Transport Evolution for the Radio Access Network (RAN) of the Future - Invited

Philippe Sehier, Philippe Chanclou, Nihel Benzaoui, David Chen, Kimmo Kettunen, Mike Lemke, Yvan Pointurier, Pascal Dom

***Abstract*— 5G systems will introduce several new requirements impacting the mobile transport network. Understanding the impact of 5G systems on transport networks, both wireless and fixed, is essential to the successful introduction of 5G. This paper provides an overview of the ongoing work supporting the specification of open interfaces between the different 5G building blocks. It also presents several technologies that can facilitate effective transport networks, describes the architectural challenges in designing flexible 5G transport networks, and provides dimensioning guidelines.**

***Index Terms*— 5G transport; backhaul; fronthaul; functional split; radio access network (RAN)**

I. Introduction

One of the main design goals for 5G has been to introduce flexibility in mobile systems to support use cases with diverse requirements ranging from enhanced Mobile BroadBand (eMBB) to massive Machine-Type Communication (mMTC) to Ultra-Reliable Low-Latency Communication (URLLC) [27]. This flexibility opens up the possibility for new use cases for mobile communication and we are already seeing many new applications emerging for 5G that are driven by different market verticals.

High-reliability, low-latency mobile communication has, in particular, raised strong interest in the industrial domain e.g., in the context of the Industrial Internet of Things (IIoT) where many use cases have communication requirements that are more stringent in terms of latency and reliability than those found in traditional consumer IoT [1]. These challenging applications are often related to the control of cyber-physical systems (CPS), in which powerful and pervasive connectivity between machines, people and objects is an essential component. 5G has the potential to meet this need, making 5G an important enabler for these emerging industrial use cases. Flexible production, advanced mobile applications for workers, and mobile robots or autonomous vehicles on the shop floor are just a few examples.

In this article, we provide an overview of the use cases and their requirements that are truly specific to 5G, with a focus on those that are not addressable effectively with 4G (section II). We report the current state of standardization of 5G transport (Section III), and describe how the 5G RAN architecture is designed to meet the wide range of 5G requirements and deployment constraints (Section IV). We explain the challenges for optical transport solutions (Section V), and provide an overview of the optical transport technologies that can enable 5G (Section VI).

Manuscript received October 30, 2018.
Ph. S., YP, and NB are with Nokia Bell Labs, France.
Ph. C. is with Orange Labs in France
DC, KK, ML and PD are with Nokia Networks.

II. 5G Use case Requirements

Many of the use cases targeted by 5G imply new requirements. These stringent requirements are challenging and meeting them requires new solutions impacting the 5G radio as well as the transport segment of the communication network.

Figure 1 positions several 5G services according to their latency, reliability, and spectral-efficiency requirements. A first observation is that the services cover a large range of combinations. Additionally, note that not all services in the URLLC category have the most demanding requirements. It is therefore necessary to have solutions that allow flexible and independent adaptation of latency and reliability characteristics. Indeed, aligning all services on the most demanding set of requirements would result in a very low spectral efficiency.

Some examples of low-latency and high-reliability applications are mobile robots and remote control [2]. These new use cases typically impose additional requirements on 5G systems, such as a need for high-accuracy positioning for mobile robots or support of isochronous communication.

Today, most communication technologies used in the manufacturing industry are non-wireless due to a lack of suitable wireless technologies that are capable of meeting the strict requirements in industrial automation. The existing wireline technologies include, among others, a variety of dedicated Industrial Ethernet technologies (e.g., Sercos®, PROFINET® and EtherCAT®) and fieldbuses (e.g., PROFIBUS®, CC-Link® and CAN®) [3]. Mobile technologies such as 5G are crucial components to enable applications where mobility and untethered connectivity provide added value for industrial communications. Examples include mobile robots, and mobile control panels with safety functions and augmented reality (AR) solutions. The latter can be used to assist shop-floor workers in their tasks by providing additional online information such as step-by-step instructions or support from remote experts [3].

Furthermore, the IEEE has defined a standard technology called Time-Sensitive Networking (TSN) to provide deterministic and isochronous communication on standard Ethernet (this will be expanded on in Section III.D). Developing 5G technologies that are compatible with TSN is another important priority.

III. RELATED STANDARDIZATION ACTIVITIES

A. 3rd Generation Partnership Project (3GPP)

1) Overview

In 3GPP Release 16, work has started to extend the 5G specification to support TSN-type deterministic and isochronous communication as well as its integration with wired TSN networks. Below is a summary of the currently ongoing key activities.

In the 3GPP Service and System Aspects Technical specification group (TSG-SA), the SA1 subgroup is responsible for service and feature requirements applicable to 3GPP technologies. It is looking at various vertical-domain use cases including factory of the future, eHealth and smart city. It is also examining specific requirements on 5G to support a wide range of industrial applications (e.g., time synchronization, low end-to-end latency, service availability, and security) [3]. Additionally, SA1 is considering suggested requirements for Ethernet transport service, e.g., support for traffic filtering and prioritization, support for industrial communication with stable latency needs, and support for synchronization with IEEE 802.1AS [4].

3GPP SA2 is the subgroup responsible for the definition, evolution and maintenance of the overall 3GPP architecture. It has started to study technical enablers to support TSN and industrial-control use cases [5].

TSG-RAN, which has the responsibility for the definition of the RAN, has recently started a study item called "New Radio (NR) Industrial IoT". It focuses on the technical enablers in the RAN for: synchronization; deterministic and low-latency delivery; optimization for periodic traffic with very short cycle times; and high-reliability requirements. The RAN TSG also supports the definition of other functional elements in line with the new 5G requirements [6].

2) RAN aspects

Since the LTE era, 3GPP and other industry associations, such as the Next Generation Mobile Networks (NGMN) alliance, have tried to define open interfaces between some of the RAN functions. The goal was to open the ecosystem and offer more deployment flexibility.

A study within 3GPP identified all possible RAN functional splits, along with the pros and cons of each. Eight options were identified, as shown in Figure 2. At the end of this study, only two options were selected to be examined and specified in more detail. These two options are expected to be sufficient to satisfy most deployment cases. They can be described as follows:

- The High Layer Split (HLS) corresponds to option 2. The split is made between the Packet Data Convergence Protocol (PDCP) and the Radio Link Control (RLC) functions. This interface is similar to conventional backhaul in terms of throughput and latency requirements. 3GPP named this interface F1.
- The Low Layer Split (LLS) corresponds to options 7 and 8. Option 8 has been widely deployed since the introduction of 3G; the interface was specified by the Common Public Radio Interface (CPRI) industry forum. Option 7 includes several sub-options; work on the interface specification is being continued outside of 3GPP.

The characteristics of the LLS interfaces, as well as the organizations in which they are specified, are presented in Section III-C. LLS is characterized by low-latency requirements, and data rates that are generally higher than in HLS.

