Deterministic Patterns for Multiple Access in Latency-Constrained Ultra-Reliable Communications

Radosław Kotaba, Member, IEEE, Roope Vehkalahti, Member, IEEE, Čedomir Stefanović, Senior Member, IEEE, Olav Tirkkonen, Senior Member, IEEE and Petar Popovski, Fellow, IEEE

Abstract

The grant-free access is envisioned as one of the enablers of the ultra-reliable low-latency communications. Yet, when there are many devices that tend to be active only intermittently, the fully orthogonal resource allocation is largely inefficient. The solution is to employ a common, shared pool of resources and account for the fact that some collisions and interference will inevitably occur. In this contribution we study the reliability aspects of such multi-user uplink communication scenario over a shared pool of channel resources, where intermittently active devices utilize multiple transmissions (K-repetition coding) to achieve diversity. We focus on two access methods – one where devices choose the K slots at random and one where the access patterns are deterministic and follow a specific code design, namely the Steiner System. We analyze the problem under two signal models that involve different complexity for the receiver. Firstly, a model which treats collisions as destructive, i.e. only those K′ among K transmissions that do not contain interference can be used and combined. Second, where receiver is capable of utilizing all K replicas and applies maximum ratio combining (MRC) treating interference as noise. Furthermore, in both cases we investigate the receiver with and without successive
interference cancellation (SIC) capabilities. As one of the main contributions of this work, we develop useful approximations and bounds for the outage probabilities in the aforementioned scenarios that match very closely the simulation results. We also show that deterministic patterns have the potential to significantly outperform fully random selection, both in terms of raw performance and by simplifying the system design.

Index Terms

grant-free, radio resource management, uplink, ultra reliable low latency communication (URLLC)

I. INTRODUCTION

The latest generation of wireless systems, 5G networks, are becoming more and more widely adopted [1], while the researchers and the industry already plan their next steps by laying ground for technologies that will come next [2]. Importantly, the shift to 5G and beyond is not just about the need for higher data rates, but is driven by a new types of use cases and applications for which the support is required from the network. Among those, particularly relevant and, simultaneously, challenging are the use cases that fall under the category of ultra-reliable low-latency communications (URLLC). These are characterized by especially stringent end-to-end (E2E) latency constraints (between $0.5 - 2$ ms) and reliability (i.e. the probability of successful delivery of a packet) of $99.999\%$ [3]. Among the most prominent URLLC applications are those that involve tactile interaction, intelligent transport and factory automation [3], [4].

The primary challenge in designing ultra-reliable systems with stringent latency constraints is to not overly compromise their spectral efficiency in the process [5]. This is particularly difficult to achieve in the uplink, which in traditional networks is centrally managed by the base station (BS) and relies on either explicit grants (more efficient but with high latency due to excessive signalling) or pre-allocation of resources (low latency but inefficient when the traffic is intermittent). As such, in order to fulfill the demanding latency and reliability targets, new uplink access protocols and modes of operation have to be devised for 5G and beyond.

One way to tackle the problem is to rely on random access based communication. Within that family, perhaps the most well-known approach is the grant-free access [6], in which user equipments (UEs) are allowed to transmit without prior, explicit scheduling. Instead, a certain portion of bandwidth is delineated and provided to a group of users who can use it whenever
they have data to send. The benefit of that is significant reduction in the signalling overhead and connected with it latency. As a matter of fact, scheduling contributes the most to the E2E delay making it the main bottleneck when designing URLLC systems [7]. However, it should be noted, that grant-free as a solution is particularly suitable for and motivated by traffic that is relatively infrequent and irregular [8]. Due to the sharing of the resources and lack of coordination it is inherently less reliable than its grant-based counterparts and without explicit control from the BS, the uplink signals of URLLC users are prone to collisions and interference. Clearly, to compensate for that, additional mechanisms which will improve the reliability are needed.

The simplest and at the same time most effective solution is to introduce redundancy through multiple transmissions [9]. This improves the reliability in two ways. Firstly, by increasing the chance that at least some of the replicas reach the BS uninterfered, and secondly, by providing diversity that allows to combat the negative effects of the fading channel. The second technique builds upon the concept of multiple transmissions. Instead of letting UEs select the resources from the pool completely at random, the idea is to structure the transmissions of the individual users into access patterns. These access patterns can be constructed in many ways and with different goals in mind, but in general they aim to provide certain reliability guarantees [10]. The drawback of such solution is that their assignment requires some coordination with the BS and signalling, which makes it less flexible than purely random selection. However, this operation can be integrated into the registration procedure that each device has to perform anyway when it first attaches to the BS or wakes up and re-synchronizes after being inactive for a prolonged period. Furthermore, even in a fully random scheme the device needs to be configured at least with the number of repetitions and portion of bandwidth where the grant-free pool is located.

Lastly, on the receiver side, there is a possibility to implement successive interference cancellation (SIC). With SIC, it is possible to iteratively decode signals by gradually removing the interference. Namely, in each round the packets that were successfully decoded in the preceding rounds can be subtracted (after proper equalization) from the received signal, thus improving the signal-to-interference-plus-noise (SINR) of the remaining ones. This is a powerful technique and especially relevant when dealing with traffic that is non-orthogonal by design [11].

A. Related work

Fundamentally, the grant-free techniques descend from one of them most well known concepts in the field of communication - slotted ALOHA [12]. Since its inception, many extensions have
been proposed. One of them is the Content Resolution Diversity Slotted ALOHA [11], [13],
which utilizes multiple transmissions (with the goal of achieving diversity) and the interference
cancellation. In [14], authors analyze a variant of this scheme - Irregular Repetition Slotted
ALOHA (IRSA) in which the number of repetitions is not fixed, but follows a certain discrete
distribution. Furthermore, in [15] the analysis is extended to the Rayleigh fading channel and
optimization of the repetition degree is presented. Another extension coined Coded Slotted
ALOHA [16] involves transmitting different coded version of the packet (redundancy versions)
rather than exact replicas, which allows to achieve better granularity in terms of transmission rate.
In [17] the author analyzes the throughput of CSA in a multichannel Rayleigh fading scenario.

In the above works, the primary focus is on the maximization of spectral efficiency of the
systems rather than achieving high reliability. The grant-free access methods as envisioned
for 5G and designed specifically for URLLC have been thoroughly researched in [18]. In its
thesis, author investigates different repetition and retransmission schemes in realistic scenarios
based on detailed system level simulator. In [19], authors focus on the combinatorial aspects
of the repetition-based 5G grant-free scheme, namely its probability of collisions, and evaluate
achievable reliability and latency levels as a function of the number of UEs, amount of pre-
allocated resources and number of replicas. Comparison to other schemes has been shown in
[20], where authors jointly evaluate repetition coding, its proactive version (where the UEs have
the possibility of early termination), and the more traditional Hybrid Automatic Repeat Request
(HARQ) based on feedback and retransmissions. Most recently, in [21] the author extends the
idea of repetition-based schemes towards network coding, making it better suited for scenarios
where devices have more than one packet to transmit at a time. A different approach is considered
in [22], where the resources are first assigned to a group of users based on sensing, and then
between themselves UEs avoid the collisions by signalling transmission announcements.

In addition to fully random grant-free transmission schemes, some authors have investigated
the access based on pre-allocated patterns and their optimal design. In [23], which is one of the
earliest works, the pattern construction is based on combinatorial design. However, authors do
not consider SIC and treat all the collisions as destructive. More recently, the designs oriented
towards interference cancellation have been considered in [24] [25] and [26]. In [24] the patterns
belong to the class of (≤ M, 1, n)-locally thin codes, in [25] are based on the (t, K, M) Steiner
Systems, while in [26] authors use LDPC codes. The idea to use deterministic access patterns
also appeared in [27], where they are used in conjunction with multiple antenna processing at
the BS. Furthermore, the channel resources are divided into high and low contention parts over which power optimization is additionally considered. Differently from the aforementioned, in [28] the patterns, are applied on the symbol-level rather than over slots.

B. Contributions

In this work we study the grant-free multiple access in which users apply access patterns, i.e. sequences consisting of multiple redundant transmissions, to achieve ultra-reliable communication. The framework involves a shared pool of resources - a short, periodic frame composed of limited number of slots, that makes our contribution relevant in scenarios with tight latency constrains, especially URLLC. We focus on the comparison between fully random selection of slots and a case of pre-assigned, deterministic patterns that are inspired by a specific code construction, known as the Steiner system. The latter is chosen due to its desirable properties, namely its construction ensures that two patterns can share at most \( t - 1 \) slots, where \( t \) is a design parameter. In other words, it provides guarantees in terms of the amount of collisions/interference.

