Ultra Reliable Communication via Optimum Power Allocation for HARQ Retransmission Schemes

Endrit Dosti, Mohammad Shehab, Hirley Alves and Matti Latva-aho

Abstract

In this work, we develop low complexity, optimal power allocation algorithms that would allow ultra reliable operation at any outage probability target with minimum power consumption in the finite blocklength regime by utilizing Karush-Kuhn-Tucker (KKT) conditions. In our setup, we assume that the transmitter does not know the channel state information (CSI). First, we show that achieving a very low packet outage probability by using an open loop setup requires extremely high power consumption. Thus, we resort to retransmission schemes as a solution, namely Automatic Repeat Request (ARQ), Chase Combining Hybrid ARQ (CC-HARQ) and Incremental Redundancy (IR) HARQ. Contrary to classical approaches, where it is optimal to allocate equal power with each transmission, we show that for operation in the ultra reliable regime (URR), the optimal strategy suggests transmission with incremental power in each round. Numerically, we evaluate the power gains of the proposed protocol. We show that the best power saving is given by IR-HARQ protocol. Further, we show that when compared to the one shot transmission, these protocols enable large average and maximum power gains. Finally, we show that the larger the number of transmissions is, the larger power gains will be attained.

I. INTRODUCTION

Modern communication systems play an essential role in everyone’s daily life. Throughout the history of the development of these systems, the goal has been to enable communication with
higher data rates. This trend is expected to be continued in the future as well. However, in the next generation systems, the vision will be to connect all devices that benefit from an Internet connection to create a data driven society and pave the road towards 6G [1], thus resulting in the creation of the Internet of Things (IoT). A key characteristic of the IoT is that most of the wireless connections are expected to be generated by the autonomous devices rather than by the human-operated terminals. To successfully implement this vision, wireless communication systems will have to support a much larger number of connected devices, and at the same time fulfill much more stringent requirements on latency and reliability that what current standards can guarantee [2].

For this purpose, 5G introduced at least two new operating modes. The first one is Massive Machine-to-Machine Communications, which is related to designing a wireless system that can support a large number of simultaneously connected devices (e.g. more than 10s of thousands). This will pave path for seamless and ubiquitous connectivity of heterogenous devices [3], [4]. The second operating mode is Ultra Reliable Low Latency Communications (URLLC). This mode refers to insuring a certain level of communication service almost 100% of the time, while satisfying very stringent delay requirements (i.e. in the order of 4 ms). [5]. URC is essential in many applications such as reliable wireless coordination among vehicles, reliable cloud connectivity, critical connections for industrial automation [6], [7].

To cope with the presence fading in wireless communication systems, several diversity schemes are implemented in order to create redundancy by transmitting the signal through several independent fading paths and then combining it accordingly at the receiver. This provides higher reliability at the cost of increased power consumption. In this context, the problem of optimum power allocation in URC scenarios has recently attracted lots of attention [8]–[10]. For example, in [9] the authors design an optimum power allocation algorithm to maximize the energy efficiency in Single-Input Multiple-Output (SIMO) systems with stringent reliability constraints. Moreover, the utilization of different retransmission schemes is a classical approach to provide larger diversity gains and achieve ultra-reliability. The most popular schemes are those which implement repetitive or parallel retransmission techniques [11]–[28]. In the former, the transmitter sends the same codewords in all the possible fading paths. While in the latter, the transmitter utilizes different and jointly designed codewords to construct the packets.

The simplest retransmission protocol which utilizes repetition coding, is Automatic Repeat Request (ARQ) [11], where transmitter sends the packets until it receives an acknowledgment
(ACK) from the receiver, or until the maximum allowed number of retransmissions is exhausted. A more robust retransmission protocol family that utilizes repetition coding is Hybrid-ARQ (HARQ) [12]. The difference is that in this family the receiver buffers all the packets and utilizes them to correctly decode the information. HARQ protocols are classified based on how the receiver combines the packets. For instance, when the receiver makes decisions based on the selected packet with the highest Signal-to-Noise Ratio (SNR), the scheme is called Selection Combining (SC) HARQ [13]. Secondly in Chase Combining (CC) HARQ [14], the receiver can do Maximal Ratio Combining (MRC) of the received signals and extract the information. All these schemes boost reliability at the cost of increased latency and complexity of the receiver design. In addition, several retransmission protocols utilize parallel coding schemes, where they utilize Incremental Redundancy (IR) [15]. Therein, the transmitter sends new information with each retransmission. This is achieved by splitting the parent codeword into several sub-codewords. The utilization of this family of retransmission protocols has been embraced by several systems, such as Worldwide Interoperability for Microwave Access (WiMAX), Long-Term Evolution (3GPP LTE), and 5G NR [16], [17].

A. Related Work

The problem of adaptive and optimal resource allocation in diversity and retransmission schemes has been investigated in several papers [9], [18]–[28]. Mainly, the optimization parameters are different performance metrics, such as outage probability, throughput, effective capacity, delay, power/energy efficiency etc. For instance, In [18] the authors derive closed form expressions for the rate-maximized throughput of ARQ for independent distributed Nakagami-\(m\) block fading interfering channels. The authors of [28] studied the trade-off between short packet transmission and utilizing HARQ in low latency communication. They showed that HARQ may significantly outperform finite blocklength for a given set of latency, reliability and bit count values. Energy-efficient adaptive power allocation for three incremental multiple-input-multiple-output (IMIMO) systems employing ARQ, CC-HARQ and IR-HARQ are considered in [19]. There, the authors formulate and provide closed form solutions for the proposed geometric programming problem (GPP) to minimize the rate outage probability. However, they did not consider the minimum power consumption problem which we target in our analysis.

An issue that has gathered much attention from the research community is power allocation between different retransmission rounds. In [20] the authors provide a closed form approximation
for the outage probability for CC-HARQ protocol. Furthermore, they formulate and solve the power allocation problem as a GPP. A limited power allocation strategy valid for a maximum of two transmissions for IR-HARQ protocol is proposed in [21]. In contrast to the work presented there, we derive closed form expression for the outage probability of IR-HARQ. Furthermore, we obtain optimal power allocation algorithms of low complexity for any number of transmissions and prove analytically that the proposed solutions are globally optimal. Also, unlike [20], [21] we provide comparisons between the performance of proposed power allocation algorithms for the three most popular retransmission protocols, respectively ARQ, CC-HARQ and IR-HARQ.

Moreover, the analysis in all the above mentioned papers is done under the assumption of asymptotically long codewords, where the length of metadata is much smaller than the actual data. However, for short packets, metadata and the actual data are almost of the same size; therefore the usage of conventional methods, such as capacity or ergodic capacity, is highly suboptimal [2]. Little work has been done for the short packets domain in the context of ARQ and short packets [24]. Herein, we show that our proposed algorithms are valid for finite length codewords. This occurs because when power levels are high enough, the dispersion term present in the maximum achievable rate looses dominance. Moreover, as shown in [29], whenever there is any type of CSI ¹, the gap to Shannon capacity that is observed due to finite blocklength is closed.

Further, in [25] the authors analyze the performance of ARQ protocol over the fading channel under very simplistic assumptions such as infinite number of transmissions, full buffer capacity, instantaneous and error free feedback. However, therein the authors do not investigate the impact of power allocation between different ARQ rounds. In [27], the authors develop a power allocation scheme for type-I ARQ protocol that minimizes the outage probability only for the case of two transmissions. However, in their scheme they do not guarantee a minimal outage probability level which would be essential in the case of URC.

B. Contributions

In this paper, we focus on Machine Type Communications (MTC) that have very stringent requirements on latency and reliability. We show that achieving ultra low latencies under the one

¹Herein we assume CSI at the Receiver side, i.e. CSIR which is a common assumption in URLLC literature and can ve obtained via channel estimation [2].
shot transmission setup would not be feasible, due to very large power consumption. Furthermore, we show that in medium and high SNR regime, the asymptotic approximation of the outage probability provides a good benchmark for further analysis.

