

Article

# Performance Evaluation of HARQ Schemes for the Internet of Things <sup>†</sup>

Lorenzo Vangelista \*  and Marco Centenaro 

Department of Information Engineering, University of Padova, 35131 Padua, Italy; centenar@dei.unipd.it

\* Correspondence: lorenzo.vangelista@unipd.it; Tel.: +39-049-827-7673

† This paper is an extended version of the conference paper, Centenaro, M.; Ministeri, G.; Vangelista, L.

A comparison of energy-efficient HARQ protocols for M2M communication in the finite block-length regime, 2015 IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB), Montreal, QC, Canada, 4–7 October 2015.

Received: 20 August 2018; Accepted: 21 September 2018; Published: 25 September 2018



**Abstract:** Hybrid Automatic Repeat reQuest (HARQ) techniques are widely employed in the most important wireless systems, e.g., the Long Term Evolution (LTE) cellular standard, to increase the reliability of the communication. Despite these schemes have been widely studied in literature in the past several years, the recent results obtained by Polyanskiy, Poor, and Verdú on the finite-blocklength regime disclosed new possibilities for the research on HARQ schemes. Indeed, new communications trends, which are usually part of the Internet of Things (IoT) paradigm, are characterized by very short packet sizes and a high reliability requirement and, therefore, they call for efficient HARQ techniques. In many scenarios, the energy efficiency of the communication plays a key role as well. In this paper, we aim at providing a comprehensive performance comparison of various kinds of HARQ schemes in the context of short-packet transmissions with energy constraints. We derive optimal power allocation strategies and we show that a minimum 50% energy saving can be achieved after very few transmission attempts if we enable packet combining at the receiver side.

**Keywords:** HARQ; energy-efficiency; finite-blocklength regime; M2M; reliable wireless communication; IoT

## 1. Introduction

To increase the reliability, many current widely-used wireless communication standards make use of Hybrid Automatic Repeat reQuest (HARQ) techniques, combining Forward Error Correction (FEC) channel codes and Automatic Repeat reQuest (ARQ) retransmission processes. Just to mention a noteworthy example in wireless networks, the current cellular standard, i.e., the Long-Term Evolution (LTE), exploits the turbo code as FEC channel code in its HARQ procedure. HARQ techniques are suitable in short packet transmissions for the Internet of Things (IoT) networks, as well, to increase the reliability of such transmissions. The effectiveness of HARQ techniques has been studied so far for coding strategies based on Shannon's theory [1], which states that a reliable communication can be guaranteed only if the information rate  $R$  is strictly lower than the channel capacity  $C$ . However, Shannon's channel coding theorem holds just in the *asymptotic regime*, i.e., only if the codewords are sufficiently long. For many practical systems, as the aforementioned LTE, this assumption is reasonable because the exchanged messages take many channel uses to be transmitted. However, for some types of communication mostly pertaining to the Internet of Things (IoT) paradigm, the assumption of an asymptotic regime does not hold and we should rather refer to the *finite-blocklength* regime. Some interesting studies by Polyanskiy–Poor–Verdú [2] and Erseghe [3], which recently appeared in the literature, provide interesting results in this direction. In these pivotal papers, for a fixed number of

channel uses  $n$ , analytic expressions are derived to link the “nominal” channel capacity by Shannon, the maximum achievable information rate, and the decoding error probability.

In this context, the aim of this paper is to tailor the analysis and optimization of the HARQ process to the case of IoT devices by exploiting the finite-blocklength theory. In particular, we will focus on Machine-to-Machine (M2M) communications, (For this reason, we will frequently swap the terms IoT and M2M throughout the entire paper.) that is, on low-cost wireless sensor nodes that collect some kind of data, e.g., temperature, pressure, pollution index, and transmit them to a unique concentrator node, which is in charge of the forwarding towards the core network. Such a *star* network topology offloads the complexity of end devices, shifting the burden to the concentrator. Internet of Things networks, although started using mesh topologies, are rapidly moving to star topologies. As a matter of fact, the emerging paradigm in the Internet of Things connectivity is the Low Power Wide Area Networks, which includes NarrowBand IoT (NB-IoT), LoRaWAN, and SigFox (see [4]).

In the following of this paper, we will derive a mathematical framework for the optimization of the HARQ procedure operation at the physical layer between the sensor node and the concentrator, considering the following *three major challenges* that M2M entails.

1. **Extremely Short Packet Size.** M2M traffic is characterized by *sporadic uplink* transmissions of *short packets* [5], in contrast to the broadband downlink traffic that is typical of traditional services as, e.g., video streaming. While the peculiar timing pattern of M2M transmissions mostly impacts the Medium Access Control (MAC) design, the packets size should be taken into account in evaluating some physical layer aspects as the Modulation and Coding Scheme (MCS). Since this paper focuses on physical layer, we will address only the latter feature of M2M traffic by utilizing the aforementioned finite-blocklength theory in the proposed mathematical framework.
2. **Energy Efficiency.** Machine-Type Devices (MTDs) are usually battery-powered, thus minimizing their energy consumption during the packet transmission process helps in prolonging their battery durations in order to reach lifespans of up to ten years [6]. Moreover, minimizing the transmission in order to save energy helps to reduce the interference with co-located MTDs can be reduced, thus increasing the overall network throughput [7,8].
3. **Diverse Performance Requirements.** On the other hand, the energy consumption can be traded for performance to support efficiently some specific IoT services. A prominent example is the emerging paradigm of Ultra-Reliable Communications (URC) [9], in which extremely low outage probability ( $\sim 10^{-5}$ ) and delivery delay ( $\sim 10^{-3}$  s) are required to support the needs of applications related to the so-called “tactile Internet” [10]. The performance evaluation results presented in this work provide useful insights on the aforementioned trade-off between energy efficiency and communication performance.

The remainder of the paper is organized as follows: in Section 2, we review the related work, explaining the novelty of our approach. In Section 3, the system model is described and the optimization problem is formulated. In Section 4, the performance of various HARQ schemes is compared and several remarks on power allocation strategies, outage probabilities, delivery delays and energy savings are made. Finally, in Section 5, we draw the conclusions of our work.

## 2. Related Work

To begin with, let us recall that, according to its mode of operation, any HARQ process belongs to one of the two following categories:

1. Type-I: the destination discards the received packet if the decoding process fails;
2. Type-II: previously received versions of the packet are combined together with the newly received copy before trying to decode the message.

As for Type-II HARQ processes, we distinguish between two further sub-classes:

- (a) Chase Combining (CC): the entire codeword is sent in each transmission attempt,

- (b) Incremental Redundancy (IR): the original codeword is divided into multiple sub-codewords, which are sent in successive transmission attempts.

