**3GPP TSG-RAN WG4 Meeting #114 R4-2500683**

**Athens, Greece, 17th – 21st February, 2025**

**Agenda item:** 7.19.5

**Source:** Moderator (Qualcomm)

**Title:** Topic summary for [114][131] NR\_AIML\_air\_part1

**Document for:** Information

# Introduction

This is the summary thread for issues related to NR AI/ML work in RAN4. WFs summarizing many topics/issues to be further studied and discussed was agreed in the previous meeting in R4-2520513 and R4-2420336.

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

This section contains the sub-topics regarding CSI prediction

## Companies’ contributions summary

|  |  |  |
| --- | --- | --- |
| **T-doc number** | **Company** | **Proposals / Observations** |
| [**R4-2500228**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500228.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**  **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. |
| [**R4-2500270**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500270.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.** |
| [**R4-2500338**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500338.zip) | MediaTek inc. | **Observation #1: 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 #1: We propose to use eType II non-predicting reporting as reference in test case definition to ensure test feasibility.**  **Proposal #2: We propose to first focus on static condition to guarantee test feasibility.**  **Observation #2: There are four existing PMI requirements for PMI prediction that can be used as starting point for Rel-19 work.**  **Observation #3: There is still room to improve existing test configuration to enable even better prediction gain.**  **Proposal #3: We propose to use existing eType II doppler codebook test configurations as a starting point for Rel-19 work.**  **Proposal #4: We propose to further study if existing eType II doppler codebook test configurations can be enhanced for better prediction opportunities.**  **Proposal #5: We propose to study prediction timing in test configuration.**  **Proposal #6: We propose to study prediction aging in test configuration.**  **Proposal #7: We propose to study CSI-RS resource allocation in test configuration.** |
| [**R4-2500512**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500512.zip) | CAICT | **Proposal 1: The Demod and PMI reporting requirements defined for TypeII-Doppler-r18 codebook could be considered as starting point for AI-based CSI prediction.**  **Proposal 2: Whether to define better requirements for AI-based CSI prediction shall be decided based on further evaluations. The test setup adopted for TypeII-Doppler-r18 codebook could be taken as starting point.** |
| [**R4-2500822**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500822.zip) | vivo | **Proposal 1: 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 2: Similar test conditions of non-AI prediction would be considered for AI-prediction as starting point.**  **Proposal 3: Table 2-1 for non-AI prediction (Table 6.3.2.1.8-1 in 38.101) can be used for AI prediction as starting point. Some parameters in Table 2-1 would be not suitable or needed for AI prediction, such as:**   * **Aperiodic CSI-RS configuration;** * **Codebook configuration for TypeII-Doppler-r18 codebook.**   Table 2-1. Test conditions for multiple PMI with 16TX Enhanced Type II codebook for predicted PMI (non-AI prediction, Table 6.3.2.1.8-1 in 38.101)   |  |  |  |  | | --- | --- | --- | --- | | **Parameter** | | **Unit** | **Test 1** | | Bandwidth | | MHz | 10 | | Subcarrier spacing | | kHz | 15 | | Duplex Mode | |  | FDD | | Propagation channel | |  | TDLA30-20 | | Antenna configuration | |  | XP Medium 16 x 2  (N1,N2) = (4,2) | | Beamforming Model | |  | As specified in Annex B.4.1 | | ZP CSI-RS configuration | CSI-RS resource Type |  | Aperiodic | | Number of CSI-RS ports (*X*) |  | 4 | | CDM Type |  | FD-CDM2 | | Density (ρ) |  | 1 | | First subcarrier index in the PRB used for CSI-RS (k0) |  | Row 5, (4) | | First OFDM symbol in the PRB used for CSI-RS (l0) |  | (9) | | CSI-RS  interval and offset | slot | Not configured | | ZP CSI-RS trigger |  | 1 in slots i, where mod(i, 5) = 1, otherwise it is equal to 0 | | NZP CSI-RS for CSI acquisition | CSI-RS resource Type |  | Aperiodic | | The number of CSI-RS resources *(K)* |  | 4 | | Number of CSI-RS ports (*X*) |  | 16 for CSI-RS resource 1,2,3,4 | | CDM Type |  | CDM4 (FD2, TD2) for CSI-RS resource 1,2,3,4 | | Density (ρ) |  | 1 for CSI-RS resource 1,2,3,4 | | First subcarrier index in the PRB used for CSI-RS (k0, k1, k2, k3) |  | Row 12, (2, 4, 6, 8) for CSI-RS resource 1,2,3,4 | | First OFDM symbol in the PRB used for CSI-RS (l0) |  | (5) for CSI-RS resource 1,2,3,4 | | CSI-RS  interval and offset | slot | Not configured for CSI-RS resource 1,2,3,4 | | aperiodicTriggeringOffset |  | 0 | | Separation between two consecutive CSI-RS resources (*m*) | slot | 2 | | CSI-IM configuration | CSI-IM resource Type |  | Aperiodic | | CSI-IM RE pattern |  | Pattern 0 | | CSI-IM Resource Mapping  (kCSI-IM,lCSI-IM) |  | (4,9) | | CSI-IM timeConfig  interval and offset | slot | Not configured | | ReportConfigType | |  | Aperiodic | | CQI-table | |  | Table 1 | | reportQuantity | |  | cri-RI-PMI-CQI | | timeRestrictionForChannelMeasurements | |  | Not configured | | timeRestrictionForInterferenceMeasurements | |  | Not configured | | cqi-FormatIndicator | |  | Wideband | | pmi-FormatIndicator | |  | Not configured | | Sub-band Size | | RB | 4 | | csi-ReportingBand | |  | 1111111111111 | | CSI-Report interval and offset | | slot | Not configured | | Aperiodic Report Slot Offset | |  | 10 | | CSI request | |  | 1 in slots i, where mod(i, 10) = 0, otherwise it is equal to 0 | | reportTriggerSize | |  | 1 | | CSI-AperiodicTriggerStateList | |  | One State with one Associated Report Configuration  Associated Report Configuration contains pointers to NZP CSI-RS and CSI-IM | | Codebook configuration | Codebook Type |  | typeII-doppler-r18 | | *paramCombination-Doppler-r18* |  | 7  (L =4, *pν* =1/2, β=1/2 ) | | R*(numberOfPMI-SubbandsPerCQI-Subband-r18)* |  | 1 | | (CodebookConfig-N1,CodebookConfig-N2) |  | (4,2) | | (CodebookConfig-O1,CodebookConfig-O2) |  | (4,4) | | CodebookSubsetRestriction |  | 0x 7FF FFFF FFFF | | RI Restriction (*typeII-RI-Restriction-r18*) |  | 0010 | | Doppler/time-domain basis vector length *(N4)* |  | 1 | | Doppler-/time-domain unit duration (*d*) | slot | 1 | | Prediction delay *(delta)* | slot | 1 | | Physical channel for CSI report | |  | PUSCH | | CQI/RI/PMI delay | | ms | 15 | | Maximum number of HARQ transmission | |  | 4 | | Measurement channel | |  | R.PDSCH.1-24.1 FDD | | PDSCH & PDSCH DMRS Precoding configuration for random Precoding | |  | Single Panel Type I, Random precoder selection updated per slot, with equal probability of each applicable i1, i2 combination, and with i1 wideband granularity and i2 subband granularity | | Note 1: When Throughput is measured using random precoder selection, the precoder shall be updated in each slot (1 ms granularity) with equal probability of each applicable i1, i2 combination. The random precoder generation shall follow 'typeI-SinglePanel' codebook configuration as specified in table 6.3.2.1.3-1.  Note 2: If the UE reports in an available uplink reporting instance at slot#n based on PMI estimation using a CSI-RS resource set in which the last CSI-RS resource is transmitted at a downlink slot not later than slot#(n-4), this reported PMI cannot be applied at the gNB downlink before slot#(n+4).  Note 3: Randomization of the dual-cluster beam directions shall be used as specified in AnnexB.2.3.2.3A. The value of relative power ratio (p) shall be fixed as 1 during the test. | | | |   **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.** |
| [**R4-2501035**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501035.zip) | CATT | **Proposal 1: Similar to AI/ML based BM, CSI prediction delay includes measurement delay for prediction and inference delay and RAN4 prioritizes the discussion on the value of inference delay.**  **Proposal 2: RAN4 to discuss the detailed reporting requirements after RAN1 has concrete agreements.** |
| [**R4-2501093**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501093.zip) | CMCC | ***Proposal 1: for CSI prediction, it is preferred to consider non-static condition, however considering the timeline and workload, non-static conditions can be discussed later if time allowed.***  ***Proposal 2: for CSI prediction, it is proposed to use intermediate KPI, e.g. SGCS, as requirements/tests metrics for LCM.*** |
| [**R4-2501499**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501499.zip) | Ericsson | [Proposal 1: RAN4 should discuss the detailed test parameters for AI/ML-based CSI prediction test (PMI reporting test with Rel-19 CSI prediction) when the WI performance part starts.](#_Toc189853742)  [Proposal 2: RAN4 should define more than one AI/ML-based CSI prediction requirements to ensure the generality of the CSI prediction model. At least RAN4 consider several channel models such as TDLA30 and TDLC300.](#_Toc189853743)  [Proposal 3: RAN4 should consider defining AI/ML-based CSI prediction requirements using CDL channel model when a suitable CDL model is available, depending on the outcome of SI spatial channel model.](#_Toc189853744)  [Proposal 4: RAN4 should wait for RAN1 conclusion on the performance monitoring for CSI prediction. Once RAN1 conclude the performance monitoring framework, RAN4 start to discuss the necessary (RRM) core requirements such as reporting delay and reporting accuracy requirements.](#_Toc189853745)  [Proposal 5: Testability of monitoring reporting delay/accuracy needs discussion.](#_Toc189853746) |
| [**R4-2501508**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501508.zip) | ZTE Corporation, Sanechips | ***Observation 1. TDL channel model doesn’t need to any gain in signal transmission.***  ***Observation 2. CDL channel model needs to consider antenna pattern gain, beamforming gain in signal transmission, which will results in the transmit power of the signal no longer being equal to 1 in link level simulation.***  ***Proposal 1. No need to consider CDL channel model for CSI prediction test.*** |
| [**R4-2501527**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501527.zip) | Huawei, HiSilicon | ***Proposal 1***: RAN4 to discuss whether introducing CDL channel model for AI CSI prediction testing, if CDL model is used to define requirement for non-AI PMI reporting.  ***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.  ***Observation 1***: Whether/How to ensure that the testing dataset aligns well with training dataset is still an open issue.  ***Proposal 4***: RAN4 to discuss training/inference consistency related testing configurations after after other WGs achieve sufficient progress. |
| [**R4-2502032**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2502032.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**  **Observation 1: Rel-19 SCM channel model study is expected to only focus on FR1 TDD DL SU-MIMO Uma scenario based on 38.827 CDL-C channel model. The testability and test repeatability for CDL channel model on demodulation and testing should be further studied.**  **Proposal 2: Prioritize to use TDL channel model for AI based CSI prediction tests in Rel-19. Further study the performance impact on AI CSI prediction with spatial channel model (e.g., CDL based channel model, or TDL based extended channel model) in Rel-20 pending on the progress of SCM channel model study in Rel-19 and specific channel model in Rel-20.**  **Proposal 3: RAN4 can consider the following baseline for AI-based CSI prediction testing**   * **For SNR test point for AI-based CSI prediction, use 90% of maximum throughput as baseline** * **For codebook configured for AI-based CSI prediction, use Rel-15 type I single panel codebook as baseline** * **For non-AI/ML based CSI prediction, use benchmark#1 of the nearest historical CSI for PMI generation as baseline for comparison as additional**   **Proposal 4: Further study the which non-static scenarios/configuration can be introduced for CSI prediction. Another alternative solution for considering non-static conditions is to define requirements with multiple scenarios under different condition to ensure the generalization aspect across different scenarios and conditions. RAN4 should study the generalization scenario for CSI prediction use cases.**  **Observation 2: Compared with RAN1 evaluation, it may be not suitable to use the Rel-18 TypeII-Doppler-r18 codebook test setup directly for AI bU**  **Proposal 5: RAN4 consider the following CSI prediction parameters as starting point,**   |  |  | | --- | --- | | **Parameters** | **Value** | | **UE speed** | **30km/h, 60km/h** | | **Input/output CSI type** | **Raw channel matrix** | | **Observation window** | **5/5ms** | | **Prediction window** | **1/5ms/5ms** | | **Performance metrics** | **Relative Throughput** | | **Codebook configuration for CSI prediction** | **Rel-15 Type I single panel codebook** |   **The test configuration for PMI test can be referred in TS38.101-4 section 6.3.2.1.3 and section 6.3.2.21.4**  **Proposal 6: Further discuss the test configuration and test framework after the sufficient progress in RAN1 for defining AI based CSI prediction requirement.**  **Proposal 7: RAN4 may need to discuss whether the reference model needed for alignment.**  **Observation 3: 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** |
| [**R4-2502070**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2502070.zip) | Nokia | 1. RAN4 should start the discussion on the requirements with N4=1. 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. |
| [**R4-2502166**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2502166.zip) | Qualcomm Incorporated | **Observation 1: 38.843 has emphasized to ensure consistency between network and UE side training. RAN1 has introduced the concept of associated ID to ensure consistency between training and inference regarding network side additional conditions.**  **Observation 2: Some companies proposed RAN4 to assume R18 eType II Doppler codebook for CSI report for AI-ML prediction.**  **Observation 3: RAN1 is still discussing whether R18 eTypeII Doppler codebook or R16 eTypeII codebook would be used for CSI report for AI-ML prediction. RAN1 allowed both Rel-18 Doppler codebook and Rel-16 eTypeII codebook to be used for evaluating CSI report of AI-ML based CSI prediction.**  **Proposal 1: RAN4 uses the same network side additional condition (scenario, configuration, antenna layout, etc.) that would be specified for tests to define the test requirements.**  **Proposal 2: RAN4 waits until RAN1’s decision to define the codebook for CSI report for AI-ML prediction tests.** |

