi_\theta(\tau)+\tau\tilde\gamma}$ on the right-hand side (RHS) of (44) is a normalizing factor that guarantees that $\vartheta_{\theta,\tau}$ is a distribution.

Let $v_{\theta,\tau}(\zeta)$ denote the MGF of the tilted random variable $V_{\ell,\theta}$, which is given by

$$v_{\theta,\tau}(\zeta) = \int_{-\infty}^{\infty} e^{\zeta x} d\vartheta_{\theta,\tau}(x)$$
$$= \frac{m_\theta(\zeta+\tau)}{m_\theta(\tau)} e^{-\zeta\tilde\gamma} \qquad (45)$$

where $m_\theta(\tau)$ is given in (13). Together with $\mathsf{E}[V_{\ell,\theta}] = v'_{\theta,\tau}(0)$, this yields

$$\mathsf{E}[V_{\ell,\theta}] = \psi'_\theta(\tau) - \tilde\gamma. \qquad (46)$$

Let now $F_\theta^{\star n}$ denote the distribution of $\sum_{\ell=1}^{n}(Z_{\ell,\theta} - \tilde\gamma)$, and let $\vartheta_{\theta,\tau}^{\star n}$ denote the distribution of $\sum_{\ell=1}^{n} V_{\ell,\theta}$. By (44) and (45), the distributions $F_\theta^{\star n}$ and $\vartheta_{\theta,\tau}^{\star n}$ again stand in the relationship (44) except that the term $e^{-\psi_\theta(\tau)}$ is replaced by $e^{-n\psi_\theta(\tau)}$ and $\tilde\gamma$ is replaced by $n\tilde\gamma$. By inverting (44), we obtain that

$$\mathsf{P}\left[\sum_{\ell=1}^{n} Z_{\ell,\theta} \geq \gamma\right] = e^{n\psi_\theta(\tau)-\tau\gamma} \int_{0}^{\infty} e^{-\tau y} d\vartheta_{\theta,\tau}^{\star n}(y). \qquad (47)$$

Furthermore, by choosing τ such that $n\psi'_\theta(\tau) = \gamma$, it follows from (46) that the distribution $\vartheta_{\theta,\tau}^{\star n}$ has zero mean. We next substitute in (47) the distribution $\vartheta_{\theta,\tau}^{\star n}$ by the zero-mean normal distribution with variance $n\psi''_\theta(\tau)$, denoted by $\mathfrak{N}_{n\psi''_\theta(\tau)}$, and analyze the error incurred by this substitution. To this end, we define

$$A_\tau \triangleq e^{n\psi_\theta(\tau)-\tau\gamma} \int_{0}^{\infty} e^{-\tau y} d\mathfrak{N}_{n\psi''_\theta(\tau)}(y) \qquad (48)$$

which for $n\psi'_\theta(\tau) = \gamma$ can be evaluated as

$$A_\tau = \frac{e^{n[\psi_\theta(\tau)-\tau\psi'_\theta(\tau)]}}{\sqrt{2\pi n\psi''_\theta(\tau)}} \int_{0}^{\infty} e^{-\tau y} e^{-\frac{y^2}{2n\psi''_\theta(\tau)}} dy$$
$$= e^{n\left[\psi_\theta(\tau)-\tau\psi'_\theta(\tau)+\frac{\tau^2}{2}\psi''_\theta(\tau)\right]} Q\left(\tau\sqrt{n\psi''_\theta(\tau)}\right). \qquad (49)$$

We continue by showing that the error incurred by substituting $\mathfrak{N}_{n\psi''_\theta(\tau)}$ for $\vartheta_{\theta,\tau}^{\star n}$ in (47) is small. To do so, we note that integration by parts [26, Ch. V.6, Eq. (6.1)] yields

$$\mathsf{P}\left[\sum_{\ell=1}^{n} Z_{\ell,\theta} \geq n\psi'_\theta(\tau)\right] - A_\tau$$
$$= e^{n[\psi_\theta(\tau)-\tau\psi'_\theta(\tau)]}\left[-\left(\vartheta_{\theta,\tau}^{\star n}(0) - \mathfrak{N}_{n\psi''_\theta(\tau)}(0)\right)\right.$$
$$\left. + \tau \int_{0}^{\infty} \left(\vartheta_{\theta,\tau}^{\star n}(y) - \mathfrak{N}_{n\psi''_\theta(\tau)}(y)\right) e^{-\tau y} dy\right]. \qquad (50)$$

We next use [26, Sec. XVI.4, Th. 1] (stated as Lemma 5 below) to assess the error committed in (50). To state Lemma 5, we first introduce the following notation. Let $\tilde Z_{1,\theta}, \ldots, \tilde Z_{n,\theta}$ be a sequence of i.i.d., real-valued, zero-mean, random variables with distribution $\tilde F_\theta$ that depends on an extra parameter $\theta \in \Theta$. We denote the k-th moment for any possible value of $\theta \in \Theta$ by $\mu_{k,\theta}$ and we denote the second moment as $\mu_{2,\theta} = \sigma_\theta^2$.

For the distribution of the normalized n-fold convolution of a sequence of i.i.d., zero-mean, unit-variance random variables, we write

$$\tilde F_{n,\theta}(x) = \tilde F_\theta^{\star n}(x\sigma_\theta\sqrt{n}). \qquad (51)$$

Note that $\tilde F_{n,\theta}$ has zero mean and unit variance. We denote by $\mathfrak{N}$ the zero-mean, unit-variance, normal distribution, and we denote by $\mathfrak{n}$ the zero-mean, unit-variance, normal pdf.

Lemma 5: Assume that the family of distributions $\tilde F_{n,\theta}$ (parametrized by θ) is nonlattice. Further assume that

$$\sup_{\theta\in\Theta} \mu_{4,\theta} < \infty \qquad (52)$$
$$\inf_{\theta\in\Theta} \sigma_\theta > 0. \qquad (53)$$

Then, for every $\theta \in \Theta$ and $x \in \mathbb{R}$,

$$\tilde F_{n,\theta}(x) - \mathfrak{N}(x) = \frac{\mu_{3,\theta}}{6\sigma_\theta^3\sqrt{n}}(1-x^2)\mathfrak{n}(x) + o\left(\frac{1}{\sqrt{n}}\right) \qquad (54)$$

where $o(1/\sqrt{n})$ is uniform in x and θ.

