Machine-type Wireless Communications Enablers for Beyond 5G: enabling URLLC via diversity under hard deadlines

Parisa Nouri, Hirley Alves, Mikko A. Uusitalo, Onel Alcaraz López, and Matti Latva-aho

***Abstract*—URLLC is a key design aim in 5G and a mandatory prerequisite in the future uncountable number of industrial applications. In this regard, cooperative relaying and diversity sources in time and frequency domains are introduced as URLLC enablers to support higher reliability and lower latency. The objective of this work is to study two URLLC performance metrics namely, probability of time underflow where the aggregated transmission time is below the time threshold, and reliability which refers to successfully delivering the message within the time window. We examine the impact of cooperative relaying and exploiting time and frequency diversities on the aforementioned performance metrics. We provide the approximated upper bound of the probability of time underflow under time and frequency diversities. In addition, we indicate the maximum achievable reliability as a function of the time threshold for a given probability of time underflow. The performance advantage of cooperative diversity compared to the single transmission to meet URLLC requirements is also highlighted.**

***Index Terms*—URLLC; cooperative diversity; time underflow.**

I. Introduction

Merging Internet-of-Things (IoT) and cyber-physical systems has opened up a new smart industry research area, so called Industry 4.0 referring to the fourth industrial revolution which is about to revolutionize the manufacturing in terms of automation and information exchange besides reducing the energy utilization [1]. Industrial IoT (IIoT) is one of the main drivers for Industry 4.0 and refers to industrial Internet which encompasses both manufacturing and non-manufacturing areas of industry [2]. IIoT requirements such as reliability and latency will be addressed via 5G NR (new radio) which connects the physical and virtual world through novel connectivity interfaces. The key requirements of 5G systems are high data rates, low latency, low energy consumption, high scalability, reliability, connectivity, and security. Authors in [3] study the aforementioned requirements and the corresponding potential solutions in more details.

5G NR is expected to enable Machine-Type Communication (MTC) [1] type-of-services which will be one of the main pillars of IoT and mostly IIoT. MTC type-of services are usually categorized as massive machine-type communications (mMTC) and ultra-reliable low-latency communications (URLLC) [5]. Although, mMTC enable a large-scale connectivity in the upcoming developments of IIoT applications, and provide promising tools in the future connectivity, in this work, we opt to focus on URLLC services which enable 5G to have a significant impact in real-time automation productivity. Furthermore, URLLC known as critical MTC bring prominent opportunities for the support of IIoT. Extreme URLLC with even tighter requirements is a central research topic also towards potentially upcoming generation 6G, which has already started with the Finish initiative 6G Flagship (http://6gflagship.com), which envisions a data-driven society, enabled by near-instant, unlimited connectivity.

The main goal of URLLC discussed in Third Generation Partnership Project (3GPP) is to minimize the latency down to 1ms[2] and guarantee at least 99.999% reliability which are essential requirements in applications such as IIoT, virtual reality and autonomous vehicles [7]. Notice that supporting and handling different types of URLLC applications which have different quality of service (QoS) requirements poses critical challenges in terms of reliability and latency requirements. Academic and industrial interest is growing as evinced by several recent works [6], [8]–[18]. With this regard, in the following section we briefly discuss the key enablers of URLLC connectivity under IIoT landscape.

II. Key URLLC Enablers

Network Slicing: The upcoming wireless networks will probably have a flexible and common framework to support various applications with different QOS requirements. Network slicing allows creating logically isolated sub-networks to intelligently support different services and business segments. The main features of network slicing are: adjustable size, robustness, secure and stable operations, and better resource allocation. However, there are still some challenges to be addressed to take the full advantage of network slicing in the upcoming wireless networks. For instance, the complexity of the networks increases since each generated slice acts as an independent and real network. Additionally, the isolation between the slices is a critical requirement to guarantee secure

P. Nouri, H. Alves, O. Alcaraz López, and M. Latva-aho are with the 6G Flagship, Centre for Wireless Communications (CWC), University of Oulu, Finland. emails:firstname.lastname@oulu.fi. Contact: parisa.nouri@oulu.fi

M. Uusitalo is with Nokia Bell Labs, Espoo, Finland. email: mikko.uusitalo@nokia-bell-labs.com

This research has been financially supported by Academy of Finland, 6Genesis Flagship (Grant no318927), Aka Prof (no 307492), and Nokia Bell Labs (project UEBE).

[1] A comprehensive overview on challenges and open research opportunities of MTC is provided in [4].

[2] In this regard, release-15 of NR 3GPP proposes the concept of mini-time slot with sizes of 2-symbols, 4-symbols, and 7-symbols to guarantee low latency communications [6].

and independent operation in singular slice and so, the slices will not cause any deficiency and security attack to each other. Additionally, security becomes critical in such a network due to the resource sharing and utilizing a common structure for supporting different services where each individual application requires a particular security level. Moreover, resource sharing and management should be taken into account in an infrastructure virtualization in order to avoid resource wasting, and possibility of error and manual modifications. Finally, this technique has to be compatible with traditional technologies and so, radio devices and their corresponding controllers can directly be connected, and extra network slices will be available in the upcoming 5G networks [19].

Intelligence at the edge: Aims at reducing the latency by performing the hard computation tasks at the network edge or close to the edge devices. Edge computing has a distributed computing infrastructure and attains a noticeable storage space by benefiting from either one or a variety of collaborating end user clients or edge device. This technique brings the storage and networking services down to the end devices to handle real-time, secure, and efficient services in the upcoming wireless networks. Nonetheless it also faces challenges in its implementation, for instance, there would be a risk of data leakage due to the data collection by Fog nodes rather than the remote data center which can be addressed by using particular encryption techniques to enhance the privacy of the collected data. Additionally, there is possibility of having fake address due to the malicious user attached to the edge node and so, having an intrusion detection system is necessary. Furthermore, having a suitable management control to handle the heterogeneity in future networks is mandatory [2].

Transmission Time Interval and HARQ Roundtrip Time In order to meet the URLLC requirements we can decrease the latency by reducing the round trip time (RTT) of HARQ mechnasim. Thus, the target reliability can be met by having adequate number of re-transmissions in a shorter time duration. Additionally, we can consider fewer OFDM symbols for each transmission time interval (TTI), albeit, we should note that less resources would be available for data transmission of other URLLC services[3]. Finally, broader subcarrier spacing diminishes the orthogonal frequency-division-multiplexing symbol duration. However, less available resources in frequency domain would be available and so, queuing delay increases [13].

