
NG Radio Access Network (NG-RAN)

Balazs Bertenyi¹, Richard Burbidge², Gino Masini³,
Sasha Sirotkin⁴ and Yin Gao⁵

¹*Chairman of 3GPP TSG-RAN, Hungary*

²*Chairman of 3GPP RAN2, Senior Wireless Systems Architect at Intel Corporation, Shrivenham, Oxfordshire, UK*

³*Chairman of 3GPP RAN3, Systems Manager – Concepts and Standards, at LM Ericsson AB, Stockholm, Sweden*

⁴*Vice Chairman of 3GPP RAN3, Senior wireless architect at Intel Corporation, Israel*

⁵*Vice Chairman of 3GPP RAN3, ZTE Corporation, China*

Received 10 April 2018;

Accepted 3 May 2018

Abstract

This paper presents an overview of the NG Radio Access Network (NG-RAN) architecture and key protocols. NG-RAN is the new RAN defined in conjunction with 5G by 3GPP. The paper presents the overall architecture, migration path options, the 5G base station architecture and key protocol components.

Keywords: 5G, NR, radio architecture, NG RAN, CU, DU, Xx, Xn, RRC, MAC, PDCP, RLC.

1 Overview of the NG-RAN Architecture

The NG-RAN represents the newly defined radio access network for 5G. NG-RAN provides both NR and LTE radio access. An NG-RAN node (i.e. base station) is either:

- a gNB (i.e. a 5G base station), providing NR user plane and control plane services;
- or,
- an ng-eNB, providing LTE/E-UTRAN services towards the UE.

Journal of ICT, Vol. 6-1&2, 59–76. River Publishers

doi: 10.13052/jicts2245-800X.614

This is an Open Access publication. © 2018 the Author(s). All rights reserved.

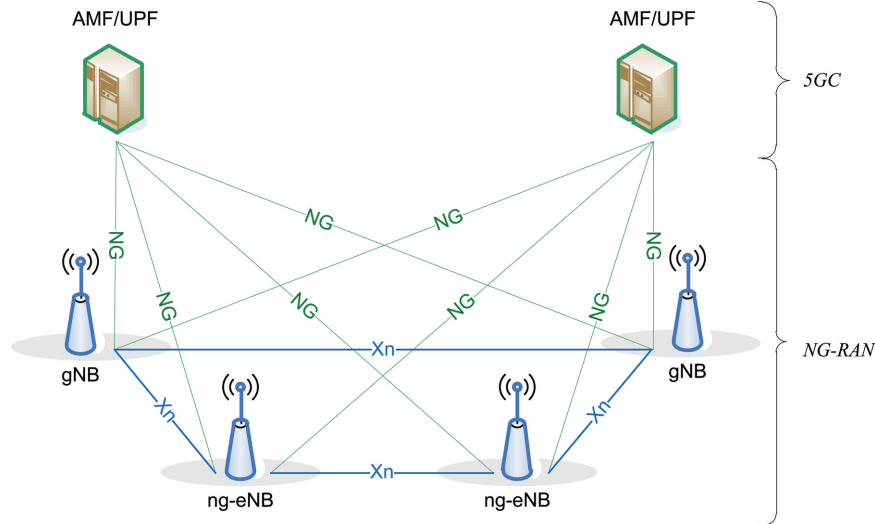


Figure 1 NG-RAN in relation to the 5G system.

The gNBs and ng-eNBs are interconnected with each other by means of the Xn interface. The gNBs and ng-eNBs are also connected by means of the NG interfaces to the 5G Core (5GC), more specifically to the AMF (Access and Mobility Management Function) by means of the NG-C interface and to the UPF (User Plane Function) by means of the NG-U interface.

The overall relation of NG-RAN in relation to the overall 5G system is shown in Figure 1.

Both the user plane and control plane architectures for NG-RAN follow the same high-level architecture scheme, as depicted in Figure 2 below. For further details of the protocol stacks see Section 4.

