Non-Orthogonal Multiple Access and Network Slicing: Scalable Coexistence of eMBB and URLLC

Eduardo Noboro Tominaga*, Hirley Alves*, Richard Demo Souza†, João Luiz Rebelatto‡, Matti Latva-aho*
*6G Flagship, Centre for Wireless Communications (CWC), University of Oulu, Finland
†Federal University of Santa Catarina (UFSC), Florianópolis, Brazil
‡Federal University of Technology - Paraná (UTFPR), Curitiba, Brazil
{firstname.lastname}@oulu.fi, richard.demo@ufsc.br, jlrebelatto@utfpr.edu.br

***Abstract*—The 5G systems feature three generic services: enhanced Mobile BroadBand (eMBB), massive Machine-Type Communications (mMTC) and Ultra-Reliable and Low-Latency Communications (URLLC). The diverse requirements of these services in terms of data-rates, number of connected devices, latency and reliability can lead to a sub-optimal use of the 5G network, thus network slicing is proposed as a solution that creates customized slices of the network specifically designed to meet the requirements of each service. Under the network slicing, the radio resources can be shared in orthogonal and non-orthogonal schemes. Motivated by Industrial Internet of Things (IIoT) scenarios where a large number of sensors may require connectivity with stringent requirements of latency and reliability, we propose the use of Non-Orthogonal Multiple Access (NOMA) to improve the number of URLLC devices that are connected in the uplink to the same base station (BS), for both orthogonal and non-orthogonal network slicing with eMBB devices. The multiple URLLC devices transmit simultaneously and across multiple frequency channels. We set the reliability requirements for the two services and evaluate the pairs of achievable sum rates. We show that, even with overlapping transmissions from multiple eMBB and URLLC devices, the use of NOMA techniques allows us to guarantee the reliability requirements for both services.**

***Index Terms*—5G, eMBB, IIoT, Network Slicing, NOMA, URLLC.**

I. Introduction

The Fifth Generation (5G) of wireless communication systems is currently under standardization and deployment around the world, and introduces three generic services: enhanced Mobile Broadband (eMBB), massive Machine-Type-Communications (mMTC) and Ultra-Reliable Low-Latency Communications (URLLC). eMBB aims to provide increased data rates, with peak rates on the order of gigabits per second to moderate rates in the order of megabits per second with high availability. The mMTC service aims at providing connectivity for a large number of cost- and energy-constrained devices (e.g. sensors) that often require low data rates. Finally, the challenging objective of URLLC is to provide ultra-reliable connectivity while operating in short block lengths, a requirement to achieve low latency demanded by time-critical applications [1].

The current 5G New Radio (NR) network is not capable yet to satisfy the very stringent requirements of reliability and latency required by URLLC applications. That is one of the reasons why the research community has already started the development of solutions for beyond-5G and 6G wireless communications systems. One of the predicted use cases for beyond-5G and 6G systems is the critical wireless factory automation that requires communication with ultra-high reliability and ultra-low latency. Moreover, it is foreseen that the number of connected devices will increase substantially for 6G, which also poses very stringent requirements in terms of spectral efficiency [2]. In this context, 6G will scale the traditional URLLC to the massive connectivity dimension, thus leading to a new service class defined massive URLLC (mURLLC), which is the merge of the traditional URLLC and mMTC services from 5G [3].

The diverse and sometimes conflicting requirements of the different 5G services and applications can lead to a sub-optimal use of the mobile network. One efficient solution is the slicing of the network in multiple virtual and isolated logical networks running on a common physical infrastructure in an efficient and economic way, thus allowing slices to be individually customized with respect to, e.g., latency, energy efficiency, mobility, massive connectivity and throughput [4]. The dynamic provisioning of network slices will be one of the features of 6G networks. Instead of having the traditional categorization into eMBB, mMTC and URLLC, some 6G applications will require dynamic service types according to the data traffic and usage patterns of the MTC network [5].

The previous generations of wireless communications systems were mostly based on the utilization of Orthogonal Multiple Access (OMA) schemes to provide connectivity to multiple users. In such schemes, users are allocated with radio resources that are orthogonal in time, frequency or code domain, and ideally no interference exists among them. However, one drawback of OMA schemes is that the maximum number of users is limited by the total amount of available orthogonal resources [6]. To meet the diverse requirements of very high data rates, ultra-reliability, low latency, massive connectivity and spectral efficiency, Non-Orthogonal Multiple Access (NOMA) emerges as a promising technology for beyond-5G and 6G. NOMA allows multiple users to share time and frequency resources in the same spatial layer via power domain or code domain multiplexing [6]. It is also predicted that in 6G scenarios there will be a need to new access schemes that can dynamically change between orthogonal and non-orthogonal multiple access schemes depending on the current

state of the network [3].