To mitigate the large power expenditures associated with the one shot transmission, we suggest the implementation of retransmission schemes. We assume that it is possible to evaluate the delay associated with one transmission round and select the maximum number of transmissions so that the delay requirement of the system is satisfied. Specifically, we focus on the problem of optimal power allocation for repetitive and parallel IR-HARQ transmission schemes under block fading channel conditions. For this purpose, we cast an optimization problem to minimize the average power expenditure needed to meet a certain target outage probability. Herein, we extend our initial analysis presented in [22], [23] for ARQ and CC-HARQ protocols, to more complex schemes such as IR-HARQ. For this scheme, we first develop a closed form approximation of the outage probability. Next, by proving the convexity of our problem, we show that the obtained solutions are globally optimal.

The main contributions of the paper are as follows

- We develop power allocation algorithms of low (linear) complexity, which would allow us to achieve any target outage probability for repetition and parallel (IR-HARQ) retransmission protocols that would enable operation in the URR and the finite block-length regime at any number of transmissions.
- We show that the optimization problem is convex, and that the globally optimal power allocation strategy suggests a transmission with increasing power in each transmission round when ultra low outage probability values are required.
- We provide a closed-form approximation for the outage probability of the IR-HARQ protocol.
- We compare the power and efficiency of the algorithms that we have developed and show that the best power saving is achieved by IR-HARQ.

C. Outline

The remainder of this paper is organized as follows. In Section II, we introduce the system model and analyze the one shot transmission in the finite block-length regime. Section III

\[^{2}\text{Notice that part of this work was initially presented in IEEE ICC'17 [22] and EUCNC'17 [23].}\]
presents the optimal power allocation algorithms for both repetitive and IR-HARQ retransmission schemes. In Section IV, we provide comparisons for the performance of the proposed algorithms through some illustrative numerical results. Finally, the main conclusions are summarized in Section V.

II. MAXIMUM CODING RATE IN FINITE BLOCKLENGTH

A. System Model

Assume a transmitter-receiver pair communicating under block-fading channel. As in [30], at the transmitter side, $b_1, b_2, \ldots, b_K$ nats\(^3\) are encoded in $c_1, c_2, \ldots, c_n$. Next, these encoded nats are interleaved and mapped to a constellation $X$. This results in the stream of modulated symbols $x_1, x_2, \ldots, x_n$. For simplicity, we assume that we map one modulated symbol per channel use. Here, $K$ and $n$ denote the number of information nats and the number of channel uses, respectively. This results in the creation of the packets that will be transmitted. The receiver then fetches the packets and tries to recover the information.

While communicating, the pair can either utilize an open loop setup, or a retransmission protocol. In the second case, the maximum number of transmissions is set to $M$. Whenever the transmitter fetches a negative acknowledgment (NACK) packet, it retransmits based on the protocol that is implemented. It will stop sending packets if it receives an ACK message from the receiver, or if the maximum number of allowed transmissions has been completed.

We consider quasi-static fading channel conditions, in which the channel gain $h$ remains constant for the duration of one packet transmission and changes independently between all the transmission rounds. The motivation is that URLLC and MTC devices usually communicate on short packets where the channel fading coefficient is almost constant through the packet duration [8]. This assumption simplifies the analysis and serves the goal of obtaining benchmark power allocation for the problem addressed. Moreover, it has been widely accommodated in the literature such as in [2]. We analyze the case when the channel coefficient is Rayleigh distributed and $h \sim CN(0, 1)$. Thus the squared envelope of the channel gain is exponentially distributed with mean one. For simplicity we denote $f_{|h|^2}(z) = e^{-z}$. We assume that the receiver has channel state information while the transmitter knows only the distribution of the channel gains and the

\(^3\)To standardize the notation, hereafter we assume that all information is encoded in nats instead of bits. Therefore, all log is the natural logarithm.
information it obtains from the feedback. Then the received signal at the $m^{th}$ round can be written as

$$ y_m = \sqrt{\rho_m} h_m x_m + w_m, $$

where $x_m$ is the transmitted signal and $w_m$ is the AWGN noise term with noise power $N_0 = 1$. The term $\rho_m$ is the packet transmitted power. Since the variance of the noise is set to unity, it corresponds to the signal-to-noise ratio (SNR).

B. One shot transmission

In this section, we briefly summarize the recent results in the characterization of the maximum channel coding rate and outage probability in the finite block-length regime. Further, we evaluate the case of the open loop setup.

For notational convenience we need to define an $(n, K, \rho, \epsilon)$ code as a collection of

- An encoder $F : \{1, \ldots, K\} \mapsto C^n$ which maps the message $k \in \{1, \ldots, K\}$ into an $n$-length codeword $c_i \in \{c_1, \ldots, c_n\}$ such that the following power constraint ($\rho$) is satisfied:
  $$ \frac{1}{n} \|c_i\|^2 \leq \rho, \forall i. \tag{2} $$

- A decoder $G : C^n \mapsto \{1, \ldots, K\}$ that satisfies the maximum error probability ($\epsilon$) constraint:
  $$ \max_{\forall i} \Pr[G(y) \neq I|I = i] \leq \epsilon, \tag{3} $$

where $y$ is the channel output induced by the transmitted codeword according to (1).

The maximum achievable rate of the code is defined as [30]

$$ R^*_{\text{max}}(n, \rho, \epsilon) = \sup \left\{ \frac{\log K}{n} : \exists (n, K, \rho, \epsilon) \text{ code} \right\}. \tag{4} $$

For the AWGN channel non-asymptotic lower and upper bounds on the maximum achievable rate have been derived in [31]. Recently, a tight approximation for $R^*_{\text{max}}(n, \rho, \epsilon)$ has been proposed for sufficiently large number of channel uses (i.e. $n > 100$) in the case of the quasi-static fading channel [30] and is given by

$$ R^*_{\text{max}}(n, \rho, \epsilon) \approx C_\epsilon + O\left(\frac{\log n}{n}\right), \tag{5} $$

where $C_\epsilon$ is the outage capacity:

$$ C_\epsilon = \sup\{R : \Pr[\log(1 + \rho \cdot z) < R] < \epsilon\}. \tag{6} $$
Then, by a channel coding rate of $R = \frac{K}{n}$ nats per channel use (ncpu), where $K$ is the information payload, the outage probability is approximated as [32]

$$
\epsilon(n, R, \rho) \approx E_z \left[ Q \left( \frac{C(\rho \cdot z) - \frac{K}{n}}{\sqrt{V(\rho \cdot z)}} \right) \right]
$$

$$
\approx \int_0^\infty e^{-z} Q \left( \frac{C(\rho \cdot z) - \frac{K}{n}}{\sqrt{V(\rho \cdot z)}} \right) dz,
$$

(7)

where $E[\cdot]$ denotes the expectation over the channel gain $z$, $Q(\cdot)$ denotes the Gaussian Q-function, $C(x) = \log(1 + x)$ denotes the channel capacity and the channel dispersion is computed as $V(x) = 1 - \frac{1}{(1+x)^2}$. However the integral in (8) does not have a closed form solution. Thus, we resort to an approximated closed-form expression as in [26]

$$
\epsilon(n, R, \rho) = 1 - \frac{\delta}{\sqrt{2\pi}} e^{-\kappa} \left( e^{\sqrt{2\delta z}} - e^{-\sqrt{2\delta z}} \right),
$$

(9)

where $\kappa = \frac{e^{R-1}}{\rho}$ and $\delta = \sqrt{\frac{n\rho^2}{e^{2R-1}} - 1}$. Note that (9) characterizes the outage probability of a single ARQ round.

