3GPP TSG-RAN WG4 Meeting #114bis R4-2504684

**Wuhan, Hubei, China, 7th – 11th April, 2025**

**Agenda item:** 7.19.5

**Source:** Moderator (Qualcomm)

**Title:** Topic summary for [114bis][126] NR\_AIML\_air\_part1

**Document for:** Information

# Introduction

This is the summary thread for issues related to the NR AI/ML WI. The topics handled in this thread are the following:

* CSI reporting requirement and testing framework for CSI prediction
* RRM core requirement and testing framework for beam management
* RRM core requirement and testing framework for Positioning accuracy enhancement

Issues related to the general part (agenda item 17.19.1) are treated in the AI/ML part 2 thread ([114bis][126])

A WF summarizing the agreements from RAN4#114 was agreed in RP-2502856. The discussion will take into account these previous agreements.

# Topic #1: CSI reporting requirement and testing framework for CSI prediction

This section contains the sub-topics regarding general issues and proposed TR updates

## Companies’ contributions summary

|  |  |  |
| --- | --- | --- |
| **T-doc number** | **Company** | **Proposals / Observations** |
| [**R4-2503252**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503252.zip) | Qualcomm Incorporated | **Observation 1: RAN4 Rel-18 non-AI-ML CSI prediction tests assumed doppler domain (DD) basis N4 = 1.**  **Observation 2: Rel-18 allowed UE to indicate via 40-3-2-1 and 40-3-2-1a that it can only support DD-basis N4 = 1 in non-AI-ML CSI prediction.**   * **Rel-19 AI-ML CSI prediction may introduce a similar UE capability.**   **Observation 3: The SGCS obtained with AI-ML based CSI prediction in both step 1 and step 2 are significantly higher than those obtained with “sample and hold” approach.**  **Table 2: SGCS results with sample-and-hold and AI-ML based prediction**   |  |  |  |  | | --- | --- | --- | --- | | Structure | Layer | SVD, SGCS1 | PC7, SGCS2 | | Sample and hold | 0 | 0.3127 | 0.297 | | 1 | 0.1734 | 0.1662 | | AI-ML based prediction | 0 | 0.9572 | 0.8653 | | 1 | 0.9173 | 0.7788 |   **Observation 4: The AI-ML model of Table 2 was trained based on 50Hz Doppler.**   * **Further investigation will be performed to see how AI-ML based CSI prediction performs in different settings corresponding to the diverse set of assumptions of RAN4 114 agreements.** * **More investigations are needed to see the performance of CSI prediction obtained via AI-ML model that was trained for a “generalized” setting (e.g., to tackle for 20Hz, 50Hz and 100Hz Doppler).**   **Observation 5: The performance of AI-ML based CSI prediction depends on the complexity of the model.**   * **The SGCS results submitted to future meetings may vary from the ones shown in Table 2 based on the complexity analysis of the AI-ML model.**   **Proposal 1: RAN4 defines Rel-19 AI-ML CSI prediction tests for, at least, doppler domain basis N4 = 1.** |
| [**R4-2503270**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503270.zip) | MediaTek inc. | **Observation #1: FDD 16TX and FDD 32TX results with the same Doppler perform quite similarly.**  **Observation #2: In lower Doppler accuracy is better than in higher Doppler.**  **Proposal #1: We propose incorporating realistic channel estimates in the next phase of evaluation.**  **Observation #3: When defining Release 18 PMI prediction test requirements, it was challenging to find test configuration that enables measurable PMI prediction gains over legacy baseline.**  **Proposal #2: We propose to use eType II non-predicting reporting as reference in test case definition to ensure test feasibility.**  **Proposal #3: We propose to first focus on static condition to guarantee test feasibility.**  **Observation #4: There are four existing PMI requirements for PMI prediction that can be used as starting point for Rel-19 work.**  **Observation #5: There is still room to improve existing test configuration to enable even better prediction gain.**  **Proposal #4: We propose to use existing eType II doppler codebook test configurations as a starting point for Rel-19 work.**  **Proposal #5: We propose to further study if existing eType II doppler codebook test configurations can be enhanced for better prediction opportunities.**  **Proposal #6: We propose to study prediction timing in test configuration.**  **Proposal #7: We propose to study prediction aging in test configuration.**  **Proposal #8: We propose to study CSI-RS resource allocation in test configuration.** |
| [**R4-2503406**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503406.zip) | CATT | **Proposal 1: Use Rel-18 Doppler codebook to define requirements and tests for AI/ML CSI prediction.**  **Observation 1: Consistency between training and test dataset is implicitly kept if test configurations include enough information to help UE vendors generate consistent training datasets.**  **Proposal 2: RAN4 to discuss necessary information that needs capturing in test configurations when defining test configurations.** |
| [**R4-2503517**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503517.zip) | Apple | **Observation 1: In the context of CSI prediction performance monitoring, RAN1 is already exploring a network-side mechanism that relies on UE-reported ground truth. This same approach can be utilized to evaluate the accuracy of CSI prediction as a performance KPI within RAN4**  **Observation 2: Significant gains are observed with AI/ML prediction over Sample and Hold for Doppler spreads up to 50 Hz. Further investigations are needed for higher Doppler spreads**  **Proposal 1:** **The γ value comparing predicted PMI to random PMI can be used in the test for CSI prediction. To infer the gains of AI/ML-based CSI prediction, the γ values can be compared between AI/ML and non-AI/ML cases.**  Proposal 2: RAN4 should assess the performance of AI/ML-based CSI prediction using the same test configurations and conditions as those defined for non-predicting eTypeII and/or eTypeII Doppler-r18.  **Proposal 3: Same mechanism used for NW-side performance monitoring can be employed for testing where the UE provides both the current channel information and the predicted channel information. Subsequently, the TE can calculate the SGCS by comparing the current channel from the latest report with the predicted channel from the previous report.**  Proposal 4: RAN4 to assume R18 eType II Doppler codebook for reporting AI/ML based CSI predictions.  **Proposal 5: RAN4 to perform simulation results alignment among companies. If simulation results don’t match, companies should be encouraged to share CSI prediction datasets and/or AI/ML model architectures.**  **Proposal 6: Simulation results show that there is a tradeoff between training and testing losses (overfitting). RAN4 to discuss the need for optimal models as a function of Doppler spread.**  **Proposal 7: For generalization purposes investigate the granularity of Doppler spread where a new AI/ML model is needed to maintain performance.**   |  |  |  |  | | --- | --- | --- | --- | | **Rank 1 (16 Tx)** | **S/H (AI/ML)**  **20 Hz** | **S/H (AI/ML)**  **50 Hz** | **S/H (AI/ML)**  **100 Hz** | | **NMSE** | **-7.15dB (-15.2dB)** | **0.21dB (-7 dB)** |  | | **SGCS** | **0.856 (0.983)** | **0.539 (0.867)** | **0.45** | |
| [**R4-2503769**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503769.zip) | CMCC | ***Proposal 1: for CSI prediction, it is proposed to use Rel-18 eType II Doppler codebook.***  ***Proposal 2: for CSI prediction, it is preferred to firstly focus on static conditions. Considering the timeline and workload, non-static conditions can be discussed later if time allowed .***  ***Proposal 3: For channel model for CSI prediction tests, it is proposed to focus on TDL channel models in Rel-19. Using CDL channel model for CSI prediction test can be further dicsussed in later release based on the progress on SCM.***  ***Proposal 4: for CSI prediction, the evaluation results for step 1 are shown as following.***   |  |  | | --- | --- | |  | SGCS | | ConvLSTM | 0.9249 | |
| [**R4-2504005**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504005.zip) | ZTE Corporation, Sanechips | ***Observation 1. CSI prediction is one side model, RAN4 can assume that the AI based CSI prediction as a black box, compared to legacy prediction algorithm, like Wiener filter, Kalman filter, which is based on the UE vendor implementation.***  ***Observation 2. Channel characteristics, model capabilities, data quality, and training methods all affect the convergence of results.***  ***Observation 3. From our simulation results, we can observe that the SGCS predicted by AI is 0.96, and non predicted is 0.82 under TDLA30-20 XP medium condition. At the same time, under the TDLA30-100 XP medium condition, the SGCS predicted by AI is 0.94 , and non predicted is 0.81.***  ***Observation 4. During the conformance testing, TE vendor generates test data based on the scenarios defined by RAN4, including frequency, power, modulation order, channel conditions, physical waveform, signaling process, etc.***  ***Proposal 1. Propose to discuss the criteria for determining whether the results have reached alignment.***  ***Proposal 2. Proposed to consider that consistency will be implicit, as the emulated channel used for testing will be highly similar to the one employed in RAN4 simulations.*** |
| [**R4-2504022**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504022.zip) | vivo | **Observation 1: Initial evaluation results for CSI prediction AI/ML model is shown in Table 2-1, for FDD and 32gNB Tx.**  Table 2-1. Initial evaluation results for CSI prediction AI/ML model, for FDD and 32gNB Tx.   |  |  |  |  | | --- | --- | --- | --- | | FDD, 32 gNB Tx | Doppler spread: 20Hz | Doppler spread: 50Hz | Doppler spread: 100Hz | | SGCS | 0.993 | 0.902 | 0.298 |   **Proposal 1: Reference AI model for CSI prediction would be needed, if it is difficult to align the results corresponding to test metric**  **Proposal 2: AI prediction and non-AI prediction could follow the similar test procedure. AI-prediction would have similar or higher gamma value compared to non-AI prediction.**  **Proposal 3: RAN4 to discuss further simulation assumptions for Tput in the RAN4#114bis, e.g., CSI-RS configuration, codebook configuration of Rel-18 eTypeII-Doppler, and SNR.**  **Proposal 4: LCM requirements (e.g., monitoring) should wait for RAN1 progress. If there is any spec impact, RAN4 to discuss whether to capture it in the RRM spec.** |
| [**R4-2504235**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504235.zip) | OPPO | **Proposal 1: The requirements for AI/ML based CSI prediction(or AI/ML based PMI prediction) could be specified in terms of the ratio:**  **In the definition of , [e.g., for 32TX PMI requirements, FFS other TX values], is 90 % of the maximum throughput obtained at [ ] using the precoders configured according to the UE reports, and is the throughput measured at [ ] with random precoding.**  **Proposal 2: could be obtained within a CSI prediction window. If a CSI prediction window contains more than one CSI prediction instance,**  **Option1: is calculated for each predicted CSI and then do the average within the CSI prediction window**  **Option2: and are calculated within the CSI prediction window first and then get the averaged value.**  **Proposal 3: 32TX PMI could be the baseline, FFS other values.**  **Proposal 4: TDL channel could be the baseline, FFS other channel conditions.**  **Proposal 5: whether tests should be conducted at a given SNR or a SNR range should be further considered.**  **Proposal 6: For CSI prediction, following test steps could be considered:**   * + **Step1: TE sends CSI-RS to DUT, within a CSI observation window (signal generating and transmitting)**   + **Step2: DUT conducts the CSI measurement within the CSI observation window (CSI-RS receiving)**   + **Step3: DUT conducts the CSI prediction within a CSI prediction window (Inference)**   + **Step4: DUT reports the predicted CSI (CSI reporting)**   + **Step5: TE calculates the for performance [and calculates the measurement/inference latency for LCM]**   **Proposal 7: Besides the TDL channel based tests(TDL could be a baseline test), CSI-related tests may be conducted under CDL channel [or other more practical channel conditions] to check a relatively generalized performance.**  **Proposal 8: After having testing cases for static configurations, RAN4 can further consider and evaluate whether to introduce non-static testing scenarios and configurations.**  **Proposal 9: Regarding the RAN4 Simulation for CSI prediction and the Reference model for CSI prediction, RAN4 should first assess the CSI prediction evaluation results across different companies before determining whether a reference model is necessary.**  **Proposal 10: Regarding the RAN4 Simulation for CSI prediction and the Reference model for CSI prediction,**   * + **If RAN4 achieves aligned CSI prediction evaluation results in Step 1 (as described in the agreed simulation assumptions), further evaluation (e.g., Step 2 comparisons) can proceed.**   + **If alignment in Step 1 proves difficult, RAN4 should identify the key factors causing discrepancies and initiate a second-round evaluation to reassess Step 1 alignment (e.g., before/during RAN4#115).**   + **If alignment remains unachievable after the second round, RAN4 may then trigger discussions and evaluations on a reference model for CSI prediction (e.g., after RAN4#115).**   **Proposal 11: Consistency between training data and test data should be ensured**   * + - **FFS: Whether the test emulated channel will be similar to the one used in RAN4 simulations**     - **FFS: Whether anything special should be done during the test**   **Observation 1: Regarding the simulation for AI/ML-based CSI prediction, we conducted the Step 1 evaluation for the CSI prediction using the simulation assumptions(highlighted in yellow in Table 1). The simulation results are shown in Table 2:**  **Table 1: Baseline Simulation assumptions for AI/ML based CSI prediction**   |  |  | | --- | --- | | **Parameter** | **Value** | | Duplex, Waveform | FDD, TDD,  OFDM | | Carrier frequency | 2GHz for FDD  4GHz for TDD | | Bandwidth | 10MHz for FDD, 40MHz for TDD | | Symbol | 1 symbol | | Number of subbands | 13 subbands, 4RB as a bundle for FDD  14 subbands, 8RB as a bundle for TDD | | Subcarrier spacing | 15kHz for 2GHz, 30kHz for 4GHz | | gNB TX antennas | Option1: 32 (N1,N2)=(4,4) as defined in Rel-15 type1  Option2: 16 (N1,N2)=(4,2) as defined in Rel-18 CSI prediction requirements | | UE RX antennas | 4 | | Channel model | TDLA30 as baseline with XP medium correlation | | Doppler spread | 20Hz, 50Hz and 100Hz | | Delay spread | 30ns | | Channel estimation | Start with Ideal DL channel estimation, for the purpose of calibration and/or comparing intermediate results (e.g., accuracy of AI/ML output CSI, etc.).  Further consider practical channel estimation in the next step   * Use a common CSI-RS configuration as in demodulation test (may need to define/choose a CSI-RS configuration) | | Rank per UE | Rank 1 and 2 | | CSI feedback assumption | Step1: check the CSI prediction (outputs Raw channel) performance, do not need to take CSI feedback into account  Step2: take CSI feedback into account (wait more related agreements in RAN4)   * Option 1: Use Rel-18 eTypeII-Doppler * Companies are encouraged to check RAN1 agreements | | CSI feedback overhead | Option1: Parameter combination = 2 (PC2)  Option2: PC4  Option3: PC6  Option4: PC7  Note: this part relates to step2, may be down selected later | | Baseline | Random PMI with Rel-15 Type I single panel codebook (for Step 2 or performance requirements) | | KPI | Average of SGCS for intermediate results over all subbands per layer (e.g. for rank 2),   * comparing the SVD of model output with SVD of the ground truth Raw channel   Note:   * Step1: check SGCS before CSI feedback (between the raw channel matrix and ground-truth at UE) * FFS Step2: check SGCS (or other metrics, e.g. TPs) after CSI feedback | | Model input type | Raw channel matrix | | Model out type | Raw channel matrix | | Observation window | Observation window (number/distance): 5/5ms as baseline  Optional for 10/5ms | | Prediction window | Prediction window (number/distance between prediction instances/distance from the last observation instance to the 1st prediction instance): 1/5ms/5ms |   **Table 2: Simulation results for AI/ML based CSI prediction**   |  |  |  | | --- | --- | --- | | Doppler spread | SGCS | NMSE | | 20Hz | 0.9994 | 0.0001 | | 50Hz | 0.9032 | 0.0335 | |
| [**R4-2504317**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504317.zip) | Huawei,HiSilicon | ***Proposal 1***: *RAN4 will not consider CDL channel model in AI CSI prediction.*  ***Proposal 2***: For generalization testing, there is no need to use non-static condition.  ***Proposal 3***: RAN4 will start to discuss whether/how to define non-static condition during testing for UE-side monitoring after other WGs achieve sufficient progress.  ***Proposal 4***: SGCS performance results of AI CSI prediction.  Table 1. SGCS performance results of AI CSI prediction   |  |  |  |  | | --- | --- | --- | --- | |  | 20Hz | 50Hz | 100Hz | | Rank 1 | 0.9985413 | 0.8673333 | 0.301114 | | Rank 2 | 0.99800384 | 0.8166052 | 0.233405 |   Note: gNB TX antennas is Option 2.  ***Observation 1***: Significant degradation in SGCS performance can be observed with 100Hz Doppler spread.  ***Proposal 5***: RAN4 will not define performance requirement for AI CSI prediction with 100Hz Doppler spread. |
| [**R4-2504346**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504346.zip) | Ericsson | Proposals:  [Proposal 1: RAN4 should discuss the simulation assumption to evaluate the follow PMI performance with AI-based PMI prediction.](#_Toc194104054)    Figure 1 Performance monitoring of AI/ML-based CSI prediction model.  [Proposal 2: RAN4 discuss the testability of monitoring reporting options, and RAN4 could inform RAN1 if RAN1 does not make decision.](#_Toc194104055)  [Proposal 3: RAN4 should consider more than one conditions to guarantee the generality of the AI-based CSI prediction model. At least RAN4 consider several channel models (e.g., TDLA30 and TDLC300) and antenna configurations (e.g., 16 CSI-RS ports and 32 CSI-RS ports).](#_Toc194104056)  [Proposal 4: RAN4 should consider defining AI/ML-based CSI prediction requirements using CDL channel model when a suitable CDL model is available.](#_Toc194104057)  Table 2 Example of PMI test with semi-static scenario/condition   |  |  |  |  | | --- | --- | --- | --- | |  | Channel model | MCS | Gamma | | T1 (e.g., 1000 slots) | TDLA30-10 | 64QAM, 0.5 | 2.0 | | T2 (e.g., 1000 slots) | TDLC300-100 | QPSK 0.3 | 1.3 | | T3 (e.g., 1000 slots) | TDLA30-50 | 16QAM 0.5 | 1.7 |   Proposal 5: For CSI prediction performance requirements, RAN4 discuss the feasibility of mixing several static condition tests to ensure the generality of the prediction model.  [Proposal 5: For CSI prediction performance requirements, RAN4 discuss the feasibility of mixing several static condition tests to ensure the generality of the prediction model.](#_Toc194104058) |
| [**R4-2504539**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504539.zip) | Samsung | **LCM related Performance Monitoring**  **Proposal 1: Further study performance monitoring related requirement, and testability according to the sufficient progress in RAN1.**  **CSI reporting requirement**  **Proposal 2: RAN4 can consider to apply the SNR test point related 90% of maximum throughput achieved with following PMI similar as legacy PMI test for AI-based CSI prediction as baseline.**  **Proposal 3: RAN4 applies Rel-18 eType II Doppler codebook for AI/ML based CSI prediction**  **Proposal 4: RAN4 should also compare the ratio of throughput obtained based on following PMI with Rel-18 eType II and random PMI with Rel-15 type I single panel codebook, when defining the requirement of AI/ML based CSI prediction with Rel-18 eType II doppler codebook.**  **Proposal 5: RAN4 should generate the training data by using the same NW-side specified for test**  **Observation 1: The whole CSI feedback process is same as the existing PMI test, excepting for CSI prediction**  **Proposal 6: RAN4 should check the throughput as KPI for Step 2. No need to check the SGCS after CSI feedback.**  **Proposal 7: RAN4 consider PC7 for AI/ML based CSI prediction with Rel-18 eType II doppler codebook**  **Proposal 8: RAN4 consider the following CSI-RS configuration for each CSI-RS resource as starting for evaluation AI/ML CSI prediction**  **For 32Tx**   |  |  |  | | --- | --- | --- | | **NZP CSI-RS for CSI acquisition** | **Number of CSI-RS ports (*X*)** | **32** | | **CDM Type** | **CDM4 (FD2, TD2)** | | **Density (ρ)** | **1** | | **First subcarrier index in the PRB used for CSI-RS (k0, k1, k2, k3)** | **Row 17,(2, 4, 6, 8)** | | **First OFDM symbol in the PRB used for CSI-RS (l0, l1)** | **Row 17,(5, 12)** |   **For 16Tx**     |  |  |  |  | | --- | --- | --- | --- | |  | **Number of CSI-RS ports (*X*)** |  | **16** | | **CDM Type** |  | **CDM4 (FD2, TD2)** | | **Density (ρ)** |  | **1** | | **First subcarrier index in the PRB used for CSI-RS (k0, k1, k2, k3)** |  | **Row 12, (2, 4, 6, 8)** | | **First OFDM symbol in the PRB used for CSI-RS (l0)** |  | **(5)** |   **Proposal 9: RAN4 should further discuss the proper test setup for AI/ML based prediction based on Rel-18 Type II-Doppler codebook. The following test setup for FDD with AI/ML CSI prediction can be considered as starting point**    **Proposal 10: RAN4 should consider the practical channel estimation (e.g., LS or LMMSE) as baseline for step 2**  **Proposal 11: RAN4 shall discuss how to choose the proper SNR range for AI/ML based CSI prediction model training and test when applying the practical channel estimation. The following method can be considered as starting point:**   * **Use the SNR value with 90% of throughput achieved based on following PMI as starting point, based on this SNR, considering ±X dB range for data collection for AI/ML training, where the SNR value is derived based on existing Rel-18 eType II Doppler codebook under the same configuration**   **RAN4 should also discuss whether the AI model is training based on dataset with specific SNR condition or mixed dataset with the selected SNR range, when applying the practical channel estimation**  **Observation 2: AI/ML based CSI prediction show better performance compared with nearest historical CSI scheme (benchmark#1), while generalization problem needs further study when defining requirement** |
| [**R4-2504541**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504541.zip) | Nokia | 1. RAN4 to support Option 2 in that data consistency between training and test data is achievable without a dedicated requirement. 2. RAN4 to study the model-ID-based LCM requirements if localized models are considered for CSI prediction. 3. RAN4 to study delay requirements in Steps 4 and 5 of the LCM procedure based on RAN2’s 5-step approach. 4. RAN4 to discuss whether the delay components of LCM operations can vary depending on the specific CSI report configurations. |