As a main contribution we provide a thorough analysis of the grant-free access based system in terms of its outage probability performance and spectral efficiency. In our analysis we consider two different signal models. One, resembling a traditional ALOHA system, where collisions are destructive, i.e. only the slots that contain a transmission of a single device can be used. Unlike in traditional ALOHA however, we allow the receiver to combine multiple collision-free replicas from a given user. In the second model, receiver is capable of utilizing all transmissions and performs their maximum ratio combining (MRC) accounting for different SINRs. Furthermore, for each of the two signal models we consider two subcases: with and without SIC at the receiver. In all of the aforementioned configurations access patterns given by a Steiner system exhibit clear gains over their fully random counterparts. Furthermore, their regular structure simplifies the overall system design. As such, we believe that Steiner systems make for a compelling solution in the design of grant-free access.

A particularly important contribution are the approximations and bounds developed for the collision model with SIC and Full MRC without SIC - two considerably non-trivial cases! The developed expressions match very closely the extensive simulation results obtained with Monte Carlo methods. The approximations are especially valuable in the context of URLLC, where relying on simulations alone is often not feasible due to the sheer number of samples required. To the best of the authors knowledge, these results are novel and have not been reported before.
Fig. 1. Example of the uplink access scenario with $K = 3$ multiple transmissions over a frame of $M = 7$ slots. There are $N = 4$ UEs out of which $U = 3$ happen to be active. Their transmissions cause collisions in slots 3 and 6.

This work extends our prior contribution [25] in several meaningful ways. Firstly, we provide in-depth analysis and develop approximations and bounds that go well beyond the results reported earlier. We also present formal proofs of the combinatorial results in [25] that treat the distribution of the number of collision-free slots and number of interferers, and which were omitted due to space constraints. Secondly, we extend the scenario by considering receiver with SIC capabilities. We also broaden the scope by considering other Steiner systems with different parameters (frame length and number of repetitions). Lastly, we discuss the limitations of access methods based on Random selection highlighting issues with their practical implementation.

The rest of the paper is organized as follows. We start by introducing the system and signal model in Section II. In Section III we introduce the two types of access patterns and discuss their properties. Then, in Section IV we consider different receiver processing techniques and provide their thorough analysis in the context of the access patterns from Section III. This is complemented by both analytical results and corresponding simulations. In Section V, we discuss the deficiencies of the Random selection approach and the challenges wrt. its practical implementation. In Section VI we compare different Steiner Systems using the analytical results derived earlier. Lastly, in Section VII we offer final conclusions that close the paper.

II. SYSTEM MODEL

We consider a communication system with a single base station (BS) serving a population of $N$ intermittently active users (UEs) transmitting in the uplink. The access channel is composed of
periodic frames, which are further broken down into $M$ access opportunities otherwise known as slots. We assume that UEs are independently activated in a frame with probability $b$, such that the total number of active devices $U$ follows a binomial distribution $f_{\text{bin}}(u; b, N) = \binom{N}{u} b^u (1-b)^{N-u}$. Whenever active, a user selects $K$ out of $M$ slots in a frame and uses them to transmit its packet, thus employing a form of $K$-repetition coding. It is further assumed that all UEs transmit with the same rate $R$ measured in bits per channel use (c.u.). The described model is visualized in the example in Fig. 1. In general $M$ is determined by the allowed latency, while $K$ is a design parameter that depends on the number of users $N$ and the activation probability $b$.

In this work we consider a Rayleigh block fading channel, where the realizations of channel coefficients are independent across slots and UEs. At the receiver, the baseband representation of the channel output in a frame is modeled as

$$y = Hx + w$$

$$= G \circ (\text{VAP})x + w \in \mathbb{C}^{M \times 1}$$

where $x \in \mathbb{C}^{N \times 1}$ is the vector of complex modulated symbols transmitted by users such that $E[|x_n|^2] = 1$, $w \in \mathbb{C}^{M \times 1}$ is the additive white Gaussian noise (AWGN) with zero mean and variance $\sigma^2$ and $H \in \mathbb{C}^{M \times N}$ are the channel gains between the $N$ users and the base station in each of the $M$ slots. In (2), $\circ$ denotes entry-wise (Hadamard) product. The channel gains $H$ can be represented as a product of $G \in \mathbb{C}^{M \times N}$ - which is zero-mean, unit-variance, circularly-symmetric complex Gaussian (ZMCSCG) and models the underlying uncorrelated Rayleigh flat fading channel, $V \in \{0,1\}^{M \times N}$ - matrix representing the access patterns of all users such that $V_{m,n} = 1$ if user $n$ is assigned to slot $m$, $A \in \{0,1\}^{N \times N}$ - the diagonal matrix designating which users are active in a frame and $P = \text{diag} \left( (P_1)^{1/2}, \ldots, (P_N)^{1/2} \right)$ which is a diagonal matrix of square roots of average powers, i.e. signal amplitudes applied by UEs.

1 Although in Fig. 1 the slots seem to be arranged in time, this is not a requirement. The slots can also represent different frequencies/groups of frequencies (subcarriers) or be 2-dimensional constructs (similar to Resource Blocks in LTE/5G).

2 We assume that the transmissions within a single frame are self contained and independent, i.e., UEs are not supposed to transmit their packets over multiple frames as that would violate the implicit latency constraint. Consequently, there is no need to introduce additional index to denote the frame number in the signal model.

3 Note that a distance-dependent path loss term is absent in the signal model. Throughout this work we assume that UEs know the long-term statistics of the channel and based on that compensate for path loss accordingly. If we were to denote by $d_n$ the path loss of user $n$, then the actual transmit power would be $d_n P_n$ such that $P_n$ represents the average received power. Due to this compensation and for the sake of simplicity, we decide to omit the path loss term altogether.
In the remainder of this work we assume that all active UEs use the same power $P_x$. Consequently, we can define the average received SNR

$$\theta = \frac{P_x}{\sigma^2}. \quad (3)$$

For a given user $n$, the choice of the slots which are used for transmission, i.e. the set of indices $\{j : V_{j,n} = 1\}$ constitute what we call an access pattern. The matrix of slot selections $V$ may be fixed, such that UEs follow access patterns that were pre-assigned to them, or it may be random. Furthermore, since each UE uses only $K$ among $M$ available slots we have that $\sum_{m=1}^{M} V_{m,n} = K$.

III. ACCESS PATTERNS: RANDOM VS. DETERMINISTIC

In this section we introduce, and later on compare, two access methods that could be employed by the devices which try to communicate over a shared pool of resources (i.e., slots). In the following we describe one which relies on random selection, and one in which users have pre-assigned, deterministic access patterns.

A. Random selection

We start with an approach in which users transmit their $K$ packet replicas over the $M$ available slots by selecting slots uniformly at random. We are interested in determining the probability of having a certain number of interference-free slots, i.e. replicas which do not experience collisions with other UEs’ replicas.

Consider a frame in which $U \geq 1$ out of a population of $N$ users is active and, without loss of generality, focus on a single, arbitrary user $u$.

**Lemma 1.** From the point of view of a user $u$, the probability that the remaining $U - 1$ users do not cause collisions to exactly $K'$ out of $K$ of its replicas (collision-free (CF)) when using random patterns (R) is given by

$$p_{CF,R}(K'|U) = \binom{K}{K'} \sum_{n=0}^{K-K'} (-1)^n a_n V_n, \quad (4)$$

where $a_n = \binom{K-K'}{n}$ and $V_n = \left( \frac{M-K'}{K} \right)^{U-1}$.

The proof of Lemma 1 can be found in Appendix B.
Another relevant metric when considering such a contention-based access and which may have an impact on the decoding is the distribution of the number of interferers $L$ in a given slot in which the reference user is active. Since users are free to select any of the $\binom{M}{K}$ possible sequences and do so independently from each other (with replacement) we have that $L \in [0, U - 1]$ and

$$p_{I,R}(L|U) = \frac{\binom{M-1}{K}^{U-1-L} \binom{M-1}{L}^{L}(U-1)}{\binom{M}{K}^{U-1}}.$$  \hspace{1cm} (5)

### B. Deterministic patterns

Another approach to the contention-based access over a shared pool of resources is the one in which UEs have fixed, pre-assigned access patterns. Such a solution is less flexible, as it requires coordination with the BS, who is responsible for assigning them, however it has the potential to greatly improve the overall reliability of the system. Typically, a pattern would be assigned when the device registers with the BS for the first time or wakes up and re-synchronizes after being in power-efficient mode. They can be also periodically updated. This is the case, for example, when the user population size changes and the resource pool needs to be adjusted; however such updates will occur relatively infrequently compared to the duration of the frame.

