Aalborg University

DENMARK

5G Wireless Network Slicing for eMBB, URLLC, and mMTC

A Communication-Theoretic View

Popovski, Petar; Trillingsgaard, Kasper Fløe; Simeone, Osvaldo; Durisi, Giuseppe

5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View

PETAR POPOVSKI[1], **(Fellow, IEEE),**
KASPER FLØE TRILLINGSGAARD[1], **(Student Member, IEEE),**
OSVALDO SIMEONE[2], **(Fellow, IEEE),**
AND GIUSEPPE DURISI[3], **(Senior Member, IEEE)**

[1]Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark
[2]King's College London, London WC2R 2LS, U.K.
[3]Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden

Corresponding author: Petar Popovski (petarp@es.aau.dk)

ABSTRACT The grand objective of 5G wireless technology is to support three generic services with vastly heterogeneous requirements: enhanced mobile broadband (eMBB), massive machine-type communications (mMTCs), and ultra-reliable low-latency communications (URLLCs). Service heterogeneity can be accommodated by network slicing, through which each service is allocated resources to provide performance guarantees and isolation from the other services. Slicing of the radio access network (RAN) is typically done by means of orthogonal resource allocation among the services. This paper studies the potential advantages of allowing for non-orthogonal sharing of RAN resources in uplink communications from a set of eMBB, mMTC, and URLLC devices to a common base station. The approach is referred to as heterogeneous non-orthogonal multiple access (H-NOMA), in contrast to the conventional NOMA techniques that involve users with homogeneous requirements and hence can be investigated through a standard multiple access channel. The study devises a communication-theoretic model that accounts for the heterogeneous requirements and characteristics of the three services. The concept of *reliability diversity* is introduced as a design principle that leverages the different reliability requirements across the services in order to ensure performance guarantees with non-orthogonal RAN slicing. This paper reveals that H-NOMA can lead, in some regimes, to significant gains in terms of performance tradeoffs among the three generic services as compared to orthogonal slicing.

INDEX TERMS 5G mobile communication, machine-to-machine communications, multiaccess communication, NOMA, wireless communication.

I. INTRODUCTION

During the past few years, there has been a growing consensus that 5G wireless systems will support three generic services, which, according ITU-R, are classified as enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultra-reliable and low-latency communications (URLLC) (also referred to as mission-critical communications) [1], [2]. A succinct characterization of these services can be put forward as follows: (*a*) eMBB supports stable connections with very high peak data rates, as well as moderate rates for cell-edge users; (*b*) mMTC supports a massive number of Internet of Things (IoT) devices,

which are only sporadically active and send small data payloads; (*c*) URLLC supports low-latency transmissions of small payloads with very high reliability from a limited set of terminals, which are active according to patterns typically specified by outside events, such as alarms. This paper studies the problem of enabling the coexistence of the three heterogeneous services within the same Radio Access Network (RAN) architecture. We describe below in more details the requirements of the three services.

eMBB traffic can be considered to be a direct extension of the 4G broadband service. It is characterized by large payloads and by a device activation pattern that remains stable

over an extended time interval. This allows the network to schedule wireless resources to the eMBB devices such that no two eMBB devices access the same resource simultaneously. The objective of the eMBB service is to maximize the data rate, while guaranteeing a moderate reliability, with packet error rate (PER) on the order of 10^{-3}.

In contrast, an mMTC device is active intermittently and uses a fixed, typically low, transmission rate in the uplink. A huge number of mMTC devices may be connected to a given base station (BS), but at a given time only an unknown (random) subset of them becomes active and attempt to send their data. The large number of potentially active mMTC devices makes it infeasible to allocate *a priori* resources to individual mMTC devices. Instead, it is necessary to provide resources that can be shared through random access. The size of the active subset of mMTC devices is a random variable, whose average value measures the mMTC traffic arrival rate. The objective in the design of mMTC is to maximize the arrival rate that can be supported in a given radio resource. The targeted PER of an individual mMTC transmission is typically low, e.g., on the order of 10^{-1}.

Finally, URLLC transmissions are also intermittent, but the set of potential URLLC transmitters is much smaller than for mMTC. Supporting intermittent URLLC transmissions requires a combination of scheduling, so as to ensure a certain amount of predictability in the available resources and thus support high reliability; as well as random access, in order to avoid that too many resources being idle due to the intermittent traffic. Due to the low latency requirements, a URLLC transmission should be localized in time. Diversity, which is critical to achieve high reliability [3], can hence be achieved only using multiple frequency or spatial resources. The rate of a URLLC transmission is relatively low, and the main requirement is ensuring a high reliability level, with a PER typically lower than 10^{-5}, despite the small blocklengths.

In 5G, heterogeneous services are allowed to coexist within the same network architecture by means of *network slicing* [4]. Network slicing allocates the network computing, storage, and communication resources among the active services with the aim of guaranteeing their isolation and given performance levels. In this paper, we are interested in the "slicing" of RAN communication resources for wireless access. The conventional approach to slice the RAN is to allocate orthogonal radio resources to eMBB, mMTC, and URLLC devices in time and/or frequency domains, consistently with the orthogonal allocation of wired communication resources. However, wireless resources are essentially different due to their shared nature. Using communication-theoretic analysis, this work demonstrates that a non-orthogonal allocation that is informed by the heterogeneous requirements of the three services can outperform the standard orthogonal approach, while still offering per-service guarantees. Importantly, the considered non-orthogonal approach multiplexes heterogeneous services, and is hence markedly distinct from the conventional Non-Orthogonal Multiple Access (NOMA) methods
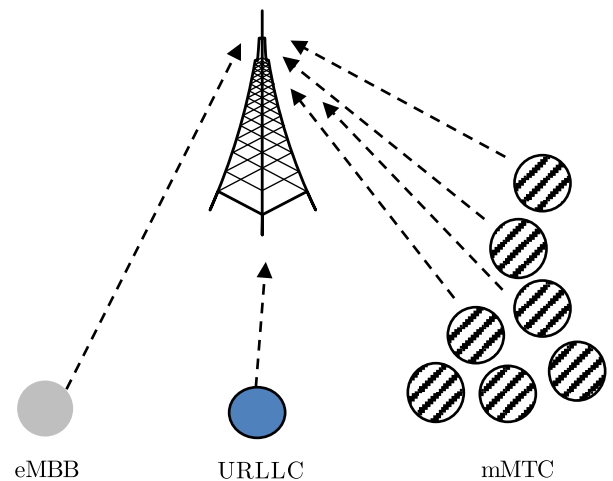

FIGURE 1. The considered scenario with uplink transmissions to a common base station (BS) from devices using the three generic 5G services.

that share radio resources only among devices of the same type (see, e.g., [5]). This is further discussed next.

A. NETWORK SLICING OF WIRELESS RESOURCES: H-OMA AND H-NOMA

Consider an uplink scenario in which a set of eMBB, mMTC and URLLC devices is connected to a common BS, as shown in Fig. 1. We note that the designing uplink access is more complex than the corresponding problem for the downlink due to the lack of coordination among users. Orthogonal and non-orthogonal slicing of the RAN among the three services are illustrated in Fig. 2(a) and (b), respectively.

The conventional orthogonal allocation depicted in Fig. 2(a) operates in the frequency domain and allots different frequency channels to eMBB, mMTC, or URLLC devices. eMBB and mMTC transmissions are allowed to span multiple time resources. In contrast, in order to guarantee the latency requirements discussed above, URLLC transmissions are localized in time and are spread over multiple frequency channels to gain diversity. Furthermore, since the URLLC traffic is bursty, the resources allocated to URLLC users may be largely unused. This is because the channels reserved for URLLC are idle in the absence of URLLC transmission.

Importantly, orthogonal slicing does not preclude the sharing of wireless resources among devices of the same type. For example, multiple eMBB users may share the allotted frequency channels in an orthogonal way, which would represent a conventional Orthogonal Multiple Access (OMA). Alternatively, they may transmit on the same frequency channels simultaneously by using NOMA [5]. Therefore, in order to distinguish orthogonality among signals originating from devices of the same type from the orthogonality among different services, we refer to the approach in Fig. 2(a) as *Heterogeneous Orthogonal Multiple Access (H-OMA)*.

As mentioned, in this work, we investigate the potential advantages of a non-orthogonal allocation of RAN resources among multiple services, which we refer to as *Het-*
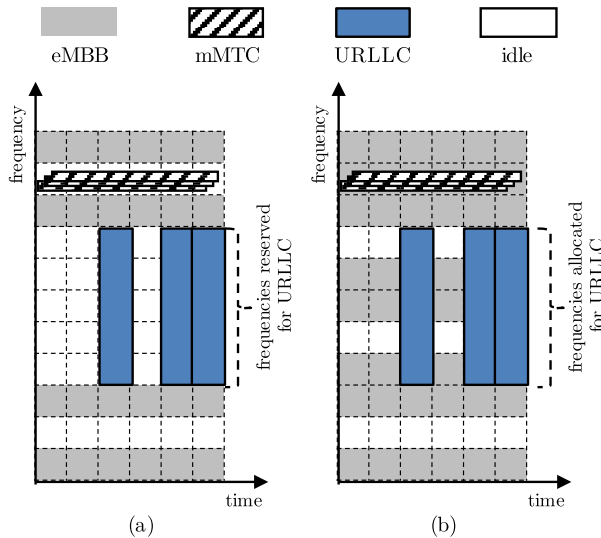
FIGURE 2. Illustration of the slicing of the wireless resources in a time-frequency frame for supporting the three generic services with: (a) Heterogeneous Orthogonal Multiple Access (H-OMA) (b) Heterogeneous Non-Orthogonal Multiple Access (H-NOMA). The idle time-frequency blocks are not used for transmission due to absence of traffic. With H-OMA, some of the frequency channels are reserved to URLLC traffic, whereas with H-NOMA the same channels are allocated to both URLLC and eMBB.

erogeneous *Non-Orthogonal Multiple Access (H-NOMA)*. Fig. 2(b) depicts an instance of H-NOMA. By comparison with the H-OMA solution in Fig. 2(a), under H-NOMA, the frequency resources that were allocated only to mMTC or URLLC traffic can also be granted to the eMBB users. In this way, H-NOMA may allow for a more efficient use of radio resources as compared to H-OMA by avoiding unused resources due to URLLC or mMTC inactivity. This may yield a higher spectral efficiency for the eMBB users that can benefit from the intermittent nature of mMTC and URLLC traffic. However, the mutual interference between eMBB and mMTC or URLLC transmissions may significantly degrade the performance for all the involved services. Ensuring desired performance levels is hence more challenging with H-NOMA.

