|  |  |
| --- | --- |
| 3GPP TR 38.835 V0.2.0 (2022-09) | |
| Technical Report | |
| 3rd Generation Partnership Project;  Technical Specification Group Radio Access Network;  NR; Study on XR enhancements for NR  (Release 18) | |
|  | |
|  |  |
|  | |
| The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP. The present document has not been subject to any approval process by the 3GPPOrganizational Partners and shall not be implemented. This Specification is provided for future development work within 3GPPonly. The Organizational Partners accept no liability for any use of this Specification. Specifications and Reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices. | |

|  |
| --- |
|  |
| ***3GPP***  Postal address  3GPP support office address  650 Route des Lucioles - Sophia Antipolis  Valbonne - FRANCE  Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16  Internet  http://www.3gpp.org |
| ***Copyright Notification***  No part may be reproduced except as authorized by written permission. The copyright and the foregoing restriction extend to reproduction in all media.  © 2022, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC).  All rights reserved.  UMTS™ is a Trade Mark of ETSI registered for the benefit of its members  3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners LTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners  GSM® and the GSM logo are registered and owned by the GSM Association |

Contents

Foreword 4

1 Scope 6

2 References 6

3 Definitions of terms, symbols and abbreviations 7

3.1 Terms 7

3.2 Symbols 7

4 Introduction to Extended Reality 7

5 XR Enhancements for NR 7

5.1 XR Awareness 7

5.2 Power Saving Techniques 7

5.2.1 Physical Layer Enhancements 7

5.2.2 Layer 2 Enhancements 7

5.3 Capacity Improvements Techniques 7

5.3.1 Physical Layer Enhancements 7

5.3.2 Layer 2 Enhancements 7

6 Conclusions 7

Annex A: Evaluation Methodology 8

Annex B: Evaluation Studies 9

Annex Z (informative): Change history 10

# Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

In the present document, modal verbs have the following meanings:

**shall** indicates a mandatory requirement to do something

**shall not** indicates an interdiction (prohibition) to do something

The constructions "shall" and "shall not" are confined to the context of normative provisions, and do not appear in Technical Reports.

The constructions "must" and "must not" are not used as substitutes for "shall" and "shall not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

**should** indicates a recommendation to do something

**should not** indicates a recommendation not to do something

**may** indicates permission to do something

**need not** indicates permission not to do something

The construction "may not" is ambiguous and is not used in normative elements. The unambiguous constructions "might not" or "shall not" are used instead, depending upon the meaning intended.

**can** indicates that something is possible

**cannot** indicates that something is impossible

The constructions "can" and "cannot" are not substitutes for "may" and "need not".

**will** indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**will not** indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**might** indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

**might not** indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

**is** (or any other verb in the indicative mood) indicates a statement of fact

**is not** (or any other negative verb in the indicative mood) indicates a statement of fact

The constructions "is" and "is not" do not indicate requirements.

# 1 Scope

The present document is intended to capture the output of the study item on XR Enhancements for NR [9], which aims at investigating power saving and capacity enhancements techniques tailored for XR services, as well as means to provide XR-awareness in RAN.

This study follows a series of earlier studies conducted in 3GPP by SA1 [2], SA4 [5] [6] [7] and RAN1 [8]. It is complemented by work in SA2 [11], SA4 [12] and SA6 [4].

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non‑specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR [21.905](http://www.3gpp.org/ftp/Specs/html-info/21905.htm): "Vocabulary for 3GPP Specifications".

[2] 3GPP TR [22.842](http://www.3gpp.org/ftp/Specs/html-info/22842.htm): "Study on Network Controlled Interactive Service (NCIS) in the 5G System (5GS)", SA1, Release 17

[3] 3GPP TR [23.748](http://www.3gpp.org/ftp/Specs/html-info/23748.htm): "Study on enhancement of support for Edge Computing in 5G Core network(5GC), SA2, Release 17"

[4] 3GPP TR [23.758](http://www.3gpp.org/ftp/Specs/html-info/23758.htm): "Study on application architecture for enabling Edge Applications", SA6, Release 17

[5] 3GPP TR [26.918](http://www.3gpp.org/ftp/Specs/html-info/26918.htm): "Virtual Reality (VR) media services over 3GPP", SA4, Release 14-15.

