Enhancing Performance of Uplink URLLC Systems via Shared Diversity Transmissions and Multiple Antenna Processing

Kotaba, Radoslaw; Manchon, Carles Navarro; Popovski, Petar

Published in:
2019 53rd Asilomar Conference on Signals, Systems, and Computers

DOI (link to publication from Publisher):
10.1109/IEEECONF44664.2019.9049079

Publication date:
2020

Citation for published version (APA):
Kotaba, R., Manchon, C. N., & Popovski, P. (2020). Enhancing Performance of Uplink URLLC Systems via Shared Diversity Transmissions and Multiple Antenna Processing. In M. B. Matthews (Ed.), *2019 53rd Asilomar Conference on Signals, Systems, and Computers* (pp. 1409-1415). [9049079] IEEE. Asilomar Conference on Signals, Systems and Computers. Conference Record https://doi.org/10.1109/IEEECONF44664.2019.9049079

Enhancing Performance of Uplink URLLC Systems via Shared Diversity Transmissions and Multiple Antenna Processing

Radosław Kotaba, Carles Navarro Manchón, and Petar Popovski
Department of Electronic Systems
Aalborg University, Denmark
Email: {rak, cnm, petarp}@es.aau.dk

***Abstract*—In this work we investigate the reliability aspects of uplink multi-user MIMO communication over a preallocated pool of time-frequency resources shared by a group of ultra-reliable low-latency devices. To achieve sufficient diversity, users perform multiple transmissions of their packets over a shared pool of time-frequency resources in a non-orthogonal manner. The preallocation allows users to employ fast, grant-free type of access, while sharing improves the overall spectral efficiency. The multiple transmit opportunities enhance the robustness of communication through incremental redundancy. On the base station side, we consider the performance of a minimum mean square error (MMSE) receiver, chosen for its relative simplicity. In addition to a baseline scheme in which devices randomly select the resources without coordination, we consider two other approaches based on preassigned access patterns: i) one in which all resources are utilized evenly and with equal power, and ii) another, where the spectrum is divided into high and low contention portions and users benefit from having few reliable transmissions and few *diversity resources*. In particular, we focus on evaluating the performance limits of the schemes as the number of antennas at the base station grows.**

I. INTRODUCTION

Ultra-reliable low latency communication (URLLC) is a new category of use cases in the latest, fifth generation cellular standard [1], and it encompasses the most demanding types of applications including (but not limited to): Industry 4.0 scenarios (factory automation, motion control) [2], tele-surgery (based on haptic feedback) and vehicular-to-anything (V2X) in the Automotive industry [3].

Achieving spectrally efficient URLLC is inherently difficult, which is the key takeaway from [4]. In fact, the solutions implemented in practical systems make simultaneous low latency and high reliability contradictory as they trade one for another. This is especially true for uplink (UL) traffic which in the classical cellular networks is fully managed by the centralized base station (BS). A comprehensive overview of the challenges faced by URLLC can be found in [5] where the authors discuss in detail various enablers and their tradeoffs.

Among the most promising and at the same time disruptive techniques is the grant-free access, which gives devices the ability to perform transmissions without prior scheduling [6]. Indeed, the requirement to perform scheduling grant handshake is one of the largest bottlenecks in the design of low latency systems. The price to pay for avoiding it is a significantly reduced control that the BS has over interference.

Since its inception the topic of grant-free access has garnered a lot of attention. Several designs and implementations were considered, earliest of which build on the legacy concept known as slotted ALOHA [7] and its extension coded random access (CRA) [8]. The actual grant-free scheme as defined for 5G NR, utilizing k-repetitions over shared resources and aperiodic traffic has been studied in [9] [10]. The former contribution focuses on the collision aspect (from the combinatorial point of view), while the latter provides a realistic assessment through system-level simulations in an outdoor urban micro scenario. In [11] a hybrid approach is studied where devices initiate the communication grant-free and switch to coordinated access for retransmissions.

The idea of introducing some coordination in the form of preassigned access patterns is discussed in several works [12] [13]. The former coins the term transmissions with shared diversity resources (TSDR) and focuses on their performance in the presence of imperfect CSI. The latter analyzes a special type of access patterns based on the code construction according to Steiner system.

In this work we analyze a multi-user URLLC system where the diversity required to achieve high reliability is provided by a combination of multiple receive antennas and multiple redundant transmissions of the packet over a shared pool of resources. We start by pointing out the shortcomings of the naive grant-free access scheme, which does not take into account potential pilot collisions, and postulate that preassigned access patterns should be used instead to avoid them. Aided by the recent work of [14], we develop an original analytical framework that allows to evaluate the outage performance of such multi-user, multi-transmission system when the BS utilizes MMSE processing. To the best of the authors' knowledge such results haven't been obtained before and until recently could only be treated under ordinary zero-forcing (ZF). We apply the developed tools to analyze two types of access patterns: uniform patterns, evenly utilizing the whole resource pool; and a generalized version of TSDR which combines slots with higher and lower amount of interference. The obtained results clearly show the superiority of preassigned patterns over the naïve grant-free in terms of outage probability. Moreover, the

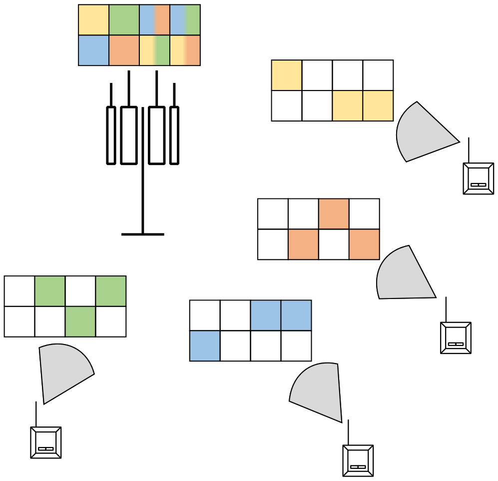


Fig. 1. Example of the grant-free access scheme with $k = 3$ multiple transmissions over a pool of $L = 8$ TF-blocks.

approach based on TSDR has a potential for reducing the complexity and effective latency compared to the uniform ones.