## Open issues summary

In the previous meeting the following agreements related to CSI prediction were reached in R4-2420513:

**Issue 2-1: Channel model for CSI prediction tests**

**Agreement:**

* For channel model for CSI prediction tests, TDL channel models will be used.
  + FFS on whether CDL channel model is needed for generalization test.

**Issue 2-2: Reference throughput**

**Agreement:**

* reference throughput is obtained based on random PMI with Rel-15 Type I single panel codebook

**Issue 2-3: Testing conditions**

**Agreement:**

* use static conditions as baseline.
  + FFS on whether to have non-static condition

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

1. Test baseline
2. Reference model for alignment
3. Training data and test consistency
4. Generalization
5. Parameters for simulations/tests

### Sub-topic 1-1

*Test baseline*

Several companies proposed the use of Rel-18 eType II Doppler codebook to define requirements/tests.

**Issue 1-1: Test baseline**

* Proposals
  + Option 1: use Rel-18 eType II Doppler codebook for AI/ML based CSI predictions
  + Option 2: use Rel-16 eTYpe II codebook for AI/ML based CSI predictions
  + Option 3: wait for RAN1 to make enough progress
  + Option 4: use Rel-15 Type I codebook for AI/ML based CSI prediction
* Recommended WF

Option 3

Qualcomm: RAN1 is discussing it.

### Sub-topic 1-2

*Reference model needed for simulation alignment and requirement definition*

RAN4 will need a simulation campaign to define the requirements and the test cases, it should be discussed whether a baseline UE side AI/ML model to be used for prediction should be agreed or there is no such need

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

* Proposals
  + Option 1: RAN4 should agree a reference AI/ML model for simulations
    - Details of the reference model need to be further discussed, companies would be invited to bring proposals
  + Option 2: There is no need for a reference AI/ML model since everything is single sided
  + Option 3: Others
* Recommended WF
  + To be discussed

Qualcomm: we see the some usefulness of option 1. Based on CSI prediction, UE vendor will implement different models. Option 2 would be OK.

Vivo: Companies may have different results. Reference is useful to alignment.

OPPO: We can do some side condition for simulation to check the performance to see whether the results are aligned or not. If we could align the results, we may not need reference.

Ericsson: In current CSI requirement, we have no reference model.

Samsung: CSI prediction is defined. It is difficult to do alignment for eType II.

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.

### Sub-topic 1-3

*Training and test data consistency*

Some companies brought up the need to have good alignment between the training data and the testing data.

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

* Proposals
  + 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
  + Option 3: others
* Recommended WF
  + To be discussed

If Option 1 is agreed, it should be further discussed what exactly needs to be studied

Qualcomm: this is one-sided. We do not specify the training data set. If we define the requirements based on TDL, we have base training data on the same condition set. Data set will not be specified.

OPPO: For both option 1 and 2, the consistency between training data and test data should be ensured.

Moderator: Vendors need generate the training data aligned with the propagation conditions for the test.

Nokia: We also think that we need ensure them. How to ensure. We can keep option 1 and option 2.

Ericsson: our concern is that some UE may choose the model only based on such propagation condition. We should avoid such case. The training data should be general. We need define multiple test cases, like CDL channel model.

Samsung: We are OK to keep the options open. We can base on RAN1 study.

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

### Sub-topic 1-4

*Generalization*

Several companies proposed to study and define some generalization tests.

**Issue 1-4: Generalization**

* Proposals
  + Option 1: RAN4 should introduce some generalization tests, companies are invited to bring concrete proposals on how to define such tests
  + Option 2: Postpone this discussion until the requirement and test definition is clear
  + Option 3: There is no need for any generalization tests
* Recommended WF
  + To be discussed

Agreement:

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

### Sub-topic 1-5

*Simulation parameters*

Several companies proposed a set of parameters to be used for defining the requirements and tests.

**Issue 1-5: Simulation parameters**

* Proposals
  + Option 1: Use 32Tx ports, discuss all other parameters separately
  + Option 2: Adopt the parameters in Table 2-1 in R4-2500822(vivo)
  + Option 3: Use the following parameters from R4-2502032(Samsung) as a starting point

|  |  |
| --- | --- |
| **Parameters** | **Value** |
| **UE speed** | **30km/h, 60km/h** |
| **Input/output CSI type** | **Raw channel matrix** |
| **Observation window** | **5/5ms** |
| **Prediction window** | **1/5ms/5ms** |
| **Performance metrics** | **Relative Throughput** |
| **Codebook configuration for CSI prediction** | **Rel-15 Type I single panel codebook** |

* + Option 4: Postpone this discussion until more decisions are made on the test baseline
  + Option 5: others
* Recommended WF
  + To be discussed

If there is agreement to start working on a set of parameters, a list combining multiple options is likely needed to be worked on.

Tentative Agreement:

* Use the assumptions for Rel-18 CSI performance requirements as the starting point for the agreement of simulation assumption, with the exception of
  + Assuming 32Tx at BS.
  + Assuming periodic CSI reporting rather than aperiodic CSI reporting
  + Assuming eType-II codebook