Fig. 1 illustrates the outage probability for the open loop setup, where the message is conveyed in a single transmission, for different channel coding rates. We have fixed the number of channel uses $n = 200$ and analyzed the case of mapping $K \in \{200, 400, 600\}$ information nats. This results in the channel coding rates $R = 1$, $R = 2$ and $R = 3$ ncpu, respectively. We can see that the integral form in (8) and the closed-form approximation in (9) match well for all the coding rates.

Furthermore, in Fig. 1 we also illustrate the performance of the asymptotic approximation (which we derive next and is given by (11)). The results show that for $\epsilon < 10^{-2}$ and in the ultra reliable region (URR), where very low outages are required (i.e., $\epsilon < 10^{-5}$), the asymptotic approximation can be used. This is due to the fact that in high SNR the maximum achievable rate (5) converges to the one with asymptotically long codewords $R^*(n, \epsilon) = C_\epsilon$, where $C_\epsilon$ is defined in (6). However, the amount of power required to achieve ultra reliability for this setup is quite high (> 50 dB). Herein, when $\rho \to \infty$ the outage probability in the $m^{th}$ round can be calculated as:

$$
\epsilon_m = \Pr[\log(1 + \rho z) < R] = 1 - e^{-\frac{e^{R-1}}{\rho m}}.
$$

(10)
The equality in (10) holds for Rayleigh fading channels. Furthermore, by using the first order of Taylor expansion $e^{-x} \approx 1 - x$ we can express the asymptotic approximation for the outage probability of the $m^{th}$ round as:

$$\epsilon_m = \frac{\phi}{\rho_m},$$

where

$$\phi = e^R - 1.$$

For ultra reliable communications, we require to have an outage probability $\epsilon$ very low, while spending as little power as possible. However, Fig. 1 shows that such low outage values are highly unlikely to be obtained when using an open loop setup. Thus, we investigate the possibility of utilizing retransmission mechanisms with optimal power allocation, in order to obtain an outage probability in the ultra reliable region.

### III. Optimal Power Allocation

In this section we evaluate the impact of repetitive and IR-HARQ retransmission schemes in the outage probability. Specifically, we focus on the analysis of ARQ, CC-HARQ and IR-HARQ protocols. We propose an optimal power allocation scheme that allows us to reach any target outage probability, assuming that we can have up to $M$-transmissions for each of the protocols.
The problem of interest is to achieve a target outage probability while spending as little power as possible for conveying the information from the transmitter to the receiver. Since we will have multiple transmissions, one approach would be to allocate equal power in each round. This implies that given a certain power budget $\rho_{\text{budget}}$, the transmit power in the $m^{th}$ round would be $\rho_m = \frac{\rho_{\text{budget}}}{M}$. However, such simplistic approach is shown to be highly inefficient for very low outage probability values [19], [27]. Thus, we propose a power allocation algorithm in order to minimize the average transmit power of the transmitter which allocates different power levels in each retransmission round. Bearing this in mind, the average transmitted power can be defined as

$$\rho_{\text{avg}} = \sum_{m=1}^{M} \rho_m E_{m-1},$$  \hspace{1cm} (13)

where $M$ is the maximum number of retransmission rounds, $\rho_m$ is the power transmitted in the $m^{th}$ round and $E_{m-1}$ is the outage probability up to the $m - 1$ round. Next, we calculate the probability that the packet can not be decoded correctly even after the maximum allowed number of retransmissions. We refer to this as packet drop probability (pdp), and it corresponds to the outage probability up to the $M^{th}$ round ($E_M$). Since we have assumed that all the transmissions of the packets experience independent fading conditions, we can express $E_M$ as

$$E_M = \prod_{m=1}^{M} \epsilon_m,$$  \hspace{1cm} (14)

where $\epsilon_m$ is the outage probability of the of the $m^{th}$ ARQ round and can be computed by (9), or asymptotically via (11). The outage probability before the first transmission, $\epsilon_0 = 1$.

Based on our system model introduced in Section II-A, we now formulate the main problem of this paper as follows:

**Problem 1.** The optimal power allocation strategy for repetitive and IR-HARQ retransmission schemes is the solution of

$$\min_{\rho_m} \rho_{\text{avg}}$$

s.t.  \hspace{0.5cm} $0 \leq \rho_m, \ 1 \leq m \leq M$

$$E_M \leq \epsilon$$

where $\epsilon$ is any target outage probability.
To obtain a globally optimal analytical solution for this problem we first verify its convexity, since it is a necessary and sufficient condition to utilize Karush-Kuhn-Tucker (KKT) conditions [33]. To do so, we propose the following lemma.

**Lemma 1.** The optimization problem (15) is convex for the protocols that will be analyzed throughout this paper.

*Proof.* Please see Appendix A.

Now, we can write the Karush-Kuhn-Tucker (KKT) conditions to obtain the optimal power allocation strategy for the convex problem (15). First, we write the Lagrangian function as

$$
\mathcal{L}(\rho_m, \mu_m, \lambda) = \sum_{m=1}^{M} \rho_mE_{m-1} + \sum_{m=1}^{M} \mu_m\rho_m + \lambda(E_M - \epsilon),
$$

(16)

where $\mu_m$ for $m = 1, \ldots, M$ and $\lambda$ are the Lagrangian multipliers. Furthermore, we express the KKT conditions as follows:

* C1 $\frac{\partial \mathcal{L}}{\partial \rho_m} = 0, \ m = 1, \ldots, M$,

* C2 $\mu_m \geq 0, \ m = 1, \ldots, M$,

* C3 $\mu_m\rho_m = 0, \ m = 1, \ldots, M$,

* C4 $E_M - \epsilon = 0$.

Note that the target is to minimize the transmit power. Thus, it is straightforward to infer that the reliability constraint is optimally achieved at equality since more power is needed to achieve lower error and higher reliability. Hereafter, we begin the solution of (15) for the repetition and IR-HARQ retransmission schemes.

**A. Repetitive retransmission schemes**

In repetitive retransmission schemes the transmitter sends the same information nats in each round. Here, we present the power allocation algorithms for two retransmission protocols that utilize repetitive schemes, respectively ARQ and CC-HARQ. Furthermore, we discuss about the main differences between them.

1) **ARQ:** Its principle is shown in Figure. 2. The transmitter sends the whole packet in each transmission round and stops when the maximum number of transmissions $M$ has been achieved, or when it receives confirmation that the packet has been successfully decoded. The receiver
makes the decisions only based on the last packet he has received, and discards the earlier packets.

For this protocol, the optimal power terms can be obtained recursively backward using the lagrangian $\lambda$ as described in Theorem 1.

**Theorem 1.** The optimal power terms for the ARQ protocol are

$$\rho_M = \sqrt{M\lambda\phi},$$  \hfill (17)  

$$\rho_m = \sqrt{2\phi\rho_{m+1}}, \quad 1 \leq m < M,$$  \hfill (18)  

where $\phi$ and $\lambda$ are given from (12) and (34) respectively.