[6] 3GPP TR [26.926](http://www.3gpp.org/ftp/Specs/html-info/26926.htm): "Traffic Models and Quality Evaluation Methods for Media and XR Services in 5G Systems", SA4, Release 17-18

[7] 3GPP TR [26.928](http://www.3gpp.org/ftp/Specs/html-info/26928.htm): "Extended Reality (XR) in 5G", SA4, Release 16.

[8] 3GPP TR [38.838](http://www.3gpp.org/ftp/Specs/html-info/38838.htm): "Study on XR (Extended Reality) evaluations for NR", RAN1, Release 17.

[9] 3GPP TS [23.700-60](https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=4007): "Study on architecture enhancement for XR and media services", SA2, Release 18.

[10] [RP-220285](http://3gpp.org/ftp/tsg_ran/TSG_RAN/TSGR_95e/Docs/RP-220285.zip): "Study on XR Enhancements for NR", RAN2, Release 18

[11] [SP-210043](http://3gpp.org/ftp/tsg_sa/TSG_SA/TSGs_91E_Electronic/Docs/SP-210043.zip): "Feasibility Study on Typical Traffic Characteristics for XR Services and other Media, SA4, Release 18"

[12] [SP-211166](http://3gpp.org/ftp/tsg_sa/TSG_SA/Workshops/2021-12-09_Rel-18_Prioritization_WorkShop/Docs/SP-211166.zip): "New SID on Study on architecture enhancement for XR and media services", SA2, Release 18

Note: hyperlinks, responsible groups and corresponding releases are used for convenience, they can be removed once the TR is presented for approval.

# 3 Definitions of terms, symbols and abbreviations

## 3.1 Terms

For the purposes of the present document, the terms given in TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in TR 21.905 [1].

**Field of view:** the angle of visible field expressed in degrees measured from the focal point.

## 3.2 Abbreviations

For the purposes of the present document, the abbreviations given in TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in TR 21.905 [1].

3DoF Three Degrees of Freedom

6DoF Six Degrees of freedom

AR Augmented Reality

DASH Dynamic Adaptive Streaming over HTTP

FoV Field of view

FPS Frames Per Second

GBR Guaranteed Bit Rate

GFBR Guaranteed Flow Bit Rate

HEVC High-Efficiency Video Coding

HMD Head-Mounted Display

HUD Heads-Up Display

PDB Packet Delay Budget

PDU Packet Data Unit

PER Packet Error Rate

QCI QoS Class Identifier

QFI QoS Flow ID

QoE Quality of EXperience

QoS Quality of Service

VR Virtual Reality

XR Extended reality

# 4 Introduction to Extended Reality

## 4.1 Reality Types

Extended Reality (**XR**) refers to all real-and-virtual combined environments and human-machine interactions generated by computer technology and wearables. XR is an umbrella term for different types of realities (see TR 26.918 [5] and TR 26.928 [7]):

- Virtual reality (**VR**) is a rendered version of a delivered visual and audio scene. The rendering is designed to mimic the visual and audio sensory stimuli of the real world as naturally as possible to an observer or user as they move within the limits defined by the application. Virtual reality usually, but not necessarily, requires a user to wear a head mounted display (HMD), to completely replace the user's field of view with a simulated visual component, and to wear headphones, to provide the user with the accompanying audio. Some form of head and motion tracking of the user in VR is usually also necessary to allow the simulated visual and audio components to be updated in order to ensure that, from the user's perspective, items and sound sources remain consistent with the user's movements.