## Open issues summary

The agreements from the previous meeting on this topic are listed below for reference:

**Issue 1-2: Reference model for CSI prediction**

**Agreement:**

* Assume no reference AI model for CSI prediction as baseline
  + If it is difficult to align the results corresponding to test metric, RAN4 will discuss the reference model.

**Issue 1-3: Training and test data consistency**

**Agreement:**

* further discuss the following options
  + Option 1: Consistency between training data and test data should be ensured
    - RAN4 to study how to ensure consistency between training data(simulation) and test data (conformance test) and whether anything special should be done during the test
  + Option 2: Consistency will be implicit as test emulated channel will be very similar to the one used in RAN4 simulations

**Issue 1-4: Generalization**

**Agreement:**

* Postpone this discussion until the requirement and test definition is clear

Also, simulation assumptions were agreed in R4-2502970.

The open issues were grouped in the following sub-topics for further discussion:

1. Codebook to be used for reporting
2. Requirement baseline
3. Test baseline
4. Test metric
5. Simulation results and next steps
6. Ues REl-18 Doppler codebook
7. LCM

### Sub-topic 1-1

*Reporting codebook*

Inline with RAN1 agreement, it should be confirmed that Rel-18 Doppler codebook will be used for AI/ML based reporting tests

**Issue 1-1: Codebook to be used in the tests**

* Proposals
  + Option 1: use Rel-18 eType II Doppler codebook for AI/ML based CSI predictions
  + Option 2: further discuss
* Recommended WF

Option 1

### Sub-topic 1-2

*Requirement baseline*

It is proposed to define requirements for Dopler domain basis N4=1, N4>1 might be introduced as an option.

**Issue 1-2: Requirement baseline**

* Proposals
  + Option 1: Requirement to be introduced for Dopler domain basis N4=1
    - N4>1 can be further discussed in the future if needed
  + Option 2: Others
* Recommended WF
  + Option 1

### Sub-topic 1-3

*Test baseline*

Several companies brought suggested taking the Rel-18 PMI prediction test setup as baseline

**Issue 1-3: Test baseline**

* Proposals
  + Option 1: Take Rel-18 PMI prediction test setup as baseline
  + Details/parameters specific to AI/ML based predictions to be further discussed
  + Option 2: Following test setup to be taken as starting point (R4-2504539):
    - 
  + Option 3: others
* Recommended WF
  + Option 1

Also discuss if the proposal in Option 2 can be taken as baseline or more discussion is needed.

### Sub-topic 1-4

*Test metric*

Several companies proposed to use relative throughputγas the metric for the requirements/tests. Some of the details also need further discussions

**Issue 1-4: Test metric**

* Proposals
  + Option 1: **The requirements for AI/ML based CSI prediction(or AI/ML based PMI prediction) could be specified in terms of the ratio:**
    - * **is 90 % of the maximum throughput obtained at [ ] using the precoders configured according to the UE reports, and is the throughput measured at [ ] with random precoding.**
        + **is obtained with Rel-15 type I single panel codebook**
      * **FFS how is calculated:** 
        + **Option 1: γ\_(AI/ML) is calculated for each predicted CSI and then do the average within the CSI prediction window**
        + **Option 2: t\_(ue\_AI/ML) and t\_rnd are calculated within the CSI prediction window first and then get the averaged γ\_(AI/ML) value.**
        + **Option 3: FFS**
  + Option 2: other proposals
* Recommended WF
  + Option 1

If Option 1 is agreed, it should be discussed whether Option 1 or 2 for how to calculate the throughput can be already agreed or this is left FFS (Option 3) for a future meeting

### Sub-topic 1-5

*Simulation results and next steps*

Several companies submitted simulation results and proposals for next steps (refinement of parameters, etc)

**Issue 1-5: Simulation results and next stepts**

* Proposals
  + Discuss the simulation results and next steps based on the simulation summary prepared by oppo
    - Simulation results
    - Refinement of parameters:and what to for the next steps
      * Introduce realistic channel estimation
      * Further discuss CSI-RS configuration, codebook configuration, Dopller, SNR, etc
      * Other proposals
* Recommended WF
  + To be discussed

Discussion on refinement of simulation parameters to be done mainly offline

### Sub-topic 1-6

*Generalization*

Several companies proposed to further study generalization issues and how to ensure that the UE performance does not degrade under different conditions.

**Issue 1-6: Generalization**

* Proposals
  + Option 1: Mix several static conditions to endure the generality of the prediction model
  + Option 2: Introduce tests with TDL-C300 or different number of CSI-RS ports, etc
  + Option 3: Introduce tests with CDL channel model if available
  + Option 4: further investigate the granularity of Doppler spread for which a new AI/ML model would be needed
* Recommended WF
  + Postpone discussion to future meetings

In the previous meeting it was agreed that generalization would be discussed later on when the requirements and tests details (e.g. conditions, etc) would become more clear. Companies can still consider the above proposals when discussing the definitions of the requirements and tests.

Ericsson: Fine with Moderator recommendation. Performance can be postponed.

### Sub-topic 1-7

*LCM*

Several companies proposed to further LCM procedures while other companies propose to wait for further progress in other WG.

**Issue 1-7: LCM**

* Proposals
  + Option 1: Wait for further progress in RAN1/2 before continuing the discussion in RAN4
  + Option 2:
    - RAN4 to study the model-ID-based LCM requirements if localized models are considered for CSI prediction.
    - RAN4 to study delay requirements in Steps 4 and 5 of the LCM procedure based on RAN2’s 5-step approach.
    - RAN4 to discuss whether the delay components of LCM operations can vary depending on the specific CSI report configurations.
  + Option 3: RAN4 to discuss the testability of monitoring reporting options, and RAN4 could inform RAN1 if RAN1 does not make decision.
* Recommended WF
  + Option 1

Once the details become clear, RAN4 will discuss whether/how to define requirements for LCM. Above proposals can be considered in future discussions

# Topic #2: RRM core requirement and testing framework for beam management

This section contains the sub-topics regarding specific issues for beam management.