### 2.1. HARQ in the Asymptotic Regime

All the aforementioned schemes have been widely studied in literature in the context of asymptotic regime: we will mention just some of them. In [11], the authors propose an information theoretic framework to compute the throughput and the average delay of HARQ. In [12], instead, a unified performance metric and analysis of practical IR and CC HARQ schemes is given. HARQ performance in correlated Rayleigh fading channels is analyzed in [13,14]. The optimal strategies for power allocation in every transmission attempt were addressed in [15,16]: the performance evaluation results show that the transmit power increases among successive transmission attempts. In [17], multiuser (i.e., with per-node and per-link constraints) resource allocation (in terms of bandwidth and power) in Type-II HARQ-based OFDMA networks is addressed.

### 2.2. HARQ in the Finite-Blocklength Regime

As mentioned in the introduction, in the last few years, a growing interest in the finite-blocklength regime began in the research community, after the seminal work by Polyanskiy et al. [2], in which the authors developed novel bounds as well as an accurate approximation, based on a quantity called *channel dispersion*, to the maximum achievable rate for a given blocklength and a probability of block error. In [3,18], Erceghe derived refined approximations and compact integral expressions for converse and achievability bounds in the finite blocklength regime. In [19], instead, Yang et al., investigated the maximal channel coding rate that can be achieved at a given blocklength and error probability when the codewords are subjected to a long-term (i.e., averaged- over-all-codeword) power constraint.

The aforementioned results triggered many new research contributions in various wireless communication scenarios. Just to mention some examples, in [20], the relaying performance under the finite-blocklength regime is analyzed and compared with the Shannon capacity regime. In [21,22], the optimal payload length and transmission energy for CC HARQ are studied to design multi-hop relaying strategies. The energy consumption of CC HARQ in multi-hop transmissions is also addressed in [23]. In [24], finite-blocklength HARQ schemes are exploited to analyze throughput and probability of error for Cloud-Radio Access Network (C-RAN) architectures.

Interesting insights in the context of energy efficient communication were obtained, instead, by Kim et al. [25] and Makki et al. [26]. In particular, the latter study is of notable interest for the optimization of HARQ processes for M2M communications because it deals with *green communications*, where the target is not the throughput maximization, like most of the previously mentioned works, rather the design of power efficient transmission schemes. In such study, the authors derive the optimal power allocations over  $M$  subsequent transmission attempts of a Type-I HARQ process in order to minimize the outage probability (that is, the probability that the packet can not be correctly delivered after  $M$  transmission attempts) under a constraint on the average transmit power  $\bar{P}$ .

### 2.3. Our Contribution

The work done in [26] is a suitable starting point for our study, since it addresses all the three challenges of IoT: short packet size, energy efficiency, and communication performance. Our paper makes one step further: in this work, indeed, we aim at measuring the performance gain provided by Type-II HARQ processes with respect to open-loop and Type-I HARQ processes. We will specifically provide a novel performance evaluation and comparison among CC HARQ, IR HARQ, and a possible mixture of these two, denoted by *mixed* HARQ, in the finite-blocklength, energy-constrained regime. The performance metrics that will be considered are: (1) outage probability, (2) delivery delay, and (3) energy efficiency. We derive also the optimal power allocation that is needed to meet such performance requirements. The final goal of this study, indeed, is to provide useful results about the achievability of given performance indices in real M2M applications, with particular reference to URC.

We believe that our analysis is useful for the development of new protocols for the IoT because one can tailor the retransmission protocol to the specific IoT use case, according to the performance metric target that has to be fulfilled.

Finally, we remark that IR HARQ processes in this context were addressed in [26], but the analytic expressions derived in [26] are not suitable for our purposes, since we make the assumption that the coherence time of the wireless channel is shorter than the length of the transmission round. Moreover, despite the large body of works appearing (even in recent times) in literature on energy-efficient HARQ (e.g., [27,28]) and IoT (e.g., [29]), none of these has yet proposed a comprehensive comparison of the various HARQ approaches in supporting a diverse set of requirements for IoT applications.

### 3. System Model

In information theory, the channel capacity  $C$  is defined as the maximum value of the mutual information  $I$  between the input  $x$  and the output  $y$  of channel, i.e.,  $C = \max_{p_x} I(x; y)$ . For Additive White Gaussian Noise (AWGN) channels with noise variance  $\sigma_w^2$ , the channel output is  $y = x + w$ , thus,  $y$  is conditionally Gaussian given the input value  $x$ , i.e.,  $p_{y|x}(y|x) \sim \mathcal{N}(x, \sigma_w^2)$ . In this context, the maximum value of mutual information is attained for Gaussian-distributed input, i.e.,  $x \sim \mathcal{N}(0, \sigma_x^2)$ , and the channel capacity (expressed in bits per channel use) is

$$C = \frac{1}{2} \log_2 \left( \frac{\sigma_y^2}{\sigma_w^2} \right) = \frac{1}{2} \log_2 (1 + \Gamma) \text{ [bpcu]}, \quad (1)$$

where  $\sigma_y^2 = \sigma_x^2 + \sigma_w^2$  and  $\Gamma = \sigma_x^2 / \sigma_w^2$  is the Signal-to-Noise Ratio (SNR). If passband modulations are considered, the capacity is doubled since we are dealing with two parallel AWGN channels:  $C = \log_2(1 + \Gamma)$  [bpcu]. This is one of the most famous expressions of information theory and, due to its simplicity, it is used in a lot of theoretical works involving coding schemes in the asymptotic regime, as discussed in the previous section.

Let us consider an encoding/decoding procedure that takes  $b$  information symbols from the alphabet  $\mathcal{M}$  and encodes them into a codeword belonging to set  $\mathcal{M}^L$ , i.e., a codeword taking  $L$  channel uses, where  $L$  is a finite number [2]. Assuming that the alphabet is binary, i.e.,  $|\mathcal{M}| = 2$ , the information rate of the codeword is

$$R = \frac{\log_2 |\mathcal{M}|^b}{L} = b \times \frac{\log_2 |\mathcal{M}|}{L} = \frac{b}{L} \text{ [bpcu]}. \quad (2)$$

The codewords are then transmitted on the wireless channel using a transmitted symbol power  $P$ . As for the wireless channel model, in the following, we will consider an AWGN channel with complex normal fading coefficients  $h$ , so that  $|h|$  is Rayleigh-distributed, as depicted in Figure 1. For the sake of simplicity and without loss of generality (We remark that the considered channel model, despite being simplified with respect to more realistic models, allows us to write the optimization problem in closed form, and solve it, gaining insights regarding real-world deployments.):

1. the fading process is assumed flat on every transmitted codeword and fading realizations among distinct blocks  $h_i$  and  $h_j$ ,  $i \neq j$ , are independent, i.e.,  $h_i \perp h_j$ ;
2. the average value of the squared fading amplitudes  $g_i = |h_i|^2$ , which are exponentially distributed, is unitary  $\forall i$ , thus  $f_{g_i}(g) = e^{-g} \forall i$ ;
3. the noise has unit power, i.e.,  $\sigma_w^2 = 1$ .

Finally, we denote as  $\epsilon$  the probability that the packet decoding process fails.

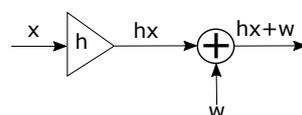


Figure 1. Block scheme of the wireless channel.

