A Vision of 6G URLLC: Physical-layer Technologies and Enablers

Azadeh Pourkabirian [a], Mohammad Sadegh Kordafshari [a], Anish Jindal [b] and Mohammad Hossein Anisi [c*]

[a] Department of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran, CO 34199-15195
[b] Department of Computer Science, Durham University, Durham, DH1 3LE, United Kingdom
[c] School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, United Kingdom
[*]Corresponding Author: E-mail: m.anisi@essex.ac.uk

***Abstract*—** The anticipated advent of 6G communication holds the promise of enabling mission-critical applications such as traffic detection, forest fire recognition, emergency search and rescue, and widespread communication broadcasting. To fulfill these services, 6G demands a robust communication framework with minimal latency. While the current fifth-generation (5G) ultra-reliable low-latency communication (URLLC), as defined by the 3rd Generation Partnership Project (3GPP), provides the requisite reliability (99.99%) and latency (<1ms) to support existing applications, many foundational aspects of URLLC remain uncertain. Therefore, upcoming networks necessitate the integration of the next-generation URLLC, known as xURLLC, to attain the pinnacle of reliability and latency. This paper presents an overview of both current 5G URLLC technologies and those anticipated for 6G. It initiates by succinctly delineating the essential URLLC requirements and scrutinizing the prevailing limitations in URLLC design. Subsequently, it spotlights xURLLC and the groundbreaking physical-layer techniques poised to define 6G URLLC. Additionally, it assesses the opportunities and challenges presented by each key xURLLC driver, engaging in a comprehensive discussion of the efficacy of each proposed solution. Finally, we put forth future research directions for the next generation of URLLC.

***Index Terms*—** 6G, reliability, latency, xURLLC, physical-layer enabler.

I. INTRODUCTION

While 5G URLLC laid the groundwork for low-latency, high-reliability communication in the current wireless landscape, 6G URLLC aims to take these capabilities to new heights, enabling transformative applications that were once only imaginable. The evolution from 5G to 6G URLLC represents a leap in connectivity possibilities and is poised to reshape industries and technologies in the coming years. Emerging new 6G services such as cargo transport, telemedicine systems, traffic detection, and extended reality pose challenges to wireless communications. On the one hand, with the deployment of massive IoT devices, the network traffic volume has significantly grown. On the other hand, these mission-critical applications need an exceptionally high degree of reliability and latency. The future communication systems are projected to support three key communications comprising ultra-reliable low-latency communication (URLLC) [1], enhanced mobile broadband (eMBB), and massive machine-type communication (mMTC) to ease these challenges. Mobile users can access high data rates (Gbits/s) with eMBB whereas URLLC is intended for extreme reliability (99.99%) and excessive low latency (1 ms) for real-time on-demand applications. In contrast, mMTC focuses on the number of connected devices in IoT (i.e., to one million connections / km^2). Fig.1 displays three main services and their features in 6G communications.

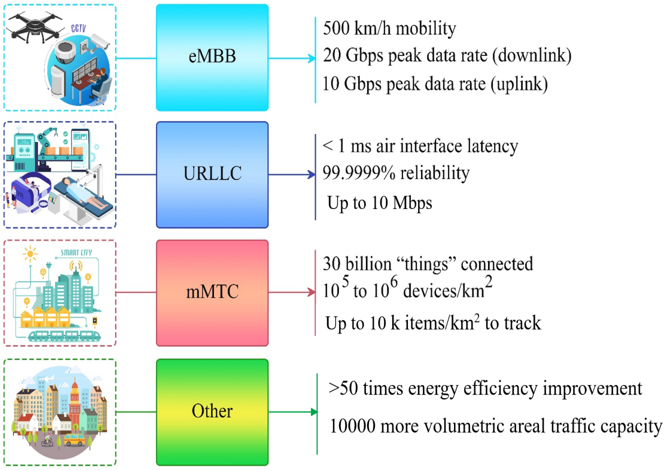


Fig. 1. Key 6G services and requirements

Compared to eMBB and mMTC, URLLC is more difficult to implement. The most challenging part is that two highly strict QoS necessities in terms of extremely low latency and very high reliability are incompatible with one another. For instance, remote motion control applications such as healthcare services require 1-10^{-5} reliability under 10ms, while search and rescue ideally should respond within 0.5 ms which is the fastest it can do. On the other hand, it is particularly difficult to deliver

URLLC traffic with an extremely low setup latency (i.e., the interval between when the devices generate the first URLLC data until the data is transmitted) since this form of traffic cannot be queued until the next slot such as alarm systems.