The 3GPP has also defined several functional entities interconnected by the interfaces presented above. Figure 3 depicts the interfaces that have been defined within the RAN for 4G and 5G architectures, as well as the RAN to core network (CN) interfaces. The RAN functional elements are:

- The Radio Unit (RU): Contains all lower physical-layer functions
- The Distributed Unit (DU): Contains the higher physical-layer functions, as well as the Medium Access Control (MAC) and RLC functions.
- The Centralized Unit (CU): Contains all RAN functions above RLC, and terminates inter-RAN interfaces (X2, Xn). The CU is subdivided into User Plane (UP) and Control Plane (CP) functions, interconnected by the E1 interface.

Functions can be deployed close to the cell site, or, on the contrary, centralized in the cloud. The placement choice depends on several criteria, the most important being the characteristics of the transport network, as well as those of the network elements (e.g., processing power and possible specialization for certain tasks).

There have been many specification efforts aimed at providing what is currently called a 'disaggregated RAN', combining elements from several commercial sources, implemented on platforms that are becoming more and more open. The possibilities offered by this specification framework are further discussed in Section IV.

3) F1 throughput assessment

The characteristics of the F1 interface are similar to backhaul. As a result, the same methodology can be used for capacity provisioning.

A widely used methodology for dimensioning the 4G/LTE backhaul capacity is described by the NGMN Alliance [7]. It specifies provisioning of "last mile" backhaul capacity based on the busy-time mean and peak backhaul traffic for single-cell and multi-cell sites. The most commonly used rule, described here for three cells, dimensions the site capacity to the sum of one peak rate plus twice the busy-time mean. Reference [7] generalizes this method to more than three cells. A sample calculation on a typical site configuration with three cells yields:

- Carrier size: 100 MHz, 8 layers, sub carrier spacing 30kHz resulting in 4.7 Gb/s peak downlink throughput per cell according to [8], table 5.1.1.1-1

- 64 Transmit/Receive (TRX) Massive Multiple-Input Multiple-Output (MIMO), resulting in an average cell rate of 2 Gb/s [8]

This results in a required link capacity for the site of:

4.7 + 2 x 2 = 8.7 Gb/s

A 10 Gb/s link could provide sufficient capacity but provisioning a 25 Gb/s link would leave headroom for sites supporting multiple radio access technologies, as well as for evolution to higher 5G capacities.

B. Common Public Radio Interface (CPRI/eCPRI)

CPRI is a cooperative industry initiative aimed at defining publicly available specifications for the key internal interfaces of radio base stations. CPRI Specification version 7.0 was published in October 2015 [9] and added support for 24G line-rate to the previously supported range of line-rates, including 10G. The CPRI specification covers split option 8. Single and multiple hops with chain, tree and ring topologies are supported. The specification supports GSM, WiMAX, UMTS, LTE and LTE-Advanced.

Three different information flows (user plane data, control and management plane data, and synchronization plane data) are carried by the interface. The specification covers layer 1 (the physical layer) and layer 2 (the data link layer. Layer 1 supports electrical as well as optical interfaces.

In August 2017, the CPRI forum released the first version of the eCPRI specification, with an update in January 2018. The scope of the eCPRI specification is to enable efficient and flexible radio data transmission via a packet based fronthaul transport network such as IP or Ethernet. eCPRI defines a protocol layer that provides various – mainly user-plane-data specific – services to the upper layers of the protocol stack.

C. XRAN / ORAN

1) History

The xRAN (extensible RAN) Forum [10] was founded in October 2016, with the goal of developing, standardizing and promoting an open alternative to the traditionally closed hardware-based RAN architectures currently available to operators. The forum was composed of both operators and vendors. The fronthaul working group within the forum was tasked with developing an open, interoperable fronthaul interface with multiple-vendor support.

In February 2018, the xRAN Forum and C-RAN Alliance announced their “intent to merge to form a world-wide, carrier-led effort to drive new levels of openness in the radio access network.” [28]. The resulting Open RAN (ORAN) Alliance intends to combine and extend the work of both C-RAN and xRAN.

In April 2018, the xRAN Forum published the first version of its interoperable fronthaul specification. This initial version covered the Control, User and Synchronization (CUS) fronthaul planes based on an intra-PHY (option 7) fronthaul split. The specification is “designed to allow a wide range of vendors to develop innovative, best-of-breed RRUs and BBUs for a wide range of deployment scenarios.” [10].

This initial version was followed in July 2018 by version 2.0 of the Control User and Synchronization (CUS) planes specification, and the addition of version 1.0 of the management-plane specification supporting this split [10].

The xRAN fronthaul specification provides:

- Efficient bandwidth scaling as a function of user throughput and spatial layers to address increasing bandwidth needs and massive MIMO deployments.
- Support for LTE and NR, with different RU product configurations, including massive MIMO beamforming antenna systems.
- Advanced receivers and co-ordination functions.
- Ethernet-based transport layer solutions.
- Extensible data models for management functions to simplify integration.

2) Interface description

The xRAN fronthaul specification defines the interface between the DU and the Radio Unit (RU). The specification is based on an intra-PHY (option 7) split, as shown in Figure 2**Error! Reference source not found.**.

The xRAN control and user interface is transported over Ethernet, using the eCPRI protocol. eCPRI provides a flexibly defined packet header and protocol structure. The xRAN interface specification adds a detailed definition of the supported eCPRI messages and the content of the messages to ensure inter-operability. Only a subset of the eCPRI defined messages are required to support the xRAN interface.

The eCPRI specification specifically states that the definition of synchronization or management plane protocols is not within the eCPRI specification scope. These are specified within xRAN based on various standards-based protocols. xRAN provides detailed specifications for the usage of these protocols to ensure vendor inter-operability.

The xRAN U-Plane packets carry frequency domain in-phase and quadrature (IQ) data based on the number of supported spatial layers. Scaling based on layers rather than on the TRX chain can greatly reduce the required data rate. The U-Plane data is segmented into groups of resource blocks (sections) sharing the same layer and beamforming characteristics.

The control plane provides scheduling and beamforming information to be applied to the U-Plane sections. The scheduling information is used to map sections to the respective sub-carriers within the symbol. This allows for the U-Plane to send only used resource blocks for any symbol, further reducing the required bandwidth. The control plane scheduling information can be sent every symbol, but is typically sent less frequently (e.g., per slot) to further reduce the require bandwidth. However, the engineering of fronthaul networks needs to carefully consider how to take advantage of statistical multiplexing gains to avoid over-subscription of the fronthaul interface.

The xRAN specifications include multiple compression methods that can be applied to the control and user planes. The xRAN support for compression does not only allow a number of different methods, but also provides a relatively dynamic approach to the application of compression. Application of compression methods can be time varying, and can be applied differently to different layers.

Ethernet, unlike CPRI, provides an asynchronous fronthaul interface. As such, the exact delivery timing of the data across the fronthaul cannot be dictated. Instead, xRAN fronthaul relies on common time references (e.g., GPS) between the DU and RU. In the downlink, the DU includes the air interface timing for data in the xRAN U-Plane packets. This timing is based on frame, sub-frame, slot, and symbol. The RU uses this symbol information in the user plane packets to determine the precise over-the-air timing. Similarly, in the uplink, the RU adds the frame, sub-frame, slot, and symbol to the received data before sending to the DU. The exchange of control and user plane packets between DU and RU must account not only for the minimum transport latency, but also the maximum variation in latency (e.g., jitter) resulting in the maximum transport latency. Packets for a layer/ symbol must be aggregated at the receiver over a period of time allowing for the maximum variation. This period is referred to as the receive window.

