Reliability analysis of Uplink Grant-Free transmission over Shared Resources

Berardinelli, Gilberto; Mahmood, Nurul Huda; Abreu, Renato Barbosa; Jacobsen, Thomas; Pedersen, Klaus I.; Kovacs, Istvan Z.; Mogensen, Preben Elgaard

Published in:
IEEE Access

DOI (link to publication from Publisher):
10.1109/ACCESS.2018.2827567

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 15, 2018, accepted March 26, 2018, date of publication April 17, 2018, date of current version May 16, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2827567

Reliability Analysis of Uplink Grant-Free Transmission Over Shared Resources

GILBERTO BERARDINELLI[1], NURUL HUDA MAHMOOD[1], RENATO ABREU[1], THOMAS JACOBSEN[1], KLAUS PEDERSEN[1,2], ISTVÁN Z. KOVÁCS[2], AND PREBEN MOGENSEN[1,2]

[1]Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark
[2]Nokia Bell Labs, 9220 Aalborg, Denmark

Corresponding author: Gilberto Berardinelli (gb@es.aau.dk)

ABSTRACT Uplink grant-free schemes have the promise of reducing the latency of a user-equipment-initiated transmission by avoiding the handshaking procedure for acquiring a dedicated scheduling grant. However, the possibility of successfully delivering a payload within a latency constraint may be severely compromised in case of grant-free operations over shared radio resources. In this paper, we study the performance of two different uplink grant-free schemes over shared resources recently discussed within the fifth generation new radio standardization, namely, a solution based on a stop-and-wait (SAW) protocol and a blind retransmission approach. Performance is evaluated assuming Rayleigh fading channels with a maximum ratio combining (MRC) multi-antenna receiver. Analytical results show the benefits of grant-free transmission with respect to the traditional grant-based approach for a tight latency constraint. A high-order receive diversity is beneficial to leverage the MRC gain and enables the possibility of achieving the 10^{-5} outage probability target set for ultra-reliable low-latency communication services. The blind retransmission approach is significantly penalized by identification and signaling errors, while a SAW solution with potentially scheduled retransmissions out of the shared bandwidth leads to the lowest outage probability, at least for frequent packet arrivals.

INDEX TERMS Multiple access, grant-free transmission, URLLC.

I. INTRODUCTION

Uplink transmissions in cellular networks are typically grant-based, i.e. an User Equipment (UE) can transmit its payload provided a scheduling grant (SG) is received from a serving base station (BS) upon request. Grant-free operations, where UEs initiate a transmission without requesting such SG, are instead considered a promising solution for the latency-critical services targeted by fifth Generation (5G) radio access technology [1]. Semi-persistent scheduling (SPS) with pre-allocation of radio resources, is the most known grant-free operation and is considered a valid approach for periodic types of traffic [2]. In case of sporadic packet arrivals, the usage of shared radio resources for a group of uplink grant-free UEs is instead advocated as a necessary solution for avoiding a prohibitive resource wastage [3].

Collision-prone transmission over shared resources has recently witnessed a regrown attention in the research community given the emergency of a plethora of novel Internet-of-Things (IoT) use cases, with a major focus on massive access [4]. The basic ALOHA and slotted ALOHA protocols [5] have known limitations in terms of maximum asymptotic cell throughput in case collisions are considered as disruptive events. However, simultaneous transmissions can in practice still be resolved, especially in case interference suppression receivers are in place [6]. An uncoordinated strategy for maximizing the throughput of a time slotted random channel is proposed in [7]. Enhanced non-orthogonal schemes such as Sparse Code Multiple Access (SCMA) [8] or Interleaved Division Multiple access (IDMA) [9] achieve robustness to collisions by using user-specific signatures, and aim at boosting the cell capacity with respect to traditional orthogonal resource allocation. Fundamental limits of collision-prone random access communication in terms of achievable rates in noise-limited and interference-limited regimes are derived in [10]. Most of the existing work focus on maximizing the number of supported users for a given set of resources, as well as in limiting the device energy consumption. However, low latency constraints are typically disregarded.

Different options for latency-critical uplink grant-free transmission are instead discussed within the 3rd Generation Partnership Project (3GPP) standardization body for the upcoming 5G New Radio (NR) [1]. The performance in terms of delay and throughput of such schemes has been recently empirically evaluated in a large network setup with extensive system level simulations [11], considering the challenging target of a 10^{-5} outage probability with 1 ms latency as targeted by Ultra-Reliable Low Latency Communication (URLLC) services. A recent contribution [12] studies instead the collision probability in shared resources considering the possibility of resolving them with repetition mechanisms and multi-user detection; the analysis is based on empirical Signal-to-Interference plus Noise Ratio (SINR) thresholds for detection derived from previous simulation studies. In general, extensive system level simulations addressing reliability performance require a larger number of snapshots compared to traditional broadband traffic studies in order to capture the low percentiles of failure probabilities targeted by URLLC services.

In this paper, we present a simple analytical model for studying the reliability of latency-constrained uplink grant-free transmission over shared resources, and evaluate its performance over Rayleigh fading channels with Maximum Ratio Combining (MRC) receivers at the BS [13]. Such receivers, though unable to suppress the interference, strengthen the power of the intended user and therefore add a tier of protection with respect to eventual collisions. We consider a traditional Stop-and-Wait (SAW) protocol and a solution based on blind retransmissions, both discussed within the 3GPP for 5G NR [1]. In particular, we derive the final outage probability upon retransmissions as a function of the occupied resources, packet arrival rate and number of receive antennas at the BS. Identification and signaling errors are also included in the analysis, since they are expected to have a significant impact when assessing reliability performance [14]. The decoding failure probability is calculated by using recent results from channel coding literature in the finite blocklength regime [15]. To the best of our knowledge, an analytical study of the reliability of the proposed uplink grant-free solutions has not been presented yet in the literature; though such type of studies has obvious limitations with respect to extensive Monte Carlo simulations, it allows to obtain fundamental insights on the potential of the techniques before running computationally heavy simulations.

The rest of the paper is structured as follows. Section II describes the scenario as well as the targeted grant-free schemes. The packet failure probabilities are derived in Section III. Performance results are presented in Section IV, while Section V resumes the conclusions and states the future work.