In this paper, we tackle this problem by developing a communication-theoretic model that aims at capturing the essential performance trade-offs and design insights for H-OMA and H-NOMA. More specifically, the main goals of this work can be illustrated using Fig. 3, as discussed next.

To start, Fig. 3(a) depicts the type of results that are of interest when studying conventional OMA and NOMA within a given service type, as done in a growing line of work [5], [6]. These results rely on the classical analysis of the multiple access channel, in which all users have identical reliability requirements and block lengths and the goal is to characterize the region of achievable rates [7] as the block length grows large.

In contrast, Fig. 3-(b,c,d) exemplify the type of results that are of interest when evaluating the performance trade-offs between heterogeneous services that are allowed by H-OMA and H-NOMA. As a first example to be further

FIGURE 3. Illustration of performance trade-offs for: (a) standard OMA and NOMA within the same traffic type; (b,c,d) H-OMA and H-NOMA between heterogeneous services.

elaborated on in the paper, Fig. 3(b) shows the trade-off between the URLLC activity, i.e., the probability of URLLC devices being active, and the eMBB transmission rate or spectral efficiency. The figure illustrates the fact that the eMBB rate is not affected by the URLLC packet arrival rate under H-OMA, while the resulting interference impairs the performance of H-NOMA. As an alternative performance evaluation, Fig. 3(c) shows the trade-off between the reliability of URLLC transmissions and the eMBB rate. The example highlights the fact that a non-trivial trade-off exists for both H-OMA and H-NOMA. In H-OMA, the increase in eMBB rate implies that more orthogonal resources are allocated to eMBB at the expense of URLLC, which decreases the reliability of URLLC. As a final illustration, Fig. 3(d) depicts the trade-off between the arrival rate of mMTC devices and the eMBB rate. In a manner similar to Fig. 3(c), this figure suggests that, even under H-OMA, the spectral efficiency of the eMBB user that shares the resources with mMTC devices can be traded off for the mMTC arrival rate by a proper allocation of radio resources.

B. FURTHER RELATED WORKS

In addition to the mentioned literature on conventional NOMA, here we briefly review works that directly tackle the coexistence of heterogeneous services. A logical architecture for network slicing in 5G in the presence of

orthogonal slicing has been presented in [4] and [8]. The downlink multiplexing of URLLC and eMBB is studied in [9] and [10]. These works investigates the dynamic scheduling of URLLC traffic over ongoing eMBB transmissions by abstracting the operation at the physical layer. Lien *et al.* [11] treat the problem of resource allocation for mMTC and URLLC in a new radio (NR) setting by focusing on the role of feedback. Orthogonal resource allocation for mMTC and eMBB users is studied in [12] by accounting for inter-cell interference. In [13], grant-free uplink transmissions are considered for the three services by considering concrete transmission/modulation/spreading methods for supporting the three services.

C. MAIN CONTRIBUTIONS
The main contributions are as follows.

- We propose a communication-theoretic model that is tractable and yet captures the key features and requirements of the three services. Unlike [13], in which the authors focus on grant-free access for all services, the proposed model takes into account the difference in arrival processes and traffic dynamics that are inherent to each individual service. The proposed model can be seen as an extension of the classical multiple access channel model that underlies the analysis of conventional NOMA in the sense that it accounts for the coexistence of heterogeneous services.
- We first analyze the performance of orthogonal slicing, or H-OMA, for all three services. We focus on achievable transmission rates for eMBB and URLLC, under the respective target reliability, and on the throughput for mMTC.
- We then consider the performance of H-NOMA. Although the modeling approach allows to study an arbitrary combination of services, in this paper we have focused on the analysis of two specific cases as illustrated in Fig. 2, namely: (*i*) slicing for URLLC and eMBB, and (*ii*) slicing for mMTC and eMBB. In the case of URLLC-eMBB slicing, among other schemes, we consider the technique of puncturing, which is currently under consideration in 3GPP [10]. It is noted that, while of interest, H-NOMA between URLLC and mMTC may be problematic due to the need to ensure reliability guarantees for URLLC devices in the presence of the random interference patterns caused by mMTC transmissions.
- Among the main conclusions, our study demonstates that non-orthogonal slicing, or H-NOMA, can achieve service isolation in the sense of ensuring performance levels for all services by leveraging their heterogeneous reliability requirements. We refer to this design principle as *reliability diversity*. As it will be discussed, the heterogeneity leveraged by reliability diversity is not only in terms of the numerical values of the reliability levels, but also in terms of very definition of reliability

across the three services. For example, the reliability metric typically considered for mMTC is the fraction of detected devices among the massive set of active users, whereas for eMBB and URLLC services one typically adopts the classical frame error rate. Our results show that, if reliability diversity is properly exploited, non-orthogonal slicing can lead, in some regimes, to important gains in terms of performance trade-offs among the three generic services.

The paper is organized as follows. The next section presents the system model and provides a performance analysis of each of the three services when considered in isolation. Section III treats the slicing of resources to support eMBB and URLLC, while Section IV is dedicated to the slicing of resources for eMBB and mMTC. Both sections provide a description of the proposed theoretical framework as well as numerical results illustrating the tradeoff between the services for both H-OMA and H-NOMA schemes. The conclusions are given in Section V-A, while Section V-B contains discussion on possible generalizations of the model considered in this paper. Two appendices, containing the technical details of some of the derivations, conclude the paper.

II. SYSTEM MODEL
We are interested in understanding how the three service described in Section I, i.e., eMBB, URLLC, and mMTC, should efficiently share the same radio resources in the uplink when communicating to a common BS. We consider F radio resources, where each resource occupies a single frequency channel and a single time slot. A radio resource, which is indexed by $f \in \{1, \ldots, F\}$, contains n symbols. The n symbols are further divided into S minislots, where each minislot consists of $n_S = n/S$ symbols. Fig. 4 shows an example of a time-frequency grid.

We assume that the transmission of an eMBB user occupies a single radio resource at a given frequency $f \in \{1, \ldots, F\}$. In contrast, due to latency constraints, a URLLC user transmits within a single minislot across a subset of $F_U \leq F$ frequency channels. An URLLC device may be active in an allocated minislot with probability a_U. Finally, the set of mMTC users is allowed to access the channel only at a specified radio resource f_M; for the example on Fig. 4 we have $f_M = F$. The number A_M of active mMTC devices in such a resource is distributed as $A_M \sim \text{Poisson}(\lambda_M)$, where λ_M is the mean value, referred to as mMTC arrival rate.

Some comments regarding the modelling choices made above are in order. First, for eMBB traffic, we focus on the standard scheduled transmission phase, hence assuming that radio access and competition among eMBB devices have been resolved prior to the considered time slot. Second, we do not model collisions among URLLC devices. We assume instead that a single URLLC device is allocated a number of minislots in the given slot, over which it is active with some probability. On the contrary, we do model the random access
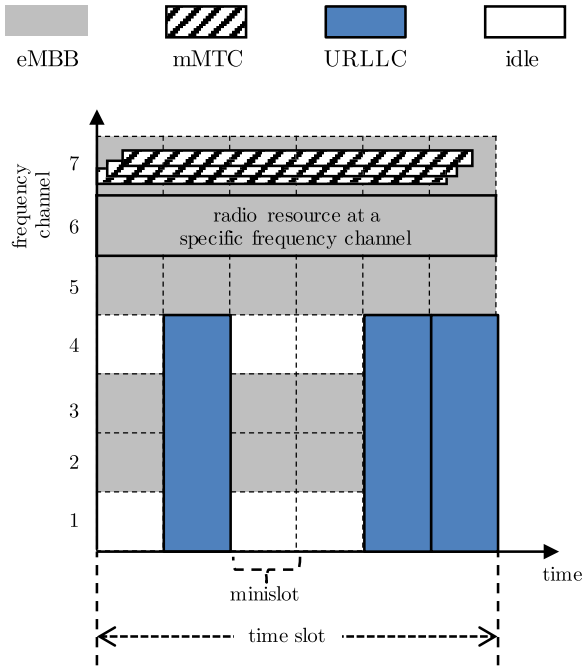
FIGURE 4. An example of H-NOMA allocation in the time-frequency grid with $F = 7$ resources and $S = 6$ minislots. A single resource (frequency channel) is allocated for mMTC transmission. Each URLLC transmission is spread over $F_U = 4$ frequency channels.

phase for mMTC traffic, since this is the key transmission phase for this type of traffic, due to the massive population of devices. Extensions of our model will be discussed in Section V-B.

Each radio resource f 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 radio resource. Furthermore, we assume that the channel coefficients fade independently across the F radio resources. The channel coefficients of the eMBB, URLLC, and the mMTC devices, which we denote by $H_{B,f}$, $H_{U,f}$, and $H_{m,f}$, $m \in \{0, \ldots, A_M\}$,[1] are independent and Rayleigh distributed, i.e., $H_{B,f} \sim \mathcal{CN}(0, \Gamma_M)$, $H_{U,f} \sim \mathcal{CN}(0, \Gamma_U)$, and $H_{m,f} \sim \mathcal{CN}(0, \Gamma_M)$ for $m \in \{0, \ldots, A_M\}$ across all radio resources $f \in 1, \ldots, F$. The channel gains for the three services in a radio resource f are denoted by $G_{B,f} = |H_{B,f}|^2$, $G_{U,f} = |H_{U,f}|^2$, and $G_{m,f} = |H_{m,f}|^2$ for $m \in 1, \ldots, A_M$.

The average transmission power of all devices is normalized to one. The differences in the actual transmission power across various users and in the path loss are accounted for through the average channel gains Γ_B, Γ_U, and Γ_M. Furthermore, the power of the noise at the BS is also normalized to one, so that the received power equals the signal-to-noise ratio (SNR) for each device. The number of symbols n_S in a minislot is assumed sufficiently large to justify an asymptotic information-theoretic analysis. Extensions of our analysis to capture finite-blocklength effects [14] will be considered

---

[1]Throughout, we use the convention that the subscripts B, U, and M indicate a quantity referring to eMBB, URLLC, and mMTC, respectively.

in future works. Due to latency and protocol constraints to be detailed later, no channel-state information (CSI) is assumed at the URLLC and at the mMTC devices. In contrast, the eMBB devices are assumed to have perfect CSI. Finally, the BS is assumed to have perfect CSI.

