Rate-Splitting Multiple Access for URLLC Uplink in Physical Layer Network Slicing With eMBB

ELÇO JOÃO DOS SANTOS JR. [1], **RICHARD DEMO SOUZA** [1], (Senior Member, IEEE),
AND JOÃO LUIZ REBELATTO [2], (Senior Member, IEEE)

[1] Department of Electrical and Electronics Engineering, Federal University of Santa Catarina (UFSC), Florianópolis, Santa Catarina 88040-900, Brazil
[2] Department of Electrical Engineering, Federal University of Technology—Paraná (UTFPR), Curitiba, Paraná 80230-901, Brazil

Corresponding author: Elço João dos Santos Jr. (e.joaojr@gmail.com)

ABSTRACT In this paper, we investigate the problem of heterogeneous service coexistence in the scope of 5G and beyond (B5G) networks, where multiple ultra-reliable low-latency communication (URLLC) and enhanced mobile broadband (eMBB) users are connected to a common base station (BS), sharing physical network resources. In contrast to the orthogonal multiple access (OMA) and non-orthogonal multiple access (NOMA) usually adopted in literature, in this work we employ rate splitting multiple access (RSMA) for URLLC transmission, where a URLLC device splits its message into two sub-messages with partial transmission power, which are potentially recovered at the BS by means of successive interference cancellation (SIC). To study the performance of such methods in the presence of eMBB users, we consider both orthogonal and non-orthogonal network slicing approaches to share the network resources between heterogeneous user profiles with diverse requirements. As a result, we show that, in general, RSMA presents an improved performance in terms of sum-rate and reliability, even when transmitting concurrently with eMBB users. Finally, our results also show that the URLLC sum-rate can be increased by properly adjusting the rate splitting factor based on the average signal-to-noise ratio (SNR), not being necessary instantaneous channel state information (CSI).

INDEX TERMS Beyond 5G, heterogeneous users, rate splitting multiple access, ultra-reliable and low latency communications.

I. INTRODUCTION

As the 5G technology deployment around the world evolves, it becomes clear how challenging are the three generic services encompassed by such technology, namely enhanced mobile broadband (eMBB), ultra-reliable and low latency communications (URLLC), and massive machine type communications (mMTC). To allow the coexistence of these heterogeneous services with diverse requirements within the same Radio Access Network (RAN) architecture, the concept of network slicing has been proposed [1], which slices the network in logical and physical sub-networks usually with customized requirements in terms of latency, energy efficiency, mobility, massive connectivity and throughput [2], aiming at guaranteeing minimum performance requirements and isolation [3], [4]. This can be performed thanks to

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Awais Javed [ID].

network softwarization and virtualization, being considered the main enabler of Resource as a Service (RaaS) for beyond-5G (B5G) [5]. In the path to B5G and 6G wireless communication systems, it is reasonable to assume that the three heterogeneous services could be divided into sub-services [6] or even combined, emerging new service classes [7]. Such services require robust multiple access methods that can combine higher spectral efficiency with strict delay and reliability requirements to attend applications like fully automated driving, where cooperation among cars for collision avoidance is vital [8], [9].

To face the massive connectivity problem, some methods have been proposed in the past few years to replace the traditional orthogonal multiple access (OMA). One of them is non-orthogonal multiple access (NOMA), a promising technology that usually exploits the power domain to allow multiple users to share the same resource block along the spectrum, time and/or code, increasing the spectral efficiency [10].

E. J. dos Santos Jr. *et al.*: Rate-Splitting Multiple Access for URLLC Uplink in Physical Layer Network Slicing With eMBB

IEEE *Access*

In order to recover the overlapped signals, the receiver of a NOMA-based communication system can apply the successive interference cancellation (SIC) algorithm, a method whose performance depends on the different power levels between the overlapped incoming signals [11]. To this end, two approaches are commonly used to guarantee such power distinctiveness: (i) user pairing; and (ii) power allocation. In (i), users with distinct channel gains are separated in groups and paired [12], [13]. It is intuitive that the complexity of such technique increases with the number of users, turning its implementation unbearable in terms of latency in scenarios with a massive number of users. In (ii) power allocation methods separate users [14], [15], even if random pairing is applied. This implies, in some cases, the need of channel state acquisition to adapt the power of transmission, which entails extra latency and a potential loss in terms of reliability.

The rate-splitting multiple access (RSMA) method has gained significant attention recently, since it enables the achievement of the entire capacity region with successive decoding [16], [17], providing superior performance over NOMA and OMA methods, like Space Division Multiple Access (SDMA) [18], [19]. In uplink RSMA, each user creates virtual users by splitting its transmission in two sub-messages. Although this procedure entails extra rounds in the SIC procedure, it automatically creates different arriving power levels among users, thus significantly reducing the implementation complexity. One of the main advantages of RSMA is the increased number of possible decoding orders, which makes it viable to reach higher capacity regions when compared to NOMA. This is also a big challenge in practical RSMA deployments and must be optimized, since the decoding order affects the achievable rate.

A. RELATED WORK

Recently, several works studied different RSMA implementations in downlink wireless networks [20]–[24], showing that RSMA can improve downlink rate and quality of service, achieving better performance than both NOMA and SDMA. For uplink RSMA systems, authors from [25], [26] study the problem of maximizing the sum-rate under proportional rate constraints for all users, by setting users transmission power and optimizing the decoding order at the BS through exhaustive search. As a result, they show that RSMA achieves better performance than NOMA and OMA techniques, such as frequency division multiple access (FDMA) and time division multiple access (TDMA). However, the proposed strategy requires *a priori* channel state information (CSI), not being in general applicable to URLLC users due to latency constraints. In [27], the authors propose the use of RSMA to reduce the scheduling complexity of NOMA, since the transmission splitting by default diversifies the arriving power at the BS, avoiding the need of user pairing. In [28], the authors apply rate splitting to a pair of users under power-domain NOMA, considering that one of them is near the BS, while the other is far from the BS. Two techniques are studied, namely, fixed rate splitting (FRS) and adaptive rate splitting (ARS), where

the power allocation factor that splits the messages of the near user can be fixed or dynamically designed based on CSI, respectively. This work is then extended in [29], adopting cyclic prefixed single carrier transmissions. In both works, rate splitting has been shown to achieve superior outage performance when compared to NOMA.

In [30], an exhaustive-search rate splitting algorithm was proposed to guarantee max-min fairness in single-input multiple-output (SIMO) NOMA networks, aiming at maximizing the minimum data rate and reduce the scheduling process. The receiver combines minimum mean squared error (MMSE) with SIC to identify the optimal detection order based on CSI. Results showed that rate splitting has higher minimum data rate and lower transmission latency than SIMO-OMA and SIMO-NOMA. The use of rate splitting in user cooperation networks is proposed in [31]. Each user transmits its signal and receives the transmitted signal of the other user in the first mini-slot and, at the second mini-slot, relays the other user's message with amplify-and-forward protocol. The rate is split between mini-slots, generating space diversity at the uplink and consequently increasing reliability. At the receiver, maximum ratio combining (MRC) is used to combine the received signals and SIC is applied to decode the superposed signal. Results prove that cooperative RSMA outperforms cooperative OMA and NOMA.

In scenarios with spectrum sharing among URLLC and eMBB services, several works compared OMA and NOMA network slicing [32]–[37]. However, none of the aforementioned works consider multiple concurrent URLLC users in the same resource block. In [38], URLLC users are assumed to share time and frequency resources through NOMA, in both OMA and NOMA slicing with eMBB service. It was shown that NOMA can leverage the URLLC sum-rate in some cases, considering that the SIC process is capable of attending the communication latency. Authors from [39] apply RSMA to URLLC in the downlink, showing its superior performance in terms of latency, allowing shorter block lengths. However, no interference from other services is considered.

