A 5G AMERICAS WHITE PAPER
JUNE 2020

Innovations in 5G Backhaul Technologies

IAB, HFC & FIBER

Contents

Executive Summary

Executive Summary

The introduction of 5G technology globally is driving radio access network (RAN) densification, new network architectures, and innovative use cases with stringent performance requirements (e.g., throughput, latency and reliability). To be successful, network operators need to deploy 5G transport technologies that can meet these new requirements in a cost-efficient and timely manner. Traditionally, point-to-point dark fiber has been the transport technology of choice for wireless networks but can quickly become cost prohibitive in certain scenarios. This paper focuses on innovative backhaul, midhaul and fronthaul transport technologies for 5G networks to address this gap. The right choice for 5G transport is driven by rigorous technical requirements and the vast array of use cases that 5G technology enables, balanced with real-world economic and operational considerations.

One of the key innovations in wireless backhaul is Integrated Access and Backhaul (IAB). IAB is standardized in 3GPP Release 16 and aims to reuse the existing 5G radio air interface for backhaul purposes as well. This technology has generated a lot of interest in the industry since IAB is expected to provide a cost-efficient and fast time-to-market backhaul solution. Several use cases of IAB include network densification, filling coverage holes, on-demand coverage, and capacity expansion. On the flip side, since IAB allows the use of access spectrum for backhaul as well, it may impact network quality due to interference or the reduction in capacity due to the multiplexing mechanism used between access and backhaul. Therefore, IAB deployments need to be carefully planned to address specific deployment scenarios and requirements. This paper addresses the technology aspects of IAB that are part of the standard, and its use cases and deployment considerations. An overview of IAB-related future evolution research and studies ongoing in the industry is also provided.

In addition, recent advances in hybrid fiber coaxial (HFC) network and passive optical network (PON) technology make these solutions equally promising options for 5G transport. Both HFC and PON are already extensively deployed in areas where 5G will be in the most demand, specifically dense urban and urban environments. Leveraging existing HFC and PON deployments significantly reduces the time-to-market and cost of deploying 5G. Other advances in Ethernet-based transport such as time-sensitive networking (TSN) and radio-over-Ethernet (RoE) are transforming 5G fronthaul networks. Additionally, wavelength division multiplexing (WDM) can be applied across a broad range of technologies to meet critical 5G transport requirements.

All 5G transport technologies discussed in this white paper have their relative advantages and disadvantages. The preferred technology choice depends on the specific application, deployment scenario, market situation, existing infrastructure, etc. For instance, IAB using mmWave spectrum is well-suited for small cell deployments where there is no existing wireline infrastructure and a low-cost, fast time-to-market deployment is a critical requirement. Similarly, HFC and PON technologies are ideal for cost-efficient and rapid 5G network deployments in areas where their respective infrastructures exist.

This paper is divided into the following chapters:
The **Introduction chapter** lays out the key requirements of 5G transport and various technology options available. It also gives an overview of business drivers for alternative transport technologies.
The **Integrated Access and Backhaul (IAB) chapter** describes this newly specified innovative wireless backhaul solution for 5G in detail. This chapter provides an overview of some of the use cases that are interesting for initial IAB deployment. It will also provide the reader with an understanding of the key technology aspects and deployment considerations for IAB.
The **Wireline Transport Technologies chapter** covers HFC, PON, Ethernet and WDM technologies in detail. For each of these technologies, this paper discusses the technology aspects, business drivers, recent advances, deployment scenarios and future trends.

1. Introduction & Background

1. Introduction and Background

As 5G networks become more complex, the increasing demand for data will introduce new requirements for 5G transport, along with the need for additional technology options to support them. This chapter lays out those options, along with an overview of the underlying business drivers that are propelling these alternative transport technologies.

1.1 5G Transport Requirements and Technologies

The introduction of 5G new radio (NR) technology is enabling new wireless use cases such as enhanced mobile broadband (eMBB), ultra-reliable low latency communications (uRLLC), massive machine type communications (mMTC), and high speed fixed wireless access (FWA). These new use cases are, in turn, placing new, more stringent, requirements on the underlying transport networks that support the 5G network. To successfully deliver a satisfying 5G user experience, future transport networks will need to provide significant improvements in peak data rates, area traffic capacity, latency, synchronization, security, automation and new interfaces.

For example, the application of massive multi-input multi-output (mMIMO) antenna technology and new coding techniques, coupled with extremely wide channel bandwidths made possible by millimeter wave (mmWave) spectrum, has produced a ten-fold increase in peak data rates, from 1 Gbps today to 10 Gbps and beyond. Similarly, the new 5G NR frame structure has drastically reduced latency from 10 milliseconds (or more) to less than 1 millisecond, compared with previous 4G technology. These and other key capabilities of 5G are captured in the ITU-R IMT-2020 framework [1] and are summarized in the figure below.

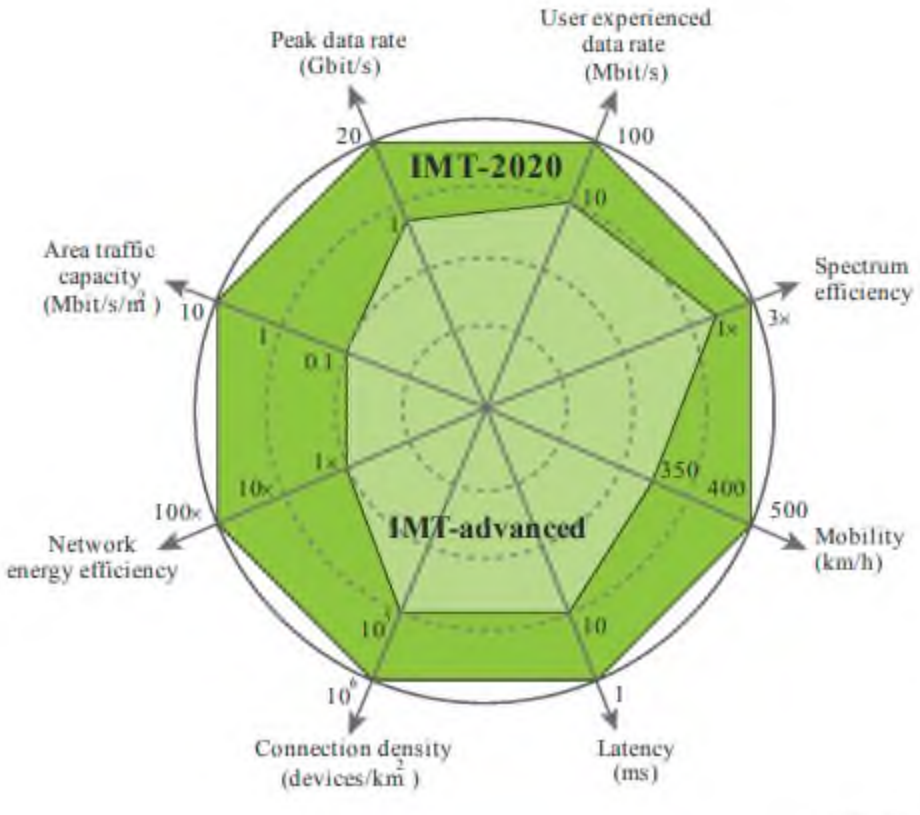


Figure 1 – Enhancement of key capabilities from IMT-Advanced to IMT-2020 [1]

To realize these new capabilities and overcome the propagation and penetration losses associated with mmWave spectrum, operators will also need to deploy much denser network topologies, requiring substantial capital investments.

In practice, actual 5G transport requirements for each radio antenna site will depend on several factors, including: the number of spectrum bands, the channel bandwidth per spectrum band; the number of MIMO layers; the maximum supported modulation scheme; the number of transmit and receive antennas; and the use cases that need to be supported by the network.

The distribution of radio access network (RAN) functions between the radio antenna site and central locations also plays a pivotal role in the transport requirements. These functions include radio frequency (RF) signal processing and other layers of the protocol stack, including: the physical (PHY); medium access control (MAC); radio link control (RLC); packet data convergence protocol (PDCP); and radio resource control (RRC) layers. Figure 2 shows the relationship between the RAN functions and the 5G core network (5GC) and end user equipment (UE).

In [2], the Third Generation Partnership Project (3GPP) defined a next generation RAN (NG-RAN) architecture where 5G NR base station (gNB) functionality is split between two logical units: a central unit (CU) and a distributed unit (DUs). In the 3GPP model, the CU is connected to the 5G core (5GC) via the NG interface and the CU is connected to the DU via the F1 interface, as shown below in Figure 3.

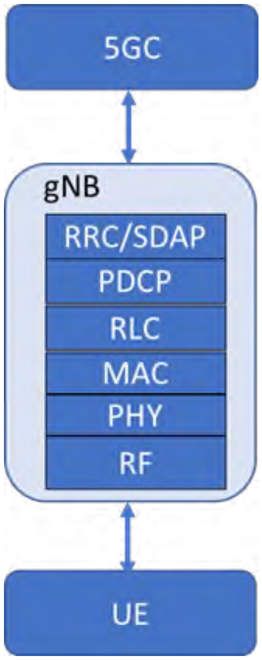


Figure 2 - Radio access network functions

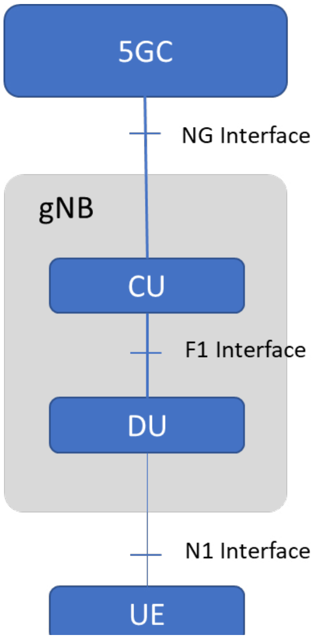


Figure 3 - 3GPP NG-RAN Architecture

The 3GPP studied several different functional splits between the CU and DU in [2]. In total, 8 possible split options were considered, including 5 high level split (HLS) options and 3 low level split (LLS) options. The different split options are shown in Figure 4 below.

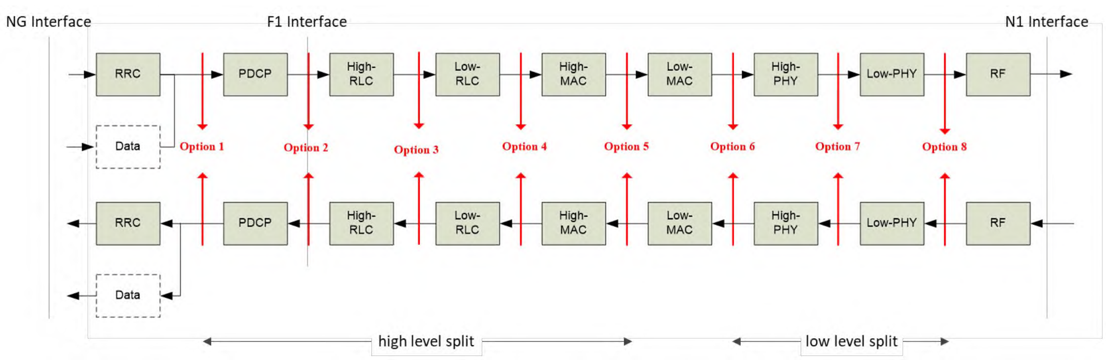


Figure 4 - Functional split between central and distributed units

As illustrated above, the radio signal processing stack in NR is a "service chain" of functions which are processed sequentially. These functions can be decomposed and isolated with defined interfaces between them to achieve disaggregation. Functions that need real-time processing are grouped within the DU, while those not requiring real-time are grouped within the CU.

The HLS options (Options 1-5) have the least demanding transport network requirements but lack the efficiencies and performance associated with more centralized approaches. Conversely, the LLS options (Options 6-8) offer higher levels of centralization and coordination across the protocol stack, which enables more efficient resource utilization and improved radio performance. The LLS options, however, have much more stringent data rate and latency requirements, which may limit network deployments in terms of network topology and available transport options. They also consume proportionately higher transport network resources, which in turn drives up transport network deployment costs.

The optimal split depends on a number of technical and business parameters, like network topology, availability of fiber, the number of users, volume of services, etc. In the end, the 3GPP selected HLS Option 2 functional split (i.e., PDCP/High RLC) for the F1 interface between the CU and DU, as specified in [3].

In a separate study [4], the ITU Telecommunication Standardization Sector (ITU-T) adopted a slightly different transport network architecture for 5G that is comprised of three logical elements: CU, DU, and remote unit (RU), as shown in Figure 5(a). In this model, the mid and lower layer functions are divided between the DU and RU. The RU implements the RF functions and, depending on the functional split between the RU and DU, possibly the low-PHY and high-PHY functions too. Depending on the network requirements, the CU, DU and RU can be grouped in different combinations to form the actual physical network elements, see Figure 5(b-d). This provides the flexibility to accommodate different network architectures, applications, and transport network requirements.

As shown in Figure 5, the transport network between the 5GC and the CU is referred to as backhaul. The backhaul network implements the 3GPP NG interface. Similarly, the transport network between the CU and DU is referred to as midhaul. The midhaul network implements the 3GPP F1 interface. Lastly, the transport network between the DU and RU is known as fronthaul. Collectively, backhaul, midhaul and fronthaul are commonly referred to as xhaul.

Several fronthaul network interfaces have been defined to date. Currently, the two most common ones are the Common Public Radio Interface (CPRI) and enhanced CPRI (eCPRI). CPRI and eCPRI are specified by the CPRI industry cooperation.

The CPRI specification was developed in 2003 as a common interface between a remote radio head (RRH) and baseband unit (BBU). The RRH is equivalent to an RU with an Option 8 functional split (i.e., RF/Low PHY) and the BBU is equivalent to a combined DU and CU. The CPRI interface was designed to transport digitized time-domain samples of the baseband signal between the RRH and the BBU. The advantages of this approach include simpler RRH equipment, lower power consumption, easier operation, and cheaper maintenance at the edge of the network.

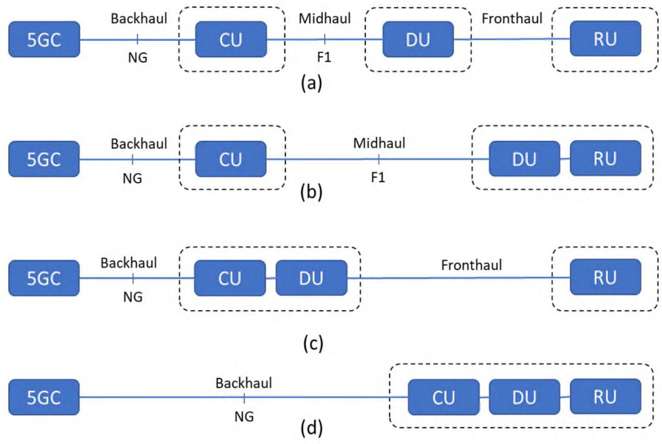


Figure 5 – Possible CU, DU and RU Combinations

The CPRI protocol supports several bit rates options, as follows:

Table 1 - CPRI Bit Rates

Level	Bit Rate
CPRI 1	614.4 Mbps
CPRI 2	2.457 Gbps
CPRI 7	9.83 Gbps
CPRI 9	12.165 Gbps

The introduction of new radio techniques such as massive MIMO drive the need to increase the capacity transported over CPRI in a way that it becomes a transport challenge due to the very high capacities demanded. For example, Figure 6 shows the required CPRI line rate (without coding) for various channel bandwidths and numbers of transmit/receive antennas. This figure clearly shows that it quickly becomes impractical to use CPRI for 5G systems with the large channel bandwidths and high transmit/receive antenna counts [5].

Consequently, the industry partners responsible for the CPRI specification developed eCPRI [6]. eCPRI reduces the demands in transport capacity via a flexible functional decomposition while limiting the complexity of the RU. eCPRI offers a ten-fold reduction in the required data rate compared with CPRI and allows packet-based transport technologies such as Ethernet to be used.

The CPRI cooperation released a new version of the eCPRI specification (2.0): introducing an interworking function to the existing RU and DU, so CPRI and eCPRI can interwork in the network. An interworking type 0 is a device located between the eCPRI transport network and one or several radio units, while interworking function type 1 and 2 devices are located between CPRI nodes and the transport network.

At the same time, Institute of Electrical and Electronics Engineers (IEEE) 1914 started working on a Next Generation Fronthaul Interface (NGFI). There are two efforts: IEEE 1914.1 covers standards for packet-based fronthaul transport networks and IEEE 1914.3 that takes care of the Radio over Ethernet (RoE) encapsulation and mappings addressing DU/CU splits 7.1/7.2 and 8.

Operators installing new NR technology collocated at existing LTE sites are dealing with the challenge of transporting CPRI and eCPRI until all traffic at the cell site is eCPRI, at which point packet switched networks can be used in the transport layer using new protocols like Time Sensitive Networking (IEEE 802.1CM) developed to deliver low latency and accurate synchronization for fronthaul traffic.

Figure 7 shows the data rate and latency requirements and distance limitations for the NG, F1, eCPRI and CPRI interfaces. The upper portion

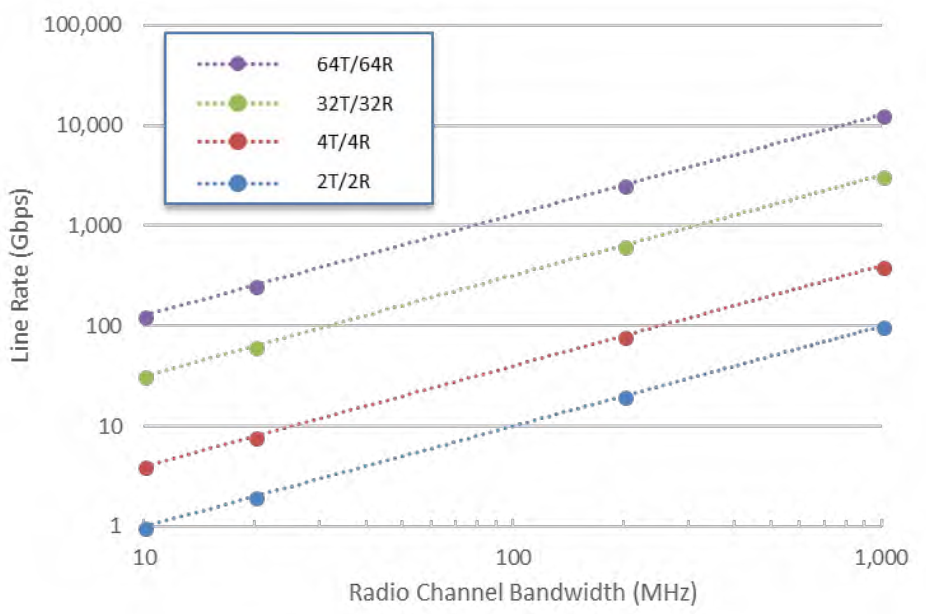
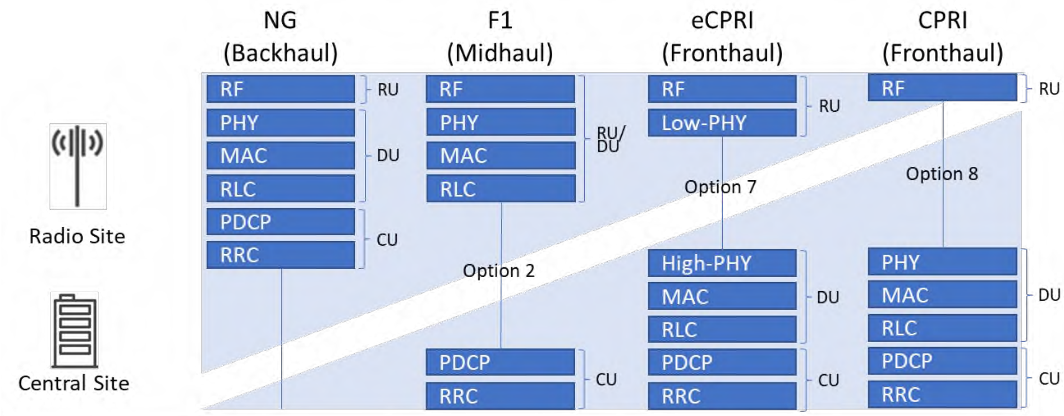


Figure 6 - CPRI Line Rates

of this figure shows the functions residing at the radio antenna site whereas the lower portion shows the functions at the central site. The transport requirements are based on a radio site with 3 sectors, 100 MHz channel bandwidth, 64 transmit/receive chains, 256 QAM, 16 MIMO layers and multiuser MIMO (MU-MIMO).



	NG (Backhaul)	F1 (Midhaul)	eCPRI (Fronthaul)	CPRI (Fronthaul)
Data Rate	10 Gbps	10 Gbps	100 Gbps	1 Tbps
Latency	< 10 ms	< 5 ms	< 200 µs	< 100 µs
Distance	200 – 500 km	100 – 400 km	< 20 km	< 10 km

Figure 7 – Transport requirements for 5G NR functional splits (Souce: Nokia)

This figure illustrates the significant difference in the required data rates and latencies between the specified backhaul, midhaul and fronthaul interfaces.

In addition to work done by the CPRI industry cooperation, the O-RAN Alliance announced in June 2018 that it would be leading efforts towards open RAN with interoperable interfaces and RAN virtualization. The O-RAN Alliance has 9 working groups looking at many topics among those L2-L3 RAN protocols for the high layer split and L1 options (e.g., eCPRI and IEEE1914) for the low layer split. O-RAN also introduced a new architecture for the 7.2 functional split. Two categories were specified: Category A and Category B. The main difference between the two is the placement of the precoding functions for the downlink. Category A devices do not have precoding functions, whereas Category B devices include precoding functions.

3GPP option 2 for the high layer split (HLS) between the Radio Link Control Protocol (RLC) and the Packet Data Convergence Protocol (PDCP) offers latency in the order of milliseconds compared to the low layer split that has it in microseconds.
Traditional backhaul solutions are seen to be required for connectivity between CU and Core, although a lot of transport will happen in the midhaul depending on the split that is chosen to be deployed by the operator.

Transport connectivity in a 5G network in general can be done using wireline and/or wireless assets. Fiber is the preferred connectivity choice, however, as we move towards a denser network specially with mmWave spectrum in the RAN, a wireless solution also makes sense. 3GPP release 16 introduces the integrated access backhaul (IAB) concept to allow the NR radio to use part of the RAN spectrum for backhaul connectivity. As a result, it is possible to use NR for a wireless backhaul link from central locations to distributed cell sites and between cell sites.

IAB can be used in any frequency band in which NR can operate. However, it is anticipated that mmWave spectrum will be the most relevant spectrum for the backhaul link. Furthermore, the access link may either operate in the same frequency band as the backhaul link (known as inband operation) or by using a separate frequency band (out-of-band operation).

When time to market is important, wireless solutions using microwave and mmWave spectrum become an option. High capacity links operating in the E-Band spectrum (70/80GHz) can today deliver

10Gbps of capacity and with low latency over a single channel. The industry is looking to opening new spectrum in mmWave, specifically in the W band (75-110GHz) and D band(110-170GHz) that will enable delivery of wireless links in the order of 100Gbps. Today, traditional Microwave can support 5G backhaul and high layer splits, while mmWave can be used for low layer splits due to the capacity and latency requirements.