Proof: See Appendix II. ∎

We next use Lemma 5 to expand (50). Note that the family of distributions $F_\theta^{\star n}$ (parametrized by θ) is nonlattice by assumption. Furthermore, as shown in Appendix III, if a family of distributions is nonlattice, then so is the corresponding family of tilted distributions. Consequently, the family of distributions $\vartheta_{\theta,\tau}^{\star n}$ (parametrized by θ) is nonlattice, too. The variable y in (50) corresponds to $x\sigma_\theta\sqrt{n}$ in (51). Hence,

applying (54) to (50) with $\vartheta_{\theta,\tau}^{\star n}(y) = \tilde{F}_{n,\theta}(y/\sqrt{n\psi_\theta''(\tau)})$ and $\mathfrak{N}_{n\psi_\theta''(\tau)}(y) = \mathfrak{N}(y/\sqrt{n\psi_\theta''(\tau)})$, we obtain

$$
\begin{aligned}
&\mathsf{P}\left[\sum_{\ell=1}^n Z_{\ell,\theta} \geq n\psi_\theta'(\tau)\right] - A_\tau \\
&= e^{n[\psi_\theta(\tau)-\tau\psi_\theta'(\tau)]}\left[\frac{\psi_\theta'''(\tau)}{6\psi_\theta''(\tau)^{3/2}\sqrt{n}}\left(-\frac{1}{\sqrt{2\pi}}+\frac{\tau^2 n\psi_\theta''(\tau)}{\sqrt{2\pi}}\right.\right. \\
&\qquad\left.\left. -\tau^3\psi_\theta''(\tau)^{3/2}n^{3/2}\Psi_\theta(\tau,\tau,n)\right) + o\left(\frac{1}{\sqrt{n}}\right)\right] \quad (55)
\end{aligned}
$$

where we used that

$$
\mathsf{Var}[V_{\ell,\theta}] = \psi_\theta''(\tau) \tag{56}
$$

$$
\mathsf{E}\left[(V_{\ell,\theta}-\mathsf{E}[V_{\ell,\theta}])^3\right] = \psi_\theta'''(\tau). \tag{57}
$$

Substituting A_τ in (49) into (55), and using that $n\psi_\theta'(\tau) = \gamma$, we establish Part 1 of Proposition 1.

B. Proof of Proposition 1, Part 2

The proof of Part 2 follows along similar lines as the proof of Part 1. Hence, we will focus on describing what is different. Specifically, the left-hand-side (LHS) of (24) differs from the LHS of (20) by the additional term $\log U$. To account for this difference, we follow the same steps as Scarlett *et al.* [28, App. E]. Since in our setting the distribution of $Z_{\ell,\theta}$ depends on the parameter θ, we repeat these steps in the following:

$$
\begin{aligned}
&\mathsf{P}\left[\sum_{\ell=1}^n Z_{\ell,\theta} \geq \gamma + \log U\right] \\
&= e^{n\psi_\theta(\tau)-\tau\gamma}\int_0^1\int_{\log u}^\infty e^{-\tau y}d\vartheta_{\theta,\tau}^{\star n}(y)du \\
&= e^{n\psi_\theta(\tau)-\tau\gamma}\left(\int_0^\infty e^{-\tau y}d\vartheta_{\theta,\tau}^{\star n}(y) + \int_{-\infty}^0 e^{(1-\tau)y}d\vartheta_{\theta,\tau}^{\star n}(y)\right)
\end{aligned}
$$
$$ \tag{58} $$

where the second step follows by changing the order of integration. We next proceed as in the proof of the previous part. The first term in (58) coincides with (47). For the second term, we substitute the distribution $\vartheta_{\theta,\tau}^{\star n}$ by the zero-mean normal distribution with variance $n\psi_\theta''(\tau)$ and analyze the error incurred by this substitution. To this end, we define

$$
\tilde{A}_\tau \triangleq e^{n\psi_\theta(\tau)-\tau\gamma}\int_{-\infty}^0 e^{(1-\tau)y}d\mathfrak{N}_{n\psi_\theta''(\tau)}(y) \tag{59}
$$

which for $n\psi_\theta'(\tau) = \gamma$ can be computed as

$$
\tilde{A}_\tau = e^{n\left[\psi_\theta(\tau)-\tau\psi_\theta'(\tau)+\frac{(1-\tau)^2}{2}\psi_\theta''(\tau)\right]}Q\left((1-\tau)\sqrt{n\psi_\theta''(\tau)}\right). \tag{60}
$$

As we did in (50), we next evaluate the error incurred by substituting $\vartheta_{\theta,\tau}^{\star n}$ by $\mathfrak{N}_{n\psi_\theta''(\tau)}$. Indeed,

$$
\begin{aligned}
&e^{n\psi_\theta(\tau)-\tau\gamma}\int_{-\infty}^0 e^{(1-\tau)y}d\vartheta_{\theta,\tau}^{\star n}(y) - \tilde{A}_\tau \\
&= e^{n[\psi_\theta(\tau)-\tau\psi_\theta'(\tau)]}\left[\frac{\psi_\theta'''(\tau)}{6\psi_\theta''(\tau)^{3/2}\sqrt{n}}\left(\frac{1}{\sqrt{2\pi}}-\frac{(1-\tau)^2 n\psi_\theta''(\tau)}{\sqrt{2\pi}}\right.\right. \\
&\qquad\left.\left. +(1-\tau)^3(n\psi_\theta''(\tau))^{3/2}\Psi_\theta(1-\tau,\tau,n)\right)+o\left(\frac{1}{\sqrt{n}}\right)\right] \quad (61)
\end{aligned}
$$

which follows by integration by parts [26, Ch. V.6, Eq. (6.1)] and by Lemma 5.

Combining (58) with (20), (47), (60), and (61), we obtain the desired result (24).

C. Proof of Corollary 2

Using (58) with (48) and (59), and using a change of variable, we obtain

$$
\begin{aligned}
&\mathsf{P}\left[\sum_{\ell=1}^n Z_{\ell,\theta} \geq \gamma + \log U\right] - A_\tau - \tilde{A}_\tau \\
&= e^{n[\psi_\theta(\tau)-\tau\psi_\theta'(\tau)]}\left[o\left(\frac{1}{\sqrt{n}}\right)+\frac{1}{\sqrt{2\pi}}\frac{\psi_\theta'''(\tau)}{6\psi_\theta''(\tau)^{3/2}\sqrt{n}}\right. \\
&\quad\times\left(\int_0^\infty(1-\tau)\sqrt{n\psi_\theta''(\tau)}(z^2-1)e^{-(1-\tau)\sqrt{n\psi_\theta''(\tau)}z-\frac{z^2}{2}}dz\right. \\
&\quad\left.\left.+\int_0^\infty\tau\sqrt{n\psi_\theta''(\tau)}(1-z^2)e^{-\tau\sqrt{n\psi_\theta''(\tau)}z-\frac{z^2}{2}}dz\right)\right]. \quad (62)
\end{aligned}
$$

