IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received December 10, 2018, accepted December 28, 2018, date of publication January 23, 2019, date of current version September 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2894263

Mission-Critical Machine-Type Communication: An Overview and Perspectives Towards 5G

NAJIB AHMED MOHAMMED, ALI MOHAMMED MANSOOR, AND RODINA BINTI AHMAD

Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia

Corresponding author: Ali Mohammed Mansoor (ali.mansoor@um.edu.my)

This work was supported by the Fundamental Research Grant Scheme (FRGS) through the Ministry of Higher Education, Malaysia, under Project FP007-2016.

ABSTRACT The machine-type communication (MTC) plays an essential role in the mobile network society and it has demonstrated its effectiveness to generate significant revenues for mobile network operators. The MTC has become the main communication paradigm for several emerging smart services, such as public safety, health care, industrial automation, drones/robotics, utilities, and transportation. The MTC requires a major model change to cope with the improvement in 5G and imposes various requirements on the enabling technology, such as ultra-reliability, low latency, and availability. The scheme that fulfills such requirements are called a mission-critical MTC (mcMTC). However, the mcMTC is still in the early-standardization phase of 5G new radio and it needs a lot of research efforts to be improved. This paper presents an extensive review and evaluations to highlight diverse challenges and future aspects of mcMTC on 5G-enabling technologies. A number of research opportunities from both academic communities and industrial partners are given as examples for encouragement purposes.

INDEX TERMS Machine-type communication, mission-critical MTC (mcMTC), MTC, URLLC, 5G, M2M.

I. INTRODUCTION

The focus of cellular communication has been historically engineered on human-centric communication, which typically provides some form of information to people from network-hosted servers, such as mobile Internet services or video streaming. Hence, the 5G is expected to play a central role in Machine-Type Communications (MTC) and Human-to-Human (H2H) communications. MTC has been formed by 3rd Generation Partnership Project (3GPP) and it is characterized by entirely automatic data generation, actuation, processing, and exchange between intelligent machines, with or without low intervention of humans [1]. MTC, often termed Machine to Machine (M2M), starts to play an increasing role in cellular networks and the efforts have been put in the latest Long-Term Evolution (LTE) releases to address MTC requirements [2] i.e., wireless system to a larger extent, will be used in the context of MTC [3]. As a result of the rapid development of embedded devices, MTC is becoming the dominant communication paradigm for various emerging smart services including industrial automation, public safety, health-care, utilities, transportation, smart metering, remote manufacturing, and numerous other applications [4], [5], [6] while mobile devices communication among humans still exists. Unlike H2H telecommunication, in which messages, voices, and videos are transmitted, most MTC devices transmit only a small amount of data. A great deal of efforts has been dedicated in Telecommunications industry to realize high-capacity and high-throughput infrastructure. However, supporting small data communications for MTC subscribers has been neglected.

The MTC in context of 5G promises huge market growth with an expectation to support connectivity of at least 100 billion devices with the connection speed of 10 Gigabits per second (Gb/s). Furthermore, to provide a zero-distance connectivity between people and connected machines with extremely low latency and response times [7]. In addition, enhanced mobile Internet services will improve consumers experience and guarantee profits for operators. MTC will also drive economic and societal growth to entirely new diversities. Thus, the emerging Industrial Internet promises a surpassing transformation of global industry, with a targeted economic boost that is more than $19 trillion within the next decade [4]. Deployment of these networks is expected to be emerged between 2020 and 2030.

The continuing growth in demand for cellular-based MTC from 2014 to 2019 motivates initiatives in the industry, mobile networks operators, academic communities, and standardization bodies to investigate evolutionary and revolutionary radio access technologies to make MTC traffic accommodate 5G. For instance, the European Union has formed an official project called ''Mobile and wireless communications Enablers for the Twenty-twenty Information Society (METIS)'' for joining 5G research and standard promotion [8]. The Chinese government established an official promotion group called ''International Mobile Telecommunications Twenty-Twenty'' (IMT2020), whose main role is to establish the foundation of 5G mobile and wireless communications system. In addition, manufacturing companies have invested potential research projects to participate to the foundation of 5G communications and to enable mission-critical applications. Some of these industrial partners are: Ericsson [9], [10], [11] Huawei [12], 5GNOW [13], 3GPP [1], NOKIA [14], FuTURE Forum of China, 5GPPP of Europe Union, ADWICS of Japan, 5G forum of Korea, 5G Novel Radio Multiservice adaptive network Architecture (5G NORMA) [15], T-Mobile and others [16]. URLLC2017 conference gathered more than ten commercial and industrial companies [17]. Figure 2 lists most of the industrial companies involved in industrial-leading demos, simulations, testing, and trials on the path to commercialization of 5G.

Although, MTC has extensive potentials to accompany improvements for both lifestyle and business, MTC is confronting several unsolved technical challenges that differ from those in H2H communications, which are related to massive connectivity, latency, reliability, availability, energy efficiency, heterogeneous devices, limited spectrum, hardware miniaturization, security and privacy [13]. In essence, Mission-critical applications require very low latency, ultra-high reliability, and high availability communications i.e. mission-critical MTC (mcMTC) [6]. Unfortunately, all current mission-critical applications use wired networks because no current wireless networks can meet their mcMTC requirement [18].

To the best of our knowledge, no work has been done yet to combine all these requirements into a theoretical framework, although the foundation has been provided in Polyanskiy work [19]. Moreover, it is not possible to satisfy these mcMTC relentless requirements with a small modification of current radio access technology [20]. Hence, This study investigates how next generation of cellular networks must be designed in order to cater to this type of traffic. Also, to study the flexibility options of mcMTC to support its various use cases which have different requirements for connectivity in terms of reliability, latency, and availability. Furthermore, it is important to explore the 5G multiplexing of mcMTC services without contradicting with other operating modes which are enhanced Mobile Broadband (eMBB) and massive Machine-Type Communications (mMTC). Other challenges aside from these three mcMTC requirements such as low power consumption and expected security level of the network are also interesting, but are often irrelevant for mcMTC [6] and are not covered in this review.

This paper presents a critical review of mcMTC requirements and challenges in the context of 5G wireless technology. It first presents the 5G operating regions and the existing MTC access methods; then it explains in detail the key requirements of mcMTC applications/services. Finally, it provides an extensive review on enabling technologies towards providing mcMTC for critical-mission applications. It shows the vision of different companies, organizations, and researchers who are pioneers in this field of study focusing on mcMTC and the proposed solutions for its requirements towards 5G. Eventually, this paper aims to outline and analyze the involved technical issues, to review recent advances, and to bring the state-of-the-art of mcMTC solutions along with future research directions.

The rest of this review paper is organized as follows. Section II presents the three 5G service regions. MTC access methods are presented in section III. Section IV discusses the key requirements of mcMTC. Enabling technologies of mcMTC which includes air interface and multiplexing schemes, network architecture, packet and frame structure, control channel design, communication diversity, proactive packet drop, and coding and modulation are discussed in section V. Finally, conclusion and future work are drawn in section VI. The organization of the paper is illustrated in Figure1, while the acronyms used in this paper are listed in Table1.

II. 5G WIRELESS OPERATING REGIONS

The evolution of cellular wireless systems in the last three decades from 2G to the current LTE-Advanced (LTE-A) has been dedicated to offer the users connectivity with a higher demand of data rates. 5G features at least two new operating modes mMTC and mcMTC and increasing the speed of existing operating mode enhanced Mobile Broadband (eMBB). These two new operating modes and the existing mode have been described in [3] and [21] as the three generic 5G services (i.e. eMBB, mMTC, and mcMTC). Among these three generic 5G services, mcMTC is expected to be a technology driver of 5G [22], [23]. In contrast to the other services, mcMTC refers to the services that provide a very high reliability and a very low latency as well as ubiquitous communication to support mcMTC applications, on which service failures have severe consequences. Table 3 summarizes the key features and the main requirements of these three 5G operating regions.

A. EMBB

eMBB initially becomes an extension to existing 4G services and therefore it is the first 5G service which evolved in the 3GPP Release 15 standards in June 2018. The standard was approved in December 2017 and already standardized

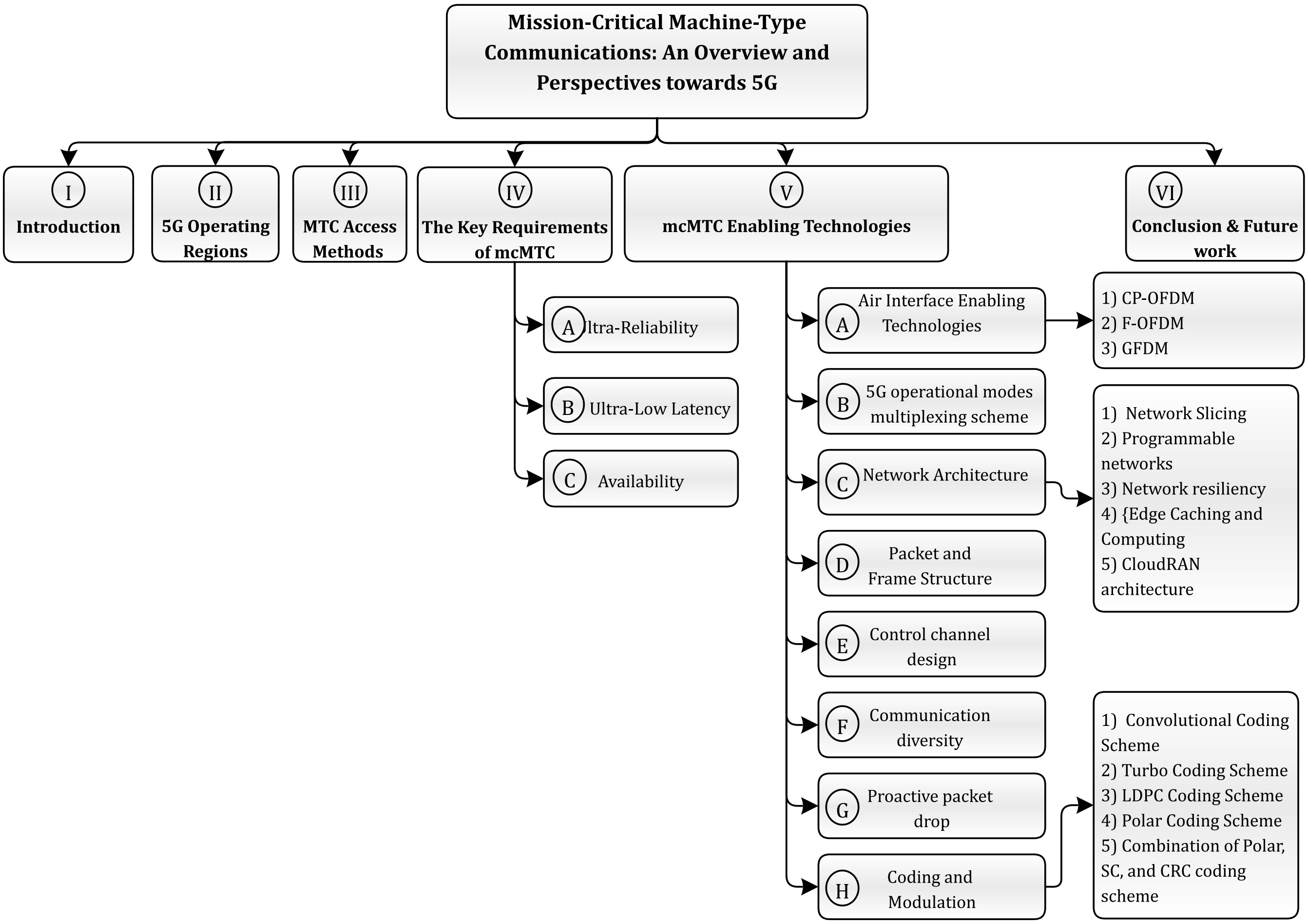

FIGURE 1. Structural organization of this review paper.