1.2 Business Drivers for Alternative Transport Technologies

Thanks to the large mmWave bandwidth, spatial beamforming and network densification, 5G promises user data rates of 1 Gbps and beyond. Small cells with less than 100 m average inter-cell distances are also expected to cover large network areas. High deployment and operational cost of large scale backhaul links, however, makes the network densification commercially infeasible. Studies show that in the U.S. alone at least $130 billion in fiber builds is required to reach the full performance that 5G promises [7]. Even if fiber xhaul was deployed today, the operational cost of backhaul would be extremely high for a large number of small cells. In addition to the financial burden for operators, the heavy cost of backhaul deployment and operation and lack of investment can increase the digital divide across the world. Both existing shared infrastructure technologies such as Hybrid Fiber Coax or Passive Optical Networks, as well as Integrated Access and Backhaul, can potentially overcome the deployment cost of the network densification to bring all promises of 5G networks.

IAB is introduced by 3GPP to overcome both CAPEX and OPEX requirements for the realization of dense cellular networks. IAB, as part of the 5G NR Access Radio Network, is expected to provide several advantages: automatic established backhaul, no additional equipment needed when the backhaul direction is within the access sector. As such, IAB is attractive as an alternative to fiber for dense street level mmWave 5G deployments.

Both HFC and PON networks are already ubiquitously deployed across the Americas, reaching virtually every building and home across the continent. Already used for backhaul, recent advancements in technology are now positioning these transport systems for 5G midhaul or even fronthaul. HFC in particular is also able to transport power, which further reduces the operational challenges of densification when using this method of transport. Leveraging these existing transport technologies can dramatically reduce CAPEX, OPEX and build time requirements of 5G densification.

2. Integrated Access & Backhaul (IAB)

2. Integrated Access & Backhaul (IAB)

Integrated Access and Backhaul (IAB) is a promising solution for successful 5G adoption. The key concept of IAB is to reuse the existing framework of 5G access link for the backhaul as well, by efficiently multiplexing access and backhaul in the time, frequency and/or space domain. While as per standard, IAB can be supported in sub-6GHz as well as above 6GHz spectrum, the availability of mmWave spectrum for 5G opens the opportunity to leverage a large amount of new access spectrum that is very well suited for IAB. The beam steering capability in massive MIMO solution may be used to allow for the spatial separation between the backhaul and the access, increasing spectrum efficiency.

This type of solution allows the operator to improve coverage by installing denser networks, without having to lay fiber or, at least, delaying the large and difficult investment of laying fiber for backhaul. In this way, IAB facilitates and reduces the costs of very dense deployments, improving cellular coverage.

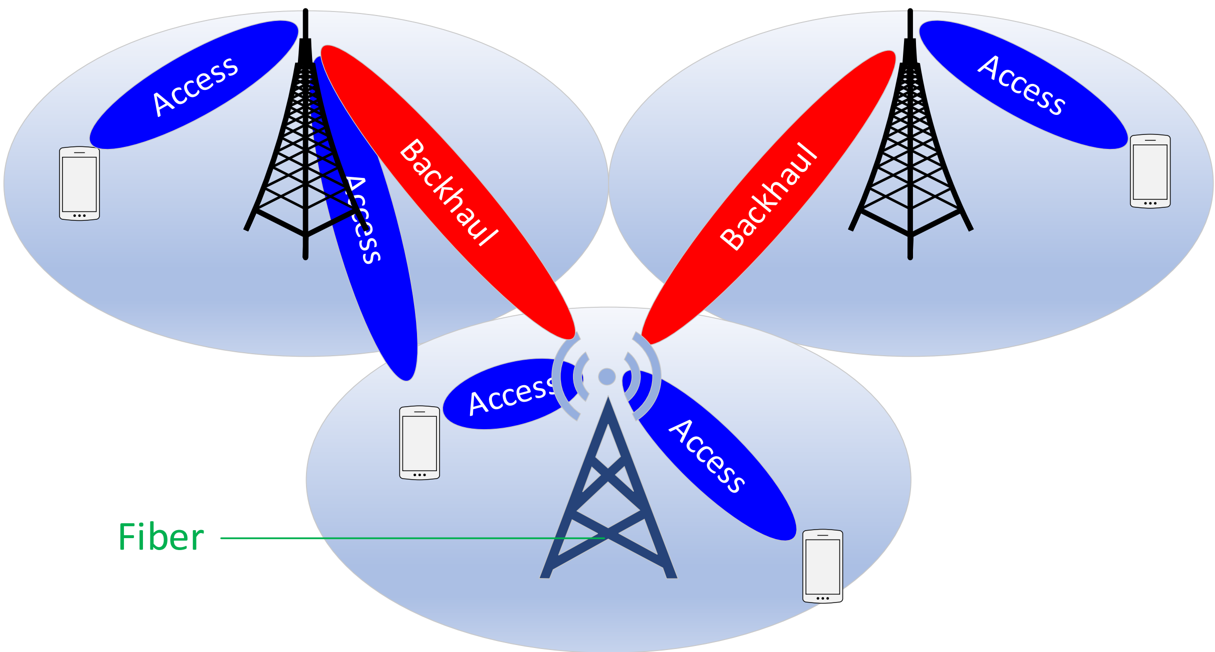


Figure 8 - Integrated access and backhaul

Some of the key characteristics of IAB are as follows:

- Leverage existing technology: IAB technology leverages the already existing NR radio interface specification between device and network (NR Uu interface) for the backhaul radio link with modifications/ extensions.
- Low deployment and operational cost: An IAB node is a combination of a gNB and a UE that plays the role of an access node as well as a backhaul relay node simultaneously. By avoiding or delaying the need for dedicated backhaul, IAB reduces the deployment and operational cost significantly.
- Frequency band flexibility: Contrary to the IEEE 802.11ad/ay standard, an IAB network can operate on multiple bands available under the 3GPP 5G NR standard.
- Efficient spectrum usage: IAB allows for a more flexible spectrum usage where the spectrum can be efficiently utilized between access and backhaul, compared to the more static allocation for conventional wireless backhaul.
- In-band or out-of-band IAB: In case of in-band IAB, access and backhaul fully or partially overlap each other in frequency domain. While out-of-band IAB implies access and backhaul have no overlap in frequency domain.
- IAB is supported in stand-alone (SA) as well as non-stand-alone (NSA) architecture in 5G. UEs can transparently connect to the network via IAB.
- Flexible and spectrally efficient range extension: IAB supports backhaul topologies with multiple hops for extended range or to support deployments in convoluted urban canyons.

- Flexible quality of service (QoS) framework: IAB allows for fine-granular end-to-end QoS support of individual traffic flows across access and backhaul links as well as QoS-class-specific traffic prioritization as applied on Ethernet or IP transport networks.
- Robustness to backhaul link failure: Support for path redundancy in the wireless backhaul topology allows for robust operation in case of individual backhaul link failures, e.g., due to moving obstructions. The topological redundancy further enables dynamic load balancing across the backhaul links to optimize backhaul capacity to time-dependent traffic load.
- Optimized management of topology, routing and resource allocation: IAB follows the software-defined-networking paradigm where crucial management functions are centrally controlled. This enables optimization of the backhaul topology and the routing paths for traffic across this topology. Further, resource allocation for backhaul and access links are centrally managed which allows accounting for duplexing constraints across multiple hops and incorporating topology-wide inter-link interference mitigation. IAB further incorporates local decision-making processes to allow for flexible and fast response to highly dynamic resource demand and to reduce control-plane latency.

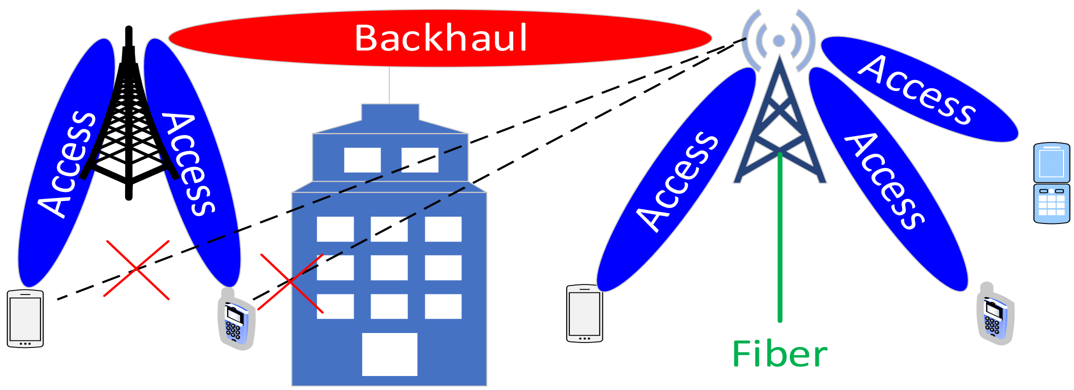


Figure 9 - Filling coverage gaps with IAB

2. 1 Standardization of IAB in 3GPP

IAB has been studied earlier in 3GPP in the scope of LTE Rel-10, under the label LTE relaying. There was already support for a wireless relay node in LTE based on efforts by the TSG Radio Access Network (TSG RAN) in Rel-9/10. However, there have been only a handful of commercial LTE relay deployments mainly because the existing LTE spectrum is too expensive to be used for backhauling, such backhauling was limited to one hop, and dynamic changes to the backhaul topology was not supported.

IAB work in 3GPP was re-initiated since 2017 with a study item followed by a normative phase in release 16. This effort was accompanied by parallel efforts in TSG SA2, in charge of developing Stage 2 of the 3GPP network standards, as well as SA3. Several design approaches were discussed, with the main criteria in consideration being that of an effective and flexible deployment of a system that allows a smooth transition and flexible integration from and to legacy deployments. At the time of this writing, the key features to be supported by the first release of 3GPP IAB network for NR backhauling (Rel-16), which is expected to be completed by June 2020, are:

- Multi-hop backhauling: to enable flexible range extension
- QoS differentiation and enforcement: to ensure that the 5G QoS of bearers is fulfilled even in a multi-hop setting
- Support for network topology adaptation and redundant connectivity: for optimal backhaul performance and fast adaptation to backhaul radio link overloads and failures
- In-band and out-of-band relaying: the use of the same (for in-band) or different (for out-of-band) carrier frequency for the access (i.e. link to UEs) and backhaul links (i.e. link to other network nodes) of the IAB node
- Support for legacy terminals: the deployment of IAB nodes should be transparent to UEs (i.e. no new UE features/standardization required)

The work item on IAB in 3GPP Release-17, currently expected by December 2021, aims to enhance Release-16 IAB in terms of robustness, spectral efficiency, latency, and end-to-end performance. Key features include:

- Simultaneous communication with parent nodes and child nodes using Spatial Division Multiplexing (SDM) or Frequency Division Multiplexing (FDM)
- Enhancements to topology adaptation and topological redundancy supporting relay migration between IAB-donors and lowering migration delay
- Enhancements to routing and transport across the backhaul for improved efficiency and performance
- Support for dual-connectivity scenarios defined in RAN2/RAN3 in the context of topology redundancy for improved robustness and load balancing [8]

2.2 IAB Architecture

The architecture of IAB networks will also represent a fundamental evolution in 5G networks. This section describes those changes.

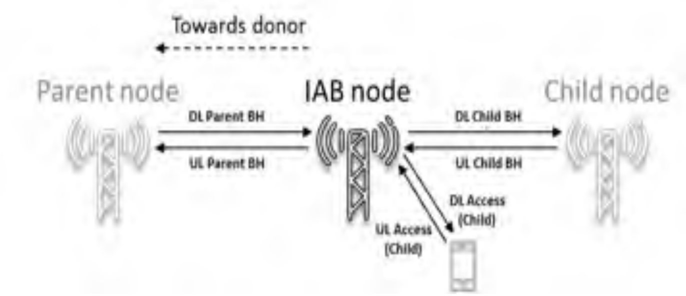


Figure 10 - IAB Parent and Child relation

Two types of links are supported in IAB networks: access links and backhaul links. An access link is a link between an access UE and an IAB node or IAB donor, while a backhaul link is a link between an IAB parent node and IAB child node (Figure 10).

IAB parent node is responsible for scheduling the DL/UL traffic for both access and backhaul links, and the IAB child node at the end of the transmission chain is responsible for scheduling the DL/UL traffic between itself and the UEs.

Figure 11 shows the architecture of IAB technology end to end.

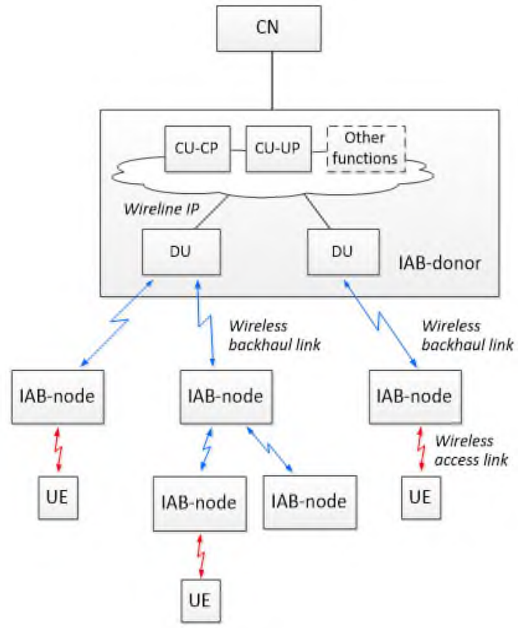


Figure 11 - IAB Architecture [1]

The IAB node can access the network using either Stand-Alone (SA) or Non-Stand-Alone (NSA) modes. In NSA mode, Evolved Universal Mobile Telecommunications System Terrestrial Radio Access New Radio (E-UTRA-NR) Dual Connectivity (EN-DC) is used. In EN-DC, the IAB-node also connects via E-UTRA to a master node (MeNB), and the IAB-donor terminates X2 as secondary node (SgNB).

These two topologies are shown in Figure 12. The standards allow several IAB-nodes to be cascaded.

IAB architecture per [9] strives to reuse existing functions and interfaces defined for access. In particular, Mobile-Termination (MT), gNB-DU, gNB-CU, user plane function (UPF), mobility management function (AMF) and session management function (SMF) as well as the corresponding interfaces NR Uu (between MT and gNB), F1, NG, X2 and N4 are used as baseline for the IAB architectures.

IAB architecture leverages CU/DU-split architecture for Radio Access Network. Figure 13 below shows the reference diagram for IAB-nodes in chain, connected to an IAB-donor for SA architecture.

IAB functionality requires two new network entities: IAB-donor, IAB-node, and a new interface. A description of each entity and interface follows in the next sections.

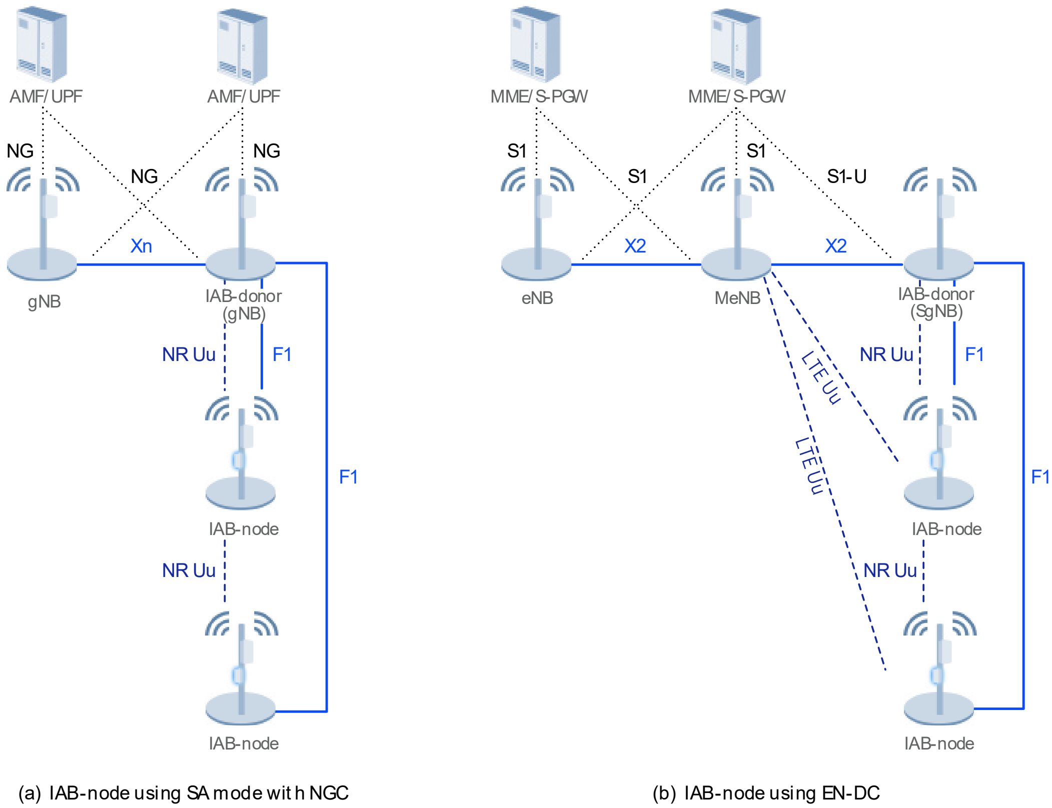


Figure 12 - IAB topologies: (a) IAB node using SA mode with next generation core (NGC); (b) IAB node using EN-DC (NSA mode)

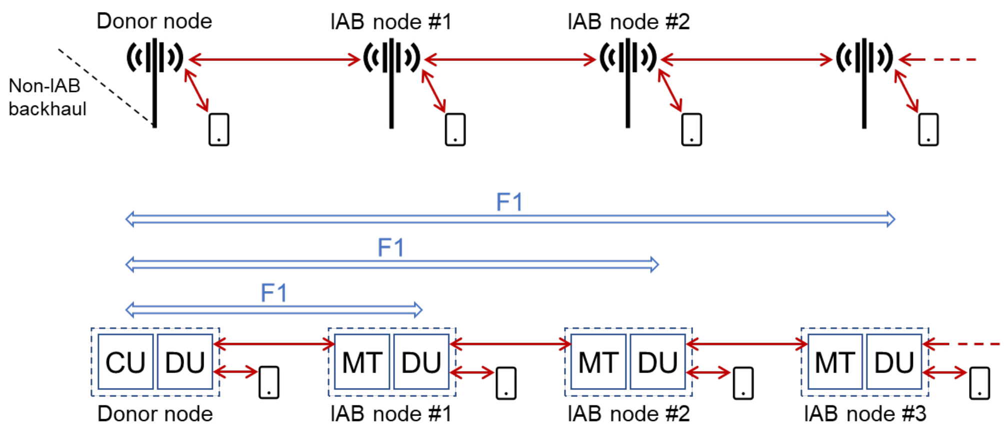


Figure 13 - IAB CU/DU architecture

2.2.1 IAB-donor

As shown in Figure 11, an IAB-donor is a gNB that provides network access to UEs via a network of backhaul and access links and consists of an IAB-donor-CU and one or more IAB-donor-DUs. The IAB-donor-CU and IAB-donor-DU communicate with each other via the F1 interface. The IAB-donor connects to the IAB-node using the 5G New Radio (NR) access interface and is connected to the Core Network. All functions specified for a gNB-DU are equally applicable for an IAB-donor-DU and all functions specified for a gNB-CU are equally applicable for an IAB-donor-CU. A Backhaul Adaptation Protocol (BAP) layer has been added above the Radio Link Control (RLC) layer in order to include routing information and allow for hop-by-hop forwarding. Details of BAP layer is mentioned in section 2.2.3.

2.2.2 IAB-node

The IAB-node connects to an upstream IAB-node or an IAB-donor-DU via a subset of the UE functionalities of the NR Uu interface (referred to as IAB-MT function of IAB-node) with some additional IAB-specific features such as support for new adaptation protocol, over the air (OTA) synchronization etc. The IAB-node provides wireless backhaul for the downstream IAB-nodes and UEs via the network functionalities of the NR Uu interface (referred to as DU function of IAB-node). IAB-nodes can be cascaded, as shown in Figure 11 above. While there is no technical limit to the number of IAB nodes that can be cascaded, it is important to keep into consideration the bandwidth and latency requirement. Hop-by-hop flow control may be required together with end-to-end congestion handling. All functions specified for a gNB-DU are equally applied for an IAB-node-DU and all functions specified for the UE context are also employed in managing the context of IAB-MT functionality.

2.2.3 Backhaul Adaptation Protocol (BAP)

Efficient multi-hop forwarding is enabled via the newly introduced IAB-specific backhaul adaptation protocol (BAP). The BAP layer is only present within the IAB network and is transparent to UEs. That is, the BAP layer is only used on the backhaul links but not on the access links.

The IAB-donor assigns a unique L2 address (BAP address) to each IAB node that it controls. In case of multiple paths, multiple route IDs can be associated to each BAP address. The BAP of the origin node (IAB-donor DU for the DL traffic, and the access IAB node for the UL) will add a BAP header to packets they are transmitting, which will include a BAP routing ID (e.g., BAP address of the destination/source IAB node and an optional path ID). Each IAB node will have a routing table (configured by the IAB-donor CU) containing the next hop identifier for each BAP routing ID.

2.3 Use Cases and Deployment Considerations

As service providers move from initial 5G market launches to building 5G capacity, they are faced with an immediate challenge of securing high bandwidth backhaul solution to the 5G sites in a fast, cost effective manner. mmWave-based 5G deployment increases the challenge of securing optimum backhaul solutions to the sites exponentially. Interestingly, mmWave-based 5G opens up a new opportunity for IAB due the very large bandwidth available in mmWave and the native deployment of massive MIMO or multi-beam system. IAB can potentially overcome some of the challenges faced by service providers planning to provide a cost-effective coverage and capacity solution. This section covers some of the foreseen use cases of IAB based on Release 16.

2.3.1 Cell Densification

In order to dramatically enhance network capacity and provision unprecedented data rates to users, overlaying mmWave 5G small cells over a macro cell is one of the best deployment scenarios that can be envisioned by the operators. But new small cells may require installation of fiber for the backhaul, which can become costly for the operator. The operator may then choose to share the mmWave spectrum for wireless backhauling thanks to the much wider bandwidth available than in lower-frequency bands. Interference mitigation techniques as well as potential resource separation between access and backhaul links such as spatial, time or frequency division can be utilized depending on the situation to minimize the negative impact of sharing the resource with backhaul link.

Depending on the deployment scenario, IAB can provide a better alternative for cell densification than wired backhaul, by connecting new cells wirelessly to backbone networks. The newly added cells increase the signal strength under their coverage, improving the overall network capacity.

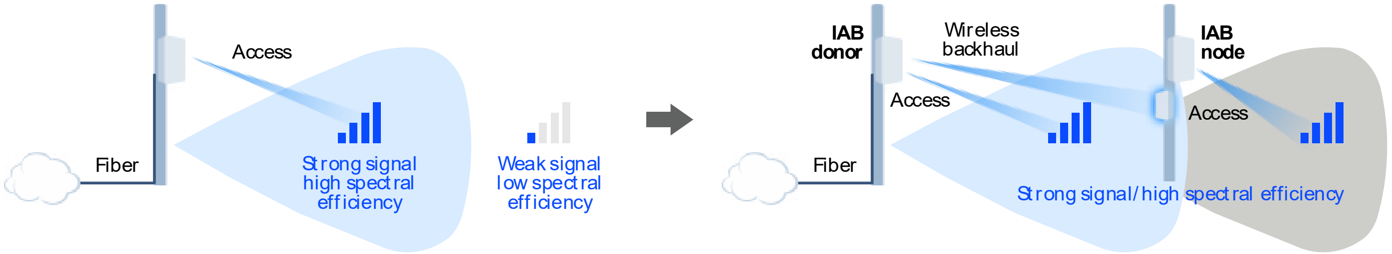


Figure 14 - Cell densification using IAB

2.3.2 Filling Coverage Holes

In 5G networks using high frequency bands, propagation is subject to high diffraction loss and pronounced shadowing, which may yield regions where the signals from the cell sites do not reach, also known as coverage holes. IAB provides a wireless backhaul link to the new cell to be added for coverage hole filling, which is in general less expensive than leasing a fiber. The coverage hole use case is depicted in the figure below.

