NGMN Overview on 5G RAN Functional Decomposition

NGMN Overview on 5G RAN Functional Decomposition

by NGMN Alliance

Version:	1.0
Date:	24-February-2018
Document Type:	Final Deliverable (approved)
Confidentiality Class:	P - Public

Project:	RAN functional split and x-haul
Editor / Submitter:	Richard MacKenzie (BT)
Contributors:	Paul Smith (AT&T); Javan Erfanian, Danick Goyette (Bell Mobility); Mike Gilson, Richard MacKenzie, Andy Sutton (BT); Jinri Huang (China Mobile); Mark Gilmour, Remus Tan (Ciena); Mark Grayson (Cisco); Kevin Linehan (Commscope); Wolfgang Stoermer (Deutsche Telekom); Peter Olanders (Ericsson); Nader Zein (NEC); Philippe Sehier, Alistair Urie (Nokia); Giwan Choi, Minsoo Na (SK Telecom); Brett Christian (Sprint); Maite Aparicio (Telefonica); Jovan Golic (TIM); John Kay (US Cellular); Manook Soghomaonian (Vodafone)
Approved by / Date:	NGMN Board, 26th February 2018

Abstract:

The 5G RAN decomposition allows for a wide range of 5G services requirements to be met by allowing for RAN functionality to be split into centralised and distributed locations. There is a trade-off in terms of performance, complexity, flexibility, and transport demands that can allow an operator to decide what RAN decomposition(s) to deploy. This work provides an overview of the various RAN functional split options and their relative trade-offs.

This work also provides an overview of the various industry activities that together form a toolbox enabling an operator to deploy the RAN functional split(s) of their choice.

The functional split options can broadly be classified as either a high layer split (HLS) or a low layer split (LLS). The HLS is progressing as a work item within 3GPP. An overview of the 3GPP activities related to the HLS is provided here. For the LLS, the story is less mature, but still making significant progress. Many industry groups including, 3GPP, eCPRI, xRAN, TIP, TTA, SCF, IEEE 1914, IEEE 802.1CM, are working on parts of the puzzle and specifications are already available or soon to be available from many of these groups. The final part of this document provides an overview of how these various aspects related to the LLS are developing.

Contents

1 INTRODUCTION

1.1 Motivation

The 5G system is being defined in phases to meet a wide range of use cases with diverse requirements, targeting a service-based, flexible, and efficient architecture, while pushing the performance limits. Fundamental aspects of the 5G flexible architecture are the CP/UP (control-plane/user-plane) split, and the RAN internal protocol-layer functional decomposition, the CU/DU (centralized-unit/distributed-unit) split. In here, we focus on the latter, though not independent of the other evolving architectural features, including CP/UP split. Furthermore, a RAN functional decomposition, in turn, implies a particular transport configuration towards the remote radio head.

It is clear that RAN functional split has a number of implications and trade-offs, in terms of meeting service requirements (e.g. latency), complexity, centralized aggregation and capability integration, distributed flexibility, defined multi-vendor interface, and bandwidth, among others. In particular, the choice of functional split will determine the transport capacity requirement and associated latency specifications and performance. This will impact the network architecture as, for example, it can determine the placement of nodes and distance between them.

There is obviously a need to evaluate the impact of the RAN functional decomposition, as applied to different use cases and deployment scenarios, as it impacts transport configuration and network slicing, and as it evolves to contribute to an end to end dynamic and reconfigurable 5G architecture. To align with the 3GPP standardization work, this work considers Option 2 and lower-layer options. However, this analysis allows us to go further and conduct a top-down review of options in terms of benefits, implications (performance, transport, scenarios), and trade-offs, for a service-based dynamic 5G architecture, and provide an overview of the industry activities that are working to define various components to form a 5G RAN.

1.2 Functional Split Options

Figure 1 shows the different possible split options. The top part was outlined initially in 3GPP Release 14 study on radio access architecture and interfaces [1] based on E-UTRA protocol stack.
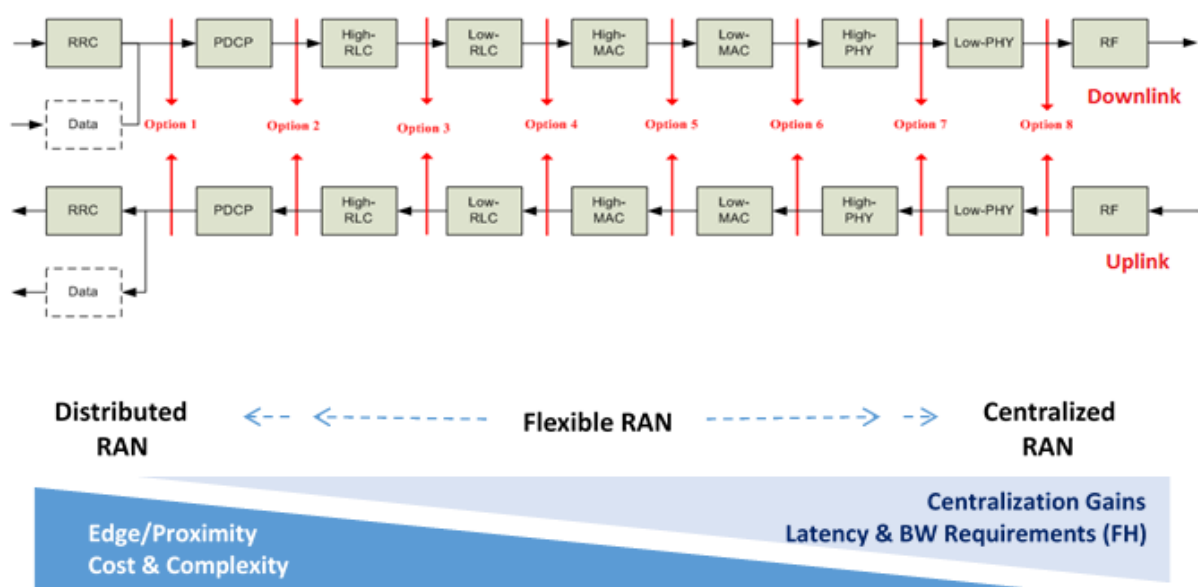
Figure 1: Functional split options

As an example, Option 8 with only the RF functionality in the distributed unit has already been well known and continues to be deployed for densification. Other options have been the subject of further evaluation and definition, in terms of CU/DU split (e.g. Options 2 and 3) and simplified DUs and transport interface options (e.g. Options 6 and 7). A current focus of 3GPP Rel 15 work item is Option 2 standardization, while lower-layer option(s), and CP/UP split for Option 2, are study items. The work on Option 2 benefits from leveraging the work already done for dual connectivity (3GPP Rel 12).

A brief comparison of functional split options is provided in initial study [1] and subsequent work, while the standardization work and further studies are underway. This document further analyses some of the candidate options as well. However, some fundamental observations are obvious. Choice of different options from a fully distributed to a fully centralized RAN architecture can address different requirements, offering the associated benefits (such as edge computing, high throughput at high mobility or centralized processing and simplified distributed coverage), within constraints (such as cost or transport requirements). Ultimately, a flexible radio can address the service or operator requirements considering these trade-offs, potentially starting with a higher layer and a lower layer option.

2   RAN ARCHITECTURES

2.1   Migration Options

A range of different 5G systems are being defined in 3GPP to offer support for multiple 5G migration scenarios. Each system is characterised by the following factors and referred to by an "option" number (see Table 1):

- Core network type: 4G EPC as defined in [2] with enhancements to support EN-DC vs. native 5G core (5GC) as defined in [3]
- NR or LTE/E-UTRA based master node which may be combined with a secondary node when the UE is connected using dual-connectivity (e.g., EN-DC with EPC or MR-DC with 5GC)
- In addition to inter RAT dual-connectivity using EN-DC and MR-DC the LTE and NR RANs may also include support for intra RAT (i.e. NR-NR and LTE-LTE) Carrier Aggregation (CA) and Dual-Connectivity (DC)

Table 1: 5G end-to-end system deployment options

Option	Core network	Master node	Secondary node	Generation
1	EPC	eNB (LTE)	-	4G
2	5GC	gNB (NR)	-	5G
3	EPC	eNB (LTE)	en-gNB (NR)	5G
4	5GC	gNB (NR)	ng-eNB (LTE)	5G
5	5GC	ng-eNB (LTE)	-	5G
7	5GC	ng-eNB (LTE)	gNB (NR)	5G

For the dual-connectivity options (3, 4 and 7) the RAN-core user plane interfaces may be terminated on either the Master node (MN) or the Secondary Node (SN). Over the radio interface (Uu) traffic may be delivered using a Master Cell Group (MCG) radio bearer or a Secondary Cell Group (SCG) radio bearer or Split radio bearer. See [4], [5], and [6] for further details.

When deploying 5G networks, operators will need to consider both their first step solutions to be used when opening 5G service and their subsequent evolution plan to add support for addition 5G use cases. Table 2 and Figure 2 illustrates a range of example migration scenarios [7].

Table 2: Example 5G migration strategy options

Migration scenario	First migration step	Subsequent migration step(s)	Remarks
3 only	• Upgrade LTE RAN to support option 3 (EN-DC features) • Add NR RAN to offer option 3 EN-DC access • Upgrade EPC to support EN-DC	• Expand coverage of option 3 EN-DC access	• Limited by lack of native 5GC services • May include support for sub 1 GHz NR carrier to improve wide area coverage
3 > 5&7	• Upgrade LTE RAN to support option 3 (EN-DC features) • Add NR RAN to offer option 3 EN-DC access • Upgrade EPC to support EN-DC	• Upgrade LTE RAN to support option 5/7 (e.g., introduce 5GC interfaces and MR-DC functionalities) • Upgrade NR RAN to support option 7 (e.g., introduce 5GC interfaces and MR-DC functionalities • Expand coverage of option 7 MR-DC access • Add 5GC core network	• Long term support for early option 3 devices may be required • May include support for sub 1 GHz NR carrier to improve wide area coverage
2 only	• Add NR RAN to offer option 2 NR only access • Add 5GC core network	• Expand coverage of option 2 NR only access	• Initially rely on inter-system handover to legacy 4G when leaving NR coverage • May include support for sub 1 GHz NR carrier to improve wide area coverage • May refarm legacy 2/3G and 4G spectrum to increase NR coverage, capacity and bitrates
3 > 2&4	• Upgrade LTE RAN to support option 3 (EN-DC features) • Add NR RAN to offer option 3 EN-DC access • Upgrade EPC to support EN-DC	• Upgrade LTE RAN to support option 4 (e.g., Xn interface and MR-DC and 5G QoS functionalities • Upgrade NR RAN to support option 2/4 (e.g., 5GC interfaces, MR-DC and 5GS QoS functionalities) • Expand coverage of option 4 MR-DC access • Add 5GC core network	• May include support for sub 1 GHz NR carrier to improve wide area coverage • Long term support for early option 3 devices may be required
2 > 2&4	• Upgrade LTE RAN to support IRAT mobility (option 1<>2) • Add NR RAN to offer option 2 NR only access • Add 5GC core network	• Upgrade LTE RAN to support option 4 (e.g., Xn interface, MR-DC and 5GS QoS functionalities) • Upgrade NR RAN to support option 2/4 (e.g., 5GC interfaces, MR-DC and 5GS QoS functionalities) • Expand coverage of option 4 MR-DC access	• May include support for sub 1 GHz NR carrier to improve wide area coverage
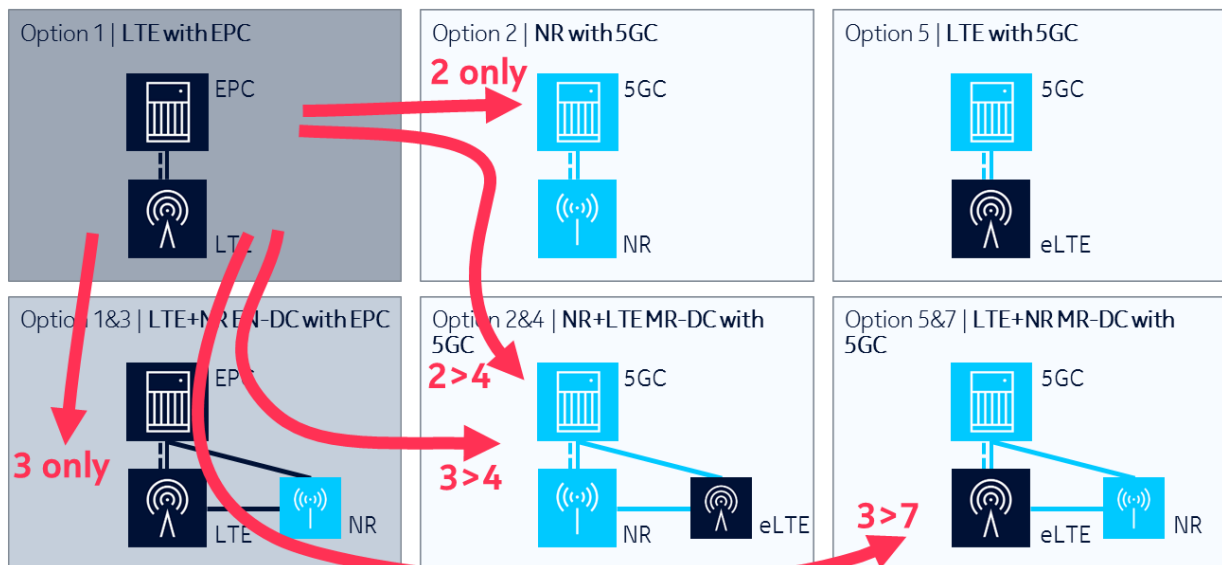
Figure 2: Example 5G migration strategy options

2.2 Functional Placement Options

Interfaces based on the RAN functional split options presented in Figure 1 may be classified into three classes:
- Low Layer Split (LLS) between radio and central RAN functions:
  - Being specified in various forums and standards bodies (i.e. IEEE 1914, CPRI forum and xRAN Forum)
  - Being studied in 3GPP for gNB, not expected to be specified in Rel 15
- High Layer Split (HLS) between distributed and central RAN functions:
  - Being specified in 3GPP as F1 for gNB [8]
  - Being studied in 3GPP as V1 for eNB [9], not expected to be specified in Rel 15
- CU-CP and CU-UP: control/user plane split within central RAN functions
  - Being specified in 3GPP as E1 for gNB [8]
- RAN-core interface:
  - Specified in 3GPP as S1 ([2] and [4]) for EPC core
  - Being specified in 3GPP as N2/N3 [3] or NG [8] for 5GC

The transport to connect functional units has a range of names within the industry, such as fronthaul, midhaul, backhaul, x-haul, etc. The usage of these terms is not consistent among groups and individuals in the industry, so can lead to confusion. For clarify, NGMN have chosen to adopt a more general naming system when discussing the transport for the various interfaces in a disaggregated RAN. When talking about the transport to connect functional units, we simply use the term "transport" and make it clear which interface is being used. For example, "the transport for the F1 interface" or "the transport for the E1 interface".

Furthermore, interfaces between RAN nodes are specified:
- X2: Between central RAN functions in eNB and eNB or en-gNB ([4] and [9])
  - In practice, split into X2-C and X2-U control plane and user plane interfaces
- Xn: Between central RAN functions in ng-eNB and/or gNB ([6])
  - In practice, split into Xn-C and Xn-U control plane and user plane interfaces

Due to overlapping use of identical terms the following terminology is adopted to describe LLS and HLS functional splits:

- RU: Radio unit. Contains all RAN functions placed below LLS interface
  - CPRI Forum uses the terms "RE" and "eRE"
  - xRAN Forum and TIP use the term "RRU"
  - SCF uses the term "PNF"
  - TTA uses the term "gNB-DU low"
- DU: Distributed Unit. Contains all RAN functions placed between LLS and HLS interfaces
  - 3GPP uses the same term "DU" to refer to all functions below HLS interface and hence refers to both RU and DU
  - CPRI Forum uses the terms "REC" and "eREC" to refer to all functions above LLS and hence refers to both DU and CU
  - xRAN forum uses the same term "DU"
  - SCF uses the term "VNF" to refer to all functions above LLS
  - TIP uses the term "vBBU" to refer to all functions above LLS for DU and CU, and "ECU" for DU
  - TTA uses the term "gNB-DU high"
- CU: Centralized unit. Contains all RAN functions above HLS interface and terminates inter-RAN interfaces (X2, Xn)
  - For the gNB, 3GPP splits this function into a single CU-CP control plane and one or more CU-UP user plane functions
  - xRAN Forum splits CU-CP into lower and upper parts and refers to CU-CP-H as xRANc with the "southbound" B1 interface to remainder of RAN
  - TIP uses the term "CAU" for all functions above the HLS for CU

Table 3: Terminology comparison

Function	RU		DU		CU-UP		CU-CP
Interface		LLS		HLS		CP-UP	
3GPP (gNB [8], [10])	gNB-DU				CU-UP		CU-CP
3GPP (eNB [9])	eNB-DU				eNB-CU		
3GPP (LLS [11])	lls-DU		lls-CU				
CPRI Forum (CPRI/eCPRI)	RE/eRE		REC/eREC				
IEEE 1914	RU		DU		CU		
xRAN	RRU		DU		CU		xRANc
SCF	PNF		VNF				
TIP vRAN fronthaul	RRU		vBBU				
TIP SI (Type 1)	RAU				CAU		
TIP SI (Type 2)	RRU		ECU		CAU		
TTA	gNB-DU-low		gNB-DU-high		gNB-CU		