# Topic #2: Requirements for Beam managements

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

## Companies’ contributions summary

|  |  |  |
| --- | --- | --- |
| **T-doc number** | **Company** | **Proposals / Observations** |
| [**R4-2500085**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500085.zip) | VIAVI Solutions | **Proposal 1:** Adopt Option 6 where both RSRP accuracy and beam prediction accuracy are used as KPIs for beam prediction.  **Proposal 2:** Adopt Option 2 where error range (or standard deviation) for the BB part is a function of SNR (or SINR). For RF part, a constant error range (or standard deviation) can be used.  **Proposal 3:** Adopt Option 2 where simplified CDL channel models are used. Put a limit for total number of clusters and constraints for angle/delay values for different clusters depending on the beamforming operation. |
| [**R4-2500225**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500225.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**  **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.** |
| [**R4-2500274**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500274.zip) | OPPO | **Proposal 1: Regarding the BM tests, the selection of metrics depends on the progress in RAN1, e.g.,**   * **if Beam information on predicted Top K beam(s) among a set of beams is the only reported content, then Beam prediction accuracy could be an test KPI accordingly** * **if both Beam information on predicted Top K beam(s) among a set of beams and RSRP of predicted Top K beam(s) among a set of beams are reported by the UE, then both Beam prediction accuracy and RSRP accuracy should be considered when defining RAN4 test KPIs**   **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 2: 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 3: Stability of performance monitoring and decision-making mechanism should be considered to mitigate the impact of random effects on monitoring outcomes.**  **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** |
| [**R4-2500513**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500513.zip) | CAICT | **Proposal 1: The principle to define the measurement period TL1-RSRP\_Measurement\_Period\_SSB** **for FR2 in current specification TS 38.133 could be reused. The detailed values of {M, N, P} needs further study.**  **Proposal 2: RAN4 to identify the impact of reduced resource configuration on existing RAN4 requirements with L1-RSRP related measurements. Table 1 could be considered as starting point.**  **Table 1**   |  |  | | --- | --- | | **RRM Req. Category** | **RRM procedure** | | Signaling characteristics | SSB /CSI-RS based Radio Link Monitoring | | Link recovery procedures (SSB/CSI-RS based beam failure indication, L1 indication, Candidate beam detection) | | Active TCI state switching delay | | Uplink spatial relation switch delay | | Pathloss reference signal switching delay | | Active downlink TCI switching delay for unified TCI | | Active uplink TCI switching delay for unified TCI | | Measurement procedure | NR intra-frequency measurements (Number of cells and number of SSB, measurement reporting requirements, etc) | | L1-RSRP measurements for reporting | | Measurement performance requirements – NR measurements | Intra-frequency RSRP accuracy requirements for FR1/FR2 | | RSRP measurement report mapping | | L1-RSRP accuracy requirements for FR1/FR2 |   **Proposal 3: Postpone discussion on reporting related requirements until other WGs define detailed reporting mechanism.**  **Proposal 4: RAN4 to discuss the delay components of LCM procedures, which may include such as,**   * **UE processing time of the signaling with model switching command** * **UE processing time of switching model** * **UE processing time of transmitting UL acknowledgement, if any** * **Signaling of RRC-reconfiguration for the target model, if any** * **Whether UE has the associated ID and/or other additional conditions for the target model.**   **Proposal 5: 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 6: Suggest to specify relative RSRP accuracy requirement with the definition in option 1.** |
| [**R4-2500593**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500593.zip) | NTT DOCOMO, INC. | **Observation 1: The requirements can be defined by two factors, the selection probability of appropriate beam and predicted RSRP value accuracy of selected beam.**  **Proposal 1: Prediction accuracy and absolute predicted L1-RSRP accuracy shall be used for beam prediction KPIs. The definitions are as follows.**   * **Prediction accuracy: the Top-1 predicted beam is one of the Top-2 strongest beams, and the Top-1 predicted beam’s ground truth RSRP value is larger than the RSRP of the strongest beam – x dB. x shall be less than 3.** * **Absolute predicted L1-RSRP accuracy: the acceptable difference between predicted RSRP value and ground truth of the selected beam. The existing absolute accuracy requirement can be the starting point.** |
| [**R4-2500612**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500612.zip) | Jio | 1. **Proposal 1**: Dynamically adjust measurement window sizes based on UE mobility, interference, and AI latency, improving resource efficiency while ensuring sufficient data for beam prediction. This enhances FR2 performance by optimizing measurement timing. 2. **Proposal 2**: Replace hierarchical reporting with QoS-aware beam priority tags (P0, P1, P2) mapped to network slices, prioritizing critical LOS beams and reducing overall reporting overhead. Focuses on slice level QoS. 3. **Proposal 3**: Implement phased model updates using small delta weights pushed from the LMF to gNB and UE, ensuring faster model adaptation and enabling a rollback mechanism in case of issues. Reduces downlink overhead. 4. **Proposal 4**: Shift from rigid thresholds to stochastic error margins, validating RSRP predictions based on confidence intervals derived from UE-assisted error distribution reports. Improves RSRP accuracy in inference for the AI engine. 5. **Proposal 5:** Dynamically partition the AI model between UE and gNB based on UE capabilities, enabling high-tier UEs to perform more inference locally, and balancing the UE level power requirements for model inferencing. |
| [**R4-2500618**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500618.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.**  **Proposal 1: For beam prediction accuracy performance metric,**   * **If beam index is used as reference, the test metric is:**   + **Top-K/N (%): the percentage of "the Top-N genie-aided beam is one of the Top-K predicted beams". FFS how N is derived.** * **If RSRP difference is used as reference, the test metric is:**   + **The maximum measured L1-RSRP of top-K predicted beams is larger than the maximum measured L1-RSRP – X dB.**   **Proposal 2: 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 3: For BM-case 2, RAN4 to discuss impact on measurement requirement, e.g. measurement time, prediction time, measurement restriction and scheduling restriction, etc.**  **Observation 3: 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 4: RAN4 may know RX beam based on UE capability.**  **Observation 4: For set B, L1-RSRP difference between best 1st beam and 8th beam is large.**  **Proposal 5: For simplicity, L1-RSRP delta for different TX beams in set B can be equal to SNR difference for beams in set B.**  **Observation 5: Under low SNR conditions, measurement error can be large.**  **Observation 6: There will be performance degradation due to measurement error.**  **Proposal 6: Impact of measurement error needs to be considered for prediction accuracy requirement.** |
| [**R4-2500823**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500823.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: For Top-K/N (%): the percentage of "the Top-N genie-aided beam is ONE of the Top-K predicted beams, the prediction becomes easy (most of the prediction accuracy is above 95%) even measurement error is considered. This may make the target too relaxed. Even if this KPI is met, it does not fully represent that the true quality of these beams is good**  **Observation 2: 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 3: 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 be more reasonable for ensuring the quality of the reported beams**  **Proposal 4: RAN4 to follow the process as above for the alignment of KPIs and definition performance requirements on BM prediction**   * **<Step 1> Discuss the definition of KPIs with the consideration of measurement error**   + **<Step 1-1> Select one or more candidate KPIs for simulation**   + **<Step 1-2> With the aligned simulation assumption, each company brings simulation results of candidate KPIs. Specifically, the measurement error modeling of simulation assumption may include**     - **Option 1: only BB measurement error is considered**     - **Option 2: only RF measurement error is considered**     - **Option 3: both BB and RF measurement error are considered**   + **<Step 1-3> Check on KPIs performance alignment from the simulation results from contributing companies**     - **repeat simulations for 2-3 meetings and check if good alignment can be achieved**       * **if good alignment of KPIs is achieved, move to step 3**       * **if not, move to step 2**   + **<Step 2> Discuss the definition of KPIs without the consideration of measurement error**     - **<Step 2-1> With the aligned simulation assumption, each company brings simulation results of candidate KPIs**     - **<Step 2-2> Discuss whether some margin be added to the aligned results of KPIs**     - **<Step 2-3> Move to next step after alignment**   + **<Step 3> RAN4 to make the final decision for KPIs definition and discuss the details of RRM requirements**   **Observation 4: There is a certain probability that the Top-K predicted beam is not any of the Top-K beams based on ground-truth RSRP.**  **Observation 5: 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 5: The absolute RSRP accuracy for AI/ML based beam prediction = 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 6: There is a non-negligible performance degradation in beam prediction accuracy when considering measurement error modeling for inference.**  **Observation 7: With the decrease of SNR, the beam prediction accuracy will also decrease. However, when the SNR reaches a certain level (e.g., 0dB), this performance decline will stabilize.**  **Observation 8: Compared with Truncated Gaussian distribution, the error derived from LLS-based modeling has a greater impact on the decline of beam prediction accuracy performance under similar SNR levels.**  **Proposal 6: RAN4 to define beam prediction accuracy requirements with consideration of the impact on measurement error.**  **Observation 9: In addition to the rate of correct beam prediction, when additional 1dB RSRP margin is considered (e.g., For the success rate of the difference between the ideal RSRP of ideal Top-1 beam and the predicted RSRP of predicted Top-K beam are within the X=1dB range), it will bring a noticeable improvement in the rate of prediction accuracy.**  **Observation 10: The probability of predicted RSRP of predicted Top-K beam falling into the range of [the RSRP of Top-1 beam± XdB] has significant improvement with the increase of X value.**  **Observation 11: When using predicted RSRP of predicted Top-1 beam and Ideal RSRP of ideal Top-1 beam as the benchmark, the probability that the difference between the RSRP of Top-1 beam and the predicted RSRP of Top-K beam is within the [X]dB range are similar.**  **Proposal 7: RAN4 to use Top-K/1 as the metric of beam prediction accuracy for further performance evaluation.**  **Observation 12: There is some impact on beam prediction and predicted RSRP value due to RF error.**  **Proposal 8: For the study on the impact of measurement error, the method of RF error modeling needs to be involved in RAN4 discussion, e.g.,**   * **An offset value in the original dataset;** * **Truncated Uniform distribution;** * **Truncated Gaussian distribution.** |
| [**R4-2500888**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500888.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 6**: 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 7**: Filtering out the data exceeds TE’s maximum Tx power has little impact on the prediction accuracy.  **Proposal 5**: 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 6**: When defining the accuracy requirements for AI/ML BM models, the impact of measurement error should be considered.    **Proposal 7**: When defining the accuracy requirements for AI/ML BM models, the impact of the dynamic range of TE should be considered.  **Proposal 8**: If some key parameter changes in HO will trigger BM model activation/switch, RAN4 should discuss whether and how to extend HO delay. |
| [**R4-2501036**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501036.zip) | CATT | **Proposal 1: For DL Tx prediction, UE Rx beam should be known to UE itself.**  **Proposal 2: Known conditions for TCI state for AI/ML BM can reuse the legacy conditions and modifications are adopted in red:**   |  | | --- | | **The TCI state is known if the following conditions are met:**  **- During the period from the last transmission of the RS resource used for the L1-RSRP measurement reporting for the target TCI state or AI/ML based inference to the completion of active TCI state switch, where the RS resource for L1-RSRP measurement is the RS in target TCI state or QCLed to the target TCI state, or the RS resources for AI/ML based inference is the RS in SetB or in target TCI state or QCLed to the target TCI state.**  **- TCI state switch command is received within 1280 ms upon the last transmission of the RS resource for beam reporting or measurement or inference**  **- The UE has sent at least 1 L1-RSRP report for the target TCI state or completes AI/ML based inference before the TCI state switch command**  **- The TCI state remains detectable during the TCI state switching period**  **- The SSB associated with the TCI state remain detectable during the TCI switching period**  **- SNR of the TCI state ≥ -3dB**  **Otherwise, the TCI state is unknown.** |   **Proposal 3: RAN4 defines unified inference delay requirements for both BM-Case1 and Case2 and discussion is based on BM-Case2.**  **Proposal 4: Inference delay requirements are defined by considering at least the follow parameters:**   * **Periodicity of measurement resources.** * **Periodicity of SMTC/MG.** * **Duration of DRX cycles.** * **F and K (For BM-Case 2).**   **Details are FFS and RAN4 can wait for more RAN1/2 inputs.**  **Proposal 5: For BM cases with NW-side AI/ML model, no AI/ML-specific RRM requirements are needed.**  **Proposal 6: For Type1 network-based performance monitoring, the time duration between the reception of the configuration/signalling for performance monitoring and UE reporting can be defined as performance monitoring delay.**  **Proposal 7: For Type2 UE-based performance monitoring, RAN4 to further discuss the necessity of performance monitoring delay requirements.** |
| [**R4-2501092**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501092.zip) | CMCC | ***Proposal 1: the relative RSRP accuracy is proposed as:***   * ***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 predicted value.***   ***Proposal 2: it is proposed that the RSRP absolute accuracy requirements apply to all predicted beams, not limit to the beam index in top-K beams.***  ***Proposal 3: For the beam in set B, it is known. For the beam not in set B, it is prefered to assume UE Rx beam is known, but we are also OK to have UE capability to move forward.***  ***Proposal 4: it is proposed that both the RSRP accuracy and beam prediction accuracy are the metrics/KPIs for beam prediction.***  ***Proposal 5: for measurement for prediction, legacy measurement delay can be reused, or at least no longer delay is expected.*** |
| [**R4-2501170**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501170.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.** |
