IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 28, 2021, accepted January 7, 2022, date of publication January 12, 2022, date of current version January 20, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3142430

Fog-RAN Enabled Multi-Connectivity and Multi-Cell Scheduling Framework for Ultra-Reliable Low Latency Communication

BINOD KHAREL, (Graduate Student Member, IEEE), ONEL L. ALCARAZ LÓPEZ, (Member, IEEE), NURUL HUDA MAHMOOD, (Member, IEEE), HIRLEY ALVES, (Member, IEEE), AND MATTI LATVA-AHO, (Senior Member, IEEE)
Centre for Wireless Communications (CWC), University of Oulu, 90570 Oulu, Finland

Corresponding author: Binod Kharel (binod.kharel@oulu.fi)

This work was supported by the Academy of Finland 6Genesis Flagship under Grant 318927.

ABSTRACT Ultra-Reliable Low Latency Communication (URLLC) is a newly introduced service class targeting emerging Internet-of-Things (IoT) application scenarios. This paper assumes an interference-limited Fog Radio Access Network (F-RAN) setup composed of multiple Remote Radio Heads (RRHs) equipped with multiple antennas serving single-antenna users. F-RAN facilitates collaborative solutions while reducing delay by pushing the network capabilities beyond the edge. By leveraging diversity, RRHs may cooperate through silencing, reducing interference, or joint transmission strategies such as maximal ratio transmission. We derive closed-form outage probability expressions and attain their diversity gain. We validate the derived analytical results through extensive numerical simulations. Furthermore, we propose a mini-slots-based scheduling framework to serve URLLC users within their fixed latency budget. In an interference-limited regime with the proposed scheduling framework, we show that a performance gain is superior when RRHs cooperate compared to when they do not. We briefly discuss the cost of reliability, i.e., the impact on the system's average sum throughput under cooperation. Moreover, numerical results verify that cooperating transmission schemes boost transmission reliability with a significantly improved latency performance at the cost of reduced system's average sum throughput.

INDEX TERMS Diversity, F-RAN, maximal ratio transmission, machine-type communication, multi-connectivity, reliability, scheduling, silencing, ultra-reliable low latency communication.

I. INTRODUCTION

Many emerging applications in the domain of the Internet of Things (IoT) require efficient machine-type communications (MTC) to interconnect wirelessly without the need for human intervention [1]. Indeed, fifth-generation (5G) wireless communication systems have categorized MTC to address two main services: massive MTC (mMTC) and URLLC [2]. In this work, we focus on the URLLC, which is the service class aiming to meet stringent reliability and latency requirements in 5G New Radio (NR) [3]. In this context, several mission-critical applications require URLLC services, e.g., factory automation [4], process automation [4], intelligent transportation systems [4], automated guided vehicles (AGV) [5] and smart grids [6]. These applications require high reliability (e.g., 10^{-5} to 10^{-9} outage probability depending upon the application) and simultaneously latency budgets of few milliseconds [6].

We can study URLLC from two aspects: Ultra-Reliable Communication (URC) and Low-Latency Communication (LLC). URC technique is possible via diversity, including Multi-Connectivity (MC) and robust physical layer design.

The associate editor coordinating the review of this manuscript and approving it for publication was Faissal El Bouanani.

In contrast, LLC via flexible numerology, grant free instant uplink, and fast processing [7]. Some significant challenges and critical technology components related to ultra-reliability are enhanced control channel reliability, link adaptation, interference mitigation, and coexistence with other higher data rate services such as Enhanced Mobile Broadband (eMBB) [8]. However, the interplay between the diverse URLLC requirements makes the physical layer design of such systems highly complex [9].

MC is identified as a critical URLLC enabler in 5G systems. MC adopts spatial diversity, which can be enabled via centralized processing in the F-RAN architecture to ensure ultra-reliability through collaborative solutions [10]. In the Fog Radio Access Networks (F-RAN), a large amount of signal processing and computing is performed in a distributed manner. At the same time, Access Points (APs), e.g., RRHs or fog APs, integrate radio frequency and signal processing functionalities which are beneficial for interference management and radio resource allocations [11]. APs are connected to a Centralized Unit (CU) in the fronthaul via high-speed optical fiber links that can support low latency and high capacity communication, both of which lead to improving network performance [12]. F-RAN has four modes of operations: global centralized mode, locally distributed mode, high power node mode, and device-to-device mode. We focus on the F-RAN enabled global centralized mode where collaborative radio signal processing and radio resource management functions are implemented centrally at the Baseband Unit (BBU) pool [13].

This paper aims to establish an MC framework through RRH cooperation strategies enabled by F-RAN. The proposed F-RAN model presented in this paper represents indoor Industrial IoT (IIoT) scenarios and is similar to the one considered by 3GPP Release 16 [14]. For example, the factory setting scenario consists of CU with BBU as a controller that enables the F-RAN with storage and computation capabilities at the edge and the User Equipment (UE) as an actuator. We exploit MC through the Maximum Ratio Transmission (MRT) scheme or reducing interference through silencing in interference-limited downlink cellular networks.

A. RELATED LITERATURE

Much literature related to URLLC appeared after the seminal work by Popvoski [15]. Furthermore, Popvoski *et al.,* discuss the communication theoretic principles for supporting URLLC in [16], [17], such as the use of various diversity resources, design of packets, and access protocols. In [18], the authors investigate different diversity sources, e.g., time, frequency, and space, to meet the challenging requirements of URLLC. However, due to critical latency and bandwidth constraints of IoT applications, the diversity gain from the frequency and time domains are very limited to overcome the possible deep fading caused by shadowing [19]. Hence, spatial diversity may often be more attractive.

The authors in [20] detail the comprehensive study and importance of spatial diversity in wireless communication. In spatial diversity, multiple antennas are physically separated from each other transmit to the user. MC adopts spatial diversity where more than one connection jointly serves the UE. The basic idea is to send replicas of the same message through more than one link. If one of them is decoded successfully, then the packet is received [21]. For instance, [22] studies MC as one of the sources to improve reliability, enabling the transmission of redundant data through multiple links using standard diversity schemes like joint decoding, selection combining, and maximum ratio combining. In [23], [24], the authors study the potential of diversity and interference management techniques to achieve ultra-reliability operation. Furthermore, [25] proposes to interface diversity where each interface is based on different technologies to offer URLLC without intervention in baseband/physical layer design. Also, cooperative diversity emerges as a workable alternative to direct communication [26]. On the other hand, such reliability gain comes at the cost of transmission of redundant packets, leading to an increase in radio-resource consumption [27]. In [28], the authors propose an MC concept in edge RAN to reduce mobility-related link failures and cell-edge degradation. Likewise, [29] discusses centralized RAN technology to support URLLC, lowering the traffic latency by a functional split of the central and radio units. The primary key difference between cloud and fog networking is that data is processed at the edge nodes in the latter. In contrast, in the former, the data is processed at CU.

Furthermore, the flexible 5G frame structure and numerology in 5G NR provides a scalable and configurable air interface design to support low-latency transmissions with the mini-slots down to two symbols in duration as defined by 3GPP Release 15 [30]. Different works [31]–[34] adopt the concept of mini-slot transmission to satisfy latency constraints for URLLC while proposing different scheduling and resource utilization techniques in multi-cell 5G networks.

In this study, we also adopt the concept of a mini slots-based URLLC user scheduling framework to reduce the data Transmission Time Interval (TTI) and schedule the URLLC users within the target latency budget. The main difference with the works above is that we analyze the system performance in an interference-limited scenario with different modes of cooperative transmission schemes enabled by F-RAN and discuss cooperation gain. Although this paper exhibits some similarities in the system model assumptions concerning [22], herein, we pay close attention to spatial diversity methods considering interference and cooperation of spatially distributed RRHs enabled by the F-RAN.

B. OUR CONTRIBUTIONS

There are three significant technical challenges from a physical layer URLLC system design perspective. They are first, minimizing the system overhead in terms of channel access, user scheduling, and allocation of resources. Second, outage probability should be reduced in the first transmission as retransmission can affect the latency. Third, the transmission of URLLC packets as soon as possible to reduce latency.

Therefore, this paper aims to design an F-RAN framework to serve a low-latency UE with an ultra-reliable transmission scheme in an interference-limited downlink scenario. The main contributions of this article are summarized as follows:

- We identify and analyze diversity techniques like silencing and MRT schemes for URC in an interference-limited downlink communications scenario. A network with fully connected RRHs, i.e., a global centralized mode, is assumed where all the RRHs are connected to CU where RRHs might cooperate or may not their transmission through BBU pool. We analytically formulate the system reliability in terms of outage probability when RRHs cooperate through BBU via silencing or MRT. Exact closed-form outage probability expressions are attained, and we verify them through numerical simulations. We show the superiority of MRT over silencing, mainly supported by the fact that a smaller number of cooperating RRHs are required to achieve the desired reliability.
- In addition, we study the asymptotic behaviour of outage probability for MRT and silencing scheme in terms of diversity gain. The analysis shows that the MRT scheme provides $k+1$ times higher diversity gain than the silencing scheme under cooperation.
- We propose a mini slots-based scheduling framework to serve URLLC users within their latency budget. We discuss performance gains of different transmission schemes in terms of percentage of user served, mean latency, and the average number of transmission/cooperation through extensive computer-level simulations. We show that adopted MRT and silencing cooperation schemes reduce the interference on F-RAN and enhance the network performance by successfully serving a higher number of URLLC users active at random positions within the given area.
- Finally, we discuss the impact of cooperation on overall network performance by analyzing the trade-off between average system sum throughput and reliability through computer-level simulations for MRT and silencing schemes. We show that with cooperation, both schemes achieve URC operation while the system's average sum throughput gradually decreases.

Section II introduces the system model. Section III details the investigated transmission schemes, and provides signal-to-interference ratio (SIR) derivations. Section IV provides the outage probability formulation and throughput reliability trade-off analysis, while Section V presents the scheduling framework. Section VI shows performance evaluation and results. Finally, Section VII concludes the paper.