Keeping only the positive part of each integral, it follows that the RHS of (62) can be upper-bounded by

$$
\begin{aligned}
&e^{n[\psi_\theta(\tau)-\tau\psi_\theta'(\tau)]}\left[o\left(\frac{1}{\sqrt{n}}\right)+\frac{1}{\sqrt{2\pi}}\frac{\psi_\theta'''(\tau)}{6\psi_\theta''(\tau)^{3/2}\sqrt{n}}\right. \\
&\quad\times\left(\int_1^\infty(1-\tau)\sqrt{n\psi_\theta''(\tau)}(z^2-1)e^{-(1-\tau)\sqrt{n\psi_\theta''(\tau)}z-\frac{z^2}{2}}dz\right. \\
&\quad\left.\left.+\int_0^1\tau\sqrt{n\psi_\theta''(\tau)}(1-z^2)e^{-\tau\sqrt{n\psi_\theta''(\tau)}z-\frac{z^2}{2}}dz\right)\right]. \quad (63)
\end{aligned}
$$

The first integral in (63) is upper-bounded by 1. The second integral is upper-bounded by

$$
e^{-(1-\tau)\sqrt{n\psi_\theta''(\tau)}}\left(1+\frac{2+2(1-\tau)\sqrt{n\psi_\theta''(\tau)}}{(1-\tau)^2 n\psi_\theta''(\tau)}\right). \tag{64}
$$

If $\tau \in [0, \min\{\zeta_0, 1-\delta\})$ for some arbitrary $\delta > 0$ independent of n and θ, then (64) vanishes faster than $1/\sqrt{n}$ uniformly in θ. We thus obtain the upper bound

$$
\begin{aligned}
&\mathsf{P}\left[\sum_{\ell=1}^n Z_\ell \geq n\psi_\theta'(\tau) + \log U\right] - A_\tau - \tilde{A}_\tau \\
&\leq e^{n[\psi_\theta(\tau)-\tau\psi_\theta'(\tau)]}\left[\frac{1}{\sqrt{2\pi}}\frac{\psi_\theta'''(\tau)}{6\psi_\theta''(\tau)^{3/2}\sqrt{n}}+o\left(\frac{1}{\sqrt{n}}\right)\right] \quad (65)
\end{aligned}
$$

thereby proving Corollary 2.

APPENDIX II
PROOF OF LEMMA 5

The proof follows along similar lines as the proof of [26, Ch. XVI.4, Th. 1]. Notice that our result holds uniformly in the parameter θ of the distribution $\tilde{F}_\theta$, which makes the conditions of our lemma slightly more restrictive. Specifically, we require the fourth moment of $\tilde{F}_\theta$ to exist, whereas in the original theorem only the third moment is required to exist.

Let us denote the characteristic function of $\tilde{Z}_{\ell,\theta} \sim \tilde{F}_\theta$ by

$$
\tilde{\varphi}_\theta(\zeta) \triangleq \mathsf{E}\left[e^{\iota\zeta\tilde{Z}_{\ell,\theta}}\right], \quad \zeta \in \mathbb{R} \tag{66}
$$

and define

$$
G_\theta(x) \triangleq \mathfrak{N}(x) - \frac{\mu_{3,\theta}}{6\sigma_\theta^3\sqrt{n}}(x^2-1)\mathfrak{n}(x), \quad x \in \mathbb{R}. \tag{67}
$$

Note that (52) implies that $\sup_{\theta\in\Theta}|\mu_{3,\theta}| < \infty$ since, by Jensen's inequality, $|\mu_{3,\theta}| \leq \mu_{4,\theta}^{3/4}$. Using this together with (53), one can show that the first derivative of G_θ is bounded in $\theta \in \Theta$. Furthermore, the characteristic function of G_θ is

$$\gamma_\theta(\zeta) = e^{-\frac{1}{2}\zeta^2}\left[1 + \frac{\mu_{3,\theta}}{6\sigma_\theta^3\sqrt{n}}(\imath\zeta)^3\right]. \tag{68}$$

It follows that G_θ satisfies the conditions of [26, Ch. XVI.3, Lemma 2], namely, that $\sup_{\theta\in\Theta, x\in\mathbb{R}}|G_\theta'(x)| \leq m$ for some positive constant m and that G_θ has a continuously-differentiable characteristic function satisfying $\gamma_\theta(0) = 1$ and $\gamma_\theta'(0) = 0$. Then, the following inequality holds for all $x \in \mathbb{R}$ and $\Delta > 0$ [26, Ch. XVI.3, Eq. (3.13)]:

$$|\tilde{F}_{n,\theta}(x) - G(x)|$$
$$\leq \frac{1}{\pi}\int_{-\Delta}^{\Delta}\left|\frac{\tilde{\varphi}_\theta^n\left(\frac{\zeta}{\sigma_\theta\sqrt{n}}\right) - \gamma_\theta(\zeta)}{\zeta}\right|d\zeta + \frac{24m}{\pi\Delta}. \tag{69}$$

Using (69) with $\Delta = a\sqrt{n}$, where the constant a is chosen sufficiently large such that $\frac{24m}{\pi} < \varepsilon a$ for some arbitrary $\varepsilon > 0$ independent of x and θ, we can write

$$|\tilde{F}_{n,\theta}(x) - G(x)|$$
$$\leq \frac{1}{\pi}\int_{-a\sqrt{n}}^{a\sqrt{n}}\left|\frac{\tilde{\varphi}_\theta^n\left(\frac{\zeta}{\sigma_\theta\sqrt{n}}\right) - \gamma_\theta(\zeta)}{\zeta}\right|d\zeta + \frac{\varepsilon}{\sqrt{n}}. \tag{70}$$

It remains to show that the RHS of (70) decays faster than $1/\sqrt{n}$ uniformly in θ. To this end, we first note that, by assumption, the family of distributions $\tilde{F}_\theta$ (parametrized by θ) is nonlattice, so $\sup_{\theta\in\Theta}|\tilde{\varphi}_\theta(\zeta)|$ is strictly smaller than 1 for every $\zeta \neq 0$. Furthermore, as we shall argue below, (52) implies that the function $\zeta \mapsto \sup_{\theta\in\Theta}\tilde{\varphi}_\theta(\zeta)$ is continuous. Consequently, there exists a number $q_{\delta,\bar{\zeta}} < 1$ (independent of θ) such that

$$\sup_{\theta\in\Theta}|\tilde{\varphi}_\theta(\zeta)| \leq q_{\delta,\bar{\zeta}}, \quad \delta \leq |\zeta| \leq \bar{\zeta} \tag{71}$$

for some arbitrary $\delta > 0$ and $\bar{\zeta} \geq a/(\inf_{\theta\in\Theta}\sigma_\theta)$.