2 Architecture Options and Migration Paths

One of the distinctive features of NG RAN is the capability to operate in both so-called “Stand-Alone” (SA) operation and “Non-Stand-Alone” (NSA) operation. In SA operation, the gNB is connected to the 5G Core Network (5GC); in NSA operation, NR and LTE are tightly integrated and connect to the existing 4G Core Network (EPC), leveraging Dual Connectivity (DC) toward the terminal. In a Dual Connectivity architecture, a Master Node (MN) and a Secondary Node (SN) concurrently provide radio resources towards the terminal for enhanced end-user bit rates. Both NSA and SA architecture options are specified as part of the phase-1 5G standards of 3GPP in Release 15.

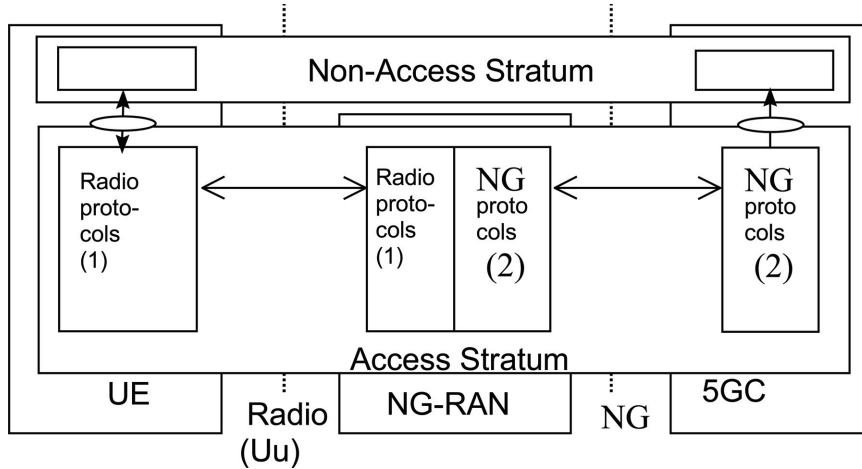


Figure 2 Overall NG-RAN architecture.

One can derive several different configuration options from the overall architecture, each of these options represent a viable deployment option for network operators. These architecture options are depicted in the subsections below. The numbering of these architecture options does not bear any particular logic or significance, it is purely historical.

2.1 NR gNB Connected to the 5GC (Option 2)

In this option, the gNBs are connected to the 5G Core Network (5GC) through the NG interface. The gNBs interconnect through the Xn interface.

2.2 Multi-RAT DC with the EPC (Option 3)

In this option, commonly known as EN-DC (LTE-NR Dual Connectivity), a UE is connected to an eNB that acts as a MN and to an en-gNB that acts as a SN, see Figure 3. An en-gNB is different from agNB in that it only implements part of the 5G base station functionality that is required to perform SN functions for EN-DC.

The eNB is connected to the EPC via the S1 interface and to the en-gNB via the X2 interface. The en-gNB may also be connected to the EPC via the S1-U interface and to other en-gNBs via the X2-U interface. The resulting architecture is shown in Figure 3 below. Notice that the en-gNB may send UP to the EPC either directly or via the eNB.

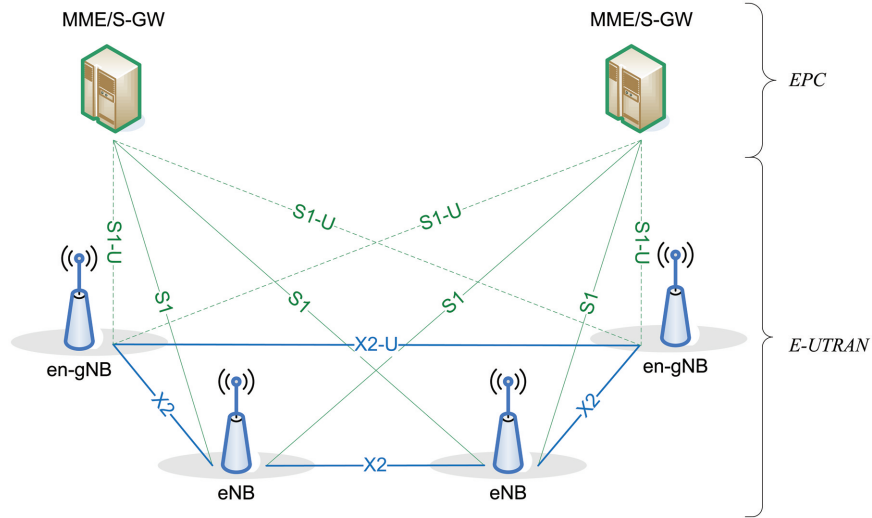


Figure 3 Overall LTE (E-UTRAN)-NR DC architecture.