Grant-Free Random Access and Channel Enhancement Random access (RA) makes the uplink connection between any cellular device and base station (BS) possible. RA is categorized into grant-based and grant-free RA protocols. In grant-based RA, BS allocates particular resources to the cellular users to send RA requests. Grant-based RA is not able to handle the latency and signaling overhead when a large number of physical devices connect to Internet and so, is not able to support URLLC connectivity over a large number of users. In contrast, grant-free RA is a potential solution to address the URLLC requirements. In grant-free RA, the users compete for the resources, and there is no reservation phase for allocating the resources to users by the BS. In order to address the reliability requirement for low latency grant-free communications, authors in [18] study an uplink transmission, and propose multiple allocation strategies that enhance reliability for grant-free transmissions for URLLC services. Authors examine different strategies such as repeated transmissions, feedback control, combinations grant-based access depending on the traffic sporadicity and networkâs implementations.

Moreover, meeting the reliability requirements for both data and control channels is needed to enable URLLC since the communication efficiency is affected by the performance of both data and control channels. The conventional approach to design the control channels is not applicable in the future URLLC connectivity due to the high diversity of the requirements in different services. With this regard, authors in [17], propose some enhancement models that allow relaxing the stringent reliability requirements of control information which will be applicable in the development of 5G systems to enable URLLC.

We pointed out key enablers for future URLLC services. Next, we focus our discussion on an illustrative use-case of use of scheduling and diversity for URLLC and on contribution of this work.

A. Packet Scheduling and Diversity Sources

Scheduling improves the reliability by enabling multiple transmission of the same data and so, increasing the possibility of successful reception of the data at the destination for a given time window. Additionally, only one network node communicates in a given time under the scheduling framework which improves the wireless network reliability. However, the importance of timing in IIoT applications has led to challenges in scheduling processes. Motivated by mission-critical and latency constraint applications of future connectivity, the concept of packet scheduling under the hard deadline in latency constraint services has been recently gained a lot of attention. For example, authors in [20] propose a scheduling policy to schedule multiple data flows with strict deadlines. Authors provide an expression for the probability of a packet to miss its random deadline. More importantly, they derive the ratio of the average number of packets missing their corresponding deadline to the average number of packets generated by all the flows per frame. Authors in [21], propose different scheduling techniques under the hard deadline and evaluate their corresponding deadline outage probabilities which indicate both channel decoding errors and violation of the hard deadline. Authors indicate the necessity of having an smart scheduling policy to combine time-division and concatenate-and-code to decrease the deadline outage probability. Moreover, packet scheduling and outage probability under the hard deadline over an erasure broadcast channel is examined in [22]. Authors propose a scheduling policy which minimizes the global deadline outage probability (which refers to the probability that at least one of the receivers has not met the hard communication deadline) if the channel erasures are known prior to transmission under the assumption of serving the user with the earliest deadline first. Authors provide

[3]This issue can be addressed by grant-free mechanism in uplink communications while we need to consider longer TTI in dense networks to meet the latency in a downlink transmission due to the non-negligible queuing delays.

the global outage probability in closed form which indicates the trade-off between the arrival rate, the hard deadlines, and the probability of meeting the corresponding deadlines. Additionally, the outage probability of a multi-user, variable-rate downlink periodic communications system for industrial applications with hard deadlines is examined in [16]. Authors examine two sources of outage in the system, $i)$ transmission error caused by a high transmission rate that the channel can not support, and $ii)$ time overflow where the aggregated transmission time duration exceeds the downlink time budget. Authors provide the probability of transmission error and time overflow in closed form besides the upper and lower bounds on the probability of time overflow. They also propose a rate-adaptation technique to eliminate the time overflow.

In addition to the packet scheduling, based on the previously done research, exploiting the diversity sources in time, frequency, and/or spatial domains supports higher reliability and rates in addition to the lower latency. With this regard, cooperative diversity is a well-known technique in the literature where multiple communication nodes help each other in the transmission process to improve the communication reliability. A large body of works has studied the advantage of cooperate relaying in the future connectivity landscape [10], [23]–[28]. For instance, authors in [23] provide a comprehensive summary of performance advantage of relaying over the direct transmission (DT) in the context of 5G URLLC. They indicate the superiority of relay-enabled networks over DT in both noise-limited and interference-limited scenarios with non-orthogonal multiple access strategy. Practical feasibility of URLLC using best relay selection mechanism is examined in [24]. Authors indicate that the cooperative diversity networks require less transmission resources to meet the target reliability compared to the systems working exclusively on time diversity. Furthermore, the advantage of using relaying technique in 5G URLLC networks with respect to reliability, latency, coding rate, power consumption, and energy efficiency over DT is examined in [10]. Please note that a comprehensive overview of cooperative communication, challenges and research directions are discussed in [29].

Contribution: in Relay-enabled URLLC networks one of the most important concerns is the frame structure and resource allocation in order to provide a suitable trade-off between communication reliability and latency. Hence, motivated by the need for packet scheduling and exploiting diversity in time or frequency domain to support URLLC transmissions, we resort to the 5G NR mini-slot so that a pool of radio resources in time and frequency domains are shared among the devices at CRAN and selected cooperating nodes are assigned based on the the availability of resources. Thus, this work considers scheduling for a relay-enabled[4] network where the cooperation happens in time or frequency domain according to the k-repetition technique [30] where a user receives the information k times. In order to increase the reliability in a URLLC system, we assume that the cooperation can take place in frequency domain if the service is time sensitive while the

[4]Notice that such network is enabled by edge cloud or mobile edge networks (e.g. [14]) where one or many remote radio heads are available to cooperate either acting as relay or for joint transmissions [11].