II. SYSTEM MODEL

We consider N users sharing a frequency band of W Hz for their uplink grant-free transmissions. The bandwidth is divided in K equally-sized frequency chunks. When a packet arrival occurs, a UE selects randomly one of the K frequency chunks for its transmission. No link adaptation is considered; every packet of B bits is mapped over a frequency chunk and a number of N_S OFDM symbols corresponding to a single Transmission Time Interval (TTI) of duration T_{TTI} seconds. We denote as λ the packet arrival rate per TTI. The UEs are assumed to be perfectly synchronized. We further assume $\lambda T_0 \ll 1$, where T_0 denotes the total time needed for transmitting a packet (including eventual retransmissions); the packet inter-arrival time is larger than the needed time for delivering a packet, such that queueing effects at each UE can be neglected. A grant-free transmission is composed of a UE identifier, mapped over, e.g. a *preamble*, followed by the payload. The usage of reference sequences such as Zadoff-Chu as preambles is currently studied within 3GPP [16] given their attractive auto/cross-correlation properties [17], which allow discriminating a number of simultaneously transmitting UEs. The preamble can also be used for estimating the channel responses of the transmitting UEs for coherent detection.

We consider the UEs operating over a flat Rayleigh fading channel, i.e. the channel response is constant over the selected frequency chunk, however, it can vary at every transmission or retransmission. The users are power controlled such that their transmissions are received at the same average power, though their instantaneous receive power may change at each transmission due to Rayleigh fluctuations.

The BS is equipped with M receive antennas and MRC receivers. It is well known that MRC receivers allow strengthening the SNR by coherently adding the multiple receive copies of the signal of interest, thus boosting its strength.

All the studied transmission procedures are depicted in Figure 1, and described next.

A. SAW PROCEDURE

Upon a packet arrival, the UE attempts to transmit its packet in one of the available grant-free resources. After a transmission, the UE will be waiting for a feedback message from the BS for a time equal to $T = 2\tau + T_P^{BS} + T_{FB}$, where τ denotes the propagation delay, T_P^{BS} is the processing time at the BS, and T_{FB} is the duration of a feedback message transmission.

The grant-free transmission is successful in case the UE is correctly identified, and its payload correctly decoded. A positive acknowledgment (ACK) will be issued from the BS and the UE will not start a retransmission in case the feedback message is received within a delay T relative to its transmission. In case of unsuccessful decoding, we distinguish the following two options:

- *SAW A: Retransmissions over shared resources*. The BS does not issue any feedback in case of unsuccessful transmission, i.e. in case the UE is not identified, or it is identified but its payload not successfully decoded. The UE will then retransmit its packet over shared resources after a time interval $\Delta T = \left\lceil \frac{T+T_P^{UE}}{T_{TTI}} \right\rceil T_{TTI}$ with respect to the first transmission, where $\lceil x \rceil$ denotes the integer equal or larger than x, and T_P^{UE} is the processing time

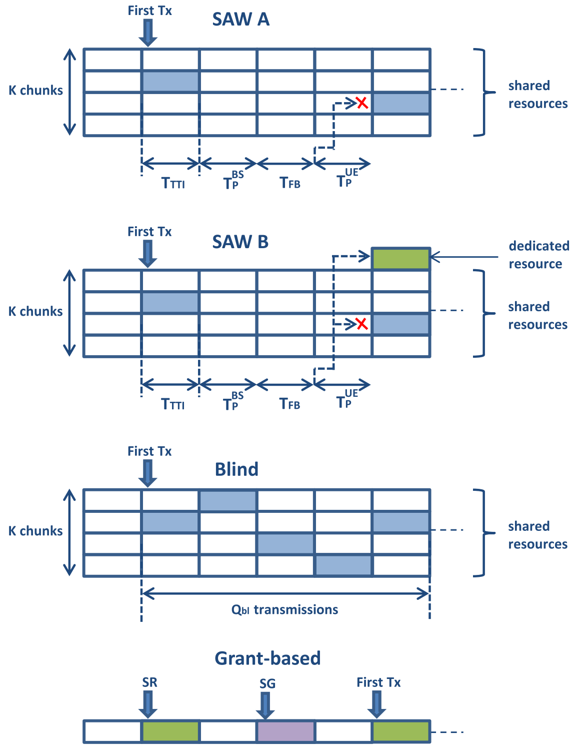

FIGURE 1. **Time diagram of the considered grant-free and grant-based schemes. Note that the propagation delay τ is not depicted for the sake of visual simplicity.**

at the UE necessary for decoding an eventual feedback message.

- *SAW B: Retransmissions over dedicated resources*. The BS is able to issue a SG for the retransmission. This happens in case the UE is correctly identified, but the detection of its payload fails. The scheduled retransmission will happen over dedicated interference-free resources out of the bandwidth allocated to grant-free transmissions. The possibility of using granted resources within the bandwidth allocated for grant-free transmission is left for future work. It is worth noticing that the UE identifier is not to be retransmitted over the granted resources. In case the UE is not identified, neither a SG or a ACK will be issued by the BS; after a ΔT interval, the UE will attempt a retransmission over shared resources similarly to the SAW A scheme. A retransmission over shared resources will also happen in case the UE is correctly identified (but its payload not successfully decoded) at a certain transmission, but it misses the reception of the SG from the BS.

The procedure can repeat until a maximum number Q_{SAW} of transmissions is reached. Q_{SAW} is to be set such that the last transmission is finalized at a time $t \leq \overline{T} - \tau - T_{BS}$, where $\overline{T}$ denotes a maximum tolerable latency. This is meant to give the BS and UE the necessary time for processing the receive data and the feedback message, respectively. In this study we assume $Q_{SAW} = 2$, i.e. a UE can rely at most on a single retransmission in case the first transmission is not identified or correctly decoded, or the SG is missed. This is consistent with the assumptions of previous studies (e.g., [11], [18]) in terms of UE/BS processing times and TTI numerologies adopted within 3GPP for the evaluation of multiple access schemes for NR [1], based on the URLLC target of $\overline{T} = 1$ ms. Such assumptions will be further discussed in Section IV. A generalization of the model for $Q_{SAW} > 2$ based on Markov chains is subject of our current research.

B. BLIND PROCEDURE

In this case, the UE transmits its payload Q_{bl} times over shared resources, in consecutive TTIs. A feedback from the BS is not required. In order to establish a fair comparison with the SAW procedure in terms of overall time dedicated to a packet transmission, we set $Q_{bl} = Q_{SAW} + \frac{\Delta T}{T_{TTI}}$. The last blind transmission is then finalized at the same time as the second transmission of the SAW procedure (as also highlighted in Figure 1).