Throughout the paper the following notation is used: $(\cdot)^{\mathrm{H}}$ to denote conjugate transpose, $(\cdot)_{i,j}$ to denote the element in i-th row and j-th column of the matrix, bold uppercase letters to denote matrices respectively. $\mathbf{I}_N$ denotes an $N \times N$ identity matrix. $\|\cdot\|_0$ denotes the ℓ_0 pseudonorm. $\mathrm{E}[\cdot]$ denotes the expected value.

II. System Model

We consider a single base station (BS) serving N URLLC-type users. The base station is equipped with M antennas while each of the devices (UEs) has a single antenna. The access channel is divided into periodic frames composed of L slots also referred to as time-frequency (TF) blocks. Each such block is further composed of K channel uses. In this work we consider the case of Rayleigh block fading channel, where the realizations of the channel coefficients are independent between different slots and UEs. We assume that all UEs transmit with the same rate R and a worst case scenario is considered where all of them are active in each frame. To harvest diversity and consequently achieve reliability each user transmits its packet using k out of L slots in a frame. The packets can be identical, constituting a form of k-repetition coding, or contain different coded symbols (redundancy versions) of the original message. Throughout the paper we will refer to the former and latter scheme as Chase Combining (CC) and Incremental Redundancy (IR) respectively.

At the receiver, the baseband representation of the channel output during l-th slot of the d-th frame can be written as:

$$\mathbf{Y}_{d,l} = \mathbf{H}_{d,l}\mathbf{X}_{d,l} + \mathbf{N}_{d,l} = \left(\mathbf{R}_{RX}\mathbf{G}_{d,l}\mathbf{P}_{d,l}\right)\mathbf{X}_{d,l} + \mathbf{N}_{d,l} \quad (1)$$

where $\mathbf{Y}_{d,l} \in \mathbb{C}^{M\times K}$ are received symbols, $\mathbf{X}_{d,l} \in \mathbb{C}^{N\times K}$ are the transmitted complex modulated symbols normalized such that $\mathrm{E}\left[\left|x_{i,j}\right|^2\right] = 1$, $\mathbf{N}_{d,l} \in \mathbb{C}^{M\times K}$ is the additive white Gaussian noise (AWGN) with zero mean and variance σ^2 and $\mathbf{H}_{d,l} \in \mathbb{C}^{M\times N}$ are the channel gains between N users and M antennas. The component $\mathbf{H}_{d,l}$ can be further represented as a product of $\mathbf{G}_{d,l} \in \mathbb{C}^{M\times N}$, which is zero-mean circularly-symmetric complex Gaussian (ZMCSCG) and models the underlying uncorrelated Rayleigh flat fading channel, $\mathbf{P}_{d,l} = diag\left((P_{d,l,1})^{1/2}, \ldots, (P_{d,l,N})^{1/2}\right)$ is a diagonal matrix of transmit powers and $\mathbf{R}_{RX}$ is a square Toeplitz matrix with parameter ρ denoting receive antenna correlation. The packet of each UE is subject to the total transmit power constraint such that independently of the total number of transmissions within a frame

$$\sum_{l=1}^{L} P_{d,l,i} = P_{tot}, \qquad \forall i \in \{1, \ldots, N\}, \forall d. \quad (2)$$

Note that some of the $P_{d,l,i} = 0$ which reflects the fact that each UE uses only k among L available slots, specifically $\sum_{l=1}^{L} \|P_{d,l,i}\|_0 = k$.

In the remainder of the paper we will omit the frame index d as the transmissions within a single frame are self contained and independent (i.e. UEs are not allowed to transmit the same packet over multiple frames as this would violate the latency constraint).

A. Receiver processing

As the use case on which we are focusing in this paper is URLLC, the relevant metric for our system is outage probability. When IR transmission mode is being used the outage can be defined as[1]

$$p_{out_i} = \Pr\left\{R > \sum_{l=1}^{L} \ln\left(1 + \mathrm{SINR}_{l,i}\right)\right\} \quad (3)$$

where R is the transmission rate (in nats) used to encode the packet. The $\mathrm{SINR}_{l,i}$ is understood as the post-processing signal-to-interference-plus-noise ratio of user i in slot l and can be computed as

$$\mathrm{SINR}_{l,i} = \frac{\left|\left(\mathbf{F}_l\mathbf{H}_l\right)_{i,i}\right|^2}{\sum_{j=1,j\neq i}^{N}\left|\left(\mathbf{F}_l\mathbf{H}_l\right)_{i,j}\right|^2 + \sigma^2\sum_{j=1}^{N}\left|\left(\mathbf{F}_l\right)_{i,j}\right|^2} \quad (4)$$

where $\mathbf{F}_l$ is the detection matrix employed by the BS. In this work we chose to focus on the minimum mean square error equalization method where $\mathbf{F}_l = \left(\mathbf{H}_l^{\mathrm{H}}\mathbf{H}_l + \sigma^2\mathbf{I}_N\right)^{-1}\mathbf{H}^{\mathrm{H}}$, in which case (4) simplifies to

$$\mathrm{SINR}_{l,i} = \frac{1}{\left(\left(\mathbf{H}_l^{\mathrm{H}}\mathbf{H}_l + \sigma^2\mathbf{I}_N\right)^{-1}\right)_{i,i}} - 1. \quad (5)$$

It is further assumed that prior to the transmission BS and UEs possess only a statistical knowledge of the channel state information (CSI). Once the packets are received, and unless otherwise stated, BS is capable of perfectly estimating $\mathbf{H}$ from the available pilots.