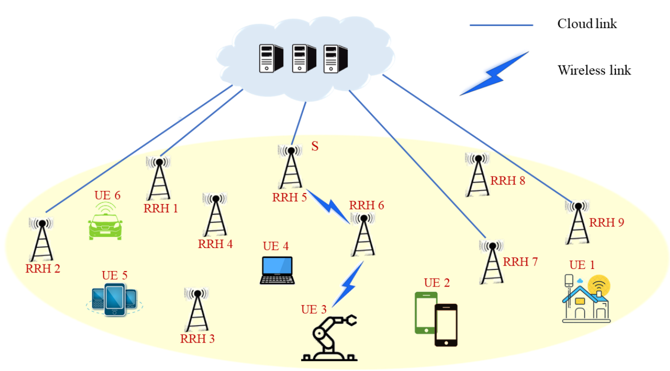

Fig. 1. System model illustration. There are multiple RRHs in the system in which some are connected to the edge cloud while others can either act as relays (time diversity) enable multi connectivity (frequency diversity) thus enhancing the URLLC link. Devices are uniformly distributed with the activation probability $\mathcal{A}$ with a fixed location
. Orthogonal resource allocation is assumed, but RHHs serve other users not illustrated in this figure. Herein we focus on the 3-node cooperative scheme depicted in the center.

cooperation happens in time domain if one frequency channel is free for the transmission, and the service could also have a larger tolerance of delay. However, both transmission scenarios are subjected to a transmission time window to guarantee latency constraint of URLLC communications. In particular,

- We examine the probability of time underflow where the aggregated transmission time duration is less or equal to the time budget.
- We provide the approximated upper bound of the probability of time underflow under studied transmission scenarios.
- We indicate the impact of cooperative diversity in time and frequency domains on the probability of time underflow subject to the hard deadline.
- We examine the probability of outage due to a high transmission rate that channel cannot support. We indicate the maximum achievable reliability as a function of the time threshold where the target probability of time underflow is met.
- We compare the performance advantage of cooperative diversity to single transmission in terms of reliability and latency for URLLC services.

The rest of the paper is organized as follows: We study the system model, and the scheduling procedure under the time and frequency domains in Section III. Additionally, We examine the impact of scheduling in time and frequency domains subject to a hard deadline on the probability of time underflow. Numerical results are provided in Section IV. Finally, Section V concludes the paper.

Notation: Throughout this paper, $\Pr[\mathrm{A}]$ denotes probability of event A, and $f_G(\cdot)$ and $F_G(\cdot)$ are the probability density function (PDF) and cumulative distribution function (CDF) of a random variable (RV) W, respectively. The outage probability is denoted by ϵ and $\mathrm{E}[\cdot]$ is the expectation. $\Gamma(\cdot)$, $\Gamma(\cdot,\cdot)$ and $\mathcal{P}(\cdot,\cdot)$ are the Gamma function [31, §6.1.1], incomplete Gamma function [31, §6.5.2] and Gamma regularized func-

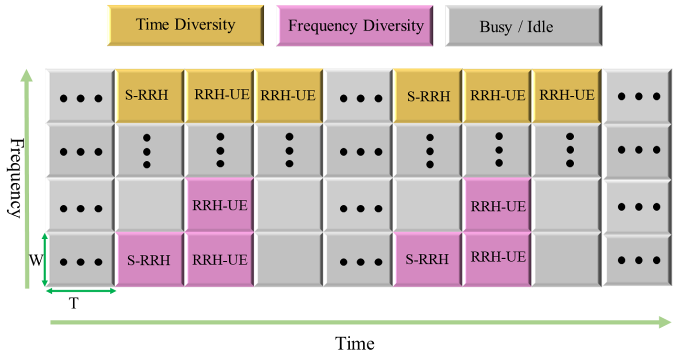

Fig. 2. Illustration of the scheduling in time and frequency domains. The time window is divided into mini-time slots to guarantee the latency requirement. Resources colored in yellow and purple are allocated to URLLC UE while the rest of the resources are free or allocated to other traffic types that are not necessarily URLLC type-of services.

tion [31, §6.5.1], respectively.

III. Cooperative Diversity under Hard deadlines

Fig.1 illustrates a decode-and-forward (DF) based cooperative scenario, where the source (S) is the remote radio head (RRH) 5 connected to the Cloud[5], and RRH 6 collaborates[6] with S to transmit the message to the user equipment (UE) 3. According to the Fig. 2 and following the rationale behind orthogonal coexistence of 5G heterogeneous services, we assume a pool of radio resources, where for simplicity a single frequency channel and a single time slot is allocated to each radio resource. In order to improve the reliability and also meet the latency constraint, we assume two re-transmissions from RRH to UE in a shorter time duration in time or frequency domain. Two transmissions happens according to the 3GPP proactive repetition strategy referred as k-repetition technique [30], where a UE receives the information k times. We have limited the repetitions to two due to the latency constraint since according to the result provided in Section IV, we observe that the high reliability attains with two repetition under tight time constraints, and so more repetitions is not necessary. Clearly, more repetitions increase the reliability but we might not be able to meet the critical time constraint. Furthermore, the resource scheduling is performed prior to the considered time window which enables support of different traffic types in time and frequency domains based on their QoS requirements. We should note that orthogonality of resource allocation and the Cloud's interference managements capabilities prevent the downlink interference among devices in a shared network. Hence, in this work, we aim at studying the impact of exploiting time and frequency diversities on the probability of time underflow and so, the reliability of the time sensitive communications. Please note that the reliability is the probability of successful transmission of a given amount of data within the time constraint. Thus, we consider that the transmissions from RRH to UE happen as follows:

- transmission is spanned over two mini time slot resources to improve the reliability by taking the advantage of time diversity in less time sensitive services. Thus, the system considers time division duplexing to transmit B bits of information over the bandwidth W; and
- the transmissions are localized in time and spread over two frequency channels and so, B bits of information is transmitted in a single mini time slot T and frequency diversity is exploited.

Both transmission schemes are under perfect channel estimation assumption[7] and the links undergo Nakagami-m fading. The squared envelop is then Gamma distributed as $X \sim \text{Gamma}(m, \frac{m}{\Omega})$, where m is the fading parameter and Ω is the scale factor, and $F_X(x; m, \Omega) = \mathcal{P}(m, \frac{mx}{\Omega})$.

