Journal of Computer Engineering
ISSN 10003428

Original Research Article
Vol 14, Issue 11, 2025 | ISSN: 1000-3428

Optimizing Subband Configurations for 5G TDD New Radio Multiplexing Across Numerologies

Dr. Thomas R. Guskey
University of Kentucky, USA.

Abstract—The 5G New Radio (NR) access technology defines multiple numerologies to support a wide range of carrier frequencies, deployment scenarios, and variety of use cases. In this paper, we consider a resource allocation problem to efficiently support multiple numerologies simultaneously. We assume frequency division multiplexing (FDM) of numerologies in a time division duplex (TDD) system with a self-contained slot format. We focus on optimizing the numerology subband (SB) configuration, as well as the duplexing ratio between downlink (DL) and uplink (UL) directions within each SB. The optimization problem minimizes the weighted sum of the normalized load (NL) for each SB in each direction. We prove that our optimization problem is convex and, furthermore, we derive the optimal closed-form expressions for the numerology SB configuration and the DL-UL duplexing ratio per SB. The effectiveness of the proposed resource allocation is validated through an end-to-end ns-3 based simulator, which shows how the optimization of the NLs is translated into an improved throughput and delay performance.

I. Introduction

The 3rd Generation Partnership Project (3GPP) is devoting significant efforts to define the 5G New Radio (NR) access technology [1], which is expected to have flexible, scalable, and forward-compatible physical layer to support a wide range of center carrier frequencies, deployment options, and variety of use cases. The variety of NR use cases are categorized into three major services, i.e., enhanced Mobile BroadBand (eMBB), massive Machine Type Communications (mMTC), and Ultra-Reliable and Low-Latency Communications (URLLC) [2]. URLLC is associated to a strict latency target with certain outage probability, eMBB requires high data rates, and mMTC targets the support of massive connections and low energy consumption.

To achieve this flexibility, one of the key features of NR is the inclusion of a flexible orthogonal frequency division multiplexing (OFDM) system by means of multiple numerologies support [3]. Each numerology in NR is characterized by a subcarrier spacing (SCS) and a cyclic prefix (CP) overhead [1]. The selection of an appropriate numerology is required to fulfill the specific requirements of each of the services [4]. For example, a large SCS is suitable to reduce latency, which is appropriate for URLLC, while a short SCS is preferred to achieve high throughput performance, as required for eMBB traffic.

Therefore, for the simultaneous support of multiple use cases with different quality-of-service (QoS) requirements within the same channel bandwidth, NR allows the multiplexing of multiple numerologies in time-domain (time division multiplexing, TDM), as well as in frequency-domain (frequency division multiplexing, FDM) to achieve better user equipment (UE) performance [5]. Note that this NR flexibility facilitates the implementation of Radio Access Network (RAN) slicing, a key component to integrate network slicing in future cellular networks [6].

In case the numerologies are multiplexed in frequency domain, each numerology occupies a part of the whole channel bandwidth [7], which we refer to as a numerology subband (SB). It is stated in [8] that in case the base station properly configures the FDM of numerologies, TDM of numerologies is rarely needed. The advantages of flexible TDM of numerologies appear in case of a sudden need of large bandwidth for URLLC; however, it involves puncturing the resources already allocated for eMBB and indicating such preemption to recover the eMBB data correctly. Therefore, in this paper, we consider the multiplexing of numerologies only in the frequency domain and assume that there is a direct mapping between numerologies and services[1]. The FDM of numerologies requires methods to find

[1]

How to cite: University of Kentucky, USA. (2025). Optimizing Subband Configurations for 5G TDD New Radio Multiplexing Across Numerologies. Journal of Computer Engineering, 14(11), 14–25. DOI: https://doi.org/10.5281/zenodo.19359114

appropriate bandwidth distributions (i.e., numerology SB configurations) for all the supported services, which is one of the objectives of this paper.

Another important feature of NR is the self-contained slot [9], which is designed for significant latency reduction, e.g., reception of UL grant (in DL control) followed by the associated UL data transmission. Different slot formats for NR have been defined in [7]. They can possibly contain all DL, all UL, or at least one DL part and one UL part. Thus, in addition to the numerology SB configuration, in a TDD system, time resources should be properly distributed between DL and UL according to the DL-UL traffic asymmetries. In this paper, we also provide a solution for resource allocation between DL and UL within each numerology SB.

The 3GPP has agreed that the bandwidth parts can be configured statically or semi-statically [5], and that the allocation between DL and UL could either vary slot by slot (dynamic) or be configured semi-statically [7]. In this paper, we consider a semi-static configuration for both. That is, the numerology SB configuration and the DL-UL duplexing ratio: *i*) are semi-static enough to adapt to a significant change in the average traffic load of a specific service and DL/UL asymmetries, but *ii*) are not adapted to a concrete set of UEs to be scheduled at a specific time instant, as this would require changing the numerology SB configuration and DL-UL duplexing ratio as soon as a UE appears or disappears in the system. An efficient method must seek to use the spectrum resources efficiently by avoiding overprovisioning of resources, avoiding situations of high resource occupancy that leads to high packet delays, and providing QoS to the UEs [10]. A suitable measure that captures these requirements is the normalized load (NL), i.e., the ratio of the average traffic load and the capacity (amount of traffic that can be served). The NL is introduced and used in [10] to perform the resource allocation in a DL multi-cell scenario, assuming that all base stations operate with the same numerology[2].