C. GRANT-BASED TRANSMISSION

As a baseline for our analysis, we consider the case of a traditional grant-based uplink transmission. Upon a packet arrival, the UE sends a scheduling request (SR) to the BS. The BS processes the SR, and responds with a SG which indicates the dedicated resources where the payload is to be transmitted. We assume that such resources are out of the shared bandwidth. In case the transmissions of the SR and SG have the same duration as payload and SG transmission in the SAW A scheme, the payload transmission will be then finalized at the same time as the second transmission in SAW A scheme, as also shown in Figure 1. This means, only one transmission is possible in order to cope with the same latency target of $\overline{T}$ seconds. In practice, scheduling delays due to multiple users simultaneously requesting dedicated resources may jeopardize the possibility of coping with the latency target in case of limited available bandwidth. In the rest of the paper, we will however neglect such scheduling delays and study our baseline grant-based scheme with the optimistic assumption that there is sufficient bandwidth for allowing multiple UEs simultaneously requesting radio resources to perform one transmission without exceeding the latency target.

III. CALCULATION OF OUTAGE PROBABILITY

The goal of an UL grant-free transmission for URLLC is to ensure that a large number of UEs can successfully complete their payload delivery within a limited time with an outage probability not higher than a certain target $P_{out,t}$. In order to calculate the outage probability, let us define the following variables:

- ϵ_I: probability of missed UE identification;
- $P_c(z)$: probability that z users select the same frequency chunk of the UE of interest for their transmissions, with $z = 0, \ldots, N-1$;

- $\epsilon_\gamma(z)$: data decoding error probability for the UE of interest received with average Signal-to-Noise Ratio (SNR) γ in case z interferers are active over the same frequency chunk where the UE is transmitting.
- ϵ_s: probability of missing a signaling message from UE or BS. We consider here the same miss probability for SG and ACK, as well as for the SR for the grant-based option.

It is clear that identifiers, e.g. preambles, need to be designed such that $\epsilon_I^{Q_{SAW}} \ll P_{out,t}$ or $\epsilon_I^{Q_{bl}} \ll P_{out,t}$, depending on the transmission procedure. In case Zadoff-Chu sequences are used as preambles, such requirement may translate to the need of using long sequences which occupy a significant amount of the radio resources [16]. Analyzing resource usage of Zadoff-Chu sequences for reliable UE identification is however out of the scope of this contribution. Note that, though both preamble and payload undergo the same instantaneous fading, their detection is based on different processing, i.e. correlation-based detection for the preamble and data demodulation plus forward error correction for the payload. In the following, we assume for simplicity that the probability of missed UE identification is independent from the data decoding error rate.

From the definitions above, the probability of correctly decoding a payload is given by:

$$P_d = \sum_{z=0}^{N-1} P_c(z)\left[1-\epsilon_\gamma(z)\right]. \tag{1}$$

Observe that, the capability of the receiver to resolve $z \geq 1$ collisions translates to $\epsilon_\gamma(z) < 1$. Collisions can be resolved thanks to the MRC gain and the possibility of eventual low channel gains of the interfering UEs. The probability of correct decoding in case of a collision-free transmission reads

$$P_d^{C_{free}} = 1-\epsilon_\gamma(0). \tag{2}$$

In the following, the final outage probability after the maximum number of transmissions is derived for the three schemes presented in Section II. Note that combining gains resulting from multiple retransmissions are not considered here.

A. SAW PROCEDURE

1) SAW A

The first transmission is successful in case the UE is correctly identified, and its payload decoded. This happens with probability $P_{co,1} = (1-\epsilon_I)P_d$. In case the first transmission is not identified, or it is identified but its payload not decoded, the UE will not receive a positive ACK and will perform a retransmission over shared resources. The probability of a correct reception of the retransmitted payload can be simply calculated as $P_{co,2} = \left(1-P_{co,1}\right)(1-\epsilon_I)P_d$. The failure probability is then given by $P_{out}^{SAW,A} = 1-P_{co,1}-P_{co,2} = [1-(1-\epsilon_I)P_d]^2$.

2) SAW B

The probability of a correct first transmission is given by $P_{co,1}$. In case the first transmission is not properly identified, a blind retransmission over shared resources will happen. The probability of a correct blind retransmission is given by $P_{co,bl} = \epsilon_I(1-\epsilon_I)P_d$. In case the first transmission is correctly identified though not correctly decoded, and the SG is correctly received, the UE will retransmit its payload in dedicated collision-free resources. In this case, the probability of a correct detection is given by $P_{co,de} = (1-\epsilon_I)(1-P_d)(1-\epsilon_s)P_d^{C_{free}}$. In case the first transmission is identified but fails, and the SG is not correctly received, an automatic retransmission will also happen. The probability of its correct detection is given by $P_{co,noFB} = (1-\epsilon_I)(1-P_d)\epsilon_s(1-\epsilon_I)P_d$.

The failure probability within first transmission and retransmission, reads then $P_{out}^{SAW,B} = 1-P_{co,1}-P_{co,bl}-P_{co,de}-P_{co,noFB}$.

Note that the SAW B scheme requires extra granted resources out of the bandwidth allocated for grant-free transmission. Such resources are used with a probability $P_{grant} = P_a(1-\epsilon_I)(1-P_d)$, where P_a denotes the transmission probability of the UE, and will be calculated in Section III-D.

B. BLIND PROCEDURE

The probability of a correct first transmission is also given by $P_{co,1}$. The receiver needs to attempt decoding the second transmission in case the first transmission is not identified, or it is identified but not decoded. The probability of a correct reception is then given by $P_{co,2} = \epsilon_I(1-\epsilon_I)P_d + (1-\epsilon_I)^2(1-P_d)P_d$. For a generic Q-th transmission, with $1 \leq Q \leq Q_{bl}$, the probability of a correct reception is then given by $P_{co,Q} = \sum_{q=1}^{Q}\epsilon_I^{q-1}(1-\epsilon_I)^{Q-q+1}(1-P_d)^{Q-q}P_d$. The final outage probability can be then calculated as $P_{out} = 1-\sum_{Q=1}^{Q_{bl}}P_{co,Q}$.