**Proof.** The outage probability for the $m^{th}$ ARQ round and can be computed by (9), or asymptotically via (11). We start the solution for problem (15) by writing the Lagrangian function, which is computed as

$$L(\rho_m, \mu_m, \lambda) = \sum_{m=1}^{M} \rho_m E_{m-1}^{ARQ} + \sum_{m=1}^{M} \mu_m \rho_m + \lambda(E_M^{ARQ} - \epsilon).$$  \hfill (19)  

Next we analyze the KKT conditions. From C1, we write the derivative of the Lagrangian function $L(\rho_m, \mu_m, \lambda)$ with respect to the power $\rho_m$ as

$$\frac{\partial L(\rho_m, \mu_m, \lambda)}{\partial \rho_m} = \left( \frac{\phi^{m-1}}{\prod_{i=1}^{m-1} \rho_i} - \sum_{i=1}^{M-m} \frac{\rho_{m+i}\phi^{m+i-1}}{\rho_m \prod_{j=1,j\neq i}^{m+i-1} \rho_j} \right) - \mu_m - \lambda \left( \frac{\phi^M}{\rho_m^2 \prod_{i=1,i\neq m}^M \rho_i} \right).$$  \hfill (20)  

![Fig. 2. The setup for ARQ protocol. The transmitter sends the first packet and waits for the ACK from the receiver. Then it sends packet 2. If the receiver can not decode the packet, it discards the packet and asks for retransmission.](image)
We relax temporarily the non-negative condition on the transmitted power terms $\rho_m$. This implies that, $\mu_m = 0$ for $m = 1, ..., M$. Now, using (20) and $\mu_M = 0$, we can write \( C1 \) for $m = M$ as

\[
\frac{\partial \mathcal{L}(\rho_m, \mu_m, \lambda)}{\partial \rho_M} = \frac{\phi_{M-1}}{\prod_{i=1}^{M-1} \rho_i} - \lambda \left( \frac{\phi_M}{\rho_M^2 \prod_{i=1, i \neq M}^{M} \rho_i} \right) = 0. \tag{21}
\]

After some algebraic manipulations of (21) we obtain the transmit power at the $M^{th}$ ARQ round as in (17). Similarly, by substituting $m = M - 1$ in (20) and using $\mu_{M-1} = 0$, we can rewrite \( C1 \) for $m = M - 1$ as

\[
\frac{\partial \mathcal{L}(\rho_m, \mu_m, \lambda)}{\partial \rho_{M-1}} = \left( \frac{\phi_{M-2}}{\prod_{i=1}^{M-2} \rho_i} - \frac{\rho_M \phi_{M-1}}{\rho_{M-1}^2 \prod_{i=1}^{M-2} \rho_i} \right) - \lambda \left( \frac{\phi_M}{\rho_{M-1}^2 \prod_{i=1, i \neq M-1}^{M} \rho_i} \right) = 0, \tag{22}
\]

which can be simplified, yielding

\[
\rho_{M-1} = \sqrt{\rho_M \phi + \frac{M \lambda \phi^2}{\rho_M}}. \tag{23}
\]

Following a similar approach, we obtain the following relationship for the case of $m = M - 2$

\[
\rho_{M-2} = \sqrt{\rho_{M-1} \phi + \frac{\rho_M \phi^2}{\rho_{M-1}} + \frac{M \lambda \phi^3}{\rho_{M-1} \rho_M}}. \tag{24}
\]

We can continue this procedure for all $m \in \{1, \ldots, M\}$ and the results can be summarized as:

\[
\rho_M = f(\lambda), \tag{25}
\]

\[
\rho_{M-1} = f(\lambda, \rho_M), \tag{26}
\]

\[
\rho_{M-2} = f(\lambda, \rho_M, \rho_{M-1}), \tag{27}
\]

\[\vdots\]

\[
\rho_1 = f(\lambda, \rho_M, \ldots, \rho_3, \rho_2). \tag{28}
\]

At this point, we can easily verify now that the obtained power values are positive and $\rho_m$ cannot be further minimized. By utilizing a method that is similar to the backward substitution approach [33, App. C.2], we can obtain a relationship between the power terms $\rho_m$ as follows:

i) by substituting $M \lambda \phi = \rho_M^2$ (see (17)) in (23) (or equivalently in (26)) we evaluate $\rho_{M-1}$ as $\rho_{M-1} = \sqrt{2 \phi \rho_M}$. ii) $\rho_{M-2}$ is evaluated by substituting $\sqrt{M \lambda \phi} = \rho_M$ and $\sqrt{2 \phi \rho_M} = \rho_{M-1}$ in (27). iii) By continuing this procedure we can express the optimal transmit power in the $m^{th}$ round as in (18).

Based on (18), we can easily verify now that the obtained power values $\rho_m$ are all positive. Further, since $\rho_M$ is a function of $\lambda$ (see (17)) and using (18), it is clear that each $\rho_m$ is a function
of $\lambda$. Thus, all that remains is to compute the Lagrangian multiplier $\lambda$. For this purpose, we utilize the outage constraint in (15) (C4). First, we substitute $\rho_m$ for $m = 1, \ldots, M$ in (14) to obtain $E_{ARQ}^M$ as

$$E_{ARQ}^M = \frac{\phi^M}{\prod_{m=1}^{M} \rho_m} = \epsilon,$$

where $\rho_m$ is given by

$$\rho_m = \sqrt{2^{a(m)}\phi^{b(m)}(M\lambda)^{c(m)}}.$$

In (30), we can compute the exponents

$$a(m) = 2 - 2^{-(M - m - 1)},$$

$$b(m) = 2 - 2^{-(M - m)},$$

$$c(m) = 2^{-(M - m)}.$$

Finally, we compute $\lambda$ by equating $E_{ARQ}^M$ to the outage target $\epsilon$ based on C4:

$$\lambda = \left( \frac{\phi^{o(M)}}{M^{o(M)}\epsilon^{p(M)}2^{q(M)}} \right)^{\frac{1}{a(M)}},$$

where

$$o(m) = 2^m - 1,$$

$$p(m) = 2^m,$$

$$q(m) = [(m - 2)2^m + 2].$$

Finally, we utilize the results presented in Theorem 1 and propose the following power allocation algorithm for ARQ protocol.

From Theorem 1, it is obvious that the analytical solution to obtain the optimal power allocation scheme for a large number of retransmissions would become a cumbersome task. Furthermore, in general the ultra reliable systems are delay-limited. For this purpose, in Appendix B, we propose a simpler analytical solution which is valid only for the case of $M = 2$ transmissions.

Another important observation we can note from Theorem 1 is the following:
Algorithm 1 Power allocation for ARQ

1: Inputs: $\phi$, $M$.

2: Compute: $a(m)$ as in (31).

3: Compute: $b(m)$ as in (32).

4: Compute: $c(m)$ as in (33).

5: Compute: $o(M)$ as in (35).

6: Compute: $p(M)$ as in (36).

7: Compute: $q(M)$ as in (37).

8: Compute: $\lambda$ as in (34).

9: Compute: $\rho_M$ as in (17).

10: while all power terms are not found do

11: Compute: $\rho_m$ as in (18).

12: Decrease: $m$ by 1.

13: end while.

14: Outputs: All the power terms $\rho_m$.

Theorem 2. In the ultra reliable region, the optimal power allocation strategy suggests a transmission with increasing power in each ARQ round.

Proof. Please refer to Appendix C.

Notice that this result fully matches the intuition. Presuming that the delay requirements are met (by setting $M$ accordingly), our goal is to achieve a target outage probability by spending as little power as possible. For this purpose, we transmit first with low power. If the channel conditions are good then the transmission will be successful, and a large amount of power is saved (as will be pointed out later in Figure 5). If it fails, then retransmissions are carried out until an ACK is received, or the maximum allowed number of transmissions is reached.

2) CC-HARQ: Another member of the repetitive retransmission schemes is CC-HARQ protocol. This protocol is widely implemented in several standards (i.e. HSDPA, LTE). Its principle is illustrated in Figure 3. Again, the entire packet is transmitted in each round. However, unlike the previous protocol we analyzed, all the received packets are buffered and the receiver performs MRC to enhance the SNR. When compared to the ARQ protocol, besides the diversity gains, this approach also provides combining gains.
Fig. 3. The setup for CC-HARQ protocol. The transmitter sends first packet 1 and waits for the ACK from the receiver. Then it sends packet 2. If the receiver cannot decode the packet, it buffers the packet and asks for retransmission. Then, when it receives packet 2 once more, it combines it with the buffered packet to extract the information.

For this protocol, in [20] the authors have derived a very tight approximation for $E_M$ as

$$E_M^{CC-HARQ} = \frac{\phi^M}{M! \prod_{m=1}^{M} \rho_m}.$$  

(38)

Bearing this in mind, we next introduce the optimal power terms for CC-HARQ in the following theorem.