IAB can also be used to extend the coverage into indoor areas where the signal does not reach due to the high penetration loss, by installing an IAB-node which is exposed to indoor and outdoor. An IAB-node can provide assistance without the need of laying cables throughout the inside of a building.

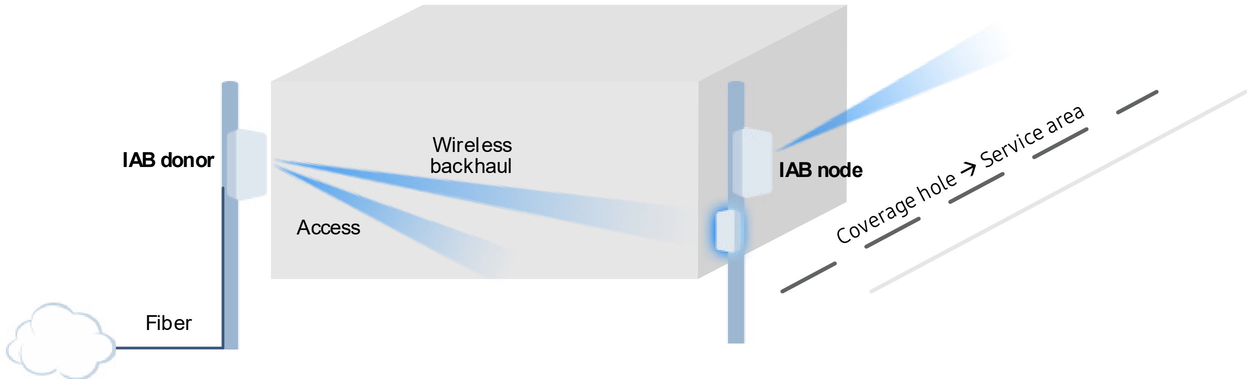


Figure 15 - Filling coverage holes using IAB

2.3.3 Coverage extension along street or highway

Another use case where IAB can help operators reduce CAPEX is by extending coverage along a street as well as around streets, which is depicted in the figure below. The signal can be relayed to the base station near the position of the user on the road through the multi-hop wireless backhaul connection provided by IAB-nodes. Only the donor nodes need fiber to the wired network, and therefore the cost for extending coverage along the road goes down.

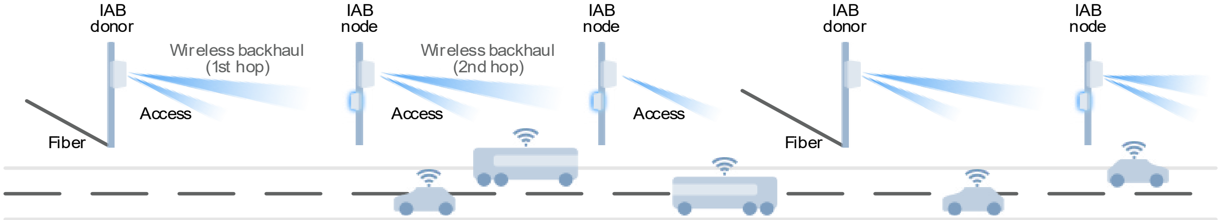


Figure 16 - Extending coverage along a linear feature using IAB

2.3.4 Infrastructure on demand

If temporary coverage or capacity needs to be added in a particular area like a stadium, concert venue, hazard zone, IAB can provide an excellent solution by allowing fast time-to-market for sites to come on air as compared to planning for a dedicated backhaul solution. IAB nodes can be opportunistically deployed/activated to deliver services to a certain geographical area for better coverage or quality of service. Due to the temporary nature of such deployment, it would also be cost effective to turn off the site and IAB once the need for additional coverage or capacity is over. IAB nodes can dynamically enter or leave the network depending on the network traffic and density of users. IAB technology consequently opens the door for a seamless realization of an infrastructure on demand network.

Figure 17 below shows a scenario of infrastructure-on-demand where a temporary IAB node provide on demand coverage for a crowded stadium.

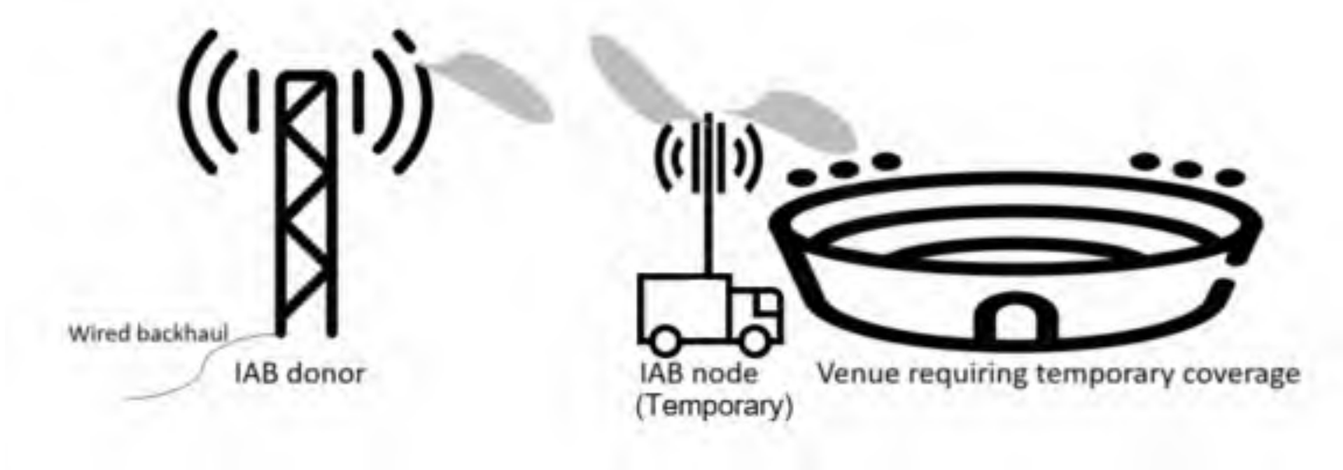


Figure 17 - IAB as an enabler of infrastructure-on-demand

2.3.5 Augmenting low-capacity indoor backhaul

In some indoor deployments, especially in small enterprises and retail stores, the existing enterprise internet backhaul connection is leveraged for radio backhaul. In these cases, the available backhaul capacity can be limited, and the strict timing sync and latency requirements required for deploying NR TDD systems cannot be always guaranteed. IAB can provide an alternate high-speed, backhaul option to the enterprise internet connection. An IAB node installed on the premises and exposed to outdoor IAB-donor for backhaul can provide high speed radio connectivity within the premises and with low CAPEX impact.

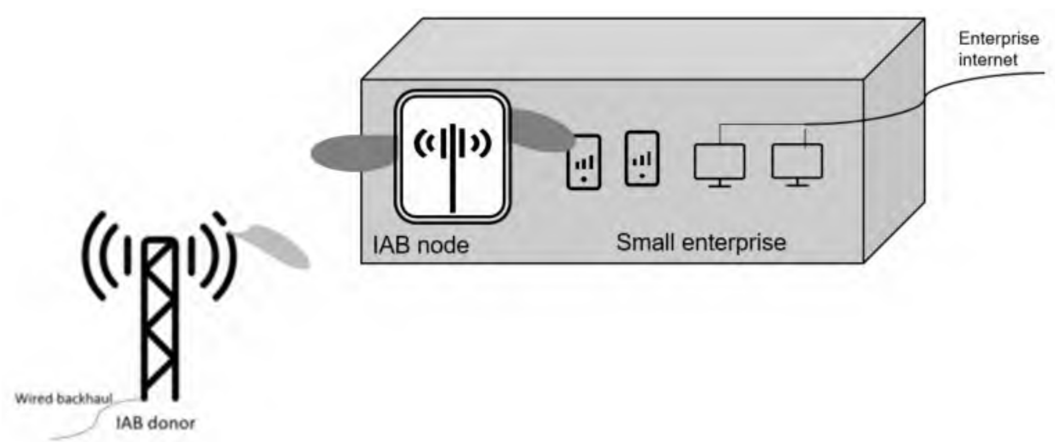


Figure 18 - IAB augmenting indoor coverage and enterprise backhaul

2.4 IAB Resource Allocation Methodologies

Additionally, IAB networks will require a different approach to how 5G networks allocate resources. The following methodologies illustrate options available in developing IAB architectures.

2.4.1 Radio Resource sharing between access and backhaul

To mitigate the cross-link interferences for in-band backhaul, different half-duplex multiplexing schemes have been designed for IAB network, such as TDM (Time Division Multiplexing), FDM (Frequency Division Multiplexing) and SDM (Spatial Division Multiplexing).

In the case of out-of-band relaying, sub-6 GHz can be considered as an access and control channel for backhaul links due to its robustness against obstacles and wide coverage area. mmWave bands can be used for high capacity backhaul links. In this case, the IAB network can operate in full-duplex mode.

Interference can occur on both access and backhaul links for the out-of-band case. In addition, cross-link interference (between access and backhaul) can occur for the in-band case. Interference management techniques, which use the channel state information, are required in order to suppress the interference among concurrent transmissions, both in access and backhaul. For the IAB node coordination, efficient signaling exchange among the MAC layer of different IAB nodes is needed, considering the rate and latency constraints of wireless backhaul links. Uplink-downlink interference is also introduced in case of asynchronous IAB node transmission mode. Adaptive intelligence algorithms can be implemented at the MAC layer of IAB and macro nodes to make adaptive decisions about the link and user/IAB child scheduling with fairness and half-duplex constraints. Training procedures can be centralized, for example at the donor nodes or distributed among some local IAB nodes.

2.4.2 Time Division Multiplexing (TDM)

In case of TDM, access and backhaul links operate at the same carrier frequency but at the different time frames. For downlink traffic, the IAB node will have to receive the signal from the parent node first before it can relay the signal further to the child node or the UEs. For uplink traffic, the IAB node will have to receive the signal from the child node or the UE first before it can relay the signal further to the parent node. In a TDM system it is necessary to divide and allocate time resources according to the resource situation of each link. Figure 19 shows one example where radio resources are allocated evenly for the child and parent links in the time domain. Note that a grey box represents a downlink (DL) slot, and a blue box represents an uplink (UL) slot.

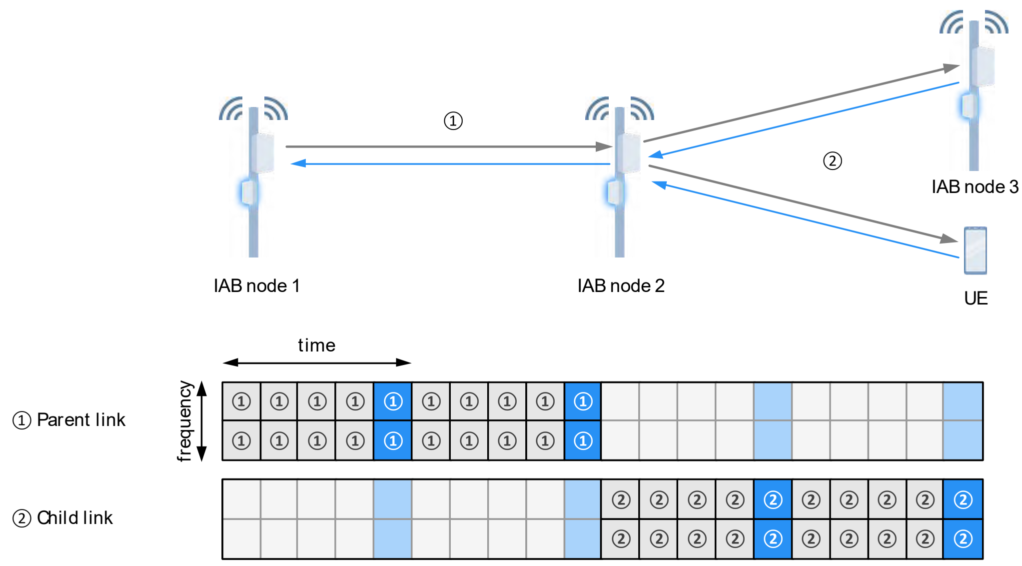


Figure 19 - Time division multiplexing

To improve the efficiency of radio resource utilization, dynamic TDM can be used. The TDM slot number and location for parent backhaul link can be flexibly configured according to the backhaul transmission capacity requirement. For each specific frame configuration, the backhaul slot can be used for access link by dynamic scheduling if the backhaul transmission is not scheduled in that slot, as shown in Figure 19. Simulation results show that dynamic TDM scheme brings significant performance improvement compared to static TDM, especially when the network resource utilization is low or medium. In the evaluations in [10], the gain for DL 50% percentile user throughput was found to be more than doubled by using dynamic TDM compared to static TDM in case of 50% resource utilization.

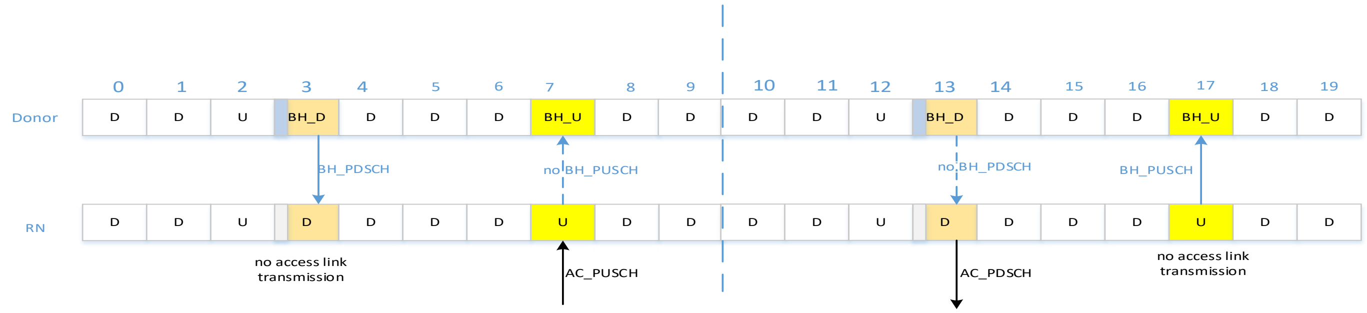


Figure 20 - Dynamic TDM scheme

2.4.3 Frequency Division Multiplexing (FDM)

If operators own enough frequency spectrum, FDM can be used to eliminate cross-link interference. Different carrier frequencies are allocated to parent backhaul, child backhaul and access links. With enough guard band between parent and child backhaul links, all the IAB nodes can transmit and receive simultaneously without introducing too much cross-link interference.

To this end, it is necessary to divide and allocate the frequency resources independently between child and parent links. Figure 21 below shows an example where the child and parent links are allocated the same amount of frequency resources.

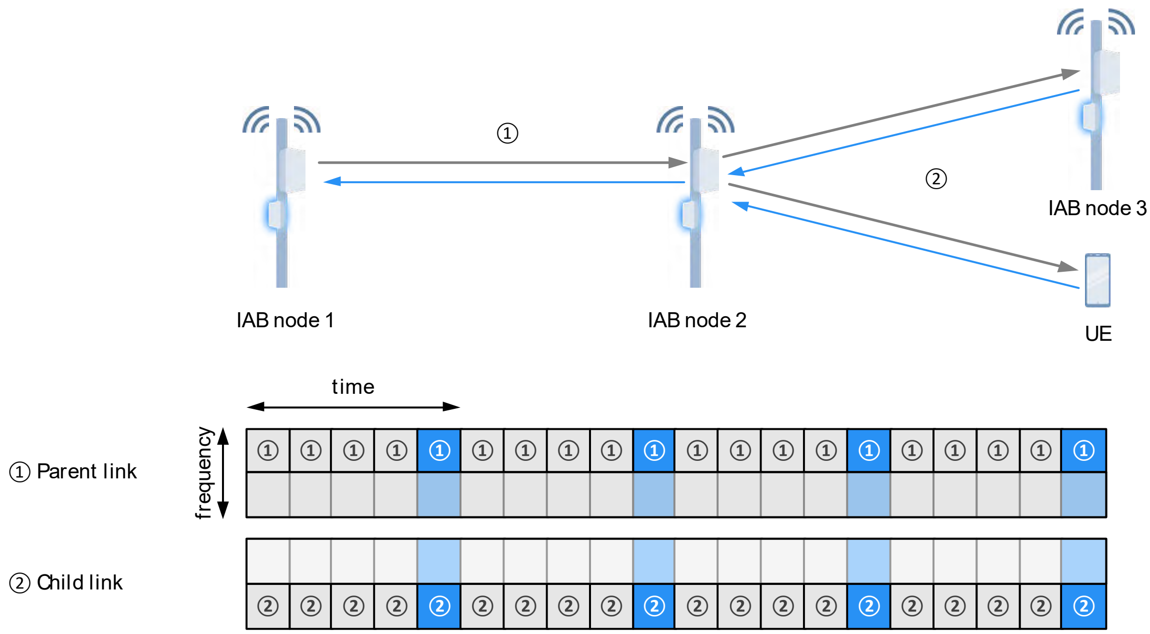


Figure 21 - Frequency division multiplexing

2.4.4 Spatial Division Multiplexing (SDM)

In addition to multiplexing schemes in time and frequency domain, spatial division multiplexing can be used to separate the backhaul and access links. Beamforming algorithms using multiple antennas can be applied on IAB nodes to separate the backhaul and access links in space.

In an SDM solution, the child and parent links exploit the spatial separation between child and parent links to minimize interference. In this case, simultaneous transmission (or reception) at the IAB-node are allowed in the same time and frequency resources, as long as the spatial separation is enough to minimize the interference between the simultaneous transmission (or reception). This will greatly reduce end-to-end transmission delay.

Figure 22 shows an example where radio resources are divided into child and parent links in the space domain.

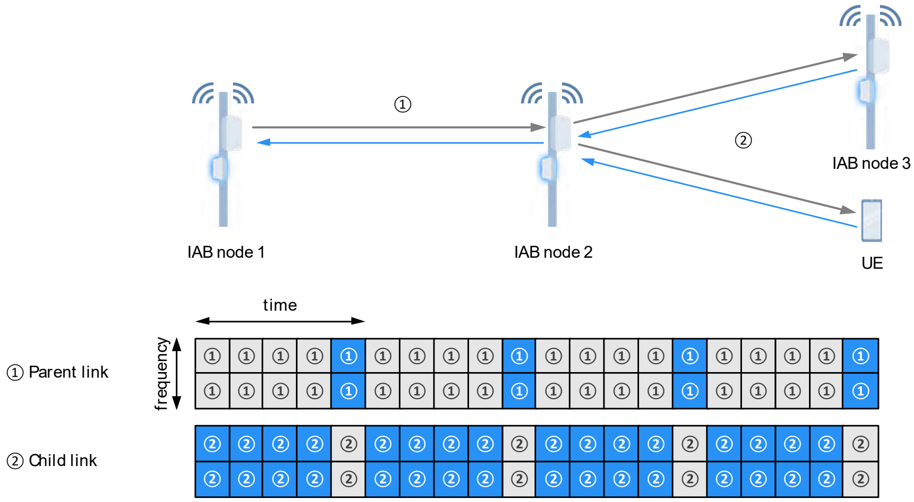


Figure 22 - Space division multiplexing

In this case, simultaneous transmission (or reception) in both child and parent links are allowed. However, even though this solution allows full exploitation of time and frequency resources, the beams are usually not narrow enough to prevent the cross-link interference if all the IAB nodes transmit and receive at the same time. As such, TDM or FDM will have to be applied together with SDM to achieve the required signal-to-interference-plus-noise ratio (SINR) for each link. Figure 23 shows different types of multiplexing of access and backhaul links in half-duplex.

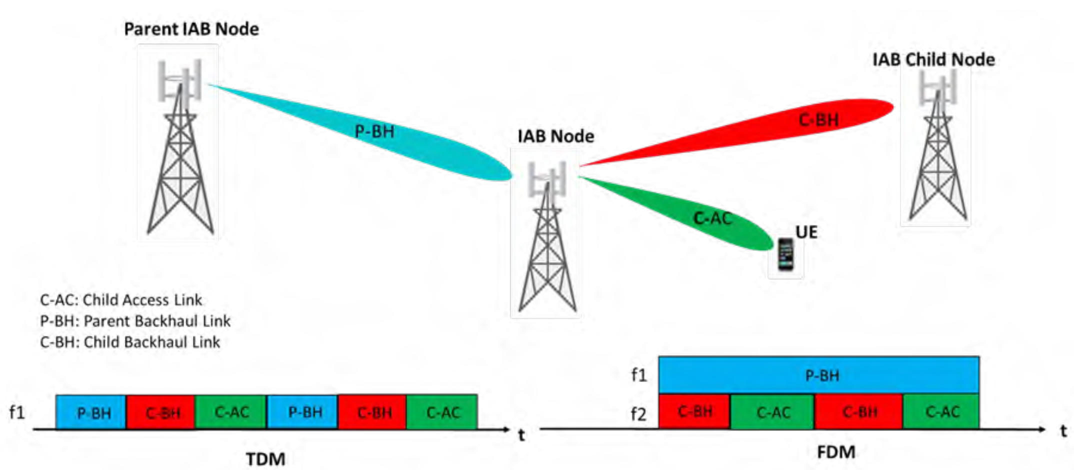


Figure 23 - Multiplexing of access and backhaul links in half-duplex

2.4.5 Full Duplex

All the above listed multiplexing schemes are subject to the half-duplex constraint and have their limitations. For example, TDM will cause additional relay delays. FDM requires much more spectrum to support the system. To improve spectral efficiency and reduce latency of the IAB network, full duplex is proposed as one of the enhancements to the IAB networks in 3GPP release 17. As shown in Figure 24 both backhaul and access links operate at the same carrier frequency, and the IAB nodes transmit and receive simultaneously, which will cause very strong self-interference between the nodes. As such, a self-interference cancellation mechanism must be implemented to address this challenge.

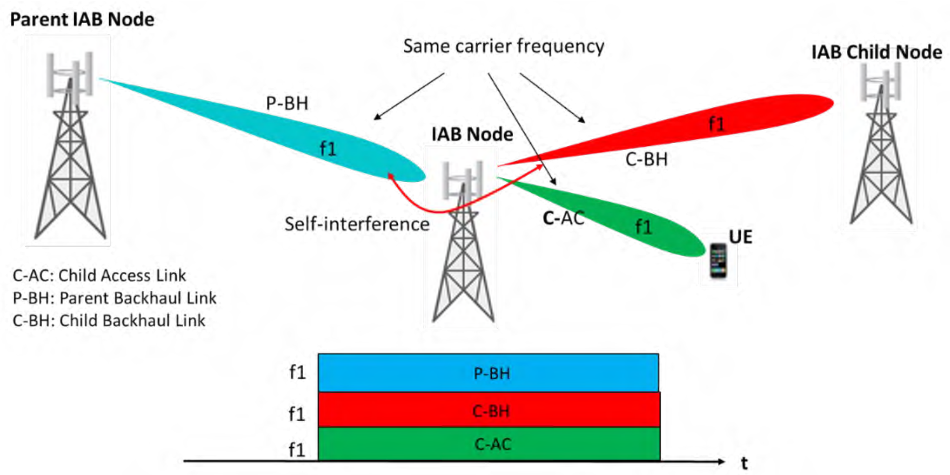


Figure 24 - Multiplexing of access and backhaul links in full duplex

One deployment scenario where full duplex may be more realistic is when there is a higher level of spatial isolation between the DU and MT parts of an IAB node, for example when IAB-based outdoor-to-indoor coverage is provided by a “distributed” IAB node with its MT part on the outside of a sufficiently isolating wall and its DU part on the inside.