3) Throughput assessment

The xRAN specification defines several control plane and user plane formats. This section presents a peak-rate evaluation of the interface in a typical configuration and compares it to that of the CPRI interface. The configuration considered is:

- Downlink direction for one cell
- Carrier bandwidth: 100 MHz
- Maximum of 16 co-scheduled devices or layers
- Most aggressive data compression method with 8 bits per symbol
- Ethernet payload maximum size: 1500 bytes

The dominant flows on the interface are the user plane and the control plane. With respect to the control plane, the main flow consists of the precoding coefficients. We assume that sets of precoding coefficients are sent every symbol time for each layer. The total peak throughput is the sum of the user plane and control plane rates:

- Peak user-plane throughput: 12.5 Gb/s
- Control Plane throughput: 0.6 Gb/s
- Total: 13.1 Gb/s

Details of the derivation are provided in Appendix A. It is important to note that the throughput scales with the number of layers, whereas the conventional CPRI throughput scales with the number of TRX radio chains. In a configuration identical to that assumed for the evaluation above, and for 64 TRX, the CPRI throughput requirements would be almost 20 times higher:

- CPRI sample rate: 3.7 Gb/s
- With overheads 16/15 x 66/64: 4.1Gb/s
- Total: 259.5 Gb/s for 64 TRX

It is well known in mobile communications that the average cell throughput is far below the peak. Sizing transport at peak would lead to a significant overestimation of the transport resources needs. Sizing methods similar to those described for the F1 interface may be used, but stringent latency requirements may require more capacity to limit the effect of jitter created by queuing. This consideration does not apply to the legacy CPRI interface for which the bit rate is constant and sized according to the peak rate.

D. IEEE

1) IEEE 802.1 Time-Sensitive Networks (TSN)

The IEEE 802.1 TSN Task Group (TG) was formed to address speed, determinism and dynamics in industrial networks (specifically, in the context of audio/video bridging standard IEEE 802.1BA [24]). TSN mechanisms to provide deterministic services through IEEE 802 networks, i.e., guaranteed packet transport with low latency, bounded jitter, and low packet loss, are being standardized in this task group. Standard Ethernet did not have any deterministic capability prior to the IEEE 802.1 TSN standards. TSN aims at providing a timely and predictable Ethernet transport for time-sensitive applications in multiple industries, e.g., cellular, automotive, aerospace, manufacturing, utilities, etc.

IEEE 802.1 TSN encompasses an evolving set of standards specifying time synchronization, traffic scheduling and shaping, as well as path selection/reservation and fault tolerance.

a) IEEE 802.1CM

802.1CM, a collaborative effort of CPRI and the IEEE 802.1 WG, specifies standard TSN profiles for fronthaul [11]. A TSN profile is a set of feature and option selections specifying aspects of bridge operation and the configuration guideline. 802.1CM profiles illustrate how to meet the fronthaul requirements in an Ethernet network. In IEEE 802.1CM-2018, two fronthaul profiles are specified. Profile A requires strict priority for user data (i.e., IQ data) mapping it to a high-priority traffic class, whereas control and management data is mapped to a lower priority traffic class. Strict priority, standardized in IEEE 802.1Q-1998, is a common quality-of-service (QoS) differentiation mechanism where the worst-case delay corresponds to the duration of a best-effort maximum transmission unit (MTU)-sized packet. Profile B leverages the frame preemption features in TSN (802.1Qbu [14] and 802.3br [25]) allowing for low priority traffic to be preempted to give preferential treatment to fronthaul traffic as high-priority express traffic.

b) IEEE 802.1AS / AS-Rev

TSN implies that the network nodes and hosts implementing TSN must share a common and accurate time-of-day. IEEE 1588 Precision Time Protocol (PTP) [26] is used to maintain a common sense of time and the PTP profiles chosen to work with TSN are IEEE 802.1AS and IEEE 802.1AS-Rev. These enable stations attached to bridged LANs to meet the respective jitter, wander, and time synchronization requirements for time-sensitive applications. This includes applications that involve multiple streams delivered to multiple endpoints. To facilitate the widespread use of bridged LANs for these applications, synchronization information is one of the components needed at each network element where time-sensitive application data is mapped or de-mapped or a time-sensitive function is performed. The implementation of 802.1AS-2011 (and in the near future 802.1AS-Rev) by all network elements is required for TSN.

c) IEEE 802.1Qbu Frame Preemption

IEEE 802.1Qbu [12] defines a preemption mechanism that suspends the transmission of a lower priority frame to allow one or more “express” frames to be transmitted before transmission of the preemptable frame is resumed. It enables minimal delay of the deterministic (express) traffic when mixed with best-effort (preemptable) traffic on an Ethernet port that supports frame preemption. By disrupting the transmission of best-effort packets when a deterministic high-priority packet arrives, worst-case packet delay is minimized for the latter. A limited amount of jitter or Packet Delay Variation (PDV) still may occur for express packets because preemption is only performed if at least 60 bytes of the preemptable frame have been transmitted and at least 64 bytes (including the frame CRC) remain to be transmitted. An issue with the preemption mechanism is that, except for the first fragment, the other packet fragments do not contain MAC-address headers. Since 802.1Qbu works hop-by-hop, fragmenting the best-effort packets and reassembling these at the next hop, preemption may only be activated in networks with bridges supporting the IEEE 802.1Qbu and IEEE 802.3br standards.

d) IEEE 802.1Qbv Enhancement for Scheduled Traffic

Frame preemption gives a strict priority to a traffic class, such that best-effort traffic interferes minimally with real-time traffic. Only one class of service is deterministic. IEEE 802.1Qbv [13] goes further by defining and scheduling traffic to guarantee delivery of critical traffic with known latency and bounded jitter. 802.1Qbv defines how a set of queues can be served by a round-robin mechanism. 802.1Qbv allows each of the queues to be served within timeslots, one-by-one in a cycle, and it schedules one or more packets in bursts from each of the queues into designated time-slots. Each 802.1Qbv bridge port runs a synchronized, repeating schedule that turns on and off each of the set of queues with up to nanosecond timescale precision.

Although both latency and jitter are reduced to a few microseconds for a relatively simple topology with 802.1Qbv [20], scalability is limited in both number of classes of service and dynamics, as it takes on the order of seconds to schedule each flow for the schedulers proposed in the literature [20, 21].

e) Other Considerations

Whether, and under which conditions, scheduling of fronthaul traffic is required, is still part of an ongoing debate. An early paper concluded that frame preemption alone was not sufficient to meet the CPRI requirement, but that Ethernet with enhancements for scheduled traffic (802.1Qbv) might be sufficient to meet the jitter constraint [14]. However, more recently, a network concept built on top of Ethernet and aimed at providing determinism for a class of traffic even in the presence of best-effort traffic (similar to 801.1Qbu in terms of functionality) called Fusion was introduced [15]. Fusion is able to insert best-effort traffic between real-time packets at the expense of a fixed additional latency at each node.