**Theorem 3.** For the CC-HARQ protocol, we can find the optimal power terms are

$$\rho_M = \sqrt{\phi \lambda},$$  

(39)

$$\rho_m = \sqrt{\frac{2\phi \rho_{m+1}}{m}},$$  

(40)

where $\phi$ and $\lambda$ are given from (12) and (34) respectively.

**Proof.** Please refer to Appendix D.

Finally, for CC-HARQ protocol we propose the following power allocation algorithm.

**B. Parallel retransmission schemes**

Here, we evaluate the effect of implementing parallel (IR-HARQ) retransmission schemes, wherein the transmitter sends different and jointly designed packets for each message. When compared to the repetitive retransmission schemes discussed in Section III-A, besides the diversity and combining gains, this approach provides also coding gains.
Algorithm 2 Power allocation for CC-HARQ

1: **Inputs:** $\phi, M$.

2: **Compute:** $a(m)$ as in (31).

3: **Compute:** $b(m)$ as in (32).

4: **Compute:** $c(m)$ as in (33).

5: **Compute:** $o(M)$ as in (35).

6: **Compute:** $p(M)$ as in (36).

7: **Compute:** $q(M)$ as in (37).

8: **Compute:** $\lambda$ as in (64).

9: **Compute:** $\rho_M$ as in (58).

10: **while** all power terms are not found **do**

11: **Compute:** $\rho_m$ as in (40).

12: **Decrease:** $m$ by 1.

13: **end while**.

14: **Outputs:** All the power terms $\rho_m$.

The final retransmission protocol that we analyze in this paper is IR-HARQ. Its operating principle is shown in Fig. 4, where we note that the transmitter splits the parent codeword into
$M$ sub-codewords of equal length. Each of these codewords is transmitted during one round. The receiver buffers all the received packets and utilizes them to recover the information. For Gaussian codebook, the mutual information that is gathered would be

$$I = \frac{1}{M} \sum_{m=1}^{M} \log (1 + \rho_m z_m).$$  \hspace{1cm} (41)

After making some manipulations we can compute the outage probability as

$$E_{IR-HARQ}^{M} = \Pr \left[ \sum_{m=1}^{M} \log (1 + \rho_m z_m) < M \tilde{R} \right].$$ \hspace{1cm} (42)

In [34] the author proves a theorem which can be used to approximate (42). There, he provides an integral form approximation of the outage probability when equal power is allocated in each retransmission round. Herein, we extend those results and provide closed-form approximation for the outage probability when different power levels are allocated in each IR-HARQ round. It can be computed as given in the following lemma.

**Lemma 2.** For the proposed setup, the outage probability of the IR-HARQ protocol can be approximated as

$$E_{IR-HARQ}^{M} \approx \frac{\psi_M}{\prod_{m=1}^{M} \rho_m}. \hspace{1cm} (43)$$

**Proof.** Please refer to Appendix E. \hfill \Box

At this point, all that remains is to obtain the optimal power allocation strategy. For this purpose, we introduce the following theorem.

**Theorem 4.** For the IR-HARQ protocol, the optimal power terms are

$$\rho_M = \sqrt{\frac{\lambda M \psi_M}{\psi_{M-1}}} \hspace{1cm} (44)$$

$$\rho_m = \sqrt{\frac{2 \rho_{m+1} \psi_m}{\psi_{m-1}}}. \hspace{1cm} (45)$$

**Proof.** Please refer to Appendix F. \hfill \Box

Finally, we can utilize the results in Theorem 4 and formulate the optimal power allocation algorithm.
Algorithm 3 Power allocation for IR-HARQ

1: **Inputs:** $R$, $M$.

2: **while** all values of $\psi_m(R)$ are not found **do**

3: **Compute:** $\psi_m$ as in (72).

4: **Increase:** $m$ by 1.

5: **end while.**

6: **Compute:** $a(m)$ as in (31).

7: **Compute:** $c(m)$ as in (33).

8: **Compute:** $d(i) = 2^{-i}$.

9: **Compute:** $o(M)$ as in (35).

10: **Compute:** $p(M)$ as in (36).

11: **Compute:** $q(M)$ as in (37).

12: **Compute:** $\lambda$ as in (81).

13: **Compute:** $\rho_M$ as in (75).

14: **while** all power terms are not found **do**

15: **Compute:** $\rho_m$ as in (45).

16: **Decrease:** $m$ by 1.

17: **end while.**

18: **Outputs:** All the power terms $\rho_m$.

IV. Numerical analysis

In this section we provide further results for the performance of the proposed optimal power allocation algorithms. First, we analyze the behavior of the power terms that will be transmitted in each round as a function of the target outage probability. Further, we show that the utilization of the power allocation algorithms provides on average, large gains when compared to the open loop setup. Also, we show that as the maximum number of transmission increases, the average power that is spent decreases. Finally, we evaluate the gains of the proposed power allocation algorithms under the assumption of maximum power expenditure.

In Fig. 5 we illustrate the variation of transmit power $\rho_m$ in each round versus the outage probability target $\epsilon$ for the case when we have a maximum of two transmissions. The channel coding rate is set to $R = 1$ ncpu. The results are obtained by implementing the protocols derived
in sections III-A and III-B. The first observation we can make is that the IR-HARQ protocol gives the best performance in terms of saving power. Further, we notice that despite the protocol that is implemented both power terms are lower than the open loop transmission which is shown in Fig. 1. Moreover, if the first round is successful (i.e. when the channel conditions are good), then the power gain with respect to the open loop setup would be very large. Approximately, we save $30 - 35$ dB (depending on the protocol) for $\epsilon = 10^{-5}$, which corresponds to the start of the ultra reliable region. Then, the more stringent the reliability requirements are, the more power we save with respect to the open loop setup. We observe that in the URR, the first power term is lower than the second power term. Notice that this result is coherent with what we obtained analytically in Theorem 2.

Fig. 6 illustrates the behaviour of the power terms as a function of the number of channel uses when $M = 2$. Here we set the number of information nats $K$ to 200 and 300 and the target error probability to $10^{-5}$.

First we observe that both power terms decrease as we increase the number of channel uses. Secondly we notice that when we increase the coding rate, we have to transmit with higher power in each transmission round. Moreover, the figure verifies that IR-HARQ is the most energy efficient scheme as it consumes the least power.

Next, in Fig. 7 we evaluate the average power as in (13) for each proposed scheme and for the scenario when the number of transmissions is increased. To attain this figure, we set $M = 2, 3$

\footnote{Note that 3GPP defines URLLC requirements for a payload of 32 bytes and 99.999% reliability.}
and $R = 1$ ncpu. From it, we notice that the amount of power that is saved per transmission on average is significant, especially when compared to the open loop setup. Further, by comparing the result of Figure 7 with the results in Figure 5, we notice that as we increase the number of transmissions, we save more power on average. For example, the average power consumption per transmission to achieve an error of $10^{-7}$ using 3 transmissions via IR-HARQ is 10 dB where the total power would be 15 dB. This power consumption level is very low compared to the power hungry 70 dB open loop setup in this case.

Fig. 8 evaluates the maximum power expenditure of our protocols in the case of $M \in \{2, 3\}$ transmissions and fixed channel coding rate $R = 1$ ncpu. To obtain the plot, we assume to have a worst case scenario, where all the transmissions are exhausted. From it, we observe that the proposed algorithms again allow us to save power when compared to the open loop setup. The largest power gains, are again given from the IR-HARQ protocol. Notice that when $M = 3$, we can save over 20 dB by implementing this protocol. Furthermore, we observe that as we increase the number of transmissions, power consumption is reduced. This happens due to the fact that the diversity and combining gains become higher. Also, notice that in the case of ARQ protocol, the power gain when the number of transmissions increases is not as large as the other two protocols. This happens due to the fact that ARQ does not benefit from combining gains.
Fig. 7. Average power required to achieve a target outage probability for rate $R = 1$ ncpu when the maximum number of transmissions is fixed to $M = 2$ and $M = 3$.

