Confidence Level for Finite Blocklength Ultra Reliable Communication over Fading Channels

Alex The Phuong Nguyen, Raphaël Le Bidan, Frédéric Guilloud

Confidence Level for Finite Blocklength Ultra Reliable Communication over Fading Channels

Alex The Phuong Nguyen
IMT Atlantique, Lab-STICC, UBL
29238 Brest, France
thephuong.nguyen@imt-atlantique.fr

Raphaël Le Bidan
IMT Atlantique, Lab-STICC, UBL
29238 Brest, France
raphael.lebidan@imt-atlantique.fr

Frédéric Guilloud
IMT Atlantique, Lab-STICC, UBL
29238 Brest, France
frederic.guilloud@imt-atlantique.fr

Abstract—We present the reliability confidence level as a way to quantify reliability for ultra reliable connections in uncertain environments subject to random block-error rate fluctuations. As ultra reliability is usually linked to the low latency constraint, we analyze the reliability confidence level using the latest results on block-error rate in the finite blocklength regime. The analysis is carried out for OFDM-based systems over Rayleigh slow frequency block-fading channels. The reliability confidence level is bounded using analytic expressions which are then applied to solve two optimization problems. We first find the minimal resources (or codeword length) to guarantee a target reliability with a given confidence. We then investigate an optimal resource sharing strategy within the context of 5G New Radio. The solutions for optimization problems are obtained without resorting to cumbersome Monte-Carlo simulations. Numerical evaluations validate the relevance as well as the tightness of the proposed bounds and approximations.

I. INTRODUCTION

Mission critical applications such as factory remote control, vehicle autopilot or telesurgery require Ultra Reliable Low Latency Communication (URLLC) which is among the intrinsic novelties of 5G networks. Reliability is defined as the probability of successfully transmitting a certain number of information bytes within a certain delay at a certain channel quality [1, Sec. 7.9]. The "success probability" is measured by the BLock Error Rate (BLER) ranging from 10^{-5} down to 10^{-9} in the Ultra Reliable (UR) context. However, the term "channel quality" must not be neglected. Indeed, if the channel quality changes randomly, the BLER is also a random variable and the assumed reliable connection may unexpectedly become unreliable. In other words, thinking only in terms of average BLER value even as low as e.g. 10^{-9} is not sufficient to assess reliability.

To address this concern, we introduce hereafter the *Reliability Confidence level* as a way to quantify reliability:

$$P_{\mathrm{R}} \triangleq \Pr\{P_{\mathrm{E}} \le \epsilon_0\} \tag{1}$$

where the BLER is a random variable denoted by P_{E} and where ϵ_0 denotes the targeted Quality of Service (QoS) BLER threshold (or simply target BLER). The random nature of BLER is induced by stochastic factors such as fading, inter-vehicle speed, mobile direction changing.

Note that P_{R} is similar in essence to the *Probably Correct Reliability* introduced in [2]: both are based on the meta-probability concept [3]. However, [2] uses asymptotic outage probability whereas the present paper resorts to recent results on channel decoding error in the Finite Blocklength (FBL) regime [4], [5] to characterize the BLER. Our approach is motivated by the fact that although the BLER can be made arbitrarily small by allowing the code length to grow arbitrarily large, such an assumption may violate the "within a certain delay" constraint inherent to the definition of UR [1, Sec. 7.9].

It is worth mentioning that P_{R} was implicitly introduced in NarrowBand-IoT (NB-IoT) [6, Sec. 7.23] as well as in the first versions of 5G [7, Sec. 8.1] under the Radio Link Monitoring (RLM) concept at link layer. Generally speaking, RLM consists of individual tests that estimate the BLER of *hypothetical* control-plane transmissions, then declare Radio Link Failure (RLF) if the estimated BLERs repeatedly exceed certain thresholds. A RLF declaration triggers in turn transmitter turning-off, cell research, and attach procedure [8, Sec. 22.7, p. 526][9]. As a consequence, even if the average BLER is small, radio link level connections may be unreliable due to these repeated connection resets.

There are other approaches to characterize reliability. A popular one is to employ queuing analysis on top of physical layer transmissions to assess the probability that the aggregate delay exceeds a given value. One such probability measure named delay violation probability (DVP) has been recently investigated in [10]. Developing the idea of DVP, but for a non-constant flow of packets and with the incorporation of data freshness notion, [11] provides the results for peak age of information, which "describes the maximum time that is elapsed since the last received update", over binary-input Additive White Gaussian Noise (AWGN) channels. In [12], the work of [11] is extended using the "parameter-s Random Coding Union bound" to characterize the BLER of short packet transmissions in block-fading channels. Other results concerning queuing-related DVP can be found in [13] where effective bandwidth is analyzed, and in [14] where the concept of effective capacity is used.