## Companies’ contributions summary

|  |  |  |
| --- | --- | --- |
| **T-doc number** | **Company** | **Proposals / Observations** |
| [**R4-2503230**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503230.zip) | CAICT | **Proposal 1: Suggest to discuss the feasibility of defining a tight requirement for measurement used to generate ground truth and how to verify the measurement performance.**  ***Observation 1:*** *Relative RSRP accuracy shows the benefit for models of beam prediction that only output beam indices****.***  ***Observation 2:*** *Definitions in option 2 and 3 require more clarification before further decision.*  **Proposal 2: Suggest to consider option 1 as the definition of relative RSRP accuracy requirement,**   * **Option 1:** **The relative RSRP accuracy for AI/ML-based beam prediction = the ground truth of L1-RSRP of the predicted beam index i - the ground truth of the L1-RSRP of the Top-1 ideal beam index j** |
| [**R4-2503257**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503257.zip) | NTU | **Observation 1: Simulation assumptions for Set B beams are missing from the previous agreement.**  **Proposal 1: Use the following BS antenna configurations and beamforming patterns:**   |  |  | | --- | --- | | BS antenna configurations | * Set A beams   + One panel: (M, N, P, Mg, Ng) = (4, 8, 2, 1, 1), (dV, dH) = (0.5, 0.5) λ as baseline.   + Number of Tx beams is 32 * Set B beams   + One panel: (M, N, P, Mg, Ng) = (2, 4, 2, 1, 1), (dV, dH) = (0.5, 0.5) λ as baseline.   + Number of Tx beams is 8 | | Beamforming characteristic of the BS pattern | Set A beams:  32 beams with grid of 4 elevation angles from [-25° to 25°] and 8 azimuth angles from [-60° to 60°]  Set B beams:  8 beams with grid of 2 elevation angles from [-25° to 25°] and 4 azimuth angles from [-60° to 60°] | |
| [**R4-2503307**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503307.zip) | Xiaomi | **Observation 1: For derive ground truth for best Top-1 beam, UE needs to measure for all beams in set A under high SNR. However, there is still some measurement uncertainty due to RF and BB. It’s challenging for UE to detect the best 1st beam even at high SNR due to the small gap between best 1st and 2nd beam.**  **Observation 2: For beam prediction accuracy, option 2 is not correct since UE can’t distinguish which is the best beam index due to measurement error.**  **Observation 3: From RAN1 performance monitoring discussion, the similar metric Top-K/M is under discussing, where at least one of the Top M beam(s) of the resource set(s) for monitoring is among Top-K predicted beam(s).**  **Proposal 1: For beam ID prediction accuracy performance metric, the KPI is**   * + **Top-K/N (%): the percentage of "at least one of the Top-N genie-aided beam is one of the Top-K predicted beams".**   **Observations 4: Differential reporting induces error propagation in absolute RSRP reconstruction:**  **- 1dB uncertainty per differential report (for 2dB step size)**  **- Cumulative effect when beam ranking changes between reports**    **Observation 5: Direct comparison of predicted vs. measured RSRP for 2nd-Kth beams becomes unreliable due to:**  **- Potential beam ranking changes**  **- Compounded quantization errors**  **Proposal 2: For BM-case 1, only define absolute RSRP accuracy for 1st-ranked predicted beam in report:**  **predicted L1-RSRP of beam index i – ground truth of L1-RSRP of beam index i.**  **where the index i is the 1st-ranked beam index in report.**  **Proposal 3: The network shall configure K=1 for the RSRP measurement report configuration, applied exclusively to the 1st-ranked predicted beam.**  **Proposal 4: For both BM case-1 and BM case-2, RAN4 to discuss whether there is impact on measurement period due to specific RX beam assumption.**  **Proposal 5: for BM-case 1, legacy measurement period can be used as baseline. FFS for the impact of specific RX beam(s).**  **Proposal 6: For BM-case 2, RAN4 to discuss impact on measurement requirement, e.g. measurement time, prediction time, measurement restriction and scheduling restriction, etc.**  **Observation 6: When best predicted TX beam is in set B, UE knows the best RX beam. When best predicted TX beam is in set A but not in set B, it’s still possible for UE to know the RX beam corresponding to the best TX beam.**  **Proposal 7: RAN4 may know RX beam based on UE capability.**  **Observation 7: SINR distribution for all UE group will span a large SINR range for all UE groups.**  **Observation 8: For K=1, measurement accuracy error will degrade performance a lot for all UE groups.**  **Observation 9: With K=4/1, when maximum SINR is larger than 0dB(UE group 1,2,3), prediction accuracy can be nearly around or higher than 90%.**  **Observation 10: When SINR<0dB(UE group 4), prediction accuracy for K/1 will degrade to 72%. If relax to K=4/2, the performance will increase to 87%.**  **Observation 11: When SINR<-5dB, prediction accuracy for all KPI is bad.**  **Proposal 8: Impact of measurement error has impact on prediction accuracy requirement, RAN4 to define SNR side condition for the best TX beam in Set B.**  **Proposal 9: Considering measurement error impact, RAN4 to relax beam index prediction KPI to Top K/N.**  **Proposal 10: For OTA test, prefer to split the test into two parts.**   * **Part 1: Verify AI model prediction algorithm, using single AOA model.** * **Part 2: Verify input data accuracy of model AI based on RX beam sweeping, using multi-AOA model.**   **Observation 12: If Emulated signal considers both TX and RX beamforming gain, RX beam pattern is UE specific. In order to reduce test complex, prefer to design a common pre-defined RX beam pattern for both TE and UE.**  **Observation 13: if a fixed RX beamforming gain is used for multi-path TX signals, there will be some mismatch between emulated signal for test and original signals for training. The mismatch impact needs to be evaluated.**  **Observation 14: For simplified channel model, L1-RSRP delta between different TX beams will increase if cluser number and AOA number is further reduced. It’s more easiler for AI to distingwish the best Top-K TX beam. The AI prediction requirement will be relaxed.**  **Proposal 11: For single AOA, RAN4 to discuss whether it’s possible to design common pre-defined RX beam pattern to generate emulated single AOA signals.**  **Proposal 12: For simplified multi-AOA CDL model, cluster number can be reduced from spatial angle offset and power aspect:**   * **Combine clusters if AOA offset between these clusters is smaller than a threshold** * **Remove cluster with low power.**   **Proposal 13: If the test metric is beam index prediction accuracy, UE may need to report ground truth of N best beam index for set A.**  **Observation 15: For deriving ground truth of set A, SNR level should satisfy that for at least Top-M beams, the measurement error is small. SNR didn’t need to guarantee that all beams in set A can be measured accurately.**  **Proposal 14: For deriving ground truth of set A, SNR level should satisfy that for Top-M beams, the measurement error is small. RAN4 to discuss how to decide M and how to set SNR.**  **Proposal 15: For BM case-1, channel doppler can set to 0 or a small value to guarantee that there is neglectable L1-RSRP variation.**  **Proposal 16: for BM case-2, channel doppler will depend on UE speed and UE trajectory. RSRP variation impact can be considered in RSRP accuracy requirement.**  **Observation 16: UE will report best predicted L1-RSRP/beam index at T1 and measured RSRP/beam index of best beam at T2 to TE, it’s easy for UE to pass the test by cheating.**  **Proposal 17: RAN4 to discuss how to solve the UE cheating issue if UE report both predicted result and ground truth.** |
| [**R4-2503407**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503407.zip) | CATT | **Proposal 1: RAN4 postpone down-selection of KPIs after having more observations and understanding based on simulation results.**  **Observation 1: For BM-Case 1, UE measures SetB beams and predict results for SetA beams in spatial domain. The legacy L1-RSRP measurement requirement can be reused for BM-Case 1.**  **Observation 2: For BM-Case 2, UE measures SetB beams and predict results for SetA beams in time domain. The legacy L1-RSRP measurement requirement can also be reused for BM-Case 2.**  **Proposal 2: RAN4 to reuse legacy L1-RSRP measurement requirement for both BM-Case 1 and BM-Case 2.**  **Proposal 3: For DL Tx prediction, UE Rx beam should be known to UE itself.**  **Observation 3: Test target of AI/ML based BM depends on the way of how performance requirements will be defined.**  **Proposal 4: For AI/ML based beam management, two options can be considered for defining performance requirements:**   * **Option 1: The legacy performance requirements are reused, regardless of overhead reduction ratio.** * **Option 2: The legacy performance requirements are relaxed with margins, and larger margin values are used for higher overhead reduction ratios.** |
| [**R4-2503515**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503515.zip) | Apple | **Observation 1: From the network operator's perspective, a number of optimal beams need to be known in order to optimize load balancing and make trade-offs in performance and complexity**  **Observation 2: Ensuring that RSRP accuracy implies beam prediction accuracy is challenging and depends on absolute RSRP accuracy and the RSRP difference between the best beams**  **Observation 3: When using neural networks for regression, we rely on the Normalized Mean Squared Error (NMSE) criterion for training. It's crucial to ensure that the accuracy of weaker beams is not compromised by the accuracy of the strongest beams**  **Observation 4: The measurement period for Set-B would depend on the accuracy needed for prediction.**  **Observation 5: There is a probability that the Top-K predicted beams may not align with the Top-K beams determined by ground-truth RSRP measurements**  **Observation 6: For RSRP absolute accuracy, if the index i can represent any beam within the Top-K beams based on ground-truth L1-RSRP, the RSRP accuracy for beams that are part of the Top-K predicted beams but not included in the Top-K beams based on ground-truth RSRP cannot be guaranteed**  **Observation 7: The legacy definition for known TCI state cannot be used when target TCI state is based on predicted beam and not transmitted to and measured & reported by UE.**  **Observation 8: RAN4 needs to discuss known TCI for UE side and NW side prediction.**  **Observation 9: RAN4's goal of aligning test conditions with real-world deployment scenarios is closely related to ensuring generalization performance. Since field deployment conditions can vary significantly across different environments, the test conditions should account for this variability**  **Observation 10: The testing environment for Beam Management (BM) case 1 and BM case 2 should replicate conditions that sufficiently capture the correlations of the transmit (Tx) and receive (Rx) beam patterns across the entire spectrum of propagation conditions in both spatial (angles of arrival and departure - AoAs and AoDs) and temporal domains.**  **Observation 11: The following conditions will introduce randomness and variations in propagation conditions across both time and spatial domains for the computation of L1-RSRP**   1. **Different AoDs with respect to the Tx antenna array** 2. **Different AoAs with respect to the Rx antenna array** 3. **Different superpositions of {AoA,AoD} pairs** 4. **Fading/Variation in time domain ( different {AoA,AoD} pairs per resolvable delay bin path)** 5. **UE movement (including rotation)**   **Observation 12: During real-world deployment, UE will encounter random radio propagations characterized by variations in both spatial and temporal domains (fading). Testing UE under similar conditions is important to reflect the realities of deployment accurately.**  **Observation 13: The following questions need to be answered to evaluate the feasibility of the FR2 OTA-based test procedure:**   1. **How can we generate multiple beams from the Set-B and Set-A Tx beams given the limitation of two AoAs?** 2. **What assumption is made regarding Rx beam sweeping? Does the UE utilize a fixed Rx beam, or does it sweep to find the optimal Rx beam?** 3. **How can we simultaneously emulate different AoAs (Rx beam) and AoD (Tx beams)?** 4. **How can we achieve dynamic variation in the AoD domain (Tx beam sweeping) for BM case 2 prediction?**   **Observation 14: For RAN 4 testing we should have the following goals:(1) Provide confidence that if a DUT passes the test it will also perform well in the field (2) Ensure reliability of using synthetic channels for test data in evaluating models trained on real data.**  **Observation 15: Rather than simulating CDL channels and [64-128] Tx beams using multiple angles of arrival (AoAs) and departure (AoDs), the spatial-selective and time-varying propagation conditions can be effectively represented by adjusting the Tx power at the test equipment (TE) for each beam at different times**  **Observation 16:** **To guarantee that the UE operates within acceptable margins, it's essential to subject it to various radio conditions and additional conditions for testing and generalization validation in RAN4**  **Observation 17:** **For BM use case the identified scenarios and configurations can be initially understood as those reported by UE through capability signaling as part of functionality identification.**  **Observation 18: The additional conditions for the AI/ML model training (which do not constitute part of UE capability) for the AI/ML-enabled feature/FG can serve as the different scenarios/configuration for defining generalization**  **Observation 19: Achieving consistency between training and inference by model monitoring could result in delays and increased complexity in model management for BM use case**  **Observation 20: If multiple models with varying generalization capabilities and requirements for network-side additional conditions are trained by different UE vendors, it would necessitate substantial standardization efforts for BM use case**  **Observation 21: Current proposals on assistance information for additional conditions and Model Identification only serve the purpose of selecting the appropriate AI/ML model. However, this approach may not be scalable due to considerations of UE implementation complexities and granualtity of conditions/additional conditions. Complexity can increase substantially, especially if condition granularity is fine.**  **Proposal 1**: **RAN4 should investigate the specification of reference AI/ML models for BM, the associated training procedures, and the training datasets to ensure alignment of simulation results. This will help pave the way for defining performance requirements for the BM use case, with careful consideration of constraints on model complexity. To initiate the discussion, we have provided a reference table as a starting point.**  **Proposal 2: We propose that RSRP accuracy be defined as the difference between the predicted RSRP and the measured (genie) RSRP associated with the same Tx beam. Accuracy requirements should apply to all predicted beams that satisfy the predefined side conditions, including SNR.**  **Proposal 3: When considering the necessity of additionally testing for beam prediction accuracy, we propose to study the additional information and significance that this test will provide, especially in light of our definition of RSRP accuracy.**  **Proposal 4: Use legacy measurement period as a starting point for Set-B for evaluation if sufficient to achieve good prediction accuracy.**  **Proposal 5: The absolute RSRP accuracy for AI/ML-based beam prediction is defined as the difference between the predicted L1-RSRP of beam index i and the ground-truth L1-RSRP of beam index i. Here, the index i can represent any beam within the Top-K beams based on predicted L1-RSRP.**  **Proposal 6: RAN4 should further evaluate whether beam prediction accuracy should be established as a metric/KPI for defining beam management requirements**  **Proposal 7: If RAN4 agrees that beam prediction accuracy should be established for defining requirements a beam rank metric should be used that considers the cumulative beak ranking other. Potential options are:**   1. **Kendall/Spearman correlation** 2. **Normalized Discounted Cumulative Gain (nDCG)** 3. **Rank-biased overlap**   **Proposal 8: For UE side model known TCI state is determined based on Set B transmission and UE report of top-K.**  **Proposal 9: For NW side model -  (1) TCI state is unknown if RS of target beam is not transmitted to UE prior to TCI state switch (2) If RS of target beam is transmitted to the UE and reported, legacy definition of known TCI state is applicable**  **Proposal 10: NW side data collection, QCL type D is signaled between set A and set B RS.**  **Proposal 11: For NW side model, TCI state is known if QCL source is signaled in both data collection and inference (to ensure UE Rx consistency), it is known. Otherwise, it is unknown.**  **Proposal 12: (for Alt 1): Define UE capability for Alt 1 of UE implementation. When UE does not support this capability, the TCI state is unknown. Unknown TCI state switching delay is applied.**  **Proposal 13: (for Alt 2): When gNB signal the QCL relationship between set A and set B, in training RS configuration, inference RS configuration and performance monitoring RS configuration (example of Alt 2 NW beam implementation of set A and set B),**  **The TCI state is known when the following is met:**  **Target TCI state is based on RS from set A beams with QCL relationship configured to a set B beam**  **The time between TCI state switch command and transmission of the QCLed RS from set B shall not exceed X ms**  **X = 1280 ms (legacy)**  **UE shall send measurement report for the QCLed RS from Set B between transmission and TCI state switch**  **TCI state and SSB of TCI state remain detectable during TCI state switching period**  **Otherwise, it is unknown.**  **Proposal 14:** **RAN4 should agree that the training dataset used in testing is strictly for testing purposes. In real-world applications, it will be the responsibility of UE implementations to determine the appropriate training dataset. This could involve using a vendor-specific dataset or creating a mixed dataset sourced from multiple vendors for training their models.**  **Proposal 15: To define the testing data for BM and provide sufficient test environments/conditions for vendors to generate the necessary training data, RAN4 should select a set of standardized test conditions. UEs will be tested across these identified conditions, and it will be up to the UE implementation to either switch between models or employ a supermodel capable of accommodating all these conditions during. Example of test conditions/configurations are:**   * **Details of the channel fading characteristics (CDL, LOS/NLOS, etc)** * **Different paterns of set B/ set A** * **Tx codebooks,Tx antenna architecture layout, antenna spacing** * **SNR conditions** * **Carrier Frequency** * **Doppler conditions, measurement window configurations and window of predicted time instances (BM case 2)**   **Proposal 16: For all potential non-tested conditions that are part of possible deployment scenarios, including additional conditions not explicitly addressed during RAN4 testing, we propose to employ Post-Deployment Procedures. These procedures will augment RAN4 conformance testing to effectively manage and assess performance across all possible deployment conditions**  **Proposal 17: RAN4 should consider adopting a realistic approach to model RF errors, as the prediction accuracy difference of the AI/ML model varies significantly between the worst-case scenario (independent RF errors) and the best-case scenario (same RF error for all Tx beams). To study the impact of measurement errors, the method of RF error modeling should be included in RAN4 discussions. Potential approaches include:**  **Using a truncated uniform distribution;**  **Using a truncated Gaussian distribution.**  **Proposal 18: To avoid the extra complexity of managing associated IDs across infra-vendors, develop a framework where the NW can equalize the differences between its own proprietary codebook and a RAN4 specified codebook for training the model. The network could employ a proprietary RSRP Equalizer to align UE-based predictions with those that account for the network's specific codebook (CB) implementations.**  **Proposal 19: Multi AoA based Testing Setup for BM: Evaluate the testability requirements for simulating time-varying input power to a sparse probe layout based on a simplified channel CDL model. The key aspects to be considered are:**   1. **How to determine the minimum number of clusters to be emulated without introducing bias in the results. Investigate a quantitative “goodness” criterion to selecting the number of probes/clusters.** 2. **Consider incorporating UE rotation for the test** 3. **How many different CDL channels we need to support (considering generalization purposes as well)**   **Proposal 20: To establish a criterion for assessing the feasibility of a simplified sparse probe layout (Option 5) for beam management (BM) testing, the following procedure can be considered:**   1. **Train a model using the reference CDL channel and test it with the simplified sparse layout-based CDL channel.** 2. **Train a model using the simplified CDL channel and test it with the reference CDL channel model.** 3. **Train and test a model using the reference CDL channel.**   **A similarity metric should be defined to compare the RSRP accuracies across these test cases. This metric would indicate the effectiveness of testing with the simplified CDL model. The similarity metric can be derived through simulations.**  **Proposal 21: For BM-Case 1 spatial prediction and for verifying RSRP accuracy we propose to use the testing setup and channel emulator functionality as described in Fig 6. (multiple static configurations)**  **Proposal 22: For testing the beam prediction/RSRP accuracy for BM-Case 2 temporal prediction we propose the framework described in this section and visualized in Fig 7, through a non static configuration of the channel emulator.**  **Proposal 23: RAN4 to consider the set of procedures/steps shown in the flowchart for adopting a single AoA BM testing setup**  A screenshot of a computer  AI-generated content may be incorrect.  **Fig. 9: Proposal for single AoA BM testing setup (flowchart)**  **Proposal 24: RAN4 to evaluate the tradeoff between testing complexity (number of different channel realizations for power measurements, different number of parameters that need to be emulated in the fading simulator) and establishing accuracy on the test metric (top-K) beam prediction accuracy**  **Proposal 25: RAN4 to evaluate if a power measurement should be performed with a single Rx beam or the optimum Rx beam should be selected from an Rx codebook to measure the power of set B.**  **Proposal 26: RAN4 to discuss Option2 (a Reference Complex UE Rx Beam/Antenna pattern) for emulation of Rx patterns in the fading simulator**  **Proposal 27: Distinguish between the descriptions of power based CDL channel model and the original complex domain CDL channel model**  **Proposal 28: RAN4 to evaluate the prediction performance difference between a reference complex UE Rx beam/antenna pattern and its own complex pattern to ensure that testing will provide confidence that UE will perform well in the field if it passes the test**  **Proposal 29: In the RAN4 core requirement, it is mandated that the consistency or association between Set B beams and Set A beams during both training and inference must be guaranteed. This serves as additional assistance information from the network side for testing UE-sided models during inference and monitoring**  **Proposal 30: RAN4 should define identified scenarios/configurations associated with the UE capability report of an AI/ML-enabled Feature FG. For defining generalization tests, the additional conditions can serve as the other identified scenarios/configurations for the BM use case**  **Proposal 31: RAN4 should investigate the feasibility of providing assistance information for the additional conditions to aid generalization and consistency across training and testing when defining requirements. Other additional conditions that are not part of UE capability can be used to define generalization tests**  **Proposal 32:** **For additional conditions that cannot be shared due to proprietary concerns, RAN4 can explore the feasibility of using a virtual ID to indicate the specific conditions under which a model was trained. This approach would assist in the proper selection of UE models to support generalization. Additionally, RAN4 should identify which additional conditions should be exclusively reserved for generalization tests.**  **Proposal 33:** **RAN4 should analyze how changes in test conditions, such as propagation conditions, affect the performance of AI/ML-enabled features. This analysis will help define the appropriate number of requirements and tests needed for each AI/ML-enabled feature and its corresponding ID combination**.  **Proposal 34: RAN4 to consider option 1 as a baseline for achieving consistency between training and inference for both set A and set B for beam spatial and temporal prediction**  **Proposal 35: Consistency should be considered in RAN4 only for static conditions and not-static conditions should test the UE ability to adapt and employ diverse AI/ML models to match the conditions (generalization)**  **Proposal 36:  In order to ease the burden for testing models with different NW additional conditions, it would beneficial to train the UE-side model with mixed dataset from various gNB settings, thus reducing the number of AI/ML models (selected by NW-side additional conditions) required to guarantee generalization and maintain the system performance for BM use case**  **Proposal 37: For achieving consistency between training and inference, if the model drifts due to misalignment of network-side additional conditions, the alignment could be achieved through an ID assigned by the network during training data collection. This ID indicates the association of the training data with the additional conditions implied to generate those data**  **Proposal 38: For achieving consistency between training and inference, if the UE lacks the ID but the network possesses the model ID, then the network can transfer the model to the UE, and the UE can update its list of models**  **Proposal 39: For achieving consistency between training and inference, if the UE has a model that partially supports the network-sided additional conditions, then the network can trigger data collection for fine-tuning the UE-sided model. Subsequently UE updates its list of model IDs.**  **Proposal 40: The UE updates its model list with a new model derived from either fine-tuning, model transfer from the NW, or monitoring to ensure the consistency of additional conditions. Then, the UE assigns an ID to the new model that supports the NW's additional conditions and shares this information with the NW. If some of the new conditions/configurations are standardized, the UE updates its capability report accordingly.**  **Proposal 41: If there is no ID available to associate training data with additional conditions, and monitoring procedure fails to guarantee the consistency of the model with the additional conditions then UE should fallback to legacy mode.**  **Proposal 42: Investigate the feasibility of enhancing the generalizability of the AI/ML model and reducing the number of AI/ML models and the testing burden for the beam management case by supplementing the core AI/ML input signals with both network (NW) and UE auxiliary information signals integral to its inference engine**    **Proposal 43: Investigate the feasibility of training the models with a mixed dataset associated with both network (NW) and UE auxiliary information signals to further enhance the generalizability of the AI/ML model for the beam management case and reduce the number of generalization tests.** |