Self-interference may not be totally suppressed by antenna separation techniques alone. Self-interference cancellation (SIC) algorithms may also be needed in both analog and digital domain. There are several studies and research ongoing. Some details of SIC algorithm is covered later in Emerging/ Future Technologies chapter.

2.4.6 IAB performance evaluations

There are different factors that should be taken into consideration for IAB deployments in order to achieve the objective of fast time-to-market, superior performance and reduced cost. Some of these considerations discussed later in this chapter are related to the impact on network performance due to IAB node introduction, multiplexing methodology, topology approach, etc. One other important aspect is the distance between the IAB donor and the IAB node and between IAB nodes. If the IAB node and IAB donor are too close to each other, the benefits of adding the IAB node are not fully realized as the coverage area is not extended by much and in addition the spectrum of the access link and backhaul link need to be shared, resulting in less capacity for the access link. On the other hand, if the two nodes are too far from each other, the backhaul link cannot sustain a high data rate, therefore also reducing the throughput that can be offered by the IAB node cell. Actual design will be dependent on the operators’ requirement. For example, the design may vary depending on the operators’ strategy to have more throughput in cell edge or just the basic coverage. In addition to the distance, other factors are at play, such as line of sight between the IAB node and IAB donor. The quality of the backhaul link will limit the throughput of the UE served by IAB-node.

This section contains evaluation results for IAB contributed during the 3GPP IAB Study Item and included in [9] based on the following assumptions:

Carrier Frequency	30 GHz
Bandwidth	400 MHz
Topology	7 Donors, 63 IAB Nodes
Channel Model	Dense Urban Micro
User distribution	Uniformly random, 100% outdoor
Antenna Array	Donor/IAB node = 16x16 array, UE 4x4 array

The table below shows the end user throughput for two different IAB resource allocation approaches: TDM only and TDM + SDM, relative to a deployment with only Donor nodes and no IAB nodes. The results illustrate that filling in coverage holes with IAB nodes can bring large improvements in user throughput for cell-edge and median users when compared to deployments without IAB nodes. Additionally, the 'TDM + SDM' multiplexing approach further increases the gains compared to the 'TDM-only' allocation approach by further increasing the efficiency of the resources split between access and backhaul links.

Table 2 - Evaluation results of different IAB allocation approaches compared to non IAB baseline

Scenarios	Multiplexing Scheme	5%-tile UPT	50%-tile UPT
7 Macro BS	N/A	0.50 Mbps	1.32 Mbps
7 Macro BS + 63 IAB nodes	TDM	6.99 Mbps	311.11 Mbps
	TDM + SDM	34.27 Mbps	522.85 Mbps

In addition to the resource multiplexing approach, the IAB topology itself can have a significant impact on the performance results. Figure 25 and Figure 26 below illustrate two example topologies for a deployment with 3 Donors and 54 IAB nodes based on the following methodologies:

- Max RSRP (baseline)
- Limit of 3 directly connected IAB nodes

As can be seen in these figures, the two methods generate very different topologies in terms of hop order distribution (with hop order 0 being the donor nodes). Each colored line indicates an IAB hop. As can be seen from the Figure 25, while max RSRP baseline produces fewer number of hop orders, but it results in multiple hops from one parent IAB node. While in case of limiting directly connected IAB node to 3, results in a spread out IAB topology as depicted in Figure 26.

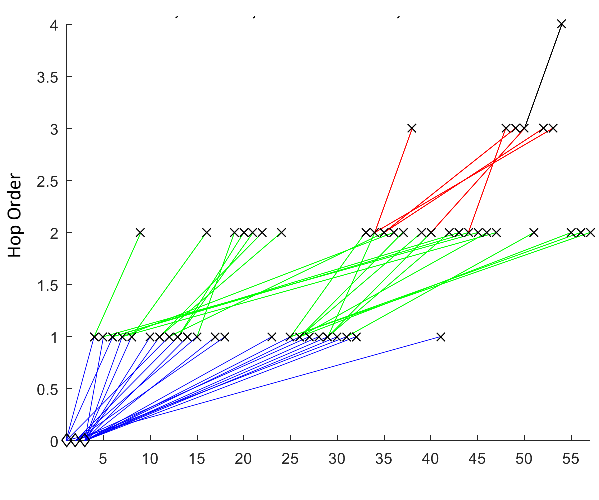


Figure 25 - Hop order distribution for Max RSRP topology formation methodology (Source: AT&T)

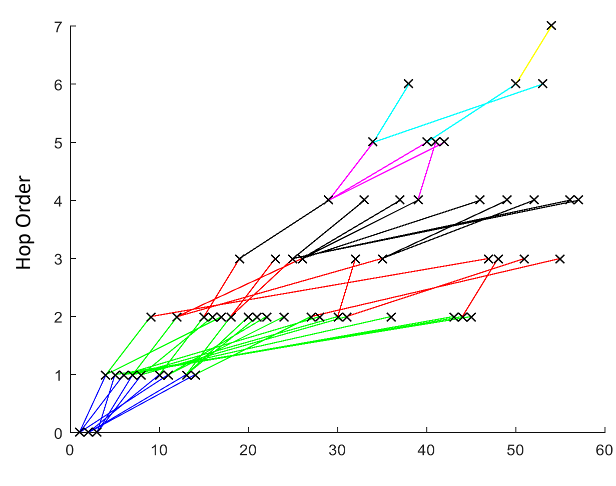


Figure 26 - Hop order distribution for Child Limit topology formation methodology (Source: AT&T)

As shown in Table 3 below, although the average number of hops is increased, the topology with the child limit of 3 IAB nodes per parent has 10x better 5th percentile user perceived throughput (UPT). It also has 75% better DL average throughput compared to the baseline method where IAB nodes connect to the parent node with the largest RSRP. This is due to less congestion on the initial backhaul hops from the wired donor nodes and the load is better balanced and spread out across the IAB topology.

Table 3 - User perceived throughput comparison of different IAB topology

Topology formation methodology	5%-tile UPT	50%-tile UPT
Max RSRP	8 Mbps	160 Mbps
Limit of 3 IAB nodes per parent	80 Mbps	280 Mbps

2.5 IAB Topology Adaptation, Routing Management & QoS Handling

There are several important elements to be considered in an IAB deployment, which include the establishment and management of the network topology, as well as the determination of Quality of Service (QoS) handling for maximum user experience.

2.5.1 IAB Topology Adaptation, Routing Management

To ensure efficient IAB network operations, the initial topology as well as the topology adaptation procedures are of utmost importance. This is mainly because the end-to-end performance of the overall network strongly depends on the IAB network topology, including: the number of hops between the donor and the IAB nodes; how many children and descendant nodes each node has to serve; and the strategies adopted for procedures such as network formation, route selection, and resource allocation.

2.5.1.1 IAB Node Integration

The IAB node integration procedure is performed in three phases. The overall procedure for IAB node integration is shown in Figure 27 below (from 3GPP TS 38.401).

In the first phase, the IAB node mobile terminal (MT) connects to the network as a normal UE. In doing so, the MT of an IAB node makes use of the synchronization signals transmitted by the already integrated nodes to estimate the channel and select its potential parents. A potential parent is identified based on an over-the-air indication from either the IAB nodes or IAB-donor-DUs, which is, for example, transmitted in the system information block (SIB). It identifies a parent node (another IAB node or an IAB donor) by performing Reference Signal Received Power (RSRP) / Reference Signal Received Quality (RSRQ) Radio Resource Management (RRM) measurements. The MT then performs random access and transmits a Radio Resource Control (RRC) connection setup request to the central unit (CU) via the parent node. Following that, the backhaul Radio Link Control (RLC) channel for carrying Control Plane (CP) traffic to and from the IAB node is established.

In phase two, a routing update is performed, which includes configuration of BAP routing identifiers and updating of routing tables of the IAB donor DU and all IAB nodes on the path to the IAB node. This contains: the configuration of the BAP address on the newly integrated IAB node; the routing identifiers for the downstream direction on the IAB-donor-DU; and the BAR routing identifiers in the upstream direction on the newly integrated IAB node's MT functionality.

In phase three, which is the IAB DU setup phase, the DU functionality of the newly integrated IAB node is configured. This consists of the transport network layer establishment and the F1-C connection setup between the IAB node and the IAB donor CU. Once this is completed, the IAB node can provide service to UEs.

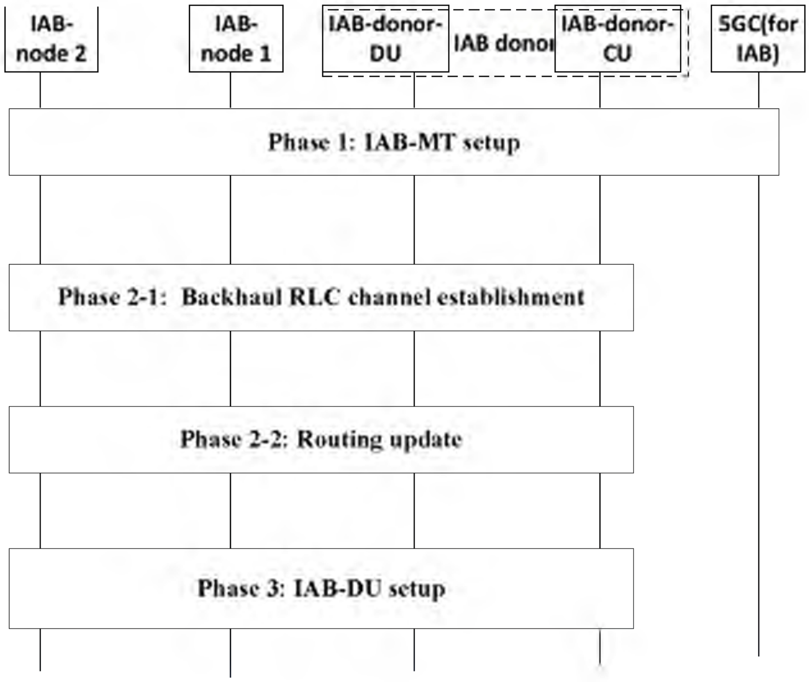


Figure 27 - Integration procedure for IAB node (TS 38.401)

Over the air (OTA) synchronization is supported in multi-hop IAB network for both Frequency Range Two (FR2), those serving frequency bands from 24.25 GHz to 52.6 GHz, as well as Frequency Range One (FR1) systems below 7.225 GHz. Considering the cell size of FR2 is expected to be smaller than FR1, timing advance (TA) of OTA synchronization signal allows up to five (5) hops. In case of FR1 TA-based OTA synchronization may not be enough to support multiple hops.

2.5.1.2 Topology adaptation

In addition to the initial access, 3GPP discusses the procedures that autonomously reconfigure the backhaul network topology. Topology adaptation has the goal to change the IAB network topology to ensure that each IAB node can continue to operate (including providing coverage and end user service continuity) even if the current active backhaul link fails. In addition, it is also desirable to minimize service disruption and packet loss during this procedure. IAB topology adaptation can be triggered by multiple incidents, including: i) the integration of a newly activated IAB node to the network; ii) the detachment of an IAB node from the topology; iii) the detection of backhaul link overload; iv) deterioration of the backhaul link quality or link failure; and v) other events such as blockage or congestion.

During topology adaptation, the network needs to determine an updated topology and then activate or deactivate links to achieve it. To this end, the following tasks are performed:

- Information collection over a sufficiently large area of the IAB topology, e.g., information on backhaul link quality, load, and signal strengths
- Topology determination: deciding on the best topology based on the information collected
- Topology reconfiguration: adjusting topology based on the result of topology determination through establishing new connections, releasing other connections, changing routes, etc.

2.5.1.3 IAB routing mechanisms

When the IAB network assumes a directed acyclic graph (DAG) topology, multiple routes can exist between two nodes in the network. This multi-connectivity or route redundancy can be used for back-up purposes, in case of backhaul link failure or node congestion. It is also possible that the redundant routes between the source and destination nodes are used concurrently, to achieve purposes like load balancing, reliability enhancement, etc. The route selection and optimization should consider the long-term network performance, but more dynamic routing decisions should also be made possible to accommodate short-term blocking and transmission of latency-sensitive traffic across backhaul links.

When new IAB-nodes connect to the network or when the topology changes, the relevant routing decisions (routing table) are also renewed by the IAB donor. From a packet's perspective, a mechanism is established within the IAB network to help forward it via multiple intermediate IAB-nodes between the IAB-donor and a specific UE. It includes the selection of route in case multiple concurrent routes exist between the source and destination, and the selection of next-hop destination at each IAB node once a route is selected.

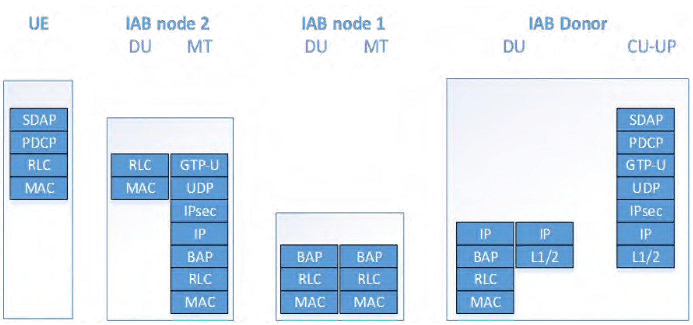


Figure 28 - IAB protocol stack. (Source: Intel)

Each IAB node is assigned a unique address by the donor (called the BAP address) and based on these node addresses a routing table is configured by the donor CU for each direction (UL/DL) at each node to direct the flow of traffic.

When an upper layer packet enters the IAB network, a BAP header is added by the first node with a BAP protocol function to form a BAP packet (that is, the donor DU for DL traffic, and the IAB access node for the UL traffic). The BAP header includes a BAP routing ID, which consists of a BAP address and a BAP path ID (see Figure 29, where the orange part of the data packet is the header).

The BAP address is used for identifying the destination node for the packet in the IAB network, which is the donor in the case of UL traffic, and the IAB access node in the case of DL traffic. The BAP path ID is used for selecting a path for the packet, from possibly multiple paths to the destination node.

At each IAB node along the selected path, if the current IAB node is already the destination of the BAP packet, then content of the packet is delivered to the upper layer. Otherwise, the BAP routing ID of the packet is used to look up the routing table to decide where it should be forwarded.

The routing table also contains a mapping that maps the BAP routing ID of a packet to the BAP address of its next hop destination. If backhaul radio link failure occurs on the selected path, then a next hop address mapped to a different BAP routing ID with the same destination BAP address can be chosen. The BAP packet is then forwarded to the selected next hop IAB node.

D/C	R	R	R	DESTINATION	Oct 1
DESTINATION (cont.)				PATH	Oct 2
PATH (cont.)					Oct 3
Data					Oct 4

Figure 29 - BAP data packet format. (Source: 3GPP TS 38.340)

The IAB routing mechanism can support multiple paths between the donor and an IAB node. The routing table can have entries with different BAP routing IDs, but the same destination BAP address mapped to different next hop BAP addresses. Thus, by setting different BAP path ID fields in the corresponding routing IDs, different packets of the same data stream can be forwarded through different paths.

2.5.1.4 Flow control over multi-hop routes

Due to the multi-hop nature of IAB networks, data congestion may occur on intermediate IAB nodes along a transmission path due to the differences of their effective link capacities. Different wireless links may have different effective SINRs, and different nodes may have different loads depending on the number of associated local UEs and child nodes, as well as their corresponding traffic patterns. Although the congestion can be handled by higher layer protocols, e.g., TCP, the scope of the impacted nodes will extend well beyond the RAN/IAB network. In addition, if packets are dropped due to congestion in the IAB network, the TCP congestion avoidance and slow start mechanisms may be triggered, and the end-to-end performances can be significantly impaired.

Therefore, flow control within IAB networks is supported for both uplink and downlink directions in order to avoid congestion-related packet drops on IAB-nodes and IAB-donor DU. For the downlink direction, this is straightforward: an upstream node can reduce its data rate toward the downstream congested node. For flow control in the uplink direction, however, data rate reduction is achieved differently: a parent node reduces the uplink resource allocation of its child node/UE along the transmission path.

Depending on the availability of congestion feedback information, the available flow control options are also different for downlink and uplink. For the downlink direction, the F1 interface between the UE's access IAB node DU and the IAB donor CU can be used to feed back the downlink data delivery status. If congestion along the path is detected, the IAB donor reduces the data volume transmitted towards the congested node. This type of flow control is called end-to-end flow control since it controls the data rates solely at the donor (the intermediate nodes do not have the feedback information), and it is not an available option for the uplink traffic.

Although alleviating the downlink data congestion problem to some extent, this end-to-end mechanism may be slow as it falls short of identifying the exact link/node on which congestion is occurring. If congestion information can be provided by the congested node to all IAB nodes along the path of transmission, e.g., through hop-by-hop forwarding of a BAP layer message, then all intermediate nodes can also react to the feedback information and perform flow control. This type of flow control is called hop-by-hop flow control and it is suitable for both downlink and uplink traffic once the required feedback mechanism is integrated into the corresponding protocol.

This mechanism can ease the congestion problem locally and reacts faster than the end-to-end method, but it requires new signaling to be added to the standards. The number of hops for the transmission of the congestion message can also be limited to constrain the number of intermediate IAB nodes performing flow control. For example, if the message can only be forwarded by one hop, then this is called one-hop flow control.

The different types of flow control can also be combined to achieve better performance, for example, the end-to-end flow control can be supplemented by one-hop flow control, or the hop-by-hop option. There could also be different granularity of the congestion feedback information, e.g. per UE radio bearer, per RLC-channel, or per backhaul link.

2.5.2 Quality of service (QoS) handling and Bearer Mapping

End-to-end management of QoS is a critical aspect of the 5G stand-alone (SA) architecture. This allows support of network slicing, thereby offering differentiated QoS treatment to different slices as well as within a slice. It is therefore essential that the IAB solution is also able to handle QoS while scheduling and mapping user data in the backhaul RLC channel.

There are two options considered for multiplexing UE data radio bearers (DRBs) to the backhaul RLC channels: one-to-one mapping and many-to-one mapping.

One-to-one mapping (Figure 30):

In this option, each UE DRB is mapped onto a separate BH RLC-channel. Further, each BH RLC-channel is mapped onto a separate BH RLC-channel on the next hop. The number of established BH RLC-channels is equal to the number of established UE DRBs.

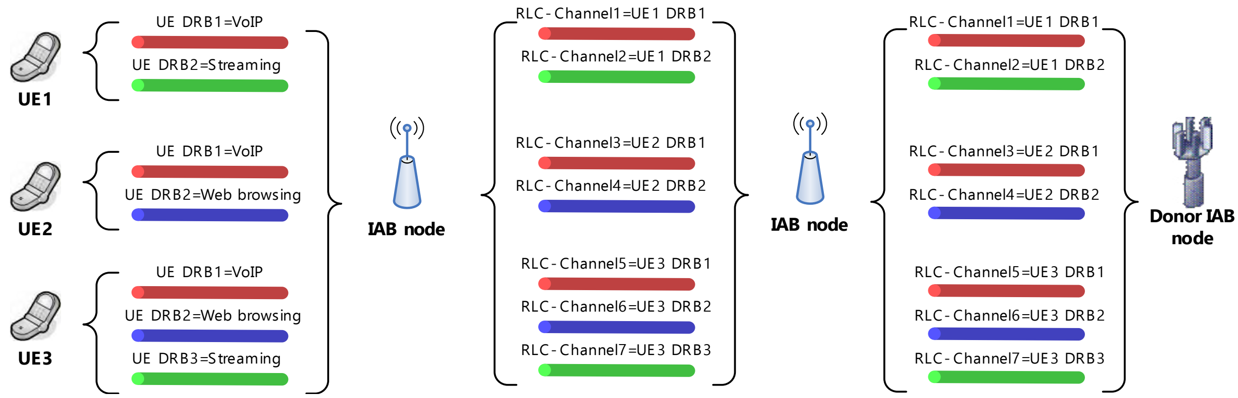


Figure 30 - Example of one-to-one mapping between UE DRB and BH RLC-Channel

Many-to-one mapping (Figure 31):

For the many-to-one mapping, several UE DRBs are multiplexed onto a single backhaul RLC-channel based on specific parameters such as bearer QoS profile. Other information such as hop-count could also be configured. The IAB-node can multiplex UE DRBs into a single BH RLC-channel even if they belong to different UEs. All traffic mapped to a single BH RLC-channel receives the same QoS treatment on the air interface.

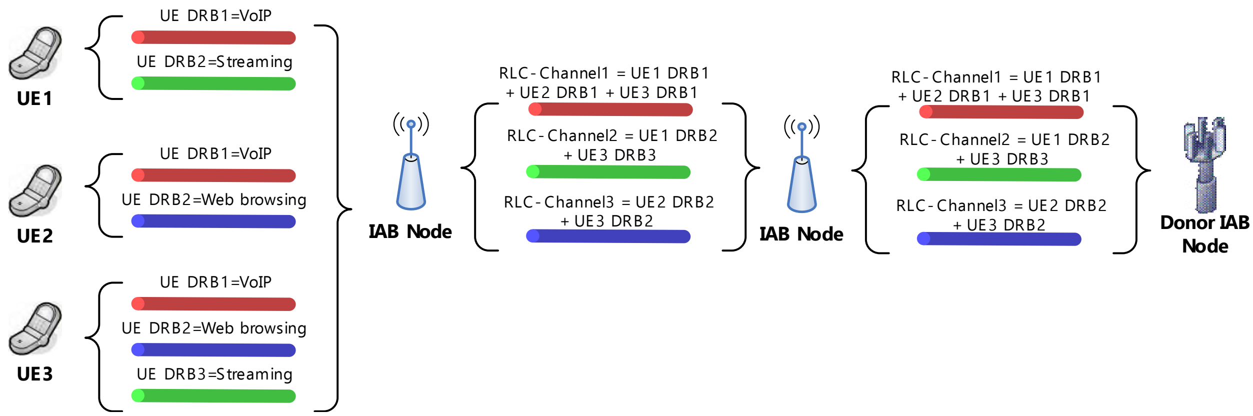


Figure 31 - Example of Many-to-one mapping between UE DRB and BH RLC-Channel

2.6 Emerging/Future Technologies

This chapter aims to cover various study or research ideas in the industry aiming to further enhance IAB performance and utility over and above what is so far planned in 3GPP study items. The solutions listed below are at research level and they may be part of the standard in future release of 3GPP.

2.6.1 Self interference Cancellation (SIC) algorithm

IAB in Full Duplex deployment scenario would need self-interference signal to be suppressed below the noise level to ensure satisfactory system performance. In addition to high level of spatial and polarization separation between access and backhaul links, added isolation can be achieved through a circulator. As shown in Figure 32, a circulator has multiple ports. Signals applied to port 1 will only exit from port 2, while signals applied to port 2 will only exist from port 3. If a transmitter is connected to port 1, and a receiver is connected to port 3, the transmitted signal will not enter the receiver.