In this paper, we develop the concept of Reliability Confidence level P_R in (1) for OFDM-based systems over Rayleigh *frequency-domain block-fading* and *time-domain slow fading* channels in the FBL regime. As P_R is almost analytically intractable, especially in the FBL regime, we resort to probabilistic bounds and then approximate them to obtain simple yet relatively tight estimates. These approximations are then used to find the smallest number of resources (codeword length) n required to guarantee a target BLER ϵ_0 with confidence α, i.e. to ensure $P_R(\epsilon_0, n) \geq \alpha$. It is worth mentioning that even with the novel method of [5], the smallest-n hunt requires i) sampling the range of possible values for n resulting in a set of candidate values denoted by $n_{\text{candidate}}$; ii) an exhaustive test of all candidate values $n_{\text{candidate}}$ and iii) for each $n_{\text{candidate}}$, a time-consuming Monte-Carlo sampling to evaluate $P_R(\epsilon_0, n_{\text{candidate}})$. Our results simplify the process by providing *analytic* expressions allowing to closely estimate the minimum value of n such that $P_R(\epsilon_0, n) \geq \alpha$.

The paper is organized as follows. The system model is described in Section II. Statement of the code length optimization problem, the Reliability Confidence level P_R analysis and the analytic approximate bound solutions are given in Section III. Section IV is dedicated to another application of our results, namely identifying an optimal resource sharing strategy within the context of 5G New Radio (NR) [15].

Notations: Upper-case letters are used to denote random variables (RV) and their realizations are written in lower-case. Boldface letters are for vectors. $F_Z(.)$ is the cumulative density function (CDF) of RV Z. $Q(.)$ is the Normal distribution Q-function. $\text{Exp}(\mu)$ denotes the Exponential distribution with mean μ. $\lceil x \rceil$ is the smallest integer number greater than or equal to x. We use $x = a^+$ as a shortcut for $\lceil x \approx a$ and $x \geq a \rceil$.

II. SYSTEM MODEL

We consider a single-antenna OFDM-based system similar to 5G NR [15]. As illustrated in Fig. 1, in the time domain, the channel is slow fading, i.e. the channel remains unchanged during a certain number of transmissions (coherence time). In the frequency domain, the channel is block-fading and a user is allocated n_c subcarriers within a fading block to form an *allocation block*. Coding is performed in the frequency domain across L such blocks; hence the codeword length is $n = n_c L$. Note that both L and n_c are interchangeable as long as the block fading assumption is still valid. The baseband received signal associated to the l–th block, $1 \leq l \leq L$, is equal to

$$H_l \mathbf{X}_l + \mathbf{W}_l \qquad (2)$$

where $\mathbf{X}_l \in \mathbb{C}^{n_c}$ contains the transmitted symbols within frequency block l and where $\mathbf{W}_l$ is the AWGN channel noise distributed according to the normal distribution
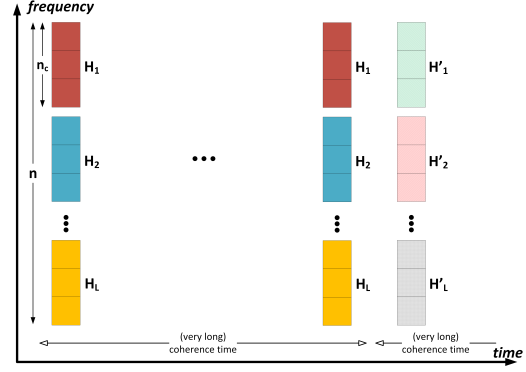

Fig. 1. System model under consideration. Coding is performed over L (frequency) allocation blocks, each composed of n_c subcarriers (block-fading). In time domain, the channel remains unchanged over a long coherence time (slow fading), then changes to other values.

$\mathcal{CN}(\mathbf{0}, \mathbf{I}_{n_c})$. The received signal is dominated by scattered diffuse components: channel coefficients H_l are distributed according to a Rayleigh fading, i.e. $H_l \sim \mathcal{CN}(0, 1)$. The transmitted symbols satisfy the power constraint $\|\mathbf{X}_l\|^2 = n_c \mu$. Let Y_l denotes the Signal to Noise Ratio (SNR) associated to the reception of block l; Y_l is exponentially distributed with mean μ. Hereafter we assume a coherent receiver with perfect channel state information (CSIR), a reasonable assumption for slow fading channels. Hence the receiver knows perfectly H_l.

The motivation for the slow fading assumption comes from the LTE standard where coherence time may be 400 times larger than one transmission slot (see [16, Sec. II]). This number can be even larger in 5G NR with mini-slots [15]. Nonetheless and departing from [16], [17] where one transmission is assumed per coherence time, many transmissions are considered in our setup. Moreover, coding in [16], [17] is performed within one coherence time and one frequency coherence block only whereas the present analysis considers codewords spanning several fading blocks in the frequency domain. Another interesting analysis for FBL coding over fading is [18]. But the latter assumes no CSI and Monte-Carlo simulations are required to assess the performance.