B. NOVELTY AND CONTRIBUTION

Motivated by the above literature, in this work we focus on increasing the URLLC spectral efficiency, allowing non-orthogonal sharing of frequency and time resources through rate-splitting for URLLC users, which we refer to U-RSMA. In the proposed scheme, we combine the benefits of RSMA, SIC decoding and frequency diversity, in both OMA and NOMA slicing with eMBB. The proposed U-RSMA scheme is then compared to the so-called U-NOMA and U-OMA schemes, where the multiple access between URLLC devices is performed by means of NOMA and OMA, respectively. To characterize the performance of eMBB and URLLC users, we evaluate each service sum-rate in different scenarios. To the best of our knowledge, this work is the first to apply RSMA to URLLC uplink transmission in

TABLE 1. List of symbols.

Symbol	Meaning
α	Power splitting factor of U-RSMA
$\bar{\Gamma}_U$	URLLC average SNR
$\bar{\Gamma}_B$	eMBB average SNR
ϵ_B	eMBB reliability
$\epsilon_U^{\text{U-OMA}}$	URLLC reliability in U-OMA
$\epsilon_U^{\text{U-NOMA}}$	URLLC reliability in U-NOMA
$\epsilon_U^{\text{U-RSMA}}$	URLLC reliability in U-RSMA
$\sigma_{n,f}$	SINR of URLLC user n in channel f
a_B	eMBB activation probability
a_U	URLLC activation probability
E_B	Event when eMBB user is in outage
E_U	Event when URLLC user is in outage
$\bar{E}_U$	Event when URLLC user is decoded
f	Channel index
F	Number of channels in the bandwidth
F_B	Number of channels allocated to eMBB
F_U	Number of channels allocated to URLLC
F'_U	Number of channels of an URLLC user in U-OMA
$G_{B,f}$	Channel gain of eMBB user in channel f
$G_{B,f}^{\min}$	Threshold SNR of eMBB user in channel f
$G_{B,f}^{\text{tar}}$	Target SNR of eMBB user in channel f
$G_{U,n,f}$	Channel gain of URLLC user n in channel f
$H_{i,f}$	Channel coefficient of user i in channel f
$\mathbb{I}_n^{\text{sum}}$	Mutual information of URLLC user n
$\mathbb{P}(E_B)$	Outage probability of eMBB user
$\mathbb{P}(E_U)$	Outage probability of URLLC user
$P(G_{B,f})$	eMBB instantaneous transmission power
$p_{G_{B,f}}(x)$	Probability density function of $G_{B,f}$
n_U	Number of URLLC users sharing the same mini-slot
r_B^{orth}	eMBB rate in OMA
$r_B^{\text{n-orth}}$	eMBB rate in NOMA
r_B^{sum}	eMBB sum-rate
$r_{U,n}$	Rate of URLLC user n
r_U^{sum}	URLLC sum-rate
S	Number of mini-slots
S_U	Number of URLLC transmissions during one time slot
$\mathbb{U}_{B,n}$	eMBB user n
$\mathbb{U}_{U,n}$	URLLC user n

a network slicing scenario, showing that RSMA can outperform OMA and NOMA methods for URLLC service even in the presence of eMBB interference, specially for very strict reliability levels.

The rest of this paper is organized as follows. Section II presents the system model. Section III introduces the outage formulation for eMBB and URLLC (for U-OMA, U-NOMA, and U-RSMA cases), for both orthogonal and non-orthogonal network slicing approaches. Numerical results illustrating the performance trade-offs between the services are given in Section IV. Finally, Section V concludes the paper.

Notation: For convenience, the list of symbols adopted in this work is summarized in Table 1.

II. SYSTEM MODEL

We evaluate the uplink of multiple eMBB and URLLC users when communicating to a common Base Station (BS) in a single-cell network with shared radio resources. The bandwidth is divided into F channels of index $f \in \{1, \ldots, F\}$ subject to independent and identically distributed (i.i.d.) Rayleigh fading. The fading realization observed by each

device is uncorrelated from another due to the assumption that all devices have a large enough spatial separation. Furthermore, the fading is considered constant during one time slot (TS), i.e., a block fading model where the TS is considered to be within the channel coherence time since its length is fairly small [40]. As we assume that the average transmission power of all devices and the noise power at the BS are normalized to one, the received power equals the signal-to-noise ratio (SNR) for each device. Moreover, the channel fading realization for user $i \in \{B, U\}$ in channel f is $H_{i,f} \sim \mathcal{CN}(0, \bar{\Gamma}_i)$, following a circular-symmetric complex Gaussian distribution, where $\bar{\Gamma}_i$ corresponds to the average SNR, being $G_{i,f} \triangleq |H_{i,f}|^2$ the channel gain, and where subscripts B and U refer to eMBB and URLLC devices, respectively. The number of channels allocated to user i is $F_i \leq F$, with $i \in \{B, U\}$. Moreover, each TS is divided into S mini-slots, as considered in low latency scenarios [41].

In accordance to [32], we assume that an eMBB user is active with probability a_B and during the entire TS, occupying one random frequency channel f among F_B available channels. Furthermore, we model only the transmission phase, assuming that radio access and competition among eMBB devices have been resolved prior to the considered time slot, as usual in wireless cellular networks. Thus, the number of eMBB devices able to transmit in such TS is equal to the number of channels F_B. Moreover, we suppose that the eMBB devices and the BS have CSI as currently implemented in wireless standards such as LTE and 5G New Radio [42]–[44]. Although channel estimation errors can occur in practice, for simplicity we consider a perfect CSI scenario in this work, as widely considered in the literature [32], [45]. In contrast, an URLLC device spreads its transmission over $F_U \leq F$ channels to increase the reliability with the aid of frequency diversity, and sends, with some activation probability a_U, the entire information in only one mini-slot (the smallest time unit in our model) that was pre-assigned to meet latency requirements. We also consider that the protocol block length, which should be considered finite given the short transmissions, is long enough to justify an asymptotic information-theoretic formulation [46]. Moreover, in each mini-slot we have a maximum number of n_U users that share the resources following three distinct methods: orthogonal (U-OMA), non-orthogonal (U-NOMA) or through rate splitting (U-RSMA) multiple access.

Different from eMBB users, we assume that the BS has no knowledge about the URLLC channel, given the high latency requirement which does not allow the exchange of reference signals for CSI acquisition. However, we do consider in U-RSMA that the BS sends (*e.g., in a synchronization mini-slot transmitted at the end of each TS*), the optimal power splitting factor based on $\bar{\Gamma}_U$ from a look-up table, which results in power adaptation for the user that performs the splitting. Despite that, the overall transmission power is the same as in U-OMA and U-NOMA cases.

A time-frequency grid is illustrated in Fig. 1, considering that the heterogeneous URLLC and eMBB traffics
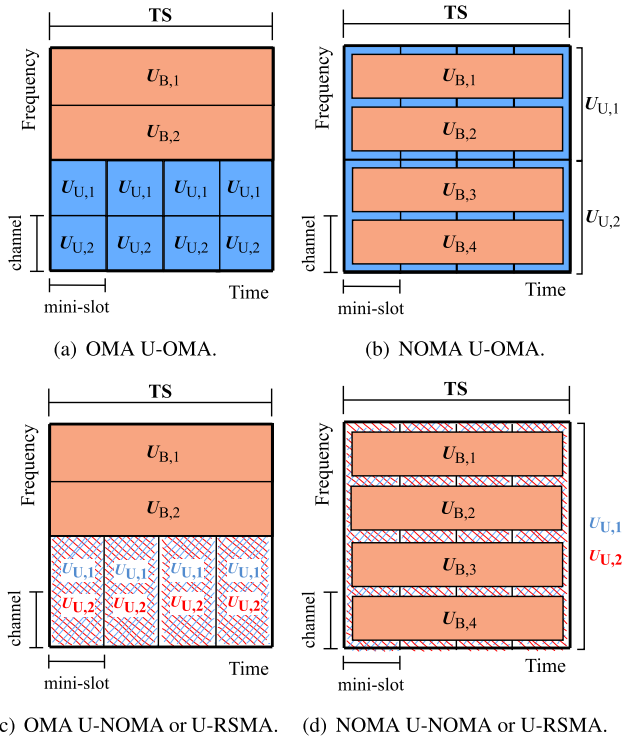
E. J. dos Santos Jr. *et al.*: Rate-Splitting Multiple Access for URLLC Uplink in Physical Layer Network Slicing With eMBB

IEEE *Access*

(a) OMA U-OMA.