A communication-theoretic framework for network slicing in 5G was presented in [7], where the same radio resources are sliced among the heterogeneous 5G services under both orthogonal and non-orthogonal strategies. However, they did not considered the use of NOMA for multiple URLLC devices, such that the maximum number of URLLC devices connected to the same Base-Station (BS) was limited by the number of minislots within the timeslot.

The coexistence of eMBB and URLLC services has also been studied in the other works. Joint scheduling of eMBB and URLLC traffic has been studied in, for example, [8], [9] and [10]. Abreu *et. al.* [11] studied the multiplexing of eMBB and URLLC traffics in the uplink using an analytical framework. They considered the cases where different bands are allocated for each service, and also the case where both services share the same band. The slicing of resources for eMBB and URLLC has been also studied in [12], where the authors proposed a risk-sensitive based formulation to allocate resources to URLLC devices while minimizing the risk of eMBB (i.e. protecting the eMBB devices with low data rate) and ensuring URLLC reliability. In [13], the authors adopted a time/frequency resource blocks approach to address the sum rate maximization problem subject to latency and slicing isolation constraints while guaranteeing the reliability requirements with the use of adaptive modulation coding. In [14], the authors analyze the coexistence of eMBB and URLLC in fog-radio architectures where the URLLC traffic is processed at the edge while eMBB traffic is handled at the cloud. In [15], the authors also study the orthogonal and non-orthogonal slicing of radio resources for eMBB and URLLC using a max-matching diversity (MMD) algorithm to allocate the frequency channels for the eMBB devices. However, none of the mentioned works studied the performance of the joint combination of NOMA, SIC decoding and frequency diversity for URLLC traffic.

Motivated by the mURLLC scenarios predicted for beyond-5G and 6G networks, and based on recent works that address the coexistence between eMBB and URLLC (in special [7]), we propose a framework that allows multiple URLLC devices to share the same radio resources with eMBB devices in a scalable manner, for both orthogonal and non-orthogonal slicing of radio resources in the uplink. The main difference of our work in comparison with [7] is that we consider the use of NOMA for multiple URLLC devices. Our contribution consists on the joint use of NOMA, Successive Interference Cancellation (SIC) and frequency diversity as a solution to improve the number of URLLC devices that can be connected to the same BS. In other words, multiple URLLC devices share the same time/frequency resource and, in order to decode the multiple URLLC signals (and also the eMBB signals in the case of non-orthogonal slicing), the BS performs SIC decoding. To characterize the performance trade-offs between eMBB and URLLC, we evaluate the pairs of achievable sum rates under predefined reliability requirements for orthogonal and non-orthogonal slicing scenarios. We show that, even with overlapping transmissions from multiple URLLC devices, the use of NOMA, SIC and frequency diversity techniques allow us to guarantee the reliability requirements of eMBB and URLLC services in both slicing schemes.

This paper is organized as follows. In the next section, we present the system model and the performance analysis of the eMBB and URLLC when they are considered in isolation. In Section III, we show how eMBB and URLLC devices can share the same radio resources for both orthogonal and non-orthogonal slicing. In Section IV, we present the numerical results illustrating the performance trade-off between the services. Finally, the conclusions are presented in Section V.

II. System Model

We consider an uplink scenario where multiple eMBB and URLLC devices transmit independent packets to a common BS, as illustrated in Fig. 1. As in [7], we also consider a time-frequency grid composed of F frequency channels indexed by $f \in \{1, \ldots, F\}$ and S minislots indexed by $s \in \{1, \ldots, S\}$. The set of S minislots composes a timeslot.

The orthogonal and non-orthogonal slicing of the radio resources for eMBB and URLLC are also based on [7] and illustrated in Fig. 2, for $F = 10$ frequency channels, from which $F_U = 5$ frequency channels allocated for the URLLC traffic, and $S = 6$ minislots. By employing the orthogonal slicing, some frequency channels are allocated exclusively for the eMBB traffic and some exclusively for the URLLC traffic. On the other hand, with non-orthogonal slicing the same frequency channels can be shared between the two services. However, differently from [7], we consider NOMA for URLLC, as indicated by the darker blue tone in Fig. 2. In other words, we allow multiple URLLC devices to transmit simultaneously in the same minislot.