| [**R4-2503553**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503553.zip) | Qualcomm Incorporated | **Observation 1: Option 2’s drawback lies in the fact that its pass/fail criteria don’t depend on whether the top predicted beams are adjacent to the strongest beams or not.**   * **If TE configures the setB and setA beams to be too close to each other and the actual measured RSRP of more than topK beams of setA lies within the range of UE’s measurement error, UE may not be able to pass option 2 focused test.** * **Note: The maximum value of K depends on the maximum number of beams that UE can report for measurement and inference. The max value of K is decided by RAN1/RAN2 and RAN4 cannot control it.**   **Observation 2: Option 3 does not suffer from the constraints of option 2 that got mentioned in observation 4. Even if more than topK beams of setA lies within the range of UE’s measurement error, UE will be able to pass option 3 tests as long as the measured RSRP of its predicted beam lies within X dB of the strongest beam.**   * **RAN4 can control the value of X dB and ensure that UE does not fail a test due to RAN4 allowed measurement error.**   **Observation 3: 3GPP is considering L1-RSRP accuracy as a potential prediction metric for AI-ML based beam management. UE has to know its RX beam to predict L1-RSRP accuracy of a particular gNB TX beam.**  **Observation 4: To compare between gNB beams of setA, UE needs to have some idea of the Rx gain associated with them. UE has to know the Rx beam of the predicted gNB beam even if it only has to predict top-1 (%) or top-K (%) beams.**  **Observation 5: In “wide-to-narrow” beam prediction, the setA CSI-RS resources whose properties that UE needs to predict are QCLed wrt type D with the setB SSBs that UE has measured. In order to receive a CSI-RS of setA beams after TCI state switch, UE can use the wide RX beam that it needs to measure the SSB that is QCLed with the CSI-RS.**  **Observation 6: In “narrow-to-narrow” beam prediction, if SSB resources are configured during the test, UE can use the RX beams corresponding to the SSBs as the RX beam for the setA CSI-RS resources if following conditions are satisfied:**   * **CSI-RS resources are configured without “repetition = ON” and QCLed wrt type D with SSB that are present during the tests.**   **Observation 7 Baseband error will reduce at high SNR.**  **Observation 8: RF error may not decrease at high SNR because UE’s Rx AGC settings might be calibrated for a particular SNR value.**  **Observation 9: Ground truth will be based on UE’s report and UE’s reported RSRP will include RF error that may not decrease with SNR.**  **Observation 10: At -3 dB SNR, the distribution of absolute baseband measurement error can be fitted with following Gaussian distributions:**   * **AWGN: Gaussian with zero mean and sigma = 0.6 dB** * **TDL-C: Gaussian with mean = -0.12 dB and sigma = 0.85 dB**   **Observation 11: The existence of relative L1-RSRP accuracy tests suggest that the total measurement error, modelled as the summation of baseband and RF error, of a particular SSB is correlated with that of the strongest SSB. This implies that the RF measurement error of a particular SSB is correlated with that of the strongest SSB.**  **Observation 12: It is not straightforward to model the exact dependency of RF measurement error of other SSBs on that of the strongest SSB.**  **Observation 13: It would be reasonable to use option 1 of RF error modeling, i.e., independent RF error for each TX beam, to generate initial simulation results for alignment among companies.**  **Proposal 1: RAN4 defines beam prediction accuracy requirement for option 3.**  **Proposal 2: After TCI state switches to a setA beam, spec mandates UE to “know” the RX beam of that setA beam in, at least, the following scenario:**   * **setA CSI-RS resources are configured without “repetition = ON”, SSB resources are configured during the test and setA CSI-RS resources are QCLed wrt type D with SSBs** * **FFS: UE’s RX beam knowledge in other scenarios.**   **Proposal 3: RAN4 investigates performance where training data includes measurement error.**  **Proposal 4: RAN4 investigates performance where ground truth during training and testing include measurement error.** |
| [**R4-2503606**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503606.zip) | MediaTek Inc. | **Observation 1**: Both DL Tx beam prediction and beam pair prediction are studied in R18 AI/ML SI, but only DL Tx beam prediction is in the scope of R19 AI/ML WI.  **Proposal 1**: Rx beam of the predicted Tx beam which is not measured can be known if the beam is QCL Type-D to a Tx beam in Set B.  **Proposal 2**: Whether Rx beam of the predicted Tx beam which is not measured can be known can be up to UE capability.  **Observation 2**: In terms of Top K/1 prediction accuracy, the performance of AI/ML BM case 1 models for different scenarios may have much difference.  **Proposal 3:** Further study and discuss whether it is possible to define scenario-agnostic performance requirements for AI/ML BM. If not, maybe we can seek for similar solutions as TDCP, i.e., define the test cases for limited scenarios instead of defining scenario-agnostic accuracy requirements.  **Proposal 4**: Further discuss which scenario among the following to define requirements and test cases for:   * Scenario 1: Set A and Set B are different (Set B is NOT a subset of Set A). E.g., use SSB beams to predict CSI-RS beams.   + Scenario 1a: RSRP with best Rx beam for each Tx beam in set B   + Scenario 1b: RSRP with specific Rx beam * Scenario 2: Set B is a subset of Set A. E.g., use part of SSB beams to predict all the SSB beams.   + Scenario 2a: RSRP with best Rx beam for each Tx beam in set B   + Scenario 2b: RSRP with specific Rx beam   **Proposal 5**: Not to define separate requirements or test cases for beam pair prediction in R19 WI.  **Observation 3**: The performance degrades a bit when all the UEs are outdoor UEs if measurement error is considered while the performance degrades significantly when most are indoor UEs.  **Observation 4**: Filtering out the data exceeds TE’s maximum Tx power has little impact on the prediction accuracy.  **Proposal 6**: Discuss which UE distribution to be used in the test and to be defined requirements for, e.g., 80% indoor+20% outdoor UEs or 100% outdoor UEs.  **Proposal 7**: When defining the accuracy requirements for AI/ML BM models, the impact of measurement error should be considered.    **Proposal 8**: When defining the accuracy requirements for AI/ML BM models, the impact of the dynamic range of TE should be considered.  **Proposal 9**: If RSRP and beam ID are reported, use RSRP accuracy and Top-K/1 (%) “the percentage of the Top-1 strongest beam is one of the Top-K predicted beams" as the KPI. If only beam ID is reported, use Top-K/1 (%) “the percentage of the Top-1 strongest beam is one of the Top-K predicted beams" as the KPI.  **Proposal 10**: Activation delay of AI/ML BM model shall also be counted into the prediction delay for aperiodic reporting and the first report of semi-persistent reporting, FFS the value.  **Proposal 11**: For AI/ML BM, further discuss whether the performance under synthetic channels can reflect the performance of AI/ML models in the real field.  **Proposal 12**: For AI/ML BM, discuss what the testing goal is, to verify the model can work well in real field or just to test UE can train a model that works well under a specific scenario.  **Proposal 13:** If the testing goal is to verify the model can work well in real field, and the performance under a specific channel/scenario cannot reflect the performance in the real field, it is necessary to design a test system that is possible to use field data in the test.  **Proposal 14**: In AI/ML BM test, emulate the spatial-selective and time-varying in propagation conditions through adjusting Tx power at TE on each beam at different time.  **Observation 5**: Single AoA test system can be applicable to all kinds of synthetic channels as well as using real field data. It can be easily extended to different scenarios and settings.  **Observation 6**: With single AoA, TE can know the expected RSRP of Tx beams at UE baseband.  **Proposal 15**: Use existing IFF or enhance IFF test systems with single AoA to test R19 DL Tx beam prediction models.  **Proposal 16**: Standardized UE radiation pattern in 38.901 can be assumed in R19 AI/ML BM test with single AoA test system instead of using test UE’s radiation pattern.  **Proposal 17**: The procedures of AI/ML BM test with single AoA:  Beam lock at test UE during the test.   * Step 1: Calibration   + TE transmit the signal with a fixed power, UE receive signal and report RSRP to TE. TE can get the relationship between TE’s transmit power and the reported RSRP. This can be done multiple times to reduce the impact of noise. * Step 2: Main test   + TE get the test dataset. The dataset can be early prepared through computer-like simulation using standardized UE radiation pattern in 38.901 and is the same for all the test UEs.   + TE transmit the calibrated power on each Tx beam in set B based on the test dataset. UE measures set B and predicts Top K or RSRP of set A.   **Proposal 18**: If to use multiple AoAs test system for AI/ML BM test, UE rotation should be considered.  **Observation 7**: The time needed to rotate UE in FR2 MIMO OTA test systems, i.e., about 1s, is not acceptable for AI/ML.  **Proposal 19**: The impact on performance of AI/ML BM models by using multiple AoA test systems should be studied at first before concluding to use multiple AoAs test system.  **Proposal 20**: How to avoid UE cheating in the test, such as reports fake values according to the previous predicted results, shall be further discussed to use the reported RSRP measurement as ground truth.  **Observation 8**: The upper bound of SNR with rough beam are much lower than that with fine beam.  **Proposal 21**: Fine beam should be used in AI/ML test.  **Observation 9**: Among all the test methods, only using peak fine beam can make sure the lowest SNR above -6dB with 30dB dynamic power range.  **Observation 10**: Multiple AoA test systems cannot use peak fine beam only.  **Proposal 22**: Using peak fine beam in AI/ML BM test with single AoA test system.  **Observation 11**: SNR dynamic range of multiple AoA test system would be at least 12dB smaller than single AoA test system. |
| [**R4-2503701**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503701.zip) | Tejas Network Limited | **Observation 1: If the gap between the best beam and the second-best beam is large, the UE may fail to meet any of the requirements in the test.**  **Observation 2: An unknown TCI state change can result in increased delay during the state transition.**  **Observation 3: When K is large, the likelihood of the Top-1 strongest beam becoming part of the Top-K predicted beams increases.**  **Observation 4: Option 3 alone is insufficient, as it only validates a single beam.**  **Proposal 1: The definition should capture the idea of relative L1-RSRP accuracy, not relative L1-RSRP values.**  **Proposal 2: The relative RSRP accuracy for AI/ML-based beam prediction = Absolute L1-RSRP Accuracy of the predicted beam index i – Absolute L1-RSRP Accuracy of the best-predicted beam.**  **Proposal 3: TCI state associated with the predicted beam should be known.**  **Proposal 4: RSRP accuracy should be one of the KPIs.**  **Proposal 5: For Option 2, the value of K should be kept small to ensure more meaningful validation and reduce the likelihood of the Top-1 strongest beam easily fitting into the Top-K predicted beams.**  **Proposal 6: Updated Option 2: Top N/K (%): Top-N/K (%): The percentage of the Top-N predicted beams that are among the Top-K strongest beams.**  **Proposal 7: RSRP accuracy should be included as one of the KPIs, combined with the updated Option 2 and Option 3.** |
| [**R4-2503711**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503711.zip) | Anritsu Corporation | ***Observation 1: RAN4 needs to clarify all the necessary conditions of associated test setup for both BM-Case 1 and Case 2 before down-selecting the test setup from 1AoA and multiple AoA.***  ***Observation 2: There are currently three kinds of presumable UE patterns from the AI model input PoV.***  UE pattern 1: No need to input Rx Beam information in the AI model.  In this case, it is expected that the UE Rx beam is pointing towards the direction of the measurement probe in an OTA chamber where the DL signals are transmitted.  UE pattern 2: One Rx Beam information is input to the AI model.  In this case, direction of the UE Rx beam is same as UE pattern 1 and that Rx Beam information is input to the AI model. TBD if the actual Rx beam pattern is necessary or it can be standardized.  UE pattern 3: Multiple Rx Beam information is input to the AI model.  In this case, for each Tx Beam ID in set B, UE will change its Rx beam direction to multiple directions and multiple L1-RSRP measurements are made per Tx Beam ID. Multiple Rx Beam information is input to AI model per Tx Beam ID.  ***Observation 3: BM-Case 1 tests with UE pattern 1 and 2 (i.e. less than or equal to one Rx Beam information is input to the AI model) can be conducted with the following conditions.***   * ***Rx beam pattern is standardized or,*** * ***Rx beam pattern is measured during the evaluation.***   ***Observation 4: During the test with UE pattern 3 in the MPAC, UE Rx beam would be swept to multiple directions even for the measurement period of one Tx Beam ID, where the detailed number of Rx Beam ID to be used for the measurement would be up to the UE implementation.***  ***Proposal 1: Agree following conditions to conduct the BM-Case 1 test with UE pattern 3 by 1AoA test setup***  Condition 1: Allow some time to move UE positions by a positioner. For example, by introducing a test mode to set a gap to run measurements of L1-RSRP, or signaling is introduced to pause and resume L1-RSRP measurements while the positioner is rotating.  Condition 2: Rx beam directions for the measurement are declared by UE vendors.  Condition 3: Like the UE pattern 1 and 2, Rx beam profile is standardized.  ***Observation 5: Measurement procedures with UE pattern 3 by 1AoA test setup would be as follows.***  Step 1: Load the Tx Beam IDs from Set B in TE.  Step 2: Rotate the positioner to one of the declared measurement directions.  Step 3: Set one of the loaded Tx Beam ID and its associated CDL channel model in TE.  Step 4: Establish the wireless (virtual) cable connection between the UE and TE.  Step 5: Create a composite DL signal from each cluster under the conditions of Step 3 on the fading simulator and transmit it from the measurement probe.  Step 6: Measure the L1-RSRP in UE.  Step 7: Repeat Step 3 to Step 6 based on the next Tx Beam ID as soon as the L1-RSRP measurement for the first Tx Beam ID is completed.  Step 8: When the measurement reaches the maximum Tx Beam ID x of Set B, rotate the positioner to the next declared measurement direction, and repeat Steps 3 to 7.  Step 9: Repeat from Step 3 to 8 until it reaches the total number of declared measurement directions.  Step 10: UE inputs the obtained Tx Beam ID, Rx Beam ID and L1-RSRP to the AI model.  Step 11: The AI model answers the TE with the best Tx Beam ID (Top-k or Top-1, etc.) as predicted by the AI model.  TBD how the UE triggers the predictions by AI or how the predicted best Tx Beam IDs are signaled.  Step 12: While switching all Tx Beam IDs as Set A, measure L1-RSRP in the same way as for Set B. However, the combination that has already been measured as Set B is skipped.  Step 13: From all the measurement results of Set A, the Tx Beam ID with the largest L1-RSRP is determined and compared with the estimation results sent in Step 11) to make a pass/fail decision.  ***Proposal 2: Allow different types of test systems depending on the UE patterns on condition if the group could confirm that the tests with the multiple AoA test setup with simplified CDL condition can ensure the AI performance in the field. i.e. 1AoA test system can be applied to the UE pattern 1 and 2 in Observation 2 and the multiple AoA test system can be applied to all the UE patterns.***  ***Observation 6: Link level simulation in RAN1 appears to apply the assumption of the channel model described in TR 38.901 clause 7.6.3 [4], i.e. Spatial consistency.***  ***Observation 7: When considering the test method of BM-Case 2 in RAN4, we need to study how we can create the channel environment by the candidate test setups (1AoA or multiple AoA) in a case that the UE moves from one point to another.***  ***Observation 8: There is a benefit that 1AoA test system with the fading simulator can simulate the channel model even with the case that UE moves, though there is a limitation that it cannot change the Rx Beam ID in a short period.***  ***Proposal 3: Companies are encouraged to provide a methodology on how MPAC can reproduce the channel environment of the moving UE condition for BM-Case 2 under the situation where probes are all fixed.*** |
| [**R4-2503768**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503768.zip) | CMCC | ***Proposal 1: it is proposed that the RSRP accuracy requirements apply to all predicted/reported beams.***  ***Proposal 2: the relative RSRP accuracy is proposed as:***   * ***Relative RSRP accuracy = (predicted L1-RSRP of beam index i - predicted L1-RSRP of beam index n) - (ground truth of L1-RSRP of beam index i - ground truth of L1-RSRP of beam index n), where the beam index n owns the largest predicted value.***   ***Proposal 3: it is proposed that the beam ID prediction accuracy is defined as Top-K/1 (%) : the percentage of "the Top-1 strongest beam is one of the Top-K predicted beams"***  ***Proposal 4: for measurement period for inference, it is proposed to take legacy L1-RSRP measurement period as baseline.***  ***Proposal 5: for beam management, it is proposed to reuse legacy L1 report mapping, i.e. L1 SS-RSRP part of Table 10.1.6.1-1 and Table 10.1.6.1-2 in TS38.133 are reused for the report of inference results.***  ***Proposal 6: for measurement error impact evaluation, it is proposed to consider a single error covering both baseband error and RF error. And it is proposed to consider baseband error under different SNR.***  ***Proposal 7: to verify beam management performance, it is proposed to use multiple AoAs test systems.***  ***Proposal 8: for beam prediction testing, it is prefered to consider CDL, and it is also OK with simplified CDL (e.g. simplified CDL with small number of clusters).*** |
| [**R4-2503786**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503786.zip) | NTT DOCOMO, INC. | **Proposal 1: RSRP accuracy of beam prediction is defined as the acceptable difference between ground truth of the predicted beam and predicted RSRP value. The ground truth of the predicted beam is the ideal measurement of RSRP on the predicted Tx beam.**  **Proposal 2: the requirement should be tightened than current absolute RSRP accuracy because prediction can include hardware characteristics, such as implementation margin.**  **Proposal 3: Beam ID prediction accuracy should be 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 and x should be less than 3.** |
