3GPP TSG-RAN WG4 Meeting #109 R4-2318489

Chicago, US, Nov13 –17, 2023

Agenda Item: 8.21.1

Source: CAICT, Qualcomm, Ericsson

Title: Proposed update for TR 38.843 with RAN4 part

Document for: Approval

# 1 Introduction

In this contribution, we provide update for TR 38.843 with RAN4 part according to the agreements in RAN4#106bis, RAN4#107, RAN4#108 and RAN4#108bis.

# 2 Text Proposal

The text proposal is mainly based around the WF agreed in previous meetings. Some proposed wording beyond we have agreed are also listed with brackets. The aim of this additional wording is to mitigate vagueness and improve the clarity of the agreements. Before we agree the proposed text as a whole, it is also proposed to spend some time to discuss and confirm the changes in bracket.

It should also be noted that section 7.4.1.2 and 7.4.1.3 are still open and further update is required after consensus to complete the TR.

Further updates can be made depending on agreements at RAN4#109

------------------------------------------- Change --------------------------------------------------------

Nokia: it is good to have testing goal in this section.

## 7.4 Interoperability and testability aspects

### 7.4.1 xxxx

In this section, the study of requirements and testing frameworks to validate AI/ML based performance enhancements and ensuring that UE and gNB with AI/ML meet or exceed the existing minimum requirements, if applicable, are documented.

The need and implications for AI/ML processing capabilities definition is considered.

### 7.4.2 Common framework

#### 7.4.2.1 General

The general requirements and testing frameworks for AI/ML based performance enhancements mainly focus on

* how to define requirements and tests for inference
* evaluate feasibility and necessity of requirements/tests for LCM
* requirements for data collection (in particular for training) could/need be defined

Requirements/tests for training will not be studied unless training procedures are defined.The design of test should ensure performance is guaranteed and avoid that a UE can pass the test but perform poorly in the field.

Apple: for LCM part, not sure if we will have link level test only for LCM. Does it mean that we will introduce the LCM test.

Vivo: we have agreed that for LCM framework, we reuse the delay test.

Agreement: capture the test goals in new section 7.4.1.1.

#### 7.4.2.2 Principles on the definition of requirements

For the definition of AI/ML requirements, the following cases related to legacy performance should be considered

* For the cases with the existing legacy performance
	+ Take the legacy performance as baseline for existing use cases/procedures/functionalities /measurements that are to be enhanced by AI/ML based methods
		- Further study may be needed on what is baseline performance in conditions different to the requirement condition but within the expected range of operation.
	+ New or enhanced performance requirements/tests could be considered for existing use cases/procedures/functionalities/measurements that are to be enhanced by AI/ML based methods
* For the cases without the existing legacy performance
	+ New performance requirements/tests could be considered for the use cases/procedures/functionalities/measurements that are carried out or are to be enhanced by AI/ML based methods

The following procedure can be considered for defining core requirements

* Performance monitoring procedure, including performance evaluation and decision-making procedure for AI/ML functionalities/models
* Functionality/Model management procedure, including functionality/model selection/activation/deactivation, and functionality/model switching/fallback/transfer/delivery/update
* Latency/interruption requirement for above procedures

The following LCM related requirements can be considered:

* Model/Functionality select/switch/activate/deactivate/fallback
* Model/Functionality monitoring
* On whether requirements for data collection (in particular for training) could/need be defined:
	+ Data collection requirements would only be defined if data collection procedure is defined in 3GPP specifications.
* On requirements for model transfer/update:
	+ Requirements would only be defined if model transfer/update would be defined in 3GPP specifications.

The legacy framework for RRC/MAC-CE/DCI based core requirements (e.g., define delay requirements based on multiple delay components) can be used as the baseline for LCM procedures if the LCM related requirements are agreed to be introduced. If new procedures which legacy framework is not applicable to are introduced, additional core requirement framework can be discussed.

LCM related tests should consider how the framework can address the possibility of updates/activation/deactivation /switching to the functionalities/models after the deployment of the devices in the field.

Qualcomm: the test goals are supposed to go into the section 7.4.1.1 or framework section 7.4.1. We prefer 7.4.1.