E. Metro Ethernet Forum (MEF)

The MEF Forum has a long history in providing Ethernet service definitions and certification for mobile backhaul supporting 2G through LTE. In January 2018, MEF approved the MEF 22.3 Mobile Backhaul Implementation Agreement [29]. The MEF 5G project extends this work by expanding into fronthaul and leveraging developments in IEEE 802.1 TSN, increased data rates and network technologies to meet the latency, isolation and capacity requirements of 5G and network slicing. The work on 5G transport is pursued via a 5G-oriented MEF Implementation Agreement. MEF has been working closely with IEEE 802.1CM and the CPRI Initiative.

As a secondary scope, MEF is examining the support of all 5G transport requirements (i.e., multiple instances of mobile backhaul and/or fronthaul) over the same Carrier Ethernet network or service (e.g., with traffic separation, QoS, etc.).

MEF Services are agnostic to the underlying transport, as provisioning 5G as the underlying layer for Ethernet and IP Services will be facilitated by native support in 3GPP-based 5G networks.

F. IETF Deterministic Networking (DetNet)

A DetNet working group was created in the Internet Engineering Task Force (IETF) in 2015. IETF DetNet is dedicated to expanding TSN beyond Ethernet bridges to include routers. The techniques developed in TSN can be extended to routed data streams. It also has a goal to scale up the TSN techniques so that they work in networks larger than those supported by Ethernet bridges.

Recently, there is a goal to integrate TSN and DetNet as the methods needed to ensure time sensitiveness of a flow are equally applicable to bridges, routers, label switches, hosts, etc. These methods should be available to both TSN and DetNet and only the traffic class selection differs (L2 priority vs. LSP priority vs. DSCP, etc.).

IV. 5G Architecture

A. Disaggregated RAN

The identification of the functional blocks in the RAN and the specification of interconnection interfaces make it possible to choose a wide range of deployment options.

The individual functional entities RU, DU, CU-UP and CU-CP may be placed at different physical locations according to operator requirements, physical site constraints, transport network topology, latency and capacity limitations, as well as compute resource availability or specialization.

Figure 4 presents a selection of example functional placement options based on the assumption that the RAN may have functionality placed at:

- Cell site

- Aggregation site (intermediate site, traditionally used for transport aggregation and may be used to host legacy baseband unit (BBU) hoteling)
- Edge site (most centralized site in RAN)

The leftmost option in the figure corresponds to the centralized RAN, in which all processing functions are co-located, except for the RUs which are, by necessity, at the cell sites. The rightmost option corresponds to the conventional deployment mode in which all of the functional elements are co-located at the radio site. The dual split RAN option combines centralized processing for services which are not highly latency sensitive, with cell-site processing for latency-sensitive services. In this option, the application servers for latency-sensitive services are located at the cell site.

Beyond flexible placement, described above, dynamic placement makes it possible to ensure load balancing, and provides failover mechanisms, by transferring the load on active processing elements in case one of them fails. The limited transport capacity and latency capabilities are generally the bottleneck for such features.

B. Practical architectures

This section proposes practical network architectures that allow the connection of a 5G antenna site to a Point-of-Presence (PoP) [16] that hosts network and service functions (e.g., wireless and wireline customer aggregation, content storage and servers, information technology infrastructure equipment). PoPs can be close to the users (i.e., at a central office (CO)) or centralized at fewer locations in pre-aggregation (primary loop or mesh topology to collect CO traffic) or aggregation (secondary loop or mesh to collect edge-node traffic) networks nodes. Three abstraction levels of the 5G transport architecture correspond to the associated planes for the networks, access terminations (fiber and radio), and users. This partition is used in Figs. 5 and 6 in different contexts to highlight the complexity of interactions between topology (Fig. 5.(a)), RAN decomposition and equipment location (Fig. 5.(b)), and end-to-end service latency (Fig. 6). In Fig. 5(a), we show the relationship between the three abstraction levels: users in medium and high density areas; fiber and radio access network terminations (e.g., passive optical fiber distribution cabinets and antenna sites, respectively); and several backhaul network segments. For this last level, network segments interconnect a fiber-to-the-home (FTTH) CO with pre-aggregation and aggregation nodes (i.e., main and core PoP, respectively) using different transport solutions. In the fiber and radio access network terminations level, the antenna site is connected with an optical fiber through the fiber distribution cabinet based on one of the following:

- a dedicated optical fiber used as dark fiber or a medium for Ethernet equipment
- a shared fiber using time division multiplexing passive optical network (TDM-PON) technology
- a shared fiber using wavelength division multiplexing (WDM)

Figure 5.b illustrates the same three abstraction levels, with potential location of 5G functions based on 3GPP RAN decomposition (cf. Figs. 3 and 4). RUs are located at the antenna site. The evolved packet core (EPC) or 5G core (5GC) functions are located at core PoPs. The placement of DUs and CUs within the CO or PoP is a function of centralization and virtualization of RAN features (typically virtual CU).

To address the challenge of providing low latency, we need to consider the 5G services levels. The 3GPP has defined five 5G Services Groups [30, 31], reproduced in the table below:

SG	**E2E latency**	**Description**
SG1	> 50ms	Conversational voice & video, Real time gaming, V2X, process automation – remote control & monitoring, IMS signaling.
SG2	≈ 25ms	Electricity distribution medium voltage
SG3	≈ 10ms	Discrete automation, low latency eMBB, Augmented reality, Intelligent transport systems
SG4	≈ 5ms	Electricity distribution – high voltage, remote control
SG5	≈ 0.5ms	Tactile interaction

Figure 6 illustrates the latency distribution for each SG, among the three abstraction levels enumerated above. In Fig. 6, the placement of the RUs, DUs, and CUs are not key design choices because we consider end-to-end latency values, with these three functional entities being intermediary points. The authors propose to allocate the latency budgets for each of the SGs as follows:

For SG1, the authors propose to partition the target latency to 4 ms between the user device and RAN equipment, and 46 ms to reach the content server which hosts the applications (APPs) and possibly the mobile edge computing (MEC) entity. Depending on the RAN decomposition, this latency distribution could be re-balanced. For SG3, the authors propose 0.5 ms and 4 ms for the first segment for support of URLLC or eMBB, respectively. For SGs 2, 4 and 5, we consider latency values based on URLLC for the first segment. The most stringent service group (SG5) is tactile interaction with 0.5 ms latency. For this scenario, the content server and RAN decomposition could be located next to the users or lightly centralized (campus scenario) due to the negligible fiber latency (5 µs/km) for either the low or high-layer RAN split.

The choice of an implementation considering a specific RU, DU and CU placement to guarantee an end-to-end latency is a function of network topology, density and transport equipment performance.

V. Transport challenges

A. Capacity at the lowest cost

Transport networks are typically designed with capacity in mind. Today's long-haul circuit-switched transport networks typically support dozens (e.g., 80) of WDM channels, with each channel carrying hundreds of Gb/s over thousands of kilometers (for example, 100 Gb/s over 4200 km, 200 Gb/s over 2500 km, 500 Gb/s over 200 km [17]) without electrical regeneration. With electrical regeneration, which drives costs up, transmission distances become virtually unlimited. The relatively high cost of standard point-to-point circuit switched networks can be, however, incompatible with the lower cost point of access segments (e.g., the costs of an access network are amortized over significantly fewer entities than the costs of a long-haul network). Thus, PONs, which are designed for the access segment, carry less data (usually tens of Gb/s) over only a few wavelengths. In either case, transport networks are designed to achieve very low Bit Error Rate (BER), typically at most 10^{-12} or 10^{-15} such that data entering the network (after switching) is guaranteed to exit an optical link error-free with very high probability. This is compliant with frame-loss rate targets of 10^{-7} for 5G fronthaul [9].