| [**R4-2503789**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503789.zip) | ZTECorporation,Sanechips | **Observation 1: The top-K beams are reported or predicted beams or top-K beams are groundtruth beams.**  **Observation 2: For the top-K beams are groundtruth beams but not the predicted beams, may be some UE reported beams could not be verified or why should we require the RSRP accuracy of the beam that the UE doesn’t even report.**  **Observation 3: What is the standard to verify the performance of AI system based on the absolute RSRP accuracy shall be considered or how to quantify the the small or large L1-RSRP difference shall be studied.**  **Observation 4: One question is that the proposed margin shall be maintained or smaller than legacy value. There are two options to consider the change of proposed margin:**   * **Option 1: Maintain the legacy measurement accuracy. In legacy, UE shall do the full beam sweeping in order to confirm the best Rx beam to receive the downlink RS and get the better communication with network. The AI system has already improve the overhead and reduce the complexity than legacy. Thus, maintain the legacy measurement accuracy could be understood.** * **Option 2: Smaller than legacy measurement accuracy. This is a more direct understanding since the AI performance shall have the better performance than legacy and the measurement accuracy shall be tightened. In this way, the performance of AI method could be displayed.**   **Proposal 1: For the absolute RSRP accuracy, the index i shall be the top-K beams based on the predicted beams or UE reported beams instead of the groundtruth beams.**  **Proposal 2: One margin for the absolute RSRP accuracy shall be studied in order to verify the performance of AI system. The proposed margin could be equal to and smaller than the legacy measurement accuracy requirements.**  **Observation 5: The relative RSRP accuracy could remove the bias and choose the optimal reported beam.**  **Proposal 3: The relative RSRP accuracy shall be considered for RSRP accuracy**.  **Observation 6: For AI beam management, only serving cell downlink Tx beam prediction is in the scope.**  **Observation 7: The groundtruth beams are based on the UE measurement which is more reliable than predicted beams.**  **Observation 8: The largest ground truth shall be as the benchmark for definition of relative RSRP accuracy.**  **Proposal 4: The formula of relative RSRP accuracy could be:**  ***Relative RSRP accuracy=predicted L1-RSRP of beam index i - ground truth of the beam index n ,***  ***where the beam index n owns the largest value***  **Observation 9: For Top-1(%), the test may be failed because of smallest prediction errors.**  **Observation 10: For Top-K/1(%), this metric will cause the mistake because of the small difference value between the best and the second ground truth beams. If the number of K is larger and larger, the probability of the best ground truth beam is in top-K predicted beams is larger and larger.**  **Observation 11: For Top-1/K(%), this metric only consider the predicted best beam, but the other predicted beams are not guaranteed, the whole system may be degraded.**  **Proposal 5: For beam prediction accuracy, Top-1(%) and Top-1/K(%) shall be set as low priority.**  **Proposal 6: For Top-K/1(%), the K shall be relaxed to K≥3.**  **Proposal 7: The Top-K/1(%) could be used together with the RSRP accuracy to verify the performance of AI system.**  **Observation 12: Option 3 only considers the predicted best beam, but the other predicted beams are not guaranteed. Although the main purpose for AI/ML beam is to select the best beam, the whole system may not work in the field and the whole system will not be ensured.**  **Proposal 8: At least the option 3 could not be precluded which could be set as low priority.**  **Proposal 9: The potential components in prediction delay contains three parts:**   * **Measurement delay: The time for measurement of SetB which is the input for inference.** * **Inference delay: The time for inference from starting performing inference to generate the prediction results.** * **Reporting delay: The time between the report triggering and the point when the UE starts to transmit the outputs over the air..**   **Proposal 10: The legacy L1-RSRP measurement period shall be as baseline or starting point.**  **Proposal 11: Postpone discussion on reporting requirements until other WGs reach the conclusion on reporting mechanism.**  **Observation 13: Different TCI state conditions mean the different TCI state switching delay.**  **Observation 14: When the UE is aware of the current RX beam it is utilizing, TCI state activation delay is reduced. Otherwise, TCI state activation delay is not reduced.**  **Observation 15: There are two situation for considering the TCI state activation:**   * **Case 1: The best predicted Tx beam ID/RSRP belongs to Set B.** * **Case 2: The best predicted Tx beam ID/RSRP doesn’t not belong to Set B.**   **Proposal 12: UE knows the Rx beam or the RS is in the Set B and the RS is also the L1-RSRP report by UE within 1280ms before TCI state activation, the TCI state is known. Otherwise, the TCI state is unknown.**  **Proposal 13: The test procedure shall be clarified and unified firstly:**   |  |  | | --- | --- | | **Training phase** | **1.TE configure resource sets for measurement of Set B and beam report of Set A** | | **2.UE measures Set B as the model input.** | | **Inference phase** | **3.Generates the test setup environment for Set B.** | | **4.UE predicts best Tx beam/ L1-RSRP in Set A.** | | **5.UE reports predicted best Tx beam/ L1-RSRP to TE in Set A.** | | **Groundtruth phase (measurement of RSRP with high SNR)** | **6.TE configures resource sets for measurement of Set A.** | | **7.Generates the test setup environment for Set A (including high SNR).** | | **8.UE measures Set A under high SNR condition as ground truth.** | | **9.UE reports groundtruth (measurement on Set A) to TE.** | | **TE evaluation phase** | **10.Evaluates the predicted results and groundtruth and judge the performance degradation** |   **Observation 16: The smaller SNR conditions, the larger measurement error.**  **Observation 17: The performance degradation will be occurred due to added measurement error.**  **Observation 18: For Top-1, the prediction accuracy is below 86% even if there is no error on test data.**  **Observation 19: The prediction is getting larger and larger when K is larger and larger whatever the measurement error is added or not.**  **Proposal 14: RAN4 shall further study the measurement error impact on different performance metrics if new metrics are introduced.**  **Proposal 15: Except measurement error impact, RF error influence shall also be considered in RAN4.** |
| [**R4-2504023**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504023.zip) | vivo | **Proposal 1: On the definition of prediction delay requirements for BM, RAN4 to consider:**   * **For Measurement for prediction, the current measurement period for L1-RSRP measurement can be the baseline and further discuss the details by considering the parameters, which at least include:**   + ***The number of samples for set B measurements (for spatial domain prediction); The number of the latest measurement instances and the time interval between two adjacent time instances (for time domain prediction)***   + ***beam sweeping factor***   + ***sharing factor N*** * **For inference delay: the reference latency and possible processing time**   **Proposal 2: For the UE Rx beam knowledge,**   * **at least for the case that the Top-K beams to be reported are in set B and set B is a subset of set A, if beam sweeping has been done during measurements on set B, UE Rx beam information is known by UE and no further beam sweeping is needed.** * **for other cases, e.g., set B is not a subset of set A, RAN4 to further discuss how to obtain Rx beam information, e.g., introducing UE capability or making the reduced beam sweeping factor assumption**   **Proposal 3: For performance monitoring for BM, RAN4 to first determine the testing goal. The following alternatives can be considered:**   * **Option 1: The testing goal is to verify whether the performance metrics can be correctly calculated by UE;** * **Option 2: The testing goal is to verify whether UE can perform monitoring and report the performance metrics within a certain latency upon performance monitoring is required;** * **Option 3: The combination of Option 1 and Option 2.**   **Observation 1: There is a certain probability that the Top-K predicted beam is not any of the Top-K beams based on ground-truth of RSRP.**  **Observation 2: For RSRP Absolute accuracy, if the index i can be any beam index in top-K beams based on ground-truth L1-RSRP, the RSRP accuracy for those beams which belong to Top-K predicted beam but not any of the Top-K beams based on ground-truth RSRP cannot be guaranteed.**  **Proposal 4: RAN4 to define absolute RSRP accuracy as one of metrics for beam management requirements/tests, the corresponding definition is:**   * ***The absolute RSRP accuracy = predicted L1-RSRP of beam index i – ground-truth of L1-RSRP of beam index i. The index i can be any beam index in top-K beams based on predicted L1-RSRP.***   **Observation 3: RF errors are more detrimental than BB errors, especially for beam-specific predictions (comparision between predicted beam and acutal beam with the same ID ).**  **Observation 4: Larger K improves resilience but may not compensate for RF-dominated errors (e.g., Case 3 in Top-5 still shows >6 dB difference).**  **Observation 5: The KPI of Top K/N mitigates the effects that the RSRPs between Top 1 and Top 2 genie-aid beams are too close (within 0.5dB), which may cause the actual Top-1 beam may be indistinct**  **Observation 6: Compared with the KPI of “the percentage of the Top-N genie-aided beam is ONE of the Top-K predicted beams”, the KPI of “the percentage of the Top-N genie-aided beam is TWO of the Top-K predicted beams” may better reflects beam reliability of the reported beams**  **Observation 7: For the metric- maximum ideal RSRP among top-K predicted beams is larger than the ideal RSRP of the strongest genie-aided beam – 1 dB, there is minimal degradation when only BB error is considered. When both RF and BB errors are considered, the performance is similar to RF error-only, suggesting RF error dominates the impact on this KPI for prediction accuracy**  **Observation 8: The simplified CDL channel with small number clusters will degrade the test dependability.**  **Observation 9: The number of probes can be greatly reduced by grouping different Tx beams together. Only one probe or one set of probes can emulate different Tx beams in one group.**  **Observation 10: Different test system with reduced cost and complexity can be introduced by using the Tx beam grouping method.**  **Proposal 5: RAN4 to consider methods to group Tx beams together to reduce the number of probes in the test system.**  **Proposal 6: RAN4 to consider the following procedures to simplify the test system:**   * **Step 1. Generate the BS beamforming power pattern of every Tx beam** * **Step 2. Add the BS beamforming power pattern into the power of different clusters in specific ZoD and AoD, and derive the new power of every cluster** * **Step 3. Use the new power of clusters to generate the power angular spectrum (PAS) of ZoA and AoA** * **Step 4. Some Tx beams with the corresponding ZoA/AoA are to be grouped together based on a certain rule (e.g., the ZoA/AoA of the strongest power derived from the PAS in step 3 are the same angle or in a pre-defined range)** * **Step 5. Using one or small number of probes to emulate one beam group.** * **Step 6. Delete the unnecessary probes and merge some adjacent probes.**   **Proposal 7: RAN4 to consider the following two approaches for design the test system**  Proposal 8: RAN4 to consider the UE Rx beamforming resolution in angular domain to perform OTA probe allocations. |
| [**R4-2504055**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504055.zip) | Qualcomm Incorporated | **Observation 1: The key requirements for AI/ML BM test setup are listed following:**   * **Requirement#1: Channel model in the test**   + **CDL or simplified CDL to properly verify beam management performance.** * **Requirement#2: Avoid the cheating**   + **Progress A: Model training for Set A’s is generated with channel parameters #1. Set B in UE measurement should have enough randomness compared to Set A’.**   + **Progress B: Set A is generated to make sure to use the same parameters (initial phase, angular spread, delay spread) as the Set B in the Progress A measurement step. NW antenna/beam configurations for Set A and Set B should be deterministic. For some prediction metrics (e.g., option 2 and 3), UE can pass the test without actual measurement if TE does not randomize the transmission of its beams.** * **Requirement#3: Support the beam number of Set A**   + **TE needs to support Set A for ground truth verification.**   **Observation 2: Verifying the AI/ML beam management performance (e.g., prediction of top K beams) in a TDL channel with single AoA without spatial characteristics makes the test too simple and inconsistent with the actual field environment. Thus it is not feasible to general proper Set A and Set B with single AoA setup .**  **Observation 3: For single AoA test setup, there are following limitations:**   * **UE has to report the UE antenna pattern with X\*Y test points that is very time consuming.** * **UE needs to support additional test function for RSRP reporting** * **If Rx beam ID is one of the input for AI/ML model, additional test mode/signalling is needed** * **UE’s beam management performance can’t be verified properly since for each of UE position there is enough dewll time and UE beam is locked.**   **Observation 9: To generate Set A and Set B, multiple AoAs/probes test setup is necessary.**  **Observation 4: There will be plenty of additional probes if BS Tx beams point to other clusters even though some of the probes could be reused.**  **Observation 5: The UE’s antenna gain gap between peak and 50%-tile spherical coverage is over 10dB for PC3. The very weak clusters would not influence the UE beam management even if they reach the UE within the UE’s spherical coverage.**  **Observation 6: The AoDs/ZoDs of very weak clusters are unlikely to influence the selection of the strongest beam in any reasonably designed codebook of Set A and Set B.**  **Observation 7: The weak clusters that have a limited impact on the beam management could be removed to reduce the number of clusters in the CDL channel model.**  **Observation 8: The clusters that have the same/similar AoAs could be merged. And the intra-cluster angle spread, e.g., CASD, CASA, CZSD, CZSA can be reduced to simplify the test setup. And in this case, a new simplified channel model is emulated by test setup.**  **Observation 10: Using Multiple AoA test setup with FR2 RRM test as the baseline can legacy the FR2 existing test setup to save the test cost and a simplified CDL channel model needs to be specified.**  **Proposal 1: Single AoA test setup incluidng the use or non-use of the two-step method should not be considerd for AI/ML BM testing. RAN4 should consider the follwoing two multiple AoAs/probes options as the test seup for AI/ML beam management testing**   * **Option 1: Multiple AoA test setup with 3D MPAC as the baseline**   + **Consider reducing the cluster number with weak power level, merging the cluster with the same/similar AoA, and decreasing intra-cluster angle spread in CDL channel model to generte a new simplified CDL channel model**   + **Define the paramters for the new simplified CDL channel model based on that 3D-MPAC test setup is used as the baseline**   + **Companeis to evlaute whether the new simiplied CDL channel model be used to properly verify AI/ML BM perforance** * **Option 2: Multiple AoA test setup with FR2 RRM test setup as the baseline**   + **Conider multiple AoA test setup with FR2 RRM test setup as the baseline emulate compareble Tx beam with simplifed CDL, e.g., generating enough Tx beam number the solutions such as combinaiton of probes transmission and power offset for each of probe can be considered.**   + **Define the paramters for the new simplified CDL channel model based on that FR2 RRM test setup is used as the baseline**   + **Companeis to evlaute whether the new simiplied CDL channel model be used to properly verify AI/ML BM perforance** |
| [**R4-2504107**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504107.zip) | Ericsson | **Observation 1: RSRP in Option3 of Beam prediction KPIs is groundtruth RSRP, i.e., the maximum RSRP among top-K predicted beams and the RSRP of the strongest beam are groundtruth RSRP.**  **Observation 2: The term ‘successful rate’ in Option 3 of Beam prediction KPIs shall be removed, as Option3 represents a mandatory requirement that must be strictly satisfied.**  Observation 3: The L1-RSRP differences of SSB with the SNR value -3 is only a few tenths of dBs so its impact would be minor as the baseband (BB) error.  **Observation 4:** **Both ±4.5 dB RF error with Gaussian distribution and uniform distribution can bring large performance degradation on prediction accuracy especially for Top-1 beam, where the impact of uniform distributed RF error is even worse.**  **Observation 5: The metric/report type of prediction results, prediction accuracy or L1-RSRP offset, for prediction and the metric for monitoring may be different since they serve different purposes.**  **Observation 6: the measurement error in datasets for training/testing may be covered by the overall measurement error.**  **Proposal 1: Metrics/KPIs for beam management requirements/tests are outlined by the below options:**   * **In case of the AI/ML model with label (beam ID) output type**   + **Beam prediction accuracy, i.e., Option 2 + Option 3 is adopted.** * **In case of the AI/ML model with label (beam ID) + RSRP output type**    + **Beam prediction accuracy and RSRP accuracy, i.e., Option1, 2 an 3, are adopted simultaneously.**   **Proposal 2: In Option2 of Beam prediction KPIs, RAN4 considers the below option.**   * **Top-K/1 (%) : the percentage of "the Top-1 strongest beam is one of the Top-K predicted beams"** * **Not introduce the parameter N.**   **Proposal 3: Maintain the current wording of Option 3 of Beam prediction KPIs without modification, as no further improvements are necessary.**  **Proposal 4: Support Option 1 of the relative RSRP accuracy, relative RSRP accuracy= predicted L1-RSRP of beam index i – predicted L1-RSRP of beam index n, where the beam index n owns the largest RSRP among the predicted beams.**  **Proposal 5: For the absolute RSRP, the index i shall be any beam index in all beams based on predicted L1-RSRP as per UE reporting.**  **Proposal 6: Regarding measurement period for prediction, for both BM-case1 and BM-case2, the legacy L1-RSRP measurement delay shall be applied, with the assumption that the measurement accuracy remains consistent.**  **Proposal 7: RAN4 to study whether specific requirements are necessary to evaluate the capability of a UE in tracking changes in L1-RSRP through measurements in BM-case 2.**  **Proposal 8: The distribution of baseband error can be fitted with zero mean Gaussian distributions,** **the sigma value can be further discussed after the BB error distribution is confirmed.**  **Proposal 9: the definition of “90%-tile L1-RSRP difference” should be clarified, such as: 90%-tile L1-RSRP difference = max (abs (95%-tile L1-RSRP difference), abs (5%-tile L1-RSRP difference)).**  **Proposal 10: The performance degradation from ±4.5 dB RF error is not neglectable, further discussion is needed for whether accept this performance degradation or not.**  **Proposal 11: RAN4 shall define measurement delay for monitoring.**  **Proposal 12: RAN4 to study the two questions regarding multiple/single AoA:**   * **What is the (maximal) difference between really tested Tx/Rx beam and the artificial Tx/Rx beam?** * **If the difference can impact AI/ML model?**   **Proposal 13: RAN4 to check if the 32 beams after simplified CDL channel can be produced by 8 or 12 probes which may achieve the balance between performance and implementation complexity.**  **Proposal 14: Measuring Tx/Rx beams using a TDM approach may be a valid method; however, the control mechanism for Rx switching must be explicitly addressed.**  **Proposal 15: The simplified CDL channel only is considered to be applied in inference, no need to be applied in training.**  **Proposal 16: The metric to estimate the “goodness” of a simplified CDL channel model may be represented by the below steps:**   * **Step 1: Test (inference) with the reference CDL channel.** * **Step 2: Test (inference) with the simplified CDL channel.** * **Step 3: The prediction difference between the results of the two tests in the above steps is less than a threshold with [95] percentile.**   **Proposal 17: RAN4 to evaluate the SNR condition for the ground truth by simulation.**  **Proposal 18: Test procedure shall at least comprise the below three steps:**   * **Acquire the predicted RSRP under the required corresponding SNR condition** * **Acquire the ideal measurement of RSRP under high SNR condition** * **Compare the two RSRPs acquired in the above steps**   **Note 1: Test configurations are kept same in the steps except for different SNR conditions.**  **Note 2: Arranging the details on the sequence and test configuration may need further investigation in performance part.**  **Proposal 19: For BM-case1, same channel is applied for measurement and inference.**  **Proposal 20: For BM-case2, plural channel instants with a series of changes of the channel parameters are necessary to reflect temporal correlation between channels. At least, it includes:**   * **Generating a set of channel instants, from channel #a+1 to channel #a+m, for measurement, and** * **Generating a set of channel instants, from channel #b+1 to channel #b+n for inference**   **Where,**   * **The channels instants comprise sequent and grade updates of a set of parameters.** * **The numbers of m and n depend on the measurement window and prediction window of BM-case2.** * **The channel parameters may comprise one or more than one of**    + **Departure angles (AOD, ZOD)**   + **Arrival angles (AOA, ZOA)**   + **UE position and/or rotation**   + **Channel coefficient**   **Proposal 21: In order to avoid overfitting UE models to the test environment only, the test (including data sets) shall be able to demonstrate the robustness of the AI/ML model and match the practical field deployment conditions.** |
| [**R4-2504216**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504216.zip) | Samsung | RAN4 Core Requirements Impact for NW-sided Model  **Observation 1:** For NW-sided model, there is no RAN1 progress on the UE reporting mechanism since RAN1#118 (August 2024), and it is reasonable to postpone the discussion on RAN4 core requirement impact for NW-sided model.  RAN4 Core Requirements for Supporting UE-sided Model  *Data collection for UE-sided model (for model training)*  **Proposal 1:** RAN4 shall confirm that no RAN4 impact is expected from data collection for UE-sided model training.  *RRM core requirement for model Inference*  **Observation 2:** RRM requirement shall be introduced to guarantee UE-side AI/ML model inference to guarantee (1) the accuracy of inference results accordingly to the agreed KPI and (2) the delay of reporting inference results (containing the inference latency).  **Proposal 2:** For absolute RSRP accuracy requirement, the requirement shall be defined as:   * The absolute RSRP accuracy requirement shall be applied to all beams in the configured top-K predication. * The existing measurement accuracy requirement is used as the baseline for prediction accuracy.   **Observation 3:** There are two root causes for relative RSRP error: (1) Predicted relative RSRP between target beam and optimal beam is estimated wrongly, and (2) The optimal beam used as reference for relative RSRP is estimated wrongly.  **Proposal 3:** In Rel-19 AI-BM, RAN4 shall not introduce any relative RSRP accuracy requirement to any relative RSRP reporting.  **Proposal 4:** If RAN4 still can’t independently conclude on Option 2 and 3 for beam prediction accuracy metrics, RAN4 shall stop the discussion to wait RAN1 conclusion on the KPI for UE-assisted performance monitoring.  **Proposal 5:** For AI-BM Case-1, the legacy L1 measurement period is used as the starting point for Set-B measurement to evaluate the achievable prediction accuracy.  **Proposal 6:** For AI-BM Case-2, the exact number of the measurement periods/occasions shall be configurable and subject to further study in RAN1.  *RRM core requirement for performance monitoring*  **Proposal 7:** At least for performance monitoring Type 1, Option 2 (UE-assisted performance monitoring), RAN4 prediction accuracy requirement shall be applicable.  *RRM core requirement for LCM*  **Observation 4:** For UE-sided model configuration, after *CSI-ReportConfig* for inference configuration in *RRCReconfiguration* signalling in Step 5:  • aperiodic CSI Report and semi-persistent CSI report can be activated/triggered by NW after *RRCReconfigurationComplete*.  • periodic CSI Report is considered as activated after *RRCReconfigurationComplete*.  **Observation 5:** For UE-sided model activation/deactivation, L1 and MAC signalling can be used for aperiodic CSI Report and semi-persistent CSI report.  **Proposal 8:** For UE-sided model “Network decision, network-initiated” LCM, RAN4 shall study the necessity of delay requirement for model activation/deactivation, by adopting SCell activation/deactivation as baseline. |