---

[1] In the case of CC experiencing independent interference, the summation appears inside the logarithm instead. Transmissions can also be processed jointly yielding single, combined post-processing SINR.

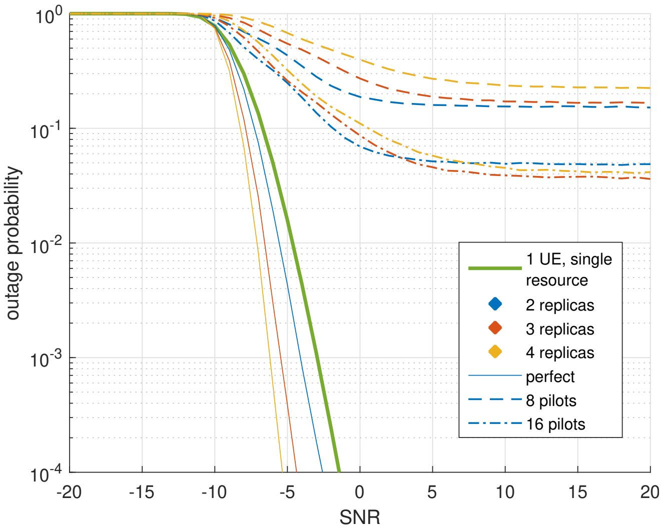


Fig. 2. Outage probability performance of idealized and pilot-limited grant-free access. The average contention levels corresponding to 2,3, and 4 replicas are 4, 6, and 8 respectively. The one-shot transmission used for comparison is equivalent to $k = 1$ and $C_l = 1$

III. Grant-Free Access

Grant-free transmissions are among the most frequently proposed enablers of the use cases requiring extremely low latencies. In its most basic version, grant-free involves dedicating a set of L TF-blocks to create a common pool of resources which are then accessed by a group of UEs in random and uncoordinated manner. Whenever a device has a packet to send, it selects randomly k out of L slots and uses them to transmit its data. A toy example of such scheme is shown in Fig. 1. In such a grant-free system with N active devices, the contention level of the lth TF-block, denoted by C_l and defined as the number of UEs using the lth TF-block, is a binomial random variable with success probability $p = \frac{k}{L}$ and N trials s.t. $\mathrm{E}\left[C_l\right] = \frac{kN}{L}$.

In principle, having such a scheme is possible as long as the BS is equipped with enough antennas and has sufficiently accurate CSI, as it can resolve the (potentially numerous) collisions with proper multi-antenna processing. In a fully uncoordinated grant-free scheme, however, the UEs need to select at random not only TF-blocks but also one out of a finite number of pilot sequences. Collisions of the latter can be more severe as they lead to pilot contamination and hinder the use of multi-antenna detectors.

To illustrate these issues we will perform the following experiment. We simulate uncoordinated, grant-free random access by considering two situations: an idealized one in which the BS always has perfect CSI, regardless of the contention levels of each slot, and a realistic one in which users select one out of D available pilot sequences in an uncoordinated manner. In the latter scenario, when pilot sequences collide in a slot, the information in the associated transmission is considered lost for the BS receiver. For additional details regarding the signal processing and issues related to the simulation of this scenario we refer the reader to the Appendix.

The results of this experiment are shown in Fig. 2. The simulations in this and other figures (unless otherwise stated) are done with $L = 12$ TF-blocks, $N = 24$ users, $M = 8$ receive antennas, $\mathbf{R}_{RX} = \mathbf{I}_M$ and for different number of replicas $k \in \{2, 3, 4\}$. The target rate is set to $R = \ln 2$ (which corresponds to 1 bit/channel use). The transmit power is $\frac{P_{tot}}{k}$ and is determined by the operating point (x-axis). From Fig. 2 we can see that when the number of pilots is limited, the outage probability is severely degraded and exhibits plateauing with the level related to the probability that all replicas are lost due to collisions. Depending on the number D, transmitting more replicas k may or may not help as it involves a trade-off between their number and increasing the chance of collision per transmission. We should also note that in practical systems increasing the number of available pilot sequences results either in a loss of spectral efficiency (larger percentage of resources dedicated to pilots) or in an increased transmission rate (and therefore reliability degradation) if a constant spectral efficiency is to be maintained.

As a reference, with thick green line we show also the performance of a single user transmitting over a dedicated, interference-free slot. Although idealized grant-free experiences much higher average contention (which translates to lower diversity order per replica) and can even lead to situations where $C_l > M$ it clearly outperforms the so called one-shot approach. It is even more surprising considering that the latter offers only half the rate of grant-free ($\frac{24R}{12}$ vs $\frac{R}{1}$).

IV. Preassigned Access Patterns

As confirmed by the experiment in the previous section, the idea of pooling resources has clearly a great potential but is reliant on the availability of CSI. To address the main flaw of grant-free access, we consider in this section the case where users have preassigned access patterns (known and assigned by the BS) rather than selecting them randomly themselves. These patterns can be considered fixed or at least changing on a much larger timescale than the duration of the frame[2].

The specific design of patterns determines the maximum contention level C_l of each TF-block. Most importantly, unlike random selection, C_l is deterministic and controlled by the BS, and can therefore be adapted to the number of available pilot sequences. In particular, since the number of simultaneously transmitting devices in a TF-block can be made lower or equal to the number of pilots, the pilot sequences can be preassigned to UEs, in a way that ensures they won't collide with each other.