Under these assumptions, the relationship among transmission rate, Shannon capacity, and error probability is derived in [2] and can be expressed as follows:

$$R = C - \sqrt{\frac{V}{L}} \times \mathcal{Q}^{-1}(\epsilon) + \frac{1}{2} \times \frac{\log_2 L}{L} + \mathcal{O}(1), \quad (3)$$

where  $C = \frac{1}{2} \log_2(1 + \Gamma)$  bpcu is Shannon capacity,  $V$  is the *channel dispersion* and is equal to

$$V = \frac{1}{2} \times \frac{\Gamma(\Gamma + 2)}{(\Gamma + 1)^2} (\log_2 e)^2 = \frac{1}{2} \times \left[ 1 - \frac{1}{(\Gamma + 1)^2} \right] (\log_2 e)^2 \quad (4)$$

and  $\mathcal{Q}(x) \triangleq (1/\sqrt{2\pi}) \int_x^\infty e^{-z^2/2} dz$ . In the case of passband channel, the Shannon capacity and the channel dispersion must be multiplied by two.

We can approximate the average decoding error probability as a function of  $L, R, P$ , taking the expectation over the fading statistics as follows:

$$\phi \approx \mathbb{E}_g[\epsilon(L, R, gP)] = \mathbb{E}_g \left[ \mathcal{Q} \left( \frac{\sqrt{L} \left( \log_2(1 + gP) + \frac{1}{2} \times \frac{\log_2 L}{L} - R \right)}{\sqrt{1 - \frac{1}{(1+gP)^2} \log_2 e}} \right) \right], \quad (5)$$

where  $\mathbb{E}$  is the expectation operator.

If the decoding process fails, the receiver does not send the acknowledgement (ACK) and the transmitter needs to send the packet again. Let us denote with  $M, P_m$ , and  $\Phi_m$  the maximum number of transmission attempts, the transmission power employed in the  $m$ -th transmission attempt, and the probability that the data is not correctly decoded until the  $m$ -th transmission attempt, respectively. Note that by definition it is  $\Phi_0 = 1$ . The *average consumed energy* during a transmission round is given by

$$\bar{\xi} = L \times \sum_{m=1}^M P_m \Phi_{m-1} \quad (6)$$

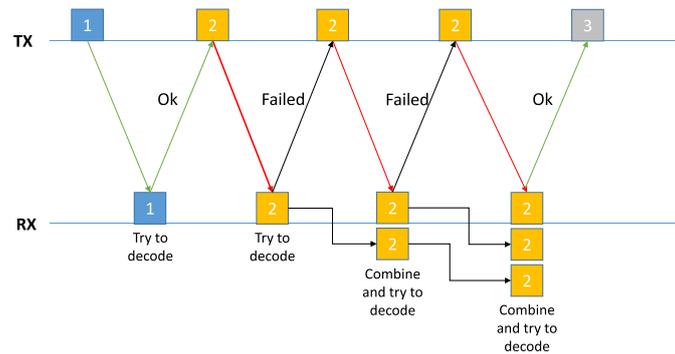
and the *expected number of channel uses* is

$$\bar{\tau} = L \times \sum_{m=1}^M \Phi_{m-1}. \quad (7)$$

Then, the *average transmission power* is computed using the renewal reward theory [30] as

$$\bar{P} = \frac{\bar{\xi}}{\bar{\tau}} = \frac{\sum_{m=1}^M P_m \Phi_{m-1}}{\sum_{m=1}^M \Phi_{m-1}}. \quad (8)$$

In [26], the power allocation value across successive transmission attempts  $P_m \forall m = 1, \dots, M$  are optimized in order to minimize the outage probability  $\Phi_M$ , i.e., the decoding failure probability after  $M$  transmission attempts, under a constraint on the average transmitted power value. In our work, we tailor the optimization problem to a more realistic scenario and we investigate the actual performance gap between Type-I HARQ, CC HARQ, IR HARQ, and mixed HARQ processes. In Figure 2, a sample flow diagram of a CC HARQ process is provided.



**Figure 2.** Sample packet flow of a CC HARQ process. At every transmission attempt failure, the receiver (RX) stores the packet received from the transmitter (TX) to perform joint decoding at the next transmission attempt.

The minimization problem we design is the following:

$$\min_{P_m} \Phi_M \tag{9a}$$

subject to

$$\bar{\zeta} = J, \tag{9b}$$

$$P_m \geq P_{\min}, \tag{9c}$$

$$P_m \leq P_{\max}. \tag{9d}$$

Note that, in contrast to what is done in [26], we are not imposing a constraint on the average transmitted power  $\bar{P}$ , but rather on the average energy budget  $\bar{\zeta}$ . Moreover, instead of letting the transmit power  $P_m$  grow to infinity, we are forcing  $P_m$  in the interval  $[P_{\min}, P_{\max}]$ , which is a reasonable constraint in a real scenario of power-constrained devices. This choice puts a constraint on the average consumed energy  $\bar{\zeta}$ , as well. The minimum energy budget  $J_{\min}$  can be easily obtained as follows:

$$\bar{\zeta} = L \times \sum_{m=1}^M P_m \Phi_{m-1} \geq L \times \sum_{m=1}^M P_{\min} \Phi_{m-1} \geq LP_{\min} \triangleq J_{\min}, \tag{10}$$

recalling that  $\Phi_0 = 1$ . A similar reasoning is applied to obtain the upper bound  $J_{\max}$ :

$$\bar{\zeta} = L \times \sum_{m=1}^M P_m \Phi_{m-1} \leq L \times \sum_{m=1}^M P_{\max} \Phi_{m-1} \leq LP_{\max}M \triangleq J_{\max}. \tag{11}$$

According to the kind of HARQ protocol that is implemented at the receiver side, the actual outage probability  $\Phi_M$  value is derived as explained in the following of this section.

### 3.1. Type-I HARQ

If the receiver tries to decode only the last received packet discarding previously received information, the outage probability is

$$\Phi_m^{(TI)} = \begin{cases} \prod_{j=1}^m \phi_j, & \text{if } m \neq 0, \\ 1, & \text{if } m = 0, \end{cases} \tag{12}$$

where  $\phi_j$  denotes the probability that the data is not decoded during the  $j$ -th transmission attempt, i.e.,  $\phi_j = \mathbb{E}_g[\epsilon(L, R, gP_j)]$ .