FIGURE 2. Industrial partners participated in 5G development and ecosystem collaborations on the path to commercialization.

in 3GPP release 15 which could be commercially available as early as 2019. Ericsson estimates that there will be one billion 5G subscriptions for eMBB globally by 2023 [24], with Asia and North America likely to be the first adopters. However, 5G Phase 2 will go beyond eMBB services to more transformational mMTC and mcMTC applications and will be included in Release 16, which is due to be completed at the end of 2019.

B. MMTC

mMTC refers to services, in which there is an enormous amount of sensors and actuators to control an environment with simple, scalable and energy efficient communication supporting massive number of devices in some areas. Moreover, mMTC devices required to remain very simple and can operate on batteries for many years. mMTC data transferred per device are typically infrequent and are accompanied by relaxed delay requirements. mMTC already emerges as an extension of the 4G LTE systems and becomes in the scope of current standardization of 3GPP and IEEE.

C. MCMTC

mcMTC is having the most crucial challenges towards making it accommodates 5G. mcMTC is also called Ultra Reliable Low Latency Communications (URLLC), but we prefer in this review to use mcMTC which is relevant to real-time control and automation of dynamic processes in automated cyber-physical systems, such as industrial process control. mcMTC is still in the early-standardization phase of 5G New Radio (NR) and is expected to be standardized on the

TABLE 1. List of Acronyms and Abbreviations.

Acronym	Description	Acronym	Description
3GPP	3rd Generation Partnership Project	M2M	Machine to Machine
4G	4^{th} generation of cellular mobile communications	MAC	Medium Access Control
5G	The fifth generation of cellular mobile communications	mcMTC	Mission-Critical Machine-Type Communications
5GNOW	5th Generation Non-Orthogonal Waveforms	MCS	Modulation Coding Scheme
ACK	Acknowledgment	MEC	Mobile Edge Cloud
AEI	Estimation and Indication	METIS	Mobile and Wireless Communications Enablers for Twenty-Twenty (2020) Information Society
ARQ	Automatic Repeat-Request	mMTC	Massive Machine-Type Communications
BER	Bit Error Rate	mmWave	millimeter-wave
BLER	Block Error Rate	MTC	Machine-Type Communications
BS	Base Station	NACK	Negative-Acknowledgment
CP	Cyclic-Prefix	NFV	Network Functions Virtualization
CP-OFDM	Cyclic-Prefix Orthogonal Frequency-Division Multiplexing	NOMA	Nonorthogonal Multiple Access
CQI	Channel Quality Indicator	NR	New Radio
CRC	Cyclic Redundancy Check	OFDM	Orthogonal Frequency-Division Multiplexing
CS	Compressed Sensing	OFDMA	Orthogonal Frequency-Division Multiple Access
CSI	Channel State Information	OSI	Open Systems Interconnection
D2D	Device-to-Device	OSTBC	Orthogonal space-time block coding
DL	Downlink	PHY	Physical Layer
E2E	End-to-End	QoS	Quality of Service
eMBB	enhanced Mobile Broadband	RAN	Radio Access Network
FDD	Frequency Division Duplex	RSC	Reliable Service Composition
FEC	Forward Error Correction	RTL	Reliable Transmission Link
F-OFDM	Filtered-OFDM	RTT	Round-Trip Time
GFDM	Generalized Frequency-Division Multiplexing	SCS	Subcarrier Spacing
GSM	Global System for Mobile Communications	SDN	Software Defined Networking
H2H	Human-to-Human	SERCOS	Serial Real-time Communication System
HARQ	Hybrid automatic repeat-request	SINR	Signal-to-Interference and Noise Ratio
HTC	Human-type Communications	TBC	Time Block Coding
IMT2020	International Mobile Telecommunications Twenty-Twenty	TDD	Time division duplex
KPI	key performance indicator	TTI	Transmit Time Interval
LDPC	Low-Density Parity-Check	UE	User Equipment
LTE	Long-Term Evolution wireless communications	UL	Uplink
LTE-A	LTE-Advanced	URLLC	Ultra Reliable Low Latency Communications Equipment

late of 2019, poses a lot of research problems that remain unsolved. This is due to the fact that mcMTC require communications with a very high reliability, high availability, and small latency going down to millisecond level [6], [9], [25].

III. MTC ACCESS METHODS

The connection of MTC devices to the network infrastructure could be either wireless or wired (i.e., cable, xDSL, and optical). Wired transmission technologies, such as Fieldbus and Ethernet-based systems, have dominated factories automation because of their reliability, fast response time and availability. However, the interest in employing wireless communications has grown recently to use the MTC connectivity for real-time applications, such as factory automation and process control in industrial networks [26], [27].

The third Generation of the Serial Real-time Communication System (SERCOS-III) interface is the international standard on the third generation open digital interface, which provides an openness, compatibility, high speed, and reliable real-time data communication among the controller and actuators [18]. SERCOS-III has been designed to provide reliable Media Access Control (MAC) sublayer schemes and to guarantee a certain level of performance (i.e. Latency $4ms$, Reliability:$10^{-8} - 10^{-10}$, and Availability 100% [18], [28]). For the sake of supporting critical mission services, Sercos-III uses the speed of fast Ethernet (i.e. IEEE 802.3 100 Mbit/s baseband) and supports full duplex Physical (PHY) layer entities. In addition, it uses Cyclic Redundancy Check (CRC) codes to detect errors in the packet and Automatic Repeat reQuest (ARQ) policy to correct errors which improve the reliability of the service. Moreover, SERCOS-III supports for either line or ring topologies which improves fault tolerance (i.e. availability) by supporting redundant data transfer. In case of a break at any point in the ring, the SERCOS-III protocol automatically switches over to a dual-line structure. However, latency is not much more affected by recovery, which occurs in 25 μs.

Although the wired solution can provide reliability, high rates and a very high response time, it is not suitable for mcMTC applications due to its expensive cost, lack of scalability and mobility. Alternatively, wireless access methods can be either cellular (i.e. LTE-A, GPRS, 3G, and WiMAX) or capillary (i.e., WLAN, IEEE 802.15.4x, and ZigBee). The wireless capillary solution, commonly used for shared short-range links, is generally cost-effective, low power, and scalable. It also allows more detailed monitoring than in wired communications, which would be economically more viable. However, it is not applicable to mcMTC applications due to the lack of coverage, low rate, weak security, and interference problem. On the other hand, the wireless cellular solution is promising to be as an infrastructure-based solution for the mcMTC applications due to its ability to provide ubiquitous coverage, mobility, roaming, and security solutions.

IV. THE KEY REQUIREMENTS OF MCMTC

Historically, wireless networks have not been designed to satisfy the mcMTC applications/services [11], while most of the efforts have been dedicated to offer the users connectivity with a higher demand for data rates. For example, LTE networks can perform a reliability rate of nearly 10^{-2} [29], while the end-to-end (E2E) latency is approximately few

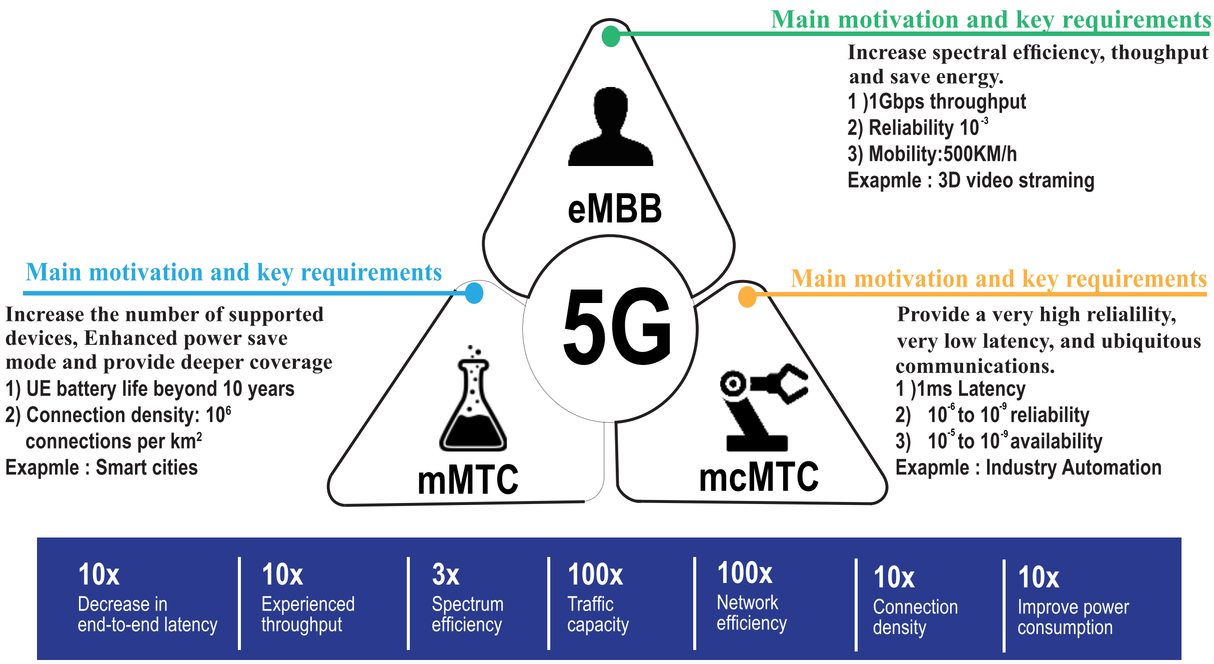

FIGURE 3. Summary of the three regions of 5G systems based on IMT2020, 3GPP key performance indicators (KPIs), and Qualcomm report 2018.

TABLE 2. Comparison of the mcMTC requirements over SERCOS-III wired network, LTE-A, and 5G.

Requirement	(Sercos-III)	LTE-A	5G
Latency	$4ms$ [18]	$\approx$ $5ms$ [33], [34]	$\leq 1ms$ [35], [3]
Reliability	10^{-10} to 10^{-12} [18], [28]	10^{-2} [35], [30], [36]	10^{-5} to 10^{-9} [35]
Availability	100% [18]	10^{-2} [37]	10^{-9} to 10^{-12} [4]

milliseconds [30]. For flexible requirements of latency, LTE can perform satisfied reliability via re-transmissions at different protocol layers. However, accomplishing a high reliability while satisfying a very-low latency presents a new challenge for the NR deployment. The key requirements of mcMTC can be considered as the ultra-high reliability, ultra-low latency and availability [6], [31], [32]. Table 2 presents a comparison of the mcMTC requirements in wired, LTE-A, and 5G networks.

In mcMTC, the required levels of reliability, latency, and availability vary according to the selected use case and application. Therefore, the NR air interface and system design should be scalable enough to efficiently use the available network resources. Figure 4 depicts the diversity of latency and reliability requirements for different mcMTC use cases and applications. For mcMTC services, the levels of availability are required to be on the order of 10^{-8} to 10^{-12} and the packet loss probability has to be on the scale of 10^{-5} to 10^{-9} [35], while the latency has to go down to millisecond level [9], [7].

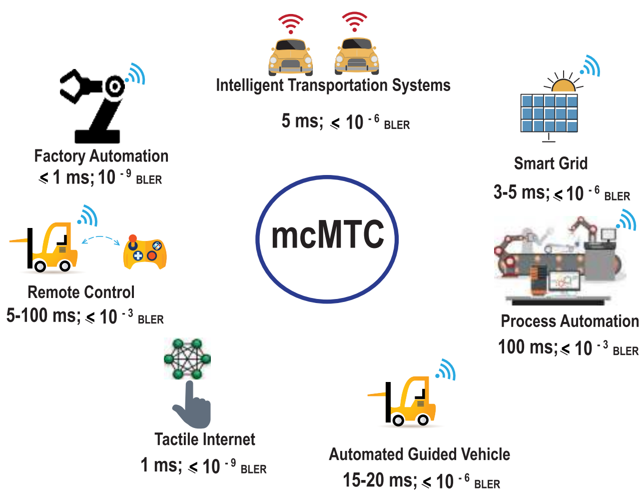

FIGURE 4. Diversity of latency and reliability requirements for mcMTC use cases and applications.