2.3 Multi-RAT DC with the 5GC, NR as Master (Option 4)

In this option, a UE is connected to a gNB that acts as a MN and to an ng-eNB that acts as an SN. This option requires the 5G Core to be deployed. The gNB is connected to 5GC and the ng-eNB is connected to the gNB via the Xn interface. The ng-eNB may send UP to the 5G Core either directly or via the gNB.

2.4 LTE ng-eNB Connected to the 5GC (Option 5)

In this option, the ng-eNBs are connected to the 5G Core Network (5GC) through the NG interface. The ng-eNBs interconnect through the Xn interface. Essentially this option allows the existing LTE radio infrastructure (through an upgrade to the eNB) to connect to the new 5G Core.

2.5 Multi-RAT DC with the 5GC, E-UTRA as Master (Option 7)

In this option, a UE is connected to an ng-eNB that acts as a MN and to a gNB that acts as an SN. The ng-eNB is connected to the 5GC, and the gNB is connected to the ng-eNB via the Xn interface. The gNB may send UP to the 5GC either directly or via the ng-eNB.

2.6 Migration Considerations

When 5G is first rolled out with NR, a likely scenario is to deploy NR on higher frequencies than for LTE. In this case, NR coverage is typically much smaller than LTE coverage, especially with frequencies above 6 GHz. Then, it is desirable to leverage the existing LTE coverage to provide continuous nationwide coverage and mobility, while boosting User-plane capacity with NR in target areas with high traffic load. Option 3 enables operators to launch the NR service in this way, building on top of their existing investments for E-UTRAN and EPC.

When operators decide to introduce the 5G Core, this will “unlock” a new set of possible deployment scenarios, among which the support for NR as a stand-alone Radio Access Technology (RAT) (Option 2), while at the same time leveraging the deployed LTE nodes as secondary nodes through Dual Connectivity (DC) (Option 4). Another possibility for the introduction of 5G Core is to keep LTE as the main “anchor”, connecting it to the 5G Core (Option 5) while still leveraging NR as secondary node through DC (Option 7).

The choice between deploying NR with 5GC as “anchor” and keeping LTE as “anchor” with the new 5GC, will typically be a business decision by each operator. It will typically depend on factors including deployed LTE network density, availability of new frequencies, rate of increase for end-user traffic demand, and relative “weight” in the business case of new functionality (such as e.g. slicing) which only the new networks can provide.

3 5G NR Base Station (gNB) Architecture

The 4G RAN architecture was based on a “monolithic” building block, the eNB. This resulted in a very simple RAN architecture, where few interactions between logical nodes need to be specified. Since the earliest phases of the NR study, however, it was felt that splitting up the gNB (the NR logical node) between Central Units (CUs) and Distributed Units (DUs) would bring additional benefits. Some benefits in this regard were in fact identified already in the early study phase, including:

- A flexible hardware implementation allows scalable cost-effective solutions.
- A split architecture allows coordination of performance features, load management and real-time performance optimization. It also enables virtualized deployments.

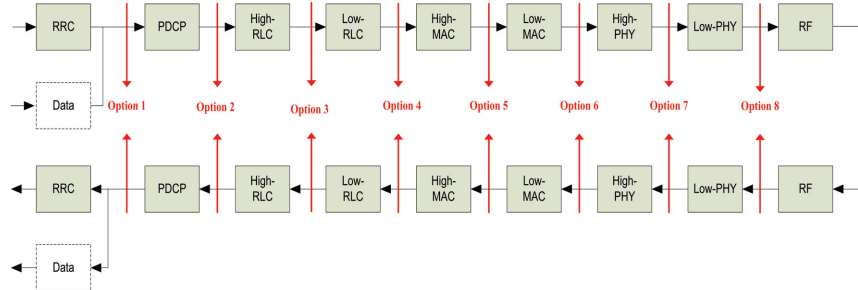


Figure 4 Function Split alternatives.

- Configurable functional splits enable adaptation to various use cases, such as variable transport latency.