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

$$\epsilon_U \ll \epsilon_B \ll \epsilon_M. \tag{1}$$

The large differences in reliability levels among the services, as well as their different definitions, which we will introduce shortly, motivate the introduction of the concept of *reliability diversity*. Reliability diversity refers to system design choices that leverage the differences among the supported services in terms of reliability requirements and definitions. For example, as we will see, strict per-packet reliability guarantees are typically enforced for eMBB and URLLC devices, whereas the notion of reliability for mMTC devices is less stringent and typically involves the computation of averages over a large group of active devices.

A. SIGNAL MODEL
To summarize the main assumptions discussed so far and to fix the notation, we assume that each eMBB user is scheduled on a single frequency channel within the considered F frequency resources; each URLLC device occupies $F_U \leq F$ frequencies resources, numbered without loss of generality as $f = 1, \ldots, F_U$, in a given minislot; and a set of mMTC devices is available for transmission in a channel frequency $f_M \in \{1, \ldots, F\}$.

Let $\mathbf{Y}_{s,f} \in \mathbb{C}^{n_S}$ denote the received vector corresponding to the minislot $s \in \{1, 2, \ldots, S\}$ and the frequency channel $f \in \{1, 2, \ldots F\}$. Based on the given assumptions, the received signal can be written as

$$\mathbf{Y}_{s,f} = H_{B,f}\mathbf{X}_{B,s,f} + H_{U,f}\mathbf{X}_{U,s,f}$$
$$+ \sum_{m=1}^{A_M} H_{[m],f}\mathbf{X}_{[m],s,f} + \mathbf{Z}_{s,f}, \tag{2}$$

where $\mathbf{X}_{B,s,f}$ is the signal transmitted by an eMBB user scheduled in the frequency resource f; $\mathbf{X}_{U,s,f}$ is the signal transmitted by a URLLC device transmitting in minislot s and frequency f; $\mathbf{X}_{[m],s,f}$ is the signal transmitted by one of the A_M active mMTC devices in frequency f; and $\mathbf{Z}_{s,f}$ represents the noise vector, whose entries are i.i.d. Gaussian with zero mean and unit variance. The notation $[m]$ for the mMTC devices, which indicates ordering, will be formally introduced in Section II-D.

We emphasize that the transmitted eMBB signal $\mathbf{X}_{B,s,f}$ in (2) is zero if no eMBB user is scheduled in frequency channel f; similarly, the URLLC signal $\mathbf{X}_{U,s,f}$ is zero if no URLLC device transmits in minislot s and frequency f, e.g., if $f > F_U$; and the mMTC signals $\{\mathbf{X}_{[m],s,f}\}$ are similarly all

equal to zero if the channel f is not allocated to mMTC traffic, i.e., if $f \neq f_M$.

As discussed, with H-OMA, resources are allocated exclusively to one of the three services, while, with H-NOMA, resources can be shared. In the remainder of this section, we study the performance of the three traffic types in an H-OMA setting, that is, in the absence of mutual interference. We also introduce the metrics that will be used to evaluate the performance of the three services.

B. EMBB

Consider a radio resource f allocated exclusively to an eMBB user. As mentioned, the eMBB is aware of the CSI $G_{B,f}$ and can use it in order to select its transmission power $P_B(G_{B,f})$. The objective is to transmit at the largest rate $r_{B,f}$ that is compatible with the outage probability requirement ϵ_B under a long-term average power constraint. This can be formulated as the optimization problem

$$
\begin{aligned}
&\text{maximize } r_B \\
&\text{subject to } \Pr\big[\log_2\big(1 + G_{B,f} P_B(G_{B,f})\big) < r_B\big] \le \epsilon_B \\
&\qquad\qquad \mathbb{E}\big[P_B(G_{B,f})\big] = 1.
\end{aligned}
\tag{3}
$$

The optimal solution to this problem is given by truncated power inversion [15]. Accordingly, the eMBB device chooses a transmission power that is inversely proportional to the channel gain $G_{B,f}$ if the latter is above a given threshold $G_{B,f}^{\min}$, while it refrains from transmitting otherwise.

Beside being theoretically justified by the mentioned rate-maximization problem, the threshold-based transmission strategy discussed above also captures the fact that eMBB devices only transmit if the current SNR is sufficient to satisfy minimal rate requirements. This is the case in most communication standards, such as LTE, in which the transmission mode is selected from a set of allowed modulation and coding schemes with given SNR constraints. As we will discuss below, the scheme has the additional analytical advantage of relating directly outage probability and probability of activation for an eMBB user. We remark that the analysis could be extended to other design criteria such as the maximization of the average transmission rate.

Based on the discussion above, the probability that the eMBB user transmits is given by

$$
a_B = \Pr\big[G_{B,f} \ge G_{B,f}^{\min}\big] = e^{-G_{B,f}^{\min}/\Gamma_B}.
\tag{4}
$$

Furthermore, in the absence of interference from other services, the only source of outage for an eMBB transmission is precisely the event that an eMBB does not transmit because of an insufficient SNR level. Hence, the probability of error equals

$$
\Pr(E_B) = 1 - a_B.
\tag{5}
$$

Imposing the reliability condition

$$
\Pr(E_B) = \epsilon_B
\tag{6}
$$

we obtain the value of the threshold SNR

$$
G_{B,f}^{\min} = \Gamma_B \ln\left(\frac{1}{1 - \epsilon_B}\right).
\tag{7}
$$

Note that, in the absence of interference and under the given assumptions, the threshold SNR $G_{B,f}^{\min}$ does not depend on the frequency channel f. This dependence is kept here in view to the extension to H-NOMA in the next sections.

Based on the power-inversion scheme, the instantaneous power $P_B(G_{B,f})$ is chosen as a function of the instantaneous channel gain $G_{B,f}$ as

$$
P_B(G_{B,f}) =
\begin{cases}
\dfrac{G_{B,f}^{\text{tar}}}{G_{B,f}} & \text{if } G_{B,f} \ge G_{B,f}^{\min} \\
0 & \text{if } G_{B,f} < G_{B,f}^{\min},
\end{cases}
\tag{8}
$$

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

$$
\begin{aligned}
1 = \mathbb{E}\big[P_B(G_{B,f})\big] &= \int_{G_{B,f}^{\min}}^{\infty} \frac{1}{\Gamma_B} e^{-x/\Gamma_B} P_B(x)\,\mathrm{d}x \\
&= \frac{G_{B,f}^{\text{tar}}}{\Gamma_B} \gamma\left(0, \frac{G_{B,f}^{\min}}{\Gamma_B}\right),
\end{aligned}
\tag{9}
$$

with $\gamma(\cdot, \cdot)$ being the lower incomplete gamma function. This implies that the target SNR is

$$
G_{B,f}^{\text{tar}} = \frac{\Gamma_B}{\gamma\left(0, \frac{G_{B,f}^{\min}}{\Gamma_B}\right)}.
\tag{10}
$$

It follows from (4)–(10) that the solution to the problem (3), which is the outage rate $r_{B,f}$ under outage probability ϵ_B, is given by

$$
r_{B,f} = \log_2\big(1 + G_{B,f}^{\text{tar}}\big), \quad \text{[bits/symbol]}.
\tag{11}
$$

We refer to the resulting rate as r_B^{orth} for reference. Note that it does not depend on f under the assumptions of this section.

C. URLLC

The URLLC device transmits data in the allocated F_U frequency channels of a minislot, with activation probability a_U. Hence, the number of URLLC transmissions during the time slot is a random variable $S_U \sim \text{Bin}(S, a_U)$. We assume that each URLLC transmission carries a different message, and that, due to the low latency requirement, each message must be decoded as soon as the relevant minislot is received. This implies that the URLLC device cannot code across multiple minislots.

Unlike eMBB users, the URLLC device is not aware of the CSI $\{G_{U,f}\}$ for the F_U allocated frequency resources. This assumption is justified by the fact that CSI at the URLLC device would require signaling exchange before transmission, which entails extra latency as well as a potential loss in terms of reliability. In fact, the high reliability constraint would enforce an even higher reliability requirement on the auxiliary procedure of CSI signaling. As a result of the lack of CSI, no power or rate adaptation is possible for URLLC devices.

We choose the rate r_U as the performance metric of choice. In the absence of interference from other services, outage occurs with probability

$$\Pr(E_U) = \Pr\left(\frac{1}{F_U}\sum_{f=1}^{F_U}\log_2\left(1 + G_{U,f}\right) < r_U\right). \quad (12)$$

Imposing the reliability condition $P(E_U) = \epsilon_U$ allows us to obtain the maximum allowed rate r_U. We will refer to this quantity as $r_U^{\text{orth}}(F_U)$ for reference. Note that 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.

D. MMTC

The key property of the mMTC traffic is that the set of mMTC devices that transmit in a given radio resource is random and unknown. An mMTC transmission has a fixed rate r_M and consumes one radio resource of n channel uses. Given the rate r_M and the reliability constraint ϵ_M, we focus on the maximum arrival rate λ_M that can be supported by the system as the performance criterion of interest. As detailed below, the probability of error measures the fraction of incorrectly decoded devices among the active ones.

SIC at the BS is a useful strategy to improve the performance of mMTC traffic. As discussed next, a SIC decoder can leverage power imbalances and other mechanisms not reviewed here (see, e.g., [16]), in order to sequentially improve the reliability of simultaneous mMTC transmissions.