The International Telecommunication Union (ITU) has set a target to achieve a $1ms$ Round-Trip Time (RTT) over-the-air communication for a single transmission [3]. This includes transmission of the payload until the corresponding acknowledgment is received.

To the best of our knowledge, no work has been done yet to combine reliability, latency, and availability into a theoretical framework, although the foundation has been provided in Polyanskiy work [19]. Also, no wireless communication systems have been proposed for systems considering mcMTC requirements of reaching $1ms$ of latency with tens to hundreds of nodes and reliability constraint less than 10^{-6}.

A. ULTRA-RELIABILITY

Reliability denotes to the capability of guaranteeing successful message transmissions from one peer to another within a given latency bound [9], [38].

The work presented in [39], the reliability is separated into two parts: link reliability and system reliability. They have defined the link reliability as the data that can be transmitted successfully within a given time frame, while system reliability is defined as when a system can precisely indicate the absence of link reliability and ensure the presence of link reliability when required by the application.

Reliable communication has been well-studied as a fundamental problem in the information theory. Shannon's landmark theoretical study shows that the possibility of communication with vanishing probability of error is at non-zero rates [40]. Reliability has been studied as well from the networking perspective to complement the techniques used at PHY layer. The reliability problem is a challenge in MTC context without human assistance, due to long distance transmissions, intermediate routing, and wireless tampering and sniffing. However, the MTC requirements differ for various types of services [9].

The most recent 4G systems can achieve a reliability of approximately 10^{-2} and measure as Typical Block Error Rate (BLER) [13], while the performance requirements of mcMTC services should be much better than this, and in fact, mcMTC in 5G services require reliability at least 10^{-5} BLER within 1 ms of latency period [41], [42]. In respect of mcMTC services, the target BLER vary from 10^{-5} to 10^{-9}. However, according to the work conducted in [13], the targeted reliability requirement of 5G will be 10^{-9} residual BLER within the latency bound. Reliability measurement (BLER) can be formulated as:

$$BLER \leq 10^{-5} \tag{1}$$

BLER is a function of Bit Error Rate (BER) given as:

$$BLER = 1 - (1 - BER)^N, BER = 1 - (1 - BLER)^{1/N} \tag{2}$$

where N is the number of bits in the block. From equation 2, it can be clearly observed that the bigger block will lead to higher BLER. Therefore, it is recommended to use short code block to support providing ultra-reliable mcMTC services.

5G tends to provide more dynamic solutions to increase the reliability instead of providing solutions to increase the bandwidth as in LTE. For instance, a selective blanking of the strongest interferers during retransmissions could significantly boost the success probability.

Considering a different level of reliability for different mcMTC services, we need a way, in which a specific communication service is composed. As suggested by [39], it is not necessary to fail the service whenever the reliability requirement is not fulfilled. Therefore, Reliable Service Composition (RSC) should offer a specific level of functionality when it is not possible to meet the full functionality. The key issue of making RSC operational in the case of MTC is by providing a reliable criteria for detecting the service level that the system should apply at a given time. The design of data/metadata for every service level should be integrated into an overall protocol, which could flexibly switch between modes as dictated by the dynamic conditions [38]. In summary, the implementation of RSC relies on a careful consideration of the requirements set by the application and the availability indicator that the communication layer provides to the application layer.

Most relevant mcMTC reliability enablers are listed below and are discussed further in section V:

1) Waveform selection
2) Multiplexing scheme
3) Channel estimation accuracy improvement.
4) Redesign the resource allocation at MAC layer
5) Channel coding schemes.
6) Frame structures and short TTI
7) Network architecture including network slicing, caching, CloudRAN, and edge computing.
8) Antenna, space, and frequency diversity schemes
9) Supporting the prompt transmission of mcMTC packets by following multiplexing approach giving the priority to mcMTC packets over eMBB and mMTC packets.
10) Using Hybrid-ARQ (HARQ) on the MAC layer and erasure coding coupled with ARQ on the upper layers.

B. ULTRA-LOW LATENCY

Latency refers to the time delay between the time at which data is generated and transmitted from one device, and the time at which the same data is correctly received by another device. According to 3GPP, latency in the network can be classified into control plane latency and user plane latency [43].

- User plane latency also known as transport delay is defined as the one-way time it takes to successfully deliver an application layer packet/message from the radio protocol layer ingress point to the radio protocol ingress point of the radio interface, in either Uplink (UL) or Downlink (DL) in the network for a given service in unloaded conditions, assuming the mobile station is in the active state.
- Control plane latency is defined as the transition time from a most "battery efficient" state (e.g., idle state) to the start of continuous data transfer (e.g. active state).
  The minimum requirements for user plane latency are 4 ms for eMBB and 1 ms for mcMTC assuming a single user while the minimum requirement for control plane latency is 20 ms [43]. Proponents are encouraged to consider lower control plane latency, e.g. 10 ms [44]. This requirement is defined for the purpose of evaluation in the eMBB and mcMTC usage scenarios.

5G community generally considers the network latency as one of the key requirements for future wireless networks enabling new applications by means of E2E latency that goes down to the millisecond level. Such latency cannot be performed by the current 4G technology. In this regard, we need drastic changes in network architecture including core and Radio Access Network (RAN) for achieving E2E latency on the order of 1 *ms* [5]. The current latency in LTE-A is about 5*ms* [33], [34]. Therefore, it is expected that the latency target for a new 5G radio access technology is going to be up to five times lower than compared with the latency of 4G networks (i.e. less or equal to 1ms). Researchers are wondering if the 1*ms* latency target will actually become a part of the 5G standard due to the restriction of fundamental laws of physics, such as the speed of light, the intrinsic properties of the fiber optics in which the light propagation within fiber optics is slower compared with the light propagation within vacuum.

E2E latency T_L is the time which includes the following components: processing time on the transmitter/receiver devices T_{proc}, over-the-air transmission delay in one way T_{trans}, and the network processing time $T_{netproc}$ (I.e. the base station (BS) and control server) [45]. Taking 1*ms* E2E latency T_L for instance, the distribution of the latency over various components can be as follows: 300µs for T_{proc}, 200µs for T_{trans} in both sides, and 500µs for $T_{netproc}$.

$$T_L = T_{proc} + 2T_{trans} + T_{netproc}$$

$$T_L \leq 1\ ms$$

The processing latency, including the processing in the device T_{proc} and the processing in the network $T_{netproc}$, includes time for receiving data symbols, acquiring channel information, extracting control (scheduling) information, decoding data packet, and checking the existence of error [20]. Another elementary delay component is the Transmit Time Interval (TTI), defined as the minimum data block length, which is involved in each transmission of grant, data, and retransmission due to errors detected in higher layer protocols. Figure 5 assumes that every 100 bit packet must be transmitted within a 1ms E2E latency budget, as aimed to 100µs for over-the-air transmission time that corresponds to 10% of the E2E latency budget [21]. However, the work presented in [7] suggests that $T_L \leq 1\ ms$ to support mcMTC.

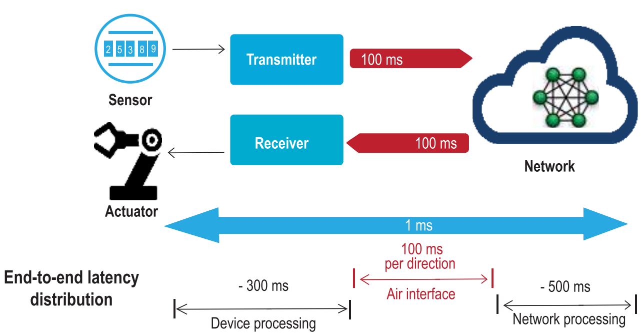

FIGURE 5. Distribution of the E2E latency assuming a 100µs air interface delay [45].

mcMTC services with ultra-low latency mostly transport very short control and command messages with the size of KBytes and Bytes and these messages are to be delivered very frequently. However, there are a few key reasons causing delay in RAN which are basically related to the following:

1) The packet retransmissions caused by channel errors and congestion.
2) The Link establishment caused by grant acquisition and random access [6].
3) Packet processing and queuing delay.
4) Channel coding delay.

Therefore, it requires careful redesign of PHY and MAC layers to overcome these reasons. A clean slate redesign includes:

- Requirement latency can be addressed based on 3GPP, by removing the retransmission by having early channel estimation and on-the-fly decoding [6].
- Network architecture including CloudRAN and new entities such as Software Defined Networking (SDN), Network Functions Virtualization (NFV), Mobile Edge Cloud (MEC), and fog network along with new backhaul based solutions.
- Caching solutions including caching placement, content delivery, centralized caching, and distributed caching [5], [46].
- Minimizing system overhead, as an example, procedures for user scheduling, channel training, and resource allocation have to be combined or removed [20].
- Since the packet re-transmission mechanism degrades the latency, the probability of packet error of the first transmission has to be reduced significantly.
- To avoid queuing delays at the radio transmitter the MAC layer should be designed to enable immediate access. This can be achieved by providing instant-access resource allocations dimensioned (i.e. real-time resource allocation) to minimize collision risks.
- Maintaining multiple connectivity links simultaneously can provide diversity and redundancy to address such stringent requirements [47].
- Reducing TTIs and widening the bandwidth of radio resource blocks to lower the latency over the radio link. This can be achieved by using fewer OFDM symbols per TTI and shortening OFDM symbols via wider Subcarrier Spacing (SCS) as well as lowering HARQ RTT [35].
- mcMTC packet should be transmitted immediately without delay by following a multiplexing mechanism giving the priority to mcMTC packets.
- For some use cases, low-latency communications are required between devices in close proximity. In this case, a direct D2D communication link can help in providing low-latency transmission [47].
- Reducing queuing time by using fewer OFDM symbol duration causing an increase of the SCS and hence fewer resource blocks are available in the frequency domain. This shortcoming can be alleviated using grant-free transmission in the UL. On the DL, longer TTI is needed at high offered loads to cope with non-negligible queuing delays [48].
- Provide a new/modified frame or packet structure, waveform designs, and multiple access techniques.

- Modulation and coding schemes to reduce processing delays.
- Control channels enhancements.
- Other enabling techniques including modifying the spacing of the sub-carrier, providing low latency symbol detection, and mmWave aggregation.

Recently, various low-latency transmission protocols have been suggested, such as grant-free, one-shot, and feedback-less protocols [20]. These are proposed to reduce the latency caused by the control signaling.

The work conducted in [46] demonstrates the first wireless broadband communication system which can achieve an E2E latency below 1*ms* and 20MHz bandwidth between two endpoints over the air. The system is introduced based on the PHY and MAC signal processing algorithms, which are implemented on a multi-core demand-side platform with a flexible software-defined radio toolkit. The demonstration shows a transmission of real data packets while evaluating E2E latency probes at the same time. However, the demonstration does not mention how to cope up achieved latency with ultra-reliability needed for mcMTC systems.

In [5], a comprehensive survey is provided for low-latency solutions in the context of 5G from three different perspectives: Core network solutions, RAN solutions, and Caching solutions. However, the study does not provide a detailed comparison of these solutions.

Lower latency concern took up its attention of 3GPP Release 15 by introducing two enhancements over LTE-A. The first is related to providing a shorter TTI. Traditional LTE TTI is 14 symbols which is 1*ms* scheduling interval. However, with TTI, both 7 symbols (0.5 *ms*) as well as 2 symbols (0.142 ms), scheduling interval is supported. The second is by reducing processing time by making the terminal respond to DL data and UL grants in 3*ms* instead of 4*ms*.