The choice of how to split NR functions in the architecture depends on radio network deployment scenarios, constraints and envisaged services. For example, it depends on the need to support specific QoS settings per offered services (e.g. low latency, high throughput, specific user density and load demand per given geographical area (which may influence the level of RAN coordination), or the need to interoperate with transport networks having different performance levels: from ideal to non-ideal.

Several possible CU-DU split options, shown in Figure 4, were considered during the study phase. The E-UTRA protocol stack, which includes PHY, MAC, RLC, PDCP, and RRC, was taken as a basis for this investigation. The investigation analyzed the possible split points across the protocol stack, these possible split points are depicted in Figure 4 as different enumerated Options.

After detailed comparison 3GPP decided to take Option 2 (based on centralised PDCP/RRC and decentralised RLC/MAC/PHY) as a basis for normative specification work. The prime reason for selecting this option was the close similarity to the protocol stack split applied in Dual Connectivity: in a DC configuration the Master Node (MN) and the Secondary Node (SN) are “split” along the same point as Option 2.

3.1 Higher Layer Split (HLS) of the gNB

The overall NG-RAN architecture with a split gNB is shown in Figure 5 below.

As shown in Figure 5, in NG-RAN a set of gNBs is connected to the 5G Core Network (5GC) through the NG interface, and they can be interconnected through the Xn interface.

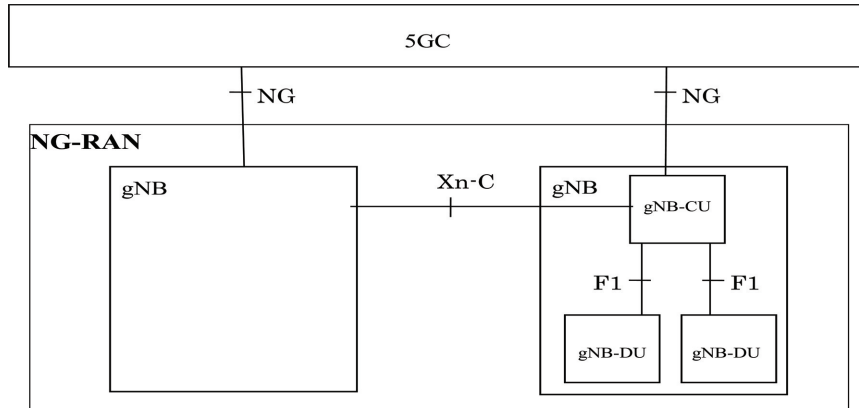


Figure 5 Higher Layer split of the gNB.

A gNB may then consist of a gNB-CU and one or more gNB-DU(s), and the interface between gNB-CU and gNB-DU is called F1. The NG and Xn-C interfaces for a gNB terminate in the gNB-CU. The maximum number of gNB-DUs connected to a gNB-CU is only limited by implementation. In 3GPP standard, one gNB-DU connects to only one gNB-CU, but implementations that allow multiple gNB-CUs to connect to a single gNB-DU e.g. for added resiliency, are not precluded. One gNB-DU may support one or more cells. The internal structure of the gNB is not visible to the core network and other RAN nodes, so the gNB-CU and connected gNB-DUs are only visible to other gNBs and the 5GC as a gNB.

The F1 interface supports signaling exchange and data transmission between the endpoints, separates Radio Network Layer and Transport Network Layer, and enables the exchange of UE-associated and non-UE-associated signaling. In addition, F1 interface functions are divided into F1-C and F1-U functions.

F1-C (Control Plane) Functions:

- F1 Interface Management Functions: These consist of F1 setup, gNB-CU Configuration Update, gNB-DU Configuration Update, error indication and reset function.
- System Information Management Functions: The gNB-DU is responsible for the scheduling and broadcasting of system information. For system information broadcasting, the encoding of NR-MIB and SIB1 is performed by the gNB-DU, while the encoding of other SI messages is performed by the gNB-CU. The F1 interface also provides

signaling support for on-demand SI delivery, enabling UE energy saving.

- **F1 UE Context Management Functions:** These functions are responsible for the establishment and modification of the necessary UE context. The establishment of the F1 UE context is initiated by the gNB-CU, and the gNB-DU can accept or reject the establishment based on admission control criteria (e.g., the gNB-DU can reject a context setup or modification request in case resources are not available). In addition, an F1 UE context modification request can be initiated by either gNB-CU or gNB-DU. The receiving node may accept or reject the modification. The F1 UE context management function can be also used to establish, modify and release Data Radio Bearers (DRBs) and Signaling Radio Bearers (SRBs).
- **RRC Message Transfer Function:** This function is responsible for the transferring of RRC messages from the gNB-CU to the gNB-DU, and vice versa.