(b) NOMA U-OMA.

(c) OMA U-NOMA or U-RSMA.

(d) NOMA U-NOMA or U-RSMA.

FIGURE 1. System model with $F = 4$ channels and $S = 4$ mini-slots, composed by eMBB and URLLC users. Services are sliced in (a) and (c) Orthogonal and (b) and (d) Non-Orthogonal multiple access schemes.

are sliced in an OMA (Figs. 1(a) and 1(c)), and NOMA (Figs. 1(b) and 1(d)) fashion. In this example, $S = 4$ is the quantity of mini-slots in the time domain, whereas $F = 4$ is the total number of channels available in the bandwidth. Considering the OMA scenario, two channels are allocated to URLLC ($F_U = 2$) and two for eMBB ($F_B = 2$). There are $n_U = 2$ URLLC active users, $\mathbb{U}_{U,1}$ and $\mathbb{U}_{U,2}$, in each mini-slot that spread their transmission over one channel, in the case of U-OMA, or over two channels when considering U-NOMA or U-RSMA, without interference from eMBB users. On the eMBB band, there are also two users, $\mathbb{U}_{B,1}$ and $\mathbb{U}_{B,2}$, connected to the BS. When considering NOMA, all four channels are available for both services ($F = F_U = F_B = 4$), which implies a multi-service interference, turning the detection at the BS more complex and prone to errors. The frequency diversity gain for URLLC users is higher in this case, and, as this device type does not necessarily transmit at every TS, the spectrum efficiency should increase because eMBB users can occupy a radio resource that might be unused for long periods, which is represented with the inclusion of new eMBB users $\mathbb{U}_{B,3}$ and $\mathbb{U}_{B,4}$.

III. OUTAGE FORMULATION AND SLICING SCHEMES

In this section, we discuss the achievable rates of the different services and slicing schemes.

A. EMBB

A given eMBB device transmits, with a certain instantaneous power and data rate, in the randomly allocated dedicated radio resource $f \in \{1, \ldots, F_B\}$, if the instantaneous channel gain is greater than a threshold SNR $G_{B,f}^{\min}$. This decision is made based on CSI. The outage probability of a point-to-point (single channel) communication is then [32]

$$\mathbb{P}(E_B) = \Pr[G_{B,f} < G_{B,f}^{\min}] = \int_0^{G_{B,f}^{\min}} p_{G_{B,f}}(x)dx, \quad (1)$$

where $p_{G_{B,f}}(x)$ is the probability density function (PDF) of $G_{B,f}$, which, due to the Rayleigh fading, is given by

$$p_{G_{B,f}}(x) = \begin{cases} \dfrac{e^{-x/\bar{\Gamma}_B}}{\bar{\Gamma}_B}, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

The eMBB outage probability is then obtained as [32]

$$\begin{aligned}
\mathbb{P}(E_B) &= \int_0^{G_{B,f}^{\min}} \frac{e^{-x/\bar{\Gamma}_B}}{\bar{\Gamma}_B}dx \\
&= \frac{1}{\bar{\Gamma}_B} \times -\bar{\Gamma}_B \times e^{-x/\bar{\Gamma}_B}\Big|_0^{G_{B,f}^{\min}} \\
&= -\left(e^{-G_{B,f}^{\min}/\bar{\Gamma}_B} - e^{-0/\bar{\Gamma}_B}\right) \\
&= 1 - e^{-G_{B,f}^{\min}/\bar{\Gamma}_B}.
\end{aligned} \quad (3)$$

Imposing the reliability condition $\mathbb{P}(E_B) = \epsilon_B$, one can obtain the threshold SNR from (3) as

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

The main objective of eMBB is to maximize its data rate, subject to the reliability requirement ϵ_B and the average power constraint $\mathbb{E}[P_B(G_{B,f})] = 1$, where $P_B(G_{B,f})$ is the instantaneous transmission power, selected using the power inversion scheme from [47] based on $G_{B,f}$, i.e.

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

This means that the eMBB device will not transmit in every slot allocated to it because of outage situations, then it is possible to increase the instantaneous power when the transmission occurs, so that the long-term average power $(P_B(G_{B,f}) = 1)$ is achieved. The target SNR $G_{B,f}^{\text{tar}}$ is then obtained by imposing the average power constraint to the expected value of the function $P_B(G_{B,f})$ of the random variable $G_{B,f}$. This is calculated using

$$\mathbb{E}\left[P_B(G_{B,f})\right] = \int_{G_{B,f}^{\min}}^{\infty} p_{G_{B,f}}(x)P_B(x)dx = 1. \quad (6)$$

After replacing (2) in (6), one has

$$\mathbb{E}\left[P_B(G_{B,f})\right] = \int_{G_{B,f}^{\min}}^{\infty} \frac{e^{-x/\bar{\Gamma}_B}}{\bar{\Gamma}_B} \frac{G_{B,f}^{\text{tar}}}{x}dx = 1$$

IEEE *Access*

E. J. dos Santos Jr. *et al.*: Rate-Splitting Multiple Access for URLLC Uplink in Physical Layer Network Slicing With eMBB

$$= \frac{G_{B,f}^{\text{tar}}}{\bar{\Gamma}_B} \underbrace{\int_{G_{B,f}^{\text{min}}}^{\infty} \frac{e^{-x/\bar{\Gamma}_B}}{x} dx}_{-\text{Ei}\left(-\frac{G_{B,f}^{\text{min}}}{\bar{\Gamma}_B}\right)} = 1, \quad (7)$$

where $-\text{Ei}\left(-G_{B,f}^{\text{min}}/\bar{\Gamma}_B\right)$ is obtained from the integral and can be classified as the upper incomplete gamma function $\Gamma(\cdot, \cdot)$ for $G_{B,f}^{\text{min}}/\bar{\Gamma}_B > 0$. Then, (7) can be rewritten as

$$\mathbb{E}\left[P_B(G_{B,f})\right] = \frac{G_{B,f}^{\text{tar}}}{\bar{\Gamma}_B} \Gamma\left(0, \frac{G_{B,f}^{\text{min}}}{\bar{\Gamma}_B}\right) = 1. \quad (8)$$

The target SNR of eMBB user $G_{B,f}^{\text{tar}}$ is then obtained from (8), resulting in

$$G_{B,f}^{\text{tar}} = \frac{\bar{\Gamma}_B}{\Gamma\left(0, \frac{G_{B,f}^{\text{min}}}{\bar{\Gamma}_B}\right)}. \quad (9)$$

Finally, one can obtain the eMBB rate as

$$r_B^{\text{orth}} = \log_2\left(1 + G_{B,f}^{\text{tar}}\right). \quad \text{(bits/s/Hz)} \quad (10)$$