Table 1 provide the list of acronyms used throughout in this paper in an alphabetical order. Throughout this paper, the superscript H denotes the complex conjugate transpose, and a boldface lowercase letter denotes a column vector. $\mathbb{C}$ denotes complex domain. $||a||$ denotes the norm of complex valued vector a, i.e., $||a|| = \sqrt{a^H a}$, while $\mathbb{E}[\cdot]$ denotes the expectation operation and $\Gamma(.)$ denotes the gamma function.

TABLE 1. Table of acronyms.

Symbol	Description
AP	Access Point
BBU	Baseband Unit
CC	Chase Combining
CSI	Channel State Information
CU	Centralized Unit
eMBB	Enhanced Mobile Broadband
F-RAN	Fog Radio Access Network
HARQ	Hybrid Automatic Request
IIoT	Industrial Internet of Things
IoT	Internet of Things
LLC	Low Latency Communication
mMTC	Massive MTC
MTC	Machine Type Communication
NR	New Radio
RTT	Round Trip Time
SIR	Signal-to-Interference Ratio
TTI	Transmission Time Interval
UE	User Equipment
URC	Ultra-Reliable Communication
URLLC	Ultra-Reliable Low Latency Communication
3GPP	3rd Generation Partnership Project
5G	Fifth Generation Networks

TABLE 2. Summary of notations.

Notation	Description
N	Total number of RRHs
k	Total number of cooperating RRHs
c	Set of cooperating RRHs defined as $0 \dots k$
i	Set of interfering RRHs defined as $k+1 \dots N$
M_t	Number of transmit antennas
$d_{i,j}$	Distance from RRH i to UE j
$\mathbf{w}_{i,j}$	Transmit beamformer vector from RRH i to UE j
$\mathbf{h}_{i,j}$	Channel gain vector from RRH i to UE j
s_j	Information symbol of UE j
$\mathbf{z}_j$	Circular symmetric complex Gaussian noise vector
α	Path loss exponent
γ	Signal-to-interference ratio
θ	Signal-to-interference ratio threshold
D	Diversity gain
TP	Average system sum throughput
R	Average sum rate of all active users in the system
K_{max}	Maximum number of available RRHs
k_{min}	The minimum number of RRHs in cooperation
N_{serv}	Total number of URLLC users to be served
N_{re-tx}	Total number of URLLC users in retransmission
τ	Critical latency threshold

Uppercase and lowercase letters denote random variables (RVs) and their realizations. The probability density function (PDF) and cumulative distribution function (CDF)

of RV X is denoted by $f_X(x)$ and $F_X(x)$, respectively. For ease of reference, Table 2 summarizes all the important notations used throughout this paper.

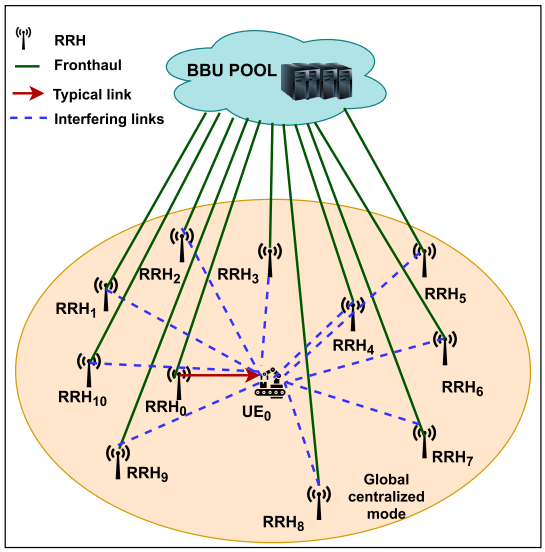

FIGURE 1. Illustration of the system architecture of F-RAN in global centralized mode with a typical link and $N = 10$ interfering RRHs.

II. SYSTEM MODEL

We consider an F-RAN system model as illustrated in Fig. 1, where all RRHs are connected to an edge cloud consisting of a baseband unit (BBU) pool via fronthaul links with the high-bandwidth and low latency communication. We assume that all RRHs are using the same spectrum resources (i.e., time and frequency) when transmitting to their corresponding UEs. In the setup, we assume that $N+1$ RRHs, i.e., RRH_0, RRH_1, ..., RRH_N are spatially distributed in a given area $\mathcal{A} \subseteq \mathbb{R}^2$. The link between RRH_0 and UE_0 is considered as a *typical link*. Meanwhile, the link from other RRH_i to UE_0 for all $i \in \{1, \ldots, N\}$ can be interfering or cooperating links to the typical link. We consider cooperating links when several RRHs cooperate to the typical link in transmission to the UE_0 while the rest of the RRHs are interfering to the typical link are interfering links.

We assume that RRHs are equipped with M_t transmit antennas and UE with single antenna. Let $d_{i,j}$ and $\mathbf{w}_{i,j} \in \mathbb{C}^{M_t \times 1}$ denote the distance and transmit beamformer from RRH_i to UE_j, respectively. Furthermore, let $\mathbf{h}_{i,j} \in \mathbb{C}^{M_t \times 1}$ denote the channel vector (small-scale fading) between RRH_i to UE_j. Then, the received signal vector at j^{th} UE is expressed as

$$\mathbf{r}_j = \underbrace{d_{0,j}^{\frac{-\alpha}{2}} \mathbf{h}_{0,j}^H \mathbf{w}_{0,j} s_j}_{\text{typical link}} + \underbrace{\sum_{c=1}^{k} d_{c,j}^{\frac{-\alpha}{2}} \mathbf{h}_{c,j}^H \mathbf{w}_{c,j} s_j}_{\text{cooperating links}} + \underbrace{\sum_{i=k+1}^{N} d_{i,j}^{\frac{-\alpha}{2}} \mathbf{h}_{i,j}^H \mathbf{w}_{i,i} s_i}_{\text{interfering links}} + \mathbf{z}_j, \quad (1)$$

where α is a path loss exponent, s_j is the information symbol, and $\mathbf{z}_j$ is circular symmetric complex Gaussian noise vector. Also (1) assumes that k closest RRHs are cooperating with the transmission of the typical link, and the rest of the RRHs are interferers. Note that k is the number of cooperating RRHs. Furthermore, assuming a dense RRHs deployment, we can neglect the noise impact since interference is dominant [35]. Hence, in the subsequent analysis of this paper, we neglect the noise.

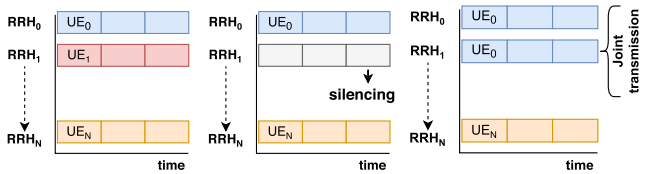

FIGURE 2. Mode of transmission with/without RRHs' cooperation.

III. TRANSMISSION SCHEMES AND SIR DERIVATION

We analyze the typical link performance when F-RAN operates under two modes when serving the UEs.

- **without cooperation:** As shown in Fig 2(a), the interfering RRHs do not cooperate with the typical link through F-RAN BBU, and each RRH transmits to its own UE. Under this condition, each UE experiences full interference from neighboring RRHs.
- **cooperation:** In this mode, neighboring RRHs cooperate with the typical link through the F-RAN BBU pool to serve the desired UE to fulfill user-centric objectives like high reliability and low latency. The main benefit of the F-RAN framework enables multi-point transmission and cooperative solutions which coordinate data transmission to the typical UE pushing network capabilities beyond the edge, reducing the delay, leveraging less burden on the BBU pool.

In the cooperation mode, we consider one of the following transmission strategies to serve the desired UE:

- **Silencing:** F-RAN silences some of the strongest interferings RRHs to mitigate the interference at the reference user, as shown in Fig. 2(b). At a silenced RRH, the data channel, control channel, and reference signals are entirely turned off, which enhances the reliability of a typical link as silencing or muting a strong interferer helps to boost the SIR in the victim cell, and it has been proposed for 5G [36], [37].
- **MRT:** Several RRHs jointly coordinate the transmission to the reference UE, as shown in Fig. 2(c). The duplicate packets are sent from multiple RRHs independently, which gives redundancy against fading, blocking, or radio link failure [38]. The joint transmission through MRT gives the best reception reliability enabling multi-point MC. Also, sending the same message through independent transmissions from multiple RRHs saves

resources, which are very limited due to the stringent latency constraints in URLLC.

Similar to the work in [39] for the above-mentioned transmission schemes, we assume that each data packet is transmitted once to enable low latency transmission (i.e., one-shot transmission is considered from RRH to the desired UE.). The optimal beamformers are computed by conjugate beamforming, assuming local Channel State Information (CSI) at each RRH, which avoids explicit control and extra signalling information exchange among cooperating RRHs. Thus, the aforementioned schemes do not incur signalling and hence are feasible for upcoming 5G systems and services [40]. We focus our attention on the above transmission schemes to enhance the reliability of the considered F-RAN system. Next, we derive the SIR received at the desired UE for the above-mentioned transmission schemes.

