

# Enabling Ultra-Reliable and Low-Latency Communications through Unlicensed Spectrum

Gordon J. Sutton, Jie Zeng, Ren Ping Liu, Wei Ni, Diep N. Nguyen, Beeshanga A. Jayawickrama, Xiaojing Huang, Mehran Abolhasan, and Zhang Zhang

## ABSTRACT

In this article, we aim to address the question of how to exploit the unlicensed spectrum to achieve URLLC. Potential URLLC PHY mechanisms are reviewed and then compared via simulations to demonstrate their potential benefits to URLLC. Although a number of important PHY techniques help with URLLC, the PHY layer exhibits an intrinsic trade-off between latency and reliability, posed by limited and unstable wireless channels. We then explore MAC mechanisms and discuss multi-channel strategies for achieving low-latency LTE unlicensed band access. We demonstrate, via simulations, that the periods without access to the unlicensed band can be substantially reduced by maintaining channel access processes on multiple unlicensed channels, choosing the channels intelligently, and implementing RTS/CTS.

## INTRODUCTION

Fifth generation (5G) wireless communications is expected to bring disruptive technologies that support ultra-reliable, low-latency communications (URLLC). The unlicensed spectrum plays an important role in 5G as it offers a significant capacity boost to licensed spectrum. Emerging 5G services can be categorized into two types: machine-type communications (MTC)-based services and human-machine interaction services. The former aims to fulfill communications tasks between connected devices in applications requiring, for example, control and safety, such as industrial control systems and automation, and autonomous vehicles and robotics. Human-machine interaction services include virtual reality and augmented reality in which almost immediate interaction between eye, body, and surrounding environment is required. The various services have different latency and reliability requirements that can be as extreme as 0.5 ms for end-to-end latency and 99.999 percent for reliability. While delivering carrier grade services in unlicensed spectrum can be a challenge, due to the shared nature of unlicensed spectrum, the unlicensed band offers additional diversity and bandwidth that can be used to achieve URLLC.

Long Term Evolution (LTE) in unlicensed spectrum (LTE-U) and licensed assisted access (LAA) have recently been proposed to exploit unlicensed band of WiFi for cellular devices/applications. Unlike LTE-U/LAA, which use the LTE carrier as an anchor (aggregated with the

unlicensed carrier via either link or carrier aggregation approaches), Qualcomm has recently proposed to operate 5G on unlicensed band through a technology called MulteFire. As the success of 5G for MTC hinges on the ability to provide highly reliable connections, it is critical (and challenging) to improve reliability for unlicensed band 5G communications (i.e., LTE-U/LAA/MulteFire and beyond) that share spectrum with WiFi access points and different network operators.

The Third Generation Partnership Project (3GPP) defines general URLLC latency and reliability targets in [1]. User plane latency is the average time to transmit a packet between radio protocol layers 2/3 without discontinuous reception. Reliability is the success probability of delivering a small packet within a time constraint. Reference [1] prescribes 0.5 ms latency per link and 99.999 percent reliability for a 32-byte packet with 1 ms user plane latency. URLLC needs to support a variety of use cases that have a variety of latency and reliability constraints. In some cases, the 3GPP latency restriction can be relaxed to 5 ms or longer. In other cases, the reliability should be guaranteed at 99.99999 percent or even higher.

Reliability can be increased by controlling power levels and scheduling resources to sub-carriers with favorable conditions. However, combining transmissions from multiple different paths provides greater protection against fading and interference. Different paths can be created by diversity in frequency, space, or connection. Whichever way, the less correlated the different paths, the greater the reliability.

Latency is bound to the frame structure. The transmission time interval (TTI) and symbol duration impose lower bounds on transmission times. Device-to-device (D2D) links can reduce the latency between two nearby devices by replacing the uplink-downlink (UL-DL) pair with a single hop. Grant-free non-orthogonal multiple access (NOMA) [2] reduces latency by allowing immediate transmissions without scheduling.

The unlicensed spectrum has largely been studied as a tappable resource that will increase network operator capacity. Instead, it could be considered another source of diversity by which to increase reliability. A flexible network structure could allow evolving multi-connectivity, where D2D links add another pathway when possible, and where unlicensed channels add another dimension as access is obtained. The unlicensed spectrum could also facilitate decreased latency by allowing a packet to be transmitted on either

the unlicensed or the licensed spectrum, whichever becomes available first.

In this article, we consider how to utilize the unlicensed spectrum for URLLC. 5G networks will have a new design, referred to as New Radio, but it is expected to be, in essence, an extension of WLAN and LTE. As such, we explore achieving URLLC with the mechanisms currently offered by WLAN and LTE. We start by reviewing physical (PHY) layer techniques pertaining to URLLC, particularly in the unlicensed spectrum. The relative merits of the PHY techniques are assessed in terms of their capacity to enable URLLC. Altering the frame structure and numerology has the potential to reduce latency. Simulations are presented which demonstrate that creating diversity, either in frequency, space, or connection, increases reliability. After the PHY layer, the medium access control (MAC) layer is considered. We discuss the possibility of gaining quick access to multiple unlicensed channels to achieve our URLLC goal. We identify a number of components of a potential URLLC scheme, including predicting traffic profiles, dynamically accessing multiple unlicensed channels, and implementing request-to-send/clear-to-send (RTS/CTS). The aim would be to reduce the periods without access between transmission opportunities on unlicensed channels, ideally achieving continuous access so that the unlicensed spectrum could be used for URLLC in its own right, or as part of a broader URLLC solution. We demonstrate that the latency to access can be significantly reduced.