Although 5G systems support URLLC, new high-stake applications of 6G request more stringent scalability, reliability, and latency demands. Therefore, next-generation communication systems require next-generation URLLC known as extreme URLLC (i.e.., xURLLC) supporting sub-millisecond latency, reliability of ≥ 99.99999%, and data transmission rate of ≥ 100 Gbps for these applications [2]. For example, data visualization demands ultra-low latency and high bandwidth to enable real-time rendering and interaction with complex datasets. 6G networks will need to provide xURLLC for immersive experiences support, such as augmented reality (AR) and virtual reality (VR), for data visualization in industries like gaming, architecture, and education. Remote surgery is another transformative use case, where surgeons can perform operations from a distance with the assistance of robotic systems. Ultra-reliable and low-latency connectivity is vital to ensure that every surgical movement is transmitted and executed without delay, guaranteeing patient safety. AI and machine learning, on the other hand, require extensive data exchange between devices and edge servers for rapid model training and decision-making. The low-latency communication of xURLLC will empower AI-powered applications across sectors, from autonomous vehicles to smart cities. Furthermore, edge computing is becoming increasingly crucial in 6G URLLC scenarios. By bringing computational resources closer to the data source, edge computing minimizes latency, supports real-time analytics, and enhances security. Industries like industrial automation and IoT will heavily depend on edge computing capabilities to achieve efficient and reliable operations. To achieve this goal, potential physical-layer techniques are presently under consideration for xURLLC support in 6G including orbital angular momentum (OAM), reconfigurable intelligent surface (RIS), terahertz (THz) communication, and holographic radio (HR).

This study provides a comprehensive outline of the enhanced physical-layer design for xURLLC. We first describe the main definition of URLLC requirements including reliability and latency and then discuss the present 5G URLLC enablers, followed by their limitations and challenges. After that, we discuss state-of-the-art physical-layer technologies for xURLLC. Besides, we identify four fundamental trade-offs for xURLLC in 6G.

The remainder of this paper is organized as follows. Section II describes the URLLC necessities. The existing 5G URLLC techniques are presented in Section III. Section IV highlights beyond 5G/6G URLLC and investigates the risks and challenges of current URLLC enablers. The next generation of URLLC known xURLLC is introduced in IV. In Section V, basic trade-offs in xURLLC are given. Finally, Section VI concludes the whole paper.

II. BASICS OF URLLC

URLLC is a critical aspect of next-generation wireless networks, such as 5G and beyond. It represents a paradigm shift in wireless communication, focusing on delivering exceptionally reliable and low-latency connections that are essential for a wide range of applications, from autonomous vehicles and telemedicine to industrial automation and IoT. URLLC ensures that data is transmitted with minimal delay and the utmost reliability, making it ideal for mission-critical tasks. Initially, we outline the core definitions of URLLC requirements, encompassing key aspects like reliability and latency.

A. Reliability

3GPP defines reliability [3] as the ability to transmit a specific amount of traffic in a given period with a high degree of success. Besides, there are other types of reliability in wireless communication systems as follows:

- Data reliability: it states the probability of data dropping, transmission error, and queuing delay violation.
- Control channel reliability: it is described as the successful decoding, the scheduling grant, or other metadata probability.
- Availability: it is defined as the probability of a given service is available. For example, 99.999% availability means one out of every ten thousand users does not receive the adequate resource.

B. Latency

3GPP targets extensive latency (< ms) for 6G communication systems and introduces the following types of latency [4] which may happen in forthcoming networks:

- User-plane latency (U-plane): Generally, this is the time when user data is delivered to the target destination successfully, either uplink or downlink, under heavy traffic conditions. URLLCs demand U-plane latency of less than 1 ms for critical-mission applications.
- Control-plane latency (C-plane): A C-plane latency which refers to the call-setup latency, describes the time it takes to transition from a battery-efficient mode to an active data transfer state.

End-to-end (E2E) latency: A latency measurement of E2E includes the overhead transmission delay (i.e., retransmissions), processing delay as well as queuing latency.

To meet URLLC latency requirement, physical-layer latency D_{URLLC} should be in the order of hundreds of microseconds [5] as follows:

$$D_{URLLC} = t_{tr} + t_{pp} + t_{pr} + t_{rtr} + t_{sig} \tag{1}$$

Where t_{tr} is the transmission time of the data, t_{pp} states the time it takes for a signal to propagate from the transmitter to the receiver, t_{pr} is the processing time for encoding, decoding, and channel estimation, t_{rtr} denotes re-transmission time, and t_{sig} typically refers to the signaling time like transmission requests, scheduling grants, channel training, and queueing latency.

III. EXISTING TECHNIQUES FOR URLLC

Here, we present an overview of some key URLLC enablers for high reliability and low latency guarantees (Table 1).