C. GRANT-BASED PROCEDURE

The single-shot grant based procedure is successful provided the SR and SG are correctly detected, and the payload is decoded. This happens with a probability $P_{co,grant} = (1-\epsilon_s)^2 P_d^{C_{free}}$. The failure probability is then simply given by $P_{out} = 1-P_{co,grant}$.

All the derived final outage probabilities are summarized in Table 1. It is worth observing that the variables and the analytical model presented in this section can also be adopted for deriving error probability of other grant-free schemes than the ones treated here, e.g. the proactive scheme presented in [11].

D. COLLISION PROBABILITY

The expressions of the achievable error probabilities presented in Table I depend on P_d and therefore on the collision probability $P_c(z)$, which is calculated here. Let us start by defining the probability that u UEs transmit a packet in the

TABLE 1. Final outage probabilities.

Scheme	P_{out}
SAW A	$[1-(1-\epsilon_I)P_d]^2$
SAW B	$1-[(1-\epsilon_I)P_d](1+\epsilon_I)+$ $-[(1-\epsilon_I)(1-P_d)]\cdot\left[(1-\epsilon_s)P_d^{C_{free}}+\epsilon_s(1-\epsilon_I)P_d\right]$
Blind	$1-\sum_{Q=1}^{Q_{bl}}\sum_{q=1}^{Q}\epsilon_I^{q-1}(1-\epsilon_I)^{Q-q+1}(1-P_d)^{Q-q}P_d$
Grant-based	$1-(1-\epsilon_s)^2 P_d^{C_{free}}$

same TTI, which is given by [19]

$$P(u\ UEs)=\binom{N-1}{u}P_a^u(1-P_a)^{N-1-u}. \quad (3)$$

The transmission probability P_a will be calculated in the next subsection.

Given such u simultaneously active UEs, the probability that p of them select the same frequency chunk for their transmission is given by:

$$P(u,p)=\binom{u}{p}\frac{(K-1)^{u-p}}{K^u} \quad (4)$$

The $P_c(z)$ probability is then given by

$$P_c(z)=\sum_{u=z}^{N-1}P(u\ UEs)P(u,z)$$
$$=\sum_{u=z}^{N-1}\binom{N-1}{u}P_a^u(1-P_a)^{N-1-u}\binom{u}{z}\frac{(K-1)^{u-z}}{K^u}. \quad (5)$$

1) TRANSMISSION PROBABILITY

In case of Poisson arrivals and single transmission by each user, $P_a = 1-e^{-\lambda}$, with λ being the packet arrival rate [19]. However, in practice P_a is affected by packet failures or missed feedback, which leads to retransmissions.

For the SAW A scheme, P_a can be expressed as $P_a^{SAW,A} = \left(1-e^{-\lambda}\right)[1+\epsilon_I+(1-\epsilon_I)(1-P_d)+(1-\epsilon_I)P_d\epsilon_s]$. The last term in the sum above refers to the case in which the packet is correctly decoded, but the ACK is missed; this does not affect the failure probability but leads to an unnecessary retransmission. Note that, since P_a is also function of P_d and therefore of $P_c(z)$, (5) is to be solved with numerical techniques.

For the SAW B scheme, we have $P_a^{SAW,B} = \left(1-e^{-\lambda}\right)[1+\epsilon_I+(1-\epsilon_I)(1-P_d)\epsilon_s+(1-\epsilon_I)P_d\epsilon_s] = \left(1-e^{-\lambda}\right)[1+\epsilon_I+(1-\epsilon_I)\epsilon_s]$.

For the blind retransmissions approach, the transmission probability is simply given by $P_a^{bl}=Q_{bl}\left(1-e^{-\lambda}\right)$.

E. DECODING ERROR PROBABILITY

The probabilities of correct/erroneous reception presented above are functions of the decoding error probabilities $\epsilon_\gamma(z)$. Such errors are ultimately depending on the transmission rate R in the occupied radio resources. In this work we consider all the UEs to operate at the same rate $R=B/n$, where n stands for the number of resource elements where the codeblock is mapped. In a practical implementation, this corresponds to the case of the same Modulation and Coding Scheme (MCS) used by all the UEs. With the assumptions defined in Section II, the number of available resource elements can be calculated as $n=\left\lceil \frac{W}{K\Delta f}\right\rceil N_s$, with Δf denoting the OFDM subcarrier spacing.

There are well-known expressions in the literature for calculating the outage probability for transmissions over fading channels as a function of the average SNR, the transmission rate R and the receiver type [20]. In case capacity-achieving codes are used, the lowest outage probability for a given R can be obtained. However, classic information theoretical results are built on the assumption that large codeblocks are used, and are not directly applicable to small codeblocks as targeted by uplink grant-free transmissions [21]. Polyanskiy *et al.* [15] have introduced the concept of *channel dispersion* which estimates the rate penalty with respect to Shannon capacity due to the limited codeblock lengths. Achievability (i.e., lower) and converse (i.e., upper) bounds of the rates as a function of the codeblock length were derived in [22] and [23], for a large set of channels including block-memoryless or quasi-static Rayleigh/Rice fading, considering multi-antenna schemes with different degrees of channel knowledge at transmitter and receiver. We characterize here the decoding error probability by considering the *normal* approximation of both achievability and converse bounds of the rate. For the case of a *quasi*-static Single-Input-Multiple-Output (SIMO) Rayleigh fading, where the channel response remains constant across the payload transmission, the following relationship between the normal approximation of the achievable rate R and the decoding error probability ϵ holds:

$$\epsilon\approx E\left[Q\left(\frac{C(\beta)-R}{\sqrt{nV(\beta)}}\right)\right], \quad (6)$$

where $E[\cdot]$ is the expectation operator, $C(\beta) = log_2(1+\beta)$, $V(\beta) = 1-\frac{1}{(1+\beta)^2}$ with β denoting the instantaneous SNR, $Q(x) = \int_x^\infty 1/\left(\sqrt{2\pi}\right)e^{-t^2/2}dt$. We then estimate the error probability $\epsilon_\gamma(z)$ in presence of z interferer by averaging over the distribution of the SINR $f_{\beta,z}$, i.e.

$$\epsilon_\gamma(z)=\int_0^\infty Q\left(\frac{C(x)-R}{\sqrt{nV(x)}}\right)f_{\beta,z}(x)\,dx. \quad (7)$$

For the case of MRC receivers and flat Rayleigh channels where all the interferers are received with the same average SNR γ, $f_{\beta,z}$ reads [24]

$$f_{\beta,z}(x)$$
$$=\frac{x^{M-1}e^{-\frac{x}{\gamma}}}{(M-1)!\gamma^{z+1}}\sum_{p=0}^{M}\binom{M}{p}\frac{\gamma^{p+z}\Gamma(z+p)}{\Gamma(z)(x+1)^{p+z}} \quad (8)$$

for $0 \leq x < \infty$, where $\Gamma(\cdot)$ denotes the Gamma function [25]. Note that the presented model can be easily generalized to other fading profiles and receiver types by using their respective SINR probability density function. It is worth mentioning that Eq.(7) is to be considered as an approximation of the error probability since it assumes that the interference is Gaussian distributed.