C. AVAILABILITY

Availability refers to the system endurance against possible outage scenarios. An mcMTC service should get a prompt response from the wireless medium to transmit the mcMTC packets immediately within a given time frame in the scheduling period (i.e., reserved resources for the mMTC and eMBB services have to be used for the mcMTC service). Based on the definition of reliability provided in [39], reliability requires a Reliable Transmission Link (RTL). An Availability Estimation and Indication (AEI) mechanism has been proposed in [39]. This mechanism could predict the availability of RTL under certain conditions. The procedure is summarized as follows: the application sends an availability request to the AEI; then, the AEI evaluates the Signal-to-Interference and Noise Ratio (SINR) and/or the Acknowledgment (ACKN)/Negative-ACKN (NACK) statistics of the retransmission protocols used at the link level. After that, the AEI provides an availability indicator for the application based on the original requirements of the availability request. However, this is at the cost of reducing the availability of RTL. Using retransmission improves the RTL but at the cost of increasing the latency.

The authors in [4] explain the need to offer an availability about 10^{-9} toward the next generation mcMTC systems, to control the probabilities of the underlying rare events. However, enhancing coverage is an essential issue for the deployment of a wireless mcMTC, in which devices and machines are installed in difficult locations. There are many solutions to enhance coverage, such as leveraging protection and redundancy capabilities which already have been deployed in most service provider networks, such as G.8032 Ethernet rings. In addition, enhanced availability can be gained by using repetition or sub-frame bundling, retransmission using HARQ and many others.

V. MCMTC ENABLING TECHNOLOGIES

The key performance requirements of mcMTC discussed above affect the design choices of every component in the communication link and the optimization process within the whole protocol stack. Thus, for enabling these requirements, the entire system needs to be redesigned; an advanced channel estimation technique, restructuring packets, frames, backhaul, storage, advanced error control coding, control signaling and corresponding multiplexing scheme must be deployed, and an advanced scheme for channel coding that is suitable for the short packet transmission should be used [20]. In addition, both the data and control planes may require significant enhancements and new technical solutions can be given from both the radio interface and network architecture aspects. These fundamental design changes promise benefits that contribute to realizing higher reliability, reducing latency, and improving availability [3], [13]. Below, we are going to review the enabling technologies of mcMTC towards 5G, such as air-interface waveforms, network architecture, frame structure, control channel design, spatial diversity, proactive packet drop, and coding schemes. Figure 6 depicts the mcMTC key requirements and its main enablers. X-axis denotes latency in *ms* while Y-axis denotes reliability from 10^{-12} to 0 BLER. It shows that the ultra-reliable communications have a restrict condition in reliability between 10^{-12} and 10^{-8}, while they have a release latency condition goes to 100*ms*. In contrast, low-latency communications have a restrict condition in latency between 0.1*ms* and 1*ms*, while they have a release condition in reliability goes to less than 10^{-2} BLER. However, the mcMTC requires restrict conditions in both latency and reliability which go from 0.1 to 1 ms and from 10^{-12} to 10^{-9} BLER, respectively.

A. AIR INTERFACE ENABLING TECHNOLOGIES

It is not possible to fulfill the challenging requirements of mcMTC with minor modifications of recent radio access technologies. For that, a new air interface is required to be designed in 5G to deal with heterogeneous traffic types and to fulfill mcMTC key requirements. These requirements may vary depending on the selected use case and application. Therefore, the new air interface and system design should

TABLE 3. Differences between LTE and 5G NR with respect to PHY layer and different channels used in these standards.

Specification	LTE	5G
Frame Structure		
Radio Frame Duration	10ms	10ms
Number of subframes in a frame	10	10
Number of slots in a frame	20 (1ms)	20 (each 1ms)
Number of symbols per slot	7 symbols for normal CP, 6 symbols for extended CP	14 symbols for normal CP, 12 symbols for extended CP 2, 4 and 7 symbols for mini-slots
Symbol duration	32.552 ns for Subcarrier Spacing (SCS) of 15 KHz and N_{FFT} of 2048; FFT is a Fast Fourier Transform	0.509 ns for SCS of 480 KHz and N_{FFT} of 4096
FFT Size	2048	4096 for 240 KHz SCS; 8192 for 480 KHz SCS
Resource Blocks	100 (maximum)	100 or more
Frequency domain		
Carrier Aggregation	5 (Rel. 10); 32 (Rel. 12)	16
SCS	Fixed: 15 KHz	Variable: $2^n \cdot 15$ KHz (Where, n = -2,0,1,....,5)
Carrier bandwidth	1.4, 3, 5, 10, 15, 20 MHz	Variable, maximum per CC is 400 MHz
Frequency bands	Under 6 GHz	Up to 100 GHz
Beamforming	Applicable to certain transmission modes	With and without DL/UL reciprocity
Modulation	Up to 256 QAM	QPSK, 16 QAM, 64 QAM, and 256 QAM
MIMO	Up to 8x8	Up to 8x8
Channel coding scheme	Turbo coding for data	NR Polar codes (control); NR LDPC (data)
PHY layer waveforms	DL: CP-OFDM, UL: DFT-S-OFDM	DL: CP-OFDM, UL: CP-OFDM or DFT-S-OFDM
CP type	Normal CP, Extended CP	Normal CP for all SCS extended CP is supported for 60 KHz SCS
Channel coding for various channels	PBCH/PDCCH: TBCC PDSCH/PUSCH: Turbo Code PUCCH: RM Block Code	PBCH/PDCCH/PUCCH: Polar code PDSCH/PUSCH: LDPC
HARQ TTI	FDD: $9ms$, TDD: $8ms$	0.25 to 16 ms
UE Bandwidth Adaptation	Not allowed	Allowed

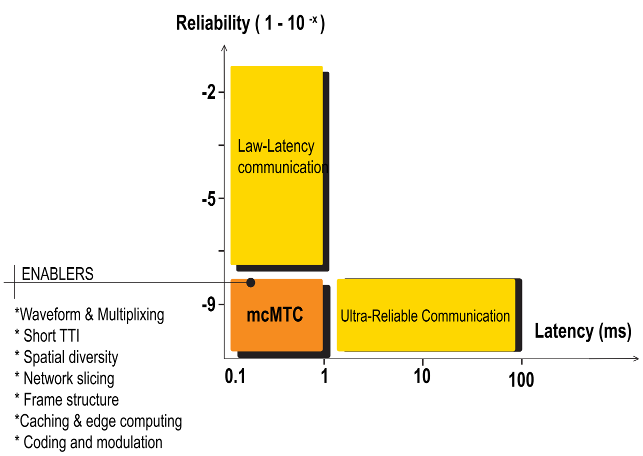

FIGURE 6. The key requirements of mcMTC and its main enablers.

be scalable enough to efficiently use the available network resources. To address the key requirements discussed above, several designs have been introduced for 5G air interface.

The 3GPP technical support presented in [49] shows that the NR access technology has been defined which in general has to add entirely new features providing a customized connection to any device from different smart devices platforms and bring the backward compatibility. The early features of this NR already standardized by 3GPP in July 2018 and the complete collection of features are to be finalized by the year 2020 (IMT2020) [20].

Several features are specific to NR in comparison with LTE shown in Table 3. These include mcMTC features, the broad range of carrier frequencies, several deployment options, configuration mechanisms, such as flexible waveform and protocols, adaptive frame structure, coding and modulation, multiple access schemes, and various use cases which is either machine-centric or human-centric and beyond eMBB. In addition, NR should support dynamic resource sharing between different latency and reliability requirements for both eMBB and mcMTC in downlink. However, the dynamic resource sharing between mcMTC and eMBB is supported by transmitting mcMTC scheduled traffic where mcMTC transmission may occur in resources scheduled for ongoing eMBB traffic. Downlink dynamic resources sharing between eMBB and mcMTC is enabled without pre-emption by scheduling the eMBB and mcMTC services on non-overlapping

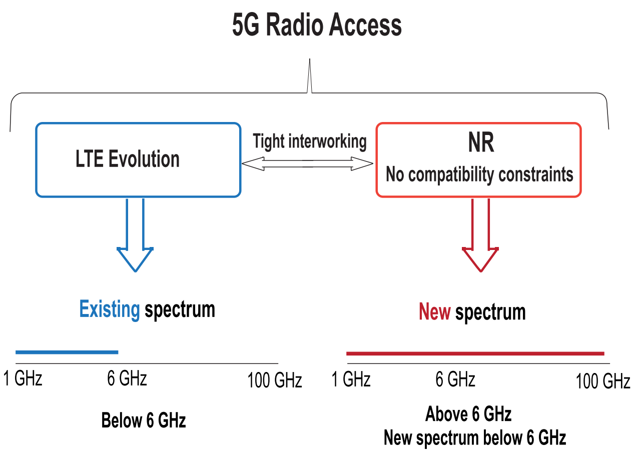

FIGURE 7. **5G Radio Access Vision [51].**

time/frequency resources [50]. As shown in Figure 7, NR and LTE are integral parts of 5G radio access. LTE is predicted to function below 6 GHz frequencies, while NR is expected to function from sub-1 GHz up to 100 GHz. The aggregation of NR and LTE traffic shall be enabled by tight integrations [51].

Huawei proposed that NR should be built upon software-defined flexible air interface and radio access virtualization [12]. In terms of air interface, it should be optimized in a specific way supporting various mcMTC use cases of applications without contradiction with other operating modes (i.e. eMBB and mMTC). In terms of virtualization of radio access, it encloses coordination and self-organization algorithms, which employ the protocols, features, and interfaces evading the limitations of the geographic "cell" construct.

To reach 5G era, the community of mobile communications has witnessed plentiful waveform proposals for NR and has come to the fact that there is no waveform that introduces all the required advantages. Several waveform proposals have been presented and the trend has been to tweaking Orthogonal Frequency-Division Multiplexing (OFDM) in any possible way, such as pulse shaping, wise filtering of sub-carriers, filtering of groups of sub-carriers, dropping cyclic-Prefix (CP), replacing CP with nulls or with other sequences, and allowing successive symbols to overlap in time. Even though OFDM has many advantages and it has been a great success, there exist several ideas for new 5G waveforms, which could add other advantages to the new cellular system under specific circumstances and conditions. However, multi-carrier waveforms are either variations of CP-OFDM (e.g., Unique Word OFDM, Pulse Shaped OFDM, Windowed-OFDM, Universally Filtered-OFDM, Filtered-OFDM) or super cases of OFDM (e.g., OFDM becomes a specialized case of a more complicated waveform, such as Filter-Bank Multi-Carrier waveforms). In addition to these discussed waveforms, others are being considered to be used with 5G. OFDM is considered as a baseline for up to 30 GHz. Unfortunately, a major drawback of OFDM systems is their large peak to average power ratio (PAPR) [52].

1) CP-OFDM

Ericsson's researchers have performed a thorough assessment of waveforms. They have chosen an OFDM based air interface in which inter-symbol interference can be kept minimal by setting CP-OFDM symbol longer than the channel delay spread. They have chosen the CP-OFDM as the most appropriate candidate for NR for many reasons listed in [51]. Two drawbacks exist in OFDM similar to all other multi-carrier waveforms which are high PAPR and less frequency localization. However, there are simple techniques that are well-established to improve frequency localization (i.e. windowing) and to reduce PAPR (e.g. with companding and clipping). These techniques could be used easily with CP-OFDM at the transmitter in a receiver agnostic way. However, Ericsson is the company, which realizes earlier that CP-OFDM has been the best potential for NR among the available candidates. However, Ericsson researchers have proposed this waveform assessment through 3GPP on April 2016 (the first 3GPP RAN 1 meeting on 5G New Radio) and have introduced CP-OFDM for 5G NR.

Ericsson has suggested some modifications to OFDM including reduced TTI and shorter OFDM symbol durations that enable fast and efficient data transmission and thus satisfy the requirement of mcMTC. Since the LTE symbol length could be too long in relation to the envisioned air interface requirements (e.g., maximal TTI is assumed to be 0.1 ms for the time-critical factory automation applications [9]), both the OFDM symbol length and TTI need to be scaled down for mcMTC. Reduced OFDM symbol length, error floor-free data coding, and higher-order antenna diversity are several examples of major technical enablers identified for the realization of mcMTC.