| [**R4-2501284**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501284.zip) | Ericsson | **Observation 1: Option 1 of the relative RSRP accuracy is somewhat closer to a beam label (beam ID) prediction than an RSRP prediction.**  Observation 2: 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 3:** **Both ±4.5 dB RF error with Gaussian distribution and uniform distribution brings large performance degradation on prediction accuracy especially for Top-1 beam, where the impact of uniform distributed RF error is even worse.**  **Observation 4: The RAN2 defined ‘the reference time of the earliest time instance for the predicted results’ for BM-Case 2 of UE-side model, which is from the prediction report, isn’t relevant to RAN4 requirements on inference delay.**  **Observation 5: For beam (ID) prediction and L1-RSRP prediction, the inference delay may be different.**  **Observation 6: 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 7: If the mapping between beam number and probe number, like 8 beams <->100 probes, is valid, it’s difficult to do multiple beam (up to 64) emulation by CDL channel.**  **Observation 8: Multi-AoA TDL, may be regarded as a particular oversimplified CDL channel, but only one cluster for each beam.**  **Observation 9: The maximum number of Tx beams in multi-AoA TDL is unclear.**  **Observation 10: Different ratios between the number of Set A and Set B demonstrates various AI/ML model characteristics.**  **Observation 11: Option 2, i.e., 8-probe OTA system, could oversimplify the test channel. For supporting such approach, we lead towards that more probes can be implemented in TE.**  **Observation 12: the measurement error in datasets for training/testing may be covered by the overall measurement error.**  **Proposal 1: For the absolute RSRP, the index i shall be any beam index in top-K beams based on predicted L1-RSRP as per UE reporting.**  **Proposal 2: For the relative RSRP, support Option 2. 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 3: RAN4 shall study the new known condition of the TCI state switch, to which the RS referred is which is ‘known’ by the prediction without any measurement.**  **Proposal 4: Knowing the Rx beams is a crucial factor in TCI state switch enhancement in terms of RS prediction, it may be supported by the UE as UE capability.**  **Proposal 5: The ‘predicted known’ conditions of the prediction-based TCI state switch shall take below conditions into account:**   * **The TCI state switch command is received during valid inference phase of AI/ML mode.** * **The predicted beam in Set A is the RS in target TCI state.** * **The predicted beam in Set A is detectable during training and inference phase, which implies the beam has been trained successfully and available in inference phase.** * **The UE knows the Rx beam for the predicted beam.**   **Proposal 6: The ‘predicted known’ conditions doesn’t apply to the RS which has been measured in prediction procedure, including measurement phase and prediction phase, either in Set A or Set B.**  **Proposal 7: 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 8: In Option2, RAN4 only considers the below option.**   * **Top-K/1 (%) : the percentage of "the Top-1 strongest beam is one of the Top-K predicted beams"**   **Proposal 9: The distribution of baseband error can be fitted with zero mean Gaussian distributions.**  **Proposal 10: 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 11: 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 12: Regarding measurement period for prediction (how long will set B be measured before making a prediction),**   * **For BM-case1, one shot L1-RSRP measurement result on Set B is enough for the UE to predict Set A.** * **For BM-case 2, the UE may measure Set B for multiple L1-RSRP measurement periods/occasions before prediction. The exact number of the periods/occasions may be configurable which is to be studied and defined in RAN1. RAN4 is only aware of measurement accuracy for each L1-RSRP measurement.**   **Proposal 13: Measurement delay is defined as the delay of single/one shot measurement occasion, which is identical to measurement delay in legacy, it is same to BM-case 1 and BM-case 2.**  **Proposal 14: RAN4 to study if any requirement is needed for evaluating the capability of a UE tracking changes of L1-RSRP by measurements in BM-case 2.**  **Proposal 15: Don’t take reporting time into the delay definition, but we can define measurement reporting requirements similar to the legacy ones, if needed.**  **Proposal 16: Inference delay is defined as the processing delay from the completion of the last (or first) measurement occasion (for BM-case 1, only one-shot measurement occasion) until the start of the first predicted reference signal occasion.**  **Proposal 17: RAN4 shall study measurement/inference delay for monitoring.**  **Proposal 18: We may set a number of probes (e.g., in the order of hundreds or dozens, if it is available) with a proper designed distribution, and TX beams are only injected to those probes, e.g., with granularized AOA/AOD, the CDL channel may be simplified (depending on the probe number) and but still can provide enough spatial characters for AI/ML model conformance test.**   |  | | --- | |  | |  |  |   **Proposal 19: For evaluating CDL channel or simplified CDL channel, RAN4 to study how many probes can be implemented in TE in a realistic way.**  **Proposal 20: For single AOA/probe, adding CDL channel and Rx gain function in TE before combined and fed into the single probe may be a compromise approach, but the effect of testing prediction performance may not be expected.**  **Proposal 21: We don’t oppose the segregate test, but before proceeding with the directive, we must respond to at least the following questions:**   1. **What’s the target of the multi-AoA test for UE Rx Beam sweeping?** 2. **What’s the test metric of the multi-AoA test for UE Rx Beam sweeping?** 3. **How to map the test metric/result of the multi-AoA test for UE Rx Beam sweeping and the test metric/result of single-AoA test for AI/ML model?**   **Proposal 22: CDL channel may be simplified by granulating the clusters in channel with respect to the probe distribution in TE.**  **Proposal 23: The simplified CDL channel only is considered to be applied in inference, no need to be applied in training.**  **Proposal 24: 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 25: For BM-case1, at least two sets of numbers of Set A and Set B are provided as test configurations. Below are two examples:**   * **For verifying the prediction model computational power/complexity of AI/ML model, the number of set A = X and the number of set B = X/4.** * **For verifying the prediction scaling capability of AI/ML model, the number of set A = X and the number of set B = X/16**   **Note 1: Different accuracy requirements may be applied for different sets of number of set A and set B.**  **Note 2: X may be the maximum number of Set A that a UE supports mandatorily.**  **Proposal 26: RAN4 shall study the mapping between the Tx beam number and the probe number, if the ratio 1:2 is feasible?**  **Proposal 27: RAN4 to evaluate the SNR condition for the ground truth through simulation.**  **Proposal 28: 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 29: How to avoid cheating in the reports on RSRP measurement and prediction is a valid issue to be studied. A simple way for the test environment is adding a procedure to manually switch the beam indexes between the prediction duration and the measurement duration.**  **Proposal 30: Same channel is applied for measurement and inference in BM-case1.**  **Proposal 31: 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 32: 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.** |
| [**R4-2501528**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501528.zip) | Huawei, HiSilicon | ***Proposal 1:*** RAN4 to define the requirement of prediction delay in AI BM, where the measurement delay and inference delay are not treated as separate requirements.  ***Proposal 2:*** For data collection in NW-sided model, use the existing core requirement for beam related information reporting as the starting point.  ***Proposal 3:*** If measurement reporting is used for the calculation of performance metric at NW during model monitoring, reuse RAN4 legacy requirement for measurement reporting.  ***Proposal 4:*** If metric reporting is introduced by RAN1 for model monitoring, RAN4 will study the related requirement. |
| [**R4-2501692**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501692.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: Accurate L1-RSRP and Top-K/1 metrics ensure reliable identification of the strongest beams.**  **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: Option 6 should be adopted in conjunction with consistency in beam prediction to ensure both accuracy and reliability in the predictions.** |
|  |  |  |
| [**R4-2501957**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501957.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   In the paper, the following Observations and Proposals were made:  **Observation 1:** Due to the nature of analog circuit design of RF, the RF uncertainty will be presenting all the time, even there are calibration tables, and algorithm, such as AGC and feedback-loop, to compensate Tx/Rx uncertainty of RF circuits, the RF error will still be presenting in the field.  **Observation 2:** RF error will be presenting as independent error for every time RF circuit is used for Tx/Rx. The error will however follow statistics distribution.  **Proposal 1: RAN4 should consider RF error model with independent error on each Tx beam.**  **Observation 3:** 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 accuracy.  **Observation 4:** 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 5:** 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.  **Observation 6:** The performance degradation of AI/ML model is high when RF error is added independently. Contrary, performance degradation is minimal when RF error is same for all TX beams. Hence, the performance degradation difference between two cases is significant.  **Proposal 2: 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.**  **Proposal 3: Given the significant degradation in prediction performance observed with independent RF error, RAN4 shall discuss:**   * **Whether such degradation of prediction performance metrics (Top-1 decreases by 13% and 90%-tile L1-RSRP difference decreases by 7 dB) with respect to the case without errors is acceptable or not.**   **Proposal 4: RAN4 to discuss the RF error model since a large difference of the prediction accuracy was observed with different RF error models. Discuss whether the RF error model that assumes the same RF error for all TX beams is feasible or not.**  **Proposal 5: RAN4 should consider, based on the study on the measurement error impacts on RSRP prediction, if worse than legacy performance for RSRP prediction can be acceptable if there is significant reduction in the measurements.**  **Proposal 6: RAN4 to consider**   **to accurately evaluate the successful rate of correct prediction.**  Where  **Observation 7:** Top-1(%) metric of Option2 is the strictest requirement and RAN4 should not consider it as a performance metric.  **Observation 8:** Top-K/1(%) metric of Option2 only verifies if actual strongest beam is among Top-K predicted beams and it doesn’t verify the accuracy of prediction for other beams.  **Proposal 7: Top-K/1(%) metric of Option 2 should be considered in the combination of Top-1/K(%).**  **Observation 9:** Top-1/K(%) metric of Option2 only verifies if Top-1 predicted beam is among Top-K actual strongest beams and it doesn’t verify the accuracy of other predictions.  **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 8: 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 9: 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 10: RAN4 should consider relaxing known conditions for TCI states (Clause 8.10.2 of TS 38.133) for BM-Case1 and BM-Case2. The TCI state switch command is received within a delay X\*1280 ms. FFS: Further analysis on defining X values.**  **Proposal 11: RAN4 should consider the impact on legacy measurement requirements for BM-Case2. It needs further analysis if legacy measurement requirements can be reused for BM-Case1.**  **Proposal 12: RAN4 should consider the impact in RAN4 requirements for the Evaluation period**  **due to the presence of predicted beams.**  **Proposal 13: 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 14: RAN4 should consider Table below as a baseline to start discussions for LCM operations related RRM core requirements.**  Table. Delay components related LCM operations for different CSI-report configurations   |  |  |  |  | | --- | --- | --- | --- | | 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  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 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  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  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  Delay related to associated ID  Delay due to CSI-RS measurement period | Delay from UE processing related to capability report  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: RAN4 should consider the impact in RAN4 requirements for performance monitoring window when UE reports performance metric or event to the NW.**  **Proposal 16: As starting point, RAN4 should analyse the delay requirements for applicable functionality reporting based on 5-step procedures agreed in RAN2.** |
| [**R4-2501964**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501964.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*  **Proposal 1:** No RAN4 impact is expected from data collection for UE-sided model training.  **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:** RAN4 shall adopt the absolute RSRP accuracy for the predicted RSRP in the report of inference results as the performance metrics for UE-sided AI-BM requirements/tests.  **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, the relative RSRP accuracy requirement is not applicable.  **Proposal 4:** If beam prediction accuracy is adopted as performance metrics, the sub option “Top-K/1 (%): the percentage of "the Top-1 strongest beam is one of the Top-K predicted beams" shall be precluded.  **Proposal 5:** For metrics for beam management requirements/tests, in addition to absolute RSRP accuracy, RAN4 shall at least adopt the same metric as RAN1 for UE-assisted performance monitoring:   * i.e., at least “Alt 1: Top 1 or Top K beam prediction accuracy (with or without margin) by comparing the prediction results and the Top 1 or Top K beam based on the measurements from a resource set/ resources for monitoring”.   + RAN4 can wait further agreement from RAN1 on “margin”, to refine RAN4 option 2 and/or option 3, or combined one.   **Observation 4:** For UE-sided model, DL Tx beam prediction are considered with the measurements from (1) the best Rx beam or (2) Tx-Rx beam pair, as the input to AI/ML network.  **Proposal 6:** RAN4 core requirement shall be general enough to cover both alternatives of UE-side model inputs:   * Alt-1: L1-measurements from the best RX beam taken as model input * Alt-2: L1-measurements from Tx-Rx beam pair taken as model input   **Observation 5:** From RAN1 perspective, the following issue has not been concluded yet:   * Whether the RS resource configured in Set A (but not in Set B) can be used as the QCL source RS for a TCI state?   **Proposal 7:** Before discussing whether the TCI state associated with a predicted Tx beam is known or unknown, RAN4 shall wait until RAN1 made conclusion or send LS to RAN1 on the following issue:   * Whether the RS resource configured in Set A (but not in Set B) can be used as the QCL source RS for a TCI state?   **Proposal 8:** After RAN1 has concluded the issue in Proposal 7, RAN4 can discuss how to define a special TCI state associated with a predicted Tx beam (within Set A but not in Set B):   * The TCI state associated with a predicted Tx beam (within Set A but not in Set B) is assumed to be a “predicted TCI state” (different from known or unknown condition based on measurement/reporting). * For TCI state switching to a “predicted TCI state”, UE is allowed to have additional measurement occasions on top of switching to a known TCI state.   **Proposal 9:** For testing UE-sided model inference and monitoring (if applicable), the consistency of the NW-side additional condition across training and inference must be ensured by applying the same associated ID as a side condition.  **Proposal 10:** 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 11:** For AI-BM Case-2, the exact number of the measurement periods/occasions shall be configurable and subject to further study in RAN1.  **Proposal 12:** At least for performance monitoring Type 1, Option 2 (UE-assisted performance monitoring), RAN4 prediction accuracy requirement shall be applicable.  **Proposal 13:** 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. |
| [**R4-2502165**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2502165.zip) | Qualcomm Incorporated | **Observation 1: RAN1’s agreed beam prediction metric without or with margin simply translates to option 2 and option 3 of RAN4’s beam prediction metric.**  **Observation 2: Support for option 1 of beam prediction metric is still TBD in RAN1.**  **Observation 3: 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 will not be able to pass option 2 focused test.** * **Note: The maximum value of K depends 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 4: 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 5: 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 6: 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 7: The evaluation results for 8 beam based setB and 64 beam based setA without any measurement error are shown in Table 2.**  Table 2 Evaluation results for BM-Case1 (Use case 1) without model generalization for DL Tx beam prediction at UE side (Results for Set B, Set B subset of Set A)   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Qualcomm | | | |  | | Assumptions | | Number of beams [beams/beam pairs] in Set A | 192 | 192 | | Number of beams [beams/beam pairs] in Set B | 24 | 24 | | Baseline scheme | **Empirical PMF-based approach** | **linear interpolation-based approach** | | AI/ML model input/output | | Model input | RSRPs of Set B beams | RSRPs of Set B beams | | Model output | Best Tx beam ID | Best Tx beam ID | | Data size | | Training | 4000 | 4000 | | Testing | 600 | 600 | | AI/ML model | | Model description | FC layer based NN | FC layer based NN | | Model complexity | 4.6K parameters | 4.6K parameters | | Computational complexity | 9.8K FLOPs | 9.8K FLOPs | | Evaluation results [with AI/ML/baseline  (Baseline-2)] | Beam prediction accuracy (%) | Top-1 (%) | 63.5 / 28.3 | 63.5 / 10.7 | | Top-2/1 (%) | 80.0 / 46.1 | 80.0 / 16.3 | | Top-5/1 (%) | 92.5 / 79.2 | 92.5 / 31.8 | | 1-dB marginal accuracy (%) | 90.4 / 59.0 | 90.4 / 32.8 | | L1-RSRP diff (otion 3 with K = 1) | Avg. L1-RSRP difference in dB | 0.36 / 1.27 | 0.36 / 4.22 | | System performance | RS overhead Reduction (%) | 87.5 | 87.5 |   **Observation 8: 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-CD: Gaussian with mean = -0.12 dB and sigma = 0.85 dB**   **Observation 9: The variance of the Gaussian distribution fitting the baseband relative error is approximately twice the variance of the Gaussian distribution fitting the baseband absolute error. This can only be possible if baseband absolute measurement error is independent across SSB samples.**  **Observation 10: RAN4 spec allows +-6.5 dB relative L1-RSRP accuracy range in FR2. Internal simulations suggest that 90 percentile baseband relative inaccuracy ranges within +-1.5 dB.**  **Observation 11: Even if we assume 1-2 dB additional margin, baseband relative inaccuracy is less than the +- 6.5 dB L1-RSRP relative inaccuracy margin provided in RAN4 RRM spec. The remaining portion of relative inaccuracy range comes from RF error.**  **Observation 12: Unlike baseband error, it is hard to derive an exact distribution of RF error. Hence, it is suitable to model RF error as a uniform random variable.**  **Proposal 1: RAN4 defines AI-ML BM requirement for option 3.**  **Proposal 2: RAN4 waits for simulation results and RAN1 progress before defining AI-ML BM requirement for option 2 and option 1 respectively.**  **Proposal 3: UE knows the RX beam paired with the predicted Tx beam from setA.**   * **FFS: TCI state of predicted TX beam.**   **Proposal 4: RAN4 discusses both of following approaches to investigate the impact of measurement error on AI-ML BM metric.**   * + **Case 1: Training data, testing data and ground truth for training and testing all include measurement error.**   + **Case 2: Training and testing data include measurement error. Ground truth for training and testing data does not include measurement error.**   **Proposal 5: RAN4 models baseband absolute measurement error as an IID Gaussian random variable across SSBs and time samples with zero mean and SD standard deviation.**   * **Note: SD may vary among companies.**   **Proposal 6: RAN4 models RF error in the following way:**   * **Absolute RF error of strongest SSB lies uniformly in [-4.5, +4.5] dB range** * **Assume that, in one instance, absolute RF error of strongest SSB is Y dB. Absolute RF error of all other SSBs (except strongest SSB) in that instances lie uniformly in [max(Y-X,-4.5 dB), min(Y+X, 4.5 dB)] range**    + **Note: X is TBD.** * **RF error does not change in time in the same band if TX power, path loss and angle of arrival don’t change** |