The determinism of C_l significantly simplifies the problem and allows to derive some analytical tools and results which are the focus of the following subsection. Then, in subsections IV-B and IV-C we consider two special cases of access patterns and apply the aforementioned tools to assess their performance.

---

[2] In practice, they could be assigned when the device first registers with the BS and then updated periodically to adapt to the varying total population of the URLLC users.

A. Outage probability analysis

Recently, the authors of [14] were able to obtain a closed-form expression for the pdf of the SINR provided by the MMSE equalizer in an uncorrelated Rayleigh fading setting with $C_l \leq M$ and equal power allocation between users. This surprisingly simple result yields

$$
\begin{aligned}
f_{\mathrm{SINR}_{l,i}}(x) = {} & \frac{x^{M-1} e^{-\frac{x}{\gamma_l}}}{(1+x)^{C_l}} \\
& \times \sum_{a=0}^{C_l-1} \binom{C_l-1}{a} \frac{\gamma_l^{a-M}}{(M-a-1)!} \left(\frac{M}{M-a} + x \right)
\end{aligned} \tag{6}
$$

where $\gamma_l = P_l/\sigma^2$ is the (same) average SNR of the UEs active in a slot l.[3]

Since in this work we define the outage criterion in terms of mutual information rather than SINR directly (cf. (3)) we need to make a simple transformation. Let $\mathrm{MI}_{l,i} = g(\mathrm{SINR}_{l,i}) = \ln(1 + \mathrm{SINR}_{l,i})$ be the mutual information (in nats) of the i-th user message obtained from the l-th TF-block. Consequently $\mathrm{SINR}_{l,i} = g^{-1}(\mathrm{MI}_{l,i}) = e^{\mathrm{MI}_{l,i}} - 1$. Using then the pdf transformation $f_Y(y) = f_X\left(g^{-1}(y)\right) \frac{d\left(g^{-1}(y)\right)}{dy}$ we obtain a new pdf

$$
\begin{aligned}
f_{\mathrm{MI}_{l,i}}(x) = {} & \frac{(e^x-1)^{M-1} e^{-\frac{e^x-1}{\gamma_l}}}{e^{x(C_l-1)}} \\
& \times \sum_{a=0}^{C_l-1} \binom{C_l-1}{a} \frac{\gamma_l^{a-M}}{(M-a-1)!} \left(\frac{a}{M-a} + e^x \right).
\end{aligned} \tag{7}
$$

Work [14] provides also the cdf of the SINR albeit the expression is slightly more complex. After adapting it to our scenario the expression reads

$$
\begin{aligned}
F_{\mathrm{MI}_{l,i}}(x) = {} & \mathcal{I}(M, M+1-C_l, \frac{1}{\gamma_l}, e^x-1) \\
& \times \sum_{a=0}^{C_l-1} \binom{C_l-1}{a} \frac{M}{(M-a)!} \gamma_l^{a-M} \\
& + \mathcal{I}(M+1, M+2-C_l, \frac{1}{\gamma_l}, e^x-1) \\
& \times \sum_{a=0}^{C_l-1} \binom{C_l-1}{a} \frac{1}{(M-a-1)!} \gamma_l^{a-M}
\end{aligned} \tag{8}
$$

where $\mathcal{I}(a,b,c,x) = \int_0^x e^{-ct} t^{a-1} (t+1)^{b-a-1} dt$. Since user i's total mutual information MI_i^{total} is a sum of contributions from the k packets transmitted by the user, we eventually rewrite the outage probability (3) as

$$
\begin{aligned}
p_{out_i} &= \Pr\left\{R > \mathrm{MI}_i^{total}\right\} = \Pr\left\{R > \sum_{l \in \mathcal{L}_i} \mathrm{MI}_{l,i}\right\} \\
&= \left(F_{\mathrm{MI}_{l_1^i,i}} * f_{\mathrm{MI}_{l_2^i,i}} * \cdots * f_{\mathrm{MI}_{l_k^i,i}} \right)(x) \Big|_{x=R}
\end{aligned} \tag{9}
$$

where $\mathcal{L}_i = \{l_1^i, l_2^i, \ldots, l_k^i\}$ is the set of indices of TF-blocks where user i transmits.

In addition to the exact outage probability given by (9), which might be cumbersome to evaluate for larger number of replicas k as it requires multiple numerical convolutions/integrations, we provide here also it's Chernoff bound. Let us start by deriving the moment generating function of the mutual information

$$
\begin{aligned}
& M_{\mathrm{MI}_{l,i}}(t) = \mathrm{E}\left[e^{t \ln(1+\mathrm{SINR}_{l,i})} \right] \\
& = \int_0^\infty (1+x)^t f_{\mathrm{SINR}_{l,i}}(x) dx \\
& = \int_0^\infty \frac{x^{M-1} e^{-\frac{x}{\gamma_l}}}{(1+x)^{C_l-t}} dx \sum_{a=0}^{C_l-1} \binom{C_l-1}{a} \frac{M \gamma_l^{a-M}}{(M-a)!} \\
& + \int_0^\infty \frac{x^{M} e^{-\frac{x}{\gamma_l}}}{(1+x)^{C_l-t}} dx \sum_{a=0}^{C_l-1} \binom{C_l-1}{a} \frac{\gamma_l^{a-M}}{(M-a-1)!} \\
& = U(M, M+1-C_l+t, \frac{1}{\gamma_l}) \sum_{a=0}^{C_l-1} \binom{C_l-1}{a} \frac{M! \gamma_l^{a-M}}{(M-a)!} \\
& + U(M+1, M+2-C_l+t, \frac{1}{\gamma_l}) \sum_{a=0}^{C_l-1} \binom{C_l-1}{a} \frac{M! \gamma_l^{a-M}}{(M-a-1)!}
\end{aligned} \tag{10}
$$