In this work we choose to focus on the patterns that are given by a specific block design known as Steiner system. A Steiner system $S(t, K, M)$ can be considered as a $M$-dimensional constant-weight code, where each codeword has $K$ ones, and for any two codewords $s_i, s_j \in S(t, K, M)$, $s_i \neq s_j$, we have $d(s_i, s_j) \geq 2K - 2(t - 1)$, where $d(\cdot, \cdot)$ is the Hamming distance. In other words, two codewords (transmission patterns) can collide on at most $t - 1$ positions.

For fixed $K$ and $M$, the lower the $t$, the smaller the codebook size and thus the number of supportable users. Specifically, we have that $C = |S(t, K, M)| = \binom{M}{t}$. Since in this work our focus is on URLLC applications, we limit our considerations to the case $t = 2$ as it provides high reliability and the support for a massive number of devices in not required.

Another property of the Steiner system is that its structure guarantees that the number of overlapping users in any given slot is at most $D = \binom{M-1}{t-1} = \binom{M}{t-1}$, i.e. there are exactly $D$ access patterns which include a certain slot, so even in the worst case scenario when all of them are active there are at most $D$ mutually interfering users. As we will elaborate later on, this is

\footnote{Technically, the highest reliability is provided when $t = 1$, however such case is trivial as it corresponds to the fully orthogonal allocation of resources.}
an important feature, as it allows to dedicate just the right amount of resources for the pilot sequences and ensure that no pilot collisions occur.

Analogously to the Random selection, we have the following results for the Steiner system.

**Lemma 2.** With devices employing access patterns from a Steiner system \( S(t, K, M) \), the probability that an arbitrary user has \( K' \) out of \( K \) collision-free slots is

\[
p_{\text{CF}, S}(K'|U) = \binom{K}{K-K'} \sum_{n=0}^{K-K'} (-1)^n a_n W_n
\]

where \( W_n = \frac{\binom{C-1}{n+K'} U-1}{\binom{C-1}{U-1}} \).

The proof of the above Lemma is provided in Appendix C.

In terms of the number of interferers \( L \), the situation is much more straightforward. In a slot in which an arbitrary user is active, there are only \( D-1 \) other patterns that could cause a collision and the selection is done without replacement due to the unique preassignment. Hence, it is the probability that \( L \) out of \( U-1 \) devices select one of them while the rest of the devices select any of the remaining \( C-D \) patterns:

\[
p_{I, S}(L|U) = \frac{\binom{D-1}{L} \binom{C-D}{U-1-L}}{\binom{C-1}{U-1}}.
\]

In Fig. 2 we compare a \( S(2,4,25) \) Steiner system (solid line) and a corresponding Random selection (dashed) with the same frame length and number of repetitions.

In terms of the number of collision-free slots \( K' \) shown in Fig. 2(a) the main conclusion is that Steiner system reduces the probability of having the best (\( K' = K \)) and worst (\( K' = 0 \)) outcome, while increasing the probability of having 'good' outcomes such as \( K' = K-1 \), \( K' = K-2 \). The ability to avoid the worst case scenarios is particularly important as the probability of \( K' = 0 \) is tied to the performance floor. Clearly, when all replicas are subject to collisions, increasing SNR is not effective and without a single packet that can be decoded, SIC cannot be applied. In that regard, the structure of the Steiner code becomes a disadvantage as the traffic intensity increases. However, as will become evident later, in most cases ultra-reliability cannot be achieved if the traffic intensity is too high, regardless if the access scheme is based on Steiner system or Random selection. As such, we note that the region of interest is primarily low and medium traffic intensity, where the average number of activated users \( bN < \frac{M}{2} \).
In Fig. 2(b) and Fig. 2(c) we compare the probability distribution of the number of interferers $L$, which has an impact on the SINR in the slots where collisions occur, as well as the utility of the SIC procedure. In other words, having more interferers makes it less likely that all of them can be removed. Once again, the Steiner system ensures that within the traffic intensities of interest, ‘good’ outcomes such as $L = 0, 1, 2$ are more likely, while really congested slots are rare. In fact, by inspecting Fig. 2(c) one can see that unless $bN = 20$ or higher, the cdf of $L$ for the Steiner system is strictly above that of the Random selection. Furthermore, as already mentioned, with Steiner system the number of interferers is strictly limited to $D - 1$, which in this case is 7.

IV. RECEIVER PROCESSING

In this section we analyze different modes and processing techniques employed at the receiver. The metric on which we are focusing is outage probability, as it is particularly relevant for URLLC use cases. We define it as

$$p_{out_i} = \Pr \{ R > \log_2 (1 + \text{SINR}_i) \}$$

where $R$ is the transmission rate in bits per channel use at which the packet is encoded and $\text{SINR}_i$ denotes the final, post-processing signal-to-interference-plus-noise ratio of user $i$’s packet. The exact definition and the means of computing it depend on the chosen scenario, which are the subject of the following subsections.
A. Collision model

We start with a simple model that entails a less computation-intensive processing method at the receiver. In the collision model, collisions are assumed to be destructive, so only the slots containing a transmission of a single device are considered. When the slot $m$ is interference free with only user $i$ transmitting, the received complex baseband signal in (2) simplifies to

$$y_m = \sqrt{P_x} g_{m,i} x_i + w_m$$

and the SNR of that signal is

$$\rho_{m,i} = \frac{P_x |g_{m,i}|^2}{\sigma^2} = \theta |g_{m,i}|^2.$$  

Since channel coefficients are Rayleigh distributed r.v.’s, the SNR of each packet follows exponential distribution $f_{\text{exp}}(\rho; \theta) = \frac{1}{\theta} e^{-\frac{\rho}{\theta}}$. As each device transmits $K$ times, there might be up to $K' \leq K$ collision-free replicas, the probability of which is given by (4) and (6), and it is possible to combine the signals to improve the overall SNR. Denote by $J_i$ the set of indices corresponding to the uninterfered transmissions of device $i$. Using maximum ratio combining (MRC), we obtain

$$\sum_{j \in J_i} g_{j,i}^* y_j = \sqrt{P_x} x_i \sum_{j \in J_i} |g_{j,i}|^2 + \sum_{j \in J_i} g_{j,i}^* w_j$$

that yields the SNR $\rho_i = \theta \sum_{j \in J_i} |g_{j,i}|^2$. As a sum of exponentially distributed r.v.’s, the total SNR after combining, conditioned on $K'$, has a gamma distribution $f_{\text{gam}}(\rho; K', \theta) = \frac{1}{\Gamma(K')} \theta^{K'} \rho^{K' - 1} e^{-\frac{\rho}{\theta}}$.

As follows from (8), the decoding is unsuccessful whenever $\rho_i < 2^R - 1$, hence

$$p_{\text{out}}(R, \theta, U) = p_{\text{CF}}(0 | U) + \sum_{K'=1}^{K} F_{\text{gam}}(2^R - 1; K', \theta)p_{\text{CF}}(K'|U)$$  

where $p_{\text{CF}}(\cdot | U)$ is given by either (4) or (6), depending on whether random or Steiner patterns are used, respectively. The equation can be further marginalized over $U$ to account for the specific activation process.

In Fig. 3 we present the results based on (11) that show the performance of deterministic patterns based on Steiner system, and random selection. The parameters chosen are: $M = 25$ slots, $K = 4$ repetitions and $N = C = 50$ users. The activation probabilities are $b = [0.02, 0.04, 0.1, 0.2]$ that translate to the mean number of active devices in a frame equal to $[1, 2, 5, 10]$ respectively. The transmission rate is $R = 2 \frac{\text{bit}}{\text{c.u.}}$. It becomes clear, that the properties of the Steiner system which were described in the previous section lead to tangible gains in terms of outage probability (up to an order of magnitude lower than Random selection). Nevertheless, even with the traffic intensity as low as 1 UE per frame, ultra-reliability is unattainable unless more sophisticated processing is applied.
Fig. 3. Outage probability performance of the system employing random and deterministic patterns as a function of the average received SNR for different mean number of active devices $bN$.

B. Collision model with successive interference cancellation

A natural extension to the model put forward in the previous subsection is to introduce successive interference cancellation (SIC). SIC involves removing from the received signal the packets which were successfully decoded, including all of their $K$ replicas. This has the potential to greatly improve the performance, as it allows to remove the interference from the slots where collisions occurred and make them usable in the subsequent iterations of the decoding process.