Table 1. Comparative analysis of URLLC techniques

URLLC/xURLLC Techniques	Key URLLC Features	Applications	Limitations	Advantages
PDCCH	Low Latency, Reliable Signaling	Autonomous Vehicles, Public Safety	Limited Spectrum Resources	Efficient Control Signaling
CQI Report	Adaptive Modulation, Quality Feedback	Telemedicine, Industrial Automation	Feedback Overhead	Adaptive Data Rate
K-Repetition	Redundancy for Reliability	Public Safety, Industrial Automation	Increased Overhead	Robust Error Recovery
HARQ	Error Recovery with Low Latency	Public Safety, Autonomous Vehicles	Latency in Retransmissions	Low Error Rate, Retransmission
SR	Uplink Resource Allocation	Industrial Automation, Telemedicine	Control Overhead	Request-Based Efficiency
URLLC/eMBB Multiplexing	Multi-Antenna Reliability	Industrial Automation, Autonomous Vehicles	Hardware Complexity	Improved Fading Resilience
Short TTI	Reduced Transmission Time	Autonomous Vehicles, Telemedicine	Control Overhead	Low Latency Data Delivery
Grant-Free Access	Reduced Scheduling Overhead	IoT, mMTC, Drone Delivery	Contentious Resource Access	Efficient IoT Access
NOMA	Spectral Efficiency, Low Latency	Industrial Automation, Telemedicine	User Interference	Efficient Resource Sharing
Network Densification	Increased Network Coverage and Capacity	Smart Grids, Emergency Communication	Deployment Complexity	Enhanced Network Performance
OAM	Spectrally Efficient Modulation	Industrial Automation, Agricultural IoT	Signal-to-Noise Ratio (SNR) Sensitivity	Enhanced Spectrum Efficiency
RIS	Signal Reflection and Control	Environmental Monitoring, THz Communication	Deployment Complexity	Dynamic Signal Enhancement
THz	Ultra-High Data Rate Communication	Space Exploration, Industrial Automation	Limited Propagation Range	Unprecedented Data Rates
HR	Enhanced Coverage and Reliability	Mining Automation, Environmental Conservation	Coordination Complexity	Coverage Extension, Reliability

A. Reliability enablers

The current enabling reliability techniques of wireless communication systems comprise hybrid automatic repeat request (HARQ) retransmission, *K*-repetition, the physical downlink control channel (PDCCH), channel quality indicator (CQI) report, finite blocklength (FBL), scheduling request (SR), multi-connectivity (MC), spatial diversity, and slicing.

HARQ retransmission [6]: It is based on conventional automatic repeat request (ARQ) in which if transmission arrives with an error, the receiver asks for retransmission by sending a negative acknowledgment (i.e., NACK). However, in contrast to ARQ, received signals with an error are stored instead of discarded, and will be combined with the next retransmission to decode the target signal. Although retransmission is only performed once the preceding transmission has failed, HARQ adds extra latency since the receiver must wait for the time duration for the acknowledgment (ACK) before retransmitting. This problem can be solved by the *K*-repetition method in which the transmitter gently uses repetitions to increase reliability.

***K*-repetition:** Another standard scheme is the *K*-repetition which sends a predetermined number of consecutive replicas of a similar packet without first waiting for the ACK. Since the transmitter obtains the required resources for *K* transmissions, it can repeatedly resend a similar packet. Although the repetition will be stopped to reduce latency when the user receives the ACK from the transmitter, the user has to monitor the feedback, which is more computationally demanding for this scheme.

PDCCH: It conveys downlink control information (DCI) which comprises power control, ACK/NACK information, and scheduling [4]. For instance, a serving base station (BS) adjusts uplink/downlink resource allocation by sending resource grant (RG) information over the PDCCH. Existing Long-Term Evolution (LTE) provides four various aggregation levels for PDCCH, which support diverse reliability levels. The highest aggregation level is used for URLLC traffic to support reliability constraints. Higher aggregation levels with lower coding rates or a lower modulation order are used to decrease BER in control information transmission.

CQI report: A transmitter requires to acquire the channel quality to adapt link conditions for downlink transmissions and grant radio resources for transmission. A CQI report indicates whether the received data was successful or unsuccessful. The receiver transmits either an ACK or a NACK signal in response to the decoded RG. Retransmission is performed by the transmitter if no ACK signal is received. UEs receive a new RG from the transmitter indicating that resources for data retransmission are being monitored. Data is retransmitted until either an ACK signal is received or it reaches its maximum number of retransmissions. The main drawback of this technique is that inaccurately decoded CQI report results in using a high transmission rate. The incorrect interpretation of ACK/NACK signals is a common kind of error under fading channels. For instance, an ACK may be wrongly decoded as a NACK leading to resource waste by forcing unnecessary data retransmission. In contrast, required retransmission is missed when a NACK is incorrectly decoded as an ACK.

FBL: One of the applicable solutions to meet the reliability requirements is FBL since a finite codeword length spans one realization of the fading channel. According to FBL, the maximum achievable rate is

$$R(m, \varepsilon) = {}^{b}\!/_{m} \tag{2}$$

where b denotes the number of transmitting bits across m channel uses, and ε is the error probability. We can determine triplets (b, m, ε) which have a (b, m, ε)-code for a given signal-to-noise ratio (SNR), and which triplets (b, m, ε) have no code based on FBL. In the short blocklength case, the received rate may be below the Shannon rate, thus, a sufficiently large channel blocklength can make the decoding error probability arbitrarily small.