To prove that $\zeta \mapsto \sup_{\theta\in\Theta}\tilde{\varphi}_\theta(\zeta)$ is continuous, note that, by [26, Ch. XV.4, Lemma 2],

$$\sup_{\theta\in\Theta}|\tilde{\varphi}_\theta'(\zeta)| \leq \sup_{\theta\in\Theta}\mathsf{E}\left[|\tilde{Z}_{\ell,\theta}|\right], \quad \zeta \in \mathbb{R}. \tag{72}$$

Moreover, for every $\zeta_1, \zeta_2 \in \mathbb{R}$,

$$\left|\sup_{\theta\in\Theta}\tilde{\varphi}_\theta(\zeta_1) - \sup_{\theta\in\Theta}\tilde{\varphi}_\theta(\zeta_2)\right| \leq \sup_{\theta\in\Theta}|\tilde{\varphi}_\theta(\zeta_1) - \tilde{\varphi}_\theta(\zeta_2)|$$
$$\leq \sup_{\theta\in\Theta}\mathsf{E}\left[|\tilde{Z}_{\ell,\theta}|\right]|\zeta_1 - \zeta_2| \tag{73}$$

where the second inequality follows from the mean value theorem [41, Th. 5.10] and (72). Since the RHS of (72) is finite by (52), it follows that for every $\epsilon > 0$ there exists a $\delta > 0$ such that $|\zeta_1 - \zeta_2| \leq \delta$ implies that the LHS of (73) is bounded by ϵ. Thus, $\zeta \mapsto \sup_{\theta\in\Theta}\tilde{\varphi}_\theta(\zeta)$ is continuous.

We next bound the integral in (70) by dividing the integration interval $|\zeta| \in (0, a\sqrt{n})$ into $|\zeta| \in (\delta\sigma_\theta\sqrt{n}, a\sqrt{n})$ and $|\zeta| \in (0, \delta\sigma_\theta\sqrt{n})$ for some arbitrary $\delta > 0$ that is independent

of x and θ. The integral over the first interval can be bounded as

$$\frac{1}{\pi}\int_{\delta\sigma_\theta\sqrt{n}<|\zeta|<a\sqrt{n}}\left|\frac{\tilde{\varphi}_\theta^n\left(\frac{\zeta}{\sigma_\theta\sqrt{n}}\right) - \gamma_\theta(\zeta)}{\zeta}\right|d\zeta$$
$$\leq \frac{2}{\pi}\log\left(\frac{a}{\delta\sigma_\theta}\right)q_{\delta,\bar{\zeta}}^n + \frac{1}{\pi}\int_{\delta\sigma_\theta\sqrt{n}<|\zeta|<a\sqrt{n}}\left|\frac{\gamma_\theta(\zeta)}{\zeta}\right|d\zeta$$
$$\leq \frac{2}{\pi}\log\left(\frac{a}{\delta\inf_{\theta\in\Theta}\sigma_\theta}\right)q_{\delta,\bar{\zeta}}^n$$
$$+ \frac{1}{\pi}\int_{|\zeta|>\delta\sqrt{n}\inf_{\theta\in\Theta}\sigma_\theta}\frac{e^{-\frac{1}{2}\zeta^2}}{|\zeta|}\left[1 + \frac{\sup_{\theta\in\Theta}\mu_{3,\theta}}{6\inf_{\theta\in\Theta}\sigma_\theta^3\sqrt{n}}|\zeta|^3\right]d\zeta \tag{74}$$

where the first inequality follows by upper-bounding the integrand using the triangle inequality and (71); the second inequality follows by upper-bounding γ_θ, defined in (68), using the triangle inequality and by optimizing over $\theta \in \Theta$. The RHS of (74) tends to zero faster than any power of $1/n$ uniformly in θ.

For the second interval, we express the integral as

$$\frac{1}{\pi}\int_{-\delta\sigma_\theta\sqrt{n}}^{\delta\sigma_\theta\sqrt{n}}\frac{e^{-\frac{1}{2}\zeta^2}}{|\zeta|}\left|e^{n\kappa_\theta\left(\frac{\zeta}{\sigma_\theta\sqrt{n}}\right)} - 1 - \frac{n\mu_{3,\theta}}{6}\left(\frac{\imath\zeta}{\sigma_\theta\sqrt{n}}\right)^3\right|d\zeta \tag{75}$$

where

$$\kappa_\theta(\zeta) \triangleq \log\tilde{\varphi}_\theta(\zeta) + \frac{1}{2}\sigma_\theta^2\zeta^2. \tag{76}$$

To bound (75), we will use that [26, Ch. XVI.2, Eq. (2.8)]

$$|e^\alpha - 1 - \beta| \leq \left(|\alpha - \beta| + \frac{1}{2}\beta^2\right)e^\gamma, \quad \gamma \geq \max(|\alpha|, |\beta|). \tag{77}$$

Recall that, by assumption (52), the fourth moment of $\tilde{F}_\theta$ is bounded. This implies that

$$\sup_{\theta\in\Theta}\int_{-\infty}^{\infty}|x|^k\mathrm{d}\tilde{F}_\theta(x) < \infty, \quad k = 1, \ldots, 4 \tag{78}$$

since, by Jensen's inequality, $\mathsf{E}[|\tilde{Z}_{\ell,\theta}|^k] \leq \mu_{4,\theta}^{k/4}$, $k = 1, 2, 3$. Then, given an $\epsilon > 0$ independent of θ and ζ, it is possible to choose a δ (independent of θ and ζ) such that, for $|\zeta| \leq \delta$,

$$\left|\kappa_\theta(\zeta) - \frac{1}{6}\mu_{3,\theta}(\imath\zeta)^3\right| < \epsilon|\zeta|^3 \tag{79}$$

and

$$|\kappa_\theta(\zeta)| < \frac{1}{4}\sigma_\theta^2\zeta^2, \quad \left|\frac{1}{6}\mu_{3,\theta}(\imath\zeta)^3\right| \leq \frac{1}{4}\sigma_\theta^2\zeta^2. \tag{80}$$

Indeed, after a Taylor series expansion of $\zeta \mapsto \kappa_\theta(\zeta)$ around $\zeta = 0$, and noting that $\kappa_\theta(0) = \kappa_\theta'(0) = \kappa_\theta''(0) = 0$, the LHS of (79) becomes

$$\left|\kappa_\theta(\zeta) - \frac{1}{6}\mu_{3,\theta}(\imath\zeta)^3\right| = \left|\frac{1}{6}\kappa_\theta'''(\tilde{\zeta})\zeta^3 - \frac{1}{6}\mu_{3,\theta}(\imath\zeta)^3\right| \tag{81}$$

for some $\tilde{\zeta} \in (0, \zeta)$. Equation (78) implies that $\tilde{\varphi}_\theta'''(0)$ exists and [26, Ch. XV.4, Lemma 2]

$$\tilde{\varphi}_\theta'''(0) = \kappa_\theta'''(0) = \imath^3\mu_{3,\theta}. \tag{82}$$