Considering the abovementioned assumptions and the fact that the channel codes used in practice are mostly AWGN codes, i.e. codes designed and optimized for AWGN channels, within each coherence time, the system model (2) is tantamount to transmitting over L independent parallel AWGN sub-channels with respective SNR $Y_1, \ldots, Y_L$ and coding across those L sub-channels. Collecting the SNRs in vector form $\mathbf{Y} = \{Y_l\}_{l=1}^{L}$, the overall block-error probability P_E for the model (2) is well approximated by the *Refined* Normal Approximation (RNA) [5]:

$$P_E(n, k, L, \mathbf{Y}) \lesssim Q\left(\sqrt{n}\frac{S - k/n + \frac{\log(2n)}{2n}}{\sqrt{V}}\right) \qquad (3)$$

with k the amount of information (in nats), $S \triangleq \frac{1}{L}\sum_{i=1}^{L}\log(1+Y_i)$, $V \triangleq \frac{1}{L}\sum_{i=1}^{L}\left(1-(1+Y_i)^{-2}\right)$. Because $\mathbf{Y}$ changes randomly every coherence time, P_E in (3) is a random variable and (1) is applicable.

It is necessary to emphasize that (3) is not the original Normal approximation given in [4] that derives from the PPV (meta) converse bound and the $\kappa\beta$ achievability bound of the same paper but without the third-order term $\frac{\log(2n)}{2n}$. The latter was first introduced in [19] where the authors, in evaluating the Random Coding Union (RCU) achievability bound which is tighter than the $\kappa\beta$ bound, not only improve the tightness of the Normal approximation of [4], but also confirm that the RNA (3) can be "taken as a reference for achievability" (see [19, inequality (9)] and the related remarks, especially the fourth one for parallel AWGN channels; also the discussion at the end of [5, Sec. IV-H]). Because of the achievable nature of (3), analyzing (1) with (3) corresponds to the worst-case scenario which is well-suited to UR applications. In order to assess the tightness of (3), the results obtained in Sections III and IV with the RNA will be compared to those obtained with the PPV converse bound [5].

III. RELIABILITY CONFIDENCE LEVEL ANALYSIS

The random nature of P_E in (1) prevents it from being upper-bounded by ϵ_0. Thus we are interested in finding the smallest codeword length n so that the probability $P_R \triangleq \Pr\{P_E \leq \epsilon_0\}$ is higher than a confidence level α_0:

OP1: $\hat{n} = \min_{n\geq 0} n$ **subject to** $P_R \geq \alpha_0$

Since latency is related to codeword length[1], **OP1** can be interpreted as the search for the *lowest possible latency* that guarantees the required reliability.

Except for $L = 1$ which is trivial, the exact calculation of P_R involves L-dimensional integrals which becomes quickly intractable. We first provide in subsection III-A upper and lower bounds on P_R. Tight approximations of these bounds are presented in subsection III-B, from which closed-form approximate solutions of problem **OP1** are derived in subsection III-C.

A. Upper and lower bounds on P_R

Substituting (3) into (1) yields:

$$P_R = \Pr\left\{Z \geq \theta \triangleq \mathcal{Q}^{-1}(\epsilon_0)\right\} \quad (4)$$

where $Z \triangleq \sqrt{\frac{n}{V}}\left(S - \frac{k}{n} + \frac{\log(2n)}{2n}\right)$. Hence $P_R(\theta)$ is the complementary CDF of the random variable Z.

[1]For example, in OFDM-based systems, increasing codeword length n by fixing the number of allocation blocks L is equivalent to increasing n_c. This requires reducing subcarrier spacing, thereby increasing the duration of OFDM symbols (latency).

Theorem III.1. *For typical setups such that $\epsilon_0 < 0.5$, P_R defined in (4) is bounded below and above by:*

$$P_R \geq P_{lo} \triangleq 1 - F_S\left(\frac{\theta}{\sqrt{n}} + \frac{k}{n} - \frac{\log(2n)}{2n}\right) \quad (5)$$

$$P_R \leq P_{up} \triangleq 1 - F_S\left(\frac{k}{n} - \frac{\log(2n)}{2n}\right) \quad (6)$$