SR: For reliable data transmission, a user needs to access the physical uplink control channel (PUCCH). Some periodic orthogonal resources should be scheduled and allocated to the user on PUCCH for uplink transmission. The user sends an SR to the serving BS through predefined radio resources. This induces a random latency before the uplink transmission. The user cannot obtain the RG for data transmission when the scheduling request is not sensed. Consequently, the user should resend the SR resulting in extra latency. A promising solution to reduce the latency is to frequently assign PUCCH radio resources to the user for every transmission time interval (TTI). However, this method wastes a high proportion of radio resources when the user produces erratic data traffic.

Multi-connectivity: MC technology enables users to simultaneously have radio resources from multiple network nodes aiming to enhance reliability and increase user throughput. This technique requires a mesh connectivity deployment where nodes can directly communicate with each other without core network management while some parts of the signaling procedure such as session establishment need to be handled by radio access network (RAN).

Spatial diversity: it is an emerging paradigm aiming to increase data reliability using sending several copies of the same signal across multiple spatial paths.

Slicing: network slicing is a standard technique in 5G that defines several virtual network slices on top of the shared physical infrastructure. The network traffic is assigned to available slices and there are always backup slices in case of overloading or failure situations. In these cases, slicing enables alternative slices and the traffic is directed to other slices to ensure transmission reliability.

B. Latency enablers

Existing wireless networks such as LTE and 5G are all included delay reduction techniques for URLLC requirements counting: short TTI, grant free (GF) access, non-orthogonal multiple access (NOMA), network densification, network coding, slicing, edge caching, and computing, URLLC/eMBB multiplexing.

Short TTI: This standard technique employs fewer OFDM symbols per TTI as well as shortening OFDM symbols through wider subcarrier spacing to reduce latency (e.g., from 1 ms in LTE to 0.125 ms in 5G new radio) [7]. In contrast, a reduction in OFDM symbol duration is further accompanied by an increase in subcarrier spacing, which results in fewer resources available in the frequency domain and consequently more congestion. Conversely, a shorter TTI increases control overhead which reduces the accessibility of resources for additional URLLC data transmissions. Grant-free transmission in the uplink can help to solve this problem.

GF access [8]: It is a promising dynamic uplink scheduling that handles burst URLLC traffic with contention-based access as opposed to persistent scheduling for regular traffic. Yet, the possibility of collisions may endanger the GF potential if a shared channel is provided to numerous users. Moreover, queuing latency can also occur for GF scheduling since a user can only send one packet per TTI.

NOMA: This is a multiple access technique designed for 5G wireless networks and aims to increase the spectral efficiency of mobile communication networks and meet challenging requirements. NOMA technique leverages transmit power or code domain multiplexing for the uplink transmission so that users can non-orthogonally access the channel. It uses successive interference cancellation (SIC), or some advanced receiver techniques such as message passing. Thus, on the same channel, multiple users with different types of traffic requests can transmit simultaneously, improving spectrum efficiency and reducing traffic congestion in massive connectivity cases. By doing so, it can potentially reduce the number of retransmissions and access latency. Nevertheless, some problems like processing latency for multiplexing, user ordering, imperfect channel state information (CSI), and other dynamics which affect latency are not well addressed.

Network densification: It focuses on adding more cellular sites such as cell towers into a network to increase the available radio frequency resources. As well as this, there is also an effort on densification in terms of reusing and re-assigning spectrum. However, it is not always easy to get permission from local governments and landowners for new transmission points.

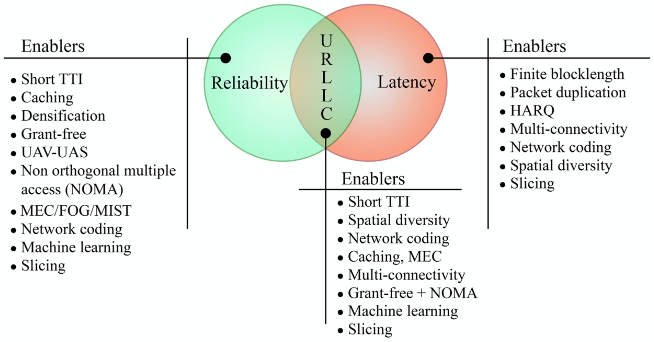
Network coding: When data passes through nodes in a network, network coding performs operations on it that leads to latency reduction.

Slicing, edge caching, and computing: There has been a significant reduction in latency when caching [9] and computing resources are pushed to the network edge. In addition to resource-intensive applications like virtual reality, other mission-critical applications will also thrive. A key part of this process will be the use of network slicing to allocate dedicated radio frequency, cashing, and processing resources.