IV. PERFORMANCE RESULTS

We analyze the performance of the presented grant-free schemes by considering a number of users transmitting packets of $B = 32$ bytes over shared radio resources according to a Poisson arrival rate. When not differently specified, we assume a bandwidth of $W = 10$ MHz shared by $N = 50$ users. The UEs are power controlled such that their average SNR per antenna when measured over the entire bandwidth is $\gamma = 3$ dB. We assume the same UE power for the different bandwidth allocation cases, i.e. different values for K; operating over a single frequency chunk leads then to an SNR equal to $K\gamma$ given the higher power spectral density. We consider a TTI duration of $N_s = 2$ OFDM symbols, corresponding to a *minislot* in the recently defined NR terminology [1]. For a $\Delta f = 15$ kHz subcarrier spacing, this leads to $T_{TTI} = 0.143$ ms when the same Cyclic Prefix duration (short configuration) as in Long Term Evolution standard [26] is considered. We assume for simplicity that propagation delay is negligible, and processing times as well as feedback transmission time have similar duration of a TTI, i.e. $T_P^{BS} \approx T_P^{UE} \approx T_{FB} \approx T_{TTI}$. As a consequence, from the definition in Section II.B we obtain $Q_{bl} = 5$. When not differently specified, we set $\epsilon_I = \epsilon_s = min\left(10^{-3}, 1 - P_d\right)$, i.e. identification and signaling errors are not larger than the payload decoding error. We further consider here that the specified bandwidth W refers to the resources allocated for the payload, i.e. the overhead for the preamble transmission is not considered in this analysis. From the definition in Section III-E, the rate per TTI turns out to be $R = 0.192 \cdot K$.

The performance is studied by using the analytical framework presented in the previous sections. Simulation results are also included for the sake of validating the analytical model. Simulation results are obtained by generating random packet arrivals and Rayleigh fading coefficients for each user at each receive antenna, and by using the known SINR expression for MRC receiver in case co-channel interferers are present [27]. The decoding error probability is then calculated with Eq.(6), i.e. the actual data coding and decoding of the user packets is not simulated.

A. COLLISION PROBABILITY

Figure 2 displays the probability of having z UEs colliding in a TTI, assuming $K = \{1, 5\}$. The collision probability obviously increases with the packet arrival rate λ. The SAW B scheme leads to the lowest collision probability since the retransmission happens on dedicated resources in case the SG is correctly received. The blind scheme suffers instead from a higher collision rate given the repeated transmissions

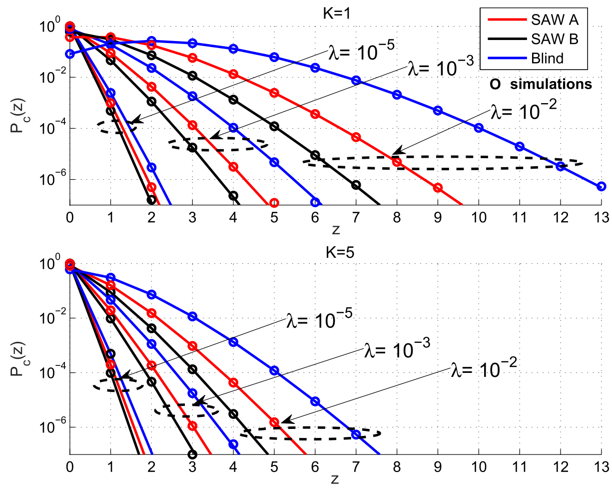

FIGURE 2. Probability of z colliding users in the same TTI.

of each packet. The gap between the different configurations increases visibly with the arrival rate. Obviously, dividing the bandwidth in $K = 5$ frequency chunks reduces significantly the collision rate, though the same trends as in $K = 1$ persist. It is worth noticing that, despite of the presence of 50 users sharing the same resources, the number of statistically relevant interferers is significantly lower even for high packet arrival rates.

B. PROBABILITY OF A DECODING FAILURE

Figure 3 shows the probability of a decoding failure $(1 - P_d)$ as a function of the number K of frequency chunks where the bandwidth is divided, considering the cases of $M = \{2, 4\}$ receive antennas. While for $M = 2$ antennas and frequent packet arrivals ($\lambda = 10^{-2}$), the dependency of the failure probability to the bandwidth allocation is rather weak, lower arrival rates and $M = 4$ antenna configuration clearly benefit from operating over the entire transmission bandwidth ($K = 1$). The advantage of using a robust transmission

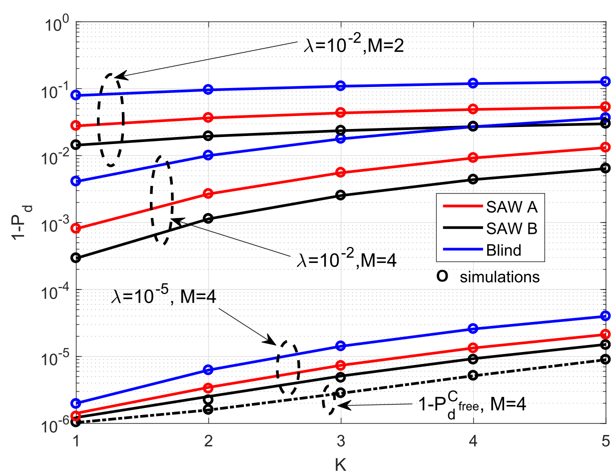

FIGURE 3. Probability of a decoding failure.

rate clearly overcomes the higher power spectral density and reduced collision probability coming from $K > 1$ configurations. The error of contention-free transmission $(1 - P_d^{C_{free}})$ is also shown in Figure 3; this corresponds to the error probability of a traditional grant-based transmission over dedicated resources provided no errors in the SR and SG happen. As expected, no major improvement is visible with respect to transmission over shared resources at low arrival rates.