where $U(a,b,c)$ is the confluent hypergeometric function of the second kind. The outage probability can be then upper-bounded as

$$
\begin{aligned}
p_{out_i} &= \Pr\left\{R > \mathrm{MI}_i^{total}\right\} \\
&= \Pr\left\{ e^{-t \sum_{l \in \mathcal{L}_i} \mathrm{MI}_{l,i}} > e^{-tR} \right\} \quad , \ t \in \mathbb{R}^+ \\
&\leq \min_{t>0} \frac{\mathrm{E}\left[e^{-t \sum_{l \in \mathcal{L}_i} \mathrm{MI}_{l,i}} \right]}{e^{-tR}} = \min_{t>0} \frac{\prod_{l \in \mathcal{L}_i} \mathrm{E}\left[e^{-t \mathrm{MI}_{l,i}} \right]}{e^{-tR}} \\
&= \min_{t>0} e^{tR} \prod_{l \in \mathcal{L}_i} M_{\mathrm{MI}_{l,i}}(-t).
\end{aligned} \tag{11}
$$

B. Uniform Patterns

We can now apply the tools developed in the preceding section to some specific cases of grant-free access with preallocated patterns. We will start with the most straightforward approach in which the patterns use the L available slots evenly (in other words, their covering is uniform). Consequently, $C_l = \frac{kN}{L}$ for all $l \in \{1, \ldots, L\}$. Similarly, we will consider the same transmit power for each replica, which yields $P_{l,i} = \frac{P_{tot}}{k}$ for all $l \in \{1, \ldots, L\}$ and $i \in \{1, \ldots, N\}$. With the slight abuse of notation, this allows to simplify (9) as

$$
p_{out} = \left(F_{\mathrm{MI}} * \underbrace{f_{\mathrm{MI}} * \cdots * f_{\mathrm{MI}}}_{k-1} \right)(x) \Bigg|_{x=R} \tag{12}
$$

where the CDF F_{MI} and all $k-1$ pdfs f_{MI} are defined as in (8), (7) and with identical parameters C_l, γ_l. In a similar manner, the Chernoff bound simplifies to

$$
p_{out} \leq \min_{t>0} e^{tR} \left(M_{\mathrm{MI}}(-t) \right)^k \tag{13}
$$

---

[3] Strictly speaking, in our scenario the parameter γ_l can have two values: either P_l/σ^2 or 0. In the latter case, the pdf should be replaced by a Dirac delta distribution (and consequently Heavyside step function for cdf).

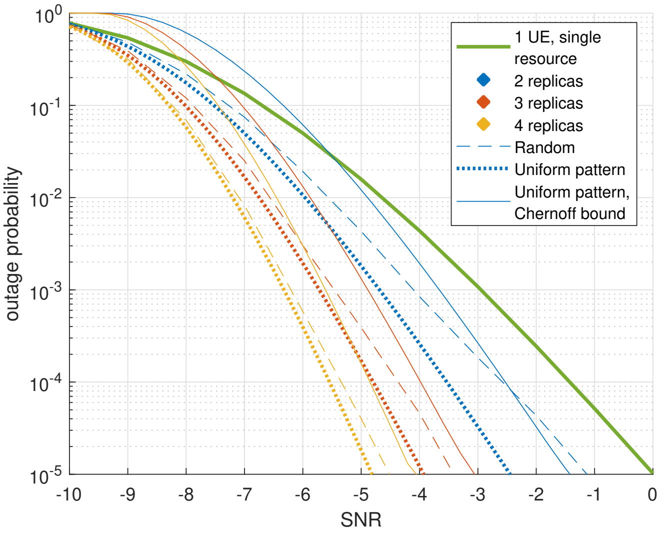


Fig. 3. Outage probability performance of uncoordinated grant-free (idealized) and with preassigned patterns.

In Fig.3 we compare the performance of the idealized grant-free with random selection (no pilot collisions) and the just described approach based on uniform patterns. The parameters used are the same as before with $L = 12$, $N = 24$, $M = 8$ and no antenna correlation. We can see that not only we were able to recover the performance of the idealized random scheme but even improve on it. This is because the ability to coordinate interference allows to avoid too heavily congested TF-blocks and protects against the most detrimental cases where $C_l > M$. With the solid lines of the appropriate color we provide also the Chernoff upper bound on the outage probability of the scheme with deterministic patterns. The bound offers reasonably good approximation by being around 1dB from the actual curve and the gap decreases with higher number of replicas.

C. Transmissions with shared diversity resources

In addition to the regular patterns evenly utilizing all resources, in this work we extend also the concept originally introduced in [12] coined transmissions with shared diversity resources. There, the main idea was to distribute available TF-blocks in such a way that each UE had one dedicated, interference-free slot and additional transmissions were performed on the remaining $L - N$ diversity resources. Clearly, such a scheme requires at least as many TF-blocks as the total number of users, which does not scale very well with the number of served UEs (though such approach was also justified by the fact that only a single receive antenna was considered).

The fact that we consider a BS with multiple antennas allows for relaxing the requirement of fully dedicated resources. To that end, we divide the TF-blocks into two parts: L_L blocks having lower contention and L_H blocks with higher contention ($L_l + L_H = L$). Consequently, each UE will be assigned an access pattern which consists of k_L transmissions located in the first part and k_H transmissions somewhere in the second part ($k_L + k_H = k$). The contention levels for the two types of slots can be calculated in a similar way as before and with these new parameters are

$$C_T = \frac{k_T N}{L_T}, \quad T \in \{L, H\}. \tag{14}$$

The example shown in Fig. 1 can be viewed as one instance of this scheme where $L_L = 4$, $L_H = 4$, $k_L = 1$, $k_H = 2$, $C_L = 1$ and $C_H = 2$.