Huawei: we have concern on the second paragraph.

#### 7.4.2.3 Reference block diagrams for testing

Reference block diagrams provide test modules/functionalities of TE/DUT and testing framework for different use cases. Both reference block diagrams for 1-sided model and 2-sided model are studied.

Qualcomm: we will update 7.4.1.2

##### 7.4.2.3.1 Reference block diagram for 1-sided model

[Reference block diagram for 1-sided model]

##### 7.4.2.3.2 Reference block diagram for 2-sided model

[Reference block diagrams for 2-sided model]

#### 7.4.2.3 Test encoder/decoder for 2-sided model

In order to determine the test encoder/decoder, the following issues are considered:

* Common assumptions for proposals of the test decoder / encoder (and the paired encoder/ decoder) for tester
* The need for and potential definition and derivation procedure of intermediate KPI for decoder evaluation and selection
* Data collection/generation for decoder evaluation, and the common assumptions/environment needed for data collection/generation
* How to minimize the impact of possible variations/differences in the test decoder/ test encoder design/implementation on UE/ gNB performance verification
* The impact of test decoder/ encoder for testing complexity to UE/gNB performance verification, and the advantage/disadvantage analysis of high/low complexity decoders.

The test decoder/encoder design should take into account complexity limitations based on e.g., feasibility of TE implementation and complexity levels considered feasible by network vendors/UE vendors for decoder/encoder deployment.

The choice of test decoder/encoder should aim as much as possible to avoid limiting the implementation choices, including e.g. complexity, back-bone model etc, of UE/gNB encoders/decoders operating in the field (this principle may not be fully achievable in practice).

Specification on the test may include some high-level parameters for the test decoder/encoder (e.g. parameters related to processing complexity, model structure, etc).

Using the above principles as reference, the considered options of test decoder are listed below

* Option 1: DUT provides the decoder
* Option 2: Infra vendor provides the decoder
* Option 3: Full decoder specification in standard
* Option 4: TE vendor provides the decoder
	+ TE vendor should be able to develop the decoder based on the specifications
	+ [Test repeatability should be ensured (variation among TE vendor implementations should be bound)
	+ Other vendors should also be able to develop such a decoder and which can deliver similar performance]

NOTE: At the current stage the framework in this session applies to CSI compression case.

Further clarifications of the four options are included in the following table:

Table for test decoder options:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Option 1** | **Option 2** | **Option 3** | **Option 4** |
| Clarification of options |
| Source of the test decoder  |  DUT vendor | Decoder vendor (infra vendor in case of testing UEs)  |  RAN4 specifications |  TE vendor, decoder developed based on RAN4 specifications |
| Source of decoder training data  | Up to DUT vendor (no need to be specified) | Up to decoder implementer (infra vendor) FFS whether coordination with encoder vendor is required | Not needed, decoder fully specified (used as part of the RAN4 procedure to specify the decoder) |  |
| DUT vendor knowledge of the test decoder | Full knowledge | No or partial or enough or full knowledge based on alignment with infra vendors or specifications  | Full knowledge based on the specifications | Partial knowledge – based on the RAN4 specification |

Apple: we are fine to capture what we discussed in the meeting. I am not sure if we have consensus that all the test goals can be achieved considering the complexity…We also need include the other options. Some wording has been removed from Option 4. We should make the texts aligned with the agreements in the meeting.

Huawei: we have concern to include encoder and decoder part, which is only related to CSI and not for other use cases.

Ericsson: To Apple, we have some edition in the ad hoc. We have no consensus that any option can achieve the test goals.

Apple: we suggest to add the paragraph to provide the summary. We did not want to give people the impression that all the goals are achievable.

Agreement:

#### 7.4.2.4 Data collection/generation for testing

Different generating methods of test dataset can be used for different tests. The following candidate methods are to be considered:

* Dataset based on TR 38.901, e.g. UMa channel, UMi channel, CDL channel, “legacy approach”, etc.
	+ “Legacy approach” refers legacy test in which a channel model is used
* Field dataset (data collected directly from field measurements)
* TE generates dataset for test based on assumptions/parameters defined by RAN4 (e.g. by defining some rules/function to generate data)
* Other methods are not precluded

Apple: this applies for single-side or two-sided or both? Training test. UE has to use label from TE.