Combining these interfaces and taking into consideration the multi-connectivity options used to combine LTE and NR access [5] the overall architecture for parallel LTE and NR RAN connected to both EPC and 5GC core networks supporting and the terminology defined above, the full range of deployment options described in section 2.1 is shown in Figure 3.
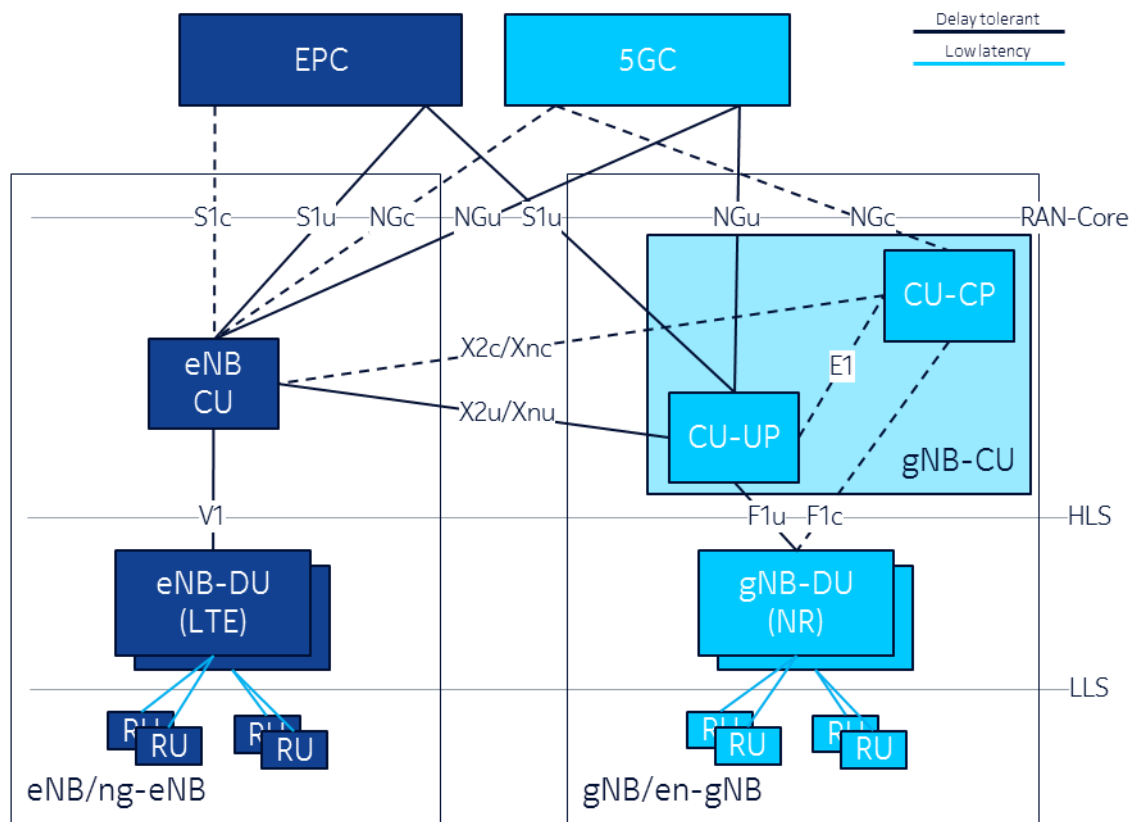
Figure 3: Overall RAN architecture

The individual functional entities RU, DU, CU-UP and CU-CP entities may be place at different physical locations according to operator requirements, physical site constraints and transport network topology, latency and capacity limitations. An additional constraint is the cardinality restrictions driven by the RAN-Core interface design assumption that a given RAN has a single control plane termination point that implies that a given RAN may only have a single CU-CP entity.

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

- Cell site
- Aggregation site (intermediate site, traditionally used for transport aggregation and may have been used to host legacy BBU hoteling)
- Edge site (most central site in RAN)

Table 4 provides additional information on each of the example functional placement options and highlights which interfaces would be used linking physically separate sites.
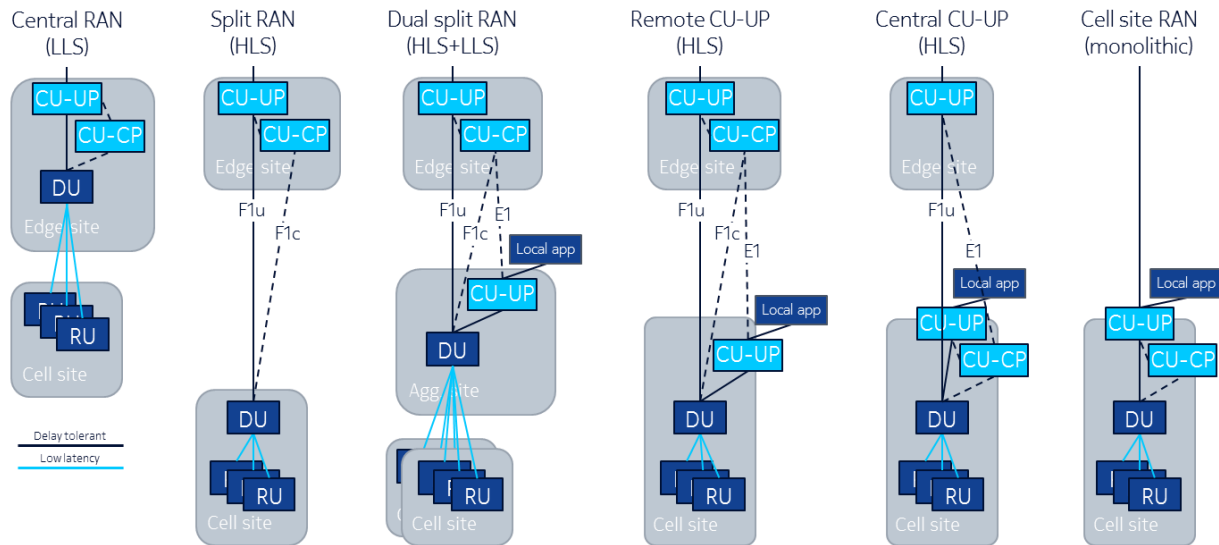
Figure 4: Example functional placement scenarios

Table 4: Summary of example functional placement scenarios

Placement option	Cell site	Aggregation site	Edge site	Inter-site interfaces
Central RAN (LLS)	RU		DU, CU	LLS
Split RAN (HLS)	RU, DU		CU	F1c, F1u
Dual split RAN	RU	DU, [CU-UP]	CU	LLS then F1c, F1u, [E1]
Remote CU-UP	RU, DU, [CU-UP]		CU-CP, CU-UP	F1c, F1u, [E1]
Central CU-UP	RU, DU, CU-CP, [CU-UP]		CU-UP	F1u, E1
Cell site RAN	RU, DU, CU			S1 and/or NG

Note: optional functions and interfaces required to local CU-UP shown as [CU-UP], [E1], etc.

2.3 RAN Latency considerations

In order to satisfy the latency requirements for 5G services, an operator must ensure the physical locations of any user plane functions (UPF) or multi-access edge computing (MEC) functions are close enough to the user. This can put a strict limit on how much functions can be centralised. The benefits of distributed UPFs are realised when content (or service in case of hosted web-apps or compute) is also local. When traffic comes from the Internet then the latency benefit is lost. At the same time there is an incentive to centralise functionality as much as possible to maximise benefits such as pooling gains. The ideal locations for functions to be placed in the network is therefore a compromise that each operator must evaluate and then ensure these physical locations have the necessary transport and compute capability.

Figure 5 shows a simplified example of a 5G RAN topology supporting a range of macro and small cells[1]. Any traffic routed from these sites towards a central entity will pass through a series of aggregation sites. Table 5 provides some indicative values of the relative number of sites. For example, for every 1000 tier 1 sites, there are only 100 tier 2 sites. If a UPF is located in a tier 1 site, then 1000 tier one sites will all have to run their own version of that UPF. If the UPF is moved into a tier 2 site, then only 100 versions of that UPF are required. This level of

---

[1] For the following discussion macro sites are considered, simply because the diagram can be modified more clearly to provide examples. Disaggregation of small cells follows the same principles, although there may be different decisions made for small cells and macro cells.

pooling benefit can save an operator considerable cost in running the UPF. However, the UPF can only be centralised so far as the latency budget is still met, plus other factors such as transport capability. Table 5 also provides some indicative values for transport latency. The one way latency from the cell site to a tier 1, tier 2 and tier 3 aggregation site are 0.6ms, 1.2ms and 4.2ms respectively[2]. Table 5 also provides some indicative values for the 5G latency if a UPF is located at these various aggregation sites. This considers the RAN latency for eMBB and URLLC, plus the round-trip times of the transport network. As a result we can see that it would be difficult to locate a 5G UPF in a tier 3 site as the service latency is probably too high. E.g. 16.4ms for eMBB. However, for a 5G service with approximately 10ms latency requirement an operator will have to look very carefully at tier 1 and tier 2 locations. While tier 1 has slightly better latency performance, tier 2 can provide significant cost benefits.
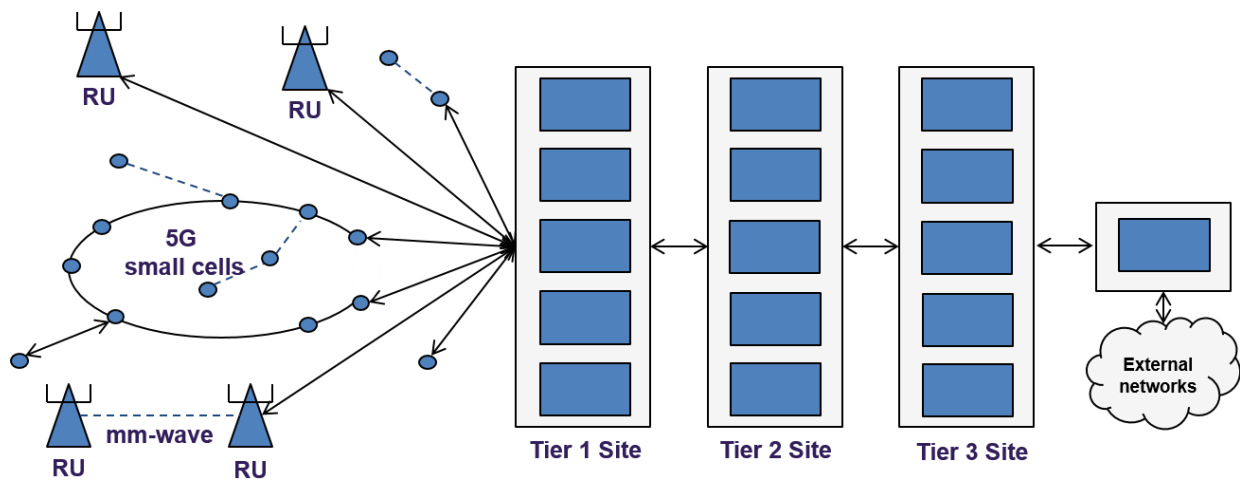
Figure 5: Example 5G RAN with multi-tier aggregation

Table 5: Example maximum latency figures for 5G RAN aggregation sites[3]

UPF Location	Tier 1	Tier 2	Tier 3
Relative Number of sites	1000	100	10
Transport Latency (1-way)[3]	0.6ms	1.2ms	4.2ms
Estimated 5G Latency (RTT)[3]	9.2ms [eMBB]	10.4ms [eMBB]	16.4ms [eMBB]
	2.2ms [URLLC]	3.4ms [URLLC]	9.4ms [URLLC]

For a disaggregated RAN, the location of the DU and CU can also follow similar considerations as the UPFs. The obvious restriction is that the DU and CU would not typically be more centralised than the UPF, otherwise unwanted tromboning effects will occur resulting in further latency increases, as well as possible transport

---

[2] Latency figures based on 95th-percentile of transmission delay (i.e. 95% of cell sites are within this) + overhead for IP

[3] Assumptions:
- Maximum Latency figures based on 95th-percentile of transmission delay (i.e. 95% of cell sites are within this) + overhead for IP
- 5G RTT assumes 8ms overhead for 5G New Radio & Next-Gen Core (eMBB case) – 1ms for URLLC (as per 3GPP release 15)

congestion.[4] Based on the values in Table 5, Figure 6, Figure 7 and Figure 8 show the most likely locations for DU and CU placement to support different 5G latency requirements.

For ultra-low latency 5G services the CU and DU, as well as any UPF or MEC functions, must all be located at or near the cell site, as shown in Figure 6. The reason being that the latency requirement is so stringent that any significant transport latency is not acceptable. In practise this does mean that ultra-low latency capability is unlikely to be provided network-wide. The capability can be added on an ad hoc basis, focussing only on those locations where such extreme latency requirements exist.
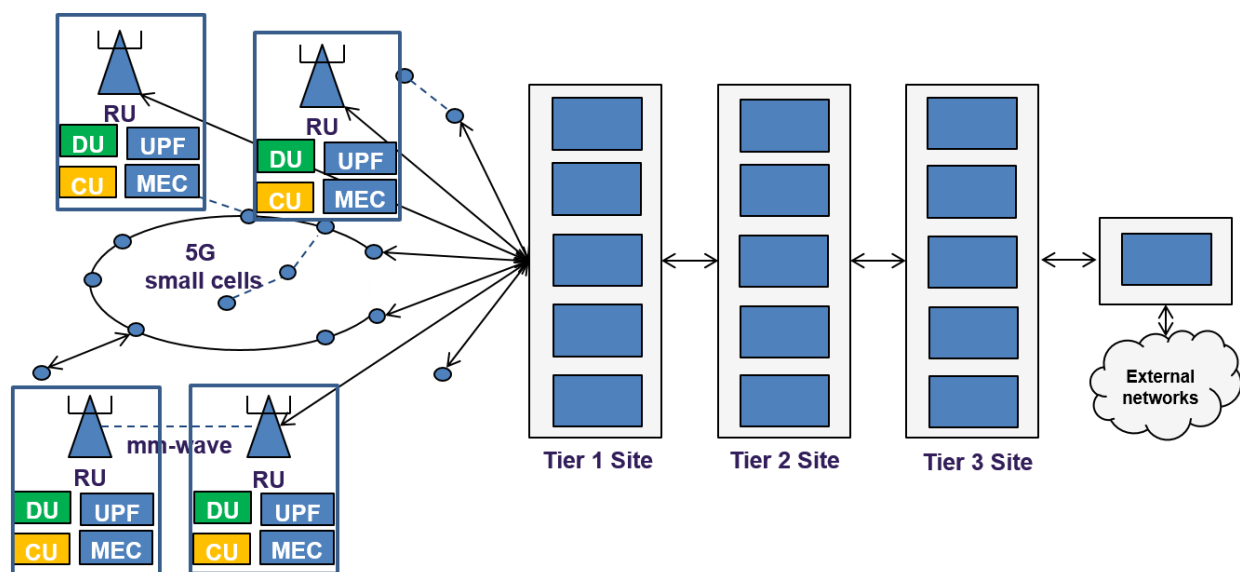
Figure 6: Example CU, DU, UPF, MEC placement for ultra-low latency (approx max. 1ms RTT)

Figure 7 shows the likely placement of DU and CU for low latency 5G services. This covers services like eMBB where the typical user plane round trip time (RTT) of approximately 10ms is targeted. It is likely that the DU would remain at the cell site, however tier 1 is a possibility. However, it seems that the CU can be located at a tier 2 site, the same as the UPF. This would achieve the best compromise between centralisation gains while still being able to provide a service with approximately 10ms RTT.

By comparing the physical locations of functions for ultra-low latency and low latency services it can be seen there is a middle ground, which might be called very-low latency. This is shown in Figure 8. While these DU and CU placements may often fail to satisfy any ultra-low latency services, and would probably fail to be a cost effective way to provide low latency services, there may be a range of 5G services which require a performance latency such as this. This architecture may also be used to offer ultra-low latency services in local areas (stadium, downtown dense urban areas, campus sites, etc.) with low latency (<<1ms) transport. Many MEC services could fall into this category. With the CU and UPF/MEC functions located at a tier 1 site a latency of 5ms RTT should be achievable.

An additional variant may be considered where the CU-UP and UPF functions are placed at the tier 1 site and the CU-CP is placed at a more central tier 2 site. This configuration would be useful for low latency user plane services that do not require rapid call establishment and so may tolerate a high latency control plane.

---

[4] With separation of control plane and user plane, it is possible that some DU and CU functions could be more centralised than a particular UPF, but they are unlikely to be involved in these UPF traffic flows.