B. URLLC
1) U-OMA
The F_U channels available for URLLC are divided in n_U orthogonal slices with F_U' channels reserved to each $\mathbb{U}_{U,n}$ user, with $n \in \{1, \ldots, n_U\}$. The outage probability of $\mathbb{U}_{U,n}$, in the absence of interference from other services, is [32]

$$\mathbb{P}^{\text{U-OMA}}(E_U) = \Pr\left(\frac{1}{F_U'} \sum_{f=1}^{F_U'} \log_2(1 + \sigma_{n,f}) < r_{U,n}\right), \quad (11)$$

where $\sigma_{n,f}$, the Signal-to-Interference-plus-Noise Ratio (SINR) of the n-th active user in frequency channel f, equals $G_{U,n,f}$, since for the moment there is no interference from other users. The target rate $r_{U,n}$ is numerically obtained by imposing the outage probability requirement $\mathbb{P}^{\text{U-OMA}}(E_U) \leq \epsilon_U$ to (11). Thus, the sum-rate of the URLLC service is given by

$$r_U^{\text{U-OMA}} = \sum_{n=1}^{n_U} r_{U,n}. \quad (12)$$

2) U-NOMA
In U-NOMA, URLLC users share the F_U channels available in each mini-slot and the BS performs SIC to decode the multiple messages, which outperforms other techniques of multi-user detection, such as puncturing and erasure decoding [32], and is a general receiver structure for non-orthogonal uplink [10]. As an user occupies more than one channel, we cannot simply define the decoding order in terms of the channel gain magnitude. Instead, the BS can order the users according to their mutual information [38]

$$\mathbb{I}_n^{\text{sum}} = \sum_{f=1}^{F_U} \log_2(1 + \sigma_{n,f}), \quad (13)$$

where $\sigma_{n,f}$ is defined as

$$\sigma_{n,f} = \frac{G_{U,n,f}}{1 + \sum_{j>n}^{n_U} G_{U,j,f}}. \quad (14)$$

The decoding procedure starts with the strongest among all the active users in the current mini-slot. If correctly decoded, it is removed from the received signal and the operation continues, until an user cannot be decoded (event that occurs with probability ϵ_U) or all users have been properly decoded. We consider that the BS is capable of decoding the n_U users within the mini-slot period, since each transmission carries a different message and the procedure must attend the latency requirement. The outage probability of the u-th user is

$$\mathbb{P}^{\text{U-NOMA}}(E_U) = \Pr\left(\frac{1}{F_U} \sum_{f=1}^{F_U} \log_2(1 + \sigma_{n,f}) < r_{U,n}\right). \quad (15)$$

The target rate $r_{U,n}$ is numerically obtained by imposing the requirement $\mathbb{P}^{\text{U-NOMA}}(E_U) \leq \epsilon_U$ to (15). Thus, the sum-rate of the URLLC service is

$$r_U^{\text{U-NOMA}} = \sum_{n=1}^{n_U} r_{U,n}. \quad (16)$$

3) U-RSMA
Either under U-OMA or U-NOMA, URLLC users directly transmit their data to the BS once they are active. However, in U-RSMA, an user may first split its information into two sub-messages, creating the concept of "virtual users". Each sub-message has transmission power defined by the so-called splitting factor $\alpha \in [0, 1]$.

As an example, let us consider the case with $n_U = 2$. In this two-user scenario, we assume that only one user, say $\mathbb{U}_{U,1}$, splits its message,[1] creating two virtual users referred to as $\mathbb{U}_{U,1,1}$ and $\mathbb{U}_{U,1,2}$. Without loss of generality, we consider that $\mathbb{U}_{U,1,1}$ is always decoded before $\mathbb{U}_{U,1,2}$. In this scenario, we have three possible decoding orders at the BS, namely: (i) $\mathbb{U}_{U,1,1} \rightarrow \mathbb{U}_{U,2} \rightarrow \mathbb{U}_{U,1,2}$; (ii) $\mathbb{U}_{U,1,1} \rightarrow \mathbb{U}_{U,1,2} \rightarrow \mathbb{U}_{U,2}$; and (iii) $\mathbb{U}_{U,2} \rightarrow \mathbb{U}_{U,1,1} \rightarrow \mathbb{U}_{U,1,2}$, such that the proper decoding order is chosen based on the sum of mutual information from (13), similarly to U-NOMA.

While the decoding orders (ii) and (iii) achieve the same results of U-NOMA with $\mathbb{U}_{U,1} \rightarrow \mathbb{U}_{U,2}$ and $\mathbb{U}_{U,2} \rightarrow \mathbb{U}_{U,1}$, respectively [48], is has been shown that (i) represents the optimal decoding order of RSMA [49]. Thus, in the SIC process, the receiver first attempts to decode a (virtual) user while regarding all the remaining messages as noise. Once the decoding is successful, its interference is removed out of the superimposed received signal, and the receiver then attempts to decode the next message following the pre-established decoding order. Upon adopting the decoding order from (i),

---

[1] Following [16], only one out of the two users needs to split its message in order to achieve the capacity region.

E. J. dos Santos Jr. *et al.*: Rate-Splitting Multiple Access for URLLC Uplink in Physical Layer Network Slicing With eMBB

IEEE *Access*

the SINR of the virtual user $\mathbb{U}_{U,1,1}$ is

$$\sigma_{1,1,f} = \frac{\alpha \, G_{U,1,1,f}}{1 + G_{U,2,f} + (1-\alpha)G_{U,1,2,f}}. \tag{17}$$

If $\mathbb{U}_{U,1,1}$ is correctly decoded and canceled from the received signal, the SINR of $\mathbb{U}_{U,2}$ becomes

$$\sigma_{2,f} = \frac{G_{U,2,f}}{1 + (1-\alpha)G_{U,1,2,f}}. \tag{18}$$

Finally, the SINR of the remaining virtual user $\mathbb{U}_{U,1,2}$, subject to the correct decoding of the previous users, is

$$\sigma_{1,2,f} = (1-\alpha)G_{U,1,2,f}. \tag{19}$$

Then, the achievable rates of U-RSMA can be calculated from (15), by substituting $\sigma_{n,f}$ with the SINRs of U-RMSA presented in (17)-(19). The final rate of user $\mathbb{U}_{U,1}$ is $r_{U,1} = r_{U,1,1} + r_{U,1,2}$. Thus, the sum-rate of the two-user U-RSMA URLLC service finally obtained as

$$r_U^{\text{U-RSMA}} = r_{U,1} + r_{U,2}. \tag{20}$$

It is worthy mentioning that, when compared to U-NOMA, U-RSMA requires an extra round in the SIC procedure, increasing the complexity of the decoding process.

C. ORTHOGONAL NETWORK SLICING

In Sections III-A and III-B we present, respectively, the achievable rates of eMBB and URLLC services when operating in standalone mode, without slicing the network resources. When such slicing between the heterogeneous eMBB and URLLC services is designed in a orthogonal fashion, they are "isolated" from each other, thus for URLLC the only source of interference are the n_U users active with probability a_U in certain mini-slot occupying all $F_U \leq F$ channels, whereas eMBB experiences an interference-free scenario since users are allocated orthogonally within the remaining $F_B = F - F_U$ channels. The OMA performance is measured in terms of the sum-rate pair $(r_B^{\text{sum}}, r_U^{\text{sum}})$, where r_B^{sum} can be defined as [32]

$$r_B^{\text{sum}} = F_B \, r_B^{\text{orth}}, \tag{21}$$

where r_B^{orth} comes from (10) and r_U^{sum} is computed as presented in Section III-B for each particular multiple access method adopted by the URLLC service.

D. NON-ORTHOGONAL NETWORK SLICING

In non-orthogonal slicing, eMBB and URLLC services simultaneously share all the F available channels, i.e., $F_B = F_U = F$. Due to latency and reliability constraints, it is assumed that the BS always attempts to decode the n_U active URLLC devices first, through SIC, while treating the eMBB traffic as interference. Therefore, the interference from URLLC transmissions into eMBB (and vice-versa) needs to be considered.