## UOpen issues summary

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

1. Beam prediction KPIs
2. Relative RSRP accuracy
3. Measurement period for inference
4. QCL source RS for TCI states
5. Requirements based on relative difference
6. Simulation results
7. Measurement error modelling

### Sub-topic 2-1

*Beam prediction KPIs*

Different metrics/KPIs have been discussed and were captured in the TR:

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

- Option 1: RSRP accuracy

- Option 2: Beam prediction accuracy

-Top-1 (%) : the percentage of "the Top-1 strongest beam is Top-1 predicted beam"

-Top-K/1 (%) : the percentage of "the Top-1 strongest beam is one of the Top-K predicted beams"

-Top-1/K (%) : the percentage of "the Top-1 predicted beam is one of the Top-K strongest beams"

- Option 3: The successful rate for the correct prediction which is considered as maximum RSRP among top-K predicted beams is larger than the RSRP of the strongest beam – x dB,

-Related measurement accuracy can be considered to determine x

- Option 4: combinations of above options

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

Proposals

* + Option 1:
    - 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.
  + Option 2: Option 2 with a combination of Top-K/1 and Top 1/K
    - Option 2a: update option 2 to Top-N/K, the percentage of "the Top-N genie-aided beam is one of the Top-K predicted beams".
  + Option 3: Option 3 as it does not have the issue of UE failing if there are multiple beams within a small RSRP margin
  + Option 4: For beam prediction accuracy: A beam ranking metric should be used that considers the cumulative beam ranking other. Potential options are:
    - **(1)   Kendall/Spearman correlation**
    - **(2)   Normalized Discounted Cumulative Gain (nDCG)**
    - **(3)   Rank-biased overlap**
  + Option 5:combinations of options, concrete proposals should be disucssed
* Recommended WF
  + To be discussed

The number of options should be downselected to make some progress or a clear path to choose an option should be discussed/agreed

### Sub-topic 2-2

*Relative RSRP accuracy*

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

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

* Proposals
  + Option 1: **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 2: ***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 3: Relative RSRP is the RSRP difference between the predicted RSRP values, FFS what the accuracy should be
  + Option 4: Others
* Recommended WF
  + To be discussed

### Sub-topic 2-3

*Measurement period for inference*

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

**Issue 2-3: 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: Differentiate case 1 and case 2
    - Reuse L1 measurement period for Case 1
    - Introduce configurable number of measurement occasions for case 2
  + Option 3: Study the measurement period based on simulations
  + Option 4: Others
* Recommended WF
  + Option 1

### Sub-topic 2-4

*QCL source RS for TCI states*

In order to further discuss the conditions for a predicted TCI state to be known/unknown, QCL conditions should be clarified

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

* Proposals
  + Option 1: send LS to RAN1 to ask whether the RS resource configured in Set A (but not in Set B) can be used as the QCL source RS for a TCI state
  + Option 2: wait for RAN1 discussion without sending any LS
  + Option 3: this information is not needed
* Recommended WF
  + Option 1

### Sub-topic 2-5

*Requirements based on relative difference*

In R4-2501957 it is proposed to define a set of requirements based on some relative difference between side conditions to ensure that the UE measurements are consistent.

**Issue 2-5: Feasibility Intermediate KPIs for CSI requirements or LCM**

* Proposals
  + Option 1: RAN4 to study how 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.
  + Option 2: RAN4 should not spend time on such a study
  + Option 3: Companies are invited to bring a more concrete proposal on how such requirement would work before RAN4 would start a study
  + Option 4: others
* Recommended WF
  + Option 3

### Sub-topic 2-6

*Simulation results*

**Issue 2-6: Simulation results**

* Proposals
  + Option 1: Discuss the simulation results based on summary – to be prepared by vivo
* Recommended WF
  + To be discussed

### Sub-topic 2-7

*Measurement error modeling*

Several companies brought proposals on how to model the RF and baseband measurement error, the model should be further discussed

**Issue 2-7: Measurement error modeling**

* Proposals
  + Option 1: Measurement error to be discussed offline together with other simulation assumptions
* Recommended WF
  + To be discussed

# Topic #3: Requirements for Positioning

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

## Companies’ contributions summary

|  |  |  |
| --- | --- | --- |
| **T-doc number** | **Company** | **Proposals / Observations** |
| [**R4-2500229**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500229.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** |
| [**R4-2500272**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500272.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.** |
| [**R4-2500514**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500514.zip) | CAICT | ***Observation 1:*** *Emulation-based tests for case 1 face issues, such as emulating sophisticated propagation conditions at different UE locations under multi-TRP scenario, discrepancy between emulated environment and realistic deployment, and etc.*  ***Observation 2:*** *Site-based tests for case 1 face issues, such as selection of scenarios, scale of site, training data collection, expenditure, and etc.*  **Proposal 1: Suggest to firstly discuss possible approaches (e.g., emulation-based, miniature testing site, etc.) and feasibility of each in RAN4 before further decision on specifying requirement for case1.** |
| [**R4-2500824**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500824.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.** |
| [**R4-2501037**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501037.zip) | CATT | **Proposal 1: For positioning measurements, legacy measurement requirements can be referred.**  **Proposal 2: RAN4 to discuss the possible values of inference delay.**  **Proposal 3: RAN4 to use the legacy reporting requirements as a starting point.**  **Proposal 4: RAN4 not to define core requirements for case 3a/3b.** |
| [**R4-2501095**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501095.zip) | CMCC | ***Proposal 1: for case 1, it is proposed to define requirements. At least positioning accuracy can be defined, which is the difference between ground truth (real position) and reported position.***  ***Proposal 2: for case 3a/3b, in addition to UL-RTOA and UL SRS-RSRPP, it is proposed to reuse legacy report mapping for UL SRS-RSRP.***  ***Observation 1: 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/3b, it is proposed to reuse legacy accuracy requirements for gNB Rx-Tx time difference.*** |
| [**R4-2501171**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501171.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.** |
| [**R4-2501529**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501529.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. |
| [**R4-2501758**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501758.zip) | Ericsson | **Observation 1**: Achievable positioning accuracy is scenario specific, and the same level of positioning accuracy cannot be observed in different scenarios.  **Proposal 1**: RAN4 to conclude that defining requirement in terms of the accuracy of estimated UE position for case 1 is not feasible.  **Proposal 2**: RAN4 to define requirements for model input for case 1. Details are FFS. |
| [**R4-2502071**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2502071.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.   **Case 2a/2b**   1. RAN4 can discuss whether/how to define the requirements on LoS/NLoS. 2. **RAN4 to discuss whether to reuse the legacy measurement period and accuracy requirements and to define any additional measurement requirement in Case 2b.**   **Case 3a/3b**   1. RAN4 to extend the applicability of existing report mappings also for UL SRS-RSRP to AI/ML based positioning in Cases 3a/3b. 2. RAN4 to establish a common understanding of the support for the various positioning techniques such as multi-RTT and discuss the RRM requirements of positioning techniques supported by other WGs. |
| [**R4-2502168**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2502168.zip) | Qualcomm Incorporated | **Observation 1: RAN2 chair has guided companies to focus on 1st priority positioning use cases. RAN plenary has suggested RAN1 not to discuss any proposal targeting 2nd priority issues in Q4 of 2024.**  **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 postpones defining requirements for 2a and 2b until RAN1 and RAN2 make further progress regarding the 2nd priority work items.**  **Proposal 2: RAN4 defines requirements for case 1, i.e., UE-based positioning with UE-side model.** |

## Open issues summary

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

Requirements for case 1

LOS/NLOS indicator requirements

Timing information reporting from UE/gNB to LMF

Report mapping for UL SRS-RSRP for Case 3a/3b

### 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)
  + 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
  + Option 3: others
  + Option 4:
* Recommended WF
  + To be discussed

CMCC: Support option2. According to RAN2 agreement, the existing LPP procedure is agreed. Ground truth vs prediction for accuracy.

ZTE: For option 2, CMCC only mention accuracy requirement. How about the other requirements.

Ericsson: We need study the feasibility. We do not think it is possible to define the requirements for case 1. Maybe we are OK to defining the requirements for measurement accuracy.

CATT: We also analyze. We prefer not to specify. Measurement accuracy requirement with controllable workload.

Nokia: We prefer option 2 as CMCC.

Qualcomm: Prefer Option2. That is prioritized topic.

NTT DOCOMO: Prefer to Option 2. RAN2 has already agreed on.

Huawei: Support option 1.

OPPO: Support Option 1. There is feasibility issue. Similar as Ericsson.

Vivo: Support option 2.

Xiaomi: Support Option 1.

CMCC: To ZTE, at least accuracy requirement needs be defined. About the feasibility, we can further check. We can consider the specific scenario for the requirements.

### Sub-topic 3-2

*LOS/NLOS indicator*

LOS/NLOS reporting was agreed in RAN1, it should be discussed whether to introduce requirements or not

**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

CMCC: Support Option 1. We can consider the LOS/NLOS hard indicator. RAN1 agreed to introduce LOS and NLOS.

Ericsson: We agree with it is more related to case 2a/2b. But they are out-of scope now.

Qualcomm: Support Option 2. We have even do not decide on legacy cases.

Huawei: In RAN1 the LOS and NLOS indicators are introduced. But there is no agreement how to differentiate them from non-AI.

OPPO: Option 2.

CATT: Option 2. This NLOS indicator is not new thing. It is not necessary to define the requirement.

Nokia: Option 2.

CMCC: RAN2 has new agreement for case 2a/2b. Need more time to check.

### Sub-topic 3-3

*Timing information reporting*

**Issue 3-3: Timing information reporting**

* Proposals
  + Option 1: Reuse the legacy requirement for reporting of timing information or timing and power information from UE/gNB to LMF
    - For Cases 2a/2b and 3a/3b
  + Option 2: Further study if the requirements can be reused
  + Option 3: No need for any requirements
  + Option 4: Others
* Recommended WF
  + Option 1

Apple: What is the legacy requirement.

CMCC: two meetings ago, we had agreement to use legacy requirement for case 2a/2b. Legacy requirements are SRS RSRP accuracy requirement.

Apple: how does the timing information link to SRS RSRP?

CMCC: RSTD, SRS RSRP… as legacy requirements.

Qualcomm: RAN1 is discussing sample based measurement in addition to legacy measurement for case 2b. At this moment, we do not need agreements.

Ericsson: The tentative agreement is not clear to us.

FFS:

* + For Case 3a, reuse the legacy accuracy requirement for gNB Rx-Tx timing difference

### Sub-topic 3-4

*Report mapping for UL SRS-RSRP*

**Issue 3-4: Report mapping for UL SRS-RSRP**

Proposals

* + Option 1: Reuse the report mapping for UL SRS-RSRP to Case 3a/3b
  + Option 2: Introduce other report mapping
  + Option 3: Mapping is not needed
  + Option 4: Other
* Recommended WF
  + Option 1

Ericsson: Option 1 is fine for Case 3a.