Proof. Let $\Omega = S - \frac{k}{n} + \frac{\log(2n)}{2n}$,

$$P_R = \Pr\{Z \geq \theta \cap \Omega > 0\} + \Pr\{Z \geq \theta \cap \Omega \leq 0\} \quad (7)$$
$$= \Pr\{Z \geq \theta \cap \Omega > 0\} \quad (8)$$

where the second term of (7) equals to 0 because $\epsilon_0 < 0.5$ and $\theta = \mathcal{Q}^{-1}(\epsilon_0) > 0$. The upper-bound P_{up} can be obtained by using the Fréchet inequality:

$$P_R = \Pr\{Z \geq \theta \cap \Omega > 0\} \leq \Pr\{\Omega > 0\} \quad (9)$$

Regarding P_{lo}, (5) is proved by letting $Z_1 = \sqrt{n}\left(S - \frac{k}{n} + \frac{\log(2n)}{2n}\right) = \sqrt{n}\Omega$, and from (8),

$$P_R = \Pr\{Z \geq \theta \mid \Omega > 0\}\Pr\{\Omega > 0\} \geq \Pr\{Z_1 \geq \theta \mid \Omega > 0\}\Pr\{\Omega > 0\} = \Pr\{Z_1 \geq \theta\} \quad (10)$$

because $\theta > 0$ and given $\Omega > 0$, we have $Z \geq Z_1$. ∎

Note that P_{up} does not depend on the threshold θ and can be considered as a variation of the classic outage probability. Also, the two bounds are consistent in the sense that

$$\lim_{n\to+\infty} P_{lo} = \lim_{n\to+\infty} P_{up} = 1 \quad (11)$$

B. Bounds approximations

1) For large L: we leverage the results of [20] (see also [21, Sec. III-B]) stating that for i.i.d. $Y_l \sim \text{Exp}(\mu)$, the $S = \frac{1}{L}\sum_{l=1}^{L}\log(1+Y_l)$ is well approximated by a Gaussian distribution $\mathcal{N}(\nu, \sigma^2)$ with:

$$\nu = e^{1/\mu}E_1(1/\mu),$$
$$\sigma^2 = \frac{1}{L}\left(\frac{2}{\mu}e^{1/\mu}G_{3,4}^{4,0}\left(1/\mu\big|_{0,-1,-1,-1}^{0,0,0}\right) - \nu^2\right) \quad (12)$$

where $E_1(x)$ denotes the exponential integral $E_1(x) \triangleq \int_1^\infty t^{-1}e^{-xt}dt$ and $G(.)$ denotes the Meijer G-function. Numerical evaluations show that this approximation is already quite accurate for $L \geq 3$ and that the accuracy improves with L.

2) For $L = 2$: because of the "low latency" and "short packet" requirements, this case is of particular interest and thus deserves a careful dedicated analysis.

Theorem III.2. *Let $S = \frac{1}{2}\left(\log(1+Y_1) + \log(1+Y_2)\right)$ with Y_1, Y_2 are i.i.d. Exp$(1/\lambda)$.*

For moderate and large $s > 0$ such that $e^{2s} \gg 1$, $F_S(s) \approx 1 - \lambda e^{2\lambda}K_{-1}(\lambda, \lambda e^{2s})$ where $K_{-1}(\lambda, \lambda e^{2s})$ is an incomplete Bessel (leaky aquifer) function [22].

For small $s = 0^+$, $F_S(s) \approx 1 - e^{\lambda - \lambda e^{2s}}(\lambda e^{2s} - \lambda + 1)$.

Sketch of proof. Let $U_1 = 1 + Y_1$, $U_2 = 1 + Y_2$ and $Z = U_1 U_2$ then $F_S(s) = F_Z(e^{2s})$. The complementary CDF of Z is

$$1 - F_Z(z) = \Pr(U_1 \geq z/U_2) = I_1 + I_2 \qquad (13)$$

where

$$
\begin{aligned}
I_1 &= \int_{u_2=1}^{\infty} \int_{u_1=z/u_2}^{\infty} f_{U_1}(u_1) f_{U_2}(u_2) \mathrm{d}u_1 \mathrm{d}u_2 \\
&= \int_{u_2=1}^{z} f_{U_2}(u_2)(1 - F_{U_1}(z/u_2)) \mathrm{d}u_2 \qquad (14) \\
&= \lambda e^{2\lambda} \int_{u=1}^{z} e^{-\lambda u - \frac{\lambda z}{u}} \mathrm{d}u
\end{aligned}
$$

$$I_2 = \int_{u_2=z}^{\infty} \int_{u_1=1}^{\infty} f_{U_1}(u_1) f_{U_2}(u_2) \mathrm{d}u_1 \mathrm{d}u_2 = e^{\lambda - \lambda z} \quad (15)$$

For large s such that $z = e^{2s} \gg 1$, $I_2 \approx 0$ and $I_1 \approx \lambda e^{2\lambda} \int_1^{\infty} e^{-\lambda u - \frac{\lambda z}{u}} \mathrm{d}u = \lambda e^{2\lambda} K_{-1}(\lambda, \lambda z)$ with $K_\nu(a,b) \triangleq \int_1^{\infty} \frac{e^{-at-b/t}}{t^{\nu+1}} \mathrm{d}t$.