F1-U (User Plane) Functions:

- **Transfer of User Data:** This function allows to transfer user data between gNB-CU and gNB-DU.
- **Flow Control Function:** This function allows to control the downlink user data transmission towards the gNB-DU. Several functionalities are introduced for improved performance on data transmission, like fast retransmission of PDCP PDUs lost due to radio link outage, discarding redundant PDUs, the retransmitted data indication, and the status report.

The following connected-mode mobility scenarios are supported in the case of CU-DU split:

- **Inter-gNB-DU Mobility:** The UE moves from one gNB-DU to another within the same gNB-CU.
- **Intra-gNB-DU inter-cell mobility:** The UE moves from one cell to another within the same gNB-DU, supported by UE Context Modification (gNB-CU initiated) procedure.
- **EN-DC Mobility with Inter-gNB-DU Mobility using MCG SRB:** The UE moves from one gNB-DU to another within the same gNB-CU when only MCG SRB is available during EN-DC operation.
- **EN-DC Mobility with Inter-gNB-DU Mobility using SCG SRB:** The UE moves from one gNB-DU to another when SCG SRB is available during EN-DC operation.

3.2 Separation of CP and UP with Higher Layer Split (HLS)

To optimize the location of different RAN functions according to different scenarios and performance requirements, the gNB-CU can be further separated into its CP and UP parts (the gNB-CU-CP and gNB-CU-UP, respectively).

The interface between CU-CP and CU-UP is called E1 (purely a control plane interface). The overall RAN architecture with CU-CP and CU-UP separation is shown in Figure 6.

The gNB-CU-CP hosts the RRC and the control plane part of the PDCP protocol; it also terminates the E1 interface connected with the gNB-CU-UP and the F1-C interface connected with the gNB-DU. The gNB-CU-CP hosts the user plane part of the PDCP protocol of the gNB-CU for an en-gNB, and the user plane part of the PDCP protocol and the SDAP protocol of the gNB-CU for a gNB. The gNB-CU-UP terminates the E1 interface connected with the gNB-CU-CP and the F1-U interface connected with the gNB-DU.

A gNB may consist of a gNB-CU-CP, multiple gNB-CU-UPs, and multiple gNB-DUs. The gNB-CU-CP is connected to the gNB-DU through the F1-C interface, and gNB-CU-UP is connected to the gNB-DU through the F1-U interface. One gNB-CU-UP is connected to only one gNB-CU-CP, but implementations allowing a gNB-CU-UP to connect to multiple gNB-CU-CPs e.g. for added resiliency, are not precluded. One gNB-DU can be connected to multiple gNB-CU-UPs under the control of the same gNB-CU-CP. One gNB-CU-UP can be connected to multiple DUs under the control of the same gNB-CU-CP.

The basic functions of the E1 interface include E1 interface management function and E1 bearer context management function.

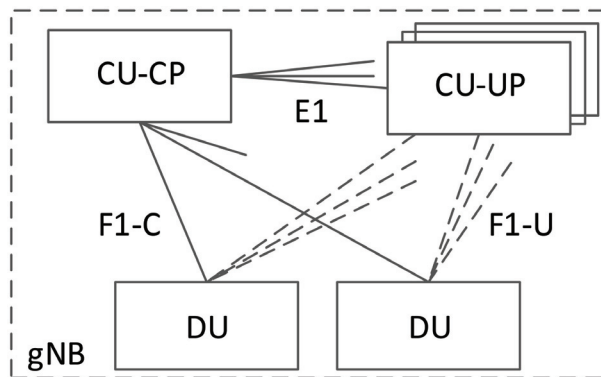


Figure 6 Overall RAN architecture with CU-CP and CU-UP separation.