Furthermore, following the steps (72)–(73), it can be shown that, for every $\xi > 0$ and $k = 0, \ldots, 3$,

$$\sup_{\theta \in \Theta, |\zeta| \leq \xi} \left| \tilde{\varphi}_\theta^{(k)}(\zeta) - \tilde{\varphi}_\theta^{(k)}(0) \right| \leq \sup_{\theta \in \Theta} \mathsf{E}\left[|\tilde{Z}_{\ell,\theta}|^{k+1} \right] \xi. \quad (83)$$

By the definition of κ_θ in (76), the k-th derivative $\kappa_\theta^{(k)}$ is given by the ratio between a combination of derivatives of $\tilde{\varphi}_\theta$ up to order k in the numerator, and $\tilde{\varphi}_\theta^k$ in the denominator. Since $\tilde{\varphi}_\theta(0) = 1$, and since, by (52), $\mathsf{E}[|\tilde{Z}_{\ell,\theta}|^k]$ is bounded in θ, it follows from (83) that, for every ϵ, there exists a $\delta > 0$ satisfying

$$\sup_{\theta \in \Theta, |\zeta| \leq \delta} |\kappa_\theta'''(\zeta) - \kappa_\theta'''(0)| \leq 6\epsilon. \quad (84)$$

Combining this with (82), we conclude that (81) can be bounded as

$$\left| \frac{1}{6} \kappa_\theta'''(\tilde{\zeta}) \zeta^3 - \frac{1}{6} \mu_{3,\theta}(\imath\zeta)^3 \right| = \frac{1}{6} |\zeta|^3 \left| \kappa_\theta'''(\tilde{\zeta}) - \imath^3 \mu_{3,\theta} \right|$$
$$\leq \epsilon |\zeta|^3, \quad |\zeta| \leq \delta. \quad (85)$$

This proves (79). The inequalities in (80) follow along similar lines.

Using (77) together with (79) and (80), and replacing ζ by $\frac{\zeta}{n\sigma_\theta}$, we obtain that, for $|\zeta| \leq \delta\sigma_\theta\sqrt{n}$, the integrand in (75) is upper-bounded by

$$\frac{e^{-\frac{1}{4}\zeta^2}}{|\zeta|} \left(\frac{\epsilon}{\sigma_\theta^3\sqrt{n}} |\zeta|^3 + \frac{\mu_{3,\theta}^2}{72n} \zeta^6 \right)$$
$$\leq e^{-\frac{1}{4}\zeta^2} \left(\frac{\epsilon}{\inf_{\theta \in \Theta} \sigma_\theta^3 \sqrt{n}} \zeta^2 + \frac{\sup_{\theta \in \Theta} \mu_{3,\theta}^2}{72n} |\zeta|^5 \right). \quad (86)$$

Integrating (86) over $\zeta \in [-\delta\sigma_\theta\sqrt{n}, \delta\sigma_\theta\sqrt{n}]$, we conclude that (75) decays faster than $1/\sqrt{n}$ uniformly in x and θ. Since the same is true for (74), and since ε is arbitrary, we obtain that the RHS of (70) is $o(1/\sqrt{n})$ uniformly in θ. Lemma 5 thus follows.

APPENDIX III
NONLATTICE DISTRIBUTIONS AND EXPONENTIAL TILTING

Lemma 6: Let φ_θ denote the characteristic function of some distribution F_θ, and let $\tilde{\varphi}_\theta$ denote the characteristic function of the tilted distribution $\vartheta_{\theta,\tau}$ (cf. (44)). Then, for every $\zeta \neq 0$,

$$\sup_{\theta \in \Theta} |\varphi_\theta(\zeta)| < 1 \implies \sup_{\theta \in \Theta} |\tilde{\varphi}_\theta(\zeta)| < 1. \quad (87)$$

Thus, if a family of distributions is nonlattice, then so is the family of tilted distributions.

Proof: The characteristic function of the tilted random variable $V_\theta \sim \vartheta_{\theta,\tau}$ can be written as

$$\tilde{\varphi}_\theta(\zeta) \triangleq \int_{-\infty}^\infty e^{\imath\zeta x} d\vartheta_{\theta,\tau}(x)$$
$$= \mathsf{E}\left[e^{(\imath\zeta+\tau)Z_\theta} \right] e^{-\imath\zeta\tilde{\gamma}} \frac{1}{m_\theta(\tau)} \quad (88)$$

where $m_\theta(\tau)$ denotes the MGF of $Z_\theta \sim F_\theta$. It follows that

$$|\tilde{\varphi}_\theta(\zeta)| = \left| \mathsf{E}\left[e^{(\imath\zeta+\tau)Z_\theta} \right] \right| \frac{1}{m_\theta(\tau)}. \quad (89)$$

We next note that there exists an $\alpha = e^{\imath\phi}$ such that

$$|\varphi_\theta(\zeta)| = \alpha \mathsf{E}\left[e^{\imath\zeta Z_\theta} \right]$$
$$= \mathsf{E}[\cos(\zeta Z_\theta + \phi)]. \quad (90)$$

Likewise, there exists an $\tilde{\alpha} = e^{\imath\tilde{\phi}}$ such that

$$|\tilde{\varphi}_\theta(\zeta)| = \tilde{\alpha} \frac{\mathsf{E}\left[e^{(\imath\zeta+\tau)Z_\theta} \right]}{m_\theta(\tau)}$$
$$= \frac{\mathsf{E}\left[e^{\tau Z_\theta} \cos\left(\zeta Z_\theta + \tilde{\phi} \right) \right]}{m_\theta(\tau)}. \quad (91)$$

Furthermore, we have

$$e^{\imath\tilde{\phi}} \mathsf{E}\left[e^{\imath\zeta Z_\theta} \right] = \mathsf{E}[\cos(\zeta Z_\theta + \tilde{\phi})] + \imath\mathsf{E}[\sin(\zeta Z_\theta + \tilde{\phi})] \quad (92)$$

so

$$\mathsf{E}\left[\cos\left(\zeta Z_\theta + \tilde{\phi} \right) \right]^2 \leq \left| e^{\imath\tilde{\phi}} \mathsf{E}\left[e^{\imath\zeta Z_\theta} \right] \right|^2$$
$$= \mathsf{E}[\cos(\zeta Z_\theta + \phi)]^2 \quad (93)$$

where the last step is due to (90). Since $\mathsf{E}[\cos(\zeta Z_\theta + \phi)]$ is equal to $|\varphi_\theta(\zeta)|$ and hence nonnegative, it follows that

$$\mathsf{E}\left[\cos\left(\zeta Z_\theta + \tilde{\phi} \right) \right] \leq \mathsf{E}[\cos(\zeta Z_\theta + \phi)]. \quad (94)$$

Let

$$f(Z_\theta) \triangleq 1 - \cos(\zeta Z_\theta + \phi) \quad (95)$$
$$\tilde{f}(Z_\theta) \triangleq 1 - \cos\left(\zeta Z_\theta + \tilde{\phi} \right). \quad (96)$$

The LHS of (87) is equivalent to $\inf_{\theta\in\Theta} \mathsf{E}[f(Z_\theta)] > 0$. Similarly, the RHS of (87) is implied by $\inf_{\theta\in\Theta} \mathsf{E}[e^{\tau Z_\theta} \tilde{f}(Z_\theta)] > 0$ because

$$1 - \sup_{\theta\in\Theta} \frac{\mathsf{E}\left[e^{\tau Z_\theta} \cos\left(\zeta Z_\theta + \tilde{\phi} \right) \right]}{m_\theta(\tau)} \geq \frac{\inf_{\theta\in\Theta} \mathsf{E}\left[e^{\tau Z_\theta} \tilde{f}(Z_\theta) \right]}{\sup_{\theta\in\Theta} m_\theta(\tau)} \quad (97)$$

and m_θ is bounded by assumption (18).