A. SILENCING

Here, the RRH_0 is the only one serving the UE_0. At the same time, k cooperating RRHs remain silent during the corresponding transmission slots, and the remaining non-cooperating RRHs are interfering. Then from (1) the received signal at UE_0 can be expressed as

$$\mathbf{r}_0 = \underbrace{d_{0,0}^{-\frac{\alpha}{2}} \mathbf{h}_{0,0}^H \mathbf{w}_{0,0} s_0}_{\text{typical link}} + \underbrace{\sum_{i=k+1}^{N} d_{i,0}^{-\frac{\alpha}{2}} \mathbf{h}_{i,0}^H \mathbf{w}_{i,i} s_i}_{\text{interfering links}}. \tag{2}$$

Thus, the SIR received at desired UE_0 using the silencing scheme is given by

$$\begin{aligned}\gamma_{\text{S}} &= \frac{\mathbb{E}_{s_0}\left[\left|\mathbf{h}_{0,0}^H \mathbf{w}_{0,0} s_0 d_{0,0}^{-\frac{\alpha}{2}}\right|\right]^2}{\mathbb{E}_{s_i}\left[\left|\sum_{i=k+1}^{N} \mathbf{h}_{i,0}^H \mathbf{w}_{i,i} s_i d_{i,0}^{-\frac{\alpha}{2}}\right|\right]^2} \\ &\overset{(a)}{=} \frac{\left|\mathbf{h}_{0,0}^H \frac{\mathbf{h}_{0,0}}{||\mathbf{h}_{0,0}||}\right|^2 d_{0,0}^{-\alpha}}{\sum_{i=k+1}^{N} \left|\mathbf{h}_{i,0}^H \frac{\mathbf{h}_{i,i}}{||\mathbf{h}_{\mathbf{i},\mathbf{i}}||}\right|^2 d_{i,0}^{-\alpha}} \\ &= \frac{||\mathbf{h}_{0,0}||^2 d_{0,0}^{-\alpha}}{\sum_{i=k+1}^{N} \left|\mathbf{h}_{i,0}^H \frac{\mathbf{h}_{i,i}}{||\mathbf{h}_{i,i}||}\right|^2 d_{i,0}^{-\alpha}},\end{aligned} \tag{3}$$

where (a) comes after algebraic simplification and assuming $\mathbb{E}[|s_0|^2] = \mathbb{E}[|s_i|^2] = 1$, we obtain (3).

B. MRT

Here the RRH_0 and k cooperating RRHs transmit simultaneously the same signal to the UE_0 while the remaining non-cooperating RRHs are interfering. Similar to [41], we assume that full channel state information (CSI) is available at centralized BBU. Then from (1) the received signal vector at UE_0 can be written as

$$\mathbf{r}_0 = \underbrace{\sum_{c=0}^{k} d_{c,0}^{-\frac{\alpha}{2}} \mathbf{h}_{c,0}^H \mathbf{w}_{c,0} s_0}_{\text{desired links}} + \underbrace{\sum_{i=k+1}^{N} d_{i,0}^{-\frac{\alpha}{2}} \mathbf{h}_{i,0}^H \mathbf{w}_{i,i} s_i}_{\text{interfering links}}. \tag{4}$$

Therefore, following the same procedure to attain (3), SIR perceived at UE_0 using MRT is given by

$$\begin{aligned}\gamma_{\text{MRT}} &= \frac{\mathbb{E}_{s_0}\left[\left|\sum_{c=0}^{k} \mathbf{h}_{c,0}^H \mathbf{w}_{c,0} s_0 d_{c,0}^{-\frac{\alpha}{2}}\right|\right]^2}{\mathbb{E}_{s_i}\left[\left|\sum_{i=k+1}^{N} \mathbf{h}_{i,0}^H \mathbf{w}_{i,i} s_i d_{i,0}^{-\frac{\alpha}{2}}\right|\right]^2} \\ &= \frac{\sum_{c=0}^{k} ||\mathbf{h}_{c,0}||^2 d_{c,0}^{-\alpha}}{\sum_{i=k+1}^{N} \left|\mathbf{h}_{i,0}^H \frac{\mathbf{h}_{i,i}}{||\mathbf{h}_{i,i}||}\right|^2 d_{i,0}^{-\alpha}}.\end{aligned} \tag{5}$$

IV. PERFORMANCE ANALYSIS

This section derives an analytical expression of the outage probability for the considered transmission schemes. Similar to the works in [16], [22], [42], we analyze the network performance in terms of outage probability as the critical performance metric evaluating the CDF of SIR derived earlier in Section III and defined as the probability that UE is in outage if

$$F_{\gamma_\rho}(\theta) = \mathbb{P}\{\gamma_\rho < \theta\} \tag{6}$$

where $\rho \in \{\text{S}, \text{MRT}\}$ and θ is an SIR threshold. We focus on the outage model defined in (6) which meets the definition of reliability in the context of URLLC, which states that a system can assure the URLLC requirements only if it can satisfy the required reliability level within the target latency budget [43]. For this reason, we analyze the system performance in terms of reliability, i.e., (1 − outage probability) for both considered transmission schemes. We also study the impact on system performance with throughput-reliability trade-off when RRHs cooperate to attain the URLLC target reliability level for the desired user. Instead of the end-to-end latency, in this work, we consider data transmission latency which is the time duration of generated URLLC packet delivered successfully to the intended UE.

A. OUTAGE PROBABILITY

We proceed to calculate the outage probability by evaluating the CDF of SIR from (6) as

$$\begin{aligned}F_{\gamma_\rho}(\theta) &= \mathbb{P}\{\gamma_\rho < \theta\} = \mathbb{P}\left\{\frac{X}{Y} < \theta\right\} = \mathbb{P}\left\{Y > \frac{X}{\theta}\right\} \\ &= \int_0^\infty f_X(x) \int_{\frac{x}{\theta}}^\infty f_Y(y) dy dx,\end{aligned} \tag{7}$$

where $X \triangleq \sum_{c=0}^{k} ||\mathbf{h}_{c,0}||^2 d_{c,0}^{-\alpha}$ for all $c \in \{0, \ldots, k\}$, and $Y \triangleq \sum_{i=k+1}^{N} \left|\mathbf{h}_{i,0}^H \frac{\mathbf{h}_{i,i}}{||\mathbf{h}_{i,i}||}\right|^2 d_{i,0}^{-\alpha}$ are the corresponding SIR for desired links and interfering links obtained in (3) and (5). We assume that the channels coefficients of $\mathbf{h}_{i,j}$ are independent, complex Gaussian normally distributed RVs with zero

mean and variance, which includes the effects of path loss i.e., $h_{i,j} \sim \mathcal{CN}(0, d_{i,j}^{-\alpha})$. Therefore, it is easy to show that the numerator X are independent RVs, which follows a gamma distribution with PDF given by

$$f_X(x) = \frac{x^{M_t-1}\exp\left(-x d_{c,0}^{\alpha}\right)}{\Gamma(M_t)(d_{c,0}^{-\alpha})^{M_t}}, \tag{8}$$

where $\Gamma(.)$ denotes gamma function [44]. Similarly, we can show that for the interference power from any i^{th} RRH to the UE_0, i.e., $y_i \triangleq \left|\mathbf{h}_{i,0}^H \frac{\mathbf{h}_{i,i}}{||\mathbf{h}_{i,i}||}\right|^2 d_{i,0}^{-\alpha}$, is an exponential distributed RV with PDF

$$f_{y_i}(y_i) = d_{i,0}^{\alpha}\exp\left(-y_i d_{i,0}^{\alpha}\right). \tag{9}$$

Remark 1: In the scenario where $N > 1$ and $d_{i,0} \neq d_{n,0}$ for $i \neq n$, the distribution of Y can be obtained as [45]

$$f_Y(y) = \sum_{i=k+1}^{N}\left[\frac{e^{\frac{-y}{d_{i,0}^{-\alpha}}}}{d_{i,0}^{-\alpha}}\prod_{\substack{n=k+1\\ n\neq i}}^{N}\frac{d_{i,0}^{-\alpha}}{d_{i,0}^{-\alpha}-d_{n,0}^{-\alpha}}\right]. \tag{10}$$

The validity of (10) is illustrated in Fig. 3. We see that analytical expression matches the corresponding simulation results perfectly. Note that for $N = 1$, $f_Y(y)$ is computed with (9). As shown, the total interfering signal power Y does not change much with increasing N because multi-antenna beamforming to serve the desired UE benefits the desired link while the additional interferers are so far away that their contribution to Y is almost negligible in the considered setup. After obtaining the corresponding PDF for desired and interference links, outage probabilities for silencing and MRT schemes are derived.

Theorem 1: The outage probability when UE_0 is served through silencing scheme is given by

$$F_{\gamma_S}(\theta) = \sum_{i=k+1}^{N}\frac{\Psi_i}{\left(1+\frac{1}{\theta}\left(\frac{d_{i,0}}{d_{0,0}}\right)^{\alpha}\right)^{M_t}}, \tag{11}$$

with $\Psi_i = \prod_{\substack{n=k+1\\ n\neq i}}^{N}\frac{d_{i,0}^{-\alpha}}{d_{i,0}^{-\alpha}-d_{n,0}^{-\alpha}}$.

Proof: Please refer to Appendix A. □

Theorem 2: Under the F-RAN cooperating mode, when the RRHs cooperate to serve the UE_0 through MRT, the outage probability is given as

$$F_{\gamma_{MRT}}(\theta) = \sum_{i=k+1}^{N}\prod_{c=0}^{k}\frac{\Psi_i}{\left(1+\frac{1}{\theta}\left(\frac{d_{i,0}}{d_{c,0}}\right)^{\alpha}\right)^{M_t}}, \tag{12}$$

with $\Psi_i = \prod_{\substack{n=k+1\\ n\neq i}}^{N}\frac{d_{i,0}^{-\alpha}}{d_{i,0}^{-\alpha}-d_{n,0}^{-\alpha}}$.

Proof: Please refer to Appendix B. □

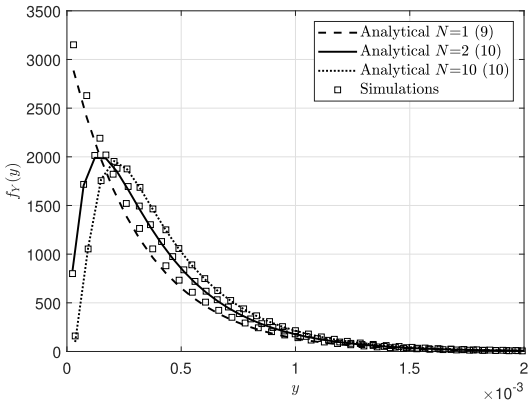

FIGURE 3. Validity of the PDF of Y for $N \in \{1, 2, 10\}$. We also set, $M_t = 4$ $d_{i,0} = 5 + 5i$, $i = 1, \ldots, N$, and $\alpha = 3.5$.