### 3.2. General Characterization of Type-II HARQ Processes

If packet combining is enabled at the receiver side, the outage probability can be computed as follows. First of all, we recall that  $M$  is the maximum number of transmission attempts and  $L$  is the total number of channel uses to transmit the codeword. Let us divide the original codeword into  $N \leq M$  sub-codewords of length

$$L_j = \frac{L}{N} \quad \forall j = 1, \dots, M. \quad (13)$$

The information rate after the  $j$ -th transmission attempt is equal to

$$R_j = \frac{b}{\min(j, N) \times L_j} = \frac{R \times N}{\min(j, N)} \quad \forall j = 1, \dots, M. \quad (14)$$

Note that  $R_j = R \forall j \geq N$ . Let  $K = \lceil M/N \rceil$  and define the  $N \times K$  matrix of channel gains  $\mathbf{G}$  as

$$\mathbf{G} = \begin{bmatrix} g_1 & g_{N+1} & \cdots & g_{(K-1)N+1} \\ g_2 & g_{N+2} & \cdots & g_{(K-1)N+2} \\ \vdots & \vdots & & \vdots \\ g_N & g_{N+N} & \cdots & g_{NK} \end{bmatrix} \quad (15)$$

and the  $K \times N$  matrix of transmit powers  $\mathbf{P}$  as

$$\mathbf{P} = \begin{bmatrix} P_1 & P_2 & \cdots & P_N \\ P_{N+1} & P_{N+2} & \cdots & P_{N+N} \\ \vdots & \vdots & & \vdots \\ P_{(K-1)N+1} & P_{(K-1)N+2} & \cdots & P_{NK} \end{bmatrix}. \quad (16)$$

Let us denote with  $\mathbf{G}^{(j)}$  and  $\mathbf{P}^{(j)}$  the channel gain and transmit power matrices having the elements of  $\mathbf{G}$  and  $\mathbf{P}$  whose subscript is less or equal to  $j$  and zeros elsewhere. Then, we define matrix  $\mathbf{\Omega}^{(j)}$  as

$$\mathbf{\Omega}^{(j)} = \mathbf{G}^{(j)} \cdot \mathbf{P}^{(j)}. \quad (17)$$

The transmission of multiple sub-codewords can be modeled as a  $N$ -parallel AWGN channel with Rayleigh fading [31]. Therefore, the decoding error probability of the  $j$ -th transmission attempt is expressed as

$$\epsilon_j(\mathbf{\Omega}^{(j)}) \simeq \mathcal{Q} \left( \frac{C_j + \frac{1}{2} \times \frac{\log_2 L_j}{L_j} - R_j}{\sqrt{V_j/L_j}} \right), \quad (18)$$

where

$$C_j = \sum_{\alpha=1}^{j \leq N} C(\Omega_{\alpha, \alpha}^{(j)}), \quad (19)$$

and

$$V_j = \sum_{\alpha=1}^{j \leq N} V(\Omega_{\alpha, \alpha}^{(j)}), \quad (20)$$

are the cumulative channel capacity and cumulative channel dispersion, respectively. Note that the subscripts  $(\alpha, \alpha)$  refer to the elements on the main diagonal of matrix  $\mathbf{\Omega}^{(j)}$ .

The decoding error probability of Type-II HARQ processes at the  $j$ -th transmission attempt  $\phi_j^{(T2)}$  is obtained averaging  $\epsilon_j$  over the channel gain distribution as follows:

$$\phi_j^{(T2)} = \mathbb{E}_{g_1, \dots, g_j} [\epsilon_j] = \underbrace{\int_0^{+\infty} \dots \int_0^{+\infty}}_{j \text{ integrals}} \epsilon_j \cdot e^{-\sum_{\alpha=1}^j g_\alpha} d\mathbf{g}. \quad (21)$$

Note that the factorization of  $f_{g_1, \dots, g_j}(g_1, \dots, g_j)$  as

$$f_{g_1, \dots, g_j}(g_1, \dots, g_j) = \prod_{\alpha=1}^j f(g_\alpha) = \exp\left(-\sum_{\alpha=1}^j g_\alpha\right) \quad (22)$$

is valid under the hypothesis of independent block fading. The outage probability at the  $m$ -th transmission attempt becomes

$$\Phi_m^{(T2)} = \begin{cases} \prod_{j=1}^m \phi_j^{(T2)}, & \text{if } m \neq 0, \\ 1, & \text{if } m = 0. \end{cases} \quad (23)$$

In the following, we tailor the preceding general characterization to the various kinds of Type-II HARQ processes.

### 3.2.1. CC HARQ

If the receiver is capable of collecting the replicas of the received packet in a buffer and combining them to increase the effective SNR, we have a CC HARQ process. The combination of the packets is done exploiting Maximal Ratio Combining (MRC) [32], thus the effective SNR after the  $j$ -th transmission attempt is

$$\Gamma_j^{(CC)} = \sum_{\alpha=1}^j \Gamma_\alpha = \sum_{\alpha=1}^j \frac{g_\alpha P_\alpha}{\sigma_w^2}. \quad (24)$$

We want to remark that this is a special case of the previously proposed characterization in which  $N = 1$ : there is not any parallel channel and just bit repetition is allowed.

### 3.2.2. IR HARQ

If the original codeword is divided into  $N = M$  sub-codewords, then the process is a pure IR HARQ. This scheme exploits a  $M$ -parallel channel, but no bit repetition is allowed.

### 3.2.3. Mixed HARQ

This approach merges the strength of IR and CC HARQ, since  $N$  is such that  $1 < N < M$ , thus both a  $N$ -parallel channel and bit repetition are considered in this process. This is the most general scheme and it is employed in the LTE cellular standard [33].

## 4. Performance Evaluation

Numerical evaluations have been performed using the parameters summarized in Table 1. The transmit power  $P_i$  is bounded in the interval  $[0, 20]$  dB, while a maximum amount of  $M = 3$  transmission attempts is allowed. A blocklength of  $L = 50$  channel uses has been chosen and the original codeword has been divided into  $N = M = 3$  segments, of length  $L_j = L/M \simeq 17$  each, and  $N = 2$  segments, of length  $L_j = L/N = 25$  each, for IR and mixed HARQ, respectively. We want to remark that the choice of these parameters has been done in order to reduce the amount of buffering required to support data retransmission, which is a really important constraint for cost- and delay-sensitive MTDs [34].

**Table 1.** Simulation parameter list.

Parameter	Value
$P_{\min}$	0 dB
$P_{\max}$	20 dB
$L$	50
$R$	1.44 bpcu
$M$	3
$N$	2