## USE CASES

In the next decade, numerous URLLC applications will come into our daily life, predictably or unpredictably. However, due to the limitation of frequency bandwidth, some URLLC applications will be implemented on unlicensed bands. In that case, terminals operate in a coexistent and competitive way, continuously monitoring channels for opportunistic access. With strict reliability and latency constraints, many scenarios and particular environments [1–4] are emerging. Figure 1 depicts the URLLC use cases we believe are the most promising for unlicensed band realization.

**Augmented Reality:** The elements of the environment are augmented by computer-generated sensory information; the latency should be maintained at an ultra-low level to avoid giddiness.

**Tactile Internet:** An Internet network enables operating as well as observing from a distance; high reliability with less than 1 ms round-trip latency should be supported.

**Vehicle-to-Vehicle (V2V):** This includes communications between vehicles to share safety messages containing location/speed/hazard information and facilitate accident avoidance and, in the long term, support fully automated driving.

**Intelligent Transport System (ITS):** Infrastructure, vehicles, and users are connected with information and communication systems to support efficient traffic; it is proposed to be deployed in the 5.9 GHz unlicensed band.

**Underground Mine Pollution Monitoring:** This involves atmospheric pollution and radiation monitoring in underground mines; reliable wireless sensor nodes need to be deployed at low-frequency unlicensed bands (915 MHz or lower).

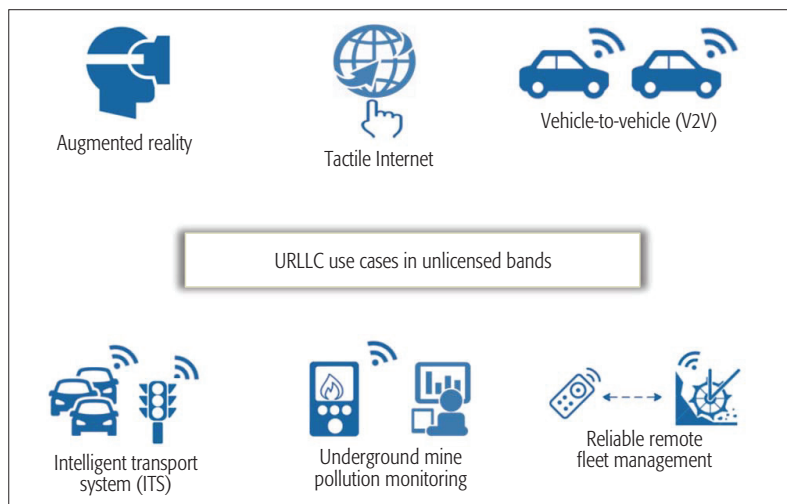


FIGURE 1. Promising URLLC use cases in unlicensed bands.

**Reliable Remote Fleet Management:** This is operated in an open-pit mine to coordinate different mining equipment to maximize capital utilization; 99.999 percent reliability should be guaranteed.

## PHY LAYER ADVANCES AND INTRINSIC LIMITATION

One key PHY technique that enhances reliability is diversity in the frequency and space domains. Diversity can be achieved through aggregation of multiple subcarriers and antennas. To enhance spatial diversity, multiple antennas can be collocated on the same device, or the aggregation can be in the form of coordinated multipoint transmission. Frequency diversity gain can be leveraged through frequency hopping techniques, which enable transmitting in different channels with a scheduled hopping pattern in a broadband rich-scattering environment. Such techniques make the links robust against deep fades and collisions. Generally, with many available channels, frequency hopping achieves diversity and power gains.

Adaptation of modulation and coding scheme (MCS) to channel conditions will play a key role in URLLC applications. A low-complexity MCS is desirable when there is a tight processing delay budget. Recently, automated rate adaptation with limited feedback has been utilized to guarantee ultra-reliability and achieve minimized transmission delay. For example, in [5], a robust link adaptation is enabled to select suitable data rates for desired reliability (99.999 percent) and delay constraints. However, automated rate adaptation relies heavily on the accuracy of channel state information (CSI). CSI measurement is critically challenging in unlicensed URLLC because of the strictly limited delay budget. Recently, some typical codes have been redesigned and optimized for short packet transmissions in MTCs, such as low-density parity check, turbo, extended Bose-Chaudhuri-Hocquenghem, polar, convolutional, and analog fountain codes [2, 3].

Further, PHY waveform design provides the potential to substantially reduce the latency and unwanted emissions that cause interference. Out-of-band emission reduction techniques, such as filtering, windowing, cancellation carrier, and