In this paper, we derive a procedure to enable RAN slicing by properly configuring the FDM of numerologies, and show its effectiveness from an end-to-end
(E2E) perspective. We consider two different services, i.e., eMBB and URLLC, and formulate a problem to jointly optimize the numerology SB configuration to each service and the DL-UL duplexing ratio per SB. To the best of our knowledge, this is the first work that performs such joint optimization based on statistical traffic and system parameters. The optimization problem minimizes the weighted sum of NLs of all SBs and directions, where the weights allow assigning different priorities to different services. Closed-form solutions are derived for the numerology SB and DL-UL configuration. Finally, the E2E performance of the proposed resource allocation is assessed through an ns-3 network simulator that supports FDM of NR numerologies [11].

The remainder of the paper is organized as follows. Sections II and III describe the system model and problem formulation, respectively. In Section IV, we derive the solution for optimal numerology SB configuration and DL-UL duplexing ratio within every SB. The simulation results are presented in Section V, and Section VI concludes the paper.

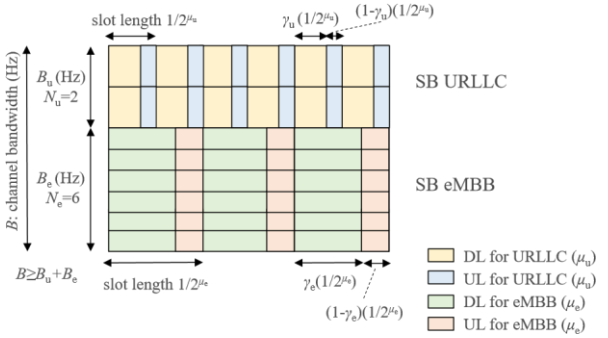

Fig. 1: Example of numerology SB configuration for a NR system that includes: *i*) FDM of two numerologies, μ_u and μ_e, to accommodate URLLC and eMBB traffics and *ii*) TDD to address DL and UL traffics per SB.

II. System Model

We consider a scenario consisting of a gNB (i.e., base station in NR) with an OFDM TDD system, where the multiplexing of different numerologies is allowed in frequency domain, and each slot is configured as selfcontained. Two different services, eMBB and URLLC are supported, each associated to one numerology configuration. However, the proposed problem formulation and the solution procedure can be easily generalized to accommodate more services with support for more than two numerologies, e.g., to accommodate eMBB, URLLC, and mMTC categories.

2

According to the 3GPP NR specifications [1], the numerology μ can take values from 0 to 4, each with a SCS of $15\times2^{\mu}$ kHz and a slot length of $1/2^{\mu}$ ms. The parameters that remain unchanged with the numerology are the number of OFDM symbols per slot, and the number of subcarriers per resource block (RB), which are set to 14, and 12, respectively. Thus, the RB width is $180\times2^{\mu}$ kHz, and the OFDM symbol length including CP overhead is equal to $1/(14\times2^{\mu})$, which are numerology-dependent. The URLLC traffic requires a short slot length to meet strict latency requirements, while the eMBB traffic requires higher throughput, which is achieved with a short SCS [3]. Therefore, among the set of supported numerologies for a specific band and deployment scenario[3], we assume that URLLC will use the numerology with the shortest slot length and eMBB will employ the numerology that is associated to the largest slot length [5].

We use μ_e and μ_u to denote the numerologies for eMBB and URLLC SBs, respectively. As shown in Fig. 1, the channel bandwidth B is split into two SBs of bandwidths B_e and B_u, with $B_e+B_u\leq B$, associated the two aforementioned numerologies, μ_e and μ_u, respectively. We denote the number of RBs in a slot for every SB as N_e for eMBB and N_u for URLLC, so that $B_e=180\times2^{\mu_e}N_e$ kHz and $B_u=180\times2^{\mu_u}N_u$ kHz. Assume that N is the number of RBs for the lowest numerology
(i.e., μ_e) that fit within the channel bandwidth B, i.e., $N=\lfloor\frac{B}{180\times10^3\times2^{\mu_e}}\rfloor$. Then, the condition $B_e+B_u\leq B$ may be translated into a condition over the number of RBs per numerology SB as:

$$N_e+N_u2^{(\mu_u-\mu_e)}\leq N. \quad (1)$$

For the example depicted in Fig. 1, $N=10$, $\mu_u-\mu_e=1$, $N_e=6$, and $N_u=2$. In this case, (1) is met with equality.

To support self-contained slots, we consider slots that are divided into a DL part and an UL part, where the partitioning may be different for every numerology SB. As shown in Fig. 1, let γ_e and γ_u denote the fraction of the slot that is allocated for DL in the eMBB SB, and URLLC SB, respectively, which satisfy $0\leq\gamma_e\leq1$ and $0\leq\gamma_u\leq1$. In both SBs, the remaining portions of the slots, i.e., $1-\gamma_e$, and $1-\gamma_u$, are allocated to UL.