2) F-OFDM

Huawei has proposed to use filtered-OFDM (F-OFDM) as it can effectively increase connectivity, improve spectral efficiency, and reduce latency. This leads to facilitate the deployment of customized scenarios that are applied to MTC and it also leads to high bandwidth-consuming scenarios, such as virtual reality. The NR utilizes two-level non-orthogonality to maximize the number of connected devices and the spectrum efficiency, and to provide flexibility supporting various services.

3) GFDM

5GNOW researchers in [53] suggest to replace the orthogonality synchronism in OFDM. They have proposed several asynchronous non-orthogonal waveforms to be deployed in the new PHY layer such as F-OFDM, universal filtered multi-carrier (UFMC), filtered bank multi-carrier (FBMC) and generalized frequency-division multiplexing (GFDM). Among all these waveforms, GFDM can be selected as the enabler for low latency due to its block structure that has an advantage in utilizing specific sequences with

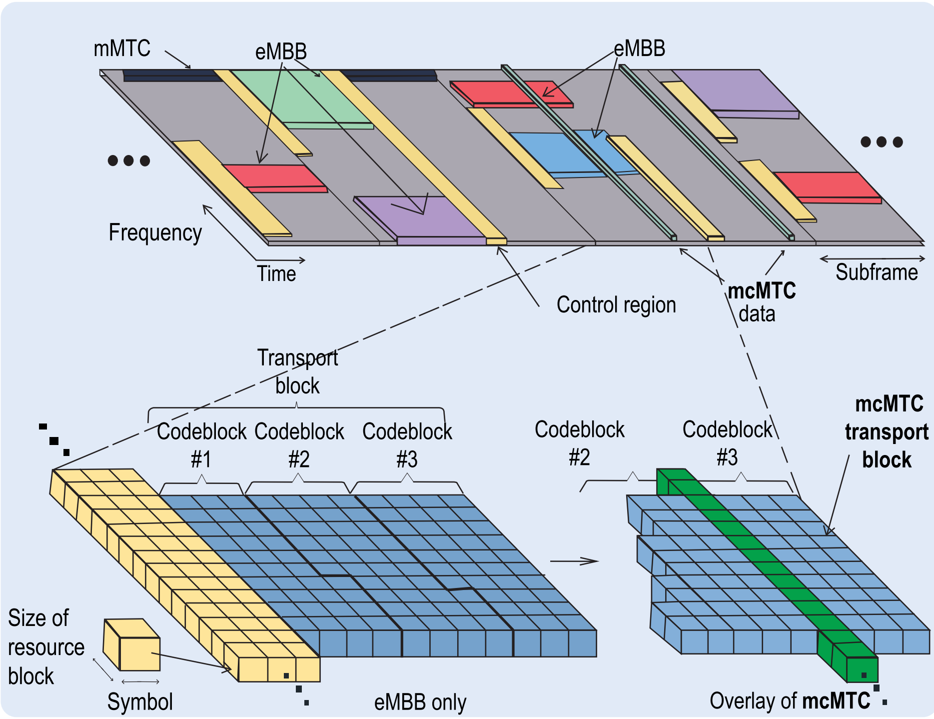

FIGURE 8. Transmission of mcMTC, eMBB, and mMTC packets at the subframe level, and scheduling of a mcMTC packet into an eMBB packet at the symbol level.

impulse self-correlation properties. In addition, GFDM framing could be designed to fulfill the time constraint 100µs [53].

4) SCMA

Sparse code multiple access (SCMA) is a new frequency domain non-orthogonal multiple-access technique. SCMA is favorable for massive connectivity due to its tolerance to overloading signals. In addition, SCMA uses blind detection technique; therefore, it can be considered as an enabler for grant-free multiple access which can effectively terminates the latency and signaling in the request-grant dynamic scheduling schemes [54]–[56].

As a result, it is anticipated that the final outcome for 5G waveforms may have an adaptive solution utilizing the optimal waveform for any introduced situation. 5G is being investigated and explored. For that, there is a good opportunity to consider the optimal waveforms for the 5G system, which shall be employed until 2040 at least. In March 2016, 3GPP has agreed to study several features of NR taking into account OFDM, unless significant gains could be presented by other waveforms [49]. In August 2016, 3GPP has agreed to use CP-OFDM for both uplink and downlink in NR. Thus, CP-OFDM takes even a greater role in 5G NR compared with 4G LTE. 3GPP has settled on CP-OFDM and most proposals [49]. Table 4 shows the different waveforms used by all wireless network generations from G1 to the upcoming 5G.

TABLE 4. Waveforms used for different cellular networks generations.

Generation	Waveforms
G1	Analog - FDMA
G2	Digital FDMA/TDMA
G3	CDMA
G4	CP-OFDM for DL; SC-FDMA for UL
G5	CP-OFDM for DL; CP-OFDM and DFTs-OFDM for UL

B. 5G OPERATIONAL MODES MULTIPLEXING SCHEME

Achieving very low latency and high system reliability for mcMTC requires intelligent scheduling and multiplexing techniques between different 5G services. When there is a mcMTC service request, whether in the scheduling period or in the middle of eMBB or mMTC transmission, mcMTC packet should be transmitted immediately [22]. In other words, to support the mcMTC packet transmission, ongoing eMBB and mMTC packets should be stopped without notice. At the same time, the eMBB traffic should not be affected much more when maximizing the mcMTC outage capacity. Otherwise, the quality of eMBB and mMTC services will degrade severely. As illustrated in figure 8, when a transport block consisting of three codeblocks is transmitted for the eMBB service, each codeblock is mapped sequentially to the scheduled time-frequency resources. Thus, when the mcMTC service is initiated in the middle of the

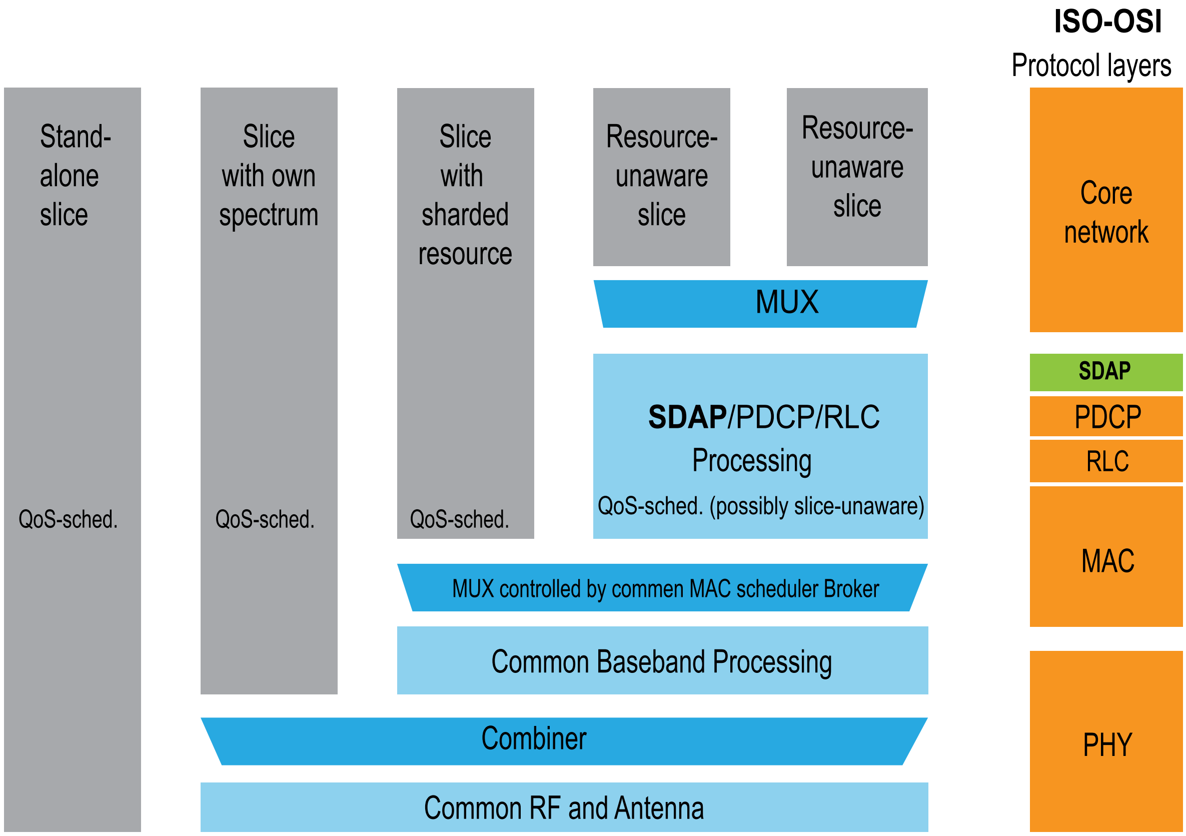

FIGURE 9. The changes of 5G on OSI protocol stack over 4G with the options for slice multiplexing.

eMBB transport block, some of the symbols in the third codeblock are replaced by the symbols of the mcMTC packet.

In the 3GPP NR, the coexistence of mcMTC with eMBB and mMTC is a serious concern to non-mcMTC services, so a proper mechanism to protect the ongoing services should be introduced [36]. However, it is also important to mention that full-Duplex transmission supports a bi-directional communications without time and frequency duplex by applying a simultaneous transmission and receiving on the same frequency at the same time. Furthermore, full-Duplex provides the potential solution to reduce the system delay and to double the system capacity. Therefore, full-Duplex provides a solution for most challenges of today's communications.

C. NETWORK ARCHITECTURE

Network architecture has been usually built for certain scenarios or use cases. As an example, LTE has been built for mobile data, and the Global System for Mobile communications (GSM) has been built mainly for voice. In general, the 3G and LTE mobile communication networks are planned to serve Human-type Communications (HTC) services, but it is complicated to supply all service characteristics of MTC. 5G needs to provide a broad range of connectivity to support extremely diverse and heterogeneous use cases, which basically works under the three 5G operation regions (eMBB, mMTC, and mcMTC) that requires a major change in network architecture. Moreover, various mcMTC use cases have different requirements for connectivity in terms of availability, latency, and reliability. However, It is expected that the 5G core network will utilize the ongoing evolution of SDN and NFV to provide a high level of flexibility and scalability when supporting 5G deployments [47]. One of the biggest projects which aims to provide a new network architecture design that could cope with the diverse and stringent 5G KPIs including network latency is the 5G NORMA project. With most of the initial implementation target determined in 3GPP as of [22] and [57]–[59], it looks that most part Radio Access Procotol overlaps with the current LTE and only small portions of new items are to be implemented as a new. The layers architecture of 5G network is almost the same as 4G based on Open Systems Interconnection (OSI) model. However, the focus of change is on PHY and Data Link layers where the 5G NR added a completely new sublayer called Service Data Adaptation Protocol (SDAP) on the top of Packet Data Convergence Protocol (PDCP) sublayer. The main role of SDAP is to apply a sophisticated Quality of Service (QoS) for each of data stream [60]. Figure 9 shows the changes of 5G on OSI model over 4G and the place of SDAP sublayer. It also depicts the slicing concept of resources, spectrum and standalone.

However, the network architecture has been studied very well in [61].

1) NETWORK SLICING

Network Slicing is one of the main concepts by which the challenges are addressed by enabling a higher degree of flexibility in the network such that each slice could be customized to meet the requirements of a certain use case in order to optimize the utilization of network resources [13], [23]. Network Slicing is a new concept to allow differentiated treatment depending on requirements of each customer [50]. The 5G network shall be designed in a way making it flexible enough for creating a virtual instance of an entire network for every different use case. Various customized virtual networks shall exist simultaneously and they shall not interfere with each other. As an example, a customized virtual network, which requires a very high throughput for video streaming, can co-exist with a customized virtual network for ultra-low latency autonomous vehicle control. Figure 10 depicts the slicing concept in 5G where it was not in 4G.