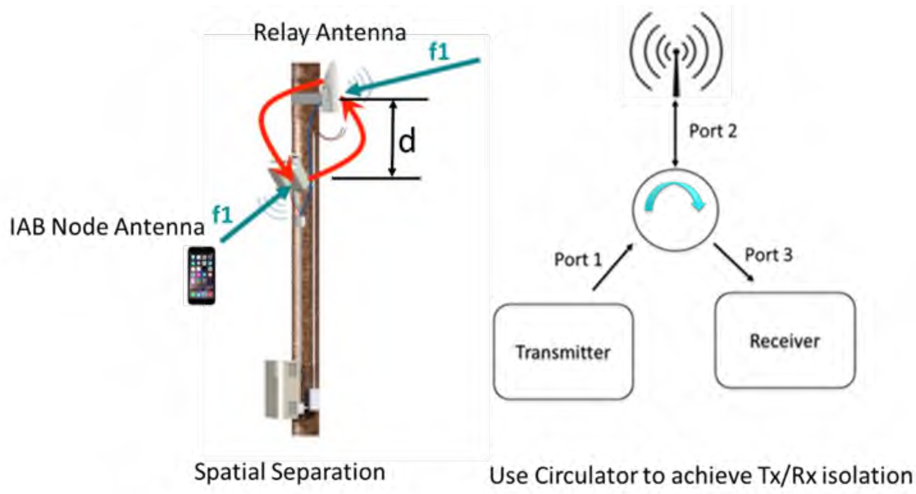


Figure 32 - Antenna separation to reduce self-interference

In general, the overall analog suppression of the self-interference signal prior to the receiver chain must be sufficient to meet the following requirements:

- To prevent receiver saturation, the power level of the residual interference should not be too high for the receiver LNA (Low-Noise Amplifier).
- The dynamic range of the ADC (Analog-to-Digital Converter) should be high enough to capture the weak received desired signal plus the residual self-interference with sufficient precision.

Most of the current available analog SIC modules are designed based on the assumption that the self-interference signal is known to the receiver because the transmitted signal is directly looped back to the receiver, so no training signal is needed.

As shown in Figure 33, the transmitted signals from multiple antennas are fed into the RF SIC module for interference signal estimation. The estimated interference signals are subtracted from the received signal directly. As another SIC strategy, a digital SIC module is implemented to further reduce the residual interference. Self-interference cancellation in the digital domain applies DSP techniques to the received signal after the ADC.

Again as shown in Figure 33, the transmitted digital signal is fed into the digital SIC module. To estimate the interference, the input signal passes through multi-tap channel filters to add the channel impacts. The output is then subtracted from the received signal in digital domain. Compared to analog filters for interference cancellation in RF domain, it is much easier to implement sophisticated algorithms to achieve better interference suppression.

However, there are also disadvantages about digital SIC. The dynamic range of the ADC greatly limits the maximum achievable interference cancellation. The quantization noise as well as phase noise will also impact the achievable cancellation. For example, a 14-bit ADC (Effective Number of Bits – ENOB of 11 bits) will have a signal-to-quantization noise ratio of 54 dB at the input of the ADC, which limits the maximum achievable interference cancellation to 54 db.

When using all three SIC mechanisms together, it is possible to meet the required interference cancellation of 131.5 db. Assume 50 dB can be achieved through antenna separation, 50 dB can be achieved by using digital SIC module, the requirement can be met if the analog SIC can achieve 31.5 dB cancellation.

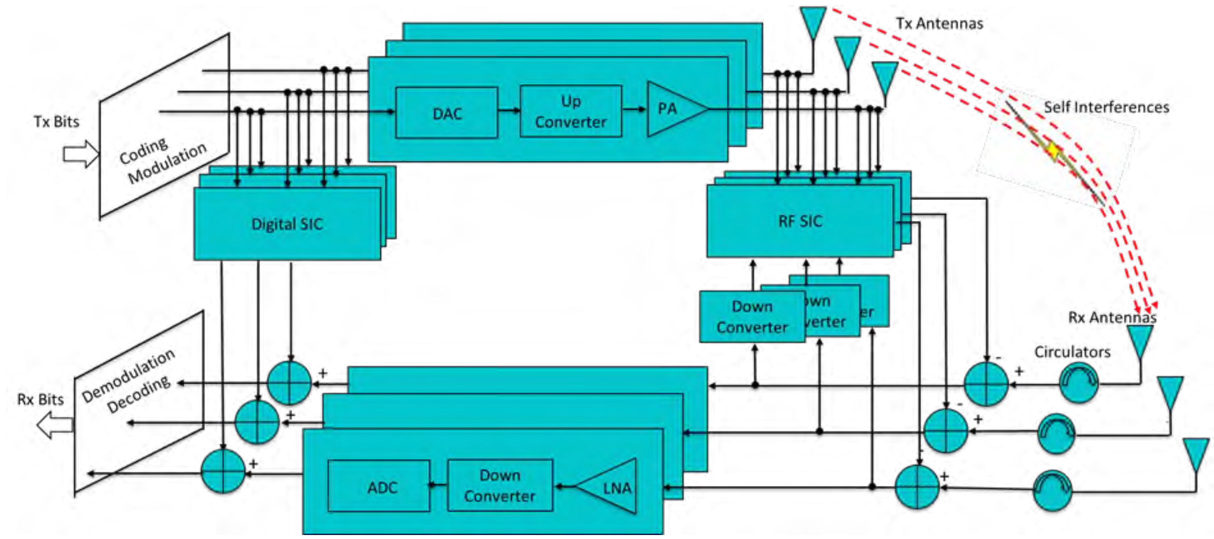


Figure 33 - Self-interference cancellation for full duplex system

To further reduce the impact the self-interference, transmit and receive beamforming algorithms can be used together with full duplex. For example, the parent IAB node can use transmit beamforming to form a narrow beam pointing directly to the child IAB node and reduce the back lobe of the beam so that the antenna gain for the desired signal can be maximized and the antenna gain for the self-interference signal can be minimized. In case of uplink transmission, the IAB node can also use IRC (Interference Rejection Combining) algorithms to maximize the received desired signal and minimized interference.

2.6.2 Network coding for reliability and latency

In the novel multi-hop, multi-route RAN environment of IAB networks, the existing solutions for the end-to-end reliability and latency guarantees are challenged. In the current 5G NR standards, retransmission-based techniques, such as Hybrid Automatic Repeat Request (HARQ) at the PHY layer and Automatic Repeat Request (ARQ) at the RLC layer are used to provide link-level reliability. While these technologies are excellent for protecting single-link packet losses, they are not designed with a consideration for multi-hop multi-route network topologies, and thus may perform sub-optimally on IAB networks.

For example, on a multi-hop path the feedback delay can accumulate over the hops and cause high end-to-end latency because, for each hop, packet retransmissions need to wait for the feedback of an Acknowledgement (ACK)/ Negative Acknowledgement (NACK) packet. Depending on the connections between nodes, there can be multiple routes between the source and destination. However, link-level techniques like HARQ and ARQ cannot make use of the multi-route diversity. Some higher-level, end-to-end, strategies such as packet duplication on the PDCP layer can make use of this multi-route diversity; however, such repetition-based schemes are not spectrally efficient.

In this context, linear network coding [11] can be considered as a solution to improve the end-to-end latency and reliability performance for IAB networks, taking advantage of the more complex network topologies. Network coding consists of adding redundancy at the upper layers and could be a superior alternative to the packet repetition schemes in such multi-hop, multi-route networks. Each data packet at the source is partitioned into equal-sized segments, to which network coding (linear combination of segments) is applied. At the destination, as long as the number of received linearly independent network encoded segments is larger than or equal to the number of segments in the original partition, the original packet can be correctly recovered, regardless of which segments are actually received.

In multi-hop/multi-route networks, network coding overcomes the limitations of the aforementioned schemes: first, on a multi-hop path, the feedback latency accumulation is eliminated by replacing ACK/NACK feedback with proactive network coding redundancy; second, we can make use of multi-route diversity, by sending network-coded segments of a packet via multiple routes to a given destination, thereby treating the multiple routes jointly as a single data pipe. Events such as blockage/congestion of a single route results in a "narrowing of the pipe" but not a complete stoppage of data transfer. If the destination receives enough encoded segments, the packet can be recovered via network decoding. It is also more spectrally efficient than PDCP duplication since the packet recovery criterion for network coding is much easier to satisfy than packet duplication. That is, network coding needs to transmit much less redundancy than PDCP duplication to achieve the same reliability target.

2.6.2.1 Machine learning and artificial intelligence

Recently, machine learning (ML) and artificial intelligence (AI) have attracted significant interest due to the increase in the size of data collected and cheap computational resources. In general, ML and AI offer an efficient alternative when conventional domain knowledge-based engineering cost and development time are high, and the problem is too complex to model and analyze. In addition, ML has the advantage of exploiting new data (change in environment) to enable easy to install, self-configuring and high performing IAB networks.

In light of the above promises, ML can be useful for the design and intersection of different protocol stacks to satisfy quality of service requirements in single and multi-hop scenarios when mixed traffic flows coexist. In addition, radio resource management (e.g., in-band and out-of-band spectra, beam management) in IAB networks between the access and backhaul links can be handled by ML to avoid congestion and interference and improve load balancing, end-to-end latency and throughput. In addition, ML/AI assisted topology adaptation (e.g., path selection, dual connectivity, multi-hop networking) methods can be designed to exploit and forecast changes in the environment (e.g., link failure, blockage) and user behavior (e.g., mobility, social events etc.).

2.6.2.2 Fully digital beamforming for mmWave

One of the main challenges of the mmWave IAB networks is the design of low power mmWave devices with multiple antennas. In the literature, hybrid phased array architectures with a small number of RF-chains is considered as a solution to enable multi-user/multi-directional beamforming, but its performance (e.g., beamforming and beam tracking capabilities) is limited by the number of RF-chains.

In contrast, fully digital mmWave architectures bring all the advantages of the digital beamforming in terms of fast beam management and beamforming optimality. Although a fully digital architecture can provide optimal performance, it has the highest ADC power consumption for a given bit resolution and sampling rate. Furthermore, power dissipation at the input/output (I/O) interface between RFIC and BBIC increases linearly with the number of ADCs and their bit resolution.

Finally, a fully digital architecture also has the highest baseband processor power consumption as the complexity of the channel estimation and multiple-input multiple-output (MIMO) processing increases linearly with the number of the RF-chains. Therefore, an efficient digital beamforming architecture with low-bit processing and spatial compression [12] can be considered in the future to have faster beam tracking (e.g., automatic pointing) and management. This architecture will enable blockage resilience, high throughput, and adaptive IAB network deployment.

3. Wireline Transport Technologies

3. Wireline Transport Technologies

5G deployments will undoubtedly drive a dramatic increase in densification and transport requirements across the Americas. While existing dark fiber assets have supported deployments of LTE macro sites well, the densification requirements of mid-band and mm-Wave 5G would require unprecedented levels of new fiber builds for 5G xHaul (i.e., front, mid- and backhaul).

Fortunately, North America already has several multi-gigabit shared infrastructure solutions that are already deployed virtually down every street and to every building on the continent. The first such network is the hybrid fiber coaxial (HFC) network that is owned and deployed by U.S. cable companies and reaches 93% of American households. With the recent release of the new data-over-cable service interface specification (DOCSIS) 4.0 standard, HFC networks will soon be able to deliver multi-gigabit capacity. In addition to HFC, fiber-based wavelength division multiplexing (WDM) and passive optical networks (PON) are also extensively deployed by many operators. WDM and PON networks have ample capacity to support current and future 5G transport needs.

Finally, new advances in Ethernet also promise to simplify and reduce cost/complexity for wireless transport deployments and builds.

3.1 Hybrid fiber coaxial (HFC) and Low Latency XHaul (LLX) Wireline Technologies

Thanks to new innovations such as low latency xHaul (LLX) and cooperative transport interface (CTI), HFC and PON are becoming viable and highly strategic alternatives to dark fiber transport for 5G. Similarly, improvements and cost-reductions in Ethernet and WDM are already proving critical for wireless densification. Leveraging these existing shared network technologies promises to dramatically reduce deployment costs and time, allowing operators to significantly accelerate deployments across the Americas. This section covers new and significant wireline transport technologies in detail.

3.1.1 Hybrid Fiber Coax (HFC)

Today's modern cable operators manage extensive hybrid fiber coax (HFC) networks, which connect virtually every building across North America. As shown in Figure 34, HFC networks use fiber optic technology to transport video, voice and data traffic from centralized headends (or data centers) to optical nodes located in the surrounding neighborhoods. At the optical node, which is typically less than 500 meters from the customer or business, the optical signal is converted to an RF signal and carried over robust, shielded coaxial cables to customer premises. Similarly, RF signals travelling in the opposite direction are converted to the optical domain and sent to the headend.

An integral part of today's HFC networks is DOCSIS, an industry standard for delivering broadband data services, primarily internet access, over HFC networks. As shown in Figure 34, a DOCSIS cable modem termination system (CMTS) is located at the headend and connected to cable modem (CM) at the customer premise via the HFC network.

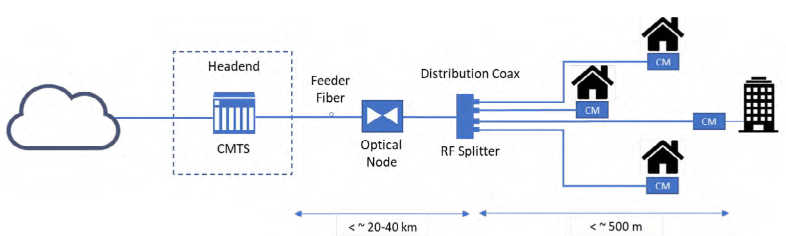


Figure 34 – HFC / DOCSIS Network Architecture

Since CableLabs first specified the DOCSIS standard in 1997 [13], the technology has evolved through six generations of progressive improvements over several key performance aspects, including capacity and latency. Table 4 shows the capabilities of the most recent DOCSIS standards. This table shows that the DOCSIS standard currently supports multi-gigabit downstream data rates with DOCSIS 3.1. With DOCSIS 4.0, downstream data rates over 10 Gbps are possible. Even higher data rates are expected in future releases of the DOCSIS standard.

Table 4 - DOCSIS Capabilities, Today and the Near Future

Requirements	DOCSIS 3.1 Today (2020)	DOCSIS Max (Future)	DOCSIS 4.0 (2023-2024)
Downstream spectrum Upstream spectrum	Shared spectrum with video 54 – 1002 MHz 5 – 42 MHz	Full spectrum through video reclamation 258 – 1218 MHz 5 – 204 MHz	Extending to 1.8 GHz, possibly 3 GHz 602 – 1794 MHz 5 – 492 MHz
DS capacity US capacity	8.5 Gbps 0.1 Gbps	8.6 Gbps 1.4 Gbps	10.8 Gbps 3.7 Gbps
Latency	Best Effort: 5 – 50 ms. With LLX / CTI: 1 – 2 ms (can be further reduced)		
Synchronization	Frequency sync only No time sync	Frequency + time sync through DTP	

A unique aspect of the DOCSIS technology is its ability to progressively expand both downstream and upstream capacity when and where it is needed. Traditionally, and even today, cable broadband services using DOCSIS share the same HFC plant as video services using frequency division multiplexing. As cable operators move away from traditional video delivery, they are in the process of reclaiming RF spectrum for broadband use. The result is higher downstream capacity as the cable industry moves towards DOCSIS 3.1 and 4.0.

The main market for DOCSIS has been residential broadband, where the bandwidth usage ratio between the downstream (DS) and upstream (US) has been asymmetrical. To efficiently utilize the limited RF spectrum, DOCSIS technology has been deployed with a corresponding asymmetrical DS-to-US capacity ratio. A similar asymmetrical DL-to-UL traffic ratio is prevalent in mobile networks, where it is also reasonable to allocate limited spectrum resources where they are needed, i.e. in the downlink. Even though today's residential broadband upstream usage demand is relatively small compared to the downstream, DOCSIS has the unique flexibility to increase the upstream capacity, as required, by re-allocating more RF spectrum to upstream traffic, as shown in the table.

Another unique advantage of DOCSIS is the ability to deploy additional capacity where it is needed. Although DOCSIS 4.0 has a theoretical capacity of around 10 Gbps, a 10-fold increase in capacity can also be achieved by node segmentation. Node segmentation is the process whereby an existing serving group is divided into smaller groups. For example, a service group with 500 households today can be segmented into serving groups with as few as 50 households. This represents as 10X increase in the average capacity per household. DOCSIS is also capable of supporting time sensitive traffic such as AR/VR applications with low latency xHaul (LLX) technology [14]. Today's DOCSIS deployment can achieve a minimum latency of 5 ms. Because the DOCSIS network is a shared medium, the latency increases as the network loading increases. The LLX aims to significantly reduce the DOCSIS latency to the 1-2 ms range, making mobile traffic less suspectable to network loading. This is discussed in further detail in Section 3.1.3.

Providing frequency, time and phase synchronization to the mobile network is another key requirement for the xHaul transport network. Today's DOCSIS network can provide frequency synchronization for LTE FDD. With additional enhancements to the DOCSIS protocol, the cable network can provide the timing and phase precision needed for 5G deployments. Synchronization is discussed further in Section 3.1.7.

3.1.2 Business Drivers for HFC Backhaul Deployments

Because of its ubiquitous deployment and the fact that coaxial cable can also deliver power, the last "few hundred meters" of coaxial cable is proving to be increasingly strategic for both today's wireless operators, as well as tomorrow's 5G networks. Already, HFC networks support hundreds of thousands of LTE small cells across the U.S. HFC provides the following strategic benefits for LTE and 5G backhaul:

- Ubiquity – HFC cables already run down virtually every street and to every home/business across North America. Being able to leverage this existing broadband backhaul offers a clear advantage in cost and deployment time compared to overbuilding the same infrastructure with fiber.
- Ample Capacity – HFC networks can already support multi-gigabit speeds using the mature DOCSIS 3.1 specification. The newly released DOCSIS 4.0 specification, now promises downstream speeds over 10 Gbps. This should provide ample capacity for most 5G transport use cases.
- Power – HFC networks offer a clear advantage over fiber in terms of power transmission. HFC networks are active, and able to draw power "from the line" to power small cell deployments. This can dramatically reduce cost and complexity of 5G densification, removing the need for separate power feeds and associated permitting.
- Construction & Permitting – Thanks to existing "strand" infrastructure, cable operators can very quickly connect small cells to their existing aerial plant. Deployment of a small cell can be accomplished in a matter of an hour, by tapping into the existing coaxial strand, and attaching a line-powered, strand-mounted small cell. This eliminates the need for any permitting and dramatically accelerates deployments.

In order to quantify these benefits, one North America cable operator completed a recent analysis of their plant to compare the construction cost and time to deploy LTE small cells using their existing HFC infrastructure, versus extending their already deep fiber infrastructure to support backhaul. They modeled the deployment of 15 small cells in a major urban center, which was highly representative of likely LTE or 5G small cell deployment locations. In their analysis, they found that all 15 preferred small cell locations were within 10 m of the existing coaxial cable plant. Because of this, the estimated civil build cost using coax was only $1,500 and could be completed in one week. Conversely, the fiber extension/build cost required to connect all of the small cells to their fiber infrastructure would have been $183,000 and would require 4-6 months to complete.

Table 5 - Backhaul build comparison for HFC vs fiber for a major North American network operator

Small Cell Count	Backhaul Option	Backbone Fibers	Estimated Civil Build Cost	Estimated Build Time
15	DWDM	1	$183k	4-6 Months
	Coax w. couplers	0	$1.5k	1 Week

This model shows that leveraging existing infrastructure can reduce build costs by up to 99% and accelerate builds by a factor of up to 20x.

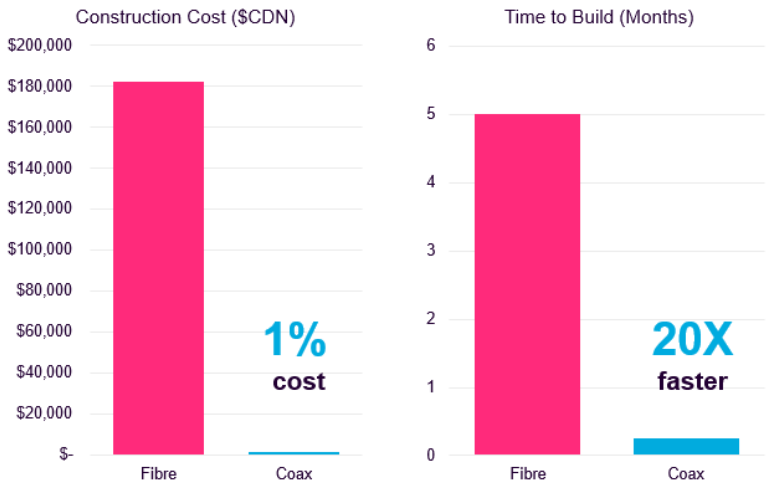


Figure 35 - Backhaul build comparison for HFC vs fiber for a major North America network operator

3.1.3 Reducing Latency for 5G Backhaul on HFC

While HFC provides many strategic advantages for 5G densification, latency and jitter are known challenges. Because HFC is a shared medium, it must schedule traffic across many endpoints. This can result in latency spikes and contention. Furthermore, DOCSIS may be challenged with the extremely high bandwidth and low latency requirements of protocols like the common public radio interface (CPRI) and enhanced CPRI (eCPRI).

The newly released low latency xhaul (LLX) specification [15] seeks to address these challenges by dramatically reducing the latency and jitter for backhaul traffic over existing HFC infrastructure.

LLX enables a range of use cases for cable operators. For a cable operator that provides wholesale xhaul transport for a mobile operator; or a converged cable operator who owns both cable and mobile operations and carries its own mobile traffic; LLX provides a means to significantly lower the latency of all traffic coming from the UE to levels comparable to fiber. Different traffic flows include signaling, IMS voice (data and signaling), low latency applications (mobile gaming), video conferencing applications (such as Zoom, Cisco WebEx and Apple FaceTime), and URLLC for 5G.

LLX has the following fundamental components:

- scheduler pipelining using the bandwidth report (BWR) message,
- a common QoS framework that matches the DOCSIS QoS to the mobile network QoS, and
- a grant sharing mechanism that allows the CM to perform real-time scheduling.

In addition, there needs to be system level configuration and operation to align the DOCSIS and mobile systems.

3.1.4 LLX Scheduler Pipelining

Scheduler pipelining is a very unique and inventive aspect of LLX and the heart of what creates a low latency transport. In a nutshell, LLX uses the decisions made by the mobile scheduler to inform the CMTS what is about to happen next.

Normally, mobile (LTE and 5G) and DOCSIS operate as two independent systems. As such, in a backhaul or xhaul situation, the end-to-end latency experienced by the mobile traffic is the sum of the two system latencies. With scheduler pipelining, however, the end-to-end latency is reduced significantly by initiating scheduling requests in parallel. This is shown in Figure 36.

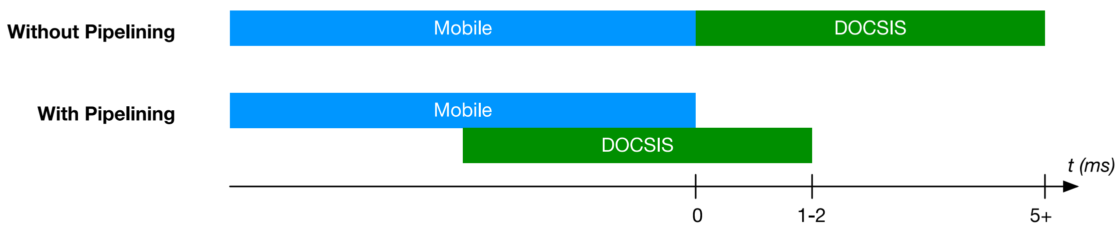


Figure 36 - Reducing Effective DOCSIS Latency

The two technologies have similar mechanisms for accessing the channel – both through a request-grant transfer loop. This is shown in the upper portion of Figure 37. In the case of LTE and 5G, the UE makes a bandwidth request (REQ-UE) to the base station and the base station scheduler computes a grant (GNT-UE) and sends it back to the UE. At the time indicated in the grant, the UE sends data to the base station, which is then forwarded to the DOCSIS CM. Similar to the mobile process, the CM makes a request (REQ-CM) and the CMTS scheduler computes a grant (GNT-CM) and sends it back to the CM. At the time indicated in the grant, the CM sends the data to the CMTS.