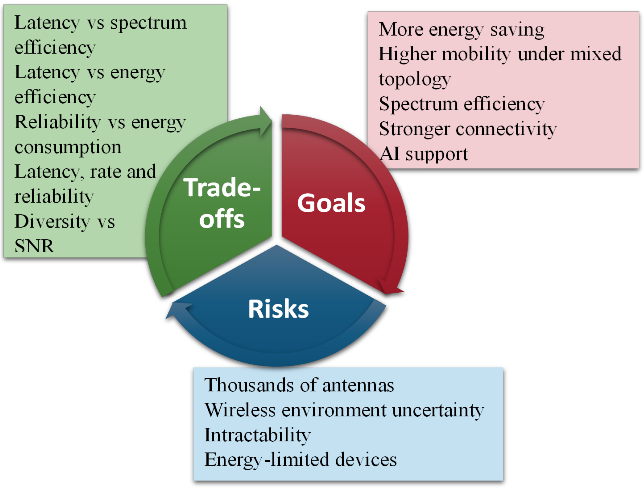
URLLC/eMBB multiplexing: Allocating additional frequency resources to a UL transmission as opposed to increasing power on narrow-band resources, can exceptionally enhance transmission reliability for URLLC traffic. This indicates that to reach low latency and high reliability for URLLC transmission, wide-band resources are required. Although a resource division [10] among URLLC and MBB transmissions might be better than a latency standpoint, it is not sufficiently efficient in resource utilization. On the other hand, when low-latency traffic arrives during the flow, a sophisticated scheduling algorithm is also required to block other scheduled traffic. Thus, dynamic URLLC/eMBB multiplexing can be a significant solution in the URLLC settings. The key enablers for reliability and latency guarantees are depicted in Fig.2a.

IV. BEYOND 5G/6G URLLC

Although 5G is projected to fulfill microseconds latency communications, there are still some challenges for URLLC. Fig. 2b. illustrates main goals, risks and trade-offs in 6G URLLC.



a. Reliability and latency enablers



b. URLLC goals, risks, and trade-offs

Fig. 2. URLLC performance enhancement

A. URLLC risks and challenges Model

- **Connected mobility:** 6G applications such as Unmanned Aerial Vehicle (UAV) control desire to take into account connected mobility as well as reliability and latency in communications. The most challenging part is that with a vehicle moving fast, there is a large Doppler frequency shift and strong inter-symbol interference, so the receiver has to adapt the carrier frequency to compensate for the Doppler shift. In addition, there will be service interruptions for URLLCs with high mobility due to frequent handovers that lead to increased latency. So, mobility in URLLC continues an open problem of 6G.
- **High-throughput under high density:** 6G is planned to empower a massive connection with a density of 107 machines/km2, 10 times of that 5G [11]. The bandwidth requirement exponentially increases with the number of devices in orthogonal multiple access technologies to support these throughput requirements. While it is certainly possible to increase bandwidth in the THz bands, since THz frequency is highly sensitive to shadow fading, low frequencies in the terahertz band suffer greater free space fading, which significantly reduces communication reliability and will have substantial impact on the network coverage. As a result, future 6G networks will require an analysis of the fundamental tradeoffs between connection density, throughput, reliability, and network coverage to be able to enable 6G use cases.
- **Stronger connectivity and security:** Due to massive connectivity, communication security is a critical issue for 6G systems. Although the security of massive data processing can be guaranteed through decentralized processing where data is processed locally. However, unintended receivers can receive information sent by a legitimate transmitter. Therefore, launching secure communications which support multi-connectivity is necessary for future 6G networks. A promising paradigm for safeguarding wireless systems is physical layer security which has grown in popularity in the past decade.
- **Integration of human intelligence with Artificial intelligence:** A new trend in developing communication networks is to integrate human intelligence with artificial intelligence at different parts of the communication system. It aims to develop computers that are capable of interacting with humans and completing tasks that they would normally have to do. Thus, AI-enabled architecture is required in 6G URLLC to exploit knowledge discovery and smart resource management.

V. NEXT GENERATION ULTRA-RELIABLE AND LOW LATENCY (xURLLC)

The next generation URLLC (xURLLC) provides novel technologies and methodologies to guarantee extra requirements on Key Performance Indicators (KPIs) of 6G applications in the context of ultra-high throughput, spectrum efficiency, network availability, and energy efficiency.

Here, we represent some promising physical-layer techniques for xURLLC support in 6G including orbital angular momentum, reconfigurable intelligent surface, terahertz communication, and holographic radio.

OAM: It can be regarded as one of the physical-layer enabling technologies for xURLLC because it offers new evolution multiplexing empowering high-speed transmissions. OAM [12] is a phase front in the propagation direction characterized by a helical shape for electromagnetic waves. This characteristic is used for multiplexing signals onto electromagnetic waves which are orthogonal to each other with multiple phase rotations and generate several orthogonal channels. Thus, it unleashes more capacity in the harshly congested frequency spectrum of communication networks. The propagating beams in OAM can be multiplexed and demultiplexed without consuming traditional resources like time or frequency. OAM has been used in two areas: optical transmission and radio frequency. In RF systems, it empowers to obtain a total bandwidth of Gbps on the millimeter wave channels with a mixture of polarization-mode multiplexing and OAM. OAM multiplexing has the advantage of enabling higher-rate transmission, however, further exploitation remains to be addressed in OAM such as OAM channel coding where there may be OAM nulls at various locations.