We next show that

$$\inf_{\theta\in\Theta} \mathsf{E}\left[e^{\tau Z_\theta} \tilde{f}(Z_\theta) \right] = 0 \implies \inf_{\theta\in\Theta} \mathsf{E}\left[\tilde{f}(Z_\theta) \right] = 0. \quad (98)$$

We then note that, by (94), $\mathsf{E}[\tilde{f}(Z_\theta)] \geq \mathsf{E}[f(Z_\theta)]$. Since $f(\cdot)$ is nonnegative, $\inf_{\theta\in\Theta} \mathsf{E}[\tilde{f}(Z_\theta)] = 0$ thus implies that $\inf_{\theta\in\Theta} \mathsf{E}[f(Z_\theta)] = 0$. Hence, by reverse logic,

$$\inf_{\theta\in\Theta} \mathsf{E}[f(Z_\theta)] > 0 \implies \inf_{\theta\in\Theta} \mathsf{E}\left[e^{\tau Z_\theta} \tilde{f}(Z_\theta) \right] > 0 \quad (99)$$

which concludes the proof of Lemma 6.

To prove (98), we first note that, for every arbitrary $\delta > 0$,

$$\mathsf{E}\left[e^{\tau Z_\theta} \tilde{f}(Z_\theta) \right] \geq \mathsf{E}\left[e^{\tau Z_\theta} \tilde{f}(Z_\theta) \mathbb{I}\{|Z_\theta| \leq \delta\} \right]$$
$$\geq \mathsf{E}\left[\tilde{f}(Z_\theta) \mathbb{I}\{|Z_\theta| \leq \delta\} \right] e^{-\tau\delta} \quad (100)$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function. Furthermore,

$$\mathsf{E}\left[\tilde{f}(Z_\theta) \right] = \mathsf{E}\left[\tilde{f}(Z_\theta) \mathbb{I}\{|Z_\theta| \leq \delta\} \right] + \mathsf{E}\left[\tilde{f}(Z_\theta) \mathbb{I}\{|Z_\theta| > \delta\} \right]$$
$$\leq \mathsf{E}\left[\tilde{f}(Z_\theta) \mathbb{I}\{|Z_\theta| \leq \delta\} \right] + 2\frac{\sup_{\theta\in\Theta} \mathsf{E}\left[Z_\theta^2 \right]}{\delta^2} \quad (101)$$

where the inequality follows because $\tilde{f}(Z_\theta)$ is upper-bounded by 2 and from Chebyshev's inequality. Combining (100) with (101), and using that $\tilde{f}(\cdot)$ is nonnegative, we thus obtain that

$$0 \leq \inf_{\theta \in \Theta} \mathsf{E}\Big[\tilde{f}(Z_\theta)\Big]$$
$$\leq e^{\tau \delta} \inf_{\theta \in \Theta} \mathsf{E}\Big[e^{\tau Z_\theta} \tilde{f}(Z_\theta)\Big] + 2\frac{\sup_{\theta \in \Theta} \mathsf{E}\big[Z_\theta^2\big]}{\delta^2}. \quad (102)$$

By assumption (18), we have that $\sup_{\theta \in \Theta} \mathsf{E}\big[Z_\theta^2\big] < \infty$. Consequently, if $\inf_{\theta \in \Theta} \mathsf{E}[e^{\tau Z_\theta} \tilde{f}(Z_\theta)] = 0$ then we obtain that $\inf_{\theta \in \Theta} \mathsf{E}[\tilde{f}(Z_\theta)] = 0$ upon letting $\delta \to \infty$. This yields (98). ∎

APPENDIX IV
THE SECOND DERIVATIVE OF THE CGF IS BOUNDED AWAY FROM ZERO

Lemma 7: Assume that Z_θ is zero mean and its MGF and CGF satisfy (18) and

$$\inf_{\theta \in \Theta} \psi_\theta''(0) > 0. \quad (103)$$

Then, for every $\zeta_0 > 0$,

$$\inf_{\theta \in \Theta, |\zeta| < \zeta_0} \psi_\theta''(\zeta) > 0. \quad (104)$$

Proof: The second derivative of the CGF is given by

$$\psi_\theta''(\zeta) = \frac{\mathsf{E}\big[Z_\theta^2 e^{\zeta Z_\theta}\big] \mathsf{E}\big[e^{\zeta Z_\theta}\big] - \mathsf{E}\big[Z_\theta e^{\zeta Z_\theta}\big]^2}{m_\theta(\zeta)^2}. \quad (105)$$

By (18), the denominator on the RHS of (105) is bounded. Thus, to obtain (104), it suffices to show that the numerator on the RHS of (105) is bounded away from zero. To shorten notation, we next define

$$A \triangleq Z_\theta e^{\frac{\zeta}{2} Z_\theta} \quad (106)$$
$$B \triangleq e^{\frac{\zeta}{2} Z_\theta} \quad (107)$$
$$\sigma_A^2 \triangleq \mathsf{E}\big[A^2\big] \quad (108)$$
$$\sigma_B^2 \triangleq \mathsf{E}\big[B^2\big] \quad (109)$$
$$\overline{\lambda}_{A,B} \triangleq \frac{A}{\sigma_A} + \frac{B}{\sigma_B} \quad (110)$$
$$\underline{\lambda}_{A,B} \triangleq \frac{A}{\sigma_A} - \frac{B}{\sigma_B}. \quad (111)$$

