IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 December 2023, accepted 28 December 2023, date of publication 1 January 2024,
date of current version 8 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3348790

RESEARCH ARTICLE

Optimum Power Allocation for HARQ-Aided NOMA With Proportional Fairness on Fading Channels

GIORGIO TARICCO, (Fellow, IEEE)
Politecnico di Torino (DET), 10129 Turin, Italy
e-mail: gtaricco@ieee.org

ABSTRACT Power-Domain NOMA is one of the enabling technologies for future wireless communication networks of the fifth and sixth generations. This work addresses some key features of Power-Domain NOMA, including the impact of block fading on interference cancellation (leading to outage events), the limited channel state information available at the transmitter (consisting in the simple statistic distribution of the channel state), the fairness of the user achievable information rates (according to the Proportional Fairness criterion), and the optimization of the outage probability in the presence of a simple or hybrid ARQ protocol. After recalling some basic results on the achievable outage information rate region, the Proportional Fairness criterion is used to optimize the power allocation rates required to achieve specific outage probabilities. This is achieved by properly choosing the outage probabilities in conjunction with the hybrid ARQ protocol. To this purpose, Maximum Ratio Combining is used to enhance the achievable rate for multiple retransmissions. The system throughput analysis and optimization resort to a Markov chain representation of the hybrid ARQ protocol. This allows to assess the impact of retransmissions on the throughput. The latency involved is assessed by evaluating the average value and the standard deviation of the packet transmission delay. Numerical results are reported for two different system models: 1) symmetric scenario, where all users have the same average SNR, which varies according to Rayleigh fading; 2) asymmetric scenario for a single-cell broadcast channel, where the users are uniformly located over a disk and their average SNR depends on the distance from the transmitting base station at the center. For the latter scenario, the base station is assumed to know the user distances, which corresponds to a partial knowledge of the channel state at the transmitter. Both scenarios are thoroughly analyzed, and the impact of several system factors is discussed in detail. The results show, among other things, that very high outage probabilities may be required to optimize the throughput in low average SNR conditions, and that the optimum power allocation at the transmitter may reach a wide dynamic range when the SNR is large.

INDEX TERMS Power-domain non-orthogonal multiple access, successive interference cancellation, outage capacity region, proportional fairness, hybrid automatic repeat request.

I. INTRODUCTION

Non-Orthogonal Multiple Access (NOMA) is one of the enabling techniques envisaged for the deployment of Fifth and Sixth Generation (5G and 6G) wireless communication systems to reach their ambitious targets (Tbps data rates, microsecond latencies, ubiquitous connectivity, and so on) [1], [2], [3]. In a nutshell, NOMA is a multiple access communication strategy that allows multiple users or devices to share the same time/frequency/code resources in a non-orthogonal manner, so that signals from different users can occupy the same resource block (RB). Among the declinations of NOMA, an important role is played by Power-Domain NOMA (PD-NOMA), which is addressed

The associate editor coordinating the review of this manuscript and approving it for publication was Adao Silva.

in this paper. PD-NOMA consists of the allocation of different power levels to the set of user signal component sharing the same RB in a broadcast channel, allowing them to distinguish their signals at the receiver based on suitable processing methods. Traditional orthogonal multiple access (OMA) techniques, like time division multiple access (TDMA), frequency division multiple access (FDMA), and code division multiple access (CDMA), assign different users distinct time slots, frequency bands, or spreading codes, respectively, in order to avoid the presence of interference [4]. However, these techniques may lead to inefficient use of the available spectrum. On the contrary, PD-NOMA allows users to share the transmission channel resources by allocating different power levels to each of their signal components. As a general rule, under a fairness criterion, users with lower channel gains (experiencing worse or weaker signal conditions) are allocated higher power levels, while users with higher channel gains (experiencing better or stronger signal conditions) are allocated lower power levels. This is the opposite of the standard water-filling power allocation, where throughput instead of user fairness is considered [5]. To achieve these goals, NOMA implements superposition encoding at the transmitter through the knowledge of the channel gains of the different users, and the receivers resort to successive interference cancellation (SIC) [6] in order to retrieve the information addressed to them.

Even though PD-NOMA addresses the overall sharing of the available system resources, alternative NOMA techniques exist focusing on complexity limitation. One of them is *user pairing* [7]. This technique divides the user population into pairs which share common channel resources. Among the pairing schemes proposed in the literature, it is worth mentioning the following ones. *i*) *Random* schemes [8], where random pairing is implemented independently of the channel conditions of the individual users. *ii*) *Adjacent* pairing schemes [9]: in this case, user pairing is based on the proximity of channel states. *iii*) *Strong-weak* pairing schemes [10], where user pairing is based on the channel state so that, with an even number of users K sorted according to their channel state as $U_1, U_2, \ldots, U_K$, user U_i is paired with U_{K+1-i} for $i = 1, \ldots, K/2$.

The background of NOMA is rooted in the information theoretical *degraded Gaussian broadcast channel* model [5]. The properties of this communication channel are well established, and its capacity region is achieved by resorting to SIC, as illustrated in [11]. With perfect channel state information at the transmitter (CSIT),[1] the outage probability and the achievable ergodic sum rate were derived in [6] under the assumption of randomly located terminals inside a circular area. However, when only the statistic distribution of the channel gains is available at the transmitter (channel distribution information at the transmitter or CDIT), ergodic achievable rates must be replaced by outage rates since SIC cannot always be implemented in the most efficient way. In this framework, the outage rate is the rate at which a communication channel fails to meet a certain predefined performance criterion or quality of service (QoS) target. This can be represented by the outage probability, which is a measure of the system's reliability in adverse conditions. By setting outage rate targets, networks and protocols can be designed to provide an acceptable QoS even in challenging environments, like those encountered by NOMA systems. The basic framework for the outage analysis of PD-NOMA has been studied in [12] and [13]. In particular, the outage event is defined as the failure of a user to cancel the interference by SIC or to decode its own data [13].

The occurrence of outages can be overcome by repeating the transmission of the data packets which have been lost, and several specific techniques have been proposed in the literature. Specifically, Hybrid Automatic Repeat reQuest (HARQ) is a communication protocol used to improve the reliability of data transmission over unreliable channels, such as wireless or wired networks. HARQ merges the application of two techniques: Automatic Repeat reQuest (ARQ) and Forward Error Correction (FEC) coding. Then, it implements a hybrid approach which enhances the reliability of data transmission at the price of increasing the latency of the transmitted information packets (see, *e.g.*, [14], [15], [16], [17], [18], [19], [20], [21], [22], [23] for a general description of HARQ). In a nutshell, HARQ assumes that the data stream is divided into packets with a unique identification number, called sequence number. These packets correspond to codewords generated according to some FEC code. The receiver attempts to decode them and, if one codeword is successfully decoded (without any errors), it acknowledges its reception by sending a positive ACK back to the transmitter, and then it processes the next packet. Otherwise, if decoding errors occur, the receiver sends a negative ACK (NACK) to the sender, indicating that the packet needs to be retransmitted. After receiving the retransmitted signal, it combines the portion of received signal corresponding to this packet with those corresponding to the other previously received erroneous packets with the same sequence number, and tries to decode their soft combination [15], [18], [19], [20]. The process continues until the receiver successfully decodes the packet or exhausts a predefined number of retransmission attempts. HARQ has commonly been used in LTE (Long-Term Evolution) and current 5G wireless systems and is envisaged for the enhancement of NOMA when perfect CSIT is not available [7], [24], [25].

A. CONTRIBUTIONS

The contributions of this work are summarized as follows. Though the outage information rate region has been studied in the literature, the impact of fairness requirements and the outage probability optimization in conjunction with a proper retransmission protocol have not yet been analyzed. Literature results assume that specific outage probabilities and power allocation are fixed to derive the corresponding

[1]Throughout this paper, perfect channel state information at the receiver (CSIR) is always assumed to be available for all users.

user information rates. Therefore, there is no attempt to implement fairness policies by using a retransmission protocol in the way proposed here. Three main contributions are claimed for this work. First, for given outage probabilities, the user information rate satisfying the proportional fairness (PF) criterion is derived with CDIT. Then, the PF information rate is optimized by considering the use of a HARQ protocol modeled by a Markov chain whose transition probabilities depend on the outage probabilities. This optimization involves the evaluation of the average number of packet transmissions with HARQ and provides the average value and the standard deviation of the transmission delay. Finally, the results are applied to two scenarios of interest. The former is based on a symmetric user arrangement with a wireless channel affected by Rayleigh block fading. The latter involves an asymmetric user scenario where the users are uniformly randomly located in a disk with the transmitting base station at the center. The average channel attenuation depends on the distance from the base station and on a certain propagation exponent. Additionally, block Rayleigh fading also affects the channel gains. For both scenarios, the ARQ and HARQ achievable rates are evaluated analytically and illustrated by numerical results, along with the corresponding optimum outage probabilities, retransmission delays, and transmitter power allocations.

B. ORGANIZATION

The paper is organized as follows. Section II introduces the system model, the concepts of CSIT and CDIT, the requirement for SIC, and the outage achievable rates derived in the literature. Section III recalls the PF approach developed in [26] in the framework of outage achievable rate analysis. Section IV provides a description of the ARQ and HARQ protocols through a Markov chain model and the corresponding transition probabilities. This description considers, for the HARQ protocol, the Maximum Ratio Combining of the sequential components of the received signal corresponding to retransmissions, along with the specific assumptions the protocol is based upon. The average number of transmissions due to HARQ is then derived, along with the mean and standard deviation of the delay and its counter-cumulative distribution function. Accordingly, the optimum PF rates with ARQ and HARQ are determined. Next, Section V discusses two scenarios of interest. The former is based on a symmetric user arrangement and the latter on an asymmetric one, both assuming that the users are uniformly randomly distributed over a disk area, and that Rayleigh fading affects the signal propagation. These scenarios are addressed analytically in detail, and numerical optimization is employed to obtain the results illustrated in the section. Finally, Section VI summarizes the contents and provides some concluding remarks on this study.