An eMBB message would not be affected by URLLC interference in two cases: (i) there are no URLLC devices

connected ($S_U = 0$) in that particular TS; or (ii) there are URLLC transmissions ($S_U > 0$), but they were decoded and removed from the signal by the SIC decoder. In case (ii), either all URLLC messages are properly decoded (event $\bar{E}_U$) or they are all incorrectly decoded (event E_U), since interference from eMBB users are constant over all mini-slots. Thus, the eMBB outage probability in the NOMA scenario depends on whether it is subjected to interference of URLLC service or not, i.e.

$$\begin{aligned} \mathcal{P}_B = {}& \Pr(S_U = 0)\Pr(E_B|S_U = 0) \\ & + \Pr(S_U > 0)\big(\Pr(E_U|S_U > 0)\Pr(E_B|E_U, S_U > 0) \\ & + \Pr(\bar{E}_U|S_U > 0)\Pr(E_B|\bar{E}_U, S_U > 0)\big), \end{aligned} \tag{22}$$

where E_B is the event of eMBB not being correctly decoded and $S_U \sim \text{Bin}(n_U S, a_U)$ is a random variable that represents the number of URLLC transmissions during the TS. The only source of outage for eMBB when there is no URLLC signal interfering is when the SNR value is below the threshold SNR ($G_{B,f}^{\min}$), which implies that the term $\Pr(E_B|S_U = 0)$ from (22) equals the outage probability for the orthogonal case $1 - a_B$, where $a_B = \exp[-G_{B,f}^{\min}/\bar{\Gamma}_B]$ for simplification purposes. Moreover, we also consider a simplified and worst case scenario where the eMBB user is in outage when the URLLC message is incorrectly decoded, i.e., such that $\Pr(E_B|E_U, S_U > 0) = 1$. Besides that, the correct decoding and subtraction of URLLC signal has the same performance effect of the case when URLLC is not transmitting, thus, $\Pr(E_B|\bar{E}_U, S_U > 0) = \Pr(E_B|S_U = 0) = 1 - a_B$. Under these assumptions,

$$\begin{aligned} \mathcal{P}_B \leq {}& (1 - a_U)^{n_U S}(1 - a_B) \\ & + \big[1 - (1 - a_U)^{n_U S} \big]\big(\epsilon_U + (1 - \epsilon_U)(1 - a_B) \big). \end{aligned} \tag{23}$$

By imposing the eMBB reliability constraint $\mathcal{P}_B \leq \epsilon_B$, one can rewrite (23) as

$$a_B \geq \frac{1 - \epsilon_B}{1 - \epsilon_U \big[1 - (1 - a_U)^{n_U S} \big]}. \tag{24}$$

Having in mind that $a_B = \exp[-G_{B,f}^{\min}/\bar{\Gamma}_B]$, it is possible to isolate the threshold SNR $G_{B,f}^{\min}$ from (24), resulting in

$$G_{B,f}^{\min} \leq -\bar{\Gamma}_B \ln\left(\frac{1 - \epsilon_B}{1 - \epsilon_U \big[1 - (1 - a_U)^{n_U S} \big]} \right). \tag{25}$$

The target SNR $G_{B,f}^{\text{tar}}$ is obtained similarly to (9) as

$$G_{B,f}^{\text{tar}} \leq \frac{\bar{\Gamma}_B}{\Gamma\left(0, \frac{G_{B,f}^{\min}}{\bar{\Gamma}_B} \right)}. \tag{26}$$

However, in the non-orthogonal case, $G_{B,f}^{\min}$ is bounded by (25). Therefore, the maximum achievable rate of an eMBB device in NOMA is $r_B^{\text{n-orth}} = \log_2(1 + G_{B,f}^{\text{tar}})$.

The threshold from (25) indicates that the impact of URLLC transmissions in the eMBB decoding should be minimal, due to the fact that, by definition, $\epsilon_U \ll \epsilon_B$, which

IEEE Access

E. J. dos Santos Jr. *et al.*: Rate-Splitting Multiple Access for URLLC Uplink in Physical Layer Network Slicing With eMBB

implies that a_B is close to $1 - \epsilon_B$. On the other hand, the eMBB interference in the URLLC traffic is supposed to be more critical, since URLLC is decoded prior to eMBB. As in [32] the outage probability of URLLC under NOMA is

$$
\mathbb{P}^{\text{NOMA}}(E_U)
$$
$$
= \Pr\left(\frac{1}{F_U} \sum_{f=1}^{F_U} \log_2\left(1 + \frac{\sigma_{n,f}}{1 + G_{B,f}^{\text{tar}}} \right) < r_{U,n} \right), \quad (27)
$$

where it is assumed that the interference of eMBB is always present in the URLLC decoding. The value of $\sigma_{n,f}$ depends on the multiple access technique used by URLLC users, as discussed in Section III-B. The URLLC achievable sum-rate r_U^{sum} is then numerically obtained by imposing the reliability constraint $\mathbb{P}^{\text{NOMA}}(E_U) \leq \epsilon_U$, where the rates are separately calculated for all n_U transmitting URLLC users.

IV. NUMERICAL RESULTS

In this section, we present some numerical results aiming at comparing the sum-rate performance of U-OMA, U-NOMA and U-RSMA under both OMA and NOMA network slicing strategies. These results were generated using Monte Carlo simulations in MATLAB, where, for each particular scenario, we average a number of 10^7 independent random runs. Herein, we consider only the case of $n_U = 2$, as having several SIC iterations would probably violate the latency constraint of a URLLC service. In U-RSMA, user $\mathbb{U}_{U,1}$ splits its transmission according to α (which is optimized in each simulation step), creating two virtual users, namely $\mathbb{U}_{U,1,1}$ and $\mathbb{U}_{U,1,2}$. Furthermore, users that belong to the same service have the same average SNR, since we consider they are running identical applications. We consider that in each mini-slot there are always two URLLC users connected, i.e., $a_U = 1$ for each one of them, thus $F'_U = F_U/2$. Also, the number of eMBB users is F_B, equaling the number of channels available for the service. Moreover, one TS is composed by $S = 5$ mini-slots and the bandwidth is divided into $F = 8$ channels. The reliability requirement of eMBB service is $\epsilon_B = 10^{-3}$. For URLLC under U-OMA, the reliability is $\epsilon_U^{\text{U-OMA}} = 10^{-5}$, however, as for U-NOMA and U-RSMA the receiver employs SIC, we follow [39] and set the reliability target as $\epsilon_U^{\text{U-NOMA}} = \epsilon_U^{\text{U-RSMA}} = 5 \times 10^{-6}$ to ensure that the overall reliability does not exceed 10^{-5}. Unless stated otherwise, we set $\bar{\Gamma}_U = 20$ dB and $\bar{\Gamma}_B = 10$. Table 2 summarizes the simulation parameters.

In Fig. 2 we plot the sum-rate pair $(r_B^{\text{sum}}, r_U^{\text{sum}})$ for OMA and NOMA network slicing with URLLC operating under U-OMA, U-NOMA, and U-RSMA schemes. Comparing the NOMA slicing curves, U-OMA presents the highest rate pair values until $r_B^{\text{sum}} \approx 7$ bits/s/Hz, from where U-RSMA outperforms the other methods. Interestingly, r_U^{sum} remains almost constant as we increase r_B^{sum} in U-RSMA and U-NOMA, making these methods a good option when we want to achieve greater eMBB rates. In OMA slicing, U-OMA is the best method until $r_B^{\text{sum}} \approx 2.3$ bits/s/Hz, after

TABLE 2. Simulation parameters.