4 NG-RAN Key Interfaces and Protocols

4.1 Xx Interface Family

NG-RAN nodes can be interconnected by means of the horizontal Xn and X2 interfaces, which are primarily used for three purposes: mobility (i.e. handover), multi-connectivity and SON (Self Optimized Networks). X2 was originally defined as the interface between two E-UTRAN nodes, later on extended to support EN-DC (i.e. as the interface between eNB and en-gNB) and will be further extended to support NR-DC. Xn follows similar design, interconnecting two gNBs.

Xn and X2 protocol stacks are similar. In the user plane, it relies on GTP-U running on top of UDP and IP. In the control plane, SCTP is used. This is illustrated in Figure 7:

The Xn-U/X2-U interface provides non-guaranteed delivery of user plane PDUs between two NG-RAN nodes to support dual/multi connectivity or mobility operation. Additionally, it supports the flow control function through Downlink Data Delivery Status procedure.

The Xn-C/X2-C interface uses Xn-AP/X2-AP protocols respectively for interface maintenance, mobility (handover, UE context retrieval, etc.) and dual/multi connectivity operation.

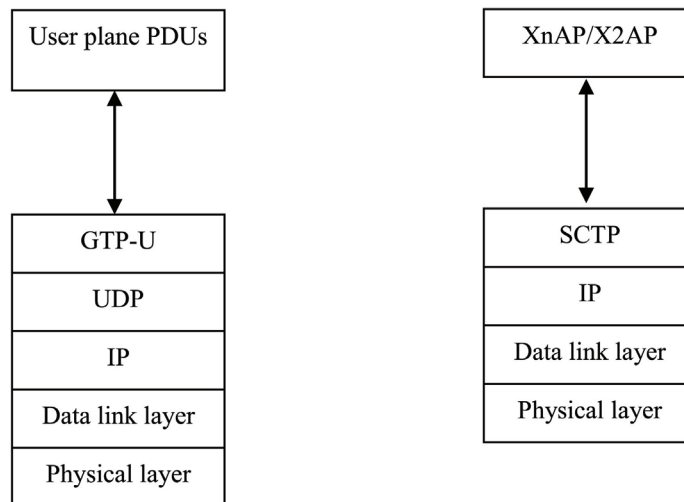


Figure 7 Xn and X2 protocol stacks.

4.2 NR Radio Interface Protocol

This section provides an overview of the architecture of the radio interface protocols that operate between NG-RAN and the UE, and then gives some details of the features of each protocol. The protocols will have some familiarity to those already knowledgeable of 4G LTE radio protocols and the description will identify some of the key differences and reasons for them.

The radio protocol architecture consists of a user plane, used for the transfer of the user data (IP packets) between the network and the UE, and a control plane that is used for control signalling between NG-RAN and the UE.

4.2.1 User plane

Figure 8 shows the user plane protocols stack within the UE and the gNB.

4.2.1.1 Service data adaptation protocol (SDAP)

The SDAP protocol is a notable difference in the user plane architecture compared to that of LTE, and it is introduced to support the new flow based QoS model of the 5G core network. With this new QoS model, the core network can configure different QoS requirements for different IP flows of a PDU session. The SDAP layer provides mapping of IP flows with different QoS requirements to radio bearers that are configured appropriately to deliver that required QoS. The mapping between IP flows and radio bearers may be configured and reconfigured by RRC signalling but it can also be changed more dynamically without the involvement of RRC signalling through a reflective mapping process.

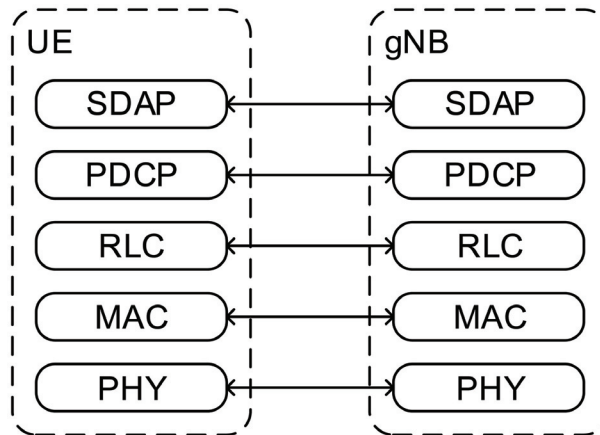


Figure 8 User plane protocol stack.