A. *Probability of Time Underflow*

In this section we study a metric referred as the probability of time underflow (TU) where the aggregated transmission time does not exceed the time budget T_B and so, the system meets the latency constraint. Accordingly, assume that the airtime needed for each transmission as $T_{i\in\{1,2,\ldots,I\}} = \frac{B}{\mathcal{R}}$, where I is the number of transmission and $\mathcal{R}$ is the coding rate equivalent to the Shannon capacity of AWGN channel with SNR of γ and bandwidth W as $\mathcal{R}= W\log_2(1+\gamma)$. Note that we are assuming packets are large enough to work under the finite blocklengh regime while the packets are also small enough that fading stays constant[8]. Thus, the probability of time underflow is

$$\Pr[\text{TU}] = \Pr\left[\sum_{i=1}^{I} \frac{B}{R_i} \leqslant T_B\right]. \tag{1}$$

Assuming transmitting one bit per Hertz, we can define a new variable as $Y = \frac{W}{\mathcal{R}}$ proportional to the airtime of each transmission and so,

$$\Pr[\text{TU}] = \Pr\left[\sum_{i=1}^{I} \frac{W}{R_i} \leqslant \frac{WT_B}{B}\right] = \Pr\left[\sum_{i=1}^{I} Y_i \leqslant \tau\right], \tag{2}$$

where $\tau = \frac{WT_B}{B}$ is the hard deadline under the time budget.

In the following, first we start with the probability of time underflow in case of a single-hop transmission from the source to RRH and then we expand the approach to the case of having transmissions in frequency and time domains to exploit the frequency and time diversities, respectively.

[5]We consider the slicing of radio access network (RAN) communication resources and so all the resource scheduling and signal processing among the BSs and devices are done in CRAN.

[6]Herein, we assumed that the cooperative RRH is selected at CRAN given load and/or channel availability and condition (e.g. highest SNR) [32], evaluation optimal cooperative selection policies are out of the scope of this work, however the reader can refer to [28], [32]–[35].

[7]As in [9] we assume that URLLC transmissions use fixed rate, though chosen to satisfy latency and reliability requirements. Noticed that pilot symbols could be used for estimation at cost of rate reduction as discussed in [27], [36], [37].

[8]Note that we opt not to use finite blocklength formulation, since we are assuming quasi-static fading. Under this regime the finite error probability performance can be tightly approximated by the outage probability, while differences in performance between these regimes only emerge for extremely small data rates or under large LOS fading components [12].

*a) **Single-hop transmission**[9]:* Probability of time underflow under hard deadline for a single transmission (ST) is

$$\Pr[\mathrm{TU}]=\Pr[Y\leqslant\tau]=F_Y(\tau)=1-\mathcal{P}\left(m,\frac{m(2^{1/\tau}-1)}{\gamma}\right). \quad (3)$$

Lemma 1. *In order to find the upper bound of (3), we resort to Chernoff bound which provides an analytically tractable expression on the tail distribution of random variables. We should note that $Y=\frac{1}{\log_2(1+X)}$, where RV X has Gamma distribution. Hence,*

$$\Pr[Y\leqslant\tau]=\Pr[X\geqslant 2^{\frac{1}{\tau}}-1]\leqslant(1-\tfrac{t}{\beta})^{-\alpha}\,\mathrm{e}^{-t(2^{\frac{1}{\tau}}-1)}, \quad (4)$$

where $0<t<\beta$, α and β are the shape and scale parameters equal to m and $\frac{m}{\Omega}$, respectively. The optimal t for the upper bound is $t_{opt}=\beta+\alpha/(1-2^{\frac{1}{\tau}})$.

Proof. According to the Chernoff bound

$$\Pr[X\geqslant a]=\Pr[\mathrm{e}^{tX}\geqslant\mathrm{e}^{ta}]\leqslant\frac{\mathrm{E}[\mathrm{e}^{tX}]}{\mathrm{e}^{ta}}$$

where $\mathrm{E}[\mathrm{e}^{tX}]$ is the Moment-Generating Function (MGF) of X equal to $(1-\frac{t}{\beta})^{-\alpha}$ for Gamma distribution. The optimal value of t is attained by finding the first derivative of the upper bound in (4) equal to zero. We should note that the bound provided in (4) is convex as the second derivative of that with respect to t is positive and so the optimal t minimizes the upper bound and results in a tight approximation.

b) Diversity sources: Now suppose that RRH cooperates with the source and sends the message to UE. In order to improve the reliability, we consider a re-transmission from RRH to UE in i) frequency domain where the aggregated transmission time duration includes the transmission time from source to RRH noted by Y_1, transmissions time from RRH to UE which takes place in the same mini time slot but at different frequency channels. Thus, the aggregated transmission time of this phase is $Y_2'=\frac{1}{\log_2(1+\sum_{i=1}^{2}X_i)}$, where X_i indicates the channel gain of each individual transmission from RRH to UE of the corresponding frequency channel, and ii) time domain where all the transmissions from source to RRH and from RRH to UE use the same frequency channel but at different mini time slots. Hence, the aggregated transmission duration is the sum of the transmission time of each link as $Y=\sum_{i=1}^{3}Y_i$.

Proposition 1. *The probability of time underflow with respect to the deadline and frequency diversity is*

$$\Pr[Y_1+Y_2'\leqslant\tau]=\int_0^{\tau}f_{Y_1}(y)F_{Y_2'}(\tau-y)dy \simeq\frac{\Gamma\left(2m,\frac{m2^{\frac{1}{\tau}}}{\gamma}\right)}{\Gamma(2m)}\frac{\Gamma\left(m,\frac{m(2^{\frac{1}{\tau}}-1)}{\gamma}\right)}{\Gamma(m)}. \quad (5)$$

Proof. First, by taking the derivative of (3) with respect to τ and substituting $\tau=y$, we find PDF of RV Y_1 as

$$f_{Y_1}(y)=\frac{2^{\frac{1}{y}}\,\mathrm{e}^{\frac{-m\left(2^{1/y}\right)}{\gamma}}\left(2^{\frac{1}{y}}-1\right)^{m-1}\left(\frac{m}{\gamma}\right)^m\ln 2}{y^2\gamma\Gamma(m)}. \quad (6)$$

[9]Notice that we opt to include single-hop transmission analysis for comparison only.