For small s, $z = 1^+$ and the interval of integration $[1,z]$ of I_1 is so small that *one of many approximations* is $I_1 \approx \lambda e^{2\lambda}(z-1)e^{-\lambda u - \frac{\lambda z}{u}}|_{u=1} = \lambda(z-1)e^{\lambda - \lambda z}$, yielding the final result. ∎

An interesting remark is that $s = 0^+$ is of particular interest for problem **OP1** where low target BLER (small ϵ_0, e.g. 10^{-9}) and high confidence (high α_0, e.g. 90%) are typically required, leading in turn to large codeword length n (hence $n \gg k$ and $n \gg \theta$, since we have $\theta = \mathcal{Q}^{-1}(10^{-9}) \approx 6$). Therefore the values s at which $F_S(s)$ is evaluated in (5) and (6) are close to 0.

C. Approximate solutions for problem OP1

Approximate solutions to **OP1** can now be obtained from the results of Section III-B by evaluating P_{lo} in (5) and P_{up} in (6) over a range of candidate values n. However, the solution would be much handier if the exhaustive search over n could be avoided. To this aim, one can solve the following equations:

$$P_{\mathrm{lo}}(n_\star) = \alpha_0, \qquad P_{\mathrm{up}}(n^\star) = \alpha_0 \qquad (16)$$

for the unknowns $n_\star$ and $n^\star$ to sandwich the optimal code length within the restricted interval $n_\star \leq n \leq n^\star$. These equations can be solved by popular root-finding algorithms, but further simplification is possible by noting that in typical low-target-BLER high-confidence scenarios, the solution n is presumably large enough to assume $\frac{\theta}{\sqrt{n}} + \frac{k}{n} \gg \frac{\log(2n)}{2n}$, so that (16) becomes

$$F_S\left(\frac{\theta}{\sqrt{n_\star}} + \frac{k}{n_\star}\right) = 1 - \alpha_0, \quad F_S\left(\frac{k}{n^\star}\right) = 1 - \alpha_0 \quad (17)$$

For large L, S is well approximated by a Normal random variable (see Section III-B), whence

$$n_\star = \left\lceil \frac{4k^2}{\left(\sqrt{\Delta_{(\geq 3)}} - \mathcal{Q}^{-1}(\epsilon_0)\right)^2} \right\rceil, \quad n^\star = \left\lceil \frac{k}{\nu + \sigma \mathcal{Q}^{-1}(\alpha_0)} \right\rceil \qquad (18)$$

where $\Delta_{(\geq 3)} = (\mathcal{Q}^{-1}(\epsilon_0))^2 + 4k(\sigma \mathcal{Q}^{-1}(\alpha_0) + \nu)$ and ν, σ are given in (12).

For small L, as discussed at the end of Section III-B, the low-target-BLER high-confidence assumption leads to $F_S(s) \approx 1 - e^{\lambda - \lambda e^{2s}}(\lambda e^{2s} - \lambda + 1)$ where $\lambda = \mu^{-1}$ (see Theorem III.2). Let $x = \lambda(e^{2s} - 1)$ and after a few mathematical manipulations,

$$n_\star = \left\lceil \frac{4k^2}{\left(\sqrt{\Delta_{(2)}} - \mathcal{Q}^{-1}(\epsilon_0)\right)^2} \right\rceil, \quad n^\star = \left\lceil \frac{2k}{\log(1 + \mu x)} \right\rceil \qquad (19)$$

where $\Delta_{(2)} = (\mathcal{Q}^{-1}(\epsilon_0))^2 + 2k\log(1 + \mu x)$ and x is the solution of $\log(1 + x) - x = \log(\alpha_0)$, which depends only on α_0 and, therefore, can be obtained efficiently by popular root-finding numerical methods.

D. Examples of numerical results for problem OP1

Approximate solutions to problem **OP1**, i.e. the smallest code length n that satisfies given target BLER and confidence requirements, are illustrated in Fig. 2 for $L = 5$ and in Fig. 3 for $L = 2$. The results are expressed in terms of the number of subcarriers per allocation block $n_c = n/L$ at fixed L, as a function of the mean SNR μ. These results are obtained for a typical URLLC setup: the target BLER is set to $\epsilon_0 = 10^{-5}$ [1, Sec. 7.9] and the confidence level is set to $\alpha_0 = 90\%$.

In both figures, we first note that the results obtained with the RNA (3) are very close to those obtained with the PPV converse bound (using the method of [5]), thereby demonstrating the accuracy of the RNA in searching for the optimal solution to problem **OP1**.

As expected, the smallest required n is reduced as the mean SNR μ increases for all curves.