Due to the introduced asymmetry we will also consider an unequal power allocation: namely users will transmit the two types of packets with powers P_L and P_H respectively. The modified outage probability (9) corresponding with the described scheme is given by

$$\begin{aligned} p_{out} &= \Pr\left\{ R > \sum_{l\in\mathcal{L}_{L_i}} \mathrm{MI}_{l,i} + \sum_{l\in\mathcal{L}_{H_i}} \mathrm{MI}_{l,i} \right\} \\ &= \left(F_{\mathrm{MI}_L} * \underbrace{f_{\mathrm{MI}_L} * \cdots * f_{\mathrm{MI}_L}}_{k_L-1} * \underbrace{f_{\mathrm{MI}_H} * \cdots * f_{\mathrm{MI}_H}}_{k_H} \right)(x)\Bigg|_{x=R} \end{aligned} \tag{15}$$

where in a similar manner as before we denote the indices of low and high contention slots of user i with $\mathcal{L}_{L_i}$ and $\mathcal{L}_{H_i}$ respectively. In the last expression, F_{MI_L} and first $k_L - 1$ f_{MI_L}'s are evaluated with parameters $C_l = C_L$, $\gamma_l = \frac{P_L}{\sigma^2}$ and the next k_H pdfs are evaluated with parameters $C_l = C_H$, $\gamma_l = \frac{P_H}{\sigma^2}$, The optimal power allocation minimizing the outage probability (15) can be found by solving the problem

$$\min_{\{P_L, P_H\}} \quad p_{out} \tag{16a}$$

$$\text{s.t.} \quad k_L P_L + k_H P_H = P_{tot} \tag{16b}$$

Across the range of scenarios considered for this paper, we found the powers P_L, P_H obtained through (16) to be only slightly different from the equal power allocation case. Namely, for the P_{tot} in the range of interest[4] (i.e. providing outage probability 10^{-3} or lower) $\frac{P_L}{P_H}$ is between 1.1 and 1.2 which corresponds to their absolute values being around 5% to 10% off from the uniform $P_l = \frac{P_{tot}}{k}$.

Next, we compare the performance of the TSDR access scheme with optimal powers to that of the uniform access patterns described in Subsection IV-B. In terms of outage probability, TSDR performs 0.1dB - 0.2dB worse than the scheme with uniform patterns which is a marginal difference.

However, the asymmetric patterns turn out to offer some other, less obvious benefits, which we will now demonstrate. For the purpose of the subsequent discussion we downselected two representative scenarios (2 and 4 replicas) which have the following parameters. In case of both 2 transmissions and 4 transmissions the split between resources is $L_L = 8$ and $L_H = 4$ while $k_L = k_H = 1$ in the former and

---

[4] For lower P_{tot} the difference between powers is more significant, however this is not the region of operation relevant to URLLC

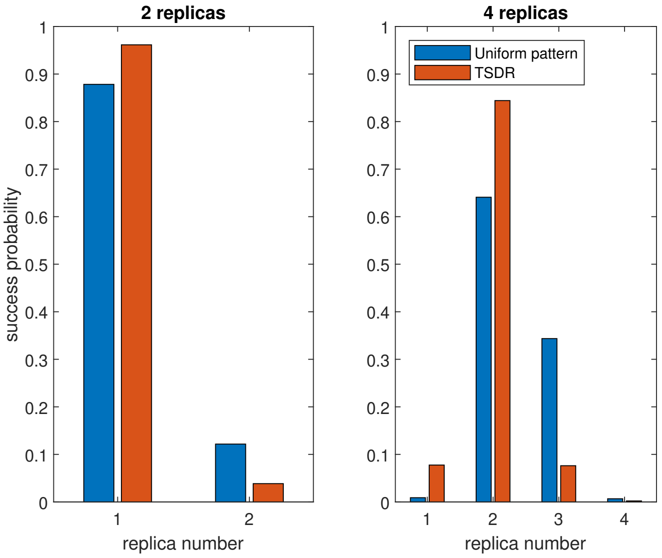


Fig. 4. Probability of decoding with a given replica number

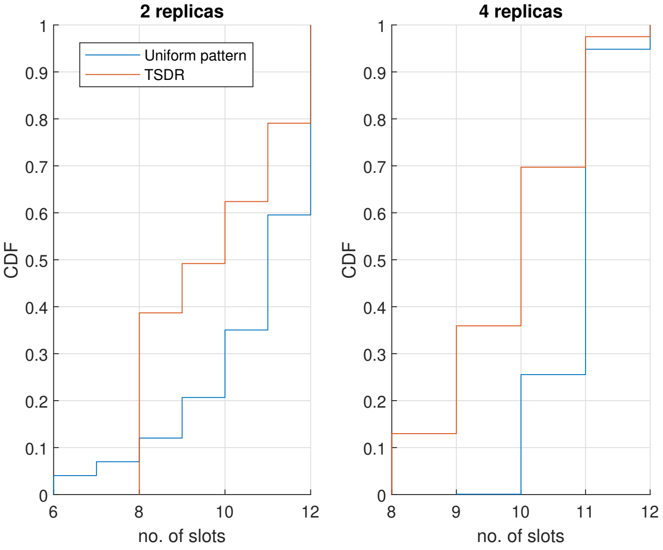


Fig. 5. CDF of the number of slots required until all packets are decoded

$k_L = k_H = 2$ in the latter scenario. Consequently, the low and high contention levels are $C_L = 3$, $C_H = 6$ and $C_L = 6$, $C_H = 12$ respectively.