4.2.1.2 Packet data convergence protocol (PDCP)

The main functions of the PDCP protocol are to provide header compression and decompression through the use of RoHC (Robust Header Compression), security functions including ciphering/deciphering and integrity protection, duplication of transmitted PDCP PDUs, and reordering and duplicate detection of received PDCP PDUs. The most significant differences in NR PDCP compared to LTE are the introduction of the data duplication over different transmission paths in order to achieve extremely high reliability for URLLC (Ultra Reliable Low Latency) applications, and the introduction of integrity protection for user plane data.

4.2.1.3 Radio link control protocol (RLC)

NR RLC is very similar in functionality to LTE RLC, with the main functions being to provide segmentation, in order to match the transmitted PDU size to the available radio resources, and error correction through ARQ. One difference compared to LTE RLC is that it does not provide concatenation of RLC SDUs, with equivalent functionality now provided by the MAC layer, and does not provide reordering, with the protocol stack instead relying only on the reordering within PDCP.

4.2.1.4 Medium access control (MAC)

Similar to LTE MAC, the functionality provided includes multiplexing and demultiplexing of data from different radio bearers to the transport blocks that are carried by the physical layer, priority handling between data from different radio bearers, and error correction through Hybrid ARQ. A notable addition compared to LTE is that the MAC protocol carries control signalling used for the purpose of beam management within the physical layer.

4.2.2 Control plane

Figure 9 shows the control plane protocol stack. The Non Access Stratum (NAS) protocols terminate in the UE and the AMF of the 5G core network and are used for core network related functions such as registration, authentication, location updating and session management. The Radio Resource Control (RRC) protocol terminates in the UE and the 5G-RAN and is used for control and configuration of the radio related functions in the UE.

A significant difference in NR RRC compared to LTE RRC is the introduction of a 3-state model with the addition of the RRC INACTIVE state, as shown in the figure below. RRC Inactive provides a state with battery efficiency similar to RRC Idle but with a UE context remaining stored within

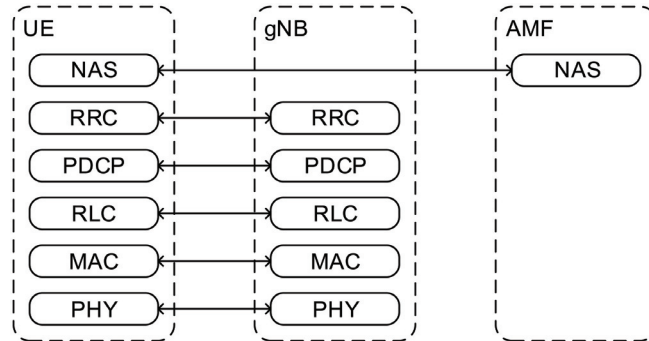


Figure 9 Control plane protocol stack.

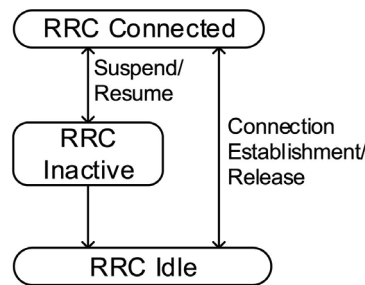


Figure 10 NR RRC state model.

the NG-RAN so that transitions to/from RRC Connected are faster and incur less signalling overhead. See Figure 10 above.

The other significant additions relative to LTE RRC are the support of an 'on demand' system information mechanism that enables the UE to request when specific system information is required instead of the NG-RAN consuming radio resources to provide frequent periodic system information broadcast, and the extension of the measurement reporting framework to support beam measurements for handover within a high frequency beam based deployment.

5 Summary

As described in this paper, the NG-RAN architecture builds on the success of the 4G LTE radio architecture, while introducing a number of key, revolutionary and forward looking concepts both on the overall architecture front as well as in protocols.

3GPP has taken several steps to specify interfaces and protocols that ease the migration of LTE-based cellular networks to 5G and NR. It is expected that these steps will help the uptake of NR and 5GC while making it easier to evolve networks in the most cost-efficient manner possible. Enhancements beyond phase-1 will address requirements and functions needed for industries beyond cellular mobile broadband: automated driving, industry automation, e-health services, etc. The 5G platform is promising to deliver the foundation for the next decade in the digital age.