A rigorous analysis of the models involving SIC is known to be inherently difficult [16] and the exact analytical results typically do not exist except for asymptotic cases and some simple special cases. Due to the multitude of possible configurations of the selected transmit patterns and different dependencies they create, the problem becomes intractable already when $U > 4$. This motivates us to look for approximate results. In the presented approach we consider the first few rounds of SIC. In each, we condition on the number of decoded users in the preceding rounds. Due to the unique structure of the Steiner system, which ensures that active users cover the frame uniformly, it is possible to simplify some of the steps by using averages. Indeed, rather than having to sum/integrate over possible outcomes of many variables, we can work with the averages in terms of the number of collision-free slots, the combined SNR, etc. This is in contrast to Random selection, where the covering is uniform with high probability only when there are many active users, while for low $U$ their replicas can be quite concentrated. Consequently, our
approach is suitable only for Steiner systems and will not provide a good approximation if the patterns follow a Random selection.

Similarly as before, the analysis will be performed from a point of view of an arbitrary user. First, consider the case that $l_1 = 1, \ldots, U - 1 - S$ users are decoded in the first iteration of SIC. Here, $S$ denotes the number of users who for some reason are excluded from the procedure and cannot be cancelled (this will be explained in detail later on). If we treat all the $U - 1 - S$ users independently, each with a probability of outage $p_{\text{out}}(R, \theta, U)$ given by (11), the distribution of $l_1$ can be approximated as binomial $f_{\text{bin}}(l_1; U - 1 - S, 1 - p_{\text{out}}(R, \theta, U))$. For the remaining users that were not successful in the first round of SIC, it is important to account for the fact that they nevertheless accumulated some of the power already. Since they failed, their SNRs, given $K'$, are drawn from a truncated gamma distribution $f_{\text{gam}}(\rho; K', \theta)$. By taking its mean and marginalizing over $K'$, we can determine the mean residual SNR, i.e. the amount of signal power that is missing before the packet can be decoded:

$$\rho_{\text{res}} = \sum_{K' = 0}^{K} \left( 2^{R} - 1 - \theta \frac{\gamma\left(\frac{2^{R} - 1}{\theta}, K' + 1\right)}{\gamma\left(\frac{2^{R} - 1}{\theta}, K'\right)} \right) p_{\text{CF}}(K'|U)$$

where $\gamma(s, x) = \int_{0}^{x} t^{s-1} e^{-t} dt$ is a lower incomplete gamma function. Furthermore, let us also define an expected number of collision-free slots per user, which is simply $\hat{K}(U) = \sum_{K' = 0}^{K} K' p_{\text{CF}}(K'|U)$. Since all the collision-free slots from the first round have been already taken into account, we are only interested in the new ones that were released after cancelling the $l_1$ successful users. On average, there will be $\hat{K}(U) - \hat{K}(U - l_1)$ new slots, so the probability of decoding a packet in the second round of SIC is

$$p_{\text{out,2}}(l_1) = F_{\text{gam}}\left(\rho_{\text{res}}; \hat{K}(U) - \hat{K}(U - l_1), \theta \right)$$

Similarly, we then consider the number of additional messages that can be decoded in the second iteration $l_2 = 0, 1, \ldots U - 1 - S - l_1$ which is given by $f_{\text{bin}}(l_2; U - 1 - S - l_1, 1 - p_{\text{out,2}}(l_1))$. We halt this procedure at the third iteration. At this point the device in focus observes the system with $U - l_1 - l_2$ devices (including itself), however, since there was no attempt to decode its

\[\text{Note that in general the decoding events are not independent. Eq. (11) is a weighted mean of all realizations of } K', \text{ however it is not possible to have a situation with } 2 \text{ active users where } K'_1 \neq K'_2.\]
packet yet, it is subject to $p_{\text{out}}(R, \theta, U - l_1 - l_2)$. By marginalizing over $l_1$ and $l_2$, the outage probability conditioned on $S$ is then

$$p_{\text{out}, \text{SIC}}(R, U | S) = p_{\text{out}}(R, \theta, U)^{U - S} + \sum_{l_1=1}^{U - S} f_{\text{bin}}(l_1; U - 1 - S, 1 - p_{\text{out}}(R, \theta, U))$$

$$\times \sum_{l_2=0}^{U - 1 - S - l_1} f_{\text{bin}}(l_2; U - 1 - S - l_1, 1 - p_{\text{out}, 2}(l_1)) p_{\text{out}}(R, \theta, U - l_1 - l_2)$$

(14)

where the first term corresponds to the case in which all users fail and SIC cannot proceed.

The expression (14) with $S = 0$ approximates very well the simulation results when the number of active devices $U$ is low. This depends on the specific Steiner system, but typically it means no more than 7. As $U$ grows, the approximation and simulations start to diverge in the high SNR regime, with the latter exhibiting plateauing. The reason is due to the existence of stopping sets [29]. It is easy to imagine a situation where the access patterns overlap in such a way that there are no collision-free slots and consequently SIC cannot be applied. Formally, a stopping set $s^{(n)}$ of order $n$ is a subset of $n$ patterns such that in every slot there is either 0, or $\geq 2$ users; that is, the decoding cannot proceed as there is no slot with a single transmission only. Furthermore, let us denote by $T^{(n)}$ the set of all stopping sets of order $n$ for a given Steiner system and by $|T^{(n)}|$ its cardinality. In order to take into account stopping sets and augment the expression (14), we need to consider three cases. If there is a stopping set of certain order $n$, then with probability $n/U$ the user in focus is its member and cannot be decoded. Conversely, with probability $1 - n/U$ the user is not involved in that stopping set and decoding is possible, however SIC is impaired since there are $S = n$ noncancelable users. Otherwise, if there are no stopping sets then $S = 0$ and there are no limitations on SIC.

Ultimately, we have the following approximate expression for the outage probability with SIC:

$$p_{\text{out}, \text{SIC}}(R, U) = \sum_{n \in \mathcal{N}} q_1(n | U) \left( \frac{n}{U} + \frac{U - n}{U} p_{\text{out}, \text{SIC}}(R, U | n) \right) + p_{\text{out}, \text{SIC}}(R, U | 0) \left( 1 - \sum_{n \in \mathcal{N}} q_1(n | U) \right)$$

(15)

where summation is over $\mathcal{N} = \{ n : T^{(n)} \neq \emptyset \}$ and $q_1(n | U) \approx f_{\text{bin}}(1; \binom{n}{U}, |T^{(n)}|/\binom{n}{U})$ is the probability that there is a stopping set of order $n$ among $U$ active users. We note that this is

---

6 While it would be more precise to write $T^{(n)}_{S(t, K, M)}$, for brevity we decide to drop the subscript $S(t, K, M)$ (as we similarly do in the case of quantities $C$ and $D$). In this paper we always consider a single Steiner system at a time so this should not lead to any confusion.
an approximation, because the $\binom{U}{n}$ tuples are not independent. Additionally, we would like to bring to the reader’s attention the fact that we are interested in “exactly one” stopping set of a given order and not “at least one”. The reason is that stopping sets are closed under union, so their combination produces another stopping set of a higher order [29]. As such, by summing over $n$ we would count some of the stopping sets multiple times.

Perhaps the most important, however, is that in (15) it is not necessary to sum over whole $\mathbb{N}$ to obtain a good approximation. In practice, performance is impacted primarily by the stopping sets of the lowest existing order, which we denote by $n’$. They are decisive for two reasons. For a given Steiner system, the outage probability cannot be made arbitrarily low regardless of the SNR whenever $U \geq n’$. Secondly, simulations show that for $i < j$, $q_1(i\,|\,U) > q_1(j\,|\,U)$ and the difference can reach several orders of magnitude, making $q_1(n’\,|\,U)$ dominant overall. This is fortunate, since finding $T^{(n)}$ requires an exhaustive search, which for high $n$ becomes prohibitive. Consequently, when generating results, for each Steiner system we use only the stopping sets of the lowest existing order $n’$, and $n’ + 1$ whenever $T^{(n’+1)}$ is not empty (if $T^{(n’+1)} = \emptyset$, $T^{(n’)}$ is sufficient).

In Fig. 4 we plot the outage probability as given by the proposed approximation (dashed lines) and compare it to the results of the corresponding simulations in which the full procedure is implemented (markers). The derived approximations prove to be very close to the exact results across the whole SNR range and different traffic intensities. The improvement compared to a system without SIC (cf. Fig. 3) is significant and allows to achieve ultra reliability at much lower
SNRs. Particularly important is the fact that Random selection exhibits a performance floor even when the mean traffic intensity is as low as 1 user/frame. This is a consequence of the stopping sets, which in a Random selection can occur already with two users if they select exactly the same pattern. This has been discussed also in [10]. Conversely, in a Steiner system the number of collisions between two patterns is strictly limited, hence there are no stopping sets as long as the number of active users is sufficiently small. It is easy to show that with a maximum of \( t - 1 \) collisions and \( K \) repetitions at least
\[
n' \geq \left\lceil \frac{K}{t-1} \right\rceil + 1
\]
users need to be active for the stopping set to occur. In practice, for some Steiner systems that number is even higher, e.g. in the used \( S(2, 4, 25) \) no subset of size \(< 7\) exist that would form a stopping set. Those issues as well as the results for other systems are further discussed in Section VI.