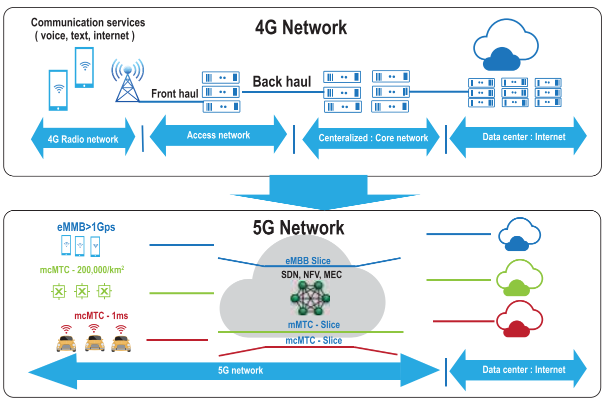

FIGURE 10. Network Transformation from 4G to 5G Networks.

2) PROGRAMMABLE NETWORKS

A flexible network shall be required to accommodate different performance requirements. A programmable infrastructure is created by software-defined functions i.e. the packets' path is not necessarily controlled by a fixed architecture and it could be programmed for optimizing latency.

Programmable network poses both scalability and latency challenges on the failure recovery process. Nokia has already offered transport solutions for all-IP covering completely integrated options for high capacity, high scalability, low latency and close synchronization meeting the requirements of a modern mobile broadband network, which are connectivity, backhaul and fronthaul requirements. Nokia researchers recommend providing programmable 5G multi-service architectures to address mcMTC requirements.

3) NETWORK RESILIENCY

Elements of a network should provide a high availability. Some core network elements are pooled and load balancing is used to guarantee that there is no service interruption when one or more core elements fail. The non-failing core elements can continue functioning, while the failing core element could be left to recover. Even when the backhaul becomes unavailable, a service shall continue to operate almost without being affected by utilizing a stand-alone operation mode [14].

4) EDGE CACHING AND COMPUTING

Pushing caching and computing resources to the edge reduce greatly latency (i.e. moving the application server and gateway closer to the radio). This provides the best and shortest path for routing traffic, which requires low latency and ensures continuity and seamless mobility at the same time. Services are not bound anymore to a single any-to-any IP connection, which evades unnecessary data forwarding to centralized mobility nodes because in this model devices communicate directly through local switching at the RAN level.

5) CLOUDRAN ARCHITECTURE

CloudRAN architecture is envisioned as one of the properties of 5G [62] to enable a flexible, and scalable architecture that can be adjusted to the needs of several use cases which run concurrently on the same infrastructure. This pattern also foreseen to significantly contribute to addressing mcMTC requirements. Providing flexibility in RAN configuration in terms of splitting the radio and baseband functionalities between central cloud and distributed entities will contribute delivery of mcMTC. In particular, latency critical MTC use cases can benefit from local computational power provided by applications running in the MEC since this reduces the physical and virtual communication distance [34]. However, network functions in 5G will increasingly be deployed based on cloud SDN and NFV. SDN is about the separation of the network control traffic (control plane) and the user specific traffic (data plane). Meanwhile, NFV is about virtualizing network functions and the functions that can run on a range of standard hardware [63]. CloudRAN, simplifies scaling and management of network infrastructure such as deep packet inspection engines and firewalls [64]. Examples of how flow-level modeling can be applied on SDN/NFV architectures may be found in [65]. However, SDN and NFV are new technology but not included in 3GPP technical specifications yet.

D. PACKET AND FRAME STRUCTURE

One of the most important objectives of 5G systems is supporting different packet lengths (1-10 MB) and various symbol numerologies (e.g., sampling rate, subcarrier spacing, symbol duration) in order to accommodate various deployment scenarios and service requirements [20]. The main aim of the design of mcMTC packet is minimizing the processing latency, which includes time for acquisition of channel information, receiving data symbols, decoding data packets, checking the existence of error, and scheduling information.

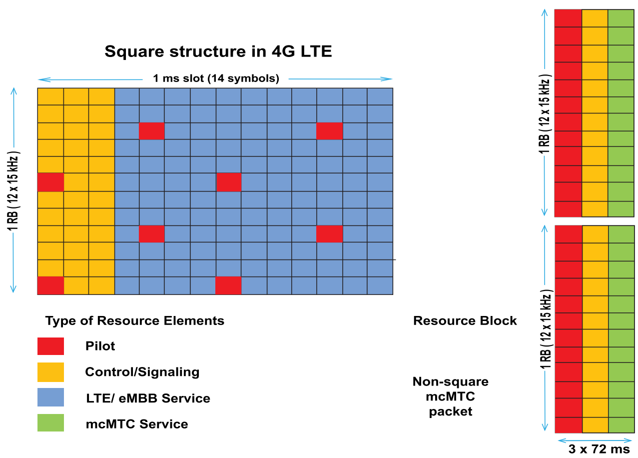

FIGURE 11. **LTE frame structure versus mcMTC packet [36].**

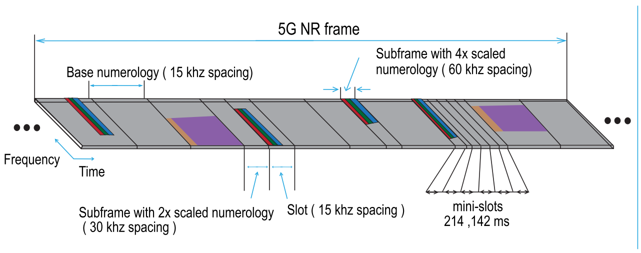

FIGURE 12. **5G frame structure.**

The frame structure is dependent on in-resource PHY layer control signaling, which follows the corresponding data transmission for each user. Comparison against the corresponding LTE design choices shows attractive benefits. In addition, enhancement of HARQ retransmissions and Flexible design entail a flexible duration of the ACK/NACK duration and configurability per user [66]. In LTE the $1ms$ TTI and $8ms$ waiting time at every retransmission results in degrading the E2E latency to more than 20 ms [14]. Hence, existing LTE frames cannot support the mcMTC service due to the structure of LTE resource blocks of $1ms$ TTI and 14 symbols size. Therefore, 5G requires a shorter frame and faster processing time at retransmissions to produce a lower latency. In Figure 11, the size of the 5G resource block is 3 symbols with $72ms$ TTI each, while in LTE the size of the resource block is 14 symbols and $1ms$ TTI. Therefore, the packet transmission time should be in the order of hundreds of microseconds to meet this stringent latency constraint [29].

In 4G systems, a rectangular-shaped packet structure is used to make an efficient utilization of the spectrum under time-frequency fading channel. In mcMTC systems, on the other hand, a non-rectangular packet is favored because employing this structure shall save time to receive the packet data. To reduce the latency further, [20] suggested that the three components of a packet (pilot, control, data part) should be grouped together and afterward transmitted sequentially. By performing this, data detection, control channel decoding, and pipelined processing of the channel estimation become possible. Furthermore, in the control part, employing a simple fast decoding scheme is preferred to relying on a time-consuming process that consists of channel decoding and blind searching.

As suggested in [38], the header and data should be enclosed into a single packet to increase the probability of successfully receiving the header and the data. This idea is very relevant to services that use short messages, such as mcMTC. This is due to the header size and data size that are of comparable size (i.e. the header cannot be sent sub-optimally by using repetition coding and a low header data rate). Popovski's solution could work in practice, but the paper does not provide simulation results or physical results to confirm the proposed idea. It also does not address the situation where the application is located at the cell edge, which may have little to no mobile Internet reception.

E. CONTROL CHANNEL DESIGN

The essential task of a control channel is to carry out the control signaling related to resource allocation for actual data transmission. Ensuring high reliability for the control channel is an important and crucial aspect of mcMTC. In LTE the protection was primarily focused on data, while ensuring high reliability for the control channel is a must. This can be done by a fairly balancing the design for both the UL and the DL control channels, such that on the DL. The Base Station (BS) can select the optimal modulation coding scheme (MCS) based on both Channel Quality Indicator $(CQI)^2$ report and remaining latency budget. However, [67] provided a design for control channel which trade-offs between providing ultra-Reliable and low-Latency service for mcMTC system.

F. COMMUNICATION DIVERSITY

In wireless communications, the most difficult case regarding performing high reliability is the Rayleigh fading channel. This is due to large fading dips in this channel. To improve the reliability and availability of signal detection and decoding, a higher-order diversity is essential. Therefore, diversity is considered to be one of the most significant techniques to achieve high reliability in a fading environment. In general, diversity is the capability to use the three variations which are namely time, frequency and/or space to exploit the channel diversity gains. Such that, the latency bounds dictated by the time-critical mcMTC applications limit the number of possible retransmissions. Moreover, the exploitation of frequency diversity would be costly due to the need for frequency resources having uncorrelated channel coefficients. Therefore, antenna diversity is identified as the major technique to overcome uncertainty and fading dips in the channel [11] and it is found beneficial in mcMTC deployments [9], [26].

In order to fully occupy the benefits of frequency diversity over the frequency domain that have uncorrelated channel coefficients, the coded bits are to be mapped on the resources.

Additionally, combining multiple diversity techniques such as frequency, antenna, and various forms of spatial diversity can offer extra gains to approach the high reliability target within stringent latency constraints [6]. Such that, antenna diversity uses more antennas to improve the quality and reliability of a wireless link, while frequency diversity allows receiving radio signals over multiple channels (different frequencies) or wide radio channel (wide frequency band) to reduce the effects of radio signal distortions such as signal fading.

G. PROACTIVE PACKET DROP

Packets can be dropped proactively by the transmitter when the channel is in a deep fade. Transmitter discards packets that cannot be transmitted even with the maximal transmit power. Similarly, packet drop can happen at the receiver when the maximum number of re-transmissions is reached. This is different than eMBB scenarios assuming infinite queue buffers. In this case either spatial diversity should be used or resources need to be increased. However, due to the substantial parity overhead, we observe that the robustness improvement scheme suffers from throughput loss (about 17 percent) over the reactive strategy [36].

H. CODING AND MODULATION

Redesign of physical channels provides a fast and reliable decoding implementation towards improving the latency, reliability and availability of signal detection and decoding mechanism. It is required that the receiver can start the processing as early as possible to do early channel estimation as well as fast decoding and demodulation. Therefore, the physical layer should be designed so that reference signals / pilots should be transmitted at the beginning of a TTI and time-domain interleaving of data should be avoided. Furthermore, coding that does not have an error floor (e.g., convolutional codes) is found as another prerequisite to ensure mcMTC [11]. In addition, channel coding techniques designed for data of NR should support info block size flexibility and codeword size flexibility. In addition, Channel coding techniques of NR should support both incremental redundancy and chase combining HARQ [50].

Recently, 3GPP has been selected LDPC and Polar codes for the eMBB data and control channels, respectively for 5G NR [68]. It is expected that LDPC codes, Polar codes, and their variants will continue their victory to be deployed in many other applications and will be included in other new standards in the future. Nevertheless, the design of such codes for mcMTC is still in its early stages due to the range of open issues waiting to be addressed.

1) CONVOLUTIONAL CODING SCHEME

Convolutional code is selected as the preferred Forward Error Correction (FEC) scheme in mcMTC use cases due to many reasons listed in [67]. In addition, the decoder is able to process the code block whilst it is being received, and thus it is able to obtain decoded bits with a very short delay. However, for better performance and manageable decoding complexity, the block codes are favorable for control channels which have short blocks lower than 10 bits. Therefore, redesign of PHY channels allows early channel estimation usage of convolutional codes (e.g., for data channels) and block codes (e.g., for control channels) [10], [11]. However, convolutional codes perform similarly to Turbo and LDPC codes for block lengths, which are up to a few hundred bits. On the contrary of convolutional codes, Turbo and LDPC codes might in specific configurations have an error floor, which makes these codes less efficient when BLER reaches very low levels. Considering latency, convolutional code decoding has a shorter delay than the iterative decoder, which is usually employed in Turbo and LDPC decoding. This is partially because of lower decoding complexity.