RIS: This technology has attracted much attention due to its potential to enhance transmission rate through passive reflections [13]. RIS is a passive reflection panel controlled by intelligent controllers that reflects received signals to their intended destinations. RIS can benefit from intelligent controllers to adjust the propagation of signals according to the environment so that received signals are enhanced and interference is minimized. By adjusting the RIS in terms of phase-shifting values and the number of elements, it is possible to minimize BER and increase the number of URLLC admitted packets. Thus, RIS can obtain extremely reliable connectivity and support the stringent URLLC constraints. To fulfill the URLLC latency constraints, the phase-shifting configuration can be proactively done at the beginning of the time slot. Each time slot is divided into several mini-time slots at which the URLLC traffic is immediately served using puncturing methods to guarantee the latency demand of the traffic. To satisfy the rate demand of URLLC, the number of serving URLLC packets should be maximized at mini-time slots. Thus, optimal frequency resource allocation plays an indispensable role in URLLC requirements guarantee besides the phase shift configuration of RIS. On the other hand, in comparison with the mini-time slot duration, adjusting an optimal configuration of RIS takes a long period, breaking URLLC's latency requirements. As a solution, several RIS configurations can be designed proactively and sent to the RIS controller at the start of every time slot before URLLC traffic arrives. At each mini-time slot, RIS controllers are instructed to switch between these configurations if needed.

TeraHertz (0.1-10 THz) transmission: Along with mmW frequencies, the development of the THz frequency band has been steered to provide an ultra-high data rate for 6G wireless users. The THz band lies between radio and optical frequencies [14] within the electromagnetic spectrum and enables photonics and electronics-based technologies to guarantee the wide-spread bandwidth resulting in extensive capacity and ultra-high data rate (Tbps). Since THz frequencies permit small antenna sizes, terminal devices can be equipped with a large array of antenna elements. Consequently, antenna directivity gain and diversity gain can be obtained over MIMO techniques. THz can be a robust candidate to establish xURLLC communications for various applications beyond 5G and 6G. However, the Doppler effect is very significant in the THz frequency band so it is not a suitable choice in mobile scenarios.

HR: it is one of the emerging physical-layer technologies for forthcoming 6G communications. HR uses EM waves and exploits antenna arrays composed of many tiny and low-cost antennas under the wavelength for continuous EM wave reception and wavefront phase measurement. It is anticipated to produce an extremely high data rate for improved mobile broadband and handle a huge number of connections through MIMO and spatial diversity. One of the promising candidates for enabling holographic radio is reconfigurable holographic surfaces (RHS) [15] consisting of many metamaterial radiation elements. RHS elements configure the amplitude of the electromagnetic waves to create object beams, which produce extreme directive gain. Despite the advantages of HR, it is still at the scientific examination stage so it currently has very few solutions.

Fig. 3 shows the URLLC enablers for 5G/6G wireless communications.

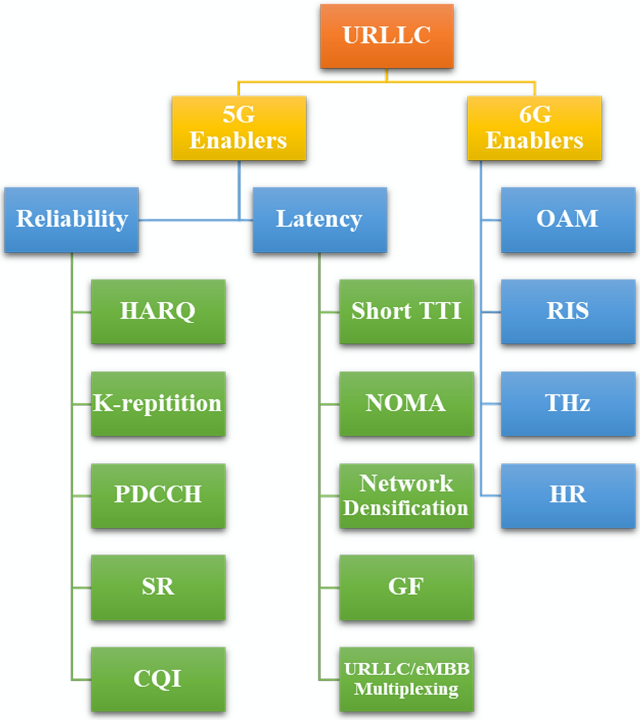


Fig. 3. Physical-layer URLLC enablers

In summary, each of the advanced techniques OAM, RIS, THz communication, and HR plays a unique and pivotal role for 6G critical mission applications. OAM excels in applications demanding spectral efficiency, such as industrial automation and agricultural IoT. RIS technology shines in scenarios like environmental monitoring, where it optimizes signal propagation. THz communication finds its calling in space exploration and industrial automation, offering unmatched data rates. Finally, HR is indispensable for applications like mining automation and environmental conservation, where it extends network coverage and enhances reliability.

VI. FUNDAMENTAL TRADE-OFFS IN XURLLC

6G is planned to feature various IIoT use cases under different quality-of-service (QoS) demands as well as xURLLC requirements (Fig. 4). Nevertheless, services that fulfill some QoS requirements may often suffer from a degradation of xURLLC demands. Thus, a trade-off between xURLLC constraint and QoS demands is of exact emphasis for effective service provision in 6G.