Fig. 8. Maximum power that will be spent to achieve a target outage probability for rate $R = 1$ ncpu when the maximum number of transmissions is fixed to $M = 2$ and $M = 3$. 
V. Conclusions and future work

In this work, we showed that operation in the URR would not be feasible under the open loop transmission setup, since very large powers are required to achieve ultra-reliability. For this reason, repetitive and parallel retransmission schemes can be implemented. Specifically, we analyzed three popular protocols that embrace these schemes, such as ARQ, CC-HARQ and IR-HARQ. In the case of IR-HARQ, we proposed a closed form approximation for the outage probability, which was later utilized in our derivations. For all three protocols, we proposed globally optimal power allocation algorithms with low complexity, that guarantee operation anywhere in the ultra reliable region while spending minimal power. We showed that the optimal power allocation strategy to operate in the URR suggests transmission with incremental power in each round. Furthermore, we showed that the best power saving is attained by the IR-HARQ.

It is obvious that when comparing IR-HARQ and repetitive retransmission schemes, there is the classical trade-off between the system complexity and performance. For example, in applications in which the block-length is very small (e.g. $n \approx 100$), repetitive protocols would be more suitable, since the utilization of IR-HARQ would reduce even more the blocklength which may lead to high outage, and thus larger number of transmissions. Besides, the existing mathematical framework for such short blocklengths does not hold ($n > 100$ [31]), thus the need for tractable and tighter approximation and bounds for this region. This is further reinforced by the fact that the nodes of such systems generally are very simple sensors. Implementing complex coding and logic in them, would result in a large increase of their cost and would severely affect their battery lifetime, which in these applications corresponds to the device lifetime. However, in other applications the data volumes to be transmitted are far larger, and the packet sizes can be larger (e.g. $n \approx 500$). Generally, the nodes communicating in these systems can afford extra complexity. Therefore, in these type of applications the utilization of IR-HARQ scheme would provide a better performance.

As future work, we intend to analyze the optimal power allocation scheme when there are limitations on the maximum power expenditure. Furthermore, it would be interesting to analyze how the utilization of multi-antenna systems would affect the power expenditure.
APPENDIX A

PROOF OF LEMMA 1.

A function is convex if both the objective function and the constraint set are convex. Based on (11), (14), (38) and (43) it is straightforward to show that proving the convexity of our problem reduces to proving the convexity of $f(\xi_1, \xi_2, \ldots, \xi_M) = \frac{1}{\xi_1 \xi_2 \cdots \xi_M}$. Notice that $f$ can be written as a composition of two functions

$$h(y) = \frac{1}{y}, \quad (46)$$

$$g(\xi_1, \xi_2, \ldots, \xi_M) = \xi_1 \xi_2 \cdots \xi_M, \text{where } \xi \in R^M. \quad (47)$$

The function $g$ is concave in $R^M$ and $h$ is convex decreasing function for $y \in R^+$. Next, we analyze the extension value extendibility $\tilde{h}$ of $h$. For this purpose, since $h$ is a convex function, we assign the value $\infty$ to all the points that are not in the domain of $h$. Note that

$$\lim_{z \to 0^-} h = \infty \geq \lim_{z \to 0^+} h, \quad (48)$$

which implies that $\tilde{h}$ is nonincreasing. Therefore, $f$ is a composition of a convex function $h$ with nonincreasing $\tilde{h}$, and a concave function $g$. Based on the composition rules [33, Ch.3 §3.11], $f$ will be a convex function in $R^M$.

APPENDIX B

A SIMPLIFIED SCENARIO FOR TWO TRANSMISSIONS.

For the scenario of $M = 2$ transmissions the optimization problem (15) simplifies to

minimize $\frac{1}{2}(\rho_1 + \rho_2 \epsilon_1)$

subject to $\frac{\phi^2}{\rho_1 \rho_2} = \epsilon \quad (49)$

To solve problem (49) we can utilize the procedure described in Section III.A. First, by using (17) and (18) we compute the power terms as functions of $\lambda$. Next, by substituting these expressions for $\rho_1$ and $\rho_2$ in C4 and solving for $\lambda$ we obtain $\lambda = \frac{\phi}{2\epsilon} \sqrt{\frac{1}{4\epsilon}}$. Finally, we compute the values of the power terms as $\rho_1 = \frac{\phi}{\epsilon} \sqrt{\frac{7}{2}}$ and $\rho_2 = \frac{\phi}{\epsilon} \sqrt{\frac{3}{2}}$.

For this specific case of $M = 2$ we can also utilize the following simpler approach to find the optimal power allocation. First, we rewrite the equality constraint as $\rho_2 = \frac{\phi^2}{\rho_1 \epsilon}$. Next, by substituting $\rho_2$ in the objective function of (49), we obtain an unconstrained optimization problem
with variable $\rho_1$. Then, we compute $\rho_1$ by setting the first derivative of the new objective function to zero as

$$\frac{1}{2} - \frac{2\phi^3}{\rho_1^3 \epsilon} = 0. \quad (50)$$

By solving (50) we find $\rho_1 = \phi \sqrt[3]{\frac{2}{\epsilon}}$, which is same as what we obtained by using the procedure described in Section III.A. Then, after substituting the first power term equation in the rewritten equality constraint we compute $\rho_2 = \frac{\phi}{\epsilon} \sqrt[3]{\frac{2}{\epsilon}}$.

As pointed out in Section III-A1, the procedure for CC-HARQ and IR-HARQ will follow a similar pattern.

**APPENDIX C**

**PROOF OF THEOREM 2.**

From (30) we can write $\rho_m$ and $\rho_{m+1}$ as:

$$\sqrt{2^{a(m)} \phi^b(m) (M\lambda)^c(m)} < \sqrt{2^{a(m+1)} \phi^b(m+1) (M\lambda)^c(m+1)}. \quad (51)$$

Next, it follows that

$$2^{a(m)} \phi^b(m) (M\lambda)^c(m) < 2^{a(m+1)} \phi^b(m+1) (M\lambda)^c(m+1), \quad (52)$$

$$\frac{2^{a(m)} \phi^b(m) M^c(m)}{2^{a(m+1)} \phi^b(m+1) M^c(m+1)} < \lambda^{c(m+1)} \lambda^{c(m)}, \quad (53)$$

$$2^{a(m)-a(m+1)} \phi^b(m)-b(m+1) M^{c(m)-c(m+1)} < \lambda^{c(m+1)-c(m)}. \quad (54)$$

In (54) the exponent of $\lambda$ is a positive number. To prove that, we can show that $c(m+1)-c(m) > 0$, which leads to

$$2 - 2^{-M+m+1} < 2 - 2^{-M+m+2}. \quad (55)$$

Inequality holds in (55) by taking the logarithm on both sides. Next, we substitute the value of $\lambda$ from (34). We observe that as $\epsilon \to 0$, the right-hand side of the inequality in (54) will tend to infinity. Since all the transformations we have done are equivalent, we can argue that $\rho_m < \rho_{m+1}$. 


APPENDIX D

PROOF OF THEOREM 3.

To obtain the optimal power allocation scheme, we solve problem (15). We start our solution by writing the Lagrangian function, which is

\[ \mathcal{L}(\rho_m, \mu_m, \lambda) = \sum_{m=1}^{M} \rho_m E_{m-1}^{CC-HARQ} + \sum_{m=1}^{M} \mu_m \rho_m + \lambda (E_M^{CC-HARQ} - \epsilon), \]

where \( E_{m}^{CC-HARQ} \) can be computed from (38).