Agreement:

* Reuse the report mapping for UL SRS-RSRP to Case 3a

# Topic #4: Testing for one sided models

This section contains the sub-topics regarding testing of one sided models, especially beam prediction

## Companies’ contributions summary

|  |  |  |
| --- | --- | --- |
| **T-doc number** | **Company** | **Proposals / Observations** |
| [**R4-2500167**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500167.zip) | Anritsu Corporation | ***Observation 1: It is necessary to clarify characteristics of the UE Rx beam knowledge and their necessity for AI model input since they have a huge impact on the 1AoA test procedure and setup with wireless (virtual) cable connection.***  ***Observation 2: AI model inputs varied per company in the evaluation results of RAN1 [2] and there are companies which used only L1-RSRPs of Set B as the model input.***  ***Observation 3: From the summary in RAN1 [3], it appears that RAN1 has not concluded types of AI model input.***  ***Proposal 1: To move forward the discussion on the AI/ML beam management test setup, RAN4 first clarify the types of AI model inputs for the prediction of beams.***  **Option 1: Mandatory AI model inputs for the beam prediction are L1-RSRP of set B and implicit Tx beam ID**  **Option 2: Mandatory AI model inputs for the beam prediction are L1-RSRP of set B and Tx beam ID**  **Option 3: Mandatory AI model inputs for the beam prediction are L1-RSRP of set B, Tx beam ID and Rx beam information (FFS for the necessary information of Rx beam)**  **Option 4: Others**  ***Observation 4: No study has been done yet regarding if there are any prediction performance differences between different types of model inputs.***  ***Proposal 2: RAN4 to study if there are any prediction performance differences between different types of model inputs.***  ***Observation 5: It is necessary to ensure that tests with simplified CDL are done and show appropriate prediction performance of the AI model in the field.***  ***Proposal 3: The decision to apply the simplified CDL should be deferred until the study has been conducted and until the prediction performance has been confirmed within RAN4.***  ***Observation 6: There is a trade-off between multi-AoA and 1AoA test setups.***  ***Observation 7: Further studies are necessary regarding applicable channel models and types of model inputs to conclude setups to be used for the AI/ML beam management tests.***  ***Observation 8: 1AoA test setup + Fading simulator with wireless (virtual) cable connection is already providing solutions to the issues listed in the WF and in other contributions.***  ***Observation 9: 1AoA test setup + Fading simulator with wireless (virtual) cable connection has no limitation to simulate channel model, which solves the concern of the degradation of evaluation or training possibly caused by the simplified CDL.***  ***Observation 10: By connecting between the DUT and test equipment by wireless (virtual) cable connection, it is possible to emulate the environment of multiple AoAs even with the 1AoA setup.***  ***Observation 11: By measuring the actual UE Rx beam patterns by the test system and using them in the fading simulator, as far as the UE beam patterns are aligned with the calculated channel model in the fading simulator, UE can use its own Rx beam information for the AI model inputs.***  ***Observation 12: Regarding the prediction of the associated Tx beam or beam pair, the UE uses only a measurement result with one UE Rx beam in its AI/ML algorithm in order to predict the Tx beam or beam pair once. FFS on the necessary number of Rx beam patterns (or UE positions) to be measured.***  ***Observation 13: For setup B, use of UE Rx beam pattern in the fading simulator can be separately considered as follows.***  AI model input case 1: UE Rx beam information is AI model input  A measurement of actual UE Rx beam pattern is necessary.  AI model input case 2: UE Rx beam information is not AI model input  Either of the following two options can be applied.   1. Use a standardized UE Rx beam pattern in the fading simulator. 2. Use an actually measured UE Rx beam pattern in the fading simulator, though it is not mandatory. |
| [**R4-2500226**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500226.zip) | Apple | **Observation 1: 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 2: 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 3: 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 4: 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 5: 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 7:**  **Advantages (multiple AoA): The channel model captures simultaneous reception across multiple AoAs/AoDs, providing a more accurate representation of CDL channels and real-world field data.**  **Disadvantages (multiple AoA): It may lead to a more complex testing setup. There could be challenges in selecting the appropriate simplified CDL channel.**  **Advantages (single AoA): The testing setup is simplified, and existing OTA set up can be reused. For beam prediction only (not RSRP prediction) it may be possible to avoid transmission of set A beams since TE could know the best beam before hand.**  **Disadvantages (single AoA): It may not accurately represent real field conditions. The reliability of using a simplified channel model for testing data may be questionable when evaluating models trained on real-world data.**  **Observation 8:** **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 9:** **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 10: 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 11: Achieving consistency between training and inference by model monitoring could result in delays and increased complexity in model management for BM use case**  **Observation 12: 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 13: 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.** ***Beam Prediction Testability Discussion (Testing set-up)*** **Proposal 1: Multipath 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 2: To establish a criterion for assessing the feasibility of a simplified sparse probe layout (Option 5 simplified CDL) 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. Below figure: CDL\_ref is the reference channel to be approximated by the simplified CDL\_TE channel during test.**  **A diagram of a training  Description automatically generated**  **Proposal 3: 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 Figure below. (multiple static configurations) (Option 5)**  **A diagram of a test  Description automatically generated with medium confidence**  **A diagram of a diagram of a diagram  Description automatically generated with medium confidence**  **Proposal 4: For testing the beam prediction/RSRP accuracy for BM-Case 2 temporal prediction we propose the framework described in this section and visualized in Figure below, through a non static configuration of the channel emulator. (dynamic channel)**  A diagram of a computer program  Description automatically generated with medium confidence  **Proposal 5: In AI/ML beam management testing, the spatial-selective and time-varying propagation conditions can be emulated by adjusting the transmit (Tx) power at the test equipment (TE) for each beam at different time intervals**  **Proposal 6: For BM spatial prediction RAN4 to consider the following options multiple AoA and single AoA:**  **Multiple AoA based Testing Setup with a simplified CDL channel and probe layout. Criterion of evaluating the feasibility of a particular simplified CDL model has been proposed.**  **Single AoA based Testing Setup**  A diagram of a circle with lines and dots  Description automatically generated with medium confidence  ***Generalization issues for BM***  **Proposal 7: 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 8: 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 9: 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 10:** **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 11:** **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**.  ***Consistency between Training and Inference***  **Proposal 12: 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 13: RAN4 to consider option 1 for achieving consistency between training and inference for only static conditions.**  **Proposal 14:  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 15: 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 16: 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 17: 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 18: 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.**  **A diagram of a model  Description automatically generated**  **Proposal 19: 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 20: 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**  A diagram of a machine learning  Description automatically generated  **Proposal 21: 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.** |
| [**R4-2500273**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500273.zip) | OPPO | **Proposal 1: 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 the number of AoAs that RAN4 needs to represent a beam** * **FFS how to map the X AoAs and corresponding X pathloss to different beams in BM set A/B**     **Proposal 2:**   * **Maximum number of set B Tx beams that test system should be able to emulate: 8** * **Maximum number of set A Tx beams that test system should be able to emulate: 64**   **Proposal 3: 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 4: The simplified CDL should be used as the channel assumption.**  **Proposal 5: The TCI state associated with a predicted Tx beam should be known to DUT.** |
| [**R4-2500403**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500403.zip) | Qualcomm Incorporated | **Observation 1: An effective test case for beam prediction use case BM-Case1 requires the following:**   * **Sufficient randomness and variation in time and spatial domain of L1-RSRP** * **DL Tx beam sweeping with enough number of Tx beams in Set B and Set A**   **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: Some companies suggested to emulate multiple beams with single AoA by using different power which is similar to ‘virtual cable’ approach used for FR2 demodulation test.The power levels will have to be defined in test configuration to enable this method. That will allow UE to predict the strongest beam through simple interpolation.**  **Observation 4: CDL with spatial properties is a more realistic and proper channel model for AI/ML beam management testing. To support CDL, multiple probes test setup is needed.**  **Observation 5: With TDL channel, to generate Set A and Set B, multiple AoAs/probes test setup is necessary.**  **Observation 6: The key requirements for AI/ML BM test setup are listed following:**   * **Requirement#1: Channel model in the test**   + **CDL or simplified CDL or TDL with multiple AoAs to properly verify beam management performance.** * **Requirement#2: Avoid the cheating**   + **Progress A: Data 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 during measurement step of progress of Figure 4.** * **Requirement#3: Support the beam number of Set A**   + **TE needs to support Set A for ground truth verification, i.e., upto 128.**   **Observation 7: 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 8: 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 9: 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.**  **Observation10: 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 11: 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.**  **Observation 12. TE vendors need to check UE positioning time to mimic simplified CDL or TDLchannel model by sequentially rotating the positioner to reduce the number of probes. Order of ~1s dwell time will lead to an unrealistic test environment for AI/ML beam management testing.**  **Proposal 1: RAN4 do NOT consider single AoA/probe test setup (with TDL or Los condition) for AI/ML beam management testing.**  **Proposal 2: RAN4 take multiple AoAs/probes test setup as the basline for AI/ML BM testing and downselet the channel model from Simplified CDL and TDL considering the following aspects:**   * **Number of Tx Beams: Upto 128 for Set A** * **Randomization:**    + **Channel parameters for data training and generation should have enough randomness**   + **During measurement step of progress B, TE randomizes the transmission of setA beams, e.g., during time t1, TE does not transmit with beam shape A1 and TE transmits using the shape of a beam whose ID is randomly selected from set A).**   **Proposal 3: RAN4 considers 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 simplify the test setup based on the evaluation of the impact on beam management performance considering the UE beam width.**  **Proposal 4: RAN4 to consider the follwoing two multiple AoAs/probes options as the test seup for AI/ML beam management testing**   * **Option 1: Simplified CDL with mutliple AoAs/probes**   + **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 simplify the test setup**   + **3D-MPAC test setup can be conisderd as the baseline** * **Option 2: TDL channel model with mutliple AoAs/probes**   + **Study how to emulate compareble Tx beam with simplifed CDL, e.g., generating enough Tx beam number for Set A, The solutions such as combinaiton of probes transmission and power offset for each of probe can be considered.**   + **Legacy FR2 RRM test setup can be considered as the baseline** |
| [**R4-2500617**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500617.zip) | Xiaomi | **Proposal 1: For OTA test, prefer to split the test motivation into two parts, RAN4 to at least test part 1 and further discuss whether to test part 2.**   * **Part 1: Verify AI model prediction algorithm** * **Part 2: Verify input data accuracy of model AI based on RX beam sweeping**   **Observation 1: If UE RX beam pattern can be known at TX side, it’s possible to use single AOA to emulate input data for AI model, the emulated data will be signal after UE RX beamforming. No simplified channel model is needed.**  **Observation 2: If a fixed RX beamforming gain is used for multi-path TX signals, there will be some mismatch between emulated signal and real signals.**  **Observation 3: 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 2: For channel modelling type in AI algorithm performance test, RAN4 to discuss whether UE RX beam pattern can be known in TX side first.**  **Proposal 3: If the test metric is beam index prediction accuracy, UE may need to report ground truth of N best beam index for set A. RAN4 to discuss how to decide number N.**  **Observation 4: 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 4: 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 5: For BM case-1, channel doppler can set to 0 or a small value to guarantee that there is neglectable L1-RSRP variation.**  **Proposal 6: 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 5: 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 7: RAN4 to discuss how to solve the UE cheating issue if UE report both predicted result and ground truth.** |
| [**R4-2500825**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500825.zip) | vivo | **Observation 1: The simplified CDL channel with small number clusters will degrade the test dependability.**  **Observation 2: 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 3: Different test system with reduced cost and complexity can be introduced by using the Tx beam grouping method.**  **Proposal 1: RAN4 to consider methods to group Tx beams together to reduce the number of probes in the test system.**  **Proposal 2: 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 3: RAN4 to consider the following two approaches for design the test system**    **approach a**    **approach b**  **Proposal 4: Companies to provide candidate probe locations for CDL-A/B/C for merging and rotating. E.g.,**   |  |  |  | | --- | --- | --- | | (a) CDL-A | (b) CDL-B | (c) CDL-C |   **Proposal 5: RAN4 to consider the UE Rx beamforming resolution in angular domain to perform OTA probe allocations.**  **Observation 4: PRUs can be deployed in the test scenario to provide measurements together with the location information.**  **Proposal 6: 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 5: At least LoS/NLoS indicator hard value can be tested since the LoS/NLoS environment is known in test system.**  **Proposal 7: 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.** |
| [**R4-2500889**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2500889.zip) | MediaTek Inc. | **Proposal 1**: 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 2**: 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 3:** 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 4**: 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 1**: 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 2**: With single AoA, TE can know the expected RSRP of Tx beams at UE baseband.  **Proposal 5**: Use existing IFF or enhance IFF test systems with single AoA to test R19 DL Tx beam prediction models.  **Proposal 6**: Standardized UE radiation pattern in 38.901 can be assumed in R19 AI/ML BM test with single AoA test system.  **Proposal 7**: If to use multiple AoAs test system for AI/ML BM test, UE rotation should be considered.  **Observation 3**: The time needed to rotate UE in FR2 MIMO OTA test systems, i.e., about 1s, is not acceptable for AI/ML.  **Proposal 8**: The impact on AI/ML BM models by using simplified CDL channels should be studied at first before concluding to use multiple AoAs test system.  **Proposal 9**: 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 4**: The upper bound of SNR with rough beam are much lower than that with fine beam.  **Proposal 10**: Fine beam should be used in AI/ML test.  **Observation 5**: Among all the test methods, only using peak fine beam can make sure the lowest SNR above -6dB with 30dB dynamic power range.  **Observation 6**: Multiple AoA test systems can not use peak fine beam only.  **Proposal 11**: Using peak fine beam in AI/ML BM test with single AoA test system.  **Observation 7**: SNR dynamic range of multiple AoA test system would be at least 12dB smaller than single AoA test system. |
| [**R4-2501038**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501038.zip) | CATT | **CSI prediction:**  **Proposal 1: CDL channel mode and non-static conditions are used only if they are proved necessary.**  **Beam management:**  **Proposal 2: Absolute RSRP accuracy for AI/ML based beam prediction = predicted L1-RSRP of beam index i – ground truth of L1-RSRP of beam index i. ~~The index i may be any beam index [in top-K beams based on ground truth L1-RSRP].~~**  **Proposal 3: The definition of relative RSRP accuracy shall be: the RSRP difference between the predicted L1-RSRP of different beams (Option 3).**  **Positioning:**  **Proposal 4: RAN4 not to define performance requirements for case 1.**  **Proposal 5: RAN4 does not discuss testability aspects related to positioning case 3a/3b.** |