C. FINAL OUTAGE PROBABILITY

The final outage probability P_{out} as a function of the arrival rate λ is shown in Figure 4 for $M = 2$ receive antennas, assuming $K = 1$. The three grant-free options as well as the grant-based approach are analyzed. The case of no identification or signaling error ($\epsilon_I = \epsilon_s = 0$) is also included, and the target outage probability $P_{out,t} = 10^{-5}$ defined by 3GPP for URLLC services is highlighted [3]. Signaling errors lead to an outage probability floor. The grant-free blind scheme offers the lowest outage probability at low/medium arrival rates. However, it is outperformed by the SAW B scheme at high arrival rates, when the higher number of collisions starts impacting the performance. No benefits of using the SAW A configuration are visible, and its performance approaches the one of grant-based scheme at very high arrival rates. Note, however, that only the blind approach is able to reach a lower failure probability than $P_{out,t}$ at low/medium arrival rates in case of signaling errors. The benefits of assuming no identification/signaling errors are rather minor for the SAW schemes, while the absence of identification errors leads to a major improvement for the blind approach.

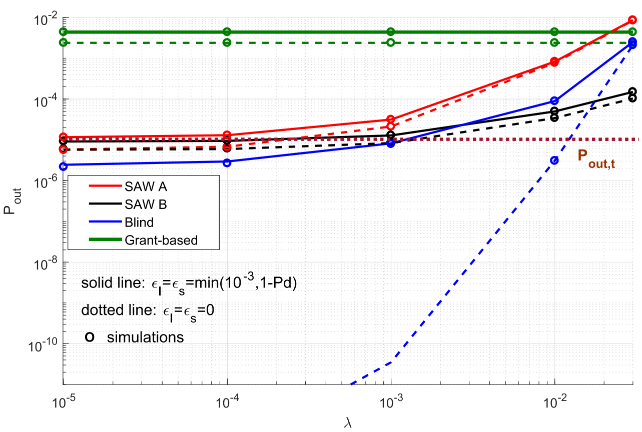

FIGURE 4. **Final outage probability considering $M = 2$ receive antennas at the BS.**

Performance for $M = 4$ receive antennas is shown in Figure 5. The higher combining gain of the MRC receiver translates to a significant reduction of the achievable failure probability. All the configurations (including the grant-based one) are able to cope with the target $P_{out,t}$ at least at low-to-medium arrival rates. The SAW A and blind scheme have a higher final outage than $P_{out,t}$ only for $\lambda > 10^{-2}$. In general, the grant-free configurations clearly overcome the grant-based one for not-too-frequent packet arrivals. Differently

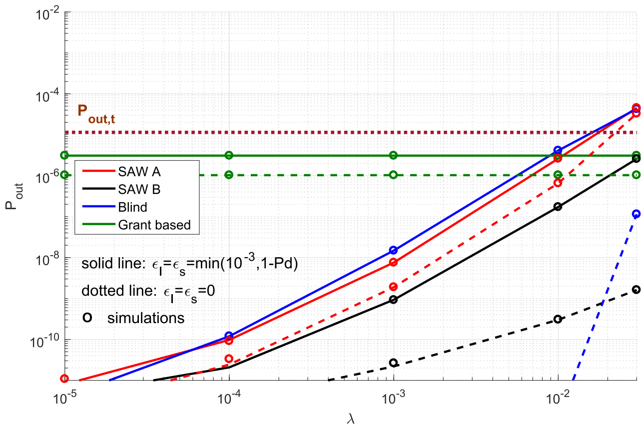

FIGURE 5. **Final outage probability considering $M = 4$ receive antennas at the BS.**

from the 2 receive antenna case, when identification and signaling errors are considered the Blind scheme is now slightly overcome by the SAW A one for $\lambda > 10^{-4}$, while SAW B leads to the lowest outage. The absence of signaling errors leads to limited performance improvement for SAW A, while SAW B and the blind scheme significantly benefit from it.

It is worth recalling that the performance benefit of the SAW B scheme comes at the expense of extra radio resources for the granted retransmissions. For $K = 1$, an amount of resources equal to $W = 10$ MHz and $T_{TTI} = 0.143$ ms is to be allocated out of the shared bandwidth assigned to grant-free operations. Figure 6 displays the probability P_{grant} of allocating such extra-resources as a function of the arrival rate. The probability of allocating an equal or larger amount of dedicated resources for the grant-based option is also shown for comparison. Operating with $M = 2$ receive antennas, leads to a more frequent necessity of allocating extra-resources than in the $M = 4$ case; this is a consequence of the higher decoding failure rate at the first transmission due to the limited MRC gain when only 2 receiver branches are used. On the other hand, the $M = 4$ case is more sensitive to the arrival rate. For $M = 2$, P_{grant} increases indeed of a factor of $\sim 10^5$ when moving from $\lambda = 10^{-5}$ to $\lambda = 0.3 \cdot 10^{-2}$,

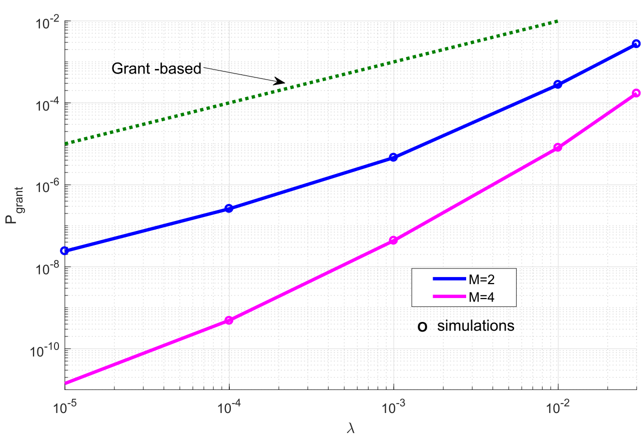

FIGURE 6. **Probability of granting radio resources out of the grant-free shared bandwidth for SAW B scheme.**

compared to a factor of $\sim 10^7$ for $M = 4$. The usage of dedicated resources is in any case significantly lower than the grant-based option even at high λ.