| [**R4-2504230**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504230.zip) | Nokia | In this paper, we provide some additional views on issues related to RRM core requirements, as listed below:   * Measurement error impact and the ground truth * KPIs for beam prediction * Indicated/activated TCI state requirements * Measurement delay requirements * Impacts on requirements related to candidate beam detection * LCM and performance monitoring related requirements * Applicable functionality reporting * Testing framework for AI/ML based Beam Management   In the paper, the following Observations and Proposals were made:  **. Observation 1:** When no measurement error, low SNR UEs experience a high 90%-tile L1-RSRP difference and high SNR UEs experience a low 90%-tile L1-RSRP difference. The same trend is observed in Top-K/1 accuracy.  **Observation 2:** With measurement error, all UEs experience degradation in beam prediction-related KPIs. However, a severe impact is observed with high SNR UEs compared to low SNR UEs.  **Observation 3:** The performance degradation of AI/ML model in presence of Gaussian RF error is lower than that of Uniform distribution as most of the RF errors are concentrated near the mean value zero, and few RF errors appear further away from the mean.  **Proposal 1: RAN4 to clarify the impact of measurement error is higher for low SNR UEs or high SNR UEs, given the significant degradation in prediction performance observed with high SNR UEs.**  **Observation 4:** Based on evaluation assumption, the observed performance degradation of AI/ML model for Set B=8 and Set A = 32 when Set B is subset of Set A is as follows,  If the BB error model with LLS (# samples = 3, SINR = -4 dB) and RF error model with Gaussian with ±4.5 dB error margin (Same for all TX beams) are considered,   * Top-1/1 without considering some margin is 83%. * L1-RSRP difference is 6 dB.   **Observation 5:** The performance degradation of AI/ML model is high when RF error is added independently.  For example, if the BB error model with LLS (# samples = 3, SINR = -4 dB) and RF error model with Gaussian with ±4.5 dB error margin (Independent for each TX beam) are considered,   * Top-1/1 without considering some margin is 70% * L1-RSRP difference is 13 dB.   **Proposal 2: RAN4 to discuss the RF error model since a large difference of the prediction accuracy was observed with different RF models. RAN4 should consider also RF error model with independent error on each TX beam.**  **Observation 6:** Acceptable absolute RSRP accuracy for AI/ML BM should meet the existing legacy requirements to avoid performance loss  **Proposal 3: In order to meet existing legacy requirements, the measured RSRP accuracy for the input to the AI/ML model should be improved. Alternatively, the acceptable predicted RSRP accuracy for AI/ML BM can be tightened such that it is close to legacy requirement**  **Proposal 4: RAN4 should consider both absolute and relative L1-RSRP accuracies for the verification of AI/ML BM functionality in case of L1-RSRP prediction.**  **Proposal 5: Relative RSRP accuracy for AI/ML BM can be defined as:**   * **relative RSRP accuracy = (reported predicted RSRP of beam i – reported RSRP of beam j) – (reported RSRP of beam i – reported RSRP of beam j) where beam i is one of the top-K predicted beams and beam j is the beam with maximum measured RSRP** * **RAN4 to further discuss whether the reported RSRP of beam j can be predicted or measured or a combination of both predicted and measured samples.**   **Observation 7:** At this, we identify that the success rate of correct prediction instead of evaluating only a single condition that checks for the highest RSRP beam amongst the top-K predicted beams meeting certain threshold, the KPI should also verify whether other beams within the predicted beam set indeed satisfy such threshold requirements.  **Observation 8:** For a fair evaluation, the threshold for ranked predicted beam, could be set as RSRP of the corresponding ranked genie-aided beam – X. Hence, the condition to be checked for the ranked beam is , where X is a fixed value for all .  **Proposal 6: The threshold for ranked predicted beam, can be considered as the RSRP of the corresponding ranked genie-aided beam, – X**  **Proposal 7: The success rate of correct prediction can be computed as a weighted average that considers all top-K predicted beams**  Where the weights can be implementation specific. For example, we use  with RSRPs here in linear scale  NOTE: All the RSRPs and their ranking in the above are determined from the ground-truth values  **Observation 9:** The simulation results confirm that the weighted success metric is robust in identifying the performance impact due to errors in comparison to option 3  **Proposal 8: RAN4 to discuss adaptation of Option 3 with following changes:**   * **All top-K predicted beams have to be considered in the success rate evaluation** * **RSRP of each top-K predicted beam is to be compared with the RSRP of its equal ranked genie-aided beam - X**   **Observation 10:** The behaviour of beam prediction functionalities can be tested without explicit definition of the ground truth by comparing the predictions to each other.  **Proposal 9: RAN4 to introduce additional requirements based on relative difference in between the predictions due to change in the conditions/inputs, i.e., without dependency on the measurement accuracy errors and uncertainties.**  **Proposal 10: Relative difference in prediction due to change in controlled conditions/input should be considered as an additional metric to test the robustness of prediction in combination of other approximated ground truth-based options.**  **Proposal 11: RAN4 should consider relaxing known conditions for TCI states (Clause 8.10.2 of TS 38.133) for BM-Case1 and BM-Case2. Two possible approaches for RAN4 to relax known TCI states conditions:**   * **The TCI state switch command is received within a delay X\*1280 ms upon the last transmission of the RS resource for beam reporting or measurement or prediction.** * **FFS: Further analysis on defining X values.**   **Proposal 12: RAN4 to further discuss the impact of inference delay on the prediction delay requirement for BM-Case1.**  **Proposal 13: RAN4 to discuss whether the legacy measurement delay requirement should be increased for BM-Case2.**  **Proposal 14: RAN4 should consider the impact in RAN4 requirements for the Evaluation period due to the presence of predicted beams.**   |  |  |  |  | | --- | --- | --- | --- | | CSI-Report Configuration | Activation | Switching | Fall-back  (From AI/ML BM to legacy) | | Periodic CSI-reporting | Delay from RRC Reconfiguration  Delay from UE processing related to capability report  (Optional) Delay related to associated ID  Delay due to CSI-RS measurement period | Delay from RRC Reconfiguration  Delay from UE processing related to capability report  (Optional) Delay related to associated ID  Delay due to CSI-RS measurement period | Delay from RRC Reconfiguration  Delay from UE processing related to capability report  Delay due to CSI-RS measurement period | | Semi-persistent CSI-reporting | Delay for activating semi-persistent CSI resource set  Delay for selecting of semi-persistent CSI resource set    Delay from UE processing related to capability report  (Optional) Delay related to associated ID  Delay due to CSI-RS measurement period | Delay for selecting of semi-persistent CSI resource set    Delay from UE processing related to capability report  (Optional) Delay related to associated ID  Delay due to CSI-RS measurement period | Delay for activating semi-persistent CSI resource set  Delay for selecting of semi-persistent CSI resource set    Delay from UE processing related to capability report  Delay due to CSI-RS measurement period | | Aperiodic CSI reporting | Delay from UE processing related to capability report  (Optional) Delay related to associated ID  Delay due to CSI-RS measurement period | Delay from UE processing related to capability report  (Optional) Delay related to associated ID  Delay due to CSI-RS measurement period | Delay from UE processing related to capability report  Delay due to CSI-RS measurement period |   **Proposal 15: For UE-side model in beam management use case, the LCM operations are based on CSI report configuration, in our view, the delay components of LCM operations should be considered for different CSI-report configurations (i) periodic CSI report (ii) semi-persistent CSI report and (iii) aperiodic CSI report for both BM-Case1 and BM-Case2.**  **Proposal 16: RAN4 should consider the Table below as a baseline for further discussion on LCM operations related RRM core requirements.**  Table : Delay components related LCM operations for different CSI-report configurations  Table : Delay components related LCM operations for different CSI-report configurations  **Proposal 17: RAN4 should consider the impact in RAN4 requirements for performance monitoring window when UE reports performance metric or event to the NW.**  **Proposal 18: According to RAN2, when an active beam prediction configuration becomes inapplicable, the UE should report Set B measurements corresponding to the newly inapplicable. RAN4 should further study**   * **delay conditions related inapplicability reporting and** * **beam failure evaluation period due to inapplicable functionality.**   **Proposal 19: RAN4 to discuss and explore the feasibility of generating Set A and Set B beams, during the conformance testing of the AI/ML-BM use case, by selecting the number of probes and controlling the transmit power of each probe. How to decide or select the required Tx power of the selected probes is FFS.**  **Proposal 20: RAN4 to discuss and decide the contents of pre-alignment check which needs to be performed between the TE and DUT before the exact start of the conformance test.**  **Proposal 21: RAN4 to discuss and explore the feasibility of performing beam sweeping in multiple iterations during the conformance testing of AI/ML based BM use case.**  **Proposal 22: For the verification/testing of generalization related aspects in RAN4 for AI/ML BM, RAN4 should define different scenarios based on parameters listed in Table below:**  Table: Parameters for Generalization Scenarios   |  |  | | --- | --- | | **Parameters** | **Description** | | UE Speed | Slow / Medium / Fast | | SINR (Deployment Scenario) | Good / Bad Radio conditions | | Propagation Model | TDL/CDL |   **Proposal 23: RAN4 needs to design a new metric, indicative of generalization capabilities of AI/ML BM functionality, to verify the generalization performance of the functionality in different scenarios.** |
| [**R4-2504238**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504238.zip) | OPPO | **Observation 1: Absolute RSRP accuracy could be utilized as metric for L1-RSRP reporting.**  **Observation 2: At least for beam ID + L1-RSRP reporting, absolute RSRP accuracy may not truly reflect the effectiveness of beam prediction in some cases.**   * **For example, if beam i is not the best choice, yet an AI/ML model designates it as the optimal beam. Despite achieving a seemingly perfect absolute RSRP accuracy score by accurately predicting the RSRP of beam i, this could lead to sub-optimal outcomes that deviate from expectations.**   **Proposal 1: Revised option3 could be utilized as BM performance metric for beam ID + L1-RSRP reporting.**   * **Option 3(revised): The successful rate for the correct prediction which is considered as maximum ground truth L1-RSRP among top-K predicted beams is larger than the ground truth L1-RSRP of the strongest beam – x dB. FFS the value of x.**   **Proposal 2: Revise option2 to avoid cases that predicting a suboptimal beam (where the suboptimal beam’s quality is close to the best beam) is incorrectly judged as a prediction failure.**  **Option 2(revised): Beam prediction accuracy with margin**   * **Top-1 (%) with margin: The successful rate for the correct prediction which is considered as maximum ground truth L1-RSRP of top-1 predicted beams is larger than the ground truth L1-RSRP of the strongest beam – x dB,** * **Top-K/1 (%) with margin: The successful rate for the correct prediction which is considered as maximum ground truth L1-RSRP among top-K predicted beams is larger than the ground truth L1-RSRP of the strongest beam – x dB,** * **Top-1/K (%) with margin: The successful rate for the correct prediction which is considered as maximum ground truth L1-RSRP of top-1 predicted beams is larger than the ground truth L1-RSRP of the K-th strongest beam – x dB.**   **Proposal 3: Regarding Metrics/KPIs for beam prediction,**   * **If RSRP and beam ID are reported, combining Option 1 and Option 2(revised) will be used as KPIs for beam prediction.** * **If only beam ID is reported, Option 2(revised) s will be used as KPIs for beam prediction.**   **Observation 3: Regarding the impact of measurement error on BM performance evaluation, the following aspects should be further considered**   * **Whether the impact of measurement error obtained from the SLS UMA channel (measuring error added from a CDL channel evaluation) can be directly applied to the OTA testing environment** * **Differences in implementation (e.g., BS antenna configurations, BS antenna element radiation pattern, Beamforming characteristic of the BS pattern) among companies may result in inconsistent evaluation results**   **Proposal 4: For BM testability, RAN4 need to:**   1. **Clarify the limitations regarding the FR2 beams or probes that TE vendors could support** 2. **To determine the test setup for BM, a potential approach could be:**  * **Assuming that TE supports X probes in a chamber** * **By utilizing X probes, X different angles of arrival (AOA) can be achieved and can be mapped to a TX beam transmission with X paths(clusters)** * **Different TX beams in BM set A or set B can be represented by setting different pathloss for the different AOA paths, to mimic the beam sweeping procedure** * **FFS the number of probes that could be supported by TE vendors** * **FFS how to map the X AoAs and corresponding X pathloss to different beams in BM set A/B**   **Proposal 5: Further consider how to ensure that the BM model constructed on the DUT side can match(or approximate match) and be utilized in the testing environment on the TE side.**  **Proposal 6: The TCI state associated with a predicted Tx beam should be known to DUT.** |
| [**R4-2504268**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504268.zip) | Rohde & Schwarz | Observation 1: (Single-AoA) CDL channel model can be emulated in a single-AoA setup, if UE Rx antenna pattern is known, while a multi-AoA setup is to be considered, if UE Rx antenna pattern is unknown.  Observation 2: (Single-AoA) An UE Rx antenna pattern assumption can be used for testing, if UE Rx antenna performance is not part of AI/ML BM requirement, while a real UE Rx antenna pattern is to be considered, if UE Rx antenna performance is part of the requirement.  Observation 3: (Single-AoA) A static UE Rx antenna pattern can be used for testing, if UE Rx beam sweeping performance is not part of AI/ML BM requirement, while an UE Rx antenna pattern definition becomes challenging, if UE Rx beam sweeping performance is part of the requirement.  Observation 4: (Single-AoA) Single-AoA testing is sufficient, if only L1-RSRP values are relevant for the UE AI/ML BM requirement, while multi-AoA testing is rather required, if also UE Rx beam index is needed (as input for the AI/ML model.  Observation 5: (Multi-AoA) The feIFF test setup supports 4 Active 2x2 MIMO probes with distinctive angular separation of up to150°, overall big quiet zone of at least 30cm, lowest MUs and ensures preservation of (RF-RRM) side conditions.  Observation 6: (Multi-AoA) The feIFF test setup is a feasible upgrade (superset) of the existing legacy eIFF setup. As such, if feIFF is concluded as capable for AI/ML BM testing, a single system covers the the whole conformance testing, ensuring continuity in the cost sensitive FR2 testing ecosystem.  Observation 7: (Multi-AoA) The required total AoA spread of 150° can be emulated with the feIFF probe layout.  Observation 8: (Multi-AoA) The simplified channel model parameters for UMi CDL-C can be implemented in the candidate test system feIFF, with 3 possible options: Option A: with flattened elevation (Table 2) and strongest cluster aligned to one probe; Option B: with flattened elevation (Table 2) and mid-point of the total angular spread of the clusters is aligned to the mid-point of the spread of the probes; Option C: with flattened elevation, AoAs aligned to probe layout and no intra‑cluster angle spread (Table 3).  Observation 9: (Multi-AoA) The evaluation of Options A and B (Table 2) becomes more complex, if the effect of the test setup / probe layout has to be included.  Observation 10: (Multi-AoA) The emulation of the relative ray phases with the feIFF is highly accurate and should satisfy the requirement for the accuracy for the cluster power fluctuations over time.  Observation 11: (Multi-AoA) For Option C (Table 3), there is no need to optimize the probe weights, and channel model parameters in Table 3 contain all the required information.  Observation 12: (Multi-AoA) Based on simulations with 8 Set B & 32 Set A beams, there is only a negligible effect (improvement of ~1%) on the prediction accuracy for Top 3 & Top 1 beams between the cases with full CDL channel model (Reference - 24 clusters) and simplified CDL channel model suitable for feIFF (Options A/B/C - 6 clusters).  Proposal 1: (Single-AoA) RAN4 to discuss and conclude whether following aspects shall be part of the UE AI/ML BM requirement and/or testing: a) UE Rx antenna pattern knowledge, b) UE Rx antenna performance, c) UE Rx beam sweeping performance, d) UE Rx beam index (as input for the AI/ML model).  Proposal 2: (Multi-AoA) Define the implementation of eIFF test setup with 4x active 2x2 MIMO probes with probes (CATRs) layout at 0°, 30°, 90°and 150° as feIFF test setup.  Proposal 3: (Multi-AoA) Consider the feIFF test setup as candidate for the AI/ML BM baseline system and analyze its suitability.  A screen shot of a video game  Description automatically generated A screen shot of a cell phone  Description automatically generated  Figure 1 – Commercial implementation of eIFF system with 4 IFF probes / CATRs ( 🡪 feIFF)  Proposal 4: (Multi-AoA) Use channel model parameters for UMi CDL-C at 28 GHz presented in Table 1 (i.e merging clusters with similar AoAs and removing weak clusters) as starting point for further channel model simplifications.  Proposal 5: (Multi-AoA) Further discuss the metric(s) to evaluate channel model implementation with the candidate solutions.  Proposal 6: (Multi-AoA) Evaluate the AI/ML BM prediction performance when using the simplified channel model parameters for Option C (Table 3).  Proposal 7: (Multi-AoA) Consider the simplified channel models for feIFF with 6 Clusters in Table 2 and Table 3 as suitable for AI/ML BM testing.  Table 2 – Channel model parameters for UMi CDL-C at 28 GHz with flat ZoA   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Cluster # | Absolute Delay [ns] | Power in [dB] | AOD in [°] | AOA in [°] | ZOD in [°] | ZOA in [°] | | 1 | 0 | -7.4318 | -30.4353 | -134.4434 | 98.9242 | 74.51134 | | 2 | 12.594 | -1.2500 | -20.9269 | 129.1633 | 99.1915 | 74.51134 | | 5 | 13.056 | -5.5318 | -28.0782 | -152.8206 | 99.5732 | 74.51134 | | 6 | 38.196 | 0.0000 | -11.6982 | 164.1145 | 99.306 | 74.51134 | | 13 | 73.71 | -8.1318 | -33.911 | 93.1719 | 100.165 | 74.51134 | | 14 | 78.498 | -9.8318 | -37.5066 | -112.0441 | 100.2604 | 74.51134 | | Per-Cluster Parameters | | | | | | | | | Parameter | CASD in [°] | CASA in [°] | CZSD in [°] | CZSA in [°] | XPR in [dB] |  | | Value | 0.799 | 10.4021 | 0.5726 | 0 | 7 |  |   Table 3 – Channel model parameters for UMi CDL-C at 28 GHz  with AoA aligned to probe layout and no intra-cluster angle spread   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Cluster # | Absolute Delay [ns] | Power in [dB] | AOD in [°] | AOA in [°] | ZOD in [°] | ZOA in [°] | | 1 | 0 | -7.4318 | -30.4353 | -114.436 | 98.9242 | 74.51134 | | 2 | 12.594 | -1.2500 | -20.9269 | 125.5639 | 99.1915 | 74.51134 | | 5 | 13.056 | -5.5318 | -28.0782 | -174.436 | 99.5732 | 74.51134 | | 6 | 38.196 | 0.0000 | -11.6982 | -174.436 | 99.306 | 74.51134 | | 13 | 73.71 | -8.1318 | -33.911 | 95.5639 | 100.165 | 74.51134 | | 14 | 78.498 | -9.8318 | -37.5066 | -114.436 | 100.2604 | 74.51134 | | Per-Cluster Parameters | | | | | | | | Parameter | CASD in [°] | CASA in [°] | CZSD in [°] | CZSA in [°] | XPR in [dB] |  | | Value | 0.799 | 0 | 0.5726 | 0 | 7 |  | |
| [**R4-2504318**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504318.zip) | Huawei,HiSilicon | ***Observation 1:*** AoA and AoD are not used as model input in AI BM Case 1 and Case 2.  ***Proposal 1:*** Use single AoA as the channel model for AI BM testing.  ***Observation 2:*** With Gaussian distribution assumption of baseband measurement error, under different SNRs, the mean values are all close to 0, while the variance increases as the SNR drops.  ***Proposal 2:*** The simulation results for different prediction accuracy regarding to KPI 1 is shown in Table 1.  **Table 1. Measurement error impact on prediction accuracy regarding to KPI 1**   |  |  |  |  | | --- | --- | --- | --- | | **KPI-1** | | **Model input w/o measurement error** | **Model input w measurement error** | | **Top K/1 w/o margin** | Top\_1\_1 | 64.42% | 42.46% | | Top\_2\_1 | 86.69% | 65.21% | | Top\_3\_1 | 93.00% | 76.97% | | Top\_4\_1 | 95.43% | 83.59% | | Top\_5\_1 | 96.75% | 87.78% | | **Top-N genie-aided beam is one of the Top-K predicted beams** | Top\_2\_2\_1 | 94.58% | 79.33% | | Top\_2\_4\_1 | 98.60% | 92.10% | | Top\_3\_2\_1 | 96.81% | 85.99% | | Top\_3\_4\_1 | 99.44% | 95.64% | | Top\_4\_4\_1 | 99.75% | 97.57% | | Top\_5\_2\_1 | 98.60% | 93.45% | | **Top-N genie-aided beam is two of the Top-K predicted beams** | Top\_2\_2\_2 | 67.09% | 42.35% | | Top\_2\_4\_2 | 92.64% | 74.61% | | Top\_3\_2\_2 | 86.39% | 63.95% | | Top\_3\_4\_2 | 96.70% | 84.85% | | Top\_4\_4\_2 | 98.31% | 90.88% | | Top\_5\_2\_2 | 94.20% | 80.73% | | **maximum RSRP among top-K predicted beams is larger than the RSRP of the strongest beam – x dB** | x = 1dB | 83.02% | 60.75% | | x = 2dB | 92.49% | 72.87% | | x = 3dB | 96.29% | 81.84% |   ***Observation 3:*** The more stringent the KPIs are, the greater the impact of the measurement error on the prediction accuracy will be.  ***Proposal 3:*** The simulation results for different prediction accuracy regarding to KPI 1 is shown in Table 2.  Table 2. Measurement error impact on prediction accuracy regarding to KPI 2   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **KPI-2** | | | **Model input w/o measurement error** | **Model input w measurement error** | | **90%-tile L1-RSRP difference between the maximum RSRP of the Top-1/ Top-3/ Top-5 predicted beam(s) and the RSRP of the genie aided strongest beam** | **Avg.** | CDF\_strong\_1 | 2.547dB | 5.871dB | | CDF\_strong\_3 | 2.345dB | 5.392dB | | CDF\_strong\_5 | 2.209dB | 4.893dB | | **w/o avg.** | CDF\_strong\_Top1 | 2.547dB | 5.871dB | | CDF\_strong\_Top2 | 2.380dB | 5.492dB | | CDF\_strong\_Top3 | 2.139dB | 4.781dB | | CDF\_strong\_Top4 | 2.049dB | 4.445dB | | CDF\_strong\_Top5 | 1.967dB | 4.044dB | | **90%-tile L1-RSRP difference between the predicted L1-RSRP of the Top-1/ Top-3/ Top-5 predicted beam(s) and the ideal L1-RSRP of the same beams** | **Avg.** | CDF\_same\_Top1 | 2.151dB | 7.250dB | | CDF\_same\_Top1&2&3 | 2.236dB | 7.406dB | | CDF\_same\_Top1&2&3&4&5 | 2.419dB | 7.444dB | | **w/o avg.** | CDF\_same\_Top1 | 2.151dB | 7.250dB | | CDF\_same\_Top2 | 2.200dB | 7.314dB | | CDF\_same\_Top3 | 2.435dB | 7.702dB | | CDF\_same\_Top4 | 2.674dB | 7.730dB | | CDF\_same\_Top5 | 2.689dB | 7.287dB |   ***Observation 4:*** Regarding RSRP prediction accuracy, measurement errors have a larger impact on the Top 1 predicted beam than on the Top 5th predicted beam.  ***Proposal 4:*** RAN4 define requirements with measurement error as model input.  ***Proposal 5:*** RAN4 will not select KPI under which acceptable performance cannot be achieved, including   * the KPI under which the prediction accuracy is less than 80%, e.g, Top K/1 with K<4 without RSRP margin. * the KPI under which the 90%-tile L1-RSRP difference is larger than 5dB, e.g, maximum RSRP among top-K predicted beams is larger than the RSRP of the strongest beam – x dB with x<3dB.   ***Proposal 6:*** Before investigating how to set up the test environment, the following questions have to be answered:   * How to ensure the consistency between model training and model testing   + Whether and how to define a common training/testing dataset.     - If define, how to capture the different UE implementations and UE behavior when constructing the dataset   + Whether UE is expected to collect training dataset and train the model before performing model performance test.     - If yes, how to resolve the test cost/time issue. * How to avoid that a UE can pass the test but performs poorly in the field, considering that some parameters used in the test set up which limit the model generalization may totally be different from that in real deployment.   ***Proposal 7:*** For test set up in AI-BM, taking the existing FR2 OTA test set up as baseline, any enhancements on top of which should be justified. |