| [**R4-2501097**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501097.zip) | CMCC | ***Proposal 1: 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 2: to verify beam management performance, it is proposed to use multiple AoAs test systems.***  ***Proposal 3: for beam prediction testing, it is prefered to consider CDL, but we are also fine with simplified CDL.***  ***Proposal 4: If the absolute RSRP is agreed to be metric for the beam management prediction, it is proposed that tests for BM at least include two parts: the part to get ground truth, and the part for prediction verification.*** |
| [**R4-2501531**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501531.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:*** Only the measured RSRP rather than the ideal RSRP can be obtained by UE both for model training and model inference.  ***Proposal 2:*** Take the measurement error into consideration for generating the labels for model training.  ***Proposal 3:*** For the impact of measurement error on BM performance evaluation, RAN4 to study how to use the evaluation results with system-level simulation assumptions for requirement definition study under link-level channel models.  ***Proposal 3:*** 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 4:*** 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. |
| [**R4-2501959**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501959.zip) | Nokia | **Proposal 1: Clarify the questions for the test setup with UE rotations: (1) How many Set A and/or Set B beams can be generated inside the test chamber at the same time? (2) How many test instance/iteration/intervals should be defined based on total number of Set A/B beams?**  **Proposal 2: UE vendors to clarify if the Set A and/or Set B beams emulated inside the test chamber can’t arrive at the DUT from different AoAs, will it be possible for the DUT to distinguish between the different beams which may have different Beam IDs but will arrive at the DUT from the same AoA.**  **Observation 1:** Reported RSRP values as approximated ground truth are error-prone and have several uncertainties.  **Proposal 3: To obtain a more accurate ground truth, RAN4 should consider the case where UL measurements received at TE/NW can be used as ground truth for DL measurements in case when Beam correspondence is ensured between UL and DL beams.**  **Observation 2:** A combination of AWGN/TDL based test system and UE rotation can be a potential candidate for the conformance testing of AI/ML based BM use case.  **Proposal 4: 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 5: 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 1.**  **Proposal 6: 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.**  **Observation 3:** When radio conditions change, UE will switch from one AI/ML model to another, and proper operation in non-static conditions needs to be ensured throughout testing.  **Proposal 7: RAN4 should define requirements/test procedures to ensure that no performance degradation occurs due to internal model switching under non-static conditions.**  **Observation 4:** When UE may have a well-generalized AIML model, no switching needs to be done for UE, and testing is needed to ensure that no performance degradation occurs.  **Observation 5:** CDL models can be used for generalization testing to ensure that practical scenarios can be simulated in a testing equipment.  **Proposal 8: RAN4 to define testing procedures/environment to ensure no performance degradation when running AIML model with TDL channel models for generalization testing.**  **Proposal 9: RAN4 needs to design a new metric, indicative of generalization capabilities of AI/ML CSI prediction functionality, to verify the generalization performance of the functionality in different scenarios**  **Observation 6:** Reporting of location estimates is supported in the current specifications.  **Observation 7:** Reuse of the reporting facility for Case 1 is possible, given that a certain location coordinate type from the existing ones is selected.  **Observation 8:** 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.  **Proposal 10: RAN4 to use existing location coordinate reporting to identify positioning accuracy requirements for Case 1.**  **Observation 9:** 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).  **Proposal 11: RAN4 should specify (i) the test mechanisms for positioning accuracy and (ii) the metric for positioning accuracy performance in Case 1.**  **Proposal 12: 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.**  **Proposal 13: 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.**  **Proposal 14: 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.**  **Proposal 15: LOS/NLOS indicator should be considered as an intermediate performance metric for case 2a (UE-assisted/LMF-based positioning with UE-side model, AI/ML-assisted positioning).**  **Observation 10:** LOS/NLOS as an intermediate metric/KPI can be considered for assisted AI/ML positioning. LOS/NLOS metric/KPI validation requires labelled data.  **Proposal 16: LOS/ NLOS metric/ KPI labelled data to be generated for Case 2a based on the ground truth extraction using the legacy methods with the help of a well calibrated device capable of precisely indicating the ratio of received LOS/NLOS signals.**  **Proposal 17: RAN4 to wait on further progress in RAN1 to address testability aspects for Case 2b related to existing path-based and, if agreed, sample-based measurements.**  **Proposal 18: RAN4 to wait on further progress in RAN1 to address testability aspects for the LOS/LOS indicator together with other measurement(s) for Case 3a.**  **Proposal 19: RAN4 to consider testability aspects for Case 3b both for path-based and sample-based measurements, for latter awaiting further details from RAN1.** |
| [**R4-2501965**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501965.zip) | Samsung | *Test procedure for AI-BM (UE-side model)*  **Proposal 1:** The conformance testing procedure for AI-BM UE-side model (using AI-BM Case1 as an example), is provided as:  - Preparation Phase:  - Step-1: TE configures resource sets for both Set A and Set B beams in *CSI-ReportConfig*, and associated ID as side condition  - Model Inference Phase:  - Step-2: TE configures the test environment to transmit Set B beams  - Step-3: DUT (a.k.a, UE) predicts Set A beams based on the measurement from Set B beams, and reports the prediction results back to TE  - Ideal Measurement Phase:  - Step-4: TE adjusts the test environment to transmit Set A beams, ensuring a certain level of SNR for the ideal measurement of RSRP  - Step-5: DUT measures the Set A beams, and report measurement results to TE  - TE Evaluation Phase:  - Step-6: TE evaluates the success of the prediction by comparing the predicted results from Step 3 with the actual measurement results from Step 5.  **Proposal 2:** The procedure to build up the dataset used for UE sided model training is out of the test procedure for AI-BM.  *Consistence between training and inference for AI-BM*  **Proposal 3:** The consistence between training and inference for AI-BM shall be guaranteed by using static condition as baseline, in which additional condition shall also be remains to be static.  *Test setup for UE-sided AI-BM testing*  **Observation 1:** The necessity of introducing spatial differentiated AoAs in AI-BM testing could depends on different AI/ML model inputs:   * For AI/ML model with TX-RX beam pair measurement as model input, the spatial differentiated AoAs is required. * For AI/ML models by TX beam measurement from the best RX beam as model input, to test this AI/ML models, the spatial AoA characteristics could be not that important.   **Proposal 4:** Taking into account the conclusion from beam or beam pair as model input, RAN4 can determine the test setup with spatial differentiated AoAs required or not.  **Proposal 5:** At least for Rel-19 AI-BM, RAN4 discussion shall focus on single TRP case, which is considered as baseline assumption for testability study.  **Proposal 6:** AI-BM OTA test setup shall generate the not only Set B of beams, but also Set A of beams for the ideal measurement of RSRP on the predicted Tx beam.  **Proposal 7:** The feasibility of enhanced 3D MPAC chamber shall consider the following factors:   1. How many beams out of Set A should be emulated?    * At least X strongest beams out of the whole beambook (e.g., 128 beams containing Set B beams) are relevant for AI-BM inference testing. 2. How many beams out of Set B should be emulated?    * At least Y strongest beams out of the whole beambook (e.g., 16 beams) are relevant for AI-BM inference testing. 3. All Tx beams from a single TRP; 4. Acceptable PAS similarity percentage based on a certain configuration of probes   **Proposal 8:** RAN4 perform system-level simulation to determine the value of X and Y (in Proposal 7):   * To determine the value of X, by allowing a reasonable AI/ML prediction performance degradation (e.g., nearly 0dB absolute error degradation for RSRP accuracy) by only measuring X strongest beams from the whole Set A beambook in the ideal Measurement Phase. * To determine the value of Y, by allowing a reasonable AI/ML prediction performance degradation (e.g., 1dB absolute error degradation for RSRP accuracy) by only measuring Y strongest beams from the whole Set B beambook. * Reasonable dynamic range of the beam signal strength level shall also be assumed. |
| [**R4-2502174**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2502174.zip) | Keysight Technologies UK Ltd | *Observation 1: A relatively simple and straightforward extension of an FR2 MIMO OTA system to 8 probes is suitable to emulate 4 dual-polarized/8 single-polarized Tx-beam RX-PAS profiles, set B Tx beams up to [16] with power offsets, and set A Tx beams up to [192] with power offsets.*  *Observation 2: The extension of an FR2 MIMO OTA system to 8 probes is suitable to support unlimited range of AoDs/ZoDs and a range of AoAs/ZoAs of 110°/25°.*  *Observation 3: Various beam management performance testing can readily be supported with a commercially available test system, i.e., complexity and time-to-market concerns should not be an issue.*  *Observation 4: The RTS methodology, applied to AI/ML Beam Management use cases, would require either standardized, measured, or vendor provided antenna patterns.*  *Observation 5: The measurement of all possible antenna patterns would be extremely time consuming*  *Observation 6: The standardized or vendor provided antenna patterns could lead to incorrect assessments/verdicts of conformance test cases*  *Observation 7: The RTS methodology, applied to AI/ML Beam Management use cases, would require mapping of antenna patterns to any given test condition/coordinate and radio/channel condition which is likely vendor proprietary.*  **Proposal 1: UE/chipset/infra vendors to provide feedback on whether modelling 4 dual-polarized/8 single-polarized RX-PAS accurately with the remaining beams modelled with power offsets is sufficient**  **Proposal 2: Consider an 8-probe OTA system with 2 additional probes when compared to the FR2 MIMO OTA system with a range of AoAs/ZoAs of 110°/25° as the starting point for AI/ML Beam management testing for 1 TRP and/or 2 TRPs with close angular proximity.**  **Proposal 3: Do not consider the single-probe OTA system approach any further for AI/ML Beam Management use cases** |
| [**R4-2501726**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_114/Docs/R4-2501726.zip) | Rohde & Schwarz | Observation 1: There are several factors to be considered in the end-to-end AI/ML BM performance.  Observation 2: The main challenges to conclude to a baseline setup are the identification of factors impacting the AI/ML BM performance, as well as the test system complexity and feasibility.  Observation 3: There seems to be two different assumptions and reference implementation for the AI/ML L1-RSRP model, with and without directional information for the L1-RSRP measurement.  Observation 4: Main drivers for a multi-AoA setup are spatial emulation of the CDL channel and the usage of directional information for the measured L1-RSRP in the prediction model.  Observation 5: (Single-AoA) The multi-AoA environment can be transformed into a single‑AoA test for AI/ML BM testing, by including the UE Rx beam/antenna pattern in the combination of BS Tx beams with the CDL channel model to be emulated, without impacting the number of BS Tx beams or the parameters of the CDL channel model.  Observation 6: (Single-AoA) There are 4 options for a multi-AoA to single-AoA test transformation and inclusion of the UE directional performance: *Option 1* use a reference fixed UE pattern for the baseband test, and a second separate test for the UE directional performance; *Option 2* use a reference complex UE pattern a the baseband test, and a second separate test for the UE directional performance; *Option 3* use a complex UE pattern provided by the UE manufacturer, no additional test required; *Option 4* use complex UE pattern measured by the TE, no additional test required;  Observation 7: (Single-AoA) L1‑RSRP accuracy is currently measured only on a single particular AoA, which in most test cases is the found UE Rx beam peak.  Observation 8: (Single-AoA) A new requirement and test case addressing the L1-RSRP spherical coverage accuracy, will be required to complement the AI/ML BM test based on single direction baseband testing, with reference UE Rx beam/antenna pattern.  Observation 9: (Single-AoA) The analysis of the single-AoA approach must consider the computational load to obtain all the resulting channel models when combining each BS Tx beam, the CDL channel model and each UE Rx beam.  Observation 10: (Single-AoA) In case of Option 2 and 3 (Reference or Manufacturer Complex UE Rx Beam/Antenna pattern), the UE Rx beam/antenna pattern must include magnitude and phase, as well as beam id information (i.e. codebook).  Observation 11: (Single-AoA) In case of Option 4 (TE measures the Complex UE Rx Beam/Antenna pattern), the UE Rx beam/antenna pattern measurement is very time consuming and must include magnitude and phase information, e.g. using an Antenna Test Function.  Observation 12: (Single-AoA) The definition of the format and granularity of the Complex UE Rx beam pattern is key to ensure consistent results among different TE vendors for the single-AoA methodology.  Observation 13: (Single-AoA) The use of fine or rough beams for AI/ML BM prediction may affect the complexity of the data to be considered for the single-AoA methodology, as well as the number of iterations to be calculated.  Observation 14: (Single-AoA) If several UE orientations towards the channel model must be assessed for the single-AoA methodology, the number of combinations and the computational effort rises exponentially.  Observation 15: (Single-AoA) The measurement timing perspective is also a critical aspect for the single-AoA methodology, because not all BS Tx beams might be measured with all UE Rx beams.  Observation 16: (Multi-AoA) Current approach to conclude to a multi-AoA baseline system for AI/ML BM testing, which is CDL channel centric, focuses on the spatial CDL emulation as per MIMO-OTA / 3D MPAC methodology, i.e. it considers a given number of probes per channel cluster and only a limited number of Tx beams.  Observation 17: (Multi-AoA) The CDL centric approach will probably not lead to a baseline system for AI/ML BM, since the system required to test the AI/ML BL requirement is not feasible, and a feasible system does not have capability to test the intended AI/ML BM requirement.  Observation 18: (Multi-AoA) The CDL centric approach is also questionable, since it is not clear what is the quantified impact of such a spatial CDL emulation on the actual UE AI/ML BM performance, which is the real scope of the test.  Observation 19: (Multi-AoA) A new AI/ML BM performance centric approach, intending to study the suitability of already feasible multi-AoA systems for the end-to-end UE AI/ML BM performance, is more realistic and leads to more solid progress towards testability assessment for multi-AoA setups.  Observation 20: (Multi-AoA) Current implementations of eIFF legacy systems, support with feasible upgrades: 4 Active 2x2 MIMO probes / AoAs with distinctive angular separation up to150°, overall big quiet zone of at least 30cm (black-box testing), lowest MUs (best L1-RSRP measurement accuracy) and ensure preservation of side conditions (RF-RRM, cross-feature).  Observation 21: (Multi-AoA) These system, have also practical advantages, since they are already matured and broadly used, support conformance testing for all features up to Rel-19, and if suitable for AI/ML BM testing, they avoid the need of an additional system, ensuring continuity in the cost sensitive FR2 testing ecosystem.    Proposal 1: (Single-AoA) UE AI/ML BM performance can be tested in single-AoA test setup, by emulating the combination of BS Tx beams, CDL channel and UE Rx beams in the baseband, based on a Reference UE Rx beam/antenna pattern, which is either fixed or complex (codebook). The UE Rx spatial/directional performance can be covered complementarily in a separate test which verifies the UE L1-RSRP spherical coverage accuracy.  Proposal 2: (Single-AoA) Study the definition of a L1-RSRP spherical coverage accuracy requirement as well as the respective test methodology, which can be used to verify the UE spatial/directional performance, complementarily to single-AoA baseband testing based on reference UE Rx beam/antenna pattern.  Proposal 3: (Multi-AoA) Change the approach for the multi-AoA baseline system analysis, from CDL channel centric to UE AI/ML BM performance centric.  Proposal 4: (Multi-AoA) Following the new UE AI/ML BM performance centric approach, analyze the AI/ML BM testing suitability for already existing multi-AoA setups (including potential feasible extensions), by assessing through simulations the UE AI/ML BM performance impact with different Set B / Set A combinations and channel model simplifications as per the concrete test setup implementation.  Proposal 5: (Multi-AoA) Define the implementation of eIFF test setup with 4x active 2x2 MIMO probes / AoAs as a candidate for the AI/ML BM baseline system and analyze its suitability (as per Proposal 4). |