What these indicative values and placements highlight is that an operator will have to consider the 5G services that they want to provide while also considering the details of their transport network, when they begin to plan where they can place the DU and CU.
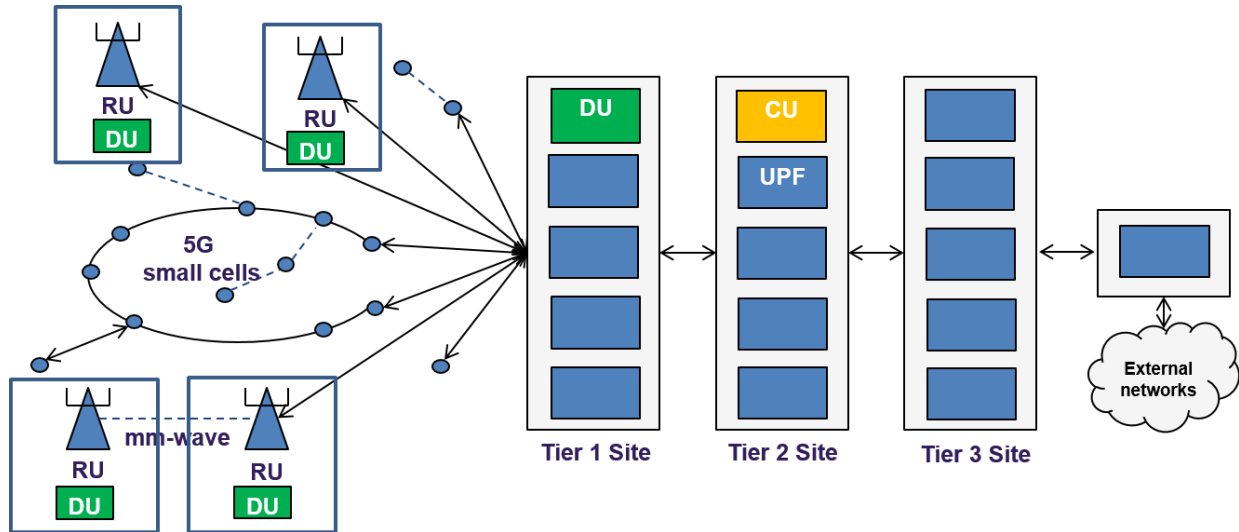

Figure 7: Example CU, DU, UPF placement for low latency (approx max. 10ms RTT)
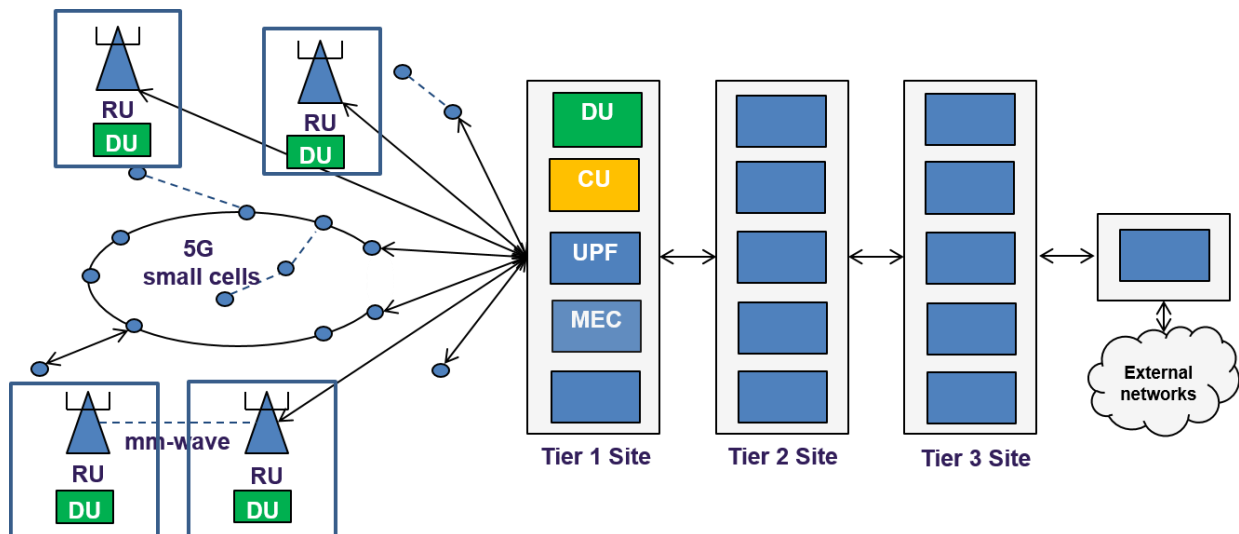

Figure 8: Example CU, DU, UPF, MEC placement for very-low latency (approx max. 5ms RTT)

2.4 Related industry activities

There are a number of industry groups that play a role in defining or implementing RAN functional splits. These are summarised below in Table 6. This section then discusses the related activities in those groups in more detail. As mentioned earlier, NGMN refers to transport for interfaces in the disaggregated RAN by using the term "transport" and specifying which interface is being used (e.g. transport for F1 interface). In this section, however, when discussing each group's activities, we adopt their terminology.

Table 6 not only provides an overview of the related activities of each industry group, it also comments on whether the interfaces being developed can be considered as open interfaces. For an interface to be considered as being "open" it should meet the following criteria:
- Specified to a level to enable multi-vendor interoperability
- Cover the following data planes
  - User plane
  - Control plane (slow and fast)
  - Management plane
  - Synchronization plane

Table 6: Summary of industry activities

Industry organisation	Summary of Activities	Status/Maturity (as of February 2018)	Open/plug-and-play
CPRI	De facto Industry standard for BB-RRU interface since 2004. eCPRI enabling use of switched packet networks and Ethernet/IP. 10 times reduction in bandwidth demand compared to CPRI.	Latest version CPRI V7.0 2015 [12]. eCPRI v1.0 published Aug 2017, V1.1 in Jan 2018 [13].	Openly available, does not cover M plane.
OBSAI	Architecture and interfaces for Cellular Base station.	Mature since 2004. Used since 2004 in high volumes.	Openly available, scope like in CPRI.
ETSI ORI	Specification for interoperable BB-RF interface based on CPRI	Closed, very limited market adoption.	Open
3GPP	High Layer Split (HLS) and Low Layer split (LLS) interfaces considered along with CP-CU split. Low Layer Split (LLS) is still a study item.	HLS interface (F1) between DU and CU approved (38.47x series) E1 interface between CU-UP and CU-CP study report [10] approved and WI (due June 2018) started LLS interface Study Item (SI) report [11] approved, possible continuation of the SI on hold	F1 and E1 are open interfaces specified by 3GPP
xRAN	Evaluated benefit, cost and technical feasibility of different fronthaul split options to develop an open, efficient & interoperable specification for DU – RU communication. Selecting the fronthaul split option for the xRAN fronthaul specification addressing 4-8T and Massive MIMO configurations for LTE & NR xRAN fronthaul whitepaper capturing Stage 1 objectives published [14] Development of detailed fronthaul specification ongoing	First draft of specification is targeted to be made available by end of Q1 2018. The first draft will include user-plane/control-plane interface functions and the architecture definition of the management-plane.	It will be an open standard specification and xRAN is responsible for version control and updates. The specification enables interoperability between DU-RU vendors. The specification aims to achieve a high degree of interoperability, but some aspects of configuration management are likely to be product specific
IEEE 1914	1914.1: Use cases, architecture and	1914.1: near-stable draft. It is	Mechanism defined in

	requirements on next generation fronthaul interface (NGFI) networks; 1914.3: a) Encapsulation & mapping of radio signal into Ethernet packets; b) Also defines objects (as well as their parameters and formats) to be transported over the RoE link for 3GPP functional split options 8, 7.1, and 7.2. To support open interface;	expected to have it publish by mid- 2018. 1914.3: technically mature enough, now at the sponsor ballot stage. The specification is expected to be published in the 1st half year of 2018. Meantime, 1914.3 is discussing to extend the support of open interface to other split options.	1914.3 is targeted at realizing open interface;
TTA	Has developed open interface between gNB-CU and gNB-DU (intra-PHY split)	Specification published on Dec. 2017 ([15], open for revision in the future) – Currently only available in Korean. English translation expected at a later date.	Fx is an open interface for lower layer split
TIP	The Solutions Integration project has developed open interfaces for two architectures for disaggregated LTE base stations, both of which use high layer split [16]. The vRAN fronthaul project is focussed on developing solutions which adopt a low layer split, that will function over non-ideal fronthaul [17]	Solutions Integration project has almost completed its activities. Open specifications are likely to be published soon The vRAN fronthaul project is just starting its lab development phase, based on a high-level specification.	High layer spec will be open. Low layer spec expected to be open
Small Cell Forum	Has developed specification for MAC/PHY (option 6) functional split of LTE [18]	Specification published, although occasionally readdressed, for LTE updates	Open spec published as nFAPI
OAI	Open source software implementation of LTE. Includes split options 8 and 7-1. Integrating SCF nFAPI Developing NR solution with F1 interface		
IEEE 802.1CM / TSN	IEEE P802.1CM is a Time-Sensitive Networking profile standard for Fronthaul (and Midhaul) transport networks. It specifies bridged transport networks over IEEE Std 802.3 Ethernet for the transport of fronthaul traffic, including user data, and control & management (C&M) data. It utilises a range of enhanced functionalities and options from IEEE 802 Ethernet standards including VLAN Bridge, Interspersing express traffic, Frame preemption specification, and Time Synchronisation. The Standard P802.1 supports low functional splits as in CPRI and eCPRI specifications but other functional splits are possible in the future with amendment to the standard. Higher functional split are relatively easier to cover since they are characterised with less stringent requirements.	The Standard specifications is stable and under its first IEEE SA Sponsor Ballot (i.e. no technical change). The Standard P802.1CM is expected to be finalised and published before the end of 2018 and possibly earlier.	This is a Profile Standard which improves interoperability and can be applied to allow time synchronisation for most types of Fronthaul and functional splits over bridged Ethernet networks.

2.4.1 CPRI

The Common Public Radio Interface (CPRI) is an industry cooperation aimed at defining publicly available specifications for the key internal interface of radio base stations. The parties cooperating to define the specification are Ericsson AB, Huawei Technologies Co. Ltd, NEC Corporation and Nokia.

The initial objective of the CPRI forum, created in 2004, was to produce a specification for digitized and serial internal base station interface between the 'Radio Equipment Control' (REC) and the 'Radio Equipment' (RE).

The CPRI Specification version 7.0 was published in October 2015 [12], and adds 24G line-rate to the previously released 10G line rate. The CPRI specification covers split option 8. Single and multiple hops with chain, tree and ring topologies are supported. The specification supports GSM, WiMAX, UMTS, LTE and LTE-Advanced.

Three different information flows (User Plane data, Control and Management Plane data, and Synchronization Plane data) are multiplexed over the interface. The specification covers layers 1 and 2. 3a. The physical layer (layer 1) supports electrical interface as well as optical interfaces.

In August 2017, the CPRI forum released the first version of the eCPRI specification, with an update in January 2018 [13].

Compared to the CPRI, eCPRI makes it possible to decrease the data rate demands between eREC and eRE via a flexible functional decomposition while limiting the complexity of the eRE.

The scope of the eCPRI specification is to enable efficient and flexible radio data transmission via a packet based fronthaul transport network like IP or Ethernet. eCPRI defines a protocol layer which provides various - mainly user plane data specific - services to the upper layers of the protocol stack.
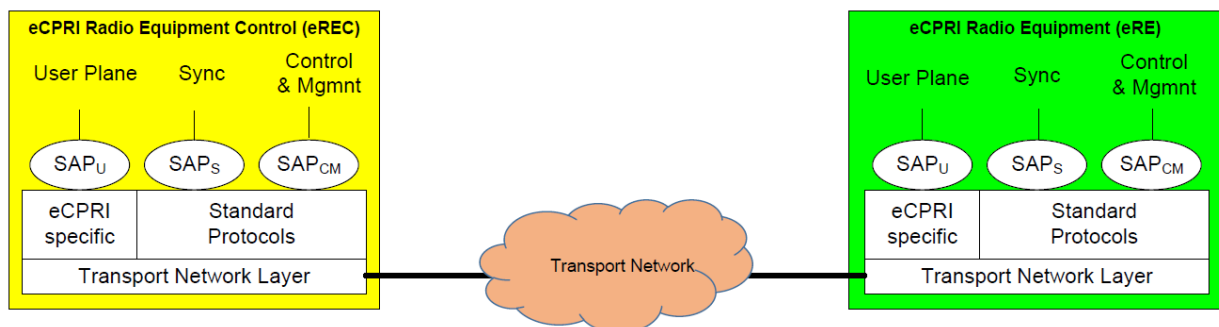
Figure 9: eCPRI system

eCPRI Protocol Overview
The eCPRI interface includes the following information flows:
1. User Plane:
   o User Data:
     User information (to be transmitted from/to the base station to/from the user equipment) with format depending on the underlying functional decomposition between the eREC and the eRE.
   o Real-Time Control data:
     Time-critical control and management information directly related to the User Data.
   o Other eCPRI services:
     eCPRI services such as User Plane support, remote reset, etc.
2. C&M Plane:

  o Control and management information exchanged between the control and management entities within the eREC and the eRE. This information flow is conveyed to the higher protocol layers and is not considered time critical.

 3. Synchronization Plane:

  o Synchronization data used for frame and time alignment.

eCPRI defines a protocol for the transfer of user plane information between eREC and eRE via a packet based fronthaul transport network. For C&M and synchronization information flows, existing protocols and standards are referenced as proposals. The interface supports Ethernet-switched or IP-routed fronthaul networks.

Functional decomposition options

Figure 1 in this document shows the protocol stack layers and possible split points for a 3GPP 4G (LTE) or 5G (NR) radio base station. Figure 10 below shows a more detailed view of the PHY layer. Two inter-layer functional splits D and E (Options 6 and 8 in 3GPP) are depicted in the figure. One additional set of intra-PHY splits named "{I_D, II_D, and I_U}" is also shown. The CPRI specification uses split E. The eCPRI defines two splits for downlink traffic (I_D and II_D) and one for uplink (I_U). Three sub options of the 3GPP split option 7 are also shown for reference.
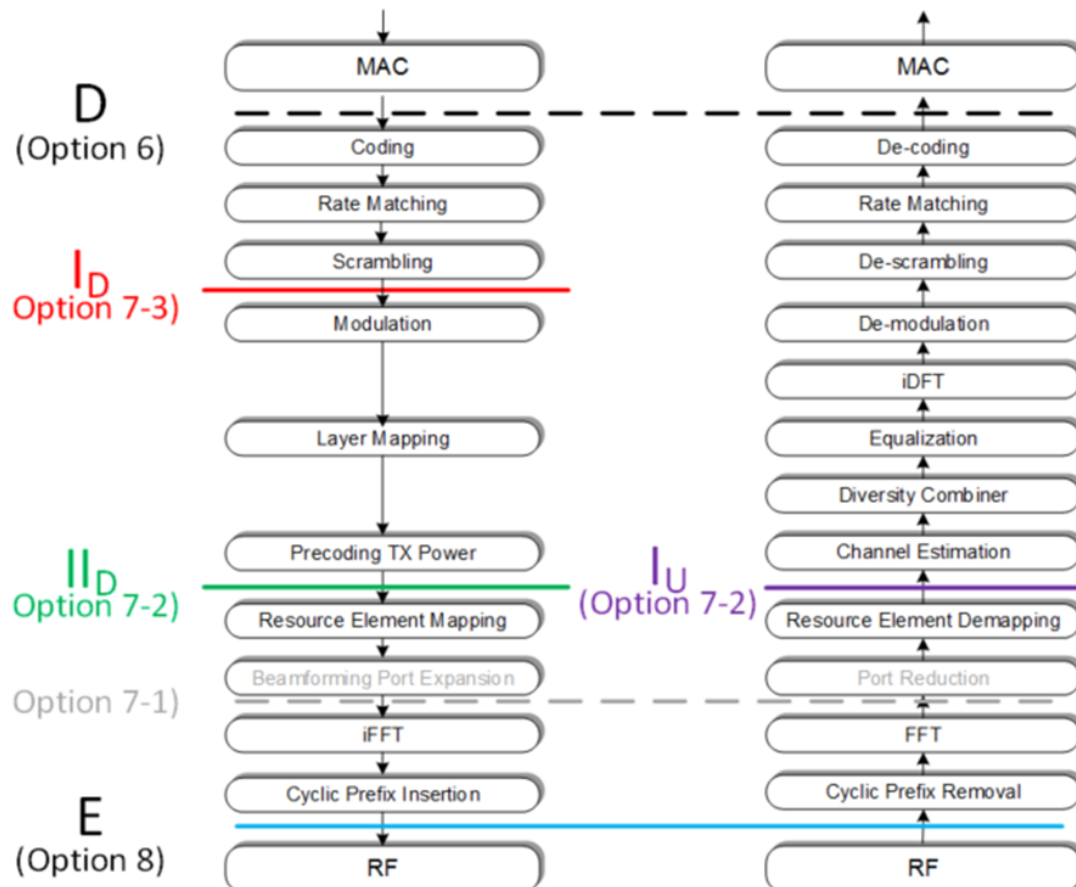

Figure 10: Functional decomposition options inside PHY layer

The advantages of the intra-PHY-split are: features such as Carrier Aggregation, Network MIMO, Downlink CoMP, and Uplink L1 CoMP Joint Processing can be efficiently supported. Some of these features might of course be supported by other splits as well.

Some disadvantages of the intra-PHY-split are: A fronthaul network with "higher" capacity and "lower" latency is required compared to higher layer splits.

As a summary, eCPRI offers the following benefits compared to original CPRI specification:
1. ~10 fold reduction of required bandwidth
2. Required bandwidth can scale flexibly according to the user plane traffic
3. Functional split inside PHY layer enables support of sophisticated coordination algorithms
4. Split in PHY keeps most of the functionality in baseband enabling new feature introduction without changes in the radio equipment
5. Enables utilization of Ethernet and IP, thus guaranteeing future evolution

In addition to the actual Specification CPRI Cooperation has also published the document "Requirements for the eCPRI Transport Network" covering the main requirements of eCPRI traffic to the packet networks [19].

2.4.2 OBSAI

The Open Base Station Architecture Initiative (OBSAI) is an industry group containing base station, module, and component manufacturers. It was established in 2002 with the aim to reduce development costs and encourage greater innovation through open base station architectures and open interfaces. OAM is defined to enable integration of multiple vendors into a base station.