C. Model with Full Maximum Ratio Combining (MRC)

In the following we will consider a more involved model, in which the receiver is capable of using the totality of all the replicas (including those experiencing interference) and combines them using maximum ratio combining (MRC).

Without loss of generality, let us consider an arbitrary active user \( i \) and one of its transmissions \( j \in \{m : V_{m,i} = 1\} \). The SINR of this signal
\[
\rho_{j,i} = \frac{P_x |g_{j,i}|^2}{\sum_{k \in \{1, \ldots, N\} \setminus \{i\}} V_{j,k} A_{j,k} P_x |g_{j,k}|^2 + \sigma^2}
\]
is a random variable. Assuming there are \( L \) interferers in slot \( j \), i.e. \( \sum_{k \in \{1, \ldots, N\} \setminus \{i\}} V_{j,k} A_{j,k} = L \), we can denote this SINR as \( \frac{X}{Y+1} \), where \( X \) follows the exponential distribution \( f_{\exp}(x; \theta) \) and \( Y \) the gamma distribution \( f_{\text{gam}}(y; L, \theta) \). Hence,
\[
P\left( \frac{X}{Y+1} > z \right) = \int_0^\infty \left( \int_0^\infty \frac{1}{\theta} e^{-\frac{y}{\theta}} dy \right) \frac{1}{\Gamma(L)\theta^L} y^{L-1} e^{-\frac{y}{\theta}} dy
\]
\[
= \frac{1}{\Gamma(L)\theta^L} \int_0^\infty y^{L-1} e^{-\frac{y}{\theta}} dy \left|_{(y+1)z} \right.
= \frac{1}{\Gamma(L)\theta^L} \int_0^\infty y^{L-1} e^{-\frac{y}{\theta}} \left|_{(y+1)z} \right. \int_0^\infty y^{L-1} e^{-\frac{y}{\theta}} dy
= \frac{e^{-\frac{z}{\theta}}}{\Gamma(L)\theta^L} \cdot \frac{\Gamma(L)\theta^L}{(z+1)^L} = \frac{e^{-\frac{z}{\theta}}}{(z+1)^L}
\]
where the last integral can be computed by integrating it by parts \( L - 1 \) times. Finally, by taking the derivative of \( 1 - P \left( \frac{X}{Y+1} > z \right) \) we obtain the pdf of the SINR given \( L \) interferers:

\[
\begin{align*}
    f_{SINR}(z; \theta | L) &= \frac{e^{-\frac{\hat{z}}{\theta}}(\theta L + z + 1)}{\theta(z + 1)^{L+1}} \\
    \text{and it can be seen that for a special case of } L = 0 \text{ the expression reduces to a simple exponential distribution.}
\end{align*}
\]

It would be tempting to consider the final SINR as a sum of \( K \) independent RV’s where SINR from a single slot is a mixture of \( f_{SINR}(z; \theta | L) \) for \( L = 0, \ldots, U - 1 \). However, such a naive approach does not provide a good approximation (especially when applied to the patterns from Steiner system) as it significantly underestimates the contribution of the interference-free slots, which contribute the most to the combined power. Instead, let us first condition on the number of interference-free slots \( K' \) (given by (4), (6)). Notice, however, that fixing \( K' \) has implications for the distribution of interferers in the remaining slots. Namely, by fixing \( K' \) we implicitly reduce the number of available patterns (in case of Steiner system) or slots (in case of random selection). Hence, we introduce the modified version of the expressions (5), (7):

\[
\begin{align*}
    p_{I,R}(L|K', U) &= \binom{M-1-K'}{K} \binom{U-1-L}{M-1-K'} \binom{L}{K-1} \binom{U-1}{L} \\
    p_{I,S}(L|K', U) &= \binom{D-1}{L} \binom{C-1-(D-1)(K'+1)}{U-1-L} \binom{K'}{U-1}. \quad (20)
\end{align*}
\]

Secondly, if the slot contains interference, then by definition \( L > 0 \) so the case \( L = 0 \) has to be excluded and the distribution re-normalized. Taking all this into account, the distribution of the SINR in the interfered slot becomes

\[
\begin{align*}
    f_{I-SINR}(z; \theta | U, K') &= \sum_{L=1}^{U-1} f_{SINR}(z; \theta | L) \frac{p_{I}(L|K', U)}{1 - p_{I}(0|K', U)}. \quad (22)
\end{align*}
\]

In order to obtain the distribution of the total SINR, we combine the above with the contribution from \( K' \) interference-free slots and marginalize:

\[
\begin{align*}
    f_{tot,SINR}(z; \theta | U) &= \sum_{K'=0}^{K} p_{CF}(K'|U) p_{gam}(z; K', \theta) * f_{I-SINR}(z; \theta | U, K')^{* (K-K')} \quad (23)
\end{align*}
\]

7To see why, consider two users employing patterns from the Steiner system. For a given slot, there is a certain probability of collision, which means that by treating them independently we include cases with \( 1, \ldots, t - 1, \ldots, K \) collisions. Meanwhile, Steiner system guarantees that two users will share at most \( t - 1 \) slots.
where \( f^{n} \stackrel{\text{def}}{=} f * \cdots * f \) and \( f^{0} = \delta \), with \( \delta \) denoting the Dirac delta distribution. Similarly, we set \( f_{\text{gam}}(z; 0, \theta) = \delta \) to keep the above expression (23) compact. Finally, we have that

\[
p_{\text{out,cap}}(R, \theta, U) = \sum_{K' = 0}^{K} p_{\text{CF}}(K'|U) F_{\text{gam}}(z; K', \theta) * f_{\text{I-SINR}}(z; \theta|U, K') \bigg| z = 2^{K} - 1 \tag{24}
\]

While not excessively complex, evaluation of (24) requires \((K - 1)(K + 2)/2\) numerical integrations. This is not an issue for the values of \( K \) considered in this work, however to address this we provide in Appendix A an even simpler approximation that avoids the convolutions altogether.

In Fig. 5(a) we present joint comparison of the simulation results and developed approximations. Once again, we consider the performance of the Steiner system and Random selection across the range of received SNRs and traffic intensities. The results from simulations, which implement the exact procedure, are given with markers. Solid and dashed lines (Steiner and Random respectively) correspond to the approximation based on eq. (24) (Approx. 1). Similarly, dotted and dash-dotted lines (Approx. 2) correspond to the simpler approximation by gamma distribution discussed in the Appendix A. Approx. 1 follows very closely the actual simulation results, however we note that in case of Steiner system the deviation is slightly larger than for Random selection. Albeit being simpler, the accuracy of Approx. 2 is not far off, especially for lower traffic intensities \( bN \), which are of primary interest when ultra-reliability is considered.

For completeness, in Fig. 5(b) we show the simulations results for the Full MRC model that leverages SIC. The improvement over non-SIC (cf. Fig. 5(a)) processing is significant as the performance does not exhibit plateauing and is capable of achieving ultra-reliability even for traffic intensities as high as \( bN = 10 \). The superiority of Steiner system over Random selection, especially for high SNRs, is diminished, however this aspect is further discussed in Section V.

V. SYSTEM DESIGN FAVORS STEINER SEQUENCES OVER RANDOM SELECTION

Until now, the implicit assumption was that perfect channel estimates have been available. In reality channel estimates are never perfect, however, the quality of estimation can be improved by making the pilot sequences longer and/or investing in them more transmit power, as long as the pilot sequences of transmitting users are kept orthogonal. Collisions in the pilot domain are particularly problematic as they lead to pilot contamination and, consequently, very poor channel estimates. Typically, that means proper equalization is not possible and the packet
Fig. 5. Outage probability in the Full MRC model (a) without SIC and (b) with SIC.

replicas involved in the collision are unusable i.e. cannot be combined with others through MRC or removed with SIC. This is a significant challenge for random access schemes that rely on fully random selection of access patterns. Since any user can be active in any slot, the only way to avoid pilot collisions, would be to assign a unique orthogonal sequence to each of the $N$ devices. In many cases, however, this is not feasible or practical (e.g. with large population of intermittently active devices similar to the scenario addressed in this work). Instead, a common approach is to provide a pool of $Q < N$ pilot sequences from which users pick one at random every time they become active and accept, that some collisions will inevitably occur.