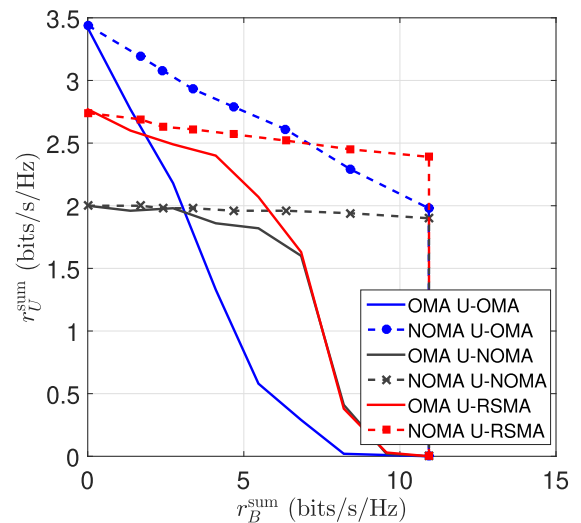| Parameter	Value
Number of URLLC users (n_U)	2
eMBB activation probability (a_B)	1
URLLC activation probability (a_U)	1
Number of channels (F)	8
Number of mini-slots (S)	5
eMBB reliability (ϵ_B)	10^{-3}
URLLC reliability in U-OMA ($\epsilon_U^{\text{U-OMA}}$)	10^{-5}
URLLC reliability in U-NOMA ($\epsilon_U^{\text{U-NOMA}}$)	5×10^{-6}
URLLC reliability in U-RSMA ($\epsilon_U^{\text{U-RSMA}}$)	5×10^{-6}
URLLC average SNR ($\bar{\Gamma}_U$)	20 dB
eMBB average SNR ($\bar{\Gamma}_B$)	10 dB

FIGURE 2. Sum-rate region in OMA and NOMA scenarios with URLLC under U-OMA, U-NOMA, and U-RSMA schemes. $\bar{\Gamma}_U = 20$ dB, $\bar{\Gamma}_B = 10$ dB, $\epsilon_U^{\text{U-OMA}} = 10^{-5}$, $\epsilon_U^{\text{U-NOMA}} = \epsilon_U^{\text{U-RSMA}} = 5 \times 10^{-6}$, $\epsilon_B = 10^{-3}$, $a_U = 1$, $F = 8$, $S = 5$, $n_U = 2$, and optimized α.

that, U-RSMA achieves higher rates, presenting almost the same results of U-NOMA for high r_B^{sum} values.

Fig. 3 shows the URLLC sum-rate for different values of power splitting factor α. Note that, as expected, in U-OMA and U-NOMA we obtain constant values, since there is no message splitting. For U-RSMA, on the other hand, it is possible to observe that, as α increases, r_U^{sum} also increases, reaching the highest value when $\alpha = 0.8$ for NOMA and $\alpha \approx 0.75$ for OMA slicing.

The rates of users $\mathbb{U}_{U,1}$ and $\mathbb{U}_{U,2}$ when operating under U-RSMA are presented in Fig. 4, for both OMA and NOMA slicing. We see that $\mathbb{U}_{U,1}$, the user that performs rate splitting, is capable of reaching higher rates when compared to $\mathbb{U}_{U,2}$. Also, NOMA slicing is the best choice for this setup, achieving higher rates.

We consider that, during one TS, each eMBB user has the same target rate, since the channel gain is constant during this period over all channels. However, for URLLC, not imposing this requirement is beneficial, since different decoding orders provided by U-RSMA enable $\mathbb{U}_{U,1}$ to reach higher rates, contributing to leverages the overall sum-rate, as shown in Figs. 5(c) and 5(d), where we plot the URLLC
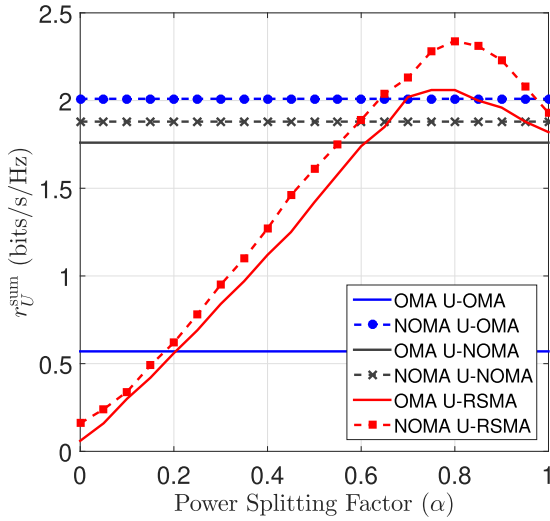
E. J. dos Santos Jr. *et al.*: Rate-Splitting Multiple Access for URLLC Uplink in Physical Layer Network Slicing With eMBB

IEEE *Access*

FIGURE 3. URLLC sum-rate under U-OMA, U-NOMA, and U-RSMA schemes in OMA and NOMA scenarios for $\alpha \in \{0, \ldots, 1\}$, $\bar{\Gamma}_U = 20$ dB, $\bar{\Gamma}_B = 10$ dB, $\epsilon_U^{U\text{-}OMA} = 10^{-5}$, $\epsilon_U^{U\text{-}NOMA} = \epsilon_U^{U\text{-}RSMA} = 5 \times 10^{-6}$, $\epsilon_B = 10^{-3}$, $a_U = 1$, $F = 8$ ($F_U = F_B = 4$ in OMA), $S = 5$, and $n_U = 2$.
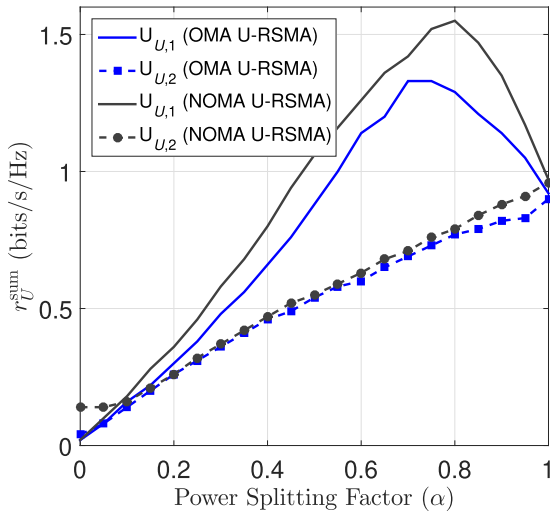

FIGURE 4. URLLC per-user-rate under U-RSMA in OMA and NOMA slicing for $\alpha \in \{0, \ldots, 1\}$, $\bar{\Gamma}_U = 20$ dB, $\bar{\Gamma}_B = 10$ dB, $\epsilon_U^{U\text{-}OMA} = 10^{-5}$, $\epsilon_U^{U\text{-}NOMA} = \epsilon_U^{U\text{-}RSMA} = 5 \times 10^{-6}$, $\epsilon_B = 10^{-3}$, $a_U = 1$, $F = 8$ ($F_U = F_B = 4$ in OMA), $S = 5$, and $n_U = 2$.

(a) OMA sum-rates.

(b) NOMA sum-rates.

(c) OMA per-user rates.

(d) NOMA per-user rates.

FIGURE 5. URLLC sum-rate and per-user rate under U-OMA, U-NOMA, and U-RSMA schemes in OMA and NOMA scenarios for $\bar{\Gamma}_U \in \{0, \ldots, 20\}$ dB, $\bar{\Gamma}_B = 10$ dB, $\epsilon_U^{U\text{-}OMA} = 10^{-5}$, $\epsilon_U^{U\text{-}NOMA} = \epsilon_U^{U\text{-}RSMA} = 5 \times 10^{-6}$, $\epsilon_B = 10^{-3}$, $a_U = 1$, $F = 8$ ($F_U = F_B = 4$ in OMA), $S = 5$, $n_U = 2$, and optimized α.

FIGURE 6. URLLC sum-rate under U-OMA, U-NOMA, and U-RSMA schemes in OMA and NOMA scenario for $F \in \{1, \ldots, 12\}$, $\bar{\Gamma}_U = 20$ dB, $\bar{\Gamma}_B = 10$ dB, $\epsilon_U^{U\text{-}OMA} = 10^{-5}$, $\epsilon_U^{U\text{-}NOMA} = \epsilon_U^{U\text{-}RSMA} = 5 \times 10^{-6}$, $\epsilon_B = 10^{-3}$, $a_U = 1$, $S = 5$, $n_U = 2$, and optimized α.

per-user rate for $\bar{\Gamma}_U \in \{0, \ldots, 20\}$ dB. Comparing U-RSMA and U-NOMA sum-rates in Figs. 5(a) and 5(b), we see that the former is capable of operating with less performance degradation as the SNR increases, due to the fact that it is capable of handling the interference better, while the latter saturates as the SIC procedure fails to eliminate the interference.