Figure 11 shows the OBSAI reference architecture. Of particular interest, for this whitepaper, is the reference point 3 (RP3) interface which defines the communication between a baseband block and RF block, then variation RP-01 which defines the communication between a baseband block and a remote RF block. This split would be option 8 split in 3GPP terminology. Other key modules in the architecture are the transport block and the control and clock block.
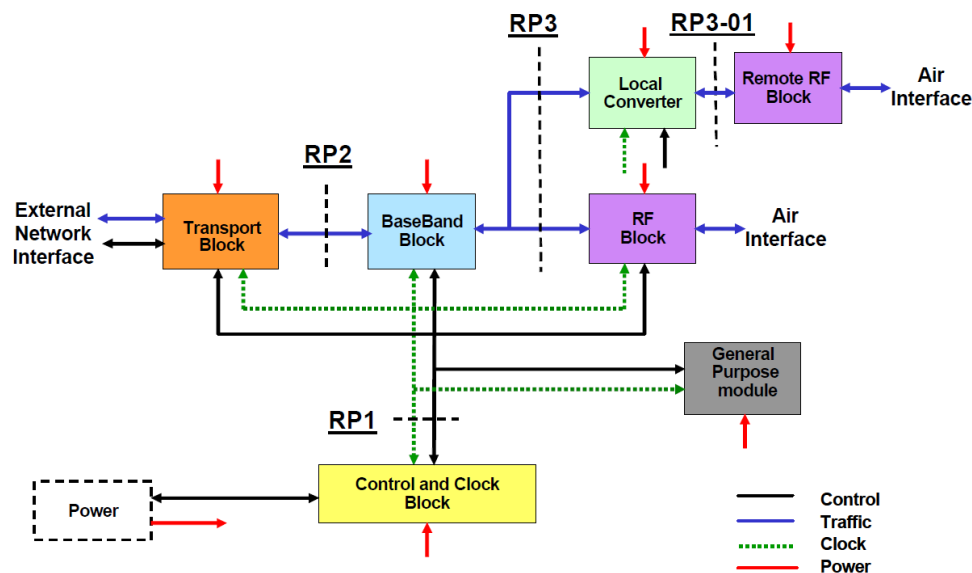

Figure 11: OBSAI architecture [20]

2.4.3 ETSI ORI

The NGMN project on open BBU-RRH interface (P-OBRI) was initiated to develop a recommendation for an open interface between BBU and RRH. Existing BBU-RRH interfaces, based on industry standards like CPRI or OBSAI do not provide a plug and play interoperability. To achieve the fully interoperable goal, options were removed and functions were added to the CPRI specification. OBRI was initially built on version 4.1 of the CPRI specification, and added a C&M layer, not included in the CPRI specification, but required for interoperability.

The preparation phase of this activity started as a task force in September 2008. The NGMN project started in May 2009 and drafted the OBRI Implementation Requirements (IR) which focuses on 3GPP radio access technologies, UTRA-FDD, E-UTRA-FDD and E-UTRA-TDD as well as multiplexing between UTRA-FDD and E-UTRA-FDD.

The OBRI project then submitted its draft recommendation to ETSI ORI as a baseline document for further development. The ETSI ORI specification work was supported by leading mobile network operators and telecommunication equipment vendors. In the continuation of the work conducted in NGMN, the ORI interface was built on top of the interface already defined by the CPRI group. The ETSI group was closed in July 2016. ORI ISG deliverables is maintained by the ETSI Technical Body RRS accessible via the ETSI Portal.

The first release of the ORI interface, published in 2012, provided a digitized radio base station interface enabling single-hop topologies. Different information flows (User Plane data, Control and Management Plane data, and Synchronization Plane data) were multiplexed over the interface. ORI covers OSI protocol Layer 1, Layer 2 up to Layer 7 and is based on CPRI version 4.1. ORI Release 1 provided support for UMTS, LTE and LTE-Advanced technologies.

ORI Release 2 added support for multi-hop topologies and RRU chaining, as well as 9.83 Gbit/s line bit rate, and is based upon CPRI version 4.2.

ORI Release 3 extended support to GSM technology and was based on CPRI version 5.0.

ORI Release 4, published in October 2014, was based on CPRI version 6.0. Release 4 adds IQ data compression for LTE and supports a line bit rate up to 10.14 Gbit/s. This is particularly useful in C-RAN type network topologies.

Figure 12 shows the document structure of the ORI specification available on the ETSI portal. The Low Layers specification (reference [2] in the diagram) covers a single ORI link. The C&M specification (reference [7] in the diagram) covers C&M communication between one REC and one RE. The present requirements specification covers a system configuration of multiple ORI links between one REC and one or multiple REs and optionally additional subsequent ORI links between those REs and other REs.
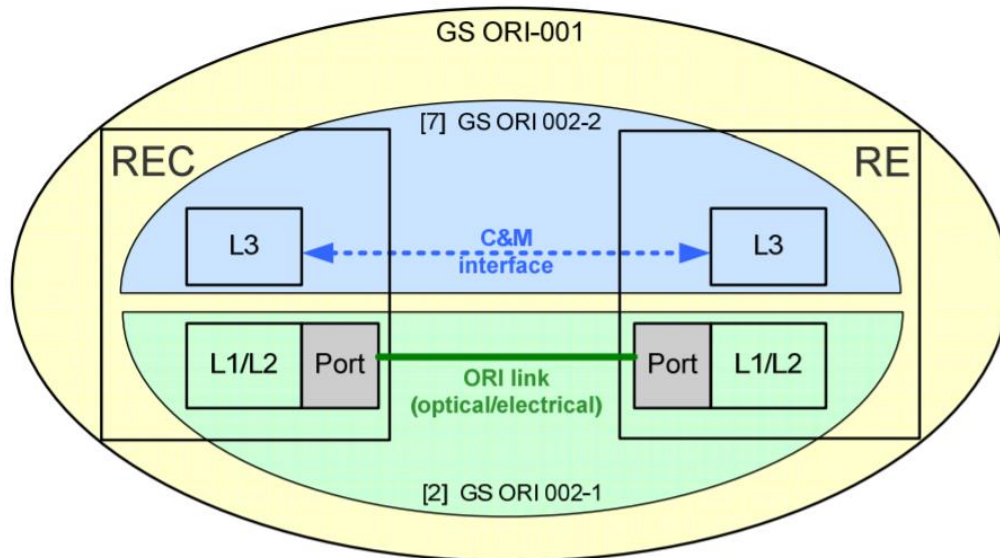
Figure 12: Document structure of the ORI specifications [21]

A number of interface tests have been performed, confirming the interoperability objectives. Despite being a well-defined open standard, the level of adoption in the industry to date remains very limited.

2.4.4   3GPP

3GPP RAN3 working group has selected Option 2 High Layer Functional Split Architecture for the evolved 5G RAN Architecture. In the evolved architecture, the protocol and functional stack of a legacy eNodeB has been split, whereby the RRC and PDCP functions are centralised in a Centralised Unit (CU) and RLC, MAC and PHY functions are left in the remote Distributed Unit (DU). This High layer split configuration allows greater centralisation of RAN functions whilst maintains tight synchronisations and integration of RLC-MAC-PHY functions.

The evolved architecture has gone through further evolution where the Control and User Planes have been split. By splitting the Control and User Planes the Signalling and Data functional block within the CU have also been split up: this functional split allows for greater flexibility of the RAN network design. Furthermore the split functional blocks allows for independent scaling of the Signalling and Data Delivery equipment with increasing traffic load in the RAN network.

The split Control and User Planes allows for three distinct deployment scenarios:
1.   Scenario 1: Centralised CU-CP and CU-UP (Figure 13)
2.   Scenario 2: Distributed CU-CP and Centralised CU-UP (Figure 14)
3.   Scenario 3: Centralised CU-CP and Distributed CU-UP (Figure 15)
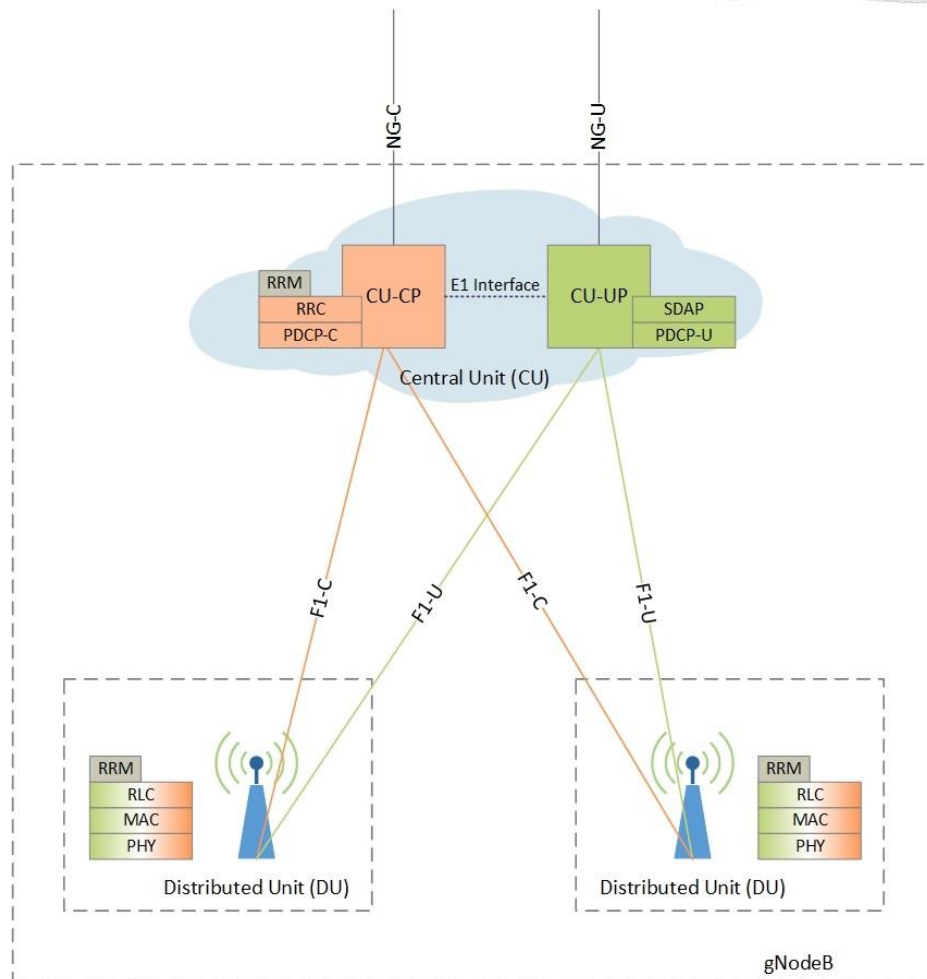
Figure 13: Scenario 1, Centralised Control Signalling and User Plane Functions
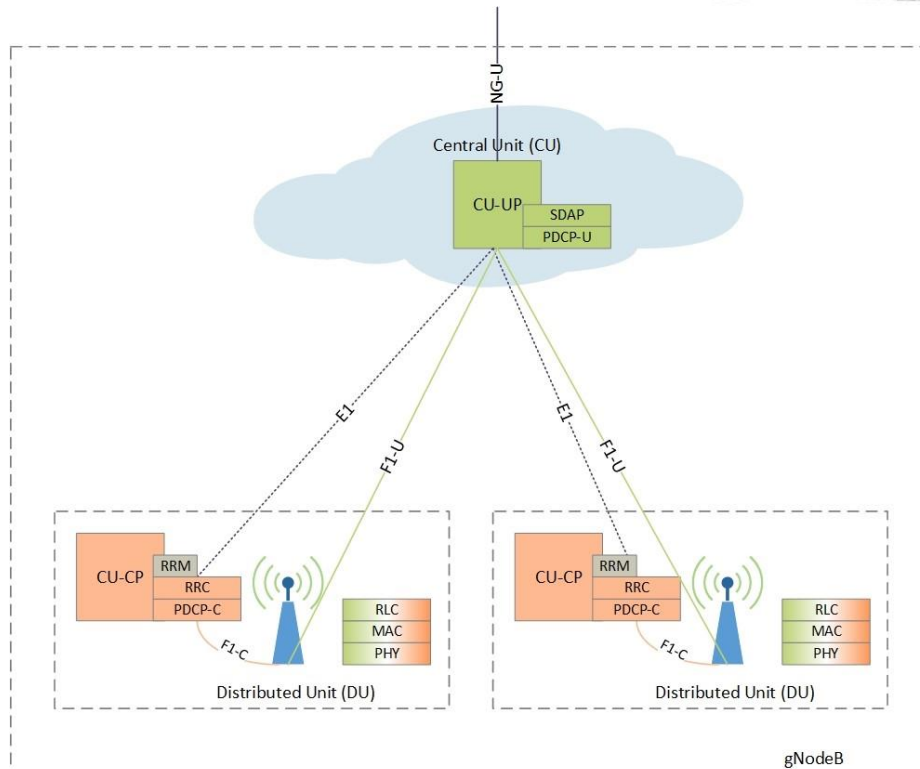
Figure 14: Scenario 2, Centralised User Plane function, CU-CP is in the DU
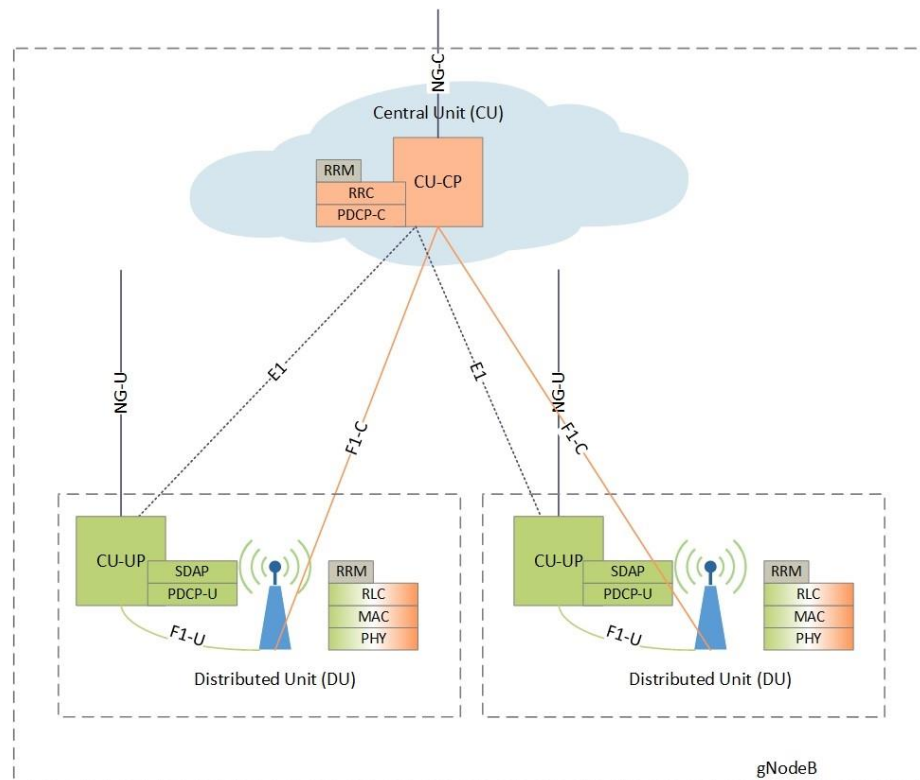
Figure 15: Scenario 3, Centralised Control Plane function, CU-UP is in the DU

Where CU-CP refers to the Control signalling functional block within the CU and CU-UP refers to the User Plane functional Block within the CU.

Each deployment scenario provides a number of advantages and allows the User Plane function or the Control Plane functions to be co-located with the DU.

Due to the significant progress in 3GPP for the HLS, chapter 3 provides a much more detailed overview of these activities.

2.4.5   xRAN

xRAN is a non-profit industry consortium that was formed by carriers to develop, standardize and promote an open alternative to the traditionally closed, hardware-based RAN architecture. The xRAN architecture is complementary to 3GPP and other standards bodies and hopes to provide a blueprint that these standards bodies can use for architecting the future RAN as the industry moves towards 5G.

Standardization and specification of the BBU-RRU interface is currently under investigation in many organizations such as 3GPP, IEEE 1914 & eCPRI. However, xRAN intends to meet the need of operators in defining a specification that can address future deployment scenarios and that can be said to be truly vendor-agnostic and interoperable. The xRAN Fronthaul working group is focused on producing an "Open" specification for next generation fronthaul between the remote radio unit (RRU) and the virtualized baseband unit (BBU), to address future use cases and deployment scenarios [14]. This work will also include the specification of the operation and management (OAM) interface to simplify interoperability between different BBU and RRU suppliers.

xRAN has identified the following operator requirements for the open interface are:
- A BBU – RRU interface that helps achieve interoperability, more specifically in the control and management plane functions
- Control and management functions of the interface should adopt a data model approach to simplify integration of diverse radio implementations as well as multiple intra-PHY split options
- The design should minimize optionality in traffic and management plane to meet key requirements and simplify development and interoperability
- The interface should cater to multiple deployment use cases as described below:
    a) Support use of advanced receivers for high performance 4G & 5G eNB and gNB deployments
    b) Support Cloud RAN deployments and associated features (CoMP, CA etc.)
    c) Scale efficiently to address large antenna configurations (Up to 64T-64R)
    d) Support specialized functions in radios to cater to listen-before-talk operations (e.g LAA)
    e) Support live migration of BBUs
- The interface should reduce bandwidth utilization relative to CPRI and scale as a function of user throughput
- The interface should utilize standardized Ethernet based technologies for transport & enable   switching using standardized switching infrastructure
- The interface should address RRU flow aggregation and improve resiliency by supporting switching traffic between BBUs
- Standards based BBU – RRU synchronization solutions should be addressed
- To meet service needs of the diversified user devices (smartphones, IoT devices, autonomous car, etc.), the fronthaul split should support Network Slicing requirements

2.4.6 IEEE 1914

The IEEE 1914 next generation fronthaul interface (NGFI) working group (WG) was officially founded in February 2016. The WG is positioned as a platform to study NGFI-related key technologies, develop relevant standards development, and ultimately accelerate a mature NGFI ecosystem. Currently there are two projects under the 1914 WG. The 1914.1 project is to describe the NGFI use case and develop the NGFI transport architecture as well as the requirements. The 1914.3 project is specifying the encapsulation format of radio signal into Ethernet packet.

The 1914.1 project has finished the study on use cases, achieved consensus on the transport architecture and begun the standards development on requirements. The requirements will be made from transport network/nodes perspective and will cover the area of performance, O&M, slicing, reliability, flexibility and so on. In addition, the NGFI node reference model and a class of service are being discussed/defined. The 1914.1 plans to have its draft specification ready by June 2018.

Another project, IEEE 1914.3 is an international standard that defines how radio information (data and control) is mapped into Ethernet frames with standardized Radio-over-Ethernet (RoE) headers. It supports the encapsulation of time-domain IQ (e.g. traditional CPRI) and frequency domain IQ including functional split options 7-1 and 7-2 defined in 3GPP.