The bounds P_{lo} and P_{up} have been evaluated by Monte-Carlo simulations. It is observed here that the higher the mean SNR μ, the closer the bounds to P_{R}. Finally, the proposed analytic approximate solutions of **OP1** given by (18) for Fig. 2 and (19) for Fig. 3, respectively, closely match in both cases the corresponding Monte-Carlo simulations of (5) and (6). The fact that P_{lo} tightly approaches P_{R} as the mean SNR μ increases follows from the fact that for large μ, V is more likely to be equal to 1, and therefore Z_1 is more likely to be Z (see the proof of Theorem III.1). Regarding now the gap between P_{R} and P_{up}, inspection of (9) suggests that the better the channel quality (higher μ), the easier it is to achieve the target BLER (event $Z > \theta = \mathcal{Q}^{-1}(\epsilon_0)$) in the non-outage case (event $\Omega = S - \frac{k}{n} + \frac{\log(2n)}{2n} > 0$).

Note that since L is fixed here, the solution of **OP1** can be expressed as $n_c = n/L$. In OFDM-based systems, increasing n_c is tantamount to reducing subcarrier spacing and, therefore, increasing the OFDM symbol duration (latency). For that reason, the smallest n, and also the smallest n_c, can be interpreted as the "lowest possible latency" that guarantee the required target BLER and
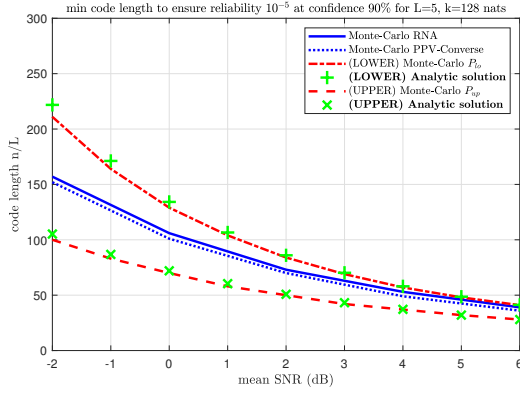
Fig. 2. Smallest code length expressed in n/L to ensure target BLER 10^{-5} at confidence 90% to transport $k = 128$ nats for $L = 5$. The analytic solution curves are obtained with (18). The Monte-Carlo P_{lo} and P_{up} curves are obtained with (5) and (6) respectively.
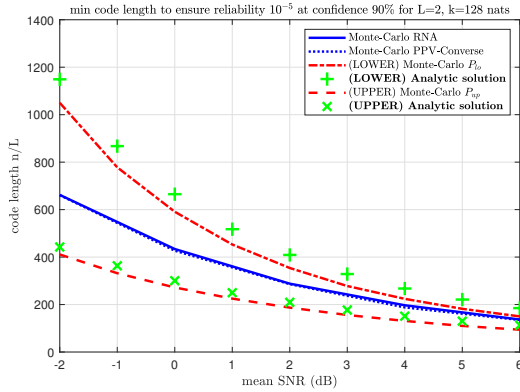

Fig. 3. Smallest code length expressed in n/L to ensure target BLER 10^{-5} at confidence 90% to transport $k = 128$ nats for $L = 2$. The analytic solution curves are obtained with (19). The Monte-Carlo P_{lo} and P_{up} curves are obtained with (5) and (6) respectively.

confidence. One may alternatively fix n_c and accordingly adjust L. This corresponds to the typical resource allocation of 5G NR and forms the basis of the resource sharing problem investigated in the next Section.

IV. RESOURCE SHARING TRADE-OFF

As a second application of the proposed analysis, we investigate optimization of resource sharing in the context of 5G NR OFDM [15]. To achieve URLLC extreme requirements, communication systems must operate in proactive manners. To this end, RLM [6, Sec. 7.23][7, Sec. 8.1] is designed to *predict* whether the current connection is reliable by regularly estimating the BLER of a hypothetical control message transmission from the monitoring of the connection quality metrics such as SNR. The main purpose of RLM, in brief, is to declare RLF if after N_{out} consecutive events [predicted BLER $> \epsilon_{\text{out}}$] (so-called out-of-sync), during the next $T310$ tests, there

are not N_{in} consecutive [predicted BLER $< \epsilon_{\text{in}}$] events[2]. More details of RLM can be found in [8, Sec. 22.7].