We assume that the numerologies for eMBB and URLLC SBs are established based on the operational band and deployment option. Moreover, for each numerology SB, the SCS, RB width, slot length, and OFDM symbol length are given. Our objective is to determine the optimal numerology SB configuration, i.e., the number of RBs (N_e and N_u) that are allocated to each SB, and the DL-UL duplexing ratio, i.e., the number of OFDM symbols that are allocated to DL and UL within each SB, so that the condition (1) is met.

To achieve that, we consider optimization of the NL, i.e., the ratio of the average traffic load and the amount of traffic that the cell can serve. In our system, the NL is separately calculated for each type of service in each direction. The NL for the sth service (s can be either 'e' for eMBB or 'u' for URLLC) in the dth direction (d can be either 'DL' or 'UL'), ρ^d_s, is given by [12, Sect. 5.5.4]:

$$\rho^{DL}_e=\frac{\lambda^{DL}_e L^{DL}_e}{N_e\gamma_e n_e C^{DL}_e},\quad \rho^{UL}_e=\frac{\lambda^{UL}_e L^{UL}_e}{N_e(1-\gamma_e)n_e C^{UL}_e}, \quad (2)$$

$$\rho^{DL}_u=\frac{\lambda^{DL}_u L^{DL}_u}{N_u\gamma_u n_u C^{DL}_u},\quad \rho^{UL}_u=\frac{\lambda^{UL}_u L^{UL}_u}{N_u(1-\gamma_u)n_u C^{UL}_u}, \quad (3)$$

where λ^d_s and L^d_s denote the mean packet arrival rate (in packets/s) and the mean packet length (in bits/packet) of the sth service in the dth direction, respectively. In the denominator, N_s and $n_s=14000\times2^{\mu_s}$ denote the number of allocated RBs and the number of OFDM symbols within one second for the sth SB[4], respectively.

[4]Note that the sth service is uniquely associated to a single numerology SB, i.e., the sth SB, that has numerology μ_s. For the sth numerology SB, every subframe has 2^{μ_s} slots of 14 OFDM symbols each, and the subframe length is 1 ms.

Furthermore, for the sth numerology SB, the average spectral efficiency in the dth direction (in bits/resource) is denoted by C^d_s, and the slot fractions used for DL and UL are given by γ_s and $(1-\gamma_s)$, respectively.

In case that $\rho^d_s<1$, then ρ^d_s equals to the resource utilization (RU), i.e., the fraction of the resources that are occupied. The RU is a measure that is widely

[3]

used in 3GPP evaluations to report the percentage of resources employed by a cell [13]. Recall also that the proposed model supports multiple UEs per service, simply by considering their total traffic load in the numerator of the NLs in (2)-(3).

For simulations, we will assume user datagram protocol (UDP) constant bit rate (CBR) traffic model. For this traffic model, each flow is characterized by a constant packet length and packet arrival rate. This is also the case in file transfer protocol (FTP) traffic model, which is widely used in 3GPP to evaluate the system performance under bursty traffic conditions [14]. Therefore, these traffic scenarios could easily be handled by the proposed model. However, this does not limit the model to only these traffic scenarios, but any kind of traffic could be accommodated.

III. Problem Formulation

We consider the minimization of the weighted sum of NLs among the numerology SBs (eMBB/URLLC) and directions (DL/UL). The weighted sum allows to assign different priorities to each numerology SB. For example, the URLLC traffic is more sensitive to delays, so that it may require a higher priority than eMBB traffic to get a lower NL (or RU).

Before proceeding, note that the term $\lambda_s^d L_s^d/(n_s C_s^d)$ in (2)-(3) can be estimated based on the average traffic load and the average spectral efficiency for the sth service in the dth direction of a gNB. The later, C_s^d, can be estimated based on the statistics of the previously served users and their attained transmission rates (see [10] for further details and expressions to estimate it). Accordingly, if we set:

$$\alpha_s^d = \frac{\lambda_s^d L_s^d}{C_s^d n_s}, \tag{4}$$

the NL in (2)-(3) can be expressed in compact form as:

$$\rho_s^{\mathrm{DL}} = \frac{\alpha_s^{\mathrm{DL}}}{N_s \gamma_s}, \quad \rho_s^{\mathrm{UL}} = \frac{\alpha_s^{\mathrm{UL}}}{N_s(1-\gamma_s)}, \tag{5}$$

which depends on the optimization variables (N_s and γ_s, s={e,u}). It can be observed in (5) that increasing the amount of resources for the sth service (N_s) leads to a low NL and hence an inefficient usage of resources, as they could be used for other purposes. On the contrary, reducing N_s leads to a high NL, which increases the packet delay and reduces the QoS of the UEs requesting the sth service. Hence, appropriate methods for the numerology SB configuration should try to balance the frequency resource distribution and avoid very different NLs for the different numerology SBs and directions.