II. SYSTEM MODEL

The system model of the downlink PD-NOMA is described in this section. There is a base station (BS) transmitting

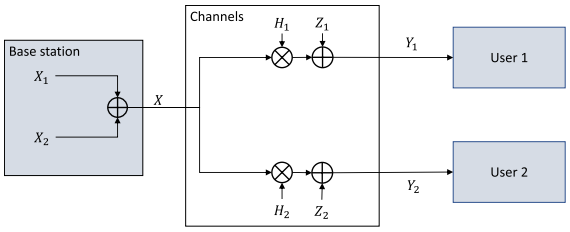

FIGURE 1. Block diagram of a two-user downlink NOMA scheme (broadcast channel).

the signals represented by the random variables $X_1, \ldots, X_K$ addressing K users with channel gains $H_1, \ldots, H_K$. The case of $K = 2$ users is illustrated by the block diagram in Fig. 1. The transmitted signal is a linear combination of the user signals, which can be implemented by superposition channel encoding.

A. DETERMINISTIC CASE WITH PERFECT CSIT

If the channel gains are known at the transmitter (BS), they can be sorted in nonincreasing order of their magnitudes, possibly after relabeling the user identifiers, i.e.,

$$|H_1| \geq |H_2| \geq \cdots \geq |H_K|. \tag{1}$$

The corresponding wireless transmission system is commonly referred to as having perfect channel state information at the transmitter (CSIT), and is a baseline reference found in classic works like [27] and [28]. The reference system model is the *Gaussian broadcast channel*, where the transmitter implements superposition encoding and the receivers implement Successive Interference Cancellation (SIC) to optimally decode their signals. The corresponding channel equations are, for each user:

$$Y_k = H_k X + Z_k, \;\; Z_k \sim \mathcal{CN}(0, 1), \;\; k = 1, \ldots, K, \tag{2}$$

where:

$$X = X_1 + \ldots + X_K. \tag{3}$$

Additionally, the available power for the user signal components in the transmitted signal is limited by the following inequalities:

$$\mathbb{E}[|X_k|^2] \leq P_k = \alpha_k P_x, \;\; \text{where} \sum_{k=1}^{K} \alpha_k = 1. \tag{4}$$

Here, P_x represents the total available transmitted power since the power components are independent with zero mean values. The optimum decoding order of the SIC receiver requires that the k-th receiver first decodes the $(k-1)$ stronger signals $X_1, \ldots, X_{k-1}$ and then its own signal X_k. Thus, the following information rates are achievable at the K different user receivers [5], [27], [28]:

$$R_k < \log_2 \left(1 + \frac{\alpha_k}{\rho_k^{-1} + \sum_{\ell=1}^{k-1} \alpha_\ell} \right), \;\; k = 1, \ldots, K. \tag{5}$$

Here, the user Signal-to-Noise power Ratios (SNRs) are given by:

$$\rho_k \triangleq |H_k|^2 P_x. \tag{6}$$

Note that the user SNRs must be ordered by the inequalities:

$$\rho_1 \geq \rho_2 \geq \cdots \geq \rho_K \tag{7}$$

to maximize the sum rate. The capacity achieving distributions are $X_k \sim \mathcal{CN}(0, \alpha_k P_x)$, where each $\alpha_k \in \mathcal{S}_K$, defined as the K-dimensional simplex:

$$\mathcal{S}_K \triangleq \{\boldsymbol{\alpha} : \alpha_k \geq 0, k = 1, \ldots, K, \sum_{k=1}^{K} \alpha_k = 1\}. \tag{8}$$

B. UNKNOWN CSIT – KNOWN CDIT

It is of great practical interest to determine the achievable information rates when the transmitter doesn't know exactly the channel gains but only their distribution. In this case, one can says that there is (perfect or ideal) channel distribution information at the transmitter (CDIT). This has been studied in [12] and [13], the latter specifically for a NOMA communication system with SIC. By this assumption, the transmitter is unable to sort the channel gains according to their magnitudes because they are unknown random variables whose values (characterizing the optimum sequential ordering of SIC) depend on the realization.

On the positive side, the propagation channels vary slowly in time (compared to the transmission rate) so that every transmitted code word is affected (approximately) by a fixed realization of the channel gains $|H_k|^2$ (this is referred to as the quasi-static fading channel model). The standard paradigm for quasi-static fading channels is resorting to the evaluation of outage probabilities, which are the probabilities that each user *fails to cancel the interference successively or fails to decode the information for itself* in the asymptotic Shannon decoding regime [13]. Following this approach, an outage probability vector is defined:[2]

$$\boldsymbol{\epsilon} \triangleq (\epsilon_1, \ldots, \epsilon_K), \tag{9}$$

along with the inverse cumulative distribution function (ICDF) of the SNR realizations $\rho_k = |H_k|^2 P_x$, i.e.,[3]

$$P(\rho_k \leq G_k(\epsilon_k)) = \epsilon_k. \tag{10}$$

The resulting achievable information rates, satisfying the outage probability vector $\boldsymbol{\epsilon}$ under the assumption that:

$$G_1(\epsilon_1) \geq G_2(\epsilon_2) \geq \cdots \geq G_K(\epsilon_K), \tag{11}$$

are given by:

$$R_k \leq \mathcal{R}_k(\boldsymbol{\alpha}, \boldsymbol{\epsilon}) \triangleq \log_2\left(1 + \frac{\alpha_k}{G_k(\epsilon_k)^{-1} + \sum_{\ell=1}^{k-1} \alpha_\ell}\right). \tag{12}$$

[2]The outage event corresponds to a single packet transmission.

[3]In case of discontinuous cumulative distribution function, the definition can be replaced by $G_k(t) \triangleq \sup_x \{x : P(\rho_k \leq x) \leq t\}$.

These equations characterize the outage capacity region of the NOMA broadcast channel [12, Th.1]. The achievability part of the theorem is easily checked because the occurrence of an outage for user k corresponds to the event:

$$\log_2\left(1 + \frac{\alpha_k}{\rho_k^{-1} + \sum_{\ell=1}^{k-1} \alpha_\ell}\right) \leq \log_2\left(1 + \frac{\alpha_k}{G_k(\epsilon_k)^{-1} + \sum_{\ell=1}^{k-1} \alpha_\ell}\right), \tag{13}$$

equivalent to $\rho_k \leq G_k(\epsilon_k)$, whose probability is ϵ_k, by the definition of $G_k(\epsilon_k)$ in (10).

III. FAIRNESS

Optimizing the achievable sum-rate in deterministic PD-NOMA corresponds to allowing only the strongest user (with maximum SNR ρ_k) to access the transmission channel. According to the encoding scheme illustrated in the previous section, this corresponds to setting $\alpha_1 = 1$ and $\alpha_k = 0$ for all $k > 1$, by inequality (7). This is considered *unfair* since fairness corresponds to the equitable allocation of resources among the users (or devices) that share a common resource, like a wireless communication channel [29].

Different types of fairness for NOMA communication systems have been proposed in the literature (see [26], [29], [30] and references therein). One of the most flexible is *proportional fairness* (PF) [30], where the users with better channel conditions are allowed to transmit more data, but not to the extent that it completely starves the other users. In other words, the goal is to strike a balance between maximizing system throughput and ensuring that all users have some level of service. More precisely, PF aims at the maximization (over the possible power allocations) of the minimum ratio between the user achievable rates R_k and certain user target rates T_k. To that purpose, the following minimum-ratio function is defined:

$$\Psi_{\min}(\boldsymbol{\alpha}, \boldsymbol{\epsilon}) \triangleq \min\left\{\frac{\mathcal{R}_1(\boldsymbol{\alpha}, \boldsymbol{\epsilon})}{T_1}, \ldots, \frac{\mathcal{R}_K(\boldsymbol{\alpha}, \boldsymbol{\epsilon})}{T_K}\right\}. \tag{14}$$

The target rates may consider the channel conditions and be higher for stronger users and lower for weaker users but nevertheless should follow a tempered dynamic distribution. When $T_k = 1$ for all $k = 1, \ldots, K$, PF reduces to MAX-MIN fairness, consisting of the maximization of the minimum achievable rate. The power allocation corresponding to PF is obtained by solving the following optimization problem [31]:

$$\begin{cases} \max & \Psi_{\min}(\boldsymbol{\alpha}, \boldsymbol{\epsilon}) \\ s.t. & \alpha_k \geq 0, \quad k = 1, \ldots, K \\ & \sum_{k=1}^{K} \alpha_k = 1 \end{cases} \tag{15}$$

This is a nonconvex optimization problem which has been addressed, for the deterministic case, in [26]. By [26, Th.1], the solution is equivalent to solving the following set of

equations:

$$\frac{\mathcal{R}_1(\boldsymbol{\alpha}, \boldsymbol{\epsilon})}{T_1} = \cdots = \frac{\mathcal{R}_K(\boldsymbol{\alpha}, \boldsymbol{\epsilon})}{T_K}. \tag{16}$$

If $T_k = 1$ (MAX-MIN fairness), these equations are equivalent to a polynomial equation of degree K in α_1 [26]. More generally, an efficient iterative algorithm is proposed in [26, Algorithm 1], which allows to solve the optimal power allocation problem for any combination of target rates.