D. RESOURCE EFFICIENCY

The analysis presented above was meant at addressing the suitability of the discussed schemes in achieving a target outage probability for a given scenario, i.e. $W = 10$ MHz and $N = 50$. As a last step for our performance evaluation, we address instead the resource efficiency of the different grant-free schemes for the same target outage probability. Figure 7 displays the minimum required bandwidth W_{min} for achieving the outage probability $P_{out,t} = 10^{-5}$ as a function of the number of users sharing the same resources, assuming $K = 1$, $\epsilon_I = \epsilon_s = min\left(10^{-3}, 1 - P_d\right)$ and $M = 4$ receive antennas. The minimum bandwidth obviously increases with the packet arrival rate. The SAW B scheme is always more efficient than the SAW A scheme in terms of allocated shared resources; as analyzed earlier, this comes at the expense of extra granted resources out of the shared bandwidth in case the SG for the retransmission is correctly received. For $\lambda = 10^{-5}$, the dependency of the required bandwidth from the number of users is negligible, while it becomes significant at higher arrival rates. The blind approach leads to the lowest minimum required bandwidth in case of sporadic arrivals, clearly benefiting from the higher number of retransmissions. At higher arrival rates, the blind approach is instead penalized by the collisions, and the minimum required bandwidth increases. The SAW B scheme becomes more efficient than the blind scheme at $\lambda = 10^{-3}$ for a sufficiently large number of users. At frequent arrival rates, the SAW B scheme is largely the most efficient solution, while the blind scheme requires same or even larger bandwidth than the SAW A scheme for a high number of users.

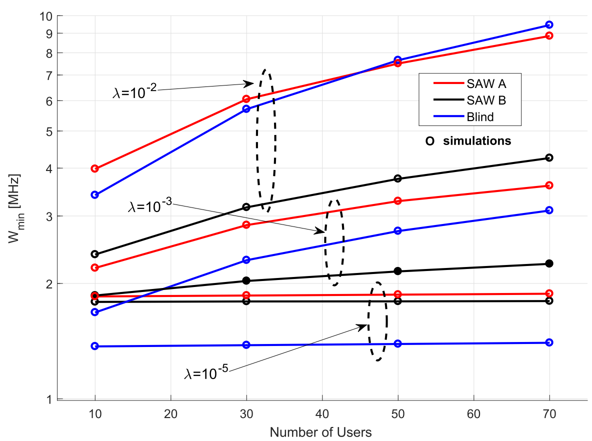

FIGURE 7. Minimum required bandwidth for achieving the target outage probability $P_{out,t} = 10^{-5}$.

V. CONCLUSIONS AND FUTURE WORK

In this paper we have studied the reliability of different uplink grant-free schemes over shared radio resources with a constraint in maximum number of transmissions, considering operations over Rayleigh fading channels and a Maximum Ratio Combining (MRC) receiver at the base station. In particular, we have considered two Stop-And-Wait (SAW) approaches and a solution based on blind retransmissions, while a traditional grant-based scheme is studied as a baseline. We have presented an analytical model for such schemes, where identification and signaling errors are also included and the decoding error probability is based on recent results on limited codeblock length transmission.

Results show that operating over a large bandwidth is beneficial in terms of packet failure rate despite of the lower power spectral density and the higher vulnerability to collisions, thanks to the lower coding rate. In case of 2 receive antennas, the grant-free schemes can hardly achieve a target outage probability of 10^{-5}, especially at high packet arrival rates. The blind approach shows the best performance at least at low and medium packet arrival rate, especially in the absence of identification and signaling errors. The failure probability is instead significantly lower than the target in case of 4 receive antennas, with the SAW scheme with scheduled retransmissions performing as the best in the presence of signaling errors. In general, the grant-free options outperform traditional grant based transmission for the same tight latency target. The SAW scheme with scheduled retransmissions is also the most efficient in terms of required shared resources at frequent packet arrival rates, given the possibility of relying on scheduled retransmissions out of the shared bandwidth.

We believe the presented model and results can be used as a reference for empirical system level analysis of uplink-grant free solutions. Future work will consider the impact of optimal combining receivers with interference suppression capabilities. Further, a detailed analysis on the signaling errors including miss detection of reference sequences will be carried out.

ACKNOWLEDGEMENTS

This research is partially supported by the EU H2020-ICT-2016-2 project ONE5G. The views expressed in this paper are those of the authors and do not necessarily represent the project views.

REFERENCES

[1] *Study on New Radio Access Technology—Physical Layer Aspects*, document 38.802 v14.2.0, 3rd Generation Partnership Project, 2017.

[2] D. Jiang, H. Wang, E. Malkamaki, and E. Tuomaala, "Principle and performance of semi-persistent scheduling for VoIP in LTE system," in *Proc. Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Sep. 2007, pp. 2861–2864.

[3] *Study on Latency Reduction Techniques for LTE*, document 36.881 v14.0.0, 3rd Generation Partnership Project, 2016.

[4] L. Dai, B. Wang, Y. Yuan, I. Han, S. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[5] L. G. Roberts, "ALOHA packet system with and without slots and capture," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 5, no. 2, pp. 28–42, 1975.

[6] E. Paolini, C. Stefanovic, G. Liva, and P. Popovski, "Coded random access: Applying codes on graphs to design random access protocols," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 144–150, Jun. 2015.