To simplify the analysis yet stay focused on the essence of (1), we assume that connections are in (so-called link level) outage as soon as the out-of-sync event occurs. Under this assumption, the optimal resource sharing problem between two users reads:

$$\textbf{OP2:} \quad \hat{n}_1 = \underset{0 \leq n_1 \leq N}{\arg\max} \; k_1^{\text{ef}} + k_2^{\text{ef}}$$

where we define the effective throughput $k_1^{\text{ef}} + k_2^{\text{ef}}$ as $k_1^{\text{ef}} \triangleq k_1 \left(1 - (\Pr\{P_{\text{E}}^{(\text{user1})} \geq \epsilon_1\})^{N_{\text{out}}} \right)$ and $k_2^{\text{ef}} \triangleq k_2 \left(1 - (\Pr\{P_{\text{E}}^{(\text{user2})} \geq \epsilon_2\})^{N_{\text{out}}} \right)$. Here, the number of available allocation blocks $N = n_1 + n_2$ is fixed and user 1 (resp. user 2) is allocated n_1 (resp. n_2) blocks to transport k_1 (resp. k_2) bits at target BLER ϵ_1 (resp. ϵ_2).

Using the results of the previous section to solve **OP2**, we illustrate the numerical solutions in Fig. 4 for two users of *the same type* (same target BLER $\epsilon_1 = \epsilon_2$ with same message length $k_1 = k_2$), and in Fig. 5 for two users of *different types* ($\epsilon_1 = \epsilon_2$ but different message lengths $k_1 \neq k_2$), respectively. We take a 5G NR frame structure: $N = 50$ available blocks, each composed of $n_c = 2 \times 12$ subcarriers (mini-slot) [15] [3]. The message lengths are approximately equal to 32 bytes [1].

It is observed that in both figures, there exists an optimal, non-necessarily unique resource sharing strategy that achieves the maximal effective throughput, and our analysis allows us to characterize such a strategy without resorting to cumbersome Monte-Carlo simulations. Note again the tightness of the results obtained with our approximate formulas compared to Monte-Carlo simulations of both the RNA and the PPV converse bound. As already mentionned, there may be more than one strategy that achieves the maximal effective throughput. Also a more tolerant reliability requirement, i.e. greater N_{out}, increases the number of such strategies. This can be intuitively explained by noting that increasing N_{out} reduces *exponentially* the link level outage probability; hence the difference between the optimal strategy and its surrounding ones becomes negligible.

In Fig. 4, two users have the same requirements hence the optimal sharing strategy is equal sharing, as expected. On the other hand, if one user needs to transmit a longer message, intuitively we need to allocate more resource blocks to that user, as confirmed by Fig. 5.

V. CONCLUSION

In this paper we have introduced the Reliability Confidence level as a way to assess reliability in Ultra

---

[2]The counters N_{out}, N_{in} and $T310$ are configured by network owners; ϵ_{out} and ϵ_{in} usually equal to 10% and 2% respectively. Here, we assume that ϵ_{out} and ϵ_{in} are much extreme.

[3]We have extended our system model to code over two adjacent OFDM symbols. Because of the slow fading assumption, this extension does not change the results of the previous sections.
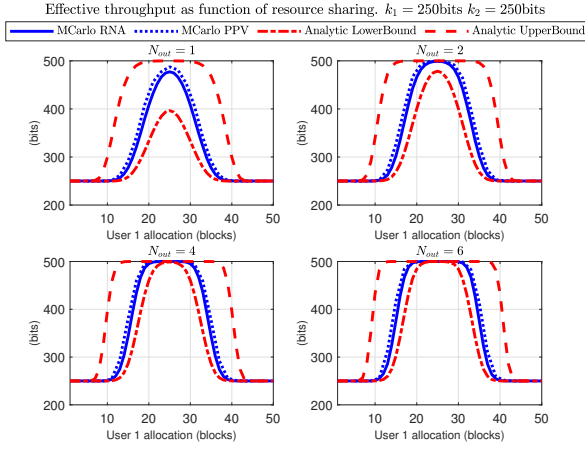
Fig. 4. Effective throughput as function of resource sharing for two users of the *same type*: ($k_1 = 250$bits, $\epsilon_1 = 10^{-5}$) and ($k_2 = 250$bits, $\epsilon_2 = 10^{-5}$). Total available blocks $N = 50$ with $n_c = 2 \times 12$ subcarriers per block. Rayleigh fading $\mu = 0$dB.
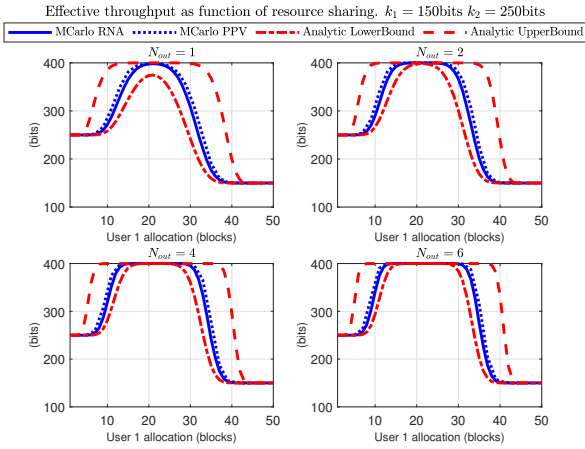

Fig. 5. Effective throughput as function of resource sharing for two users of *different types*: ($k_1 = 150$bits, $\epsilon_1 = 10^{-5}$) and ($k_2 = 250$bits, $\epsilon_2 = 10^{-5}$). Total available blocks $N = 50$ with $n_c = 2 \times 12$ subcarriers per block. Rayleigh fading $\mu = 0$dB.