## Open issues summary

Previous agreements on the definition of core requirements are listed below:

R4-2417212:

**Issue 2-5: Requirements for beam predictions**

Agreement:

* + Requirements to be defined:
    - Measurement period for prediction (how long will set B be measured before making a prediction)
      * FFS on whether the legacy measurement period can be reused or not.
    - Prediction accuracy
      * FFS how/what to define depending on the chosen KPI (e.g. predicted RSRP, beam ID, Top-K/1 or Top 1/K, etc)
    - TCI state known/unknown
  + FFS on whether the other requirements are needed

R4-2420513:

**Issue 3-5: Prediction delay requirements**

**Agreement:**

* Prediction delay includes:
  + Measurement for prediction
    - FFS on whether the legacy measurement delay can be reused
  + Inference delay
  + Other components are FFS

Further agreements from the previous meeting on this topic are listed below for reference:

**Issue 2-1: Metrics/KPIs for beam prediction**

If RSRP and beam ID are reported:

* RSRP accuracy will be one of the KPIs.
  + How option 1 is applied is FFS
* Beam ID prediction accuracy will be one of the KPIs
  + FFS whether beam ID prediction accuracy refers to a combination of Option 2 and/or Option 3 or modified Option 2/3 or other options

If only beam ID is reported:

* Beam ID prediction accuracy will be used as the KPI
  + FFS whether beam ID prediction accuracy refers to a combination of Option 2 and/or Option 3 or modified Option 2/3 or other options

**Issue 2-4: QCL Source RS for TCI states**

Agreement:

wait for RAN1 discussion to conclude without sending any LS

**Issue 4-1: Baseline test setup**

**Agreement:**

* Discuss both single AoA based and multiple AoA based set up
  + Continue to study the two multiple AoA candidate setups
    - One option to continue the discussion proposed as following:
      * Multi AoA setup – evaluation of already feasible systems for AI/ML BM testing suitability
        + AI/ML BM performance centric, not CDL channel emulation centric.
        + Define test system candidates which are already feasible
      * Perform system level simulations including the probe layout of candidate systems to address the impact on the UE AI ML BM performance.
  + Start with the set up with Set A = 32 beams, Set B = 8 beams, and develop the test procedures with single AoA and multiple AoAs

**Issue 4-2: Channel model**

**Agreement:**

* Use CDL-based channel model as starting point

The open issues were grouped in the following sub-topics for further discussion:

Beam prediction KPIS, revised/modified option 3 for beam ID KPI, postpone?

Absolute RSRP application

1. Beam ID prediction KPI
2. Absolute RSRP accuracy applicability
3. Relative RSRP accuracy definition
4. Measurement period for inference
5. Simulation results and next steps
6. Test system basic setup
7. Test system channel model

### Sub-topic 2-1

*Beam ID prediction KPIs*

The options for the beam prediction KPIs were downselected in the previous meeting, the KPIs for beam IDs still have to be agreed. The previously tentative agreement

- 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

**Issue 2-1: Metrics/KPIs for beam ID prediction**

Proposals

* + Option 1: Adopt only Option 3 as is
  + Option 2:Modified Option 3:
    - All top-K predicted beams have to be considered in the success rate evaluation
    - RSRP of each top-K predicted beam is to be compared with the RSRP of its equal ranked genie-aided beam - X
  + Option 3: Option 2 only with Top-K/1 (%)
  + Option 4: Option 2 and Option 3
  + Option 5: Wait for more simulation results to understand how the different options work
  + Option 6: Others (e.g. different distributions, etc): Normalized Discounted Cumulative Gain (nDCG)
* Recommended WF
  + Option 5

### Sub-topic 2-2

*Absolute RSRP accuracy applicability*

**Issue 2-2: Absolute RSRP accuracy**

* Proposals
  + Option 1: Absolute RSRP accuracy requirement applies to all predicted beams (all beams in the report)
  + Option 2: Absolute RSRP accuracy requirement applies to only a subset of predicted beams
  + Option 3: Others
* Recommended WF
  + To be discussed

### Sub-topic 2-3

*Relative RSRP accuracy*

Relative RSRP accuracy has been discussed for a few meetings without any clear agreement

**Issue 2-3: Relative RSRP accuracy**

* Proposals
  + Option 1: The relative RSRP accuracy for AI/ML-based beam prediction = the ground truth of L1-RSRP of the predicted beam index i - the ground truth of the L1-RSRP of the Top-1 ideal beam index j
* Option 2: ***The relative RSRP accuracy for AI/ML-based beam prediction = Absolute L1-RSRP Accuracy of the predicted beam index i – Absolute L1-RSRP Accuracy of the best-predicted beam.***
  + Option 3: Relative RSRP accuracy = (predicted L1-RSRP of beam index i - predicted L1-RSRP of beam index n) - (ground truth of L1-RSRP of beam index i - ground truth of L1-RSRP of beam index n), where the beam index n owns the largest predicted value.
  + Option 4: Relative RSRP accuracy=predicted L1-RSRP of beam index i - ground truth of the beam index n ,where the beam index n owns the largest value
  + Option 5: **relative RSRP accuracy= predicted L1-RSRP of beam index i – predicted L1-RSRP of beam index n, where the beam index n owns the largest RSRP among the predicted beams.**
  + Option 6: Relative RSRP accuracy for AI/ML BM can be defined as:
    - relative RSRP accuracy = (reported predicted RSRP of beam i – reported RSRP of beam j) – (reported RSRP of beam i – reported RSRP of beam j) where beam i is one of the top-K predicted beams and beam j is the beam with maximum measured RSRP
    - RAN4 to further discuss whether the reported RSRP of beam j can be predicted or measured or a combination of both predicted and measured samples.
  + Option 7
* Recommended WF
  + To be discussed

### Sub-topic 2-4

*Measurement period for inference*

Requirements for the measurement period are discussed by multiple companies and should be defined as core requirements

**Issue 2-4: Measurement period for inference**

* Proposals
  + Option 1: Use “legacy” L1-RSRP measurement as baseline
    - FFS whether same values for M,N,P should be reused
  + Option 2: Split case 1 and case 2
    - Reuse L1 measurement period for Case 1
    - Introduce configurable number of measurement occasions for case 2
  + Option 3: Measurement period should be configurable
  + Option 4: Others
* Recommended WF
  + To be discussed

### Sub-topic 2-5

*Simulation results*

**Issue 2-5: Simulation results**

* Proposals
  + Option 1: Discuss the simulation results based on summary – to be prepared by vivo
    - Simulation results
    - Refinement of simulation assumptions (e.g. error modelling, etc)
* Recommended WF
  + To be discussed

### Sub-topic 2-6

*Test system basic setup*

The basic test setup has been discussed for several meetings without much progress.

**Issue 2-6: Test system setup**

* Proposals
  + Option 1: Single AoA with UE antenna pattern knowledge
  + Option 2: Single AoA without any UE antenna pattern knowledge
  + Option 3: Multi AoA based on MPAC
  + Option 4: Multi AoA based on enhanced IFF RRM testing setup
    - Exact definition of enhanced IFF is FFS
* Recommended WF
  + To be discussed

Aspects to consider in the discussion based on the companies’ inputs:

**Test objectives and data to be used**

* whether the performance under synthetic channels can reflect the performance of AI/ML models in the real field.
* discuss what the testing goal is, to verify the model can work well in real field or just to test UE can train a model that works well under a specific scenario.
* If the testing goal is to verify the model can work well in real field, and the performance under a specific channel/scenario cannot reflect the performance in the real field, it is necessary to design a test system that is possible to use field data in the test.

**UE models and impact on testing setup and performance:**

***There are currently three kinds of presumable UE patterns from the AI model input PoV.***

UE pattern 1: No need to input Rx Beam information in the AI model.

In this case, it is expected that the UE Rx beam is pointing towards the direction of the measurement probe in an OTA chamber where the DL signals are transmitted.

UE pattern 2: One Rx Beam information is input to the AI model.

In this case, direction of the UE Rx beam is same as UE pattern 1 and that Rx Beam information is input to the AI model. TBD if the actual Rx beam pattern is necessary or it can be standardized.

UE pattern 3: Multiple Rx Beam information is input to the AI model.

In this case, for each Tx Beam ID in set B, UE will change its Rx beam direction to multiple directions and multiple L1-RSRP measurements are made per Tx Beam ID. Multiple Rx Beam information is input to AI model per Tx Beam ID.

***Observation based on above models:***

***BM-Case 1 tests with UE pattern 1 and 2 (i.e. less than or equal to one Rx Beam information is input to the AI model) can be conducted with the following conditions.***

* ***Rx beam pattern is standardized or,***
* ***Rx beam pattern is measured during the evaluation.***

**UE beam management related issues**

RAN4 to discuss and conclude whether following aspects shall be part of the UE AI/ML BM requirement and/or testing:

a) UE Rx antenna pattern knowledge,

b) UE Rx antenna performance,

c) UE Rx beam sweeping performance

d) UE Rx beam index (as input for the AI/ML model).

### Sub-topic 2-7

*Test system channel model*

In the previous meeting it was agreed that a CDL-based channel model will be taken as starting point. Further details need to be discussed

**Issue 2-7: Test system channel model**

* Proposals
  + Option 1: “full” CDL model to be emulated in the test
  + Option 2: simplified CDL should be used
  + Option 3: others
* Recommended WF
  + To be discussed

If Option 2 is chosen, simplification methods should be discussed.