To characterize the performance achievable with SIC, we let $[m]$ denote the index of the mMTC device with the m-th largest channel gain $\{G_{m,f_M}\}$ for the allocated frequency $f = f_M$. In the rest of this section, we drop the dependence on f_M for simplicity of notation. By definition, we then have the inequalities $G_{[1]} \geq G_{[2]} \geq \ldots \geq G_{[A_M]}$. In the absence of interference from eMBB and URLLC traffic, the SINR $\sigma_{[m_0]}$ available when decoding the signal of the m_0−th mMTC device, under the additional assumption that the devices with indices $[1], \ldots, [m_0 - 1]$ are correctly decoded, depends only on its channel gain $G_{[m_0]}$ and on the channels gains of the other active mMTC devices as

$$\sigma_{[m_0]} = \frac{G_{[m_0]}}{1 + \sum_{m=m_0+1}^{A_M} G_{[m]}}. \quad (13)$$

The m_0−th mMTC device is correctly decoded if the inequality $\log_2(1+\sigma_{[m_0]}) \geq r_M$ holds; and, if decoding is successful, the signal from the device is subtracted from the received signal. We let D_M be the random number of mMTC devices in outage, i.e., D_M is the largest integer in $\{0, \ldots, A_M\}$ satisfying, for all $k \in \{1, \ldots, D_M\}$, the inequality

$$\log_2(1 + \sigma_{[k]}) \geq r_M \quad (14)$$

The error rate of the mMTC devices is then quantified as the ratio

$$\Pr(E_M) = \frac{\mathbb{E}[D_M]}{\lambda_M} \quad (15)$$

between the average number of users in outage, namely $\mathbb{E}[D_M]$, and the average number λ_M of active users. The maximum rate λ_M that can be supported under the reliability condition $\Pr(E_M) = \epsilon_M$ is defined for reference as

$$\lambda_M^{\text{orth}}(r_M) = \max\{\lambda_M : \Pr(E_M) \leq \epsilon_M\}. \quad (16)$$

This quantity can be computed by means of Monte Carlo numerical methods.

III. SLICING FOR EMBB AND URLLC

In this section, we consider the coexistence of eMBB and URLLC devices, while assuming that there is no mMTC traffic, i.e., that the mMTC arrival rate is $\lambda_M = 0$. We first briefly recall, using the results in the previous section, how the performance of the two services can be evaluated for the case of H-OMA, and then analyze the more complex scenario of H-NOMA.

A. ORTHOGONAL SLICING: H-OMA FOR EMBB AND URLLC

In the case of orthogonal slicing, i.e., under H-OMA, we assume that F_U out of the F frequency radio resources for all minislots in the given radio resource are allocated to the URLLC transmissions, while the remaining $F_B = F - F_U$ radio resources are each allocated to one eMBB user. Note that, in each minislot, the probability that the F_U frequency channels allocated to URLLC traffic are unused is the complement of the activation probability, i.e., $1 - a_U$.

The performance of the system is specified in terms of the the pair (r_B, r_U) of eMBB sum-rate r_B and URLLC rate r_U achievable at the given reliability levels (ϵ_B, ϵ_U). The eMBB sum-rate is obtained as

$$r_B = (F - F_U)r_B^{\text{orth}}, \quad \text{[bits/symbol]} \quad (17)$$

where r_B^{orth} is obtained as explained in Section II-B. The URLLC rate r_U is computed from (12) by imposing the equality $\Pr(E_U) = \epsilon_U$ as detailed in Section II-C.

B. NON-ORTHOGONAL SLICING: H-NOMA FOR EMBB AND URLLC

We now consider non-orthogonal slicing, or H-NOMA, whereby all F frequency channels are used for both eMBB and URLLC transmissions. Hence, $F_U = F_B = F$. With non-orthogonal slicing, eMBB and URLLC transmissions interfere, and, hence, the rate pair (r_B, r_U) cannot be directly obtained from the analysis in Section II. We next describe different decoding architectures, and derive corresponding achievable pairs r_B and r_U for non-orthogonal slicing.

Decoding Architectures: A key observation in the design of decoding schemes is that, due to latency constraints, the decoding of a URLLC transmission cannot wait for, and hence depend upon, the decoding of eMBB traffic. In fact, decoding of a URLLC transmission can only rely on the signal received in the given minislot. This constraint prevents SIC decoders whereby eMBB transmissions are decoded first

and canceled from the received signal prior to decoding of the URLLC messages. Note also that, because of the heterogeneity of reliability requirements, decoding eMBB first and then URLLC in a SIC fashion would require decoding the eMBB traffic at the same level of reliability needed for the URLLC traffic. As a result of these considerations, in H-NOMA with SIC the URLLC transmissions should be decoded while treating eMBB signals as an additional noise.

In contrast to URLLC traffic, eMBB requirements are less demanding in terms of latency, and hence eMBB decoding can wait for URLLC transmissions to be decoded first. This enables a SIC mechanism whereby URLLC messages are decoded and then canceled from the received signal prior to decoding of the eMBB signal. Since the reliability of URLLC is two or more orders of magnitude higher than eMBB, the performance of eMBB under the described SIC decoder is expected to be close to the ideal orthogonal case in which no interference from URLLC traffic is present. This design choice is an instance of reliability diversity.

That being said, the SIC decoder may be ruled out by considerations such as complexity. In such circumstances, one could adopt another decoding approach that is, in a sense, diametrically opposite to SIC in its treatment of URLLC interference. Such a decoder treats any minislot that contains a URLLC transmission as erased or *punctured*. This option of H-NOMA with puncturing is currently being considered within the 5G community [10]. Note that this approach requires the decoder at the BS to be able to detect the presence of URLLC transmissions, e.g., via energy detection.

Encoding: If H-NOMA with SIC is used, we set the eMBB rate to

$$r_{B,f}^{\text{SIC}}(G_{B,f}^{\text{tar}}) = \log_2(1 + G_{B,f}^{\text{tar}}), \quad \text{[bits/symbol]} \quad (18)$$

where $G_{B,f}^{\text{tar}}$ represents the target SNR for eMBB transmission, which is to be determined.

In contrast, in H-NOMA with puncturing and erasure decoding, the eMBB device applies an outer erasure code with rate $1 - k/S$, which is concatenated to the codebook used in the physical-layer transmission of the eMBB encoder. Thanks to the erasure code, the decoder is able to correct $k \leq S$ erased minislots, while, if the number of URLLC transmissions is larger than k, the decoding process fails. The parameter k needs to be designed so as to satisfy the target error rate ϵ_B for eMBB users. The resulting data rate for eMBB transmission in frequency channel f is

$$r_{B,f}^{\text{pun}}(G_{B,f}^{\text{tar}}, k) = \left(1 - \frac{k}{S}\right) \log_2(1 + G_{B,f}^{\text{tar}}). \quad (19)$$

Regarding the selection of the target SNR $G_{B,f}^{\text{tar}}$, we recall that in the orthogonal case, as shown in (10), the variable $G_{B,f}^{\text{tar}}$ is uniquely determined by the error probability target ϵ_B via the threshold SNR $G_{B,f}^{\text{min}}$ defined in (7). In contrast, with non-orthogonal slicing, it may be beneficial to choose a smaller target SNR than the one given by (10), so as to reduce the interference caused to URLLC transmissions. This yields the

inequality

$$G_{B,f}^{\text{tar}} \leq \frac{\Gamma_B}{\gamma\left(0, \frac{G_{B,f}^{\text{min}}}{\Gamma_B}\right)}. \quad (20)$$

Summarizing the discussion so far, decoding of URLLC traffic cannot leverage SIC and treats eMBB transmissions as noise. In contrast, the eMBB decoder at the BS can either leverage SIC by decoding URLLC traffic first, or rather treat any minislot occupied by URLLC traffic as erased. These are two extreme points among all possible eMBB decoders.

Rate Region: The objective of the analysis is to determine the rate region (r_B, r_U) for which the target error probabilities of the two services are satisfied. To this end, we fix the URLLC rate $r_U \in [0, r_U^{\text{orth}}(F)]$, and compute the maximum attainable eMBB rate r_B. We recall that the available degrees of freedom in the design are the target SNR $G_{B,f}^{\text{tar}}$ and the minimum channel gain $G_{B,f}^{\text{min}}$ at which an eMBB device is active (or equivalently the activation probability (4)), as well as the erasure code parameter k if a puncturing approach is adopted for eMBB decoding. We also emphasize that, unlike the orthogonal case, the target SNR $G_{B,f}^{\text{tar}}$ and the minimum SNR $G_{B,f}^{\text{min}}$ are separate degrees of freedom, which are related by the inequality (20).

We start by imposing the reliability constraint for the URLLC user, which yields the following condition for both SIC and erasure decoder:

$$\Pr(E_U) = \Pr\left(\frac{1}{F_U}\sum_{f=1}^{F_U}\log_2\left(1 + \frac{G_{U,f}}{1 + \delta_f G_{B,f}^{\text{tar}}}\right) < r_U\right) \leq \epsilon_U. \quad (21)$$

Here, $\{\delta_f\}_{f=1}^{F_U}$ are independent Bernoulli random variables with parameter a_B given in (4). Recall that a_B is a function of $G_{B,f}^{\text{min}}$. The term $\delta_f G_{B,f}^{\text{tar}}$ represents the interference power caused by an eMBB transmission on frequency channel f to the URLLC traffic. The inequality (21) imposes a joint constraint on both $G_{B,f}^{\text{tar}}$ and $G_{B,f}^{\text{min}}$. Next, we impose the reliability constraint for eMBB traffic by considering separately SIC and erasure decoders.

1) SIC DECODER

Under H-NOMA the decoding of an eMBB message is generally affected by the interference from the URLLC users. However, this interference is not present if: (*i*) there are no URLLC transmissions, i.e., $S_U = 0$; or (*ii*) if URLLC transmissions are present, i.e., $S_U > 0$, but the corresponding signals are decoded successfully and canceled by the SIC decoder. As for the latter event, since the interference from eMBB users and the fading gains are constant across the minislots, either all URLLC transmissions are decoded incorrectly (event E_U) or they are all correctly decoded (event $\overline{E}_U$).

Based on the discussion above, we can bound the eMBB error probability by distinguishing the case in which the eMBB transmission is subject to interference from URLLC signals, and the case in which is not, using the law of total

probability, as presented through the equations (22)-(24), as shown at the bottom of the this page.

Here, equality (22) holds because the only source of outage for eMBB in absence of URLLC interference is the instantaneous SNR being below the minimum SNR, which implies $\Pr(E_B|S_U = 0) = 1 - a_B$; moreover, (23) follows by using that $\Pr(E_B|E_U, S_U > 0) \leq 1$ and that canceling URLLC interference results in the same performance achievable in the absence of URLLC transmissions, so that we have the equality $\Pr(E_B|\bar{E}_U, S_U > 0) = \Pr(E_B|S_U = 0) = 1 - a_B$. Imposing the reliability condition $\Pr(E_B) \leq \epsilon_B$ and using (24) we obtain the inequality

$$a_B \geq \frac{1 - \epsilon_B}{1 - \epsilon_U(1 - (1 - a_U)^S)}. \tag{25}$$

As already pointed out, this equivalently imposes a constraint on $G_{B,f}^{\min}$ through (4).