Reliability context. As Ultra Reliability is usually linked with Low Latency constraint, we analyze the Reliability Confidence level using the latest results on block error rate in the Finite Blocklength regime. The analysis is carried out for Rayleigh slow frequency block-fading channels. We have first obtained lower and upper bounds on the Reliability Confidence level that are tight at high SNR, and then proposed their analytic approximations.

These approximations have been used to solve two optimization problems. The first one consists in finding the minimum codeword length required to meet a given target block-error-rate with a given confidence and the second one is characterizing the optimal resource sharing between two users in a typical 5G New Radio communication scenario. Solutions to both problems are obtained very fast with our closed-form approximate formulas, without the need of cumbersome Monte-Carlo simulations. In addition, the approximate solutions have been shown to accurately match the results predicted by well-known bounds in the two applications considered here.

REFERENCES

[1] *5G; Study on scenarios and requirements for next generation access technologies*, 3GPP Std. 38.913 v15.0.0, Sep. 2018.

[2] M. Angjelichinoski, K. F. Trillingsgaard, and P. Popovski, "A statistical learning approach to Ultra-Reliable Low Latency communication," *arXiv preprint arXiv:1809.05515*, Sep. 2018.

[3] M. Haenggi, "The meta distribution of the SIR in Poisson bipolar and cellular networks," *IEEE Trans. Wireless Comm.*, vol. 15, no. 4, pp. 2577–2589, 2016.

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

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

[6] *LTE; E-UTRA; Requirements for support of radio resource management*, 3GPP Std. 38.133 v15.4.0, Jan. 2019.

[7] *5G; NR; Requirements for support of radio resource management*, 3GPP Std. 38.133 v15.4.0, Dec. 2018.

[8] S. Sesia, I. Toufik, and M. Baker, "Radio Link Monitoring Performance," in *LTE - The UMTS Long Term Evolution: From Theory to Practice*. John Wiley & Sons, 2011, ch. 22, p. 525.

[9] *5G; NR; Radio Resource Control (RRC); Protocol specification*, 3GPP Std. 38.331 v15.3.0, Oct. 2018.

[10] S. Schiessl, H. Al-Zubaidy, M. Skoglund, and J. Gross, "Delay performance of wireless communications with imperfect CSI and finite-length coding," *IEEE Trans. Comm.*, vol. 66, no. 12, pp. 6527–6541, 2018.

[11] R. Devassy, G. Durisi, G. C. Ferrante, O. Simeone, and E. Uysal, "Reliable transmission of short packets through queues and noisy channels under latency and peak-age violation guarantees," *IEEE Journal on Selected Areas in Comm.*, 2019.

[12] J. Östman, R. Devassy, G. Durisi, and E. Uysal, "Peak-age violation guarantees for the transmission of short packets over fading channels," *arXiv preprint arXiv:1903.06771*, Mars 2019.

[13] C. She, C. Yang, and T. Q. Quek, "Cross-layer Transmission Design for Tactile Internet," in *2016 IEEE Global Comm. Conference (GLOBECOM)*. IEEE, 2016, pp. 1–6.

[14] M. C. Gursoy, "Throughput analysis of buffer-constrained wireless systems in the finite blocklength regime," *EURASIP Journal on Wireless Comm. Networking*, vol. 2013, no. 1, p. 290, 2013.

[15] *5G; Study on New Radio (NR) access technology*, 3GPP Std. 38.912 v15.0.0, Sep. 2018.

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

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

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

[19] V. Y. F. Tan and M. Tomamichel, "The third-order term in the normal approximation for the AWGN channel," *IEEE Trans. Info. Theory*, vol. 61, no. 5, pp. 2430–2438, 2015.

[20] M. R. McKay, P. J. Smith, H. A. Suraweera, and I. B. Collings, "On the mutual information distribution of OFDM-based spatial multiplexing: exact variance and outage approximation," *IEEE Trans. Info. Theory*, vol. 54, no. 7, pp. 3260–3278, 2008.

[21] P. Wu and N. Jindal, "Coding versus ARQ in fading channels: How reliable should the PHY be?" *IEEE Trans. Comm.*, vol. 59, no. 12, pp. 3363–3374, 2011.

[22] F. E. Harris, "Incomplete Bessel, generalized incomplete gamma, or leaky aquifer functions," *Journal of Computational and Applied Mathematics*, vol. 215, no. 1, pp. 260–269, 2008.