Remark 1: The application of PF to a wireless scenario depends very much on the user SNR dynamics. If the SNR range is limited, PF can be implemented by setting $T_k = 1$ (or any constant), which leads to MAX-MIN fairness. Otherwise, in case of a very wide SNR range, MAX-MIN fairness implies that all users achieve a very low rate. Then, the stronger users may be exceedingly damaged by the presence of weaker users. A simple alternative (see [26] for the perfect CSIT case) is setting the target rates T_k equal to the achievable rates in the case of single user transmission, i.e., $T_k = \log_2(1 + \rho_k)$. A generalization of this approach may consist of setting $T_k = \{\log_2(1 + \rho_k)\}^\beta$ for some exponent $\beta \in [0, 1]$. This corresponds to a continuous shift from MAX-MIN fairness ($\beta = 0$) to single-user achievable rate PF ($\beta = 1$).

The previous developments and most of the literature are based on the selection of specific values for the outage probabilities, which determine the outage achievable rates. However, it is not clear from the literature how one should choose these *parameters*, i.e., the vector $\boldsymbol{\epsilon}$, and the following sections shed a light on the subject.

IV. ARQ AND HYBRID ARQ

This section considers the use of ARQ and hybrid ARQ (HARQ) to recover transmission losses due to decoding errors. This is done by using the automatic retransmission scheme discussed in Section I and in the references cited therein. Within this framework, assume that retransmission occurs whenever any user is unable to decode the corresponding codeword by using the current and (possibly) previous received signals. Assume that the current time slot corresponds to the first transmission of the codewords and that the codeword symbols are represented by the random variables $X_k, k = 1, \ldots, K$. Then, if r transmissions occur, the following channel equations hold (independent replicas of eqs. (2)):

$$Y_{k,i} = H_{k,i}(X_1 + \ldots + X_K) + Z_{k,i}, \tag{17}$$

for $i = 1, \ldots, r, k = 1, \ldots, K$. Here, the channel gains $H_{k,i}$ and the noise samples $Z_{k,i}$ are independent, and the SNRs

$$\rho_{k,i} \triangleq |H_{k,i}|^2 P_x \tag{18}$$

are defined. The different characteristics of the ARQ and HARQ schemes are summarized in following.

In both cases, ARQ and HARQ, retransmission occurs whenever any user is unable to decode its codeword and affects all the transmitted user codewords. However, with ARQ, decoding considers only the current received signal and ignores the previous ones. Therefore, retransmission occurs if any of the following inequalities holds:

$$R_\ell > \log_2\left(1 + \frac{\alpha_\ell}{\rho_{k,i}^{-1} + \sum_{m=1}^{\ell-1} \alpha_m}\right), \tag{19}$$

for $1 \le \ell \le k \le K$.

On the other hand, with HARQ, decoding also takes into account the signal fragments received during the previous time slots corresponding to the same codewords. By assuming that there have been r retransmissions, exploiting the perfect CSIR assumption, and applying Maximum Ratio Combining (MRC) [18] to the r received signal blocks at each of the K receivers, the equivalent channel equations are:

$$\tilde{Y}_{k,r} = \frac{\sum_{i=1}^{r} H_{k,i}^* Y_{k,i}}{\tilde{H}_{k,r}} = \tilde{H}_{k,r}(X_1 + \ldots + X_K) + \tilde{Z}_{k,r}. \tag{20}$$

Here, the equivalent channel gains

$$\tilde{H}_{k,r} \triangleq \left\{\sum_{i=1}^{r} |H_{k,i}|^2\right\}^{1/2} \tag{21}$$

are obtained by applying MRC and the equivalent noise components, denoted by $\tilde{Z}_{k,r}$, are independent circularly symmetric complex Gaussian random variables distributed according to $\mathcal{N}(0, 1)$. The transmission of multiple copies of the same codeword reduces the error probability because it increases the SNR after MRC. Meanwhile, it increases the latency and reduces the system throughput. If r transmissions of the same codeword occurred and setting

$$\tilde{\rho}_{k,r} \triangleq |\tilde{H}_{k,r}|^2 P_x. \tag{22}$$

the codeword is still retransmitted if any of the following inequalities holds:

$$R_\ell > \log_2\left(1 + \frac{\alpha_\ell}{\tilde{\rho}_{k,r}^{-1} + \sum_{m=1}^{\ell-1} \alpha_m}\right), \tag{23}$$

for $1 \le \ell \le k \le K$. That is, if any user k fails to decode the codeword intended for another user ℓ with $1 \le \ell < k$ or to decode its own codeword, according to the definition in [13]. This definition extends the previous one given in eq. (18). The random process is described by the state diagram illustrated in Fig. 2. In this representation, each state is labeled by the number of transmissions which have occurred at a given time. This number includes the first transmission and, possibly, the number of retransmissions of the same codeword obtained by superposition encoding. State 1 is the initial state corresponding to the first transmission. A state labeled by r (corresponding to the first transmission plus $(r - 1)$ retransmissions) has two possible transitions: *i*) to the state labeled $r+1$ because for at least one user the SNR after MRC, $\tilde{\rho}_{k,r}$, is not sufficiently large for some user, i.e., $\tilde{\rho}_{k,r} < G_k(\epsilon_k)$ for at least one index k with $1 \le k \le K$; *ii*) to the state labeled ∞ when the opposite condition $\tilde{\rho}_{k,r} > G_k(\epsilon_k)$ for all k with $1 \le k \le K$ occurred. Since $\tilde{\rho}_{k,r}$ grows monotonically with the number of transmissions r, eventually the state ∞ is

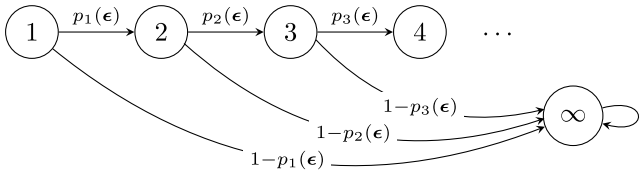

FIGURE 2. Markov chain description of the HARQ error probabilities. The state labels represent the number of transmissions. The state labeled "∞" corresponds to the eventual correct reception of all the user codewords (after a sufficient number of transmissions).

reached. Then, the evolution of the HARQ random process is derived by using the following probabilities:

$$\pi_{k,r}(\boldsymbol{\epsilon}) \triangleq P(\tilde{\rho}_{k,r} > G_k(\epsilon_k)), \tag{24}$$

where $\pi_{k,1}(\boldsymbol{\epsilon}) \equiv 1-\epsilon_k$. Since the equivalent random gains $\tilde{H}_{k,r}$ (and hence the SNR's after MRC $\tilde{\rho}_{k,r}$) are independent for different user indexes k, the HARQ process continues with probability[4]

$$p_r(\boldsymbol{\epsilon}) \triangleq 1 - \prod_{k=1}^{K} \pi_{k,r}(\boldsymbol{\epsilon}) \tag{25}$$

Denoting by $N(\boldsymbol{\epsilon})$ the number of HARQ transmissions (including the first one, hence ≥ 1), the average transmission rate with HARQ is reduced by the average value of $N(\boldsymbol{\epsilon})$, which can be calculated as follows:

$$\begin{aligned} \nu(\boldsymbol{\epsilon}) \triangleq \mathbb{E}[N(\boldsymbol{\epsilon})] &= \sum_{r=1}^{\infty} r(1-p_r(\boldsymbol{\epsilon})) \prod_{i=1}^{r-1} p_i(\boldsymbol{\epsilon}) \\ &= \sum_{r=1}^{\infty} r[q_r(\boldsymbol{\epsilon}) - q_{r+1}(\boldsymbol{\epsilon})] = \sum_{r=1}^{\infty} q_r(\boldsymbol{\epsilon}), \end{aligned} \tag{26}$$

where $q_r(\boldsymbol{\epsilon}) \triangleq \prod_{i=1}^{r-1} p_i(\boldsymbol{\epsilon})$ and $q_1(\boldsymbol{\epsilon}) = 1$. Accordingly, the achievable rate with HARQ for the k-th user becomes:

$$\mathcal{R}_{k,HARQ}(\boldsymbol{\alpha}, \boldsymbol{\epsilon}) = \frac{\mathcal{R}_k(\boldsymbol{\alpha}, \boldsymbol{\epsilon})}{\nu(\boldsymbol{\epsilon})}. \tag{27}$$

Since the scaling factor $\nu(\boldsymbol{\epsilon})$ is independent of k, the PF-HARQ optimization problem is equivalent to determining:

$$\max_{\boldsymbol{\epsilon}\in(0,1)^K} \left\{ \frac{1}{\nu(\boldsymbol{\epsilon})} \max_{\boldsymbol{\alpha}\in\mathcal{S}_K} \left[\min_{1\leq k\leq K} \frac{\mathcal{R}_k(\boldsymbol{\alpha}, \boldsymbol{\epsilon})}{T_k} \right] \right\}. \tag{28}$$

The previous analysis can be simplified by using an upper bound on the average number of packet transmissions $\nu(\boldsymbol{\epsilon})$. By the definition from eq. (24), $\pi_{k,r}(\boldsymbol{\epsilon}) \geq \pi_{k,1}(\boldsymbol{\epsilon}) = 1-\epsilon_k$. Thus,

$$p_r(\boldsymbol{\epsilon}) \leq 1 - \prod_{k=1}^{K} (1-\epsilon_k). \tag{29}$$

[4]Though it is not immediately clear by the definition, these probabilities are conditional on the originating state since the conditioning clause is embedded in the definition of the SNR $\tilde{\rho}_{k,r}$.