With scheduler pipelining, the CM request (REQ-CM) is replaced by a bandwidth report (BWR) message. The BWR is sent from the base station to the CM and contains information about the UE grant (GNT-UE), including the amount of data and time slot(s) allocated to the UE. Consequently, the two grant requests (i.e., REQ-UE and BWR) and the corresponding grants (i.e., GNT-UE and GNT-CM) take place in rapid succession. As a result, the request-grant latency is incurred only once by the end-to-end system. This is shown in the bottom portion of Figure 37.

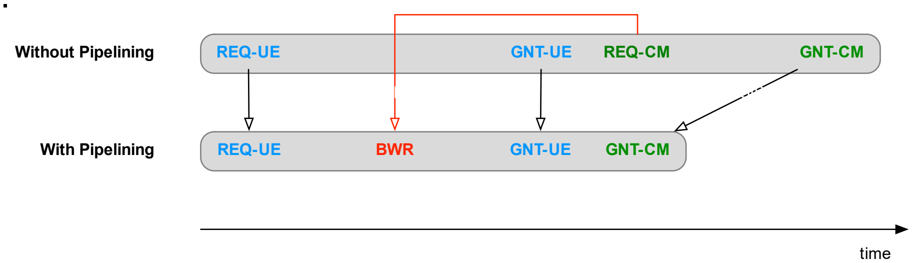


Figure 37 - Pipelining Requests and Grants

Figure 38 shows scheduler pipelining in more detail for an LTE system that is backhauled over a DOCSIS network. The two systems are both multi-point to point in the upstream, and both systems have upstream paths that are centrally scheduled. This means that both systems have an inherent latency due to the request-grant delay in the upstream. Without pipelining, there are two independent request-grant cycles, one in mobile and one in DOCSIS. The scheduler pipelining leverages the knowledge the mobile scheduler has in the terms of how many bytes it is expecting at a future point in time. The results of the mobile request-grant process, in the form of a BWR message, are then passed to the DOCSIS system so the CMTS can grant capacity to the CM directly without waiting for a native DOCSIS request.

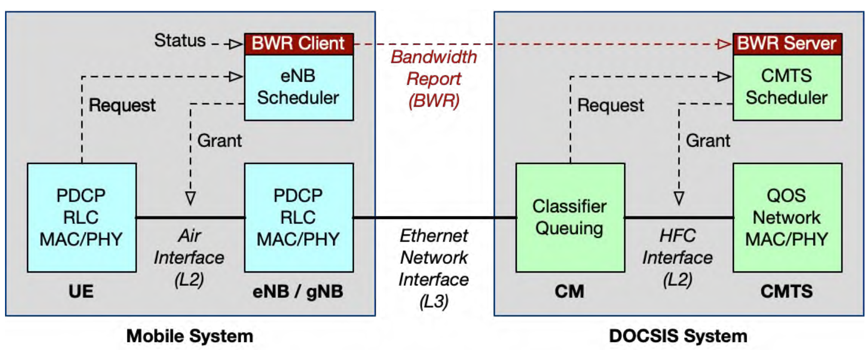


Figure 38 – Linking DOCSIS and LTE Schedulers with BWR

The scheduling pipelining can be illustrated through an example where the UE needs to send 1000 bytes of data. The UE sends a request to the eNB scheduler. The eNB scheduler responds with a grant indicating 1000 bytes 8 ms from a reference time. At the same time, the eNB scheduler knows that it can expect 1000 bytes in 8 ms and decides what will happen across the network interface it shares with the DOCSIS system. The eNB then sends a BWR message to the CMTS scheduler, indicating 1000 bytes will arrive on the shared network port 9 ms from the refence time, after adding a 1 ms of engineering margin to cover buffering or other internal delays. The CMTS scheduler now knows when the bytes will arrive at the CM, and determines when it needs to send a grant to the CM. The grant will arrive at the CM just in time as the UE data arrives at the CM from the eNB.

This concept can be applied to midhaul and fronthaul systems as well where the full stack eNB or gNB is replaced with a radio unit (RU), a distributed unit (DU), and a central unit (CU). In principle, the transport system can also be either a DOCSIS system or a PON system. In O-RAN Alliance terminology, a transport unit (TU) can be either a CM or an optical network unit (ONU), and a transport node (TN) can be either a CMTS or an optical line terminal (OLT). These network elements are shown in Figure 39.
For backhaul, the mobile scheduler is contained locally in the eNB or gNB, and the BWR messages travel over the transport network. For midhaul, the CU is centralized. This case is similar to backhaul in that the BWR messages travel through the transport network.

For fronthaul, the CU and DU are moved centrally. The scheduler is now centralized and the BWR messages do not traverse the same network that the data packets do. In the fronthaul case, the transmission time for the BWR message from the BWR client to the BWR server is typically much less compared to backhaul. This is good as the fronthaul case is usually associated with more stringent latency requirements.

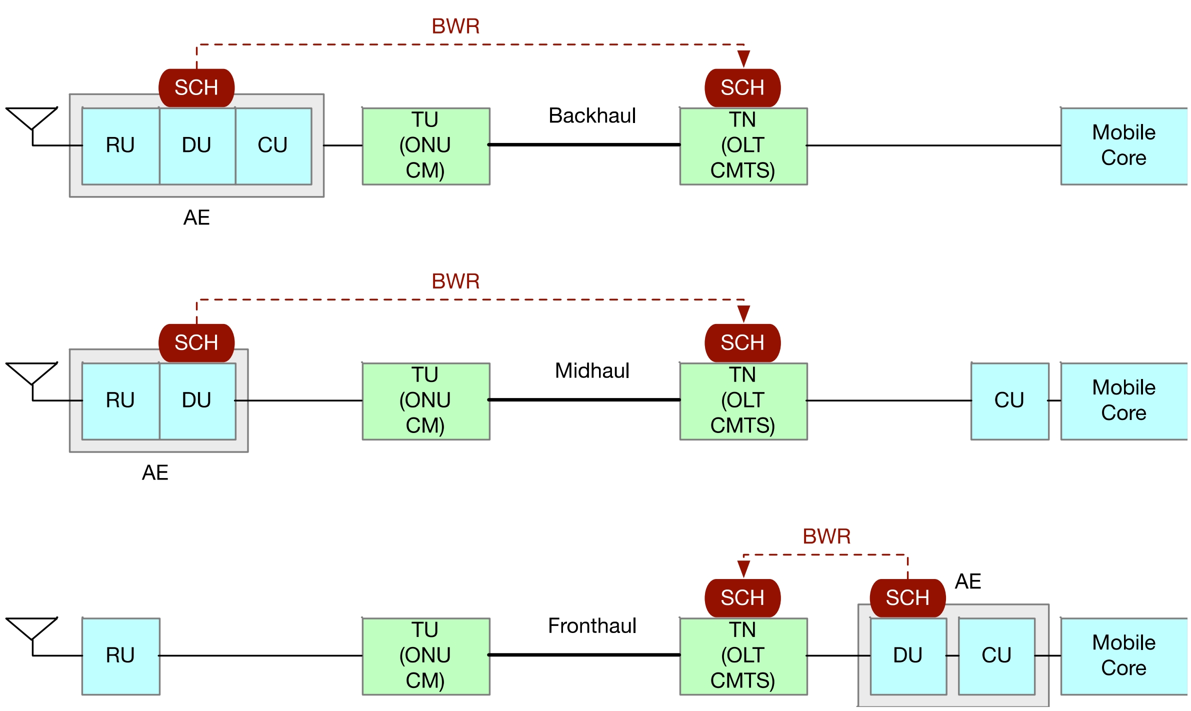


Figure 39 – Mobile Xhaul over DOCSIS Using BWR

Therefore, it is established that it is possible to take a flow of packets that are crossing the mobile air interface and move those across the transport network with lower latency. Next, we go into the detail at another level and see what happens when that stream of bytes and packets may be composed of multiple flows with multiple queues and multiple scheduling mechanisms.

3.1.5 LLX Common QoS Framework

If all traffic moves across an interface together, then all high priority traffic, such as signaling, and low priority traffic, such as a file transfer, would experience the same latency. If the latency of that interface was infinitely low and the bandwidth was infinitely high, then there would not be a problem.

But that is rarely the case. The mobile bandwidth is limited as is the bandwidth of the transport network. In such systems, the file transfers can saturate a system, causing buffers to fill up and system latency to increase. For this fundamental reason, it is common to separate traffic into multiple flows that are sorted or classified into multiple queues. This is a fundamental property of QoS. With a QoS mechanism, services such as mobile signaling or 5G URLLC can be provisioned to have the absolute minimum latency.

The air interface for LTE has four layer-2 request queues known as logical channel groups (LCGs). The UE requests capacity per LCG, the eNB grants capacity based on the requested bytes per LCG. Each eNB grant result is tracked separately within the BWR message as a BWR flow. In 5G, the number of LCGs per UE expands from 4 to 8 for even more granular QoS treatment. Ironically, in the fronthaul case, these LCGs are combined into one eCPRI flow, although there is work underway at O-RAN to separate some of the traffic out of the main eCPRI flows.

Consider the following scenario: all mobile traffic flows are aggregated together onto a common Ethernet and passed to the transport network. How does the transport network segregate the flows and put the right flows on the right queues? How are these queues associated with the BWR flows? How many transport grants and what level of priority should each queue get?

To address these system-level design problems, LLX proposes a common QoS framework between the mobile system and the transport system. There are many variations of how this could be done, but the most fundamental and pure system is to do the following:

- Use the same number of queues in the transport system as there is in the mobile system
- Use the same classifier mechanism in the transport system as there is in the mobile system
- Use the same policy/queue-weighting mechanism in the transport system as there is in the mobile system

To illustrate these design principles for LTE and DOCSIS backhaul example from the previous section, let's follow a packet through this system, as shown in Figure 40.

An application in the UE creates a packet. That application has an associated QoS class indicator (QCI). The packet and the QCI marking are placed into a radio bearer and into a logical channel. The logical channels are placed into one of four LCGs. Each LCG has its own request-grant exchange with the eNB scheduler and its own flow entry in the BWR message. When the packets are received at the eNB, they are placed into GPRS Tunnelling Protocol (GTP) tunnels. The GTP packet is marked with a differentiated services code point (DSCP) that is chosen based on the QCI.

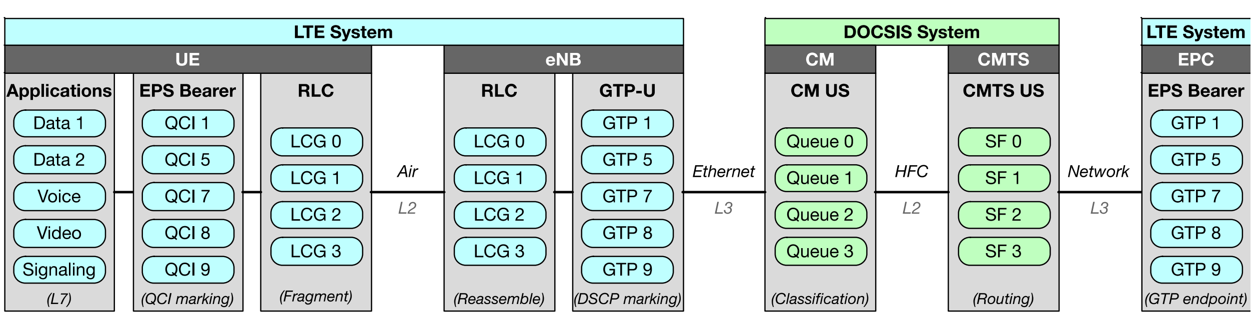


Figure 40 – QoS Model of LTE System with DOCSIS Xhaul

When the packets arrive at the DOCSIS system, they need to be classified into different service flows. If the CM is provided with the correct DSCP to service flow mapping, the contents of the original four LCGs can be recreated on the four DOCSIS service flows.

3.1.6 LLX Performance

The LLX technology was jointly developed by CableLabs and Cisco. As part of the project, CableLabs and Cisco built a proof of concept testbed using a Cisco cBR-8 CMTS and the open air interface (OAI) LTE RAN platform. We have previously published numerous test results with this test setup and reported that BWR achieves 1-2 ms of DOCSIS US latency with a low to medium channel loading on the DOCSIS network [15] [16] [17] [18].

In July 2019, Shaw Communications, Cisco, Sercomm, and CableLabs jointly conducted a BWR trial on a LTE system backhauled over a DOCSIS network. The LTE portion of the network used a Sercomm CBRS F208 small cell with BWR software that gathers the LTE scheduler outputs and generates the BWR messages. The DOCSIS portion of the network used a Cisco cloud native broadband router (cnBR) with a BWR API.

Various backhaul scenarios were tested, particularly for when the DOCSIS network was subject to different levels of utilization. Figure 41 shows DOCSIS latency under 70% of utilization (high). The blue curve is the DOCSIS latency cumulative distribution function (CDF) without BWR. The orange curve is the DOCSIS latency CDF with BWR enabled. At the 95th percentile, BWR reduces DOCSIS upstream latency almost an order of magnitude, from 22 ms to 2.5 ms.

The trial was conducted on the DOCSIS 3.0 Advanced Time Division Multiple Access (ATDMA) channel. With the newer DOCSIS 3.1 Orthogonal Frequency Division Multiple Access (OFDMA) channel, it is expected the absolute latency achievable by BWR can be reduced much further.

3.1.7 Synchronization over HFC

The mobile network is by nature synchronous. To achieve the goal of sharing a common clock, base stations often utilize synchronization from the Global Navigation Satellite System (GNSS). An equivalent global clock signal can also be transported over the IP network using Precision Time Protocol (PTP), which is described as IEEE 1588-2008.

The PTP is a two-step time transfer protocol that requires a time-delay symmetric network for the PTP slave to accurately derive the timing offset from the PTP master. Unfortunately, the DOCSIS network is asymmetrical, mainly due to the large upstream queueing delay relative to downstream delays and

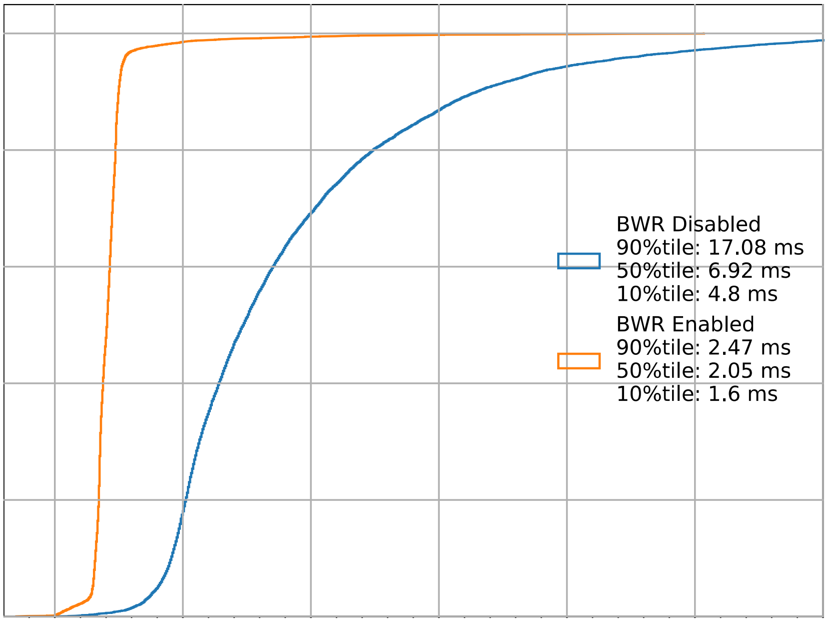


Figure 41 – Shaw BWR Trial – DOCSIS Latency with 70% (High) Loading

the difference in interleaver depths between the upstream and downstream.
If PTP messages are sent over the top of the DOCSIS network, the messages can experience variable buffer delay, which in turn can cause packet delay variation (PDV) and large time transport errors. While mitigation techniques such as assigning the PTP packets with higher priority DSCP can result in a workable solution for FDD deployments, a better way to propagate timing over the DOCSIS network is through the DOCSIS Time Protocol (DTP).

3.1.7.1 DOCSIS Time Protocol (DTP)

The high-level end-to-end timing system through a DOCSIS network is shown in Figure 42. Timing from the GNSS system is received by a primary reference time clock, which acts as a grand master clock, which generates PTP messages. The PTP messages are sent through one or more Ethernet switches which operate as telecom boundary clocks (T-BC).

The PTP is terminated when it arrives at the DOCSIS domain. The DOCSIS system is already a synchronous network with its own timestamp. The DTP algorithm is run between the cable modem termination system (CMTS) to determine the one-way downstream delay. The cable modem (CM) then adds the timing offset to its timestamp. The CMTS and the CM each behave as the equivalent of a T-BC and regenerate the necessary PTP functionalities. The CM appears as a PTP master to the mobile radio downstream and passes on the recalculated timestamp.

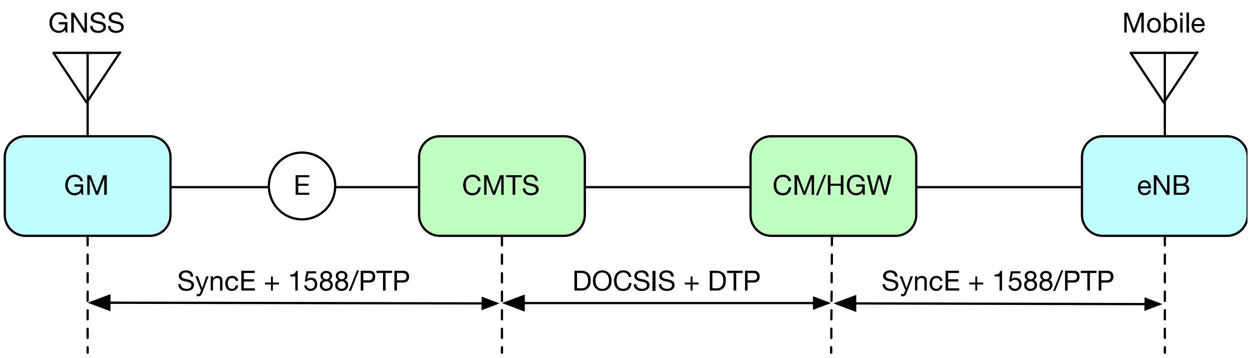


Figure 42 - Network Timing Deployment over a DOCSIS Network

CableLabs, the cable industry's international standardization organization, has completed the first version of the synchronization over DOCSIS specification [19], which includes the synchronization architecture and requirements specified for the DOCSIS network equipment. HFC equipment vendors have demonstrated the feasibility of DTP in proof of concepts.

3.1.8 Next Steps

LLX technology has been proven through commercial proof of concept trials to capable of achieving 1-2 ms of latency on the DOCSIS upstream and may be even lower with appropriate tuning of parameters.
CableLabs has standardized the LLX protocol, with the publication of LLX I01 [20] specification in June 2019, and a revision of the LLX specification I02 is expected in May 2020.

3.2 Passive Optical Networks

Like HFC, PON access networks have been deployed extensively throughout the Americas over the past couple of decades. PON is a fiber-optic network technology that uses a point-to-multipoint topology and optical splitters to deliver data from a centralized location to multiple user endpoints. In PON, only passive components are used in the end-to-end path. As such, electrical power is only required at the endpoints.

Figure 43 shows a typical PON architecture. A PON consists of an optical line terminal (OLT), an optical distribution network (ODN) and optical network terminals (ONTs), also referred to as an optical network units (ONUs). The OLT is located at a central office (or headend) and is connected to optical splitters in the field via optical feeder fibers. Distribution fibers connect the optical splitter to optical network terminals (ONTs) located in the customer premise, which is usually located within 500 m of the splitter. The optical link budget for a typical PON system limits distances between the central office and the customer premise to approximately 20km.

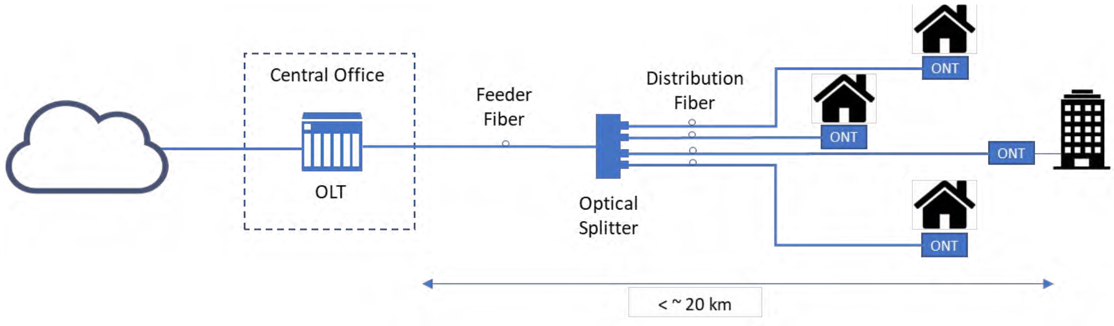


Figure 43 - PON Architecture

Two standards organizations have been responsible for developing PON standards: the ITU-T and the IEEE. The first ITU-T PON standard, gigabit PON (GPON), was approved in 2003. The first IEEE PON standard, ethernet PON (EPON), was ratified a year later. Although the initial standards supported gigabit speeds, both standards organizations have since issued new PON standards that offer speeds up to 10 Gbps and beyond, as shown in Figure 44 below.

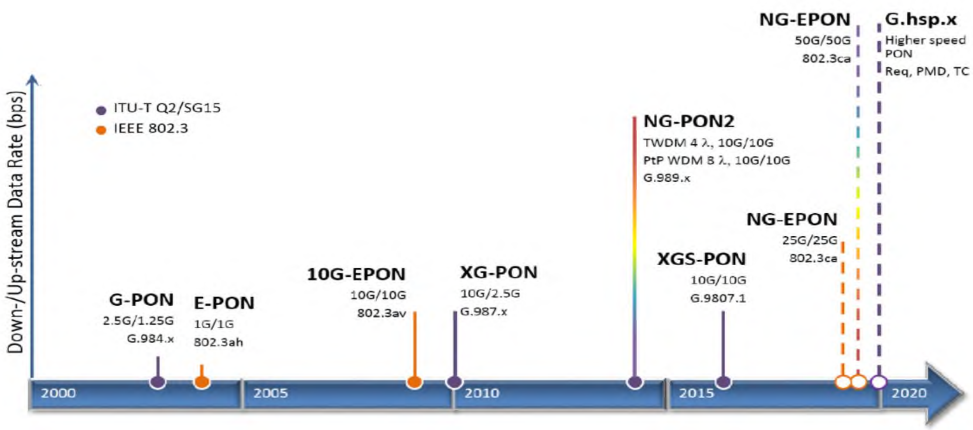


Figure 44 - PON Standards

The major differences between the ITU-T and IEEE PON standards lie in the line rates, Forward Error Correction (FEC) type, optical budget coding gain, packet segmentation (there is no segmentation of Ethernet frames in EPON), and other differences in the EPON MAC (Media Access Control) layer, which is analogous to the GPON transmission convergence (TC) layer.