Then assuming $Z=\tau-y$, and $X_1=\mathrm{Gamma}\,(m_1,\beta_1)$ and $X_2=\mathrm{Gamma}\,(m_2,\beta_2)$, sum of two Gamma distributed RVs X_1 and X_2 is a Gamma distributed RV as $\mathrm{Gamma}\,(m_1+m_2,\beta)$. Notice that the same message is transmitted to UE via two frequency channels at the same time. Hence, we can use maximum ratio combining (MRC) at the destination which requires equal channel gains of the transmitted messages. Thus, channel gains equivalent to the inverse of β on both frequency channels are equal which results in $F_{X_1+X_2}(x;m,\Omega)=\mathcal{P}(m_1+m_2,\frac{mx}{\Omega})$. Assuming $m_1=m_2=m$ and $\beta_1=\beta_2=\beta$, we attain

$$F_{Y_2'}(Z)=1-\mathcal{P}\left(2m,\frac{m(2^{\frac{1}{Z}}-1)}{\gamma}\right) \quad (7)$$

By plugging (6) and (7) into (5) and changing the variable $\zeta=2^{\frac{1}{y}}$, after some algebraic manipulations we attain

$$\Pr[Y_1+Y_2'\leqslant\tau]=\frac{(\frac{m}{\gamma})^m}{\Gamma(m)}\int_{2^{\frac{1}{\tau}}}^{\infty}\frac{\zeta\,\mathrm{e}^{\frac{-m(\zeta-1)}{\gamma}}(\zeta-1)^{m-1}\ln 2}{(\frac{\ln 2}{\ln\zeta})^2} \times\frac{\Gamma\left(2m,\frac{m(2^{\frac{1}{(\tau-\frac{\ln 2}{\ln\zeta})}}-1)}{\gamma}\right)}{\Gamma(2m)}\frac{\ln 2}{\zeta(\ln\zeta)^2}d\zeta. \quad (8)$$

Unfortunately (8) does not have a closed-form expression and so, in order to simplify the integral, we approximate

$$2^{\frac{1}{\tau-\frac{\ln 2}{\ln\zeta}}}-1<2^{\frac{1}{\tau-\frac{\ln 2}{\ln\zeta}}}$$

as follows [16]

$$2^{\frac{1}{\tau-\frac{\ln 2}{\ln\zeta}}}=2^{\frac{\log\zeta}{\tau\log\zeta-1}}=\zeta^{\frac{1}{\tau\log\zeta-1}}\overset{(a)}{\geqslant}\zeta^{\frac{1}{\tau\log\zeta}}=2^{\frac{1}{\tau}}, \quad (9)$$

where (a) comes from the fact that $\tau\log\zeta\gg 1$, then after some simplifications and by plugging the approximation term $2^{\frac{1}{\tau}}$ into (8), the integral simplifies into (5).

Proposition 2. *The probability of time underflow with respect to the deadline and time diversity is*

$$\Pr\left[\sum_{i=1}^{3}Y_i\leqslant\tau\right]=\Pr[Y_4+Y_3\leqslant\tau]=\int_0^{\tau}f_{Y_3}(y)F_{Y_4}(\tau-y)dy \simeq\frac{1}{\Gamma(m)^3}\Gamma\left(m,\frac{m2^{\frac{1}{\tau}}}{\gamma}\right)^2\Gamma\left(m,\frac{m(2^{\frac{1}{\tau}}-1)}{\gamma}\right). \quad (10)$$

Proof. PDF of RV Y_3 is equal to (6) since RVs are i.i.d. Then, assuming $Y_1+Y_2=Y_4$ and $Z=\tau-y$, we find the CDF of RV W as (11), where PDF of RV Y_1 and CDF of RV Y_2 attain from (6) and (3), respectively.

$$F_{Y_4}(Z)=\int_0^{Z}f_{Y_1}(s)F_{Y_2}(Z-s)ds \simeq\frac{1}{\Gamma(m)^2}\Gamma\left(m,\frac{m2^{\frac{1}{Z}}}{\gamma}\right)\Gamma\left(m,\frac{m(2^{\frac{1}{Z}}-1)}{\gamma}\right), \quad (11)$$

By plug in (11) into (10), and multiplying by $f_{Y_3}(y)$, after some manipulations and changing the variable $\zeta=2^{\frac{1}{y}}$ and following the same steps as described earlier in Proposition 1, we attain the probability of time underflow as (10).

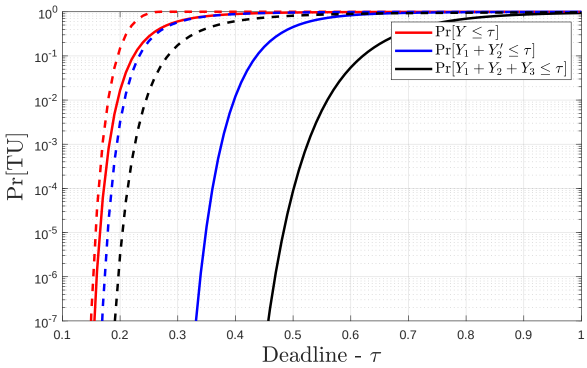

Fig. 3. Probability of time underflow as a function of hard deadline τ,where $Y_i, i \in \{1,2,3\}$ represents the time duration of each transmission phase, Y_2' is time duration of second transmission phase via frequency diversity, and time duration of a single-hop transmission is indicated by Y. Dashed lines indicate the upper bounds of each scenario.

Please notice that in both above mentioned propositions, we approximate the upper bound of the probability of time underflow since we are not able to find the MGF of sum of generated RVs, and Chernoff bound is not applicable in these two scenarios. Additionally, by substituting (9) in the integrals, we attain the upper bound which is indicated with dashed lines in the numerical results. This upper bound is tight on the right tail where the probability of the time underflow goes to one.