The minimization of the weighted sum of NLs among numerology SBs and directions, subject to the channel bandwidth constraint (1), is formulated as:

$$\underset{\{N_e, N_u, \gamma_e, \gamma_u\}}{\text{minimize}} \quad w_e\left(\rho_e^{\mathrm{DL}}+\rho_e^{\mathrm{UL}}\right) + w_u\left(\rho_u^{\mathrm{DL}}+\rho_u^{\mathrm{UL}}\right) \tag{6}$$

$$\text{subject to} \quad N_e + N_u 2^{(\mu_u-\mu_e)} \le N,$$
$$0 \le \gamma_e \le 1, 0 \le \gamma_u \le 1,$$

where w_e and w_u are the weights given to eMBB and URLLC traffic, respectively, and $\rho^{\mathrm{DL}}{}_e$, $\rho^{\mathrm{UL}}{}_e$, $\rho^{\mathrm{DL}}{}_u$, $\rho^{\mathrm{UL}}{}_u$ are given by (5).

The problem in (6) is a combinatorial optimization task that involves high complexity [15]. Note that NR Rel-15 supports up to 275 RBs within the channel bandwidth [7], which may likely be extended in NR Rel16 when moving to frequency bands with higher central carrier frequencies. This leads to many combinations under an exhaustive search method (brute force) that checks all plausible states. For that reason, we focus on solving the relaxed optimization problem with continuous variables (corresponding to the numerology SB configuration, $\{N_s\}$, and the DL-UL duplexing ratio, $\{\gamma_s\}$) and then we discretize the obtained result, as discussed in the next section.

Proposition 1: For continuous variables, the problem in (6) is jointly convex with respect to $\{N_e, N_u, \gamma_e, \gamma_u\}$.

Proof: Let us denote as f_s the part of the objective function in (6) related to the sth service, i.e., $f_s = w_s(\rho_s^{\mathrm{DL}}+\rho_s^{\mathrm{UL}}) = \frac{w_s}{N_s}\left(\frac{\alpha_s^{\mathrm{DL}}}{\gamma_s} + \frac{\alpha_s^{\mathrm{UL}}}{1-\gamma_s}\right)$. The Hessian matrix of f_s, H_s, is:

H

$$s = \begin{bmatrix} \frac{s}{N_s^2}\left(\frac{\alpha_s^{\mathrm{DL}}}{\gamma_s} + \frac{\alpha_s^{\mathrm{UL}}}{1-\gamma_s}\right) & \frac{s}{(1-\gamma_s)^2} \\ w\left(\alpha^{\mathrm{DL}} - \alpha^{\mathrm{UL}}\right) & 2w\left(\alpha^{\mathrm{DL}} + \alpha^{\mathrm{UL}}\right) \end{bmatrix} \quad (7)$$

For $N_s \geq 0$ and $0 \leq \gamma_s \leq 1$, H_s in (7) is a 2×2 real-valued matrix with non-negative diagonal elements and equal off-diagonal elements, which leads to a positive semidefinite matrix H_s by construction. Therefore, f_s is jointly convex with respect to N_s and γ_s. As the objective function in (6) is a separable function, say $f=f_e+f_u$, f is jointly convex with respect to $\{N_e,N_u,\gamma_e,\gamma_u\}$ within the constraint set. Finally, as the constraint set in (6) is linear and the objective function is jointly convex, the problem in (6) is jointly convex with respect to $\{N_e,N_u,\gamma_e,\gamma_u\}$. ■

IV. Numerology Subband Configuration and DL-UL Duplexing Ratio Optimization

Since the problem in (6) is jointly convex with respect to all the optimization variables, a single optimal solution exists. To solve (6), we use a two-step optimization:

1) DL-UL duplexing ratio optimization (time resources) within every SB, γ_e and γ_u, and
2) numerology SB configuration optimization (frequency resources) for each service, N_e and N_u.

As we will see, the DL-UL duplexing ratio per SB can be optimally obtained and turns out to be independent of the numerology SB configuration. Intuitively speaking, for any given system parameters and traffic statistics, the DL-UL duplexing ratio remains unchanged even though the SB configuration changes, since DL and UL traffics within a numerology SB have the same priority. Then, we will see that, by including the optimal DL-UL duplexing ratio into the global problem formulation in (6), the optimal numerology SB configuration can also be derived in closed-form.

A. DL-UL Duplexing Ratio

Assume $\{N_s\}$ is fixed per numerology SB. Since the DL-UL duplexing ratios of each SB (γ_e, γ_u) are not coupled between different numerology SBs (see (6)), they can be independently obtained for each SB. In particular, the optimal repartition of the time resources for the sth SB to problem (6), $\gamma_s^\star$, is obtained as the solution to:

$$\begin{aligned} \underset{\gamma_s}{\text{minimize}} \quad & \frac{w_s}{N_s}\left(\frac{\alpha_s^{\mathrm{DL}}}{\gamma_s} + \frac{\alpha_s^{\mathrm{UL}}}{(1-\gamma_s)}\right) \quad (8) \\ \text{subject to} \quad & 0 \leq \gamma_s \leq 1. \end{aligned}$$

As the objective function in (8) is the sum of two convex functions in the interval $\gamma_s \in [0,1]$, and the constraint is linear, the problem in (8) is convex with respect to γ_s [16]. Therefore, the solution can be obtained by setting the derivative of the objective function equal to 0, which leads to a closed-form expression for the DL-UL duplexing ratio on the sth SB ($s=\{e,u\}$):

$$\gamma_s^\star = \frac{\sqrt{\alpha_s^{\mathrm{DL}}}}{\sqrt{\alpha_s^{\mathrm{DL}}} + \sqrt{\alpha_s^{\mathrm{UL}}}}, \quad (9)$$

which lies within the constraint set. The optimal DL-UL ratio is independent of the SB configuration, $\{N_s\}$.