3.2.1 Business Drivers for PON Backhaul Deployments

Both ITU-T and IEEE PON standards are used to provide fiber to the home (FTTH) and fiber to the premise (FTTP) and are widely deployed in many urban areas. In general, PONs shares many of the same advantages as HFC networks, including ubiquity, ample capacity, and ease of construction and permitting, due to the presence of existing support structures and access agreements.

A recent analysis by Bell Labs [21] showed that leveraging existing FTTH networks can decrease 5G transport costs by more than 50% compared to traditional solutions such as microwave and P2P dark fiber. The study considered the introduction of 5G small cells in areas where FTTH was already deployed and had sufficient spare fibers for small cell backhaul. The total cost of ownership (TCO) was calculated over a 5-year period for various backhaul technologies. The main findings of the study included:

- PON-based transport was the most cost-efficient 5G transport option.
- Compared with new construction of P2P fiber, PON transport was about 10-times less expensive.
- The cost advantage was the largest for operators that own the FTTH network. But even for operators that do not own the FTTH network, the cost savings 40% or more can still be realized.

3.2.2 GPON, XGPON, XGSPON, NGPON2 and Beyond

Since the initial GPON standard was issued over 15 years ago, three additional PON standards have been developed by the ITU-T. The four standards are listed in Table 6 along with their corresponding downstream (DS) and upstream (US) data rates.

The ITU-T PON standard enables a smooth capacity upgrade (evolution to higher capacity PONs) by using standardized system elements (e.g., co-existence elements in ITU-T G.984.5). The upstream and downstream wavelength plan also allows for the provision of 10 Gbps point-to-point wavelengths (P2P WDM) on the same ODN as the PON. Bitrates in the access network of 25 Gbit/s and above is still a research topic and are now being considered in ITU-T Study group 15 Question 2.

As noted in Section 1.1, depending on the functional split between the CU, DU and RU, the need for high bandwidth and low latency can be dramatically higher compared to the traditional backhaul case (i.e., all-in-one gNB). As shown in Figure 45, PON is an applicable technology for several function splits, including backhaul, HLS midhaul and LLS fronthaul.

In the past, PON solutions were not able to transport conventional CPRI fronthaul (or split option 8 interfaces) because legacy CPRI fronthaul is a stream-based interface scaling with the number of radio transmit/receive chains, with a constant and very high throughput, which requires P2P dark fiber transport.

The new 3GPP Release 16 and 17 services (i.e., URLLC and mMTC) require an even more agile topology based on virtualization of these RAN processes (CU or CU/DU). Mobile network operators (MNOs) are also migrating to highly programmable and scalable cloud-native services architectures with network functions disaggregation to enable more flexible services deployment. As a result, their network functions must be able to reliably communicate using multiple types of services over a common transport infrastructure. Agile service (re-) configuration and SLA guarantees will be vital enablers for this vision.

3.2.3 EPON, 10G EPON, NG-EPON

Like the ITU-T, the IEEE has also been actively developing new PON standards since the initial standard was issued in 2004. The two most recent standards are 10G EPON and NG-EPON. Table 7 lists the three IEEE PON standards and their corresponding downstream and upstream data rates.

Table 6 – ITU-T PON Standards

Variant	ITU-T Standard	Issued	DS Data Rate	US Data Rate
GPON	G.984x	2003	2.5 Gbps	1 Gbps
XG-PON	G.987	2010	10 Gbps	2.5 Gbps
NG-PON2	G.989	2015	4x10 Gbps	4x10 Gbps
XGS-PON	G.9807.1	2016	10 Gbps	10 Gbps

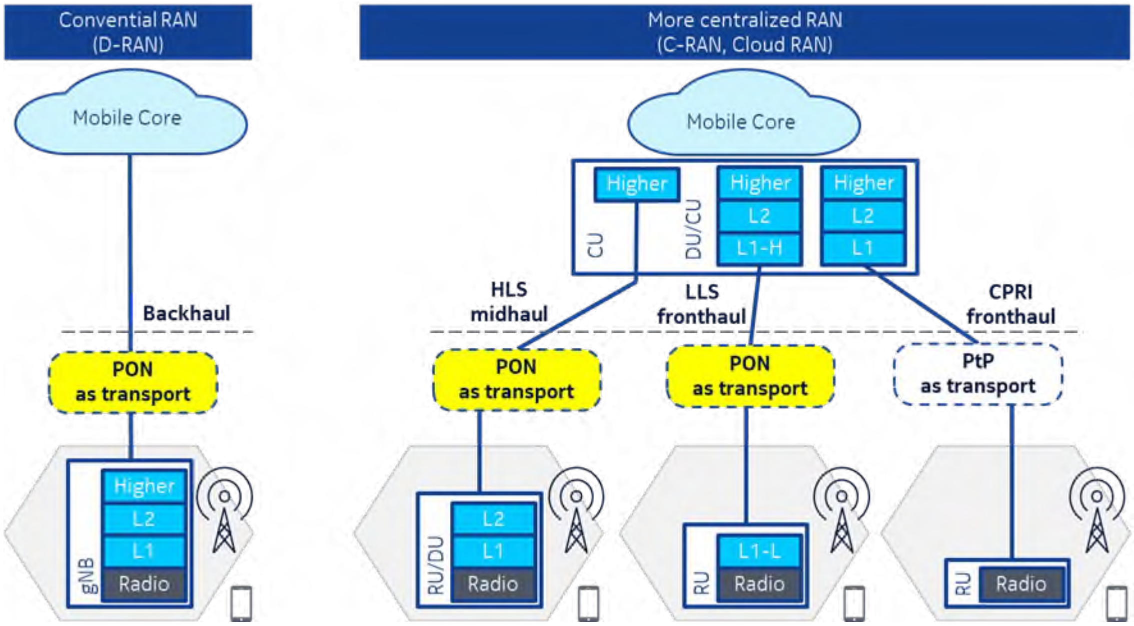


Figure 45 – Functional split of gNB; xHaul architectures

Table 7 – IEEE PON Standards

Variant	IEEE Standard	Issued	DS Data Rate	US Data Rate
EPON	802.3ah	2004	1 Gbps	1 Gbps
10G EPON	802.3av	2009	10 Gbps	10 Gbps
NG-EPON	802.3ca	2020	25/50 Gbps	25/50 Gbps

All three IEEE PON standards use time division multiplexing (TDM) and point-to-multipoint (P2MP) technologies, as is the case with the ITU-T PONs. Optical link budgets for IEEE PONs are also similar, with link distances for 32 ONT split ratios ranging from 20 to 30 km, depending on optical module ratings. The recent NG-EPON standard, also known as 25G/50G EPON, has passed all balloting stages and is now in the final IEEE approval and expected to be released by mid-2020. This latest version consists of either one 25Gb/s PON or 2x25 Gb/s PONs on different wavelengths, and now also supports channel bonding between the two 25 Gb/s PONs for a peak aggregate rate capacity of 50 Gb/s.

3.2.4 Throughput Dimensioning

Both 4G and 5G small cells are being deployed in higher densities. As shown in ITU-T G.sup66, a 1 Gbit/s interface is usually sufficient for 4G/LTE backhaul and midhaul (i.e., the F1 interface), while 10 Gbit/s (aggregated) is usually sufficient transport for 5G NR backhaul and midhaul. Several small cells may use IAB wireless links (using high layer split 'midhaul') to aggregate to the nearest fiber point-of-presence. Therefore, a single PON Optical Network Terminal (ONT) for midhaul may be aggregating the traffic for several 5G cells.

Other 5G fronthaul options are being specified (e.g., O-RAN's low-layer split fronthaul) where the throughput at the interface scales with the cell load and the maximum number of MIMO layers, regardless of the number of radio transmit/receive chains. This makes it possible to combine multiple variable rate flows on the same shared PON capacity using statistical multiplexing.

If an operator is interested in using PON to provide transport for the O-RAN low-layer split fronthaul (LLS-FH), the increased throughput requirements mean that a 10 Gbit/s interface will already be limiting for this LLS-FH. Thus, the interface for cell sites with LLS-FH, however, would need to be 25 Gbit/s. The O-RAN specifications still define several control plane and user plane formats, making a generic throughput assessment for LLS-FH difficult.

Today, the cost of 25 Gbit/s optical interfaces of LLS-FH might seem like a significant barrier to wide-scale deployment in a PON system. However, given the volumes anticipated for 25 Gbit/s in data centers, the costs are likely to drop over time. This could allow for 25 Gbit/s interfaces to be considered for PON in the future.

3.2.5 Latency

The traffic latency & jitter requirements will also change with 5G. There are two types of latency. First, there is the latency at transport level needed to have proper xHaul operations across the mobile functional split interface, regardless of the application being transported. This will be especially true for fronthaul (between RU and DU, latency T_{FX} in Figure 46). Note that the propagation delay over fiber must be included in fronthaul latency (T_{FX}) as well.

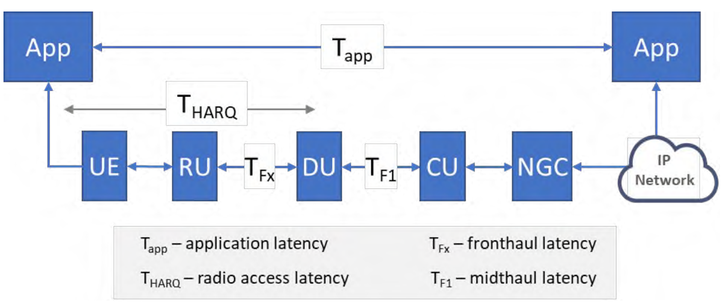


Figure 46 - Latency requirements for Mobile xHaul

Second, the application-level latency requirement (T_{app}) will need to be very low for applications like ultra-reliable low-latency communications (URLLC) services, irrespective of the xHaul split being used. Consequently, the latency requirements for transport of xHaul cannot be captured in a single value. Different standards have different ways to describe various latency classes, depending on the application or the equipment capabilities for given CU/DU or DU/RU functional split. The latency classes can be interpreted as follows:

- At the application-level (T_{app}), the requirements for one-way latency vary between 4ms for eMBB and 0.5ms for URLLC [22].
- Another significant challenge for transporting LLS-FH over PON is the latency of PON, which may be too high for most of some common LLS interface options.

IEEE P1914.1 [23] converts application-level requirements into one-way latency requirements for T_{FX} (i.e., for RAN control and management, transport network control and management, data plane) between 50µs and 100ms. Some implementations of network function application platform interface (nFAPI) option 6 and Telecom Infra Project (TIP) implementations of 7-3 should be possible. If lower latency solutions do become available for PON, this will make it more open to support an LLS more widely for LLS Fronthaul.

High traffic burstiness and increased traffic flows will require different treatment and the PON platform should be aware of the underlying xHaul split as well as the applications that are being carried. PON technology is by nature perfectly suited to bursty service traffic.

In general, the xHaul transport network must use well-known traffic management techniques to differentiate between types of traffic priorities (e.g., voice/video, gaming, signaling, 3G/4G/5G best effort and high priority data). Real-time traffic such as voice and video and xHaul are delay and jitter sensitive and will not benefit from deep buffers. Traffic of this type requires priority scheduling to minimize delay and jitter. The current PON technologies support midhaul and/or backhaul with latency from 1ms up ("Medium latency"). Further improvements for low latency next generation PONs are a research topic.

Additionally, specifically for PON, there are two sources of extra upstream latency in compared to a point-to-point technology. The first is the dynamic bandwidth allocation (DBA), which reacts to monitored traffic and buffer status read-outs. This introduces a delay between a change in traffic and adaptation of the allocated bandwidth, which must be buffered in the ONTs. However, O-RAN specifications are being defined for a cooperative transport interface (CTI) to align mobile scheduling at a RAN with the DBA algorithm. Via the O-RAN CTI, the PON DBA scheduler effectively follows the RAN scheduler. The aim is to restrain upstream latency while allowing for statistical multiplexing over the shared distribution network. This is analogous to the DOCSIS LLX scheduler pipelining technique described in Section 3.1.4.

The second source of latency is the interruptions of upstream traffic that are conventionally used to detect and initialize newly activated ONTs. In the most general case, where the fiber distance of the ONT from the OLT is not known, this delay can be as large as a few 100 µs. This can, however, be reduced by different techniques.

With 5G bringing fast-changing architectures and traffic patterns, OAM tools will be critical to offer and assure new SLAs. This will use a mix of performance monitoring by signaling, sending of telemetry data to external controllers, and dynamic reaction of the network to congestion or latency problems.

3.2.6 Synchronization

The synchronization of end nodes within a 5G network is an essential function. Although GNSS is commonly used for synchronization, 5G transport solutions may also need to deliver precision timing to protect against GNSS failures or in situations where the base station is unable to reliably receive GNSS signals (e.g., indoor locations).

One option for delivering precision timing and synchronization over a PON is to implement distributed GNSS using full timing support (FTS) with hybrid synchronous Ethernet (SyncE) and precision timing protocol (PTP) with IEEE 1588v2. FTS requires that each node in the end-to-end path act as a telecom boundary clock (T-BC), as defined in [24]. In this case, the OLT receives precise time from an external PTP timing source and functions as a concatenation of one T-BC (NT) and a (distributed) dual T-BC (LT – ONT).

All three ITU-T PON standards (i.e., GPON, XGS-PON, and NG-PON2) support T-BC functionality with a time error accuracy of 70 ns. The transport of precise network synchronization over IEEE PONs operates in a similar manner and is described in IEEE 802.1as, clause 13.

3.3 Ethernet

This section provides an overview of Ethernet's capabilities, as well as a discussion on radio encapsulation techniques and time sensitive networking.

3.3.1 Introduction

As operators densify their networks to achieve the promise of greater speeds and capacity with 5G, they need to optimize the total cost of ownership (TCO) of the RAN. As such, many service providers plan to leverage cloud RAN (C-RAN) architectures to centralize baseband processing functions. This provides a more efficient utilization of baseband processing resources and improves operational efficiency at cell sites. C-RAN also enables operators to address different use cases by locating compute and storage resources at different locations within the network as needed for specific use cases. For example, operators can support latency-sensitive applications using cloud edge solutions that are located closer to the serving cell site.

But moving process-intensive functions to an aggregation site requires the stricter transport bandwidth and latency requirements associated with fronthaul networks. The 4G fronthaul networks of today are typically implemented using semi-proprietary protocols such as CPRI over dark fiber. While these techniques meet the necessary transport requirements, they are costly to build and maintain and limit the ability to support multiple services.

With recent innovations in time-sensitive networking (TSN) and radio encapsulation techniques, it is now possible to implement fronthaul networks over Ethernet technology. Ethernet has become the de facto standard for cell site backhaul because of its low cost, multi-vendor interoperability, service flexibility, and ubiquity. Using Ethernet into fronthaul networks, offers the prospect of significantly reducing the cost of C-RAN architectures. Migrating to Ethernet transport also enables operators to leverage the latest innovations in network automation and orchestration, and extensively deploy robust monitoring and network control. In addition, migrating to Ethernet introduces the possibility of converging all 5G xHaul networks (i.e., fronthaul, midhaul and backhaul) and other fixed services onto a single transport network, as shown in Figure 47.

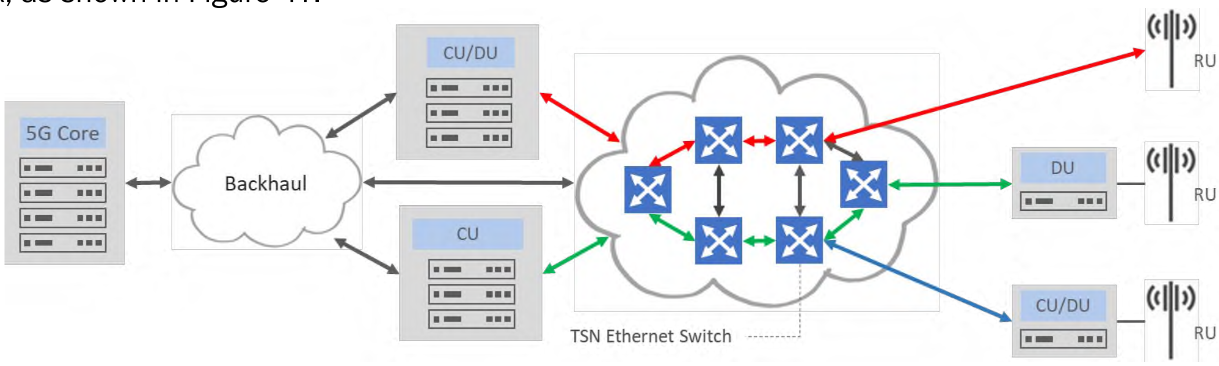


Figure 47 – Converged Ethernet xHaul network

Figure 48 shows the relationship between the various xHaul interface types (e.g., F1, CPRI, eCPRI), radio encapsulation techniques such as Radio over Ethernet (RoE), and TSN Ethernet and other transport platforms.

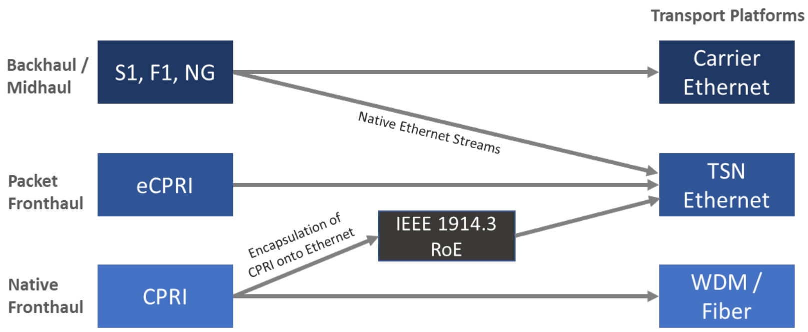


Figure 48 - xHaul transport options

These topics are discussed in further detail in the following subsections.

3.3.2 Radio Encapsulation Techniques

Using Ethernet for fronthaul transport requires the radio signals from/to the RU to be encapsulated into a standard Ethernet frame at the Layer-2 (MAC) level. There are two encapsulation methods for Ethernet fronthaul defined to date: Enhanced Common Public Radio Interface (eCPRI); and Radio-Over-Ethernet (RoE). While the details differ, the underlying principles of both encapsulation methods are similar, and both require low delay and packet delay variation (PDV).

As noted in Section 1.1, the eCPRI specification was developed by the CPRI cooperation group to address some of the limitations of the previous CPRI specification. Unlike CPRI, the bandwidth requirements for eCPRI scale proportionally with user traffic and, as such, eCPRI is roughly 10x more efficient than CPRI. eCPRI is a Layer-3 (and above) protocol that relies on Ethernet MAC and PHY functions. The protocol stack is shown in Figure 49.

The synchronization and control and management (C&M) planes are not covered in the eCPRI specification but instead require existing methods such as precision time protocol (PTP) and/or synchronous Ethernet (SyncE). Although eCPRI is a relatively open framework, key elements are vendor specific in the way they are implemented, which limits multi-vendor interoperability.

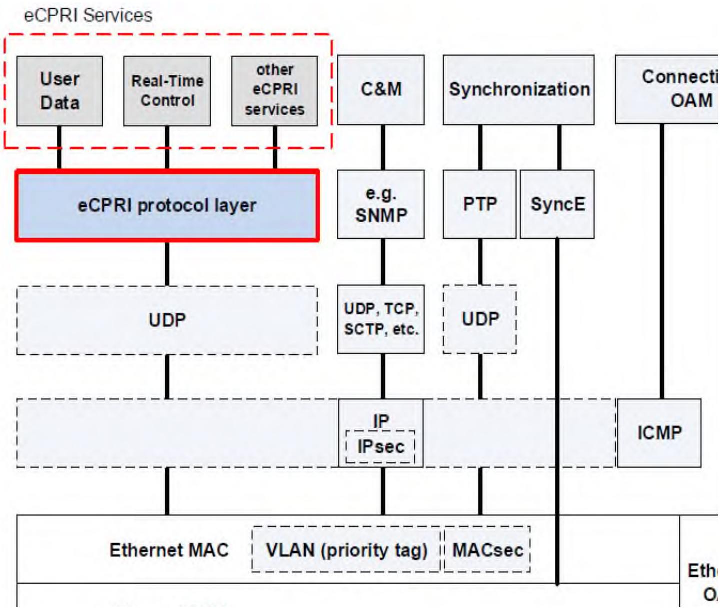


Figure 49 - eCPRI protocol stack over IP / Ethernet [6]

RoE, on the other hand, is defined by the IEEE 1914.3 working group [25]. RoE is an open standards effort to specify a transport protocol and an encapsulation format for transporting time-sensitive wireless radio related application streams between two endpoints over Ethernet-based transport networks. RoE defines a native encapsulation header format for transporting time-sensitive "radio data" and "control data" streams.

In the first release of the 1914.3 standard, three mapping methods are defined: structure agnostic; structure aware; and native mode. Native mode contains two sub-mapping methods for different functional splits. Structure aware and agnostic modes were defined to ease the evolution towards packet-based fronthaul by allowing RoE to be used in existing CPRI-based systems without any modifications to the BU or RU. The native modes, on the other hand, require changes to the hardware but will result in a more efficient fronthaul.

A key requirement for both encapsulation methods is strict latency and packet delay variation (PDV) control. For example, the one-way delay requirement for eCPRI is 100 μs, including the fiber propagation delay.

The next section describes recent advances in TSN that address the requirement for deterministic latency and PDV in Ethernet networks.

3.3.3 Time Sensitive Networking

The IEEE TSN Task Group has published a new standard (IEEE 802.1CM) that addresses TSN for fronthaul networks [26]. The purpose of this standard is to enable the transport of time-sensitive fronthaul streams in Ethernet bridged networks. The standard defines profiles that select features, options, configurations, defaults, protocols and procedures of bridges, stations and LANs that are necessary to build networks that are capable of transporting fronthaul streams, which are time sensitive.

Ethernet networks complying with IEEE 802.1CM will provide deterministic transport of eCPRI and RoE streams by controlling traffic scheduling, timing synchronization, and system reliability. Since Ethernet networks are a shared medium, it is important to prioritize fronthaul packets over other lower-priority packets.

The TSN Task Group has addressed this need with a standard (IEEE 802.1Qbu) that enables express packets to preempt lower priority packets. For example, fronthaul traffic will be able to preempt other "best effort" traffic on the same Ethernet port even after transmission has started. Separate flows are also assigned to fronthaul and synchronization traffic. In addition, because failure detection is important for network resiliency, new redundancy and failure detection capabilities are also being studied.

Two 802.1CM profiles have been defined thus far, both of which are applicable to the CPRI and eCPRI specifications. Profile A is based on strict priority where user data (radio signal samples) are given higher priority over control and management data. In Profile B, 802.1Qbu frame preemption is added to strict priority. In both cases, the maximum frame size for all traffic is 2000 octets.

By using standards-based approaches to enhance Ethernet, operators will get deterministic network performance that meets the stringent throughput and latency requirements of fronthaul networks. Operators will also gain the flexibility, traffic efficiency, and the openness of packet Ethernet networks in a technology that is proven and well understood.

3.4 Wavelength Division Multiplexing (WDM)

Wavelength-division multiplexing (WDM) is a technology which multiplexes a number of optical carrier signals onto a single optical fiber by using different wavelengths of light. This technique enables bidirectional communications over one strand of fiber, as well as multiplication of capacity. This section provides a short introduction to WDM, as well as provides examples of how WDM can be applied to xHaul.