B. Latency

Latency in optical transport networks is usually driven by the propagation delay in fiber (5 µs/km). In typical long-haul networks based on Reconfigurable Optical Add-Drop Multiplexers (ROADMs) and with point-to-point traffic, fiber propagation is the main latency component, followed by Forward Error Correction (FEC) [18]. If electrical switching is present (e.g., in an IP router), queuing latency is incurred as well. Point-to-multipoint (downstream) or multipoint-to-point (upstream) networks such as PONs require a multi-user access scheme. Thus these networks require a MAC to avoid data collision, resulting in additional delay. Overall, it is generally accepted that distances between antennas and processing units do not need to be larger than 10-20 km, corresponding to a bidirectional propagation latency of 100-200 µs. Over such distances, FEC and queuing delays may no longer be negligible relative to the propagation delay.

In addition, as noted earlier, for network elements (including base stations), a synchronization protocol such as IEEE 1588/PTP is needed. Asymmetries in links lengths or even within network equipment make the required timing precision of a few hundreds of nanoseconds difficult to achieve.

C. Jitter

A jitter constraint of +/-130 ns, including +/-100 ns for transport is generally accepted [19]. However, positioning, as well as other TSN-related services, may require lower jitter excursion. Optical circuit switching techniques that dedicate a full wavelength to a point-to-point service enable an essentially jitterless (ns-timescale) connection. However, multiplexing in the time domain is often used to improve the utilization of the optical spectrum. Such multiplexing can be through packet switching or Time Division Multiplexing (TDM). Packet switching is the technology best suited for highly dynamic environments as data is switched per-packet, with a typical packet duration of only a few ns at 10 or 100 Gb/s. However, packet switching incurs stochastic queuing at the switching nodes, which results in jitter that accumulates at each node. TDM can implement virtual circuit switching by allowing the resources to be reserved end-to-end at a sub-wavelength granularity within a network or even across networks segments. Jitter accumulation is then prevented, and jitter is limited to the insertion node and can be compensated at the reception node.

VI. Transport Solutions

A. Ethernet

Standard Ethernet is a very mature switching technology relying on packet switching to efficiently use the underlying physical medium. Ethernet scales to multi-Tb/s or even Pb/s networks inside data centers but supports only limited QoS functions. In particular, latency determinism is impossible to achieve without extensions, some of which are described below. It should be noted that even use of an over-dimensioned Ethernet network will not offer guarantees of any kind as contention is always possible with Ethernet, and results in jitter.

B. Ethernet with Time Sensitive Networking (TSN)

Employing Ethernet in mobile fronthaul poses a new level of performance requirements, especially for delay, delay variation, packet loss, and reliability parameters. Delay requirements are challenging for Ethernet as it was not originally designed for delay-sensitive networks or real-time applications. The Ethernet standard was therefore extended through the TSN effort, as described earlier (in particular, in 802.1CM; see Section III.D). The maximum end-to-end one-way delay for fronthaul is set at 100 µs, including fiber and Ethernet bridge delay. Jitter control is limited to a small number of service classes even with the TSN extensions, e.g., two classes with 802.1Qbu [12] and eight with 802.1Qbv [13].

C. Passive Optical Networks (PONs)

TDM-PONs were developed for FTTH as a low-cost multiple-access optical tree topology network. TDM-PON technologies use a scheduled broadcast to all Optical Network Units (ONUs) in the downstream direction. In the upstream direction, a Time-Division Multiple Access (TDMA) scheme is used in concert with Dynamic Bandwidth Allocation (DBA) in 125 µs frames. Worldwide mass deployments are currently using Gigabit PON (GPON) (2488 Mb/s downstream, 1248 Mb/s upstream) defined by ITU-T G.984 series [33]. The high cost of point-to-point fibers has led operators to increasingly use PON solutions to cope with the need for increasing capacity and densification of their networks. GPON was also successfully deployed for 4G small cell, where a single GPON is shared by several fiber access points and statistical multiplexing is adopted to converge wireline and wireless services.

Questions remain whether shared-bandwidth TDM-PONs can meet rising bandwidth demand. Deployment of 10 Gb/s class PONs are currently ramping up: e.g., 10G EPON with 1 or 10 Gb/s upstream, and 10 Gigabit Symmetrical XGS-PON (ITU-T G.9807.1) [32]. Capacity expansion via multiple wavelengths was proposed for future PONs. A point-to-point WDM over a point-to-multipoint tree topology, named "WDM-PON", dedicates a full wavelength for each cell site, resulting in a network of point-to-point dedicated links – highly deterministic, static and over-dimensioned. These three drawbacks are removed by adding TDM to WDM, resulting in TWDM-PON, standardized for example in NG-PON2 defined in ITU-T G.989 [34]. NG-PON2 offers up to 4 channels of 10 Gb/s capacity downstream and upstream in TWDM-PON mode.

With statistical multiplexing, transport requirements can be reduced; it was shown in ITU-T G.SUP.5GP that 10G PON has adequate capacity for 5G F1 interface transport. Transport slicing can be implemented through resource reservation (in time and/or wavelength.)

WDM PON technology is being considered for CPRI fronthaul transport as it provides low-latency transport. New LLS based architectures also require very low latency. With TDMA upstream, PONs typically have larger latency, on the order of 2 ms, due to the DBA, which allows for statistical multiplexing. With TDM-PONs, the latency can be driven down to 100 µs per direction (this includes a few km of propagation), and jitter can be virtually removed by using a fixed bandwidth allocation (FBA) scheduler, rather than a DBA scheduler [22]. Of course, using FBA offsets the benefits of statistical multiplexing and reduces the dynamics of the network, which is then essentially static. This problem is being addressed by a "cooperative DBA" (CO-DBA) scheme whereby the 5G DU scheduling process informs the OLT precisely when an RU requires resources. Efforts to standardize the CO-DBA protocol have begun in ITU-T SG 15 Q2. Furthermore, the conventional ranging process (OLT – ONU distance measurement) requires a "quiet" window on an order of magnitude of 100 µs, during which ONUs must buffer upstream traffic. Alternative methods to perform ranging must be investigated to reduce this added delay. Additionally, TDM-PON technologies will need to support the key performance indicators as laid out by TSN (see section above).

D. Optical Transport Network (OTN) and FlexE

Optical Transport Network (OTN) enables the multiplexing of several circuits in the time and wavelength domain over an optical transport infrastructure. OTN is designed to scale to several hundreds of Gb/s per wavelength, with up to 100 wavelengths per link. However, each circuit is statically allocated sub-wavelength capacity through time division multiplexing. OTN supports any mesh network topology, and advanced resilience mechanisms, and latency is typically dominated by propagation delays. However, OTN typically requires a FEC module, which can consume up to 10 µs of latency, depending on the data rate, type of FEC, etc. [18]. Slicing is enabled via time and wavelength reservation and static allocation of resources preventing support of highly time-varying traffic and lack of fast reconfigurability in general.