Then, the average number of HARQ transmissions can be upper bounded as:

$$\nu(\boldsymbol{\epsilon}) \leq \sum_{r=1}^{\infty} \left\{ 1 - \prod_{k=1}^{K} (1-\epsilon_k) \right\}^{r-1} = \prod_{k=1}^{K} \frac{1}{1-\epsilon_k}. \tag{30}$$

Therefore, the achievable rate with HARQ for the k-th user is lower bounded by:

$$\begin{aligned} \mathcal{R}_{k,HARQ}(\boldsymbol{\alpha}, \boldsymbol{\epsilon}) &\geq \mathcal{R}_{k,ARQ}(\boldsymbol{\alpha}, \boldsymbol{\epsilon}) \\ &\triangleq \mathcal{R}_k(\boldsymbol{\alpha}, \boldsymbol{\epsilon}) \prod_{k=1}^{K} (1-\epsilon_k). \end{aligned} \tag{31}$$

The lower bound is the achievable rate with ARQ, according to which the earlier received signal components are not used to improve the chances of correct decoding. The variance of the number of HARQ transmissions $N(\boldsymbol{\epsilon})$ can be of interest, too, and can be obtained in a similar way as the average value. Then,

$$\begin{aligned} \mathbb{E}[N(\boldsymbol{\epsilon})^2] &= \sum_{r=1}^{\infty} r^2 (1-p_r(\boldsymbol{\epsilon})) q_r(\boldsymbol{\epsilon}) \\ &= \sum_{r=1}^{\infty} r^2 q_r(\boldsymbol{\epsilon}) - \sum_{r=1}^{\infty} r^2 q_{r+1}(\boldsymbol{\epsilon}) \\ &= \sum_{r=1}^{\infty} [r^2 - (r-1)^2] q_r(\boldsymbol{\epsilon}) = \sum_{r=1}^{\infty} (2r-1) q_r(\boldsymbol{\epsilon}). \end{aligned} \tag{32}$$

Hence, the variance is

$$\sigma^2(\boldsymbol{\epsilon}) = 2\sum_{r=1}^{\infty} r q_r(\boldsymbol{\epsilon}) - \nu(\boldsymbol{\epsilon}) - \nu(\boldsymbol{\epsilon})^2. \tag{33}$$

The mean and variance of $N(\boldsymbol{\epsilon})$ provide an estimate of the HARQ delay range. Additionally, the counter-cumulative probability distribution is given by the probabilities:

$$\begin{aligned} P(N(\boldsymbol{\epsilon}) \geq r) &= \sum_{i=r}^{\infty} (1-p_i(\boldsymbol{\epsilon})) \prod_{j=1}^{i-1} p_j(\boldsymbol{\epsilon}) \\ &= \sum_{i=r}^{\infty} (q_i(\boldsymbol{\epsilon}) - q_{i+1}(\boldsymbol{\epsilon})) = q_r(\boldsymbol{\epsilon}). \end{aligned} \tag{34}$$

This result can also be derived by inspection of the Markov chain in Fig. 2.

V. APPLICATIONS

This section reports two selected applications of the previous results. First, a symmetric scenario is considered, where all users have the same average gain and outage probability requirements. The outage achievable rates are derived explicitly in the case of iid Rayleigh fading with $H_k \sim \mathcal{CN}(0,1)$. Next, the HARQ gain is obtained by optimizing the outage probability.

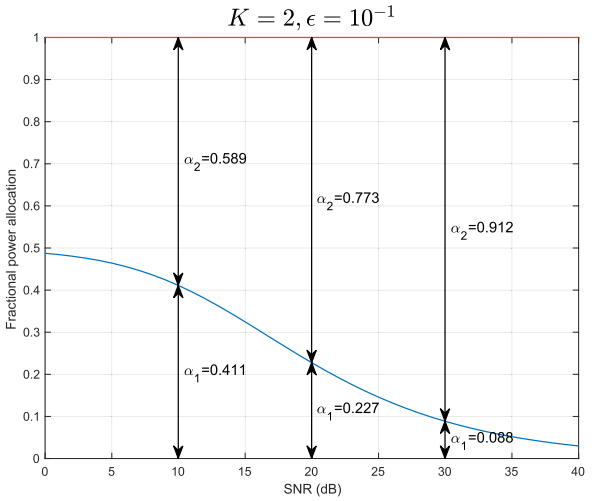

FIGURE 3. Fractional power allocation versus SNR optimizing the outage achievable rate with MAX-MIN fairness criterion for $K = 2$ users with Rayleigh fading and outage probability $\epsilon = 10^{-1}$.

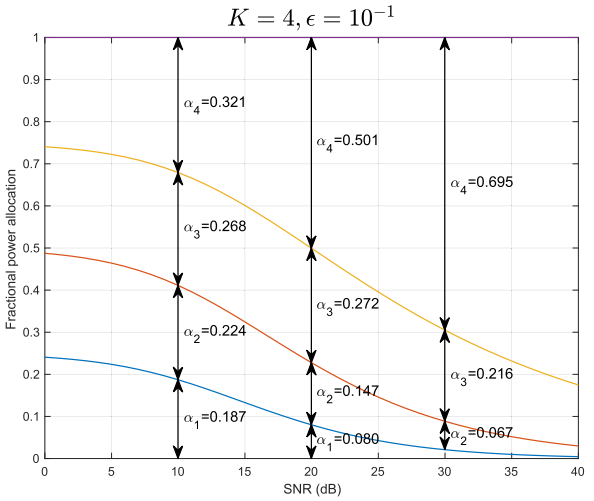

FIGURE 4. Same as Fig. 3 but for $K = 4$ users.

A. SYMMETRIC USER SCENARIO

This case assumes that all users have the same channel gain distribution, i.e.,

$$P(|H_k|^2 \le x) = \tilde{F}(x) \tag{35}$$

Therefore, $G_k(t) = P_x \tilde{G}(t)$, where $\tilde{G}(t)$ is the ICDF of $|H_k|^2$. Since the noise powers and the average channel power gain are all equal to 1, P_x is also the SNR. The ICDF $\tilde{G}(t)$ can be evaluated analytically in some cases, as illustrated in Table 1 on page 2334. For fixed outage probabilities, i.e., $\epsilon_k = \epsilon$ for $k = 1, \ldots, K$, the rate functions become

$$\mathcal{R}_k(\boldsymbol{\alpha}, \epsilon) = \log_2 \left(1 + \frac{\alpha_k}{G(\epsilon)^{-1} + \sum_{\ell=1}^{k-1} \alpha_\ell} \right). \tag{36}$$

When $T_k = 1$ (MAX-MIN fairness), the optimum power allocation derives by solving the following set of nonlinear equations, for $k = 1, \ldots, K-1$:

$$\frac{\alpha_k}{G(\epsilon)^{-1} + \sum_{\ell=1}^{k-1} \alpha_\ell} = \frac{\alpha_{k+1}}{G(\epsilon)^{-1} + \sum_{\ell=1}^{k} \alpha_\ell}. \tag{37}$$

The solution of these equations yields the optimum power allocation coefficients:

$$\alpha_{k,\mathrm{opt}} = \frac{[1 + G(\epsilon)]^{\frac{k}{K}} - [1 + G(\epsilon)]^{\frac{k-1}{K}}}{G(\epsilon)} \tag{38}$$

for $k = 1, \ldots, K$. Correspondingly, the outage MAX-MIN achievable rate per user is:

$$\frac{1}{K} \log_2[1 + G(\epsilon)]. \tag{39}$$

Figs. 3 and 4 illustrate the optimum power allocation with Rayleigh fading, $K = 2$ and 4 users, respectively, and outage probability $\epsilon = 10^{-1}$. It can be noticed that the optimum power allocation dynamics (the ratio α_K/α_1) increases with the SNR and decreases with the number of users K.

The following optimization step requires to determine the outage probability ϵ maximizing the achievable HARQ rate. To this purpose, the probabilities $p_r(\epsilon)$ are required. In the case of Rayleigh fading, $G(\epsilon) = P_x \tilde{G}(\epsilon) = -P_x \ln(1-\epsilon)$ and[5]

$$P(\tilde{H}_{k,r}^2 > x) = \tilde{\gamma}(r, x) \triangleq \frac{\Gamma(r, x)}{\Gamma(r)}. \tag{40}$$

Hence,

$$\pi_{k,r}(\boldsymbol{\epsilon}) = \tilde{\gamma}(r, -\ln(1-\epsilon)), \tag{41}$$

$$p_r(\boldsymbol{\epsilon}) = 1 - \pi_{k,r}(\boldsymbol{\epsilon})^K, \tag{42}$$

$$q_r(\boldsymbol{\epsilon}) = \prod_{i=1}^{r-1} p_i(\boldsymbol{\epsilon}), \quad \nu(\epsilon) = \sum_{r=1}^{\infty} q_r(\boldsymbol{\epsilon}). \tag{43}$$

By using this result, the HARQ MAX-MIN throughput per user is given by:

$$\mathcal{R}_{HARQ} = \frac{1}{K} \max_{0 \le \epsilon \le 1} \frac{\log_2(1 - P_x \ln(1-\epsilon))}{\nu(\epsilon)}. \tag{44}$$

With ARQ,

$$\mathcal{R}_{ARQ} = \frac{1}{K} \max_{0 \le \epsilon \le 1} (1-\epsilon)^K \log_2(1 - P_x \ln(1-\epsilon)). \tag{45}$$

Plainly, $\mathcal{R}_{ARQ} \le \mathcal{R}_{HARQ}$.

Fig. 5 shows the HARQ and ARQ MAX-MIN throughput for the Rayleigh fading symmetric user scenario considered with $K = 4$ users. The optimum outage probabilities are shown in Fig. 6 for ARQ and HARQ. As a simple optimality check, Fig. 5 also reports the throughput corresponding to the fixed outage probabilities $\epsilon = 0.2, 0.9$ for HARQ and $\epsilon = 0.1, 0.2$ for ARQ. The throughput curves corresponding to these outage probabilities are all below the optimal curves. For example, considering an SNR $\approx$ 0 dB, the optimum outage probability is ≈ 0.9 so that the optimal curve and the

[5]The function $\gamma(r, x) \triangleq \int_0^x u^{r-1} \mathrm{e}^{-u} du$ is the incomplete Gamma function [33, 8.35].