[7] H. S. Dhillon, H. C. Huang, H. Viswanathan, and R. A. Valenzuela, "Fundamentals of throughput maximization with random arrivals for M2M communications," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 4094–4109, Nov. 2014.
[8] H. Nicopur and H. Baligh, "Sparse code multiple access," in *Proc. 24th IEEE Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 332–336.
[9] L. Ping, L. Liu, K. Wu, and W. K. Leung, "Interleave division multiple-access," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 938–947, Apr. 2006.
[10] D. Malak, H. Huang, and J. G. Andrews, "Fundamental limits of random access communication with retransmissions," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.
[11] T. Jacobsen *et al.*, "System level analysis of uplink grant-free transmission for URLLC," in *Proc. IEEE Globecom*, Dec. 2017, pp. 1–6.
[12] B. Singh, O. Tirkkonen, Z. Li, and M. A. Uusitalo, "Contention-Based Access for Ultra-Reliable Low Latency Uplink Transmissions," *IEEE Wireless Commun. Lett.*, vol. 7, no. 2, pp. 182–185, Apr. 2010.
[13] M. K. Simon and M. S. Alouini, *Digital Communications Over Fading Channels*. 2nd ed. Hoboken, NJ, USA: Wiley, 2005.
[14] H. Shariatmadari, Z. Li, S. Iraji, M. U. Uusitalo, and R. Jäntti, "Control channel enhancements for ultra-reliable low-latency communications," in *Proc. Int. Conf. Workshop Commun. (ICC)*, May 2017, pp. 504–509.
[15] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
[16] *Preamble Design for UL Grant-Free Transmission*, document R1-1708526, 3GPP, TSG-RAN WG1 Meeting #89, May 2017.
[17] B. M. Popovic, "Generalized chirp-like polyphase sequences with optimum correlation properties," *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1406–1409, Jul. 1992.
[18] R. Abreu, P. Mogensen, and K. I. Pedersen, "Pre-scheduled resources for retransmissions in ultra-reliable and low latency communications," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2017, pp. 1–5.
[19] J. Proakis, *Digital Communications*, 4th ed. New York, NY, USA: McGraw-Hill, 2001.
[20] R. Janaswamy, *Radiowave Propagation and Smart Antennas for Wireless Communications*. Orlando, FL, USA: Academic, 2000.
[21] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Aug. 2016.
[22] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, Jul. 2014.
[23] G. Durisi, T. Koch, J. Östman, Y. Polyanskiy, and W. Yang, "Short-packet communications over multiple-antenna Rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 618–629, Feb. 2016.
[24] V. A. Aalo and J. Zhang, "Performance analysis of maximal ratio combining in the presence of multiple equal-power cochannel interferers in a Nakagami fading channel," *IEEE Trans. Veh. Technol.*, vol. 50, no. 2, pp. 497–503, Mar. 2001.
[25] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York, NY, USA: McGraw-Hill, 1991.
[26] H. Holma and A. Toskala, *LTE for UMTS: OFDMA and SC-FDMA Based Radio Access*. Hoboken, NJ, USA: Wiley, 2009.
[27] J. Cui and A. U. H. Sheikh, "Outage probability of cellular radio systems using maximal ratio combining in the presence of multiple interferers," *IEEE Trans. Commun.*, vol. 47, no. 8, pp. 1121–1124, Aug. 1999.

GILBERTO BERARDINELLI received the bachelor's and master's degrees (*cum laude*) in telecommunication engineering from the University of L'Aquila, Italy, in 2003 and 2005, respectively, and the Ph.D. degree from Aalborg University, Denmark, in 2010. He is currently an Associate Professor with the Wireless Communication Networks Section, Aalborg University. He is also in tight cooperation with Nokia Bell Labs. He is currently part of the EU funded research project ONE5G that focuses on end-to-end-aware optimizations and advancements for the network edge of fifth-generation (5G) new radio. His research interests are mostly focused on physical layer, medium access control, and radio resource management design for 5G systems.

NURUL HUDA MAHMOOD was born in Chittagong, Bangladesh. He received the M.Sc. degree in mobile communications from Aalborg University (AAU), Denmark, in 2007, and the Ph.D. degree in wireless communications from the Norwegian University of Science and Technology, Norway, in 2012. He has been with the Wireless Communication Networks Section, AAU, since 2012. He is also contributing to the EU funded research project ONE5G. He has authored or co-authored over 50 peer-reviewed publications. His current research interests include resource optimization techniques for fifth generation and modeling and performance analysis of wireless communication systems.

RENATO ABREU received the M.Sc. degree in electrical engineering from the Federal University of Amazonas, Brazil, in 2014. From 2008 to 2016, he was with the Nokia Technology Institute (INDT), Brazil, where he developed industrial test systems for mobile devices and researched wireless systems. Since 2016, he has been a Ph.D. Fellow with the Wireless Communication Networks Section, Aalborg University, Denmark, where he is in cooperation with Nokia Bell Labs. His research focus is on ultra-reliable and low-latency communication for fifth-generation networks.

THOMAS JACOBSEN received the M.Sc. degree in engineering (network and distributed systems) from Aalborg University, Denmark, in 2015. He is currently a Ph.D. Fellow with the Wireless Communication Networks Section, Aalborg University, where he is cooperating closely with Nokia Bell Labs. His research focus is on radio resource mechanisms for ultra-reliable low-latency communications. He drives a student society focusing on autonomous vehicle communications and system development for demons, one which was present at MWC 2018.

KLAUS PEDERSEN received the M.Sc. degree in electrical engineering and the Ph.D. degree from Aalborg University, Aalborg, Denmark, in 1996 and 2000, respectively. He is currently leading the Nokia Bell Labs Research Team, Aalborg, and he is a part-time Professor with the Wireless Communications Network Section, Aalborg University. He has authored or co-authored approximately 160 peer-reviewed publications on a wide range of topics and holds several patents. His current research is related to fifth-generation (5G) new radio standardization and related concept research with special emphasis on radio resource management and mechanisms that offer improved end-to-end performance delivery. He is currently part of the EU funded research project ONE5G.

ISTVÁN Z. KOVÁCS received the B.Sc. degree from the "Politehnica" Technical University of Timi oara, Romania, in 1989, the M.Sc.E.E. degree from the École Nationale Supérieure des Télécommunications de Bretagne, France, in 1996, and the Ph.D.E.E. degree in wireless communications from Aalborg University, Denmark, in 2002. He is currently a Senior Wireless Networks Specialist with Nokia, Aalborg, Denmark, where he conducts research on radio connectivity evolution for machine-type and aerial vehicle communications in LTE and fifth-generation networks.

PREBEN MOGENSEN received the M.Sc. and Ph.D. degrees from Aalborg University in 1988 and 1996, respectively. Since 2000, he has been a Full Professor and is currently leading the Wireless Communication Networks Section, Department of Electronic Systems, Aalborg University. He has supervised over 35 successfully finalized Ph.D. candidates and has co-authored over 300 papers in various domains of wireless communication. Since 1995, he has also been part-time associated with Nokia, and currently a Principal Engineer with Nokia Bell Labs. His current research focus is on fifth generation and machine-type communication /Internet of Things.