For this particular choice of the parameters, in the CC scheme, it is

$$\mathbf{G}^{(1)} = [g_1 \ 0 \ 0], \mathbf{G}^{(2)} = [g_1 \ g_2 \ 0], \mathbf{G}^{(3)} = [g_1 \ g_2 \ g_3];$$

$$\mathbf{P}^{(1)} = [P_1 \ 0 \ 0]^T, \mathbf{P}^{(2)} = [P_1 \ P_2 \ 0]^T, \mathbf{P}^{(3)} = [P_1 \ P_2 \ P_3]^T;$$

$$\mathbf{\Omega}^{(1)} = [g_1 P_1], \mathbf{\Omega}^{(2)} = [g_1 P_1 + g_2 P_2], \mathbf{\Omega}^{(3)} = [g_1 P_1 + g_2 P_2 + g_3 P_3],$$

and, therefore, the cumulative capacity values are

$$C_1 = C(g_1 P_1), C_2 = C(g_1 P_1 + g_2 P_2), C_3 = C(g_1 P_1 + g_2 P_2 + g_3 P_3);$$

the same holds for dispersion  $V_j$ . For the IR scheme, instead, we have

$$\mathbf{G}^{(1)} = [g_1 \ 0 \ 0]^T, \mathbf{G}^{(2)} = [g_1 \ g_2 \ 0]^T, \mathbf{G}^{(3)} = [g_1 \ g_2 \ g_3]^T;$$

$$\mathbf{P}^{(1)} = [P_1 \ 0 \ 0], \mathbf{P}^{(2)} = [P_1 \ P_2 \ 0], \mathbf{P}^{(3)} = [P_1 \ P_2 \ P_3];$$

$$\mathbf{\Omega}^{(1)} = \begin{bmatrix} g_1 P_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{\Omega}^{(2)} = \begin{bmatrix} g_1 P_1 & \cdot & 0 \\ \cdot & g_2 P_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{\Omega}^{(3)} = \begin{bmatrix} g_1 P_1 & \cdot & \cdot \\ \cdot & g_2 P_2 & \cdot \\ \cdot & \cdot & g_3 P_3 \end{bmatrix},$$

where the matrix elements denoted with the highlighted dot are not explicitly calculated because their values do not affect the computation of  $\Omega_{\alpha,\alpha}^{(j)}$ . Therefore, the cumulative capacity values are

$$C_1 = C(g_1 P_1), C_2 = C(g_1 P_1) + C(g_2 P_2), C_3 = C(g_1 P_1) + C(g_2 P_2) + C(g_3 P_3);$$

the same holds for dispersion  $V_j$ . Finally, for the mixed HARQ case, it is

$$\mathbf{G}^{(1)} = \begin{bmatrix} g_1 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{G}^{(2)} = \begin{bmatrix} g_1 & 0 \\ g_2 & 0 \end{bmatrix}, \mathbf{G}^{(3)} = \begin{bmatrix} g_1 & g_3 \\ g_2 & 0 \end{bmatrix};$$

$$\mathbf{P}^{(1)} = \begin{bmatrix} P_1 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{P}^{(2)} = \begin{bmatrix} P_1 & P_2 \\ 0 & 0 \end{bmatrix}, \mathbf{P}^{(3)} = \begin{bmatrix} P_1 & P_2 \\ P_3 & 0 \end{bmatrix};$$

$$\mathbf{\Omega}^{(1)} = \begin{bmatrix} g_1 P_1 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{\Omega}^{(2)} = \begin{bmatrix} g_1 P_1 & \cdot \\ \cdot & g_2 P_2 \end{bmatrix}, \mathbf{\Omega}^{(3)} = \begin{bmatrix} g_1 P_1 + g_3 P_3 & \cdot \\ \cdot & g_2 P_2 \end{bmatrix},$$

and, therefore, the cumulative capacity values are

$$C_1 = C(g_1 P_1), C_2 = C(g_1 P_1) + C(g_2 P_2), C_3 = C(g_1 P_1 + g_3 P_3) + C(g_2 P_2);$$

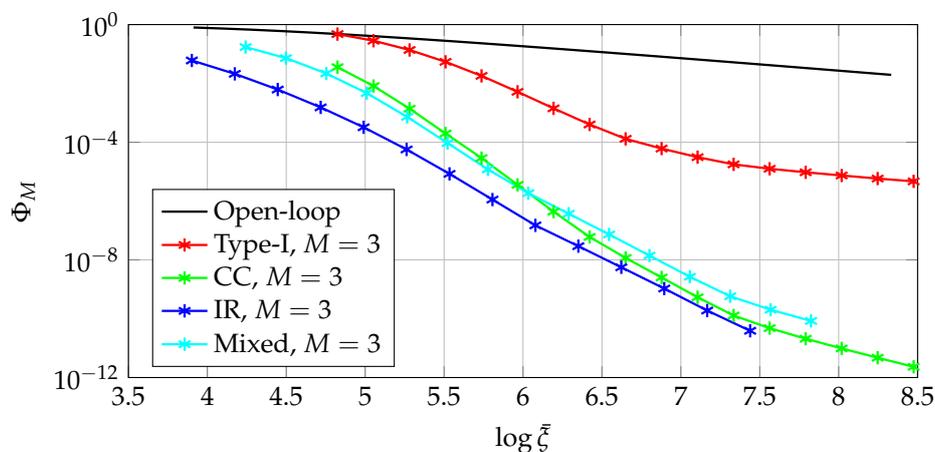
the same holds for dispersion  $V_j$ .

In the following, we will discuss the performance of the various HARQ schemes in terms of outage probability, power allocation, delivery delay, and energy efficiency.

#### 4.1. Outage Probability

The performance of the various HARQ schemes in terms of outage probability  $\Phi_M$  as a function of the average consumed energy  $\bar{\zeta}$  can be found in Figure 3. The open-loop system, i.e., the system with no retransmissions ( $M = 1$ ), is reported as baseline. As expected, all the HARQ schemes outperform the open-loop system thanks to the retransmission process employed. Moreover, all the Type-II HARQ schemes outperform the Type-I HARQ scheme: even if the number of transmission attempts we allow is very low, the performance of Type-II HARQ dramatically improves thanks to the effect of packet combination. Focusing on the performance of Type-II HARQ processes, the reader can infer that IR HARQ has the best performance in terms of outage probability with respect to the other two schemes. It is worth noticing that, for low energy budgets, i.e.,  $\log \bar{\zeta} < 6$ , mixed HARQ behaves better than CC HARQ, as expected because it is a mixture between IR and CC approaches. However, when  $\log \bar{\zeta} \geq 6$ , the CC scheme outperforms the mixed one. The reason for this has to do with the upper bound on the transmission power and will be explained in the following of the section. Finally, note that, as the energy budget  $\bar{\zeta}$  increases, the performance gap between Type-II HARQ processes narrows.