From \( C_1 \) we compute the derivative of the Lagrangian function \( \mathcal{L}(\rho_m, \mu_m, \lambda) \) with respect to the power \( \rho_m \) as

\[ \frac{\partial \mathcal{L}(\rho_m, \mu_m, \lambda)}{\partial \rho_m} = \left( \frac{\phi^{m-1}}{(m-1)! \prod_{i=1}^{m-1} \rho_i} - \sum_{i=1}^{M-m} \frac{\rho_{m+i} \phi^{m+i-1}}{(m+i-1)! \prod_{j=1, j \neq m}^{m+i-1} \rho_j} \right) - \mu_m - \lambda \left( \frac{\phi^M}{M! \rho_m \prod_{i=1, i \neq m}^{M} \rho_i} \right). \]

Following a similar procedure as in the proof of Theorem 1, we allow \( \mu_m = 0 \). By taking the derivatives with respect to \( \rho_M, \rho_{M-1}, \ldots, \rho_1 \) and making some mathematical manipulations we obtain the following structure

\[ \rho_M = \sqrt{\phi \lambda}, \]

\[ \rho_{M-1} = \sqrt{\frac{2 \rho_M \phi}{M - 1}}, \]

\[ \rho_{M-2} = \sqrt{\frac{2 \rho_{M-1} \phi}{M - 2}}, \]

\[ \vdots \]

\[ \rho_{M-m} = \sqrt{\frac{2 \rho_{M-m+1} \phi}{M - m}}. \]

From the structure above, we observe that we can still apply the back-substitution approach and obtain \( \rho_m \) as in (40).

Based on (40), it is straightforward to show that the power terms are positive. Further, all the power terms are computed as a function of \( \lambda \). To find the value of the equality Lagrange multiplier we utilize \( C_4 \) and write:

\[ E_M^{CC-HARQ} = \frac{\phi^M}{M! \prod_{m=1}^{M} \rho_m} = \epsilon, \]
where $\rho_m$ is calculated as

$$\rho_m = \frac{2a \phi^b \lambda^c}{\sqrt{m}}. \quad (63)$$

In (63), the exponents $a(m)$, $b(m)$ and $c(m)$ are defined in (31), (32) and (33). Finally, we compute the Lagrangian multiplier $\lambda$ as:

$$\lambda = \left( \frac{\phi^{(M)}}{(M!)^e p(M) \prod_{m=1}^{M-1} \frac{2}{M-m} p^{(M)-p(m)}} \right)^{1/o(M)} \quad , \quad (64)$$

where $o(m)$ and $p(m)$ are given in (35) and (36) respectively.

**APPENDIX E**

**PROOF OF LEMMA 2.**

In [34] the author shows that

$$\lim_{s \to \infty} s^{m+1} \Pr \left[ u_m + v_m < R \right] = \int_0^R \psi(R-x)\gamma'(x)dx. \quad (65)$$

where $\psi(R)$ and $\gamma(R)$ are monotone and increasing and integrable functions, $\gamma'(R)$ is integrable and $u_m$ and $v_m$ are independent random variables that satisfy the following conditions

$$\lim_{s \to \infty} s \Pr \left[ u_m < R \right] = \gamma(R),$$

$$\lim_{s \to \infty} s^m \Pr \left[ v_m < R \right] = \psi(R).$$

In (42) we set $u_m = \log(1 + \rho_m)$. It is straightforward to show that, when the channel gains are Rayleigh distributed

$$\Pr \left[ u_m < R \right] = 1 - e^{\left(\frac{s^{R-1}}{\rho_m}\right)}, \quad (66)$$

and

$$\lim_{s \to \infty} s \Pr \left[ u_m < R \right] = e^R - 1. \quad (67)$$

Letting $\psi_0(R) = 1$ and $\psi_1(t) = \gamma(t)e^R - 1$, and recursively applying the theorem we obtain

$$\lim_{s \to \infty} s^2 \Pr \left[ \sum_{m=1}^{2} u_m < R \right] = \int_0^R \psi_1(R-x)f'(x)dx \quad \quad (68)$$

$$= R(e^R - 1) + 1. \quad (69)$$
The expression computed in (69) corresponds to $\psi_2(R)$. By continuing this, we obtain the recursive integral
\[
\lim_{s \to \infty} s^M \Pr \left[ \sum_{m=1}^{M} u_m < R \right] = \int_{0}^{R} \psi_{M-1}(R - x) \gamma'(x) dx
\]  
(70)
\[
= \psi_M(R).
\]  
(71)
The recursive integral in (71) provides each of the terms $\psi_1(R), \psi_2(R), \ldots, \psi_M(R)$. This integral converges to the series shown in (72). Thus, we compute each of the terms $\psi_m(R)$ as
\[
\psi_m(R) = (-1)^m \left( 1 + \sum_{i=0}^{m} \frac{(-1)^i}{(i - 1)!} e^{R(R - 1)} \right), \quad m = 1, \ldots, M.
\]  
(72)
Finally, by using these results, we can approximate the outage probability as shown in (43).

**APPENDIX F**

**PROOF OF THEOREM 4.**

To compute the power allocation algorithm, again we resort to problem (15). First, we write the Lagrangian function as
\[
\mathcal{L}(\rho_m, \mu_m, \lambda) = \sum_{m=1}^{M} \rho_mE_{IR-HARQ}^{m-1} + \sum_{m=1}^{M} \mu_m \rho_m + \lambda(E_{IR-HARQ}^M - \epsilon),
\]  
(73)
By analyzing C1, we write the derivative of the Lagrangian function $\mathcal{L}(\rho_m, \mu_m, \lambda)$ with respect to the power $\rho_m$ as
\[
\frac{\partial \mathcal{L}(\rho_m, \mu_m, \lambda)}{\partial \rho_m} = \left( \frac{\psi_{m-1}}{\prod_{i=1}^{m-1} \rho_i} - \sum_{m=1}^{M-m} \frac{\rho_{m+i} \psi_{m+i-1}}{\prod_{i=1}^{m+i-1} \rho_i} \right) - \mu_m - \lambda \left( \frac{\psi_M}{\prod_{i=1}^{M} \rho_m \prod_{i=1,i \neq m} \rho_i} \right)
\]  
(74)
Similarly as in the repetitive schemes, we relax the non-negativity constraint for the power terms. Next, by derivating with respect to $\rho_M, \rho_{M-1} \ldots \rho_1$ and making some mathematical operations we obtain
\[
\rho_M = \sqrt{\frac{\lambda M \psi_M}{\psi_{M-1}}},
\]  
(75)
\[
\rho_{M-1} = \sqrt{\frac{2 \rho_M \psi_{M-1}}{\psi_{M-2}}},
\]  
(76)
\[
\rho_{M-2} = \sqrt{\frac{2 \rho_{M-1} \psi_{M-2}}{\psi_{M-3}}},
\]  
(77)
\[
\vdots
\]
\[
\rho_1 = \sqrt{\frac{2 \rho_2 \psi_1}{\psi_0}}.
\]  
(78)
Again, we notice that the back-substitution approach is applicable, and the power term of the $m^{th}$ round can be written as in (45).

At this point, we argue that the power terms in (45) are non-negative. Further, we compute each of the power terms as a function of the equality Lagrangian multiplier, $\lambda$. To derive a closed-form expression for it, we utilize C4:

$$E_M^{1R-HARQ} = \frac{\psi^M}{M! \prod_{m=1}^{M} \rho_m} = \epsilon, \quad (79)$$

where the power term in the $m^{th}$ round can be found as

$$\rho_m = 2^{a(m)}(m\lambda)c(m) \prod_{i=1}^{m} \frac{\psi_i}{\psi_{i-1}} \cdot \psi_1 \cdot d(i). \quad (80)$$

In (80), $a(m)$ and $c(m)$ can be found from (31) and (32), respectively. Further, the exponent $d(i)$ is found as $d(i) = 2^{-i}$. Finally, we can find the value of $\lambda$ as

$$\lambda = \left( \frac{\psi^M}{M^{o(M)}e^{p(M)}2^{q(M)} \prod_{m=1}^{M} \frac{1}{\psi_{m-1}^{2q(M)-m}}} \right)^{\frac{1}{2q(M)}}, \quad (81)$$

where $o(m)$, $p(m)$ and $q(m)$ are given from (35), (36) and (37) respectively.