Indeed, while the electronic switching matrix in OTN can provide fast switching (i.e., sub-second), OTN is often used in combination with ROADMs to take advantage of optical bypass so that the number of line cards deployed in the network can be reduced. In such a configuration, OTN provides a limited amount of network dynamics – service establishment can take seconds or minutes due to the physics of the underlying components (such as laser tuning or ROADM reconfiguration).

FlexE [23] is a connection-oriented technology that implements TDM on top of an Ethernet network (and which in turn can be mapped to an OTN infrastructure for transport), adds additional features such as bandwidth allocation with 5 Gb/s granularity, bonding of channels, sub-rating of links, etc. It is conceptually similar to OTN and brings similar advantages and drawbacks. In particular, FlexE is fully scheduled with static allocation and does not support best-effort traffic other than through static scheduling.

VII. Conclusion

Regardless of the architecture, currently envisioned 5G networks already have placed stringent requirements on transport networks, both in terms of capacity and latency/jitter. These requirements, together with mature optical transport technologies that are able to meet the current specifications, were reviewed in this paper.

The 3GPP has developed a flexible 5G RAN architecture. The specification of building blocks and related interconnection interfaces make it possible to choose a wide range of deployment options, in a more open ecosystem. This evolution provides the flexibility to deploy 5G networks over existing transport networks to provide basic services. However, future applications will introduce more stringent requirements, and current technologies may encounter capacity, latency/jitter or scalability hurdles. Such issues are well understood by standardization bodies. Current research is addressing these challenges, and the industry is active in developing solutions, but more work is needed to convert those efforts into commercial products.

Appendix

A. xRAN interface peak throughput assessment:

We explain how the xRAN interface rates presented in III.C.3) were obtained. The assumptions are listed below:

- Carrier bandwidth: 100 MHz, 273 physical resource blocks (PRBs), Sub Carrier Spacing = 30kHz
- 64 TRX massive MIMO antenna
- Class B RU (precoding done in the RU)
- Maximum of 16 layers
- Maximum of 16 frequency-division multiplexed (FDM) users per TTI (or sub-bands)
- Symbol index transmitted: 8 bits per symbol to support up to 256 QAM
- Precoding coefficient sets are sent on the interface every transmission time interval (TTI) for all user equipment (UE) scheduled in that TTI. Each complex precoding coefficient is coded on 8+8 bits.
- Maximum size of the Ethernet payload: 1500

bytes

1) User plane (UP)

Each packet includes several headers: Ethernet, radio transport (e.g., eCPRI header) as well as an application header (defined in the xRAN specification). There are up to 256 UEs co-scheduled in each TTI, each supported by a dedicated segment. Due to the payload size limitation, several packets are needed to transport the user plane data of one symbol. The parameters used for the evaluation are listed in the table below:

Ethernet overhead (byte)	A	42
Radio transport header (byte)	B	8
Application header (byte)	C	4
Per segment header UP (byte)	D	4
Max ethernet payload (byte)	E	1500
Max. number of layers	F	16
TTI length (ms)	G	0.5
Max number of PRBs	H	273
Max number of FDM users	I	16
Sample size (byte)	J	1

The user plane throughput is derived as follows:

Section size (bytes)	K	208	$D + 12.J.\lceil H/I \rceil$
# sections in a Eth. Packet	L	7	$\lfloor (E - C - B)/K \rfloor$
#sections / symbol	M	256	$F.I$
Number of symbols/s	N	28000	$(14/G).1000$
# packets/second	O	1E+06	$\lceil M/L \rceil.N$
Packet size (bytes)	P	1510.0	$A + B + C + L.K$
Peak total throughput UP (Gb/s)		**12.5**	$O.P.8.10^{-9}$

2) Control Plane (CP)

The dominant flow for the control plane consists of the precoding coefficients. In a similar way to the user plane, the control plane information is organized in segments. The segments related to one precoding set needs to be fragmented over several packets due to the payload size limitation. The control plane throughput can be derived as indicated in the table below:

FCP per segment overhead (byte)	Y	12	
Number of TRX	X	64	
Coefficients quantization (byte)	W	2	
Number of sets of precoding coefficients	Q	256	$F.I$
Precoding set size (byte)	R	128	$X.W$
Precoding set size with overhead (byte)	S	140	$Y + R$
Number of sets/packet	T	10	$\lfloor (E - C - B)/S \rfloor$
Number of packets/TTI	U	26	$\lceil Q/T \rceil$
Packet size (byte)	V	1454	$A + B + C + S.T$
CP throughput (Gb/s)		**0.6**	$U.V.8.10^{-6}/G$

REFERENCES

[1] Industrial Internet Consortium (IIC), "The Industrial Internet of Things Volume G8: Vocabulary" v2.1, August 2018

[2] 5G Alliance for Connected Industries and Automation (ACIA) "5G for Connected Industries and Automation", https://www.5g-acia.org/publications/5g-for-connected-industries-and-automation-white-paper/, April 2018

[3] 3GPP TR 22.804 Study on Communication for Automation in Vertical Domains.(Release 16)"

[4] 3GPP TR 32.821. Telecommunication management; Study of SON related Operations, Administration and Maintenance (OAM) for Home Node B

[5] 3GPP TR 32.734 Telecommunication management; Service Specific Core Network (CN) IP Multimedia Subsystem (IMS) Network Resource Model (NRM) Integration Reference Point (IRP): Common Management Information Protocol (CMIP) Solution Set (SS)

[6] 3GPP TR 38.825. Study on NR industrial Internet of Things (IoT)

[7] Guidelines for LTE Backhaul Traffic Estimation 3rd July 2011. NGMN

[8] 3GPP TR 37.910. Study on self evaluation towards IMT-2020 submission

[9] Common Public Radio Interface. CPRI/eCPRI specification www.cpri.info/spec.html

[10] xRAN Next Generation RAN Architecture. www.xran.org

[11] IEEE 802.1CM-2018 - IEEE Standard for Local and metropolitan area networks -- Time-Sensitive Networking for Fronthaul, https://standards.ieee.org/standard/802_1CM-2018.html

[12] IEEE 802.1Qbu-2016 - IEEE Standard for Local and metropolitan area networks -- Bridges and Bridged Networks -- Amendment 26: Frame Preemption, https://standards.ieee.org/standard/802_1Qbu-2016.html

[13] IEEE 802.1Qbv-2015 - IEEE Standard for Local and metropolitan area networks -- Bridges and Bridged Networks - Amendment 25: Enhancements for Scheduled Traffic, https://standards.ieee.org/standard/802_1Qbv-2015.html

[14] T. Wan and P. Ashwood-Smith, "A performance study of CPRI over Ethernet with IEEE 802.1Qbu and 802.1Qbv enhancements," in Proc. IEEE Global Communications Conference (GLOBECOM), Dec. 2015.