From (25), we see that, unlike the orthogonal case, with non-orthogonal slicing the eMBB activation probability a_B is larger than $1 - \epsilon_B$. This is because URLLC interference may cause an eMBB decoding error even when the eMBB's SNR is above the threshold. However, the impact of URLLC interference is typically minimal. Indeed, the high reliability requirements for URLLC, which are reflected by the very small value of ϵ_U, imply that a_B is close to $1 - \epsilon_B$.

To summarize, for a given feasible URLLC rate $r_U \in [0, r_U^{\text{orth}}(F)]$, the maximum eMBB rate is obtained by maximizing $\log_2(1 + G_{B,f}^{\text{tar}})$ subject on the constraints on G_B^{tar} and $G_B^{\min}$ implied by (20), (21), and (25). This maximization requires the use of a two-dimensional numerical search. Note that the activation probability a_B is typically very close to 1, and hence, when solving this problem, one can conservatively assume that the eMBB interference is always present in (21), i.e., $\delta_f = 1$. In contrast, the dependence of the right-hand side of (21) on G_B^{tar} causes a non-trivial interdependence between r_U and r_B.

2) PUNCTURING AND ERASURE DECODER
Turning now to the erasure decoder, we can write the probability of error for an eMBB user by means of the law of total probability as

$$\Pr(E_B)$$
$$= \Pr(S_U \leq k)\Pr(E_B|S_U \leq k) + \Pr(S_U > k)\Pr(E_B|S_U > k), \tag{26}$$

where we have distinguished the case in which the erasure code is able to correct the erasures caused by URLLC transmissions, i.e., $S_U \leq k$, and the case in which an error is
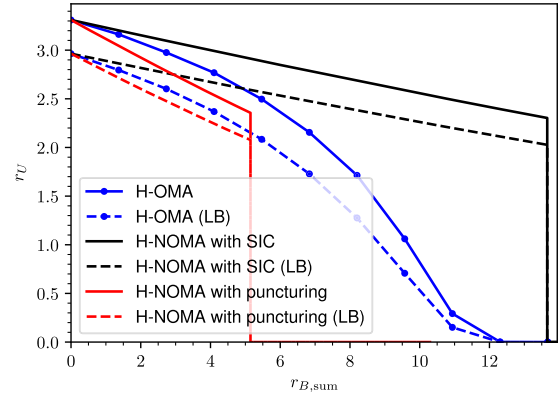

FIGURE 5. Rate region $(r_{B,\text{sum}}, r_U)$ for the eMBB rate $r_{B,\text{sum}}$ and the URLLC rate r_U when $\Gamma_U = 20$ dB, $\Gamma_B = 10$ dB, $S = 5$, $a_U = 0.1$, $F = 10$, $\epsilon_U = 10^{-5}$, $\epsilon_B = 10^{-3}$. H-NOMA is present with two variants, SIC and puncturing. The lower bound (LB) is derived in Appendix A. The axes' units are bits per channel use [bpcu].

instead declared, i.e., $S_U > k$. When the latter event occurs, a decoding error occurs, and hence we have $\Pr(E_B|S_U > k) = 1$. In contrast, when $S_U \leq k$, the only source of outage is the instantaneous SNR being below threshold, which results in $\Pr(E_B|S_U \leq k) = 1 - a_B$. Overall, the resulting eMBB reliability requirement is

$$\Pr(E_B) = \Pr(S_U \leq k)(1 - a_B) + \Pr(S_U > k) \leq \epsilon_B. \tag{27}$$

Imposing equality in (27), we determine the parameter a_B and, hence, $G_{B,f}^{\min}$ via (4). Given the desired feasible URLLC rate $r_U \in [0, r_U^{\text{orth}}(F)]$, we then obtain the target SNR $G_{B,f}^{\text{tar}}$ and, hence, the eMBB rate (19) from the URLLC reliability condition (21).

C. NUMERICAL ILLUSTRATION
Here we present simulation results for the rate region (r_B, r_U) for H-OMA as well as H-NOMA, with both SIC and puncturing decoders. In addition to the results obtained from the previous analysis, we also show curves obtained from the expressions derived in Appendix A, which are easier to evaluate and are shown to provide a performance lower bound ("LB").

In Figs. 5 and 6, we plot the rate regions for $S = 5$, $a_U = 0.1$, $F = 10$, $\epsilon_U = 10^{-5}$, $\epsilon_B = 10^{-3}$. Fig. 5 considers the case $\Gamma_U > \Gamma_B$ with $\Gamma_U = 20$ dB and $\Gamma_B = 10$ dB, while Fig. 6 focuses on the complementary set-up with $\Gamma_B > \Gamma_U$ when $\Gamma_U = 10$ dB and $\Gamma_B = 20$ dB. For both figures, H-NOMA with puncturing uses the optimal puncturing parameter k.

$$\Pr(E_B) = \Pr(S_U = 0)\Pr(E_B|S_U = 0) + \Pr(S_U > 0)\big(\Pr(E_U|S_U > 0)\Pr(E_B|E_U, S_U > 0)$$
$$+ \Pr(\bar{E}_U|S_U > 0)\Pr(E_B|\bar{E}_U, S_U > 0)\big) = \tag{22}$$
$$= (1 - a_U)^S(1 - a_B) + [1 - (1 - a_U)^S](\epsilon_U \Pr(E_B|E_U, S_U > 0) + (1 - \epsilon_U)\Pr(E_B|\bar{E}_U, S_U > 0)) \tag{23}$$
$$\leq (1 - a_U)^S(1 - a_B) + [1 - (1 - a_U)^S](\epsilon_U + (1 - \epsilon_U)(1 - a_B)). \tag{24}$$
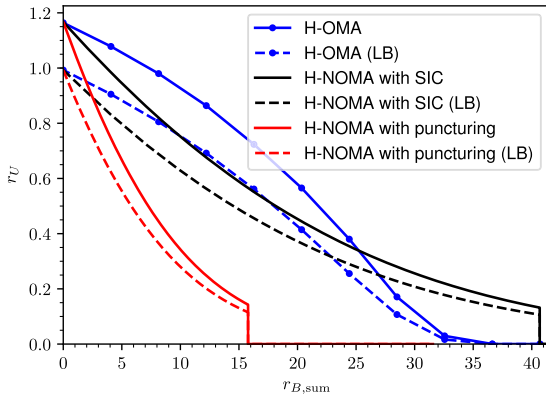
FIGURE 6. Rate region $(r_{B,\text{sum}}, r_U)$ for the eMBB rate $r_{B,\text{sum}}$ and the URLLC rate r_U when $\Gamma_U = 10$ dB, $\Gamma_B = 20$ dB, $S = 5$, $a_U = 0.1$, $F = 10$, $\epsilon_U = 10^{-5}$, $\epsilon_B = 10^{-3}$. H-NOMA is present with two variants, SIC and puncturing. The lower bound (LB) is derived in Appendix A. The axes' units are bits per channel use [bpcu].

For both set-ups considered in the figure, H-MONA with puncturing is outperformed by both H-OMA and H-NOMA with SIC. Furthermore, when $\Gamma_U > \Gamma_B$, we see from Fig. 5 that the SIC region dominates the region achievable by orthogonal slicing. This is thanks to the capability of the BS to decode and cancel URLLC transmissions by leveraging reliability diversity.

In contrast, when $\Gamma_U < \Gamma_B$, Fig. 6 shows that orthogonal slicing can attain pairs (r_B, r_U) that are not attainable by H-NOMA with SIC. In particular, H-OMA is preferable if one wishes to obtain large values of the URLLC rate. This is due to the difficulty of ensuring high reliability in the presence of eMBB transmissions when $\Gamma_U < \Gamma_B$. We recall that this is a consequence of the impossibility to decode and cancel eMBB transmissions prior to URLLC decoding owing to the URLLC latency constraint. In contrast, if one is interested in guaranteeing large eMBB sum-rates, H-NOMA offers significant performance gains. This is because non-orthogonal transmission allows eMBB users to operate over a larger number of spectral resources while not being significantly affected by URLLC interference.

We see that the lower bound is able to capture the shape of the region obtained through more accurate and time-consuming Monte-Carlo simulations.

IV. SLICING FOR EMBB AND MMTC

In this section, we treat the slicing of wireless resources to jointly support eMBB and mMTC services, while assuming that the URLLC traffic, if present, has been allocated orthogonal resources. Analogously to the case of eMBB-URLLC coexistence, we consider separately orthogonal slicing (H-OMA) and non-orthogonal slicing (H-NOMA). We shall focus without loss of generality on the case $F = 1$, since the mMTC users are assumed to be active on a single frequency channel. The extension to the case $F > 1$, in which the mMTC devices are allowed to randomly access all F channels, is rather straightforward, as further elaborated in

Section V-B. Since a single channel is considered, in this section we omit all frequency indices f.

A. ORTHOGONAL SLICING: H-OMA FOR EMBB AND MMTC

For the case of orthogonal slicing, we assume that the eMBB and the mMTC devices use the frequency radio resource in a time-sharing manner. Let $\alpha \in [0, 1]$ and $1 - \alpha$ be fraction of time in which the resources are allocated to the eMBB device and the mMTC devices, respectively. We aim at characterizing the region of pairs (r_B, λ_M) of eMBB rate r_B and mMTC arrival rate λ_M that can be supported by orthogonal slicing for a given mMTC transmission rate requirement r_M and probability of error ϵ_M.

For a given time-sharing factor α, the achievable pair of eMBB rate r_B and mMTC arrival rate λ_M can be written in terms of the quantities derived in Sections II-B and II-D as

$$r_B = \alpha r_B^{\text{orth}} \tag{28}$$

$$\lambda_M = \lambda_M^{\text{orth}} \left(\frac{r_M}{1 - \alpha} \right), \tag{29}$$

respectively, where r_B^{orth} is obtained as explained in Section II-B and $\lambda_M^{\text{orth}}(\cdot)$ is defined in (16). In fact, with orthogonal slicing, both the achievable eMBB rate r_B and the achievable mMTC transmission rate (specified on the right-hand-side of (14)) are scaled according to the fraction of time resources allocated to the service.

B. NON-ORTHOGONAL SLICING: H-NOMA FOR EMBB AND MMTC

In H-NOMA, the eMBB device is allowed to use the radio resource at the same time as the mMTC devices.