B. Numerology SB Configuration

Given the optimal DL-UL duplexing ratio expressions in (9), the optimization problem in (6) can be written as:

$$\begin{aligned} \underset{\{N_e \geq 0, N_u \geq 0\}}{\text{minimize}} \quad & w_e\frac{\beta_e}{N_e} + w_u\frac{\beta_u}{N_u} \quad (10) \\ \text{subject to} \quad & N_e + N_u 2^{(\mu_u-\mu_e)} \leq N, \end{aligned}$$

where β_e and β_u depend on the parameters $\{\alpha_s^d\}$ in (4), and are given by:

$$\beta_e = \frac{\alpha_{e?}^{\mathrm{DL}}}{\sqrt{\gamma_{e?}}} + \frac{\alpha_{e?}^{\mathrm{UL}}}{\sqrt{(1-\gamma_{e?})}}, \quad \beta_u = \frac{\alpha_{u?}^{\mathrm{DL}}}{\gamma_{u?}} + \frac{\alpha_{u?}^{\mathrm{UL}}}{(1-\gamma_{u?})}. \quad (11)$$

The problem in (10) is jointly convex with respect to N_e and N_u, since the objective function is jointly convex and the constraint is linear [16]. Furthermore, the optimal solution can be derived in closed-form by following the Lagrange duality method. Assume that the Lagrange multiplier is denoted φ, then, the Lagrangian function of the problem in (10) is given by [16]:

$$L = w_e\frac{\beta_e}{N_e} + w_u\frac{\beta_u}{N_u} + \phi\left(N_e + N_u 2^{(\mu_u-\mu_e)} - N\right). \quad (12)$$

Taking the derivatives of the Lagrangian function L in (12) and making them equal to 0 leads to:

$$\frac{\delta \mathcal{L}}{\delta N_{\mathrm{e}}} = -\frac{w_{\mathrm{e}}\beta_{\mathrm{e}}}{N_{\mathrm{e}}^2} + \phi = 0, \qquad (13)$$

$$\frac{\delta \mathcal{L}}{\delta N_{\mathrm{u}}} = -\frac{w_{\mathrm{u}}\beta_{\mathrm{u}}}{N_{\mathrm{u}}^2} + \phi 2^{(\mu_{\mathrm{u}}-\mu_{\mathrm{e}})} = 0 \qquad (14)$$

which gives the solution for $\{N_s\}$ as:

$$N_{\mathrm{e}} = \sqrt{\frac{w_{\mathrm{e}}\beta_{\mathrm{e}}}{\phi}}, \quad N_{\mathrm{u}} = \sqrt{\frac{w_{\mathrm{u}}\beta_{\mathrm{u}}}{\phi 2^{(\mu_{\mathrm{u}}-\mu_{\mathrm{e}})}}}. \qquad (15)$$

By including (15) into the constraint of problem (10), and setting the equality[4], we obtain φ:

$$\phi = \left(\frac{\eta}{N}\right)^2, \quad \eta = \sqrt{w_{\mathrm{e}}\beta_{\mathrm{e}}} + \sqrt{w_{\mathrm{u}}\beta_{\mathrm{u}}2^{(\mu_{\mathrm{u}}-\mu_{\mathrm{e}})}}. \qquad (16)$$

Therefore, by combining (15) and (16), the optimal solution for the numerology SB configuration, $N_s^?$, is given by:

$$N_{\mathrm{e}}^? = \frac{\sqrt{w_{\mathrm{e}}\beta_{\mathrm{e}}}}{\eta} N, \qquad N_{\mathrm{u}}^? = \frac{\sqrt{w_{\mathrm{u}}\beta_{\mathrm{u}}}}{2^{(\mu_{\mathrm{u}}-\mu_{\mathrm{e}})}\eta} N. \qquad (17)$$

Note that the optimal solution in (17) gives more resources to the SB experiencing a higher weighted ratio between the average traffic load and the average spectral efficiency (i.e., higher $w_s\beta_s$). Also, a penalizing term, $2^{(\mu_{\mathrm{u}}-\mu_{\mathrm{e}})}$, in the optimal $N_{\mathrm{u}}^?$ appears because the RB width for URLLC SB is larger than the RB width for eMBB SB.