The transmission of an eMBB device occupies a single frequency channel f and extends over the entire timeslot. Moreover, for eMBB traffic, we model only the standard scheduled transmission phase, hence assuming that radio access and competition among eMBB devices have been solved prior to the considered time slot. A URLLC device, in turn,

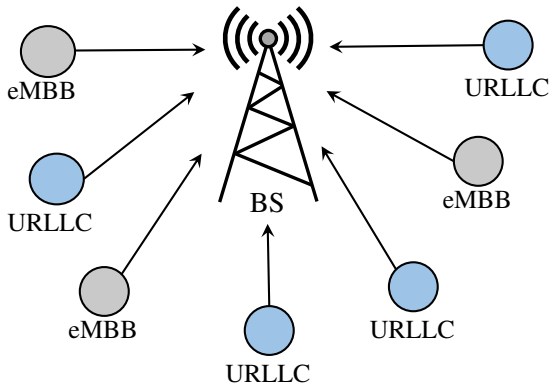

Fig. 1: Uplink transmissions to a common base station (BS) from multiple eMBB and URLLC devices.

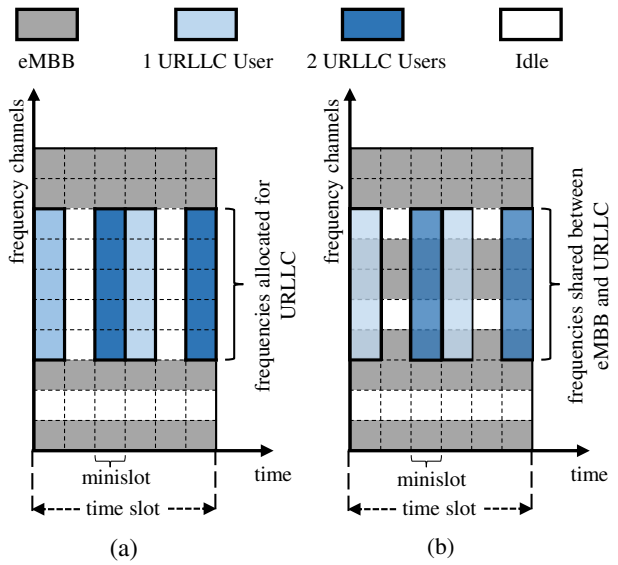

Fig. 2: Illustration of the time-frequency grid used for the network slicing for the eMBB and URLLC services in the (a) orthogonal and (b) non-orthogonal scenarios. The darker blue tone indicates the overlap of URLLC transmissions.

transmits within a single minislot across a subset of $F_U \leq F$ frequency channels, as a mean to achieve frequency diversity and meet the reliability requirements [7]. Due to the low latency requirement, each URLLC packet must be decoded within the duration of a minislot, not being allowed to span over multiple minislots.

Each radio resource is assumed to be within the time- and frequency-coherence interval of the wireless channel, so that the wireless channel coefficients are constant within each minislot. We also assume that the coefficients fade independently across the radio resources. The channel envelops as seen by the eMBB and the URLLC traffics, which we denote by $h_{i,f}$ with $i \in \{B, U\}$, are independent and complex Gaussian distributed, i.e., $h_{i,f} \sim \mathcal{CN}(0, \Gamma_i)$, thus the channel gains $G_{i,f} = |h_{i,f}|^2$ are exponentially distributed with average Γ_i. The noise power at the BS is normalized to one, such that the average channel gain Γ_i can be also interpreted as the average received Signal-to-Noise Ratio (SNR). Moreover, no Channel-State Information (CSI) is assumed at the URLLC devices, whereas the eMBB devices and BS are assumed to have perfect CSI as in [7].

The outage probabilities of the eMBB and URLLC devices are denoted as $\Pr(E_B)$ and $\Pr(E_U)$, respectively, and must satisfy the reliability requirements $\Pr(E_B) \leq \epsilon_B$ and $\Pr(E_U) \leq \epsilon_U$.