# Topic #3: RRM core requirement and testing framework for Positioning accuracy enhancement

This section contains the sub-topics regarding specific issues for positioning

## Companies’ contributions summary

|  |  |  |
| --- | --- | --- |
| **T-doc number** | **Company** | **Proposals / Observations** |
| [**R4-2503254**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503254.zip) | Qualcomm Incorporated | **Observation 1: RAN plenary removed case 2a and 2b from the scope of AI-ML positioning in the latest WID.**  **Observation 2: UE based positioning with UE side model is a 1st priority item of AI/ML positioning.**  **Observation 3: Legacy spec has previously defined the signalling for reporting UE’s location via DL-TDoA IE in non-AI-ML scenario.**  **Observation 4: RAN2 has agreed to LTE positioning protocol (LPP) based UE reporting of inferred location for LCM of case 1.**  **Proposal 1: RAN4 will not define any requirement for case 2a/2b.**  **Proposal 2: RAN4 defines requirements for case 1, i.e., UE-based positioning with UE-side model.** |
| [**R4-2503408**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503408.zip) | CATT | **Proposal 1: RAN4 do not discussion case 2a/2b related issues in Rel-19.**  **Proposal 2: RAN4 not to define performance requirements for case 1.**  **Proposal 3: RAN4 not to define requirements for LOS/NLOS indicator reporting.** |
| [**R4-2503518**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503518.zip) | Apple | **Proposal 1: RAN4 to further discuss the feasibility and how to define requirements for positioning accuracy for case 1. Defining requirements is also dependent on RAN1/2 conclusion regarding whether UE will report positioning**  **Proposal 2: For Assisted AIML Positioning, the KPI test metric (e.g., LOS/NLOS) needs to be considered for validating the positioning accuracy. RAN4 should inevstiagte how to establish requirements for LOS/NLOS indicators, considering their potential impact on performance and accuracy in positioning systems.**  **Proposal 3: For AIML positioning, RAN4 should take as baseline the requirements on the measurements supported in legacy positioning (RSTD, PRS-RSRP, and PRS-RSRPP) and define the requirements on the new measurements (e.g., CIR, DP, and PDP) based on RAN1 progress.**  **Proposal 4: RAN4 to study defining performance accuracy requirements for any reported UE measurements defined by other groups.**  **Proposal 5: RAN4 should define measurement requirements for both path-based and sample-based reporting in relation to use case 2b.**  **Proposal 6: Before RAN4 investigates testing Case 2b in AI/ML for positioning, it is important to first examine the relationship between the measurement accuracy reported by the UE and the resulting positioning accuracy at the LMF.**  **Proposal 7: RAN4 to define performance accuracy requirements for use case 3a for the measurements reported by gNB. The specifics and scope of the core and performance requirements for use case 3a are contingent on the outcomes of discussions within RAN1 and RAN2**  **Proposal 8: RAN4 to define performance accuracy requirements for measurements performed at gNB for use case 3b. The specifics and scope of the core and performance requirements for use case 3a are contingent on the outcomes of discussions within RAN1 and RAN2** |
| [**R4-2503771**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503771.zip) | CMCC | ***Proposal 1: for case 1, it is proposed to define positioning accuracy requirement, which is the difference between ground truth (real position) and reported position.***  ***Proposal 2: for case 1, it is proposed to define delay related requirements, e.g. prediction delay, LCM related delay.***  ***Observation 1: for AI/ML assisted positioning case 3a, RAN1 agreed that UL RTOA, gNB Rx-Tx time difference are supported for reporting.***  ***Observation 2: RAN1 common understanding is that RAN1 proceed with the investigation assuming each case is deployed individually, which means that Case 2a and Case 3a can be used individually.***  ***Proposal 3: for case 3a, it is proposed to take legacy accuracy requirements for gNB Rx-Tx time difference as baseline.*** |
| [**R4-2503790**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503790.zip) | ZTECorporation,Sanechips | **Observation 1: There is no need for RAN4 to consider model inference for case 2b since the AI/ML model is deployed at LMF side which the output UE location on the LMF side depends on the network implementation.**  **Observation 2: RAN4 would not define any positioning accuracy requirements because positioning is LMF based.**  **Proposal 1: RAN4 shall not define the positioning accuracy requirements for case 2b, it is up to network implementation.**  **Observation 3: In legacy, there are two kinds of location request: UE-triggered location request and NW-triggered location request. The latest agreement contains the implicit location request which is the NW triggers the location request, otherwise no reporting scheme is needed.**  **Proposal 2: RAN4 shall wait for RAN1 agreements on defining reporting scheme.**  **Observation 4: For AI/ML based positioning, the main difference compared to the legacy is that the AI model resides within the LMF.**  **Observation 5: RAN4 would not define any positioning accuracy requirements since positioning is LMF based. Apart from the positioning accuracy requirements, the reported metrics can be studied for case 2a.**  **Proposal 3: The intermediate features such as RSTD, RSRP, UE Rx-Tx time diference etc. or some enhancements shall be considered.**  **Observation 6: From request to reporting time, there are two periods of time which are PRS measurement and model inference. If there is no limitation on these two periods, LMF would take a long time to receive an inference outcome or the intermediate features. It has the impact on the performance accuracy since the accuracy of model output will also decrease over time.**  **Proposal 4: RAN4 shall define the delay requirements when considering PRS measurement and AI model processing time.**  **Observation 7: RAN1 just agreed at least LOS/NLOS indicator is supported for reporting.**  **Proposal 5: Although LOS/NLOS could be the intermediate feature to be reported, RAN4 shall consider whether and how to define the requirements for LOS/NLOS.**  **Proposal 6: RAN4 shall deprioritize the study of ToA until RAN1 supports UE report ToA as model output for case 2a.**  **Observation 8: If the ground truth is known to test equipment, the accuracy can be quantified compared to defined ground truth, such as AI/ML positioning error. Then the model performance can be evaluated and the subsequent operation of model can be clarified.**  **Observation 9: The UE location is the coordinate and the PRU will be set on anywhere, if we set PRU randomly, we do not think the accuracy can be accurately evaluated.**  **Observation 10: If UEs have the GNSS capability, they also can be the reference UEs. However, if we choose the reference GNSS UEs, the GNSS error shall be considered when evaluating the AI/ML positioning accuracy error.**  **Proposal 7: The ground truth can be the PRU and the test map shall be considered. Reference GNSS UEs shall be considered.**  **Observation 11: RAN1 deems that LOS/NLOS indicator can be the input for model inference at LMF side. However, how to verify the indicator is one concern since RAN4 did not discuss the propagation conditions before.**  **Proposal 8: Whether the single path or multi channels shall be considered and which kind of channels shall be used shall be studied.**  **Proposal 9: A single path for LOS is enough. How many multiple channels will be used for NLOS shall be discussed based on which kind of channels will be used.** |
| [**R4-2503927**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2503927.zip) | Ericsson | **# Use case 1**  **Observation 1**: Performance requirement in terms of the accuracy of the location estimated by the UE is not specified for UE based positioning supported by the NR specifications.  **Observation 2**: Achievable positioning accuracy is scenario specific, and the same level of positioning accuracy cannot be achieved in different scenarios.  **Observation 3**: Producing a test setup that mimics for example InF or IOO or UMi or UMa with lot more cells than in the legacy setup is hard to be achieved in a lab.  **Proposal 1**: RAN4 to conclude that defining requirement in terms of the accuracy of estimated UE position for case 1 to be met in any scenario is not feasible.  **# Use case 3a**  **Proposal 2**: Extend applicability of existing report mapping table for gNB Rx-Tx time difference measurement to report gNB Rx-Tx time difference measurement for AI/ML based positioning use cases 3a.  **# Use case 3b**  **Observation 4**: Depending on the request from LMF, gNB can report path-based measurements to the LMF. For such measurement reporting, no enhancement to Rel. 18 reporting has been introduced.  **Observation 5**: Depending on the request from LMF, gNB can report sample-based measurements to the LMF. For such measurement reporting, report mapping table needs to be specified in RAN4 specification.  **Observation 6**: For sample-based reporting, gNB may either report timing measurement or timing and power measurements associated to Nt´ samples of the estimated channel response.  **Observation 7**: Reporting granularities for sample-based timing measurement reporting is same as the reporting granularities supported for legacy measurements UL-RToA and gNB Rx-Tx time difference measurement.  **Proposal 3**: Reporting range for sample-based timing measurements is from -985024´Tc to 985024´Tc.  **Proposal 4**: Table 1 is used to define measurement report mapping table for sample-based timing measurement reporting for reporting granularity corresponding to k = 0.  **Proposal 5**: Table 2 is used to define measurement report mapping for sample-based timing measurement reporting for reporting granularity corresponding to k = 1.  **Proposal 6**: Table 3 is used to define measurement report mapping for sample-based timing measurement reporting for reporting granularity corresponding to k = 2.  **Proposal 7**: Table 4 is used to define measurement report mapping for sample-based timing measurement reporting for reporting granularity corresponding to k = 3.  **Proposal 8**: Table 5 is used to define measurement report mapping for sample-based timing measurement reporting for reporting granularity corresponding to k = 4.  **Proposal 9**: Table 6 is used to define measurement report mapping for sample-based timing measurement reporting for reporting granularity corresponding to k = 5. |
| [**R4-2504024**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504024.zip) | vivo | **Proposal 1: RAN4 to define delay requirement for case 1 at least considering the following procedures: PRS measuring, AI processing, UE reporting location to LMF.**  **Proposal 2: RAN4 to consider the straight-line distance (unit: m) between the real position and the estimated position as the metric for accuracy requirement for case 1.**  **Proposal 3: RAN4 to postpone the discuss of requirement for positioning case 2b.**  **Proposal 4: RAN4 to define delay requirement for case 2a at least considering the following procedures: PRS measuring, AI processing, UE reporting intermediate features to LMF.**  **Proposal 5: RAN4 to reuse the existing accuracy requirements for the timing-related intermediate features: RSTD, UE Rx-Tx time difference for positioning case 2a.**  **Observation 1: It is maybe not possible to define requirement for the LoS/NLoS indicator soft value.**  **Proposal 6: RAN4 to define performance requirement for LoS/NLoS indicator hard value, the accuracy requirement can be the probability of maximum LoS/NLoS misestimation.**  **Proposal 7: RAN4 not to define requirements for positioning case 3.**  **Proposal 8: RAN4 to wait for more progress from other WGs to consider define delay requirements for monitoring for positioning case 1.**  **Observation 2: PRUs can be deployed in the test scenario to provide measurements together with the location information.**  **Proposal 9: RAN4 to consider deploy PRUs in conformance or the pre-deployment test so as to derive the ground truth value of UE position and intermediate features.**  **Observation 3: At least LoS/NLoS indicator hard value can be tested since the LoS/NLoS environment is known in test system.**  **Proposal 10: RAN4 to discuss the key factors that affect the performance and consider introduce reference scenario/model/dataset for AI/ML based positioning test for generalization.** |
| [**R4-2504237**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504237.zip) | OPPO | **Proposal 1: For case1, RAN4 will not define positioning accuracy requirements in R19 WI.**  **Observation 1: For case1(AI/ML assisted positioning), the feasibility of using intermediate results(output of AI/ML models with non-linear processing) for RAN4 tests is not clear.**  **- how to get the label data for intermediate results(e.g. RSTD, identification of LoS/NLoS and other metrics that agreed in RAN1) and test these intermediate results is not clear.** |
| [**R4-2504319**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504319.zip) | Huawei,HiSilicon | ***Proposal 1***: For Case 2b and 3b, reuse the legacy requirement for reporting of timing information or timing and power information from UE/gNB to LMF.  ***Proposal 2***: For Case 2a and 3a, RAN4 to discuss whether to reuse the legacy requirement for reporting of timing information from UE/gNB to LMF, if there is no explicit indicator introduced to distinguish whether the timing information is obtained by legacy method or by Rel-19 AI/ML.  ***Proposal 3***: RAN4 to discuss the requirement for Rel-19 AI/ML based timing information reporting if introduced by other WGs.  ***Proposal 4:*** RAN4 to discuss requirements for LCM procedure especially performance monitoring for AI/ML positioning, based on RAN1 conclusion on performance monitoring schemes and also RAN4 conclusion on the requirements for inference. |
| [**R4-2504542**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114bis/Docs/R4-2504542.zip) | Nokia | **Case 1**   1. For AIML positioning in Case 1, RAN4 to define the requirements on the measurements supported in legacy positioning (e.g., RSTD, PRS-RSRP, and PRS-RSRPP) and define the requirements on the new measurements (e.g., CIR, DP, and PDP) based on RAN1 progress. 2. RAN4 to study the necessity of any requirements for the input of the AIML model in Case 1. 3. RAN4 requirements (e.g., in Case 1) should be firstly defined with respect to single PFL. 4. The consistency between training and inference should be achieved by having a requirement on consistent configuration (e.g., same PFL) when UE-sided models are used. 5. RAN4 to study a new requirement on a quality indicator for the ground-truth accuracy on positioning label in data collection. 6. Positioning accuracy should be considered as the performance metric in Case 1. 7. Even when ground truth, i.e., accurately known position of the UE, for Case 1 AI/ML-based positioning is not available in the testing setup for absolute positioning accuracy calculation, the new functionality should not be left without any requirements or tests. 8. RAN4 to consider relative positioning accuracy metric for Case 1, i.e., requirement on the possible change in AI/ML-based position due to a change/state of the environment/configuration.   If an LCM action is required and it is not taken in a timely manner, the performance degradation for AI/ML enabled Positioning use case may be degraded to undesirable level.   1. RAN4 to define the time latency limit on UE’s LCM actions indicated by network. 2. RAN4 to define RRM requirements for LCM, prioritizing Activation, Deactivation, Switching, Fallback to non-AI operation, and Performance monitoring. 3. Until performance monitoring aspects are agreed in RAN1, RAN4 will wait to discuss the necessity of UE / PRU requirements for performance monitoring.   Reporting of location estimates is supported in the current specifications.  Reuse of the reporting facility for Case 1 is possible, given that a certain location coordinate type from the existing ones is selected.  The UE is enabled to report its local 3D coordinates with origin in a known reference location together with its estimated location uncertainty, both inferred through its AI/ML model.   1. RAN4 to use existing location coordinate reporting to identify positioning accuracy requirements for Case 1.   Positioning accuracy can be verified based on the ground truth which may consist of the location points with known positioning co-ordinates (e.g., PRU or GNSS based).   1. RAN4 should specify (i) the test mechanisms for positioning accuracy and (ii) the metric for positioning accuracy performance in Case 1. 2. RAN4 to specify a test grid map with known location points in indoor space and an appropriate set of TRPs for the positioning measurement in Case 1. 3. RAN4 to select either error calculation method 2 (without location uncertainty) or 3 (applying worst case uncertainty error) for deriving the performance metric for Case 1. 4. RAN4 to select the three-dimensional error e of the absolute reported positioning co-ordinates relative to the evaluated ground truth as performance metric for the positioning accuracy in Case 1.   Case 3a/3b   1. RAN4 to study whether to reuse the legacy performance requirements for the measurement types (B) and whether to specify separate performance requirements for type (B) measurement. 2. RAN4 to wait on further progress in RAN1 to address testability aspects for the LOS/NLOS indicator together with other measurement(s) for Case 3a. 3. RAN4 to consider testability aspects for Case 3b both for path-based and sample-based measurements, for latter awaiting further details from RAN1. |

## Open issues summary

The open issues were grouped in the following sub-topics for further discussion:

Case 1 requirements

LOS/NLOS indicator

gNB Rx-Tx time difference reporting requirements

Report mapping for gNB Rx-Tx time difference

Sample based timing measurement reporting

### Sub-topic 3-1

*Requirements for case 1*

It has been discussed for several meetings whether to define requirements for case 1 without any progress

**Issue 3-1: Requirements for case 1**

Proposals

* + Option 1: RAN4 will not define positioning accuracy requirements for case 1(direct UE positioning)
    - Why it is not feasible?
  + Option 2: RAN4 to further study how to define the requirements, companies are invited to bring concrete proposals on a requirements framework and next steps on how to define
    - Concrete proposals on how the requirements would be defined, how to determine the core/performance requirements and testability
    - At least reporting delay(including inference delay) requirements will be defined
  + Option 3: others
* Recommended WF
  + Option 2

If no concrete proposals are shown in the next meeting, discussion should not continue.

Ericsson: The question is the feasibility. We think it is not feasible. We prefer to Option 1. It is related to both testability and requirements.

Nokia: Prefer Option 2. What kind of requirements to be defined? We want minimal set of requirements.

CATT: We did simulation analysis. It is difficult to define the general requirements. Share the view from Ericsson. Option 1.

Moderator: better to discuss it in May meeting further. We should have more clear view on the proposals. Unless we see the clear proposal, we won’t discuss it more. We need finish core part soon.

### Sub-topic 3-2

*LOS/NLOS indicator*

It should be discussed whether to introduce requirements or not for LOS/NLOS indicator

**Issue 3-2: LOS/NLOS indicator**

Proposals

* + Option 1: RAN4 to introduce requirements for LOS/NLOS indicator
    - Companies should bring detailed proposals on how to define requirements and potential tests
  + Option 2: It is not feasible to introduce requirements for this indicator
  + Option 3: Requirements are not needed
  + Option 4: Others
* Recommended WF
  + To be discussed

Qualcomm: this is related to case 2a and 2b, which are removed from WID.

Ericsson: This is also our view.

Conclusion: the LOS/NLOS indicator is out of scope of WID. Stop discussions on it.

### Sub-topic 3-3

*Case 3a – Rx-Tx time difference reporting*

**Issue 3-3: Rx-Tx time difference**

* Proposals
  + Option 1: Introduce gNB Rx-Tx time difference reporting accuracy requirements
    - Take legacy accuracy requirements as baseline
  + Option 2: Do not define any requirements
  + Option 3: others
* Recommended WF
  + Option 1

Ericsson: OK to reuse the legacy requirements.

Agreement:

* Introduce gNB Rx-Tx time difference reporting accuracy requirements
  + - Reuse legacy accuracy requirements

### Sub-topic 3-4

*Report mapping for gNB Rx-Tx time difference*

**Issue 3-4: Report mapping for gNB Rx-Tx time difference**

Proposals

* + Option 1: Reuse the existing report mapping for gNB Rx-Tx time difference measurements
  + Option 2: Introduce other report mapping
* Recommended WF
  + Option 1

Samsung: For those two sub-topics, we are OK with WF. For perf, we need discuss the bandwidth for testing. Is that common understanding?

Moderator: this is nothing about the performance. Samsung comment is correct.

Agreement:

* Reuse the existing report mapping for gNB Rx-Tx time difference measurements

### Sub-topic 3-5

*Sample based timing measurements*

**Issue 3-5: Sample based timing measurements**

Proposals

* + Option 1: Introduce sample based timing measurement report mapping as follows:

**Proposal 3**: Reporting range for sample-based timing measurements is from -985024´Tc to 985024´Tc. **Proposal 4**: Table 1 is used to define measurement report mapping table for sample-based timing measurement reporting for reporting granularity corresponding to k = 0.   
**Proposal 5**: Table 2 is used to define measurement report mapping for sample-based timing measurement reporting for reporting granularity corresponding to k = 1.  
**Proposal 6**: Table 3 is used to define measurement report mapping for sample-based timing measurement reporting for reporting granularity corresponding to k = 2.  
**Proposal 7**: Table 4 is used to define measurement report mapping for sample-based timing measurement reporting for reporting granularity corresponding to k = 3.  
**Proposal 8**: Table 5 is used to define measurement report mapping for sample-based timing measurement reporting for reporting granularity corresponding to k = 4.

**Proposal 9**: Table 6 is used to define measurement report mapping for sample-based timing measurement reporting for reporting granularity corresponding to k = 5.

* + Option 2: Introduce other report mapping
* Recommended WF
  + To be further discussed

Qualcomm: Is this for Case 3b?

Ericsson: this is for 3b.

Nokia: We can check it more.