One of the key features of 1914.3 specification is that it is open and allows all compliant equipment to interoperate. For example, for split option 7-1, the WG has defined extensive control and management messages which the WG thinks could support open (option 7-1) split to large extent.

In 1914.3 several mapping methods are defined:
- Structure agnostic
  - An entire stream of radio information, of any format, is mapped into Ethernet at the RoE mapper and restored to its original form at the RoE demapper. This mapping method has a tunneling mode, where the entire stream is mapped, and a line-coding aware mode, where 8B/10B or 64B/66B line-encoding and scrambling is removed before the information is mapped and is reinserted after the information is demapped.
  - CPRI streams can be carried over Ethernet using the structure agnostic mapping method.
- Structure aware
  - The structure of the protocol that carries radio information is known to the structure aware mapper. I/Q data and control messages extracted from the protocol are mapped into unique Ethernet frames.
  - Specific methods have been defined for the structure aware mapping of CPRI.
- Native RoE
  - Time-domain or frequency-domain I/Q information (the corresponding function splits include option 8, 7-1 and option 7-2 defined in 3GPP) is mapped into Ethernet by the Native RoE mappers. Control information is transported via TLVs (type length value encoding) in control messages.

This standard also defines mechanisms that enable the demapped data to be presented to the radio equipment at the desired time.

Figure 16 may give readers a basic idea of RoE packet format.
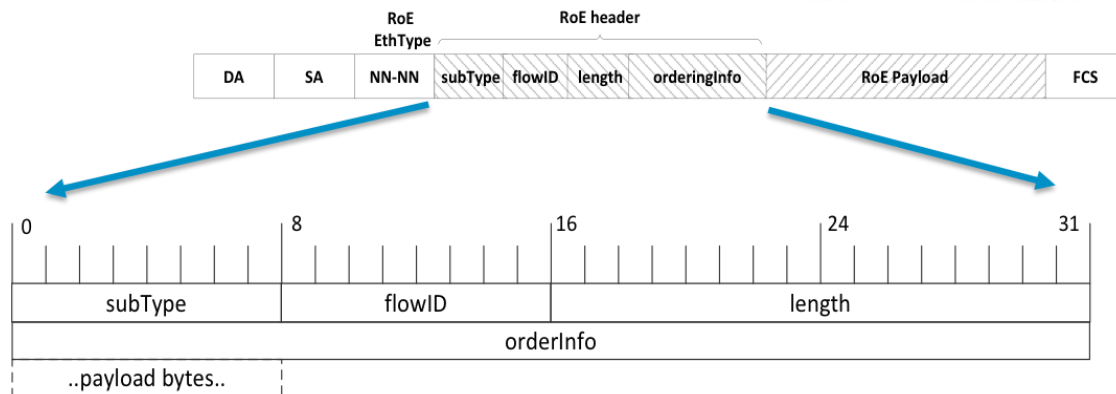
Figure 16: RoE packet format (http://sites.ieee.org/sagroups-1914/)

As a general encapsulation & mapping method, the RoE specification could apply to any functional split option for both the control plane and the data plane. Currently, to support an open interface, the RoE specification defines specific objects (as well as their parameters and formats) to be transported over the RoE link for 3GPP functional split options 8, 7.1, and 7.2. Extension to other functional split options and other control plane objects are expected in the next-stage work. With the RoE specification, it is expected that an open interface can be realized to enable multi-vendor interoperability.

2.4.7   TTA

The Telecommunication Technology Association of Korea (TTA), establish, revise, and distribute Korean ICT standards in communication networks, ICT convergence, security, software, broadcasting, radio/mobile communications. They also test and certify ICT products and services in accordance with the domestic and international standards, and develop testing standards. They also provide support to enhance the competitiveness of small and medium-sized enterprises.

Relevant to this discussion is working group (SPG31), whose role covers 5G radio standardisation. For 5G they transpose 3GPP technical specifications into TTA standards and also develop and maintain 5G RAT standards. TTA also have international collaborations with related industry groups such as 3GPP and NGMN.

Figure 17 shows TTA's 5G RAN architecture with a high layer split (option 2 in 3GPP terminology) and a lower layer split (option 7 in 3GPP terminology). In December 2017 the 5G open interface was published [15]. This provides the specification for the Fx interface between the gNB-CU and gNB-DU. This specification can apply to all variants of the option 7 split. The specification covers user plane, control & management plane, and synchronisation.

This specification has tried to stay in alignment with other industry activities. The architecture itself is in alignment with that discussed within 3GPP. The packet format is based on IEEE 1914.3 RoE, and includes mapping for the CPRI specification.

NG-RAN
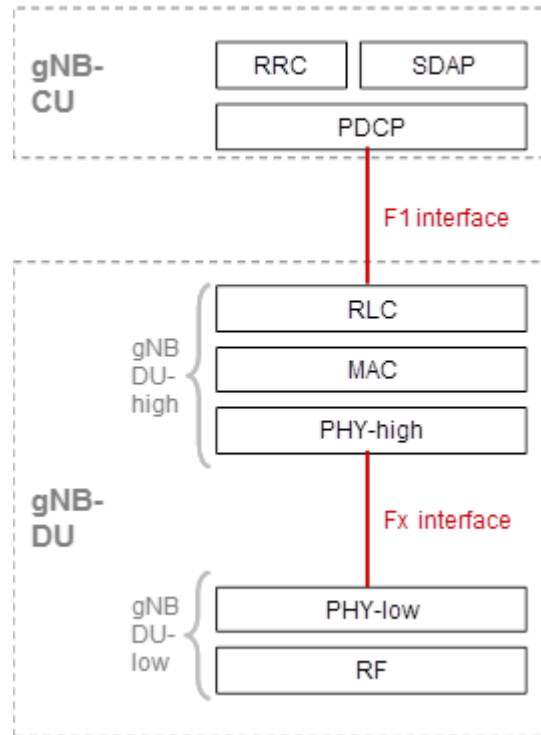
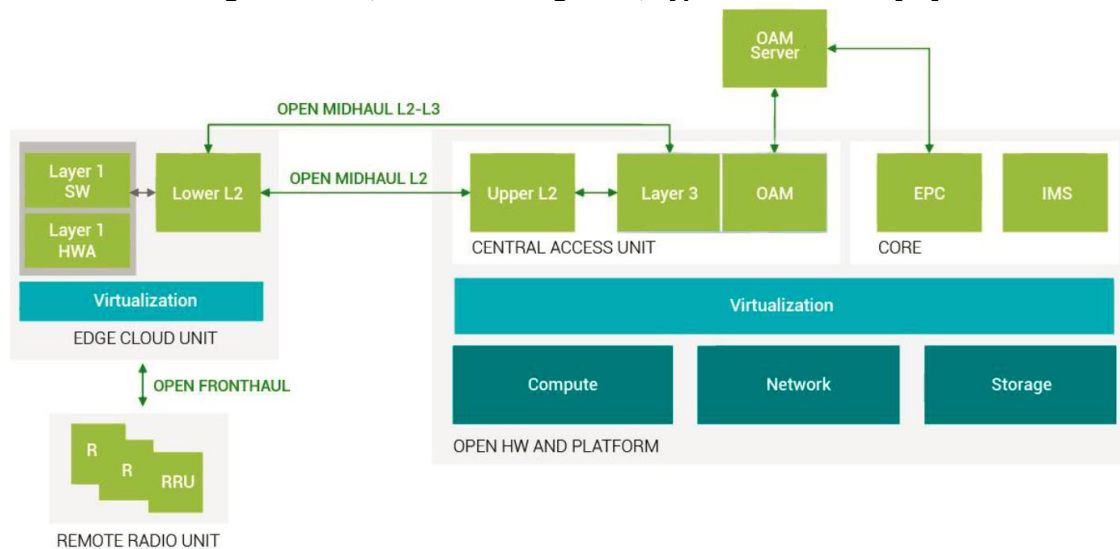Figure 17: TTA architecture

2.4.8 TIP

The Telcom Infra Project (TIP) was founded in February 2016 and is an engineering-focused initiative driven by operators, suppliers, developers, integrators, and start-ups to disaggregate the traditional network deployment approach [17].

2.4.8.1 Solutions Integration

This project aims to develop an "Unbundled RAN" architecture [16]. This means splitting the base station into functional blocks and defining open specifications for those functional blocks to communicate, thus enabling multi-vendor interoperability. The activity has focussed initially on LTE solutions and two RAN architectures, as shown below. The common theme for both of these architectures are the OAM components and the PDCP-RLC split (option 2 split, in 3GPP terminology) in the LTE stack. The main variation among the two architectures is that type 1 architecture includes a single unit, the remote access unit, to contain the RLC/MAC/PHY/RF, while the type 2 architecture has two units: the edge cloud unit which contains RLC/MAC/PHY, and the remote radio unit containing the RF (this is an option 8 split in 3GPP terminology). The connection between the edge cloud unit and remote radio unit will be based on CPRI or ETSI ORI. The open APIs are due to be published shortly.
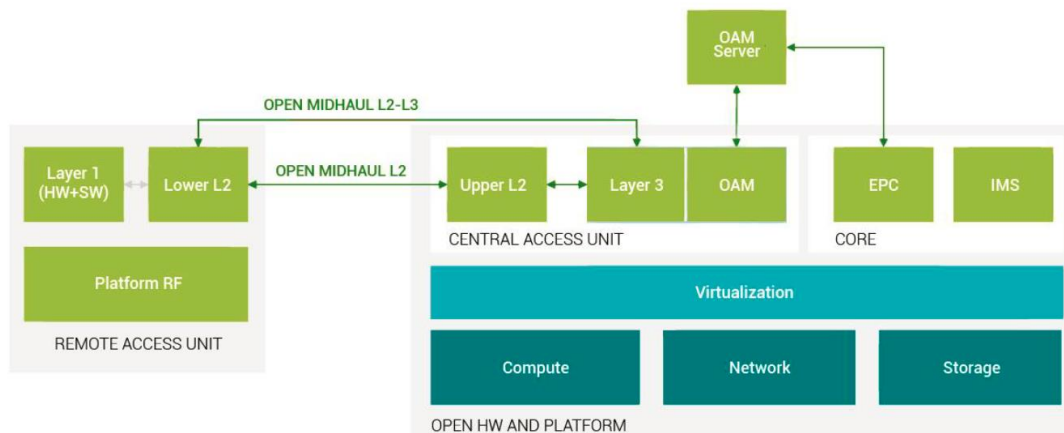
Figure 18: TIP, Solutions Integration, Type 1 architecture [16]

Figure 19: TIP, Solutions Integration, Type 2 architecture [16]

2.4.8.2 vRAN with non-ideal fronthaul

The goal of the project is to create an ecosystem for virtualised RANs with a clear path towards commercial deployments. As, made clear in the project name, there is a focus on vRAN solutions that will work over fronthaul that is "non-ideal". This includes transport technologies such as DOCSIS, G.Fast, microwave and shared-Ethernet, meaning, in comparison to ideal transport, there is a wide range of challenging performance characteristics.

The project is focused on building and developing working solutions in TIP community labs that can be commercially deployed. Each lab activity is focussed on an operator-supported use case and is developing a solution that is truly multi-vendor. This means that the RRU and vBBU must be developed by different vendors.
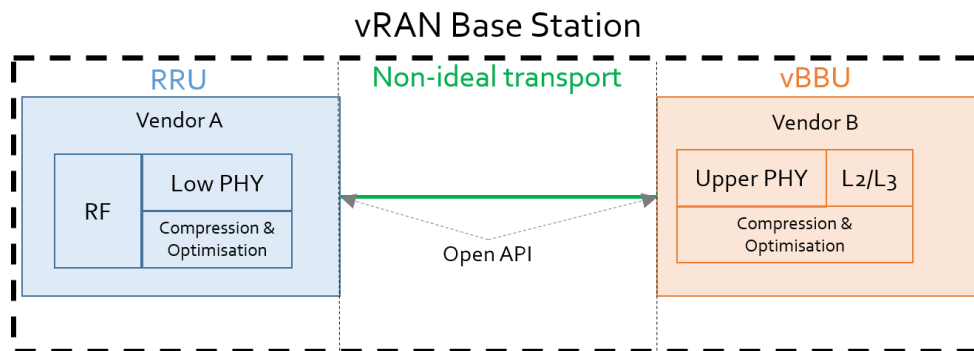
vRAN Base Station

Figure 20: TIP, vRAN architecture [17]

A high level reference design has been agreed within the group, to allow activities to be initiated. This high level design considers a functional split in the physical layer that promises:

- Data compression gains over the fronthaul relative to CPRI
- A fronthaul load that scales with user traffic and is resilient to latency and capacity fluctuations
- The use of open interfaces to allow for multi-vendor interoperability

The benefits of this high level reference design have already been demonstrated in early proof-of-concepts with single-vendor implementations. Over the following months the project will develop more commercial-grade and truly multi-vendor solutions, demonstrating the feasibility of the vRAN architecture for a range of real-world solutions. A high level design for an intra-PHY fronthaul interface has been agreed within the group, to enable the activities to progress. The idea hardware platform being developed should enable flexibility, such that alternative fronthaul interfaces can be reconfigured by software/firmware upgrades.

2.4.9 Small Cell Forum

Recognizing that several small cell equipment vendors were already achieving great success by centralizing and virtualizing some small cell base station functionality, the Small Cell Forum has defined a MAC/PHY split (option 6 with 3GPP terminology) aimed at avoiding industry fragmentation around the increasing adoption of virtualized small cell base stations. This open interface called nFAPI is evolved from the SCF's previous FAPI MAC/PHY platform interface that has been successfully used to support the multi-vendor small cell ecosystem [18].
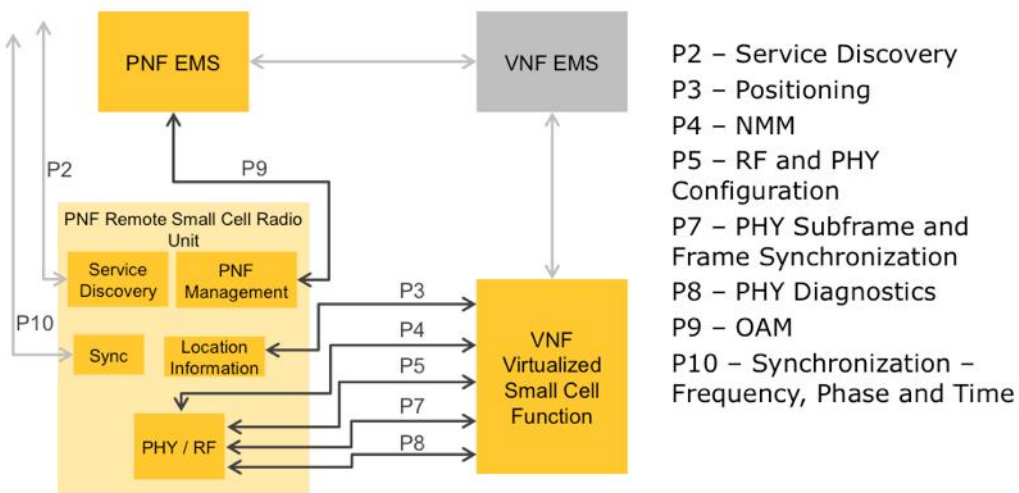

P2 – Service Discovery
P3 – Positioning
P4 – NMM
P5 – RF and PHY Configuration
P7 – PHY Subframe and Frame Synchronization
P8 – PHY Diagnostics
P9 – OAM
P10 – Synchronization – Frequency, Phase and Time

Figure 21: nFAPI Interfaces

From a management perspective, nFAPI has borrowed from the established DAS model in terms of how to partition shared physical network function (PNF) resources, and enable a common physical unit to be able to provide services from multiple virtual network functions (VNF)s that in turn may be operated by different service providers [22].
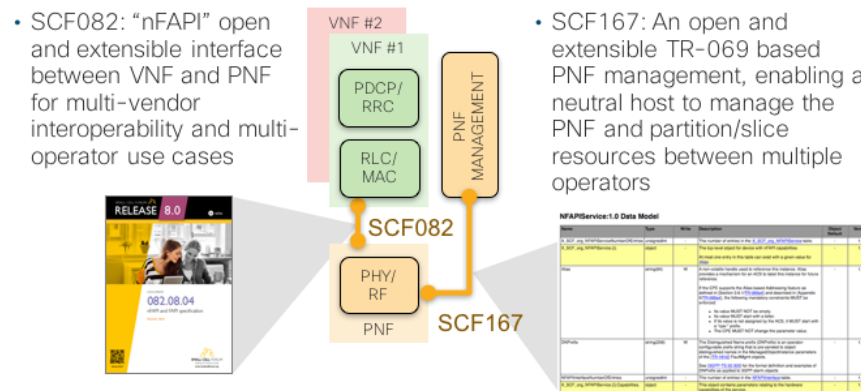

Figure 22: SCF's nFAPI RAN Architecture

One of the challenges introduced with operating a split small cell over a fronthaul link is associated with the latency and in particular the limited delay budget necessary to meet HARQ processing requirements. Small Cell Forum's document SCF160 discusses the use of HARQ interleaving that uses standardized signalling to defer buffer emptying, enabling higher latency fronthaul links to be used with nFAPI [23].

While HARQ interleaving functionally is standardized, its operation does add additional latency to the overall transmission flow in both the downlink and uplink, and may, when a split small cell is serving only a limited number of UEs, limit the peak cell throughput.