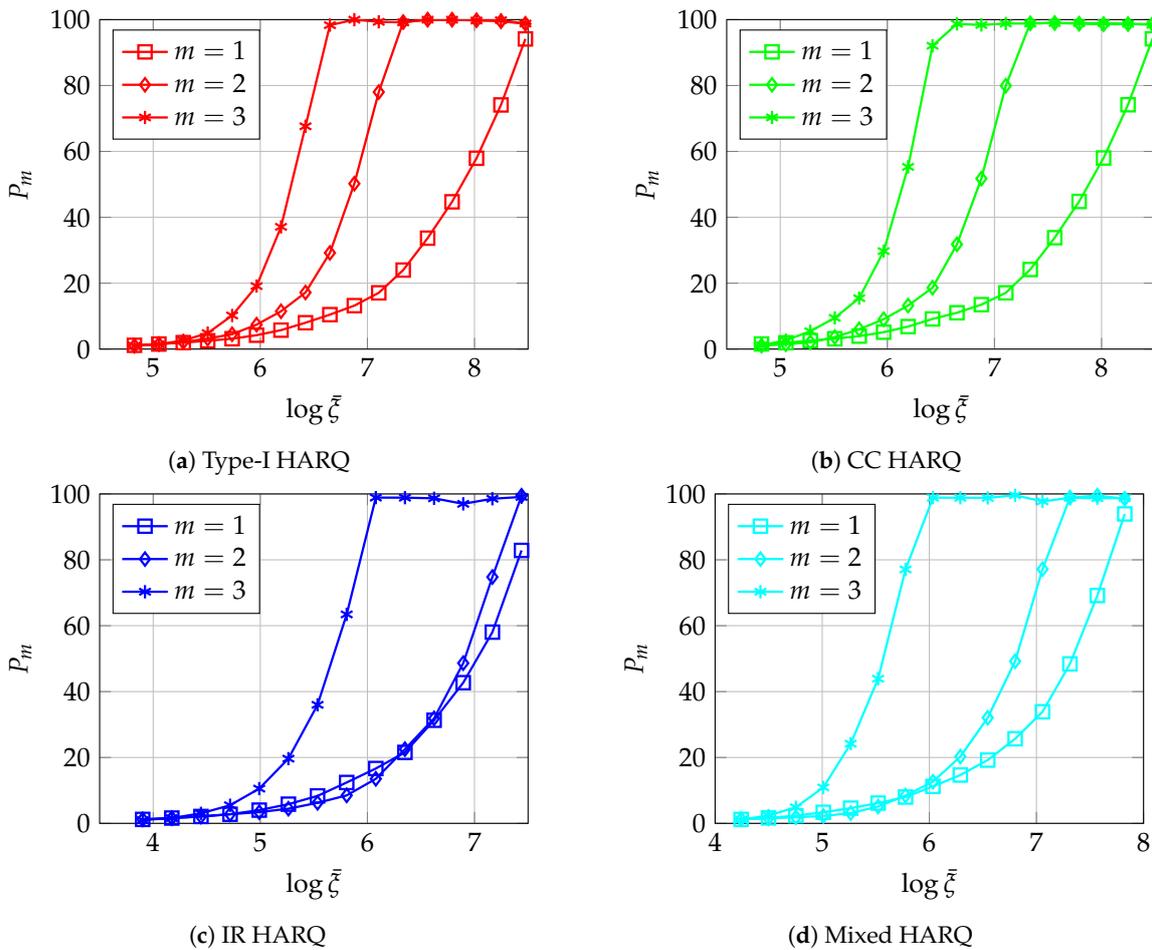
We remark that the  $10^{-5}$  outage probability target for URC is met soon by all kinds of Type-II HARQ, with an energy budget  $\bar{\zeta}$  in the interval  $[5.5, 6]$  (on a logarithmic scale). On the other side, Type-I HARQ is hardly capable of meeting the same target, achieving it only for high energy budgets. We also observe that the open-loop system cannot meet the requirements of URC, since it does not use multiple transmission attempts.



**Figure 3.** Outage probability  $\Phi_M$  vs. average consumed energy  $\bar{\zeta}$ . Note that  $\bar{\zeta}$  is represented on a logarithmic scale.

#### 4.2. Power Allocation

Figure 4 shows the power allocation  $P_m \forall m = 1, \dots, M$  vs. the average consumed energy  $\bar{\zeta}$ . In Figure 4a, it can be seen that, for Type-I HARQ processes, the optimal power allocation consists in always employing increasing transmission powers for successive transmission attempts, i.e.,  $P_1 < P_2 < P_3$ , confirming the results reported in [26]. A similar power allocation policy can be inferred from Figure 4b for the CC HARQ process, with the only difference that  $P_3$  grows faster than in the Type-I before the saturation point since the combining procedure of CC HARQ allows for allotting less energy to the first transmission attempt.



**Figure 4.** Power allocation  $P_m$  vs average consumed energy  $\bar{\xi}$ . Note that  $\bar{\xi}$  is represented on a logarithmic scale. (a–d) show the trends for the Type-I HARQ, CC HARQ, IR HARQ, and mixed HARQ case, respectively.

In the case of IR HARQ and mixed HARQ the strategy changes: the two schemes with variable rate show different trends of the power allocation. Figure 4c,d illustrate the power allocation results for IR HARQ and mixed HARQ, respectively. In the IR case, for  $\log \bar{\xi} < 6$ , i.e., before  $P_3$  saturates, the best power allocation strategy consists in using more power in the first transmission attempt than in the second one, yielding  $P_2 < P_1 < P_3$ . Indeed, since the first packet is transmitted using the highest rate, it is the least resilient and, therefore, more power is necessary. The optimal power allocation strategy, thus, suggests to use more power in the first transmission to minimize the outage probability. On the other hand, when  $\log \bar{\xi} \geq 6$ , it is better to allocate increasing power in successive transmission attempts. Regarding the mixed scheme, a similar policy can be inferred: before the saturation of  $P_3$ ,  $P_1$  is slightly higher than  $P_2$ ; on the other hand, after the saturation, one should employ increasing power levels again. Note that the gap between  $P_1$  and  $P_2$  when  $\log \bar{\xi} < 6$  is tighter than in the IR case. This is due to the effect of bit repetition: indeed, in mixed HARQ, the first transmission attempt is still the least resilient but will be transmitted again in the final transmission attempt. Therefore, thanks to the combination of two version of the first sub-codeword at the receiver side,  $P_1$  is lower than the IR case.

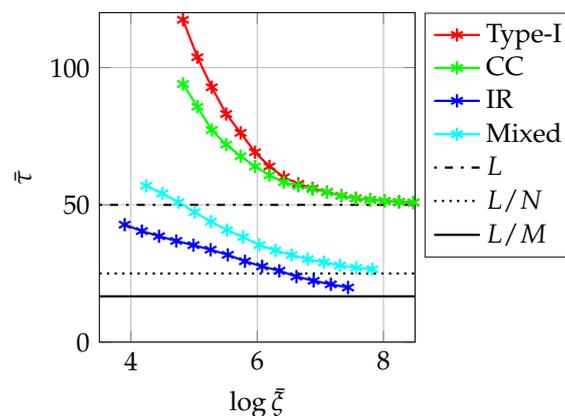
#### 4.3. Delivery Delay

In Figure 5, the average delivery delay  $\bar{\tau}$  of the four HARQ schemes is represented. Let us remark that this performance metric can also be considered as the *average buffer occupation* on the receiver side.

The curves show that IR HARQ provides the best performance in terms of delay, employing a lower number of channel uses with respect to the other HARQ schemes. On the contrary, the trends of Type-I and CC HARQ curves, instead, overlap as the energy budget  $\bar{\zeta}$  grows, since they employ the same number of channel uses in every transmission attempt and the outage probability  $\Phi_M$  tends to zero for both the schemes. Finally, the mixed HARQ provides an interesting performance in terms of delay, which is much lower than the CC trend.