1) DIVERSITY GAIN

The main goal of URLLC systems is to enhance reliability; therefore outage probability curve provides a benchmark for performance evaluation. However, to further investigate the behaviour of outage probability expression at infinite SIR, we assess the diversity gain that can be achieved from Silencing and MRT schemes. Similar to [46], we define diversity gain D as

$$D_\rho = -\lim_{\theta\to\infty}\frac{\log F_{\gamma_\rho}(\theta)}{\log\theta}, \tag{13}$$

where $F_{\gamma_\rho}(\theta)$ is the outage probability obtained in (11) and (12) and $\rho \in \{S,MRT\}$ for silencing and MRT schemes. Next, the diversity gain for silencing and MRT schemes at infinite SIR regime is investigated as follows:

Silencing: Using (13) the diversity gain for silencing scheme is expressed as

$$\begin{aligned} D_S &= -\lim_{\theta\to\infty}\frac{\log F_{\gamma_S}(\theta)}{\log\theta},\\ &\overset{(a)}{=} -\lim_{\theta\to\infty}\frac{\log\left(\sum_{i=k+1}^{N}\Psi_i\left(\frac{d_{i,0}}{d_{0,0}}\right)^{-\alpha M_t}\left(\frac{1}{\theta}\right)^{-M_t}\right)}{\log\theta},\\ &\overset{(b)}{=} -\lim_{\theta\to\infty}\left(\frac{\log\theta^{M_t}}{\log\theta}+\frac{\log\left(\sum_{i=k+1}^{N}\Psi_i\left(\frac{d_{i,0}}{d_{0,0}}\right)^{-\alpha M_t}\right)}{\log\theta}\right),\\ &\overset{(c)}{=} M_t. \end{aligned} \tag{14}$$

where (a) follows from CDF of silencing scheme (11) and using the binomial approximation of $\left(1+\frac{1}{\theta}\left(\frac{d_{i,0}}{d_{0,0}}\right)^{\alpha}\right)^{M_t}$ and taking only the dominant term at high SIR. Meanwhile, (b) comes from exploiting $\log(ab) = \log a + \log b$, and (c) follows immediately after taking the limit.

MRT: Following the similar procedure for attaining (14), the diversity gain for MRT is obtained as

$$D_{MRT} = -\lim_{\theta\to\infty}\frac{\log F_{\gamma_{MRT}}(\theta)}{\log\theta},$$

$$\stackrel{(a)}{=} -\lim_{\theta\to\infty} \frac{\log\left(\sum_{i=k+1}^{N}(\frac{1}{\theta})^{-(k+1)M_t}\prod_{c=0}^{k}\Psi_i\left(\frac{d_{c,0}}{d_{i,0}}\right)^{-\alpha M_t}\right)}{\log\theta},$$

$$\stackrel{(b)}{=} M_t(k+1). \tag{15}$$

Note that comparing (14) and (15), MRT asymptotic diversity gain is $k+1$ times higher than silencing scheme.

2) COMPLEXITY ANALYSIS

The derived closed-form expression in (11) and (12) include simple algebraic scalar operation of the product, addition, and division terms. Note that the complexity of MRT increases with the number of cooperating RRHs k, while the complexity of silencing does not scale up. However, the diversity gain of MRT is also scaled $k+1$ times compared to silencing as shown in (14) and (15). All in all, the derived closed-form solutions do not involve computational severe efforts at BBU of the proposed F-RAN network concerning the number of RRHs and users.

B. THROUGHPUT-RELIABILITY TRADE-OFF

We define throughput-reliability trade-off as the cost the considered reliability-oriented system model has to bear on its average system sum throughput when increasing the number of cooperating RRHs to meet URLLC service requirements. We implement a system-level simulation to study the fundamental trade-offs between reliability and average system sum throughput for the considered transmission schemes in the F-RAN cooperating mode. We assess the reliability of the typical link, which is given by

$$\mathrm{Rel}_{\rho}^{ref} = 1 - F_{\gamma_\rho}(\theta), \tag{16}$$

where γ_ρ is the SIR evaluated from (3) and (5) with $\rho \in \{\mathrm{S},\mathrm{MRT}\}$ for silencing and MRT schemes, respectively. Meanwhile, the average system sum throughput is given by

$$\mathrm{TP}_\rho = R_\rho^{ref} + R_\rho^{A}, \tag{17}$$

where R_ρ^{ref} is the achievable reliable rate at the typical link, and R_ρ^{A} is the corresponding average sum rate of non-cooperating RRHs active users. Then,

$$\begin{aligned} R_\rho^{ref} &= \mathrm{Rel}_{\rho}^{ref}\log_2(1+\theta), \\ R_\rho^{A} &= \mathbb{E}\Big[\sum_{i=k+1}^{N}\log_2(1+\gamma_\rho^{i})\Big]. \end{aligned} \tag{18}$$

Finally, from (17) and (18) we have that

$$\mathrm{TP}_\rho = \mathrm{Rel}_{\rho}^{ref}\log_2(1+\theta) + \mathbb{E}\Big[\sum_{i=k+1}^{N}\log_2(1+\gamma_\rho^{i})\Big]. \tag{19}$$

Note that when the number of cooperating RRHs k increases, the average system sum throughput decreases.

V. SCHEDULING FRAMEWORK FOR URLLC USERS

This section proposes a scheduling framework suitable for URLLC users. 3GPP introduces a flexible frame structure for 5G NR with different options to shorten TTI as compared to LTE [30], we consider mini slots of duration 0.125 ms for transmission in short TTI to meet low latency requirements as in [47]. We use different transmission modes for serving the typical URLLC users. Fig. 4 shows the time frame structure of downlink URLLC with the different transmission modes. As shown, the initial latency budget of 6 TTIs for the URLLC user leaves a sufficient time budget for a maximum of two re-transmission since a hybrid automatic request (HARQ) round trip time consumes 2 TTIs [48]. In the case of re-transmission(s), multiple re-transmitted packets are combined using chase combining (CC), boosting the desired signal power [49].

Note that there are multiple active URLLC users simultaneously competing for the resources. Hence, the overall objective of the given scheduling framework is to serve the maximum number of active URLLC users within the latency budget. Three different performance metrics are considered to evaluate the performance of the scheduling framework:

1) *Percentage of users served*: Accounts for all the successfully delivered packet.
2) *Mean latency*: An average latency (difference between start time and end time) of all successful users.
3) *Average number of transmissions/cooperations*: Average total number of transmissions for all successful users in case of re-transmissions with HARQ. While in the case of cooperating schemes, we average the total number of cooperating RRHs.

To elaborate more on the concept of the proposed scheduling framework and serving strategies for URLLC users. Fig. 5 shows a high-level flow diagram of the scheduling framework at the CU and typical URLLC users, respectively. The scheduling framework works as follows: at each time slot, CU checks if there are any active URLLC users in the network. The data corresponding to the active URLLC users is added to the transmission buffer. If the maximum available RRHs K_{max} is smaller than the number of active URLLC users to be served N_{serv} it reschedules newly active URLLC users (N_{rs}) from the current slot by updating the corresponding latency budget to the next scheduling time. CU has information regarding the availability of RRHs, which are currently not serving any URLLC user, including pending HARQ re-transmissions and global CSI of corresponding UEs. CU assigns the new active URLLC users to the closest available RRH and predicts the SIR determined by corresponding location and fading. Once the SIR is determined, the CU may serve the URLLC user using one out of four transmission modes.

1) *Without cooperation/re-transmission (Baseline)*: Only the typical RRH serves to the UE. CU serves desired UE if SIR is greater than or equal to the given SIR threshold. Otherwise, it drops the corresponding UE.

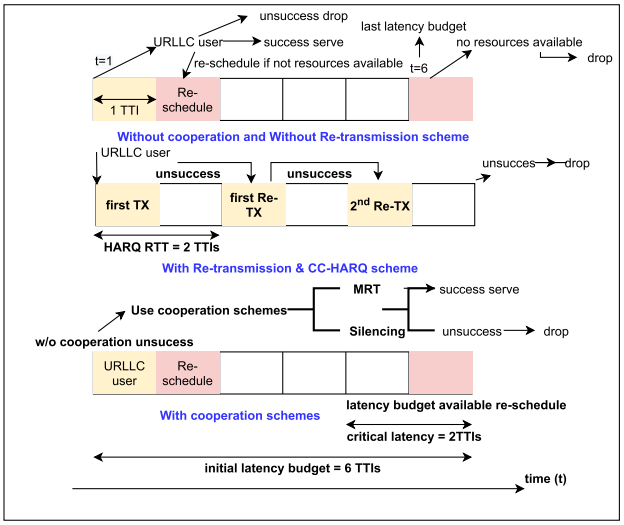

FIGURE 4. Illustration of downlink data transmission for URLLC users with four transmission modes. We assume a two HARQ re-transmissions for URLLC latency budget of one ms, corresponding to the mini-slot duration of 0.125 ms TTI.