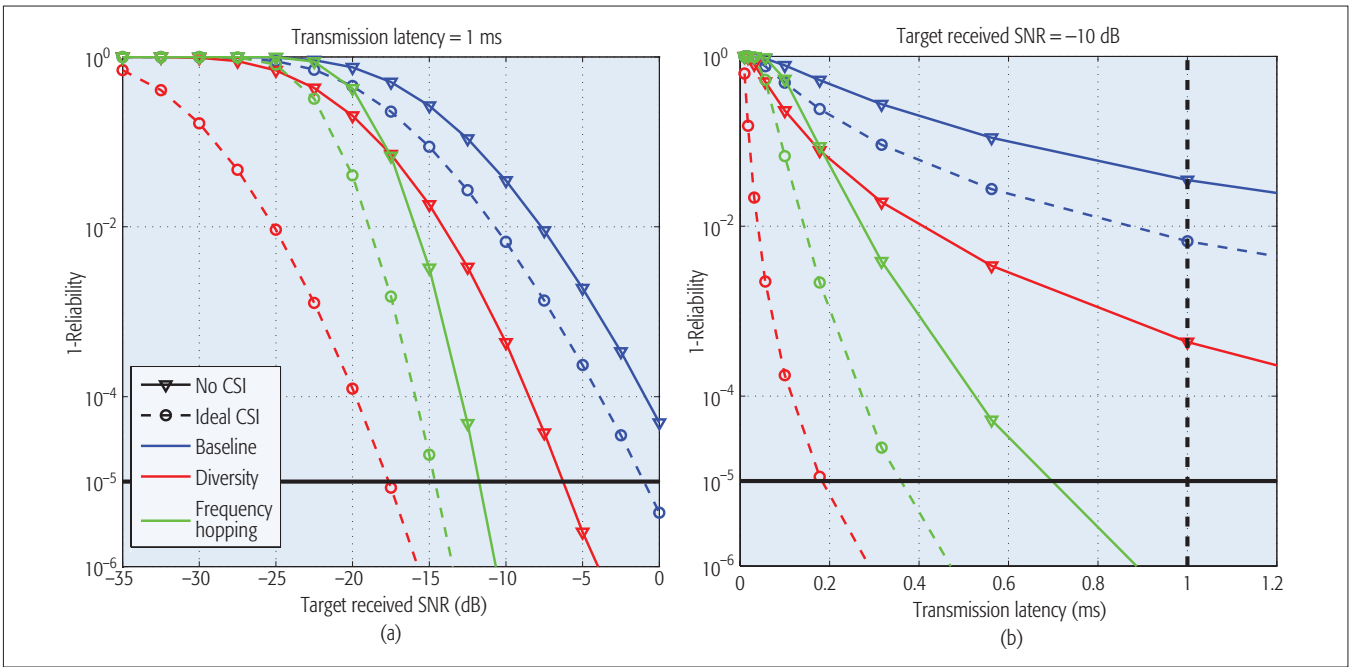


FIGURE 2. Relationship among target received SNR, transmission latency, and reliability: a) reliability vs. target received SNR under 1 ms transmission latency constraint; b) reliability vs. transmission latency under -10 dB target received SNR.

spectral precoding techniques [6], can be utilized at the transmitter to avoid interference to other users and systems.

Other enabling technologies include designing a frame structure to meet 1 ms user plane latency [3], wideband spectrum sensing techniques [7], and allowing grant-free access [2]. Recent improvements in self-interference suppression (or in-band full-duplex) can be used to reduce the sensing delay for URLLC. Additionally, grant-free NOMA, which benefits from advanced receiver design, can remove the grant-request and scheduling processes, thereby reducing the user plane latency without remarkably decreasing reliability. However, such techniques need to be thoroughly evaluated for their ability to meet URLLC requirements.

Simulations are performed to demonstrate the reliability of different PHY techniques under given transmission latency and target received signal-to-noise ratio (SNR). It is assumed that a 32-byte packet is appropriately encoded by a dedicated channel coding scheme to be transmitted in a channel experiencing block Rayleigh fading with shadow fading, in which the block size of block Rayleigh fading is  $20 \text{ MHz} \times 1 \text{ ms}$ , and the shadow fading obeys the log-normal distribution with 4 dB shadowing standard deviation. The transmission latency is defined as the total transmission time, equaling the sum of UL and DL transmission symbol durations. The target received SNR is defined as the long-run average received SNR over 1 s of transmission, which is much longer than the coherence time.

In each iteration of the simulation, the target received SNR and transmission latency are given as constraints, and a realization of the Rayleigh channel is generated independently. The selected PHY techniques are implemented in turn and assessed on reliability, that is, the proportion of time the 32-byte packet is transmitted error-free through both UL and DL. Theoretical reliability

probabilities are also calculated and compared to the simulation results.

In the "baseline" scenario, two antennas are equipped at each of the source and destination user equipment (UE), and two antennas are equipped at the base station. In the "diversity" scenario, four antennas are equipped at the source and destination UE, respectively, while 16 antennas are equipped at the base station, creating a 16-fold diversity gain over the baseline. For the frequency hopping scheme, each transmission hops between five independent 20 MHz channels. We assume that when ideal CSI is acquired at the transmitter side, perfect transmission beamforming is used, resulting in an  $M$ -fold power gain, where  $M$  is the number of antennas at the transmitter side. Otherwise, equal power transmission is used when no CSI is acquired.

Figure 2a plots (1 - reliability) vs. target received SNR under 1 ms transmission latency. Diversity and frequency hopping provide noticeable gains due to frequency/space diversity. CSI measurement also brings remarkable improvement from their power gains when ideal CSI or precise CSI is obtained at the transmitter. While frequency hopping seems to be the most efficient way to increase reliability, diversity provides the largest reduction in the target received SNR required to achieve a given reliability, requiring approximately 16 dB less than the baseline for  $1 - 10^{-5}$  reliability. Figure 2b instead plots (1 - reliability) vs. transmission latency when the target received SNR is -10 dB. Diversity and frequency hopping with ideal CSI can achieve the  $1 - 10^{-5}$  reliability target when transmission delay is very short. At the same time, frequency hopping without CSI could fulfill the 1 ms total latency requirement when propagation and processing delay are very short. Thus, it can be concluded from the figures that while reliability can be consistently enhanced by increasing the transmission power, under given transmission latency and bandwidth

constraints, from the perspective of avoiding severe interference, diversity and frequency hopping techniques are more energy-efficient and can be coexistence-friendly, so are suitable for deployment in unlicensed bands.

The relevance to various PHY techniques to URLLC are compared in Table 1 under three performance measures: latency, reliability, and interference. The interference is assessed on a combination of the in-band emission, which affects the coexisting system and out-of-band emission, which affects the adjacent channels. Relevance refers to having an impact, either positive or negative.

## URLLC IMPLICATIONS AT THE MAC LAYER

In this section, we explore MAC layer access mechanisms that impact URLLC in the unlicensed band. To that end, we start with the MAC layer access mechanisms in the licensed band.