Note that, as the energy budget grows, the average delay of the various schemes asymptotically converges to  $L$ ,  $L/N$ , and  $L/M$ , i.e., the delay of a single transmission attempt.

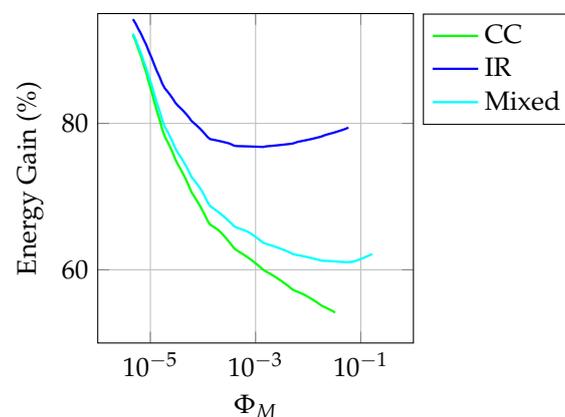
As for the low latency requirement in URC, the plots clearly state that the lowest delivery delay is provided by the IR and mixed variants of Type-II HARQ. Since CC HARQ does not fragment the original codeword, it provides a much higher delivery delay.



**Figure 5.** Average number of channel uses  $\bar{\tau}$  for Type-I, CC, IR, and mixed HARQ vs. average consumed energy  $\bar{\zeta}$ .

#### 4.4. Energy Efficiency

Figure 6 depicts the energy gain trend of Type-II HARQ processes with respect to Type-I HARQ, i.e., the energy gain provided by combining previously received packets before decoding. The reader can infer that an energy gain from 50% to a considerable 90% can be obtained using Type-II HARQ. The highest gain is provided by the IR scheme, as expected. Moreover, it can be seen that the energy gain decreases as the outage probability increases, i.e., as the energy budget  $\bar{\zeta}$  is lower, and that all the gains tend to converge to 90% as  $\Phi_M$  tends to zero.



**Figure 6.** Energy gap of CC, IR, and mixed HARQ with respect to Type-I HARQ vs. outage probability  $\Phi_M$ .

Finally, comparing these results with what we obtained in our previous work [35], we can state that the energy gain depends on the number of transmission attempts  $M$ . Indeed, increasing the value

of  $M$ , the achievable outage probability given a certain energy budget by Type-II HARQ schemes decreases. Therefore, the energy gain they can provide with respect to Type-I HARQ is higher if we increase  $M$ .

#### 4.5. Final Observations and Remarks

Based on the results obtained solving the optimization problem (9), we can make some considerations.

First of all, let us observe that Type-I and CC HARQ, which are constant-rate schemes, employ increasing transmission powers for all values of  $\bar{\zeta}$ . IR and mixed HARQ, which are, instead, variable-rate schemes, employ increasing transmission powers *only when*  $P_3$  saturates, i.e.,  $P_3 = P_{\max}$ . Therefore, we can state that introducing constraint (9d) in the optimization problem influences the power allocation of variable-rate schemes: there exists a threshold energy budget  $\bar{\zeta}^*$  (in this case, equal to  $e^6$ ) such that the power allocation policy changes. Moreover, we observe that, for all HARQ schemes, when  $P_3$  saturates, the slope of outage curves changes and they become less steep. This intuition can be exploited in the implementation of real transmission systems.

Secondly, we may observe that, despite providing the best performance in terms of outage and delivery delay, the IR schemes is the least flexible, since it cannot support a variable number of transmission attempts, while, given  $N$ , the mixed approach supports an arbitrary amount  $M$  of attempts. Intuitively, the performance of this two schemes converges for  $N/M \rightarrow 1$ . On the other hand, CC HARQ is the approach with the lowest implementation complexity at the terminal device, as it does not even need to segment codewords.

Finally, note that for extremely low energy budgets, the performance of Type-I and CC HARQ becomes very poor, while IR and mixed ones still work.

Table 2 summarizes the simulation results.

**Table 2.** Best Type-II HARQ approach given a target performance metric.

	CC	IR	Mixed
Outage probability		✓	
Delivery delay		✓	
Energy efficiency		✓	
Transmission flexibility			✓
Implementation complexity	✓		

## 5. Conclusions

In this paper, a novel comparison between all the different kinds of HARQ processes in the context of finite-blocklength, energy-efficient M2M communication has been provided. The reference scenario is really challenging and extremely timely, since low-complexity, energy-constrained devices will be employed in many IoT applications like, e.g., smart metering or environmental monitoring. As expected, Type-II HARQ approaches outperform the Type-I approach in terms of outage probability due to the combining process, providing up to a considerable 90% energy saving. Advantages and disadvantages of the various kinds of Type-II HARQ processes have been discussed, with a special reference to the emerging paradigm of URC. Power allocation policies have been derived solving the proposed optimization problem and some useful insights have been suggested for the implementation of practical transmission systems.

**Author Contributions:** Conceptualization, L.V. and M.C.; Formal analysis, L.V. and M.C.; Investigation, L.V. and M.C.; Methodology, L.V. and M.C.; Software, M.C.; Supervision, L.V.; Writing—Original Draft, L.V. and M.C.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AWGN	Additive White Gaussian Noise
CC	Chase Combining
C-RAN	Cloud-Radio Access Network
FEC	Forward Error Correction
HARQ	Hybrid ARQ
IoT	Internet of Things
IR	Incremental Redundancy
LTE	Long-Term Evolution
M2M	Machine-to-Machine
MAC	Medium Access Control
MCS	Modulation and Coding Scheme
MRC	Maximal Ratio Combining
MTD	Machine-Type Device
SNR	Signal-to-Noise Ratio
URC	Ultra-Reliable Communications