Qualcomm: We agree with Apple some clarification is needed in WI phase. These texts are copied-pasted from the previous agreements. We can discuss separately the open issues.

OPPO: this is about test data rather than training data.

Qualcomm: We do think there is any contradiction. You can use something what you want rather than standardized ones.

Apple: for the encoder part, UE passes the test based on training data for infra vendor A, but it cannot pass the test for infra vendor B.

#### 7.4.2.5 Data collection/generation for training

#### 7.4.2.6 Generalization/scalability aspects

The necessity and feasibility of defining requirements or test to verify the generalization of AI/ML is studied.

The goals of generalization test are to verify whether the performance gain/minimum level of performance of AI/ML functionality/model can be achieved/maintain under the identified scenarios and/or configurations, while the performance won’t be significantly degraded in other scenarios and/or configurations. The following aspects should be considered for generalization/scalability related testing:

* details about the scenarios and/or configurations for test and the corresponding AI/ML models/functionality
* what the minimum level performance for each identified scenario and/or configuration is
* what the significant degradation for other scenarios and/or configurations is

It should also be considered that generalization and/or scalability related requirements for different scenarios/ configurations can be implicitly handled in the test case definition.

As for the handling of generalization tests, the following option is considered as baseline:

Signaling based LCM procedures and performance monitoring are considered in dedicated test cases and are excluded in tests verifying generalization. RAN4 may define multiple tests with different conditions. In each of the test, TE configures the same specified UE configuration, and therefore the same specified UE configuration is tested under different conditions to verify its generalizability. (environment differs in each test but not changing dynamically during the test)

* Specified UE configuration includes functionality and/or model ID if defined

#### 7.4.2.7 AI/ML processing capability

The practical processing capability and implementation complexity for device under test should be assumed when specifying RAN4 requirements.

* The UE capability may be needed to handle different complexity for one side and two-side models.
* The complexity of UE should also be studied when making assumption on BS side model, and vice versa.

### 7.4.3 CSI feedback enhancement

Both time domain CSI prediction and spatial-frequency domain CSI compression are considered.

PMI reporting framework (follow PMI vs. random PMI test, use of γ as criteria, etc.) is taken as starting point for CSI related tests. Other metrics/framework is not precluded.

For metrics for CSI requirements/tests, the following test metrics are identified:

* Option 1: Throughput/relative throughput
* Option 2: SGCS, NMSE
* Option 3: CSI prediction accuracy

Option 1 should be used as baseline. For option 3, further discuss is needed on the feasibility to define the CSI prediction accuracy in WI. For metrics for CSI monitoring, further discussion is needed in WI.

### 7.4.4 Beam management

Both spatial-domain DL beam prediction and temporal DL beam prediction are considered.

For metrics for beam management requirements/tests, the following test metrics are identified and could be considered

* Option 1: RSRP accuracy
* Option 2: Beam prediction accuracy
	+ Top-1 (%) : the percentage of “the Top-1 strongest beam is Top-1 predicted beam”
	+ Top-K/1 (%) : the percentage of “the Top-1 strongest beam is one of the Top-K predicted beams”
	+ Top-1/K (%) : the percentage of “the Top-1 predicted beam is one of the Top-K strongest beams”
* Option 3: The successful rate for the correct prediction which is considered as maximum RSRP among top-K predicted beams is larger than the RSRP of the strongest beam – x dB,
	+ Related measurement accuracy can be considered to determine x
* Option 4: combinations of above options

The overhead/latency reduction should be considered for the requirements as the side condition.

### 7.4.5 Positioning accuracy enhancements

Both direct AI/ML positioning and AI/ML assisted positioning are considered.

For metrics for positioning requirements/tests, the candidate options include

* Option 1: positioning accuracy: Ground truth vs. reported
	+ only option available for direct positioning
* Option 2: CIR/PDP, channel estimation accuracy
* Option 3: ToA, RSTD and RSRP, and RSRPP
* Option 4: others (e.g., intermediate KPIs, LoS/NLoS)/combinations of the above

The feasibility and testability of different options should be further justified in WI.