A. Signal Model

Let $n_U \in \{\mathbb{N}^+\}$ denote the maximum number of URLLC devices transmitting simultaneously in the same minislot. The baseband signal received at the BS in minislot s and frequency channel f is

$$y_f[s] = h_{B,f} x_{B,f}[s] + \sum_{u=1}^{n_U} h_{U_u,f} x_{U_u,f}[s] + w_f[s], \quad (1)$$

where $x_{B,f}[s] \in \mathbb{C}$ is the symbol transmitted by an eMBB device scheduled in the frequency channel f, $x_{U_u,f}[s] \in \mathbb{C}$ is the symbol transmitted by the u-th URLLC device in the frequency channel f, and $w_f[s]$ corresponds to an Additive White Gaussian Noise (AWGN) sample with zero mean and unit variance, i.e. $w_f[s] \sim \mathcal{CN}(0, 1)$.

To compute the pairs of maximum achievable sum rates, we evaluate the performance of a worst case scenario where there is always an eMBB device transmitting in each frequency channel f and the n_U URLLC devices transmitting in all minislots.

B. Performance Analysis of eMBB

In this subsection, we present the performance analysis of eMBB that was originally presented in [7]. The eMBB device aims at transmitting at the highest rate r_B that is compatible with the outage probability requirement ϵ_B under a long-term average power constraint[1]. This can be formulated as the optimization problem

$$\begin{aligned} &\text{maximize } r_B, \\ &\text{subject to } \Pr\left\{\log_2(1 + G_{B,f} P_B(G_{B,f})) \leq r_B\right\} \leq \epsilon_B \\ &\qquad \text{and } \mathbb{E}\left\{P_B(G_{B,f})\right\} = 1, \end{aligned} \quad (2)$$

where $P_B(G_{B,f})$ is the instantaneous transmit power that is a function of the instantaneous channel gain $G_{B,f}$. The optimal solution to this problem is given by the truncated power inversion scheme. The eMBB device chooses a transmission power that is inversely proportional to $G_{B,f}$ if the latter is above a given threshold $G_{B,f}^{\min}$, while it refrains from transmitting otherwise.

In the absence of interference from other services, the only source of outage for an eMBB transmission is the event that an eMBB device does not transmit because of insufficient SNR. The probability that $G_{B,f}$ is below $G_{B,f}^{\min}$ is

$$\Pr(E_B) = \Pr\left\{G_{B,f} < G_{B,f}^{\min}\right\} = 1 - \exp\left(-\frac{G_{B,f}^{\min}}{\Gamma_B}\right). \quad (3)$$

Imposing the reliability requirement $\Pr(E_B) = \epsilon_B$, the threshold SNR becomes

$$G_{B,f}^{\min} = \Gamma_B \ln\left(\frac{1}{1 - \epsilon_B}\right). \quad (4)$$

The instantaneous power $P_B(G_{B,f})$ is chosen as a function of the channel gain $G_{B,f}$ as

$$P_B(G_{B,f}) = \begin{cases} \dfrac{G_{B,f}^{\text{tar}}}{G_{B,f}} & \text{if } G_{B,f} \geq G_{B,f}^{\min} \\ 0 & \text{if } G_{B,f} < G_{B,f}^{\min}, \end{cases} \quad (5)$$

[1]Notice that full CSI of eMBB device is assumed as in [7]. Since eMBB transmissions are scheduled, devices have sufficient time to undergo through CSI acquisition procedures [7], [10].

where $G_{B,f}^{\text{tar}}$ is the target SNR, which is obtained by imposing the average power constraint as

$$\mathbb{E}\left\{P_B(G_{B,f})\right\} = \frac{G_{B,f}^{\text{tar}}}{\Gamma_B}\Gamma\left(0, \frac{G_{B,f}^{\min}}{\Gamma_B}\right) = 1.$$

This implies that the target SNR is

$$G_{B,f}^{\text{tar}} = \frac{\Gamma_B}{\Gamma\left(0, \frac{G_{B,f}^{\min}}{\Gamma_B}\right)}, \tag{6}$$

where $\Gamma(a,z) = \int_z^\infty t^{a-1}e^{-t}\mathrm{d}t$ is the upper incomplete gamma function. Finally, the outage rate achieved by the eMBB device is

$$r_B^{\text{orth}} = \log_2(1 + G_{B,f}^{\text{tar}}). \tag{7}$$

C. Performance Analysis of the URLLC device

Since the URLLC devices are assumed to have no CSI, and for mathematical tractability and simplicity, herein, we consider that all of them transmit with the same data rate r_U. We also adopt a worst case assumption that there are always $n_U \in \{\mathbb{N}^+\}$ URLLC devices transmitting in all minislots, thus there is always NOMA interference when $n_U > 1$. The BS performs SIC[2] decoding to decode the multiple overlapping signals.