B. Reliability with respect to Time Budget

Another source of system outage is transmission error due to the high transmission rate that the channel can not support. In this section, we aim at finding the maximum coding rate as a function of the deadline ($\mathcal{R} = \frac{1}{\tau}$) which satisfies the latency constraint $\Pr[\mathrm{TU}] = T_{\mathrm{th}}$ in the interval of interest ($0.99 < T_{\mathrm{th}} < 0.99999$). We plot the maximum feasible reliability in the system while the target probability of time underflow is met. The reliability is $1-\epsilon$, where ϵ is the outage probability. The outage probability is a function of the coding rate $\mathcal{R}$, fading parameter m and SNR γ, where under Nakagami-m fading condition for S-RRH link is denoted by $\epsilon_{\text{S-RRH}} = \mathcal{P}(m, \frac{m}{\gamma}\left(2^{\mathcal{R}}-1\right))$. Notice that for both frequency and time diversity schemes, the outage probability of RRH-UE link is $\epsilon_{\text{RRH-UE}} = \mathcal{P}(2m, \frac{m}{\gamma}\left(2^{\mathcal{R}}-1\right))$ due to the sum of two Gamma distributed channel gains at the transmitter or destination side. Thus, the system is in outage if the transmission from S to RRH or the transmissions from RRH to UE fail. The outage probability in a DF scenario is

$$\epsilon_{\text{DF}} = \epsilon_{\text{S-RRH}} + (1-\epsilon_{\text{S-RRH}})\epsilon_{\text{RRH-UE}}, \qquad (12)$$

Thus, we first find the optimal threshold τ^* where $\Pr[\mathrm{TU}] = T_{\mathrm{th}}$. Thereafter, we update the coding rate in (12) by the inverse of τ^* to attain the maximum achievable reliability while the time constraint is met. Hence, this technique results in meeting the hard deadline and attaining the maximum feasible reliability corresponding to the target latency indicated in the next section along with the corresponding discussion.

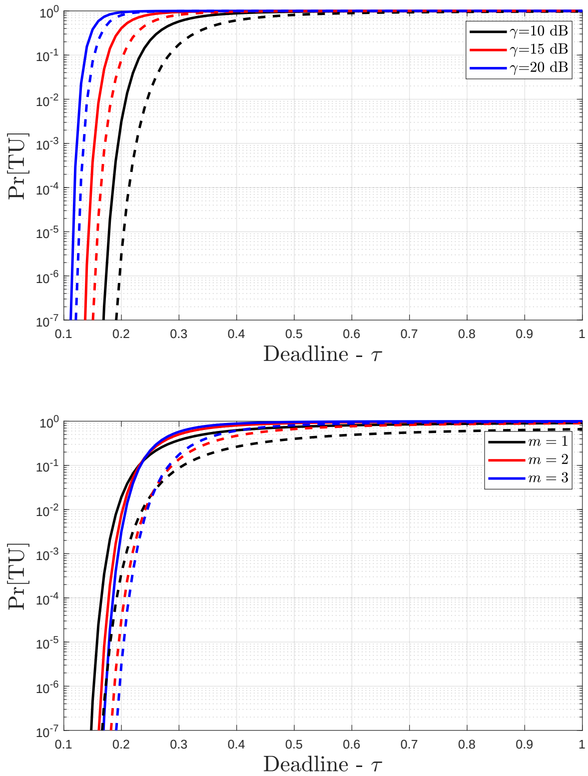

Fig. 4. Probability of time underflow as a function hard deadline with different values of SNR γ and fading parameter as m. Solid lines indicate the frequency diversity and time diversity is illustrated by dashed lines.

IV. Numerical Results

In this section, we provide some numerical results along with the corresponding discussions regarding the scheduling in time and frequency domains under the hard deadline constraint. We particularly, study two performance metric namely, probability of time underflow and reliability. First, we examine the impact of exploiting time and frequency diversities on the probability of time underflow as a function of a hard deadline. Moreover, we illustrate the impact of SNR γ and m on the probability of time underflow. Additionally, we indicate the maximum achievable reliability when the target time constraint is met. Unless stated, otherwise, assume $\gamma = 10$ dB, fading parameter $m = 3$.

Fig.3 represents the probability of time underflow as a function of the deadline in case of having one link, and time and frequency diversity sources. We compare the probability of time underflow to its upper bound approximation. The solid and dashed lines indicate the numerical and the approximated bound on the probability of time underflow, respectively. As we expected by increasing the time budget, the probability of not exceeding the latency constraint increases and by increasing the aggregated duration of transmissions the probability of time underflow decreases. Moreover, although the time diversity technique might not be an appropriate solution for URLLC type-of-services having a stringent latency constraint, the number of users might not be also scalable with frequency diversity. Hence, depending on our system design requirements

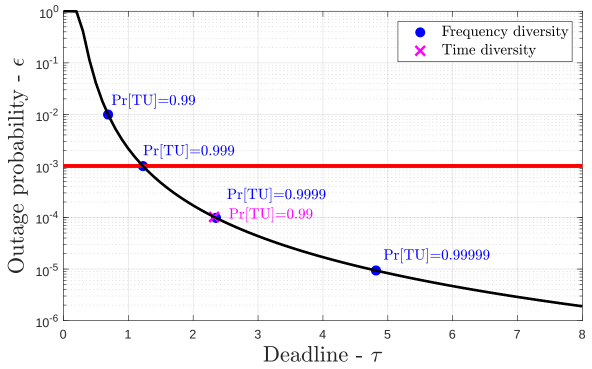

Fig. 5. Outage probability as a function of the deadline τ. The star points and the circle point indicate the maximum achievable reliability where the target latency as $\Pr[\text{TU}] = T_{\text{th}}$ (as indicated in the figure) subject to frequency and time allocation strategies, respectively is met.

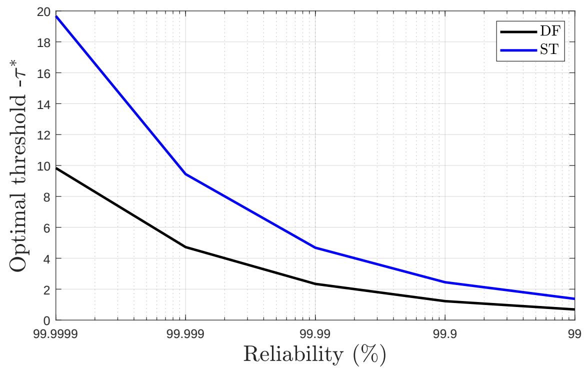

Fig. 6. Comparing the optimal permissive value of the deadline τ^* to meet the target reliability in the interval of interest as $99 \leqslant 1-\epsilon \leqslant 99.9999$ (indicated in percentage in the figure) for ST and DF cooperative scenarios,

and constraints we can work in time or frequency domain. For example, we can take the advantage of time diversity in less time sensitive services while frequency diversity is more suitable for time critical applications if there is enough available frequency channels.