## References

1. Shannon, C. A mathematical theory of computation. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
2. Polyanskiy, Y.; Poor, H.; Verdú, S. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory* **2010**, *56*, 2307–2359. [[CrossRef](#)]
3. Erseghe, T. On the evaluation of the Polyanskiy-Poor-Verdú converse bound for finite block-length coding in AWGN. *IEEE Trans. Inf. Theory* **2015**, *61*, 6578–6590. [[CrossRef](#)]
4. Centenaro, M.; Vangelista, L.; Zanella, A.; Zorzi, M. Long Range Communications in Unlicensed Bands: The Rising Stars in the IoT and Smart City Scenarios. *IEEE Wirel. Commun.* **2016**, *23*, 60–67. [[CrossRef](#)]
5. Centenaro, M.; Vangelista, L. A study on M2M traffic and its impact on cellular networks. In Proceedings of the 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT), Milan, Italy, 14–16 December 2015; pp. 154–159.
6. Vodafone Group Plc. *New Study Item on Cellular System Support for Ultra Low Complexity and Low Throughput Internet of Things*; Technical Report GP-140421, 3GPP TSG GERAN#62; 3GPP: Sophia Antipolis, France, 2014.
7. Zhang, X.; Gao, Q.; Zhang, J.; Wang, G. Impact of transmit power on throughput performance in wireless ad hoc networks with variable rate control. *Comput. Commun.* **2008**, *31*, 3638–3642. [[CrossRef](#)]
8. Krunch, M.; Muqattash, A.; Lee, S.-J. Transmission power control in wireless ad hoc networks: Challenges, solutions and open issues. *IEEE Netw.* **2004**, *18*, 8–14. [[CrossRef](#)]
9. Popovski, P.; Nielsen, J.J.; Stefanovic, C.; de Carvalho, E.; Strom, E.; Trillingsgaard, K.F.; Bana, A.; Kim, D.M.; Kotaba, R.; Park, J.; et al. Wireless Access for Ultra-Reliable Low-Latency Communication: Principles and Building Blocks. *IEEE Netw.* **2018**, *32*, 16–23. [[CrossRef](#)]
10. Fettweis, G.P. The tactile Internet: applications and challenges. *IEEE Veh. Technol. Mag.* **2014**, *9*, 64–70. [[CrossRef](#)]
11. Caire, G.; Tuninetti, D. The throughput of hybrid-ARQ protocols for the Gaussian collision channel. *IEEE Trans. Inf. Theory* **2001**, *47*, 1971–1988. [[CrossRef](#)]
12. Cheng, J.F. Coding performance of hybrid ARQ schemes. *IEEE Trans. Commun.* **2006**, *54*, 1017–1029. [[CrossRef](#)]
13. Shi, Z.; Ma, S.; Tam, K.W. Outage analysis on Type I HARQ over time-correlated Rayleigh fading channels. In Proceedings of the 2015 IEEE/CIC International Conference on Communications in China (ICCC), Shenzhen, China, 2–4 November 2015; pp. 1–6.
14. Harsini, J.S.; Lahouti, F.; Levorato, M.; Zorzi, M. Analysis of non-cooperative and cooperative type II hybrid ARQ protocols with AMC over correlated fading channels. *IEEE Trans. Wirel. Commun.* **2011**, *10*, 877–889. [[CrossRef](#)]
15. Caire, G.; Taricco, G.; Biglieri, E. Optimum power control over fading channels. *IEEE Trans. Inf. Theory* **1999**, *45*, 1468–1489. [[CrossRef](#)]

16. Tumula, C.; Larsson, E.G. Optimal power allocation for hybrid ARQ with Chase combining in i.i.d. Rayleigh fading channels. *IEEE Trans. Commun.* **2013**, *61*, 1835–1846.
17. Ksairi, N.; Ciblat, P.; Martret, C.J.L. Near-optimal resource allocation for type-II HARQ based OFDMA networks under rate and power constraints. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 5621–5634. [[CrossRef](#)]
18. Erseghe, T. Coding in the finite-blocklength regime: Bounds based on Laplace integrals and their asymptotic approximations. *IEEE Trans. Inf. Theory* **2016**, *62*, 6854–6883. [[CrossRef](#)]
19. Yang, W.; Caire, G.; Durisi, G.; Polyanskiy, Y. Optimum power control at finite blocklength. *IEEE Trans. Inf. Theory* **2015**, *61*, 4598–4615. [[CrossRef](#)]
20. Hu, Y.; Gross, J.; Schmeink, A. On the capacity of relaying with finite blocklength. *IEEE Trans. Veh. Technol.* **2016**, *65*, 1790–1794. [[CrossRef](#)]
21. Ge, S.; Xi, Y.; Zhao, H.; Huang, S.; Wei, J. Energy efficient optimization for CC HARQ over block Rayleigh fading channels. *IEEE Commun. Lett.* **2015**, *19*, 1854–1857. [[CrossRef](#)]
22. Han, J.; Ge, S.; Xi, Y.; Wei, J. An energy-efficient design for multihop relay network adopting CC HARQ scheme. In Proceedings of the 2016 8th International Conference on Wireless Communications & Signal Processing (WCSP), Yangzhou, China, 13–15 October 2016; pp. 1–6.
23. Manhas, E.B.; Pellenz, M.E.; Brante, G.; Souza, R.D.; Rosas, F. Energy efficiency analysis of HARQ with chase combining in multihop wireless sensor networks. In Proceedings of the 2014 IEEE Symposium on Computers and Communications (ISCC), Madeira, Portugal, 23–26 June 2014; pp. 1–6.
24. Khalili, S.; Simeone, O. Uplink HARQ for Cloud RAN via separation of control and data planes. *IEEE Trans. Veh. Technol.* **2016**, *66*, 4005–4016. [[CrossRef](#)]
25. Kim, S.H.; Sung, D.K.; Le-Ngoc, T. Performance analysis of incremental redundancy type hybrid ARQ for finite-length packets in AWGN channel. In Proceedings of the 2013 IEEE Global Communications Conference (GLOBECOM), Atlanta, GA, USA, 9–13 December 2013; pp. 2063–2068.
26. Makki, B.; Svensson, T.; Zorzi, M. Finite block-length analysis of the incremental redundancy HARQ. *IEEE Wirel. Commun. Lett.* **2014**, *3*, 529–532. [[CrossRef](#)]
27. Han, J.-M.; Xi, Y.; Wei, J.-B. Performance Analysis of Multihop CC HARQ over Lognormal Shadowed Nakagami-m Fading Channels. In Proceedings of the 2017 International Conference on Information Science and Technology (IST 2017), Beijing, China, 18–20 October 2017; Volume 11.
28. Kim, S.; Yu, H. Energy-Efficient HARQ-IR for Massive MIMO Systems. *IEEE Trans. Commun.* **2018**, *66*, 3892–3901. [[CrossRef](#)]
29. Shariatmadari, H.; Duan, R.; Iraj, S.; Li, Z.; Uusitalo, M.A.; Jantti, R. Resource Allocations for Ultra-Reliable Low-Latency Communications. *Int. J. Wirel. Inf. Netw.* **2017**, *24*, 317–327. [[CrossRef](#)]
30. Ross, S.M. *Stochastic Processes*, 2nd ed.; Wiley: Hoboken, NJ, USA, 1995.
31. Polyanskiy, Y.; Poor, H.; Verdú, S. Dispersion of Gaussian channels. In Proceedings of the 2009 IEEE International Symposium on Information Theory, Seoul, Korea, 28 June–3 July 2009; pp. 2204–2208.
32. Goldsmith, A. *Wireless Communications*; Cambridge University Press: Cambridge, UK, 2005.
33. Centenaro, M.; Vangelista, L. HARQ in LTE uplink: A simple and effective modification suitable for low mobility users. In Proceedings of the 2015 IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015; pp. 3763–3769.
34. 3GPP. *Cellular System Support for Ultra-Low Complexity and Low Throughput Internet of Things (CIoT)*; Technical Report 45.820; 3GPP: Sophia Antipolis, France, 2015.
35. Centenaro, M.; Ministeri, G.; Vangelista, L. A comparison of energy-efficient HARQ protocols for M2M communication in the finite block-length regime. In Proceedings of the 2015 IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB), Montreal, QC, Canada, 4–7 October 2015; pp. 1–6.