TABLE 1. ICDF for different random fading processes. Here, $Q(x) \triangleq \int_x^\infty \mathrm{e}^{-u^2/2}\frac{du}{\sqrt{2\pi}}$, $\Gamma(m,x) \triangleq \int_x^\infty u^{m-1}\mathrm{e}^{-u}du$, and $\Gamma(m) \triangleq \Gamma(m,0)$ [33, 8.35].

Fading type	Characterization	Normalized ICDF $\tilde{G}(t)$
Rayleigh	$H_k \sim \mathcal{CN}(0,1)$	$-\ln(1-t)$
Log-Normal	$20\log_{10}\lvert H_k\rvert \sim \mathcal{N}(0,\sigma^2)$	$Q^{-1}(\frac{10\log_{10}t}{\sigma})$
Nakagami-m	$f_{\lvert H_k\rvert^2}(x) = \frac{m^m}{\Gamma(m)}x^{m-1}\mathrm{e}^{-mx}$	$\tilde{F}(x) = \frac{\Gamma(m,mx)}{\Gamma(m)}, \tilde{G}(t) = \tilde{F}^{-1}(t)$

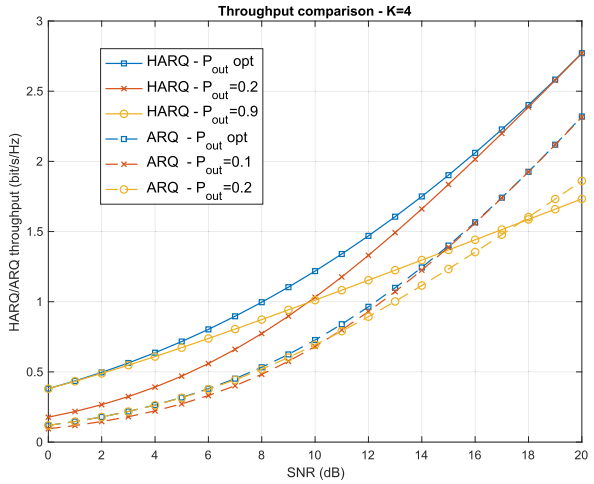

FIGURE 5. Plot of the MAX-MIN throughput with HARQ and ARQ for the symmetric Rayleigh fading user scenario with $K = 4$. Curves at non-optimum outage probabilities are also reported.

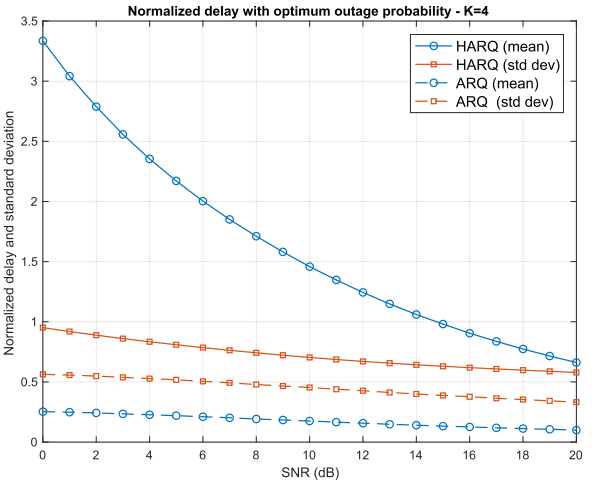

FIGURE 7. Plot of the normalized delay with optimum outage probabilities with $K = 4$ users.

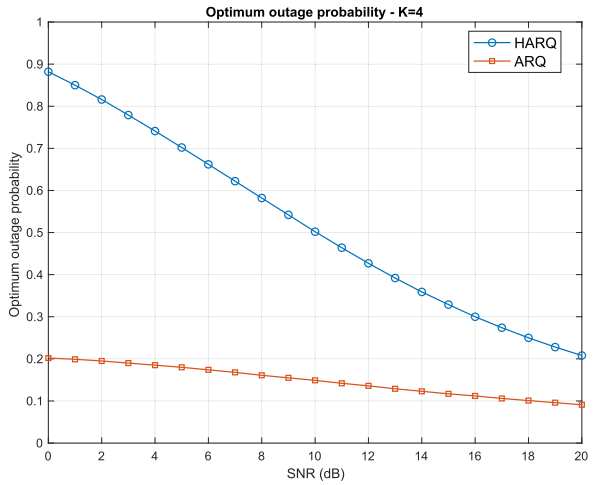

FIGURE 6. Plot of the optimum outage probabilities with $K = 4$ users.

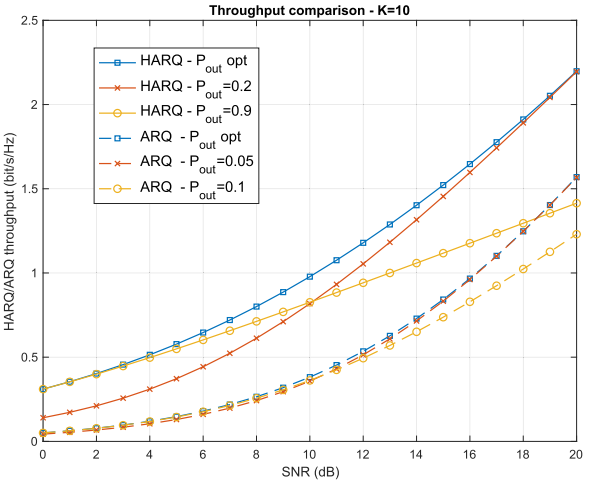

FIGURE 8. Same as Fig. 5 but with $K = 10$ users.

one for $\epsilon = 0.9$ are very close at the lower SNR values but tend to diverge as the SNR increases. Similar considerations hold for the other cases.

Notice that the large outage probabilities required to maximize the MAX-MIN throughput entail the occurrence of large transmission delays. This is illustrated, for example, in Fig. 7, which shows that the transmission delay can reach an average value of $\approx$ 3.3 packet times and a standard deviation of $\approx$ 1.0 packet times.[6]

The scaling of these results with the number of users is illustrated in Figs. 8 to 10, which report the HARQ and ARQ achievable sum-rates, optimum outage probabilities,

[6]The total delay may further increase because of the scheduling policy of the retransmission protocol. The delay considered here assumes that the packets are retransmitted immediately after the decoding failure, without any processing, propagation, or scheduling delays.

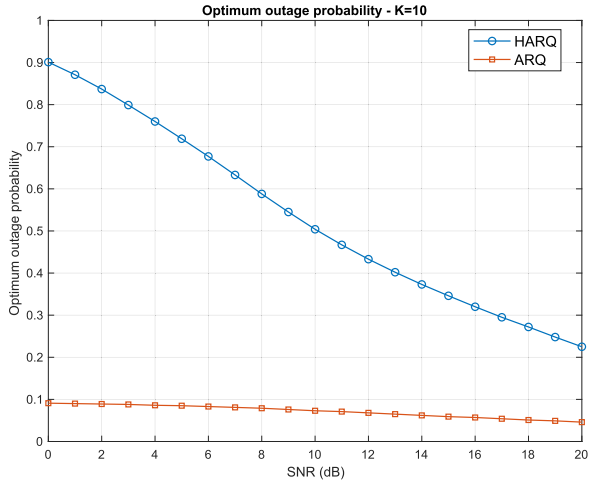

FIGURE 9. Same as Fig. 6 but with $K = 10$ users.

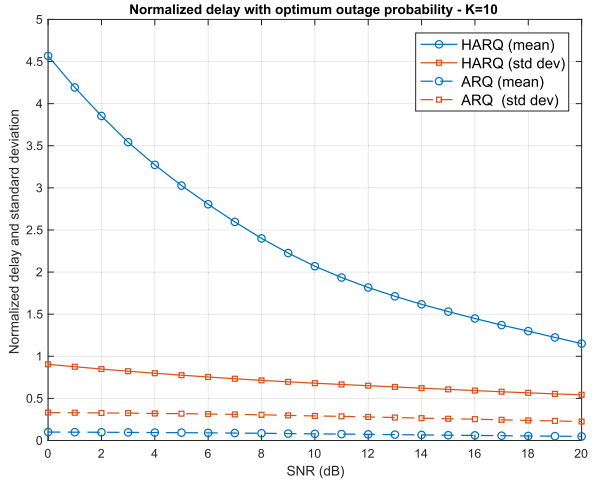

FIGURE 10. Same as Fig. 7 but with $K = 10$ users.

and normalized delays, respectively, with $K = 10$ users. Notice that the sum-rates are lower than with $K = 4$ users (Fig. 8 versus 5) because of the greater number of constraints to be satisfied during the SIC. Notice that, as the number of users increases, the advantage of using HARQ rather than ARQ increases as well. The optimum outage probabilities are similar (Fig. 9 versus 6). The optimum delays are worse with $K = 10$ than with $K = 4$ (Fig. 10 versus 7).