- Augmented reality (**AR**) is when a user is provided with additional information or artificially generated items or content overlaid upon their current environment. Such additional information or content will usually be visual and/or audible and their observation of their current environment may be direct, with no intermediate sensing, processing and rendering, or indirect, where their perception of their environment is relayed via sensors and may be enhanced or processed.

- Mixed reality (**MR**) is an advanced form of AR where some virtual elements are inserted into the physical scene with the intent to provide the illusion that these elements are part of the real scene.

Other terms used in the context of XR are Immersion as the sense of being surrounded by the virtual environment as well as Presence providing the feeling of being physically and spatially located in the virtual environment. The sense of presence provides significant minimum performance requirements for different technologies such as tracking, latency, persistency, resolution and optics.

This document uses the acronym XR throughout to refer to equipment, applications and functions used for VR, AR and MR. Examples include, but are not limited to HMDs for VR, optical see-through glasses and camera see-through HMDs for AR and MR and mobile devices with positional tracking and camera. They all offer some degree of spatial tracking and the spatial tracking results in an interaction to view some form of virtual content.

## 4.2 Human Perception and Tracking

For providing XR experiences that make the user feel *immersed* and *present*, several relevant quality of experience factors have been collected (see TR 23.748 [7]). Presence is the feeling of being physically and spatially located in an environment. Presence is divided into 2 types: Cognitive Presence and Perceptive Presence. Cognitive Presence is the presence of one's mind. It can be achieved by watching a compelling film or reading an engaging book. Cognitive Presence is important for an immersive experience of any kind. Perceptive Presence is the presence of one's senses. To accomplish perceptive presence, one's senses, sights, sound, touch and smell, have to be tricked. To create perceptive presence, the XR Device has to fool the user's senses, most notably the audio-visual system. XR Devices achieve this through positional tracking based on the movement. The goal of the system is to maintain your sense of presence and avoid breaking it. Perceptive Presence is the objective to be achieved by XR applications.

The Human field of view (**FoV**) is defined as the area of vision at a given moment (with a fixed head). It is the angle of visible field expressed in degrees measured from the focal point. The monocular FoV is the angle of the visible field of one eye whereas the binocular FoV is the combination of the two eyes fields (see TR 26.918 [5]). The binocular horizontal FoV is around 200-220°, while the vertical one around 135°. The central vision is also called the comfort zone where sensibility to details is the most important. Although less sensitive to definition, the peripheral vision is more receptive to movements. Figure 4.2-1 and 4.2-2 below summarize the viewing angles.