Let u denote the index of the URLLC device with channel gain $G_{U_u,f}$ in the allocated frequency channel f. Besides, let $\{1, \cdots, n_U\}$ denote the SIC decoding ordering. The Signal-to-Interference-plus-Noise Ratio (SINR) in the frequency channel f when decoding the u-th active URLLC device, and assuming that the devices with indexes $\{1, \cdots, u-1\}$ have been correctly decoded, reads

$$\sigma_{u,f} = \frac{G_{U_u,f}}{1 + \sum\limits_{j=u+1}^{n_U} G_{U_j,f}}. \tag{8}$$

The active URLLC device is decoded successfully if

$$\frac{1}{F_U}\sum_{f=1}^{F_U}\log_2(1+\sigma_{u,f}) \geq r_U. \tag{9}$$

During the SIC decoding procedure, the BS first attempts to decode the strongest user among all the active URLLC devices in a minislot. If this user is correctly decoded, its interference is subtracted from the received signal, then the BS attempts to decode the next user in the order of strongest users, and so on. The SIC decoding procedure ends when the decoding of one active URLLC device fails or after all the active URLLC devices have been correctly decoded. We assume that all SIC decoding steps can be realized within the time duration of a minislot.

For simulation purposes, and in order to emulate the behaviour of the BS while performing the SIC decoding of the URLLC devices, we define the SIC decoding ordering according to their sum of mutual information across the F_U frequency channels, which is defined by

$$I_u^{\text{sum}} = \sum_{f=1}^{F_U}\log_2(1+\sigma_{u,f}). \tag{10}$$

Given the reliability requirement $\Pr(E_U) \leq \epsilon_U$, the objective is to obtain the maximum rate r_U that is a function of the number of frequency channels allocated for URLLC traffic. Then, the URLLC sum rate is given by

$$r_U^{\text{sum}} = n_U\, r_U. \tag{11}$$

Increasing F_U enhances the frequency diversity and, hence, makes it possible to satisfy the reliability target ϵ_U at a larger rate r_U [7]. Besides, during the computation of r_U, the error probabilities for all URLLC devices are computed individually.

III. Slicing for eMBB and URLLC

In this section, we consider the coexistence of eMBB and NOMA URLLC devices for both orthogonal and non-orthogonal slicing.

A. Orthogonal Slicing between eMBB and URLLC

Under the orthogonal slicing scenario, F_U out of F frequency channels for all minislots are allocated for URLLC traffic, while the remaining $F_B = F - F_U$ channels are each allocated to one eMBB device. The performance of the system is specified in terms of the pair $(r_B^{\text{sum}}, r_U^{\text{sum}})$ of eMBB sum-rate r_B^{sum} and URLLC sum-rate r_U^{sum}.

The eMBB sum-rate is given by [7]

$$r_B^{\text{sum}} = (F - F_U)r_B^{\text{orth}}, \tag{12}$$

where r_B^{orth} is given by (7). Given a number F_U of frequency channels allocated for URLLC, we compute the maximum URLLC rate r_U that guarantees the reliability constraint $\Pr(E_U) \leq \epsilon_U$ for all n_U URLLC devices transmitting simultaneously, as detailed in Section II-C.

B. Non-Orthogonal Slicing between eMBB and URLLC

In the non-orthogonal slicing scenario, all F frequency channels are used for both eMBB and URLLC services. Hence, $F_U = F_B = F$. Due to the latency constraints, the decoding of a URLLC transmission cannot wait for the decoding of eMBB traffic. The eMBB requirements are less demanding in terms of latency, and hence eMBB decoding can wait for the URLLC transmissions to be decoded first. This enables a SIC mechanism whereby URLLC packets are successive decoded and then canceled from the received signal prior to the decoding of the eMBB signal [7]. As a consequence, during the decoding attempts of the URLLC packets, the interference of the eMBB traffic is always present.

In the orthogonal case, as shown in (6), the variable $G_{B,f}^{\text{tar}}$ is uniquely determined by the error probability target ϵ_B and the threshold SNR $G_{B,f}^{\min}$ defined in (4). For the non-orthogonal slicing, it may be beneficial to choose a smaller target SNR

[2]Note that, as presented in [7], SIC outperforms other techniques of multi-user detection, such as puncturing.

than the one given if (6), so as to reduce the interference caused to URLLC transmissions. This yields the inequality [7]

$$G_{B,f}^{\text{tar}} \leq \frac{\Gamma_B}{\Gamma\left(0, \frac{G_{B,f}^{\min}}{\Gamma_B}\right)}. \tag{13}$$

The maximum allowed SNR for the eMBB devices, which we denote by $G_{B,\max}^{\text{tar}}$, is set by the inequality in (13). Consequently, the maximum allowed eMBB data rate is given by $r_B^{\max} = \log_2(1 + G_{B,\max}^{\text{tar}})$.