From Fig. 6, considering the case of NOMA slicing, we conclude that U-OMA needs more bandwidth to outperform other methods, which is a limiting factor. Moreover, U-RSMA is the better choice for smaller chunks of spectrum, resulting in higher spectral efficiency since we can transmit more data with less bandwidth. In OMA, U-RSMA is better than other methods in all the evaluated range.
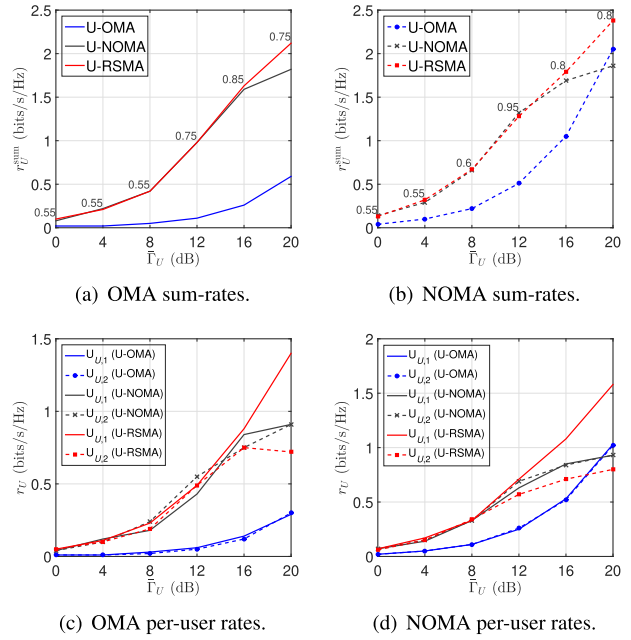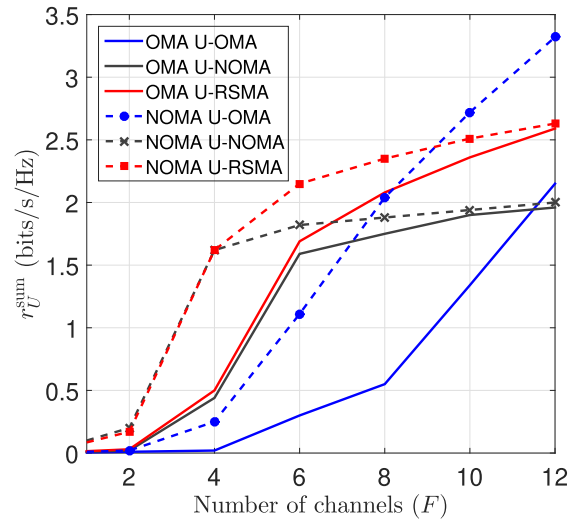
V. FINAL COMMENTS

In this paper, we considered the problem of radio resource slicing between eMBB and multiple URLLC devices. We evaluated the sum-rate performance of three multiple access methods for URLLC, namely U-OMA, U-NOMA, and U-RSMA, when operating under both OMA and NOMA network slicing strategies. Our results show that U-RSMA is capable of achieving higher rates when the power splitting factor is properly configured, even with strict reliability requirements. Moreover, we show that non-orthogonal

IEEE *Access*

E. J. dos Santos Jr. *et al.*: Rate-Splitting Multiple Access for URLLC Uplink in Physical Layer Network Slicing With eMBB

network slicing is capable of reaching the highest pair of rates for URLLC and eMBB simultaneously. This leads us to show another interesting scenario in which combining U-RSMA and NOMA is a powerful tool for attending 6G demands. The practical implementation of RSMA is still evolving, however, for our scenario, limiting the number of users in each mini-slot is a good strategy to reduce the decoding complexity and delay. Also, it is necessary to add a few bits of information in the message intended to the user that splits its data, sent in the synchronization slot to set the power allocation factor based on the average SNR. As some future research topics, the impact of imperfect CSI and applying rate-splitting among users of different services could be investigated.

REFERENCES

[1] H. Zhang, N. Liu, X. Chu, K. Long, A. Aghvami, and V. C. M. Leung, "Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 138–145, Aug. 2017.

[2] GSMA Association. *An Introduction to Network Slicing*. Accessed: 2017. [Online]. Available: https://www.gsma.com/futurenetworks/wp-content/uploads/2017/11/GSMA-An-Introduction-to-Network-Slicing.pdf

[3] X. Li, H. A. Chan, D. Bhamare, L. Gupta, C. Guo, and R. Jain, "Network slicing for 5G: Challenges and opportunities," *IEEE Internet Comput.*, vol. 21, no. 5, pp. 20–27, Sep. 2017.

[4] A. E. Kalor, R. Guillaume, J. J. Nielsen, A. Mueller, and P. Popovski, "Network slicing in Industry 4.0 applications: Abstraction methods and end-to-end analysis," *IEEE Trans. Ind. Informat.*, vol. 14, no. 2, pp. 5419–5427, Dec. 2018.

[5] F. Tariq, M. R. A. Khandaker, K.-K. Wong, M. A. Imran, M. Bennis, and M. Debbah, "A speculative study on 6G," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 118–125, Aug. 2020.

[6] N. H. Mahmood, S. Böcker, A. Munari, F. Clazzer, I. Moerman, K. Mikhaylov, O. Lopez, O. S. Park, E. Mercier, H. Bartz, and R. Jäntti, "White paper on critical and massive machine type communication towards 6G," 2020, *arXiv:2004.14146*.

[7] 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, Oct. 2020.

[8] P. Popovski, J. J. Nielsen, C. Stefanovic, E. De Carvalho, E. Strom, K. F. Trillingsgaard, A. S. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sorensen, "Wireless access for ultra-reliable low-latency communication: Principles and building blocks," *IEEE Netw.*, vol. 32, no. 2, pp. 16–23, Mar./Apr. 2018.

[9] H. Chen, R. Abbas, P. Cheng, M. Shirvanimoghaddam, W. Hardjawana, W. Bao, Y. Li, and B. Vucetic, "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 119–125, Dec. 2018.

[10] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

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

[12] X. Chen, A. Benjebbour, A. Li, and A. Harada, "Multi-user proportional fair scheduling for uplink non-orthogonal multiple access (NOMA)," in *Proc. IEEE 79th Veh. Technol. Conf. (VTC Spring)*, May 2014, pp. 1–5.

[13] A. Rauniyar, P. Engelstad, and O. N. Österbö, "An adaptive user pairing strategy for uplink non-orthogonal multiple access," in *Proc. IEEE 31st Annu. Int. Symp. Pers., Indoor Mobile Radio Commun.*, Aug. 2020, pp. 1–7.

[14] M. A. Sedaghat and R. R. Müller, "On user pairing in uplink NOMA," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3474–3486, May 2018.

[15] I. Azam, M. B. Shahab, and S. Y. Shin, "User pairing and power allocation for capacity maximization in uplink NOMA," in *Proc. 42nd Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2019, pp. 690–694.

[16] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 364–375, Mar. 1996.

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

[18] B. Clerckx, H. Joudeh, C. Hao, M. Dai, and B. Rassouli, "Rate splitting for MIMO wireless networks: A promising PHY-layer strategy for LTE evolution," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 98–105, May 2016.

[19] Y. Mao, B. Clerckx, and V. O. K. Li, "Energy efficiency of rate-splitting multiple access, and performance benefits over SDMA and NOMA," in *Proc. 15th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2018, pp. 1–5.

[20] J. Cao and E. M. Yeh, "Asymptotically optimal multiple-access communication via distributed rate splitting," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 304–319, Jan. 2007.

[21] H. Joudeh and B. Clerckx, "Robust transmission in downlink multiuser MISO systems: A rate-splitting approach," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6227–6242, Dec. 2016.

[22] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847–4861, Nov. 2016.

[23] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting for multi-user multi-antenna wireless information and power transfer," in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2019, pp. 1–5.