2) TURBO CODING SCHEME

Turbo codes have been used as FEC for data in several modern communication systems, such as LTE. Turbo code encoder is built using a parallel concatenation of two recursive systematic convolutional codes and the associated decoder, using a feedback decoding rule. Turbo codes have low error probability performance within a 1dB fraction from the Shannon limit and relatively low complexity which is about 10^{-6} BLER [52]. Although it is being used in LTE, it does not satisfy the performance requirements of 5G services for all the code rates and block lengths as the implementation complexity is too high for higher data rates. In addition, due to the error floor observed in this scheme. Turbo codes generally have a low encoding complexity and high decoding complexity.

3) LDPC CODING SCHEME

Since LDPC codes were rediscovered in 1993. LDPC codes have attracted growing interests in both academia and industry because it can significantly improve the performance of both wired and wireless communication systems. LDPC codes are linear codes and have a sparse parity check matrix consisting low density which have relatively simple and practical decoding algorithms. The iterative nature of LDPC decoding algorithms increases the accuracy in each iteration while the number of iterations is decided based on the requirement of the application. Compared with turbo codes, LDPC codes can achieve better performance and faster decoding such that LDPC codes have an excellent ability to achieve almost 95% of the theoretical limits of channel capacity [69]. In addition, LDPC codes have the ability to trade-off between latency, reliability performance which make it a good candidate for future communications supporting mcMTC. However, modern LDPC decoders work with soft decision algorithms which further enhance the decoder gain.

In eMBB the channel coding scheme for data is the flexible LDPC and is the same as the single channel coding scheme for all block sizes. In addition, the channel coding scheme is Polar Coding with exception to very small block lengths

TABLE 5. A comparison between the most popular channel coding schemes.

Channel Coding Scheme	Pros	Cons
Convolutional Code	- Convolutional code has a shorter delay than the iterative decoder because of its lower decoding complexity.	- Less error correction performance, therefore, it has a poor performance in supporting mcMTC requirements. - The throughput is less than the other codes due to the less BLER performance.
Turbo Codes	- Has low complexity encoders. - Suitable to simple rate adaptation. - Low complexity iterative decoding schemes. - Works well with long messages as its iterative decoders rely on exchanging extrinsic information between two constituent decoders. - Performance for short lengths is still comparable with other modern coding schemes. - Has a good throughput for medium block length and code rate around 0.75. - Is the channel coding schemes for the 3G and 4G mobile communication standards.	- Error floor makes it less efficient when the BLER is less than 10^{-9} - The computationally most challenging part of the Turbo decoder. - The power and area efficiency of Turbo decoders deteriorates very fast when increasing the block size.
LDPC	- Performs typically well for large blocks and high rates greater than 0.5. - Almost achieved 95% of theoretical channel capacity. - Has high throughput for long block lengths and high code rates. - Has been selected for a control channel coding in 5G.	- Error floor makes it less efficient when the BLER is less than 10^{-9} - Has a difficulty to support fine granularity of code rates and code lengths. - Across all considered message lengths, both codes have a gap of 1 to 2.5dB to the finite length bounds which can be especially important for short message lengths. - Performance in terms of BLER and throughput is poor for small blocks and high rates less than 0.5. Such code rate range can be the most common scenario in eMBB case.
Polar	- Has the best performance in terms of BLER and throughput for control channels that have short block lengths. - Low encoding algorithm complexity and manageable decoding complexity. - No error floor. - Almost achieved 95% of channel capacity. - Suitable to be used in a wide range of scenarios with diverse requirements. Due to its encoder ability to work with the different types of decoders. - It uses a Small-list decoder or Successive Cancellation decoder for low-power realization. Has been selected for a data channel coding in 5G.	- The performance of encoding long block size is less than LDPC codes. - Less performance decoding complexity. - The performance in terms of BLER and throughput for control channels that have long block lengths is less than LDPC codes.

where repetition/block coding is preferred. However, LDPC codes are currently being used in many communication systems such as 802.11n (Wi-Fi allowing MIMO) and 802.16e (Mobile WiMAX) etc. However, LDPC coding has been selected for a control channel coding in 5G [68].

4) POLAR CODING SCHEME

In coding theory, polar codes are considered as a big advancement. When the code block size is large enough, polar codes are able to perform Shannon capacity with a simple encoder and a simple Successive Cancellation (SC) decoder. Polar codes have raised much interest so that several research works have been mainly conducted on code design and decoding algorithm [70]. Polar codes outperform all the codes, which are currently used in the 4G systems, specifically for short code length [71]. Hence, polar codes are treated as great potentials for the FEC module in 5G air interface design [12]. Many variants of Polar codes have been proposed to improve the performance of Polar codes, some concatenated coding and combined decoding schemes are proposed. Such as CRC-concatenated Polar codes (CA-Polar Code) with single-parity-check code concatenation and multi CRC-concatenation Parity-check concatenated Polar code (PC-Polar Code). Furthermore, the performance of Polar Codes keeps improving along with increasing list size of the software Successive Cancellation List (SCL) decoder. The flexibility of SCL decoder enables to explore trade-offs between throughput, latency decoding performance [72].

5) COMBINATION OF POLAR, SC, AND CRC CODING SCHEME

It is essential that establishing a very low latency could be gained by using a simple encoder and fast decoding algorithms. Several performance simulations show that the concatenation between Polar codes and CRC and an adaptive SC decoder are able to overcome turbo/LDPC codes for short and moderate code block sizes [12]. The SC decoding algorithm is among the most important decoding algorithms and it is able to perform same as the optimal maximum-likelihood decoding algorithm with a list size of 32 for moderate code block sizes [73]. Reference [13] suggested that an advanced error control coding such as a concatenation of Polar codes with CRC codes could be applied to improve the reliability and promises to overcome turbo or LDPC codes. Also, reliability should be enhanced for both the control and data parts. For the control part, a new strategy, which does not depend on the CRC and channel decoding is required. For the data part, an advanced channel coding scheme that functions with the short-packet regime could be used (i.e., polar code) [20].

The study in [19] provides a method analyzing of channel capacity with finite block length called a Polyansky-Poor-Verdu (PPV). PPV has provided the trade-offs between delays, throughput, and reliability on Gaussian channels and fixed rate block codes, by introducing a new fundamental parameter called "channel dispersion"; this analysis shows that there is a severe capacity loss at short block-lengths. There are no known codes that achieve the PPV limit. LDPC codes and polar codes have been reported to achieve almost 95% of the PPV bound at BLERs as low as 10^{-7} for block lengths of a few hundred symbols [69]. However, their main drawback is the large decoding latency [74]. From the other size, Self-adaptive codes appear as a promising solution to mcMTC [74]. Self-adaptive codes, also known as rateless codes, can adapt the code rate to the channel variations by sending an exact amount of coded symbols needed for successful decoding. This self-adaptation does not require any CSI at the transmitter side, thus eliminating the channel estimation overhead and delay [75]. Table 5 summarizes the popular coding schemes and their contributions to mcMTC in 5G.

VI. CONCLUSION AND FUTURE WORK

The aim of this work is to investigate the recent works that have been conducted on mcMTC. This paper highlighted the challenges and future aspects toward 5G. It is anticipated that within the next two years, mobile networks will experience major changes in comparison to the current state. Transmission rate is expected to be higher, the number of users and associated machines are predicted to increase by 10-100 times, and the traffic volume is anticipated to increase up to 1000 times. It is important to take the entire field into account, as it provides new opportunities for service providers, network operators, and users in the value chain. Achieving low-latency and reliable communications have been mainly studied in the context of MTC systems on a theoretical level.

This article has outlined an overview of mcMTC main requirements, its enabling techniques include 5G air interface design, network architecture, packet and frame structure, diversity, and coding schemes. The goal is to design an air interface, which is resilient to diverse services, and to support mcMTC future applications.

The future work in this regard is to enhance existing channel coding schemes including, LDPC codes, Polar codes, LDPC convolutional codes and spatially-coupled LDPC codes, rateless codes, and their variants. This improvement should take into consideration the complexity, error rate, and hardware implementation. In addition, development trends and challenges for turbo codes is still a hot topic. Moreover, it is beneficial to investigate control overhead additionally incurred by TTI. In addition, investigation in MEC in order to bring the Evolved Packet Core (EPC) closer to the end devices for mcMTC applications is recommended. To further explore the beamforming strategies for control and data part and the reconfigurable mcMTC protocol. Next, study of advanced transceiver architecture to support dynamic numerology adaptation and simultaneous decoding is highly needed. Furthermore, investigating of deployment strategies of 5G functions based on SDN and NFY is recommended.

Finally, a trade-off between a high reliability and tight latency constraints while maintaining availability to fulfill the minimum requirements of mcMTC towards 5G is a bit challenging and needs further improvement, especially with the advanced of 5G second phase which is more related to mcMTC standardization in 5G NR.

REFERENCES

[1] *Service Requirements for Machine-Type Communications (MTC); Stage 1 (Release 10)*, document 3GPP V0.3.0, 2009.

[2] E. G. Ström, P. Popovski, and J. Sachs. (2015). "5G ultra-reliable vehicular communication." [Online]. Available: https://arxiv.org/abs/1510.01288

[3] *IMT Vision—Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond*, document ITU M.2083-0, 2015.

[4] A. Orsino *et al.*, "Effects of heterogeneous mobility on D2D- and drone-assisted mission-critical MTC in 5G," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 79–87, Feb. 2017.

[5] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai. (2017). "A survey on low latency towards 5G: RAN, core network and caching solutions." [Online]. Available: https://arxiv.org/abs/1708.02562

[6] 3GPP, "Feasibility study on new services and markets technology enablers for critical communications, release 14," 3GPP, Sophia Antipolis, France, Tech. Rep. 9, 2016.

[7] T. Wen and P. Y. Zhu, "5G: A technology vision," Huawei, Shenzhen, China, White Paper, 2013. [Online]. Available: https://www.huawei.com/5gwhitepaper