2) *Re-transmission with CC-HARQ*: In this non-cooperating mode, CU checks for any pending HARQ re-transmission. If any pending HARQ re-transmission, then it updates the chase combined SIR (γ_{cc}) of intended UE. In this framework, chase combined SIR denoted as γ_{cc} is computed combining the SIR from the previous transmission as $\gamma_{cc} = \sum_{t=1}^{Tx} \gamma_t$, where γ_t is the SIR at current time slot for re-transmission users, and Tx is the maximum allowable re-transmission. CU serves intended UE if $\gamma_{cc} \geq \theta$, otherwise it drops the respective users if only if the available latency budget, i.e., t_{ava} is less than the critical latency threshold (τ). If $t_{ava} \geq \tau$, CU re-schedules the intended UE by updating the latency budget to the next scheduling time.
3) *Silencing*: In this cooperating mode, CU forces some of the available cooperating RRHs to remain silent during the transmission slots and computes the minimum RRH (k_{min}) required to meet the given SIR threshold from (3) as $\gamma_S \geq \theta$. CU serves respective URLLC users if $\gamma_S \geq \theta$ and $k_{min} \leq K_{max} - N_{serv}$, where K_{max} is the maximum number of available RRHs and N_{serv} is the number of RRH assigned to URLLC users during that transmission slots. Otherwise, it drops the respective URLLC users.
4) *MRT*: In this cooperating mode, some of the cooperating RRHs jointly cooperate to transmit to the UE during that transmission slot. CU computes the minimum RRH (k_{min}) required to satisfy the given SIR threshold from (5) as $\gamma_{\text{MRT}} \geq \theta$. CU serve respective URLLC users if $\gamma_{\text{MRT}} \geq \theta$ and $k_{min} \leq K_{max} - N_{serv}$ during that transmission slots, otherwise it drops the respective URLLC users.

In this scheduling framework, there are two practical constraints for the RRHs cooperation:

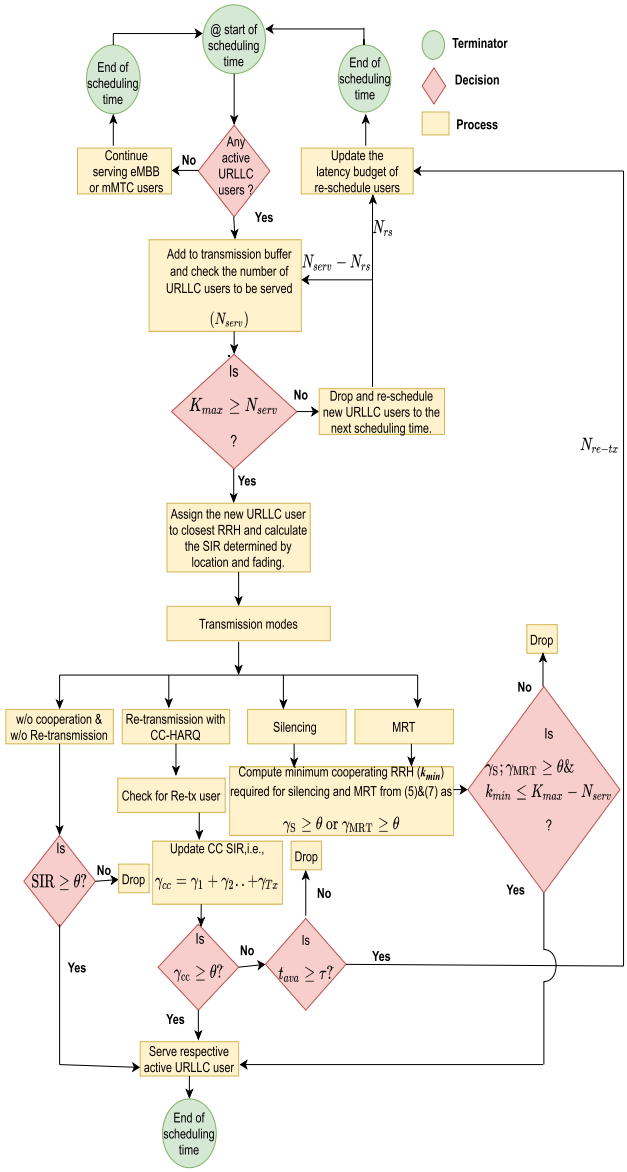

FIGURE 5. Flow diagram of the scheduling framework.

- C1: In a given time instant, RRHs are available for cooperation only if not serving any active URLLC users.
- C2: Cooperating RRHs are selected based on their distance to intended active URLLC users. If available, the closest RRH is given priority, and so on.

The detailed performance evaluation for the above-mentioned metrics is presented and analyzed in Section VI-D.

VI. NUMERICAL RESULTS

This section presents numerical and simulation results related to the system performance under the discussed transmission schemes. In the analysis, we set $\alpha = 3.5$ based on practical radio propagation measurement in the industrial setups [50]. The plots presented in the Section VI-C and VI-D are generated using a system-level simulation, where we adopt the

parameters summarized in Table 3. We generate Rayleigh fading channel realizations and random users locations over run time. In particular, our Monte Carlo simulations comprise 10^7 runs such that the performance results are accurate for targeted reliability of up to $1-10^{-5} = 0.99999$ (five 9's) corroborating our analytical expressions provided in Section IV.

A. IMPACT OF THE NUMBER OF TRANSMIT ANTENNAS

Fig. 6 shows the outage probability of both MRT and silencing schemes as a function of the SIR threshold for the different number of antennas at the RRHs. As shown, the outage probability under both schemes improves with an increasing number of transmit antennas because a higher number of transmit antennas gives more transmit diversity and array gain. It is also demonstrated that with an increase in M_t, both MRT and silencing schemes attain the URC regime at a higher SIR threshold because a large M_t with the beamforming boosts the desired signal power. However, for the same number of M_t and k, the MRT scheme has a higher θ value in the URC regime than the silencing because cooperating RRHs jointly transmit to the UE benefits the signal power in the SIR expression. The analytical results match perfectly with simulation results.

B. IMPACT OF THE NUMBER OF COOPERATING RRHS

Fig. 7 shows the improvement over the outage probability while increasing the number of cooperating RRHs k. As shown, without cooperation ($k = 0$), the system approaches the URC regime at low θ. In contrast, with cooperation $k > 0$, URC is attained at higher values of θ because increasing k reduces the value of interference term in the SIR expression of (3) and (5) for both schemes, which results in boosting the SIR. However, MRT schemes approach a higher SIR threshold in the URC regime with the same number of k than the silencing scheme because MRT also benefits the signal power in the SIR expression. Comparing the results presented in Fig. 6 with that in Fig. 7, it is interesting to observe that even with fewer transmit antennas, RRHs' cooperation provides a significant improvement in the outage probabilities. The analysis shows a solid potential to boost the system's reliability with the cooperation of considered transmission schemes.

C. THROUGHPUT-RELIABILITY TRADE-OFF

In this section, we present the throughput-reliability trade-off results discussed in Section IV-B. As shown in Fig. 8, there is a clear trade-off between average system sum throughput and reliability when we increase cooperating RRHs to achieve the target reliability. For example, as the number of cooperating RRHs increases, there is substantial improvement in reliability for both schemes. However, the average system sum throughput is reduced. Meanwhile, the average system sum throughput comparatively reduces in the case of MRT compared to silencing. For example, with MRT, the average system sum throughput for $k = 1$ and $M_t = 2$ is around 67 bps/Hz, and silencing is around 69 bps/Hz. The increase in silencing throughput is because the cooperating RRHs in silent mode reduce interference factors for reference UE and other active UEs, boosting their corresponding SIRs and increasing average system sum throughput. With the MRT schemes, interference with other active users persists. However, the average system throughput does not increase with k under silencing because ongoing transmission to the other users is interrupted, impacting overall system sum throughput. Furthermore, with an increase in the transmit antennas at the RRHs, average system sum throughput and reliability increase for the same number of cooperating RRHs, which suggests that the number of cooperating RRHs can be reduced by increasing the number of transmit antennas per RRH. This reduction in the number of cooperating RRHs improves throughput, reducing overhead consumption. Furthermore, MRT operates at the URLLC target regime with smaller k for the same M_t compared to silencing, thus allowing better utilization of the network resources.

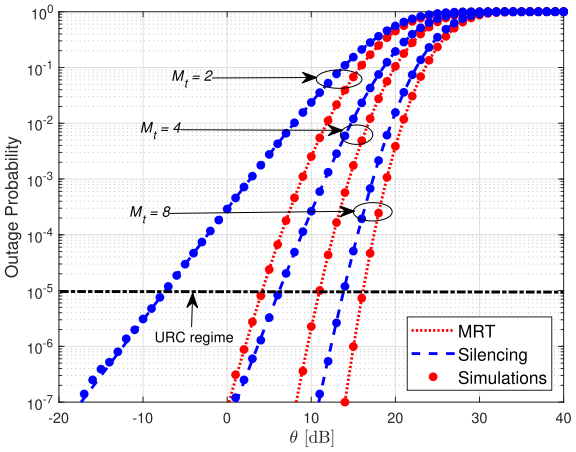

FIGURE 6. Outage probability as a function of θ for $M_t \in \{2, 4, 8\}$ and $k = 2$ cooperating (via MRT or silencing) RRHs. We also set $N = 15$, $d_{0,0} = 15$, $d_{i,0} = 10 + 15i$, $i = 1 \ldots N$.

D. PERFORMANCE ANALYSIS WITH THE SCHEDULING FRAMEWORK

This section shows the performance analysis of the scheduling framework as discussed in Section V. We consider a 0.25 km^2 communication area where RRHs are randomly deployed, and URLLC users get activated with a specific activation rate. We set the number of URLLC users to be equal to 10. We test the performance metrics of the proposed scheduling framework as a function of transmit antennas M_t, activation rate, and SIR threshold θ for the different transmission schemes. Note that we only consider the time duration of successfully delivered packets in mean latency analysis, i.e., dropped packets are not considered.