In that case, it is possible to provide a reasonable lower bound for the model utilizing SIC as $\theta \to \infty$. Let us again focus on an arbitrary user $u$ and one of its slots. We can distinguish two types of events. In the first case, whenever one or more interferers select the same pilot sequence as the user of interest $u$, the packet replica is lost. This is given by

$$p_{(1)} = \sum_{L=1}^{U-1} p_{t}(L|U) \left( 1 - \left( 1 - \frac{1}{Q} \right)^L \right).$$

(25)

The second type of event, is when user $u$’s pilots are intact, however there are some pilot

\[ p_{out} \]
collisions among the interferers. The probability that the slot is of this second kind is

\[ p_{II} = \sum_{L=2}^{U-1} p_I(L|U) \left( 1 - \prod_{i=0}^{L-1} \frac{Q - i}{Q^L} \right) \left( 1 - \left( 1 - \frac{1}{Q} \right)^L \right). \]  

(26)

In that case, even if the packets of interfering users can be decoded based on their other replicas, due to the lack of channel knowledge SIC cannot be applied, so the SINR is limited to at most

\[ \frac{X}{Y_1 + \cdots + Y_{L'} + 1}, \]

where \( L' \) is the number of mutually colliding (in the pilot domain) interferers. For the lower bound, we can fix \( L' = 2 \) and as \( \theta \to \infty \) the ’1’ in the denominator can be dropped.

Along with the fact that \( X, Y_1, Y_2 \) are exponentially distributed with the same scale parameter \( \theta \), the SINR of user \( u \)'s replica follows a beta prime distribution

\[ f_{BP}(x; \alpha, \beta) = \frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{B(\alpha,\beta)} \]

with \( \alpha = 1 \) and \( \beta = 2 \). Considering there can be \( n = 0, \ldots, K \) such replicas (with the remaining ones being of the first type), the outage probability can be bounded by

\[ p_{\text{bound}}(R, U) = p_{II}^K + \sum_{n=1}^{K} \binom{K}{n} p_{II}^{K-n} p_{II}^n F_{BP}(2R - 1; n, 2) \]  

(27)

where \( F_{BP}(\cdot; \cdot, \cdot) \) is the CDF of beta prime distribution and we leverage the fact that sum of its \( n \) i.i.d variables is also beta-prime distributed, in this case, with \( \alpha_n = n, \beta_n = 2 \).

Unlike Random selection, the Steiner system guarantees that at most \( D \) users can be active in any given slot. Furthermore, since each user has to be assigned a specific access pattern for their packet replicas, it can be simultaneously instructed which pilot sequence to use in which slot, thus eliminating any possibility of collisions with just \( Q = D \) orthogonal pilot sequences.

Recalling that in a Steiner system the number supportable users is \( N = C = \binom{M}{t} / \binom{K}{t} \), while \( D = \binom{M-1}{t-1} / \binom{K-1}{t-1} \), we have that \( D = N \frac{K}{M} \), i.e. the number of pilot sequences required to ensure no collisions is reduced by a factor \( K/M \) compared to the Random selection.

Another caveat is that, clearly, the receiver must know where each replica of each user is located in order to perform combining through MRC. Because with the Steiner system there is an association between pilot sequences and user IDs, observing a certain pilot sequence in a given slot automatically indicates which user is active and where to look for its remaining replicas. This is not the case when Random selection scheme is used so additional procedures might be required. One possibility is to look for the correlation between signals in different slots and combine those with the highest correlation score. However, such a solution is not perfect as it might miss some of the replicas or introduce false positives. Furthermore, it entails exhaustive search and, hence, high complexity. Alternatively, a unique ID that can be decoded independently
Fig. 6. Impact of pilot collisions and realistic MRC processing on Random selection scheme. Although not shown here for the sake of readability, the performance of Steiner system matches, and for low $\theta$ even exceeds, the perfect Random scheme.

of the rest of the payload could be added to each packet or, each slot could be preceded by an activity-indication phase. Clearly, the downside of this solution is the introduction of overhead.

Lastly, we remark that MRC processing is impeded when using Random selection. Note, that the SINR of the combined packet is equal to the sum of the SINRs of its constituents only when the interference in each is uncorrelated. This will not be the case if a given user collides with another in more than one slot. Whether the resulting SINR will be higher or lower than the sum will depend on whether the interference adds destructively or constructively. Interestingly, even though both are equally likely, the performance of the SIC is ultimately impaired. To circumvent that, a more computationally-heavy equalization method such as zero-forcing (ZF)

$^9$Consider the case with two users colliding in more than one slot and let us establish a baseline where the combined SINR is equal to the sum of SINRs in individual slots. As already mentioned, in reality, the SINR after MRC will be lower (we can call it negative MRC) or higher (positive MRC) than that baseline. Importantly, if the combining is negative(positive), it is negative(positive) for both users. Consider now 4 possible decoding outcomes than can happen in the baseline scenario. If both users are successful, then positive MRC would have no effect. Similarly, if only one of them succeeds but the other fails, positive MRC also wouldn’t change anything since the successful user can be cancelled through SIC anyway. Only when both users fail the positive MRC can make a difference - that is if it makes at least one of them decodable and triggers SIC. Now let us consider negative MRC. When both users fail it has no effect. If there is one successful user, it can happen that negative MRC turns it into an undecodable one, thus making SIC impossible. Similarly, if both users are successful, negative MRC could make them both undecodable (in this case it is not enough to turn just one of them, as the SIC could still be applied). In the end, even though negative and positive MRC are equally probable, if the system uses SIC, the negative MRC has the potential to be detrimental in three out of four cases, while the latter can only help in one of them.
or minimum mean squared error (MMSE) would have to be used. Conversely, in a Steiner system with \( t = 2 \) each user is guaranteed to collide with another at most once so the interference is always uncorrelated.

In Fig. 6 we demonstrate the impact of the above described issues. The markers represent the performance of the idealized version of the Random selection scheme and serve as a reference. The dotted lines show the actual performance of MRC in the presence of correlated interference. The dashed lines depict the scenario with finite pool of pilot sequences \( Q = 24 \) and the solid horizontal line is the corresponding lower bound as given by (27). Lastly, dash-dotted curves take into account both detrimental factors. The Fig. 6 reveals that, in a more realistic setting, the performance of the Random selection would be significantly impaired and cannot match that of Steiner system (cf. Fig. 5(b)).

VI. PERFORMANCE EVALUATION: CHOICE OF FRAME PARAMETERS \( M \) AND \( K \)

Lastly, we consider Steiner systems with different configurations of the frame length \( M \) and number of repetitions \( K \). In Table I we provide the relevant parameters for the systems used in this work, namely the number of patterns \( C \), maximum number of interferers per slot \( D \), order of the smallest existing stopping sets as well as their number. The patterns themselves can be found in [30]. The objective is to determine the highest supported rate \( R \) for a given traffic
intensity $bN$ and fixed mean SNR $\theta$, that fulfills certain target reliability $\epsilon_{tar}$:

$$\max \quad R$$

$$\text{s.t.} \quad \sum_{u=0}^{N} f_{\text{bin}}(u, b, N) p_{\text{out}}(R, \theta, u) \leq \epsilon_{tar} \quad (28)$$

Complementary to rate which is per UE, we define also the spectral efficiency of the system given by $bN \cdot R/M$. In the figures, the results are plotted as a function of the absolute mean traffic intensity $bN$. We note that since $N$ is different for each Steiner System, so is the activation probability $b$. Furthermore, in order to jointly compare Steiner Systems with different number of repetitions $K$, we decide to normalize their mean received SNRs. The rationale is that, with $\theta$ being the same in each case, the systems with higher $K$ would use proportionally more energy and thus have an advantage. Consequently, we perform our evaluations by fixing $\theta K = 25$ dB. Lastly, as we are considering ultra-reliability we set $\epsilon_{tar} = 10^{-5}$.

We focus on two cases a) the Full MRC model without SIC and b) the collision model with SIC. The results are obtained based on the derived approximations (24) and (15) respectively which are applied to (28). To improve the readability of the figures, we do not show the results for Random selection schemes in this section, noting that they are always strictly worse than the corresponding Steiner system (cf. earlier discussions and Fig. 3-5).