Then, the numerator on the RHS of (105) becomes

$$\sigma_A^2 \sigma_B^2 - \mathsf{E}[AB]^2. \quad (112)$$

By following the proof of the Cauchy-Schwarz inequality [42, Th. 3.3.1], it can be shown that

$$|\mathsf{E}[AB]| \leq \sigma_A \sigma_B \mathsf{K} \quad (113)$$

where

$$\mathsf{K} \triangleq \max\left\{\left(1 - \frac{1}{2} \inf_{\theta \in \Theta, |\zeta| < \zeta_0} \mathsf{E}\big[\underline{\lambda}_{A,B}^2\big]\right)^+,\right.$$
$$\left.\left(1 - \frac{1}{2} \inf_{\theta \in \Theta, |\zeta| < \zeta_0} \mathsf{E}\big[\overline{\lambda}_{A,B}^2\big]\right)^+\right\} \quad (114)$$

and $(x)^+ \triangleq \max\{0, x\}$. Using (113), we can lower-bound (112) as

$$\sigma_A^2 \sigma_B^2 - \mathsf{E}[AB]^2 \geq (1 - \mathsf{K}^2) \inf_{\theta \in \Theta, |\zeta| < \zeta_0} \sigma_A^2 \inf_{\theta \in \Theta, |\zeta| < \zeta_0} \sigma_B^2$$
$$\geq (1 - \mathsf{K}^2) \inf_{\theta \in \Theta, |\zeta| < \zeta_0} \sigma_A^2 \quad (115)$$

where the second inequality follows because Z_θ is zero mean by assumption, so Jensen's inequality gives $\sigma_B^2 = \mathsf{E}\big[e^{\zeta Z_\theta}\big] \geq 1$. Thus, in order to show that the numerator on the RHS of (105) is bounded away from zero, it remains to show that

$$\inf_{\theta \in \Theta, |\zeta| < \zeta_0} \sigma_A^2 > 0 \quad (116)$$
$$\inf_{\theta \in \Theta, |\zeta| < \zeta_0} \mathsf{E}\big[\underline{\lambda}_{A,B}^2\big] > 0 \quad (117)$$
$$\inf_{\theta \in \Theta, |\zeta| < \zeta_0} \mathsf{E}\big[\overline{\lambda}_{A,B}^2\big] > 0. \quad (118)$$

To prove (116), recall that $\sigma_A^2 = \mathsf{E}[Z_\theta^2 e^{\zeta Z_\theta}]$. We next show that

$$\inf_{\theta \in \Theta, |\zeta| < \zeta_0} \mathsf{E}\big[Z_\theta^2 e^{\zeta Z_\theta}\big] = 0 \implies \inf_{\theta \in \Theta} \mathsf{E}\big[Z_\theta^2\big] = 0. \quad (119)$$

Since $\mathsf{E}\big[Z_\theta^2\big] = \psi_\theta''(0)$, it follows by assumption (103) that the RHS of (119) cannot be true. Hence, by reverse logic, $\inf_{\theta \in \Theta, |\zeta| < \zeta_0} \mathsf{E}[Z_\theta^2 e^{\zeta Z_\theta}] > 0$, which is (116).

To prove (119), we follow along the lines of (100)–(102). Indeed, for every arbitrary $\delta > 0$,

$$\mathsf{E}\big[Z_\theta^2 e^{\zeta Z_\theta}\big] \geq \mathsf{E}\big[Z_\theta^2 \mathbb{I}\{|Z_\theta| \leq \delta\}\big] e^{-\zeta \delta}. \quad (120)$$

Furthermore,

$$\mathsf{E}\big[Z_\theta^2\big] \leq \mathsf{E}\big[Z_\theta^2 \mathbb{I}\{|Z_\theta| \leq \delta\}\big]$$
$$+ \frac{\sqrt{\sup_{\theta \in \Theta} \mathsf{E}[Z_\theta^4] \sup_{\theta \in \Theta} \mathsf{E}[Z_\theta^2]}}{\delta} \quad (121)$$

by the Cauchy-Schwarz inequality and Chebyshev's inequality. Combining (120) with (121), and using that Z_θ^2 is nonnegative, it follows that

$$0 \leq \inf_{\theta \in \Theta} \mathsf{E}\big[Z_\theta^2\big]$$
$$\leq e^{\zeta_0 \delta} \inf_{\theta \in \Theta, |\zeta| < \zeta_0} \mathsf{E}\big[Z_\theta^2 e^{\zeta Z_\theta}\big]$$
$$+ \frac{\sqrt{\sup_{\theta \in \Theta} \mathsf{E}[Z_\theta^4] \sup_{\theta \in \Theta} \mathsf{E}[Z_\theta^2]}}{\delta}. \quad (122)$$

By assumption (18), we have that $\sup_{\theta \in \Theta} \mathsf{E}\big[Z_\theta^k\big] < \infty$ for $k = 2, 4$. Consequently, if $\inf_{\theta \in \Theta, |\zeta| < \zeta_0} \mathsf{E}\big[Z_\theta^2 e^{\zeta Z_\theta}\big] = 0$ then we obtain that $\inf_{\theta \in \Theta} \mathsf{E}\big[Z_\theta^2\big] = 0$ upon letting $\delta \to \infty$. This yields (119).

We next prove (117). To do so, we write $\mathsf{E}[\underline{\lambda}_{A,B}^2]$ as $\mathsf{E}[(A - B\eta_{\theta,\zeta})^2]/\sigma_A^2$, where $\eta_{\theta,\zeta} \triangleq \sigma_A/\sigma_B$. It can be shown that $\sigma_A^2 = m_\theta''(\zeta)$. Consequently, σ_A is bounded by assumption (18). To prove (117), it thus remains to show that

$$\inf_{\theta \in \Theta, |\zeta| < \zeta_0} \mathsf{E}\Big[(A - B\eta_{\theta,\zeta})^2\Big]$$
$$= \inf_{\theta \in \Theta, |\zeta| < \zeta_0} \mathsf{E}\Big[e^{\zeta Z_\theta}(Z_\theta - \eta_{\theta,\zeta})^2\Big]$$
$$> 0. \quad (123)$$

To this end, we first note that the steps (120)–(122) with Z_θ^2 replaced by $(Z_\theta - \eta_{\theta,\zeta})^2$ yield

$$\inf_{\theta \in \Theta, |\zeta| < \zeta_0} \mathsf{E}\left[e^{\zeta Z_\theta}(Z_\theta - \eta_{\theta,\zeta})^2\right] = 0$$
$$\implies \inf_{\theta \in \Theta, |\zeta| < \zeta_0} \mathsf{E}\left[(Z_\theta - \eta_{\theta,\zeta})^2\right] = 0. \quad (124)$$

Since Z_θ is zero mean, we further have that

$$\mathsf{E}[(Z_\theta - \eta_{\theta,\zeta})^2] \geq \mathsf{E}\left[Z_\theta^2\right] \quad (125)$$

so if $\mathsf{E}[(Z_\theta - \eta_{\theta,\zeta})^2]$ is zero, then so is $\mathsf{E}\left[Z_\theta^2\right]$. Since, by assumption (103), we have $\inf_{\theta \in \Theta} \mathsf{E}\left[Z_\theta^2\right] > 0$, the inequality in (123) follows from (124) by reverse logic.