Fig. 9 shows the performance of the scheduling framework in terms of the percentage of users served metric as a function of the number transmit antennas M_t, activation rate, and SIR

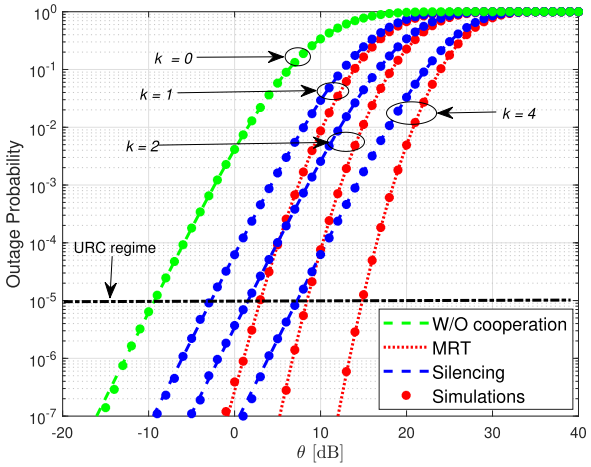

FIGURE 7. Outage probability as a function of θ for $k \in \{0, 1, 2, 4\}$ cooperating (via MRT or silencing) RRHs and $M_t = 3$. We also set $N = 15$, $d_{0,0} = 15$, $d_{i,0} = 10 + 15i$, $i = 1 \ldots N$.

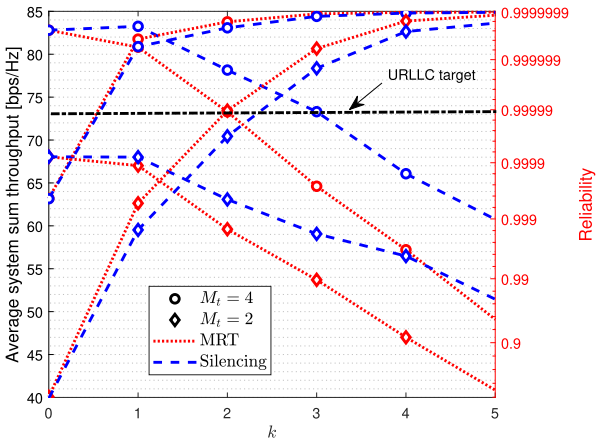

FIGURE 8. The average system sum throughput and reliability analysis as a function of the number of cooperating RRHs k with MRT and silencing schemes. We also set $M_t \in \{2, 4\}$, and θ =15 [dB].

TABLE 3. System configuration for simulations.

Parameter	Value
Minimum inter RRH distance	100m
RRH cell geometry	Regular circular grid
Service area	500m × 500m
Minimum user distance from RRH	5m
Number of fading samples	10^5
Number of user location samples	10^2
Total number of RRHs	10
Path loss exponent	3.5
TTI size	0.125ms
HARQ round trip time	2 TTIs
Initial latency budget	6 TTIs
Critical latency threshold	$\tau = 2$ TTIs
Number of URLLC users	10

threshold θ. Note that the percentage of users served increases with M_t while decreasing with the activation rate and θ. More transmit antennas per RRH increases the desired received power at intended UE, boosting SIR and more URLLC users satisfying the target SIR. However, when the activation rate grows, more URLLC users are activated in the system; thus, URLLC users demand more resources to satisfy the target, SIR. HARQ scheme performs poorly when activation rate is more than 0.1 since more users in re-transmission hold resources during the re-transmission, which results in unavailability of resources to the other new users; hence few numbers of the user is served. We observe that at low, e.g., $\theta < 0$ dB, cooperating and re-transmission with HARQ schemes perform similarly regarding the percentage of users served. Cooperating transmission schemes with MRT and silencing performance seems better than schemes without cooperation and re-transmission. However, we observe that the silencing scheme performs better than MRT in all three cases because in multi-user URLLC scenarios, silencing some RRHs can reduce overall interference in the system, which favors the SIR at all the active URLLC users. However, MRT benefits intended users with MC while interference continues for other active users in the system.

In Fig. 10, it is interesting to observe that the mean latency of baseline and the cooperating scheme is approximately 0.125 ms since there is a single-shot transmission for the intended active URLLC user. However, the mean latency of re-transmission with HARQ decreases when increasing M_t because most users satisfy the target θ at first transmission by exploiting antenna array gains. Meanwhile, the mean latency for re-transmission with the HARQ scheme increases with increasing activation rate and θ. Due to more active URLLC users, the demand for re-transmission is more to satisfy target θ, thus latency increases.

Fig. 11 evaluates the performance of the scheduling framework in terms of a number of transmission/cooperation required for HARQ and cooperation schemes of all successful users as a function of transmit antennas M_t, activation rate, and θ. The results reveal that the number of transmissions with the HARQ scheme slightly decreases with increasing M_t. A large M_t promotes high diversity gain, so the target θ is satisfied in a first transmission. Meanwhile, we see that number of cooperating RRHs is more in silencing than MRT because the diversity gained from silencing is less than MRT. Thus, more cooperating RRHs need to be silenced to fulfill the target, SIR. As shown, when θ increases, the average number of cooperation decreases, but when $\theta > 8$ [dB], the number of cooperation rises because users demand more cooperating RRHs to attain the target, SIR. These results suggest that with the proposed framework, one-shot transmission with cooperation seems more appropriate than re-transmission and without cooperation to support more URLLC users and overcome the hard latency deadline.

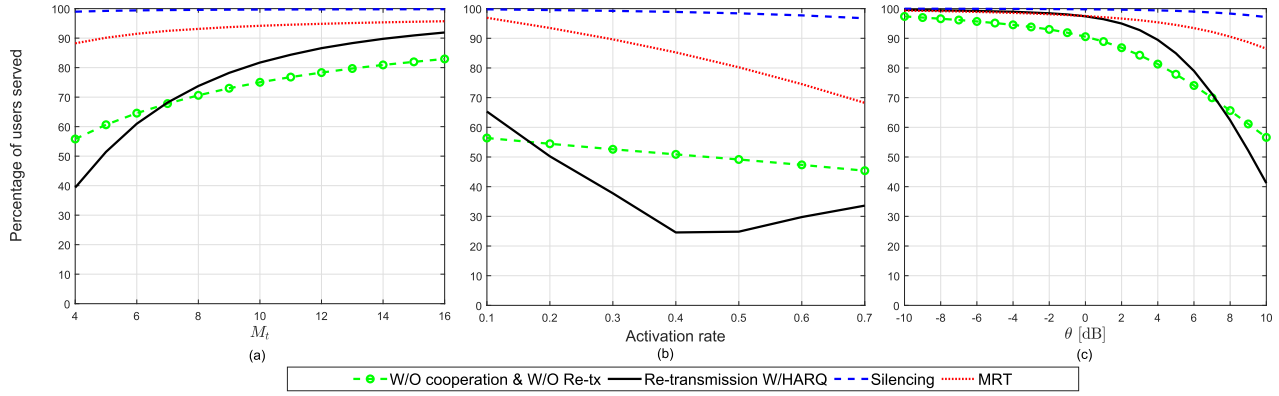

FIGURE 9. Performance evaluation of the scheduling framework in terms of percentage of users served as function of (a) $M_t \in \{4 \dots 16\}$ for $\theta = 10$ dB and activation rate to be equal to 0.3 (left), (b) activation rate$\in \{0.1 \dots 0.7\}$ for $\theta = 10$ dB and $M_t = 4$ (middle), and (c) $\theta \in \{-10 \dots 10\}$ dB for $M_t = 4$, and activation rate to be equal to 0.3 (right) for different transmission schemes.

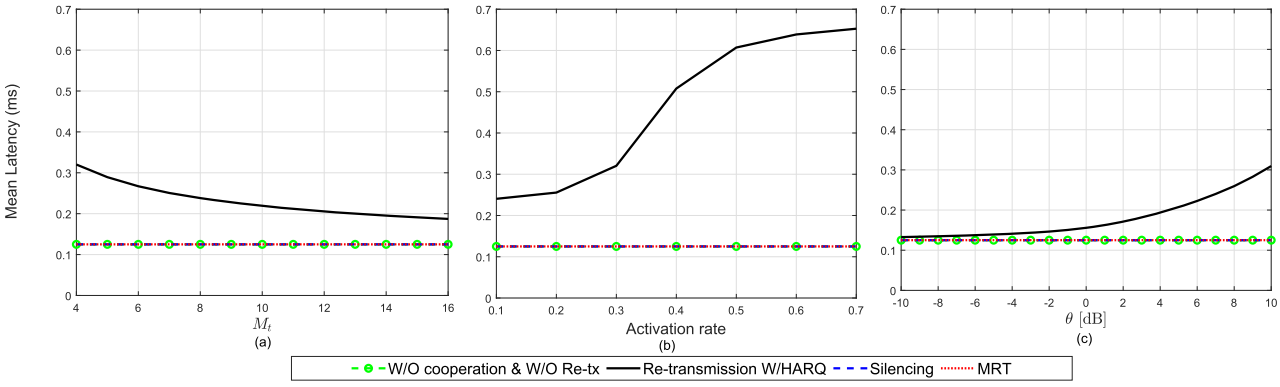

FIGURE 10. Performance evaluation of the scheduling framework in terms of mean latency as function of (a) $M_t \in \{4 \dots 16\}$ for $\theta = 10$ dB and activation rate to be equal to 0.3 (left), (b) activation rate$\in \{0.1 \dots 0.7\}$ for $\theta = 10$ dB and $M_t = 4$ (middle), and (c) $\theta \in \{-10 \dots 10\}$ dB for $M_t = 4$, and activation rate to be equal to 0.3 (right) for different transmission schemes.