Decoding Architecture. As argued in Section II-D, a SIC decoder may enhance the reliability of mMTC decoding. Furthermore, when radio resources are allocated exclusively to mMTC devices, optimal decoding follows the order of descending channel gains. The situation is more complicated in the presence of an interfering eMBB transmission.

In light of the higher reliability requirements of eMBB transmissions as compared to mMTC traffic, i.e., $\epsilon_B \ll \epsilon_M$, one may be tempted to consider decoding the eMBB traffic before attempting to decode any mMTC traffic. This appears to be in line with the discussion in the previous section concerning SIC for eMBB and URLLC coexistence. However, this approach is suboptimal, since it neglects to account for the different definition of reliability of mMTC traffic. In fact, the probability of error (15) measures the fraction of incorrectly detected active users and not a per-device decoding probability. As such, some of the active mMTC devices may well have very high channel gains, hence, causing large interference, making it beneficial to decode and cancel them prior to decoding the eMBB signal. Selecting a SIC decoder that accounts for this important feature of mMTC traffic is another example of a design choice that utilizes reliability diversity.

Based on this discussion, we assume that, at each decoding step, the BS decodes either the eMBB device, provided that it has not been decoded yet, or the next available mMTC device in order of decreasing channel gains. Note that this implies that the decoding step at which the eMBB device is decoded is random, as it depends on the realization of the channel gains. The process ends when no more transmissions can be reliably decoded.

As in non-orthogonal slicing of eMBB and URLLC, the eMBB rate is set to

$$r_B = \log_2(1 + G_B^{\text{tar}}) \quad (30)$$

where G_B^{tar} is the target SNR for the eMBB transmission, which is to be determined. Similar to the eMMB-URLLC coexistence case (see Section III-B), this quantity needs to satisfy

$$G_B^{\text{tar}} \leq \frac{\Gamma_B}{\gamma\left(0, \frac{G_B^{\text{min}}}{\Gamma_B}\right)}. \quad (31)$$

Again, as in Section III-B, we allow the eMBB device not to use the maximal power since it may be beneficial to use a value of G_B^{tar} lower than the right-hand side of (31) in order to control the impact of eMBB interference on the overall SIC procedure.

Next, we formalize the SIC decoding procedure. When the eMBB is inactive because of an insufficient SNR, i.e., $G_B < G_B^{\text{min}}$, the SIC decoding procedure is equivalent to the procedure described in Section II-D, namely, the mMTC devices are decoded in the order of decreasing channel gains. When the eMBB is active, the SIC procedure runs as follows. Starting from $m_0 = 1$, the receiver computes the SINR for the m_0-th mMTC device as

$$\sigma_{[m_0]} = \frac{G_{[m_0]}}{1 + G_B^{\text{tar}} + \sum_{m=m_0+1}^{A_M} G_{[m]}}. \quad (32)$$

If $\log_2(1 + \sigma_{[m_0]}) \geq r_M$, the m_0-th mMTC is decoded, canceled, m_0 is incremented by one, and the procedure starts over. Otherwise, the receiver attempts to decode the eMBB user. To this end, it computes the SINR of the eMBB transmission as

$$\sigma_B = \frac{G_B^{\text{tar}}}{1 + \sum_{m=m_0}^{A_M} G_{[m]}} \quad (33)$$

and decodes and cancels the eMBB if the condition $\log_2(1 + G_B^{\text{tar}}) \geq r_B$ is satisfied. If the eMBB is decoded successfully, the decoding procedure continues as in Section II-D. If $\log_2(1 + G_B^{\text{tar}}) < r_B$, the procedure terminates.

Let $D_M \in \{0, \ldots, A_M\}$ and $D_B \in \{0, 1\}$ be the random variables denoting the number of decoded mMTC and eMBB devices. With this notation, the probabilities of error for mMTC and eMBB users are given as $\Pr(E_M) = 1 - \mathbb{E}[D_M]/\lambda_M$ and $\Pr(E_B) = 1 - \mathbb{E}[D_B]$, respectively.

In order to characterize the achievable pairs (r_B, λ_M), we evaluate the maximum supported mMTC arrival rate λ_M

as a function the eMBB rate r_B as

$$\lambda_M^{\text{non-orth}}(r_B)$$
$$= \max\left\{\lambda_M \geq 0 : \exists G_B^{\text{tar}} \text{ and } G_B^{\text{min}}\right.$$
$$\left. \text{s.t. } \mathbb{E}[D_M]/\lambda_M \geq 1 - \epsilon_M \text{ and } \mathbb{E}[D_B] \geq 1 - \epsilon_B\right\}. \quad (34)$$

We remark that, the probability distributions of D_M and D_B depend on the parameters λ_M, G_B^{tar}, G_B^{min}, r_B, and r_M. The computation of (34) requires Monte Carlo simulations.

C. NUMERICAL ILLUSTRATION

We present numerical simulation results illustrating the trade-offs between the eMBB rate r_B and the mMTC arrival rate λ_M for orthogonal and non-orthogonal slicing. In addition to numerical results obtained by solving (34) through Monte Carlo methods, we also report results obtained upper and lower bounds on $\lambda_M^{\text{non-orth}}(\cdot)$ in (34), which are easier to evaluate and are derived in Appendix B. Throughout this section, we set $\epsilon_M = 10^{-1}$, and $r_M = 0.04$.

In Fig. 7, we plot the maximum mMTC arrival rate λ_M for both orthogonal and non-orthogonal slicing as a function of r_B when $\Gamma_M = 5$ dB, $\Gamma_B = 25$ dB, and $\epsilon_B = 10^{-3}$. When orthogonal slicing (H-OMA) is used, the supported mMTC arrival rate λ_M is seen to decrease in an approximately linear fashion with the eMBB rate. As for non-orthogonal slicing (H-NOMA), we observe three fundamentally different regimes as r_B changes from zero towards is maximal value r_B^{orth}.

The *first regime* consists of very small values of the eMMB rate r_B, for which the supported arrival rate $\lambda_M^{\text{non-orth}}$ is almost constant. At such values of r_B, the eMBB device can be reliably decoded before the mMTC devices. Therefore, interference from eMBB user can be cancelled, and the performance of mMTC traffic is unaffected by small increases in r_B. The *second regime* spans intermediate values of r_B. In this case, the eMBB signal can only be decoded after some of the strongest mMTC signals are decoded and canceled. Hence, the mMTC performance is reduced by the interference from eMBB transmissions. Also, the SIC decoder tends to stop the decoding process after detecting the eMBB user, and decoding typically fails while detecting an mMTC device because of the interference from the other, yet undecoded, mMTC devices. In the *third regime*, eMBB decoding fails with a probability comparable to that of the weaker mMTC devices due to the mutual interference between the two services. As a result, in this regime, the supported mMTC arrival rate decays to zero as the eMBB rate r_B increases.

The first and the third regime identified in Fig. 7 can also be understood with the help of the lower and upper bounds derived in Appendix B. In particular, when r_B is very low, as mentioned, it is almost always possible to decode the eMBB transmission before decoding any mMTC device. This is the premise of the lower bound, which, as shown in Fig. 7, agrees with the simulation results in the first regime. On the contrary, the upper bound is computed by first identifying the
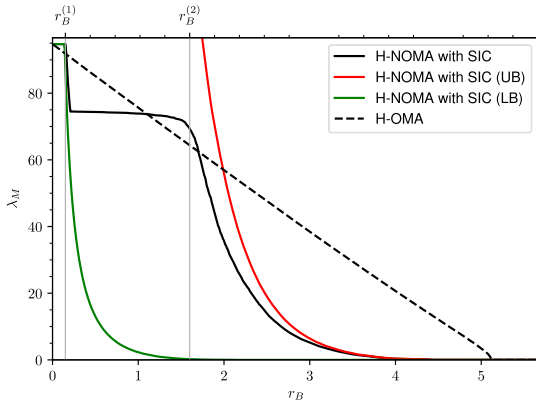
FIGURE 7. Arrival rates λ_M^{orth} and $\lambda_M^{\text{non-orth}}$ for mMTC traffic under H-OMA and H-NOMA, respectively, as a function of the eMBB rate r_B. The upper bounds (UB) and lower bounds (LB) on the H-NOMA arrival rate $\lambda_M^{\text{non-orth}}(r_B)$ derived in Appendix B are also shown. The parameters are $\Gamma_M = 5$ dB, $\Gamma_B = 25$ dB, $\epsilon_M = 10^{-1}$, $\epsilon_B = 10^{-3}$, and $r_M = 0.04$. The unit of r_B is bits per channel use [bpcu], while $\lambda_M^{\text{non-orth}}$ is [arrivals/slot].

subset of mMTC devices whose channels are so weak that the additive noise and the eMBB interference alone make their decoding impossible. The upper bound is seen to agree the simulation results in the third regime, i.e., when r_B is large.

In Figs. 8 and 9, we plot the supported mMTC arrival rate λ_M as a function of r_B for different values of the eMBB SNR Γ_B and $\epsilon_B = 10^{-3}$, and for different eMBB reliability levels ϵ_B and $\Gamma_B = 20$ dB, respectively. This figures allows us to assess the impact of the average eMBB gain Γ_B and of the eMBB reliability ϵ_B on the operation of the system in the three regimes identified above and on relative performance of orthogonal and non-orthogonal slicing, as discussed next.

As it pertains to three regimes, we observe from Fig. 8 that the rate at which the transition from the second to the third regime occurs does not change as the eMBB average channel gain Γ_B is increased from 20 dB to 30 dB. On the contrary, the r_B value corresponding to the transition from the first to the second regime becomes larger with this increase in Γ_B. This increase, in fact, allows the eMBB transmission to be decoded earlier in the SIC process for a larger set of values of r_B. From Fig. 9, we observe that the eMBB error probability constraint significantly affects the supported mMTC arrival rate in the second and third regimes. In fact, in these case, the higher eMBB transmission power required to ensure a higher reliability impairs the decoding of mMTC users via interference.