In Fig. 4 we show the probability of decoding a packet with a given replica number. The immediate observation is that TSDR, which employs asymmetric patterns, yields a much higher chance to decode the packet early. For instance, when UEs transmit 4 times, BS will need more than two replicas only $\sim 7\%$ of the time when TSDR is used compared to $\sim 35\%$ with the uniform scheme. From the practical point of view, these results translate to lower effective latency of individual packets for TSDR. The two factors responsible for this effect are the higher P_L and (simultaneously) lower C_L.

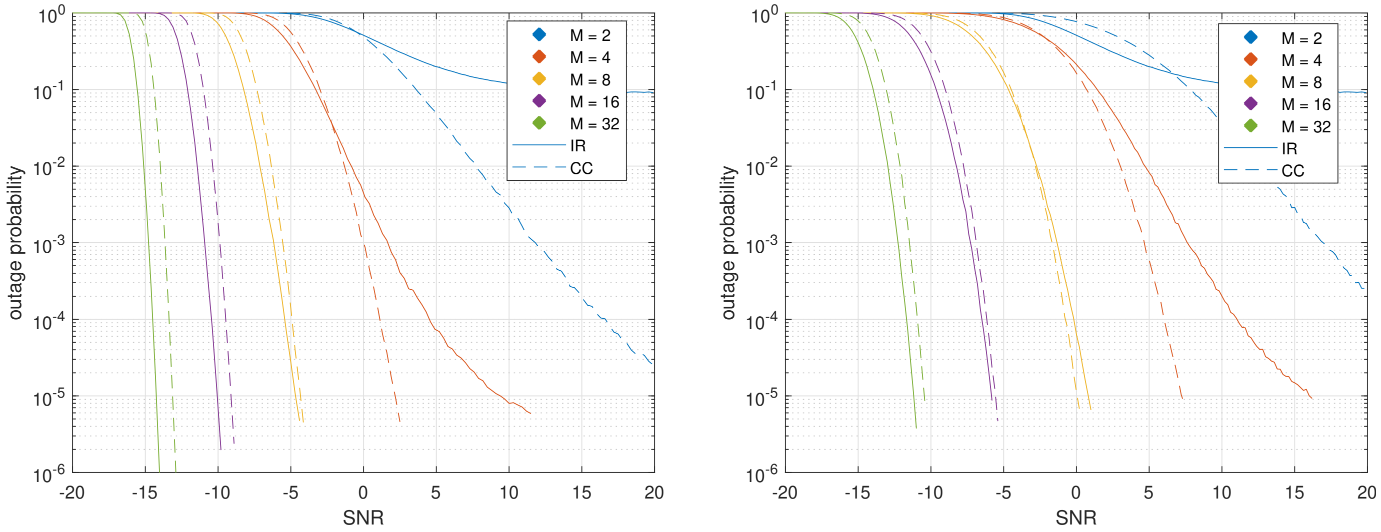
In Fig. 5 we present the CDF of the number of received TF-blocks required to decode all packets. This type of performance can be viewed as an indicator of two other metrics. One is the total, system-wide latency and the other is the complexity as each additional TF-block entails more processing. Again, the approach based on asymmetric power and patterns offers tangible gains.

Lastly, in Fig. 6 we investigate the impact of the antenna correlation and their total number on the TSDR scheme. The chosen scenario is the one with 4 replicas, and non-uniform access patterns with contention levels $C_L = 6$ and $C_H = 12$. With low number of antennas (less than the contention level) the performance is degraded and exhibits plateauing, which is to be expected as the BS receiver does not have the required degrees of freedom to separate all transmissions. An interesting case is the one with $M = 8$ as we arrive at a situation where for some transmissions $C_l = C_H > M$, and yet the penalty in terms of outage probability is not as prominent as one would expect.

In the situations with fewer antennas (or high correlation between them, which reduces their effective number) the performance can be recovered to some extent by switching from Incremental Redundancy to the Chase Combining mode of operation. Since in CC all replicas of the packet are identical, then instead of considering matrices $\mathbf{H}_l$ from each slot individually, they can be stacked together to obtain a single $\mathbf{H} \in \mathbb{C}^{LM \times N}$ similarly as in [12]. This way, the transmissions are processed jointly based on a total of LM measurements rather than by solving many underdetermined problems. By comparing Fig. 6(a) with 6(b) to assess the impact of correlation, we can conclude that the degradation at 10^{-5} outage ranges from 3dB for $M = 32$ antennas, up to 7dB for $M = 4$. Again, the case with $M = 8$ is the most interesting as it shows that with a reduction in the effective number of antennas, CC becomes preferable to IR.

V. CONCLUSIONS

In this work we have studied the reliability aspects of MIMO URLLC systems with grant-free access, operating in either uncoordinated manner or with preassigned channel resources. Using recently derived results on the SINR distribution of MMSE multi-antenna receivers, we have derived analytical results and bounds on the outage probability of the studied schemes under different conditions. Our results show that, although a totally uncoordinated scheme performs well when perfect CSI is assumed, preallocation of the channel resources provides an effective way to avoid pilot sequence collisions and to limit the maximum contention levels in each slot. In addition, we have also found that dividing the resource pool into two types of resources, with low and high contention levels, can help reducing the receive processing latency with virtually no loss in terms of reliability. Overall, we conclude that the combination of multi-antenna processing at the receiver with intelligent design of the preallocated resources can significantly boost the performance of URLLC systems, even in the presence of strong receive antenna correlation.