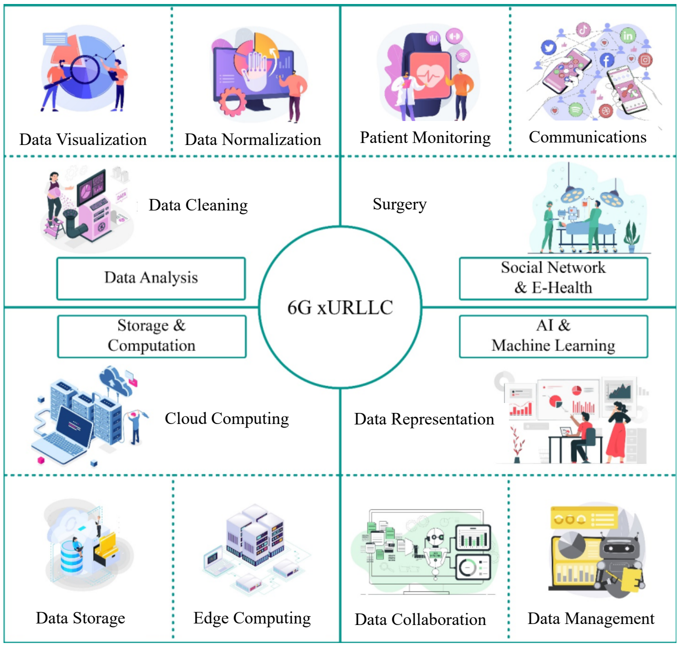


Fig. 4. xURLLC use cases in 6G

Some fundamental trade-offs in xURLLC are as follows:

A. Latency versus spectrum efficiency

Multi-access and broadcast systems lack a trade-off between latency and spectrum efficiency under bursty traffic and fading channels. Because of shorter TTI, meeting latency constraints destroys spectrum efficiency.

B. Latency versus energy efficiency

It is crucial to understand how energy consumption and latency can be traded-off. Wireless devices go to sleep mode when they are not sending or receiving to consume less energy. A device should periodically wake up to see whether packets have been sent to it from the network. Latency and energy consumption are both determined by the frequency at which the device checks for incoming packets. A higher frequency results in lower latency, but higher energy consumption.

C. Reliability versus energy consumption

Reliability versus energy consumption is an essential trade-off that should be characterized. Multiple low-power transmissions are more reliable than one high-power transmission. Nevertheless, this relies on whether independent or correlated fading is included, as well as the diversity order that can be achieved.

D. Latency, rate, and reliability

Due to retransmissions, higher reliability generally leads to higher latency, however, both can sometimes be optimized. When it comes to data rates, larger rates indeed come with less latency and vice versa.

E. Diversity versus SNR

Diversity involves sending data over independent paths using several transmit/receive antennas. The transmit power is divided among several antennas to achieve independent fading paths by transmitter diversity. Without increasing transmitter power or bandwidth, independent paths cannot be realized. These independent paths are combined in some way to minimize the fading of the resultant signal. Combining the diversity signals at the receiver causes higher SNR at the receiver which makes it less reliable than a single receive antenna. Although a higher SNR ratio increases signal quality; however, extra mathematical signal processing, as well as RF resource, would be required to have diversity. Thus, it is essential to address some research questions such as how does reliability change with more links or nodes? How do diversity order and network nodes affect the requirements for SNR? How much SNR is required to account for fast-fading channels, time-varying events, etc? Fig. 5 shows basic xURLLC balances in 6G applications.