[15] R. Veisllari, S. Bjørnstad, J. P. Braute, K. Bozorgebrahimi, and C. Raffaeli, "Field-trial demonstration of cost efficient sub-wavelength service through integrated packet/circuit hybrid network," IEEE/OSA Journal of Optical Communications and Networking, vol. 7, no. 3, pp. A379–A387, Mar. 2015.

[16] S. Gosselin, A. Pizzinat, X. Grall, D. Breuer, E. Bogenfeld, S. Krauss, J. A. Torrijos Gijon, A. Hamidian, N. Fonseca, B. Skubic, "Fixed and mobile convergence: which role for optical networks?", IEEE/OSA Journal of Optical Communications and Networking, Vol. 7, Issue: 11, 2015, pp.: 1075-1083

[17] T. Zami, B. Lavigne, and M. Bertolini, "How 64 GBaud Optical Carriers Maximize the Capacity in Core Elastic WDM Networks With Fewer Transponders per Gb/s," IEEE/OSA Journal of Optical Communications and Networking, vol. 11, no. 1, pp. A20–A32, Jan. 2019.

[18] Microsemi, "Enabling C-RAN: the case for OTN mobile fronthaul," Nov. 2017, white paper PMC-2143908. [Online]. Available: http://pmcs.com/cgi-bin/download_p.pl?res_id=277541

[19] H. Li, L. Han, R. Duan, and G. M. Garner, "Analysis of the synchronization requirements of 5G and corresponding solutions," IEEE Communications Standards Magazine, vol. 1, no. 1, pp. 52–58, First Quarter 2017.

[20] N. G. Nayak, F. Dürr, and K. Rothermel, "Time-sensitive software defined networks (TSSDN) for real-time

applications," in Proc. International Conference on Real-Time Networks and Systems (RTNS), Oct. 2016.
[21] N. J. Gomes, P. Sehier, H. Thomas, P. Chanclou, B. Li, D. Münch, P. Assimakopoulos, S. Dixit, and V. Jungnickel, "Boosting 5G through Ethernet," IEEE Vehicular Technology Magazine, vol. 13, no. 1, pp. 74–84, Mar. 2018.
[22] T. Pfeiffer, "Next generation mobile fronthaul and midhaul architectures," IEEE/OSA Journal of Optical Communications and Networking, vol. 7, no. 11, pp. B38–B45, Nov. 2015.
[23] Optical Internetworking Forum, "Flex Ethernet implementation agreement," Mar. 2015.
[24] IEEE 802.1BA-2011 - IEEE Standard for Local and metropolitan area networks--Audio Video Bridging (AVB) Systems, https://standards.ieee.org/standard/802_1BA-2011.html
[25] IEEE 802.3br-2016 - IEEE Standard for Ethernet Amendment 5: Specification and Management Parameters for Interspersing Express Traffic, https://standards.ieee.org/standard/802_3br-2016.html
[26] IEEE 1588-2008 - IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems, https://standards.ieee.org/standard/1588-2008.html
[27] ITU-R M.2083-0 IMT Vision - "Framework and overall objectives of the future development of IMT for 2020 and beyond"
[28] Press release: https://www.o-ran.org/news/
[29] MEF https://www.mef.net/ technical-specifications-contribution-to-ce-2-0-and-mef-3-0
[30] 3GPP TS 23.203 "Policy and charging control architecture". September 2018
[31] 3GPP TS 22.261: "Service requirements for next generation new services and markets". December 2018
[32] G.9807.1 10-Gigabit-capable symmetric passive optical network (XGS-PON). https://www.itu.int/rec/T-REC-G.9807.1-201606-I/en
[33] G.984 series: Gigabit-capable Passive Optical Networks (GPON): General characteristics". ITU-T. 2003–2012
[34] G.989 : 40-Gigabit-capable passive optical networks (NG-PON2)

Philippe Sehier received his engineer degree from Ecole Supérieure d'Electricité (Supelec) in 1984. He is leading the physical research department in Bell Labs. His scope includes fifth-generation (5G) definition. Previously, he has been involved in various activities in wireless telecommunications, including satellite, fixed, and mobile wireless in third- and fourth-generation technologies as a team or project manager.

Philippe Chanclou received the Ph. D. and Habilitation degrees from Rennes University, France in 1999 and 2007, respectively. He joined Orange Labs in 2004 where he was engaged in research on the next generation optical access networks. He is the manager of a fixed access network unit and is an active contributor to FTTx studies concerning Next Generation-Passive Optical Network and fronthaul for Cloud and Virtual Radio Access Network. He received the SPIE Rudolf Kingslake Prize in 2007 and the Global Telecom Business innovation awards in 2011.

Kimmo Kettunen has master's degree in mathematics from University of Helsinki in 1992 and doctoral degree in signal processing from Helsinki University of Technology (now Aalto) in 2003. Currently he works as a Senior Architect at Nokia Bell Labs.

David Chen has more than twenty years of industrial experiences in advanced technologies and the wireless industry. He is currently the Senior Architect at Nokia with a focus on 5G & Cloud RAN transport architecture and X-haul Proof of Concept product design. He received his bachelor's degree in Electrical Engineering from National Taiwan University and both his Master degree and PhD in Electrical Engineering from Northwestern University (USA).

Mike Lemke is a Senior Cloud Architect at Nokia with more than 30 years telecommunications experience. He is also a key contributor to the xRAN fronthaul specification in the area of fronthaul delay. He received his bachelor's degree in Computer and Electrical Engineering from Purdue University (USA) and a master's degree in Computer Science from the University of Minnesota (USA).

Yvan Pointurier (S02, M06, SM12) received a Ph.D. from the University of Virginia, USA in 2006. Between 2006 and 2009, he was a postdoctoral fellow at McGill University in Montreal and then a senior researcher at AIT, Greece. In 2009, Dr. Pointurier joined Alcatel-Lucent (now Nokia) Bell Labs as a research engineer, and is now the head of the "Dynamic Optical Networking and Switching" department at Nokia Bell Labs. His team is working on circuit and optical packet switched networks, with activities ranging from the physical layer to planning algorithms. Dr. Pointurier has authored or co-authored more than 15 European and US patents, and over 100 technical papers in leading journals, key conferences (OFC, ECOC, ACM Internet Measurement Conference, IEEE Infocom, ICC, Globecom), several of them top-scored, and book chapters. He received a best paper award at the IEEE ICC conference in 2006 and an IEEE Communication Letters Exemplary Reviewer award in 2014 and 2016

Nihel Benzaoui received the engineering degree from "Institut National des Telecommunications et des Technologies de l'Information et de la Communication", Oran, Algeria in 2010, and a Master Specialisé in Conception et Architecture Réseau from "Telecom ParisTech", Paris, France in 2012. She joined Nokia Bell Labs, France, as a Ph.D. student in 2012. In 2015 she was appointed as a permanent research engineer position. Her current research interests are oriented toward architecture design, proposal and evaluation of multi-layer mechanisms for optical networks in support for time sensitive 5G networks and beyond.