We start with the rate $R$ of the Full MRC model shown in Fig. 7(a). As expected, for a given
mean number of active users, the larger the frame and the number of repetitions, the higher the rate. Increasing the frame size decreases the chance of collisions, while increasing $K$ makes the transmission more robust and allows to harvest more diversity. The relationship, however, is not as straightforward when it comes to spectral efficiency shown in Fig. 7(b). For a given number of repetitions $K$, increasing the frame length actually reduces the spectral efficiency, as the increase in rate is not enough to offset the extra resources (cf. $S(2, 5, 25)$ vs $S(2, 5, 41)$ or $S(2, 4, 25)$ vs $S(2, 4, 28)$ vs $S(2, 4, 37)$). Furthermore, even though higher $K$ itself is generally beneficial, the system with high $M$ and $K$ might be less spectrally efficient than the one with lower parameters when $bN$ is low (cf. $S(2, 5, 41)$ vs $S(2, 4, 25)$ or $S(2, 4, 28)$).

To provide further insight, in Fig. 7(b) we also mark the traffic intensity $bN$ beyond which orthogonal resource allocation becomes more spectrally efficient than Steiner system. To find this value, first, we note that when resources are orthogonal, the maximum rate does not depend on the traffic intensity and is given by $R_{orth} = \log_2(F^{-1}_{\text{gam}}(\epsilon_{\text{tar}}; K, \theta) + 1)$, where $F^{-1}$ is the inverse CDF (quantile function). Since in a Steiner system the maximum number of users is $N = C$, the equivalent orthogonal allocation requires a frame of length $M_{orth} = CK$. For the sake of readability, we plot the spectral efficiency curve of the orthogonal system only for one representative case for each $K = 3, 4, 5$ (dashed curve) and for the rest simply mark the point at which $bN \cdot R/M$ intersects with $bN \cdot R_{orth}/M_{orth}$.

In Fig. 8 the results corresponding to the collision model with SIC are shown. There are
several notable differences. First, the maximum rate does not exhibit such a smooth degradation as in the Full MRC case. Instead, it stays very high and close to its absolute maximum (i.e. \(\log_2(F_{\text{gam}}^{-1}(\epsilon_{\text{tar}}; K, \theta) + 1)\) as in the orthogonal allocation) and then goes abruptly to 0 once it reaches certain cut-off traffic intensity. The maximum supportable \(bN\) of a Steiner system is a non-trivial function of the order and number of its stopping sets, frame length \(M\), and number of patterns \(C\). While the general trend is preserved, i.e. higher \(K\) and \(M\) lead to higher rates, there are some exceptions. Comparing \(S(2, 3, 25)\), \(S(2, 3, 33)\) and \(S(2, 3, 39)\), one can see that the case with \(M = 33\) actually performs the worst. This is tied to the particularly high number of stopping sets (cf. Table I). On the other hand, \(S(2, 4, 37)\) can sustain higher traffic intensity than \(S(2, 5, 41)\) despite shorter frame length. In this case, the reason lies in the lower \(K\) and, consequently, higher number of patterns (111 vs 82). Even though the order and number of stopping sets is similar, the probability of their occurrence is effectively lower in \(S(2, 4, 37)\).

VII. Conclusions

In this work we have proposed and investigated the usage of deterministic access patterns to provide ultra-reliable communication for a group of intermittently active users sharing a pool of resources. The patterns, which are a realization of the Steiner system, aim to control the number of collisions and interference among users. This feature leads to significant gains in terms of outage probability compared to an approach were the choice of channel resources is fully random. In our evaluations we have considered two different signal models - based on destructive collisions and Full MRC, and two receiver processing techniques - with and without SIC. As the second main contribution of this work, we have developed simple approximations for the outage probability in a collision model with SIC and Full MRC model without SIC that closely match the simulation results. Such approximations are particularly important in the context of ultra-reliable systems where the number of required samples/simulations needed to properly assess the performance is often infeasible.

Appendix A

A relatively good approximation for the expression (23) can be obtained by first approximating (22) with a gamma distribution, i.e. finding \(f_{\text{gam}}(z; k_{K'}, \alpha_{K'}) \approx f_{I-SINR}(z; \theta|U, K')\). This can be done e.g. by solving the following optimization problem to find the suitable coefficients:
arg min \( k_{K'}, \alpha_{K'} \) \( \int_A \left[ f_{I-SINR}(z; \theta|U, K') - f_{\text{gam}}(z; k_{K'}, \alpha_{K'}) \right]^2 dz \)

\[ \text{s.t.} \quad k_{K'} > 0, \quad \alpha_{K'} > 0 \]

where \( A = [0, 2^R - 1] \), since we are interested in the outage probability and hence concerned with a good fit only in that region. In Fig. 9 we show as an example the results of such fitting for the Random selection in the SINR range \([0, 2^2 - 1]\). Goodness of fit tends to be the lowest for a medium number of interferers which is also reflected in the final approximation in Fig. 5(a) as it becomes less tight the higher the traffic intensity \( bN \).

![Fig. 9. Results of least-squares approximation of (22) with gamma distribution.](image)

Since the SINRs of individual interfered slots are i.i.d and for a given \( K' \) each is now represented by a gamma distribution with parameters \( k_{K'}, \alpha_{K'} \), we have simply that

\[ f_{\text{gam}}(z; k_{K'}, \alpha_{K'})^*(K-K') = f_{\text{gam}}(z; (K-K')k_{K'}, \alpha_{K'}) \]

\[ \text{(30)} \]

Lastly, to combine (30) with the contribution from the interference-free slots \( f_{\text{gam}}(z; K', \theta) \) we can use the result in [31] which provides the analytical expression for the sum of two gamma distributed RVs with arbitrary parameters:

\[ \tilde{f}_{\text{tot}, SINR}(z; \theta|U) = \sum_{K'=0}^{K} p(K'|U) f_{\text{gam}}(z; K', \theta) * f_{\text{gam}}(z; (K-K')k_{K'}, \alpha_{K'}) \]

\[ = \sum_{K'=0}^{K} p(K'|U) \left( \frac{\alpha_{K'}}{\theta} \right)^{K'} \frac{1}{\Gamma(K')} e^{-\frac{z}{\alpha_{K'}}} \left( K'; \frac{z}{\alpha_{K'}} \right) \]

\[ \Gamma(K) F_1 \left( K'; \kappa; \left( \frac{z}{\alpha_{K'}} - \frac{z}{\theta} \right) \right) \]

\[ \text{(31)} \]
where $\kappa = K' + (K - K')k_{K'}$ and $\mathbf{1}_{1} F_1 (\cdot; \cdot; \cdot)$ is a Kummer's confluent hypergeometric function. The CDF in this case becomes

$$
\tilde{F}_{\text{tot}, \text{SINR}} (z; \theta | U) = \sum_{K' = 0}^{K} p(K'|U) F_{\text{gam}} (z; K', \theta) \ast f_{\text{gam}} (z; (K - K')k_{K'}, \alpha_{K'})
$$

$$
= \sum_{K' = 0}^{K} p(K'|U) \left( \frac{\alpha_{K'}}{\theta} \right)^{K'} \left[ \frac{1}{\alpha_{K'}} \right]_{2} \mathbf{1}_{2} F_1 \left( \kappa, K' \kappa; \left( 1 - \frac{\alpha_{K'}}{\theta} \right) \right)
$$

(32)

where $[\cdot]_{2} F_1 (\cdot; \cdot; \cdot)$ is the incomplete Gauss hypergeometric function.

**Appendix B**

**Proof of Lemma**

*Proof.* Let us define a probability space where the samples are different ways the $U - 1$ other users can select the locations of their packets inside the frame. We will call these samples configurations. An event is then a set of configurations and has a probability.

Let us now fix $K'$ of the $K$ slots of a given user $u$. Denote with $A_i$ an event that consists of all the configurations where the fixed set $K'$, and at least one of the remaining $K - K'$ slots, denoted by $i$, are free of interference. This gives us $K - K'$ different sets $A_i$.

Given now an $n$-element subset $J$ of $\{1, \ldots, K - K'\}$, then the probability for an event $\bigcap_{j \in J} A_j$ to appear is $P(\bigcap_{j \in J} A_j) = \left( \frac{M - K' - n}{M} / \left( \frac{M}{K} \right) \right)^{U - 1} = V_n$, independent of the selected $J$ and $K'$. This can be directly seen as $\bigcap_{j \in J} A_j$ consists of all the configurations that leave at least fixed $n + K'$ slots of user $u$ interference free.