[8] H. Tullberg *et al.*, "The METIS 5G system concept: Meeting the 5G requirements," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 132–139, Dec. 2016.
[9] O. N. C. Yilmaz, Y.-P. E. Wang, N. A. Johansson, N. Brahmi, S. A. Ashraf, and J. Sachs, "Analysis of ultra-reliable and low-latency 5G communication for a factory automation use case," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 1190–1195.
[10] O. N. C. Yilmaz, "Ultra-reliable and low-latency 5G communication," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, 2016.
[11] N. A. Johansson, Y.-P. E. Wang, E. Eriksson, and M. Hessler, "Radio access for ultra-reliable and low-latency 5g communications," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 1184–1189.
[12] Huawei. (2015). *5G: New Air Interface and Radio Access Virtualization*. [Online]. Available: https://www.huawei.com/minisite/has2015/img/5g_radio_whitepaper.pdf
[13] Q. Zhang and F. H. P. Fitzek, "Mission critical IoT communication in 5G," in *Proc. Future Access Enablers Ubiquitous Intell. Infrastruct.* Springer, 2015, pp. 35–41.
[14] NOKIA. (2016). *5G for Mission Critical Communication*. [Online]. Available: http://www.hit.bme.hu/~jakab/edu/litr/5G/Nokia_5G_for_Mission_Critical_Communication_White_Paper.pdf
[15] A. Colazzo, R. Ferrari, and R. Lambiase, "Achieving low-latency communication in future wireless networks: The 5G norma approach," in *Proc. Eur. Conf. Netw. Commun*, 2016, pp. 1–5.
[16] S. Zhang, X. Xu, Y. Wu, and L. Lu, "5G: Towards energy-efficient, low-latency and high-reliable communications networks," in *Proc. IEEE Int. Conf. Commun. Syst. (ICCS)*, Nov. 2014, pp. 197–201.
[17] (2017). *Ultra Reliable Low Latency Communications*. [Online]. Available: http://urllc2017.executiveindustryevents.com/Event/
[18] M. G. Weiner, "Low-latency, high-reliability wireless networks for control applications," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, Berkeley, CA, USA, 2015.
[19] 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.
[20] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim. (2017). "Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects." [Online]. Available: https://arxiv.org/abs/1704.05565
[21] METIS2020. (2015). *Mobile and Wireless Communications Enablers for the Twenty-Twenty Information Society*. [Online]. Available: https://www.metis2020.com
[22] 3GPP, "Study on new radio access technology—Physical layer aspects (release 14)," Sophia Antipolis, France, Tech. Rep. 38.802, 2017.
[23] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi. (2018). "5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view." [Online]. Available: https://arxiv.org/abs/1804.05057
[24] N. Nikaein and S. Krea, "Latency for real-time machine-to-machine communication in LTE-based system architecture," in *Proc. 11th Eur. Wireless Conf. 2011-Sustain. Wireless Technol. (Eur. Wireless)*, 2011, pp. 1–6.
[25] A. Kumbhar, F. Koohifar, I. Güvenç, and B. Mueller, "A survey on legacy and emerging technologies for public safety communications," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 97–124, 1st Quart., 2017.
[26] N. Brahmi, O. N. C. Yilmaz, K. W. Helmersson, S. A. Ashraf, and J. Torsner, "Deployment strategies for ultra-reliable and low-latency communication in factory automation," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2015, pp. 1–6.
[27] H. Shariatmadari *et al.*, "Machine-type communications: Current status and future perspectives toward 5G systems," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 10–17, Sep. 2015.
[28] V. N. Swamy, P. Rigge, G. Ranade, A. Sahai, and B. Nikolic. (2018). "Network coding for real-time wireless communication for automation." [Online]. Available: https://arxiv.org/abs/1803.05143
[29] T. Z. Oo and C. S. Hong, "URLLC service for LTE radio access networks," in *Proc. Korean Inf. Sci. Soc.*, 2018, pp. 308–310.
[30] S. Parkvall, A. Furuskar, and E. Dahlman, "Evolution of LTE toward IMT-advanced," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 84–91, Feb. 2011.
[31] 3GPP, "Feasibility study on new services and markets technology enablers for critical communications," 3GPP, Sophia Antipolis, France, Tech. Rep. 9, 2016.
[32] 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, Sep. 2016.
[33] M. Keskinen. (2015). *Mobile Network Evolution*. [Online]. Available: http://www.oulu.fi/sites/default/files/content/Keskinen-5Glecture_0.pdf
[34] 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, pp. 70–78, Feb. 2017.
[35] M. Bennis, M. Debbah, and H. V. Poor. (2018). "Ultra-reliable and low-latency wireless communication: Tail, risk and scale." [Online]. Available: https://arxiv.org/abs/1801.01270
[36] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 124–130, Jun. 2018.
[37] Qualcomm. (2014). *LTE MTC: Optimizing LTE Advanced for Machine-Type Communications*. [Online]. Available: https://www.qualcomm.com/media/documents/files/lte-mtc-optimizing-lte-advanced-for-machine-type-communications.pdf
[38] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *Proc. 1st Int. Conf. 5G Ubiquitous Connectivity (5GU)*, 2014, pp. 146–151.
[39] H. D. Schotten, R. Sattiraju, D. Gozalvez Serrano, Z. Ren, and P. Fertl, "Availability indication as key enabler for ultra-reliable communication in 5G," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, 2014, pp. 1–5.
[40] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay trade-off in wireless networks," in *Proc. INFOCOM 23rd Annu. Joint Conf. IEEE Comput. Commun. Soc.*, vol. 1, Mar. 2004, p. 475.
[41] *Technical Specification Group Radio Access Network; Study on Scenarios and Requirements for Next Generation Access Technologies (Release 14)*, document 3GPP;TR38.913, 2016.
[42] T. Fehrenbach, R. Datta, B. Göktepe, T. Wirth, and C. Helge. (2018). "URLLC services in 5G—Low latency enhancements for LTE." [Online]. Available: https://arxiv.org/abs/1808.07034
[43] *Service Requirements for the 5G System, Version 16.0.0*, document 3GPP;TS22.261, 3GPP, Jun. 2017.
[44] *Minimum Requirements Related to Technical Performance for IMT2020 Radio Interface(s)*, document ITU-R M.2410-0, ITU-R, Nov. 2017.
[45] G. Fettweis, H. Boche, T. Wiegand, E. Zielinski, H. Schotten, P. Merz, S. Hirche, A. Festag, W. Häffner, M. Meyer, and E. Steinbach, "The tactile Internet-ITU-T technology watch report," Int. Telecommun. Union, Geneva, Switzerland, White Paper, 2014.
[46] J. Pilz, M. Mehlhose, T. Wirth, D. Wieruch, B. Holfeld, and T. Haustein, "A tactile internet demonstration: 1ms ultra low delay for wireless communications towards 5G," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2016, pp. 862–863.
[47] E. Dahlman *et al.*, "5G wireless access: Requirements and realization," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 42–47, Dec. 2014.
[48] C.-P. Li, J. Jiang, W. Chen, T. Ji, and J. Smee, "5G ultra-reliable and low-latency systems design," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2017, pp. 1–5.
[49] 3GPP, "Study on nr new radio access technology," 3GPP TSG RAN, Sophia Antipolis, France, Tech. Rep. RP-160671, 2016.
[50] 3GPP-TSG, "Study on new radio (NR) access technology," 3GPP, Sophia Antipolis, France, Tech. Rep. TR 38.912, 2017.
[51] A. A. Zaidi *et al.*, "Waveform and numerology to support 5G services and requirements," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 90–98, Nov. 2016.
[52] H. N. Gamage, "Waveforms and channel coding for 5G," M.S. thesis, Fac. Inf. Technol. Elect. Eng., Dept. Commun. Eng., Univ. Oulu, Oulu, Finland, 2017.
[53] G. Wunder *et al.*, "5GNOW: Non-orthogonal, asynchronous waveforms for future mobile applications," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 97–105, Feb. 2014.
[54] K. Au *et al.*, "Uplink contention based SCMA for 5G radio access," in *Proc. Globecom Workshops (GC Wkshps)*, 2014, pp. 900–905.
[55] A. Bayesteh, E. Yi, H. Nikopour, and H. Baligh, "Blind detection of SCMA for uplink grant-free multiple-access," in *Proc. 11th Int. Symp. Wireless Commun. Syst. (ISWCS)*, 2014, pp. 853–857.
[56] Y. Zhou, H. Luo, R. Li, and J. Wang, "A dynamic states reduction message passing algorithm for sparse code multiple access," in *Proc. Wireless Telecommun. Symp. (WTS)*, 2016, pp. 1–5.
[57] *Study on New Radio Access Technology: Radio Access Architecture and Interfaces*, document 3GPP 38.801, 2017.
[58] *Study on New Radio Access Technology, Radio Frequency (RF) and Co-Existence Aspects (Release 14)*, document 3GPP 38.803, 2017.
[59] *Study on New Radio Access Technology, Radio Interface Protocol Aspects (Release 14)*, document 3GPP 38.804, 2017.

[60] *NR and NG-RAN Overall Description, Release 15*, document 3GPP TSG RAN, TS3300, 2018.
[61] 5GPPP Architecture Working Group, "View on 5G architecture," 5GPPP, Heidelberg, Germany, White Paper, Dec. 2017.
[62] (2016). 5G Network Architecture: A High-Level Perspective. HUAWEI Technologies CO. Ltd. [Online]. Available: https://www.huawei.com/minisite/hwmbbf16/insights/5G-Nework-Architecture-Whitepaper-en.pdf
[63] Ericsson, "5G systems—Enabling the transformation of industry and society," Ericsson, Stockholm, Sweden, White Paper, UEN 284 23-3251, 2017. [Online]. Available: https://www.ericsson.com/assets/local/publications/white-papers/wp-5g-systems.pdf
[64] G. Mountaser, M. Condoluci, T. Mahmoodi, M. Dohler, and I. Mings, "Cloud-RAN in support of URLLC," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2017, pp. 1–6.
[65] K. Mahmood, A. Chilwan, O. Østerbø, and M. Jarschel, "Modelling of openflow-based software-defined networks: The multiple node case," *IET Netw.*, vol. 4, no. 5, pp. 278–284, 2015.
[66] K. I. Pedersen, G. Berardinelli, F. Frederiksen, P. Mogensen, and A. Szufarska, "A flexible 5G frame structure design for frequency-division duplex cases," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 53–59, Mar. 2016.
[67] S. A. Ashraf, F. Lindqvist, R. Baldemair, and B. Lindoff, "Control channel design trade-offs for ultra-reliable and low-latency communication system," in *Proc. IEEE GLOBECOM Workshops (GC WKSHPS)*, Dec. 2015, pp. 1–6.
[68] *5G; NR; Multiplexing and Channel Coding, Release 15*, document 3GPP TS38.212, Jul. 2018.
[69] 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.
[70] C. Leroux, I. Tal, A. Vardy, and W. J. Gross, "Hardware architectures for successive cancellation decoding of polar codes," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 1665–1668.
[71] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015.
[72] M. Léonardon, A. Cassagne, C. Leroux, C. Jégo, L.-P. Hamelin, and Y. Savaria. (2017). "Fast and flexible software polar list decoders." [Online]. Available: https://arxiv.org/abs/1710.08314
[73] M. Sybis, K. Wesolowski, K. Jayasinghe, V. Venkatasubramanian, and V. Vukadinovic, "Channel coding for ultra-reliable low-latency communication in 5G systems," in *Proc. IEEE 84th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2016, pp. 1–5.
[74] H. Chen *et al.* (2017). "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches." [Online]. Available: https://arxiv.org/abs/1709.00560
[75] M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Near-capacity adaptive analog fountain codes for wireless channels," *IEEE Commun. Lett.*, vol. 17, no. 12, pp. 2241–2244, Dec. 2013.

NAJIB AHMED MOHAMMED received the B.Sc. degree in computer science from Taiz University, Yemen, in 2007, and the M.Sc. degree in computer networks from University Putra Malaysia, in 2015. He is currently pursuing the Ph.D. degree in wireless networks with the Faculty of Computer Science and Information Technology, University of Malaya, Malaysia. He was as a Lecturer Assistant with IBB University, Yemen, from 2008 to 2013. During his master's degree, he has received a patent entitled Method for a Placement of the Internet Gateway in a Backbone Wireless Mesh Networks. His research interests are in LTE-A, 5G, and machine-type communications.

ALI MOHAMMED MANSOOR received the B.Sc. degree in computer science and distributed computing from Amman University, Jordan, in 2005, and the M.Sc. degree in computer science and distributed computing and the Ph.D. degree in communications and network engineering from University Putra Malaysia, Malaysia, in 2008 and 2014, respectively. Since 2009, he has been a Lecturer with the Department of Computer Engineer, Faculty of Engineering, Aden University, Yemen. He was involved in broadband networks at Wireless Communications Cluster, MIMOS Berhad, Malaysia. He is currently a Senior Lecturer with the Faculty of Computer Science and Information Technology, University of Malaya, Malaysia. His main research interests are wireless communication and networking, resource management and QoS, the Internet of Things (IoT), and emerging wireless technologies in 4G and 5G, such as edge computing, D2D communication, and LTE-advance networks.

RODINA BINTI AHMAD received the bachelor's degree and the master's degree in computer science, in 1984 and 1989, respectively, under the Petronas Scholarship in USA, and the Ph.D. degree in information systems management from the National University of Malaysia, in 2006. She started teaching at the Computer Science Department, University Technology Malaysia, in 1991. She moved to the University of Malaya to continue her teaching career, in 1993, where she has been involved in various research activities in the areas of requirements engineering and organizational analysis and modeling.