Furthermore, by increasing SNR, the upper bound approximation of the time underflow indicated in solid and dash lines for frequency diversity and time diversity, respectively, enhances as indicated in Fig.4. It can be clearly seen that, frequency diversity technique outperforms time diversity with respect to the probability of time underflow as a function of the SNR for a given deadline. However, very tight deadlines can be also met by time diversity if we increase the SNR γ, although having high SNR might not be optimal/applicable in the practical systems. Moreover, the impact of improving the line-of-sight (LOS) whose effect can be assessed by increasing the fading parameters m is also illustrated in the same figure. By increasing the value of m, the probability of time underflow improves at higher deadlines which can be also concluded from the fact that better LOS improves the reliability which means that the coding rate is small and so, the time threshold is large. However, in order to meet the target latency e.g P[TU]= 0.01 at smaller deadlines, greater LOS results in better reliability and so, lower coding or higher time threshold which can be clearly seen in Fig.4. Furthermore, even by improving the LOS, time diversity can not meet the target probability of time underflow e.g $\Pr[\text{TU}] = 0.1$ with stringent latency constraint compared to the frequency diversity.

Fig.5 indicates the outage probability ϵ as a function of the time budget. The most loose constraint is denoted with a red line where the outage probability is 10^{-3}, thus 99.9% reliability is feasible. The circle points indicated in blue, and the $\times$ point in magenta represents the maximum reliability as function of the optimal value of the time threshold τ^*, where the target latency $\Pr[\text{TU}] = T_{\text{th}}$ is met. As it can be clearly seen, the reliability increases by having a looser time constraint due to the fact that the coding rate is equivalent to the inverse of the time threshold as $\mathcal{R} = \frac{1}{\tau}$. Thus by increasing the deadline, the coding rate decreases and so, the reliability increases as higher coding rates result in higher outage probabilities and so, lower communication reliability. Moreover, it can be noted that under the time diversity, we have only one value $\tau^* = 2.3$ where the system meets the target latency $T_{\text{th}} = 0.99$, and 99.99% reliability is possible. This conveys this message that ultra-reliable communication is possible under time diversity albeit it requires a large time budget to satisfy the time constraint which is not compatible with time critical systems having stringent latency requirements. The optimal values of τ^* highly increases to meet higher T_{th} which is not indicated in this plot. On the other hand, in order to meet the target latency $T_{\text{th}} = 0.99$ with frequency diversity, we attain $\tau^* = 0.6$ which supports lower reliability as 99% since the coding rate in this case is higher compared to the time diversity, thus the higher coding rate leads to lower reliability. Although, in the scenario the reliability increases by time diversity compared to the frequency diversity technique, it requires a large tolerance of latency to meet the time constraint which is not suitable in latency critical systems while supports systems with stringent reliability requirements.

Furthermore, Fig.6 compares the optimal feasible latency where the target reliability in the interval of interest is met for a ST and DF transmission. As it can be clearly seen, the communication reliability increases by having a larger tolerance of latency since higher value of latency constraint result in lower coding rate. Thus, the lower coding increases the reliability or in other words decreases the outage probability. Moreover, ST requires larger tolerance of latency to meet the target reliability in comparison to DF cooperative scheme which highlights the necessity of having cooperative diversity in the future connectivity to meet URLLC requirements. Moreover, Fig.7 indicates the optimal allowable latency as a function of SNR γ (dB) to meet the target reliability 99.99% for ST and DF transmission scenarios. To meet the target reliability, the latency is smaller at high SNR regime since we can communicate with larger communication rate and so, τ^* reduces. As we expected, DF transmission has lower latency compared to the ST to meet the target reliability.

V. Conclusion and Final Remarks

This work overviews some URLLC enablers for IIoT applications for the future connectivity. We particularly discuss the

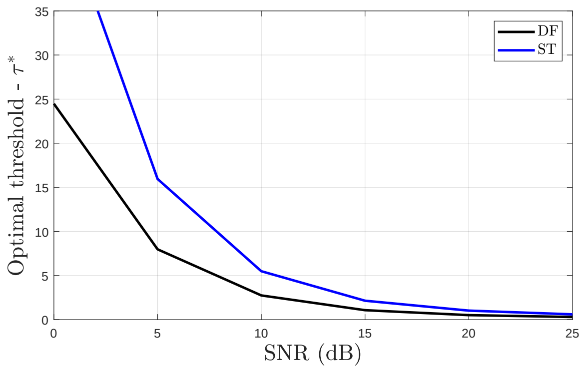

Fig. 7. Comparing the optimal permissive value of the deadline τ^* for target reliability 99.99% as function of SNR γ (dB) for ST and DF cooperative scenarios.

packet scheduling and diversity sources in URLLC networks. We examine two performance metrics as $i)$ probability of time underflow which indicates that the aggregated transmission time is below a hard deadline, and $ii)$ reliability which indicates the probability of successful transmission of the message within the deadline. Furthermore, we indicate the impact of scheduling in time and frequency domains on the probability of time underflow, and attain the approximated upper bound of that. Additionally, we illustrate the maximum achievable reliability as a function of the deadline or in other words the coding rate where the target latency is met. We observe that although time diversity supports higher reliability compared to frequency diversity for a given target latency T_{th}, it requires a large tolerance of latency to meet the time constraint which is not compatible with latency critical type of services. Additionally, we highlight the key role of cooperative diversity technique as a potential solution to meet reliability and latency requirements of future connectivity. In a nutshell, through a suitable packet scheduling in a pool of time and frequency resources, reliability and latency requirements of different type of URLLC services can be met. Finally, in our future work, we will consider the mobility of the users in order to make the scheduling compatible with the dynamic networks.

References

[1] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial internet of things: Challenges, opportunities, and directions," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4724–4734, 2018.

[2] M. Aazam, S. Zeadally, and K. A. Harras, "Deploying fog computing in industrial internet of things and industry 4.0," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4674–4682, 2018.

[3] I. F. Akyildiz, S. Nie, S.-C. Lin, and M. Chandrasekaran, "5G roadmap: 10 key enabling technologies," *Computer Networks*, vol. 106, pp. 17–48, 2016.

[4] O. A. Amodu and M. Othman, "Machine-to-machine communication: An overview of opportunities," *Computer Networks*, vol. 145, pp. 255–276, 2018.