### MECHANISMS IN LICENSED BAND

Access delays start when establishing a default bearer, either initially, or after a period of inactivity. Congestion during random access adds to access delay and is controlled, or contained, by using (extended) access class barring, segmented preamble sets, and dedicated preambles. The 3GPP Allocation and Retention Priority helps control congestion by restricting bearer access, and also allows quick cancelling of bearers to make way for higher-priority users in an emergency. Default bearers maintain a connection and are scheduled as best effort. Additional bearers can be established with a quality of service (QoS) class identifier that defines priority, latency, and packet error rate requirements, and whether a guaranteed bit rate is required. Bearers can be configured with dynamic scheduling or semi-persistent scheduling. The UE reports the state of the channel through the CSI, including the channel quality indicator, which reports the MCS selection and resources required, thereby increasing reliability. Alternatively, when UE are closely located, D2D bearers can be established.

In dynamic scheduling, resource blocks are often allocated each TTI based on maximizing per-bearer metrics [8]. Metrics can be designed to maximize fairness, maximize resource efficiency, meet quality of service (QoS) constraints, or balance multiple objectives. For URLLC, satisfying QoS constraints takes highest priority. Advanced schedulers aim to optimize metrics and meet QoS constraints over a longer period than just one TTI. One such scheduler is the frame level scheduler (FLS) [9], which initially schedules over a 10 ms frame, aiming to meet QoS constraints; then, for each TTI, it incorporates channel quality feedback to try to improve the efficiency of the allocations. Once QoS constraints are met, it also allocates resources to other bearers, aiming to maintain fairness. In the unlicensed band, a similar two-tier idea can be applicable. Specifically, resources can be scheduled over multiple frames, based on access predictions, and then as access occurs, the scheduling is revised.

D2D communications allow direct short-range links, instead of communicating through the evolved Node B (eNB). D2D communications in the overlay mode, where the eNB schedules dedicated D2D resources, is possible in the un-

Access delays start when establishing a default bearer, either initially, or after a period of inactivity. Congestion during random access adds to access delay and is controlled, or contained, by using (extended) access class barring, segmented preamble sets, and dedicated preambles.

	Latency	Reliability	Interference
Frame structure	3	1	1
Waveform design	2	2	3
Diversity	1	3	2
MCS	2	2	1
CSI measurement	2	2	2
Frequency hopping	1	3	3
Grant-free NOMA	3	2	2

TABLE1. Relevance of various PHY techniques to URLLC in unlicensed bands (1: low relevance; 2: middle relevance; 3: high relevance).

censed spectrum if scheduled and completed during the eNB's channel occupancy time (COT). Such scheduling would approximately halve the latency by reducing the number of hops from two to one, thereby helping facilitate URLLC.

### Wi-Fi MAC LAYER IMPLICATIONS

**PCF/HCF:** The point coordination function (PCF) and hybrid coordination function (HCF) use polling with priority defer periods, PCF interframe spaces (PIFSs), to create contention-free periods (CFPs). The HCF defines priority classes as part of the enhanced distributed channel access (EDCA) mechanism, where higher-priority access classes have shorter backoff processes. The highest two access classes also have a specified maximum transmission opportunity (TXOP), during which multiple packets may be sent. From the perspective of LTE accessing the unlicensed band, long CFPs pose an access problem. To accommodate the CFPs, the eNB can include a Wi-Fi receiver to decode the beacons that announce the duration of each CFP. The eNB could then use the beacon information to decide whether to defer from channels during CFPs, or to compete, using high-priority access classes that have defer periods equal to the PIFS. Moreover, the information would make access to the unlicensed band more predictable, which in turn makes utilizing the unlicensed band for URLLC more feasible.

**NAV, RTS/CTS:** The RTS/CTS mechanism is a two-way handshake, aiming to alleviate the hidden node problem. Specifically, stations inform nearby nodes about their incoming transmissions, and the nodes set their network allocation vectors (NAVs) to the transmission duration as a form of virtual carrier sensing. If all transmitting devices use RTS/CTS, the cost of collisions can be vastly reduced, especially with long aggregated packets. With a Wi-Fi transmitter, an eNB could send CTS-to-self frames, advertising to Wi-Fi receiving devices its intent to transmit; however, this would eat into the eNB's COT. With a Wi-Fi receiver, the eNB could use any advertised packet durations to help build a traffic profile and predict when the channel would be available.



Technique	LAA application
EDCA	LAA priority classes, to gain shorter access more often
Carrier aggregation	To enable access to multiple unlicensed channels
Diversity	Multi-unlicensed-channel diversity to reduce access discontinuities
RTS/CTS	Used to reduce collision time; recipient UE selected randomly or based on its distance from eNB
TTI reduction	To allow DL control and UL transmission within a COT (e.g., 2 ms)
Flexible frame structure	To accommodate different balance of control, DL and UL
Scheduling (FLS)	Unlicensed channel traffic monitoring, apply FLS over LTE frames

TABLE 2. Techniques with potential to reduce discontinuity of access to unlicensed bands.

**Carrier Aggregation:** In the very high throughput mode, Wi-Fi can aggregate carriers and transmit on up to 160 MHz of spectrum, being either contiguous or in two 80 MHz bands. This poses a problem of simultaneous access block out on a potentially large number of channels for other devices, including the eNB. As such, when building a traffic profile, the eNB should ideally account for possible correlations between channel occupancies. Again, traffic profiling could be facilitated by reading the Wi-Fi headers.

#### LTE ACCESS TO THE UNLICENSED SPECTRUM

Load-based listen-before-talk has been the consensus mechanism for LTE access to the unlicensed spectrum. There are currently two slightly different versions, 3GPP [10] and European Telecommunications Standards Institute (ETSI) [11]. Both are backoff-based with four access priority classes, and are similar to the Wi-Fi EDCA mechanism. Higher priority classes have shorter minimum contention windows, fewer backoff stages, and shorter defer periods, so access occurs more frequently. To counter the advantage, a shorter COT is allowed each access. In DL, the highest priority class defer period equals a PIFS, as used by a Wi-Fi point/hybrid coordinator.

