3GPP TSG-RAN WG4 Meeting #109 R4-2315351

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.

[The final input to TR will clean up all FFS, either by deleting or by adding notes.]

# 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

## 7.4 Interoperability and testability aspects

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.

### 7.4.1 Common framework

The study of general requirements and testing frameworks for AI/ML based performance enhancements mainly focuses on

* study how to define requirements and tests for inference
* evaluate feasibility of requirements/tests for LCM

Requirements/tests for training are not studied, since the training procedure is not defined. Dataset to be used for DUT model training is left to implementation.

#### 7.4.1.1 Principles on the definition of requirements

The high-level test design principle for all tests to be considered is to avoid that a DUT can easily pass the test but perform poorly in the field.

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
  + [FFS how to define “legacy performance” (whether on meeting/exceeding existing RAN4 requirements, or a wider criterion taking into account generalization]
* 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 legacy framework for RRC/MAC-CE/DCI based core requirements (e.g., define delay requirements based on multiple delay components) should be used as the baseline for LCM procedures. If legacy framework for future specified procedures is not applicable, additional core requirement framework may be discussed.

The following potential aspectes are identified if RAN4 studies 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 potential aspects are identified if RAN4 studies LCM related requirements

* Model/Functionality select/switch/activate/deactivate/fallback
* Model/Functionality monitoring
* [FFS if requirements for data collection (in particular for training) could/need be defined]
* [FFS if requirements for transfer/delivery/update]

#### 7.4.1.2 Test Dataset

Different generating methods of test dataset are identified with potential down selection:

* 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

#### 3

The intention to consider generalization is 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 is not significantly degraded in other scenarios and/or configurations.

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

Note: Gcanbe

The may es.

* Note:

DUT

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

In order to determine the test encoder/decoder, the following aspects are identified.

* 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).

Based on the above principles, the potential 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

[Table for potential options of determining the test decoder]



### 7.4.2 CSI feedback enhancement

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

The following potential options are identified if studying how to define test metrics for AI/ML CSI feedback enhancement.

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

[FFS: Feasibility to define the CSI prediction accuracy need discussion in WI]

### 7.4.3 Beam management

The following potential options are identified if studying how to define test metrics for beam management.

* 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.

### 7.4.4 Positioning accuracy enhancements

The following potential options are identified if studying how to define test metrics for positioning accuracy enhancement.

* 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

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