The objective of the analysis is to determine the rate pair $(r_B^{\text{sum}}, r_U^{\text{sum}})$ for which the reliability requirements of the two services are satisfied. To this end, first we fix an eMBB data rate $r_B \in [0, r_B^{\max}]$ and then we compute the maximum achievable rate r_U. During this computation, for a given value of r_U, we search for the minimum value of the SNR $G_B^{\text{tar}} \in [G_B^{\min}, G_{B,\max}^{\text{tar}}]$ that can be used for all eMBB devices. The error probabilities for all eMBB and URLLC devices are computed individually.

Considering the SIC decoding ordering $\{1, \cdots, n_U\}$, the SINR in the frequency channel f while decoding the u-th active URLLC device, and assuming that the devices with indexes $\{1, \cdots, u-1\}$ have been correctly decoded, is given by

$$\sigma_{u,f} = \frac{G_{U_u,f}}{1 + G_{B,f}^{\text{tar}} + \sum_{j=u+1}^{n_U} G_{U_j,f}}. \tag{14}$$

The u-th URLLC device is correctly decoded if the condition given by (9) holds, but now considering the SINR given by (14).

The SIC decoding performed in the non-orthogonal slicing is similar to the procedure performed in the orthogonal scenario and described in Section II-C. The only difference is that in the non-orthogonal slicing there is always interference from the eMBB devices in all F_U frequency channels during the SIC decoding of the URLLC devices. The BS tries to decode the eMBB packets only after all the URLLC devices have been correctly decoded. If an error occurs during the SIC decoding of the URLLC devices, the packets from the eMBB devices are lost.

IV. Numerical Results

In this section we present Monte Carlo simulation results for the orthogonal and non-orthogonal slicing of radio resources between eMBB and URLLC. For the sake of tractability, we consider the cases where the maximum number of URLLC devices transmitting simultaneously in the same minislot is $n_U \in \{1, 2, 3, 4\}$, where $n_U = 1$ is the scenario from [7]. Besides, we consider a time-frequency grid with $F = 10$ frequency channels and reliability requirements $\epsilon_U = 10^{-5}$ and $\epsilon_B = 10^{-3}$ [7]. Besides, we assume that the URLLC and eMBB devices are located in a similar environment, e.g. within a sector of a factory floor, aggregated by their proximity to the BS. This renders the same average received SNR to all the devices belonging to the same service class.

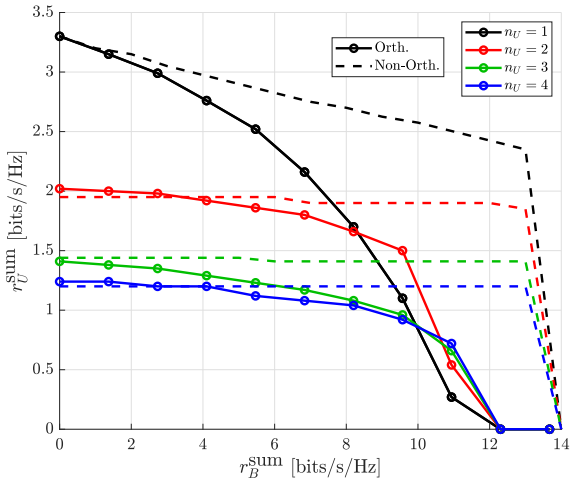

Fig. 3: eMBB sum rate r_B^{sum} versus URLLC sum rate r_U^{sum} for the the orthogonal and non-orthogonal slicing when $\Gamma_U = 20$ dB, $\Gamma_B = 10$ dB, $F = 10$, $\epsilon_U = 10^{-5}$ and $\epsilon_B = 10^{-3}$.