ETSI defines initiating and responding devices. Responding devices, such as UE, can transmit within the COT obtained by an initiating device, separated by short breaks, which is one mechanism for UL transmission. UE can also obtain access via their own backoff processes. There are again four access priority classes, which are similar to the downlink set, the main difference being that an additional slot is added to the defer periods of the highest two priority classes.

3GPP has Type A and Type B multi-carrier access. In Type A, a separate backoff process is maintained for each carrier. This appears to have potential for reducing the periods without access. However, it is implied that once one backoff process gains access and a transmission starts, the other backoff processes are paused, which reduces the potential gain. In the ETSI equivalent, Option 1, the backoff processes may be independent if the devices are capable of maintaining independent transmissions on separate carriers.

Type B multi-carrier access has a single backoff process that is maintained on one carrier and used for all carriers. Just prior to transmission, all carriers in the multi-carrier set are sensed to assess channel activity. Transmissions may be made on

all carriers that are found to be idle. The channel on which the backoff process is performed is either selected randomly after each transmission or selected arbitrarily, but no more often than every second. As such, although the Type B multi-carrier option potentially increases capacity, it does not reduce the periods without access to the unlicensed spectrum. The ETSI equivalent, Option 2, is like 3GPP Type B.

#### POTENTIAL URLLC RESEARCH DIRECTIONS

The implementation of URLLC in the unlicensed band is hampered by discontinuous access. We explore combining both PHY and MAC techniques to alleviate the access discontinuity, and propose using the unlicensed spectrum as another source of diversity. Table 2 presents techniques that might work together to this end and explains their contributions.

We next explore the challenges of achieving multichannel diversity in the unlicensed band, and then present simulation results showing the potential of multi-unlicensed-channel diversity.

#### MULTICHANNEL DIVERSITY

Achieving URLLC over unlicensed channels is challenging. This is because the access to unlicensed channels is usually based on contention and not always guaranteed. That can lead to longer latency or even transmission disruptions, affecting the reliability. To mitigate the above, one can leverage the canonical diversity and frequency hopping techniques that were used to combat fading or jammers. Specifically, instead of relying on a single unlicensed channel, a set of unlicensed channels can be monitored simultaneously, and the link can opportunistically and proactively hop onto these channels to avoid transmission disruption. To that end, we need to address several challenges.

First, one needs to design flexible channel bonding methods [12] that allow unlicensed channels to be quickly and efficiently subdivided and aggregated. We believe recent advances in multi-carrier communications, for example, non-contiguous orthogonal frequency-division multiplexing, are very suitable. However, the challenge here is how to realize channel subdivision and aggregation in a mixed pool of unlicensed and licensed channels [13]. Second, a mechanism that allows the transmitter and receiver to rendezvous with low time and signaling overhead must be developed. Unfortunately, although various rendezvous methods have been proposed for the licensed bands in the literature, their performance has not been investigated/confirmed in the unlicensed bands. Additionally, a fast frequency hopping method can also be recruited. To guarantee a given reliability level, we also need to quantify the impact of the number of unlicensed channels, their traffic profiles, and transmission durations on the resulting reliability. To that end, the theory of optimal stopping point (OSP) [14] can be a robust tool. Specifically, given the underlying probabilistic properties of each channel, OSP allows us to decide when to hop to another channel to maximize a given utility function (either minimizing latency or maximizing reliability).

In a cellular system, the base station has to be equipped with the capability to learn and profile

multiple unlicensed channels simultaneously. It then can forecast its transmission opportunities in different bands/channels, and hence notify the UE in advance to rendezvous. How to estimate the traffic on each channel and then how to predict the next channel to produce an uncontested access opportunity are challenging research topics. A potential approach is to rely on Q-learning [15]. A trade-off is expected between reliability and average performance, where minimizing the 95th quantile would be preferable to minimizing the average. In the following, we investigate the efficacy of such a method in LAA.

### EVALUATION OF MULTI-CHANNEL ACCESS MECHANISMS

Simulations are performed to explore whether using multiple unlicensed channels can deliver almost continuous access to the unlicensed spectrum, and thereby reduce the access delay to being comparable with the licensed spectrum.

We consider a framework where the eNB monitors all the unlicensed channels of interest, learning the traffic in each channel and maintaining a separate backoff process for each channel. At decision points, the eNB estimates the order in which channels are likely to provide uncontested access, based on the traffic profile of each channel and the current state of each backoff process. The eNB then instructs the UE, via the licensed band, to monitor one or multiple channels, waiting for a header with their identifier. Four options were compared by simulation:

1. The eNB selects a single channel for UE to monitor for the whole run.
2. The eNB instructs UE to monitor all channels; the eNB transmits on first available channel.
3. This is the same as option 2, except the eNB uses RTS/CTS to cut short frames that collide.
4. The eNB instructs UE to monitor all channels and can independently transmit on each channel.

Option 4 uses the ETSI multi-carrier access rules [11], allowing independent transmissions on each unlicensed channel, and thus has an advantage over Options 1–3.

Each of between 1–10 channels supports the same homogeneous traffic, comprising 10 saturated Wi-Fi stations transmitting 2 ms packets and maintaining a backoff process with contention window (CW) sizes {15, 31, ..., 511}. The packets are transmitted with RTS/CTS, so that in the case of a collision, the channel is occupied for 156  $\mu$ s. The eNB uses two options, representing LAA access priority class 1 and 3 (AC1 and AC3). For AC1, the eNB obtains 2 ms transmission opportunities and uses CW sizes {3, 7}. For AC3, the eNB obtains 6 ms transmission opportunities, and uses CW sizes {15, 31, 63}. When the eNB implements RTS/CTS, LTE collisions are communicated after 156  $\mu$ s. To keep access fair, all devices use a distributed inter-frame space (DIFS) for their defer periods, and all devices are within transmission range.