Another key capability that is impacted by latency budgets it Release 13 License Assisted Access (LAA) capability. In particular, as it relates to FAPI and nFAPI, LAA introduces Listen Before Talk (LBT) procedures that need to complete before any signal is transmitted. In particular, the timing of the LBT procedures occur on the micro-second basis, e.g., a defer duration is defined as (16μs + n x 9μs). Although LBT can be logically considered a Medium Access Control functionality, the tight timing requirements mean that moving the LAA Sensing Engine below the FAPI/nFAPI interface, ensures that nFAPI transport delay budgets are not impacted by LAA adoption.
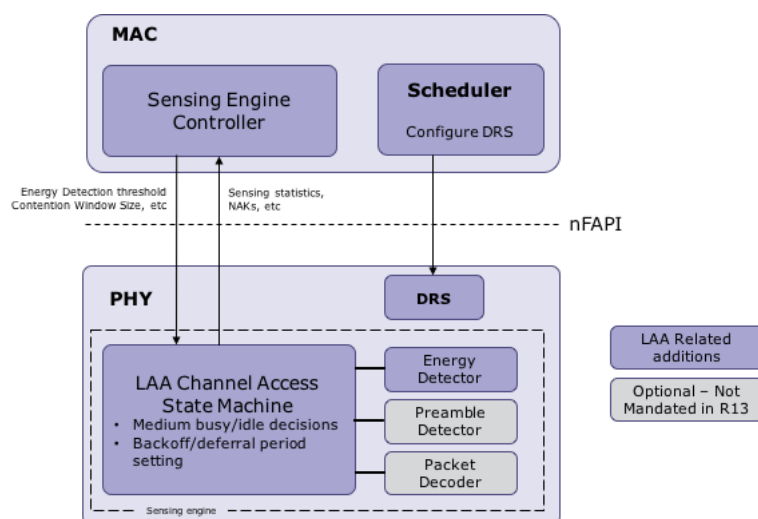

Figure 23: nFAPI support for LAA's Listen Before Talk functionality [24]

2.4.10 Open Air Interface

The Open Air Interface (OAI) software alliance is a non-profit consortium fostering a community of industrial as well as academic contributors for open source software and hardware development for the core network (EPC), access network and user equipment (EUTRAN) of 3GPP cellular networks. The software can interface with commercial SDR platforms such as USRP B210 for over-the-air (OTA) experiments with commercial devices.

The current OAI EUTRAN implementation supports time domain and frequency domain IQ splits (3GPP splits 8 and 7.1). More recently, the OAI code has been re-architected to enable support for other split options, as illustrated in Figure 24. In particular, OAI has overhauled their MAC-PHY interface, leveraging the nFAPI structures defined in Cisco's open-nFAPI repository for supporting split option 6 [25]. Additionally, the PDCP/RRC functionality has been decoupled from the Radio-Access Unit (RAU) functionality, ready for split 2 support.
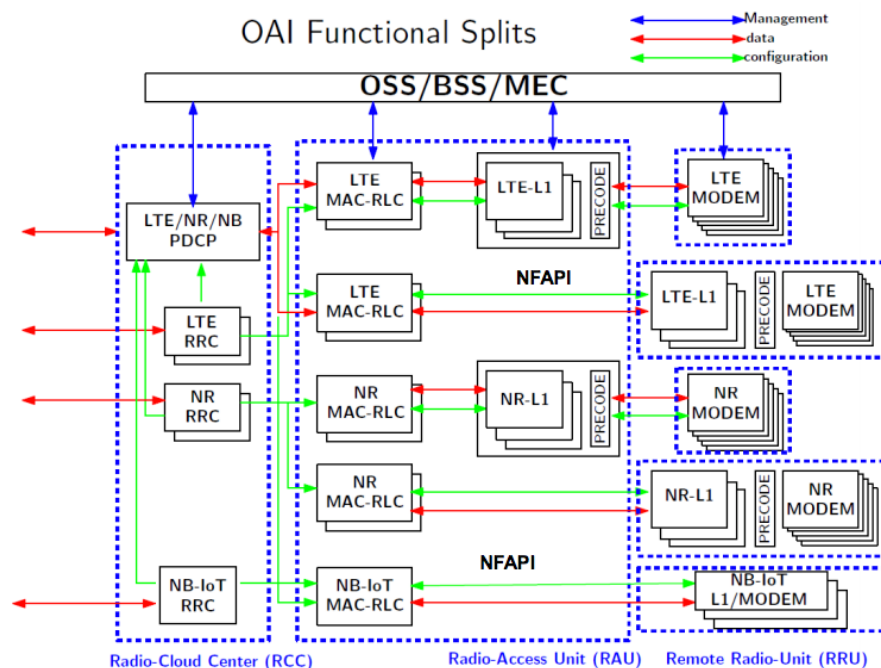
Figure 24: OAI Functional Splits

At the OAI fall workshop in November 2017, results were shared of progress in implementing a CU/DU split based on 3GPP's F1 interface [26]. This project is aimed at developing the protocol for the full communication between the CU and DU units and is targeted at supporting multi-vendor applications (e.g., OAI CU and third party DUs) as well as multi-tier splits (e.g., CU/DU/RRU).

In January 2018, the OAI merged the nFAPI split 6 functionality into their develop branch [27].

2.4.11 802.1CM

There are multiple IEEE 802.1 standards and ongoing projects relevant for 5G transport. Most of these standards are being developed in the Time-Sensitive Networking (TSN) Task Group, e.g., IEEE 802.1Qbv Enhancements for Scheduled Traffic, IEEE 802.1Qci Per-Stream Filtering and Policing, IEEE 802.1Qch Cyclic Queueing and Forwarding, and IEEE 802.1CB Frame Replication and Elimination.
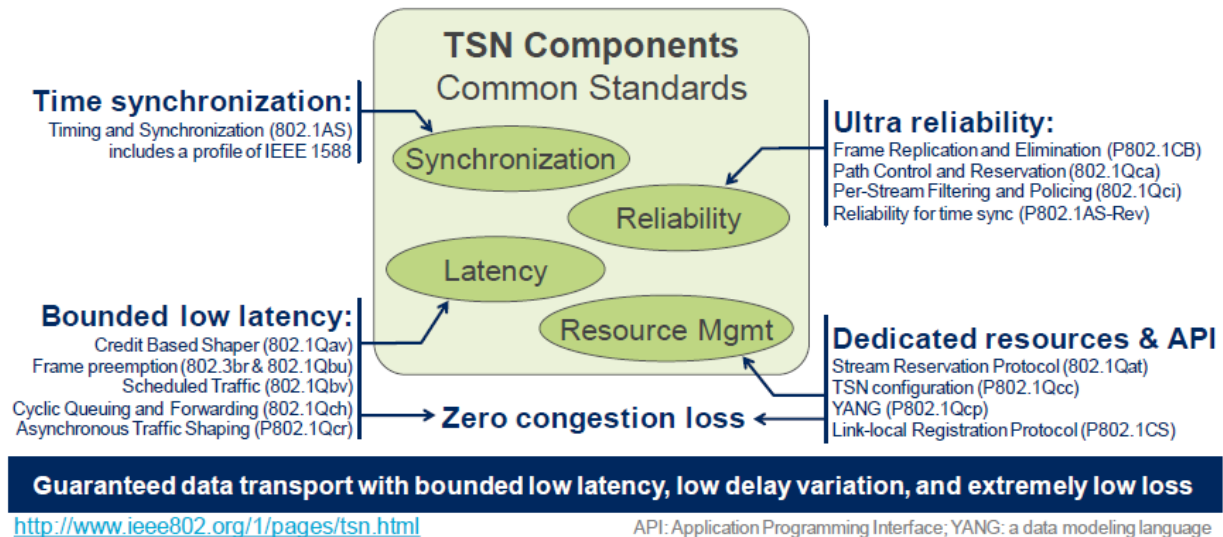
IEEE 802.1 Time-Sensitive Networking

Figure 25: Scope of IEEE 802.1 TSN activities

The ongoing IEEE P802.1CM Time-Sensitive Networking for Fronthaul project in the IEEE 802.1 Time-Sensitive Networking (TSN) Task Group is relevant to fronthaul. The P802.1CM objective is to specify bridged transport networks over IEEE Std 802.3 Ethernet for the transport of fronthaul traffic, including user data, and control & management (C&M) data. The work includes the fronthaul interfaces where the split of the radio base station is according to the Common Public Radio Interface (CPRI 7.0) specification, and the intra-PHY splits are supported by the eCPRI V1.0 fronthaul interface specification. P802.1CM has completed its first Working Group ballot on draft 1.0.

The Fronthaul network considered by the IEEE P802.1CM is illustrated in the figure below for a CPRI example.
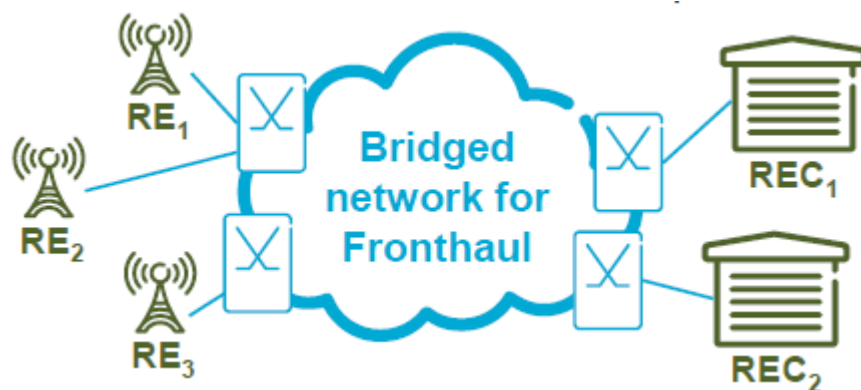
Figure 26: IEEE 802.1CM example fronthaul network

The IEEE 802.1CM specification provides guidance for meeting fronthaul requirements, which includes selecting 802.1 TSN features in order to build networks capable of transmitting fronthaul streams like decomposed CPRI. It also describes how the selected TSN features and components can be combined, configured and used in order to meet fronthaul requirements.

The standards from which features and options are selected by this standard are as follows:

a) The VLAN Bridge specification in IEEE Std 802.1Q.
b) Interspersing express traffic specification in IEEE Std 802.32.
c) Frame preemption specification IEEE Std 802.1Q3.
d) The time synchronization standard IEEE Std 1588, and one of its suitable PTP profiles.
e) The MAC and PHY standards specified for the various MAC/PHY technologies as specified by IEEE Std 802.3.

The IEEE P802.1CM standard is scheduled for completion by end of 2019, however it is very likely to be completed much earlier within a year.

3   HIGH LAYER SPLIT (HLS)

3.1   Introduction

In order to reduce OPEX and to provide a flexible RAN architecture, the functionality of a typical base-station was analysed with a view to centralize a number of its functions.

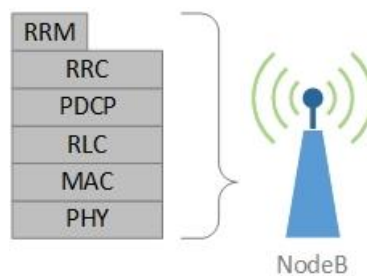A typical base-station or NodeB is illustrated in the following Figure 27:

Figure 27: Internal Protocol Stack of a typical base-station

A typical NodeB consists of a number of protocol layers as shown in Figure 27, and it was generally agreed in 3GPP RAN3 working group that lower protocol layers, such as RLC, MAC and PHY layers must be co-located to maintain tight synchronisation between each layers; this leaves RRC and PDCP layers to be separated away from the NodeB and moved into a centralised location.

3.2   Architecture Option 2

A number of split options were analysed as discussed in [1]: By reference to Table A-1 shown in [1] document and the Study Item on the Lower Layers split carried out in RAN3, it becomes apparent as we split further down the protocol stack, that the transmission requirements become very significant, very costly and impractical for large mobile networks. Thus a balance had to be struck between network function centralisation and practicality.

Following discussions in RAN3 on the advantages and the disadvantages of various functional splits, Option 2 was selected as the High Layer Functional Split Architecture. See following Figure 28.
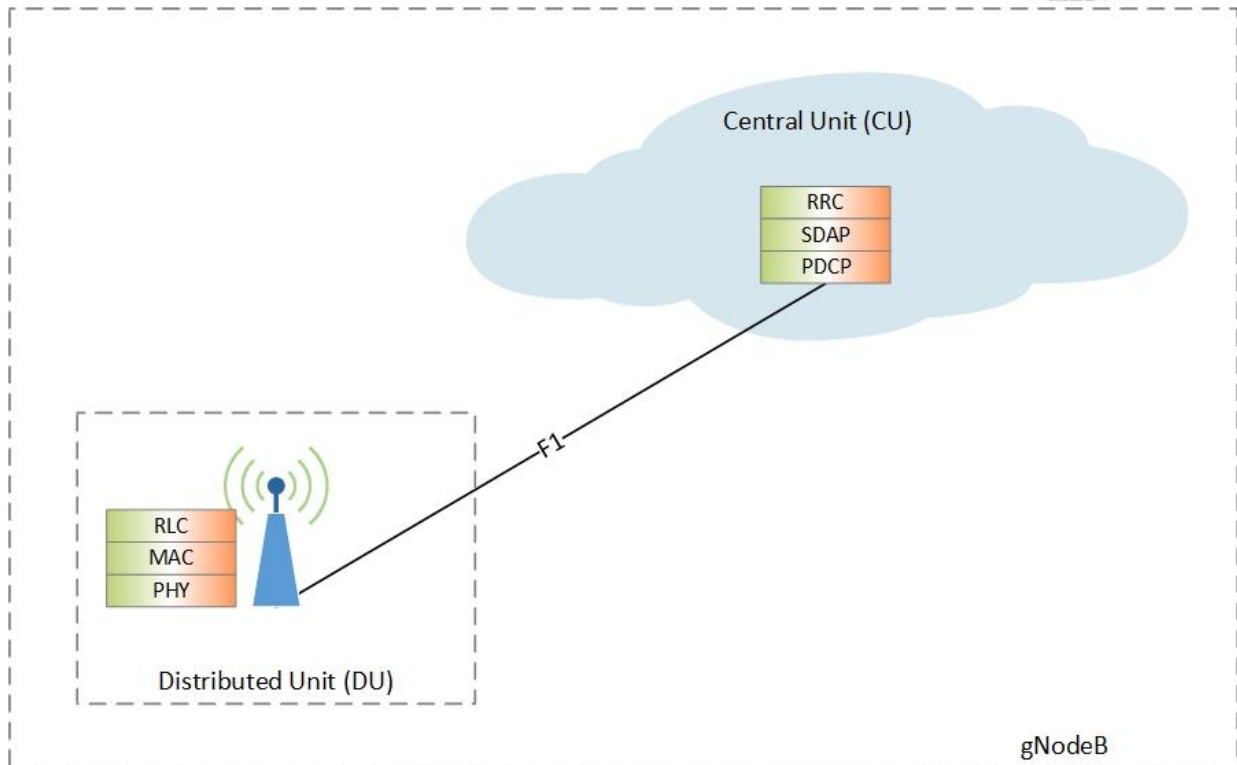
Figure 28: Architecture Option 2 - Split base-station

With the Option 2 Architecture, the typical base-station is now "disaggregated" into two units:
1- Centralised Unit (CU), housing RRC, PDCP and SDAP functions
2- Distributed Unit (DU) housing RLC, MAC and Physical layer functions

This functional split offers:
- Flexible hardware implementations which can reduce network deployment and OPEX cost
- Lower transmission requirements, compared to low layer split options
- Less susceptible to network latencies and poor network synchronisation
- Better network integration
- Load management, real-time performance optimization, and NFV/SDN features can be better co-ordinated

With the separation of the RRC, PDCP and SDAP from RLC-MAC-PHY layers a new interface was opened up, which was named by RAN3 as F1 interface. This interface would carry control signalling and user plane data between CU and the DU.

3.3   Separation of Control and User Planes

During the RAN3 work, the Option 2 Architecture was further split into Control and User Planes such that the elements in the CU were split into Control Plane Signalling (CU-CP) and one or more User Plane (CU-UP) logical entities. Similarly, the F1 interface was split into control and user plane, F1-C and one or more F1-U respectively. With the separation of control and user plane functions within the CU, the different instances of the PDCP function can also be split such to allow Control PDCP (PDCP-C) and User Plane (PDCP-U) to flow independently over F1-C and F1-U, respectively. As a consequence of control and user plane separation, another new interface was opened up which was names as E1 interface. This E1 interface would carry signalling information between the Control (CU-CP) and a given instance of the User Plane (CU-UP) logical entities, as shown in Figure 29.
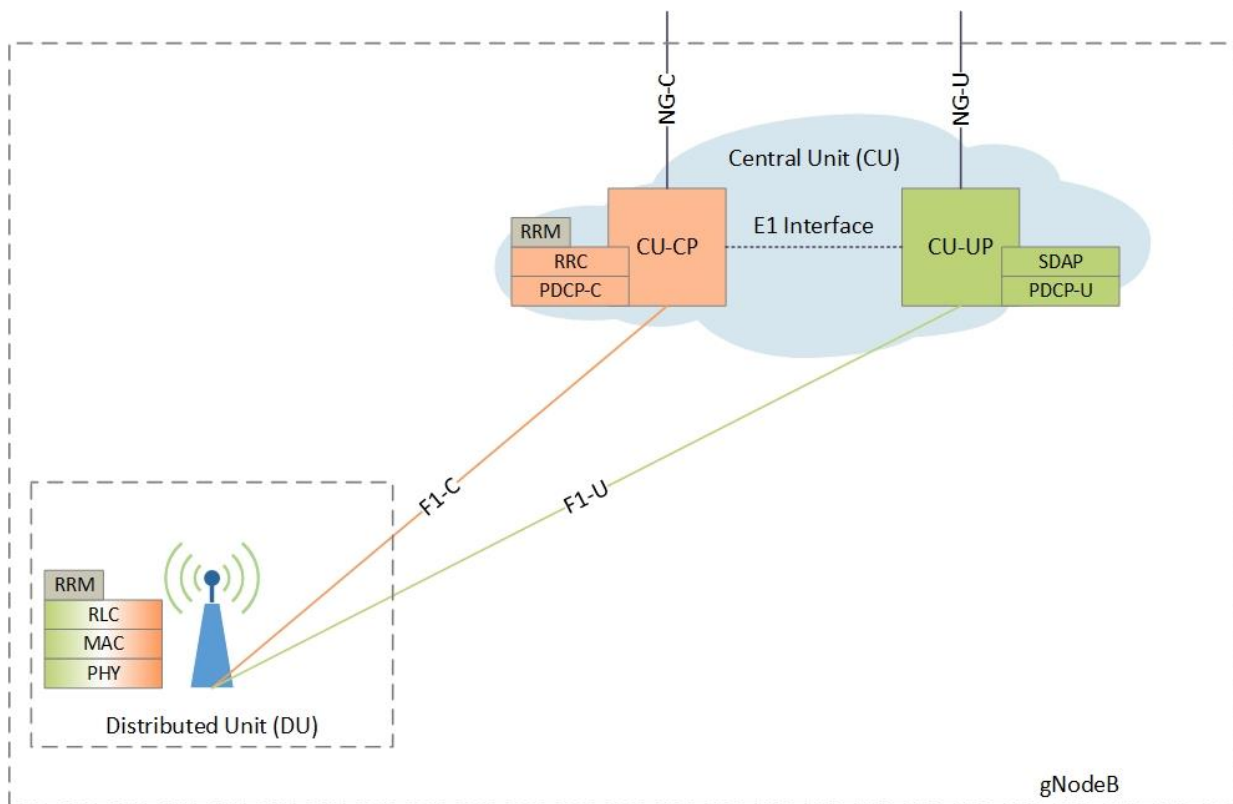
Figure 29: Architecture Option 2 with separated Control and User planes logical entities

3.4 Deployment Considerations

With the separation of the Control and User Plane functions in the CU and the separation of F1 into control and user planes, a number of network deployment options are created. These deployment options give the network operator flexibility to deploy a variant of the Option 2 Architecture to suit a particular use.