C. Mapping the solution into real resources

The optimal SB configuration and DL-UL duplexing ratios have been derived as continuous portions of the available frequency spectrum and slot lengths. However, in practical systems the resource distribution is done as an integer multiplier of the minimum allowed unit of the resource allocation. Thus, the optimal values found in the previous subsections ($N_{\mathrm{e}}^?$,$N_{\mathrm{u}}^?$ in (17) and $\gamma_{\mathrm{e}}^?$,$\gamma_{\mathrm{u}}^?$ in (9)) should be mapped accordingly.

In the frequency domain, the RBs for eMBB and URLLC SBs (N_{e},N_{u}) should be integer numbers that satisfy the constraint $N_{\mathrm{e}}+N_{\mathrm{u}}2^{(\mu_{\mathrm{u}}-\mu_{\mathrm{e}})} \le N$. In the time domain, since there are 14 OFDM symbols in a slot to be split between DL and UL on every SB, the slot fractions γ_{e},γ_{u} should be a multiple of 1/14 and satisfy $0 \le \gamma_s \le 1$, $s=\{\mathrm{e},\mathrm{u}\}$. The round down (i.e., take the floor) would always satisfy the conditions. However, any other rounding that satisfies the conditions is also valid.

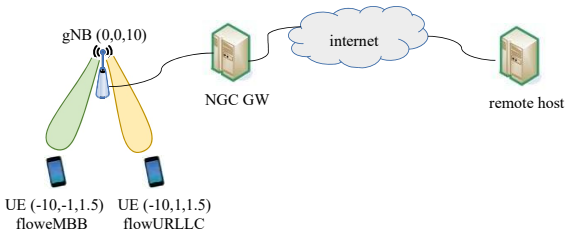

Fig. 2: Scenario for E2E evaluation in ns-3.

V. Simulation Results

In order to evaluate the performance of the proposed solutions in Section IV, we use ns-3 [17], an opensource discrete-event network simulator that allows fullstack simulations. In particular, we use a simulator developed by CTTC that was built over the mmWave ns3 module [18] to address communications in mmWave bands through beamforming methods, which is also extended to support the NR frame structure and FDM of numerologies (see implementation details in [11]). The ns-3 simulator provides abstraction at the physical layer and high-fidelity models at higher layers. We use UDP CBR traffic and RLC-UM (unacknowledged mode) to avoid retransmissions at both link and transport layers, and then simulate different loads by tuning the traffic intensity of the UDP flows. Please note that the proposed solution in Section IV considers both DL and UL traffic. However, in this section, for better understanding of the results, we assess the performance with only DL traffic and focus on evaluating the optimal SB configuration.

As shown in Fig. 2, we consider a gNB that provides wireless access to two UEs. One UE demands eMBB traffic, and the other UE requests an URLLC traffic flow. Each service is mapped into a SB with a different numerology, which are fixed to: μ_{e}=2 (SCS=60 kHz, normal CP), μ_{u}=4 (SCS=240 kHz, normal CP) [1]. A channel bandwidth B=200 MHz at 28 GHz carrier frequency (mmWave band) is used. It gives N=277 RBs with μ_{e}=2 to fit within

[4]

the channel bandwidth. Equal average spectral efficiencies within the two numerology SBs is assumed, i.e., $C_e^{DL}=C_u^{DL}$.

We use two different fixed eMBB loads, for which the mean packet length is $L^{DL}_e=1280\times8$ bits/packet and the mean packet arrival rate is fixed either to $\lambda^{DL}_e=12500$ packets/s or $\lambda^{DL}_e=37500$ packets/s. The former leads to an eMBB load of 128 Mbps, and the later to 384

Mbps. For each fixed eMBB load, we vary the URLLC payload, for which $L_u^{DL}=128\times8$ bits/packet and λ_u ={1250,62500,125000,250000,375000,500000} packets/s. We assume that the first OFDM symbol and the last OFDM symbol in a slot are reserved for DL control and UL control, respectively. Since only the DL traffic is emulated, we fix $\gamma_e=12/14$, $\gamma_u=12/14$, and focus on optimizing the SB configuration for the two services. We consider two strategies to compare our results:

- uniform SB configuration (uniSB): the total bandwidth is uniformly distributed between the two SBs, i.e., $B_e=100$ MHz ($N_e=138$ RBs) and $B_u=100$ MHz ($N_u=34$ RBs).
- optimized SB configuration (optSB): the total bandwidth is split between URLLC and eMBB traffics according to the optimization in (17), for which $B_u^?=180\times2^{\mu_u}N_u^?$ kHz and $B_e^?=180\times2^{\mu_e}N_e^?$ kHz. We use the following weighting coefficients: $w_e=0.4$, $w_u=0.6$ (i.e., higher priority to URLLC).

The optimization of the SB configuration in (17) is performed beforehand, based on μ_s, N, $L^{DL}{}_s$, $\lambda^{DL}{}_s$, C_s^{DL}, γ_s, w_s, s={e,u}, and then the system performance is assessed through ns-3 E2E evaluations with the obtained SB configuration.