[24] B. Clerckx, Y. Mao, R. Schober, and H. V. Poor, "Rate-splitting unifying SDMA, OMA, NOMA, and multicasting in MISO broadcast channel: A simple two-user rate analysis," *IEEE Wireless Commun. Lett.*, vol. 9, no. 3, pp. 349–353, Mar. 2020.

[25] Z. Yang, M. Chen, W. Saad, W. Xu, and M. Shikh-Bahaei, "Sum-rate maximization of uplink rate splitting multiple access (RSMA) communication," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[26] Z. Yang, M. Chen, W. Saad, W. Xu, and M. Shikh-Bahaei, "Sum-rate maximization of uplink rate splitting multiple access (RSMA) communication," *IEEE Trans. Mobile Comput.*, early access, Nov. 16, 2020, doi: 10.1109/TMC.2020.3037374.

[27] Y. Zhu, X. Wang, Z. Zhang, X. Chen, and Y. Chen, "A rate-splitting non-orthogonal multiple access scheme for uplink transmission," in *Proc. 9th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2017, pp. 1–6.

[28] H. Liu and K. S. Kwak, "Adaptive rate splitting for uplink non-orthogonal multiple access systems," in *Proc. 11th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Jul. 2019, pp. 158–163.

[29] H. Liu, T. A. Tsiftsis, K. J. Kim, K. S. Kwak, and H. V. Poor, "Rate splitting for uplink NOMA with enhanced fairness and outage performance," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4657–4670, Jul. 2020.

[30] J. Zeng, T. Lv, W. Ni, R. P. Liu, N. C. Beaulieu, and Y. J. Guo, "Ensuring max–min fairness of UL SIMO-NOMA: A rate splitting approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11080–11093, Nov. 2019.

[31] O. Abbasi and H. Yanikomeroglu, "Rate-splitting and NOMA-enabled uplink user cooperation," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Mar. 2021, pp. 1–6.

[32] 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. 55765–55779, 2018.

[33] A. Anand, G. De Veciana, and S. Shakkottai, "Joint scheduling of URLLC and eMBB traffic in 5G wireless networks," in *Proc. IEEE Conf. Comput. Commun. (IEEE INFOCOM)*, Apr. 2018, pp. 1970–1978.

[34] 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 *Proc. IEEE 89th Veh. Technol. Conf. (VTC-Spring)*, Apr. 2019, pp. 1–6.

[35] M. Alsenwi, N. H. Tran, M. Bennis, A. K. Bairagi, and C. S. Hong, "EMBB-URLLC resource slicing: A risk-sensitive approach," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 740–743, Apr. 2019.

[36] 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 *Proc. IEEE 24th Int. Workshop Comput. Aided Modeling Design Commun. Links Netw. (CAMAD)*, Sep. 2019, pp. 1–5.

[37] 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. 13035–13049, 2019.
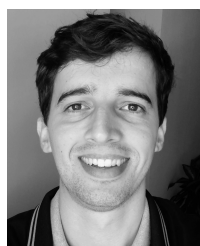
E. J. dos Santos Jr. *et al.*: Rate-Splitting Multiple Access for URLLC Uplink in Physical Layer Network Slicing With eMBB

IEEE *Access*

[38] E. N. Tominaga, H. Alves, R. D. Souza, J. L. Rebelatto, and M. Latva-Aho, "Non-orthogonal multiple access and network slicing: Scalable coexistence of eMBB and URLLC," in *Proc. IEEE 93rd Veh. Technol. Conf. (VTC-Spring)*, Apr. 2021, pp. 1–6.

[39] O. Dizdar, Y. Mao, Y. Xu, P. Zhu, and B. Clerckx, "Rate-splitting multiple access for enhanced URLLC and eMBB in 6G: Invited paper," in *Proc. 17th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Sep. 2021, pp. 1–6.

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

[41] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 124–130, Jun. 2018.

[42] K. Takeda, H. Harada, and R. Osawa. (Accessed: Sep. 2021). *NR Physical Layer Specifications in 5G*. Accessed: 2019. [Online]. Available: https://www.nttdocomo.co.jp/english/binary/pdf/corporate/technology/rd/technical_journal/bn/vol20_3/vol20_3_007en.pdf

[43] A. Zaidi and Z. Wang. (Accessed: Sep. 2021). *Designing for the Future the 5G NR Physical Layer*. Accessed: 2017. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/designing-for-the-future-the-5g-nr-physical-layer

[44] 3GPP. (Accessed: Sep. 2021). *5G, NR, Physical Layer Procedures for Data*. Accessed: 2018. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138200_138299/138214/15.02.00_60/ts_138214v150200p.pdf

[45] A. A. Zaidi, R. Baldemair, M. Andersson, S. Faxér, V. Molés-Cases, and Z. Wang, "Designing for the future: The 5G NR physical layer," Ericsson Technol. Rev. Articles, Stockholm, Sweden, Tech. Rep., Jul. 2017. [Online]. Available: https://www.ericsson.com/en/ericsson-technology-review/archive/2017/designing-for-the-future-the-5g-nr-physical-layer

[46] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, Jun. 2014.

[47] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1468–1489, Jul. 1999.

[48] H. Liu, T. A. Tsiftsis, K. J. Kim, K. S. Kwak, and H. V. Poor, "Rate splitting for uplink NOMA with enhanced fairness and outage performance," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4657–4670, Jul. 2020.

[49] Z. Yang, M. Chen, W. Saad, W. Xu, and M. Shikh-Bahaei, "Sum-rate maximization of uplink rate splitting multiple access (RSMA) communication," *IEEE Trans. Mobile Comput.*, early access, Nov. 16, 2020, doi: 10.1109/TMC.2020.3037374.

RICHARD DEMO SOUZA (Senior Member, IEEE) received the D.Sc. degree in electrical engineering from the Federal University of Santa Catarina (UFSC), Brazil, in 2003. From 2004 to 2016, he was with the Federal University of Technology—Paraná (UTFPR), Brazil. Since 2017, he has been with UFSC, where he is a Professor. His research interests include in the areas of wireless communications and signal processing. He was a co-recipient of the 2014 IEEE/IFIP Wireless Days Conference Best Paper Award, the supervisor of the awarded Best Ph.D. Thesis in Electrical Engineering in Brazil, in 2014, and a co-recipient of the 2016 Research Award from the Cuban Academy of Sciences. He has served as an Editor or an Associate Editor for the *Journal of Communication and Information Systems* (SBrT), the IEEE COMMUNICATIONS LETTERS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and the IEEE INTERNET OF THINGS journal.

ELÇO JOÃO DOS SANTOS JR. was born in Florianópolis, State of Santa Catarina, Brazil, in 1994. He received the B.Sc. degree in electronics engineering and the M.Sc. degree in electrical engineering from the Federal University of Santa Catarina (UFSC), Brazil, in 2019 and 2020, respectively. He is currently a Researcher at the SENAI Institute of Innovation in Embedded Systems (ISI-SE) and a Ph.D. student at UFSC. His research interests include wireless communications, multiple-access strategies, URLLC, random access protocols, and signal processing.

JOÃO LUIZ REBELATTO (Senior Member, IEEE) was born in Lapa, State of Paraná, Brazil. He received the B.Sc. degree in electrical engineering from the Federal University of Technology—Paraná (UTFPR), Curitiba, Brazil, in 2006, and the D.Sc. degree in electrical engineering from the Federal University of Santa Catarina (UFSC), Florianópolis, Brazil, in 2010. From 2009 to 2010, he was a visiting Ph.D. student with The University of Sydney, NSW, Australia. From 2011 to 2012, he held a postdoctoral position at UFSC. Since 2011, he has been with the Department of Electronics, UTFPR, where he is currently an Associate Professor. His research interests include in the area of coding and information theory, with applications to wireless communications systems. He was a co-recipient of the 2011 Brazilian Symposium on Telecommunications (SBrT) Best Paper Award, the 2014 IEEE/IFIP Wireless Days Conference Best Paper Award, and the 2016 Research Award from the Cuban Academy of Sciences.

● ● ●