(a) (b)

Fig. 6. The effect of varying number of receive antennas M in case of (a) no correlation and (b) $\rho = 0.85$

APPENDIX

Let $d_{l,i}$ denote a specific sequence chosen by user i who is active in slot l. Furthermore, let us denote by $\mathcal{J}_l$ a subset of indexes of the active UEs who selected the same sequence as some other UE (e.g. if users 1,2,3,5,7,9 transmitted in the same TF-block l and: 1, 3, 7 used the same sequence a; 2 and 5 used sequence b; 9 used sequence c then $\mathcal{J}_l = \{1, 2, 3, 5, 7\}$). More formally $\mathcal{J}_l = \{i : (\exists j)\,[d_{l,i} = d_{l,j} \ \wedge \ P_{l,i}, P_{l,j} > 0]\}$. We assume that the receiver is not able to estimate the channel coefficients of the users involved in pilot collisions. As a consequence, the corresponding transmissions are lost and become a part of the noise.

From the signal processing point of view, we deal with this case by defining a new matrix $\mathcal{H}_l$ which is the original $\mathbf{H}_l$ with columns $\mathcal{J}_l$ set to $\mathbf{0}$. Note, that the optimal MMSE detector in this case is also different [15] and has a form $\mathbf{F}_l = \left(\mathcal{H}_l{}^{\mathrm{H}}\boldsymbol{\Sigma}_l^{-1}\mathcal{H}_l + \mathbf{I}_N\right)^{-1}\mathcal{H}_l{}^{\mathrm{H}}\boldsymbol{\Sigma}_l^{-1}$ where $\boldsymbol{\Sigma}_l = \sigma^2\mathbf{I} + \sum_{a\in\mathcal{J}_l}(\mathbf{h}_l)_a(\mathbf{h}_l)_a^{\mathrm{H}}$ is the new $\mathbb{C}^{M\times M}$ noise covariance matrix with $(\mathbf{h}_l)_a$ being the columns of $\mathbf{H}_l$. This matrix is in fact unknown due to the assumption stated earlier, however in a simplified scenario with no antenna correlation $\boldsymbol{\Sigma}_l$ becomes diagonal with i-th diagonal element equal to $\sigma^2 + \sum_{a\in\mathcal{J}_l}|(\mathbf{H}_l)_{i,a}|^2$ (which requires from the BS only the knowledge of the total magnitude of the combined noise-plus-interference).

ACKNOWLEDGEMENT

The work of Radosław Kotaba and Petar Popovski was funded by the European Research Council (ERC), Consolidator Grant Nr. 648382 (WILLOW), under the European Union Horizon 2020 research and innovation program.

REFERENCES

[1] 3GPP TS 38.300, "NR;NR and NG-RAN Overall description; Stage-2," Sep. 2019, v:15.7.0

[2] P. Schulz et al., "Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture" in *IEEE Commun. Mag.,* vol. 55, no. 2, Feb. 2017

[3] O. Kaiwartya et al., "Internet of Vehicles: Motivation, Layered Architecture, Network Model, Challenges, and Future Aspects" in *IEEE Access,* vol. 4, pp. 5356-5373, Sep. 2016

[4] B. Soret, P. Mogensen, K. I. Pedersen and M. C. Aguayo-Torres, "Fundamental Tradeoffs among Reliability, Latency and Throughput in Cellular Networks," *IEEE GLOBECOM*, Dec. 2014

[5] M. Bennis, M. Debbah, H. V. Poor, "Ultra-Reliable and Low-Latency Wireless Communication: Tail, Risk and Scale," in *Proceedings of the IEEE,* vol. 106, no. 10, Oct. 2018

[6] 3GPP TR 36.881, "Study on latency reduction techniques for LTE," 2016, v:14.0.0

[7] B. Singh, O. Tirkkonen, Z. Li, M. A. Uusitalo "Contention-Based Access for Ultra-Reliable Low Latency Uplink Transmissions" in *IEEE Wireless Communications Letters,* vol 7, no. 2, Apr. 2018

[8] E. Paolini, C. Stefanović, G. Liva, Petar Popovski, "Coded random access: applying codes on graphs to design random access protocols" in *IEEE Commun. Mag.,* vol. 53, no. 6, Jun. 2015

[9] M. C. Lucas-Estañ, J. Gozalvez and M. Sepulcre. "On the Capacity of 5G NR Grant-Free Scheduling with Shared Radio Resources to Support Ultra-Reliable and Low-Latency Communications" in *Sensors* 2019, 19, 3575.

[10] T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, P. Mogensen, I. Z. Kovacs, T. K. Madsen, "System Level Analysis of Uplink Grant-Free Transmission for URLLC" in *Proceedings of the 2017 IEEE Globecom Workshops (GC Wkshps)*, Singapore, 4–8 Dec. 2017;

[11] N. H. Mahmood, R. Abreu, R. Böhnke, M. Schubert, G. Berardinelli, T. H. Jacobsen, "Uplink Grant-Free Random Access Solutions for URLLC services in 5G New Radio" in *2019 16th International Symposium on Wireless Communication Systems (ISWCS),* Aug. 2019

[12] R. Kotaba, C. N. Manchón, T. Balercia, P. Popovski, "Uplink Transmissions in URLLC Systems With Shared Diversity Resources" in *IEEE Wireless Communications Letters,* vol. 7, no. 4, Aug. 2018.

[13] C. Boyd, R. Kotaba, O. Tirkkonen, P. Popovski "Non-Orthogonal Contention-Based Access for URLLC Devices with Frequency Diversity" in *2019 IEEE 20th Int. Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jul. 2019

[14] H. Lim, D. Yoon, "On the Distribution of SINR for MMSE MIMO Systems" in *IEEE Transactions on Communications* , vol. 67, no. 6, Jun. 2019.

[15] T. Brown, E. De Carvalho, P. Kyritsi, "Practical Guide to the MIMO Radio Channel," in *John Wiley & Sons Ltd*, 2012