Three main deployment scenarios were identified [10]:
1- Scenario 1: Centralised CU-CP and CU-UP
2- Scenario 2: Distributed CU-CP and Centralised CU-UP
3- Scenario 3: Centralised CU-CP and Distributed CU-UP

3.4.1 Scenario 1: Centralised CU-CP and CU-UP

This scenario represents the basic case where the CP and UP logical functions are co-located in a centralised data centre, housing core network elements and application layer elements. In this deployment, RRC and PDCP-C are housed in the CU-CP and the PDCP-U and SDAP functions are housed in the CU-UP.

The Control Plane Signalling Entity (CU-CP) and the User Plane Entity (CU-UP) are both centralised and consequently all signalling and data to DUs flow down from CU-CP and CU-CP respectively, as shown in Figure 30.
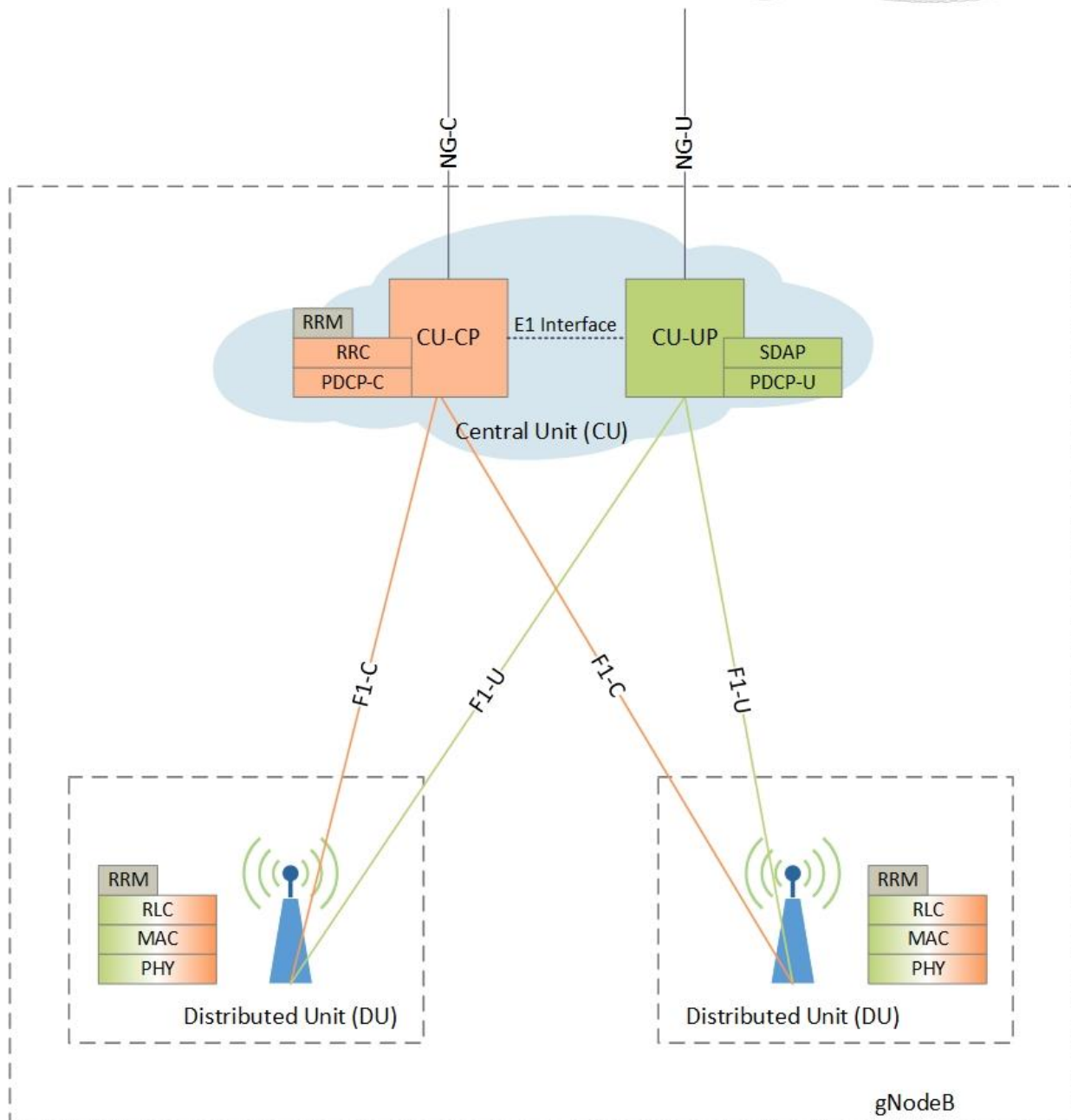
Figure 30: Scenario 1, Centralised Control Signalling and User Plane Functions

The Advantages and Disadvantages of this particular deployment options are as follows:

Advantages:
1- Centralized CU-CP potentially provides efficient load balancing and radio coordination of several DUs
2- By implementing Cloud Technologies, the CU-CP and CU-UP functions can be virtualised
3- In scenarios where the CP and UP functions are co-located the E1 interface would be internal to the CU
4- User plane, CU-UP functions can be increased independently of the CP as the network expands and as the user data volume increases
5- The user plane latencies are reduced for inter-site dual connectivity deployment
6- Parallel CU-UP instances may be placed at different locations (i.e. cell site, tier 1 and tier 2 sites)

7- Inter-site dual connectivity (NSA) use cases (i.e. options 3, 4 and 7) may be supported without introducing excessive transport latency

8- User plane transport latency increase is insignificant as long as CU-UP is placed at a suitable location between cell site and UPF site

Disadvantages:

1- Control plane functions are away from the DUs which may introduce performance impacts when transport latencies are well in excess of limits

2- For a network with DU separated several miles from the CU, network latencies could cause issues for some applications which require low latencies, this may be avoided for low latency services using multiple CU-UP locations

3.4.2   Scenario 2: Distributed CU-CP and Centralised CU-UP

In this scenario, the user plane functions and the CU-UP logical entity are located in the Centralised units and the Control Plane Signalling function, CU-CP is located in the DU.

In this configuration the Control signalling provides rapid signalling to the UEs for low latency applications.

Furthermore, the F1-C is internal to the DU and no control signalling flows over the transmission network.

With the separation of the CU-CP and CU-UP, the E1 interface flows over the transmission network between the CU and DU as shown in Figure 31:
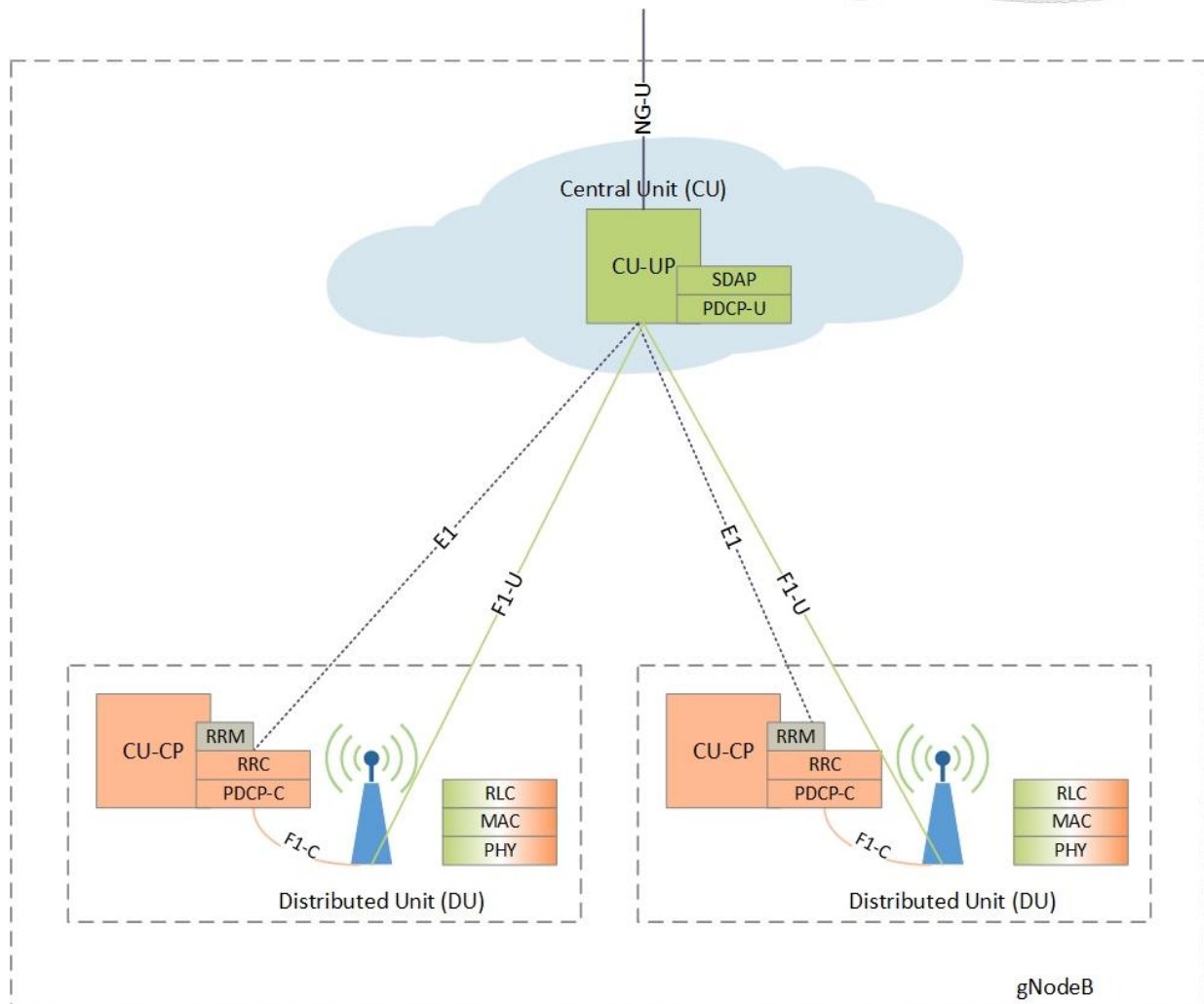
Figure 31: Scenario 2, Centralised User Plane function, CU-CP is in the DU

The Advantages and Disadvantages of this particular deployment options are as follows:

Advantages:
1- By placing the CP logical entity in the DU site allows more "soft" radio information to be used, and/or, an easier development path from an LTE eNodeB that has an integrated CP logical entity
2- For UEs using multi-site dual connectivity, there are transmission capacity, cost and latency benefits from placing the UP functionality 'high' up in the network
3- Control signalling are delivered rapidly to the UEs, suitable for low latency applications
4- Cloud technologies can be used to provide an efficient user data delivery and management
5- Load on the transmission network is reduced as only E1 flows over the transmission network

Disadvantages:
1- As E1 control signalling flows between the DU and CU, depending on the network topology and design, some latencies could occur. However, these latencies can be minimised by careful network design and efficient transport network infrastructure

3.4.3   Scenario 3: Centralised CU-CP and Distributed CU-UP

In this deployment scenario, the user plane function CU-UP logical entities are located in the DUs whilst the Control Plane Signalling Logical Entity, CU-CP, is centrally housed in the CU.

In this configuration, the CU-UP provides a rapid access to user plane data through the NG-U interface or for some applications from a locally cached data centre. Furthermore, the DUs are centrally controlled by CU-CP and all control plane signalling flow over the F1-C interface, as shown in Figure 32.
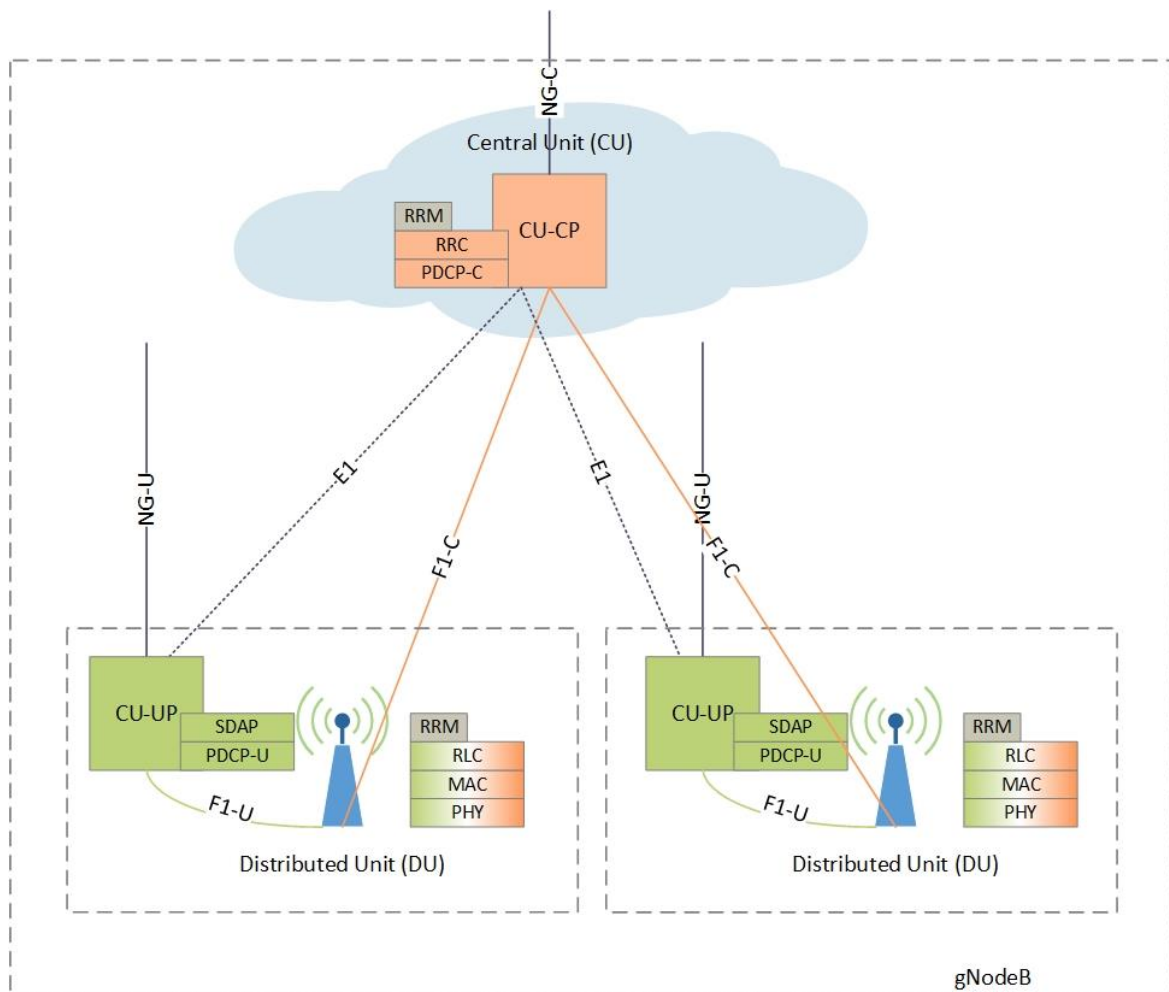
Figure 32: Scenario 3, Centralised Control Plane function, CU-UP is in the DU

The Advantages and Disadvantages of this particular deployment options are as follows:

Advantages:
1- Provides a flexible architecture for placing user data close to the UE or applications, such as factory automation or MTC
2- Control signalling functions can be centralised in the CU by using NFV/SDN technologies
3- Frequently used user or application data can be cached close to the DU to reduce data fetch cycle

Disadvantages:
1- This deployment scenario may introduce additional latency on the control plane, however by careful network design and topology this can be minimised

2- This deployment scenario will introduce excessive transport latency for Xn interface when used to carry split bearer flows in inter-site NSA dual-connectivity

3.5 Latency

Latencies in general vary from one network to another and it is difficult to quantify these parameters due to differences in network topologies and performance of equipment along the way, e.g. from DU to CU etc. However in all RAN3 discussion the one-way latency was assumed to be around 5-8ms: 5ms was the average agreed latencies for modern transmission networks and 8ms for legacy networks. Moreover, 8ms network latency between CU and DU eases the co-location of the CU with the Serving Gateway and other application platforms.