3.4.1 Introduction

Today, fiber-optic networks play a crucial role in mobile backhaul, supporting over 65% of all macro cell and small cell connections in North America [27]. As 5G deployments continue to accelerate, service providers will need to augment their current fiber networks to stay ahead of growing 5G capacity demands. For operators with their own fiber assets, this may require costly and time-consuming new fiber builds as spare fibers along existing routes are exhausted. Similarly, operators leasing dark fiber links will see increased leasing costs and potential time-to-market delays due to construction. As a result, there has been significant interest of late in leveraging wavelength division multiplexing (WDM) technology for 5G xHaul [28], [29] & [30].

WDM technology was first developed for and deployed in long-haul fiber-optic networks to maximize system capacity. Figure 50 shows the basic operation of a point-to-point WDM system. Here, multiple optical signals, each on a separate optical wavelength, are combined at one end of the link and sent to the other end over a single fiber. At the far end of the link, a WDM filter separates the composite optical signal into the individual wavelengths.

Using WDM, up to 100 or more wavelengths can be combined (or multiplexed) onto a single fiber, effectively increasing system capacity by 100 times. Over time, as WDM technology has matured it has migrated from long-haul networks to metro networks, and more recently into access and data center networks.

In addition to minimizing network build costs and time-to-market, WDM technology offers several other advantages. Specifically, industry standards for WDM technologies have been developed by several organizations and, as such, a huge selection of standard components (e.g., filters, multiplexers, pluggable optical transceivers, etc.) are available from a multitude of suppliers. In turn, costs have steadily declined due to large economies of scale. A WDM system is also largely transparent to the optical signals it carries. WDM is protocol agnostic and introduces little to no latency, a feature that is particularly important for certain 5G applications (URLCC) and architectures (e.g., fronthaul). Depending on the wavelength grid and the optical transceivers employed, WDM is also capable of supporting individual wavelengths with data rates up to 100 Gbps or more. Distances of 20 to 30 km are routinely achieved with passive WDM systems and even greater distances are possible with active systems.

Figure 50 – Wavelength division multiplexing

3.4.2 WDM Applications for XHaul

There are several ways in which WDM technology can be leveraged for use in 5G transport networks, depending on the RAN architecture adopted by the operator. For example, in a partially distributed RAN architecture, as illustrated in Figure 51, WDM can be used to increase effective fiber capacity in both the backhaul and midhaul. Although a point-to-point WDM system is shown between the CU and the 5G core network, a 1+1 WDM system or WDM ring could also be used for added network resiliency.

In this case of the fronthaul link, the CU located at the hub site connects to a cluster of DUs (co-located at the antenna site with the RU) via DWDM. Here, one end of the WDM link is located at the hub site, while the other end is typically located in a splice enclosure or equipment cabinet near the antenna sites. This architecture significantly reduces the dark fiber requirements between the antenna sites and the hub site, an area where fiber resources can be scarce, particularly for ultra-dense 5G deployments.

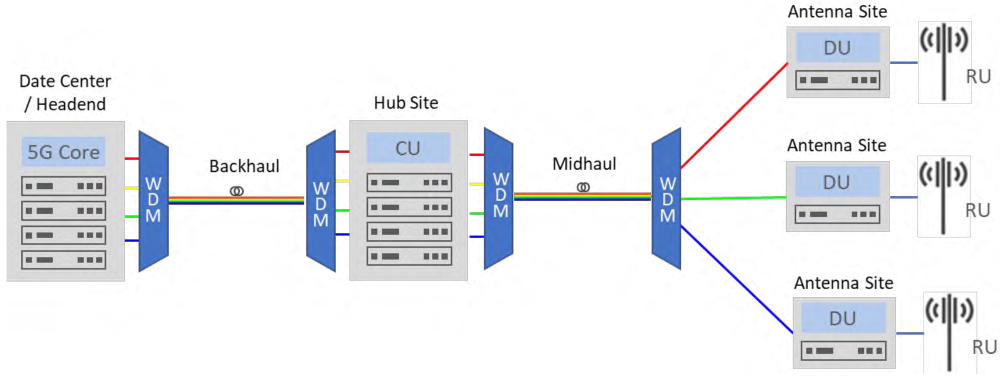


Figure 51 – Wavelength division multiplexing of backhaul and midhaul links

This same WDM architecture can be used to combine fronthaul links from multiple RUs onto a single fiber for transport back to the DU/CU, as shown in Figure 52. WDM is particularly well-suited for fronthaul applications because it introduces virtually no additional latency, an important requirement for CPRI and eCPRI links, as noted earlier in this paper.

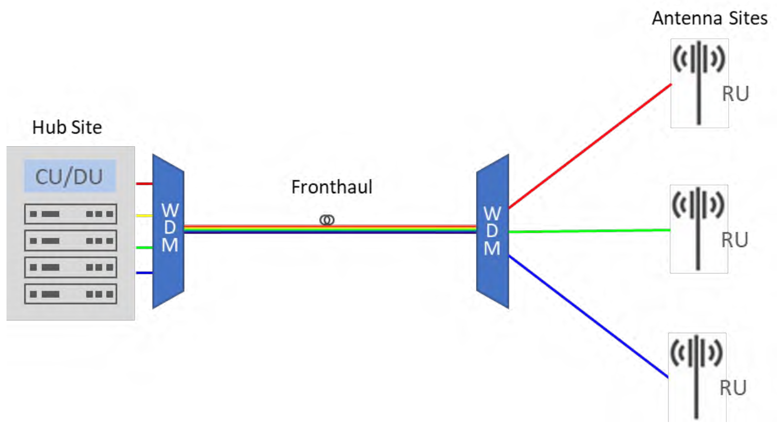


Figure 52 - Wavelength division multiplexing of fronthaul links

WDM can also be combined with time division multiplexing (TDM) based PON systems to augment system capacity. In fact, two of the PON standards noted in Section 3.2 (i.e., NG-EPON and NG-PON2) make use of WDM to increase system capacity by 2-4 times. Figure 53 illustrates the use of WDM PON for fronthaul. In the downstream direction, light from four fixed wavelength OLT lasers are combined by a WDM multiplexer (mux). The light is then sent over fiber to a mux, where the wavelengths are split out and sent to each ONT. At the ONT, the light is filtered with an actively tunable filter that passes the desired downstream wavelength to its receiver and then onto the RU. In the upstream direction, tunable lasers at each ONT are dynamically assigned to a wavelength. The light from all ONTs are combined by the passive mux and sent to the DU via the OLT.

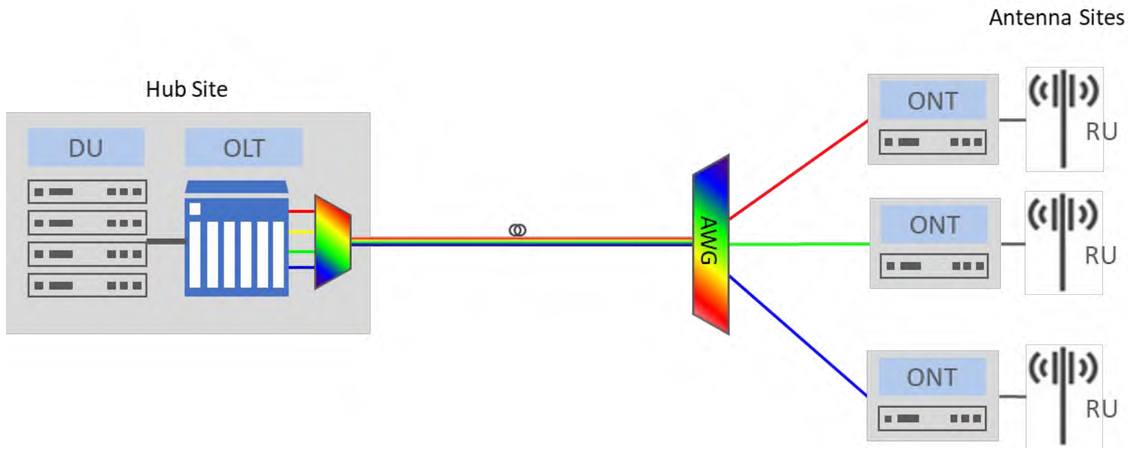


Figure 53 - WDM PON

Similarly, WDM can be used in HFC networks to aggregate traffic from multiple optical nodes onto a single fiber, particularly in areas where additional fiber strands are not available. This is especially relevant for “fiber deep” optical node architectures, as shown in Figure 54. In this example, an original optical fiber node is replaced by four new optical nodes extended deeper into the HFC network to increase capacity. The optical nodes are then assigned unique wavelengths, which are combined by a WDM mux and sent over a single fiber (or fiber pair) to the headend or hub site.

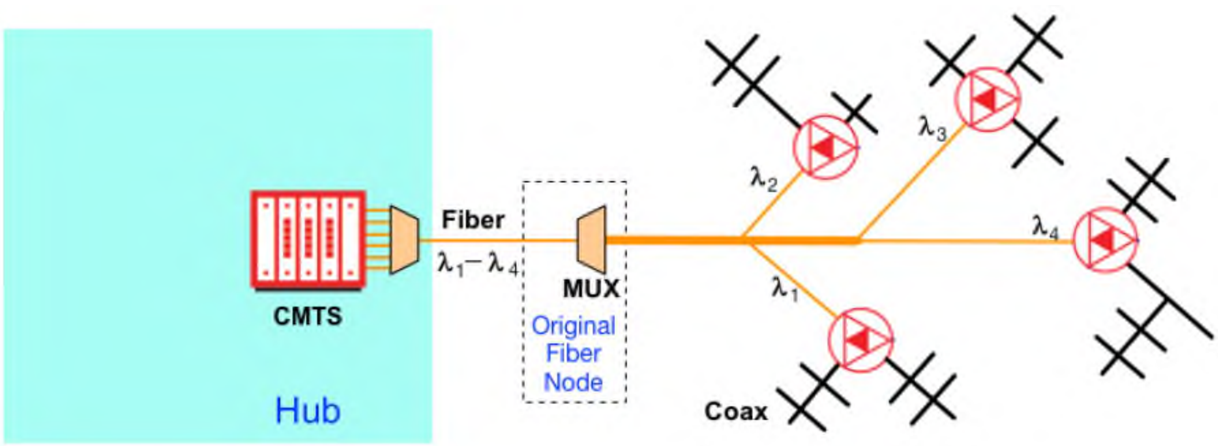


Figure 54 - Example of WDM for fiber deep node architectures [31]

Another application of WDM for mobile xHaul is shown in Figure 55. In this case, fixed passive optical filters are used along the fiber path to add and drop traffic from small cells. Each small is assigned a unique upstream and downstream wavelength. At the central location, a passive mux/demux separates the traffic from each small cell. Note that this architecture can be used for fronthaul, midhaul or backhaul.

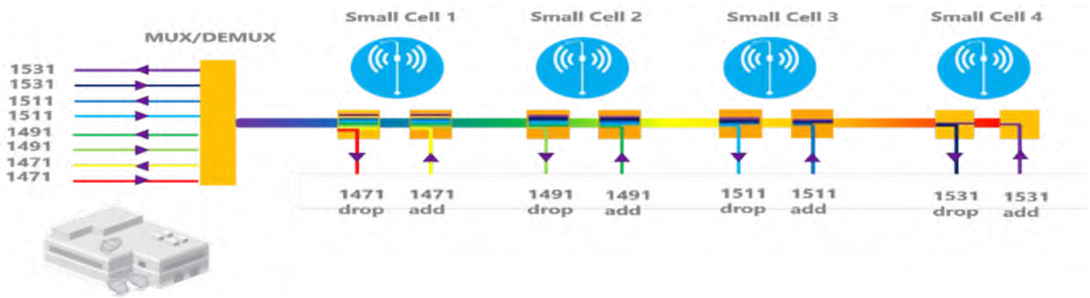


Figure 55 - Add-drop WDM system

As the above examples illustrate, WDM technology offers operators a compelling way to extend the capacity of existing fiber assets to deliver 5G fronthaul, midhaul and backhaul services in a cost-effective and timely manner. Combined with its other advantages such as high capacity, low latency, transparency, etc., WDM technology is considered as an important enabler of future 5G deployments.

Conclusion

Conclusion

Just as mobile wireless technology continues to improve, so to does the technical capabilities and possibilities of backhaul transport. This paper described innovative wireless and wireline backhaul solutions for 5G in detail and the use cases that these technologies could support. Specifically, it covered Integrated Access and Backhaul (IAB) as well as evolving wireline transport technologies (HFC, PON, Ethernet, WDM). Each of these technologies has both business and technical drivers advantages and disadvantages that lead to carriers choosing which to deploy. However, overall the innovation of the technologies continues to be both evolutionary and revolutionary with the goal of attaining the full promise of 5G technology services and applications for customers.

Appendix

Appendix

Acronyms

3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation Mobile Networks
5GC	Fifth Generation Core
AI	Artificial Intelligence
AMF	Access and Mobility Management Function
BF	Beamforming
CA	Carrier Aggregation
CIoT	Cellular IoT
CLI	Cross-Link Interference
cMTC	Critical Machine Type Communications
CN	Core Network
CoMP	Coordinated Multi-Point Transmission and Reception
CP	Control Plane / Cyclic Prefix
CU	Control/ User Plane OR Central Unit
D2D	Device-to-Device
DC	Dual Connectivity
DCI	Downlink Control Indicator
DL	Downlink
DU	Distributed Unit
E2E	End-to-End
EB	Enhanced Beam forming
eMBB	Enhanced Mobile Broadband
EN-DC	E-UTRAN New Radio Dual Connectivity
eNodeB	Evolved NodeB
EPC	Evolved Packet Core also known as System Architecture Evolution (SAE)
EPC/SAE	Evolved Packet Core/System Architecture Evolutions
ePDG	Evolved Packet Data Gateway
EPS	Evolved Packet System
ETSI	European Telecommunications Standards Institute
FCC	Federal Communications Commission
FD	Frequency Division
FD	Full Dimension as in FD-MIMO
FDD	Frequency Division Duplex
FDM	Frequency-Division Multiplexing
FR1	Frequency Range 1 (410 MHz – 7125 MHz)
FR2	Frequency Range 2 (24250 MHz – 52600 MHz)
FWA	Fixed Wireless Access
GERAN	GSM EDGE Radio Access Network

GPRS	General Packet Radio Service
HSS	Home Subscriber Server
IAB	Integrated Access Backhaul
ID	Identity
IEEE	Institute of Electrical and Electronics Engineers
IMS	Internet Protocol Multimedia Subsystem
IMT	International Mobile Telecommunications
IIoT	Industrial Internet of Things
IoT	Internet of Things
IP	Internet Protocol
I-SMF	Intermediate SMF
ITU	International Telecommunications Union
LTE	Long Term Evolution
MEC	Multi-access Edge Computing
MIMO	Multiple-Input Multiple-Output
mMTC	Massive Machine Type Communications
mmWave	Millimeter Wave
MTC	Machine Type Communications
MU-MIMO	Multi-User Multiple-Input Multiple-Output
NEF	Network Exposure Function
NF	Network Functionality
NG	Next Generation
NID	Network ID
NPN	Non-Public Network
NR	New Radio
NRF	Network Repository Function
NR-U	NR Unlicensed
NSA	Non-Standalone
NSSAA	Network Slice-Specific Authentication and Authorization
OCB	Occupied Channel Bandwidth
OFDM	Orthogonal Frequency Division Multiplexing
OTA	Over-The-Air
P-CSCF	Proxy Call Session Control Function
PDCCH	Physical Downlink Control Channel
PDCP	Packet Data Convergence Protocol
PDSCH	Physical Downlink Shared Channel
PLMN	Public Land Mobile Network
PMCH	Physical Multicast Channel
PNI NPN	Public Network Indicated NPN
PRACH	Physical Random-Access Channel
ProSe	Proximity Services

PS	Packet Switched
PSBCH	Physical Sidelink Broadcast Channel
PSCCH	Physical Sidelink Control Channel
PSFCH	Physical Feedback Control Channel
PSSCH	Physical Sidelink Shared Channel
PTRS	Phase-Tracking Reference Signal
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
QoS	Quality-of-Service
RACH	Random Access Channel
RACS	Radio Capabilities Signaling Optimization
RAN	Radio Access Network
RIM	Remote Interference Management
RIT	Radio Interface Technology (IMT-2020 proposal)
RMSI	Remaining Minimum System Information
RNTI	Radio Network Temporary Identity
RRC	Radio Resource Control
RRM	Radio Resource Management
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RSSI	Received Signal Strength Indication
RSTD	Received Signal Time Differenfce
RTOA	Relative Time of Arrival
RTT	Round Trip Time
SA	Stand-Alone
SBA	Service-Based Architecture
Scell	Secondary cell
SCI	Sidelink Control Indicator
SCG	SeNB Cell Group
SCG	Secondary Cell Group
SCP	Service Communication Proxy
SC-PTM	Single-Cell Point-to-Multipoint
SFN	Single Frequency Network
SIM	Subscriber Identity Module
SINR	Signal-to-Interface-and-Noise Ratio
SL	Sidelink
SMF	Session Management Control Function
SNPN	Stand-Alone NPN
SON	Self-Optimizing or Self-Organizing Network
SRIT	Set of Radio Interface Technologie(s)
SRS	Sounding Reference Signal
SRVCC	Single Radio Voice Call Continuity
TA	Time Alignment
TDD	Time-Division Duplex
TDM	Time-Division Multiplexing
TDOA	Time Difference Of Arrival

TSC	Time Sensitive Communication
TTI	Transmit Time Travel
Tx	Transmit
UAV	Unmanned Ariel Vehicles
UC	UniCast
UCI	Uplink Control Indicator
UCMF	UE (radio) Capability Management Function
UDM	Unified Data Management
UE	User Equipment
UL	Uplink
UPF	User-Plane Function
URLLC	Ultra-Reliable Low Latency Communications
V2P	Vehicular-to-Pedestrian
V2V	Vehicular-to-Vehicular
V2X	Vehicle-to-Everything
VN	Virtual Network
WG	(3GPP) Working Group
WI	Work Item
WID	Work Item Description

Appendix

References

[1] 3GPP, “Study on New Radio Access Technology: Radio Access Architectures and Interfaces. 3GPP Technical Specification,” 3GPP, 2018.

[2] 3GPP, “TR 38.801, Technical Specifications Group Radio Access Network: Study on new radio access technology: Radio access architecture and interfaces,” 3GPP, 2017.

[3] 3GPP, “TS 38.401, NG-RAN Architecture description,” 3GPP, 2019.

[4] ITU-T, “Technical Report GSTR-TN5G - Transport network support of IMT-2020/5G,” ITU-T, 2018.

[5] ITU-T, “Technical Report GSTR-TN5G, Transport network support of IMT-2020/5G,” 2018.

[6] CPRI, “eCPRI Specification V2.0, “Common Public Radio Interface: eCPRI Interface Specification”,” CPRI, 2019.

[7] G. Macri, “Deploying-5G-Will-Cost-at-Least-130-Billion-in-Fiber-Study-Says,” 10 July 2017. [Online]. Available: https://www.govtech.com/network/Deploying-5G-Will-Cost-at-Least-130-Billion-in-Fiber-Study-Says.html.

[8] 3GPP, “RP-193251, Enhancements of Integrated Access and Backhaul,” Dec 9 - 12, 2019.

[9] 3GPP, “Study on integrated access backhaul,” 3GPP, 2019.

[10] 3GPP, “R1-1812199, System Performance Evaluation in Multi-Hop IAB network,” Nov 2018.

[11] S.-Y. R. Li, R. W. Yeung and N. Cai, “Linear Network Coding,” IEEE Transactions on Information Theory, vol. 49, no. 2, pp. 371-381, 2003.

[12] O. Orhan, H. Nikopour, J. Nam, N. Naderializadeh and S. Talwar, “A Power Efficient Fully Digital Beamforming Architecture for mmWave Communications,” in 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring), Kuala Lumpur, 2019.

[13] CableLabs, “Data-Over-Cable Service Interface Specifications - DOCSIS 1.0,” CableLabs, 1997.

[14] CableLabs, “Low Latency Mobile Xhaul over DOCSIS Technology,” CableLabs, 2019.

[15] CableLabs, “Low Latency Mobile Xhaul over DOCSIS Technology, CM-SP-LLX-I01-190628,” CableLabs, Boulder, CO, 2019.

[16] J. Andreoli-Fang and J. T. Chapman, “Mobile-aware scheduling for low latency backhaul over DOCSIS,” in Proc. of IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), Montreal, 2017.

[17] J. Andreoli-Fang and J. T. Chapman, “Latency reduction for mobile backhaul over DOCSIS through

pipelining," in Proc. of IEEE Globecom, Singapore, 2017.
[18] J. T. Chapman, J. Andreoli-Fang, M. Chavin, E. C. Reyes, L. Zheng, D. Liu, J. Padden and A. Bernstein, "Low latency techniques for mobile backhaul over DOCSIS," in Proc. of IEEE Wireless Communication and Networking Conference (WCNC), Barcelona, 2018.

[19] CableLabs, "Synchronization Techniques for DOCSIS Technology Specification, Version I01," CableLabs, Boulder, 2020.

[20] CableLabs, "Low Latency Mobile Xhaul over DOCSIS Technology," CableLabs, Louisville, 2019.

[21] Bell Labs, "Quantifying the cost benefits of FTTH for 5G transport," Bell Labs, 16 April 2020. [Online]. Available: https://www.nokia.com/blog/quantifying-cost-benefits-ftth-5g-transport/. [Accessed 20 May 2020].

[22] 3GPP, "Study on scenarios and requirements for next generation access technologies," 3GPP, 2018.

[23] IEEE, "IEEE P1914.1, IEEE Draft Standard for Packet-based Fronthaul Transport Networks," IEEE, 2019.

[24] ITU-T, "G.8273.2 : Timing characteristics of telecom boundary clocks and telecom slave clocks," ITU-T, 2019.

[25] IEEE, "1914.3-2018 IEEE Standard for Radio over Ethernet Encapsulations and Mappings," IEEE, 2018.

[26] IEEE, "IEEE 802.1CM - Time-Sensitive Networking for Fronthaul," IEEE, 2018.

[27] ABIresearch, "Mobile backhaul options -Spectrum analysis and recommendations," GSMA, 2018.

[28] C. W. M. E. Jim Zou, "Optical fronthauling for 5G mobile: A perspective of passive metro WDM technology," in 2017 Optical Fiber Communication Conference and Exhibition (OFC), Los Angeles, CA, USA, 2017.

[29] Commscope, "How WDM Helps With 5G Deployment," Commscope, 17 07 2017. [Online]. Available: https://www.commscope.com/blog/2017/how-wdm-helps-with-5g-deployment/. [Accessed 25 05 2020].

[30] NGOF, "5G-Oriented OTN Technology," NGOF, 2018.

[31] CableLabs, "P2PCO-SP-ARCH-I02-190311 - P2P Coherent Optics Architecture Specification," CableLabs, Louisville, CO, 2019.

[32] J. Clayton, "Crafting a Powerful Executive Summary," Harvard Business School, 8 Sept. 2003. [Online]. Available: https://hbswk.hbs.edu/archive/crafting-a-powerful-executive-summary. [Accessed 13 Feb. 2020].

[33] ITU-R, "Recommendation M.2083-0, IMT Vision - Framewrok and overall objectives of the future development of IMT for 2020 and beyond," ITU-R, 2015.

Acknowledgments

5G Americas' Mission Statement: 5G Americas facilitates and advocates for the advancement and transformation of LTE, 5G and beyond throughout the Americas.

5G Americas would like to recognize the significant project leadership and important contributions of group leaders Ranjeet Bhattacharya of Ericsson and Dave Morley of Shaw Communications, along with many representatives from member companies on 5G Americas' Board of Governors who participated in the development of this white paper.