B. SINGLE-CELL SCENARIO

This scenario follows from [6], under the assumption that the base station knows only the user distances and not the fading channel gains (so that only CDIT is available instead of CSIT). The base station is located at the center of a disk $\mathcal{D}$ of ray $\mathcal{R}_{\mathcal{D}}$ and K users are uniformly distributed over the disk area. The k-th user gain is characterized by:

$$H_k \mid d_k \sim \mathcal{CN}\left(0, \frac{1}{1+d_k^{\mu}}\right), \tag{46}$$

where d_k is the (random) distance from the base station at the center and μ is the path loss exponent. Accordingly, since d_k^2 is uniformly distributed over $(0, \mathcal{R}_{\mathcal{D}}^2)$, the channel gain CDF is given by:[7]

$$\begin{aligned} F_{|H|^2}(x) &= 1 - \frac{1}{\mathcal{R}_{\mathcal{D}}^2}\int_0^{\mathcal{R}_{\mathcal{D}}^2} \mathrm{e}^{-x(1+u^{\mu/2})}du \\ &= 1 - \frac{2\mathrm{e}^{-x}}{\mu\mathcal{R}_{\mathcal{D}}^2 x^{\frac{2}{\mu}}}\gamma\left(\frac{2}{\mu}, \mathcal{R}_{\mathcal{D}}^{\mu}x\right), \end{aligned} \tag{47}$$

where $\gamma(r, x)$ is the incomplete Gamma function [33, 8.35]. Then, the average SNR over the cell area can be calculated by using [33, 6.455.2]:

$$\begin{aligned} \overline{\mathrm{SNR}} &= P_x\mathbb{E}[|H|^2] = P_x\int_0^{\infty}[1 - F_{|H|^2}(x)]dx \\ &= P_x\int_0^{\infty}\frac{2\mathrm{e}^{-x}}{\mu\mathcal{R}_{\mathcal{D}}^2 x^{\frac{2}{\mu}}}\gamma\left(\frac{2}{\mu}, \mathcal{R}_{\mathcal{D}}^{\mu}x\right)dx \\ &= \frac{P_x}{1+\mathcal{R}_{\mathcal{D}}^{\mu}}\,{}_2F_1\left(1, 1; \frac{2}{\mu}+1; \frac{\mathcal{R}_{\mathcal{D}}^{\mu}}{1+\mathcal{R}_{\mathcal{D}}^{\mu}}\right). \end{aligned} \tag{48}$$

Now, SIC can be implemented by considering the *average* user SNR, which decreases with the user distance. Then, SIC proceeds from the strongest user (at minimum distance) to the weakest (at maximum distance), so that it is convenient to arrange the distances $d_k, k = 1, \ldots, K$ in increasing order, i.e., $d_{(1)} \le \ldots \le d_{(K)}$, where $d_{(k)}$ denotes the k-th ordered random distance. Accordingly, by using the method of order statistics (see, *e.g.*, [32]),

$$\begin{aligned} F_{d_{(k)}^2}(x) &\triangleq P(d_{(k)}^2 \le x) \\ &= \sum_{\ell=k}^{K}\binom{K}{\ell}[F_{|d|^2}(x)]^{\ell}[1 - F_{|d|^2}(x)]^{K-\ell} \\ &= \sum_{\ell=k}^{K}\binom{K}{\ell}\left(\frac{x}{\mathcal{R}_{\mathcal{D}}^2}\right)^{\ell}\left(1 - \frac{x}{\mathcal{R}_{\mathcal{D}}^2}\right)^{K-\ell} \end{aligned} \tag{49}$$

for $x \in (0, \mathcal{R}_{\mathcal{D}})$. To evaluate the transition probabilities $p_r(\boldsymbol{\epsilon})$ of the Markov chain describing the HARQ random process, it is assumed that the user positions are fixed throughout the successive codeword transmissions (quasi-static assumption). Then, the k-th user normalized ICDF $\tilde{G}(\epsilon_k)$ is found by solving eq. (10):

$$\epsilon_k = 1 - \int_0^{\mathcal{R}_{\mathcal{D}}^2} \mathrm{e}^{-G(1+x^{\mu/2})}dF_{d_{(k)}^2}(x) \tag{50}$$

with respect to G. Here,

$$dF_{d_{(k)}^2}(x) = k\binom{K}{k}\left(\frac{x}{\mathcal{R}_{\mathcal{D}}^2}\right)^{k-1}\left(1 - \frac{x}{\mathcal{R}_{\mathcal{D}}^2}\right)^{K-k}\frac{dx}{\mathcal{R}_{\mathcal{D}}^2}. \tag{51}$$

Moreover, by definitions (22) and (24),

$$\pi_{k,r}(\boldsymbol{\epsilon}) = \int_0^{\mathcal{R}_{\mathcal{D}}^2}\tilde{\gamma}(r, \tilde{G}(\epsilon_k)(1 + x^{\mu/2}))dF_{d_{(k)}^2}(x) \tag{52}$$

[7] A closed form of this distribution is used here, instead of its numerical approximation considered in [6].

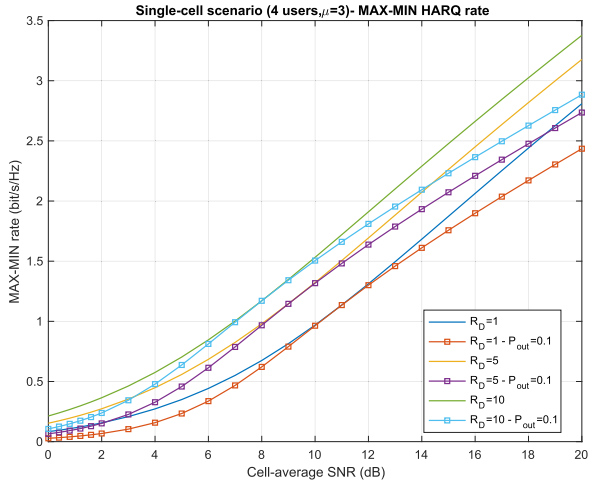

FIGURE 11. Plot of the HARQ MAX-MIN user throughput versus the average SNR over the cell $\overline{\text{SNR}}$ (in dB) for the single-cell scenario with $K = 4$ users, path loss exponent $\mu = 3$, and cell rays $= 1, 5, 10$ (normalized units). The solid lines correspond to the optimum HARQ rates obtained by calculating (14) for constant $T_k = 1$ (MAX-MIN fairness), including the simultaneous optimization over the power allocation vector α and the outage probability ϵ.

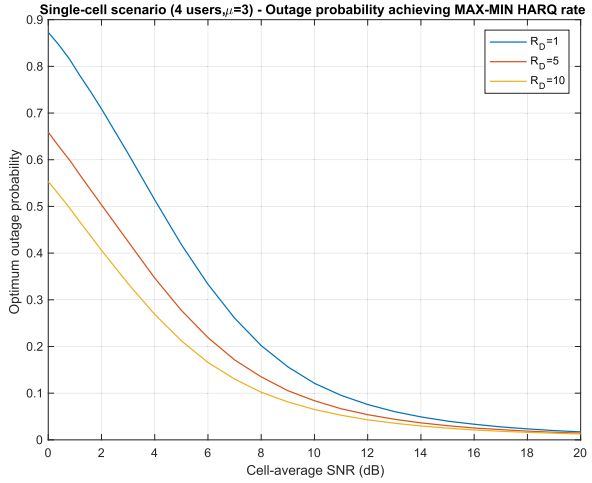

FIGURE 12. Plot of the optimum outage probabilities corresponding to the single-cell scenarios described in Fig. 11.

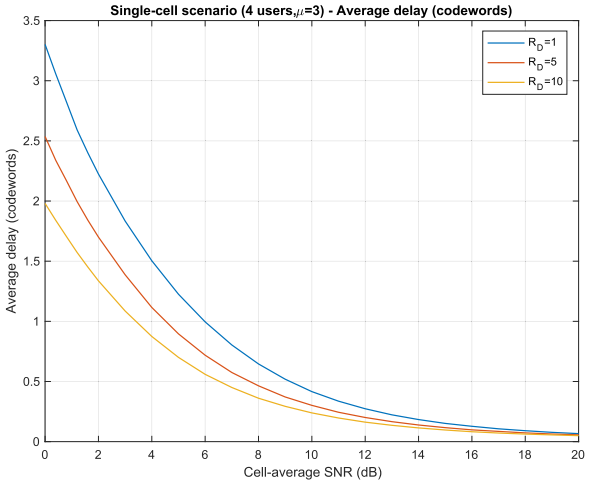

FIGURE 13. Plot of the average HARQ delay corresponding to the optimum outage probabilities reported in Fig. 12 and the single-cell scenarios described in Fig. 11.

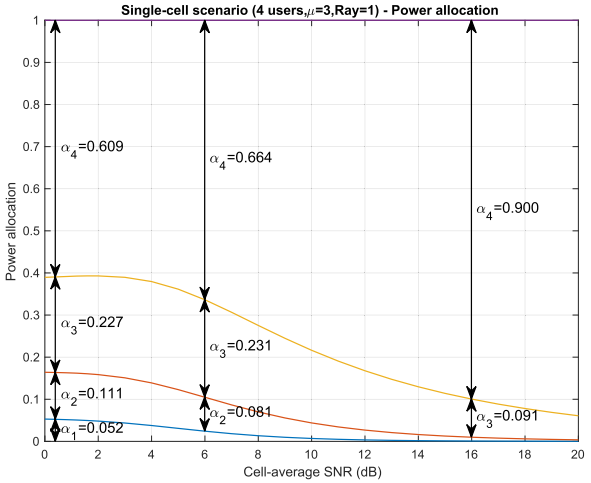

FIGURE 14. Plot of the optimum power allocation for the single-cell scenarios described in Fig. 11 with $\mathcal{R}_{\mathcal{D}} = 1$.

for $k = 1, \ldots, K$ and $r = 1, 2, \ldots$, yield the full set of Markov chain probabilities $p_r(\boldsymbol{\epsilon})$ defined in eq. (25).