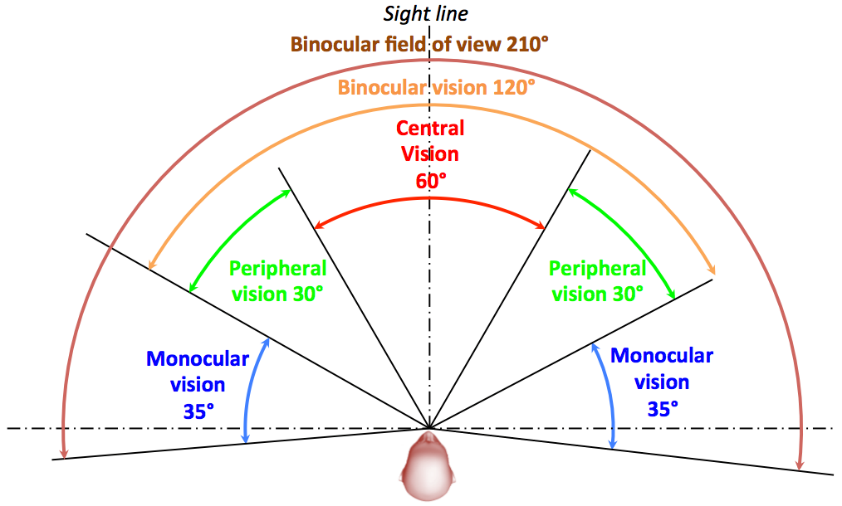


Figure 4.2-1: Horizontal FoV

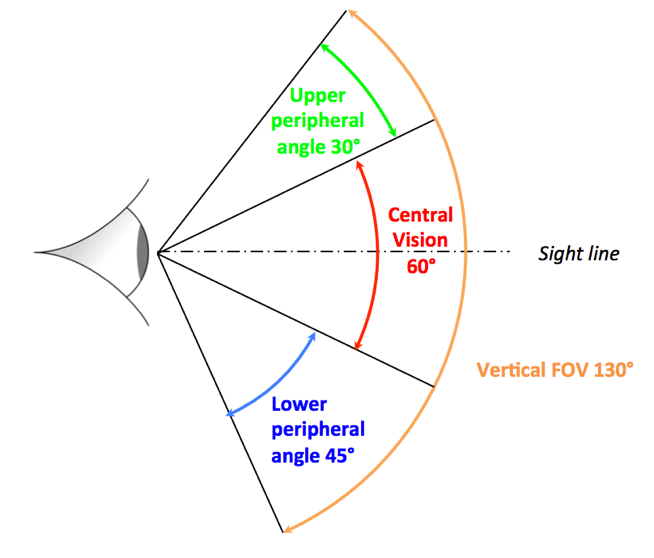


Figure 4.2-2: Vertical FoV

In XR, actions and interactions involve movements and gestures. Thereby, the Degrees of Freedom (**DoF**) describe the number of independent parameters used to define movement in the 3D space (see TR 26.928 [7]):

- 3DoF: three rotational and un-limited movements around the X, Y and Z axes (respectively pitch, yaw and roll). A typical use case is a user sitting in a chair looking at 3D 360 VR content on an HMD.



Figure 4.2-3: 3DoF

- 6DoF: 3DoF with full translational movements along X, Y and Z axes. Beyond the 3DoF experience, it adds (i) moving up and down (elevating/heaving); (ii) moving left and right (strafing/swaying); and (iii) moving forward and backward (walking/surging). A typical use case is a user freely walking through 3D 360 VR content (physically or via dedicated user input means) displayed on an HMD.



Figure 4.2-4: 6DoF

An **XR View** describes a single view into an XR scene for a given time. Each view corresponds to a display or portion of a display used by an XR device to present the portion of the scene to the user.

An **XR Viewport** describes a viewport, or a rectangular region, of a graphics surface. The XR viewport corresponds to the projection of the XR View onto a target display. An XR viewport is predominantly defined by the width and height of the rectangular dimensions of the viewport.

An **XR Pose** describes a *position* and *orientation* in space relative to an XR Space. An essential element of XR is the spatial tracking of the viewer pose.

## 4.3 Capture, Encoding and Delivery

### 4.3.1 Video

XR content may be represented in different formats, e.g. panoramas or spheres depending on the capabilities of the capture systems. Since modern video coding standards are not designed to handle spherical content. projection is used for conversion of a spherical (or 360°) video into a two-dimensional rectangular video before the encoding stage. After projection, the obtained two-dimensional rectangular image can be partitioned into regions (e.g. front, right, left, back, top, bottom) that can be rearranged to generate "packed" frames to increase coding efficiency or viewport dependent stream arrangement.

There are mainly three approaches that can be considered for 360 video delivery (see TR 26.918 [5]):

- Single stream approach: the single stream approach consists in providing the full 360 video and showing the interesting part only. Solutions that lie within this group have the drawback that either they may not be scalable or they may impose a big challenge in terms of required network resources (high bitrate of high resolution video) and required processing at the client side (decode a very high resolution video).

- Multi-stream approach: the multi-stream approach consists of encoding several streams, each of them emphasizing a given viewport and making them available for the receiver, so that the receiver decides which stream is delivered at each time instance.

- Tiled stream approach: the tiled stream approach consists in emphasizing the current user viewport through transmitting non-viewport samples with decreased resolution. The tiles can be provided as one common bitstream (using motion-constrained HEVC tiles) or as separate video streams.