3.6 Synchronisation

Partial gNodeB synchronization has been specified in section 9 of [8]. This refers to CU-DU synchronization. gNodeB and 5G core network synchronization is still under study.

Network transmission requirements such as packet loss, jitter and latencies have not yet been discussed.

In order to satisfy the synchronization requirements for 5G services specific frequency and phase requirements need to be met at the air interface of the Radio Unit (RU). The Telecommunications area of the International Telecommunications Union (i.e. ITU-T) has been working on synchronization. Within ITU-T, Study Group 15 (SG15) works on networks, technologies and infrastructure for transport, access and home use. The functional area within SG15 dealing with matters concerning synchronization is covered by Question 13 (Q13) covers all matters relating to the generation and transport of timing and synchronization signals.

Synchronous Ethernet and the Telecom Profiles of the Precision Timing Protocol (PTP) based on the IEEE1588 standard form the basis of transporting synchronization to meet mobile requirements. ITU is continuing to enhance both Synchronous Ethernet and PTP to address 5G mobile requirements in the suite of ITU-T G.826x and ITU-T G.827x series of recommendations. In addition these recommendations also describe the generation of timing using references derived from Global Navigation Satellites System (GNSS) signals and from atomic clock sources.

3.7 Encryption and Integrity Protection

Security considerations depend on the network deployment scenario and the related configurations. The security threats for the created F1 interface are generally different from those existing for the air interface. Basic security features include confidentiality, integrity and anti-replay protection, for Control Plane traffic CP on the F1-C interface and for User Plane traffic UP on the F1-U interface. Apart from RRC signalling traffic, F1-C also includes F1 management traffic. For simplicity, anti-replay protection is considered here as a type of integrity protection, where replayed messages are detected/prevented through integrity protection and the use of nonces.

According to 3GPP TS 33.501 [28], the 5G network mandatorily supports confidentiality and integrity protection of both CP and UP, in terms of the corresponding encryption and integrity keys, derived from the UE subscription credential, respectively.

The use of integrity protection for RRC signalling on F1-C, over PDCP-C, is mandatory, and the use of encryption for RRC signalling on F1-C, over PDCP-C, is optional.

The use of both encryption and integrity protection of UP on F1-U, over PDCP-U, is optional. Note that the integrity protection increases the packet size, which may have a considerable impact on UP, especially for the air interface.

The use of both encryption and integrity protection for F1 management traffic on F1-C is mandatory, due the sensitivity of these signalling data. To this end, it is necessary to establish a new security channel between DU and CU, possibly offering protection over the network layer via IPsec and key exchange IKEv2 protocols. Both traditional and virtualised implementations of this channel are possible. Alternative solutions over the transport layer (e.g., TLS over SCTP) or the application layer may also be possible.

Basic security functions on F1 also include authentication and key management. Mutual authentication between gNB-CU and gNB-DU should be strong and (possibly) based on public-key certificates. In particular, key management also includes the secure provisioning of UE-specific PDCP secret keys to gNB, for protecting CP and UP between UE and gNB.

3.7.1   Scenario 1 with Centralised CU-CP and CU-UP

Scenario 1, with both CU-CP and CU-UP centralised, enables a security solution whereby both the encryption and integrity protection over F1 can be realized by the PDCP tunnel between UE and CU, more precisely, by PDCP-U for UP and PDCP-C for CP RRC signalling. For this purpose, the corresponding UE-specific keys are used. They are derived from the gNB key received from the Access and Mobility Function (AMF) function. The same keys are locally derived in the UE. Accordingly, no additional security channel is required for the protection of UP and CP RRC signalling over F1.

Provided that the CU is placed in a secure environment, an advantage of this scenario is that UE-specific PDCP secret keys are then managed only in the CU, and not in the distributed DUs, which are more likely to be placed in less secure environments.

### 3.7.2   Scenario 2 with Centralised CU-UP

Scenario 2, with centralised CU-UP, enables a security solution whereby the encryption and integrity protection over F1-U can be realized by the PDCP-U tunnel between UE and CU-UP for UP. For this purpose, the corresponding UE-specific keys derived from the gNB key (only for PDCP-U) are used. Accordingly, no additional security channel is needed for the UP protection, over F1-U.

Since F1-C is located in the DU, the corresponding UE-specific derived PDCP-C keys need to be securely provisioned to and processed in the DU. PDCP-C integrity protection is mandatory, while optional PDCP-C encryption then essentially only serves for the air interface.

3.7.3   Scenario 3 with Centralised CU-CP

Scenario 3, with centralised CU-CP, enables a security solution whereby the encryption and integrity protection over F1-C can be realized by the PDCP-C tunnel between UE and CU-CP for CP RRC signalling. For this purpose, the corresponding UE-specific keys derived from the gNB key (only for PDCP-C) are used. Accordingly, no additional security channel is needed for the CP protection, over F1-C. PDCP-C necessarily includes the integrity protection. For illustration, this scenario is depicted in Figure 33.

Since F1-U is located in the DU, the corresponding UE-specific derived PDCP-U keys need to be securely provisioned to and processed in the DU. Optional PDCP-U protection then essentially only serves for the air interface.
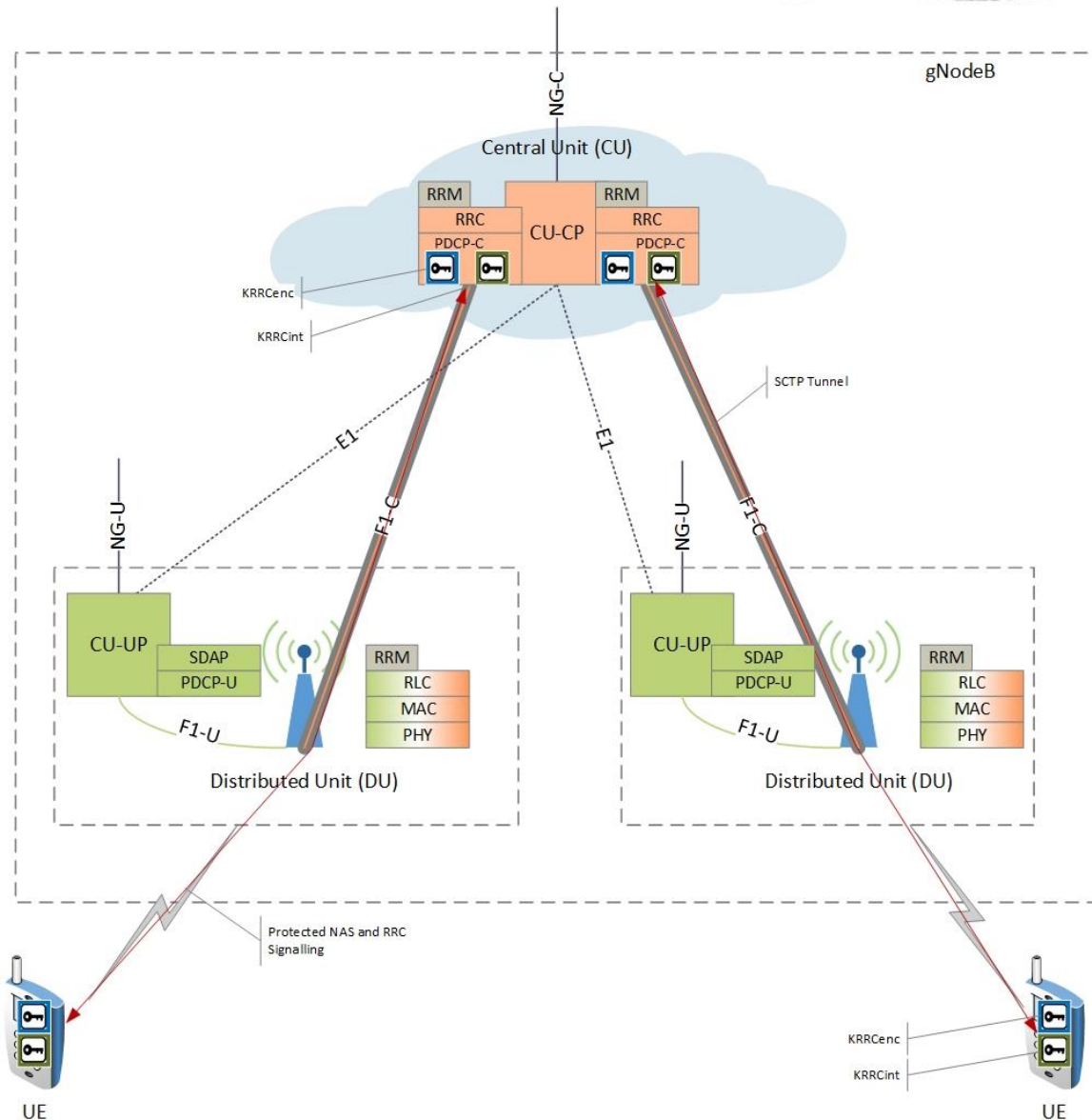
Figure 33: Control Plane Security Solution with Scenario 3 Deployment

3.8 Network Dimensioning

Network dimensioning is generally dependent on following:
- User Traffic Profile
- Signalling Load per user and per cell
- Number of Cells connected to one DU
- Control Signalling load over F1-C
- User Traffic load over F1-U
- Number of DUs connected to one CU
- Network equipment capability
- Network Topology
- Etc.

At the time of writing this document, no dimensioning data was available.

3.9   Summary

The work on Option 2 Architecture has been completed and baseline changes have now been agreed to 36.xxx, 37.xxx and 38.xxx series of 3GPP Standards.

For future architectural work, depending on the application and network requirements, the three deployment scenarios discussed in this chapter can be used.

4 LOW LAYER SPLIT (LLS)

The Low Layer Split (LLS), using 3GPP terminology, refers to RAN functional split options from option 6 (MAC/PHY split) down to option 7 (PHY split). It should be noted that 3GPP did not discard option 8, but it agreed that no further work is needed on such option. Within 3GPP it was concluded that the split options 6 and 7 are possible implementations of a low layer split architecture. Namely variations to such options are possible, as specified in [29]. Within option 7, there are 3 sub-options identified (3 for downlink, but only 2 of those apply to uplink). For options with sufficient levels of low layer function centralization, the value of the LLS comes from the improved coordination among adjacent cells and pooling gains. For options where low layer functions are kept to a higher extend in decentralized nodes, the gain is a considerably less demanding transport requirements in comparison to, e.g., an option 8 split, which can therefore scale to massive MIMO more easily. It needs to be however observed that transport bandwidth depends greatly on the particular split option, as concluded in [11].

Within 3GPP there is a study item on the LLS for NR [11]. This analysis provided information about the various LLS options, and concluded that all of the identified lower layer split options are feasible depending on the particular radio configuration. For this reason, it was concluded in [11] that there is a preference for 3GPP to be open to all of the identified lower layer split options (and even further to the variants thereof). At the time of writing, this 3GPP activity is on hold, to be reassessed in March 2018.

Whilst the Small Cell Forum has already published their split option 6 specification, other industry groups have also been evaluating the LLS options, notably TTA, xRAN, TIP and IEEE 1914 WG. xRAN are currently specifying an open interface for a LLS [14]. According to the latest public information, this interface will support 2 split options for downlink (7-1 and 7-3) and 2 split options for uplink (7-1 and 7-2). TIP project "vRAN with non-ideal fronthaul" are looking to develop a vRAN ecosystem that supports an open option 7 split. The work is starting with a high level reference design where the fronthaul interface is designed to cope with non-ideal transport (relatively low throughput and relatively high latency). At the time of writing, neither xRAN nor TIP have published their fronthaul interfaces.

Meantime, although some other organizations, e.g. IEEE 1914, eCPRI, are not directly specifying the LLS interface itself, their work is closely relevant. For example, the IEEE 1914.3 project "Radio Over Ethernet" does specify mechanisms to support open interfaces with LLS option 7-1 and 7-2 to allow multi-vendor interoperability. At the time of writing, the 1914.3 RoE specification is at the sponsor ballot stage and is very likely to be published in the first half year of 2018.

Finally, TTA's recently published fronthaul interface [15], can support all variants of option 7 split and tries to align with other industry activities (e.g. IEEE 1914.3 and 3GPP). At the time of writing, this specification is only available in Korean, reflecting where earliest adoption may occur, but may be available to a wider audience once translations are provided in the near future.

Initial analysis of the possible location of the LLS (option 6 down to option 7-1) shows a significant variation in transport requirements depending on where the functional split occurs [1], so the chosen split location can greatly change the transport design and overall RAN architecture. Furthermore, the analysis done to date is based around an LTE stack, so NR may have some differences.

NGMN will track these activities with interest and try to offer guidance. In order to avoid fragmentation of the industry, these related groups should collaborate and try to align where their solutions are similar or can support each other.

ABBREVIATIONS

5GC	5G core
BBU	Base band unit
C&M	Control and management
CP	Control plane
CU	Centralized unit
CU-CP	Centralized unit-control plane
CU-UP	Centralized unit-user plane
CPRI	Common public radio interface
DC	Dual connectivity
DU	Distributed unit
eRE	eCPRI radio equipment
eREC	eCPRI radio equipment control
eMBB	enhanced mobile broadband
EPC	Evolved packet core
FAPI	Functional application programming interface
gNB	5G NodeB
HARQ	Hybrid automatic repeat request
HLS	High layer split
IoT	Interference of things
LAA	License assisted access
LLS	Low layer split
LTE	Long term evolution
MAC	Medium access control
MEC	Multi-access edge computing
MIMO	Multiple input multiple output
mMTC	massive machine type communication
MN	Master node
nFAPI	Network functional application programming interface
NGCN	Next generation core network
NGFI	Next generation fronthaul interface
NGMN	Next generation mobile networks
NR	New radio
NSA	Non-standalone
OAI	Open air interface
OBRI	Open BBU RRH interface
OBSAI	Open base station architecture initiative
ORI	Open radio equipment interface
PDCP	Packet data convergence protocol
PDCP-C	PDCP control plane
PDCP-U	PDCP user plane
PHY	Physical layer
RAN	Radio access network
RLC	Radio link control
RoE	Radio over ethernet
RRC	Radio resource control
RRH	Remote radio head
RRU	Remote radio unit
RTT	Round trip time
RU	Radio Unit
SA	Standalone
SCF	Small cell forum
SDAP	Service data adaptation protocol
SN	Secondary node
TIP	Telecom infra project
TSN	Time sensitive networking
UP	User plane
UPF	User plane function

uRLLC   ultra-reliable low latency communications
vBBU   Virtualized base band unit

REFERENCES

[1]   3GPP TR 38.801, V14.0.0, "Study on new radio access technology: Radio access architecture and interfaces", Mar. 2017.

[2]   3GPP TS 23.401, V15.2.0, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 15)", Dec. 2017.

[3]   3GPP TS 23.501, V15.0.0, "System Architecture for the 5G System", Dec. 2017.

[4]   3GPP TS 36.300, V15.0.0, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description", Dec. 2017.

[5]   3GPP TS 37.340, V15.0.0, "Evolved Universal Terrestrial Radio Access (E-UTRA) and NR; Multi-connectivity", Dec. 2017.

[6]   3GPP TS 38.300, V15.0.0, "NR; NR and NG-RAN Overall Description", Dec. 2017.

[7]   Nokia, "5G Migration Strategies", NGMN Forum, Oct. 2017.

[8]   3GPP TS 38.401, V15.0.0, "NG-RAN; Architecture description", Dec. 2017.

[9]   3GPP TR 36.756, V1.0.0, "Study on architecture evolution for Evolved Universal Terrestrial Radio Access Network (E-UTRAN)", Sep. 2017.

[10]   3GPP TS 38.806, V15.0.0, "Study of separation of NR Control Plane (CP) and User Plane (UP) for split option 2", Dec. 2017.

[11]   3GPP TR 38.816, V15.0.0, "Study on CU-DU lower layer split for NR", Dec. 2017.

[12]   CPRI, V7.0, "CPRI interface specification", Oct. 2015.

[13]   CPRI, V1.1, "eCPRI interface specification", Jan. 2018.

[14]   XRAN Forum, "Fronthaul working group", Oct. 2017.

[15]   TTA TTAK.KO-06.0461, "5G open fronthaul interface", Dec. 2017.

[16]   TIP, "Unbundled RAN Architecture: An open RAN architecture for multi-vendor interoperability", 2017.

[17]   TIP, "Virtualizing the network, whatever the transport", RAN USA, Dec. 2017.

[18]   Small Cell Forum, SCF082, "FAPI and nFAPI specifications", May 2017.

[19]   CPRI, D1.0, "Requirements for the eCPRI transport network", Aug. 2017

[20]   OBSAI, "BTS system reference document, Version 2.0" Nov. 2016.

[21]   ETSI GS ORI 001, V4.1.1, "Requirements for Open Radio Interface (ORI)", Oct. 2014.

[22]   Small Cell Forum, SCF167, "nFAPI service: 1.0 service object definition", Feb. 2017.

[23]   Small Cell Forum, SCF160, "Coverage and capacity impacts of virtualization", Jun. 2015.

[24]   Small Cell Forum, SCF108, "Virtualized small cells - Overview", Oct. 2016.

[25]   Cisco, "open-nFAPI repository", available online at: https://github.com/cisco/open-nFAPI [Accessed 24 Feb. 2018]

[26]   Makris, Basaras, Korakis, Nikaein, "Implementation of the CU/DU split functionality over OAI", OAI workshop, Nov. 2017.

[27]   OAI, "Nfapi howto", available online at: https://gitlab.eurecom.fr/oai/openairinterface5g/wikis/nFAPI-howto [Accessed 24 Feb, 2018]

[28]   3GPP TS 33.501, V0.6.0, "Security Architecture and Procedures for 5G System", Dec. 2017.

[29]   3GPP R1-1719203, "Reply LS on NR L1 processing chain", Oct. 2017