Finally, (118) follows from the same steps as the ones used to show (117), but with $\eta_{\theta,\zeta}$ replaced by $-\eta_{\theta,\zeta}$. ∎

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] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55 765–55 779, Sep. 2018.

[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] 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.

[5] A. Martinez and A. Guillén i Fàbregas, "Saddlepoint approximation of random-coding bounds," in *Proc. Inf. Theory and Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2011.

[6] 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.

[7] B. Makki, T. Svensson, and M. Zorzi, "Finite block-length analysis of the incremental redundancy HARQ," *IEEE Wireless Commun. Lett.*, vol. 3, no. 5, pp. 529–532, Oct. 2014.

[8] ——, "Finite block-length analysis of spectrum sharing networks using rate adaptation," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2823–2835, Aug. 2015.

[9] ——, "Wireless energy and information transmission using feedback: Infinite and finite block-length analysis," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 5304–5318, Dec. 2016.

[10] M. Haghifam, M. Robat Mili, B. Makki, M. Nasiri-Kenari, and T. Svensson, "Joint sum rate and error probability optimization: Finite block-length analysis," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 726–729, Dec. 2017.

[11] Y. Yu, H. Chen, Y. Li, Z. Ding, and B. Vucetic, "On the performance of non-orthogonal multiple access in short-packet communications," *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 590–593, Mar. 2018.

[12] X. Sun, S. Yan, N. Yang, Z. Ding, C. Shen, and Z. Zhong, "Short-packet downlink transmission with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4550–4564, Jul. 2018.

[13] L. Zhang and Y. Liang, "Average throughput analysis and optimization in cooperative IoT networks with short packet communication," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11 549–11 562, Dec. 2018.

[14] H. Wang, Q. Yang, Z. Ding, and H. V. Poor, "Secure short-packet communications for mission-critical IoT applications," *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 2565–2578, May 2019.

[15] C. Li, N. Yang, and S. Yan, "Optimal transmission of short-packet communications in multiple-input single-output systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 7199–7203, Jul. 2019.

[16] Y. Polyanskiy and S. Verdú, "Scalar coherent fading channel: Dispersion analysis," in *Proc. IEEE Int. Symp. Theory (ISIT)*, Saint Petersburg, Russia, Jul. 2011, pp. 2959–2963.

[17] 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.

[18] 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.

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

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

[21] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Feb. 1999.

[22] X. Wu and R. Srikant, "Asymptotic behavior of error exponents in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1310–1325, Apr. 2007.

[23] H. Shin and M. Z. Win, "Gallager's exponent for MIMO channels: A reliability-rate tradeoff," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 972–985, Apr. 2009.

[24] I. Abou-Faycal and B. M. Hochwald, "Coding requirements for multiple-antenna channels with unknown Rayleigh fading," Bell Labs., Lucent Technologies, Tech. Rep., 1999.

[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] W. Feller, *An Introduction To Probability Theory And Its Applications*, 2nd ed. New York, NY, USA: Wiley, 1971, vol. II.

[27] J. L. Jensen, *Saddlepoint Approximations*. Clarendon Press, Oxford, UK, 1995.

[28] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2647–2666, May 2014.

[29] J. Font-Segura, G. Vazquez-Vilar, A. Guillén i Fàbregas, and A. Lancho, "Saddlepoint approximations of lower and upper bounds to the error probability in channel coding," in *Proc. Conf. Inf. Sci. and Sys. (CISS)*, Princeton, NJ, USA, Mar. 2018.

[30] G. Vazquez-Vilar, A. Guillén i Fàbregas, T. Koch, and A. Lancho, "Saddlepoint approximation of the error probability of binary hypothesis testing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Vail, CO, USA, Jun. 2018, pp. 2306–2310.

[31] T. Erseghe, "On the evaluation of the Polyanskiy-Poor–Verdú converse bound for finite block-length coding in AWGN," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6578–6590, Dec. 2015.

[32] ——, "Coding in the finite-blocklength regime: Bounds based on Laplace integrals and their asymptotic approximations," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6854–6883, Dec. 2016.

[33] 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.

[34] 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.

[35] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb 2005.

[36] 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.

[37] A. Lancho, "Fundamental limits of short-packet wireless communications," Ph.D. dissertation, Universidad Carlos III de Madrid, Spain, Jun. 2019.

[38] R. G. Gallager, *Information Theory and Reliable Communication*, 1st ed. New York, NY, USA: John Wiley & Sons, 1968.

[39] Y. Altuğ and A. B. Wagner, "Refinement of the random coding bound," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6005–6023, Oct. 2014.

[40] J. Honda, "Exact asymptotics of random coding error probability for general memoryless channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Vail, CO, USA, Jun. 2018, pp. 1844–1848.

[41] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York, NY, USA: McGraw-Hill Book Co., 1976, International Series in Pure and Applied Mathematics.

[42] A. Lapidoth, *A Foundation in Digital Communication*, 2nd ed.. Cambdrige, UK: Cambridge University Press, 2017.

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, Gothenburg, Sweden. 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.

Johan Östman (S'16) received the B.Sc. and the M.Sc. degrees in electrical engineering from Chalmers University of Technology, Gothenburg, Sweden, in 2011 and 2014, respectively. He is currently pursuing the Ph.D. degree in electrical engineering at Chalmers University of Technology, Gothenburg, Sweden.

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.

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.

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. He served as Vice Chair of the Spain Chapter of the IEEE Information Theory Society from 2013–2016, and he is serving as the chapter's Chair in 2020–2021. Since 2020, he is also serving as Associate Editor for Communications for the IEEE TRANSACTIONS ON INFORMATION THEORY. His research interests are in digital communication theory and information theory.

Gonzalo Vazquez-Vilar (S'08–M'12) is a Visiting Professor with the Signal Theory and Communications Department of the Universidad Carlos III de Madrid, Spain. He received the Telecommunication Engineering degree from the University of Vigo, Spain, in 2004, the Master of Science degree from Stanford University, U.S., in 2008 and the Ph.D. in Communication Systems from the Universidade de Vigo, Spain, in 2011. He held appointments as visiting researcher at Stanford University, U.S., University of Cambridge, U.K., and Princeton University, U.S, at different points in his career. From 2011 to 2014, he was a post-doctoral fellow in the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Spain, and he joined the Signal Processing Group at Universidad Carlos III de Madrid, Spain, in September 2014.

Gonzalo Vazquez-Vilar was an FPI Fellow during his Ph.D. studies and a Juan de la Cierva Fellow in the period 2015-2018 (both granted by the Spanish Ministry of Economy and Competitiveness). He served as Chair of the Spain Chapter of the IEEE Information Theory Society from 2017–2019. His research interests lie in the field of Shannon theory, with emphasis on finite-length information theory and communications.