To assess the four options, we consider the durations between the starts of consecutive successful LTE frames, which we refer to as consecutive-frame-start intervals (CFS intervals). For Options 1–3, the CFS intervals must exceed the LTE frame duration. For Option 4, they can be

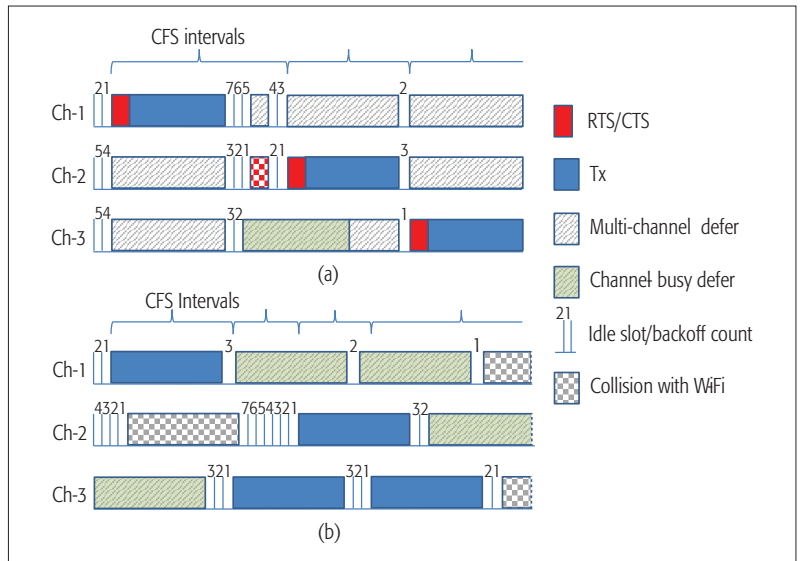


FIGURE 3. Multi-channel access frame timing: a) option 3: dependent backoff processes, one LTE transmission at a time, with RTS/CTS; b) option 4: independent backoff processes, allowing overlapping frames.

shorter, such that LTE frames in different channels overlap. Example frame timings for multi-channel access Option 3 and Option 4, and the resulting CFS intervals, are depicted in Fig. 3.

The relative performance of the four options are compared in Fig. 4. Figure 4a has settings representative of AC1, and Fig. 4b has settings representative of AC3. The 95th quantile of the CFS intervals is plotted against the number of unlicensed channels considered by the eNB.

In Fig. 4a, the square markers are for Option 1, with no channel switching, and set a baseline of 12 ms for the 95th quantile of the CFS intervals. With 2 ms LTE frames (indicated by the red solid line), 12 ms equates to a period without access between LTE frames of 10 ms. By transmitting on the first available channel after the previous LTE frame finishes (circles), the 95th quantile of the CFS intervals reduces as more channels are monitored, falling to below 8.5 ms when the UE monitors 10 channels. By additionally implementing RTS/CTS, the 95th quantile of the CFS intervals approximately halves, reducing to 6 ms with just 2 channels monitored and to 3.5 ms with 10 channels monitored. When transmitting on all channels, with independent frame starts, the 95th quantile of the CFS intervals falls below 2 ms with six channels monitored. This means that at least 95 percent of the LTE frames overlap the next LTE frame and produce continuous access. We note that although the reliability increases when progressing from Option 1 through Option 4, the required overheads also increase. This trade-off is a potential topic for future research.

To explore the sensitivity of the 95th quantile of the CFS intervals to traffic density, Options 3 and 4 are simulated with 5 and then 20 stations (STAs) per channel. With (5, 10, 20) STAs per channel, the collision probability in each channel without the eNB is (0.33, 0.42, 0.52). For AC1, under Option 4, and monitoring 10 channels, the resulting 95th quantile of the CFS intervals, with (5, 10, 20) STAs per channel, is (1.0, 1.3, 1.8) ms, or (–22, 0, +37) percent. With between