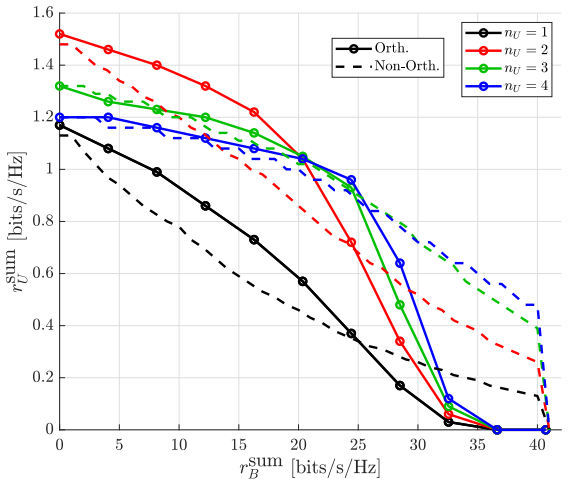

Fig. 4: eMBB sum rate r_B^{sum} versus URLLC sum rate r_U^{sum} for the the orthogonal and non-orthogonal slicing when $\Gamma_U = 10$ dB, $\Gamma_B = 20$ dB, $F = 10$, $\epsilon_U = 10^{-5}$ and $\epsilon_B = 10^{-3}$.

Fig. 3 shows the pairs of achievable sum rates considering $\Gamma_U = 20$ dB and $\Gamma_B = 10$ dB, that is, a scenario where the URLLC devices have better channel conditions than the eMBB devices. In this situation, due to the high levels of interference among the URLLC devices, the highest values of r_U^{sum} are achieved when only one URLLC device is transmitting in the minislot, that is, $n_U = 1$. Moreover, as we increase n_U, we decrease the achievable values of r_U^{sum}, which renders a trade-off between data rate and number of connected users. Another important conclusion is that the non-orthogonal slicing outperforms its orthogonal counterpart for the whole range of r_B^{sum}, since the former allows us to achieve pairs of sum rates that are not possible to achieve using the latter. Note also that, when adopting the non-orthogonal slicing and for $n_U > 1$, the values of r_U^{sum} are approximately constant for the whole

range of r_B^{sum}.

Fig. 4 shows the pairs of achievable sum rates for an opposite scenario where $\Gamma_U = 10$ dB and $\Gamma_B = 20$ dB, that is, when the URLLC devices have worse channel conditions than the eMBB devices. Compared to the previous case, now we can achieve much higher eMBB sum rates at the cost of lower URLLC sum rates. Moreover, we observe that the the URLLC sum rates achieved with $n_U > 1$ are higher compared to the case with $n_U = 1$ due to the lower levels of interference among the URLLC devices. For $r_B^{\text{sum}} < 25$ bits/s/Hz, the orthogonal slicing outperforms the non-orthogonal strategy for all values of n_U, since it allows us to achieve higher values of r_U^{sum}. The non-orthogonal slicing becomes advantageous for $r_B^{\text{sum}} > 25$ bits/s/Hz, since it allows us to achieve URLLC sum rates that are not possible to achieve with the orthogonal strategy. In this higher range, increasing n_U also increases r_U^{sum}, which represent gains in both data rates and number of connected devices. Note also that, in both Figs. 3 and 4, the curves obtained for $n_U = 1$ match those in [7].

Considering both slicing strategies, we conclude that the best compromise between number of NOMA users and data rates is achieved for limited number of users, in this case $n_U \leq 2$, which corroborates with other related works, e.g. [16]. We also conclude that the non-orthogonal slicing is the best choice in applications where URLLC devices coexist with eMBB devices that require very high data rates rates. Note that NOMA with SIC decoding increases the receiver complexity and yields higher delays in processing times, which must be taken into consideration in practical applications.

V. Conclusions

We proposed the joint use of NOMA, SIC decoding and frequency diversity as a solution to improve the number of URLLC devices that are connected in the uplink to the same BS and when they share the same RAN with eMBB devices. The radio resources are shared between the two services by employing orthogonal or non-orthogonal network slicing strategies. Resorting to Monte Carlo simulations, we showed that the proposed method allow multiple eMBB and URLLC devices to transmit overlapping signals to same BS while their reliability requirements are still met. We also demonstrated that when the URLLC devices have better channel conditions than the eMBB devices, the non-orthogonal slicing is advantageous over the orthogonal slicing for the whole range of eMBB sum rates. However, when the eMBB devices have better channel conditions than the URLLC devices, the non-orthogonal slicing outperforms the orthogonal slicing only for very high values of eMBB sum rates.

As stated in [2], the current 5G NR network is not yet capable of meeting the very stringent latency and reliability requirements of URLLC applications. Meeting these requirements requires hyper-flexible networks where technologies such as Artificial Intelligence (AI) and Machine Learning (ML) can be used to determine the optimal radio resource allocations for BSs and users. The framework developed in this work can be used in the specification of such 6G networks for mURLLC scenarios, where a large number of devices used for the control and/or monitoring of critical processes may require URLLC connectivity in the coexistence with other applications that require the high data rates provided by eMBB, e.g. video surveillance.