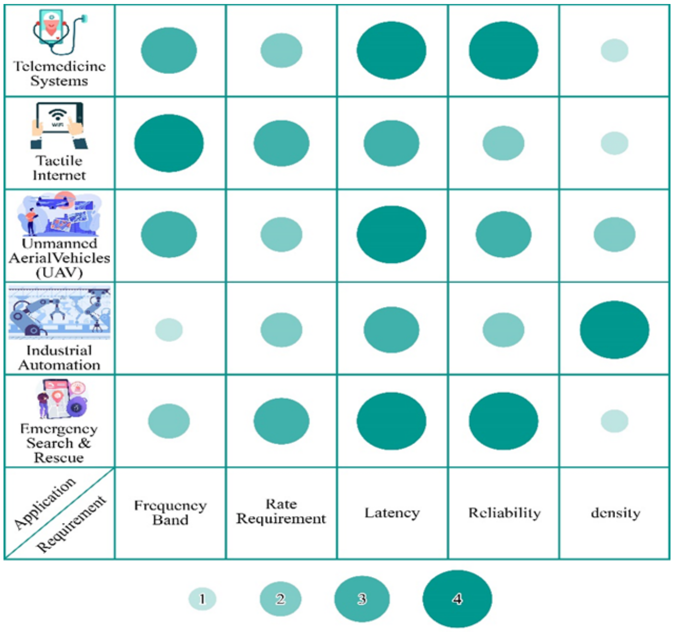


Fig. 5. Basic trade-offs in 6G xURLLC use cases

VII. CONCLUSION

6G xURLLC aims to redefine reliability and latency benchmarks, pushing beyond conventional expectations. Despite the wealth of practical knowledge gained from extensive studies on 5G URLLC, understanding the core principles and navigating the trade-offs of 6G xURLLC remains a complex challenge. This paper provided a comprehensive review of existing 5G URLLC technologies, shedding light on their associated risks and challenges in the context of future 6G communication systems. We introduced cutting-edge physical-layer xURLLC enablers, offering insights into their potential applications in the next-generation 6G landscape. Furthermore, we conducted a rigorous evaluation of each technique and explored critical trade-offs between xURLLC constraints and other vital QoS requirements in upcoming communication systems. As a future study, we plan to delve into the development of energy-efficient techniques and protocols specifically designed for the demands of xURLLC scenarios. with the increasing number of connected devices, the importance of energy-efficient communication has gained greater prominence than ever before. Furthermore, as terahertz (THz) communication is expected to play a significant role in 6G, future work could focus on developing accurate channel models for THz frequencies. This will be essential for designing effective xURLLC techniques tailored to THz communication.

REFERENCES

[1] She, Changyang, Trung Q. Duong, Saeed R. Khosravirad, Petar Popovski, Mehdi Bennis, and Tony QS Quek. "URLLC: Faster, Higher, Stronger, and Together." Ultra-Reliable and Low-Latency Communications (URLLC) Theory and Practice: Advances in 5G and Beyond 2023.

[2] Wang, Y., Wu, H., Jhaveri, R.H. and Djenouri, Y., "DRL-Based URLLC-Constraint and Energy-Efficient Task Offloading for Internet of Health Things," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[3] 3GPP. "Service requirements for the 5G system." *3rd Generation Partnership Project (3GPP), Technical Specification (TS) 22.261,* 2019.

[4] Le, Trung-Kien, Umer Salim, and Florian Kaltenberger. "An overview of physical layer design for Ultra-Reliable Low-Latency Communications in 3GPP Releases 15, 16, and 17." IEEE access 9 , 433-444, 2020.

[5] Ji, Hyoungju, Sunho Park, Jeongho Yeo, Younsun Kim, Juho Lee, and Byonghyo Shim. "Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects." IEEE Wireless Communications 25, no. 3, 124-130, 2018.

[6] Almarshed, Saleh, Dionysia Triantafyllopoulou, and Klaus Moessner. "Swift HARQ Based on Machine Learning for Latency Minimization in URLLC." IEEE Access, 2023.

[7] Zhang, Jingxuan, Xiaodong Xu, Kangjie Zhang, Bufang Zhang, Xiaofeng Tao, and Ping Zhang. "Machine learning based flexible transmission time interval scheduling for eMBB and uRLLC coexistence scenario." IEEE Access 7, 65811-65820, 2019.

[8] Ding, Jie, Mahyar Nemati, Shiva Raj Pokhrel, Ok-Sun Park, Jinho Choi, and Fumiyuki Adachi. "Enabling grant-free URLLC: An overview of principle and enhancements by massive MIMO." IEEE Internet of Things Journal 9, no. 1, 384-400, 2021.

[9] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5g wireless networks," Communications Magazine, IEEE, vol. 52, no. 8, pp. 82–89, Aug. 2014.

[10] Almekhlafi, Mohammed, Mohamed Amine Arfaoui, Mohamed Elhattab, Chadi Assi, and Ali Ghrayeb. "Joint resource allocation and phase shift optimization for RIS-aided eMBB/URLLC traffic multiplexing." IEEE Transactions on Communications 70, no. 2, 1304-1319, 2021.

[11] Shahjalal, Md, Woojun Kim, Waqas Khalid, Seokjae Moon, Murad Khan, ShuZhi Liu, Suhyeon Lim et al. "Enabling technologies for AI empowered 6G massive radio access networks." *ICT Express,* 2022.

[12] Affan, Affan, Shahid Mumtaz, Hafiz M. Asif, and Leila Musavian. "Performance analysis of orbital angular momentum (oam): A 6g waveform design." IEEE Communications Letters 25, no. 12, 3985-3989, 2021.

[13] Basar, Ertugrul. "Reconfigurable intelligent surface-based index modulation: A new beyond MIMO paradigm for 6G." IEEE Transactions on Communications 68, no. 5 (2020): 3187-3196.

[14] Song, Ho-Jin, and Namyoon Lee. "Terahertz communications: Challenges in the next decade." *IEEE Transactions on Terahertz Science and Technology* 12, no. 2 , 105-117, 2021.

[15] C. Huang, S. Hu, G. C. Alexandropoulos, A. Zappone, C. Yuen, R. Zhang, M. Di Renzo, and M. Debbah, ''Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends,'' IEEE Wireless Commun., vol. 27, no. 5, pp. 118–125, Oct. 2020.