We now elaborate on the comparison between orthogonal and non-orthogonal slicing. The presented figures emphasize the fact that there are points in the rate region (r_B, λ_M) that can be attained by non-orthogonal slicing and not by orthogonal slicing, and vice versa. Specifically, non-orthogonal slicing is seen to be beneficial when r_B is across the second and the third regimes, especially for not too large reliability levels ϵ_B (see Fig. 8). For such values, the eMBB rate is large, and yet low enough not to hamper the decoding of the mMTC users. Once again, reliability diversity is crucial to ensure the
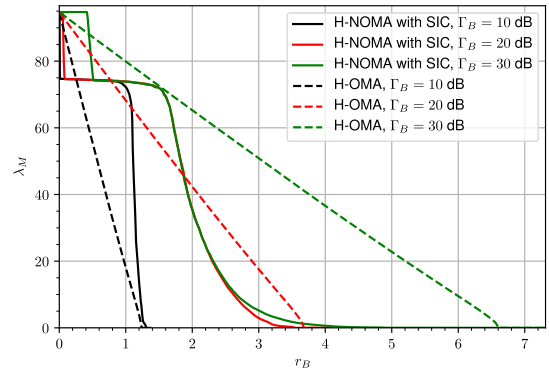

FIGURE 8. Arrival rates λ_M^{orth} and $\lambda_M^{\text{non-orth}}$ for mMTC traffic under H-OMA and H-NOMA, respectively, as a function of the eMBB rate r_B for $\Gamma_B \in \{10, 20, 30\}$ dB. The parameters are $\Gamma_M = 5$ dB, $\epsilon_M = 10^{-1}$, $\epsilon_B = 10^{-3}$, and $r_M = 0.04$. The unit of r_B is bits per channel use [bpcu], while $\lambda_M^{\text{non-orth}}$ is [arrivals/slot].
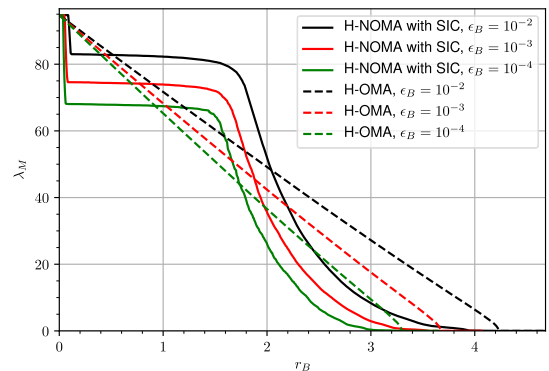

FIGURE 9. Arrival rates λ_M^{orth} and $\lambda_M^{\text{non-orth}}$ for mMTC traffic under H-OMA and H-NOMA, respectively, as a function of the eMBB rate r_B for $\epsilon_B \in \{10^{-2}, 10^{-3}, 10^{-4}\}$. The parameters are $\Gamma_M = 5$ dB, $\Gamma_B = 20$ dB, $\epsilon_M = 10^{-1}$, and $r_M = 0.04$.

effectiveness of non-orthogonal slicing. In contrast, for large values of the rate r_B, when non-orthogonal slicing is deeply in the third operating regime, orthogonal slicing is always superior. This is because in this regime the performance is limited by the interference caused by eMBB users.

V. DISCUSSION, CONCLUSIONS AND FUTURE WORK
A. DISCUSSION AND CONCLUSIONS

In this work, we have presented a communication-theoretic model that enables the investigation of the fundamental trade-offs associated with the sharing of the wireless resources among the three 5G traffic types, namely eMBB, mMTC and URLLC. Albeit simple, the model accounts for the differences among the services in reliability, latency, and number of supported devices. Specifically, we have considered the slicing of resources among the services in the uplink over a shared multiple access resource. We have utilized the term "slicing" in order to emphasize the heterogeneous performance requirements that need to be satisfied for each service as well as the performance isolation among services. Two slicing paradigms have been investigated, orthogonal

and non-orthogonal and the respective transmission schemes H-OMA and H-NOMA, where the latter is inherently possible only in shared wireless channels.

We have applied the model to the study of the slicing for two services in two different cases: *(i)* eMBB and URLLC and *(ii)* eMBB and mMTC. In both cases, we have shown that, in order to be effective, the design of non-orthogonal slicing solutions must be guided by reliability diversity. For the case of eMBB-URLLC coexistence, reliability diversity dictates that, in H-NOMA with SIC, the URLLC device should be decoded first, as its decoding cannot depend on the decoding of eMBB, whose reliability and latency requirements are much looser compared to URLLC. The implications of reliability diversity are more subtle in the case of non-orthogonal slicing between eMBB and mMTC. In this case, considering the fact that the number of active mMTC devices is large with high probability, it is natural to introduce a reliability metric that accounts for the fraction of correctly decoded transmissions. The analysis demonstrated that there are regimes in which the decoding of eMBB should be performed after the decoding of one or multiple mMTC devices in order to benefit from non-orthogonal slicing.

Our numerical results show that there are regimes in which H-NOMA is advantageous over H-OMA and vice versa. In the case of eMBB-URLLC, H-NOMA with SIC is always beneficial when the eMBB rate is very large. In the case of eMBB-mMTC, non-orthogonal slicing is beneficial when the eMBB rate takes values that are small enough not to hamper the decoding of the mMTC devices.

B. GENERALIZATIONS AND FUTURE WORK

The analysis presented in this paper is based on some simplifying assumptions. However, the basic model and the methodology developed here can be extended to more general models and other operation regimes, as briefly discussed here.

Starting with eMBB-URLLC coexistence, one could devise another H-NOMA scheme, where eMBB and URLLC users are allowed to access partially non-orthogonal resources, so that only a subset of frequency channels potentially occupied by URLLC traffic may be interfered by eMBB transmissions. Another direct generalization is to assume that the minislots are pre-allocated to different URLLC devices. Recall that, when all transmissions are made by the same URLLC device, the block fading model dictates that either all or none of the transmissions in a minislot are decoded correctly. If each URLLC transmission is carried out by a different device, then then error decoding events are independent across the minislots. As a more involved extensions of the model, one may consider the impact of frequency diversity also for eMBB traffic, and the performance under alternative decoding strategies, such as treating interference as noise.

As for the coexistence of mMTC and eMBB services, an interesting extension is to allow multiple channels for mMTC traffic. In particular, mMTC devices may be allowed to use frequency hopping, and the number of allocated frequency channels may depend on the reliability requirements.

Another aspect that deserves study is the impact of the arrival process, which here has been assumed to be Poisson. Namely, the higher burstiness of the arrival process can potentially improve the gain that one can obtain with non-orthogonal slicing. Finally, following the approach in NB-IoT systems, the transmission of a single mMTC device may consist of replicas of the same packet in multiple time slots. This makes the non-orthogonal slicing of mMTC and eMBB even more relevant, as it is not feasible to reserve resources exclusively for replicas of packets generated by sporadically active mMTC devices.

APPENDIX A
LOWER BOUND ON THE URLLC RATE r_U
By setting $\Pr(E_U) = \epsilon_U$, we can upper bound (21) as

$$\epsilon_U = \Pr\left[\sum_{f=1}^{F_U}\log_2\left(1 + \frac{G_{U,f}}{1 + \delta_f G_B^{\text{tar}}}\right) < F_U r_U\right] \quad (35)$$

$$= \Pr\left[\prod_{f=1}^{F_U}\left(1 + \frac{G_{U,f}}{1 + \delta_f G_B^{\text{tar}}}\right)^{-t} \geq 2^{-r_U F_U t}\right] \quad (36)$$

$$\leq \frac{\mathbb{E}\left[\prod_{f=1}^{F_U}\left(1 + \frac{G_{U,f}}{1 + \delta_f G_B^{\text{tar}}}\right)^{-t}\right]}{2^{-r_U F_U t}} \quad (37)$$

$$= \frac{\mathbb{E}\left[\left(1 + \frac{G_{U,1}}{1 + \delta_1 G_B^{\text{tar}}}\right)^{-t}\right]^{F_U}}{2^{-r_U F_U t}}. \quad (38)$$

We obtained (35) by multiplying both terms in the inequality by $-t$ and then by exponentiating them; (36) follows from Markov inequality; and (37) holds because $\{G_{U,f}\}$ and $\{\delta_f\}$, $f \in \{1, \ldots, F_U\}$ are i.i.d., and hence we can set $f = 1$ in (38) without loss of generality. The inequality in (38) can be rewritten as

$$r_U \geq \frac{1}{tF_U}\log_2\epsilon_U - \frac{1}{t}\log_2\mathbb{E}\left[\left(1 + \frac{G_{U,1}}{1 + \delta_1 G_B^{\text{tar}}}\right)^{-t}\right]$$

$$> \frac{1}{tF_U}\log_2\epsilon_U - \frac{1}{t}\log_2\mathbb{E}\left[\left(1 + \frac{G_{U,1}}{1 + G_B^{\text{tar}}}\right)^{-t}\right] \quad (39)$$

where the strict lower bound follows by assuming that the eMBB interference is always present, i.e., $\Pr(\delta_1) = 1$. The expectation in (39) can be calculated by a Monte Carlo simulation and the value of $t \geq 0$ in (39) is chosen such as to maximize the lower bound.

APPENDIX B
BOUNDS FOR NON-ORTHOGONAL SLICING FOR EMBB AND MMTC
1) A FIRST UPPER BOUND
The idea behind the bound is as follows: if the eMBB decoding fails, then the decoding of all mMTC devices for which $G_{[m]}/(1 + G_B^{\text{tar}}) \leq 2^{r_M} - 1$ must also fail. We obtain next a lower bound on the eMBB error probability, which will give us the desired upper bound on λ_M, by assuming that

$$\Pr(E_B) \geq 1 - a_B + a_B \Pr\left[\frac{G_B^{\text{tar}}}{1 + \sum_{m=1}^{A_M} G_{[m]} \mathbb{1}\left\{\frac{G_{[m]}}{1+G_B^{\text{tar}}} \leq 2^{r_M} - 1\right\}} \leq 2^{r_B} - 1\right]$$

$$\geq 1 - a_B + a_B \Pr\left[\frac{G_B^{\text{tar}}}{2^{r_B} - 1} - 1 \leq \underbrace{\sum_{m=1}^{A_M} G_{[m]} \mathbb{1}\left\{\frac{G_{[m]}}{1 + G_B^{\text{tar}}} \leq 2^{r_M} - 1\right\}}_{=\chi}\right]. \quad (40)$$

all mMTC devices that do not satisfy $G_{[m]}/(1 + G_B^{\text{tar}}) \leq 2^{r_M} - 1$ are decoded correctly and cancelled before eMBB decoding, and that the remaining ones, which are not decoded correctly, cause interference to the eMBB. This yields the result described in (40), as shown at the top of the this page, where $\mathbb{1}\{\mathcal{A}\}$ is the indicator function of the event $\mathcal{A}$.