Acknowledgment

This research was financially supported by 6Genesis Flagship project (grant no. 318927), FIREMAN project (grant no. 326201) and Academy Professor project from Academy of Finland (grant no. 307492).

References

[1] H. Tullberg, P. Popovski, Z. Li, M. A. Uusitalo, A. Hoglund, O. Bulakci, M. Fallgren, and J. F. Monserrat, "The METIS 5G System Concept: Meeting the 5G Requirements," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 132–139, December 2016.

[2] M. Latva-Aho and K. Leppänen, "Key Drivers and Research Challenges for 6G Ubiquitous Wireless Intelligence," in *6G Wireless Summit, Levi, Finland*, Mar 2019.

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

[4] GSM Association, "An Introduction to Network Slicing," Tech. Rep., 2017. [Online]. Available: https://www.gsma.com/futurenetworks/wp-content/uploads/2017/11/GSMA-An-Introduction-to-Network-Slicing.pdf

[5] N. H. Mahmood, H. Alves, O. A. López, M. Shehab, D. P. M. Osorio, and M. Latva-Aho, "Six Key Features of Machine Type Communication in 6G," in *2020 2nd 6G Wireless Summit (6G SUMMIT)*, 2020, pp. 1–5.

[6] L. Dai, B. Wang, Y. Yuan, S. Han, C. I, and Z. Wang, "Non-Orthogonal Multiple Access for 5G: Solutions, Challenges, Opportunities, and Future Research Trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[7] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View," *IEEE Access*, vol. 6, pp. 55 765–55 779, 2018.

[8] Z. Wu, F. Zhao, and X. Liu, "Signal Space Diversity Aided Dynamic Multiplexing for eMBB and URLLC Traffics," in *2017 3rd. Int. Conf. Comput. Commun. (ICCC)*, Dec 2017, pp. 1396–1400.

[9] A. Anand, G. De Veciana, and S. Shakkottai, "Joint Scheduling of URLLC and eMBB Traffic in 5G Wireless Networks," in *2018 IEEE Int. Conf. Comput. Commun. (INFOCOM)*, April 2018, pp. 1970–1978.

[10] A. A. Esswie and K. I. Pedersen, "Opportunistic Spatial Preemptive Scheduling for URLLC and eMBB Coexistence in Multi-User 5G Networks," *IEEE Access*, vol. 6, pp. 38 451–38 463, 2018.

[11] R. Abreu, T. Jacobsen, G. Berardinelli, K. Pedersen, N. H. Mahmood, I. Z. Kovacs, and P. Mogensen, "On the Multiplexing of Broadband Traffic and Grant-Free Ultra-Reliable Communication in Uplink," in *2019 IEEE 89th Veh. Technol. Conf. (VTC2019-Spring)*, 2019, pp. 1–6.

[12] M. Alsenwi, N. H. Tran, M. Bennis, A. Kumar Bairagi, and C. S. Hong, "eMBB-URLLC Resource Slicing: A Risk-Sensitive Approach," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 740–743, April 2019.

[13] P. K. Korrai, E. Lagunas, S. K. Sharma, S. Chatzinotas, and B. Ottersten, "Slicing Based Resource Allocation for Multiplexing of eMBB and URLLC Services in 5G Wireless Networks," in *2019 IEEE Int. Workshop Comput. Aided Model. Des. Commun. Links Netw. (CAMAD)*, Sep. 2019, pp. 1–5.

[14] R. Kassab, O. Simeone, P. Popovski, and T. Islam, "Non-Orthogonal Multiplexing of Ultra-Reliable and Broadband Services in Fog-Radio Architectures," *IEEE Access*, vol. 7, pp. 13 035–13 049, 2019.

[15] E. J. dos Santos, R. D. Souza, J. L. Rebelatto, and H. Alves, "Network Slicing for URLLC and eMBB With Max-Matching Diversity Channel Allocation," *IEEE Commun. Lett.*, vol. 24, no. 3, pp. 658–661, 2020.

[16] O. L. Alcaraz López, H. Alves, P. H. Juliano Nardelli, and M. Latva-aho, "Aggregation and Resource Scheduling in Machine-Type Communication Networks: A Stochastic Geometry Approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4750–4765, July 2018.