A panoramic or 360 video can be delivered in unicast, multicast or broadcast mode via DASH (see TR 26.918 [5]). With DASH, the client requests appropriate segments depending on the viewport position, available network throughput, device capabilities and service requirements - e.g. for multi-stream approach, the DASH client requests the stream (representation) that matches best to the expected viewport position (subject to network latency and user movement).

### 4.3.2 Audio

For Audio, we can distinguish channel-based and object-based representations (see TR 26.918 [5]):

- Channel-based representation using multiple microphones to capture sounds from different directions and post-processing techniques are well known in the industry, as they have been the standard for decades.

- Object-based representations represent a complex auditory scene as a collection of single audio elements, each comprising an audio waveform and a set of associated parameters or metadata. The metadata embody the artistic intent by specifying the transformation of each of the audio elements to playback by the final reproduction system. Sound objects generally use monophonic audio tracks that have been recorded or synthesized through a process of sound design. These sound elements can be further manipulated, so as to be positioned in a horizontal plane around the listener, or in full three-dimensional space using positional metadata.

## 4.4 XR Engines and Rendering

XR engines provide a middleware that abstracts hardware and software functionalities for developers of XR applications (see TR 26.928 [7]). Typical components include a rendering engine for graphics, an audio engine for sound, and a physics engine for emulating the laws of physics. In the remainder of this Technical Report, the term *XR engine* is used to provide any type of typical XR functionalities as mentioned above.

The processing of an XR engine is not exclusively carried out in the device GPU. In power and resource constrained devices, it can be assisted or split across the network through edge computing (see TR 22.842 [2]): the UE sends the sensor data in uplink direction to the cloud side in a real time manner and when the cloud side receives the sensor data, it performs rendering computing and produces the multimedia data and then sends back to the user devices for display. This is where NR can play an essential role.

## 4.5 Requirements

### 4.5.1 Video

According to TR 26.918 [5], the **latency** of action of the angular or rotational vestibulo-ocular reflex is known to be of the order of 10 ms or in a range from 7-15 milliseconds and it seems reasonable that this should represent a performance goal for XR systems. This results in a motion-to-photon latency of less than 20 milliseconds, with 10ms being given as a goal.

Regarding the **bit rates**, between 10 and 200Mbps can be expected for XR depending on frame rate, resolution and codec efficiency (see TR 26.926 [6] and 26.928 [7]).

### 4.5.2 Audio

According to TR 26.918 [5], due to the relatively slower speed of sound compared to that of light, it is natural that users are more accustomed to, and therefore tolerant of, sound being relatively delayed with respect to the video component than sound being relatively in advance of the video component. Recent studies have led to recommendations of an accuracy of between 15 ms (audio delayed) and 5 ms (audio advanced) for the **synchronization**, with recommended absolute limits of 60 ms (audio delayed) and 40 ms (audio advanced) for broadcast video.

### 4.5.3 Pose Information

To maintain a reliable registration of the virtual world with the real world, as well as to ensure accurate tracking of the XR Viewer pose, XR applications require highly accurate, low-latency tracking of the device at about 1kHz sampling frequency. The size of a XR Viewer Pose associated to time, typically results in packets of size in the range of 30-100 bytes, such that the generated data is around several hundred kbit/s if delivered over the network (see TR 23.748 [7]).

Pose information has to be delivered with ultra-high reliability, therefore, similar performance as URLLC is expected i.e. packet loss rate should be lower than 10E-4 for uplink sensor data – see TR 22.842 [2].

Editor's Note: the relationship between pose information sampling frequency and video frame rate should be clarified by SA4.

# 5 XR Enhancements for NR

## 5.1 XR Awareness

In both uplink and downlink, XR-Awareness relies at least on the notion of PDU set (see TR 23.700-600 [9]): a group of packets which have inherent dependency on each other in the media layer, and thus decoded/handled as a whole, for instance a video frame that may only be decoded in case all packets carrying the video frame are successfully delivered.