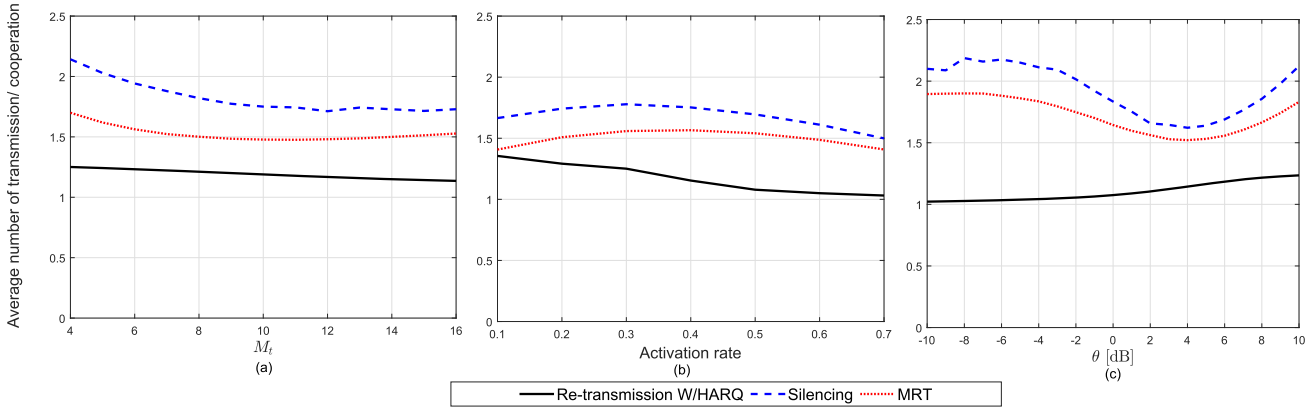

FIGURE 11. Performance evaluation of the scheduling framework in terms of number of transmission/cooperation as function of (a) $M_t \in \{4 \dots 16\}$ for $\theta = 10$ dB and activation rate to be equal to 0.3 (left), (b) activation rate$\in \{0.1 \dots 0.7\}$ for $\theta = 10$ dB and $M_t = 4$ (middle), and (c) $\theta \in \{-10 \dots 10\}$ dB for $M_t = 4$, and activation rate to be equal to 0.3 (right) for different transmission schemes.

VII. CONCLUSION

We studied the performance of the F-RAN-enabled framework for URLLC with MRT and silencing diversity schemes in the interference-limited downlink scenarios. We attained accurate closed-form expressions for the outage probability for MRT and silencing schemes. The analysis presented herein demonstrates that the outage probability performance improves with the number of cooperating RRHs, transmitting antennas at the RRHs, and diversity schemes. We studied the asymptotic behavior of outage probability expressions for MRT and silencing scheme with diversity gain analysis. The result showed that MRT provides $k+1$ times higher diversity gain than silencing. Furthermore, we proposed a mini-slots-based scheduling framework to serve URLLC users under hard latency deadlines. The analysis showed that cooperating schemes like MRT and silencing serve more URLLC

users under the hard latency deadline than with no cooperation and re-transmission with HARQ schemes. Besides, we evaluated the impact on average system sum throughput when increasing the number of cooperating RRHs to ensure URLLC. The results showed that MRT and silencing schemes enhance the system performance in terms of reliability but at the cost of reduced average system sum throughput. Overall, the extensive numerical investigations showed that with the RRHs cooperation, diversity schemes like MRT and silencing could achieve URLLC in an F-RAN network. We intend to extend our analysis for efficient multiplexing of users with heterogeneous quality-of-service requirements as future work.

APPENDIX A: PROOF OF THEOREM 1

In silencing scheme, X is distributed as (8). Then from the definition of moment generating function (MGF), we get the MGF of X as

$$\mathrm{M}_{\mathrm{X}}(t) = \mathbb{E}[\exp(tX)] = \frac{1}{(1 - td_{0,0}^{-\alpha})^{M_t}}. \tag{20}$$

In fact, $\int_0^\infty f_Y(y)dy=1$, thus, we can obtain $\sum_{i=k+1}^{N} \Psi_i = 1$. Using (7) and (10) we have that

$$\begin{aligned} F_{\gamma_{\mathrm{S}}}(\theta) &= \int_0^\infty f_X(x) \int_{\frac{x}{\theta}}^\infty \sum_{i=k+1}^{N} \frac{\Psi_i}{d_{i,0}^{-\alpha}} \exp\left(\frac{-y}{d_{i,0}^{-\alpha}}\right) dy dx \\ &= \int_0^\infty f_X(x) \sum_{i=k+1}^{N} \Psi_i \exp\left(\frac{-x}{\theta d_{i,0}^{-\alpha}}\right) dx \\ &= \sum_{i=k+1}^{N} \Psi_i \mathbb{E}\left[\exp\left(\frac{-1}{\theta d_{i,0}^{-\alpha}} X\right)\right]. \end{aligned} \tag{21}$$

Substituting (20) with $t = \frac{-1}{\theta d_{i,0}^{-\alpha}}$ into (21), we attain (11), thus, concluding the proof.

APPENDIX B: PROOF OF THEOREM 2

In MRT, we have several RRHs jointly transmitting to the desired UE. The MGF of the sum of independent RVs X_c can be expressed through a product of MGFs of each RVs X_c as

$$\begin{aligned} \mathrm{M}_{(\sum_{c=0}^{k} X_c)}(t) &= \prod_{c=0}^{k} \mathrm{M}_{X_c}(t) = \prod_{c=0}^{k} \mathbb{E}[\exp(tX_c)] \\ &= \prod_{c=0}^{k} \frac{1}{(1 - td_{c,0}^{-\alpha})^{M_t}}. \end{aligned} \tag{22}$$

Then, using (7) and (10) we have that

$$\begin{aligned} F_{\gamma_{\mathrm{MRT}}}(\theta) &= \int_0^\infty f_{X_c}(x_c) \int_{\frac{x_c}{\theta}}^\infty \sum_{i=k+1}^{N} \frac{\Psi_i}{d_{i,0}^{-\alpha}} \exp\left(\frac{-y}{d_{i,0}^{-\alpha}}\right) dy dx_c \\ &= \int_0^\infty f_{X_c}(x_c) \sum_{i=k+1}^{N} \Psi_i \exp\left(\frac{-x_c}{\theta d_{i,0}^{-\alpha}}\right) dx_c \\ &= \sum_{i=k+1}^{N} \Psi_i \mathbb{E}\left[\exp\left(\frac{-1}{\theta d_{i,0}^{-\alpha}} X_c\right)\right]. \end{aligned} \tag{23}$$

Substituting (22) with $t = \frac{-1}{\theta d_{i,0}^{-\alpha}}$ into (23), we reached (12), thus, concluding the proof.

REFERENCES

[1] N. H. Mahmood, O. L. A. López, O. S. Park, I. Moerman, K. Mikhaylov, E. Mercier, A. Munari, F. Clazzer, S. Böcker, and H. Bartz, "White paper on critical and massive machine type communication towards 6G," *6G Res. Visions*, no. 11, pp. 1–34, 2020. [Online]. Available: http://jultika.oulu.fi/files/isbn9789526226781.pdf

[2] *IMT Vision–Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond*, document Rec. ITU-R 2083, 2015.

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

[4] B. Holfeld, D. Wieruch, T. Wirth, L. Thiele, S. A. Ashraf, J. Huschke, I. Aktas, and J. Ansari, "Wireless communication for factory automation: An opportunity for LTE and 5G systems," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 36–43, Jun. 2016.

[5] M. Gharba, H. Cao, S. Gangakhedkar, J. Eichinger, A. R. Ali, K. Ganesan, V. Jain, S. Lapoehn, T. Frankiewicz, T. Hesse, Y. Zou, C. Tang, and L. Gu, "5G enabled cooperative collision avoidance: System design and field test," in *Proc. IEEE 18th Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Jun. 2017, pp. 1–6.

[6] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Müller, T. Elste, and M. Windisch, "Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 70–78, Feb. 2017.

[7] M. Bennis, M. Debbah, and H. V. Poor, "Ultra-reliable and low-latency wireless communication: Tail, risk, and scale," *Proc. IEEE*, vol. 106, no. 10, pp. 1834–1853, Oct. 2018.

[8] G. Pocovi, H. Shariatmadari, G. Berardinelli, K. Pedersen, J. Steiner, and Z. Li, "Achieving ultra-reliable low-latency communications: Challenges and envisioned system enhancements," *IEEE Netw.*, vol. 32, no. 2, pp. 8–15, Mar. 2018.

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

[10] G. J. Sutton, J. Zeng, R. P. Liu, W. Ni, D. N. Nguyen, B. A. Jayawickrama, X. Huang, M. Abolhasan, Z. Zhang, E. Dutkiewicz, and T. Lv, "Enabling technologies for ultra-reliable and low latency communications: From PHY and MAC layer perspectives," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2488–2524, 3rd Quart. 2019.

[11] H. Zhang, Y. Qiu, X. Chu, K. Long, and V. C. M. Leung, "Fog radio access networks: Mobility management, interference mitigation, and resource optimization," *IEEE Wireless Commun.*, vol. 24, no. 6, pp. 120–127, Dec. 2017.

[12] N. Alliance, "Further study on critical C-RAN technologies," *Next Gener. Mobile Netw.*, pp. 1–93, Mar. 2015.

[13] M. Peng, S. Yan, K. Zhang, and C. Wang, "Fog-computing-based radio access networks: Issues and challenges," *IEEE Netw.*, vol. 30, no. 4, pp. 46–53, Jul./Aug. 2016.

[14] *3GPP Specification Group Radio Access Network: Study on Channel Model for Frequencies From 0.5 to 100 GHz (Release 16)*, document Rec. 3GPP TR 38.901v, 3GPP, 2019.

[15] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *Proc. 1st Int. Conf. 5G Ubiquitous Connectivity*, 2014, pp. 146–151.

[16] P. Popovski, C. Stefanovic, J. J. Nielsen, E. de Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A.-S. Bana, "Wireless access in ultra-reliable low-latency communication (URLLC)," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5783–5801, Aug. 2019.