[5] S. Zhang, Y. Wang, and W. Zhou, "Towards secure 5G networks: A survey," *Computer Networks*, vol. 162, p. 106871, 2019.

[6] Y. Alfadhli, Y.-W. Chen, S. Liu, S. Shen, S. Yao, D. Guidotti, S. Mitani, and G.-K. Chang, "Latency performance analysis of low layers function split for urllc applications in 5G networks," *Computer Networks*, vol. 162, p. 106865, 2019.

[7] M. Alsenwi, N. H. Tran, M. Bennis, A. K. Bairagi, and C. S. Hong, "embb-urllc resource slicing: A risk-sensitive approach," *IEEE Communications Letters*, 2019.

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

[9] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55 765–55 779, 2018.

[10] P. Nouri, H. Alves, and M. Latva-aho, "Performance analysis of ultra-reliable short message decode and forward relaying protocols," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, p. 202, 2018.

[11] B. Kharel, O. L. A. López, H. Alves, and M. Latva-aho, "Achieving ultra-reliable communication via cran-enabled diversity schemes," in *2019 European Conference on Networks and Communications (EuCNC)*. IEEE, 2019, pp. 320–324.

[12] O. L. Alcaraz Lopez, H. Alves, and M. Latva-aho, "Joint Power Control and Rate Allocation Enabling Ultra-Reliability and Energy Efficiency in SIMO Wireless Networks," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5768–5782, 2019.

[13] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, 2018.

[14] R. Kassab, O. Simeone, P. Popovski, and T. Islam, "Non-orthogonal multiplexing of ultra-reliable and broadband services in fog-radio architectures," *IEEE Access*, vol. 7, pp. 13 035–13 049, 2019.

[15] R. P. Antonioli, E. B. Rodrigues, D. A. Sousa, I. M. Guerreiro, C. F. e Silva, and F. R. Cavalcanti, "Adaptive bearer split control for 5g multi-rat scenarios with dual connectivity," *Computer Networks*, vol. 161, pp. 183 – 196, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128619308527

[16] R. Jurdi, S. R. Khosravirad, H. Viswanathan, J. G. Andrews, and R. W. Heath, "Outage of periodic downlink wireless networks with hard deadlines," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1238–1253, 2019.

[17] H. Shariatmadari, Z. Li, S. Iraji, M. A. Uusitalo, and R. Jäntti, "Control channel enhancements for ultra-reliable low-latency communications," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2017, pp. 504–509.

[18] B. Singh, O. Tirkkonen, Z. Li, and M. A. Uusitalo, "Grant-free transmissions for ultra-reliable and low latency uplink communications."

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

[20] M. Kashef and N. Moayeri, "Random-deadline missing probability analysis for wireless communications in industrial environments," in *14th International Workshop on Factory Communication Systems (WFCS)*. IEEE, 2018, pp. 1–8.

[21] D. Tuninetti, B. Smida, N. Devroye, and H. Seferoglu, "Scheduling on the gaussian broadcast channel with hard deadlines," in *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018, pp. 1–7.

[22] Z. Ovaisi, N. Devroye, H. Seferoglu, B. Smida, and D. Tuninetti, "On erasure broadcast channels with hard deadlines," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2018, pp. 1–7.

[23] Y. Hu, M. C. Gursoy, and A. Schmeink, "Relaying-enabled ultra-reliable low-latency communications in 5g," *IEEE Network*, vol. 32, no. 2, pp. 62–68, 2018.

[24] M. Serror, S. Vaaßen, K. Wehrle, and J. Gross, "Practical evaluation of cooperative communication for ultra-reliability and low-latency," in *19th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*. IEEE, 2018, pp. 14–15.

[25] P. Nouri, H. Alves, R. Demo Souza, and M. Latva-aho, "Ultra-reliable short message cooperative relaying protocols under nakagami-m fading," in *2017 International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2017, pp. 287–292.

[26] P. Nouri, H. Alves, and M. Latva-aho, "On the performance of ultra-reliable decode and forward relaying under the finite blocklength," in *2017 European Conference on Networks and Communications (EuCNC)*, June 2017, pp. 1–5.

[27] P. Nouri, H. Alves, R. D. Souza, and M. Latva-aho, "In-band pilot overhead in ultra-reliable low latency decode and forward relaying," in

2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring). IEEE, 2019, pp. 1–5.

[28] Y. Hu, C. Schnelling, M. C. Gursoy, and A. Schmeink, "Multi-relay-assisted low-latency high-reliability communications with best single relay selection," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7630–7642, Aug 2019.

[29] J. W. Raymond, T. O. Olwal, and A. M. Kurien, "Cooperative communications in machine to machine (m2m): solutions, challenges and future work," *IEEE Access*, vol. 6, pp. 9750–9766, 2018.

[30] Z. Li, M. A. Uusitalo, H. Shariatmadari, and B. Singh, "5g urllc: Design challenges and system concepts," in *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2018, pp. 1–6.

[31] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. US Government printing office, 1948, vol. 55.

[32] Y. Hu, C. Schnelling, Y. Amraue, and A. Schmeink, "Finite blocklength performance of a multi-relay network with best single relay selection," in *2017 International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2017, pp. 122–127.

[33] A. Saha, A. Ghosh, and W. Hamouda, "Learning-based relay selection for cooperative networks," in *2014 IEEE Global Communications Conference*, Dec 2014, pp. 386–391.

[34] S. Sangeetha, R. Dhaya, and R. Kanthavel, "Improving performance of cooperative communication in heterogeneous manet environment," *Cluster Computing*, pp. 1–7, 2018.

[35] K. Geng, Q. Gao, L. Fei, and H. Xiong, "Relay selection in cooperative communication systems over continuous time-varying fading channel," *Chinese Journal of Aeronautics*, vol. 30, no. 1, pp. 391–398, 2017.

[36] M. Mousaei and B. Smida, "Optimizing pilot overhead for ultra-reliable short-packet transmission," in *2017 IEEE International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–5.

[37] S. Schiessl, H. Al-Zubaidy, M. Skoglund, and J. Gross, "Delay performance of wireless communications with imperfect csi and finite-length coding," *IEEE Transactions on Communications*, vol. 66, no. 12, pp. 6527–6541, Dec 2018.