Pascal Dom is a Senior Product Manager at Nokia with more than 25 years telecommunications experience. He is today a contributor to the PON specifications in the area of mobile transport. He received his master's degree in Computer Sciences from VUB Brussels Free University (Belgium) and an executive master's degree in Business Administration (eMBA) from the University of Antwerp (Belgium

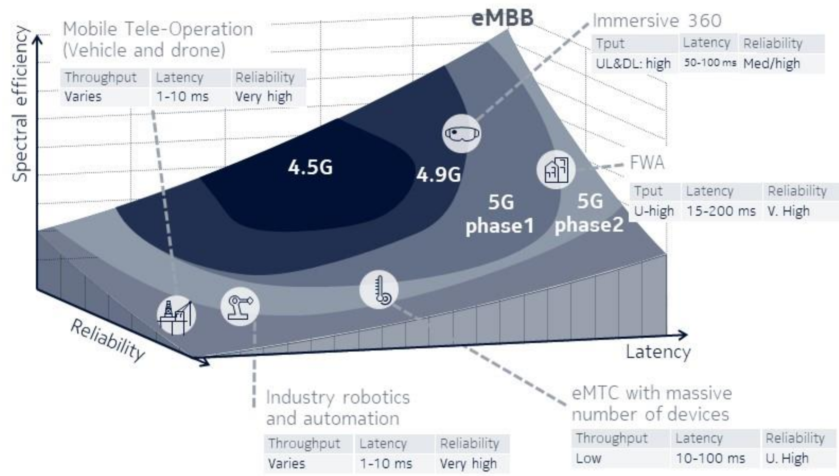


Figure 1: Relation between latency, reliability and spectral efficiency for several applications

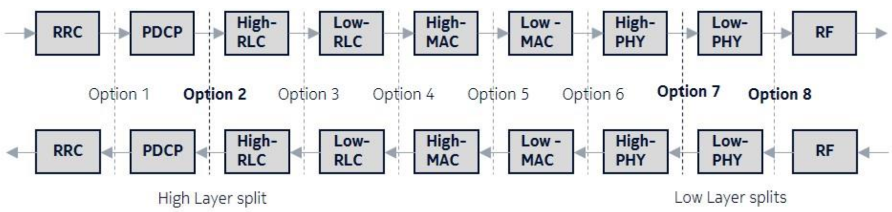


Figure 2: RAN split points

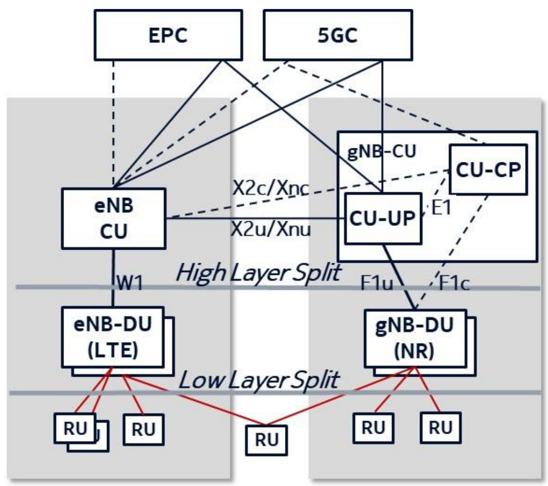


Figure 3:Functional RAN decomposition and mapping to physical entities

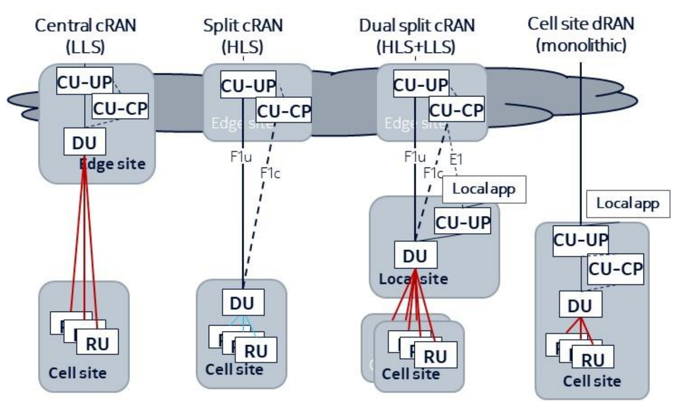
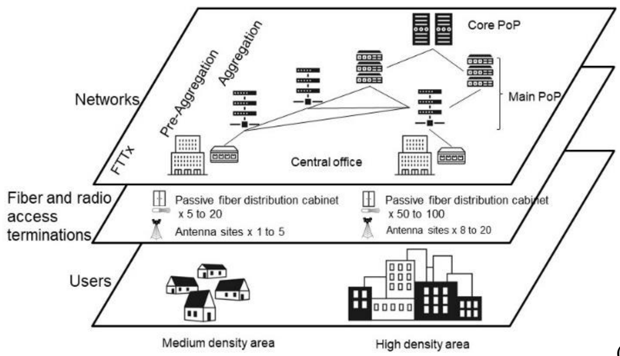
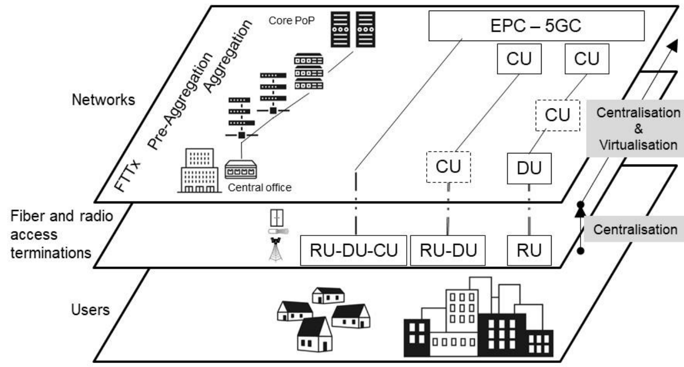


Figure 4: Disaggregated RAN, selected functions placement options



(a)



(b)

Figure 5: Network architectures considering users, fiber and radio access networks terminations and different networks segments for wireless and wireline backhauling (a). and with localization of 3GPP RAN decomposition equipment (b)

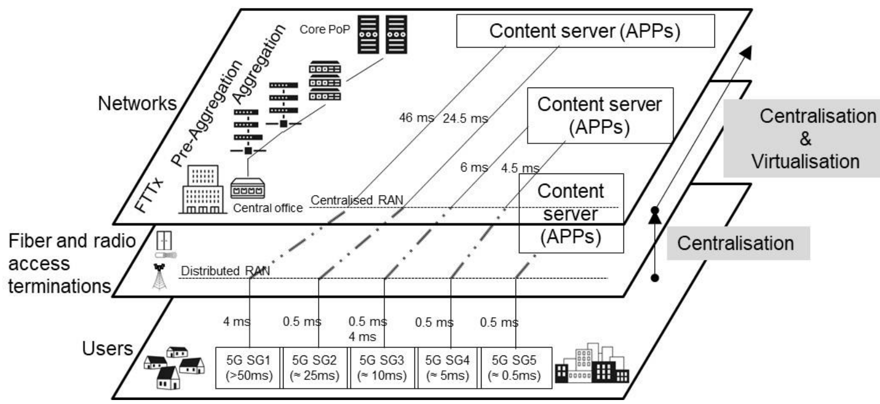


Figure 6: Latency 5G services group distribution