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

[18] D. Ohmann, A. Awada, I. Viering, M. Simsek, and G. P. Fettweis, "Diversity trade-offs and joint coding schemes for highly reliable wireless transmissions," in *Proc. IEEE 84th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2016, pp. 1–6.
[19] J. Zeng, T. Lv, R. P. Liu, X. Su, Y. J. Guo, and N. C. Beaulieu, "Enabling ultrareliable and low-latency communications under shadow fading by massive MU-MIMO," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 234–246, Jan. 2020.
[20] S. N. Diggavi, N. Al-Dhahir, A. Stamoulis, and A. R. Calderbank, "Great expectations: The value of spatial diversity in wireless networks," *Proc. IEEE*, vol. 92, no. 2, pp. 219–270, Feb. 2004.
[21] M.-T. Suer, C. Thein, H. Tchouankem, and L. Wolf, "Multi-connectivity as an enabler for reliable low latency communications—An overview," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 156–169, 1st Quart. 2020.
[22] A. Wolf, P. Schulz, M. Dörpinghaus, J. C. S. S. Filho, and G. Fettweis, "How reliable and capable is multi-connectivity?" *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1506–1520, Feb. 2019.
[23] B. Soret, K. I. Pedersen, N. T. K. Jørgensen, and V. Fernández-López, "Interference coordination for dense wireless networks," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 102–109, Jan. 2015.
[24] G. Pocovi, B. Soret, M. Lauridsen, K. I. Pedersen, and P. Mogensen, "Signal quality outage analysis for ultra-reliable communications in cellular networks," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2015, pp. 1–6.
[25] J. J. Nielsen, R. Liu, and P. Popovski, "Ultra-reliable low latency communication using interface diversity," *IEEE Trans. Commun.*, vol. 66, no. 3, pp. 1322–1334, Mar. 2018.
[26] O. L. Alcaraz Lopez, R. Demo Souza, H. Alves, and E. M. G. Fernandez, "Ultra reliable short message relaying with wireless power transfer," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.
[27] M. Centenaro, D. Laselva, J. Steiner, K. Pedersen, and P. Mogensen, "System-level study of data duplication enhancements for 5G downlink URLLC," *IEEE Access*, vol. 8, pp. 565–578, 2020.
[28] F. B. Tesema, A. Awada, I. Viering, M. Simsek, and G. P. Fettweis, "Mobility modeling and performance evaluation of multi-connectivity in 5G intra-frequency networks," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2015, pp. 1–6.
[29] G. Mountaser, M. Condoluci, T. Mahmoodi, M. Dohler, and I. Mings, "Cloud-RAN in support of URLLC," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2017, pp. 1–6.
[30] *5G NR; Physical Channels and Modulation*, document Rec. TS 38.211, V.15.2.0, 3GPP, Jul. 2018.
[31] A. Karimi, K. I. Pedersen, N. H. Mahmood, J. Steiner, and P. Mogensen, "5G centralized multi-cell scheduling for URLLC: Algorithms and system-level performance," *IEEE Access*, vol. 6, pp. 72253–72262, 2018.
[32] N. H. Mahmood, A. Karimi, G. Berardinelli, K. I. Pedersen, and D. Laselva, "On the resource utilization of multi-connectivity transmission for URLLC services in 5G new radio," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshop (WCNCW)*, Apr. 2019, pp. 1–6.
[33] G. Pocovi, K. I. Pedersen, and P. Mogensen, "Joint link adaptation and scheduling for 5G ultra-reliable low-latency communications," *IEEE Access*, vol. 6, pp. 28912–28922, 2018.
[34] A. A. Esswie and K. I. Pedersen, "Opportunistic spatial preemptive scheduling for URLLC and eMBB coexistence in multi-user 5G networks," *IEEE Access*, vol. 6, pp. 38451–38463, 2018.
[35] X. Zhang and J. G. Andrews, "Downlink cellular network analysis with multi-slope path loss models," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1881–1894, May 2015.
[36] V. Hytonen, Z. Li, B. Soret, and V. Nurmela, "Coordinated multi-cell resource allocation for 5G ultra-reliable low latency communications," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2017, pp. 1–5.
[37] B. Soret and K. I. Pedersen, "On-demand power boost and cell muting for high reliability and low latency in 5G," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Jun. 2017, pp. 1–5.
[38] G. Pocovi, M. Lauridsen, B. Soret, K. I. Pedersen, and P. Mogensen, "Ultra-reliable communications in failure-prone realistic networks," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Sep. 2016, pp. 414–418.
[39] A. Anand and G. de Veciana, "Resource allocation and HARQ optimization for URLLC traffic in 5G wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 11, pp. 2411–2421, Nov. 2018.
[40] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks—A technology overview," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 405–426, 1st Quart. 2015.
[41] R. Kassab, O. Simeone, P. Popovski, and T. Islam, "Non-orthogonal multiplexing of ultra-reliable and broadband services in fog-radio architectures," *IEEE Access*, vol. 7, pp. 13035–13049, 2019.
[42] B. Kharel, O. L. Alcaraz Lopez, H. Alves, and M. Latva-aho, "Achieving ultra-reliable communication via CRAN-enabled diversity schemes," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2019, pp. 320–324.
[43] *Service Requirements for the 5G System*, document Rec. TS 22.261, V.16.0.0, 3GPP, Jun. 2017.
[44] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55. Washington, DC, USA: US Government printing office, 1948.
[45] Y. Li, L. Zhang, L. J. Cimini, and H. Zhang, "Statistical analysis of MIMO beamforming with co-channel unequal-power MIMO interferers under path-loss and Rayleigh fading," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3738–3748, Aug. 2011.
[46] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[47] Y. Liu, Y. Deng, M. Elkashlan, A. Nallanathan, and G. K. Karagiannidis, "Analyzing grant-free access for URLLC service," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 741–755, Mar. 2021.
[48] N. H. Mahmood and H. Alves, "Dynamic multi-connectivity activation for ultra-reliable and low-latency communication," in *Proc. 16th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2019, pp. 112–116.
[49] P. Frenger, S. Parkvall, and E. Dahlman, "Performance comparison of HARQ with Chase combining and incremental redundancy for HSDPA," in *Proc. IEEE 54th Veh. Technol. Conf.*, vol. 3, Oct. 2001, pp. 1829–1833.
[50] D. A. Wassie, I. Rodriguez, G. Berardinelli, F. M. L. Tavares, T. B. Sorensen, and P. Mogensen, "Radio propagation analysis of industrial scenarios within the context of ultra-reliable communication," in *Proc. IEEE 87th Veh. Technol. Conf. (VTC Spring)*, Jun. 2018, pp. 1–6.

BINOD KHAREL (Graduate Student Member, IEEE) was born in Kapilvastu, Nepal. He received the B.Sc. degree from the Himalaya College of Engineering, Tribhuvan University, Nepal, in 2012, and the M.Sc. degree from the University of Oulu, in 2018, where he is currently pursuing the Ph.D. degree with the Centre for Wireless Communications (CWC). His research interests include machine type communications, specifically in ultra reliable low latency communication (URLLC) and edge-enabled network architecture for future networks. He was a recipient of the 2019 IEEE European Conference on Networks and Communications (EuCNC) Best Student Paper Award.

ONEL L. ALCARAZ LÓPEZ (Member, IEEE) was born in Sancti-Spritus, Cuba, in 1989. He received the B.Sc. degree (Hons.) in electrical engineering from the Central University of Las Villas (UCLV), Cuba, in 2013, the M.Sc. degree in electrical engineering from the Federal University of Paraná (UFPR), Brazil, in 2017, with a grant from CAPES/CNPQ, and the Ph.D. degree from the University of Oulu, in 2020. From 2013 to 2015, he worked as a Specialist in telematics at Cuban Telecommunications Company (ETECSA). He is currently an Assistant Professor with the University of Oulu. His research interests include wireless communications, specifically in ultra-reliable, low-latency communications, energy harvesting setups, and efficient access techniques for massive machine-type communications. He was a co-recipient of the 2019 IEEE European Conference on Networks and Communications Best Student Paper Award and a collaborator to the 2016 Research Award given by the Cuban Academy of Sciences.

NURUL HUDA MAHMOOD (Member, IEEE) was born in Bangladesh. He is currently a Senior Research Fellow of 6G Flagship with the Center for Wireless Communication, University of Oulu, where he joined, in December 2018. Prior to that, he was an Associate Professor with the Department of Electronics Systems, Aalborg University, Denmark. His current research interests include resource optimization techniques with focus on ultra reliable low latency communication (URLLC), Industrial Internet of Things communication and modeling, and performance analysis of wireless communication systems.

HIRLEY ALVES (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Federal University of Technology-Paraná (UTFPR), Brazil, in 2010 and 2011, respectively, and the dual D.Sc. degree from the University of Oulu, Oulu, Finland, and UTFPR, in 2015. In 2017, he was an Adjunct Professor in machine-type wireless communications with the Centre for Wireless Communications (CWC), University of Oulu. In 2019, he joined CWC, as an Assistant Professor. He is currently the Head of the Machine-Type Wireless Communications Group. He is actively working on massive connectivity and ultra-reliable low latency communications for future wireless networks, 5G and 6G, full-duplex communications, and physical-layer security. He leads the URLLC activities for the 6G Flagship Program. He was a co-recipient of the 2017 IEEE International Symposium on Wireless Communications and Systems (ISWCS) Best Student Paper Award and a co-recipient of the 2016 Research Award from the Cuban Academy of Sciences. He was the General Chair of the ISWCS'2019 and the Co-Chair of the 1st 6G Summit, Levi 2019.

MATTI LATVA-AHO (Senior Member, IEEE) received the M.Sc., Lic. (Tech.), and Dr. (Tech.) (Hons.) degrees in electrical engineering from the University of Oulu, Finland, in 1992, 1996, and 1998, respectively. From 1992 to 1993, he was a Research Engineer at Nokia Mobile Phones, Oulu, Finland, after where he joined CWC. From 1998 to 2006, he was the Director of CWC and the Head of the Department for Communication Engineering, until August 2014. He serves as an Academy of Finland Professor, in 2017 and 2022. He is the Director for 6Genesis-Finnish Wireless Flagship for the period 2018–2026. He has published more than 350 conferences or journal articles in the field of wireless communications. His research interests include mobile broadband communication systems and currently his group focuses on 5G and beyond systems research. He received Nokia Foundation Award for his achievements in mobile communications research, in 2015.