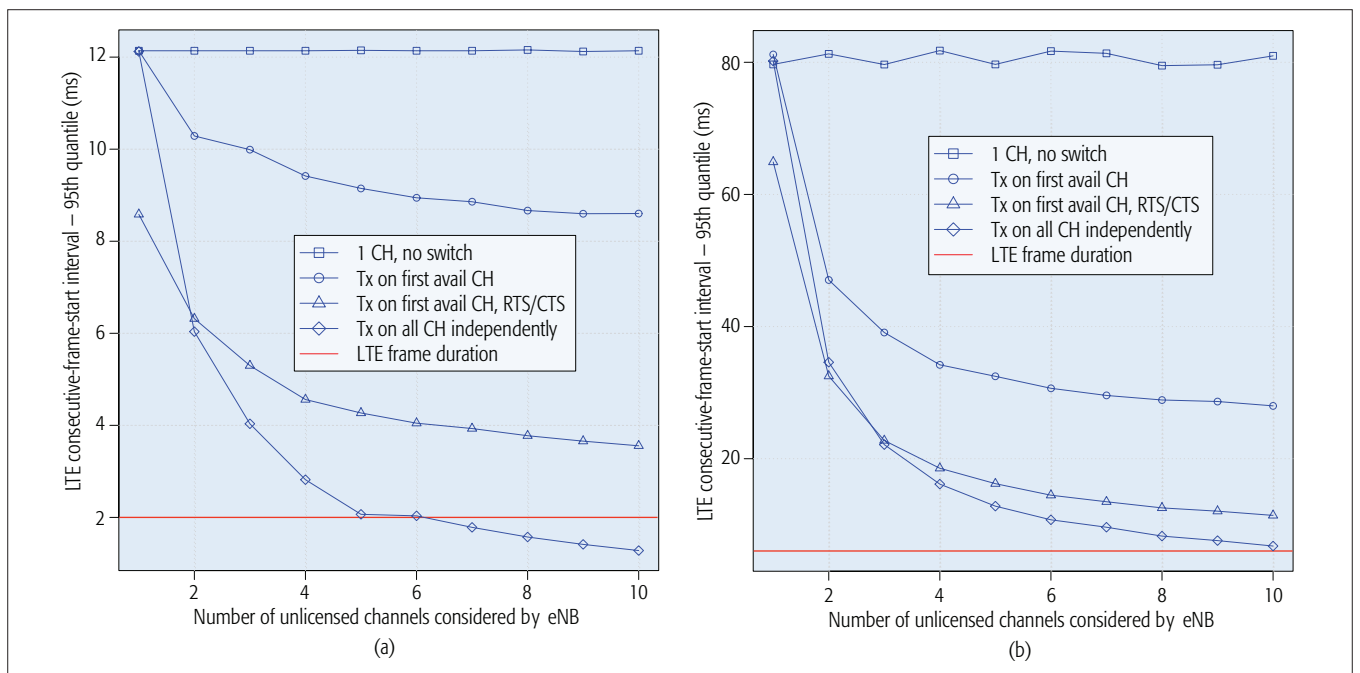


FIGURE 4. Comparison of different LAA multi-channel access strategies: a) LAA CW sizes {3, 7}, LTE frame duration = 2 ms; b) LAA CW sizes {15, 31, 63}, LTE frame duration = 6 ms.

1–10 channels monitored, the percentage change ranged between –33 and +64 percent. For AC1, under Option 3, the percentage changes were approximately half those for Option 4, ranging from –27 to +31 percent.

In Fig. 4b the eNB uses a longer 6 ms LTE frame, but also longer CWs. With no channel switching, the baseline 95th quantile of the CFS interval is 80 ms. Monitoring two channels and transmitting on the first available channel each time gives a 40 percent reduction, and with 10 channels, a 65 percent reduction. Again, by additionally implementing RTS/CTS, the 95th quantile of the CFS interval is approximately halved, achieving 11.5 ms with 10 channels monitored. When transmitting on all channels, with independent frame starts, the 95th quantile of the CFS intervals almost reduces to the 6 ms LTE frame duration.

These simulations demonstrate that the periods without access to the unlicensed band can be substantially reduced by maintaining channel access processes on multiple unlicensed channels, choosing the channels intelligently, and implementing RTS/CTS. This reduction, to almost continuous frames, sets the foundation for utilizing the unlicensed band in a URLLC solution, whether the unlicensed band delivers all the data, or takes a supporting role within an unlicensed-band assisted URLLC regime. A thorough characterization of reliability and performance gains under different traffic and channel conditions is an open research topic. Characterizing the effective capacity would provide reliability guarantees that account for input buffers and scheduling, as well as the distribution of CFS intervals.

## CONCLUSION

In this article, we have aimed to demonstrate that the unlicensed spectrum can be part of a URLLC solution, not just a capacity booster. The obvious impediment to using the contention-based unlicensed spectrum is discontinuous and uncertain

access, due to having to share the spectrum. We have explored the possibility of utilizing multiple unlicensed channels to reduce periods without access. The periods without access can be significantly reduced through signaling design and coordination of multiple unlicensed channels. Simulations showed that by maintaining backoff processes on multiple channels and transmitting on the first available channel, the 95th quantile of the duration between the starts of consecutive successful LTE frames is considerably reduced compared to remaining with one channel. Our simulation results demonstrate that almost continuous access to the unlicensed spectrum was achieved, with over 95 percent of the frames, being sourced from multiple channels, overlapping the next frame. Our analysis reveals that licensed assisted access is the key to achieving URLLC in 5G unlicensed, and the integration of licensed and unlicensed access calls for new frame formats and new scheduling designs.

## ACKNOWLEDGMENT

This work has been supported by the Huawei Innovation Research Program, subject No. HIRPO20160105.

## REFERENCES

- [1] 3GPP TR 38.913 V14.3.0, "Study on Scenarios and Requirements for Next Generation Access Technologies (Release 14)," June 2017.
- [2] H. Chen *et al.*, "Ultra-Reliable Low Latency Cellular Networks: Use Cases, Challenges and Approaches," arXiv preprint arXiv:1709.00560, 2017.
- [3] I. Parvez *et al.*, "A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions," arXiv preprint arXiv:1708.02562, 2017.
- [4] P. Schulz *et al.*, "Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture," *IEEE Commun. Mag.*, vol. 55, no. 2, Feb. 2017, pp. 70–78.
- [5] U. Oruthota, F. Ahmed, and O. Tirkkonen, "Ultra-Reliable Link Adaptation for Downlink MISO Transmission in 5G Cellular Networks," *Information*, vol. 7, no. 1, 2016, pp. 1–18.