The random variable χ in (40) is the sum of a Poisson-distributed number of truncated, exponential-distributed random variables, which allows for an efficient numerical evaluation of the probability term in (40). Next, we set the right-hand side of (40) equal to ϵ_B, and find the values of G_B^{min} and G_B^{tar} that result in the largest λ_M. This value is precisely the desired upper bound on $\lambda_M^{\text{non-orth}}(r_B)$.

2) A LOWER BOUND AND AN ALTERNATIVE UPPER BOUND

We upper-bound the eMBB error probability by considering the following suboptimal decoding scheme. We force the decoder to always decode the eMBB first and subsequently decode the mMTC devices. The maximal supported arrival rate with this modified decoder is clearly a lower bound on $\lambda_M^{\text{non-orth}}(r_B)$. While deriving this bound, we shall derive as by-product also an alternative upper bound on $\lambda_M^{\text{non-orth}}(r_B)$.

As already mentioned, decoding the eMBB as the first device results in an upper bound on the eMBB error probability. Mathematically,

$$\Pr(E_B) \leq 1 - a_B + a_B \Pr\left[\frac{G_B^{\text{tar}}}{1 + \sum_{m=1}^{A_M} G_{[m]}} \leq 2^{r_B} - 1\right] \quad (41)$$

$$= 1 - a_B + a_B \Pr\left[\frac{G_B^{\text{tar}}}{2^{r_B} - 1} - 1 \leq \sum_{m=1}^{A_M} G_{[m]}\right]. \quad (42)$$

Here, the random variable $\sum_{m=1}^{A_M} G_{[m]}$ follows an Erlang distribution. It will turn out convenient to denote by $q_B(r_B, \lambda_M)$ the right-hand side of (42). Let now $\mathcal{E}_M^{\text{orth}}(\lambda_M)$ be the mMTC error probability as a function of mMTC arrival rate λ_M in the absence of the eMBB device. By the law of total probability, the mMTC error probability when the eMBB is present can be upper- and lower-bounded as

$$\mathcal{E}_M^{\text{orth}}(\lambda_M) \leq \Pr[E_M] = 1 - \frac{\mathbb{E}[D_M]}{\lambda_M}$$
$$\leq \mathcal{E}_M^{\text{orth}}(\lambda_M) + \epsilon_B. \quad (43)$$

Set now

$$\lambda_{M,\text{lb}} = \max\left\{\lambda_M \geq 0 : \mathcal{E}_M^{\text{orth}}(\lambda_M) \leq \epsilon_M - \epsilon_B\right\} \quad (44)$$

$$\lambda_{M,\text{ub}} = \max\left\{\lambda_M \geq 0 : \mathcal{E}_M^{\text{orth}}(\lambda_M) \leq \epsilon_M\right\}. \quad (45)$$

In words, these are the largest arrival rates for which the right-hand side and the left-hand side of (43) are smaller than ϵ_M, respectively. Furthermore, let r_B^{low} be given by

$$r_B^{\text{low}} = \max\left\{r_B \geq 0 : q_B(r_B, \lambda_{M,\text{lb}}) \leq \epsilon_B\right\}. \quad (46)$$

It follows that the pair $(r_B^{\text{low}}, \lambda_{M,\text{lb}})$ is achievable. Furthermore, we have that

$$\lambda_{M,\text{lb}} \leq \lambda_M^{\text{non-orth}}(r_B) \leq \lambda_{M,\text{ub}} \quad (47)$$

for all $r_B \in [0, r_B^{\text{low}}]$. When $r_B \in (r_B^{\text{low}}, r_B^{\text{orth}}]$, we obtain a lower bound on $\lambda_M^{\text{non-orth}}(r_B)$ by finding the largest value of λ_M for which $q_B(r_B, \lambda_M) \leq \epsilon_M$ and by taking the smallest between this value and $\lambda_{M,\text{lb}}$. We conclude by noting that the upper bound on $\lambda_M^{\text{non-orth}}(r_B)$ on the right-hand-side of (47), which holds for all r_B, can be combined with the upper bound resulting from (40) to tighten it when r_B is small.

REFERENCES

[1] *ITU-R M.[IMT-2020.TECH PERF REQ]—Minimum Requirements Related to Technical Performance for IMT–2020 Radio Interface(s)*, document ITU-R M.2410-0, International Telecommunication Union—Recommendations, Nov. 2017.

[2] *Study on New Radio (NR) Access Technology Physical Layer Aspects*, document TR 38.802, 3GPP, Mar. 2017.

[3] G. C. Ferrante, J. Ostman, G. Durisi, and K. Kittichokechai, "Pilot-assisted short-packet transmission over multiantenna fading channels: A 5G case study," in *Proc. 52nd Annu. Conf. Inf. Sci. Syst. (CISS)*, Princeton, NJ, Mar. 2018, pp. 1–6.

[4] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. 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.

[5] Z. Wu, K. Lu, C. Jiang, and X. Shao, "Comprehensive study and comparison on 5G NOMA schemes," *IEEE Access*, vol. 6, pp. 18511–18519, 2018.

[6] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. 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] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.

[8] W. Guan, X. Wen, L. Wang, Z. Lu, and Y. Shen, "A service-oriented deployment policy of end-to-end network slicing based on complex network theory," *IEEE Access*, vol. 6, pp. 19691–19701, 2018.

[9] Z. Wu, F. Zhao, and X. Liu, "Signal space diversity aided dynamic multiplexing for eMBB and URLLC traffics," in *Proc. 3rd IEEE Int. Conf. Comput. Commun. (ICCC)*, Dec. 2017, pp. 1396–1400.

[10] A. Anand, G. de Veciana, and S. Shakkottai. (2017). "Joint scheduling of URLLC and eMBB traffic in 5G wireless networks." [Online]. Available: https://arxiv.org/abs/1712.05344

[11] S.-Y. Lien, S.-C. Hung, D.-J. Deng, and Y. J. Wang, "Efficient ultra-reliable and low latency communications and massive machine-type communications in 5G new radio," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–7.

[12] N. H. Mahmood, M. Lauridsen, G. Berardinelli, D. Catania, and P. Mogensen, "Radio resource management techniques for eMBB and mMTC services in 5G dense small cell scenarios," in *Proc. IEEE 84th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2016, pp. 1–5.

[13] L. Tian, C. Yan, W. Li, Z. Yuan, W. Cao, and Y. Yuan, "On uplink non-orthogonal multiple access for 5G: Opportunities and challenges," *China Commun.*, vol. 14, no. 12, pp. 142–152, Dec. 2017.

[14] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.

[15] 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.

[16] C. Bockelmann *et al.*, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.

PETAR POPOVSKI (S'97–A'98–M'04–SM'10–F'16) received the Dipl.Ing. and Magister Ing. degrees in communication engineering from the Saints Cyril and Methodius University of Skopje, Skopje, and the Ph.D. degree from Aalborg University in 2005. He is currently a Professor of wireless communications with Aalborg University. He has over 300 publications in journals, conference proceedings, and edited books. He holds over 30 patents and patent applications. His research interests are in the area of wireless communication and communication theory. He received an ERC Consolidator Grant in 2015, the Danish Elite Researcher Award in 2016, the IEEE Fred W. Ellersick Prize in 2016, and the IEEE Stephen O. Rice Prize in 2018. He is currently a Steering Committee Member of IEEE SmartGridComm and previously served as a Steering Committee Member of the IEEE INTERNET OF THINGS Journal. He is currently an Area Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He is the General Chair for the IEEE SmartGridComm 2018 and the IEEE Communication Theory Workshop 2019.

KASPER FLØE TRILLINGSGAARD received the B.Sc. degree in electrical engineering, the M.Sc. degree in wireless communications, and the Ph.D. degree in electrical engineering from Aalborg University, Denmark, in 2011, 2013, and 2017, respectively. From 2017 to 2018, he was a Post-Doctoral Researcher at Aalborg University. He was a Visiting Student at the New Jersey Institute of Technology, NJ, USA, in 2012, and at the Chalmers University of Technology, Sweden, in 2014. In 2018, he joined InCommodities as a Data Scientist. His research interests are in the areas of information and communication theory.

OSVALDO SIMEONE (F'16) received the M.Sc. degree (Hons.) and the Ph.D. degree in information engineering from the Politecnico di Milano, Milan, Italy, in 2001 and 2005, respectively. From 2006 to 2017, he was a Faculty Member with the Electrical and Computer Engineering Department, New Jersey Institute of Technology, where he was affiliated with the Center for Wireless Information Processing. He is currently a Professor of information engineering with the Centre for Telecommunications Research, Department of Informatics, King's College London. He is a co-author of two monographs, an edited book published by Cambridge University Press, and more than one hundred research journal papers. His research interests include wireless communications, information theory, optimization, and machine learning. He is a fellow of the IET. He was a co-recipient of the 2017 JCN Best Paper Award, the 2015 IEEE Communication Society Best Tutorial Paper Award, and the Best Paper Awards of IEEE SPAWC 2007 and IEEE WRECOM 2007. He was a recipient of the Consolidator Grant by the European Research Council (ERC) in 2016. His research has been supported by the U.S. NSF, ERC, the Vienna Science and Technology Fund, as well by a number of industrial collaborations. He currently serves on the Editorial Board of the *IEEE Signal Processing Magazine*, and he is a Distinguished Lecturer of the IEEE Information Theory Society.

GIUSEPPE DURISI (S'02–M'06–SM'12) received the Laurea *(summa cum laude)* and Ph.D. degrees from the Politecnico di Torino, Italy, in 2001 and 2006, respectively. From 2006 to 2010, he was a Post-Doctoral Researcher at ETH Zürich, Zürich, Switzerland. In 2010, he joined the Chalmers University of Technology, Gothenburg, Sweden, where he is currently a Professor and Co-Director of the Chalmers information and communication technology Area of Advance.

His research interests are in the areas of communication and information theory and machine learning. He was a recipient of the 2013 IEEE ComSoc Best Young Researcher Award for the Europe, Middle East, and Africa Region, and has co-authored a paper that won the Student Paper Award at the 2012 International Symposium on Information Theory and a paper that won the 2013 IEEE Sweden VT-COM-IT Joint Chapter Best Student Conference Paper Award. In 2015, he joined the Editorial Board of the IEEE TRANSACTIONS ON COMMUNICATIONS as an Associate Editor. From 2011 to 2014, he served as the Publications' Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY.

• • •