Abbreviations

5GC	5G Core Network
AMF	Access and Mobility Management Function
CP	Control Plane
CU	Central Unit
DC	Dual Connectivity
DU	Distributed Unit
EN-DC	LTE-NR Dual Connectivity
E-UTRA	Evolved Universal Terrestrial Radio Access
MAC	Medium Access Control
NG-RAN	NG Radio Access Network
NGAP	NG Application Protocol
NR	NR Radio Access
PDCP	Packet Data Convergence Protocol
RRC	Radio Resource Control
RLC	Radio Link Control
SDAP	Service Data Adaptation Protocol
SMF	Session Management Function
UE	User Equipment
UP	User Plane
UPF	User Plane Function
URLLC	Ultra-Reliable and Low Latency Communications
XnAP	Xn Application Protocol
Xn-C	Xn-Control plane
Xn-U	Xn-User plane

Biographies



Balazs Bertenyi received an M.Sc. Degree in Computer Science and Telecommunications in 1998 at the Technical University of Budapest.

Balazs joined Nokia in 1998 and started to work on circuit switched mobile switching.

In 1999 he joined the research group on IMS (IP Multimedia Subsystem) and soon started to work in 3GPP standardization on IMS architecture.

- 2000–2003 Representative of Nokia in 3GPP SA2 (Architecture Working Group)
- 2003–2007 Head of delegation for Nokia in 3GPP SA2
- 2007–2011 Chairman of 3GPP SA2
- 2011–2015 Chairman of 3GPP TSG-SA (Technical Specification Group – Services and System Aspects)

In 2015 Balazs moved to the radio standardization group within Nokia to focus on 5G matters. He started attending TSG-RAN in 3GPP in 2015 covering 5G topics.

In 2017 Balazs was elected as Chairman of TSG-RAN for a 2-year term with a potential to be re-elected for one additional term.

Balazs has led various key projects for Nokia on Mobile Broadband architecture and standards strategy.



Richard Burbidge is a Senior Principal Engineer with Intel Corporation. With 25 years of experience in the wireless industry, he has been involved with the development of cellular standards since 1998, and has been a 3GPP delegate since its inception in 1999. Over this time he has contributed to 3 generations of cellular technology, firstly UMTS, then LTE, and now NR, 3GPP's 5G technology. He is currently serving as the chairman of the 3GPP TSG RAN 2 committee that is responsible for the specification of the layer 2 and 3 radio interface protocols for LTE and 5G/NR, having previously served as the vice chairman from 2005–2009. Richard received a Bachelor's degree in Electrical and Information Sciences from the University of Cambridge in 1993.



Gino Masini, MBA received his M.Sc. degree in Electronics Engineering from Politecnico di Milano and his MBA from SDA Bocconi School of Management in Milano, Italy. He started his career as a researcher in the Department of Electronics at Politecnico di Milano, working on microwave propagation and satellite telecommunications for the European Space Agency and the Italian Space Agency. He joined Ericsson in 1999, working at first on microwave antennas and network planning for microwave radio links, and later with MMIC design. Since 2009 he works with 4G and 5G radio network architecture, interfaces, and protocols. He is active in standardization since 2001: he attended ETSI, ITU, CEPT, and occasionally the Small Cell Forum.

He is active in 3GPP since 2009, and he is currently serving as 3GPP RAN WG3 Chairman. He has more than 30 patents granted, he authored or co-authored more than a dozen scientific publications, and he holds a “Six Sigma” certification.



Sasha Sirotkin is a senior wireless architect in the Next Generation and Standards group at Intel Corporation. Sasha has over 15 years of experience in wireless communications, in product development, system architecture, standardization and research. Sasha is currently serving as the vice chairman of 3GPP TSG RAN 3 committee that is responsible Radio Access Network architecture. Sasha holds B.Sc. degrees in Computer Science and Physics and M.Sc. degree in Applied Statistics and Astrophysics from Tel-Aviv University, Israel.



Yin Gao received the Master degree in Circuit and System from Xidian University, Xi’an, China, in 2005. She attended the Xidian University where she received his B.Sc. in Electronic Engineering in 2002. Since 2005 she has been with research center of ZTE and is engaged in the study of 3G/4G/5G technology, especially in both LTE and 5G RAN architecture and higher layer signaling procedures. From August, 2017, she was elected as 3GPP RAN3 Vice Chairman.