- [6] X. Huang, J. A. Zhang, and Y. J. Guo, "Out-of-Band Emission Reduction and a Unified Framework for Precoded OFDM," *IEEE Commun. Mag.*, vol. 53, no. 6, June 2015, pp. 151–59.
- [7] Y. Ma et al., "Reliable and Efficient Sub-Nyquist Wideband Spectrum Sensing in Cooperative Cognitive Radio Networks," *IEEE JSAC*, vol. 34, no. 10, 2016, pp. 2750–62.
- [8] F. Capozz et al., "Downlink Packet Scheduling in LTE Cellular Networks: Key Design Issues and a Survey," *IEEE Commun. Surveys & Tutorials*, vol. 15, no. 2, 2013, pp. 678–700.
- [9] G. Piro et al., "Two-Level Downlink Scheduling for Real-Time Multimedia Services in LTE Networks," *IEEE Trans. Multimedia*, vol. 13, no. 5, Oct. 2011, pp. 1052–65.
- [10] 3GPP TS 36.213 v14.1.0, Tech. Spec. Group Radio Access Network, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures (Release 14)," Dec. 2016.
- [11] ETSI Draft EN 301 893 v2.0.7, Harmonized European Standard, 5 GHz RLAN, "Harmonised Standard Covering the Essential Requirements of Article 3.2 of Directive 2014/53/EU," Nov. 2016.
- [12] Z. Khan et al., "Distributed and Coordinated Spectrum Access Methods for Heterogeneous Channel Bonding," *IEEE Trans. Cogn. Commun. Net.*, 2017.
- [13] J. Qadir et al., "Resource Pooling," *Wireless Networks: Solutions for the Developing World*, ACM SIGCOMM Comp. Commun. Rev., vol. 46, no. 4, 2016, pp. 30–35.
- [14] T. Shu and M. Krunz, "Throughput-Efficient Sequential Channel Sensing and Probing in Cognitive Radio Networks Under Sensing Errors," *Proc. ACM MobiCom*, Sept. 2009, pp. 37–48.
- [15] N. Morozs, T. Clarke, and D. Grace, "Distributed Heuristically Accelerated Q-Learning for Robust Cognitive Spectrum Management in LTE Cellular Systems," *IEEE Trans. Mobile Comp.*, vol. 15, no. 4, Apr. 2016, pp. 817–25.

## BIOGRAPHIES

GORDON J. SUTTON received a B.Sc. (mathematics) in 1993, a B.E. (systems engineering) in 1995, and a Ph.D. (control theory) in 1999, all from the Australian National University. He has been with the University of Technology Sydney (UTS) since 2015, modeling Wi-Fi/LTE coexistence protocols. He previously worked in optic fiber connector design (ADC Australia), time series analysis (Australian Bureau of Statistics), statistical modeling (CSIRO), and Bayesian statistics and chemometrics (University of New South Wales).

JIE ZENG [M'09, SM'16] received his B.S. and M.S. degrees in electronic engineering from Tsinghua University in 2006 and 2009, respectively. His research interests include novel network architecture, ultra-dense networks, and novel multiple access. He has published 3 books and over 100 conference and journal papers. He holds more than 20 Chinese and 5 international patents. He obtained the Science and Technology award of Beijing in 2015, and the Best Cooperation Award of Samsung Electronics in 2016.

REN PING LIU [M'09, SM'14] is a professor and head of the Network and Cybersecurity discipline at UTS. Prior to that he was a principal scientist and research leader at Commonwealth Scientific and Industrial Research Organisation (CSIRO), where he led wireless networking research activities. He specializes in

protocol design and modeling, and has delivered networking solutions to a number of government agencies and industry customers. His research interests include 5G, VANET, IoT, cybersecurity, and blockchain.

WEI NI is a team and project leader at CSIRO, Australia. Prior to this, he was a deputy project manager at Alcatel/Lucent Bell Labs (2005–2008), a postdoctoral research fellow at Shanghai Jiaotong University (2005–2008), and a senior researcher at Devices R&D, Nokia (2008–2009). He received B.E. and Ph.D. degrees in electronic engineering from Fudan University, China, in 2000 and 2005, respectively. His research interests include optimization, game theory, and graph theory, as well as their applications to network and security.

DIEP N. NGUYEN is a faculty member at UTS. He received M.E. and Ph.D. degrees in electrical and computer engineering from the University of California, San Diego and the University of Arizona, respectively. Before joining UTS, he was a DECRA research fellow at Macquarie University, Sydney, Australia, a member of technical staff at Broadcom (California), ARCON Corporation (Boston, Massachusetts), consulting the Federal Administration of Aviation on turning detection of UAVs and the Air Force Research Lab on anti-jamming.

BEE SHANGA ABEWARDANA JAYAWICKRAMA received his B.E. degree in telecommunications engineering (Hons. I) and Ph.D. degree in electronic engineering from Macquarie University in 2011 and 2015, respectively. He is currently a lecturer with the School of Electrical and Data Engineering, UTS. His research interests include resource allocation, cognitive radio, and signal processing.

XIAOJING HUANG is currently a professor of information and communications technology with the School of Electrical and Data Engineering, UTS, Australia. Before joining UTS, he was a principal research scientist at CSIRO, and the project leader of CSIRO microwave and mmWave backhaul projects since 2009, an associate professor at the University of Wollongong, Australia, since 2004, and a principal research engineer at Motorola Australian Research Centre since 1998.

MEHRAN ABOLHASAN [SM] completed his B.E. in computer engineering and Ph.D. in telecommunications on 1999 and 2003, respectively, at the University of Wollongong. He is currently an associate professor and deputy head of the School of Electrical and Data Engineering at UTS. He has authored over 120 international publications and has won over AU\$3 million dollars in research funding. His current research interests are in software defined networking, IoT, wireless mesh, wireless body area networks, cooperative networks, 5G networks and beyond, and sensor networks.

ZHANG ZHANG received his B.Eng. and Ph.D. degrees from Beijing University of Posts and Telecommunications, China, in 2007 and 2012, respectively. From May 2009 to June 2012, he also served as a research assistant for the Wireless and Mobile Communications Technology R&D Center, Tsinghua University, Beijing, China. He is currently with the Department of Radio Access Network (RAN) Research, Huawei, Shanghai, China. His current research interests include information theory and radio access technologies.