## Open issues summary

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

Baseline test setup

Channel model

Set B and Set A beams

Other requirements for the test setup

Positioning test feasibility for case 1

### Sub-topic 4-1

*Baseline test setup*

The test setup has been discussed for multiple meetings without much progress. Some decisions should be made to limit the number of candidate setups.

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

* Proposals
  + Option 1: Single AoA setup (different Tx beams emulated only with power difference?)
    - Is it possible to test if UE can also make a prediction when set B beams come from different directions?
    - How many Tx beams can be emulated when the spatial domain properties are lost?
    - Will UE need just a simple interpolation model or is a more complex model necessary to pass such a test?
    - Can additional (L1-)RSRP measurement accuracy requirement (e.g. spherical coverage) be used to complement Single AoA testing with directional/spatial UE performance?
  + Option 2: Single AoA setup with “two stage approach”
    - Will a UE which uses Rx beam information in the prediction model have to also implement a model which does not use the Rx beam information?
    - Are there any implementation constraints on a UE to execute/pass this test?
    - How will the TE know the UE Rx beam pattern?
  + Option 3: multi AoA setup– RRM test setup as baseline (including enhancement for up to 4AoA active probes)
    - How many Tx beams can be emulated with such a setup, what is the relationship between number of probes and number of set B Tx beams?
    - Can AoAs between the probes also be emulated (is it possible to “make it look like” signals are coming from other angles than where the probes are placed?)
    - Can a simplified CDL be emulated or only TDL from each probe?
  + Option 4: multi AoA setup – 3D MPAC system as baseline
    - How many Tx beams can be emulated with the baseline setup?
    - 4Tx beams with dual pol(set B?) can be emulated with 2 additional probes

Pros&cons of different test setups

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Single AoA | Single AoA with two stage | Multi AoA – RRM test setup as baseline | Multi AoA 3D MPAC |
| Tests that UE uses spatial information(UE actually makes measurement in multiple directions to make a prediction) | No | Maybe? | Yes | Yes |
| Relevance for field performance | Low | Maybe? | Good/High | Good/High |
| Simplified CDL emulation | No | Yes |  | Yes |
| Test setup complexity | Low | Low | Medium | Medium/High? |
| Test method complexity | Low | High | Medium | Medium/High? |
| Reuse of test setup for other tests | Medium (multi AoA testing already exists in RRM) |  | High? | Low/medium?  Can only reuse the test setup for FR2 MIMO OTA? |
|  |  |  |  |  |

Companies are invited to propose further criteria which are relevant to BM prediction testing.

* Recommended WF
  + Do not further discuss single AoA test setups, continue to study the two multi AoA candidate setups

One option to continue the discussion proposed by R＆S is the 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.

R&S: continue discussions with single AoA and multi AoA.

Keysight: We do not believe much use to look at single AoA.

Mediatek: we have different view as Keysight. We have no common understanding on the table. If we check RSRP received at UE, it can follow CDL channel. We realize multiple AoA is not possible. Single AoA system is preferable.

Qualcomm: We support considering multiple AoA setup. For single AoA, UE has to lock the beam and UE cannot do beam sweeping. For single AoA with measurement pattern, it is quite time consuming. We can prioritize multi-AoA. For single AoA, how to address beam locking issue.

Nokia: Both single AoA and multiple AoA are feasible. Multiple AoA is a good starting point. I would like to make simple test cases and then discuss the procedure how to do the test.

Samsung: Back to table. For some field there are some problems.

Moderator: the table is for discussions. People can correct the entry.

Apple: it is pre-mature to preclude single AoA. Many companies are talking about Rx beam sweeping. It is not true to say the performance is not tested without beam sweeping. The main purpose is to test UE performance to based on set B input to find output in Set A. Table is a good start. Some of entries may represent some companies’ view. There are other aspects need be considered: cost. We cannot draw conclusion at this stage.

Anritsu: Support carrying out comparison the simplified CDL and CDL. We should not reduce the possibility of single AoA and multiple AoA at this moment.

R&S: In Rel-15 the system is designed with two probes. We use those two probes for some Rel-15 feature testing. Now we can further enhance to enable 4 probes. That system is available in Rel-15 features.

Keysight: To availability of test equipment, they are available. Different location in the chamber.

Nokia: Can we make simple test case? We define both single AoA set method and multiple AoA test method.

Qualcomm: We can study the test procedure for both single AoA and multiple AoA. With two parallel procedure. One procedure to test first.

Ericsson: Do not preclude single AoA. There would be spatial filter before AoA. Spatial filter based on artificial or real beam sweeping are different. We should consider the real sweeping. It is really time-consuming.

Nokia:

Mediatek:

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

### Sub-topic 4-2

*Channel model*

The channel model to be used has also been discussed for several meetings but not much progress has been made

**Issue 4-2: Channel model**

* Proposals
  + Option 1: simplified CDL
  + Option 2: TDL
  + Option 3: others
* Recommended WF
  + To be discussed

If simplified CDL is to be further considered, more detailed proposals on what to study/analyze further should be discussed (e.g. RX PAS, number of relevant clusters to emulate, dependency on number of clusters and number of probes in the chamber, etc)

Agreement:

* Use CDL-based channel model as starting point

### Sub-topic 4-3

*Set B and Set A beams*

The number of beams which needs to be emulated will be a requirement on the test setup

**Issue 4-3: Set B and Set A beams**

* Proposals
  + Option 1: set B: 16 beams, set A 128 beams. How many beams are need simultaneously?
  + Option 2: RAN4 perform system-level simulation to determine the value of X and Y (in Proposal 7):
    - To determine the value of X, by allowing a reasonable AI/ML prediction performance degradation (e.g., nearly 0dB absolute error degradation for RSRP accuracy) by only measuring X strongest beams from the whole Set A beambook in the ideal Measurement Phase.
    - To determine the value of Y, by allowing a reasonable AI/ML prediction performance degradation (e.g., 1dB absolute error degradation for RSRP accuracy) by only measuring Y strongest beams from the whole Set B beambook.
    - Reasonable dynamic range of the beam signal strength level shall also be assumed.
  + Option 3: Others
* Recommended WF
  + To be discussed

### Sub-topic 4-4

*Test setup requirements*

The test setup has been discussed for multiple meetings without much progress. Some decisions should be made to limit the number of candidate setups

**Issue 4-4: Option 4 for 2-sided model**

* Proposals
  + Option 1: UE rotation during the test
    - multiple TRP (or reflections from another direction)
    - AoA – to be discussed later
    - AoD – to be discussed later
  + Option 2:
    - Other paramters
* Recommended WF
  + To be discussed

### Sub-topic 4-5

*Positioning test feasibility for case 1*

There is a debate on whether requirements for positioning case 1 should be defined or not, the testability of case 1 is also part of this discussion

**Issue 4-5: Positioning case 1 testing**

* Proposals
  + Option 1: Case 1 cannot be tested because UE does not report the position
  + Option 2: Case 1 cannot be tested because the ground truth(actual UE position) cannot be derived
  + Option 3: Testing is feasible by using PRUs
  + Option 4: others
* Recommended WF
  + To be discussed