For the evaluation, a deployment of one gNB at position (0,0,10) and two UEs at positions (−10,1,1.5) and (−10,−1,1.5) is considered, where (x,y,h) indicate $x$position, y-position, and h-height (in meters), as shown in Fig. 2. An Urban Micro (UMi) propagation model is used at 28 GHz band. The number of antennas at gNB and UEs are set to 64 and 16, respectively. The gNB has a total available power of 4 dBm that is uniformly distributed within the B=200 MHz channel bandwidth. Therefore, for the uniform SB configuration strategy, 1 dBm is available for each SB. For the optimized SB configuration strategy, the power available per SB is adjusted according to the optimized SB configuration. An optimal beamforming method (i.e., long-term covariance matrix method) is assumed, for which the beamforming vectors are taken as the maximal eigenvectors of the channel covariance matrices [18]. Adaptive modulation and coding scheme is used. The L2-L1 (layer 2 and layer 1) processing delays are fixed to 2 slots, and the UE decoding time is set to 100 us. RLC queues are of 1 GB. The simulation time is 10 s.

Two UDP CBR flows in DL are configured (from remote host to UE), one for each UE, with the aforementioned mean packet lengths and packet arrival rates. As key performance indicators we consider the throughput per UDP flow (in Mbps) and the mean delay of the packets per UDP flow (in ms).

Fig. 3 and Fig. 4 show the throughput and the mean delay per UDP flow for different URLLC loads, respectively, for an eMBB load of 128 Mbps. Fig. 5 and Fig. 6 depict the same performance metrics with a larger eMBB load, i.e., 384 Mbps.

For the uniform SB configuration strategy, the performance of the eMBB flow does not vary with the

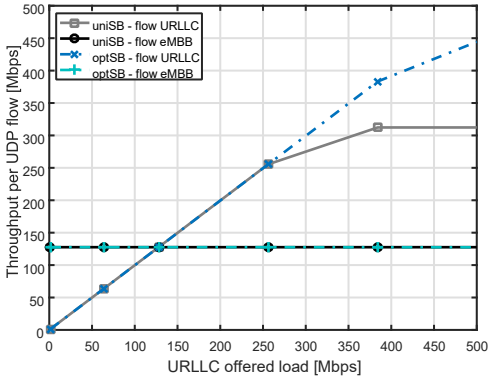

Fig. 3: Throughput per UDP flow vs. URLLC load. eMBB load=128 Mbps.

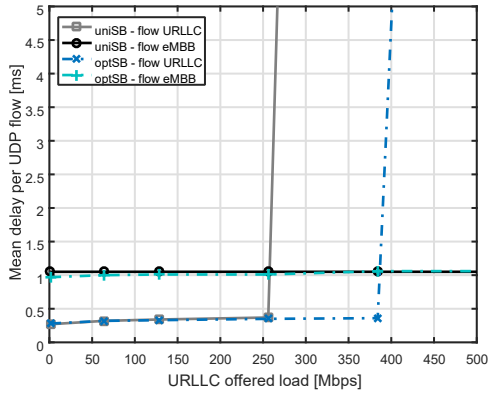

Fig. 4: Mean delay per UDP flow vs. URLLC load. eMBB load=128 Mbps.

URLLC load due to the static spectrum distribution and fixed eMBB load. Under this strategy, for every eMBB load, a different system behavior is observed. For an eMBB load of 128 Mbps, eMBB traffic can fit within the bandwidth allocated to eMBB SB (see low delay in Fig. 4). However, for an eMBB load of 384 Mbps, eMBB traffic can not fit within the bandwidth for eMBB SB, which becomes saturated, and leading to high delays (see Fig. 6). Note that 'uniSB - flow eMBB' delay does not appear in Fig. 6, since it is much larger than 5 ms.

When the eMBB SB is not saturated with the uniform SB configuration strategy (see Fig. 3 and Fig. 4), we observe that increasing the URLLC load leads to saturation of the URLLC SB. In this situation, the optimized SB configuration strategy provides an improved throughput and reduced mean delay for URLLC traffic, since the proposed strategy is able to properly redistribute the spectrum to fit all the traffic loads.

When the eMBB SB is already saturated with the uniform SB configuration strategy (see Fig. 5 and Fig. 6), at low URLLC loads, the attained throughput and mean delay of the eMBB traffic are improved thanks to the proposed optimized SB configuration strategy. Notably, a significant delay reduction is obtained for eMBB traffic (see Fig. 6), which means that none of the

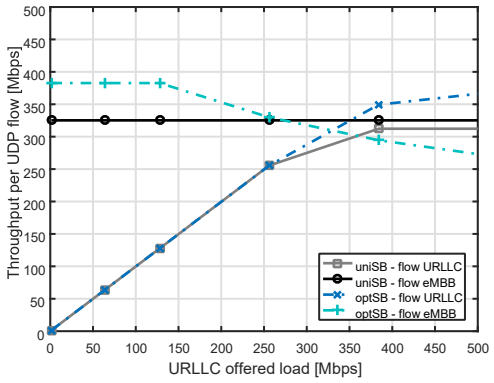

Fig. 5: Throughput per UDP flow vs. URLLC load. eMBB load=384 Mbps.