REFERENCES

[1] M. Latva-aho and K. Leppanen, “Key drivers and research challenges for 6G ubiquitous wireless intelligence,” 6G Flagship, University of Oulu, Finland, Sep 2019.
[2] G. Durisi, T. Koch, and P. Popovski, “Toward massive, ultrareliable, and low-latency wireless communication with short packets,” Proceedings of the IEEE, vol. 104, no. 9, pp. 1711–1726, Sept 2016.
[3] M. R. Palattella, M. Dohler, A. Greico, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, “Internet of Things in the 5G Era: Enablers, Architecture, and Business Models,” IEEE Journal on Selected Areas in Communications, vol. 34, no. 3, pp. 510–527, March 2016.
[4] Nokia Networks, “5G for mission critical communication: Achieve ultra-reliability and virtual zero latency,” Nokia networks white paper, 2016.
[5] P. Popovski, “Ultra-reliable communication in 5G wireless systems,” in 1st International Conference on 5G for Ubiquitous Connectivity, Nov 2014, pp. 146–151.
[6] O. N. C. Yilmaz, Y. P. E. Wang, N. A. Johansson, N. Brahim, S. A. Ashraf, and J. Sachs, “Analysis of ultra-reliable and low-latency 5G communication for a factory automation use case,” in 2015 IEEE International Conference on Communication Workshop (ICCW), June 2015, pp. 1190–1195.
[7] N. H. Mahmood, H. Alves, O. L. A. López, M. Shehab, D. P. M. Osorio, and M. Latva-aho, “Six key enablers for machine type communication in 6G,” CoRR, vol. abs/1903.05406, 2019. [Online]. Available: http://arxiv.org/abs/1903.05406
[8] M. Shehab, H. Alves, and M. Latva-aho, “Effective capacity and power allocation for machine-type communication,” IEEE Transactions on Vehicular Technology, vol. 68, no. 4, pp. 4098–4102, April 2019.
[9] O. L. A. Lopez, H. Alves, and M. Latva-aho, “Joint power control and rate allocation enabling ultra-reliability and energy efficiency in simo wireless networks,” *IEEE Transactions on Communications*, pp. 1–1, 2019.

[10] M. Shehab, E. Dosti, H. Alves, and M. Latva-aho, “On the effective energy efficiency of ultra-reliable networks in the finite blocklength regime,” in *2017 International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2017, pp. 275–280.

[11] L. Le and E. Hossain, “An analytical model for arq cooperative diversity in multi-hop wireless networks,” *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, pp. 1786–1791, May 2008.

[12] M. Jabi, M. Benjillali, L. Szczecinski, and F. Labeau, “Energy efficiency of adaptive harq,” *IEEE Transactions on Communications*, vol. 64, no. 2, pp. 818–831, Feb 2016.

[13] Y. Deng, L. Wang, M. Elkashlan, K. J. Kim, and T. Q. Duong, “Generalized selection combining for cognitive relay networks over nakagami- m fading,” *IEEE Transactions on Signal Processing*, vol. 63, no. 8, pp. 1993–2006, April 2015.

[14] K. Wang and Z. Ding, “Diversity Integration in Hybrid-ARQ With Chase Combining Under Partial CSI,” *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2647–2659, June 2016.

[15] C. Ji, D. Wang, N. Liu, and X. You, “On power allocation for incremental redundancy hybrid arq,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1506–1518, March 2015.

[16] “IEEE Standard for Air Interface for Broadband Wireless Access Systems– Amendment 3: Multi-tier Networks, 2015 ,” *IEEE 802.16q*.

[17] 3GPP, “3GPP TS 38.300: Technical Specification Group Radio Access Network; NR; NR and NG-RAN Overall Description; Stage 2;” *Rel 15*, vol. 15.5.0, pp. 1–97, 2019.

[18] P. Larsson, L. K. Rasmussen, and M. Skoglund, “Analysis of rate optimized throughput for arq in fading interference channels,” in *2014 IEEE International Conference on Communications (ICC)*, June 2014, pp. 5926–5931.

[19] T. V. K. Chaitanya and T. Le-Ngoc, “Energy-efficient adaptive power allocation for incremental mimo systems,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2820–2827, April 2016.

[20] T. V. K. Chaitanya and E. G. Larsson, “Optimal power allocation for hybrid arq with chase combining in i.i.d. rayleigh fading channels,” *IEEE Transactions on Communications*, vol. 61, no. 5, pp. 1835–1846, May 2013.

[21] ——, “Outage-optimal power allocation for hybrid arq with incremental redundancy,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 7, pp. 2069–2074, July 2011.

[22] E. Dosti, U. L. Wijewardhana, H. Alves, and M. Latva-aho, “Ultra reliable communication via optimum power allocation for Type-I ARQ in finite Block-Length,” in *IEEE ICC 2017 Wireless Communications Symposium (ICC’17 WCS)*, Paris, France, May 2017, pp. 5019–5024.

[23] E. Dosti, M. Shehab, H. Alves, and M. Latva-aho, “Ultra reliable communication via CC-HARQ in finite Block-Length,” in *2017 European Conference on Networks and Communications (EuCNC): Physical Layer and Fundamentals (PHY) (EuCNC2017 - PHY)*, Oulu, Finland, Jun. 2017.

[24] S. H. Kim, D. K. Sung, and T. Le-Ngoc, “Performance analysis of incremental redundancy type hybrid ARQ for finite-length packets in AWGN channel,” in *2013 IEEE Global Communications Conference (GLOBECOM)*, Dec 2013, pp. 2063–2068.

[25] R. Devassy, G. Durisi, P. Popovski, and E. G. Strom, “Finite-blocklength analysis of the ARQ-protocol throughput over the Gaussian collision channel,” in *2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, May 2014, pp. 173–177.

[26] B. Makki, T. Svensson, and M. Zorzi, “Finite block-length analysis of the incremental redundancy harq,” *IEEE Wireless Communications Letters*, vol. 3, no. 5, pp. 529–532, Oct 2014.
[27] B. Makki, T. Svensson and M. Zorzi, “Green communication via Type-I ARQ: Finite block-length analysis,” in 2014 IEEE Global Communications Conference, Dec 2014, pp. 2673–2677.

[28] J. Ostman, R. Devassy, G. C. Ferrante, and G. Durisi, “Low-latency short-packet transmissions: Fixed length or harq?” in 2018 IEEE Globecom Workshops (GC Wkshps), Dec 2018, pp. 1–6.

[29] G. Durisi and T. Koch and J. Ostman and Y. Polyanskiy and W. Yang, “Short-packet communications over multiple-antenna rayleigh-fading channels,” IEEE Transactions on Communications, vol. 64, no. 2, pp. 618–629, Feb 2016.

[30] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, “Quasi-static multiple-antenna fading channels at finite blocklength,” IEEE Transactions on Information Theory, vol. 60, no. 7, pp. 4232–4265, July 2014.

[31] Y. Polyanskiy, H. V. Poor, and S. Verdu, “Channel coding rate in the finite blocklength regime,” IEEE Transactions on Information Theory, vol. 56, no. 5, pp. 2307–2359, May 2010.

[32] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, “Block-fading channels at finite blocklength,” in ISWCS 2013; The Tenth International Symposium on Wireless Communication Systems, Aug 2013, pp. 1–4.

[33] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, UK: Cambridge University Press, 2004.

[34] J. N. Laneman, “Limiting analysis of outage probabilities for diversity schemes in fading channels,” in Global Telecommunications Conference, 2003. GLOBECOM ’03. IEEE, vol. 3, Dec 2003, pp. 1242–1246 vol.3.