Let us denote with $S$ the set consisting of all the configurations where at least the fixed $K'$ slots are interference free. By the complementary form of the inclusion-exclusion principle we then have that

$$
P \left( S \setminus \bigcup_{i=1}^{K-K'} A_i \right) = \sum_{n=0}^{K-K'} (-1)^n \binom{K - K'}{n} V_n,
$$

(33)

where $V_0 = P(S) = \left( \frac{M - K'}{M} / \left( \frac{M}{K} \right) \right)^{U - 1}$. Here the set $\bigcup_{i=1}^{K-K'} A_i$ is an event that contains all the configurations that leave the fixed $K'$ slots and at least one other of the slots occupied by the user $u$ free of interference. Then the set $S \setminus \bigcup_{i=1}^{K-K'} A_i$ is an event that consists of all the configurations, where the fixed $K'$ slots are free, but none of the other $K - K'$ occupied by the user $u$. In other words, it is the set of those configurations, where exactly $K'$ of $K$ slots are free. One can place $K'$ packets to $K$ available slots in $\binom{K}{K'}$ different ways. Hence the final result is gotten by multiplying (33) with the term $\binom{K}{K'}$.  

$\square$
Appendix C

Proof of Lemma \[2\]

Proof. The probability that a specific set of $K'$ slots selected by the user $u$ is not occupied by the packets of remaining $U - 1$ users is

$$\binom{C-1-K'(D-1)}{U-1} \binom{C-1}{U-1},$$

because there are $C$ patterns in total (including pattern of user $u$) and $K'(D-1)$ of them share a slot with the $K'$ slot set of pattern of user $u$. The proof follows that of Lemma \[1\] verbatim, after realizing that now

$$P(\bigcap_{j \in J} A_j) = \binom{C-1-(D-1)(n+K')}{U-1} \binom{C-1}{U-1},$$

for any $n$ elements set of $J$. \qed

References

[1] “Ericsson Mobility report,” Tech. Rep., Nov. 2021. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/november-2021

[2] W. Saad, M. Bennis, and M. Chen, “A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems,” IEEE Network, vol. 34, no. 3, pp. 134–142, 2020.

[3] 3GPP, “Service requirements for the 5G system; Stage 1,” 3rd Generation Partnership Project (3GPP), TS 22.261, Dec. 2021, v18.5.0.

[4] “Verticals URLLC Use Cases and Requirements,” NGMN Alliance, Tech. Rep., July. 2019. [Online]. Available: https://www.ngmn.org/publications/verticals-urllc-use-cases-and-requirements.html

[5] S. Schiessl, “Performance Trade-offs for Ultra-Reliable Low-Latency Communication Systems,” Ph.D. dissertation, KTH Royal Institute of Technology, 2019. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-251650

[6] G. Berardinelli, N. Huda Mahmood, R. Abreu, T. Jacobsen, K. Pedersen, I. Z. Kovács, and P. Mogensen, “Reliability Analysis of Uplink Grant-Free Transmission Over Shared Resources,” IEEE Access, vol. 6, pp. 23 602–23 611, 2018.

[7] N. Patriciello, S. Lagen, L. Giupponi, and B. Bojovic, “The Impact of NR Scheduling Timings on End-to-End Delay for Uplink Traffic,” in 2019 IEEE Global Communications Conference (GLOBECOM), 2019, pp. 1–6.

[8] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, “Grant-Free Non-Orthogonal Multiple Access for IoT: A Survey,” IEEE Communications Surveys Tutorials, vol. 22, no. 3, pp. 1805–1838, 2020.

[9] G. Choudhury and S. Rappaport, “Diversity ALOHA - A Random Access Scheme for Satellite Communications,” IEEE Transactions on Communications, vol. 31, no. 3, pp. 450–457, 1983.

[10] C. Boyd, R. Vehkalahti, and O. Tirkkonen, “Interference cancelling codes for ultra-reliable random access,” International Journal of Wireless Information Networks, vol. 25, pp. 1–12, 12 2018.

[11] E. Casini, R. De Gaudenzi, and O. Del Rio Herrero, “Contention Resolution Diversity Slotted ALOHA (CRDSA): An Enhanced Random Access Scheme for Satellite Access Packet Networks,” IEEE Transactions on Wireless Communications, vol. 6, no. 4, pp. 1408–1419, 2007.
[12] L. G. Roberts, “ALOHA Packet System with and without Slots and Capture,” vol. 5, no. 2, 1975. [Online]. Available: https://doi.org/10.1145/1024916.1024920

[13] G. Liva, “Graph-Based Analysis and Optimization of Contention Resolution Diversity Slotted ALOHA,” IEEE Transactions on Communications, vol. 59, no. 2, pp. 477–487, 2011.

[14] M. Ghanbarinejad and C. Schlegel, “Irregular repetition slotted ALOHA with multiuser detection,” in 2013 10th Annual Conference on Wireless On-demand Network Systems and Services (WONS), 2013, pp. 201–205.

[15] F. Clazzer, E. Paolini, I. Mambelli, and C. Stefanović, “Irregular repetition slotted ALOHA over the Rayleigh block fading channel with capture,” in 2017 IEEE International Conference on Communications (ICC), 2017, pp. 1–6.

[16] E. Paolini, G. Liva, and M. Chiani, “Coded Slotted ALOHA: A Graph-Based Method for Uncoordinated Multiple Access,” IEEE Transactions on Information Theory, vol. 61, no. 12, pp. 6815–6832, 2015.

[17] J. Choi, “Throughput Analysis for Coded Multichannel ALOHA Random Access,” IEEE Communications Letters, vol. 21, no. 8, pp. 1803–1806, 2017.

[18] R. Abreu, “Uplink grant-free access for ultra-reliable low-latency communications in 5g: Radio access and resource management solutions,” Ph.D. dissertation, 2019, PhD supervisor: Prof. Preben Mogensen, Aalborg University Assistant PhD supervisors: Assoc. Prof. Gilberto Berardinelli, Aalborg University Prof. Klaus Pedersen, Aalborg University.

[19] M. C. Lucas-Estañ, J. Gozálvez, and M. Sepulcre, “On the Capacity of 5G NR Grant-Free Scheduling with Shared Radio Resources to Support Ultra-Reliable and Low-Latency Communications,” Sensors (Basel, Switzerland), vol. 19, 2019.

[20] Y. Liu, Y. Deng, M. Elkashlan, A. Nallanathan, and G. K. Karagiannidis, “Analyzing Grant-Free Access for URLLC Service,” IEEE Journal on Selected Areas in Communications, vol. 39, no. 3, pp. 741–755, 2021.

[21] J. Choi and J. Ding, “Network Coding for K-Repetition in Grant-Free Random Access,” IEEE Wireless Communications Letters, vol. 10, no. 11, pp. 2557–2561, 2021.

[22] M. Lucas-Estan and J. Gozálvez, “Sensing-based Grant-Free Scheduling for Ultra Reliable Low Latency and Deterministic Beyond 5G Networks,” IEEE Transactions on Vehicular Technology, pp. 1–1, 2022.

[23] G. T. Peeters, R. Bocklandt, and B. Van Houdt, “Multiple Access Algorithms Without Feedback Using Combinatorial Designs,” IEEE Transactions on Communications, vol. 57, no. 9, pp. 2724–2733, 2009.

[24] C. Boyd, R. Vehkalahti, and O. Tirkkonen, “Combinatorial code designs for ultra-reliable IoT random access,” in 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 2017.

[25] C. Boyd, R. Kotaba, O. Tirkkonen, and P. Popovski, “Non-Orthogonal Contention-Based Access for URLLC Devices with Frequency Diversity,” in 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2019, pp. 1–5.

[26] E. Paolini, G. Liva, and A. Graell i Amat, “A structured irregular repetition slotted aloha scheme with low error floors,” in 2017 IEEE International Conference on Communications (ICC), 2017, pp. 1–6.

[27] R. Kotaba, C. N. Manchón, and P. Popovski, “Enhancing Performance of Uplink URLLC Systems via Shared Diversity Transmissions and Multiple Antenna Processing,” in 2019 53rd Asilomar Conference on Signals, Systems, and Computers, 2019, pp. 1409–1415.

[28] W. Tang, S. Kang, B. Ren, and X. Yue, “Uplink grant-free pattern division multiple access (GF-PDMA) for 5G radio access,” China Communications, vol. 15, no. 4, pp. 153–163, 2018.

[29] C. Di, D. Proietti, I. Telatar, T. Richardson, and R. Urbanke, “Finite-length analysis of low-density parity-check codes on the binary erasure channel,” IEEE Transactions on Information Theory, vol. 48, no. 6, pp. 1570–1579, 2002.

[30] La Jolla Covering Repository. Steiner Systems. [Online]. Available: https://ljcr.dmgordon.org/cover.php.

[31] F. D. Salvo, “A characterization of the distribution of a weighted sum of gamma variables through multiple hypergeometric functions,” Integral Transforms and Special Functions, vol. 19, no. 8, pp. 563–575, 2008.