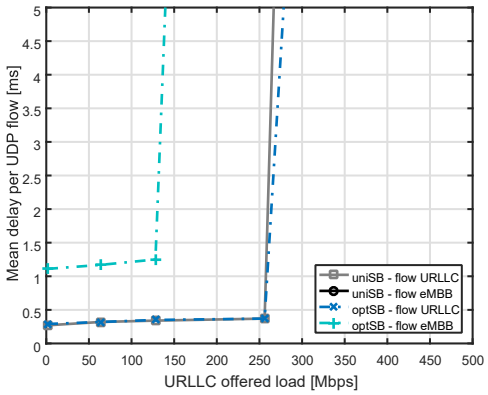

Fig. 6: Mean delay per UDP flow vs. URLLC load. eMBB load=384 Mbps.

SBs are saturated and all the load that arrives to gNB is successfully delivered without saturating the RLC buffers. As the URLLC load increases, the optimized SB configuration strategy gives more resources to the URLLC SB due to its higher priority. Therefore, in this case, the URLLC traffic obtains a higher throughput and a lower mean delay as compared to the uniform bandwidth distribution case (see Fig. 5 and Fig. 6). It comes at the cost of having a lower throughput and higher delay for eMBB, because the system is fully saturated. In this regime, the weighting coefficients allow trading-off in between the throughput/delay performance of the different SBs.

VI. Conclusions

This paper proposes a procedure to optimize the numerology SB configuration and the DL-UL duplexing ratio in a TDD self-contained NR system that has multiple SBs to multiplex different numerologies and accommodate different services (e.g., eMBB and URLLC). We focus on minimizing the weighted sum of the normalized loads, for which the optimal distribution of the resources per SB and direction are obtained. The E2E simulation results through ns-3 network simulator show that the proposed optimization of the NLs translates into an improvement of the throughput and an effective reduction of mean delay per UDP flow, when a SB is saturated with a uniform spectrum allocation strategy. This demonstrates the effectiveness of a properly configured FDM of numerologies from an E2E perspective, which could help to implement RAN slicing and deploy 5G NR networks in the future.

Future work includes the optimization of the transmit power used for every SB, as well as a mixed optimization with a semi-static configuration of the numerology SBs and a dynamic update of the DL-UL duplexing ratio per SB. Also, an interesting research area is to extend the framework to incorporate QoS requirement of the different users for every slice.

REFERENCES

[1] 3GPP TS 38.300, *TSG RAN; NR; NR and NG-RAN overall description; Stage 2*, Release 15, v15.1.0, Mar. 2018.

[2] 3GPP TR 38.913, *TSG RAN; Study on scenarios and requirements for next generation access technologies*, Release 14, v14.3.0, June 2017.

[3] A. A. Zaidi *et al.*, "Waveform and numerology to support 5G services and requirements," *IEEE Commun. Mag.*, vol. 54, pp. 90–98, Nov. 2016.

[4] A. Yazar and H. Arslan, "A flexibility metrix and optimization methods for mixed numerologies in 5G and beyond," *IEEE Access*, vol. 6, pp. 3755–3764, Jan. 2018.

[5] 3GPP R1-1707238, TSG RAN WG1 88bis Meeting, Vivo, *Discussion on NR resource allocation*, May 2017.

[6] 3GPP TR 28.801, *TSG SSA; Telecommunication management; Study on management and orchestration of network slicing for next generation network*, Release 15, v15.1.0, Jan. 2018.

[7] 3GPP TS 38.211, *TSG RAN; NR; Physical channels and modulation*, Release 15, v15.1.0, Mar. 2018. [8] 3GPP R1-1611515, TSG RAN WG1 87 Meeting, Ericsson, *Puncturing sTTI in legacy TTI*, Nov. 2016.

[9] "The 5G unified air interface," *Qualcomm Technologies, Inc.*, Nov. 2015.

[10] S. Lagen, O. Munoz, A. Pascual, J. Vidal, and A. Agustin, "Long-˜ term provisioning of radio resources based on their utilization in dense OFDMA networks," *IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun.*, Sep. 2016.

[11] B. Bojovic, S. Lagen, and L. Giupponi, "Implementation and evaluation of frequency division multiplexing of numerologies for 5G new radio in ns-3," in *Proceedings of the 2018 Workshop on ns-3*, June 2018.

[12] A. Kumar, D. Manjunath, and J. Kuri, *Communication Networking: an analytical approach*. Morgan Kaufmann Publishers, Elsevier, 2004.

[13] 3GPP TR 36.872, *TSG RAN; Small cell enhancements for EUTRA and E-UTRAN - physical layer aspects*, Release 12, v12.1.0, Dec. 2013.

[14] 3GPP TR 36.814, *E-UTRA; Further advancements for E-UTRA physical layer aspects*, Release 9, v9.2.0, Mar. 2017.

[15] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey," *IEEE Commun. Surveys & Tutorials*, vol. 3, pp. 10–31, Apr. 2000.

[16] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, 2004.

[17] "ns-3 Network Simulator." http://code.nsnam.org/ns-3-dev, 2018.

[18] M. Mezzavilla *et al.*, "End-to-end simulation of 5G mmWave networks," *IEEE C ommun. Surveys & Tutorials*, Apr. 2018.