1) NUMERICAL RESULTS

Fig. 11 shows the HARQ MAX-MIN throughput versus the average SNR over the cell, denoted by $\overline{\text{SNR}}$, for the single-cell scenario described in this section with $K = 4$ users, path loss exponent $\mu = 3$, and cell rays $= 1, 5, 10$ (normalized units). The figure reports the HARQ MAX-MIN throughput (derived by calculating (14) for constant $T_k = 1$) when all users have the same outage probabilities $\epsilon_k = \epsilon$. The unoptimized HARQ MAX-MIN user information rates corresponding to a fixed outage probability $\epsilon = 0.1$ are also reported for comparison. It can be noticed that the outage probability optimization is effective in the lower end of the average cell SNR range considered. The optimum outage probability (common to all users) is reported versus the average cell SNR in Fig. 12. It can be noticed that, in the lower end of the average cell SNR range considered, the optimum outage probabilities are noticeably different from 0.1, which justifies the performance gap illustrated in Fig. 11. The corresponding HARQ transmission delay is reported in Fig. 13 for $\mathcal{R}_{\mathcal{D}} = 1$, 5, and 10. The higher outage probabilities required to optimize the information rates imply the higher delays in the lower end of the average cell SNR range considered. Nevertheless, it is worth noticing that these delays are much shorter than what they would be with ARQ, i.e., $1/(1-\epsilon)^K$. Finally, Figs. 14 to 16 illustrate the optimum power allocation with $\mathcal{R}_{\mathcal{D}} = 1$, 5, and 10, respectively. As expected, the stronger users are assigned a progressively lower power at the transmitter in order to attain

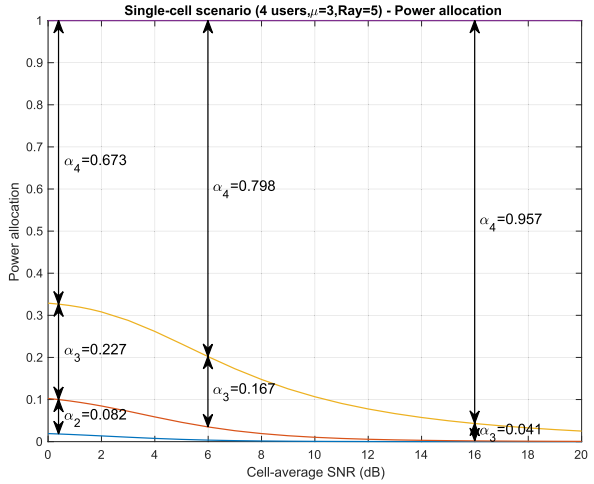

FIGURE 15. Same as Fig. 14 but with $\mathcal{R}_{\mathcal{D}} = 5$.

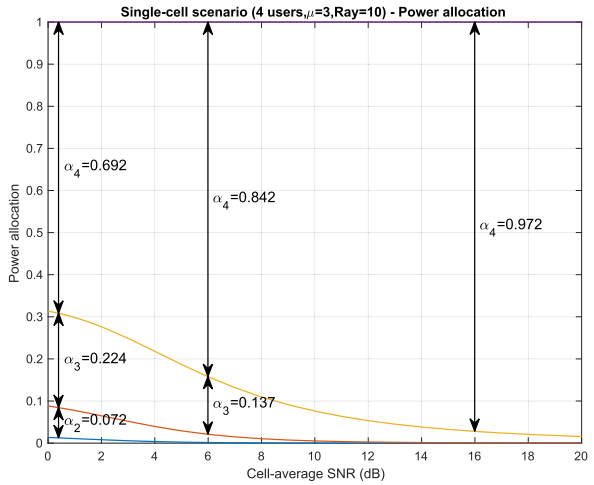

FIGURE 16. Same as Fig. 14 but with $\mathcal{R}_{\mathcal{D}} = 10$.

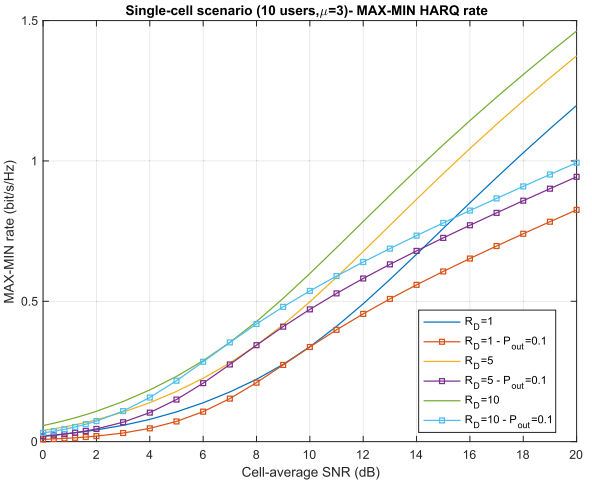

FIGURE 17. Same as Fig. 11 but for $K = 10$.

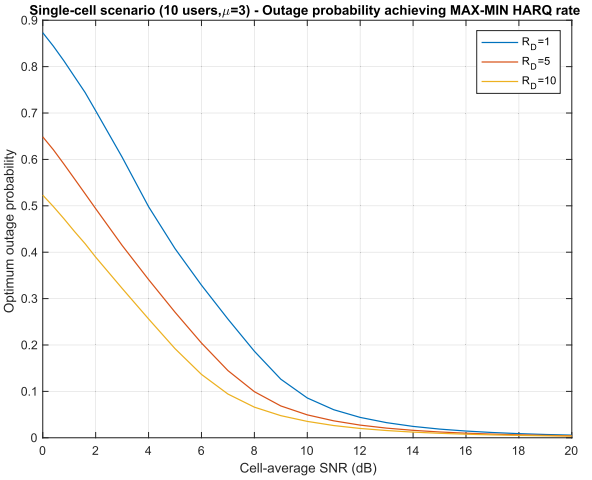

FIGURE 18. Same as Fig. 12 but for $K = 10$.

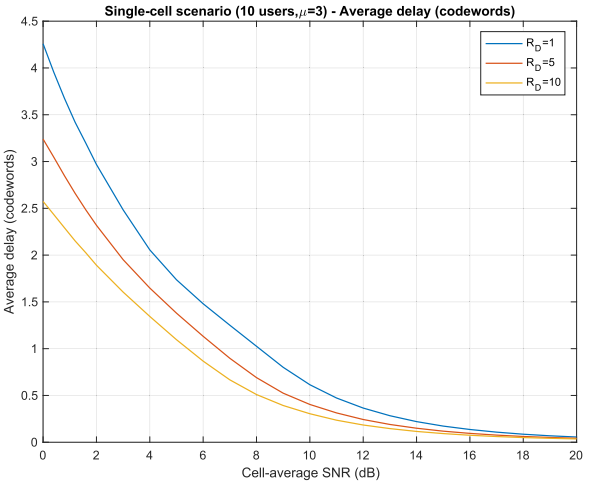

FIGURE 19. Same as Fig. 13 but for $K = 10$.

the MAX-MIN fairness condition. This may be tempered by turning to PF optimization (see Remark 1), which limits the dynamics of the different user power allocation factors.

2) SCALABILITY WITH THE NUMBER OF USERS

The diagrams reported in this section extend the previous analysis from $K = 4$ to $K = 10$ users to illustrate the scalability of the results. Fig. 17 shows the HARQ MAX-MIN user throughput versus the average SNR over the cell $\overline{\text{SNR}}$ for the single-cell scenario described in this section with $K = 10$ users, path loss exponent $\mu = 3$, and cell rays equal to 1, 5, 10 (normalized units), Fig. 18 reports the optimum outage probability (common to all users), and Fig. 19 reports the corresponding HARQ transmission delay. Figs. 20 to 22 illustrate the optimum power allocation with $\mathcal{R}_{\mathcal{D}} = 1, 5$, and 10, respectively with $K = 10$ users.

All results are in line with those relevant to the 4-user scenario, except, possibly, for the nonmonotonic behavior of some power allocation fractions with respect to the cell-average SNR.

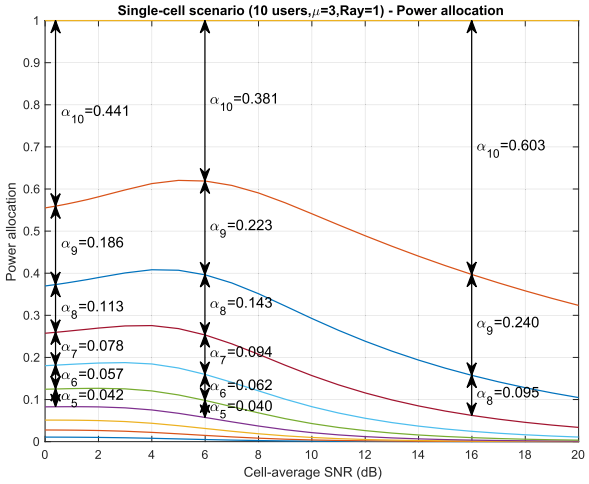

FIGURE 20. Same as Fig. 14 but for $K = 10$.

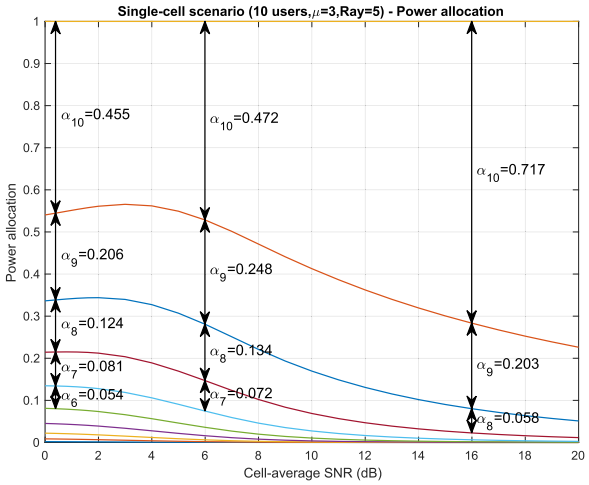

FIGURE 21. Same as Fig. 15 but for $K = 10$.