Editor's Note: once a definition is adopted by SA2.

## 5.2 Power Saving Techniques

### 5.2.1 Physical Layer Enhancements

### 5.2.2 Layer 2 Enhancements

## 5.3 Capacity Improvements Techniques

### 5.3.1 Physical Layer Enhancements

### 5.3.2 Layer 2 Enhancements

# 6 Conclusions

Note: this clause will capture the conclusions of the SI.

Annex A:  
Evaluation Methodology

Note: this Annex will capture the deployment scenarios, traffic models and KPI.

Annex B:  
Evaluation Studies

Note: this Annex will capture the evaluation studies.

Annex C (informative):  
RAN2 Agreements

# C.1 RAN2#119-e

Agreements from RAN2#119-e meeting:

- RAN2 does not intend to ask RAN1 to change their simulation assumptions;

- RAN2 should take SA2/SA4 work into account.

- RAN2 assumes that PDU Set based parameters and PDU Set related information may be used for better support of XR services. RAN2 can consider both UL and DL directions.

- RAN2 will study PDU Set based parameters and PDU Set related information handling in Network and UE.

- RAN2 to adopt the current SA2 definition of PDU Set as an application media unit as working assumption, subjected to further guidance from SA2 and SA4.

- XR awareness discussion in RAN2 should consider PDU set characteristics and how to use the information available on those (for UL and/or DL). Can also consider how to handle data bursts.

- RAN2 can study e.g. periodicity, arrival time, jitter and frame-size variations for XR awareness to enable power savings and capacity enhancements. Can study also how often such parameters change (i.e. how dynamic they are).

- RAN2 can consider how PDU sets can be mapped to DRBs (FFS if SA2 discussion on PDU set mapping to QoS (sub-)flows impacts this).

- RAN2 to focus on the following issues for power saving, as well necessary parameters XR-awareness to support such enhancements, i.e.:

- DRX enhancements to address the issues of DRX cycle mismatch and jitter;

- Identify necessary parameters from CN for XR-awareness for power saving.

- Enhancements to Rel-17 PDCCH adaptation can be discussed based on RAN1 feedback, if they have any RAN2 impact.

- RAN2-specific aspects can be studied based on contributions (e.g. multiple XR traffic flows with different periodicities, SFN wrap-around, RAN2-specific CDRX aspects, …).

- As starting point, RAN2 can further discuss the solutions in TR 38.838 that can impact on L2 operation (e.g., BSR, LCP, assistance information for scheduling, packet discarding, prioritization) for XR-specific capacity improvement. RAN2-specific solutions are not precluded (even if RAN1 hasn’t discussed them before).

- Enhancement to SPS/CG should be justified for XR scheduling and should be evaluated against dynamic grant (DG) scheduling which should be considered as baseline. Should justify why enhancements are needed.

- RAN2 considers SPS enhancements may not be needed in Rel-18 XR since PDCCH capacity is not assumed to be a problem for XR. FFS if SPS has some power consumption benefits.

Annex Z (informative):  
Change history

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Change history | | | | | | | |
| Date | Meeting | TDoc | CR | Rev | Cat | Subject/Comment | New version |
| 2022-04 | RAN1#109 | R1-2204673 |  |  |  | Initial Skeleton | 0.0.1 |
| 2022-08 | RAN2#119 | R2-2207373 |  |  |  | Initial Skeleton | 0.0.1 |
| 2022-08 | RAN2#119 | R2-2207374 |  |  |  | Updated Sructure | 0.0.2 |
| 2022-08 | RAN2#119 | R2-2208748 |  |  |  | First Endorsed Baseline | 0.1.0 |
| 2022-08 | RAN2#119 | R2-2208749 |  |  |  | Table of Content updated | 0.1.1 |
| 2022-09 | RAN2#119 | R2-220xxxx |  |  |  | Overview and first RAN2 agreements included | 0.2.0 |