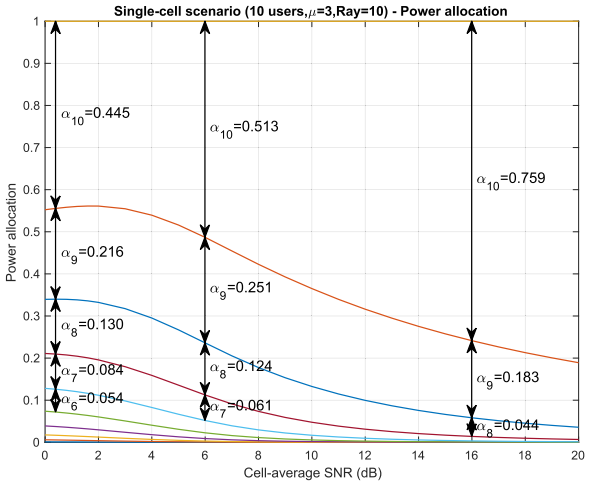

FIGURE 22. Same as Fig. 16 but for $K = 10$.

VI. CONCLUSION

In this work, PD-NOMA has been investigated focusing on a block fading scenario where transmission outages play a basic role for reliability. Many analytic results have been obtained capitalizing on earlier literature works. Specifically, the optimum information rates with PF have been derived in [26], and the outage rates have been derived in [30]. These analytic results have been used to optimize the system throughput by leveraging on the outage level design and the power allocation at the transmitter. By allowing an optimum number of outages, the throughput is maximized at the expense of an increased transmission delay, evaluated in terms of average, standard deviation, and counter-cumulative probability distribution. The scenarios considered provide an extensive illustration of possible wireless networks.

In the first one (symmetric user scenario), Rayleigh fading determines the instantaneous path loss of each user but the average gain is fixed for all them. This assumption implies that the transmitter has only CDIT and not CSIT. This symmetry allows for several analytic simplifications, like the independence of the ICDF from the user and the outage probabilities. A closed form expression for the power allocation coefficients has been obtained. Numerical results illustrate the power allocation in some cases, depending on the outage probability, the HARQ and ARQ MAX-MIN information rates, the optimum outage probabilities, and the mean and standard deviation of the transmission delays.

The second scenario follows the from the literature [6] and consists of a set of users uniformly located over a disk area where the base station is placed at the disk center. Each user's path loss depends on the distances from the base station and on a given propagation exponent, with superimposed Rayleigh fading. The base station is assumed to know the user distances but not the realization of the Rayleigh fading components. In this setting, the MAX-MIN throughput is studied in conjunction with the use of a HARQ protocol. The optimum outage probabilities, the average transmission delays, and the optimum power allocation at the transmitter are illustrated by an extensive set of numerical examples. The effectiveness of optimizing the outage level is emphasized by comparing the optimum results against those corresponding to sub-optimum outage probability values. These results are reported versus the average SNR over the cell area, which has been calculated analytically.

Possible extensions of this work may address more sophisticated scheduling techniques for the HARQ protocol, aimed at maximizing the exploitation of the received signal fragment by reducing the number of retransmissions. Moreover, an analysis based on actual FEC codes may provide additional insight on the actual throughput which can be achieved in a real system, in consideration of the fact that this study is based on information theoretical results.

REFERENCES

[1] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017, doi: 10.1109/JSAC.2017.2725519.

[2] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[3] O. Maraqa, A. S. Rajasekaran, S. Al-Ahmadi, H. Yanikomeroglu, and S. M. Sait, "A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2192–2235, 4th Quart., 2020, doi: 10.1109/COMST.2020.3013514.

[4] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York, NY, USA: McGraw-Hill, 2008.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2006.

[6] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014, doi: 10.1109/LSP.2014.2343971.

[7] S. Wu, Z. Deng, A. Li, J. Jiao, N. Zhang, and Q. Zhang, "Minimizing age-of-information in HARQ-CC aided NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 2, pp. 1072–1086, Feb. 2023, doi: 10.1109/TWC.2022.3201115.

[8] Z. Ding, P. Fan, and H. V. Poor, "User pairing in non-orthogonal multiple access downlink transmissions," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–5.

[9] S. M. R. Islam, M. Zeng, O. A. Dobre, and K.-S. Kwak, "Resource allocation for downlink NOMA systems: Key techniques and open issues," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 40–47, Apr. 2018.

[10] L. Zhu, J. Zhang, Z. Xiao, X. Cao, and D. O. Wu, "Optimal user pairing for downlink non-orthogonal multiple access (NOMA)," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 328–331, Apr. 2019.

[11] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[12] W. Zhang, S. P. Kotagiri, and J. N. Laneman, "On downlink transmission without transmit channel state information and with outage constraints," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4240–4248, Sep. 2009.

[13] Z. Tang, J. Wang, J. Wang, and J. Song, "On the achievable rate region of NOMA under outage probability constraints," *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 370–373, Feb. 2019, doi: 10.1109/LCOMM.2018.2870584.

[14] S. Lin and P. Yu, "A hybrid ARQ scheme with parity retransmission for error control of satellite channels," *IEEE Trans. Commun.*, vol. COM-30, no. 7, pp. 1701–1719, Jul. 1982, doi: 10.1109/TCOM.1982.1095643.

[15] Y.-M. Wang and S. Lin, "A modified selective-repeat type-II hybrid ARQ system and its performance analysis," *IEEE Trans. Commun.*, vol. COM-31, no. 5, pp. 593–608, May 1983, doi: 10.1109/TCOM.1983.1095873.

[16] S. Lin, D. J. Costello, and M. J. Miller, "Automatic-repeat-request error-control schemes," *IEEE Commun. Mag.*, vol. 22, no. 12, pp. 5–17, Dec. 1984.

[17] R. Comroe and D. Costello, "ARQ schemes for data transmission in mobile radio systems," *IEEE J. Sel. Areas Commun.*, vol. SAC-2, no. 4, pp. 472–481, Jul. 1984, doi: 10.1109/JSAC.1984.1146084.

[18] D. Chase, "Code combining—A maximum-likelihood decoding approach for combining an arbitrary number of noisy packets," *IEEE Trans. Commun.*, vol. COM-33, no. 5, pp. 385–393, May 1985, doi: 10.1109/TCOM.1985.1096314.

[19] S. Kallel, "Analysis of a type II hybrid ARQ scheme with code combining," *IEEE Trans. Commun.*, vol. 38, no. 8, pp. 1133–1137, Aug. 1990.

[20] B. A. Harvey and S. B. Wicker, "Packet combining system based on the Viterbi decoder," *IEEE Trans. Commun.*, vol. 42, pp. 1544–1557, Feb./Mar./Apr. 1994.

[21] D. J. Costello, J. Hagenauer, H. Imai, and S. B. Wicker, "Applications of error-control coding," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2531–2560, Oct. 1998.

[22] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1971–1988, Jul. 2001, doi: 10.1109/18.930931.

[23] S. Sesia, G. Caire, and G. Vivier, "Incremental redundancy hybrid ARQ schemes based on low-density parity-check codes," *IEEE Trans. Commun.*, vol. 52, no. 8, pp. 1311–1321, Aug. 2004, doi: 10.1109/TCOMM.2004.833022.

[24] J. Choi, "On HARQ-IR for downlink NOMA systems," *IEEE Trans. Commun.*, vol. 64, no. 8, pp. 3576–3584, Aug. 2016, doi: 10.1109/TCOMM.2016.2585651.

[25] D. Cai, Z. Ding, P. Fan, and Z. Yang, "On the performance of NOMA with hybrid ARQ," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 10033–10038, Oct. 2018, doi: 10.1109/TVT.2018.2854184.

[26] G. Taricco, "Fair power allocation policies for power-domain non-orthogonal multiple access transmission with complete or limited successive interference cancellation," *IEEE Access*, vol. 11, pp. 46793–46803, 2023, doi: 10.1109/ACCESS.2023.3274470.

[27] T. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.

[28] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[29] G. Gui, H. Sari, and E. Biglieri, "A new definition of fairness for non-orthogonal multiple access," *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1267–1271, Jul. 2019.

[30] J. Choi, "Power allocation for max-sum rate and max-min rate proportional fairness in NOMA," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2055–2058, Oct. 2016.

[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[32] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. Hoboken, NJ, USA: Wiley, 2003.

[33] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 8th ed. New York, NY, USA: Academic, 2007.

GIORGIO TARICCO (Fellow, IEEE) received the Laurea degree (cum laude) in ingegneria elettronica from Politecnico di Torino, Italy, in 1985.

He was a Researcher with Italian Telecom Laboratories (CSELT), from 1985 to 1987, where he was involved in the design process of the GSM mobile telephony standard (channel coding). Since 1991, he has been with the Department of Electronics and Telecommunications (DET), Politecnico di Torino, as an Assistant Professor, where he has been a Full Professor, since 2010. In 1996, he was a Research Fellow with ESA/ESTEC, The Netherlands. He has coauthored more than 200 papers published in international journals and international conferences, several book chapters, and three international patents. His research interests include information theory, error-control coding, multiuser detection, space-time coding, MIMO communications, cognitive radio networks, sensor networks, the IoT satellite networks, and MIMO relay channels. He was a co-recipient of the Best Paper Awards from the WPMC 2001 Conference and the *Journal of Communications and Networks* (Special Issue on Coding and Signal Processing for MIMO Systems), in 2003. He was on the ISI highly cited researcher list in the category of computer science, from 2008 to 2013. He has participated in several committees of IEEE conferences (as a treasurer). He has been an Associate Editor of the IEEE COMMUNICATIONS LETTERS, the IEEE TRANSACTIONS ON INFORMATION THEORY, and the *Journal on Communications and Networks*. He was also a Senior Associate Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS.

Open Access funding provided by 'Politecnico di Torino' within the CRUI CARE Agreement